Development, Pesticide Exposure and Repeatability of an Aquatic, Tri-trophic Laboratory Microcosm

Verena Maria Riedl

Doctor of Philosophy

University of York

Environment and Geography

August 2018

Abstract

The environmental risk of pesticides is routinely assessed in single-species tests. Multi-species systems are only employed once effect concentrations (divided by safety factors) compared to predicted environmental concentrations give reason for concern. However, direct chemical effects at the individual level often don't directly translate into impacts observed at higher levels of organization because species interactions play an important role in mediating indirect chemical effects. In this thesis, I aimed to develop a tool that combines the advantages of single-species tests (repeatability, interpretability) and multi-species tests (ecological realism) for the repeatable study of indirect chemical effects mediated by ecological interactions. I show the standardization and testing of a tri-trophic laboratory scale microcosm (*Pseudokirchneriella subcapitata*, *Ceriodaphnia dubia*, *Hydra viridissima*) to better understand and quantify the effects of multiple stressors (e.g. chemicals, food availability and predation) on organisms and their interactions. We found close repeatability of system dynamics in the short term indicating the system's ability to detect small pesticide effects and bottom-up and top-down effect propagation. Yet, inter-experimental differences between dynamics in controls were found in the long term. An investigation of the influences of a variety of experimental factors showed that deviations from standardized population dynamics were likely caused by medium related factors that acted on algal populations and led to bottom up effects. These likely masked the effects of a herbicide in exposure experiments and I did not gain conclusive results on direct and possibly indirect herbicide effects on grazer and predator populations. My findings demonstrate that considerable consistency and in-depth understanding of the characteristics of all system components are required to achieve repeatability even in apparently simple multi-species systems. My work illustrates possible pitfalls of tools aimed at the generation of repeatable effect data on ecologically relevant endpoints and identifies future research needs to achieve repeatable dynamics in the tri-trophic microcosm and to enhance its applicability.

List of Contents

Abstract..I

List of Contents...III

List of Tables..VII

List of Figures (short titles)...IX

Acknowledgements XIII

Author's declaration XIV

Chapter One: **General Introduction 1**

1.1 Environmental risk assessment of PPPs 3

1.1.1 Lower tier tests and why they might be insufficient to estimate risk ... 5

1.1.2 Higher tier, or multi-species tests 7

1.1.3 Microcosms for chemical risk assessment in the aquatic environment - State of the science 10

1.2 Aims and objectives 12

1.3 Model species 13

1.3.1 Pseudokirchneriella subcapitata (Korshikov) F. Hindák 1990............. 13

1.3.2 Ceriodaphnia dubia Richard, 1894 14

1.3.3 Hydra viridissima Pallas, 1766 15

1.4 Test substance 17

1.5 Breakdown of chapters 18

Chapter Two: **A standardized tri-trophic small-scale system (TriCosm) for the assessment of stressor induced effects on aquatic community dynamics 21**

2.1 Abstract 21

2.2 Introduction 22

2.3 Materials and Methods 25

2.3.1 Test organisms 25

2.3.2 The TriCosm 26

2.3.3 Monitoring of the populations 26

2.3.4 Assessing a suitable community composition 27
2.3.5 Validation of an optimal experimental setup 28
2.3.6 Population dynamics and interactions 28
2.3.7 Statistical analyses 29
2.4 Results 30
2.4.1 Optimizing the experimental setup 30
2.4.2 Validation of the test setup 32
2.4.3 Population dynamics and interactions 34
2.4.4 Statistical analyses 36
2.5 Discussion 36
2.6 Conclusion 41
2.7 Acknowledgement 41
Chapter Three: **Species interactions and indirect effects in a standardized tri-trophic laboratory microcosm exposed to a pesticide 43**
3.1 Abstract 43
3.2 Introduction 44
3.3 Materials and Methods 47
3.3.1 Test organisms 47
3.3.2 TriCosm setup, sampling and environmental parameters 48
3.3.3 Test substance, dosing and chemical analysis 49
3.3.4 Statistical analysis 50
3.4 Results 51
3.4.1 Linuron analysis and solvent control effects 51
3.4.2 Experiments 1 and 2 52
3.4.3 Experiments 3 and 4 54
3.4.4 Experiments 5 and 6 55
3.4.5 Control population dynamics in exposures and the TriCosm 57
3.5 Discussion 58

3.5.1 Indirect effects and analysis of multi-species data 61

3.5.2 TriCosm repeatability and the influence of the medium 63

3.6 Conclusion 66

3.7 Acknowledgement 66

***Chapter Four*: Factors affecting the growth of *Pseudokirchneriella subcapitata*: lessons on the experimental design and consequences for the reproducibility of a multi-trophic laboratory microcosm 67**

4.1 Abstract 67

4.2 Introduction 68

4.3 Materials and methods 71

4.3.1 Algal culture and general experimental conditions 71

4.3.2 Which experimental components limit algal growth? 72

4.3.3 Does increased light exposure affect algal dynamics? 73

4.3.4 Does the condition of the exposure vessels affect algal dynamics? ... 73

4.3.5 Does the test medium impact algal dynamics? 74

4.3.6 Statistical analyses 76

4.4 Results 81

4.4.1 Algal growth rates: comparison among treatments 81

4.4.2 Algal population dynamics: medium, shaking speed and light exposure 83

4.4.3 Algal population dynamics: vessel type/condition and metal concentration/composition 84

4.4.4 Algal population dynamics: inoculation density 85

4.4.5 Algal population dynamics: medium buffering, aseptic preparation and aeration 86

4.4.6 Medium pH 89

4.5 Discussion 89

4.5.1 What caused alterations to population dynamics? 90

4.5.2 What does this mean for the reproducibility of multi-species systems? 92

4.5.3 With so many ways of getting it wrong – how shall we get it right? .. 93

4.6 Conclusion 95

4.7 Acknowledgement 95

***Chapter Five*: General Discussion and Conclusions 97**

5.1 Summary of thesis aims and results 97

5.2 Limitations and future outlook 101

5.3 Conclusion 109

Appendices 111

Appendix A – Supporting Data for Chapter Two 111

Appendix B – Supporting Data for Chapter Three 125

Appendix C – Supporting Data for Chapter Four 133

Abbreviations 143

Full list of references 145

List of Tables

Table 1: Publication status of chapters presented in this thesis. XV

Table 1 - 1: Micro- and mesocosm size and volume...8

Table 2 - 1: The TriCosm community composition at the beginning of 4 test setups. 27

Table 3 - 1: Experimental dates, treatments, duration, solvent acetonitrile and Na_2SiO_3 (for medium preparation) concentrations in experiments 1 - 6. 51

Table 4 - 1: Conditions in different experimental treatments with the alga *P. subcapitata*. Reported are the factors investigated, experimental duration, characteristics of the medium, light exposure, shaking speed, algal inoculation density, vessel type and replicate number of each experiment and treatment. Factor variations among treatments and experiments are indicated in bold. 77

Table A - 1: Chemical composition, preparation and storage of stock solutions for the TriCosm medium (ASTM E1366-11, 2011). 112

Table A - 2: Volumes (mL/L) of stock solutions needed for the preparation of the TriCosm medium (ASTM E1366-11, 2011). 113

Table A - 3: Plate reader settings for in-vivo fluorescence analyses. 113

Table A - 4: Assembly of the multi-species TriCosm with sequential addition of species and regular monitoring (here experimental set up on Thursday). 123

Table A - 5: Generation (G) renewal of cultured *C. dubia* females for continuous supply of neonates from third or fourth broods. 124

Table B - 1: Count settings, corrections and background subtraction during liquid scintillation counting to analyse ^{14}C linuron concentrations in aquatic samples. ...126

Table B - 2: Method for HPLC analysis to determine unlabelled linuron concentrations in aquatic stock solutions. 127

Table B - 3: Classification scheme for *Hydra* toxicity according to morphology scores (1 - 10). Scores ≥ 6 are considered sub-lethal and reversible whereas scores ≤ 5 are considered lethal. Redrafted from Quinn et al., 2012 from Wilby, 1988. 128

Table B - 4: Sodium metasilicate (Na_2SiO_3) effect concentrations are reported for *P. subcapitata, C. dubia* and *H. viridissima*. Endpoint, age, duration and concentration range are indicated. 132

Table C - 1: Plate reader settings for in-vivo fluorescence analyses. 135

Table C - 2: Chemical composition, preparation and storage of stock solutions for the medium T82MV (from ASTM E1366-11, 2011). 136

Table C - 3: Preparation and storage of the medium T82MV (from ASTM E1366-11, 2011). 137

Table C - 4: Chemical composition, preparation and storage of Keating’s metal solution. Only trace metals that are not added present in the medium T82MV are added. Once Keating’s trace metals are added to T82MV medium, the medium is called T86MVK (from ASTM E1366-11, 2011). 137

Table C - 5: Sodium metasilicate (Na_2SiO_3) effect concentrations are reported for *P. subcapitata, C. dubia* and *H. viridissima*. Endpoint, age, duration and concentration range are indicated. 137

List of Figures (short titles)

Figure 1 - 1: Schematic presentation of the tiered risk assessment approach of acute and chronic effects of PPPs in the aquatic environment (redrafted from EFSA 2013). 5

Figure 1 - 2: Morphology of the unicellular green alga *P. subcapitata.* 14

Figure 1 - 3: a) Adults and juveniles of *C. dubia* and b) schematic representation of *C. dubia* (scale bar 100 µm; from Munger et al., 1998). 15

Figure 1 - 4: *H. viridissima* with buds and sexual organs. 17

Figure 2 - 1: Algal concentrations and total number of *Ceriodaphnia dubia* and *Hydra viridissima* over 21 d. 31

Figure 2 - 2: Coefficients of variation of algal concentrations (cells/mL) and total abundance of *Ceriodaphnia dubia* and *Hydra viridissima* at each sampling event... 32

Figure 2 - 3: Abundance of (A) *Pseudokirchneriella subcapitata*, (B) *Ceriodaphnia dubia*, and (C) *Hydra viridissima* at each sampling point over 21 days. 33

Figure 2 - 4: Algal population trajectories compared between (A) ungrazed and grazed systems and (B) grazed systems with and without predation. Population dynamics of *Ceriodaphnia dubia* with and without predation as (C) total individual number and (D) juveniles and adults. 35

Figure 3 - 1: Population dynamics of *P. subcapitata* (log_{10} cells/mL), *C. dubia* and *H. viridissima* numbers (top to bottom row, respectively) in experiments 1 and 2. 52

Figure 3 - 2: Population dynamics of *P. subcapitata* (log_{10} cells/mL) and *C. dubia* in experiments 3 and 4. 55

Figure 3 - 3: Population dynamics of *P. subcapitata* (log_{10} cells/mL), *C. dubia* and *H. viridissima* numbers (top to bottom row, respectively) in experiments 5 and 6. 56

Figure 4 - 1: Specific algal growth rates (day^{-1}) are ranked from highest to lowest (top to bottom) in treatments of experiments 1 - 20. 82

Figure 4 - 2: Algal population dynamics in experiments 2 - 5 with differing light exposure (photoperiod- intensity (par µmol m^{-2} s^{-1}) and vessel type 84
Figure 4 - 3: Algal population dynamics in experiments 6, 8 – 11 and 14 with differing inoculation concentrations. 86
Figure 4 - 4: Algal population dynamics in experiments 15 - 20 with differing buffer salt concentration, type and/or medium pH adjustment via aeration (h) or HCl addition 88

Figure A - 1: Spearman correlation between suspended and total algal concentrations (cells/mL).. 118
Figure A - 2: Coefficients of variation (CV) of (A) algal concentrations (cells/mL), (B) total number of *C. dubia*, (C) *C. dubia* adults, (D) *C. dubia* juveniles and (E) *H. viridissima* over 21 days. 119
Figure A - 3: Comparison of coefficients of variation (CV) of algal concentrations (cells/mL), the total number of *C. dubia* and *H. viridissima* in Experiment 1 and Experiment 2, respectively 120

Figure B - 1: Population dynamics of adult and juvenile *C. dubia* in experiments 1, 2 and 6. 128
Figure B - 2: Water parameter pH in controls and treatments in experiments 1 - 6 over time. 130
Figure B - 3: Water dissolved oxygen (mg/L) in controls and treatments in experiments 1 - 6 over time... 131

Figure C - 1: Short-term (over three months) and long-term (over one year) reproducibility of the population dynamics of *P. subcapitata*, *C. dubia* and *H. viridissima* in the standardized TriCosm 134
Figure C - 2: Algal population dynamics in experiment 1. Medium type, light exposure (photoperiod- intensity) and vessel type among treatments. 138

Figure C - 3: Algal population dynamics in experiments 6, 7, 12 – 14 with differing medium sterility, vessel condition and type and trace metal composition and concentration. 138
Figure C - 4: Algal concentration differences between measurements taken before and after mixing the water columns in experiments 7 and 15 139
Figure C - 5: Water pH in experiments 1 and 2 139
Figure C - 6: Water pH in experiments 3 - 6. 140
Figure C - 7: Water pH in experiments 6, 8 - 11, 14 among treatments with different inoculation concentrations 140
Figure C - 8: Water pH in experiment 12 and 14 141
Figure C - 9: Water pH in experiments 15 – 17, 19, 20 in algal treatements and blanks 141
Figure C - 10: Water pH in experiment 18 in media blanks and algal treatments ..142
Figure C - 11: Algal growth dynamics compared among treatments with T82MV medium in crystallizing dishes, 2 x 10^4 cells/mL starting concentration and 65 rpm orbital shaking in experiments 1 – 20 142

Acknowledgements

First, I express sincere gratitude to Dr. Roman Ashauer, whose persistent support, availability, encouragement and love for Gantt Charts meant that this PhD was completed within time and to scientific rigour. He always provided me with opportunities and helped me gain new perspectives and enthusiasm throughout this project. I am thankful to Dr. Rachel Benstead for her critical questions, professional expertise, support and refusal to let me give up on the TriCosm. She kept me on track and helped making this thesis become, in parts, a nerdy detective story I'm somewhat excited to have investigated. Thanks go to Dr. Annika Agatz for her one-step-at-a-time policy and continuous support and Prof. Dave Raffaelli and Prof. Frieda Taub for many helpful discussions. Without their excellent guidance and convincing reassurance of my worth as a researcher and that of my work I couldn't have developed as much, both professionally and personally.

Throughout our four year love-hate relationship I wouldn't have considered it an option but I also thank the system I developed, the TriCosm; without it I wouldn't know just how stubborn I can be and would have missed a rock-hard lesson on exercising personal perseverance.

I profoundly thank my parents Renate and Martin, sisters Mirjam and Julia, my partner Ben, brother-in-law Robert, nieces Elena and Sofia and grandparents Louise, Sepp and Christl. Thank you for believing in me, your endless love, support, encouragements and visits in multiple locations during the past 10 years of my life as a student abroad.

Thanks go to my friends who supported me during this PhD; naming them and their contribution would exceed the space dedicated to this section but they'll know who I mean. The PhD is a unique and often lonely journey and heartfelt thanks go to my friends and colleagues from the University of York, Fera Science Ltd. and everyone in York who made me feel welcome. Finally, I thank the Department of Environment and Geography, Fera Science Ltd. and Dr. Roman Ashauer for funding.

Author's declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. Sources are acknowledged as References.

The presented work was supervised by Dr. Roman Ashauer and Dr. Rachel Benstead. I performed the presented work enrolled as a full-time PhD student at the Department of Environment and Geography at the University of York, York, and practical work was performed at the Centre for Chemical Safety and Stewardship at Fera Science Ltd., York.

The chapters are written as research articles for publication in peer-reviewed journals. Other than in published versions, the figures and tables were inserted close to their first reference in the main text but the content of published or submitted chapters was not changed. The Supplemental Data for each chapter was presented as Appendix. For consistency and to follow the formatting requirements of the presented thesis, citations were adapted and a full list of references is presented collectively at the end of the thesis.

I played the leading role in data collection, analysis and authoring all chapters but they were improved through advice and editing by co-authors and anonymous reviewers in case of publication. *Chapters Two, Three* and *Four* are co-authored and subject to joint authorship and the publication status of the presented work is given in Table 1.

Table 1: Publication status of chapters presented in this thesis.

Title	**Authors**	**Journal**	**History**	***Chapter***
A standardized tri-trophic small-scale system (TriCosm) for the assessment of stressor induced effects on aquatic community dynamics	Verena Riedl Annika Agatz Rachel Benstead Roman Ashauer	Environmental Toxicology and Chemistry	Submitted 24-06-2017 Revised 15-09-2017 Accepted 03-11-2017 Published 31-03-2018	*Two*
Species interactions and indirect effects in a standardized tri-trophic laboratory microcosm exposed to a pesticide	Verena Riedl Annika Agatz Rachel Benstead Roman Ashauer	To be decided		*Three*
Factors affecting the growth of *Pseudokirchneriella subcapitata*: lessons on the experimental design and consequences for the reproducibility of a multi-trophic laboratory microcosm	Verena Riedl Annika Agatz Rachel Benstead Roman Ashauer	Environmental Toxicology and Chemistry	Submitted 11-06-18 Revised	*Four*

Chapter One: General Introduction

Projections estimate the global population will grow to a number between 9.6 and 12.3 billion individuals by 2100 (Bradshaw and Brook 2014; Gerland et al., 2014) whereas the food demand will increase by up to 50 % within the next 15 years (Gnacadja, 2012). Given that the availability of arable land is restricted, this puts pressure on the agricultural industry to increase yield. Pesticides, or plant protection products (PPPs) are chemicals with different modes of action to increase the crop production efficiency and the agricultural sector is the leading user of these substances (Fairchild et al., 1994). Besides PPPs, non-chemical methods such as biopesticides or genetically modified crops can be employed to enhance yield production. Biopesticides are, for instance, bacteria, fungi or animals that act negatively on pests (e.g. through predation, parasitism) whereas genetic modification of crops is used to increase their resistance to insects and pathogens. However, both practices have drawbacks because of elevated cost and limited efficacy to the broad spectrum of threats (e.g. transgenic crops exhibit resistance only to specific agents). Less selective pesticides are thus still heavily used and appear as the most rapid and cost effective mean to intensify agriculture (Arias-Estévez et al., 2008).

PPPs can be applied with sprayers, via injection to treat seeds or with the irrigation system, depending on the characteristics of the active substance, the treated area, the frequency and the time of treatment. During and after application, however, the chemical substances may reach other compartments (soil, air or water) than the areas intended for treatment and cause environmental contamination. The main pathways for the contamination of water bodies are pesticide drift after spray application (e.g. with strong air movement), leaching through soil and surface runoff after rainfall or irrigation events (Arias-Estévez et al., 2008).

Because pesticides were extensively used over the past decades their occurrence is widely reported in surface waters (Stehle and Schulz, 2015; Malaj et al.,

2014; Spycher et al., 2018). There, they frequently affect not only the targeted species, but also non-target organisms and cause negative impacts on biodiversity (Beketov et al., 2013) and ecosystem services, for instance, when primary producers (Cardinale et al., 2011), pollinators (Davies et al., 2009; Blacquière et al., 2012) and nutrient cycling soil organisms (Eisenhauer et al., 2009) are affected. Based on the mode of action, or the mechanism by which the substance acts on the pest, pesticides often target traits common to different taxa. For instance, in the case of photosynthesis-inhibiting herbicides (e.g. atrazine, linuron), the electron transport inhibiting substance may impact the photosynthetic activity of all autotrophic plants and algae present in a treated area (Daam et al., 2009). Similarly, acetylcholinesterase inhibiting insecticides may affect various organism groups relying on this enzyme for neurotransmission (Fulton and Key, 2001).

Besides species sensitivity, the environmental context plays an important role in how and to what extent natural systems are influenced by toxicants (Relyea and Hoverman, 2006; Stampfli et al., 2011; Stampfli et al., 2013). Other than direct effects (e.g. on reproduction, growth and survival of organisms), indirect effects are likely to occur when interactive relationships with other system components are affected. Increasing evidence suggests that indirect effects are even more common and more complex than direct effects (Fleeger et al., 2003; Stampfli et al., 2011; Schäfer et al., 2016). In complex ecosystems where all components are interconnected directly or indirectly by mutual relationships, this could cause a variety of unexpected effects on ecosystem composition, structure and dynamics. Sudden changes in species composition could have, for instance, implications for the functioning of ecosystems and lead to unexpected system shifts (Scheffer and Carpenter, 2003). Knowledge on what mechanisms lead to structural alterations, what triggers them and quantifying effects on species interactions is, hence, important to evaluate pesticide impacts on the structure and stability of communities (Preston, 2002) and gain understanding of their role in maintaining and promoting resilience and recovery after disturbance events (Foit, Kaske, Wahrendorf, et al., 2012).

In this context, this PhD focuses on aspects of the environmental risk assessment of pesticides in the aquatic environment and in particular on the effect assessment of pesticides at ecologically relevant endpoints i.e. at the population level and on species interactions (European Food and Safety Authority (EFSA) 2013).

1.1 Environmental risk assessment of PPPs

To estimate the potential risks of PPPs to non-target species and their environment, the Organisation for Economic Co-operation and Development (OECD) introduced a chemical assessment program in 1977. It proposed guidelines regarding standardized testing procedures to be conducted prior to the commercialisation of new chemicals (Hushon et al., 1979). Standardised tests have been since adopted at an international level and are used by governmental and industrial stakeholders to evaluate and categorize chemical substances potentially harmful to biotic systems. OECD guidelines are followed for the fulfilment of data requirements in the European Union and similar guidelines were released in the US by the Environmental Protection Agency (EPA office of "Chemical Safety and Pollution Prevention"). Both EPA and OECD guidelines set regulatory testing standards and indicate how data necessary for the registration of pesticides can be obtained (EFSA, 2013). Much effort has been made to standardise and optimise the guidelines to minimize the variability between results obtained among different laboratories. For that reason, the OECD released a Council decision on the Mutual Acceptance of Data stating that test results obtained in independent laboratories are accepted in all OECD member countries and adherent, non-OECD countries, providing the tests are performed in accordance with the OECD test guidelines and under the Principles of Good Laboratory Practice (OECD, 2015).

The risk evaluation of compounds with potentially adverse effects on the environment is based on an integrated approach where measured or predicted exposure concentrations are compared with the results obtained during effect assessment tests.

Exposure assessment is carried out retrospectively when chemicals are already in use and problems were detected in accordance with the Water Framework Directive (WFD) (European Commission (EC), 2000) or prospectively before the chemicals are placed on the market in accordance with the legislation Regulation (EC) No 1107/2009 (EC, 2009). While, the WFD focuses prevalently on larger European water basins within the riverine context, the Regulation (EC) No 1107/2009 applies to smaller edge-of-field surface waters in agricultural settings. It is intended exclusively for the authorisation of PPPs proposed for the use within the European Union. The WFD requires the calculation of predicted no effect concentrations (PNEC; the concentration of a compound considered as harmless for a certain community) that are employed for the derivation of environmental quality standards (EQS). EQS are subsequently compared to monitored exposure values to assess the probability of adverse chemical effects to an environmental compartment and its ecological community of concern. To evaluate the risk of pesticides, the exposure assessment is carried out prospectively and predicted environmental concentrations (PECs) (derived with e.g. FOCUS scenarios and models for a given environmental compartment) are compared to regulatory acceptable concentrations (RACs; EFSA, 2013).

Effect assessment follows a tiered risk assessment scheme (for the aquatic environment; Figure 1 - 1) to estimate adverse effects of potentially hazardous chemicals on non-target organisms (Commission Regulation (EU) No 546/2011 (EC, 2011; EFSA, 2013). It is performed for the derivation of RACs that are obtained by dividing acute and/or chronic toxicity values by safety or assessment factors (AF). AFs are used as a hypothetical means to account for the uncertainty that arises for data limitations (data are incomplete, obtained only in laboratory tests, available for few species) and when the effects are extrapolated from one context to another (e.g. from a laboratory setting to natural ecosystems; Chapman et al., 1998).

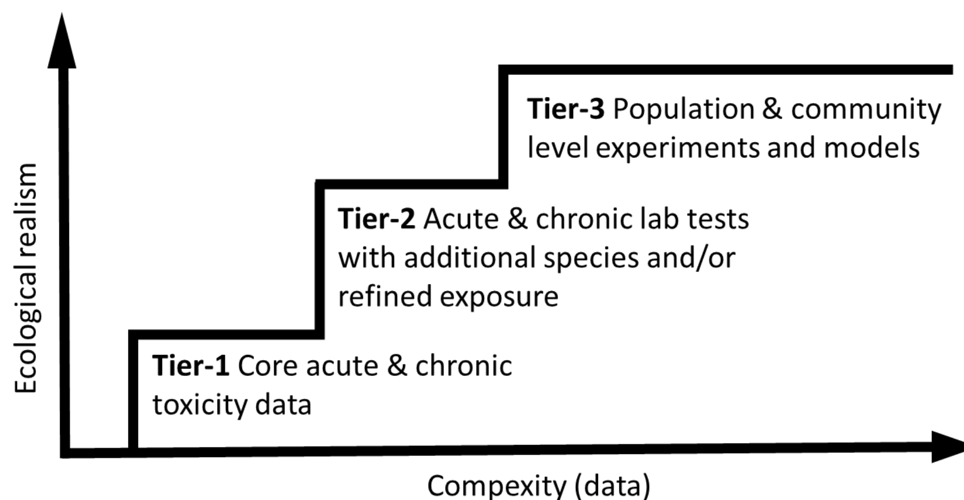

Figure 1 - 1: Schematic presentation of the tiered risk assessment approach of acute and chronic effects of PPPs in the aquatic environment (redrafted from EFSA 2013).

A tiered-testing system is used worldwide (Bednarska et al., 2013) and generally incorporates simple lower tier studies (Tier-1, Tier-2) and more complex higher tier studies (Tier-3). A comprehensive exposure or effect assessment is performed at each tier and assessment endpoints (RAC and PEC) are derived (EFSA, 2013).

1.1.1 Lower tier tests and why they might be insufficient to estimate risk

Lower tier, or single-species testing, is performed with a base-set of algal, crustacean and fish model species under standardized and optimized testing conditions for basic toxicity screening. Effects on species-specific endpoints (e.g. reproduction, growth, survival) are assessed and EC_x values (concentration at which x % of the population shows effects) are determined. For instance, during the algal growth inhibition test, impacts on the growth rate are assessed over 72 hours (OECD, 2011) or 96 hours (USEPA, 2012). The toxicity results are then used to determine whether refined exposure data with additional species or higher tier tests are necessary (EFSA, 2013).

When compared to higher, ecologically more relevant tiers, lower tier testing is considered as cost-effective for industry and regulatory stakeholders because it requires fewer resources (i.e. effort, time and money). The endpoints (e.g. LC_x, EC_x,

lethal or effective concentrations, respectively) can be directly compared among compounds and different studies and the results are accepted among all OECD adherent countries. However, the results might lead to insufficient information for the understanding of the magnitude of chemical impacts on natural sites (Campos et al., 2014; Dalinsky et al., 2014; Taub, 1997b). Even though AFs are used, general extrapolation of results to the complex environment (e.g. from one species to another with different morphology, physiology, sensitivity and distribution, from acute to chronic exposure and from the laboratory to the field) is hard because additionally influencing factors (e.g. indirect effects and species interactions) are not accounted for (Bednarska et al., 2013).

Direct effects on the reproduction or survival of sensitive species may lead to altered population densities and, in turn, indirectly affect other system components when food availability, predation and/ or competition pressure change. For instance, resistant species could be favoured indirectly under toxicant exposure to perform better in interspecific competition and out-compete more sensitive species. The gradual loss of sensitive species could lead to the reduction of species diversity, but might be concealed at first by compensatory mechanisms of functionally redundant, more resistant species. In other words, the functioning of the ecosystem as a whole initially might not be compromised because more resistant species with similar ecosystem functions may become more abundant and compensate for the role of lost sensitive species (Duffy et al., 2007).

Indirect effects are generally harder to assess and understand because they are mediated by interactions between system components (Fleeger et al., 2003). Hence, in interconnected natural systems, even components that are not directly affected could be impacted and lead to different toxicant-induced effects than could be predicted with lower tier test results that don’t consider species interactions (Taub, 1997b; Van de Perre et al., 2018; Preston, 2002). For instance, contrasting predator sensitivity to a contaminant might influence the competition outcome of competing prey species. If a predator shows higher sensitivity and decreases in abundance, a competing predator may increase in number and alter its predation

pressure on its preferred prey species. This could have indirect beneficial effects on any species competing directly with the more intensively preyed species or any food source of the latter (Bednarska et al., 2013). Consequently, the effects of toxicants on communities can hardly be predicted a priori if the relationships among species are not sufficiently understood (Preston, 2002). Even if the PNEC is higher than the reported EC_x, (EFSA, 2013), the toxic substance might still have an effect by lowering the tolerance to other stressors present in their environment. Stampfli et al., (2011) showed that zooplankton communities were ten times more sensitive at low food availability and high competition strength than when food was abundant and competition was weak. The environmental context, the sensitivity of species and the direct and indirect interactions with other system components need to be understood to predict the possible effect range of contaminants on communities in natural ecosystems (Preston, 2002; Taub, 1997b; Relyea and Hoverman, 2006; Rohr et al., 2006). Thus, the simplified representation of the environment in single-species tests and absence of stressor combinations which are prevalent in natural systems are a considerable source of uncertainty. Chemicals might be more dangerous in the natural environment than what would be expected according to lower tier test results or vice versa.

1.1.2 Higher tier, or multi-species tests

Higher tier tests, or micro- and mesocosms, are employed when the effect concentrations determined in lower tier tests compared to PECs raise concern. Multi-species systems are used to investigate direct and indirect effects of pesticides on species and their ecological interactions under a variety of different conditions (Liess, 2002; Relyea and Hoverman, 2008; Stampfli et al., 2011). Spatially and temporally restricted units may represent simplified subsets of environmental conditions and, because not all levels of organization are reproduced, the interpretation of examined processes is facilitated (EFSA, 2013; Van den Brink et al., 2005). The experimental set-up (e.g. species composition and abundance or abiotic factors), the biological complexity and the size of the systems can be varied based on the hypothesis in question. The distinction between micro- and mescocosms was

internationally defined based on the volume of the systems in the early ’90s (Crossland et al., 1992) but the terms are still often used interchangeably (Table 1 - 1).

Table 1 - 1: Micro- and mesocosm size and volume according to different authors.

Microcosm		**Mesocosm**		
Size	Volume	Size	Volume	Reference
< 15m^3 or < 15m length	< 15,000L	> 15m^3 or > 15m length	> 15,000L	Crossland et al., 1992
0.01 - 10m^3	10 - 10,000L	> 1m^3	> 1000L	Cooper and Barmuta, 1993
	< 1000L		> 1000L	Lasserre, 1990
	< 1L		1 - 100L	Srivastava et al., 2004

Multi-species systems can account for ecological effects at the population and/or community level (Bednarska et al., 2013; EFSA, 2013). Effect data on key processes can be studied to understand how basic processes operate in similar larger ecosystems (MacDonald, 2016), for instance, by extrapolating the results to bigger systems with ecological models (De Laender et al., 2008). However, while the larger volume of mesocosms and large microcosms allows a variety of interacting factors and increased environmental realism that captures greater complexity, system dynamics can easily diverge and complicate the extrapolation of causal mechanisms, i.e. associate an observed stress response to what has caused it (Bednarska et al., 2013). They require more effort to be established, maintained and analysed which makes them costly and time consuming and often limits the number of replicates to a handful of units. Complex and divergent system dynamics further frequently increase the replicate variability and decrease the statistical power which is another factor complicating the interpretation of effects. Therefore, abiotic factors and the composition and dynamics of species can often be controlled more easily in smaller and less complex microcosms (Foit, Kaske, Wahrendorf, et al., 2012).

A system of intermediate complexity that bridges simple single-species tests and complex microcosms could help filling knowledge gaps on the effects of toxicants on basic community processes such as species interactions. Common processes and

effect/response relationships can be studied in systems with a lower level of complexity and extrapolated to more complicated systems with similar principles. Classic examples of ecological processes that were studied in experimental microcosms to understand processes in natural ecosystems and more complex situations are the studies of Gause (1934) and Huffaker (1958). Gause (1934) examined what is known today as the competitive exclusion principle by studying the competition between species of yeast and paramecium for the same food resource. Huffaker (1958) examined the coexistence of predator and prey species by focusing on their oscillations. These experiments contributed to the understanding of both the exclusion mechanisms of species competing for the same resource and the processes that enable predators and prey to coexist in larger ecosystems.

Similarly, simple systems were used more recently to understand the effects of chemicals on species interactions and ecosystem functions in the aquatic environment. For instance, Englert et al., (2012) reported that field relevant concentrations (0.50 - 1.00 µg/L) of an insecticide significantly increased the predation of *Gammarus fossarum* on insect nymphs. This increase was not sufficient to compensate for a decreased leaf litter consumption at increasing insecticide concentrations and findings showed how pesticides might affect trophic interactions and possibly alter ecosystem functions. Viaene et al., (2015) investigated effects of different levels of intra- and interspecific competition and predation in conjunction with pulses to the hydrocarbon pyrene on *Daphnia magna* populations. Pyrene itself mostly affected juvenile individuals of *D. magna* individuals, however, species interactions within and between populations altered population responses to chemical exposure. The authors reported antagonistic effects between the test substance and both predation and competition but among all interactions tested, predation had the largest negative effect on population densities. Interactions among species might thus considerably influence chemical effects at higher levels of organization and are important factors to be considered in ecological risk assessments.

1.1.3 Microcosms for chemical risk assessment in the aquatic environment - State of the science

Many aquatic microcosms of variable dimension, test duration and complexity (e.g. species richness, composition, trophic levels) have been used to study ecotoxicological effects of pesticides (Liebig et al., 2008; Viaene et al., 2015; Taub, 1997b; Foit, Kaske, Wahrendorf, et al., 2012; Barry and Davies, 2004; van Wijngaarden et al., 2010; Fleeger et al., 2003). However, few of them describe impacts on communities in controlled systems smaller than 10L (hereafter small-microcosms; Taub, 1989; Taub, 1997a; Barry and Davies, 2004; Liebig et al., 2008; Englert et al., 2012; Foit et al., 2012b; Del Arco et al., 2015) and they rarely employ more than 5 replicates (Englert et al., 2012; Taub, 1997b; Foit, Kaske and Liess, 2012; Stampfli et al., 2013; Taub, 1989). Few of them describe both direct (reduced abundance) and indirect effects (increased or reduced abundance after an alteration of competition or predation pressure or food availability; Fleeger et al., 2003) and allow an understanding of causal mechanisms (Del Arco et al., 2015; Foit, Kaske, Wahrendorf, et al., 2012; Stampfli et al., 2013; Viaene et al., 2015; Taub, 1997b). Many of them focus on species either belonging to the same or two trophic levels to study effects on intra- or interspecific competition (Foit, Kaske and Liess, 2012; Del Arco et al., 2015) or consumer-resource relationships (Liebig et al., 2008; Englert et al., 2012; Viaene et al., 2015). Few microcosms consider effects at the community level and on three or more trophic levels (Harrass and Taub, 1985; Van Donk et al., 1995; Daam et al., 2009) in systems smaller than 10 L (bacteria-flagellate-ciliate community (Fuma et al., 2000); alga-grazer-bacteria community (Taub, 1976); algae-grazer-predatory zooplankton community (Barry and Davies, 2004)).

Altermatt et al., (2015), for instance, published a broad overview of methods for protist microcosm experiments with the aim of improving the quality and standardization of microcosm research. The authors propose their application in fields of evolutionary biology (e.g. adaptation) and ecology (e.g. competition and predation) to obtain mechanistic insights to system processes. However, only one small standardized aquatic microcosm (SAM; Taub, 1993) is known to the authors

that was standardized and developed for regulatory pesticide testing (ASTM E1366-11, 2011). An extrapolation of causal mechanisms and effect/response relationships in the SAM might, however, be difficult because ten algal and five invertebrate species interact through numerous pathways. The time scale of this and other macroinvertebrate small-microcosms is relatively long (3 months (Barry and Davies, 2004); 2 months (Taub, 1989); 1.5 months (Foit, Kaske and Liess, 2012)) and the reproduction times of the predators involved usually exceed the testing times so that effects on predator population dynamics generally cannot be assessed (*Gammarus fossarum* (Englert et al., 2012); *Chaoborus sp.* Larvae (Viaene et al., 2015); notonectids (Barry and Davies, 2004)).

Although the SAM was tested for repeatability and successfully passed inter-laboratory testing, multi-species tests are no longer required for regulatory pesticide testing in the US since the early 1990s (Taub, 1997a). In Europe, instead, multi-species tests are used at higher tiers of ERA and scientific working groups of the European Commission indicated the need of ecologically more relevant endpoints to improve the assessment of effects of pesticides at the population level (Scientific Committee on Health Environmental Risks et al., 2013). Yet, chemical effects derived in ecologically relevant systems are often not used for regulatory purposes because the data do not fulfil core regulatory requirements (standardization, repeatability, interpretability; Hanson et al., 2017).

Standardized multi-trophic systems tested for repeatability that allow the assessment of pesticide effects on species interactions and population dynamics could help quantify the effects of chemicals on species interactions. Even so, experimental systems of intermediate ecological complexity, simple design, short test duration and with known causal mechanisms are lacking. De Laender et al., (2011) described that exposure data obtained in multi-species tests is often reported as abundance changes of species but it is less often reported how and to what extent these responses affect associated interactions and functions. Similarly, Rohr et al., (2006) reported that overlooking indirect effects of contaminants might be one of the reasons why adverse effects of toxicants are still observed. Systems designed to

investigate the extent of chemical effects on species interactions could thus improve our understanding of the mechanisms by which multi-species systems are affected. In that manner, not only chemical-induced effects but also species-mediated effects of chemicals on ecosystem processes could be predicted. Key processes such as toxic effects on interactions and whether these influence aquatic community dynamics could be determined in an aquatic community with primary producer, grazer and predator species.

1.2 Aims and objectives

Given the research need described above, the aim of this PhD project was to design and test a standardized multi-species system for the repeatable assessment of direct and indirect pesticide effects on ecological interactions. To achieve this aim, the objectives were as follows:

(a) Optimising the seeding density and addition sequence of three species to standardize a multi-species test system. The focus will be on the interactions between species and minimizing the variability among experimental units.
(b) Understanding and characterizing the population and community dynamics of the trophic chain to allow predictions on the extent of system responses to alterations of its components.
(c) Assessing the repeatability of the system dynamics in the short and long-term.
(d) Testing how and to what extent pesticides alter interactions among species and the dynamics of the test community.
(e) Investigating the influence of different experimental factors on the algal population dynamics to refine the range of conditions under which the tri-trophic microcosm can be reproduced.

1.3 Model species

The test species used in this thesis are the green alga *Pseudokirchneriella subcapitata,* the cladoceran *Ceriodaphnia dubia* and the cnidarian *Hydra viridissima.* These species were selected based on their sensitivity to biotic and abiotic stressors, trophic relationship with each other, small adult size and short generation time.

All species are interconnected through consumer-resource relationships and can be found in calm waters (e.g. ponds and lakes) in warm and temperate regions. While *P. subcapitata* and *C. dubia* are planktonic species living in the water column (Stewart and Konetsky, 2008; Yamagishi et al., 2017), *H. viridissima* is an epibenthic species that can be found attached to substrata or water plants (Massaro and Rocha, 2008). The species show small adult size (important to limit the requirements for space and costs for culture maintenance), high growth and reproduction rates and short generation time allowing the assessment of chemical effects on several generations and life stages.

1.3.1 Pseudokirchneriella subcapitata (Korshikov) F. Hindák 1990

P. subcapitata (Figure 1 - 2; also known as *Selenastrum capricornutum* (Printz, 1914), *Ankistrodesmus subcapitata* (Korshikov 1953) and *Raphidocelis subcapitata*) is currently the most frequently used green alga in the regulatory risk assessment framework (OECD, 2011) due to its high sensitivity to toxicants. This species takes the base level of the trophic web as autotrophic phytoplankton and is an important contributor to primary production and acts as a food source for many filter-feeding species. The cells are horseshoe shaped and average 8-14 µm in length and 2-3 µm in width, proliferate via multiple fission to two, four or eight autospores (daughter cells) and an average population growth rate of up to 1.95 per day was reported (Yamagishi et al., 2017). The unicellular algae can exhibit grazer-induced colony formation once triggered by chemical cues released by foraging herbivorous zooplankton to impede or alter the foraging behaviour of filter feeders such as daphnids (Kampe et al., 2007).

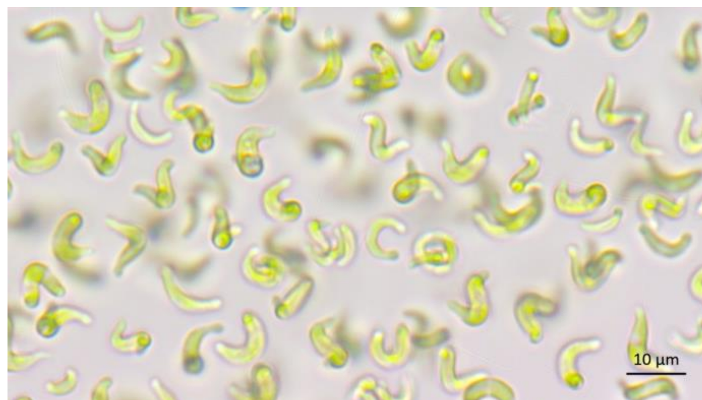

Figure 1 - 2: Morphology of the unicellular green alga *P. subcapitata.*

1.3.2 Ceriodaphnia dubia Richard, 1894

C. dubia is (besides *Daphnia magna*) currently one of the most frequently used crustacean species in standardized toxicological studies (USEPA, 2002) to estimate acute and chronic toxicant effects (Figure 1 - 3; Vasquez et al., 2014). It shows high sensitivity to toxicants (Rose et al., 2002b; Stewart and Konetsky, 2008) and plays a dominant role in the functioning and structure of in many freshwater bodies. As part of the zooplankton, they occupy a central position in the food web as an important food source for predatory species (e.g. carnivorous invertebrates and fish) and their grazing action is essential for the density control of algal species. An alteration of the grazing pressure could lead to increased primary production and algal blooms with effects on the whole community (e.g. water turbidity may affect the photosynthesis of plants and a lack of oxygen leads to hypoxic conditions for animals; Hanazato, 2001). This filter feeder performs diel migrations moving vertically up and down the water columns depending on seasonal differences in food supply (e.g. bacteria, algae, particulate organic matter, protozoans) and predator abundance. Under favourable conditions, individuals in *C. dubia* populations are all female that rapidly reproduce via parthenogenetic asexual reproduction. Depending on environmental parameters (e.g. food availability, temperature), an average life span of 36.1 days and average body size of 0.36 and 0.87 mm for neonates and adults, respectively, and on average 12 broods (7.3 individuals each) for regularly fed animals were reported (Anderson

and Benke, 1994). Females commonly start reproducing at the age of approx. 4 days and produce one batch (group of offspring) approx. every 2 days throughout their lifetime (Stewart and Konetsky, 2008). These traits make *C. dubia* an excellent test organism for ecotoxicological testing because the use of large batches and numerous genetically identical neonates decreases the variability caused by inter-individual response differences. Under stress conditions (e.g. high competition, predator abundance or scarce food availability) female cladocerans shift to sexual reproduction and produce haploid eggs and sexual males. When haploid eggs are fecundated (ephippia) these enter a dormant stage and allow cladoceran populations to endure and survive unfavourable conditions (Hobaek & Larsson, 1990; Schön et al., 2009).

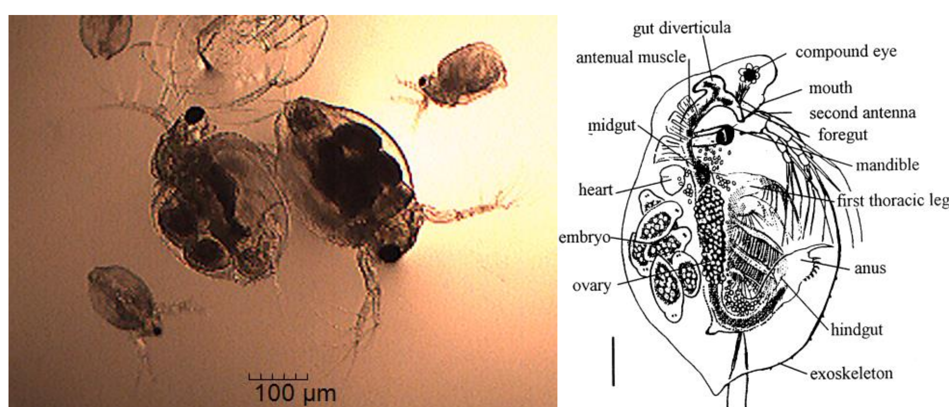

Figure 1 - 3: a) Adults and juveniles of *C. dubia* and b) schematic representation of *C. dubia* (scale bar 100 µm; from Munger et al., 1998).

1.3.3 Hydra viridissima Pallas, 1766

H. viridissima (Figure 1 - 4; formerly known as *Chlorohydra viridissima;* Massaro and Rocha, 2008) is not used in standardized toxicity tests but is increasingly used to assess the impacts of environmental contaminants due to its sensitivity to metals (Quinn et al., 2012), the endocrine disruptor nonylphenol (Pachura-Bochet et al., 2006), pesticides (Kovacevic et al., 2009; Demetrio et al., 2012), mycotoxins (Brown et al., 2014) and pharmaceuticals (Quinn et al., 2008). Physiological and behavioural (Kovacevic et al., 2007; Kovacevic et al., 2009) endpoints can be used for toxicological

testing and morphological impacts can be measured with a morphology score index (Wilby, 1988). The index specifies the degree of morphological damage caused by chemicals and is measured according to toxicant induced changes of the animal's tentacles and trunks (Wilby, 1988; Quinn et al., 2012; Tökölyi et al., 2014). *H. viridissima* forms symbiotic interactions with photosynthetic green algae (*Chlorella vulgaris*) commonly referred to as zoochlorellae that inhabit the vacuoles of digestive cells and give *H. viridissima* its green colour (Massaro et al., 2013; Tökölyi et al., 2014). The transfer of nutritional products from algae to *H. viridissima* is favourable, especially during conditions of stress (Kovacevic et al., 2007; Kovacevic et al., 2009) and enables this species to survive starvation longer than other *Hydra* species without symbionts (Habetha et al., 2003). *Hydra* are carnivorous predators and take an intermediate position in the food web because they prey on zooplankton but are predated by other carnivores such as insect larvae, flatworms and small fish (Massaro et al., 2013). The sessile species feeds when small invertebrates such as cladocerans, annelids and copepods touch their tentacles. The animals paralyse the prey with poison from nematocysts, coil all tentacles around it and move the prey towards the mouth for ingestion in the centre of their tentacles (Massaro and Rocha, 2008). Under favourable conditions, reproduction occurs asexually via the formation of buds i.e. miniature polyps that detach from the mother animal. As for *C. dubia*, growth and reproduction rates depend on environmental parameters and population doubling rates can vary between 3, 14 and 95 days when polyps are fed daily, once per week or once every other week, respectively (Habetha et al., 2003). This species commonly starts budding (reproducing asexually) at an age of 5 days and at an average length of 2.2 mm and new polyps separate from the mother *Hydra* at an average length of 1.2 mm (Massaro and Rocha, 2008). *Hydra* population growth is hence assumed to respond rapidly to variation of *C. dubia* population size and chemically induced population effects can likely be measured within short (few weeks) test durations (Quinn et al., 2012; Massaro et al., 2013). The hermaphrodite species shifts to sexual reproduction when exposed to stress and produces ovaries and testis with free swimming gametes that fertilize the eggs of other individuals (Habetha et al., 2003).

Figure 1 - 4: ***H. viridissima*** **with buds and sexual organs.**

1.4 Test substance

The PPP used for the exposure of the test system is the herbicide linuron. It was first introduced in the 1960's, belongs to the urea substance group and is a selective and systemic herbicide inhibiting the photosystem II. It is used for the pre- and post-emergence control of annual grass and broad-leaved weeds such as chickweed, prickly lettuce or goose grass and is applied on crops such as carrots, tea, rice, beans or cereals (Lewis et al., 2016). Besides being moderately persistent, linuron is reported to easily enter surface waters through agricultural runoff, spray drift or leaching (Cuppen et al., 1997) and was selected based on the following properties (Lewis et al., 2016; Garthwaite et al., 2017):

- high solubility in water (63.8 mg/L),
- potential bioaccumulation (Log P 3.0),
- long degradation time via aqueous hydrolysis (1460 days) and photolysis (> 30 days at 20 °C and pH 7.0),
- ecotoxicity to the green alga *P. subcapitata* (EC_{50} 16 µg/L) and
- high application quantities on crops (102,582 tonnes in the UK in 2016).

Based on the degradation time, the herbicide concentration was expected to remain constant throughout the experimental duration of 21 days. Compared to its

extensive usage in agriculture, measurements of surface water concentrations of linuron are scarce but concentrations in the range from ng/L to µg/L were described (Webster et al., 2015).

1.5 Breakdown of chapters

The thesis is presented as five main chapters. *Chapter Two, Chapter Three* and *Chapter Four* are written as scientific publications that can be understood independently from the rest. Each chapter contributes to the thesis aim as specified above and briefly outlined below. While there is some repetition between chapters, this format was chosen to ease the publication of research findings. A full list of references is given at the end of the thesis and not as per paper-style format at the end of each chapter.

In *Chapter One,* I give a general introduction to current approaches of environmental risk assessment of pesticides in Europe focusing on the aquatic environment. The state of current science is outlined to identify emerging research needs and give direction for the following chapters.

In *Chapter Two,* I summarise the work to develop and standardise the repeatable aquatic, multi-species system TriCosm. I show how variations of addition sequence and timing of the three interacting populations (*P. subcapitata*, *C. dubia*, and *H. viridissima*) influence the variability between replicates and assess the repeatability of population dynamics in control systems. This chapter was published in Environmental Toxicology and Chemistry in 2018 and the supporting data are provided in *Appendix A*.

Chapter Three is directly based on *Chapter Two*. I summarize six herbicide exposure experiments in the TriCosm. The purpose of this chapter was to assess whether small indirect effects could be detected in the system and to what extent the population dynamics and species interactions in the trophic chain were altered. Inter-experimental variations of population dynamics in control systems were

compared to population dynamics in the standardized TriCosm to evaluate the long-term repeatability of control systems. This chapter contains additional supporting data in *Appendix B*.

In *Chapter Four,* I investigate how a series of experimental factors influence the population trajectories of the algal trophic layer in control systems and which matter the most. The purpose of this work was to experimentally refine the range of conditions under which the tri-trophic microcosm can be reproduced and to critically evaluate the findings in light of current debates over a reproducibility crisis in empirical research. This chapter was submitted to Environmental Toxicology and Chemistry and contains additional supporting data in *Appendix C*.

In *Chapter Five*, I delineate the principal outcomes and conclusions of all the work submitted as part of this dissertation. I outline how the research makes an original contribution to our understanding regarding the repeatability and ecological realism of multispecies tests in the context of the environmental risk assessment of pesticides. Research limitations are presented and areas for improvement and future perspectives are indicated.

Chapter Two: A standardized tri-trophic small-scale system (TriCosm) for the assessment of stressor induced effects on aquatic community dynamics

Published in Environmental Toxicology and Chemistry (2018), 37, 1051–1060

Verena Riedl,[a,b,*] Annika Agatz,[a] Rachel Benstead,[b] and Roman Ashauer[a]

[a]*Environment Department, University of York, Heslington, York, United Kingdom*

[b]*Fera Science Ltd., Centre for Chemical Safety and Stewardship, Sand Hutton, York, United Kingdom*

* Address correspondence to verena.riedl@york.ac.uk

2.1 Abstract

Chemical impacts on the environment are routinely assessed in single-species tests. They are employed to measure direct effects on non-target organisms but indirect effects on ecological interactions can only be detected in multi-species tests. Micro- and mesocosms are more complex and environmentally realistic, yet, they are less frequently used for environmental risk assessment because resource demand is high while repeatability and statistical power are often low. Test systems fulfilling regulatory needs (i.e. standardization, repeatability and replication) and the assessment of impacts on species interactions and indirect effects are lacking. Here we describe the development of the TriCosm, a repeatable aquatic multi-species test with three trophic levels and increased statistical power. High repeatability of community dynamics of three interacting aquatic populations (algae, *Ceriodaphnia*, *Hydra*) was found with an average coefficient of variation of 19.5% and the ability to determine small effect sizes. The TriCosm combines benefits of both single-species tests (fulfillment of regulatory requirements) and complex multi-species tests (ecological relevance) and can be used, for instance at an intermediate tier in environmental risk assessment. Furthermore, comparatively quickly generated

population and community toxicity data can be useful for the development and testing of mechanistic effect models.

2.2 Introduction

The thorough assessment of environmental risks is essential for chemicals that could potentially be released into the environment. Agricultural pesticides, for instance, are used to enhance crop production but due to their toxic nature they may have negative effects on organisms other than the targeted species (Benton et al., 2007; Rockström et al., 2009; Beketov et al., 2013; Stehle and Schulz, 2015).

The risks of pesticides to non-target organisms are routinely assessed in i) simple single-species tests at lower tiers and, if lower tier assessments raise concern, in ii) complex microcosms or mesocosms (EFSA, 2013). However, systems that bridge the gap between the two alternatives to an intermediate level of complexity are lacking. The former require less effort and rapidly deliver large amounts of highly repeatable data on the performance of individual non-target organisms. The information obtained is, however, often of low ecological relevance as it is not directly relevant at the population and community level (Fleeger et al., 2003; Liebig et al., 2008). In contrast, micro-/mesocosms are environmentally more realistic, yet, they are less frequently used in environmental risk assessment. Unlike single-species tests, they are resource, time and effort demanding. A variety of ecologically interacting factors can rapidly lead to divergent system dynamics and increase the variance between replicates impeding the understanding of dose-response relationships (Landis et al., 1997).

The importance of integrating environmental complexity into testing approaches has been acknowledged as a priority for the assessment of chemical safety (Landis et al., 1997; Landis, 2003; Scientific Committee on Health Environmental Risks et al., 2013; Bednarska et al., 2013). Chemical exposure could trigger indirect effects through interactions with the environmental context such as

the hydrological regime (Stampfli et al., 2013), temperature (Moe et al., 2013), food quality (Campos et al., 2014) or other organisms (Viaene et al., 2015; Del Arco et al., 2015); indirect effects have important implications for the sensitivity of communities (Fleeger et al., 2003).

Organisms living in a contaminated environment may be pushed towards the boundaries of their ecological niche and become more susceptible to additional stressors (Van Straalen, 2003; Bednarska et al., 2013). Food chain processes, such as competition for food and altered predation were shown to be particularly relevant to determine the magnitude of toxic effects (Heugens et al., 2001; Bednarska et al., 2013; Kattwinkel et al., 2015). For instance, the no observed effect concentration (NOEC) of the herbicide prometryn to ciliates was found to be approximately 145 times lower in a bi-trophic microcosm compared to single-species tests. The lower threshold was likely caused in response to an indirect and toxicant induced reduction of food (Liebig et al., 2008). Intraspecific competition can also change the sensitivity to pesticides (Foit, Kaske and Liess, 2012; Viaene et al., 2015) and indirectly altered predation rates can lead to cascading effects on other trophic interactions and ecosystem functions (Englert et al., 2012; Agatz et al., 2014; Viaene et al., 2015). Multi-species testing using environmentally more relevant approaches, i.e. at the population and community level is clearly needed to assess indirect toxicant effects such as shifts in ecological interactions (Fleeger et al., 2003; Benton et al., 2007).

The necessity towards an inclusion of ecological interactions in chemical impact testing was described 10 years ago when a review on 14 years of pesticide studies in freshwater test systems was published (Relyea and Hoverman, 2006). At the time, the authors found only 133 studies with at least two potentially interacting species of which only 17 studies focused on three trophic levels with producers, herbivores and carnivores.

Yet, microcosms that describe impacts on populations and/or communities in systems smaller than 10 L are rare (Metcalf et al., 1971; Daam and Van Den Brink, 2007; Liebig et al., 2008; Englert et al., 2012; Foit, Kaske and Liess, 2012; Dolciotti et

al., 2014; Viaene et al., 2015; Del Arco et al., 2015). Mostly they were used to focus on impacts on intra- or interspecific competition (one trophic level) (Foit, Kaske and Liess, 2012; Dolciotti et al., 2014; Del Arco et al., 2015; Viaene et al., 2015) or on consumer-resource relationships (two trophic levels) with herbivore-producer (Daam and Van Den Brink, 2007) or predator-prey interactions (Barry and Davies, 2004; Liebig et al., 2008; Englert et al., 2012). Tri-trophic systems are frequently used in terrestrial research, for example in plant-herbivore-parasite systems (Bredeson et al., 2015; Uhl et al., 2015) but few small test systems exist to assess direct and indirect impacts at the population and community level in the aquatic environment. Test formats include simulations of microbial detritus food chains (producer-consumer-decomposer (Fuma et al., 2000; Dawoud et al., 2017)) and producer-consumer communities with either invertebrate predator (Barry and Davies, 2004) or vertebrate predator (Metcalf et al., 1971). Microbial tests were often conducted in culture flasks (250 mL) (Fuma et al., 2000; Dawoud et al., 2017) and small macroinvertebrate community tests were performed in systems of few litres, for example in 10 L (Barry and Davies, 2004) and 7 L systems (Metcalf et al., 1971).

Still, single-species systems appear convenient because they fulfill the regulatory needs for international standardization of test procedures, comparability of effect data, repeatability and replication (Liebig et al., 2008). Standardized and repeatable multispecies systems of intermediate complexity that bridge the simplicity of single species tests and the complexity of microcosms, yet fulfill regulatory requirements, are rare. To our knowledge, there is only one standardized microcosm (Taub, 1989) available that falls into this category. The aquatic system was registered for pesticide testing (American Society for Testing of Materials (ASTM), 2011) and effects on two trophic levels covering ten primary producer and five primary consumer species can be assessed. It is, however, rarely used for standardized effect assessment, perhaps due to its relative complexity and the lack of mechanistic understanding of the interactions between species involved.

We developed a new test system with species interacting across three trophic levels and increased statistical power (i.e. standardization and low replicate

variability). The system was designed to be cost-effective, rapid, repeatable with well understood population dynamics to i) allow the detection of small changes in population dynamics due to direct and indirect interactions, and ii) link observed effects to known system processes. Here we describe the standardized aquatic tri-trophic microcosm (hereafter TriCosm) focusing on system design and variability in the control treatment.

2.3 Materials and Methods

2.3.1 Test organisms

The TriCosm comprises populations of the green alga *Pseudokirchneriella subcapitata*, the cladoceran *Ceriodaphnia dubia* and the cnidarian *Hydra viridissima.* This dynamic food-chain is subject to fluctuating but predictable changes in food supply and intraspecific competition and is interconnected through consumer-resource relationships. The species were chosen based on their rapid life cycles and their sensitivity to toxicants. The green alga *P. subcapitata* and the cladoceran *C. dubia* are routinely used for tests in the regulatory risk assessment framework (Organization for Economic Co-operation and Development (OECD), 2004; OECD, 2006; OECD, 2012).

P. subcapitata stock cultures were obtained from the Culture Collection of Algae and Protozoa (CCAP, Scotland, UK) and used to initiate a culture line prior to each study and cultured in OECD media (OECD, 2006). *C. dubia* were obtained from Unilever (Safety and Environmental Assurance Centre, Bedford, UK) and cultured as age specific cultures in moderately hard, synthetic freshwater (United States Environmental Protection Agency (USEPA), 2002). They were fed five times per week with a suspension of yeast, cerophyl® and trout-chow (~3.5 mL) and *P. subcapitata* (~11 x 10^7 cells/day) (USEPA, 2002). *H. viridissima* were obtained from the Department of Evolutionary Zoology (University of Debrecen, Hungary), cultured in modified T82MV medium (modified after ASTM E1366-11, 2011; *Appendix A* Table A

- 1, Table A - 2) and fed with newly hatched *Artemia salina* three times per week ad libitum. Both animal cultures were kept at 25 ± 1 °C and 12/12h light/dark cycle.

2.3.2 The TriCosm

TriCosms consist of Pyrex® crystallizing dishes (Sigma-Aldrich, UK) filled with 500 mL of T82MV medium (ASTM E1366-11, 2011; *Appendix A* Table A - 1, Table A - 2) that was determined as suitable for each species. The systems were covered with transparent watch glasses (diameter 125 mm; Sigma-Aldrich, UK) and positioned on an orbital laboratory shaker (Adolf Kuehner AG Switzerland, Type LS-W) set at 65 rpm throughout the test. The experiments were set up for 21 days at 25 ± 1 °C, 12/12h light/dark, 1100 lux at the water surface with cool white fluorescent light tubes 58.5 W (approx. 1.3 m above the test vessels). TriCosms were started with *P. subcapitata* cells from a culture in exponential growth phase, *C. dubia* neonates (≤ 24h age) from the third or fourth brood of cultured mothers and *H. viridissima* without visible buds (≤ 2 d age). Water parameters (pH, dissolved O_2) and animal and algal populations were monitored throughout the test duration two, three and five times per week, respectively.

2.3.3 Monitoring of the populations

The systems were placed on an orbital shaker and slow shaking kept the algal cells suspended. Only suspended algae were measured and no stirring was necessary before sampling as preliminary studies showed significant correlation between suspended and total algal concentrations (cells/mL) ($r_S = 0.98$, $p < 0.01$, $n = 90$, *Appendix* A Figure A - 1). In-vivo fluorescence activity of water subsamples (5 x 200 µl) was measured with a plate reader (Tecan® Infinite 200 PRO, settings *Appendix A* Table A - 3) to determine the algal concentration (cells/mL).

C. dubia and *H. viridissima* were monitored with non-invasive methods to avoid impacts on population dynamics and counted by eye three times per week. *C. dubia* were visually grouped in two age-classes, juveniles and adults based on their dimensional similarity with individuals in cultures aged younger or older than 4 d. All

manual counts were repeated until count differences did not exceed 20 % of the lower value.

2.3.4 Assessing a suitable community composition

Tests with different setups were performed to optimize replicate variability, test duration, addition times and densities for each species. A full factorial design for density and timing was not feasible due to a too high number of possible combinations. Hence, preliminary tests were performed to determine which algae-grazer combination in terms of organism abundance would prevent both algal blooms and the death of grazers due to starvation. No preliminary tests were done to determine the impact of *Hydra* predation on *C. dubia* numbers prior to the test outlined in Table 2 - 1.

Table 2 - 1: The TriCosm community composition at the beginning of 4 test setups.

	P. subcapitata		*C. dubia*		*H. viridissima*		
	Cells/mL	Day	Individuals	Day	Individuals	Day	Replicates
Setup 1	2×10^4	0	10	1	3	6	8
Setup 2	2×10^4	0	10	0	6	5	8
Setup 3	4×10^4	0	10	0	6	4	7
Setup 4	4×10^4	0	20	0	6	4	7

Two organism densities and different addition times were chosen based on preliminary testing and four different setups were conducted simultaneously (Table 2 - 1). *C. dubia* were added on the same day as the green algae in all experiments, except for setup 1 where grazers were added 1 day later to allow short acclimation of the algae to test conditions. Dependent on food concentrations, *C. dubia* matured later in setups 1, 2 than in setups 3, 4 hence *H. viridissima* were introduced to the systems 5 and 4 days, respectively, after *C. dubia* were added. The predators were added only once the grazers started reproducing to prevent variable numbers of *C. dubia* reproducers and neonates early on in the systems that could lead to noticeable

impacts on community dynamics and replicate variability. Replicate numbers differed between setups 1, 2 and 3, 4 due to space constraints on the shaker platform.

2.3.5 Validation of an optimal experimental setup

The coefficient of variation (CV) was calculated as a standardized measure of variance between replicates. It was expressed as a percentage and indicates the magnitude of the standard deviation in comparison to the mean. Thus, if the inherent variability between replicates is large compared to the size of the measured endpoint (e.g. animal abundance) a significant treatment effect could only be detected if the response was very large (Sanderson, 2002).

We computed the CVs for the algal concentration (cells/mL) and the total number of *C. dubia* and *H. viridissima* on every sampling day. The values were then compared within and between setups to monitor replicate variation over time and to determine an optimal setup in terms of low variance. The experimental setup with the lowest replicate variability (Experiment 1) was determined and repeated (Experiment 2) to assess the reproducibility of system dynamics and replicate variability.

2.3.6 Population dynamics and interactions

In the interacting system, the intermediate trophic layer is directly affected by both variations of food availability and predation strength, while indirect effects between the bottom and top trophic level regulate a bottom up or top down controlled system. A comparison of algal and grazer dynamics between systems where grazers are subject to i) variations of food but not to predation and ii) a combination of food limitation and predation can thus yield information on species interaction strength and whether the system is controlled by bottom up or top-down effects. Hence, we performed additional tests with i) only algae (Experiment 2A, n = 8) and ii) algae and grazers (Experiment 2B, n = 8). The experiments were carried out simultaneously to Experiment 2 and according to the experimental conditions and test setup used as for Experiments 1 and 2 (see Table 2 - 1 for details). Experiments 2A and 2B were

then compared to determine direct impacts of grazers on algal dynamics and Experiments 2 and 2B were compared to assess i) indirect effects of predators on algal growth and ii) direct effects of predators on *C. dubia* population trajectories.

2.3.7 Statistical analyses

An *a priori* power analysis was performed to estimate minimum detectable response sizes between control and treated TriCosm populations and increase the reliability and transparency of the derived endpoints (EFSA, 2013). The minimum detectable difference (MDD), i.e. the size of a variation between sample averages required to be detected as significantly different, is dependent on the chosen Type I error value α, the number of replicates employed and on the inherent variance such as replicate variability and/or sampling error (Brock et al., 2015). Here, we computed the MDD using the CVs assuming similar variance among controls and treatments. We hypothesized the use of 8 replicates and estimated the sensitivity of the TriCosm to reveal chemical effects for each population and at each sampling point. The MDD was calculated as described by Brain et al. (2005) from Sokal and Rohlf (1995):

$$\mathrm{MDD} = \frac{\sqrt{2}\left(\mathrm{t}_{\alpha,\nu} + \mathrm{t}_{\beta,\nu}\right) \mathrm{x\ CV}}{\sqrt{n}}$$

where $t_{\alpha,v}$ and $t_{\beta,v}$ are the t-values for α and β set to 0.05 and 0.2, respectively, for a confidence level of 95% and a power of 80% at ν degrees of freedom. CV is the coefficient of variation and n is the number of replicates used.

The degrees of freedom were computed as $\nu = k\,(n\text{-}1)$ and the number of groups k was set to 2, e.g. to compare each treatment to the control. The calculated MDDs were compared to MDD classes as proposed by EFSA (2013) that grouped MDD sizes into five classes and described the likely ability of effect detection (Class 0: MDD > 100% = no effect detection, Class I: MDD 90 – 100% = only large effects, Class II:

MDD 70 – 90% = large to medium effects, Class III: MDDs 50 – 70% = medium effects, Class IV: MDD < 50% small effects).

To assess species interactions between system components, population dynamics were compared graphically between experiments and significant differences were assumed where 95% confidence intervals did not overlap.

2.4 Results

2.4.1 Optimizing the experimental setup

The community dynamics (Figure 2 - 1) and the coefficients of variation differed among the four setups (Figure 2 - 2) and over time (*Appendix A* Figure A - 2). In general, the algal concentration (cells/mL) peaks were followed by *C. dubia* abundance peaks and a constant increase of *H. viridissima* populations. The highest *C. dubia* peak 172 (± 10) individuals (mean ± 95% confidence interval range) was found in Setup 1 on day 14 following an algal peak on day 6 when an average of 6.86 (± 0.64) x 10^5 cells/mL was measured (Figure 2 - 1 A). The highest algal peak of 11.51 (± 5.59) x 10^5 cells/mL appeared in Setup 3 on day 11 with increasing variance in terms of organism numbers between replicates over time (Figure 2 - 1 C). *H. viridissima* populations showed steady growth during the test duration and increased in numbers by an average of 42 (± 6), 45 (± 8), 72 (± 18) and 54 (± 15) individuals in Setup 1, 2, 3 and 4, respectively (Figure 2 - 1 A-D).

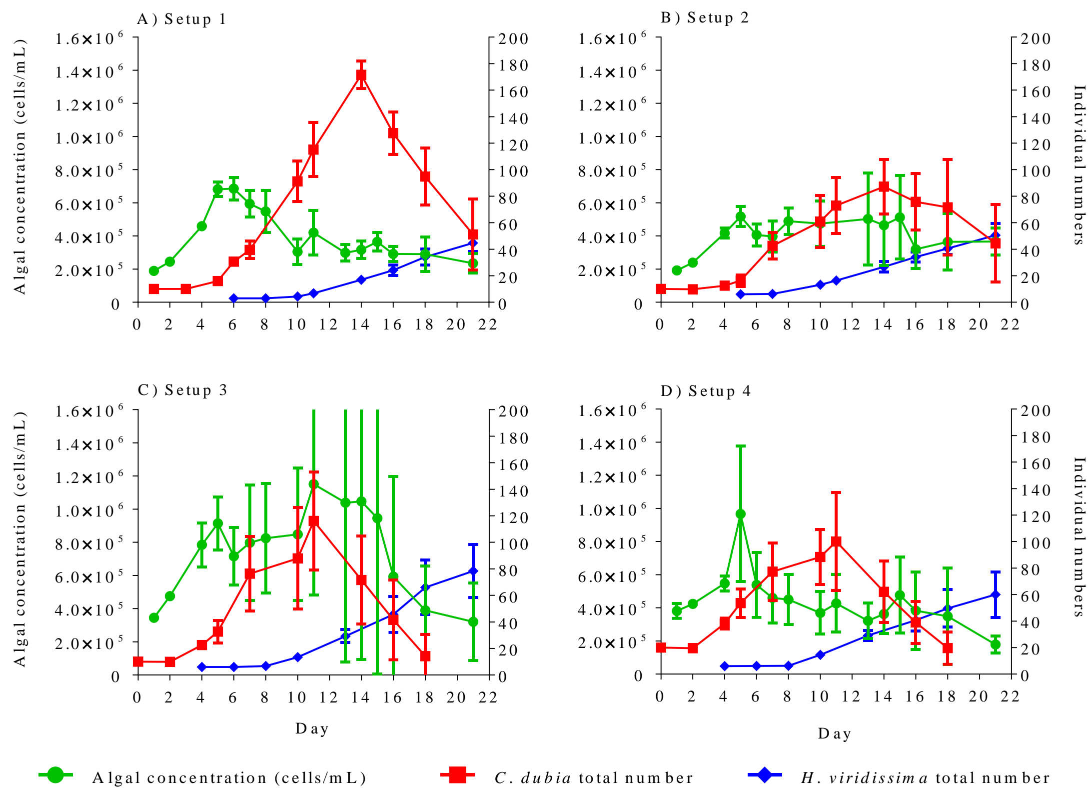

Figure 2 - 1: Algal concentrations and total number of *Ceriodaphnia dubia* and *Hydra viridissima* over 21 d. Shown are means ± 95% confidence intervals in 4 test setups (A–D; see Table 2 - 1 for details).

Final counts differed due to different addition numbers, timings and food availability (*C. dubia* abundances), however, *Hydra* populations showed the smallest replicate variance when compared to algae and *C. dubia* (Figure 2 - 2).

The CVs of all test variables in each setup increased over time (*Appendix A* Figure A - 2) with setup 3 showing the highest replicate variability (with the exception of *H. viridissima*) and setup 1 showing the lowest average CV (with the exception of *H. viridissima)* (Figure 2 - 2). High CVs observed in setups 2 - 4 indicated reduced ability to detect treatment related system alterations. Therefore, we selected setup 1 (Experiment 1) as the most appropriate setup procedure (*Appendix A*) and repeated the test (Experiment 2) to evaluate the repeatability of the system.

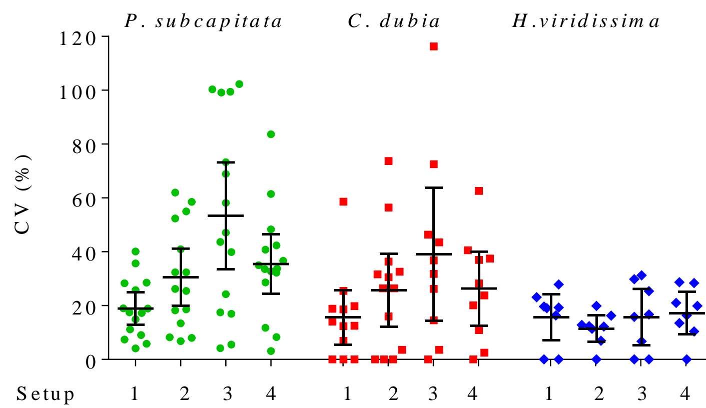

Figure 2 - 2: Coefficients of variation of algal concentrations (cells/mL) and total abundance of *Ceriodaphnia dubia* and *Hydra viridissima* at each sampling event. Black horizontal lines indicate 95% confidence intervals in setups 1 to 4.

2.4.2 Validation of the test setup

The population dynamics of experiment 1 and the repeated experiment 2 were similar (Figure 2 - 3). The algal populations peaked on day 6 and day 5 with average algal concentrations of 6.86 (± 0.64) x 10^5 and 8.83 (± 0.90) x 10^5 cells/mL in experiments 2 and 1, respectively (Figure 2 - 3 A).

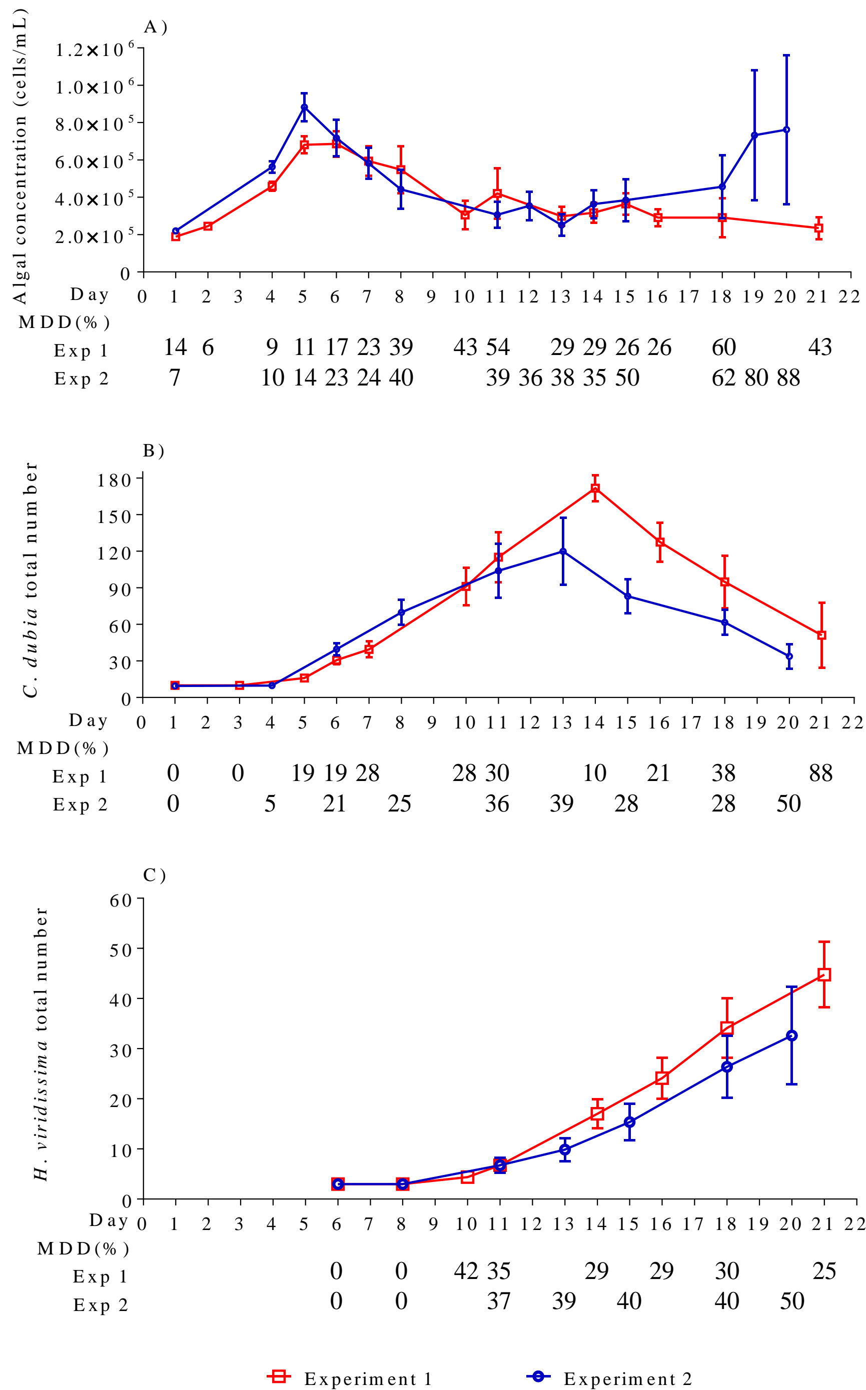

Figure 2 - 3: Abundance of (A) *Pseudokirchneriella subcapitata*, (B) *Ceriodaphnia dubia*, and (C) *Hydra viridissima* at each sampling point over 21 days. Shown are means ± 95% confidence intervals and minimum detectable differences below the x-axis of experiments 1 and 2. MDD = minimum detectable difference.

The highest *C. dubia* abundance was measured 8 days after the algal peaks in both cases. Grazer numbers declined 1 day earlier in experiment 2 and individual counts were lower due to smaller juvenile numbers of 147 (± 10) and 97 (± 24)

juveniles in experiment 1 and 2, respectively (Figure 2 - 3 B). The dynamics of *H. viridissima* populations were similar between experiments (Figure 2 - 3 C) but experiment 1 showed a slightly steeper population increase with a larger final population of 45 (± 6) and 33 (± 9) individuals in experiment 1 and 2, respectively.

Due to slightly different sampling frequencies, we computed 15, 11 and 8 CV values in experiment 1 and 14, 9 and 7 CVs in experiment 2 for algal concentrations (cells/mL), *C. dubia* and *H. viridissima*, respectively. As observed for population dynamics, replicate variance was similar between populations in both experiments (*Appendix A* Figure A - 3). The CVs of algal concentrations (cells/mL) increased by day 7, 14 and 21 to an average of 9, 26 and 26% in experiment 1 and 10, 25 and 47% in experiment 2, respectively. The CVs calculated for *C. dubia* populations increased from 9 to 15 and 33% on average in experiment 1 and from 4 to 23 and 24% on average in experiment 2. *H. viridissima* were added on day 6, so the replicate variability was 18, 18% and 17, 29% by day 14 and 21 in experiment 1 and 2, respectively.

2.4.3 Population dynamics and interactions

Significant reductions of algal concentrations (cells/mL) by 33.4% were found on the first sampling day after *C. dubia* addition, on day 4. An average algal concentration of 9.18 (± 0.48) x 10^5 cells/mL was found in Experiment 2A (only algae), while grazed algae in Experiment 2B (algae and grazers) reached an abundance of 6.11 (± 0.43) x 10^5 cells/mL by day 4 and stayed significantly lower throughout the experimental duration (Figure 2 - 4 A). Grazed algal concentrations in systems with and without predator (Experiments 2 and 2B, respectively) showed similar trajectories until day 14. After day 14, algal concentrations in Experiment 2B stayed moderately constant with an average of 2.12 (± 0.21) x 10^5 cells/mL until day 20. On the contrary, algal concentrations in Experiment 2 (grazers and predators) increased to 7.63 (± 0.37) x 10^5 cells/mL by day 20 exceeding average algal abundances of Experiment 2B by 75.0% (Figure 2 - 4 B).

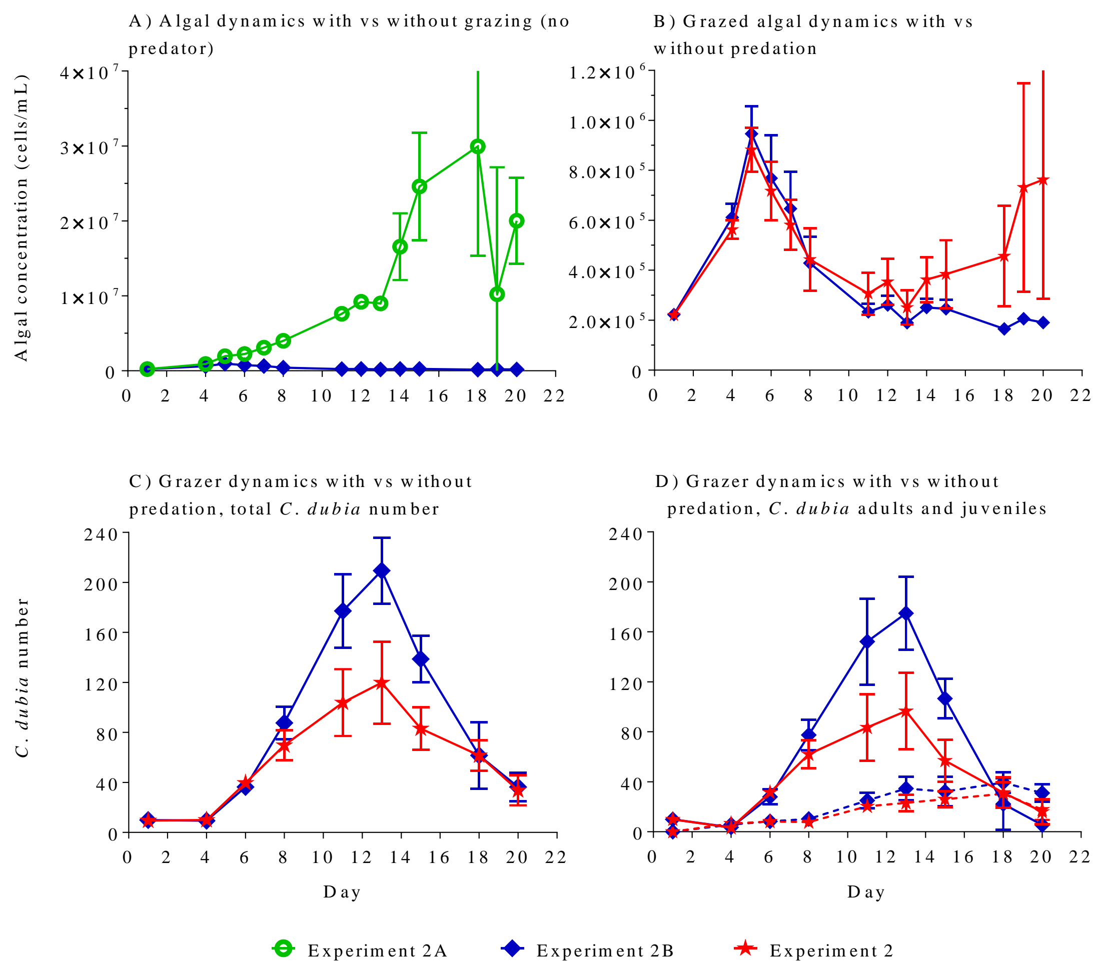

Figure 2 - 4: Algal population trajectories compared between (A) ungrazed (green circles) and grazed (without predation, blue diamonds) systems and (B) grazed systems with (red stars) and without (blue diamonds) predation. Population dynamics of *Ceriodaphnia dubia* with (red stars) and without (blue diamonds) predation as (C) total individual number and (D) juveniles (continuous line) and adults (dotted line).

Grazer population dynamics were similar in Experiment 2 and 2B until day 6 when predators were added to Experiment 2 (Figure 2 - 4 C). Population numbers peaked in both experiments on day 13 but steeper population growth curves in Experiment 2B lead to an on average 39.1% larger peaking population of 210 (± 21) individuals when compared to 128 (± 25) individuals in Experiment 2. A similarly steeper population decrease in experiment 2B resulted in similar total *C. dubia* counts of 36 (± 9) and 37 (± 10) individuals in systems without and with predators, respectively, by day 20. Population dynamics of total grazer numbers largely reflected the trajectories of *C. dubia* juveniles that rapidly increased until day 13 to

175 (± 23) and 107 (± 23) individuals constituting 88.3% and 90.1% of the total *C. dubia* population in experiments 2B and 2, respectively. By day 20, juvenile numbers dropped to 6 (± 4) and 15 (± 6) individuals while adult grazers showed a moderate but constant increase throughout the test and constituted 84.9% and 59.7% of the total *C. dubia* populations in experiments without and with predators, respectively (Figure 2 - 4 D).

2.4.4 Statistical analyses

Minimum detectable differences (MDDs) of hypothetical TriCosm exposures were calculated according to control variance and were similar between experiments 1 and 2. The TriCosm became less sensitive over time as replicate variation and MDDs increased. When variances between controls and treatments are similar, the TriCosm is estimated to be sufficiently sensitive to identify differences of 12% (± 4), 36% (± 7) and 50% (± 17) for *P. subcapitata* and 9% (± 7), 31% (± 4) and 38% (± 19) for *C. dubia* populations in week 1 (day 1 – 6), week 2 (day 7 – 13) and week 3 (day 14 – 21), respectively (averaged MDDs between Experiment 1 and 2). Averaged MDDs for *H. viridissima* were 25% (± 14) and 35% (± 7) in week 2 and 3, respectively (Figure 2 - 3). The MDDs for critical endpoints in the TriCosm can be assigned to MDD classes III (50 – 70%) and IV (< 50%), indicating the ability to determine small and medium sized effects, respectively (EFSA, 2013).

2.5 Discussion

The assessment of chemical effects with single-species tests fulfills regulatory requirements; however, primary goals of protecting populations and ecosystems might not be adequately addressed. That is because information obtained at the individual level is often not ecologically relevant since there are neither directly proportionate relationships between direct and indirect effects nor amongst responses at the individual, population and community level. An understanding of impacts on interactions in ecologically relevant test settings is thus critical and a

priority for chemical safety assessment as unexpected shifts in community profiles cannot be predicted in single-species tests (Fleeger et al., 2003; Benton et al., 2007; Liebig et al., 2008; SCHER et al., 2013).

We designed the TriCosm as a rapidly cycling, tri-trophic system with a producer-herbivore-carnivore community of small size for the purpose of quick detection of chemical impacts on species interactions. Our system is comparatively smaller (0.5 L) than many other multi-trophic systems (Metcalf et al., 1971; Daam and Van Den Brink, 2007; Foit, Kaske, Wahrendorf, et al., 2012; Dolciotti et al., 2014; Del Arco et al., 2015) and all system components exhibit rapid generation times so that treatment effects can be measured on several generations and at different life stages during short test durations (21 days compared to 80 days (Metcalf et al., 1971) and 33 days (Barry and Davies, 2004) in other tritrophic macroinvertebrate communities). Also the predator *Hydra* is a rapid reproducer with generation times of only three days under favourable conditions (Habetha et al., 2003). Chemical impacts on population dynamics can thus be detected not only at the producer-consumer level but also at a higher trophic level. The choice of a small and rapidly reproducing predator has further the advantage that it can be added at an early experimental stage (day 6) when compared to vertebrate predators that are often introduced shortly before test termination as they quickly consume remaining invertebrate prey (Metcalf et al., 1971; Harrass and Taub, 1985).

All multi-species systems have ecologically interacting components that are not independent in statistical terms as they constantly adapt to changing conditions in a dynamic environment. In fact, it has been frequently reported that even though communities are set up identically as replicates, minor variations at the beginning and/or throughout the experiments (e.g. starting conditions or uneven sample removal) can quickly lead to the development of unique properties in each replicate (Landis et al., 1997; Sanderson, 2002; Van Straalen, 2003). Indeed, different population dynamics and replicate variability were observed in four different TriCosm setups and indicated strong sensitivity to starting conditions and interaction strength. The statistical quality (in terms of interpretability, reproducibility and

replicability) of environmentally more realistic data obtained in multi-species tests is thus often reason for concern in the registration procedure of pesticides (Sanderson, 2002).

The repeatability and reproducibility of the TriCosm were thus given major consideration during test development. Initial properties and sampling techniques were adjusted and confirmed as optimized when experiments conducted at different times showed low coefficients of variation (CVs) and high reproducibility of system dynamics. Desynchronized population dynamics were observed between experiments that can be attributed to random fluctuations in test conditions (e.g. quality of the animals) and could occur even if procedures are standardized. For these reasons we assessed the repeatability by comparing CVs and not the total organism abundances. Nonetheless, a comparison of total abundances or derived variables (e.g. population growth rates) is also appropriate when chemical effects are assessed since differences between population trajectories are most likely and primarily due to chemical impacts rather than fluctuations of test conditions.

When the TriCosm is used for chemical effect assessment, two factors of major importance are i) the presence of interactions rather than the exact timing when these occur and ii) low CVs so that treatment responses can be interpreted with greater certainty and distinguished from unexplained sample variability (Sanderson, 2002). The ability to detect significant effects does depend on the magnitude of an effect but also on the ability of the test system to detect responses and that is in turn determined by the inherent variance among replicates. Test variables with coefficients of variation (CV) in the range of up to 30% have been theorized as acceptable and manageable in terms of practicality and costs (Kraufvelin, 1998). According to a review (Sanderson, 2002) that analyzed two decades of pesticide studies with micro/mesocosms, the values of CVs appear to be generally higher. The author reported an average of 45% (32% in smaller and less realistic indoor systems) with larger values in studies where animals were involved and an average use of 3.5 replicates. The average CV of 19.5% measured in the tri-

trophic system on the contrary showed smaller variance and was determined with a higher number of replicates (n = 8).

The CVs were further used for the calculation of theoretically detectable minimum differences (MDDs) between controls and treatments under the assumption of similar variances. It is to be mentioned, however, that the variance could increase, decrease and/or remain similar in treated systems (Kraufvelin, 1998; Sanderson, 2002). A modification of the number of replicates, groups or treatments, though, can decrease MDDs and allow the detection of desired effect sizes. Due to often large variability in micro-/mesocosms, EFSA may still regard endpoints with MDD classes I and II (70 – 100%) relevant but considers the exceeding of class II ideal (i.e. MDDs < 70%) (EFSA, 2013). Most projected MDDs of critical endpoints in the aquatic system correspond to effect class IV (i.e. < 50%) (with exception of algae and grazers in week 3) and confirm the ability to reveal small toxicant induced effect sizes (EFSA, 2013), distinguishing the TriCosm from other multi-trophic systems.

As expected, variations of population trajectories were observed as a result of interactions with other system components. Algal concentrations (cells/mL) and predation both directly impacted on the middle trophic layer while they indirectly impacted on the top and the bottom level, respectively. An initially small grazing pressure of juvenile *C. dubia* allowed algal populations to grow exponentially which in turn favored the development of grazer populations. As a consequence of an increasing grazing pressure by maturing and reproducing *C. dubia,* the algal concentrations dropped, yet the grazer population numbers further increased for approximately one week after food availability became limiting. The continuing population growth is attributable to a rise of juvenile numbers as adult *C. dubia* most likely matured eggs and stored energy before algal concentrations decreased. Peaking *C. dubia* populations thus coincided with lows of food availability and caused the decrease of grazer numbers.

Algae stabilized and remained at relatively constant levels as concentrations were most likely too low to be further reduced if maximum grazer filtering rates were

reached. Grazer population numbers decreased due to food shortage and independently of whether predators were present or not. While predation did not cause *C. dubia* populations to crash, it directly reduced grazer numbers, intraspecific competition among them and indirectly favored algal populations to recover. An increase of algal concentrations after grazing release was, however, not observed in Experiment 1 where grazer populations reached larger abundances but decreased later and might be due to a different quality of neonates used to start the experiments. Algal populations in Experiment 1 were thus subject to a higher and prolonged grazing pressure impeding the recovery of algal abundances within the experimental duration. An indirect effect after grazing release by *Hydra* could, however, likely be expected if the test duration was prolonged. Bottom up and top down processes are thus both likely regulating population dynamics in the TriCosm. When the system is exposed to chemicals it will thus depend on the mode of action of the toxicant impacting on one or more trophic levels leading to direct, indirect or both effects on the trajectories of interacting populations.

The focus during system development was not on achieving a steady state community and impacts on resilience (here defined as the extent of disturbance a system can tolerate before shifting to an alternative state with altered community composition and processes) cannot be assessed, neither can system shifts be detected. Nonetheless, it can indicate the recovery potential of species after stressor removal and detect small changes in interactions as the system moves through a single cycle of the middle trophic layer. Ecological impacts of toxicants rapidly propagate in an interacting system and the grazer level is directly influenced by variations in food availability and predation. Toxicant impacts on the population dynamics of this critical and key trophic layer will therefore yield important information on the ecological relevance and protectiveness of data obtained in single-species tests. Population responses to combined stressor exposures, e.g. to toxicants, predation and/or food fluctuations, could be used to facilitate both the development and the testing of mechanistic effect models. Measured community responses in terms of individual abundance changes and population trajectories

could be employed for the calibration and parameter fitting of ecological models. In turn, chemical effects on interactions within a simple freshwater community can be measured and quantified in the TriCosm and provide empirical benchmarking to estimate and test model prediction accuracy and power.

There is no doubt that the complexity of the TriCosm community is low when compared to natural systems. But besides offering higher statistical power when compared to larger and / or more complex microcosms, the impacts on system processes can be quantified as interactions change. This makes it possible to assess the effects of environmental contaminants on i) species interactions, ii) indirect effects and iii) at the population and community level. An understanding of which and to what extent processes are affected may also give insights into responses of more complex systems (Benton et al., 2007; Daam and Van Den Brink, 2007; Boonstra et al., 2011).

2.6 Conclusion

The TriCosm is a novel aquatic test system and could be a tool to address shifts in ecological interactions. It suggests that a cost-effective approach of chemical environmental safety testing with more ecological relevance whilst being statistically powerful is feasible. It can provide important insights into chemical safety in multi-trophic systems and facilitate the development and testing of mechanistic effect models for environmental risk assessment. Even so, a careful examination of the replicability of the TriCosm both within and between laboratories with and without chemical exposure is needed.

2.7 Acknowledgement

The present study was financed by the Environment Department, University of York (York, UK), and the Centre for Chemical Safety and Stewardship, Fera Science Ltd.

(York, UK). The authors thank J. Turton and T. McGowan (Centre for Chemical Safety and Stewardship) for technical support and 2 anonymous reviewers for their comments that helped to improve the manuscript.

Chapter Three: Species interactions and indirect effects in a standardized tri-trophic laboratory microcosm exposed to a pesticide

Verena Riedl,[a,b,*] Annika Agatz,[a,c] Rachel Benstead,[b] and Roman Ashauer[a]

[a]*Environment Department, University of York, Heslington, York, United Kingdom*

[b]*Fera Science Ltd., Centre for Chemical Safety and Stewardship, Sand Hutton, York, United Kingdom*

[c]*Current address: Ibacon GmbH, Institute for Biological Analytics and Consulting, Roßdorf, Germany.*

* Address correspondence to verena.riedl@york.ac.uk

3.1 Abstract

Microcosms are useful tools to better understand the indirect effects associated with the exposure to chemicals. Impacts on ecologically relevant processes and at the population level can be assessed. Yet, compared to standardized and repeatable single-species tests, their complexity often leads to increased variability within and between experiments and limits the extrapolation of cause-effect relationships. Here, we used the tri-trophic laboratory scale microcosm TriCosm (*Pseudokirchneriella subcapitata, Ceriodaphnia dubia, Hydra viridissima*) developed with the intention of limiting these variations, to assess indirect effects of the herbicide linuron. We found variability of system dynamics between six experiments in controls and treatments exposed to regulatory acceptable (0.5 µg/L) and environmentally relevant concentrations (1.2 µg/L, 10 µg/L). Bottom-up effects caused early depression of grazer and predator populations but observed effects were likely caused by experimental factors other than the herbicide. For this reason, no conclusions on indirect pesticide induced effects could be drawn. Our results demonstrate that full reporting of experimental data is important to assess the acceptability of results, limit uncertainty and make data valuable to others. First results suggest the worth of the TriCosm for the detection of indirect chemical effects

but further research is required to refine the test conditions under which the TriCosm population dynamics are repeatable.

3.2 Introduction

Pesticides have been used at a global scale over the past decades and their occurrence has been reported in surface waters (Stehle and Schulz, 2015; Spycher et al., 2018). Yet, information on effects on non-target organisms at the population and community levels which are aimed to be protected is often scarce (Van Straalen, 2003; Relyea and Hoverman, 2006). The environmental risk assessment (ERA) of pesticides in Europe consists of a tiered approach with lower tier tests (single-species tests focus on individual stressor effects at the organism level) and higher tier tests (multispecies systems, i.e. micro- and mesocosms). The latter are conducted once effect results from lower tier experiments compared to predicted environmental concentrations give reason for concern about the safety of a compound (EFSA, 2013). Results by Stehle and Schulz (2015) who conducted a meta-analysis of 838 peer-reviewed studies are, however, concerning because a 30 % reduction of regional aquatic biodiversity at regulatory acceptable pesticide concentrations (RACs) was reported. Similarly, findings by Peters et al. (2013) who reviewed 122 peer-reviewed studies suggest that negative impacts on ecosystem functions (leaf litter breakdown, primary production or community respiration) may be found at concentrations of up to 1000 times below the toxic units considered protective for aquatic ecosystems. The authors, hence, concluded that the safety factors applied to standard single species tests may not be sufficient to protect the functional endpoints investigated in their study. For the derivation of RACs, effect concentrations are divided by safety factors to account for the uncertainty encountered when results are extrapolated e.g. across levels of biological organization and from laboratory to field conditions. Yet, safety factors are generally not specifically calculated based on the context of exposure but standard values are applied and derived RACs may thus be under- or overprotective (Chapman et al., 1998).

Depending on the mechanism by which pesticides control the targeted pest species, effects on a multitude of non-target species are likely and a multitude of unexpected and indirect effects in interconnected and complex ecosystems may occur. However, toxicological studies are primarily conducted with individual organisms and impacts on interactive relationships among non-target organisms are assessed to a lesser extent. Single-species tests are reproducible, less expensive and quicker than higher tier tests and the interpretation of results is generally straightforward. Yet, they solely focus on the assessment of direct effects and exclude ecologically relevant processes mediating indirect effects.

Under natural conditions, the typical exposure scenario is given by a multitude of interacting stressors and chemicals constitute a substantial risk of causing ecological effects. Fleeger et al. (2003) indicated that indirect effects are likely more common and complex than direct effects. Direct impacts may affect an organism's performance but abundance changes of populations usually indirectly lead to increased or decreased interaction strengths and abundances of competing and/or predator species (Gergs et al., 2013; Moe et al., 2013; Schäfer et al., 2016; Van de Perre et al., 2018). For instance, species A could benefit from a release of competition pressure after a competing more sensitive species B experienced direct chemically induced effects. However, when species A and B are prey for species C, a reduction of species B could cause an increase of predation pressure on species A and mask positive effects from a release of competition. The effects and mechanisms acting in this simple example with three species on two trophic levels might appear straightforward when the mode of action of the chemical is known but a quantitative prediction is still difficult (Rohr et al., 2006). In a more complex scenario such as typically found in the environment, chemical effects may lead to altered sensitivity and vulnerability of species to other stressors (Stampfli et al., 2011; Foit, Kaske and Liess, 2012) and effects may proliferate across three or more trophic levels affecting the composition and structure of communities (Beketov et al., 2013). Chemical risks in interconnected ecosystems can thus hardly be evaluated if the assessment of risk is solely based on single-species tests. The understanding and prediction of stress

responses resulting from a combination of chemical and natural pressures and their ecological impacts on aquatic ecosystems is thus important for the assessment of chemical safety (Van Straalen, 2003; Relyea and Hoverman, 2006; Moe et al., 2013; Gergs et al., 2013; Schäfer et al., 2016).

Microcosms have a long history in ecology providing insights to ecological processes that are applicable at larger environmental scales (Huffaker, 1958; Taub, 1997b; Benton et al., 2007). In the context of ERA, micro- and mesocosms are used to evaluate how chemicals affect environmentally more realistic systems and provide information beyond that derived with single-species tests (Taub, 1997b; Landis et al., 1997; Benton et al., 2007). In multi-species systems, all components are interlocked by reciprocal cause-effect pathways and species constantly adapt and react to changing conditions. Hence, microcosms typically become more variable with increasing complexity because minor variations of any system component will affect population trajectories. The variability of systems with higher levels of complexity is thus often high between replicates and between studies setting limitations to the derivation of concentration-effect relationships (Peters et al., 2013; Poisot et al., 2015).

In the EU, 492 active substances are currently approved for use and 37 more substances are pending registration (European Union, 2016). Because pesticides contain at least one active substance, this leads to potentially thousands of different formulations. Thus, there is a great need to develop tools to assess and quantify their effects on system processes and to extrapolate concentration-effect relationships to a bigger scale (Rohr et al., 2006; Benton et al., 2007). The tri-trophic, laboratory scale microcosm TriCosm (*Pseudokirchneriella subcapitata, Ceriodaphnia dubia, Hydra viridissima*; Riedl et al., 2018) was standardized to minimise variability and allow the repeatable assessment of small pesticide effects on ecological processes. Trophic relationships and intraspecific competition make the TriCosm more complex than single-species tests, yet, it is much less complex than mesocosms used at higher tiers of regulatory ERA.

Here, we used the TriCosm to assess and quantify direct and indirect effects of the herbicide linuron on species interactions. Linuron was applied at the regulatory acceptable concentration (RAC, 0.5 µg/L), the concentration causing an effect to 10% of the most sensitive algal population (EC_{10} of *Scenedesmus acutus,* 1.2 µg/L) and the highest concentration at which no effect was observed for the algal species used in this study (NOEC of *Pseudokirchneriella subcapitata*, 10 µg/L; Crane et al., 2007). Our aim was to target the base trophic layer (the algae) and assess abundance changes of all trophic layers (algae, daphnids, *Hydra*) and evaluate the environmental safety of the herbicide at concentrations considered safe based on single-species tests.

Although the repeatability of TriCosm dynamics was previously demonstrated in the short-term (Riedl et al., 2018; see Chapter Two), different population dynamics were found in control systems in the long-term. For this reason, we slightly varied the treatment conditions between experiments (six linuron exposures are reported in chronological order) with the purpose of improving test medium conditions to obtain similar population dynamics in controls as shown by Riedl et al., (2018).

3.3 Materials and Methods

3.3.1 Test organisms

The test organisms *Pseudokirchneriella subcapitata* (green alga), *Ceriodaphnia dubia* (cladoceran grazer) and *Hydra viridissima* (cnidarian predator) are interconnected through consumer-resource relationships, display rapid life cycles (Riedl et al., 2018) and an overall high sensitivity to chemicals (Rose et al., 2002a; Quinn et al., 2008; Aruoja, 2011; Quinn et al., 2012). Stock cultures of *P. subcapitata* were obtained from the Culture Collection of Algae and Protozoa (Scotland, UK) and cultured in OECD medium (OECD, 2011). *C. dubia* were obtained from the College of Life and Environmental Sciences at the University of Birmingham (Birmingham, UK) and age specific cultures (approx. 50 organisms per 450 mL) were maintained in moderately hard, synthetic freshwater (USEPA, 2002). They were fed three times per week with approx. 3.5 mL of a suspension of yeast, cerophyl® and trout-chow and approx. 11 x

10^7 cells *P. subcapitata* (USEPA, 2002). *H. viridissima* were obtained from the Department of Evolutionary Zoology at the University of Debrecen (Debrecen, Hungary) and cultured in T82MV medium (ASTM E1366-11, 2011). They were fed with newly hatched *Artemia salina* (Ocean Nutrition™ Brine Shrimp Eggs; Ocean Nutrition Europe, Belgium) three times per week ad libitum and all animal cultures were kept at 25 (± 1) °C, 12:12 h light:dark, ca. 13 µmol m^{-2} s^{-1}.

3.3.2 TriCosm setup, sampling and environmental parameters

The test setup, test conditions, sequential assembly of species and sampling techniques are described in detail by Riedl et al. (2018). Eight replicates were prepared for each treatment in six experiments, with exception of negative controls in experiment 4 where 3 replicates were used due to space constraints. Tests were performed in a controlled environment (25 (± 1) °C, 12:12 h light:dark, orbital shaking at 65 rpm, ca. 16 µmol m^{-2} s^{-1} at the water surface), test vessels (500 mL Pyrex® crystallizing dishes; Sigma-Aldrich, UK) were baked (121 °C) or acid washed (10 % HCl) and covered with watch glasses (diameter 125 mm; Sigma-Aldrich, UK). *P. subcapitata* (2 x 10^4 cells/mL) were added on day 0. *C. dubia* (10 neonates aged ≤ 24 h) and *H. viridissima* (3 animals aged ≤ 2 d without visible buds) were added to each replicate on days 1 and 6, respectively. The test medium T82MV (ASTM E1366-11, 2011) was prepared with reduced concentrations of the buffering salt sodium metasilicate (0.02 g/L Na_2SiO_3) in experiments 1, 2, 3 and 4, wereas Na_2SiO_3 concentrations were increased (to 0.23 and 0.11 g/L, respectively) in experiments 5 and 6 to improve medium stability and prevent excessive pH fluctuations. Experiments were terminated once *C. dubia* populations were extinct or when grazers did not reproduce (for details see Table 3 - 1).

T82MV was aerated for ca. 24 h (10 mL Fisherbrand™ glass pipette (Fisher Scientific, UK) attached to a Whisper® Aquarium Air Pump (TetraFish, UK) and adjusted to pH 7.0 (± 0.1) with HCl (5 %) or NaOH before the test, either before or after a 24-h aeration period. Water parameters (pH and dissolved O_2) were analysed two times per week with an Orion Star™ multiparameter meter (ThermoFisher

Scientific, UK) whereas animal and algal populations were monitored three and five times per week, respectively. The water column was not mixed before algal sampling performed at the centre of the vessel at variable depths beneath the water surface. Algal cell density was determined via in-vivo fluorescence activity measurements (200 µl subsample, 5 replicates, Tecan® Infinite 200 PRO). *C. dubia* and *H. viridissima* were counted manually until count differences did not exceed 20 % of the lower value. Stress indicating morphological changes such as the occurrence of ephippia were documented in *C. dubia.* For *Hydra*, we recorded the overall morphology score according to Wilby (1988) which is measured based on chemically induced morphological changes of the animal's tentacles and trunks and indicates the degree of observable stress damage (Wilby, 1988; Quinn et al., 2012; Tökölyi et al., 2014). The concentration of ^{14}C labelled linuron was determined three times per week (1 mL samples in 4 mL Ultima Gold (PerkinElmer, UK) via liquid scintillation counting (Tri-Carb® 4810TR Liquid Scintillation analyser, PerkinElmer, UK; see *Appendix B* Table B - 1 for details on count settings).

3.3.3 Test substance, dosing and chemical analysis

Linuron is a systemic herbicide that disrupts photosynthesis via an inhibition of the photosystem II and is reported to occur in the range of ng/L to µg/L in surface waters (Webster et al., 2015). Based on its aqueous degradation times (degradation of 50% of the compound in 1460 days via hydrolysis and > 30 days via photolysis at 20 °C and pH 7.0; Lewis et al., 2016) the herbicide concentration was expected to remain constant throughout the test duration. A total of six linuron exposures were set up with negative (without solvent) and/or solvent (acetonitrile) controls and treatment T RAC (regulatory acceptable concentration, 0.5 µg/L). In addition, treatment T 1.2 (1.2 µg/L) was prepared in experiments 1, 2, 3 and 4, whereas treatment T 10 (10 µg/L) was prepared in experiments 5 and 6. The total linuron concentration was reached by addition of ^{14}C linuron in T RAC or a combination of ^{14}C linuron and unlabelled compound in T 1.2 and T 10 to limit the volume of solvent acetonitrile added with ^{14}C linuron. Depending on the concentration of ^{14}C linuron, the added volume of solvent varied among experiments (Table 3 - 1) but remained below the

OECD recommended threshold level for *Daphnia magna* (0.1 mL/L OECD, 2012) in all experiments. Even so, an experiment with *C. dubia* was conducted to verify that solvent used at these concentrations did not affect endpoints measured in our experiments (for details see *Appendix B*).

Linuron stock solutions were measured via liquid scintillation counting (^{14}C radiolabelled compound, Tri-Carb® 4810TR Liquid Scintillation analyser, PerkinElmer, UK; *Appendix B* Table B - 1) and HPLC analysis (cold compound, Agilent 1100 Series, Agilent Technologies, US; *Appendix B* Table B - 2). The efficiency to detect ^{14}C compounds via liquid scintillation counting was calculated (*Appendix B*) for each replicate per sampling day in linuron treatments and variations between experiments 1 to 6.

3.3.4 Statistical analysis

GraphPad Prism (Version 6.07 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com) was used to compare *C. dubia* neonate numbers and algal densities between treatments and the control. A d'Agostino and Pearson omnibus test for normality was performed followed by a Kruskal-Wallis test and Dunn's multiple comparisons test with 0.05 significance level or by ANOVAs (one-way and two-way for repeated measurements) followed by a Dunnet test with 0.05 significance level. Graphs show organism numbers as means ± 95 % confidence intervals. Control population dynamics in two TriCosm experiments during system standardization (Riedl et al., 2018) are indicated in graphs with stars to allow comparisons with controls from earlier experiments. *C. dubia* and *H. viridissima* count averages and confidence intervals were rounded to the nearest integer.

3.4 Results

3.4.1 Linuron analysis and solvent control effects

The detection efficiency of ^{14}C linuron was 93.5 % (± 0.02 %; n = 304) and measured ^{14}C linuron concentrations deviated on average by 0.03 (± 0.01) µg/L from nominal ^{14}C linuron concentrations (Table 3 - 1).

Table 3 - 1: Experimental dates, treatments, duration, solvent acetonitrile and Na_2SiO_3 (for medium preparation) concentrations in experiments 1 - 6.

Exp	Date	Treat-ment[1]	Duration (days)	Na_2SiO_3 (g/L)	Solvent (mL/L)	linuron concentration (µg/L) Total	^{14}C nominal	^{14}C measured
1	06.10-19.10.2016	NC	13	0.02	0	0	0	0
		T RAC			0.02	0.5	0.5	0.51 (0.45-0.56)
		T 1.2			0.04	1.2	1.2	1.25 (1.16-1.39)
2	17.11.-30.11.2016	SC	13	0.02	0.04	0	0	0
		T RAC			0.04	0.5	0.5	0.54 (0.47-0.58)
		T 1.2			0.04	1.2	0.5	0.55 (0.50-0.59)
3	02.02-06.02.2017	SC	4	0.02	0.04	0	0	0
		T RAC			0.04	0.5	0.5	0.48 (0.46-0.50)
		T 1.2			0.04	1.2	1.2	1.21 (1.13-1.25)
4	13.02-18.02.2017	SC	5	0.02	0.02	0	0	0
		T RAC			0.02	0.5	0.5	0.49 (0.45-0.52)
		T 1.2			0.02	1.2	1.2	1.19 (1.14-1.27)
5	07.09-12.09.2017	SC	5	0.23	0.07	0	0	0
		T RAC			0.07	0.5	0.5	0.47 (0.44-0.50)
		T 10			0.07	10	2.0	1.90 (1.82-1.97)
6	28.09-11.10.2017	NC	12	0.11	0	0	0	0
		SC			0.02	0	0	0
		T RAC			0.02	0.5	0.5	0.49 (0.46-0.53)
		T 10			0.02	10	0.5	0.50 (0.46-0.55)

[1] NC: negative control (without solvent); SC: solvent control; T RAC: regulatory acceptable concentration 0.5 µg/L linuron; T 1.2: 1.2 µg/L linuron; T 10: 10 µg/L linuron.

In solvent tests, *C. dubia* neonate numbers did not significantly differ between controls and treatments but among treatments (Kruskal-Wallis: $H = 7.8$, $p < 0.05$). Dissolved oxygen and pH generally increased with time and pH was higher in experiments 5 and 6 where higher buffering salt (Na_2SiO_3) concentrations were used (*Appendix B* Figure B - 2).

3.4.2 Experiments 1 and 2

Experiments 1 and 2 (0.5 and 1.2 µg/L linuron, 0.02 g/L Na_2SiO_3) were terminated after 13 days because the population trajectories in the control deviated considerably to those observed in the standardized TriCosm (Figure 3 - 1).

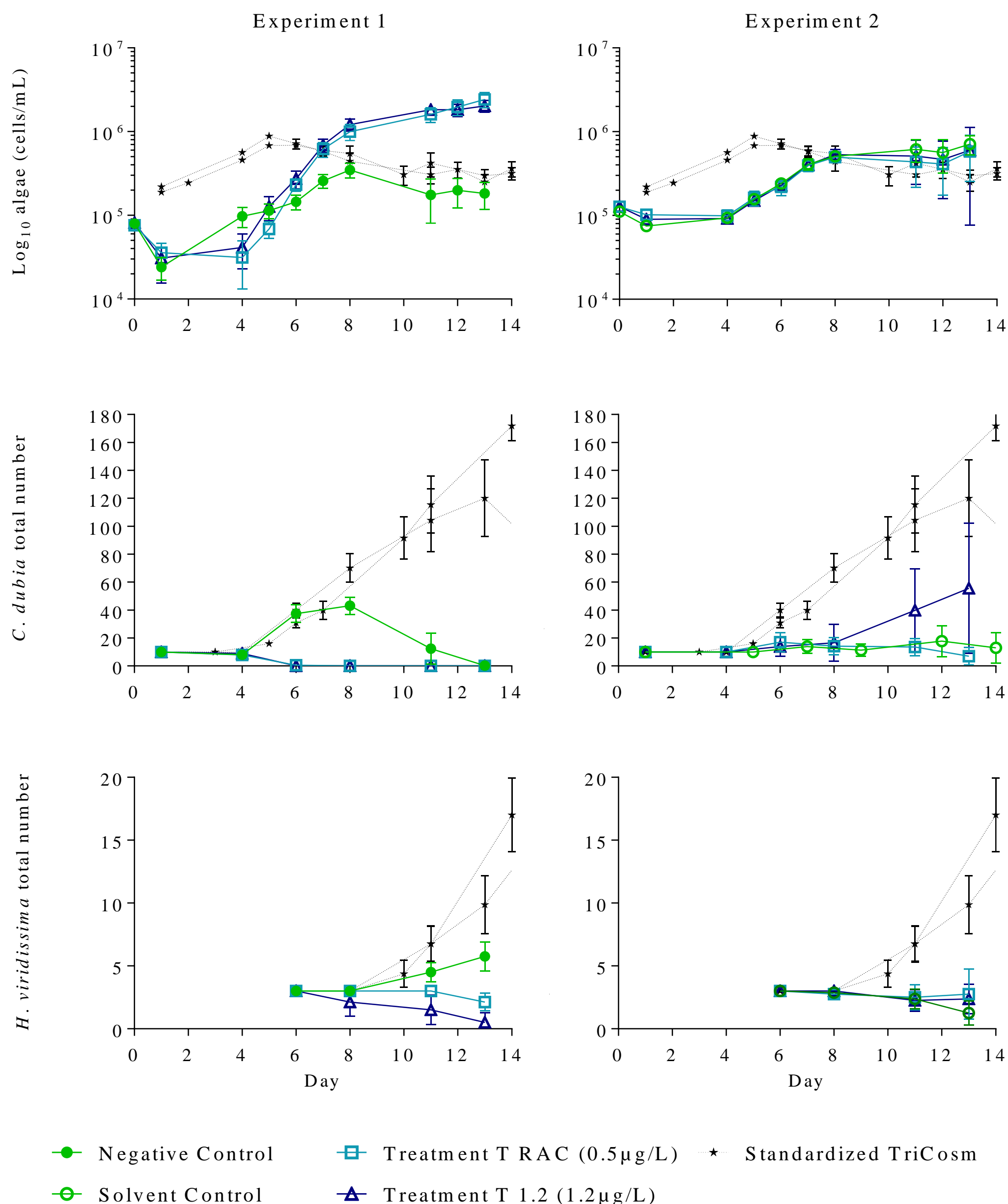

Figure 3 - 1: Population dynamics of *P. subcapitata* (log_{10} cells/mL), *C. dubia* and *H. viridissima* numbers (top to bottom row, respectively) in experiments 1 and 2. Population dynamics measured in earlier TriCosm experiments (Riedl et al., 2018) are indicated for reference with stars. Shown are mean values ± 95 % confidence intervals.

In experiment 1, algal concentrations decreased within 24 h by an average of 61.7 % (range: 53.0 - 69.8 %) to 3.02 x 10^4 cells/mL in the control and treatments (Figure 3 - 1). On day 4, similarly low concentrations were observed in T RAC and T 1.2 whereas significantly higher algal concentrations (9.8 x 10^4 cells/mL; one-way ANOVA: $F (2, 21) = 16.04$, $p < 0.01$) were measured in controls. After day 4, exponential growth was measured in the control and treatments with steeper trajectories in T RAC and T 1.2. At the end of the test (day 13), control algal populations (1.8 (± 0.5) x 10^5 cells/mL) were approx. 12 times smaller than in T RAC and T 1.2 (24.2 (± 3.8) x 10^5 and 20.3 (± 2.5) x 10^5 cells/mL, respectively). *C. dubia* populations in controls reached peak numbers on day 8 (17 (± 3) adults and 26 (± 5) juveniles; *Appendix B* Figure B - 1) but decreased thereafter to a total of 12 (± 10) individuals on day 12. Contrary, no reproduction was observed in linuron treatments and *C. dubia* populations were extinct on day 6. No stress-induced reproductive structures were observed in *C. dubia* populations. *Hydra* numbers doubled within one week (day 13) to 6 (± 1) individuals in controls whereas populations in T RAC and T 1.2 decreased to 2 (± 1) and 1 (± 1) individuals, respectively. *Hydra* were attributed an average morphology score of 7 in T RAC and T 1.2 indicating reversibly shortened tentacles or bodies whereas control organisms showed normal morphological conditions at the end of the experiment. Morphology scores rank between 0 and 10 while scores ≥ 6 are considered sub-lethal and reversible whereas scores ≤ 5 are considered lethal. Information on the morphology scores according to Quinn et al. (2012) are provided in *Appendix B* Table B - 3

In experiment 2, algal populations decreased by an average of 27.2 % within 24 h in the control and treatments (range: 19.6 - 32.6 %) and an exponential increase was measured after day 4 (Figure 3 - 1). Controls were not significantly different to treatments throughout the test duration (two-way ANOVA for repeated measurements: $F (2, 21) = 0.40$, $P = 0.68$) with comparable final concentrations of 7.1 (± 1.4) x 10^5, 5.8 (± 2.5) x 10^5 and 6.0 (± 4.1) x 10^5 cells/mL on day 13 in controls and treatments, respectively. *C. dubia* did not produce ephippia but reproduction was small in the control and T RAC (*Appendix B* Figure B - 1). Individual counts remained

approximately constant throughout the experiment (13 (± 9) and 7 (± 5) individuals, respectively, on day 13). In T 1.2, *C. dubia* populations increased to 56 (± 36) individuals by day 13. *Hydra* populations in controls decreased within one week after addition to 1 (± 1) individuals whereas animal counts in T RAC and T 1.2 (3 (± 2) and 2 (± 1) individuals on day 13, respectively) remained similar to the starting conditions. At the end of the experiment, *Hydra* were attributed an average morphology score of 7 in T RAC and T 1.2 indicating reversibly shortened tentacles or bodies. Control organisms were attributed an average morphology score of 5 indicating totally contracted bodies and damaged but visible tentacles (*Appendix B* Table B - 3).

3.4.3 Experiments 3 and 4

Experiments 3 and 4 (0.5 and 1.2 µg/L linuron, 0.02 g/L Na_2SiO_3) were terminated on days 4 and 5, respectively, because grazer populations were in poor condition (weak swimming activity of few surviving individuals, Experiment 3) or extinct by day 4 (Experiment 4; Figure 3 - 2) and *Hydra* were not added to the systems. Algal populations decreased within 24 h after test initiation by an average of 17.2 % (range: 13.3 - 20.1 %) and 75.2 % (range: 74.1 - 76.7 %) in controls and treatments, respectively.

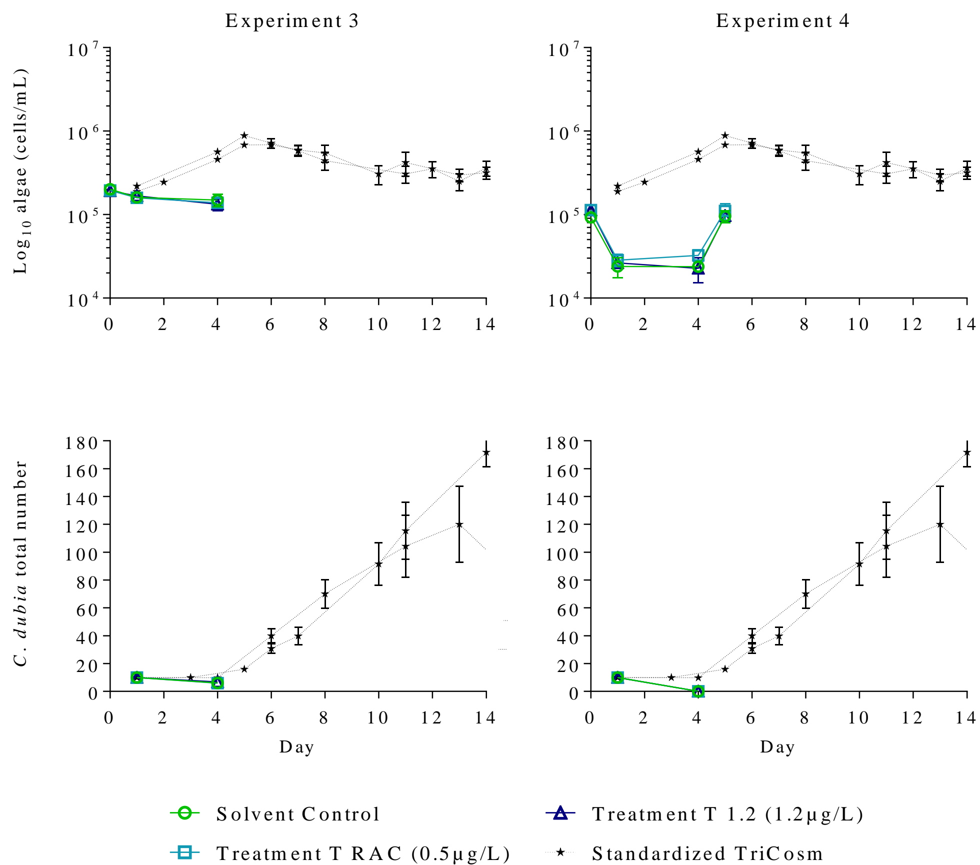

Figure 3 - 2: Population dynamics of *P. subcapitata* (log_{10} cells/mL) and *C. dubia* in experiments 3 and 4. Population dynamics measured in earlier TriCosm experiments (Riedl et al., 2018) are indicated for reference with stars. Shown are mean values ± 95 % confidence intervals.

3.4.4 Experiments 5 and 6

Experiments 5 and 6 (0.5 and 10 µg/L linuron, 0.23 and 0.11 g/L Na_2SiO_3, respectively) were terminated on days 5 and 12, respectively. Algal populations in experiment 5 increased throughout the test duration and more than quintuplicated by day 5 but grazer populations were extinct by day 4 and *Hydra* were not added before the test was terminated (Figure 3 - 3).

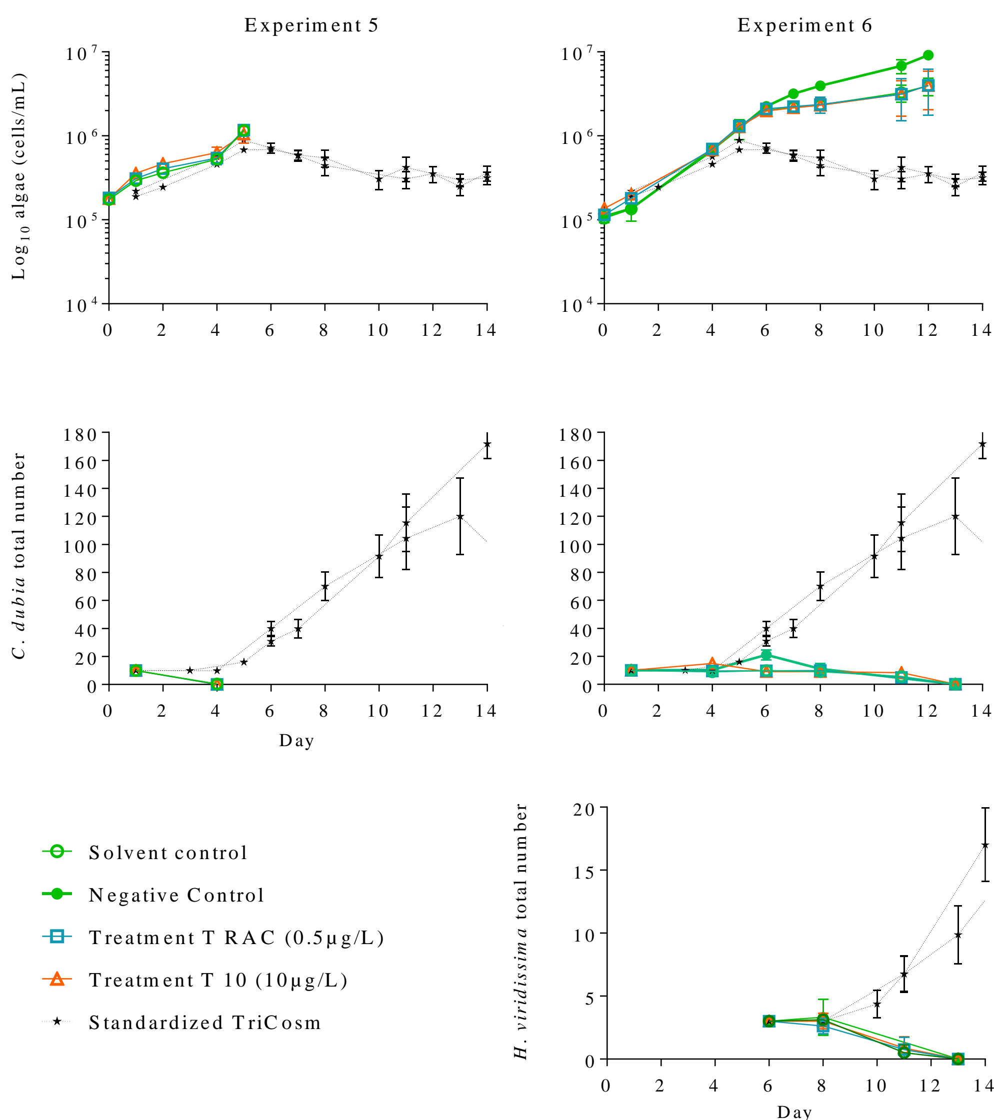

Figure 3 - 3: Population dynamics of *P. subcapitata* (log_{10} cells/mL), *C. dubia* and *H. viridissima* numbers (top to bottom row, respectively) in experiments 5 and 6. . Population dynamics measured in earlier TriCosm experiments (Riedl et al., 2018) are indicated for reference with stars. Shown are mean values ± 95% confidence intervals.

In experiment 6, algal populations increased by an average of 39.2% within 24 h in all treatments (range: 23.1 - 57.6%; Figure 3 - 1). Algal population trajectories were similar between solvent controls, T RAC and T 10 throughout the study duration with comparable final concentrations (39.4 (± 8.7) x 10^5, 39.9 (± 20.6) x 10^5 and 39.8 (± 17.9) x 10^5 cells/mL, respectively) whereas, in the negative controls, algal populations were more than twice as large (91.2 (± 5.1) x 10^5 cells/mL) at the end of the experiment. *C. dubia* population dynamics were similar among treatments and

decreased throughout the study to 6 (± 1), 5 (± 2) and 8 (± 2) individuals in solvent controls, T RAC and T 10, respectively. Grazer development and egg maturation was observed until day 4 in controls and treatments but the condition of grazers in solvent controls, T RAC and T 10 quickly deteriorated and eggs were aborted to empty brood pouches on day 6. Grazer populations in negative controls were in good condition throughout the test duration and reproduced 11 (± 4) neonates on day 6 (*Appendix B* Figure B - 1). Adult *C. dubia* did not show signs of stress by producing ephippia but the total population decreased to 4 (± 1) individuals on day 11. *Hydra* populations decreased within 5 days (day 11) to 0 (± 1) individuals in negative controls and 1 (± 1) individuals in solvent controls, T RAC and T 10. Animals were attributed low average morphology scores of 5 in all treatments, shown as contracted bodies and tentacle damage that impeded feeding (*Appendix B* Table B - 3).

3.4.5 Control population dynamics in exposures and the TriCosm

Algal and animal population dynamics in experiments 1 to 6 differed from TriCosm population trajectories measured in earlier experiments (Figure 3 - 1, Figure 3 - 2, Figure 3 - 3). Algal population dynamics in experiments 5 and 6 were most similar to trajectories measured in the standardized TriCosm until day 5. Contrary to standardized TriCosm experiments where algal populations decreased after peak concentrations (6.8 (± 0.4) x 10^5 and 8.8 (± 0.7) x 10^5 cells/mL) on day 5, algal populations in experiment 6 were by this time twice as abundant (12.3 (± 1.5) x 10^5 cells/mL) and growth was continuous throughout the experiment. The largest grazer population was counted in controls in experiment 1 but these abundances were still smaller than observed in the standardized TriCosm tests until day 9. The population peak (43 (± 6) individuals on day 8) was approximately 3 times smaller and occurred 4.5 days earlier than in the standardized TriCosm tests (172 (± 10) and 120 (± 26) individuals on days 13 and 14, respectively). *Hydra* were added in only three of six linuron experiments (experiments 1, 2 and 6) and populations increased only in controls of experiment 1. Final populations (6 (± 1) individuals on day 13) were

approximately 2 times smaller than in the standardized TriCosm tests (10 (± 2) and 17 (± 3) individuals).

3.5 Discussion

In linuron experiments 1 and 2, similarly small algal populations were measured throughout the first few days after test initiation. Nonetheless, approx. three times smaller algal populations in treatments compared to negative controls in experiment 2 on day 4 might indicate negative effects of the herbicide. This might have led to an indirect effect of linuron on *C. dubia* populations through a direct reduction of food availability. Grazer populations with higher food levels increased in controls and, in turn, lead to higher food availability for *Hydra. Hydra* populations in controls thus increased whereas populations in linuron treatments decreased as an indirect effect of linuron and mediated by *C. dubia*. Even though *C. dubia* populations increased in the controls, populations were considerably smaller than in the standardized TriCosm most likely because higher algal concentrations with higher nutritional value during early developmental stages supported earlier and healthier grazer development and maturation in the standardized TriCosm (Rose et al., 2002b).

In experiments 3 and 4 (0.5 and 1.2 µg/L linuron), similar algal concentrations in treatments and the control indicated no direct linuron effects on *P. subcapitata*. The concentrations used in both exposures were substantially below the reported EC_{50} values of 16 µg/L (72 h exposure; Lewis et al., 2016) and 43 µg/L - 67 µg/L (120 h exposure; Crane et al., 2007). Similarly, direct impact of linuron on *C. dubia* neonates were unlikely because the reported chronic no-observed-effect-concentration is considerably higher than linuron concentrations used in our experiments (180 µg/L for *Daphnia magna*, 21d study by Lewis et al., 2016). Low food availability likely affected grazers since individuals display high metabolic rates in early life stages and *C. dubia* growth, fecundity and survival are greatly affected by food levels (Rose et al., 2002b). Unless the filtering rates are increased, the ingestion rates decrease with decreasing food density and the energetic gain is too low to meet

higher metabolic rates caused by increased filtering rates with possibly negative effects on survival (Rodgher et al., 2008). Rodgher and Luiz Gaeta Espíndola (2008) fed *C. dubia* neonates with 1 x 10^4 cells/mL *P. subcapitata* each day for one week (compared to 16 x 10^4 and 2.6 x 10^4 cells/mL on average in experiments 3 and 4, respectively) and reported 100 % survival (n = 10) and three neonates on average per female. Similarly, Gama-Flores et al. (2011) observed growth and reproduction when *C. dubia* neonates were fed with 1 x 10^4 cells/mL *Chlorella vulgaris*. It is thus likely that negative effects observed in our study resulted from a dietary quality effect because the quality of food in terms of its nutritional value is at least as important for the development, growth and survival of Cladocerans as the quantity of food they receive (Locke and Sprules, 2000; Rodgher and Luiz Gaeta Espíndola, 2008).

Low algal concentrations and thus small *C. dubia* populations were unlikely caused by the herbicide but possibly by acidic medium conditions or pH fluctuations in the medium. In experiments 3 and 4, the pH of the test medium was adjusted to pH 7.0 (± 0.1) with HCl before medium aeration on day -1. However, values below pH 7.0 (pH 6.4 and pH 6.5, respectively) were measured on day 0 and pH had to be re-adjusted to pH 7.0 (± 0.1) with NaOH before test start. As will be described further in detail in *Chapter Four*, an aeration period of 24 h might be too short to allow for chemical equilibrium in the test medium which is made from seven salt solutions, one metal and one vitamin solution in deionised water. Opposing pH adjustments before and after medium aeration to reach pH 7.0 might, hence, have further destabilized medium pH and negatively affected *P. subcapitata* populations for which a range of pH 7.5 to pH 8.1 is recommended in standard algal media (OECD, 2011). Fluctuating pH values of the test medium could have also directly affected the condition of *C. dubia*. Locke and Sprules (2000), for instance, found negative effects of low levels of pH on *Daphnia pulex* egg production and the lipid-ovary index when individuals were kept in acidic water or fed with acidic phytoplankton. Belanger and Cherry (1990) reported 50 % mortality after 48 h when *C. dubia* neonates from mothers cultured at pH 8.0 were exposed to medium with pH 4.4 - 4.7. Here, *C. dubia* mothers were reared in moderately hard synthetic water (USEPA, 2002) at pH 7.8 (±

0.2) and transferred to T82MV (test medium) one day before test begin for the release of neonates. Neonates might thus have experienced pH fluctuations during the change of medium and increased their sensitivity to low food levels.

In experiments 1 - 4, the test medium was prepared with low buffering salt (Na_2SiO_3) concentrations whereas higher concentrations were used in experiments 5 and 6 (0.5 and 10 µg/L linuron) to improve the buffering capacity of the medium. As observed in experiments 3 and 4, also in experiments 5 and 6, similar algal concentrations in control and treated replicates indicated no direct linuron effects on the base trophic layer. However, likely as a result of increased medium buffering, the algal population dynamics closely resembled those measured in the standardized TriCosm. The food availability for grazers was high and the volume of solvent used (0.07 mL/L) did not cause effects in terms of survival; nonetheless, grazer populations remained small. Other than in experiments 1 – 4, where negative impacts of sodium metasilicate on animal species were excluded, higher buffering salt concentrations likely negatively impacted animal species. In experiments 5 and 6, the reported concentration at which an effect was observed on 50% of *C. dubia* populations (immobilization test with sodium silicate as SiO_2; Warne & Schifko, 1999) was exceeded by approx. ten and five times, respectively (*Appendix* B Table B - 4). Decreased mobility (i.e. swimming activity) of filter-feeding *C. dubia* likely caused too low food intake to cover their metabolic demand and lead to their death in experiment 5 whereas negative impacts on reproduction were found in experiment 6. We observed reproduction only in the negative control, which might indicate combined effects of the solvent and the buffering salt on grazers in solvent controls and the treatments. Further, increased buffering salt concentrations likely caused *Hydra* body contraction and tentacle damage with negative effects on feeding and survival. Similar findings were reported by Šimičev et al., (2016) who found negative effects on the morphology, behaviour and reproduction of *H. viridissima* 72 h after an exposure to sodium metasilicate in the range of 0.05 - 0.39 g/L.

3.5.1 Indirect effects and analysis of multi-species data

Bottom up effects were observed in all experiments between the base and the top trophic layers i.e. chemical effects on algal populations indirectly regulated *Hydra* population dynamics through the mediating intermediate grazer level. Top down effects where *Hydra* populations indirectly influence algal populations were not observed because a general lack of *C. dubia* reproduction did not support large enough *Hydra* populations to affect lower trophic layers through predation effects. Possibly effects of linuron on species interactions could only be distinguished between treatments in experiment 1.

The distinction of grazers in age classes showed that in experiments 2 and 6 solely the initially added 10 neonates matured to adults whereas the adult group in experiment 1 was larger than 10 individuals. Hence, neonates born throughout the experiment reached adult age and contributed to the total grazer number via reproduction. The reporting of the age structure of populations can, for instance, give further indication on the impact of grazing on algal populations because the filtering rates differ between life stages (Rose et al., 2002a). Agatz et al. (2012), instead, tested mixture toxicity of two dispersing agents and reported positive effects at the individual level due to accelerated reproduction but negative effects at the population level due to size selective mortality. Size selective effects might, hence, affect similarly abundant populations in different ways depending on their population age/size structure. Morphological and/or behavioural (e.g. swimming speed) changes of individuals can, for instance, give information on causal relationships and help understanding observed population patterns. For instance, the morphology of *Hydra* (i.e. tentacle damage) following an exposure to increased sodium metasilicate concentrations points to the reason for decreasing population numbers (i.e. feeding impediment and starvation).

Other than taking into account chemical effects on morphological and behavioural aspects of test organisms, the replication and non-selective reporting of outcomes are important to distinguish chemical effects from random noise and

impacts caused by other factors. Considerable variation among experiments might result from factors other than the test substance. In our case, consideration of only experiment 1 might have led to false conclusions on the environmental safety of linuron at the used concentrations. Because we used environmentally relevant concentrations, smaller algal populations in treatments than in controls on day 4 could suggest direct linuron effects on primary producers that caused indirect effects along the food chain via a limitation of food. Van Den Brink et al. (1997) found a lowest NOEC of 0.5 µg/L based on direct effects in macrophyte-dominated freshwater microcosms after chronic (28 d) exposure. The authors reported that endpoints related to linuron's photosynthesis-inhibiting properties (photosynthesis efficiency, growth inhibition, densities of primary producers, oxygen and pH metabolism) were more sensitive than responses of invertebrates. Based on our results, we couldn't conclude whether linuron affects species interactions in the tri-trophic system and gain additional information on the protectiveness of the regulatory acceptable concentration.

Variations of population trajectories in multi-species systems after chemical exposure give information on the effects of chemicals on common patterns or classes of interactions. Yet, because of the multivariate nature of responses of ecological systems, datasets generally require different statistical methods for analysis than data in single-species tests where properties of each variable can be analysed separately. Ecological datasets frequently contain missing data and measurements are linked, correlated and not independent when system components are measured repeatedly in time and affect each other (Eisenhauer et al., 2015). Other than univariate methods such as ANOVA and its derivatives (e.g. Interval of Nonsignificant Difference; ASTM International, 2011), multivariate statistical methods are more suitable to analyse ecological data obtained in multi-species systems because they allow the detection of patterns and testing of hypotheses on the entirety of variables. Commonly used multivariate analysis techniques are, for instance, principal component, principal response curve and redundancy analysis (Landis et al., 1997). Generalized linear mixed models and structural equation models are used to obtain

causal understanding in complex ecological networks and assess effects on direct versus indirect relationships (Eisenhauer et al., 2015). The TriCosm exposure data were not analysed with multivariate analysis due to inconsistent experimental results. Essentially most of our observed differences between treatments and experiments can be explained by small, but with hindsight relevant, differences in the experimental protocol or by speculative post hoc explanations. Therefore, we think that in this case multivariate analyses cannot add useful information or interpretation. Generally, sufficient replication and low inherent variability among replicates are critical aspects of multispecies toxicity tests and increase the power of whichever analysis method is used. Low replicate variability was observed in the experiments presented here and is desirable. Other than small variations between replicates, the repeatability of control dynamics (control populations follow similar patterns when experiments are repeated in time) is an important property to increase the confidence in and certainty of experimental results and procedures.

3.5.2 TriCosm repeatability and the influence of the medium

The repeatability of ecological experiments is often difficult because the variability among test units rapidly increases with complexity. Even in experiments with apparently identical set up, the system dynamics easily diverge over time and different patterns are observed among experiments in time (Hines et al., 2014). The variations of community dynamics among TriCosm experiments were likely caused by medium related factors and our findings show the influence medium characteristics can have on organism performance and experimental outcomes. Other than linuron, the concentration of the buffering salt and the instability of the medium likely affected the population dynamics in all experiments.

It is widely reported in the literature and discussed previously that pH fluctuations can have large impacts on species performance (Belanger and Cherry, 1990; Locke and Sprules, 2000; Rendal et al., 2012). The buffering capacity of the medium must therefore be given close attention to obtain chemical stability and maintain pH values within species tolerance ranges. Knowledge about the buffering

capacity and resulting pH stability of the medium are important, for instance, when metal toxicity or ionizable compounds are tested and organisms might influence the water pH either through respiration (CO_2 increase) or photosynthetic activity (CO_2 decrease; Rendal et al., 2012).

As part of standard single-species test guidelines, the Organisation for Economic Co-operation and Development (OECD) and Environmental Protection Agency in the United States (USEPA) provide detailed guidance on the preparation of test media for the culturing and the experimental exposure of test species. These test media are optimized for the specific requirements of each species. For instance, medium recommended for algal growth inhibition tests (OECD, 2011) contains high concentrations of metals and EDTA to maximize algal growth whereas media used in cladoceran immobilization and reproduction tests (*C. dubia* (USEPA, 2002)*, D. magna* (OECD, 2004)) usually contain more calcium required for carapace formation (Giardini et al., 2015) but no trace metals (with exception of magnesium). The latter are not necessary for cladoceran development and toxic at higher concentrations (Rodgher and Luiz Gaeta Espíndola, 2008). The formulation of a standardized test medium for multi-species tests is challenging because it must be chemically optimised to satisfy the requirements of all species. In other words, multi-species test media must incorporate a variety of trace metals, salts and vitamins and be chemically balanced at concentrations high enough to sustain species' growth and reproduction but low enough to not negatively impact any of them.

Yet, the choice of chemically defined and standardized media for multiple species belonging to different organism groups (e.g. algae, invertebrates) is rather limited because there is a lack of standardized microcosms. For instance, the guideline for the site-specific aquatic microcosm (USEPA, 1996) indicates the use of indigenous water, i.e. water from a specific water body that is mimicked. Other authors working on the assessment of chemical effects on species interactions in multi-species studies often report the use of membrane filtered water sampled from uncontaminated ditches (Van Wijngaarden et al., 2005; Boonstra et al., 2011), ponds (Van de Perre et al., 2018), streams (Englert et al., 2012), wells (Jarvis et al., 2014)

and the tap (Barry and Davies, 2004; Stampfli et al., 2011). Existing standardized media were used for microcosms with the same organism groups such as Elendt M7 medium (OECD, 2004) for microcosms with *D. magna* and *Culex pipiens* (Foit et al., 2012; Dolciotti et al., 2014). The only standardized multi-species media known to the authors are the 'standard reference water' (Freeman, 1953) and T82MV (ASTM E1366-11, 2011) that were developed for repeatable microcosms, the model ecosystem with aquatic-terrestrial interface (Metcalf et al., 1971) and the standardized aquatic microcosm (SAM; ASTM International, 2011). Medium T82MV was used in the present study and is recommended on the basis of interlaboratory testing (Taub et al., 1986; Taub, 1993). It was designed for the specific requirements of the species in the SAM and has low metal chelation and low pH buffering capacity (ASTM E1366-11, 2011). Although the SAM and the TriCosm contain similar initial algal concentrations (approx. 1 - 2 x 10^4 cells/mL), the stabilizing effects through photosynthesis of 10 algal species in the SAM were likely different to one algal species in the TriCosm and might explain why problems with low buffering capacity of the medium were not reported before.

Reproducible microcosms are still not employed for ERA in Europe and in the US, multi-species data are not required for most pesticide decisions since 1992 (Taub, 1997a). Hence, if standardized multi-species tests should be used to address the research needs on the improvement of ERA as outlined by the European commission (Scientific Committee on Health Environmental Risks et al., 2013), further work is needed on the development and/or refinement of chemically defined and balanced test media for multiple organism groups. In the case of the TriCosm, small differences of the medium composition between experiments likely caused large impacts on population dynamics and test findings. Improving the medium conditions by, for instance, an increase of the buffering concentration might solve the problem for one species (algae). However, in multi-species systems more complex relationships with other species must be considered and the stability of the medium could, for instance, be improved by non-chemical means such as thorough aeration (≥ 48 h) before test start (see *Chapter Four*, submitted for review).

3.6 Conclusion

Indirect effects of chemicals on species interactions can be detected in the aquatic tri-trophic microcosm TriCosm. Yet, the population responses and system repeatability are sensitive to the stability of the experimental medium and setup and medium preparation requires more understanding and control than previously thought. Efforts to achieve the standardization and repeatability of multispecies systems and their use in ERA are, however, worth pursuing to allow the integration of ecologically more relevant endpoints to routine environmental risk assessment tests. Further research on the development or refinement of test media for multiple organism groups is necessary for repeatable population dynamics in standardized multi-species tests.

3.7 Acknowledgement

The authors thank F. B. Taub for numerous discussions. This study was financed by the Department for Environment and Geography, University of York, Heslington, York, UK, and the Centre for Chemical Safety and Stewardship, Fera Science Ltd., Sand Hutton, York, UK. The authors have no conflict of interest to declare.

Chapter Four: Factors affecting the growth of *Pseudokirchneriella subcapitata*: lessons on the experimental design and consequences for the reproducibility of a multi-trophic laboratory microcosm

Verena Riedl,[a,b,*] Annika Agatz,[a,c] Rachel Benstead,[b] and Roman Ashauer[a]

[a]*Environment Department, University of York, Heslington, York, United Kingdom*

[b]*Fera Science Ltd., Centre for Chemical Safety and Stewardship, Sand Hutton, York, United Kingdom*

[c]*Current address: Ibacon GmbH, Institute for Biological Analytics and Consulting, Roßdorf, Germany.*

** Address correspondence to* verena.riedl@york.ac.uk

4.1 Abstract

The need for an integrated risk assessment at an ecologically relevant scale (e.g. at the population/community levels and on species interactions) has been acknowledged. Multi-species systems (e.g. microcosms) with increased ecological complexity, however, are difficult if not impossible to reproduce. The laboratory scale microcosm TriCosm (*Pseudokirchneriella subcapitata*, *Ceriodaphnia dubia*, *Hydra viridissima*) of intermediate complexity was developed for the reproducible assessment of chemical effects at the population and community levels. The system dynamics were repeatable in the short term but inter-experimental variation of algal dynamics in the long term triggered knock on effects on grazer and predator populations. Here we present 20 experiments with a total of 60 treatments to assess the effects of twelve factors (test medium, vessel type, shaking speed, light intensity, light regime, vessel condition, inoculation density, medium preparation components, metal concentration, metal composition, buffering salt type and salt concentration) on growth of the green alga *P. subcapitata* in the TriCosm (grazer and predator species excluded). Algal growth rates overall varied between -2.94 (±0.18) and 1.10 (±0.08) day^{-1} and generally were greatest with increased shaking speed, light exposure, medium buffer or aeration time whereas treatments conducted in dishes

with aseptically prepared, scarcely buffered and/or hardly aerated medium resulted in low algal growth rates. We found that inter-experimental variation of algal dynamics in the TriCosm was caused by a modification of medium preparation with the aim of reducing microbial contamination, i.e. the omission of medium aeration. Our findings highlight that consistency in experimental procedures and in-depth understanding of system components are indispensable to achieve repeatability even in simple systems. We stress the importance of performing and publishing reproducibility studies to decrease concerns over the value of ecotoxicological data to regulators.

4.2 Introduction

Reproducibility is a key driver for scientific progress, but evidence suggests that a number of peer-reviewed experiments cannot be reproduced (Ioannidis, 2005; Hines et al., 2014; Harris and Sumpter, 2015; Baker, 2016). Concerns over a 'reproducibility crisis' in scientific research were first raised when investigations attempting to replicate recently published key results in the biomedical literature failed (Begley and Ellis, 2012; Hunter, 2017; Hanson et al., 2017). Figures from a survey conducted by *Nature* in 2016 indicated that over 70% of researchers (n = 1,576) failed at reproducing other scientists' studies. Insufficient or selective reporting, pressure to publish, insufficient replication of the original work, low statistical power and inappropriate analyses were listed among the reasons for irreproducible results (Baker, 2016), and reproducibility rates of only 11% were reported in preclinical cancer research (Begley and Ellis, 2012). Less is known about the reproducibility and robustness of ecotoxicological studies which is concerning because regulators often rely on empirical data from the peer reviewed literature to inform and underpin decisions regarding the environmental safety of chemicals (Harris and Sumpter, 2015; Attanasio, 2016; Hanson et al., 2017).

The effects of chemical substances at higher levels of organization (i.e. populations, communities) can be studied in multi-species systems (e.g. micro- and

mesocosms) where organisms experience competition, trophic interactions and/or other stressors (Taub, 1997b). Assessments made in the presence of multiple trophic layers (i.e. in microcosms) are potentially more comprehensive measures than isolated effects studied in single-species tests, where the conditions for that species are optimized. Interacting populations are more likely to be vulnerable to chemical exposure and the combination of stressors has important implications for their sensitivity. Hence, indirect effects could have implications for threshold-concentrations (e.g. Regulatory Acceptable Concentration, Predicted No Effect Concentration, Environmental Quality Standard) that are considered environmentally safe and mainly based on the direct effects measured in single-species tests (Van Straalen, 2003; Fleeger et al., 2003; Bednarska et al., 2013; Kattwinkel et al., 2015).

While the assessment of chemical effects in more environmentally relevant test systems would be beneficial, empirical results obtained in complex systems often cannot be used for regulatory purposes because the data do not fulfil core regulatory requirements, often due to high variability (Hanson et al., 2017). Variability has differing definitions in the literature (OECD, 2005; Goodman et al., 2016; Leek & Jager 2017; Riedl et al., 2018) and here we describe differences within experiments as intra-test variability, small differences between experiments within the same laboratory as repeatability and differences between experiments amongst laboratories as reproducibility. Stressor effects can only be distinguished statistically from natural variations when intra-test variability is low and repeatability is high. Yet, the variability between replicates often increases with greater environmental realism because system properties (e.g. community dynamics) emerge from independent but ecologically interacting elements. Interlocked reciprocal cause-effect pathways (components constantly adapt and react to changing conditions) often lead to non-linear dynamics and make complex systems difficult to understand and system evolution extremely sensitive to initial conditions (Ladyman et al., 2013).

In multi-species tests, system behaviour is regulated by processes among organisms (e.g. competition, trophic interactions, symbiosis) and between organisms

and the abiotic environment (e.g. nitrogen fixation, nutrient cycling, photosynthesis; Poisot et al., 2015). However, to achieve repeatability and reproducibility, realism and experimental control must be at balance and the standardization of physical, chemical, and biological parameters and initiation procedures are vital (Harris et al., 2014; Hanson et al., 2017). A full understanding of the properties of experimental components and the behaviour of control dynamics must be gained. For instance, the characteristics of the test species (e.g. culturing history, lifespan, reproductive output or morphological condition; Lithgow et al., 2017) and the chemical suitability of the aquatic test medium for all organisms in multiple species systems (e.g. nutrient load, salinity or pH) are critical factors that can impact reproducibility (Taub, 1997b; Rendal et al., 2012). Even in apparently simple experiments, small variations in space or time (e.g. starting conditions or uneven sampling) can cause divergent replicate behaviour or inconsistencies among experiments even if the set up was seemingly identical (Hines et al., 2014; Poisot et al., 2015; Lithgow et al., 2017). For instance, Lithgow et al. (2017) worked for four years to determine possible sources of inconsistency of their own findings previously published in *Science* (Melov et al., 2000). Similarly, Hines et al. (2014) described an occasion where two groups of scientists collaborated for over a year before finding that different sample stirring patterns had led to irreproducible results.

Standardized laboratory microcosms of intermediate ecological complexity are valuable tools to integrate ecological realism into chemical safety assessment (Taub, 1997b; Riedl et al., 2018). However, the aquatic microcosm SAM (10 primary producer and 5 primary consumer species) is the only standardized, multi-species system known to the authors that successfully passed inter-laboratory testing (Taub, 1997b). Similarly, the standardized, tri-trophic microcosm TriCosm (*Pseudokirchneriella subcapitata*, *Ceriodaphnia dubia*, *Hydra viridissima*) was optimized to minimize intra-test variability and fulfilled repeatability in the short term (over three months; Riedl et al., 2018). However, when the long-term repeatability of the system was tested (unpublished data, see *Appendix C* Figure C -

1) the population dynamics of the primary producer differed to those observed during standardization and caused direct and indirect impacts across the food chain.

The purpose of the work reported here is to determine the factor(s) causing inter-experimental variations of algal population dynamics because grazer and predator species and the overall system dynamics are influenced by this base trophic layer. We specifically focus on factors impacting algal populations in the TriCosm without the presence of grazer and predator populations to exclude confounding top-down impacts by these trophic layers. We investigated how different experimental conditions impact the growth of *P. subcapitata* and determined which factors mattered most. The aims are to i) empirically refine the range of conditions under which the dynamics of *P. subcapitata* can be reproduced and ii) discuss our findings in relation to concerns over the reproducibility multi-species test systems.

We investigated the effects of twelve factors (test medium, vessel type, shaking speed, light intensity, light regime, vessel condition, inoculation density, medium preparation components, metal concentration, metal composition, buffering salt type and salt concentration) on algal growth and selected a total of 60 factor combinations (experimental details are summarized in Table 4 - 1) because a full factorial design was not feasible.

4.3 Materials and methods

4.3.1 Algal culture and general experimental conditions

The green alga *Pseudokirchneriella subcapitata* is the primary producer in the tri-trophic laboratory microcosm TriCosm and influences the grazer *Ceriodaphnia dubia* directly as a food source and has indirect impacts on the population dynamics of the predator *Hydra viridissima.* Prior to each test, new algal batch cultures (250 mL Duran® Erlenmeyer flasks, ca. 150 mL OECD medium (OECD, 2006) with ca. 3.5×10^5 cells (stock cultures from the Culture Collection of Algae and Protozoa, Scotland, UK)) were incubated for 72h at 110 rpm orbital shaking, 23 ± 1°C, 24h light and an average

of 70 µmol m^{-2} s^{-1}. *P. subcapitata* from these batch cultures in the exponential growth phase were used to inoculate the experimental replicates positioned on an orbital shaker. Shaking kept the algal cells suspended and, unless indicated otherwise, water columns were not mixed before sampling. Algal population size was determined at either four or five sampling points via in-vivo fluorescence activity measurements (5 x 200 µl subsamples) with a plate reader (Tecan® Infinite 200 PRO; for settings see *Appendix C* Table C - 1) to determine the algal concentration. Vessels were sterilized (either baked at 121 °C or acid washed in 10% HCl), dishes were covered with watch glasses (diameter 125 mm; Sigma-Aldrich, UK), flasks were closed with autoclaved Azpack™ non-absorbent cotton wool (Fisher Scientific, UK) and experiments were performed at 25 (± 1) °C, 65 rpm orbital shaking and ca. 16 µmol m^{-2} s^{-1}. Water pH was measured with an Orion Star™ Multiparameter meter (Thermo Fisher Scientific, UK) either daily or on day 0 and at the end of the experiments.

4.3.2 Which experimental components limit algal growth?

Two experiments (Exp) were performed to investigate impacts of test medium, vessel type, shaking speed, light intensity and photoperiod on algal dynamics (Exp1 and Exp2, Table 4 - 1). Factors were set to i) standard TriCosm test conditions (i.e. TriCosm medium T82MV (*Appendix C* Table C - 2, Table C - 3) in 500 mL Pyrex® crystallizing dishes (Sigma-Aldrich, UK) at 65 rpm orbital shaking, 25 ± 1°C, 12/12h light/dark and ca. 16 µmol m^{-2} s^{-1}; Riedl et al., 2018) or ii) standard OECD algal growth inhibition test conditions (i.e. OECD medium in e.g. 250 mL Duran® Erlenmeyer flasks (Sigma-Aldrich, UK) at 150 rpm orbital shaking, 23 ± 1°C, 24h light, ca. 70 µmol m^{-2} s^{-1}; OECD, 2011).

The test medium, vessel type and overall incubation condition (i.e. shaking speed, temperature and light intensity/photoperiod) were varied in 8 treatments in Exp1 (for details see Table 4 - 1). The water columns were mixed with a sterile pipette prior to sampling because the shaking speed differed between treatments (Exp1). High shaking speed is not suitable for the animal species in the TriCosm and, in Exp2, four treatments were performed at 65 rpm with different light intensity (either ca.

16 or 60 - 80 µmol m^{-2} s^{-1} (position dependent) and 12/12h photoperiod) and vessel type (Erlenmeyer flasks or crystallizing dishes). Dishes and flasks were covered with cling film and autoclaved cotton wool, respectively, to reduce airborne contamination and were filled with 450 or 200 mL of media, respectively, to obtain similar shaking patterns (due to different volume/surface ratio) at a low shaking speed (65 rpm). At high shaking speed (150 rpm), all vessels were filled with approx. 200 mL to avoid spillage.

4.3.3 Does increased light exposure affect algal dynamics?

Three experiments were performed to investigate impacts of light intensity (ca. 70 or 300 µmol m^{-2} s^{-1}), vessel type (dish or flask) and preadaptation of algae to the test medium (Exp3, Exp4 and Exp5, Table 4 - 1) on algal population dynamics.

To allow the algae to adapt and reduce stress related to media change, *P. subcapitata* starter lines were cultured in T82MV instead of OECD medium (for details on incubation conditions see section on Algal culture and general experimental conditions). The light intensity was increased to 70 µmol m^{-2} s^{-1} in Exp3 and Exp4 and 300 µmol m^{-2} s^{-1} in Exp5 during the first 24 h (acclimation phase) and decreased to ca. 16 µmol m^{-2} s^{-1} until the end of the study. Test vessels were covered with cling film to reduce airborne contamination.

4.3.4 Does the condition of the exposure vessels affect algal dynamics?

Two experiments were performed to investigate impacts of cleaning procedure and vessel condition (previously used or newly purchased vessels) on algal population growth (Exp6 and Exp7, Table 4 - 1).

Previously used dishes were hand washed and subsequently i) soaked in demineralized water (ca. 24h), ii) acid washed (10% HCl) or iii) baked at 120 °C (ca. 24h) (4 replicates each). Four treatments were prepared per cleaning practice with dishes inoculated with one of four algal concentrations (2 x 10^4, 5 x 10^4, 10 x 10^4, 20 x 10^4 cells/mL). In addition, new crystallizing dishes were purchased to exclude the

impacts of algal growth due to residues on glass surfaces. Two treatments were prepared with 2 x 10^4 cells/mL algae in flasks that served as controls (as per standard OECD algal growth inhibition test protocol (OECD, 2011)). This experiment was conducted in dishes covered with cling film to minimize contamination. Populations in Exp7 were subsampled before and after mixing the water columns with a sterile pipette, and concentrations were compared to assess whether any cell aggregations not visible by eye were affecting cell measurements.

As a result of poor growth and algal clumps found in Exp7 after 72h, the light intensity was increased to 300 µmol m^{-2} s^{-1} in dishes and 70 µmol m^{-2} s^{-1} in flasks. The incubation conditions were improved in this manner to determine if population growth within the test units was limited by external (e.g. light intensity) or by internal factors (e.g. bacteria).

4.3.5 Does the test medium impact algal dynamics?

Three sets of experiments were performed to assess if algal dynamics are influenced by the components used for medium preparation, in particular the composition and the concentration of trace metals and the buffering capacity of the medium.

4.3.5.1 Preparation components

Four experiments were performed to assess if biotic or abiotic waterborne contamination during media preparation or the quality of starter algal source limited the population dynamics of the algae (Exp8 to Exp11, Table 4 - 1). Media were prepared i) in a different medium preparation container, ii) with demineralized water from a different source, iii) using newly acquired algal starter source or iv) by autoclaving the basal medium and adding 0.22 µm membrane filtered stock solutions. Three or four treatments were prepared for each experiment with different algal concentrations. Medium blanks were set up in two experiments to monitor pH over time.

4.3.5.2 Trace metals

Three experiments were performed to investigate impacts of trace metal concentration and composition on algal population dynamics (Exp12 toExp14, Table 4 - 1). To improve metal availability, media were prepared with i) standard or twofold trace metal concentration, ii) EDTA as Ethylenedinitrilotetraacetic (DN), Ethylenediaminetetraacetic disodium (Na_2) and Ethylenediaminetetraacetic tetrasodium (Na_4) salt or iii) with the addition of Keating's metal solution (*Appendix C* Table C - 4) to limit the precipitation of phosphates possibly leading to algal aggregation (ASTM E1366-11, 2011). Two or three treatments were set up per experiment.

4.3.5.3 Buffering capacity

Because insufficiently buffered medium can lead to fluctuating pH and limit algal growth (Rendal et al., 2012), five experiments were performed to assess impacts of quantities of buffering salt and the pH adjustment method on medium stability and algal growth (Exp15 to Exp20, Table 4 - 1). Media were prepared with different i) buffering salt concentrations (0.23 mg/L, 0.05 mg/L and 0.02 mg/L Na_2SiO_3) below the toxic range for TriCosm test species (*Appendix C* Table C - 5) and/or ii) pH adjustment methods (HCl and/or aeration for 24 or 48h) and one, two or six treatments were prepared per experiment.

In Exp18, treatments were prepared with un-aerated (T1 to T5) and aerated media (T6). In Exp18 T6, medium pH was not further adjusted after aeration (approx. 24h) while in T1 to T5, medium pH was adjusted to one of five pH values (pH 8.0, pH 7.5, pH 7.0, pH 6.5, pH 6.0) with HCl to compare algal growth between treatments with different starting pH and pH adjustment method. In Exp20, two types of T82MV were prepared with i) $NaHCO_3$ (0.04 g/L) or ii) Na_2SiO_3 (0.02 mg/L) and aerated (48h) to compare algal growth and pH change between media buffered with different salts and aerated for 48h. The water pH was monitored daily or at two or three time points during the experiments and compared between blanks and treatments to determine possible effects of algae on medium pH through microbial respiration and photosynthetic activity (for details see Table 4 - 1).

4.3.6 Statistical analyses

The replicate number was low in order to maximize the total number of experiments feasible and explore which experimental factor limits algal growth in the TriCosm. Exponential growth models were not fitted because algal concentrations declined in most experiments before increasing densities could be measured. Algal dynamics and growth rates were compared with values computed in the standardized TriCosm (n = 8, without grazing pressure as indicated for reference in graphs with stars). Algal growth rates were computed between day 0 and, depending on the sampling time point, day 4, day 5 or day 6 and compared among treatments. Average specific growth rates were calculated according to OECD (2011).

$$\mu_{i-j} = \frac{\ln X_j - \ln X_i}{t_j - t_i} \ (\mathrm{day}^{-1})$$

In this equation $\mu_{i\text{-}j}$ is the specific growth rate averaged from time t_i to t_j and X_i and X_j are cell concentrations (cells/mL) at time t_i and t_j, respectively.

Table 4 - 1: Conditions in different experimental treatments with the alga *P. subcapitata*. Reported are the factors investigated, experimental duration, characteristics of the medium, light exposure, shaking speed, algal inoculation density, vessel type and replicate number of each experiment and treatment. Factor variations among treatments and experiments are indicated in bold.

				Medium						**Light**					
Factors	**Exp**	**Treatment**	**Duration (d)**	**Type**	**Aseptic** [1]	**Buffer**[2] **(g/L)**	**pH adjust**[3]	**Air**[4] **(h)**	**Metals (mL/L)**	**Photoperiod (h; light/dark)**	**Intensity (µmol m^{-2} s^{-1})**	**Shaking (rpm)**	**Inoculation (cells/mL)**	**Vessel** [5]	**Replicates**
		T1		T82MV	Yes	0.02	HCl	-	0.05	12/12	16	65	2×10^4	Dish	3
		T2		T82MV	Yes	0.02	HCl	-	0.05	**24/0**	**70**	**150**	2×10^4	Dish	3
		T3		T82MV	Yes	0.02	HCl	-	0.05	12/12	16	65	2×10^4	**Flask**	3
		T4		T82MV	Yes	0.02	HCl	-	0.05	**24/0**	**70**	**150**	2×10^4	**Flask**	3
Medium/		**T5**		**OECD**	Yes		HCl	-		12/12	16	65	2×10^4	Dish	3
vessel type,		**T6**		**OECD**	Yes		HCl	-		**24/0**	**70**	**150**	2×10^4	Dish	3
shaking		**T7**		**OECD**	Yes		HCl	-		12/12	16	65	2×10^4	**Flask**	3
speed, light	**1**	**T8**	4	**OECD**	Yes		HCl	-		**24/0**	**70**	**150**	2×10^4	**Flask**	3
		T1		T82MV	Yes	0.02	HCl	-	0.05	12/12	16	65	2×10^4	Dish	3
		T2		T82MV	Yes	0.02	HCl	-	0.05	12/12	**70**	65	2×10^4	Dish	3
		T3		T82MV	Yes	0.02	HCl	-	0.05	12/12	16	65	2×10^4	**Flask**	3
	2	**T4**	5	T82MV	Yes	0.02	HCl	-	0.05	12/12	**70**	65	2×10^4	**Flask**	3
										24/0 (0-24h)	**70 (0-24h)**				
		T1		T82MV	Yes	0.02	HCl	-	0.05	12/12 (≥24h)	16 (≥24h)	65	2×10^4	Dish	3
Vessel type,										**24/0 (≤24h)**	**70 (≤24h)**				
light	**3**	**T2**	4	T82MV	Yes	0.02	HCl	-	0.05	12/12 (≥24h)	16 (≥24h)	65	2×10^4	**Flask**	3
Vessel type,										**24/0 (≤24h)**	**70 (≤24h)**				
algae pre-		**T1**		T82MV	Yes	0.02	HCl	-	0.05	12/12 (≥24h)	16 (≥24h)	65	2×10^4	Dish	3
adapted,										**24/0 (≤24h)**	**70 (≤24h)**	65			
light	**4**	**T2**	5	T82MV	Yes	0.02	HCl	-	0.05	12/12 (≥24h)	16 (≥24h)		2×10^4	**Flask**	3

Factors	Exp	Treatment	Duration (d)	Medium: Type	Aseptic [1]	Buffer [2] (g/L)	pH adjust [3]	Air [4] (h)	Metals (mL/L)	Light: Photoperiod (h; light/dark)	Intensity (µmol m^{-2} s^{-1})	Shaking (rpm)	Inoculation (cells/mL)	Vessel [5]	Replicates
		T1		T82MV	Yes	0.02	HCl	-	0.05	**24/0 (≤24h)** 12/12 (≥24h)	**300 (≤24h)** 16 (≥24h)	65	2 x 10^4	Dish	3
		T2		T82MV	Yes	0.02	HCl	-	0.05	**24/0 (≤24h)** 12/12 (≥24h)	**300 (≤24h)** 16 (≥24h)	65	2 x 10^4	**Flask**	3
Vessel type, light	**5**	**Blank**	4	T82MV	Yes	0.02	HCl	-	0.05	**24/0 (≤24h)** 12/12 (≥24h)	**300 (≤24h)** 16 (≥24h)	65	**Blank**	Dish	1
		T1		T82MV	**No**	0.02	HCl	-	0.05	12/12	16	65	2 x 10^4	Dish	3
		T2		T82MV	**No**	0.02	HCl	-	0.05	12/12	16	65	**5 x 10^4**	Dish	3
		T3		T82MV	**No**	0.02	HCl	-	0.05	12/12	16	65	**10 x 10^4**	Dish	3
Vessel cleaning	**6**	**T4**	5	T82MV	**No**	0.02	HCl	-	0.05	12/12	16	65	**20 x 10^4**	Dish	3
Vessel condition		**T1**		T82MV	Yes	0.02	HCl	-	0.05	12/12	16	65	2 x 10^4	**Dish, new**	3
(new dish)	**7**	**T2**	4	T82MV		0.02	HCl	-	0.05	12/12	16	65	2 x 10^4	**Flask**	3
		T1		T82MV	Yes	0.02	HCl	-	0.05	12/12	16	65	2 x 10^4	Dish	3
		T2		T82MV		0.02	HCl	-	0.05	12/12	16	65	**5 x 10^4**	Dish	3
Medium		**T3**		T82MV		0.02	HCl	-	0.05	12/12	16	65	**10 x 10^4**	Dish	3
container,		**T4**		T82MV		0.02	HCl	-	0.05	12/12	16	65	**20 x 10^4**	Dish	3
Algal density	**8**	**Blank**	5	T82MV		0.02	HCl	-	0.05	12/12	16	65	**Blank**	Dish	1
		T1		T82MV	Yes	0.02	HCl	**2**	0.05	12/12	16	65	2 x 10^4	Dish	3
		T2		T82MV		0.02	HCl	**2**	0.05	12/12	16	65	**5 x 10^4**	Dish	3
H_2O source		**T3**		T82MV		0.02	HCl	**2**	0.05	12/12	16	65	**10 x 10^4**	Dish	3
changed,		**T4**		T82MV		0.02	HCl	**2**	0.05	12/12	16	65	**20 x 10^4**	Dish	3
Algal density	**9**	**Blank**	4	T82MV		0.02	HCl	**2**	0.05	12/12	16	65	**Blank**	Dish	1
	10	**T1**	5	T82MV	Yes	0.02	HCl	-	0.05	12/12	16	65	2 x 10^4	Dish	3

Factors	Exp	Treatment	Duration (d)	Medium: Type	Aseptic[1]	Buffer[2] (g/L)	pH adjust[3]	Air[4] (h)	Metals (mL/L)	Light: Photoperiod (h; light/dark)	Intensity ($\mu mol\ m^{-2}\ s^{-1}$)	Shaking (rpm)	Inoculation (cells/mL)	Vessel[5]	Replicates
New starter		**T2**		T82MV		0.02	HCl	-	0.05	12/12	16	65	**5×10^4**	Dish	3
culture, Algal		**T3**		T82MV		0.02	HCl	-	0.05	12/12	16	65	**10×10^4**	Dish	3
density		**T4**		T82MV		0.02	HCl	-	0.05	12/12	16	65	**20×10^4**	Dish	3
Autoclaved		**T1**		T82MV	Yes	0.02	HCl	-	0.05	12/12	16	65	2×10^4	Dish	3
medium,		**T2**		T82MV		0.02	HCl	-	0.05	12/12	16	65	**10×10^4**	Dish	3
Algal density	**11**	**T3**	4	T82MV		0.02	HCl	-	0.05	12/12	16	65	**20×10^4**	Dish	3
		T1		T82MV	Yes	0.02	HCl	-	0.05	12/12	16	65	2×10^4	Dish	3
	12	**T2**	5	T82MV		0.02	HCl	-	**0.10**	12/12	16	65	2×10^4	Dish	3
		T1		T82MV	**No**	0.02	HCl	-	0.05 **DN**[6]	12/12	16	65	2×10^4	Dish	2
		T2		T82MV		0.02	HCl	-	0.05 **Na_2**[6]	12/12	16	65	2×10^4	Dish	2
Trace metals	**13**	**T3**	5	T82MV		0.02	HCl	-	0.05 **Na_4**[6]	12/12	16	65	2×10^4	Dish	2
Algal density,		**T1**		T82MV	Yes	0.02	HCl	-	0.05	12/12	16	65	2×10^4	Dish	3
Keating's		**T2**		T82MV		0.02	HCl	-	0.05	12/12	16	65	**10×10^4**	Dish	3
metals	**14**	**T3**	4	T82MV		0.02	HCl	-	0.05	12/12	16	65	**20×10^4**	Dish	3
Increased		**T1**		T82MV	Yes	**0.23**	HCl	-	0.05	12/12	16	65	2×10^4	Dish	3
buffer	**15**	**Blank**	6	T82MV		**0.23**		-	0.05	12/12	16	65	2×10^4	Dish	1
Increased		**T1**		T82MV	Yes	**0.23**	HCl	-	0.05	12/12	16	65	2×10^4	Dish	3
buffer,		**T2**		T82MV	Yes	**0.23**	**-**	**48**	0.05	12/12	16	65	2×10^4	Dish	3
medium		**Blanks**					see	see							
aeration	**16**	**T1,T2**	4	T82MV	Yes	**0.23**	T1,T2	T1,T2	0.05	12/12	16	65	2×10^4	Dish	2x1
								24+24							
	17	**T1**	5	T82MV	No	**0.05**	HCl	**equil.**	0.05	12/12	16	65	2×10^4	Dish	3

Factors	Exp	Treatment	Duration (d)	Medium						Light		Shaking (rpm)	Inoculation (cells/mL)	Vessel[5]	Replicates
				Type	Aseptic[1]	Buffer[2] (g/L)	pH adjust[3]	Air[4] (h)	Metals (mL/L)	Photoperiod (h; light/dark)	Intensity (µmol m^{-2} s^{-1})				
		T1		T82MV	Yes	0.02	**6.0±0.1**		0.05	12/12	16	65	2 x 10^4	Dish	3
		T2		T82MV		0.02	**6.5±0.1**		0.05	12/12	16	65	2 x 10^4	Dish	3
		T3		T82MV		0.02	**7.0±0.1**		0.05	12/12	16	65	2 x 10^4	Dish	3
		T4		T82MV		0.02	**7.5±0.1**		0.05	12/12	16	65	2 x 10^4	Dish	3
		T5		T82MV		0.02	**8.0±0.1**		0.05	12/12	16	65	2 x 10^4	Dish	3
		T6		T82MV		0.02	-	24	0.05	12/12	16	65	2 x 10^4	Dish	3
Initial pH value	**18**	**Blanks T1-T6**	5	T82MV		0.02	see T1-T6	see T1-T6	0.05	12/12	16	65	**Blank**	Dish	6x1
Medium aeration	**19**	**T1**	6	T82MV	**No**	0.02	**-**	**48**	0.05	12/12	16	65	2 x 10^4	Dish	3
		T1	8	T82MV	**No**	0.02	**-**	**48**	0.05	12/12	16	65	2 x 10^4	Dish	3
		T2		T82MV	**No**	**0.04 $NaHCO_3$**	**-**	**48**	0.05	12/12	16	65	2 x 10^4	Dish	3
Buffer salt, medium aeration	**20**	**Blanks T1,T2**		T82MV	**No**	as per T1,T2	**-**	**48**	0.05	12/12	16	65	**Blank**	Dish	2x1

[1] Aseptic: if yes, autoclaved or 0.22µm membrane filtered stock solutions were added to autoclaved demineralized water, no aeration.
[2] Buffer: Na_2SiO_3 ($NaHCO_3$ in Exp20 T2 and Exp20 blank T2).
[3] pH adjustment to pH 7.0 ± 0.1 with HCl (10%).
[4] Aeration with a 10 mL Fisherbrand™ glass pipette (Fisher Scientific, UK) attached to a Whisper® Aquarium Air Pump (TetraFish, UK).
[5] Dish: 500 mL Pyrex® crystallizing dishes, filled with 400 – 450 mL (ca. 200 mL in Exp1 T2, T5); Flask: 250 mL Duran® Erlenmeyer conical flasks, filled with 150 – 200 mL.
[6] EDTA DN: Ethylenedinitrilotetraacetic acid salt, Na_2: Ethylenediaminetetraacetic acid disodium salt, Na_2: Ethylenediaminetetraacetic acid tetrasodium salt.

4.4 Results

4.4.1 Algal growth rates: comparison among treatments

Average specific algal growth rates (Figure 4 - 1) ranged from - 2.94 (± 0.18) in Exp18 T2 (T82MV, 65 rpm, 12h-16 µmol m^{-2} s^{-1}, 2 x 10^4 cells/mL, dish) to 1.10 (± 0.08) in Exp1 T8 (OECD medium, 150rpm, 24h-70 µmol m^{-2} s^{-1}, 2 x 10^4 cells/mL, flask). The highest algal growth rates (range: 0.26 (± 0.26) to 1.10 (± 0.08), mean ± 95% confidence interval) were found in treatments with high shaking speed (150 rpm), high light exposure (24h-70 µmol m^{-2} s^{-1}), high buffer concentration (0.23 g/L), aerated medium (48h), OECD medium and/or conducted in flasks. Intermediate growth rates (range: 0.19 (± 0.10) to 0.11 (± 0.04)) were found in treatments with increased algal inoculation densities (10 x 10^4 and 20 x 10^4 cells/mL) or conducted in flasks. The lowest growth rates (range: 0.11 (± 0.12) to -2.94 (± 0.18)) were found in treatments conducted in dishes with low buffer concentration and aseptically prepared medium, or short medium aeration time (≤ 24h) and the variability among replicates increased.

Growth rates comparable to those measured in the standardized TriCosm (0.47 (± 0.04) and 0.79 (± 0.08) on days 4 and 5, respectively), were found in treatments conducted in dishes with increased buffer concentration or medium aerated for 48h or increased light exposure (photoperiod or intensity).

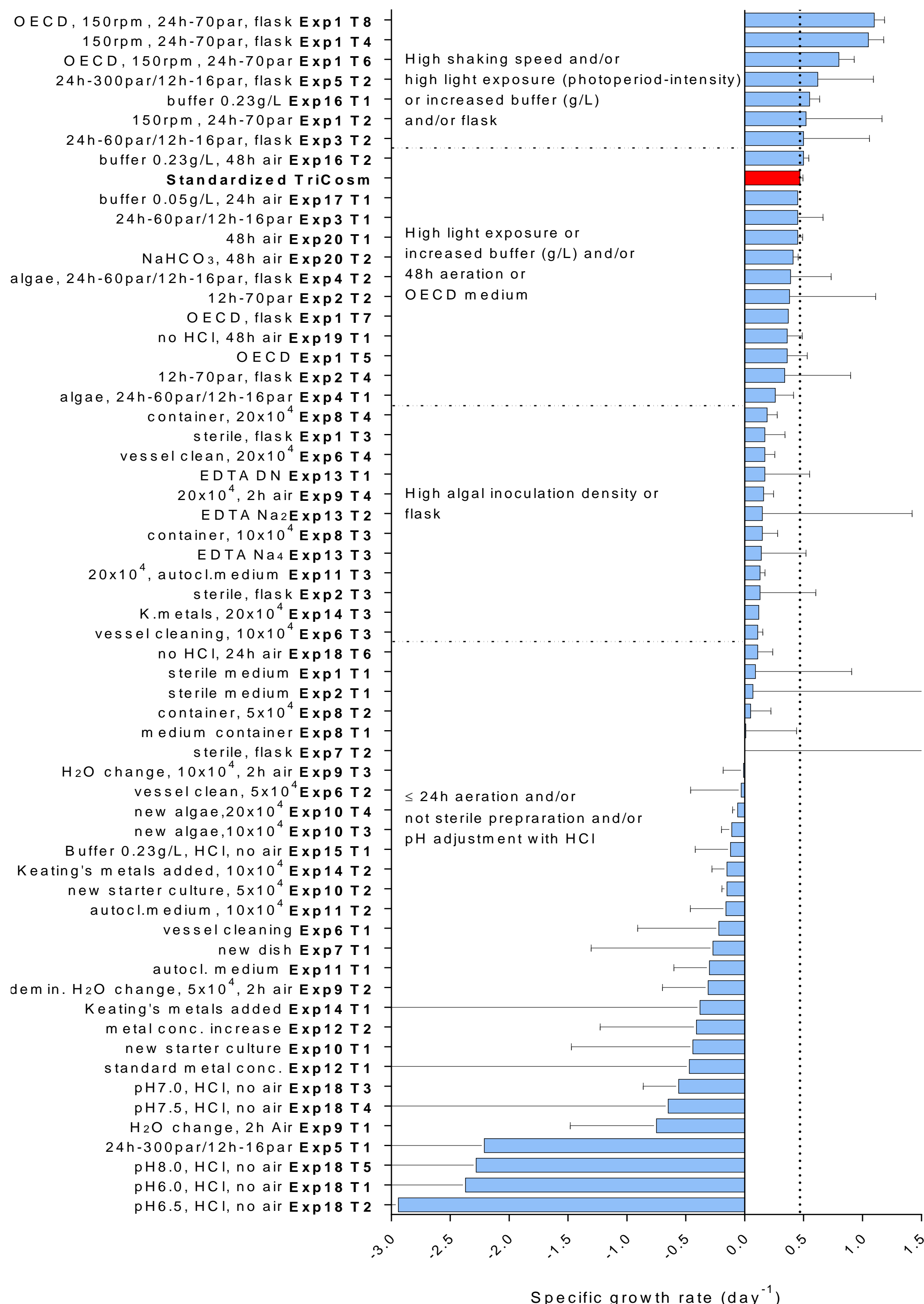

Figure 4 - 1: Specific algal growth rates (day^{-1}) are ranked from highest to lowest (top to bottom) in treatments of experiments 1 - 20 (experimental details see Table 4 - 1). Labels indicate factor variations among treatments and conditions different to the standardized TriCosm (T82MV, 12h-16par µmol m^{-2} s^{-1}, EDTA DN, buffer 0.02 g/L, 48h medium aeration, dish). The algal growth rate in the standardized TriCosm without grazing pressure on day 4 is indicated in red for reference. Shown are mean values + 95% confidence intervals.

4.4.2 Algal population dynamics: medium, shaking speed and light exposure

Cell concentrations generally decreased during an initial lag phase (day 0 - 2) in Exp1 - Exp5 where the type of test medium, the shaking speed and/or light exposure (photoperiod-intensity) were varied. Algal growth was supported in both types of media, but algal populations were generally larger in treatments with OECD medium (Exp1 T5 - T8) than in treatments with T82MV medium (Exp1 T1 - T4) when compared within the incubation environment (shaking speed, photoperiod, light intensity and vessel type; *Appendix* C Figure C - 2). Similarly, algal growth was supported at low (65 rpm) and high (150 rpm) shaking speed but, irrespective of medium and vessel type, algal concentrations were generally higher in treatments with increased shaking speed and light exposure (photoperiod-intensity; Exp1 T2, T4, T6, T8) than in treatments with low shaking speed and shorter photoperiod (Exp1 T1, T3, T5, T7) (*Appendix* C Figure C - 2). At constant shaking speed and within vessel type, algal concentrations were higher in treatments where the light intensity was increased (Exp2 T2, T4: 12h photoperiod throughout the experiment; Exp3, Exp4, Exp5: for the first 24h after test start) compared to treatments where the light intensity and photoperiod were constantly low (Exp2 T1, T3; Figure 4 - 2).

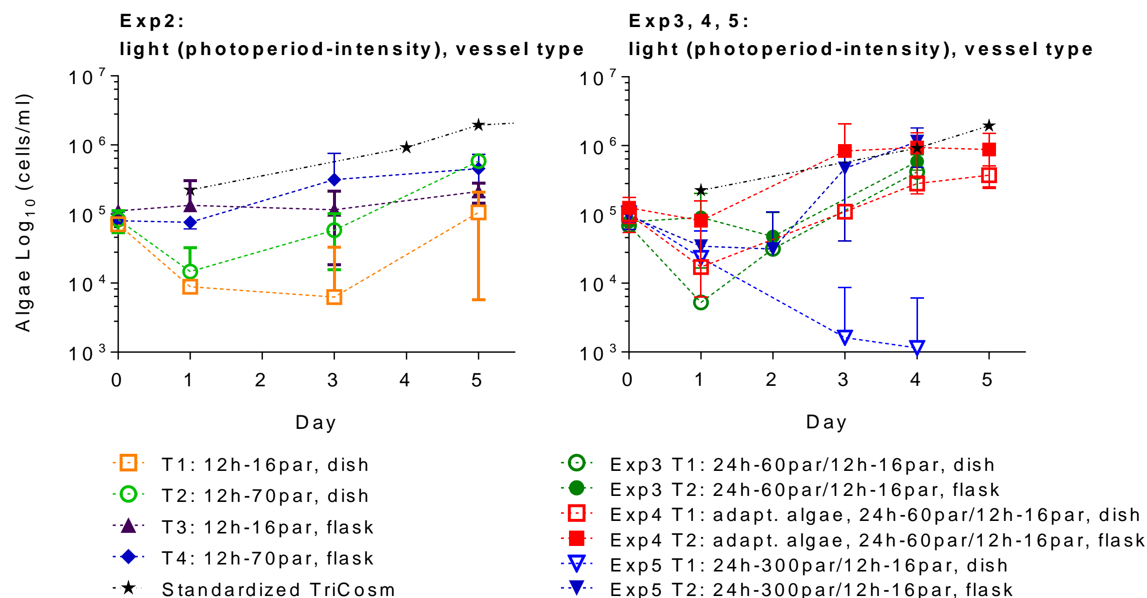

Figure 4 - 2: Algal population dynamics in experiments 2 - 5 with differing light exposure (photoperiod- intensity (par µmol m^{-2} s^{-1}) and vessel type. Labels indicate factor variations among treatments (for details see Table 4 - 1). Algal dynamics observed in the standardized TriCosm without grazing pressure are indicated with stars for reference. Shown are mean values (n = 3) ± 95% confidence intervals.

4.4.3 Algal population dynamics: vessel type/condition and metal concentration/composition

Low algal concentrations were measured in both vessel types during an initial lag phase (day 0 - 2) and algal concentrations in dishes were on average 4.2, 1.6, 1.4 and 2.4 times smaller than in flasks in Exp1, Exp2, Exp3 and Exp4, respectively when compared within the same incubation environment (T1; Figure 4 - 2, *Appendix C* Figure C - 2). No algal growth and low concentrations were measured in dishes in Exp5. The use of differently pre-treated (Exp6) and new test vessels (Exp7) did not positively impact algal growth (*Appendix C* Figure C - 3).

Algal concentrations were measured before and after the water columns were mixed (Exp7, Exp15) and differences (= cells/mL $_{\text{after mixing}}$ – cells/mL $_{\text{before mixing}}$) correlated in flasks and dishes (*Appendix C* Figure C - 4) suggesting that uneven algal detection due to sampling procedures is unlikely a cause for low algal population measurements.

No substantial variations of algal growth were found between populations incubated in test media with i) increased metal concentration (Exp12), ii) disodium- (Na_2), tetrasodium- (Na_4) or dinitrilo- (DN) EDTA (Exp13) or iii) addition of Keating's metals (Exp14). Algal cell densities decreased within 24h after test start and remained low (Exp12) or increased to concentrations similar to inoculation densities (Exp13, 14, T1; *Appendix C* Figure C - 3).

4.4.4 Algal population dynamics: inoculation density

Growth dynamics varied among treatments with different inoculation concentrations, but were similar among experiments (Figure 4 - 3). Extended lag phases and low densities throughout the experimental duration were observed at low inoculation concentrations (2 x 10^4, 5 x 10^4 cells/mL) whereas moderate to exponential growth was measured at higher inoculation concentrations (10 x 10^4 and 20 x 10^4 cells/mL; Figure 4 - 3). The steepest population increase and largest cell densities were measured in experiments where the basal medium was autoclaved (Exp11) and Keating's metals were added (Exp14), however, the measured inoculation concentrations (day 0) averaged among treatments was approx. 3.5 times higher than in other experiments.

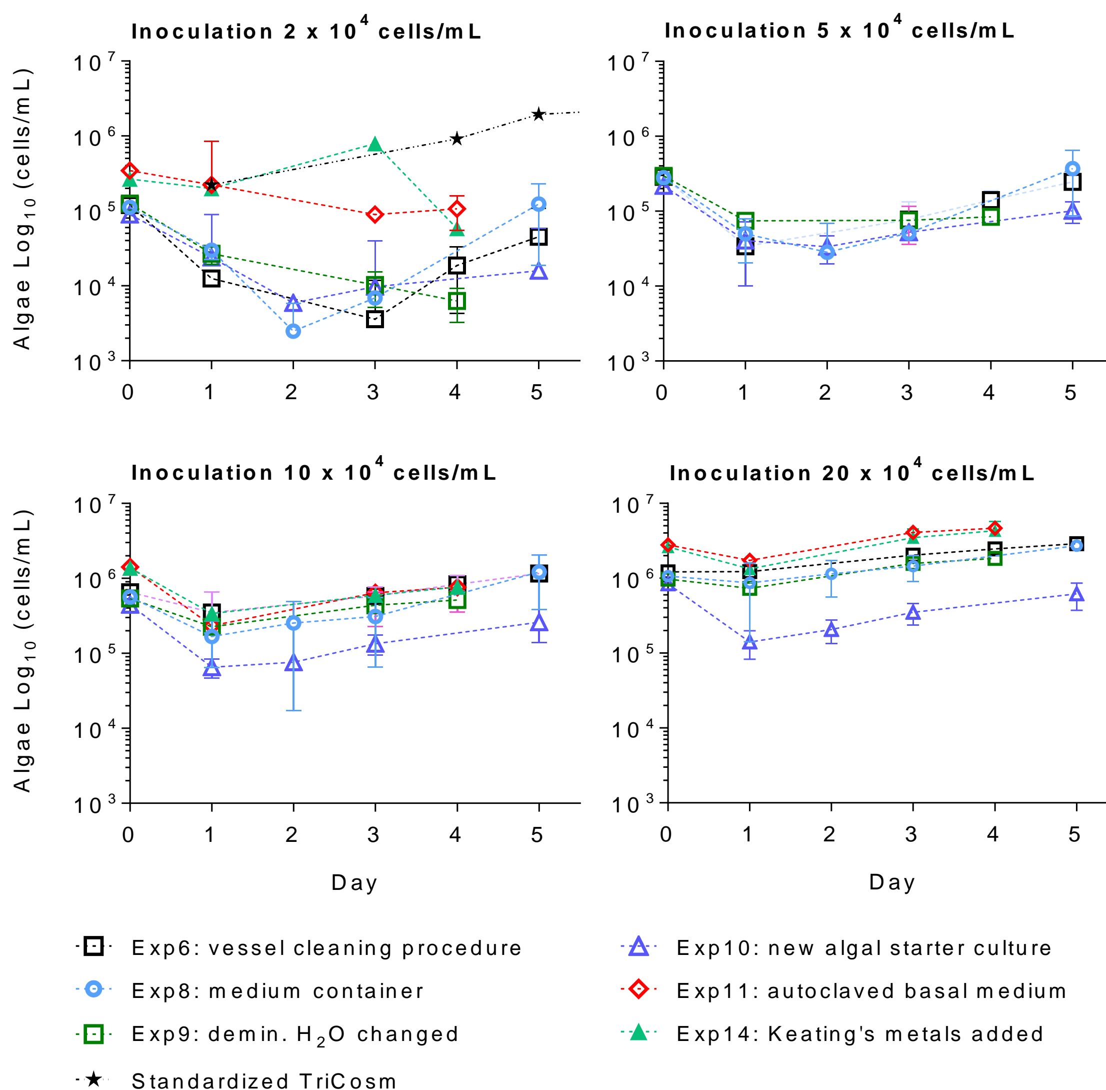

Figure 4 - 3: Algal population dynamics in experiments 6, 8 – 11 and 14 with differing inoculation concentrations. Labels indicate variations among experiments (for details see Table 4 - 1). Algal dynamics observed in the standardized TriCosm without grazing pressure are indicated with stars for reference. Shown are mean values (n = 3) ± 95% confidence intervals.

4.4.5 Algal population dynamics: medium buffering, aseptic preparation and aeration

Exponential algal growth was found irrespective of buffering salt (Na_2SiO_3) concentration (i.e. 0.23 g/L in Exp15, Exp16; 0.05 g/L in Exp17; 0.02 g/L in e.g. Exp19, Exp20) and in medium prepared aseptically (Exp16) and not (e.g. Exp17, Exp19,

Exp20; Figure 4 - 4). In experiments performed with highly buffered medium (Exp15, Exp16; 0.23 g/L), algal concentrations decreased within 24h after test start and remained low until test termination (Exp15) or grew exponentially to high final algal concentrations (Exp16). Similarly, in experiments performed with poorly buffered medium (0.02 g/L), exponential growth was measured in Exp19, Exp20 (Figure 4 – 4) whereas limited population growth was found in treatments with similar conditions in e.g. Exp2, Exp6, Exp8 - Exp10 (dish, inoculation 2 x 10^4 cells/mL and performed under standard light exposure (12 h - 16 µmol m^{-2} s^{-1}); Figure 4 - 2, Figure 4 - 3).

Algal concentrations generally declined after the test start for 2 - 3 days and recovered slowly thereafter, whereas populations in Exp16, Exp17, Exp19 and Exp20 increased exponentially throughout the test duration with similar concentrations to the standardized TriCosm populations (Figure 4 - 4). Other than in most experiments where the test medium was prepared aseptically and/or used shortly after preparation (i.e. not aerated), in Exp16, Exp17, Exp19 and Exp20, medium was aerated for ≥ 24h and used approx. 48h after medium preparation.

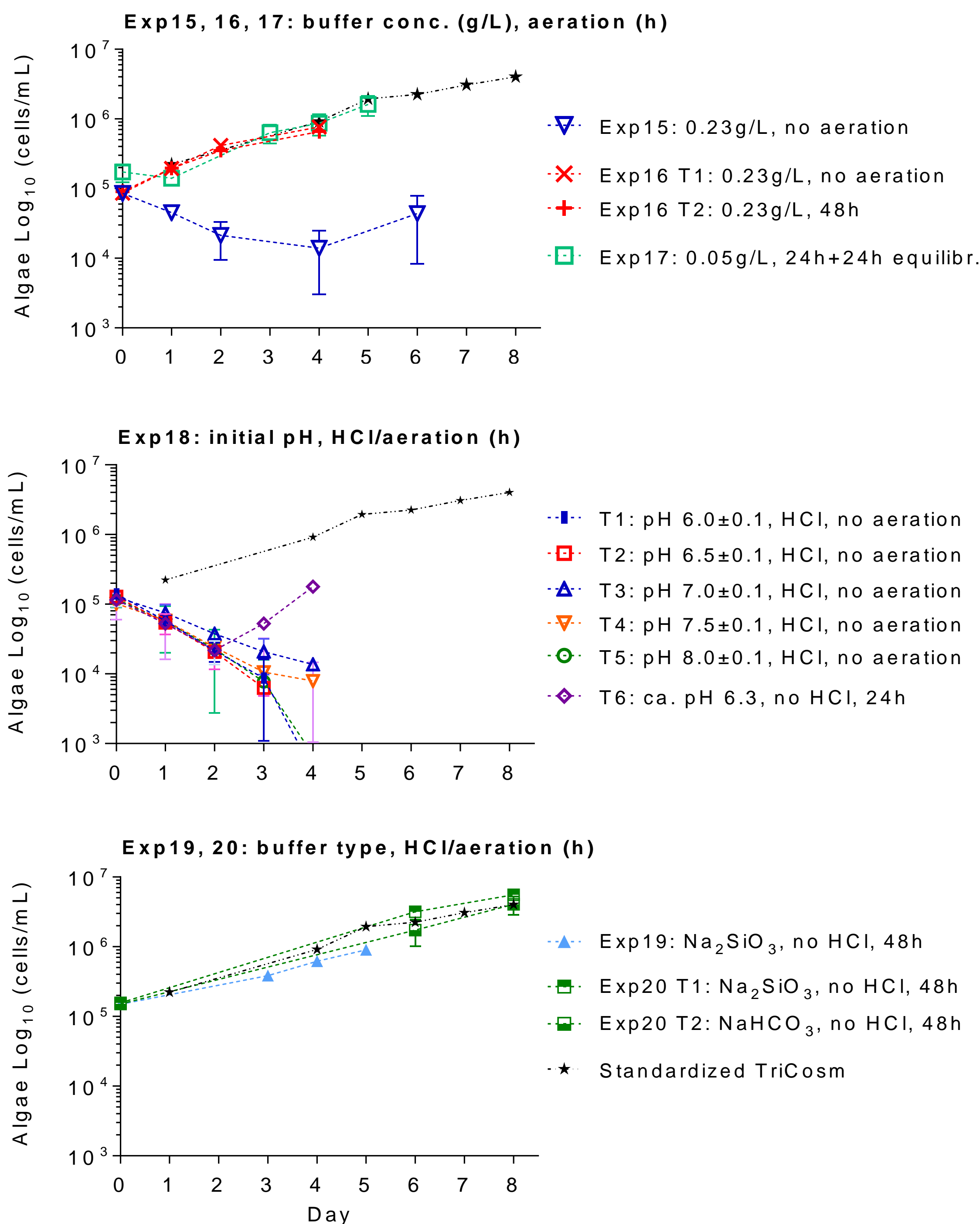

Figure 4 - 4: Algal population dynamics in experiments 15 - 20 with differing buffer salt concentration, type and/or medium pH adjustment via aeration (h) or HCl addition. Labels indicate factor variations among treatments and experiments (for details see Table 4 - 1). Algal dynamics observed in the standardized TriCosm without grazing pressure are indicated with stars for reference. Shown are mean values (n = 3) ± 95% confidence intervals.

4.4.6 Medium pH

Water pH at the end of the test duration was generally higher in treatments with higher algal concentrations (*Appendix C* Figure C - 5 to Figure C - 10). In media blanks, water pH decreased by an average of 2.2 pH units through 48h aeration following medium preparation (Exp16, Exp19) but increased throughout the experimental duration in these treatments (range: pH 7.24 - pH 8.50 at the end of the tests in Exp16, Exp 19, Exp 20; *Appendix C* Figure C - 9) whereas it decreased in experimental treatments with medium aerated for shorter time (range: pH 5.72 - pH 6.81 at the end of the tests in Exp4, Exp8, Exp9, Exp15, Exp18; *Appendix C* Figure C - 6, Figure C - 7, Figure C - 9 and Figure C - 10). In Exp18 (with exception of T6) where the initial water pH was adjusted to pH between pH 6.0 and pH 8.0 with HCl, pH and algal concentrations decreased within 24h after test initiation (*Appendix C* Figure C - 10).

4.5 Discussion

Repeatability and reproducibility are important elements to judge the reliability of empirical research. Growing evidence, however, indicates that some scientific research cannot be reproduced. In ecotoxicological research, this raises concerns considering its limited value to regulators who often use data from the peer-reviewed literature to inform and underpin regulatory decisions (Ioannidis, 2014; Harris and Sumpter, 2015; Attanasio, 2016; Hanson et al., 2017).

Standardized and reproducible multi-species tests with known population dynamics could be useful to generate ecologically relevant results for chemical safety testing or ecological research. Small perturbations to individual components can, however, affect the population dynamics of all species and the overall system (Riedl et al., 2018; Taub, 1997b). This means that when unexpected alterations of population dynamics are observed, a high number of different parameter combinations (depending on the experimental complexity) might have to be tested to determine the factor causing irreproducible results. In the TriCosm, where

variations of *P. subcapitata* populations triggered shifts in grazer and predator populations, a series of experiments was conducted to determine which factors could have impacted this basal layer of the tri-trophic system.

4.5.1 What caused alterations to population dynamics?

The combination of high shaking speed, light intensity, OECD medium and flasks as in the standardized algal growth inhibition test protocol (OECD, 2011) unsurprisingly lead to the highest population growth rate in *P. subcapitata*. However, high shaking speed creating strong water movement and OECD medium are not suitable for the animal species in the TriCosm. The importance of light for algal growth is expected and well evidenced (Singh and Singh, 2015), nonetheless, when the light intensity and photoperiod were optimized (Exp2 - 5), the algal concentrations remained generally lower than those measured in the standardized TriCosm. Experiments were performed in a controlled environment and excessive temperature fluctuations were excluded. As a result, we hypothesized that internal, medium related factors (e.g. chemical or biological properties) rather than external factors (e.g. light or temperature) impacted algal population dynamics.

Effects of reduced metal availability (Raven et al., 1999) or biotic (Grover, 2000; Ramanan et al., 2016) and abiotic (Pavlić et al., 2005) contamination of the test medium are further reported to influence algal population dynamics, but could be ruled out because the algal concentrations still remained considerably lower than in the standardized TriCosm when these factors were tested. For instance, when contamination impacts were assessed, algal populations increased exponentially solely in treatments with high but not with low inoculation concentrations. If abiotic contamination (e.g. due to dirty glassware or contaminated stock solutions) were the impacting factor then algal growth would likely be limited in all treatments and independently of starting densities. Population size dependent algal growth rather suggests biotic competition between algae and microorganisms due to impacts through competitive interactions (Ramanan et al., 2016). Finally, biotic contamination of this kind (e.g. due to contaminated algal starter stocks, originating

in aerated media or from insufficient removal of biofilms from test vessels) was, however, excluded because delayed population dynamics were observed even when experiments were carried out in newly purchased test vessels and the medium was prepared aseptically.

Only once the buffering salt concentration was increased were similar dynamics as measured in the standardized TriCosm observed. This was an indication of a low buffering capacity in the test medium and suggested that population dynamics had changed due to drifting pH values (Rendal et al., 2012). This hypothesis was confirmed when pH measurements were performed on a daily basis and, irrespective of the starting pH, values drifted below pH 7.0 within 24h (*Appendix C* Figure C - 9). Hence, it is likely that lower pH values at the end of the tests were, other than previously assumed, not only a result of lower respiratory and photosynthetic activity in small algal populations but, *vice versa*, smaller populations were a consequence of unfavorable media conditions.

While the medium was aerated before use when the TriCosm was developed and standardized, in many of the experiments here it was prepared in a manner intended to preserve sterility, and therefore it was not aerated to limit microbial contamination. Yet, a period of aeration increased water movement and possibly accelerated the chemical reactions between salt solutions, equilibrated the medium and consequently reduced pH fluctuations during the test. Aeration likely further promoted the transfer of CO_2 from the air to the medium, increased its inorganic carbon content (e.g. HCO_3^-) and decreased water pH naturally reducing the need for further pH adjustment with HCl. Increased water movement through aeration could also explain why larger algal concentrations were measured in flasks when compared to dishes. Different volume/surface ratios between vessel types possibly caused different shaking patterns, accelerated the CO_2 transfer and the creation of chemical equilibrium in the medium in flasks. Sampling bias between vessel types (e.g. caused by different cell distributions at low shaking speed) was excluded when measurements were compared between vessels before and after the water columns were mixed (*Appendix C* Figure C - 2). The measurement of the inorganic carbon

content could have helped determine whether algal growth was limited by this factor. However, in less buffered systems with low alkalinity, limited availability of inorganic carbon (e.g. HCO_3^-) would likely negatively affect treatments with high inoculation densities the most but we generally found higher algal growth with higher inoculation concentrations.

When the medium was subsequently prepared with low buffering salt concentration (as used in the TriCosm) and aerated prior to test initiation, algal populations showed similar dynamics as in the standardized TriCosm. Indeed, when algal dynamics were compared between treatments in experiments 1 to 20 with aerated and not aerated medium and similar test conditions to the standardized TriCosm (T82MV medium in crystallizing dishes, 2 x 10^4 cells/mL starting concentration and 65 rpm orbital shaking; *Appendix C* Figure C - 11), population growth was measured only in treatments with aerated medium. This means that the modification of medium preparation with the aim of reducing microbial contamination, i.e. the omission of medium aeration, was most likely the source of variability that lead to variations of algal dynamics and knock on effects on grazer (*C. dubia*) and predator (*H. viridissima*) population dynamics in the TriCosm. Hence, medium equilibration by aeration (48h) and close control of pH development over time is necessary to achieve reproducible population dynamics of the green alga *P. subcapitata* in the TriCosm.

4.5.2 What does this mean for the reproducibility of multi-species systems?

Our findings confirm that even minor changes in experimental procedures can have large impacts on system dynamics. In multi-species systems this means that perturbations to any system component are likely to affect the overall system reproducibility. The achievement of predictable system dynamics and subsequent regulatory needs are thus progressively challenging in systems with increasing complexity, because all components are interlocked by reciprocal cause-effect pathways (Poisot et al., 2015).

The standardized aquatic microcosm SAM (ASTM E1366-11, 2011) is a remarkable example of a multi-species system where inter-laboratory reproducibility was tested. Considerable effort was invested in standardizing and studying system behaviour prior to conducting exposure experiments, but different timing and magnitude of biological responses were measured among experimental controls in the host laboratory and three independent laboratories. Although the ecological experiments in the SAM were statistically closer within laboratories than between laboratories, similar enough patterns emerged to draw the same conclusions across laboratories (Taub et al., 1986; Taub, 1993). While variations in system dynamics likely have been caused by differing rearing history of the grazers between laboratories in the SAM (Taub, 1993), in the TriCosm the lack of repeatability was medium related. The same medium (T82MV) is used in the SAM and the TriCosm but while the medium is autoclaved and not aerated prior to use in the SAM, the medium was unstable without aeration in the TriCosm. No similar problems were reported during the inter-laboratory reproducibility test of the SAM (Taub, 1993; Lithgow et al., 2017).

Considering the effort needed to identify which factor directly impacted the base trophic layer with consequent indirect repercussion in an interconnected system such as the TriCosm indicates the crucial role of consistency among a multitude of experimental factors.

4.5.3 With so many ways of getting it wrong – how shall we get it right?

Multi-species systems are certainly resource demanding when reproducibility problems arise. As reported before (Hines et al., 2014; Lithgow et al., 2017) and demonstrated here, experimental investigations to find sources of variability can take considerable time and effort even in simple systems. In the case of the TriCosm, for instance, approx. one person-year was necessary to address the source of variation, although all experiments were performed in the same laboratory and by the same researchers that developed the standardized TriCosm.

In the light of the current debate regarding a ‘reproducibility crisis’ and the demand for the use of ecologically more relevant test systems in Europe, we stress that documentation, compliance to standardized procedures and repeatability testing are increasingly important with increasing system complexity. Knowledge of the dynamics and behaviour of system components is crucial to define the range of acceptable deviations, to develop rejection criteria and to create a basis for reproducibility. In the case of the TriCosm, for instance, advanced understanding of the water chemistry of the medium could have saved much time and effort. Information on typical pH ranges of standard test media is not only important for species performance, but are also a consideration when the effects of e.g. chemicals are assessed since the ionization and the toxicity of the test compounds might be changed by pH (Rendal et al., 2012).

Not all experiments can be performed in an ideal way (e.g. in standardized test setting and with well-established methods), and novelty might often be prioritized over reliability. Nonetheless, the quality of empirical research and its relevance to regulators could easily be increased (Ioannidis et al., 2014). For instance, if principles of sound ecotoxicology are followed (Harris et al., 2014) and the performance of reproducibility studies was incentivized by funders or during the peer review, then confidence in experimental outcomes would be increased. Non-selective documentation as well as the independent replication and reproduction of experiments are at least as important as the reporting of experimental results (Harris and Sumpter, 2015; Harris et al., 2017; Hanson et al., 2017).

By sharing this laborious experience in our search for repeatability, we aim to communicate that the dynamic tri-trophic system TriCosm may take more understanding and control than first assessed. Efforts to achieve the standardization and reproducibility of dynamic systems are, however, well worth pursuing. They have much to offer, granting insights to indirect chemical effects, advantages of statistical power, speed of study performance and above all, a better understanding of the integration of natural and anthropogenic (chemical) stressors.

4.6 Conclusion

Standardized laboratory, multi-species tests allow the cost-effective assessment of chemical impacts on species interactions, but system control and the prediction of population dynamics take considerably more understanding than in simple systems. The reproducibility of system dynamics can be greatly affected by small variations of the test protocol. In the case of the TriCosm, equilibration of the medium (by aeration) is critical to achieve pH stability and reproducible *P. subcapitata* population dynamics. Greater consideration must be given to reproducibility studies, non-selective reporting and publication of unexpected outcomes, both by funding bodies and during peer review to increase the transparency and value of empirical research to regulators.

4.7 Acknowledgement

The authors thank F. B. Taub for numerous discussions and her advice on medium preparation. This study was financed by The Environment Department, University of York, Heslington, York, UK, and the Centre for Chemical Safety and Stewardship (CCSS), Fera Science Ltd., Sand Hutton, York, UK. The authors have no conflict of interest to declare.

Chapter Five: General Discussion and Conclusions

5.1 Summary of thesis aims and results

In the presented PhD thesis I addressed one of the challenges for the improvement of the environmental risk assessment (ERA) of pesticides outlined by scientific working groups of the European commission (EC, 2013). I sought to gain and further the knowledge on how to assess the impacts of chemical and natural pressures on organisms and their ecological interactions in a repeatable manner.

In *Chapter One,* I reviewed the literature on the use and environmental risk of plant protection products and illustrate current approaches in place to identify environmental risk of pesticides with a focus on non-target organisms in the aquatic environment. My first aim was to identify knowledge gaps and limitations to give direction for research addressed in the following chapters. Generally, single-species tests are rapid, require less resources and results are commonly easier to repeat than system responses in multi-species systems. Micro- and mesocosms incorporate greater environmental realism and account for chemical impacts on species interactions at the population level but are time-, resource-, and effort-demanding. Scientific working groups of the European commission (SCHER, SCENHIR, SCCS) outlined that *'results should be replicable and reproducible with acceptable approximation'* to be valuable for regulators and fulfil the requirements for the applicability of current procedures for risk characterisation. Yet, integrating ecologically more relevant endpoints to routine tests of environmental risk assessment (ERA) of pesticides was also regarded as a relevant research need (Scientific Committee on Health Environmental Risks et al., 2013). In this context, I focused on addressing concerns regarding the lack of environmentally relevant endpoints such as species interactions in standardized and routinely performed tests in ERA (i.e. lower tier tests). I identified a lack of standardized systems of intermediate complexity that bridge simple single-species tests and complex microcosms. Such a system could help study the effects of toxicants on common

processes and be used to extrapolate effect/response relationships to more complicated systems with similar principles.

Based on the research need identified in *Chapter One*, the research aim in *Chapter Two* was the development of a new aquatic multi-species system for the repeatable detection of pesticide effects on population dynamics and species interactions. I designed the TriCosm as a rapidly cycling, small-scale laboratory microcosm and tool to assess chemical effects on consumer-resource interactions across three trophic levels during short experimental duration (21 days). The tritrophic system comprises a producer, herbivore and carnivore species (*Pseudokirchneriella subcapitata*, *Ceriodaphnia dubia*, and *Hydra viridissima*) with rapid generation times to obtain quick population responses to system alterations. Population dynamics in the TriCosm generally changed following interactions with other system components and the middle trophic layer (*C. dubia*) was directly affected by food availability (algal concentrations) and predation and indirectly transmitted effects to top and bottom levels. Bottom-up and top-down processes both likely regulated population dynamics in the TriCosm. The population trajectories, interaction strength and coefficients of variation were strongly sensitive to starting conditions (species addition timing and quantity) but we found good repeatability of population dynamics when the setup procedure with the smallest overall coefficient of variation (CV) was repeated. An average CV of 19.5% compared to an average CV of 32% in small indoor systems and larger CVs in studies involving animals (Sanderson, 2002) indicated the ability to detect treatment-related system alterations. Calculated minimum detectable differences (MDDs) for critical endpoints between controls and the repeated systems were similar when the system was first developed and when its repeatability was assessed in the short term (after approximately three months). The calculated MDDs were assigned to MDD classes III (50–70%) and IV (<50%) indicating the ability to detect small and medium-sized effects, respectively, on the trajectories of interacting populations once the TriCosm is exposed to chemicals (European Food Safety Authority, 2013). *Chapter Two* demonstrated that population dynamics can be standardized in simple multi-species

systems and allow a cost-effective and statistically powerful approach to the repeatable assessment of chemical risks on species interactions. *Chapter Two* was published in Environmental Toxicology and Chemistry in 2018 and the supporting data are provided as published in *Appendix A*. This Chapter formed the basis for *Chapter Three* and *Chapter Four*.

In direct response to *Chapter Two*, *Chapter Three* aimed at testing whether subtle chemically-induced effects could be detected on species interactions and if control system dynamics in the TriCosm were repeatable in the long term. In six independent experiments at regulatory acceptable and environmentally relevant concentrations, system dynamics were sensitive to effects on ecological interactions. I measured bottom up effects in all experiments i.e. alterations of algal population dynamics indirectly regulated *Hydra* populations through the mediating intermediate grazer level. Low food availability likely impacted *C. dubia* neonates and controlled grazer population growth keeping *Hydra* populations too small to trigger top-down processes. *P. subcapitata* populations were unlikely affected by direct effects of the herbicide linuron in five of six experiments but system responses were likely affected by changes in medium preparation as was further addressed in *Chapter Four*. Experiments were performed with different concentrations of buffering salt (Na_2SiO_3) in the medium and at highly buffered and more stable medium in terms of pH, algal population dynamics closely resembled those measured in the standardized TriCosm. Still, grazer and predator population growth differed to dynamics measured in the standardized TriCosm (*Chapter Two*) and morphological changes of individuals, e.g. *Hydra* tentacle damage impeded feeding, helped understanding causal mechanisms behind observed population patterns. Variations between algal populations in control and treated systems likely due to herbicide effects were found only in one experiment with low buffering capacity but aerated test medium. Smaller algal populations in treatments might have led to an indirect effect of linuron on *C. dubia* populations through food limitation but even though algal and *C. dubia* populations increased in the controls, populations were considerably smaller compared to those when the system was standardized (*Chapter Two*). My results

were highly inconsistent and, for this reason, no statistical analyses were performed to compare control and treated system dynamics for differences in species interactions. False conclusions on the environmental safety of linuron at the concentrations employed for the tests would have likely been drawn had I only considered one out of six experiments. In hindsight, most of the differences between treatments and experiments could be explained by small, but important, differences in the experimental protocol. Inter-experimental variation in *Chapter Three* gave reason for further research as outlined in *Chapter Four* aimed to investigate and understand which experimental factor had to be controlled more carefully and caused the variations in TriCosm system dynamics. Supporting Data for *Chapter Three* is provided in *Appendix B.*

In direct response to *Chapter Three*, *Chapter Four* aimed at determining the factors leading to inter-experimental variability and at refining the range of conditions under which the TriCosm population dynamics can be reproduced. I focused on the assessment of effects of twelve experimental factors (test medium, vessel type, shaking speed, light intensity, light regime, vessel condition, inoculation density, medium preparation components, metal concentration, metal composition, buffering salt type and concentration) combined to 60 treatments on *P. subcapitata* growth. The focus was specifically on the base trophic layer because it directly determines the dynamics of the grazer and predator populations via bottom-up effects during early experimental stages in the TriCosm. When the growth rates of algae in all treatments were compared, algal growth rates were higher in treatments with increased shaking speed, light exposure, medium buffer, or aeration time. *P. subcapitata* populations in treatments with aseptically prepared, scarcely buffered and/or shortly aerated medium conducted in dishes, generally, showed low algal growth rates. The shaking speed, light exposure, vessel type were not changed between the standardized TriCosm (*Chapter Two*) and linuron exposure experiments (*Chapter Three*). Hence, my findings strongly indicated that inter-experimental variations and different algal growth trajectories among control systems in experiments shown in *Chapter Two*, *Chapter Three* and *Chapter Four* were caused by

the modification of medium preparation. The medium was aerated thoroughly before use when the TriCosm was developed whereas in many linuron exposures and experiments in this *Chapter Four*, the medium was not aerated with the aim to limit microbial contamination. We hypothesised that the scarcely buffered test medium was likely equilibrated by aeration (48h) and promoted CO_2 transfer from the air to the medium which possibly decreased water pH naturally, reduced later pH fluctuations and limited the need for further pH adjustment. *Chapter Four* evidenced that the aeration of the medium and close control of pH development over time is critical to achieve reproducible population dynamics of the green alga *P. subcapitata* in the TriCosm. In relation to current concerns over a 'reproducibility crisis' in scientific research, *Chapter Four* highlighted the crucial role of consistency among a multitude of experimental factors to obtain reproducible system dynamics and ecological interactions even in apparently simple systems. *Chapter Four* was submitted and revised for publication in Environmental Toxicology and Chemistry and supporting data is provided in *Appendix C*.

5.2 Limitations and future outlook

Throughout this thesis, I described that the TriCosm could be a valuable tool to assess chemical effects on species interactions and help us gain mechanistic understanding on ecological processes in larger systems. However, the validation and testing of the multi-species system still requires more attention and a series of limitations, including ecological, practical and conceptual aspects should be acknowledged and/or addressed before it can be used. In the following section, I will describe limiting factors and elaborate on future research opportunities for the improvement and application of the standardized microcosm in ERA.

Ecological limitations – The TriCosm consists of an alga-grazer-predator community and effects on the population dynamics of one species per trophic level can be assessed. However, it was not assessed how the system, and in particular the

algal trophic level, was influenced by associated microbes. As outlined in a recent review by Ramanan et al., (2016), bacterial communities play an important role in algae-bacteria interactions, in the chemical cycle of e.g. nutrients and in defining primary productivity in many ecosystems. Bacterial communities might interact with algae in a variety of modes, from impeding to enhancing algal growth and the lack of knowledge on which bacteria were present in the TriCosm and how these might have influenced primary production is, hence, an important shortcoming of this system. The TriCosm is not closed and the biota is open to exchange with the surrounding environment. This means that microbial communities present in the laboratories might colonize the systems and influence dynamics of the algal trophic layer and, in turn, the grazer and predator populations. This might have implications for the reproducibility of the TriCosm between experiments and among laboratories. The non-closed nature of the system allowing atmospheric exchange further does not allow the study of chemical cycles such as, for instance, how photosynthesis and respiration regulate CO_2 and O_2 in an aquatic community that is limited to starting resources and closed to external atmospheric exchange (Taub and McLaskey, 2013).

The low complexity of the TriCosm when compared to other microcosms or natural environments further excludes the existence of a steady-state community, and impacts on resilience or system shifts cannot be assessed. The main aim of the project was to develop a standardized multi-species system and the number of species had to be reduced to keep the system as simple as possible but to still include ecological interactions. A trade-off between increasing the number of species on the same trophic level and the total number of trophic levels was made. With the inclusion of three trophic levels, indirect chemical effects via bottom up and top down effects can likely be assessed but with only one species per trophic level, inter-specific competition cannot be detected and indirect effects might be more pronounced in the TriCosm than in structurally more complex microcosms due to a lack of functional redundancy. In microcosms with only one species per trophic level, small impacts on one population will quickly trigger shifts in interactions between the species on other trophic levels because the lost functions performed by impacted

organisms cannot be replaced. In ecologically more realistic systems, higher species diversity allows the existence of functional groups (i.e. several organisms perform similar ecosystem functions) and chemical impacts will likely be concealed at first by compensatory mechanisms of functionally redundant species. Hence, the effects on other trophic levels might be dampened and the functioning of the system as a whole might not be compromised because more resistant species with similar ecosystem functions may become more abundant and compensate for the role of lost sensitive species (Duffy et al., 2007; Daam et al., 2009).

Practical limitations – The population dynamics in the TriCosm can be rapidly determined for algae via in-vivo fluorescence analysis and the conversion of fluorescence intensity to cell densities by means of a calibration curve. Yet, the analysis of grazer population dynamics is rather time intensive. Small (ca. 0.3 - 1 mm), semi-transparent individuals of *C. dubia* move in a three-dimensional space and counting them is exhausting for the eyes. In an experimental setup with negative and solvent controls, two treatments and eight replicates each, the operator has to count grazer populations at least 64 times (4 x 8, counted twice). This roughly adds up to three and a half hours of continuous counting (4 min/count) and populations frequently have to be counted more than twice when populations are numerous. The development of automated measurements with e.g. image analysis would thus be more time-efficient and the measurements could be repeated and/or performed at a later time. With automated image analysis not only animal numbers but possibly also individual body size could be measured and chemical effects could be assessed on other endpoints, e.g. animal growth.

The test medium T82MV employed for TriCosm tests was designed with low buffering capacity (ASTM E1366-11, 2011). Even though chemical equilibrium can likely be reached once the medium was aerated for 48h as described in *Chapter Four*, fluctuations in water pH might affect the sensitivity of species to toxicants (Belanger and Cherry, 1990; Berezina, 2001; Fettweis et al., 2018) and the toxicity of test substances to test species (e.g. ionisable compounds; Anskjær et al., 2013). Refining

the standardized test medium T82MV to obtain a more constant and known pH range would thus limit confounding effects of the water chemistry and reduce the magnitude of a variation needed to be detectable as an effect on population dynamics.

The TriCosm is a small static system with no medium renewal throughout the experimental duration. In the case when algal populations are temporarily released from grazing pressure (e.g. via feeding inhibition from an insecticide), algal blooms could lead to overgrown systems limiting the transparency of the water column and impede the assessment of grazer and predator population dynamics. Fluctuations in animal numbers could thus hardly be determined and population abundances assessed only once the experiment is terminated.

Personal limitations/reflection – To achieve the aim of the project, i.e. the development and testing of a reproducible and standardized multi-species system, profound understanding on the interconnectedness of system components and the sensitivity of system dynamics to small experimental alterations is necessary. A limitation of the project might thus have been its conceptualization as a PhD project because, as such, it had a dual aim: creating new knowledge and training a researcher. New information furthering our understanding on the standardization of multi-species systems was created and limitations regarding the experimental design were described in previous chapters. My skills and knowledge were expanded and I developed considerably as a researcher. Nonetheless, the testing and application of the TriCosm would likely be at a more advanced stage, had the project been executed by a more experienced researcher to start with. When the TriCosm was developed, I considered factors such as the culturing history of test species (e.g. reproductive output, exact number of animals per beaker, animal life span, brood numbers and neonates per brood) less important and invested little time in the understanding of chemical characteristics of the test medium. Hence, disregard of, in hindsight, important factors lead to much time and effort required to determine the reasons

for inter-experimental alterations and might still be needed before the TriCosm can be used in the future.

Other limitations – Although we likely determined the source of inter-experimental variation among TriCosm control dynamics (i.e. impacts of chemical instability in the test medium on algal population dynamics causing knock on effects on animal populations), the TriCosm was not repeated to confirm this assumption. The ease by which the system dynamics can be reproduced by other operators and the consistency of animal counts between operators was not estimated. Additional TriCosm experiments within and among laboratories with and without chemical exposure are thus needed to evaluate i) the reproducibility of TriCosm control dynamics and ii) the sensitivity of the TriCosm to detect small effect sizes as predicted in *Chapter Two*.

Experiments should address new knowledge regarding the medium preparation procedure and attention must be paid to pH stability throughout the experimental duration. The documentation of the culturing history, individual performance (e.g. reproductive output, survival) of grazers and predators and continuous monitoring of the water parameter pH can add key information to delimit acceptability ranges of system components before experiments are started and for system dynamics throughout the experimental duration. Additional information can consequently be used to complete and refine the Standard Operating Procedure for the TriCosm as developed in *Chapter Two* (*Appendix A*) and facilitate the use of the system by other researchers. Experimental data should be compared within and between laboratories to evaluate the reproducibility of system dynamics and possibly identify procedures requiring improvement or standardization. Data on pesticide exposure could further be used to determine the concrete range of minimum detectable differences (MDD's) between controls and treated systems. In *Chapter Two*, I calculated theoretical MDD's under the assumption of similar variances between control and treated systems, however, variances could decrease, increase or remain similar in treatments (Kraufvelin, 1998; Sanderson, 2002).

Future application and research opportunities – Once the reproducibility of the TriCosm is optimized, it could be used, for instance, at an intermediate tier in the environmental risk assessment of pesticides. Similarly, to single-species tests, there is great uncertainty involved if concentration-response relationships observed at the population levels in the TriCosm were extrapolated to realistic field conditions but it could allow a rapid evaluation of chemical effects on species interactions and be employed to confirm the protectiveness of regulatory acceptable concentrations derived from single species tests. Nonetheless, we mainly envisioned its application as part of a future approach to environmental risk assessment of pesticides i.e. for the development and testing of ecological effect models.

When the amount of possible combinations between active substances, biotic and abiotic factors interacting in natural systems is considered, the number of experiments necessary to assess chemically induced effects via factorial design becomes unfeasibly high. For this reason, ecological modelling techniques are employed so that a broad range of exposure scenarios can be addressed without the need to empirically test all of them in microcosm studies. Mechanistic effect models can link different factors (e.g. environmental complexity, species physiology and interactions), help with the interpretation, understanding and extrapolation of processes and allow the prediction of probabilities for unacceptable effects to occur (Hommen et al., 2010).

Applied to the pesticide regulation, suitable modelling approaches could possibly replace or complement current practices for the derivation of the RAC (EC_x are divided by AFs; EFSA, 2013). The integration of mechanistic effect models to ERA has much potential to decrease the uncertainty related to ERA, increase its ecological relevance and make it more cost-effective, comprehensive and consistent (Hommen et al., 2016). Yet, ecological effect models are not an integral part of the regulatory risk assessment of pesticides. This is mostly because the number of properly tested and validated models for regulatory purposes is still limited but their use is receiving growing attention especially for the risk assessment of pesticides under the European

Directive 91/414/EEC (European Commission, 2002). Ecological effect models often lack validation with different data sets to assess their prediction ability under new conditions and are often tested based on their capacity to discriminate between responses rather than the extent to which they estimate accurate probabilities (Chivers et al., 2014).

The Tricosm could be conceived as complementary to a conceptual and computational ecological model representing a tri-trophic microcosm. Repeatable population responses to combined stressor exposures (e.g., toxicants, predation, and/or food fluctuations) in the Tricosm could be used to facilitate both the development and the testing of mechanistic effect models. Measured community responses in terms of individual abundance changes and population trajectories could be employed for the calibration and parameter fitting of ecological models used to extrapolate effects to different ecological and environmental scenarios. In turn, chemical effects on interactions within a simple freshwater community can be measured and quantified in the TriCosm and provide empirical benchmarking to estimate and test model prediction accuracy and power. Other than for pesticides, the application of the TriCosm for the development and testing of predictive models could be extended to biocides, pharmaceuticals and industrial chemicals to gain information on recovery processes and possible indirect effects in communities.

An interesting application of the experimental TriCosm and an associated ecological effect model could be the assessment of ecological relevance of effects determined at the individual level for the prediction of population level effects. In other words, the TriCosm model could be parameterized with effect data determined at the individual level (e.g. survival and reproduction of *C. dubia* in standard single-species tests) and model predictions be compared to experimental results in the microcosm. If and to which extent the population dynamics in the simple tri-trophic system were predicted by the model could then be evaluated and indicate which processes (e.g. species interactions) must be accounted for in ecological models and are important to assess the uncertainty of model predictions. Direct effects on endpoints at the individual level are often determined by a variety of pathways not

included in the model. The endpoint itself often includes responses of interacting processes and, for instance, reproduction is influenced by the availability of food, depends on the nutritional history of organisms and on how energy is allocated (e.g. to metabolic maintenance, growth and/or reproduction). The reproductive output and population dynamics will, hence, likely change with effects on any of these processes (Agatz and Brown, 2013). For instance, Martin et al., (2014) found that effects of hypothetical chemicals with different modes of action on *D. magna* reproduction determined at the individual level are not sufficient to predict toxicity at the population level in experimental systems. However, when the authors parameterized an individual based model rooted in dynamic energy budget theory with data on individual growth and reproduction at four food levels, growth rates and peak densities of *D. magna* could be closely predicted. However, the decline phase could not be assessed until assumptions on food-dependent mortality of juveniles were included in the model (Martin et al., 2013). Preuss et al., (2010) parameterized an individual- based *Daphnia magna* population model with 3,4-dichloroaniline effect data on mortality and reproduction at the individual level and found similar population dynamics when predicted results were compared to data derived from population experiments.

Although the extrapolation of effect data at the individual level is often successful to predict dynamics at the population level, few studies exist where predictions from ecological effect models calibrated with single-species tests are compared to population dynamics in empirical systems with multiple trophic levels. Swartzman et al., (1989), for instance, modelled a microcosm community with several algal and grazer species and compared model predictions with experimental data after exposing the microcosm to streptomycin with model predictions. The authors found satisfactory overall agreement between the model and data, however, the parameters were calibrated based on single species, paired species and multiple algal species (without grazing) growth tests and the authors, hence, accounted for effects of competitive interactions on algal growth. Fettweis et al., (2018) recently reported good accordance between empirical results in a microcosm with one algal

and one grazer species and predictions from a model calibrated solely with data obtained in standard species tests.

The results documented by Fettweis et al., (2018) and a combination of individual effect data combined with modelling techniques in general, appears promising and can help interpreting responses and underlying processes in simple microcosms. With regard to current approaches to increase the ecological realism of ERA of pesticides, ecological effect models will likely play an integrative part in the future. Extensive experimental setups would require much time and resources and more research on the predictive power of multi-species models calibrated solely with single species test data is still needed.

5.3 Conclusion

To the best of my knowledge, the TriCosm is the only aquatic, standardized microcosm including macroscopic invertebrate species with short reproduction times and documented repeatability (in the short term) of population dynamics across three trophic levels. As such, my research has advanced the field of ERA because it demonstrated that direct and indirect impacts on system processes can be measured in a cost-effective and statistically powerful approach increasing the ecological realism compared to single-species tests. System dynamics in multi-species systems are generally highly sensitive to experimental conditions and their control take considerably more understanding and effort than is evident at first. Careful documentation and consistency of experimental procedures, and thorough knowledge about system components are required for the achievement of repeatability even in simple systems. In the light of the current demand for standardized and repeatable systems with ecologically more relevant endpoints in Europe, I conclude that the development or refinement of test media, suitable for microcosms with multiple organism groups required for this purpose, is lagging behind. Similarly, incentives for the performance and publication of reproducibility

studies addressing a ‘reproducibility crisis’ of empirical research could decrease concerns over the value of empirical data to regulators.

Much work is still needed to understand all system components (e.g. medium chemistry) of the TriCosm and its reproducibility within the same and among different laboratories must still be improved and tested. Yet, empirical data presented in this thesis suggests that it is a useful system worth further pursuit with much potential for future application. As such, repeatable measures of chemical effects on population dynamics could be employed for the parameterization, testing and validation of mechanistic effect models and to assess the uncertainty of model predictions when variables from endpoints at the individual level are extrapolated across three trophic levels. A comparison between empirical and modelled microcosm behaviour can help identify which processes require inclusion in a simple tri-trophic food chain model, discover gaps in current understanding of individual-level interactions and promote the development of theory.

Appendices

Appendix A – Supporting Data for Chapter Two

A standardized tri-trophic small-scale system (TriCosm) for the assessment of stressor induced effects on aquatic community dynamics

Published in Environmental Toxicology and Chemistry (2018), 37, 1051–1060

Verena Riedl* Annika Agatz, Rachel Benstead, and Roman Ashauer

** Address correspondence to* verena.riedl@york.ac.uk

Material and Methods: Preparation of the test medium

Table A - 1: Chemical composition, preparation and storage of stock solutions for the TriCosm medium (ASTM E1366-11, 2011).

Stock	Compound	g/L	mL	Preparation and storage
1	$NaNO_3$	21.25		Autoclaved at 121 °C for 1 h; stored in the fridge (ca. 4 °C).
2	$MgSO_4$ $7H_2O$	12.33		
3	$CaCl_2$ $2H_2O$	73.50		
4	NaCl	43.80		
5	$Al_2(SO_4)_3$ $18H_2O$	1.60		
6	Na_2SiO_3 $9H_2O$	11.38		0.22 µm membrane filter sterilized and stored in the fridge (ca. 4 °C).
7	NaOH	0.640		
	KH_2PO_4	2.72		
8 Metals	Solution 8a		250	Stored in the fridge (ca. 4 °C).
	Solution 8b		500	
	Solution 8c		60	
	Deionized water		190	
8a	NaOH	10.70		Solution should be allowed to cool after each step. Preparation as follows: - 2.80 g NaOH are dissolved in 260 mL of deionized water - Addition of 26.1 g EDTA - Addition of 24.9 g FeSO4 7H2O Aeration overnight, used for stock 8.
	EDTA (Ethylenedinitro tetraaceticacid)	26.10		
	$FeSO_4$ $7H_2O$	24.90		
8b	H_3BO_3	1.85		Used for stock 8.
	$ZnSO_4$ $7H_2O$	0.290		
	$MnCl_2$ $4H_2O$	1.98		
	Na_2MoO_4 $2H_2O$	0.240		
	$CuSO_4$ $5H_2O$	0.045		
	$Co(NO_3)_2$ $6H_2O$	0.029		
8c	NaOH	12.00		Solution should be allowed to cool prior to EDTA addition. Used for stock 8.
	EDTA	29.00		
9 Vitamins	NaOH	~0.2		Before Vitamin addition, NaOH is dissolved in 1L of deionized water; 0.22 µm membrane filter sterilized and stored in freezer (ca. -20 °C in 2 - 5mL tubes).
	Calcium pantothenate	1.750		
	Thiamin (B1)	0.150		
	Riboflavin (B2)	0.100		
	Nicotinamide	0.325		
	Folic acid	0.825		
	Biotin	0.075		
	Choline	1.250		
	Inositol	2.750		
	Pyridoxine (B6) monohydrochloride	1.250		
	Solution 9a		1	
9a	Cyanocobalamin (B12)	0.075		Used for stock 9.

Table A - 2: Volumes (mL/L) of stock solutions needed for the preparation of the TriCosm medium (ASTM E1366-11, 2011).

Stock	Addition volume (mL) per litre of deionized water
1	2
2	2
3	2
4	2
5	2
6	2
7	2
8 – Metals	0.05
9 – Vitamin solution	0.4

The Medium is aerated for ca. 24 h and, if necessary, adjusted to pH 7.2 ± 0.2. The pH value should not drop below pH 7.0 and be monitored throughout the test.

Table A - 3: Plate reader settings for in-vivo fluorescence analyses.

Plate Reader	
Model:	Tecan® Infinite 200PRO
Plate Description:	[GRE96ut] - Greiner 96 U Transparent
Range:	B2:F11
Shaking	
Duration:	30 sec
Mode:	Linear
Amplitude:	1 mm
Frequency:	886.9 rpm
Fluorescence Intensity	
Excitation Wavelength:	430 nm
Excitation Bandwidth:	20 nm
Emission Wavelength:	670 nm
Emission Bandwidth:	25 nm
Reading Mode:	Top
Lag Time:	0 µs
Integration Time:	20 µs
Number of Reads:	25
Settle Time:	0 ms
Gain:	Manual
Gain Value:	59
Mirror:	Automatic
Mirror:	Dichroic 510 (e.g. fluorescein)

Material and Methods: Protocol on materials, procedure and measurement to operate the TriCosm

Equipment for two treatments and control (8 replicates each)

- Orbital shaker (accepting loads ≥ 15 kg; platform dimension to hold ≥ 24 crystallising dishes (500 mL))
- Dissolved oxygen meter
- pH-meter
- Equipment for the control of the light/dark regime
- Equipment for the measurement of the light intensity
- Equipment for the temperature control
- Plate reader (with function to detect in-vivo fluorescence intensity)
- 24 crystallizing dishes (500 mL)
- 24 watch glasses (diameter 125 mm)
- 60 polystyrene 96 well plates (flat bottom, transparent with lid)
- 4 polystyrene 12 well plates (flat bottom, transparent with lid)
- Pipettor (for volumes of 200 µl)
- Pipette tips
- Cling film
- Light box
- 2 transparent clear PVC sheets
- Handheld counter
- Deionized water
- *Pseudokirchneriella subcapitata*
- *Ceriodaphnia dubia*
- *Hydra viridissima*

Procedures for Test Setup

<u>Conditions of exposure</u>

Test medium.

The test medium T82MV (as in ASTM E1366-11 2011) should be prepared 1 - 2 days before the test initiation, aerated for 48h and adjusted to pH 7.0 ± 0.1.

Test vessels and duration.

The crystallizing dishes are soaked for at least 24h in deionized water to remove detergent residues and labelled. On day 0 they are filled with 500 ± 5 mL T82MV and positioned on the shaker. The shaker platform can be wrapped with cling film to guarantee a good adhesion of the dishes. The replicates are evenly distributed on the platform and they are rotated every other day to equalize the lighting conditions between replicates. The orbital shaker is set to a shaking speed of 65 rpm and only stopped for observations or measurements. The test systems are monitored for 20 days.

Test organisms, addition sequence and densities

Pseudokirchneriella subcapitata.

Addition day: Day 0

Initial concentration per replicate: 2.0 x 10^4 cells/mL

Total number for 2 treatments and control: 24 x 10^7 cells

P. subcapitata is added to the replicates on day 0 to obtain an initial cell concentration of 2.0 x 10^4 cells/mL per replicate (approx. 10 x 10^6 cells in 500 mL). A culture of *P. subcapitata* is set up 3 - 4 days prior to test initiation to obtain high concentrations of viable algal cells. Once the exponential growth phase is reached, the *P. subcapitata* culture is used for the experiments. Cultures with different initial densities are prepared to guarantee the availability of healthy algal cells for the experiment start. Cultures with higher algal concentrations might reach the death phase early and cannot be used for the experiments. The culturing procedure for *P. subcapitata* has been adapted after OECD (2011) (see *Culturing procedures*). The volume of algal culture to be added to each replicate is calculated as follows:

$$\mathrm{V1} = \frac{(\mathrm{C2\ x\ V2})}{\mathrm{C1}} \qquad \mathrm{V1} = \frac{(2.0\ \mathrm{x}\ 10^4\ \mathrm{cells/mL\ x\ 500\ mL})}{\mathrm{C1}}$$

Where:

$C1$ = cell concentration (#/mL) in the algal culture

$V1$ = volume (mL) of algal solution to be added to each replicate

$C2$ = initial cell concentration (#/mL)

$V2$ = volume (mL) of test medium per replicate

Ceriodaphnia dubia.

Addition day: Day 1

Initial number per replicate: 10 neonates (age ≤ 24h)

Total number for 2 treatments and control: 240 individuals

C. dubia neonates (age ≤ 24h) from the third or fourth brood of cultured females are added to the replicates (10 individuals each) on day 1. The culture is set up at least 14 days prior to the test initiation and new animal cultures are set up with neonates from mature females each week to renew the culture and keep animals of known age (see *Appendix* 1). A feeding regime is followed and the culturing medium is prepared as in EPA, Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms (USEPA, 2002). Cultured females (age 12 - 17 days) are separated from their brood on day 0, fed with algae and YCT and half of their culture medium is renewed with T82MV. On day 1, the neonates (aged < 24h) are collected and used for the experiments.

Hydra viridissima.

Addition day: Day 6

Initial number per replicate: 3 juveniles (age 1 – 2 days)

Total number for 2 treatments and control: 72 individuals

H. viridissima juveniles are added on day 6 and each system receives 3 individuals (age 1 - 2 days, without buds). The culture is set up at least 14 days prior to the test initiation and a culturing procedure is followed (see *Appendix* A). On day 4, 72 individuals with buds are placed individually into 12 well plates (prior soaked in deionized water and filled with T82MV) and fed with *Artemia salina*. On day 5 and on day 6, the neonate *Hydra* are removed and individuals aged 1 day are used for the

tests. Individuals aged 2 days are used if less than 240 individuals with age 1 day are available.

Measurement and Analysis

P. subcapitata

The algal concentration is measured on day 1 or earlier and at least every second day throughout the test duration. For the analyses, 5 subsamples of 200 µl each (1 mL total) are taken while the replicates are still positioned on the rotating shaker. The water column is sampled homogeneously without sucking deposited algae from the bottom of the vessel. The subsamples are pipetted individually into 96 well plates and analysed with a Tecan® Infinite 200 PRO plate reader (plate reader settings see Table A - 3). The in-vivo fluorescence intensity of chlorophyll *a* is measured and the median value between the 5 subsamples per test vessel is converted to cell concentration (cells/mL) with a calibration curve.

C. dubia

C. dubia individuals are counted from day 4 and three times per week (e.g. Monday, Wednesday and Friday) throughout the test duration. To ease the count, a cross is drawn on a PVC sheet and positioned under the test vessel for each count on the light box. With the counting area divided in 4 quadrants, the number of *C. dubia* can be counted systematically. The number of both adults and juveniles is counted manually twice and if the difference of the counts exceeds 20% of the lower value, the count is repeated. The two counts are averaged for both adults and juveniles and the total *C. dubia* number is computed as the sum of both averages. Juveniles and adults are distinguished based on the presence/absence of eggs and the size of the animal (larger/smaller than an individual aged 5 days).

H. viridissima

Individuals of *Hydra* are counted from day 8 and three times per week (e.g. Monday, Wednesday and Friday) throughout the test duration. The sessile individuals are counted manually on a light box and the total number of individuals, the total

number of buds on all individuals and the overall morphology score are recorded. A morphology score ≥ 6 is considered reversible and sub-lethal while scores ≤ 5 are lethal (Quinn et al., 2012).

Water parameters

The dissolved oxygen content (DO, mg/L) and the pH value are determined at the start and at the end of the test and twice per week. The room temperature and the light regime are monitored constantly throughout the test duration.

Results: Correlation total/suspended algal concentration, coefficients of variation, culturing procedures and schedules

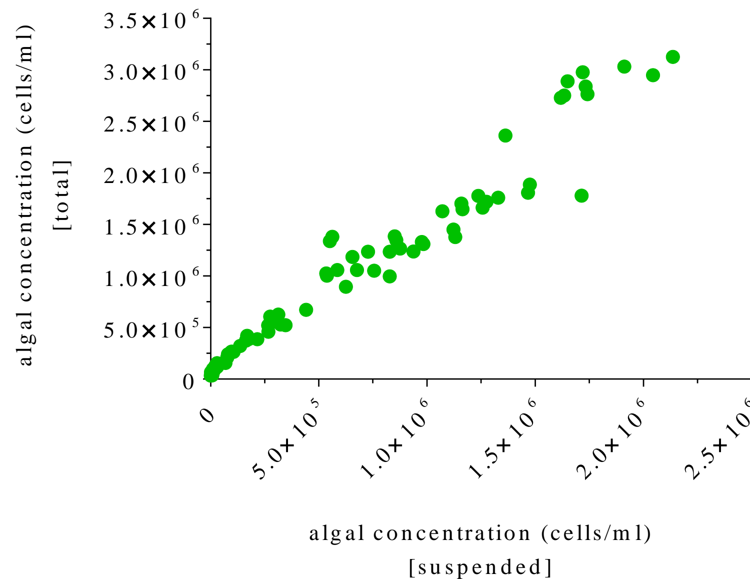

Figure A - 1: Spearman correlation between suspended and total algal concentrations (cells/mL) ($r_S = 0.98$, $p < 0.01$, $n = 90$).

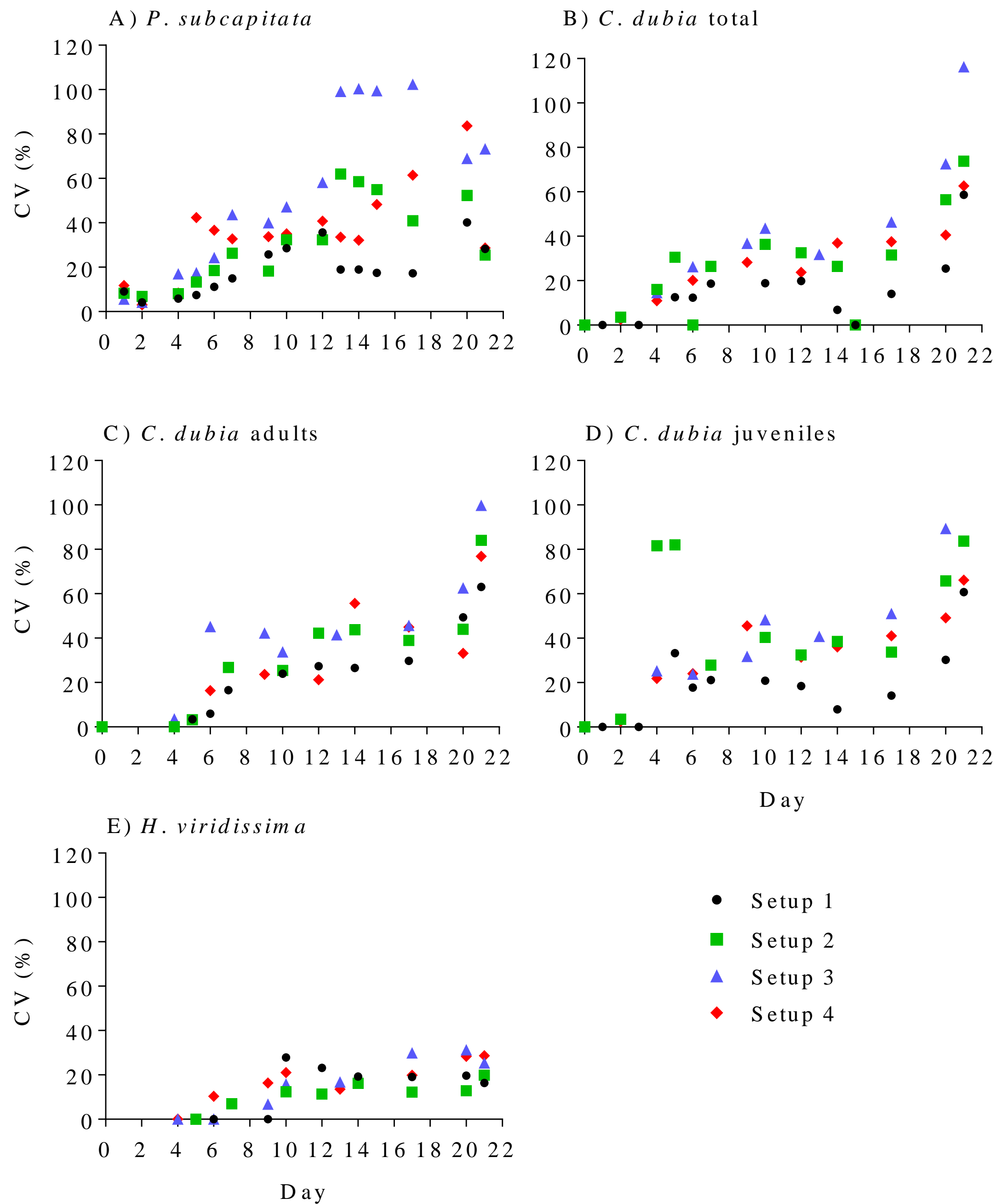

Figure A - 2: Coefficients of variation (CV) of (A) algal concentrations (cells/mL), (B) total number of *C. dubia*, (C) *C. dubia* adults, (D) *C. dubia* juveniles and (E) *H. viridissima* over 21 days. Shown are the coefficients in setup 1, setup 2, setup 3 and setup 4.

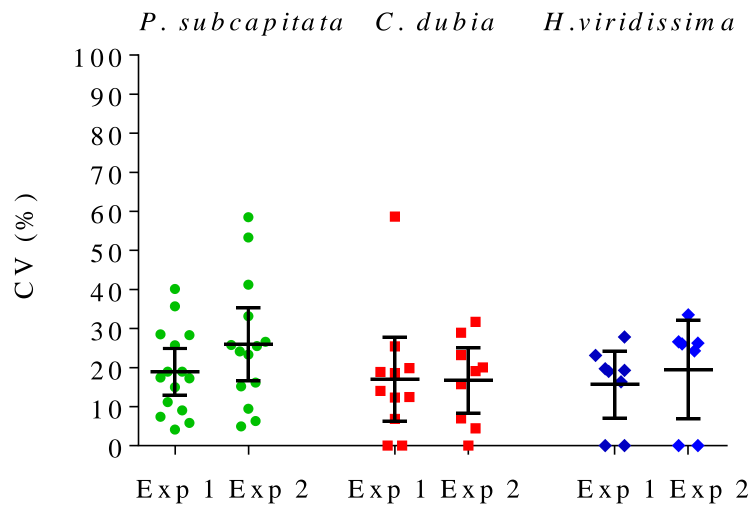

Figure A - 3: Comparison of coefficients of variation (CV) of algal concentrations (cells/mL) (green points), the total number of *C. dubia* (red squares) and *H. viridissima* (blue rhombi) in Experiment 1 and Experiment 2, respectively. Shown are Black horizontal lines indicate means ± 95 % confidence intervals.

Culturing procedures

Pseudokirchneriella subcapitata culturing

Adapted after OECD Freshwater Alga and Cyanobacteria Growth Inhibition Test (OECD, 2011). A concentrated starter culture of *P. subcapitata* can be obtained commercially. All working steps to set up *P. subcapitata* subcultures are performed under sterile conditions. The concentration is determined and aliquots are pipetted in 250 mL conical flasks with OECD medium (OECD, 2011). The subcultures are incubated at 25.0 ± 1°C, 24h light (approx. 9000 lux) on an orbital shaker (110 rpm) for 3 - 4 days. Approx. 3.5 - 4 x 10^7 cells are incubated in 150 mL and 200 mL each. After 3 - 4 days, the concentration of the 150 mL algal culture is determined and the cells are observed under the microscope. If the colour of the culture is yellowish or the cells don't look normal, the culture is discarded and the procedure is repeated for the 200 mL culture. If the colour of the subculture is vivid green and the cells have their normal horseshoe shape the culture is used for the experiments.

C. dubia culturing

The culture can be started with *C. dubia* ephippia or adult individuals from other sources. Cultures of known age are kept at 25.0 ± 1°C and a 12:12h light:dark (approx. 1000 lux) in moderately hard fresh water. The medium is aerated to saturation prior to usage and used within 1 week from preparation date. Each Friday, the neonates of females aged approx. 17 days are collected and new cultures are set up (approx. 50 individuals in 500 mL medium, at least 2 replicates). Each beaker contains adults/juveniles/neonates of approx. the same age and 2 - 3 age groups are kept simultaneously. The neonates produced in cultures of mature females have to be removed 3 times per week as neonates rapidly grow to the size of adults and then can't be distinguished. Removed neonates are discarded with the exception of individuals obtained on Fridays from adult females aged 16 - 17 days. The culture medium is renewed and the vessels are exchanged once or twice a week. Adult females aged approx. 17 days are discarded but their brood is transferred to new medium and establishes a new generation (G). Broods from cultures with adults aged 12 - 17 days are third or fourth broods and can be used for the experiments (see Table A - 5).

Hydra viridissima culturing

Hydra cultures are kept at 25.0 ± 1°C at 12:12h light:dark regime in T82MV. The medium is aerated to saturation prior to usage and used within 1 week from preparation date. Individuals are cultured in glass vessels that are loosely covered with transparent PVC sheets to avoid excessive water evaporation. The cultures are fed 3 times per week with *Artemia salina* previously rinsed in deionised water or culturing medium. *Hydra* are fed *ad libitum* to allow 10 *Artemia* per individual. The culture medium is changed after digestion (approx. 7h at 25 °C) to remove excess and regurgitated *Artemia* and avoid algal or fungal growth. The test vessels are changed after approx. 10 days and for the transfer, individuals can be carefully detached from the bottom either with a pipette or with the finger wearing powder free gloves.

Artemia salina culturing

A beaker (500 mL) is filled with approx. 350 mL of deionized water, 10 g of aquarium salt and a tablespoon of *Artemia* cysts. The beaker is covered with tin foil to avoid excessive evaporation, aerated and incubated at 25 - 30 °C. After approx. 48h, most of the cysts have hatched and neonate *Artemia* can be used for feeding. The aeration tube is removed and unhatched cysts either float on the surface or sink to the bottom separating from hatched *Artemia* that can be collected with a pipette.

Schedules

Table A - 4: Assembly of the multi-species TriCosm with sequential addition of species and regular monitoring (here experimental set up on Thursday).

Day	**Mo**	**Tue**	**We**	**Thu** **0 -START**	**Fri** **1**	
Procedure	Set up algal culture			Addition of algae, transfer *C. dubia* ♀ to 50% T82MV and feed	Addition of *C. dubia* neonates	**Week 1**
Measure				DO, pH	Algae	
Day	**4**	**5**	**6**	**7**	**8**	
Procedure		Transfer 72 *Hydras* to T82MV and feed	Addition of *Hydra* juveniles			**Week 2**
Measure	Algae, *C. dubia* DO, pH	Algae.	Algae *C. dubia*	Algae	Algae *C. dubia* *Hydra* DO, pH	
Day	**11**	**12**	**13**	**14**	**15**	
Measure	Algae, *C. dubia* *Hydra,* DO, pH	Algae	Algae *C. dubia* *Hydra*	Algae	Algae *C. dubia* *Hydra* DO, pH	**Week 3**
Day	**18**	**19**	**20 - END**			
Measure	Algae *C. dubia* *Hydra* DO, pH	Algae	Algae *C. dubia* *Hydra*			**Week 4**

Table A - 5: Generation (G) renewal of cultured *C. dubia* females for continuous supply of neonates from third or fourth broods.

Mo	Tue	Wed	Thu	Fri	Sat	Sun	
G 1 (age 4-5d)	G 1 (age 5-6d)	G 1 (age 6-7d)	G 1 (age 7-8d)	G 1 (age 8-9d) keep neonates	G 1 (age 9-10d)	G 1 (age 10-11d)	**Week 1**
				G 2 (age 1-2d)	G 2 (age 2-3d)	G 2 (age 3-4d)	
Mo	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**	**Sun**	
G 1 (age11-12d)	G 1 (age 12-13d)	G 1 (age 13-14d)	G 1 (age 14 15d)	G 1 (age 16-17d) discard adults, keep neonates			**Week 2**
				G 3 (age 1-2d)	G 3 (age 2-3d)	G 3 (age 3-4d)	
G 2 (age 4-5d)	G 2 (age 5-6d)	G 2 (age 6-7d)	G 2 (age 7-8d)	G 2 (age 8-9d)	G 2 (age 9-10d)	G 2 (age 11-12d)	
Mo	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**	**Sun**	
G 2 (age 12-13d)	G 2 (age 13 14d)	G 2 (age 14-15d)	G 2 (age 15 16d)	G 2 (age 16-17d) discard adults, keep neonates			**Week 3**
				G 4 (age 1-2d)	G 4 (age 2-3d)	G 4 (age 3-4d)	

Appendix B – Supporting Data for Chapter Three

Species interactions and indirect effects in a standardized tri-trophic laboratory microcosm exposed to a pesticide

Verena Riedl, * Annika Agatz, Rachel Benstead, and Roman Ashauer

** Address correspondence to* verena.riedl@york.ac.uk

Materials and Methods: Analysis methods used to quantify linuron concentrations via liquid scintillation counting and HPLC analysis

Table B - 1: Count settings, corrections and background subtraction during liquid scintillation counting to analyse ^{14}C linuron concentrations in aquatic samples.

Count Conditions		**Background Subtraction**	
Scintillation Cocktail	Ultima Gold (PerkinElmer, UK)	Background Subtract	Off
		Low CPM Threshold	Off
Quench indicator	tSIE/AEC		
Pre-Count Delay (min)	0	**Count Corrections**	
Count time (min)	5.00	Static Controller	On
Count Mode	Normal	Coloured Samples	Off
Assay Count Cycles	1	Coincidence Time (nsec)	18
#Vials/Sample	1	Luminescence Correction	n/a
Repeat Sample Count	1	Heterogeneity Monitor	n/a
Calculate % Reference	Off	Delay Before Burst (nsec)	75

The efficiency to detect ^{14}C compounds via liquid scintillation counting was calculated for each replicate per sampling day in linuron treatments. The counting efficiency (%) was calculated as described by Lee and Kim, (2006).

$$\text{Efficiency (\%)} = \frac{\text{cpm}}{\text{dpm}} \text{ x } 100$$

In this equation, cpm (counts per minute) is the number of light impulses the instrument detected per minute; dpm (disintegrations per minute) is the activity of the sample measured as decays per minute.

Table B - 2: Method for HPLC analysis to determine unlabelled linuron concentrations in aquatic stock solutions.

Autosampler		**Pump**	
Model Type	G1313A	Model Type	G1311A
Draw Speed	200 µL/min	Max Flow up/down	100 mL/min^2
Ejection Speed	200 µL/min	Diode Array Detector	
Flush Volume	250 mL		
Cells and Flow Rate		**Solvent Information**	
Cell Volume	500 µL	Solvent	50/50
Injection Volume (µL)	50		Acetonitrile/Water

C. dubia sensitivity to acetonitrile

A solvent limit of 0.1 mL/L is suggested for toxicity testing with *D. magna* (OECD, 2012). To exclude effects due to interspecific variations e.g. due to smaller size*, C. dubia* were exposed to five acetonitrile concentrations (0 mL/L, 0.02 mL/L, 0.05 mL/L, 0.08 mL/L and 0.1 mL/L) with two replicates per treatment (60 mL Duran® crystallizing dishes Sigma-Aldrich, UK; with 50 mL aerated T82MV (0.02 µg/L Na_2SiO_3; ASTM E1366-11, 2011) 65 rpm orbital shaking, 12:12 h light:dark, 25 (± 1) °C, approx. 16 par µmol m^{-2} s^{-1}). Three *C. dubia* neonates (age ≤ 24 h) were used per replicate and fed with ca. 6.6 x 10^6 cells *P. subcapitata* on days 0 and 2. The water parameter pH and the number of neonates were recorded on days 0, 2 and 4 (when the experiment was terminated) to assess acetonitrile effects on reproduction.

Results:

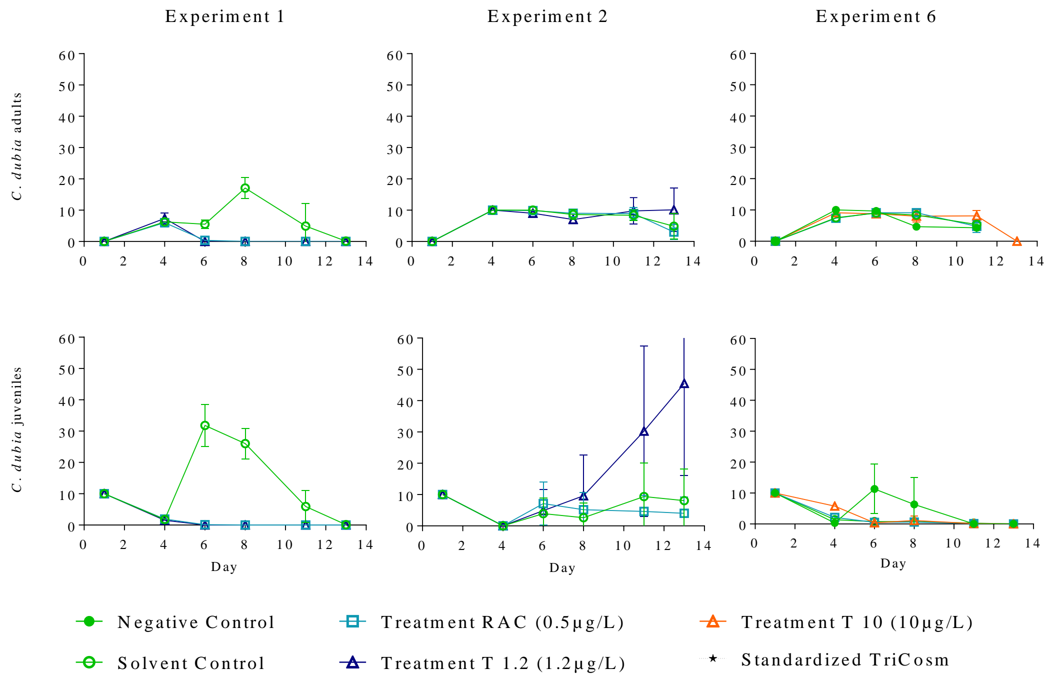

Figure B - 1: Population dynamics of adult (top row) and juvenile (bottom row) *C. dubia* in experiments 1, 2 and 6. Shown are means ± 95 % confidence intervals.

Table B - 3: Classification scheme for *Hydra* toxicity according to morphology scores (1 - 10). Scores ≥ 6 are considered sub-lethal and reversible whereas scores ≤ 5 are considered lethal. Redrafted from Quinn et al., 2012 from Wilby, 1988.

General morphology	Score	Detailed morphology
	10	Extended tentacles, body reactive
Normal	9	Partially contracted, slow reactions
	8	Clubbed tentacles, body slightly contracted
Clubbed/bulbed tentacles	7	Shortened tentacles, body slightly contracted
Shortened tentacles	6	Tentacles and body shortened
	5	Totally contracted, tentacles visible
	4	Totally contracted, no visible tentacles
Tulip	3	Expanded, tentacles visible
	2	Expanded, no visible tentacles
Loss of regulation	1	Dead but intact
Disintegration	0	Disintegrated

Results: pH and dissolved oxygen

The water parameter pH generally increased with time in all experiments (Figure B - 2). Values in experiments 5 and 6 were higher compared to pH in experiments 1, 2, 3 and 4 and the standardized TriCosm. Average pH values were similar between treatments and experiments 1, 2, 3 and 4 (controls: pH 7.11 (± 0.05), T RAC: pH 7.22 (± 0.06), T 1.2: pH 7.18 (± 0.06)) whereas higher values and steeper pH increases were measured in experiments 5 and 6 (controls: pH 7.91 (± 0.09), T RAC: pH 7.92 (± 0.08), T 10: pH 7.93 (± 0.08)).

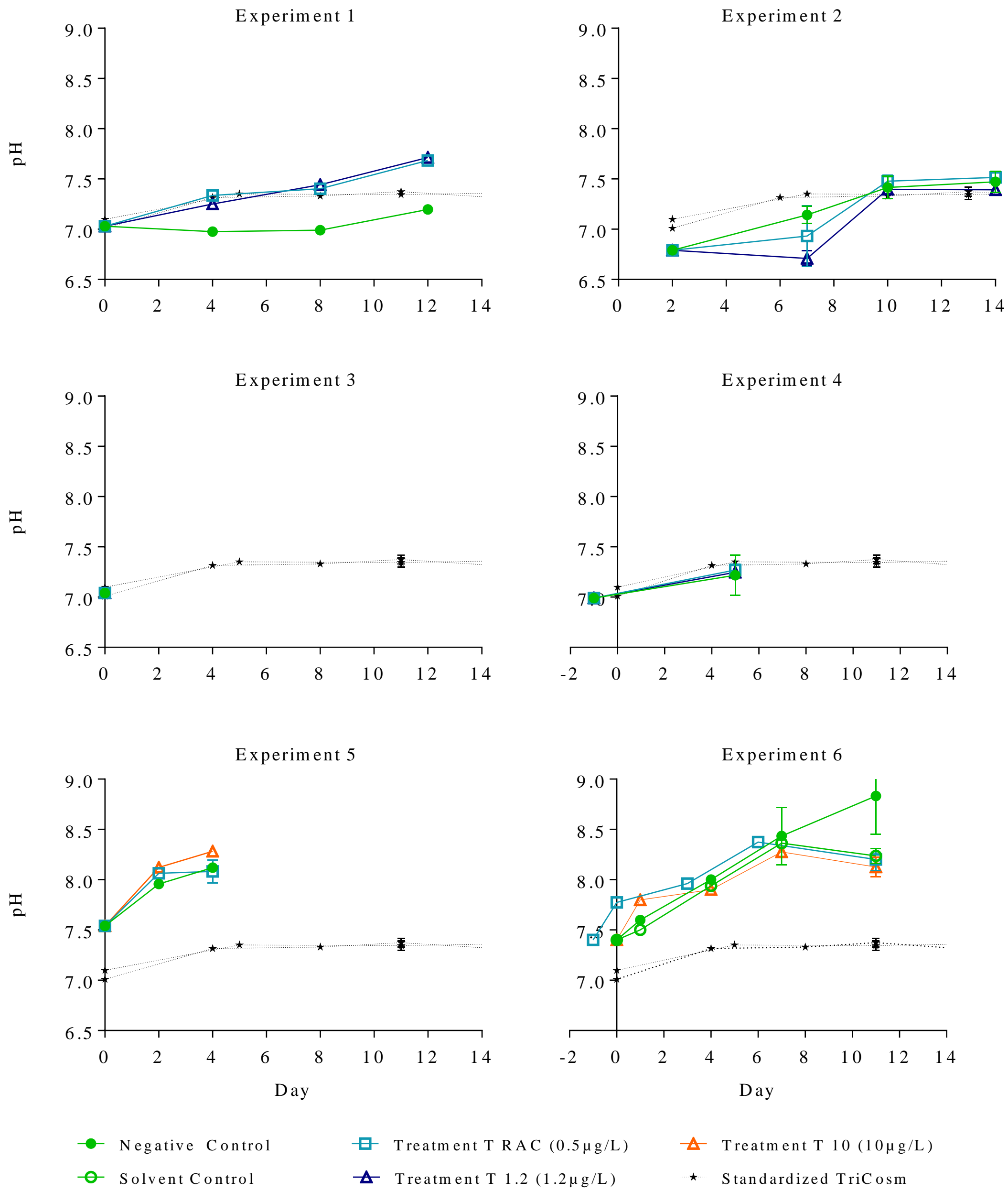

Figure B - 2: Water parameter pH in controls and treatments in experiments 1 - 6 over time. PH measurements in the standardized TriCosm are indicated for reference with stars. Shown are mean values ± 95 % confidence intervals.

Dissolved oxygen concentrations showed small increases over time and were similar between treatments and among experiments 1, 2 and 6 (Figure B - 3). In experiment 5, a lower average of 7.78 (± 0.06) mg/L was measured among treatments over time whereas only one starting measurement was taken in experiment 3.

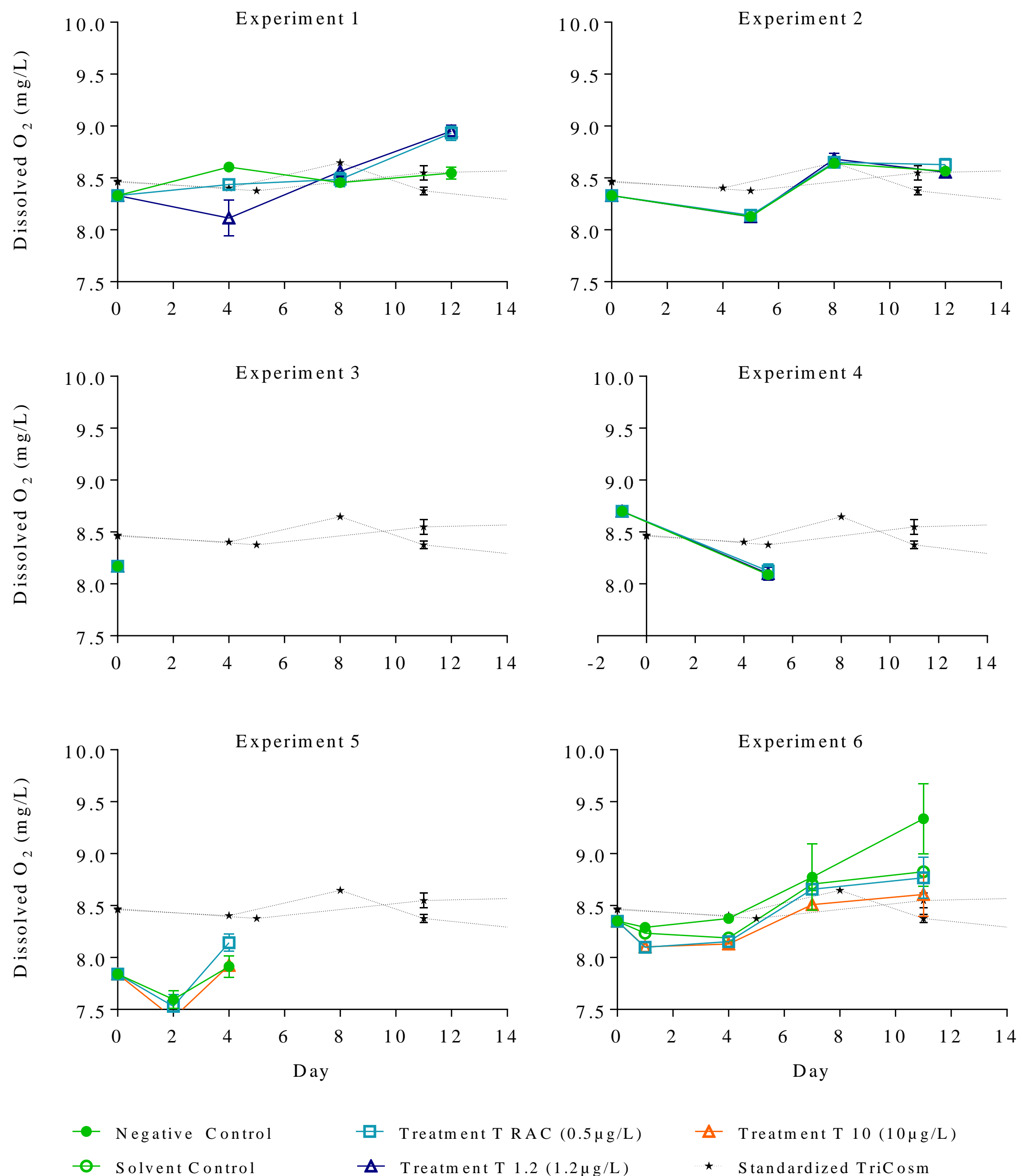

Figure B - 3: Water dissolved oxygen (mg/L) in controls and treatments in experiments 1 - 6 over time. Measurements of dissolved oxygen in the standardized TriCosm are indicated for reference with stars. Shown are mean values ± 95 % confidence intervals.

Results: Sodium metasilicate toxicity data

Table B - 4: Sodium metasilicate (Na_2SiO_3) effect concentrations are reported for *P. subcapitata, C. dubia* and *H. viridissima*. Endpoint, age, duration and concentration range are indicated.

Species	Endpoint		Age	Duration (h)	g/L	Reference
P. subcapitata	Growth	EC_{10}	-	72	0.228	Van Hoecke et al., 2008
		EC_{20}	-	72	0.234	
C. dubia	Mobility	EC_{50}	< 24h	48	0.023 - 0.049	Warne and Schifko, 1999
H. viridissima	Morphology, reproduction	-	-	72	0.36 - 0.39	Šimičev et al., 2016

Appendix C – Supporting Data for Chapter Four

Factors affecting the growth of *Pseudokirchneriella subcapitata*: lessons on the experimental design and consequences for the reproducibility of a multi-trophic laboratory microcosm

Verena Riedl,* Annika Agatz, Rachel Benstead, and Roman Ashauer

** Address correspondence to* verena.riedl@york.ac.uk

Introduction

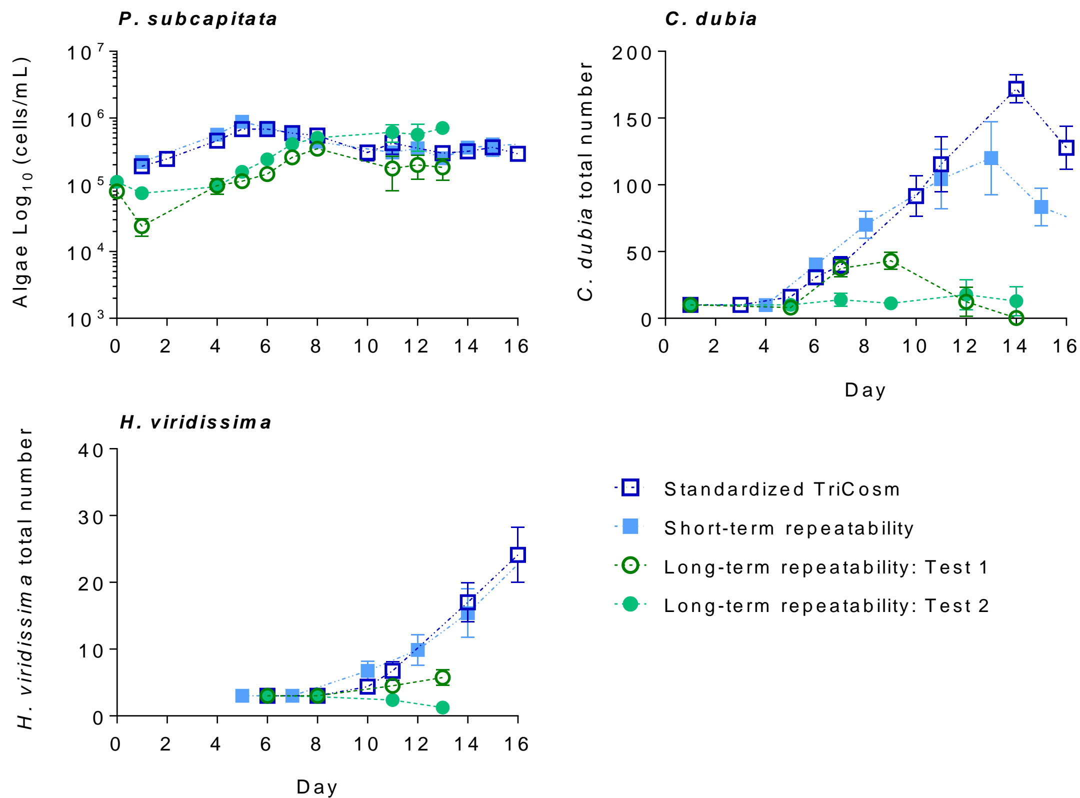

Figure C - 1: Short-term (over three months) and long-term (over one year) reproducibility of the population dynamics of *P. subcapitata*, *C. dubia* and *H. viridissima* in the standardized TriCosm. The short-term reproducibility was described previously by Riedl et al., (2018) and Test 1 and Test 2 were conducted to assess the long-term repeatability of control population dynamics in the standardized TriCosm. Shown are mean (n = 8) ± 95 % confidence intervals.

Materials and methods

Table C - 1: Plate reader settings for in-vivo fluorescence analyses.

Model	
Plate reader:	Tecan® Infinite 200PRO
Plate Description:	[GRE96ut] - Greiner 96 U Transparent
Range:	B2:F11
Shaking	
Duration:	30 sec
Mode:	Linear
Amplitude:	1 mm
Frequency:	886.9 rpm
Fluorescence Intensity	
Excitation Wavelength:	430 nm
Excitation Bandwidth:	20 nm
Emission Wavelength:	670 nm
Emission Bandwidth:	25 nm
Reading Mode:	Top
Lag Time:	0 µs
Integration Time:	20 µs
Number of Reads:	25
Settle Time:	0 ms
Gain:	Manual
Gain Value:	59
Mirror:	Automatic
Mirror:	Dichroic 510 (e.g. fluorescein)

Table C - 2: Chemical composition, preparation and storage of stock solutions for the medium T82MV (from ASTM E1366-11, 2011).

Stock	Compound	g/L	mL	Preparation and storage
1	$NaNO_3$	21.25		**Stocks 1-8:** store in the
2	$MgSO_4$ $7H_2O$	12.33		refrigerator (ca. 4 °C) and renew
3	$CaCl_2$ $2H_2O$	73.50		regularly (ca. 6 months).
4	NaCl	43.80		
5	$Al_2(SO_4)_3$ $18H_2O$	1.60		
6	Na_2SiO_3 $9H_2O$	11.38		**Stocks 1-5:** autoclave at 121°C for
7	NaOH	0.640		1h;
	KH_2PO_4	2.72		
8-Trace	Solution **8a**		250	**Stocks 6-7:** sterilize via 0.22µm
Metals	Solution **8b**		500	membrane filtration;
8	Solution **8c**		60	
	Deionized water		190	
8a	NaOH	10.70		
	EDTA	26.10		
	$FeSO_4$ $7H_2O$	24.90		**Stock 8a:** Dissolve 10.7g NaOH in
8b	H_3BO_3	1.85		260mL deionized water, add 26.1 g
	$ZnSO_4$ $7H_2O$	0.290		EDTA, add 24.9g $FeSO_4$ $7H_2O$, fill to
	$MnCl_2$ $4H_2O$	1.98		1L with deionized water and
	Na_2MoO_4 $2H_2O$	0.240		aerate overnight.
	$CuSO_4$ $5H_2O$	0.045		**Stock 8a, 8b:** let solutions cool
	$Co(NO_3)_2$ $6H_2O$	0.029		after adding NaOH, use EDTA
8c	NaOH	12.00		(Ethylenedinitrilotetraacetic acid
	EDTA	29.00		form).
9-	NaOH	0.2		
Vitamins	Calcium pantothenate	1.750		**Stock 9:** dissolve 0.2g NaOH in 1L deionized water, add vitamins and
	Thiamin (B1)	0.150		sterilize solution via 0.22 µm
	Riboflavin (B2)	0.100		membrane filtration.
	Nicotinamide	0.325		Fill solution in 2 - 5mL tubes and
	Folic acid	0.825		store in the freezer (ca. -20 °C).
	Biotin	0.075		
	Choline	1.250		
	Inositol	2.750		
	Pyridoxine (B6) monohydrochloride	1.250		
	Solution **9a**		1	
9a	Cyanocobalamin	0.075		

Table C - 3: Preparation and storage of the medium T82MV (from ASTM E1366-11, 2011).

Stock	mL/L	Preparation and storage
1	2	Add stocks to 1L deionized water, aerate for at least 48h to reach chemical equilibrium.
2	2	If necessary, adjust to pH 7.2 ± 0.2 with sterile HCl (5%). The pH value must not drop below pH 7.0 and pH should be monitored throughout the experimental duration.
3	2	
4	2	
5	2	
6	2	
7	2	
8 – Metals	0.05	
9 – Vitamin solution	0.4	

Table C - 4: Chemical composition, preparation and storage of Keating's metal solution. Only trace metals that are not added present in the medium T82MV are added. Once Keating's trace metals are added to T82MV medium, the medium is called T86MVK (from ASTM E1366-11, 2011).

Compound	mg/L	Preparation and storage
NaBr	64.40	Add compounds to 1L distilled water in a glass container, autoclave and store in refrigerator (ca. 4 °C).
$SrCl_2 6H_2O$	304.00	
RbCl	141.50	
LiCl	611.00	
KI	6.50	
SeO_2	1.41	1 mL is added per Litre of T82MV.
$NH_4 VO_3$	1.15	

Table C - 5: Sodium metasilicate (Na_2SiO_3) effect concentrations are reported for *P. subcapitata, C. dubia* and *H. viridissima*. Endpoint, age, duration and concentration range are indicated.

Species	Endpoint		Age	Duration (h)	g/L	Reference
P. subcapitata	Growth	EC_{10}	-	72	0.228	Van Hoecke et al., 2008
		EC_{20}	-	72	0.234	
C. dubia	Mobility	EC_{50}	< 24h	48	0.023 - 0.049	Warne & Schifko, 1999
H. viridissima	Morphology, reproduction	-	-	72	0.36 - 0.39	Šimičev et al., 2016

Results - Algal population dynamics

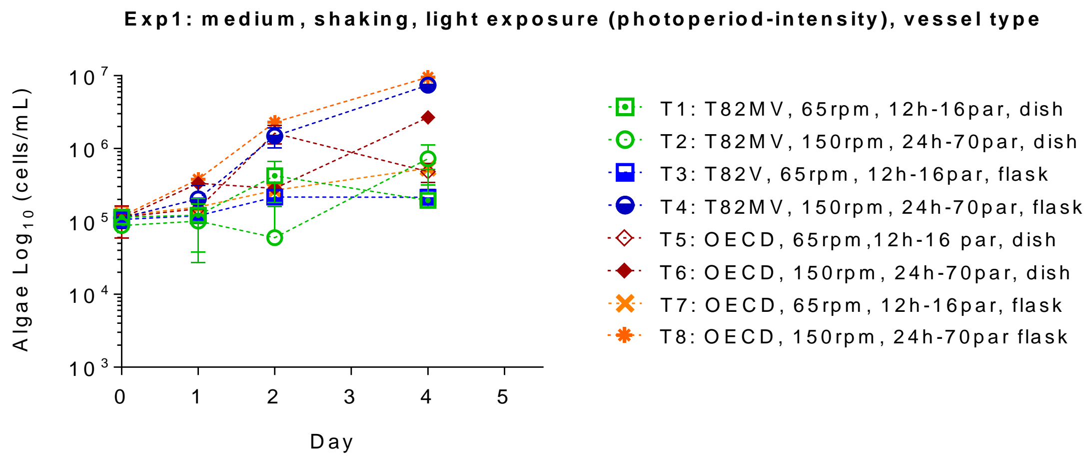

Figure C - 2: Algal population dynamics in experiment 1. Medium type, light exposure (photoperiod- intensity) and vessel type among treatments (for details see Table 4 - 1). Shown are mean (n = 3) ± 95 % confidence intervals. Exp = experiment, T = treatment.

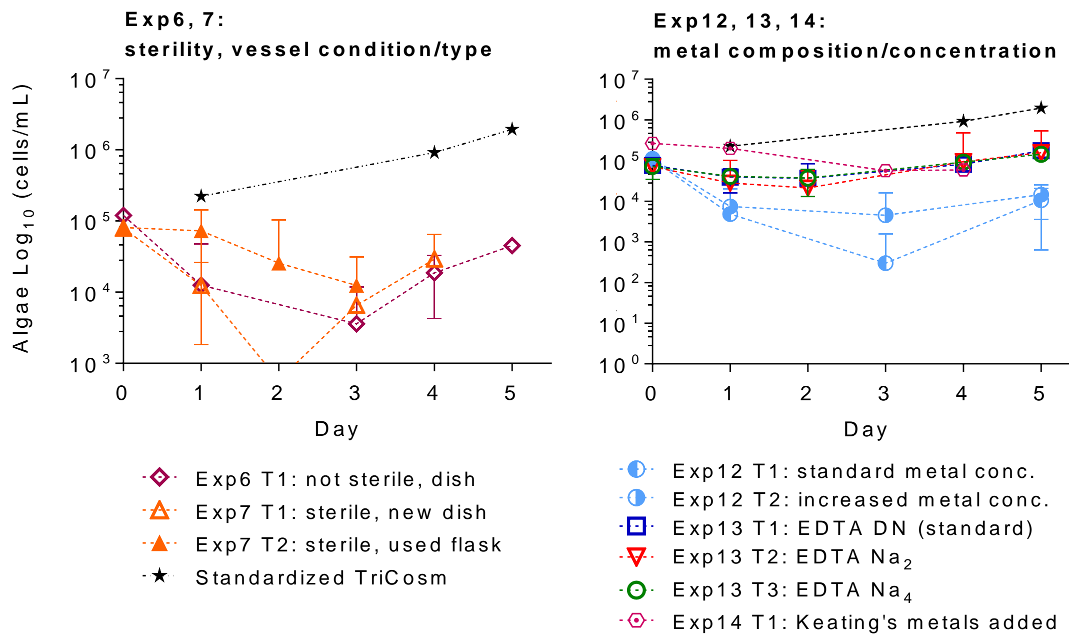

Figure C - 3: Algal population dynamics in experiments 6, 7, 12 – 14 with differing medium sterility, vessel condition and type and trace metal composition and concentration. Labels indicate factor variations among treatments (for details see Table 4 - 1). Algal dynamics observed in the standardized TriCosm without grazing pressure are indicated with stars for reference. Shown are mean (n = 3) ± 95 % confidence intervals. Exp = experiment, T = treatment.

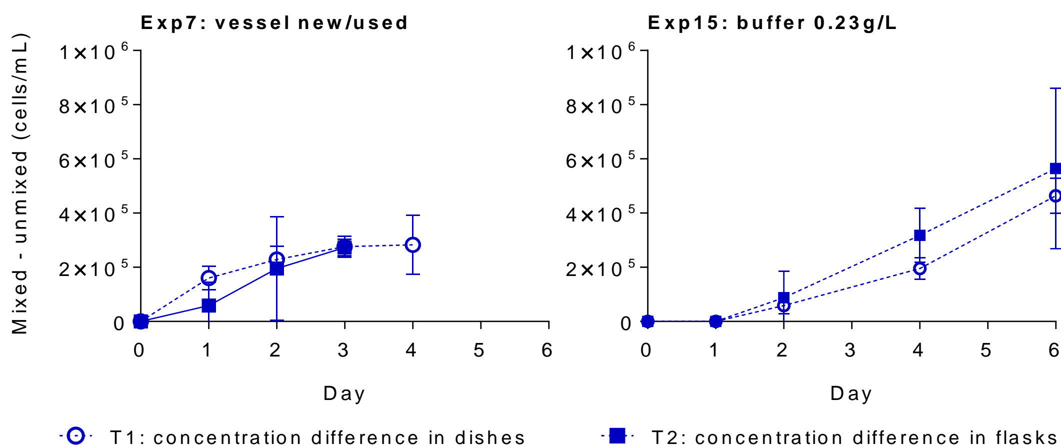

Figure C - 4: Algal concentration differences between measurements taken before and after mixing the water columns in experiments 7 and 15 (A-D; full square = flask, empty circle = dish; see Table 4 - 1 for details). Shown are mean (n = 3) ± 95 % confidence intervals. Exp = experiment, T = treatment.

Results – pH values in different treatments

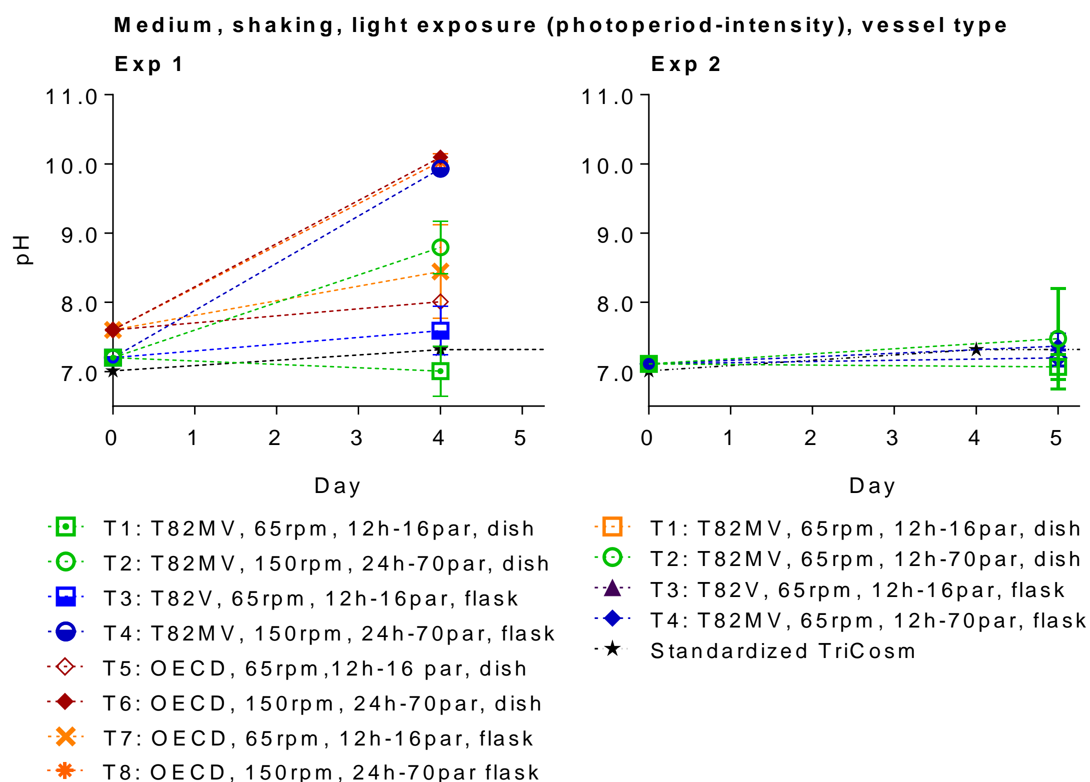

Figure C - 5: Water pH in experiments 1 and 2. Shown are mean (n = 3) ± 95 % confidence intervals; pH (n = 8) measured in the standardized TriCosm is indicated for reference. Exp = experiment, T = treatment.

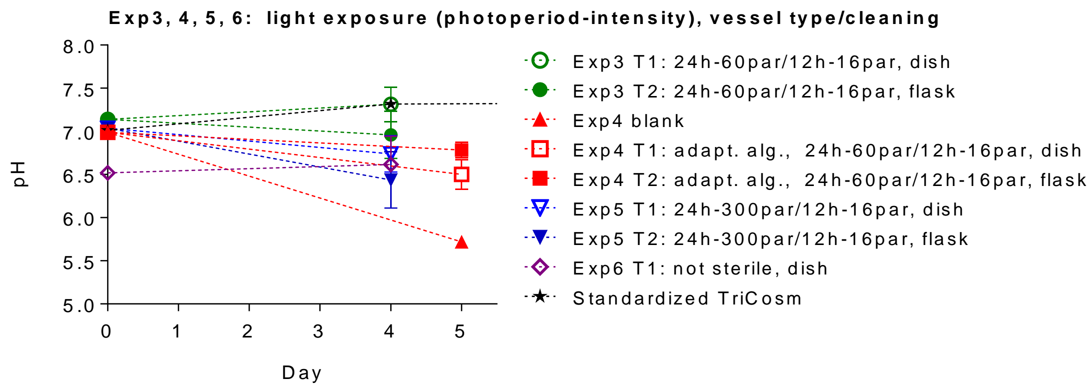

Figure C - 6: Water pH in experiments 3 - 6 (empty symbol = dish, full symbol = flask; see Table 4 - 1 for details). Shown are mean (n = 3) ± 95 % confidence intervals; pH (n = 8) measured in the standardized TriCosm is indicated for reference. Exp = experiment, T = treatment.

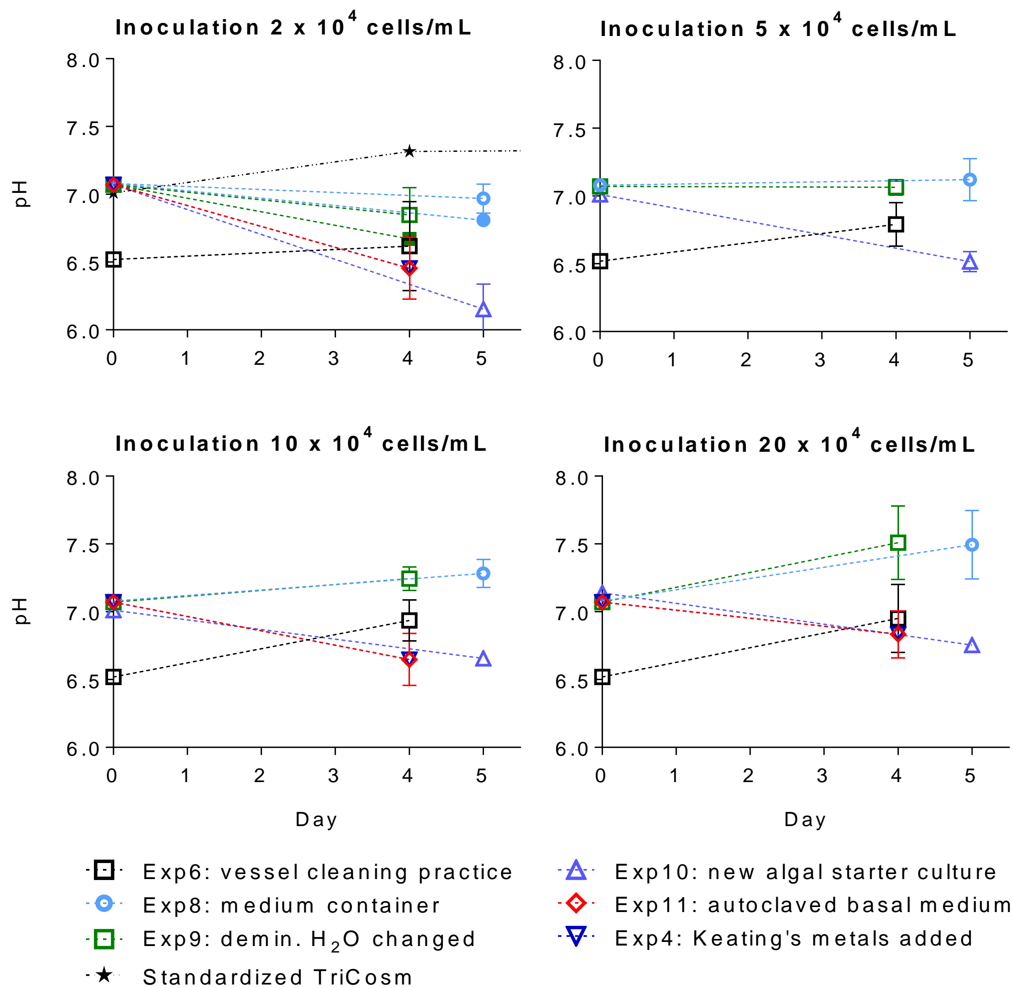

Figure C - 7: Water pH in experiments 6, 8 - 11, 14 among treatments with different inoculation concentrations (A-D; see Table 4 - 1 for details). Shown are mean (n = 3) ± 95% confidence intervals; pH (n = 8) measured in the standardized TriCosm is indicated for reference. Exp = experiment.

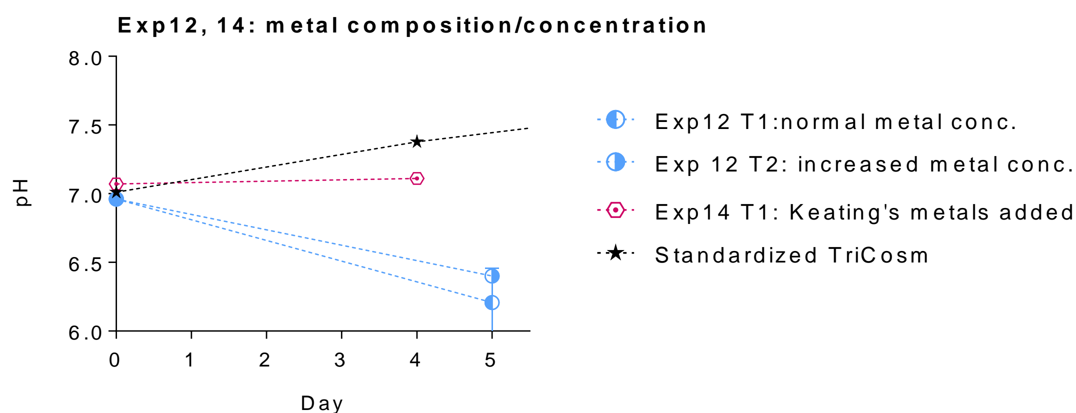

Figure C - 8: Water pH in experiment 12 and 14 (see Table 4 - 1 for details). Shown are mean (n = 3) ± 95 % confidence intervals; pH (n = 8) measured in the standardized TriCosm is indicated for reference. Exp = experiment; T = treatment.

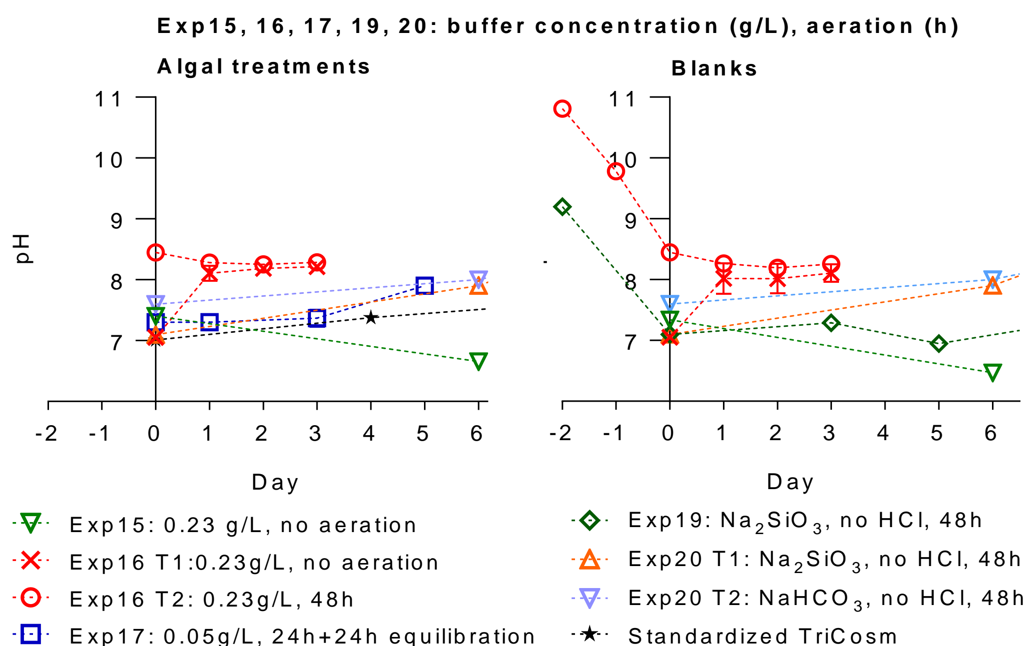

Figure C - 9: Water pH in experiments 15 – 17, 19, 20 in algal treatements and blanks. No media blank was set up. Shown are mean (n = 3) ± 95 % confidence intervals; pH (n = 8) measured in the standardized TriCosm is indicated for reference. Exp = experiment; T = treatment.

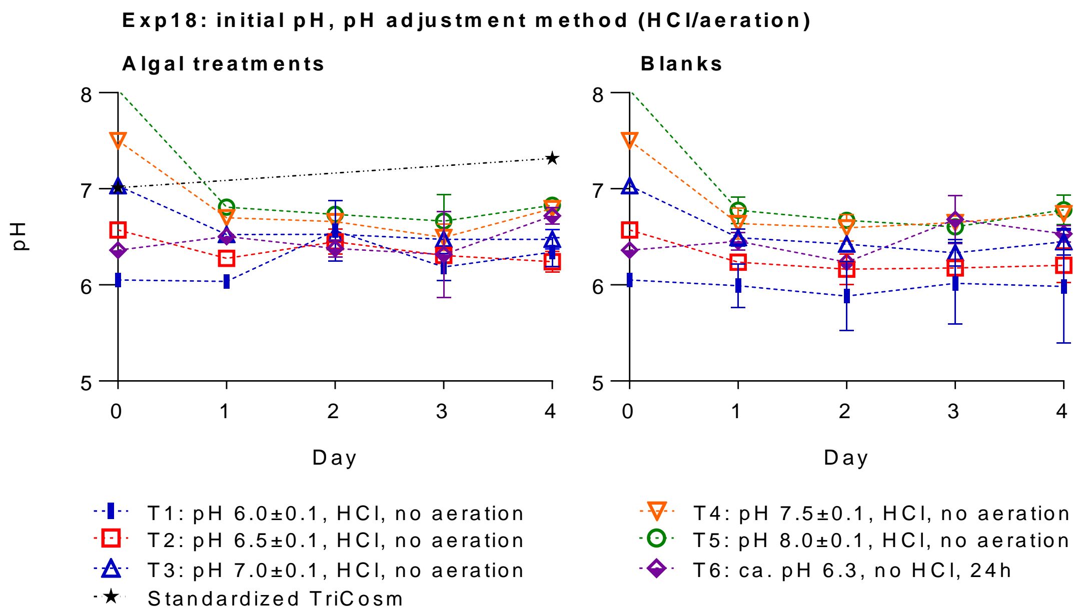

Figure C - 10: Water pH in experiment 18 in media blanks and algal treatments (A-B; see Table 4 - 1 for details). Shown are mean (n = 3) ± 95 % confidence intervals; pH (n = 8) measured in the standardized TriCosm is indicated for reference. Exp = experiment; T = treatment.

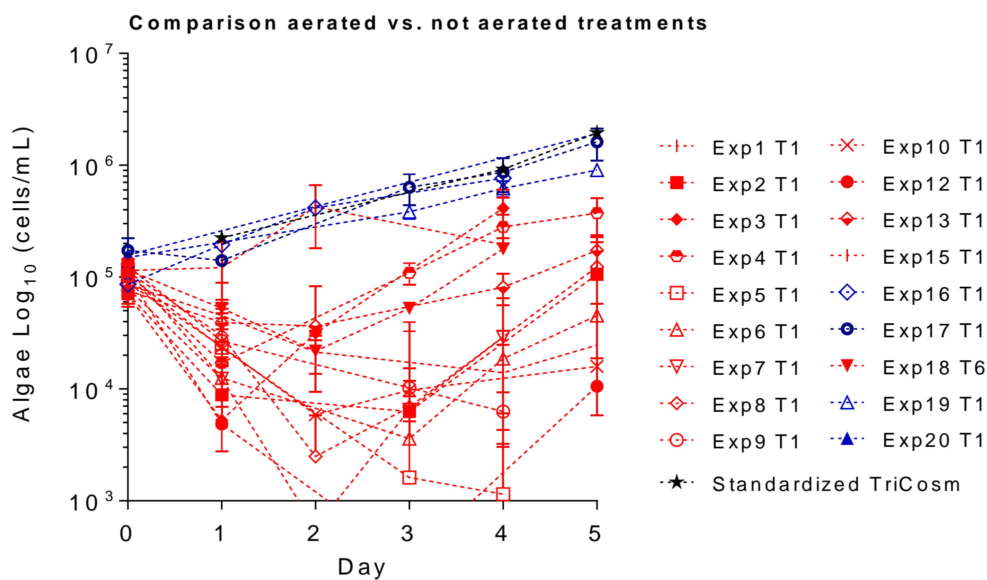

Figure C - 11: Algal growth dynamics compared among treatments with T82MV medium in crystallizing dishes, 2 x 10^4 cells/mL starting concentration and 65 rpm orbital shaking in experiments 1 – 20 (see Table 4 - 1 for details). Colours distinguish experiments where medium was prepared aseptically as possible and/or used shortly after preparation (red) and experiments where medium was aerated for ≥ 24h and only used approx. 48h after medium preparation (blue). Shown are mean values ± 95 % confidence intervals.

Abbreviations

AF	Assessment Factor
ASTM	American Society for Testing of Materials
CCAP	Culture Collection of Algae and Protozoa
CV	Coefficient of Variation
EC	European Commission
EC_x	x % of the population shows an effect at a given concentration
EDTA DN	Ethylenedinitrilotetraacetic salt
EDTA Na_2	Ethylenediaminetetraacetic disodium salt
EDTA Na_4	Ethylenediaminetetraacetic tetrasodium salt
EFSA	European Food and Safety Authority
EPA	Environmental Protection Agency
EQS	Environmental Quality Standard
ERA	Environmental Risk Assessment
Exp	Experiment
MDD	Minimum Detectable Difference
NOEC	No Observed Effect Concentration
OECD	Organisation for Economic Co-operation and Development
PEC	Predicted Environmental Concentration
PNEC	Predicted No Effect Concentration
PPP	Plant Protection Product
RAC	Regulatory Acceptable Concentration
rpm	Revolutions Per Minute
SAM	Standardized Aquatic Microcosm
SCCS	Scientific Committee on Consumer Safety
SCENHIR	Scientific Committee on Emerging and Newly Identified Health Risks
SCHER	Scientific Committee on Health and Environmental Risks
T82MV	Standardized test medium in the SAM and the TriCosm
USEPA	United States Environmental Protection Agency
WFD	Water Framework Directive

Full list of references

Agatz, A., Ashauer, R. and Brown, C.D. 2014. Imidacloprid perturbs feeding of *Gammarus pulex* at environmentally relevant concentrations. *Environmental Toxicology and Chemistry*. **33**(3), pp.648–653.

Agatz, A. and Brown, C.D. 2013. Evidence for links between feeding inhibition, population characteristics, and sensitivity to acute toxicity for *Daphnia magna*. *Environmental Science and Technology*. **47**(16), pp.9461–9469.

Agatz, A., Hammers-Wirtz, M., Gabsi, F., Ratte, H.T., Brown, C.D. and Preuss, T.G. 2012. Promoting effects on reproduction increase population vulnerability of *Daphnia magna*. *Environmental Toxicology and Chemistry*. **31**(7), pp.1604–1610.

Altermatt, F., Fronhofer, E.A., Garnier, A., Giometto, A., Hammes, F., Klecka, J., Legrand, D., Mächler, E., Massie, T.M., Pennekamp, F., Plebani, M., Pontarp, M., Schtickzelle, N., Thuillier, V. and Petchey, O.L. 2015. Big answers from small worlds: A user's guide for protist microcosms as a model system in ecology and evolution. *Methods in Ecology and Evolution*. **6**(2), pp.218–231.

Anderson, D.H. and Benke, A.C. 1994. Growth and reproduction of the cladoceran *Ceriodaphnia dubia* from a forested floodplain swamp. *Limnology and Oceanography*. **39**(7), pp.1517–1527.

Anskjær, G.G., Rendal, C. and Kusk, K.O. 2013. Effect of pH on the toxicity and bioconcentration of sulfadiazine on *Daphnia magna*. *Chemosphere*. **91**(8), pp.1183–1188.

Del Arco, A.I., Parra, G., Rico, A. and Van den Brink, P.J. 2015. Effects of intra- and interspecific competition on the sensitivity of aquatic macroinvertebrates to carbendazim. *Ecotoxicology and Environmental Safety*. **120**, pp.27–34.

Arias-Estévez, M., López-Periago, E., Martínez-Carballo, E., Simal-Gándara, J., Mejuto, J.C. and García-Río, L. 2008. The mobility and degradation of pesticides in soils and the pollution of groundwater resources. *Agriculture, Ecosystems and Environment*. **123**(4), pp.247–260.

Aruoja, V. 2011. *Algae* Pseudokirchneriella subcapitata *in environmental hazard*

evaluation of chemicals and synthetic nanoparticles. PhD thesis. Estonian University of Life Sciences.

ASTM E1366-11 2011. *Standard Practice for Standardized Aquatic Microcosms: Fresh Water*. West Conshohocken, PA: ASTM International.

Attanasio, R. 2016. The rise of evidence-based ecotoxicology. *Integrated Environmental Assessment and Management*. **12**(2), p.215.

Baker, M. 2016. Reproducibility crisis? *Nature*. **533**, p.26.

Barry, M.J. and Davies, W. 2004. Effects of invertebrate predators and a pesticide on temporary pond microcosms used for aquatic toxicity testing. *Environmental Pollution*. **131**(1), pp.25–34.

Bednarska, A.J., Jevtić, D.M. and Laskowski, R. 2013. More ecological ERA: incorporating natural environmental factors and animal behavior. *Integrated environmental assessment and management*. **9**, pp.1–8.

Begley, C.G. and Ellis, L.M. 2012. Drug development: Raise standards for preclinical cancer research. *Nature*. **483**(7391), pp.531–533.

Beketov, M. a, Kefford, B.J., Schäfer, R.B. and Liess, M. 2013. Pesticides reduce regional biodiversity of stream invertebrates. *Proceedings of the National Academy of Sciences of the United States of America*. **110**(27), pp.11039–43.

Belanger, S.E. and Cherry, D.S. 1990. Interacting Effects of pH Acclimation, pH, and Heavy Metals on Acute and Chronic Toxicity to *Ceriodaphnia dubia* (Cladocera). *Journal of Crustacean Biology*. **10**(2), pp.225–235.

Benton, T.G., Solan, M., Travis, J.M.J. and Sait, S.M. 2007. Microcosm experiments can inform global ecological problems. *Trends in Ecology and Evolution*. **22**(10), pp.516–521.

Berezina, N.A. 2001. Influence of ambient pH on freshwater invertebrates under experimental conditions. *Russian Journal of Ecology*. **32**(5), pp.343–351.

Blacquière, T., Smagghe, G., Van Gestel, C. a M. and Mommaerts, V. 2012. Neonicotinoids in bees: A review on concentrations, side-effects and risk assessment. *Ecotoxicology*. **21**(4), pp.973–992.

Boonstra, H., Reichman, E.P. and Van Den Brink, P.J. 2011. Effects of the veterinary pharmaceutical ivermectin in indoor aquatic microcosms. *Archives of*

Environmental Contamination and Toxicology. **60**(1), pp.77–89.

Brain, R.A., Bestari, K., Sanderson, H., Hanson, M.L., Wilson, C.J., Johnson, D.J., Sibley, P.K. and Solomon, K.R. 2005. Aquatic microcosm assessment of the effects of tylosin on *Lemna gibba* and *Myriophyllum spicatum*. *Environmental Pollution*. **133**(3), pp.389–401.

Bredeson, M.M., Reese, R.N. and Lundgren, J.G. 2015. The effects of insecticide dose and herbivore density on tri-trophic effects of thiamethoxam in a system involving wheat, aphids, and ladybeetles. *Crop Protection*. **69**(March), pp.70–76.

Van den Brink, P.J., Hartgers, E.M., Fettweis, U., Crum, S.J.H., Van Donk, E. and Brock, T.C.M. 1997. Sensitivity of Macrophyte-Dominated Freshwater Microcosms to Chronic Levels of the Herbicide Linuron. *Ecotoxicology and Environmental Safety*. **38**, pp.13–24.

Van den Brink, P.J., Tarazona, J. V, Solomon, K.R., Knacker, T., Van den Brink, N.W., Brock, T.C.M. and Hoogland, J.P. 2005. The use of terrestrial and aquatic microcosms and mesocosms for the ecological risk assessment of veterinary medicinal products. *Environmental toxicology and chemistry*. **24**(4), pp.820–829.

Brock, T.C.M., Hammers-Wirtz, M., Hommen, U., Preuss, T.G., Ratte, H.T., Roessink, I., Strauss, T. and Van den Brink, P.J. 2015. The minimum detectable difference (MDD) and the interpretation of treatment-related effects of pesticides in experimental ecosystems. *Environmental science and pollution research international*. **22**(2), pp.1160–1174.

Brown, K., Mays, T., Romoser, A., Marroquin-Cardona, A., Mitchell, N., Elmore, S. and Phillips, T. 2014. Modified hydra bioassay to evaluate the toxicity of multiple mycotoxins and predict the detoxification efficacy of a clay-based sorbent. *Journal of Applied Toxicology*. **34**(1), pp.40–48.

Campos, D., Alves, A., Lemos, M.F.L., Correia, A., Soares, A.M.V.M. and Pestana, J.L.T. 2014. Effects of cadmium and resource quality on freshwater detritus processing chains: A microcosm approach with two insect species. *Ecotoxicology*. **23**(5), pp.830–839.

Cardinale, B.J., Matulich, K.L., Hooper, D.U., Byrnes, J.E., Duffy, E., Gamfeldt, L.,

Balvanera, P., O'Connor, M.I. and Gonzalez, A. 2011. The functional role of producer diversity in ecosystems. *American Journal of Botany*. **98**(3), pp.572–592.

Chapman, P.M., Fairbrother, A. and Brown, D. 1998. A critical evaluation of safety (uncertainty) factors for ecological risk assessment. *Environmental toxicology and chemistry*. **17**(1), pp.99–108.

Chivers, C., Leung, B. and Yan, N.D. 2014. Validation and calibration of probabilistic predictions in ecology. *Methods in Ecology and Evolution*. **5**(10), pp.1023–1032.

Cooper, S.D. and Barmuta, L.A. 1993. Field Experiments in Biomonitoring. *In*: D. M. Rosenberg and V. H. Resh, eds. *Freshwater Biomonitoring and Benthic Macroinvertebrates*. New York: Chapman and Hall, pp.399–441.

Crane, M., Maycock, D., Watts, C., Atkinson, C. and Johnson, I. 2007. *Proposed EQS for Water Framework Directive Annex VIII substances: linuron*. Bristol, UK: Environment Agency.

Crossland, N.O., Heimbach, F., Hill, I.., Boudou, A., Leeuwangh, P., Personne, G. and P., M. 1992. *Summary and recommendations of the European Workshop on Freshwater Field Tests (EWOFFT), Potsdam, Germany, June 25–26*.

Daam, M. and Van Den Brink, P.J. 2007. Effects of chlorpyrifos, carbendazim, and linuron on the ecology of a small indoor aquatic microcosm. *Archives of Environmental Contamination and Toxicology*. **53**(1), pp.22–35.

Daam, M., Rodrigues, A.M.F., Van den Brink, P.J. and Nogueira, A.J. 2009. Ecological effects of the herbicide linuron in tropical freshwater microcosms. *Ecotoxicology and Environmental Safety*. **72**(2), pp.410–423.

Dalinsky, S.A., Lolya, L.M., Maguder, J.L., Pierce, J.L.B., Kelting, D.L., Laxson, C.L. and Patrick, D.A. 2014. Comparing the Effects of Aquatic Stressors on Model Temperate Freshwater Aquatic Communities. *Water, Air, & Soil Pollution*. **225**(6), p.2007.

Davies, K.F., Holyoak, M., Preston, K.A., Offeman, V.A. and Lum, Q. 2009. Factors controlling community structure in heterogeneous metacommunities. *Journal of Animal Ecology*. **78**(5), pp.937–944.

Dawoud, M., Bundschuh, M., Goedkoop, W. and McKie, B.G. 2017. Interactive effects

of an insecticide and a fungicide on different organism groups and ecosystem functioning in a stream detrital food web. *Aquatic Toxicology*. **186**, pp.215–221.

Demetrio, P.M., Bulus Rossini, G.D., Bonetto, C.A. and Ronco, A.E. 2012. Effects of pesticide formulations and active ingredients on the coelenterate *Hydra attenuata* (Pallas, 1766). *Bulletin of Environmental Contamination and Toxicology*. **88**(1), pp.15–19.

Dolciotti, I., Foit, K., Herkelrath, A. and Liess, M. 2014. Competition impedes the recovery of *Daphnia magna* from repeated insecticide pulses. *Aquatic Toxicology*. **147**, pp.26–31.

Van Donk, E., Prins, H., Voogd, H.M., Crum, S.J.H. and Brock, T.C.M. 1995. Effects of nutrient loading and insecticide application on the ecology of *Elodea*-dominated freshwater microcosms. I. Responses of plankton and zooplanktivorous insects. *Archiv für Hydrobiologie*. **133**, pp.417–439.

Duffy, J.E., Cardinale, B.J., France, K.E., McIntyre, P.B., Thebault, E. and Loreau, M. 2007. The functional role of biodiversity in ecosystems: incorporating trophic complexity. *Ecology Letters*. **10**(6), pp.522–538.

Eisenhauer, N., Bowker, M.A., Grace, J.B. and Powell, J.R. 2015. From patterns to causal understanding: Structural equation modeling (SEM) in soil ecology. *Pedobiologia*. **58**(2), pp.65–72.

Eisenhauer, N., Klier, M., Partsch, S., Sabais, A.C.W., Scherber, C., Weisser, W.W. and Scheu, S. 2009. No interactive effects of pesticides and plant diversity on soil microbial biomass and respiration. *Applied Soil Ecology*. **42**(1), pp.31–36.

Englert, D., Bundschuh, M. and Schulz, R. 2012. Thiacloprid affects trophic interaction between gammarids and mayflies. *Environmental Pollution*. **167**, pp.41–46.

European Commission 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. *Official Journal of the European Communities*. **L 269**(September 2000), pp.1–93.

European Commission 2009. *Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC*

and 91/414/EEC.

European Commission 2011. Regulation (EC) No 546/2011 of 10 June 2011 implementing Regulation (EC) No 1107/2009 of the European Parliament and of the Council as regards uniform principles for evaluation and authorisation of plant protection products.

European Commission 2002. Working Document Guidance Document on Aquatic Ecotoxicology in the context of Directive 91/414/EEC. *European Commission, Health and Consumer Protection Directorate General*. **4**(October), pp.1–62.

European Food and Safety Authority (EFSA) Panel on Plant Protection Products and their Residues 2013. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. *EFSA Journal*. **11**(7), p.3290.

European Union 2016. Active substances, EU - Pesticides database. *European Commission*.

Fairchild, J.F., Point, T.W. and Schwartz, T.R. 1994. Effects of an herbicide and insecticide mixture in aquatic mesocosms. *Archives of Environmental Contamination and Toxicology*. **27**(4), pp.527–533.

Fettweis, A., De Schamphelaere, K. and Smolders, E. 2018. Zinc toxicity to *Daphnia magna* in a two-species microcosm can be predicted from single-species test data: The effects of phosphorus supply and pH. *Environmental Toxicology and Chemistry*. **37**(8), pp.2153–2164.

Fleeger, J.W., Carman, K.R. and Nisbet, R.M. 2003. Indirect effects of contaminants in aquatic ecosystems. *Science of The Total Environment*. **317**(1–3), pp.207–233.

Foit, K., Kaske, O. and Liess, M. 2012. Competition increases toxicant sensitivity and delays the recovery of two interacting populations. *Aquatic Toxicology*. **106–107**, pp.25–31.

Foit, K., Kaske, O., Wahrendorf, D.S., Duquesne, S. and Liess, M. 2012. Automated Nanocosm test system to assess the effects of stressors on two interacting populations. *Aquatic Toxicology*. **109**, pp.243–249.

Freeman, L. 1953. A Standardized Method for Determining Toxicity of Pure Compounds to Fish. *Sewage and industial wastes*. **25**(7), pp.845–848.

Fulton, M.H. and Key, P.B. 2001. Acetylcholinesterase inhibition in estuarine fish and invertebrates as an indicator of organophosphorus insecticide exposure and effects. *Environmental toxicology and chemistry*. **20**(1), pp.37–45.

Fuma, S., Takeda, H., Miyamoto, K., Yanagisawa, K., Inoue, Y., Ishii, N., Sugai, K., Ishii, C. and Kawabata, Z. 2000. Simple aquatic microcosm for ecotoxicity screening at the community level. *Bulletin of Environmental Contamination and Toxicology*. **65**(6), pp.699–706.

Gama-Flores, J.L., Huidobro-Salas, M.E., Sarma, S.S.S. and Nandini, S. 2011. Somatic and population growth responses of *Ceriodaphnia dubia* and *Daphnia pulex* (Cladocera) to changes in food (*Chlorella vulgaris*) level and temperature. *Journal of Environmental Biology*. **32**(4), pp.489–495.

Garthwaite, D., Barker, I., Ridley, L., Mace, A., Parrish, G., Macarthur, R. and Lu, Y. 2017. *Pesticide Usage Survey Report 271 Arable Crops in the United Kingdom 2016*.

Gause, G.F. 1934. The Struggle for Existence. *The Yale journal of biology and medicine*. **7**(6), p.609.

Gergs, A., Zenker, A., Grimm, V. and Preuss, T.G. 2013. Chemical and natural stressors combined: from cryptic effects to population extinction. *Scientific Reports*. **3**, p.2036.

Gerland, P., Raftery, A.E., Ševčíková, H., Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B.K., Chunn, J., Lalic, N., Bay, G., Buettner, T., Heilig, G.K. and Wilmoth, J. 2014. World population stabilization unlikely this century. *Science*. **346**(6206), pp.234–237.

Giardini, J., Yan, N.D. and Heyland, A. 2015. Consequences of calcium decline on the embryogenesis and life history of *Daphnia magna*. *Journal of Experimental Biology*. **218**(13), pp.2005–2014.

Gnacadja, L. 2012. *Moving to Zero-Net Rate of Land Degradation*. United Nations Convention to Combat Desertification.

Goodman, S.N., Fanelli, D. and Ioannidis, J.P.A. 2016. What does research reproducibility mean? *Science Translational Medicine*. **8**(341), 341ps12.

Grover, J.P. 2000. Resource competition and community structure in aquatic micro-

organisms: experimental studies of algae and bacteria along a gradient of organic carbon to inorganic phosphorus supply. *Journal of Plankton Research*. **22**(8), pp.1591–1610.

Habetha, M., Anton-Erxleben, F., Neumann, K. and Bosch, T.C. 2003. The *Hydra viridis/Chlorella symbiosis*. Growth and sexual differentiation in polyps without symbionts. *Zoology*. **106**(2), pp.101–108.

Hanazato, T. 2001. Pesticide effects on freshwater zooplankton: an ecological perspective. *Environmental Pollution*. **112**, pp.1–10.

Hanson, M.L., Wolff, B.A., Green, J.W., Kivi, M., Panter, G.H., Warne, M.S.J., Ågerstrand, M. and Sumpter, J.P. 2017. How we can make ecotoxicology more valuable to environmental protection. *Science of the Total Environment*. **578**, pp.228–235.

Harrass, M.C. and Taub, F.B. 1985. Effects of small fish predation on microcosm community bioassays *In*: R. D. Cardwell, R. Purdy and R. C. Bahner, eds. *Aquatic Toxicology and Hazard Assessment: Seventh Symposium, ASTM STP 854*. Philadelphia: American Society for Testing and Materials, pp.117–133.

Harris, C.A., Scott, A.P., Johnson, A.C., Panter, G.H., Sheahan, D., Roberts, M. and Sumpter, J.P. 2014. Principles of sound ecotoxicology. *Environmental Science and Technology*. **48**(6), pp.3100–3111.

Harris, C.A. and Sumpter, J.P. 2015. Could the Quality of Published Ecotoxicological Research Be Better? *Environmental Science and Technology*. **49**(16), pp.9495–9496.

Harris, M.J., Huggett, D.B., Staveley, J.P. and Sumpter, J.P. 2017. What training and skills will the ecotoxicologists of the future require? *Integrated Environmental Assessment and Management*. **13**(4), pp.580–584.

Heugens, E.H.W., Hendriks, A.J., Dekker, T., Straalen, N.M. van and Admiraal, W. 2001. A Review of the Effects of Multiple Stressors on Aquatic Organisms and Analysis of Uncertainty Factors for Use in Risk Assessment. *Critical Reviews in Toxicology*. **31**(3), pp.247–284.

Hines, W.C., Su, Y., Kuhn, I., Polyak, K. and Bissell, M.J. 2014. Sorting out the FACS: A devil in the details. *Cell Reports*. **6**(5), pp.779–781.

Van Hoecke, K., De Schamphelaere, K.A.C., Van Der Meeren, P., Lucas, S. and Janssen, C.R. 2008. Ecotoxicity of Silica Nanoparticles to the green alga *Pseudokirchneriella subcapitata*: importance of surface area. *Environmental toxicology and chemistry*. **27**(9), pp.1948–1957.

Hommen, U., Baveco, H., Galic, N. and Brink, P.J. Van Den 2010. Potential Application of Ecological Models in the European Environmental Risk Assessment of Chemicals I: Review of Protection Goals in EU Directives and Regulations. *Integrated environmental assessment and management*. **6**(3), pp.325–337.

Hommen, U., Forbes, V., Grimm, V., Preuss, T.G., Thorbek, P. and Ducrot, V. 2016. How to use mechanistic effect models in environmental risk assessment of pesticides: Case studies and recommendations from the SETAC workshop MODELINK. *Integrated Environmental Assessment and Management*. **12**(1), pp.21–31.

Huffaker, C.B. 1958. Experimental Studies on Predation: Dispersion Factors and Predator-Prey Oscillations. *Hilgardia*. **27**(14), pp.795–835.

Hunter, P. 2017. The reproducibility 'crisis'. *EMBO reports*. **18**(9), pp.1493–1496.

Hushon, J., Clerman, R. and Wagner, B. 1979. Tiered testing for chemical hazard assessment. *Environmental Science and Technology*. **13**, pp.1202–1206.

Ioannidis, J.P.A. 2014. How to Make More Published Research True. *PLoS Medicine*. **11**(10), p.e1001747.

Ioannidis, J.P.A. 2005. Why most published research findings are false. *PLoS Medicine*. **2**(8), p.e28.

Ioannidis, J.P.A., Greenland, S., Hlatky, M.A., Khoury, M.J., Macleod, M.R., Moher, D., Schulz, K.F. and Tibshirani, R. 2014. Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet*. **383**(9912), pp.166–175.

Jarvis, A.L., Bernot, M.J. and Bernot, R.J. 2014. The effects of the psychiatric drug carbamazepine on freshwater invertebrate communities and ecosystem dynamics. *Science of The Total Environment*. **496**, pp.461–470.

Kampe, H., König-Rinke, M., Petzoldt, T. and Benndorf, J. 2007. Direct effects of *Daphnia*-grazing, not infochemicals, mediate a shift towards large inedible colonies of the gelatinous green alga *Sphaerocystis schroeteri*. *Limnologica*.

37(2), pp.137–145.

Kattwinkel, M., Liess, M., Arena, M., Bopp, S., Streissl, F. and Römbke, J. 2015. Recovery of aquatic and terrestrial populations in the context of European pesticide risk assessment. *Environmental Reviews*. **23**(4), pp.382–394.

Kovacevic, G., Davor, Ž., Horvatin, K. and Kalafati, M. 2007. Morphological features and comet assay of green and brown hydra treated with aluminium. *Symbiosis*. **44**, pp.145–152.

Kovacevic, G., Kalafati, M. and Ljubesic, N. 2009. Effects of Norflurazon on Green and Brown Hydra. *Folia biologica (Kraków)*. **57**(1–2), pp.91–96.

Kraufvelin, P. 1998. Model ecosystem replicability challenged by the 'soft' reality of a hard bottom mesocosm. *Journal of Experimental Marine Biology and Ecology*. **222**, pp.247–267.

Ladyman, J., Lambert, J. and Wiesner, K. 2013. What is a complex system? *European Journal for Philosophy of Science*. **3**(1), pp.33–67.

De Laender, F., De Schamphelaere, K.A.C., Vanrolleghem, P.A. and Janssen, C.R. 2008. Validation of an ecosystem modelling approach as a tool for ecological effect assessments. *Chemosphere*. **71**(3), pp.529–545.

De Laender, F., Taub, F.B. and Janssen, C.R. 2011. Ecosystem functions and densities of contributing functional groups respond in a different way to chemical stress. *Environmental Toxicology and Chemistry*. **30**(12), pp.2892–2898.

Landis, W.G. 2003. Twenty years before and hence; Ecological risk assessment at multiple scales with multiple stressors and multiple endpoints. *Human and Ecological Risk Assessment*. **9**(5), pp.1317–1326.

Landis, W.G., Matthews, R.A. and Matthews, G.B. 1997. Design and analysis of multispecies toxicity tests for pesticide registration. *Ecological Applications*. **7**(4), pp.1111–1116.

Lasserre, P. 1990. Marine Microcosms: Small-Scale Controlled Ecosystems *In*: C. M. Lalli, ed. *Enclosed Experimental Marine Ecosystems: A Review and Recommendations. Coastal and Estuarine Studies*. Coastal and Estuarine Studies. New York, NY: Springer, pp.20–60.

Lee, J. and Kim, G. 2006. A simple and rapid method for analyzing radon in coastal

and ground waters using a radon-in-air monitor. *Journal of Environmental Radioactivity*. **89**(3), pp.219–228.

Leek, J.T. and Jager, L.R. 2017. Is Most Published Research Really False? *Annual Review of Statistics and Its Application*. **4**(1), pp.109–122.

Lewis, K.A., Tzilivakis, J., Warner, D.J. and Green, A. 2016. An international database for pesticide risk assessments and management. *Human and Ecological Risk Assessment*. **22**(4), pp.1050–1064.

Liebig, M., Schmidt, G., Bontje, D., Kooi, B.W., Streck, G., Traunspurger, W. and Knacker, T. 2008. Direct and indirect effects of pollutants on algae and algivorous ciliates in an aquatic indoor microcosm. *Aquatic Toxicology*. **88**(2), pp.102–110.

Liess, M. 2002. Population response to toxicants is altered by intraspecific interaction. *Environmental toxicology and chemistry*. **21**(1), pp.138–142.

Lithgow, G.J., Driscoll, M. and Phillips, P. 2017. A long journey to reproducible results. *Nature News*. **548**(7668), p.387.

Locke, A. and Sprules, W.G. 2000. Effects of acidic ph and phytoplankton on survival and condition of *Bosmina longirostris* and *Daphnia pulex*. *Hydrobiologia*. **437**, pp.187–196.

MacDonald, J. 2016. How Microcosms Help Us Understand Ecology. *JSTOR Daily*. [Online]. [Accessed 10 November 2018]. Available from: https://daily.jstor.org/how-microcosms-help-us-understand-ecology/.

Malaj, E., von der Ohe, P.C., Grote, M., Kühne, R., Mondy, C.P., Usseglio-Polatera, P., Brack, W. and Schäfer, R.B. 2014. Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. *Proceedings of the National Academy of Sciences of the United States of America*. **111**(26), pp.9549–54.

Martin, B.T., Jager, T. and Nisbet, R. 2014. Limitations of extrapolating toxic effects on reproduction to the population level. *Ecological Applications*. **24**(8), pp.1972–1983.

Martin, B.T., Jager, T., Nisbet, R.M., Preuss, T.G. and Grimm, V. 2013. Predicting Population Dynamics from the Properties of Individuals: A Cross-Level Test of Dynamic Energy Budget Theory. *The American Naturalist*. **181**(4), pp.506–519.

Massaro, F.C., Negreiros, N.F. and Rocha, O. 2013. A search for predators and food selectivity of two native species of *Hydra* (Cnidaria: Hydrozoa) from Brazil. *Biota Neotropica*. **13**, pp.35–40.

Massaro, F.C. and Rocha, O. 2008. Development and population growth of *Hydra viridissima* Pallas, 1766 (Cnidaria, Hydrozoa) in the laboratory. *Brazilian journal of biology*. **68**(2), pp.379–383.

Melov, S., Ravenscroft, J., Malik, S., Gill, M.S., Walker, D.W., Clayton, P.E., Wallace, D.C., Malfroy, B., Doctrow, S.R. and Lithgow, G.J. 2000. Extension of life-span with superoxide dismutase/catalase mimetics. *Science*. **289**(5484), pp.1567–1569.

Metcalf, R.L., Gurcharan, K.S. and Inder, P.K. 1971. Model Ecosystem for the Evaluation of Pesticide Biodegradability and Ecological Magnification. *Environmental Science and Technology*. **5**(8), pp.709–713.

Moe, S.J., De Schamphelaere, K., Clements, W.H., Sorensen, M.T., Van den Brink, P.J. and Liess, M. 2013. Combined and interactive effects of global climate change and toxicants on populations and communities. *Environmental Toxicology and Chemistry*. **32**(1), pp.49–61.

Munger, C., Hare, L., Craig, A. and Charest, P.M. 1998. Influence of exposure time on the distribution of cadmium within the cladoceran *Ceriodaphnia dubia*. *Aquatic Toxicology*. **44**(3), pp.195–200.

Organisation for Economic Co-operation and Development 2005. Giudance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. *OECD series on testing and assessment*. **34**(90), pp.1–96.

Organisation for Economic Co-operation and Development 2011. Guideline 201: Freshwater Alga and Cyanobacteria, Growth Inhibition Test *In*: *OECD guidelines for the Testing of Chemicals*. Paris, France, pp.1–25.

Organisation for Economic Co-operation and Development 2004. Guideline 202: *Daphnia sp*. Acute Immobilisation Test *In*: *OECD Guidelines for the Testing of Chemicals*. Paris, France, pp.1–12.

Organisation for Economic Co-operation and Development (OECD) 2015. *Decision of*

the Council concerning the Mutual Acceptance of Data in the Assessment of Chemicals.

Organisation for Economic Co-operation and Development (OECD) 2012. Guideline 211: *Daphnia magna* Reproduction Test *In*: *OECD Guidelines for the Testing of Chemicals*. Paris, France, pp.1–21.

Pachura-Bochet, S., Blaise, C. and Vasseur, P. 2006. Toxicity of Nonylphenol on the Cnidarian *Hydra attenuata* and Environmental Risk Assessment. *Environmental toxicology*. **21**(4), pp.388--394.

Pavlić, Ž., Vidaković-Cifrek, Ž. and Puntarić, D. 2005. Toxicity of surfactants to green microalgae *Pseudokirchneriella subcapitata* and *Scenedesmus subspicatus* and to marine diatoms *Phaeodactylum tricornutum* and *Skeletonema costatum*. *Chemosphere*. **61**(8), pp.1061–1068.

Van de Perre, D., Roessink, I., Janssen, C.R., Smolders, E., De Laender, F., Van Den Brink, P.J. and De Schamphelaere, K.A.C. 2018. The Combined and Interactive Effects of Zinc, Temperature and Phosphorus on the Structure and Functioning of a Freshwater Community. *Environment Toxicology and Chemistry*. **37**(9), pp.2413–2427.

Peters, K., Bundschuh, M. and Schäfer, R.B. 2013. Review on the effects of toxicants on freshwater ecosystem functions. *Environmental Pollution*. **180**, pp.324–329.

Poisot, T., Stouffer, D.B. and Gravel, D. 2015. Beyond species: Why ecological interaction networks vary through space and time. *Oikos*. **124**(3), pp.243–251.

Pratt J. R., Melendez A. E., Barreiro R., B.N.J. 1997. Predicting the Ecological Effects of Herbicides. *Ecological Applications*. **7**(4), pp.1117–1124.

Preston, B.L. 2002. Indirect Effects in Aquatic Ecotoxicology: Implications for Ecological Risk Assessment. *Environmental Management*. **29**(3), pp.311–323.

Preuss, T.G., Hammers-Wirtz, M. and Ratte, H.T. 2010. The potential of individual based population models to extrapolate effects measured at standardized test conditions to relevant environmental conditions-an example for 3,4-dichloroaniline on *Daphnia magna*. *Journal of environmental monitoring*. **12**, pp.2070–2079.

Quinn, B., Gagne, F. and Blaise, C. 2012. Hydra, a model system for environmental

studies. *International Journal of Developmental Biology*. **56**(6–8), pp.613–625.

Quinn, B., Gagné, F. and Blaise, C. 2008. An investigation into the acute and chronic toxicity of eleven pharmaceuticals (and their solvents) found in wastewater effluent on the cnidarian, *Hydra attenuata*. *Science of the Total Environment*. **389**(2–3), pp.306–314.

Ramanan, R., Kim, B.H., Cho, D.H., Oh, H.M. and Kim, H.S. 2016. Algae-bacteria interactions: Evolution, ecology and emerging applications. *Biotechnology Advances*. **34**(1), pp.14–29.

Raven, J.A., Evans, M.C.W. and Korb, R.E. 1999. The role of trace metals in photosynthetic electron transport in O_2-evolving organisms. *Photosynthesis Research*. **60**(2–3), pp.111–149.

Relyea, R. a. and Hoverman, J.T. 2008. Interactive effects of predators and a pesticide on aquatic communities. *Oikos*. **117**(11), pp.1647–1658.

Relyea, R. and Hoverman, J. 2006. Assessing the ecology in ecotoxicology: A review and synthesis in freshwater systems. *Ecology Letters*. **9**(10), pp.1157–1171.

Rendal, C., Trapp, S. and Kusk, K.O. 2012. Critical evaluation and further development of methods for testing ecotoxicity at multiple pH using *Daphnia magna* and *Pseudokirchneriella subcapitata*. *Environmental Toxicology and Chemistry*. **31**(8), pp.1843–1852.

Riedl, V., Agatz, A., Benstead, R. and Ashauer, R. 2018. A standardized tritrophic small-scale system (TriCosm) for the assessment of stressor-induced effects on aquatic community dynamics. *Environmental Toxicology and Chemistry*. **37**(4), pp.1051–1060.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., de Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P. and Foley, J. 2009. Planetary boundaries: Exploring the safe operating space for humanity. *Ecology and Society*. **14**(2), p.32.

Rodgher, S., Lombardi, A.T., Da Graça Gama Melão, M. and Tonietto, A.E. 2008.

Change in life cycle parameters and feeding rate of *Ceriodaphnia silvestrii* Daday (Crustacea, Cladocera) exposure to dietary copper. *Ecotoxicology*. **17**(8), pp.826–833.

Rodgher, S. and Luiz Gaeta Espíndola, E. 2008. Effects of interactions between algal densities and cadmium concentrations on *Ceriodaphnia dubia* fecundity and survival. *Ecotoxicology and Environmental Safety*. **71**(3), pp.765–773.

Rohr, J.R., Kerby, J.L. and Sih, A. 2006. Community ecology as a framework for predicting contaminant effects. *Trends in Ecology and Evolution*. **21**(11), pp.606–613.

Rose, R.M., Warne, M.S.J. and Lim, R.P. 2002a. Food Concentration Affects the Life History Response of *Ceriodaphnia cf. dubia* to Chemicals with Different Mechanisms of Action. *Ecotoxicology and Environmental Safety*. **51**(2), pp.106–114.

Rose, R.M., Warne, M.S.J. and Lim, R.P. 2002b. Some life history responses of the cladoceran *Ceriodaphnia cf. dubia* to variations in population density at two different food concentrations. *Hydrobiologia*. **481**, pp.157–164.

Sanderson, H. 2002. Pesticide Studies: Replicability of Micro/Mesocosms. *Environmental science and pollution research international*. **9**(6), pp.429–435.

Schäfer, R.B., Kühn, B., Malaj, E., König, A. and Gergs, R. 2016. Contribution of organic toxicants to multiple stress in river ecosystems. *Freshwater Biology*. **61**(12), pp.2116–2128.

Scheffer, M. and Carpenter, S.R. 2003. Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends Ecol Evol*. **18**(12), pp.648–656.

Schön, I., Martena, K. and Van Dijk, P. (eds.). 2009. *Lost Sex The Evolutionary Biology of Parthenogenesis*. Netherlands: Springer.

Scientific Committee on Health Environmental Risks, S., Scientific Committee on Emerging and Newly Identified Health Risks, S. and Scientific Committee on Consumer Safety, S. 2013. *Addressing the New Challenges for Risk Assessment*. Brussels, Belgium: European Commission.

Šimičev, A., Kovačević, G. and Gračan, R. 2016. *Ecotoxicological effect of sodium metasilicate on green (Hydra viridissima Pallas, 1766) and brown Hydra (Hydra*

oligactis Pallas, 1766). PhD thesis. University of Zagreb.

Singh, S.P. and Singh, P. 2015. Effect of temperature and light on the growth of algae species: A review. *Renewable and Sustainable Energy Reviews*. **50**, pp.431–444.

Sokal, R.R. and Rohlf, J.F. 2012. *Biometry The Principles and Practices of Statistics in Biological Research* 4th ed. (J. Correa, ed.). New York: W. H. Freeman and Company.

Spycher, S., Mangold, S., Doppler, T., Junghans, M., Wittmer, I., Stamm, C. and Singer, H. 2018. Pesticide Risks in Small Streams - How to Get as Close as Possible to the Stress Imposed on Aquatic Organisms. *Environmental Science and Technology*. **52**(8), pp.4526–4535.

Srivastava, D.S., Kolasa, J., Bengtsson, J., Gonzalez, A., Lawler, S.P., Miller, T.E., Munguia, P., Romanuk, T., Schneider, D.C. and Trzcinski, M.K. 2004. Are natural microcosms useful model systems for ecology? *Trends in Ecology and Evolution*. **19**(7), pp.379–384.

Stampfli, N.C., Knillmann, S., Liess, M. and Beketov, M. a. 2011. Environmental context determines community sensitivity of freshwater zooplankton to a pesticide. *Aquatic Toxicology*. **104**(1–2), pp.116–124.

Stampfli, N.C., Knillmann, S., Liess, M., Noskov, Y. a., Schäfer, R.B. and Beketov, M. a. 2013. Two stressors and a community - Effects of hydrological disturbance and a toxicant on freshwater zooplankton. *Aquatic Toxicology*. **127**, pp.9–20.

Stehle, S. and Schulz, R. 2015. Agricultural insecticides threaten surface waters at the global scale. *Proceedings of the National Academy of Sciences*. **112**(18), pp.5750–5755.

Stewart, A. and Konetsky, B. 2008. Longevity and reproduction of *Ceriodaphnia dubia* in receiving waters. *Environment Toxicology and Chemistry*. **17**(6), pp.1165–1171.

Van Straalen, N.M. 2003. Ecotoxicology becomes stress ecology. *Environmental science and technology*. **37**, pp.324–330.

Swartzman, G., Rose, K., Kindig, A. and Taub, F. 1989. Modeling the direct and indirect effects of streptomycin in aquatic microcosms. *Aquatic Toxicology*. **14**(2), pp.109–129.

Taub, F.B. 1997a. Are Ecological Studies Relevant to Pesticide Registration Decisions? *Ecological Applications*. **7**(4), pp.1083–1085.

Taub, F.B. 1976. Demonstration of pollution effects in aquatic microcosms. *International Journal of Environmental Studies*. **10**, pp.23–33.

Taub, F.B. 1989. Standardized Aquatic Microcosms. *Environmental Science and Technology*. **23**(9), pp.1064–1066.

Taub, F.B. 1993. Standardizing an aquatic microcosm test *In*: P. Soares, AMV and Calow, ed. *Progress in Standardization of Aquatic Toxicity Tests*. SETAC Special Publications Series. Lewis Publishers Inc, pp.159–188.

Taub, F.B. 1997b. Unique Information contributed by Multispecies Systems: Examples from the Standardized Aquatic Microcosm. *Ecological Applications*. **7**(4), pp.1103–1110.

Taub, F.B., Kindig, A.C. and Conquest, L.L. 1986. Preliminary Results of Interlaboratory Testing of a Standardized Aquatic Microcosm *In*: C. John Jr., ed. *Community Toxicity Testing.* Philadelphia, Pennsylvania, USA: ASTM International, pp.93–120.

Taub, F.B. and McLaskey, A.K. 2013. Pressure, O_2, and CO_2, in aquatic Closed Ecological Systems. *Advances in Space Research*. **51**(5), pp.812–824.

Tökölyi, J., Rosa, M., Bradács, F. and Barta, Z. 2014. Life history trade-offs and stress tolerance in green hydra (*Hydra viridissima* Pallas 1766): the importance of nutritional status and perceived population density. *Ecological Research*. **29**(5), pp.867–876.

Uhl, P., Bucher, R., Schäfer, R.B. and Entling, M.H. 2015. Sublethal effects of imidacloprid on interactions in a tritrophic system of non-target species. *Chemosphere*. **132**, pp.152–158.

United States Environmental Protection Agency (USEPA) 2012. *Ecological Effects Test Guidelines, OCSPP 850.4500: Algal Toxicity*.

United States Environmental Protection Agency (USEPA) 1996. *Ecological Effects Test Guidelines OPPTS 850.1925 Site-Specific Aquatic Microcosm Test, Laboratory*.

United States Environmental Protection Agency (USEPA) 2002. *Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to*

Freshwater Organisms. Washington, DC.

Vasquez, M.I., Lambrianides, A., Schneider, M., Kummerer, K. and Fatta-Kassinos, D. 2014. Environmental side effects of pharmaceutical cocktails: what we know and what we should know. *Journal of Hazardous Materials*. **279**, pp.169–189.

Viaene, K.P.J., De Laender, F., Rico, A., Van den Brink, P.J., Di Guardo, A., Morselli, M. and Janssen, C.R. 2015. Species interactions and chemical stress: Combined effects of intraspecific and interspecific interactions and pyrene on *Daphnia magna* population dynamics. *Environmental Toxicology and Chemistry*. **34**(8), pp.1751–1759.

Warne, M.S.J. and Schifko, A.D. 1999. Toxicity of Laundry Detergent Components to a Freshwater Cladoceran and Their Contribution to Detergent Toxicity. *Ecotoxicology and Environmental Safety*. **44**, pp.196–206.

Webster, U.T.M., Perry, M.H. and Santos, E.M. 2015. The herbicide linuron inhibits cholesterol biosynthesis and induces cellular stress responses in brown trout. *Environmental Science and Technology*. **49**(5), pp.3110–3118.

Van Wijngaarden, R.P., Arts, G.H.P., Belgers, J.D.M., Boonstra, H., Roessink, I., Schroer, a. F.W. and Brock, T.C.M. 2010. The species sensitivity distribution approach compared to a microcosm study: A case study with the fungicide fluazinam. *Ecotoxicology and Environmental Safety*. **73**(2), pp.109–122.

Van Wijngaarden, R.P., Brock, T.C.M. and Douglas, M.T. 2005. Effects of chlorpyrifos in freshwater model ecosystems: The influence of experimental conditions on ecotoxicological thresholds. *Pest Management Science*. **61**(10), pp.923–935.

Wilby, O.K. 1988. The *Hydra* regeneration assay *In*: *Proceedings of workshop organised by Association Francaise de Teratologie*., pp.108–124.

Yamagishi, T., Yamaguchi, H., Suzuki, S., Horie, Y. and Tatarazako, N. 2017. Cell reproductive patterns in the green alga *Pseudokirchneriella subcapitata* (=*Selenastrum capricornutum*) and their variations under exposure to the typical toxicants potassium dichromate and 3,5-DCP. *PLoS ONE*. **12**(2), pp.1–12.