Null model detection of multiple stressor interactions in aquatic ecosystems

Benjamin Joshua Burgess

A thesis submitted for the degree of

Doctor of Philosophy of University College London (UCL)

August 2021

Primary supervisor

David J. Murrell	(UCL)

Secondary supervisors

Drew W. Purves	(DeepMind)
Georgina M. Mace	(UCL)

Institution

Centre for Biodiversity and Environment Research (CBER)

Department of Genetics, Evolution, and Environment,

University College London,

London,

WC1E 6BT

I, Benjamin Joshua Burgess, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Where co-authors have contributed to specific chapters, this has been detailed in the thesis outline.

Abstract

Ecosystems are subjected to a wide range of stressors, many of which are anthropogenic in origin. However, far from being impacted by isolated threats, these ecosystems are affected by multiple co-occurring stressors. Currently, there is little understanding of how, or whether, these stressors interact to affect individuals, populations, or communities. Indeed, studies vary in whether they find co-occurring stressors to interact in an additive, antagonistic, or synergistic manner. However, attempts to determine general ecological covariables which may explain these disparate findings have so far failed to do so. Here, I use meta-analytical and theoretical approaches to better understand how stressors can be expected to interact, with a particular emphasis on freshwater ecosystems. Firstly, I simulate food chains which are subjected to co-occurring stressors and compare these results to the findings of the largest multiple stressor meta-analysis, here focusing on freshwater densities. Both approaches illustrate that null (i.e., additive) classifications dominate for individual interactions; although, overall stressors interact to affect density in an antagonistic manner. Secondly, I analyse the statistical tools frequently used to classify multiple stressor interactions. I illustrate that many results which are ascribed ecological importance instead arise due to statistical artefacts of these tools. In turn, I highlight that many experimental designs, commonplace to multiple stressor ecology, lack the statistical power necessary to detect the interactions of co-occurring stressors. Thirdly, I collate and analyse the datasets of seven aquatic multiple stressor meta-analyses under a single consistent framework. I illustrate that the current absence of generalities from multiple stressor meta-analyses primarily arises due to methodological, not ecological, variation. In turn, removing methodological differences results in generalities becoming apparent. Finally, I collate the findings of the above chapters and outline the potential implications for multiple stressor ecology. In doing so, I explore current challenges facing the field alongside future directions.

Impact statement

Given that multiple stressors affect nearly all ecosystems, it is imperative that the interactions of these stressors are understood. However, there is currently limited knowledge of stressor interactions, with similar studies often finding highly disparate results. Accordingly, there is an absence of consistent generalities apparent across multiple stressor ecology. This has ramifications both for the field and the insights which multiple stressor ecology can provide to allied disciplines (e.g., conservation science).

This thesis utilises a range of different approaches to provide novel insights into the interactions of co-occurring stressors. Ultimately, all of the chapters of this thesis seek to answer the question of how stressors interact. Here, I implement various empirical, theoretical, and analytical approaches to address this question, often from new perspectives. Indeed, I consider how ecological, experimental, and methodological properties may influence the conclusions researchers reach regarding interacting stressors. Overall, the findings of this thesis have implications for experimental designs, analytical choices, and the interpretation of any results. In turn, such findings may also be subsequently considered by conservationists, as understanding how stressors interact is crucial when implementing management actions.

Each of the chapters of this thesis substantially builds on existing research. For example, Chapter Two (recently published in *Global Change Biology*) combines meta-analytical and theoretical approaches to better understand stressor interactions. In doing so, I highlight the value of implementing ecological theory within multiple stressor ecology, a currently under-utilised approach. Likewise, Chapter Three is the first substantial exploration of the statistical tools frequently used to classify stressor interactions. In doing so, I illustrate that the design of many (if not most) experiments within multiple stressor ecology is inadequate when seeking to determine whether, or how, stressors interact. Finally, Chapter Four is the first analysis in which the results of different multiple stressor meta-analyses are compared. This analysis has revealed that methodology, not ecology, is likely responsible for the absence of generalities within multiple stressor ecology. As such, this chapter has wide-ranging consequences for the field of multiple stressor ecology. Indeed, I illustrate that an advanced ecological understanding of multiple stressor interactions can only be gained once analytical methods are understood and correctly implemented.

Table of contents

Main figures and tables

Chapter One

Chapter Two

Chapter Three

Chapter Four

Thesis outline

Chapter One - *Introduction*

This chapter provides background information on the current understanding of how stressors interact to affect ecosystems. Literature gaps and questions in the field of multiple stressor ecology are highlighted alongside the aims of this thesis. This chapter is primarily general or freshwater-centric in its focus.

Chapter Two - *Classifying ecosystem stressor interactions: Theory highlights the data limitations of the additive null model and the difficulty in revealing ecological surprises*

In this chapter, I combine theoretical ecology with a large-scale freshwater meta-analysis to better understand multiple stressor interactions and determine whether general insights could be determined from across these different approaches. This work was conducted in collaboration with David Murrell, Drew Purves, and Georgina Mace. The study was conceived by myself and DM; I conducted the analysis and wrote the first draft. All collaborators contributed to the editing of this chapter and manuscript. Work from this chapter has been presented at various conferences including multiple British Ecological Society (BES) conferences and BES Aquatic Special Interest Group meetings. This chapter has been recently published in *Global Change Biology* and in subsequent chapters may be referred to as Burgess et al., (2021).

Chapter Three - *Multiple stressor null models frequently fail to detect interactions due to low statistical power*

In this chapter, I examine the additive and multiplicative null models from a statistical perspective, determining how the results of these approaches may vary under experimentally relevant conditions (e.g., sample sizes). This work was conducted in collaboration with DM and Michelle Jackson (University of Oxford). Myself and DM conceived the idea for this project and designed methods. I conducted the analysis and wrote the chapter with input from DM and MJ.

Chapter Four - *Methodological variation obscures generalities across multiple stressor meta-analyses*

In this chapter I analyse the influence which differing methods have upon the conclusions of seven multiple stressor meta-analyses. In doing so, I attempt to outline whether this source of variation is responsible for the absence of generalities to emerge from these studies. This work was primarily conducted in collaboration with DM, MJ, and Charlie Loewen, with DP and GM additionally providing advice. Data for this analyses was also provided by MJ, Kristy Kroeker, Ben Halpern, Rachel Przeslawski, Ben Harvey, and Kristina Lange. This study was conceived by myself, DM, and MJ. I conducted all analyses following discussions with DM, MJ, and CL. I wrote this chapter with input from DM.

Chapter Five - *Discussion*

This chapter collates the findings of this thesis and evaluates the contribution it makes to multiple stressor ecology. Furthermore, this chapter outlines future directions for the field and current obstacles that must be overcome.

Acknowledgements

Firstly, a thank you to my supervisors. David Murrell, for discussions, suggestions, and constant support over many wide-ranging obstacles. Drew Purves, for encouragement and big picture perspectives throughout this time. Georgina Mace, for critical insights, both general and ecological, though much missed over the last year. Likewise, my thanks to Tim Newbold for acting as my thesis chair and Michelle Jackson for her advice and insights on multiple stressors across this thesis.

Thank you to both the National Environment Research Council (NERC) and University College London for supporting this research, each over several years.

My thanks to all those in CBER for making the last few years an enjoyable and interesting experience, it is hard to imagine a better place to be a PhD student. In particular, thanks to Amy and Chris for anything and everything UCL related, and to fellow PhDs including Joe, Bo, Sean, and Adam for conversations and more besides, whether ecological or otherwise. Likewise, my thanks to all those in CBER who kept spirits high, or at the very least came to the bar and kept the spirits flowing.

Outside of CBER, my thanks to my friends both from London and further afield for their support. A special mention to those from Nottingham, particularly over the last eighteen months, and to Brett both for his encouragement and getting me involved with his wild tulips project.

Finally, a heartfelt thank you to my family, without whom none of this would have been possible.

Chapter One – Introduction

Humans have altered both ecological and environmental systems over many hundreds of years (Lewis & Maslin, 2015), leading to wide-spread impacts to ecosystems across the globe (Dirzo et al., 2014). However, following a recent intensification of anthropogenic activities, ecosystems are subjected to a plethora of different stressors (Best, 2019; Brook et al., 2008; Halpern et al., 2015). Stressors (sensu drivers, factors, or threats) are external drivers that can affect both biodiversity and ecosystem functioning (Orr et al., 2020). Although stressors may occur due to natural processes, they are frequently anthropogenic in origin (Beauchesne et al., 2021). Furthermore, the impacts of these stressors can be severe. Indeed, anthropogenically induced climate change is likely responsible for increasing the extent of forest fires (Abatzoglou & Williams, 2016), while ocean acidification has been shown to enable coral reefs to be dominated by macroalgae (Enochs et al., 2015). Similarly, water abstraction has resulted in the shrinking of the freshwater Aral Sea (Dudgeon, 2019), while mass mortality events are occurring, across realms, at an increased frequency as a consequence of a variety of stressors (Fey et al., 2015). However, since the 1990s there has been an appreciation that not only do stressors damage *nature* (e.g., natural habitat intactness) but also the goods and services which ecosystems provide (Mace, 2014). These services are essential, including crop pollination (Rader et al., 2014) and the provisioning of food resources (Bernhardt & O'Connor, 2021).

While covering 2.3% of the Earth's surface (Lehner & Döll, 2004) and containing approximately one third of all vertebrate species (Tickner et al., 2020), freshwater ecosystems are some of the most threatened on the planet (He et al., 2019). On average, freshwater species have experienced a decline in population abundances greater than those of species in either the terrestrial or marine realms (Reid et al., 2019; Sala et al., 2000; WWF, 2020). Furthermore, population abundances of freshwater megafauna have declined by up to 99% in some regions (He et al., 2019). As such, freshwater ecosystems are subjected to a wide range of anthropogenically induced stressors (Dudgeon et al., 2006; Reid et al., 2019). Examples of such threats include global stressors such as climate change (Heino et al., 2009; Knouft & Ficklin, 2017), and more localised stressors including pollution (Malaj et al., 2014;

Woodward et al., 2012), non-native species (Erős et al., 2020; Gallardo et al., 2016), habitat alteration (Grill et al., 2019; Reidy Liermann et al., 2012) or infectious diseases (Fisher & Garner, 2020; Johnson & Paull, 2011). Indeed, stressors are capable of affecting all levels of ecological organisation, from organisms through to ecosystems and their functioning (Li et al., 2020; Woodward et al., 2010). Similarly, stressors may act through differing physiological pathways (Segner et al., 2014), though knowledge of mechanisms through which stressors act is often limited (Spears et al., 2021). Whilst the majority of research has, to date, focussed on the effects of individual stressors (Jackson et al., 2021; O'Brien et al., 2019), it is increasingly evident that aquatic ecosystems are predominately acted upon by multiple co-occurring stressors (Ormerod et al., 2010; Halpern et al., 2015; Reid et al., 2019). Indeed, 50% of European surface freshwater bodies are affected by multiple stressors (EEA, 2018); though these stressors may induce either a positive or negative effect upon a species (Kroeker et al., 2017). For instance, nutrient enrichment (i.e., eutrophication) may increase the biomass of algae (i.e., a beneficial effect for algal species) (Wurtsbaugh et al., 2019), though this algal bloom may subsequently increase the mortality of other species (e.g., fishes) through a reduction in oxygen levels (i.e., a detrimental effect for non-algal species) (Anderson, 2009; Breitburg et al., 2018).

The presence of multiple stressors gives rise to a number of serious issues. Firstly, how do ecosystems respond to the presence of multiple stressors? For instance, can the effect of two co-occurring stressors on a species' density simply be calculated by summing the effects of each stressor individually? The assumption of additivity regarding stressor effects has been widely adopted, with any subsequently observed deviation from additivity being termed an 'ecological surprise' (sensu Paine et al., 1998). However, there is growing evidence that such ecological surprises are more widespread than previously thought (Jackson et al., 2016), with such ecological surprises capable of manifesting in three different ways (Figure 1.1). Firstly, the effects of multiple stressors can be greater than anticipated under the assumption of additivity (hence termed the additive null model), with a *synergistic* interaction occurring between stressors. Secondly, the combined effect of co-occurring stressors can be less than anticipated by the additive null model, with an *antagonistic* interaction occurring between stressors. Finally, there may be instances where the observed effect of multiple stressors is of a different polarity to the expected effect under the additive null model (e.g., the additive null model predicts that the combined effect of two stressors will have a negative effect on a population, but the observed effect of the co-occurring stressors is instead positive). In such a situation, a *reversal* interaction is occurring between the stressors. Where the assumption of additivity is met, this is referred to as a *null* interaction (although this may be referred to as an *additive* interaction elsewhere, e.g., Jackson et al., 2016; Orr et al., 2020).

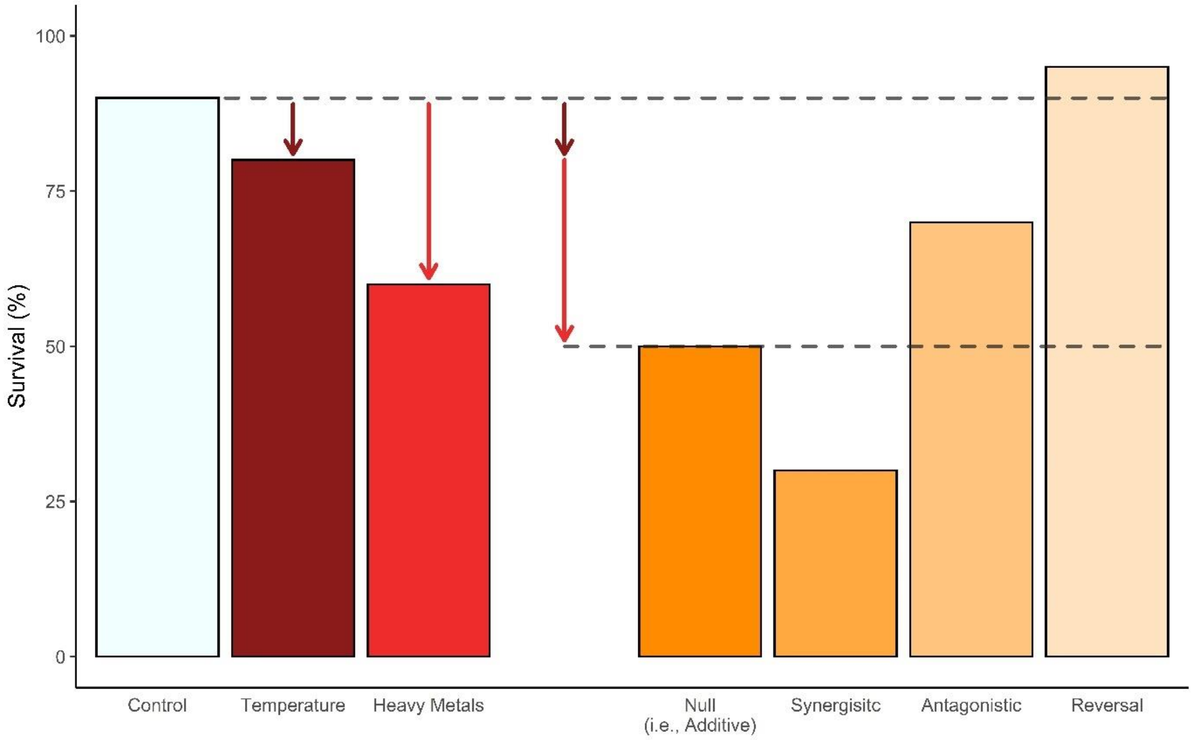

Figure 1.1: Graphical illustration of interaction classifications for the response metric of survival. The white bar denotes the survival rate of a given species under control conditions (i.e., absence of stressors). The dark red (red) bar denotes the survival rate of the given species under the stressor of temperature (heavy metals), with arrows illustrating the relative change in survival from that under control conditions. Orange bars illustrate examples of the four different interaction classifications: null (i.e., additive), synergistic, antagonistic, and reversal interactions.

Given the potential for stressors to combine in numerous different ways, there are many questions regarding multiple stressor interactions that are important to answer. For instance, how often can we expect the additive null model predictions to be met? How frequently do ecological surprises occur? Are some interaction classifications more prevalent than others? Accordingly, with freshwater ecosystems, there has been an increasing quantity of research which seeks to address these questions. Such freshwater research may be empirical, seeking to determine how a particular combination of stressors affects a specific system (e.g., Juvigny-Khenafou et al., 2021, Richardson et al., 2019); theoretical, using fundamental ecology to attempt to explain stressor interactions (e.g., Belarde & Railsback, 2016; Galic et al., 2018); or meta-analytical, attempting to find generalities in stressor interactions across multiple different experiments (e.g., Bancroft et al., 2008, Jackson et al., 2016). Indeed, such generalities are important to determine as they may allow for the prediction of stressor interactions which cannot be (or are not) measured. However, to date, the focus of multiple stressor research has primarily been on determining *how* stressors interact; although there is an increasing appreciation for the need to understand *why* stressors may interact in a given way.

One of the main implications of multiple stressor research, is its potential to aid in the conservation or management of ecosystems affected by co-occurring stressors. Indeed, an understanding of stressor interactions can inform management practitioners when to implement any conservation measures (Côté et al., 2016; Kath et al., 2018). Furthermore, it has previously been shown that a failure to consider how stressors interact can lead to management actions having a limited, negligible, or even detrimental impact on the ecosystems they are intended to conserve (Brown et al., 2013). However, at present, there have been no studies which have used knowledge of how stressors interact to successfully remediate two, or more, co-occurring stressors in a freshwater ecosystem (Spears et al., 2021). As such, despite potential benefits for conservation, predicting multiple stressor interactions is distinctly complex (Hodgson & Halpern, 2019), which potentially prevents management planning from incorporating knowledge of stressor interactions. Ultimately, understanding how multiple co-occurring stressors interact represents one of the grand challenges currently facing ecologists (Fleishman et al., 2011; Hodgson et al., 2017).

1.1 Stressor effects across organisation levels

While the overall aim of multiple stressor research may be to determine how co-occurring stressors interact to affect ecosystems, it is nevertheless important to consider the effects which an individual stressor is capable of inducing. Importantly, it has also been suggested that the manner in which stressors interact is in part due to the ecological mechanisms which the individual stressors affect (e.g., Galic et al., 2018). As such, stressors are capable of affecting individuals, populations, or even entire communities (Jackson et al., 2016). However, these differing levels of ecological organisation are not affected in isolation (Woodward et al., 2010), with effects at one organisation level capable of transitioning to other levels. Firstly, stressors predominately act at the individual level (Maltby, 1999; Schäfer & Piggott, 2018), with stressor impacts at this level broadly able to be grouped into behavioural, physical, and physiological effects. Behavioural impacts include changes to anti-predator behaviours (e.g., Johansen et al., 2017), while physical effects may include body size alterations (e.g., Shrimpton et al., 2007; Yvon-Durocher et al., 2011) or an increased likelihood of developmental abnormalities (e.g., Reeves et al., 2010). In contrast physiological impacts include changes in metabolic rate (Jackson et al., 2021) or disruption to the endocrine system (Besson et al., 2020; Roccuzzo et al., 2021).

Species populations are similarly able to be affected by stressors, with these effects manifesting in a myriad of different ways. At the population level, perhaps some of the most commonly observed stressor impacts are changes in the abundance, or density, of a species (e.g., Mebane et al., 2017). Indeed, the relative declines in the abundances of many species due to climate change and other stressors are now well known (WWF, 2020). Furthermore, stressors are capable of affecting many response metrics at the population level. As such, population survival (Bancroft et al., 2008), reproduction (e.g., Muyssen et al., 2010) or growth rates (Seifert et al., 2020) may likewise be affected by stressors; although a change in one response metric (e.g., density) does not preclude a change in another (e.g., survival) from occurring. Stressors are similarly capable of impacting communities (Bruder et al., 2019). Such effects may be declines in the species richness of community (e.g., Kratina et al., 2012) or alterations to the interactions between species (e.g., feeding rates; Shears & Ross, 2010). Stressors are also capable of affecting other community level metrics such as biomass (e.g., Zhao et al., 2013). As such, it may be assumed that declines in individual species biomasses, or abundances, will be mirrored at the community level. However, compensatory effects between individual species may mean that a decline in the biomass of one species may be offset by an increase in the biomass of another (Ives & Cardinale, 2004). For example, Christensen et al. (2006) found that freshwater zooplankton community biomass increased in the presence of stressors, though this increase in biomass was driven by a single stress-tolerant species, whereas all other zooplankton species experienced biomass declines.

As alluded to above, it is likewise important to note that a single stressor is capable of inducing multiple different effects, either directly or indirectly, across multiple different organisation levels. For example, following the metabolic theory of ecology (Brown et al., 2004), an increase in temperature may not only affect an individual's metabolic rate (Gillooly et al., 2001), but similarly induce changes in a species' generation time (a population level response metric) (Gillooly, 2000; Jackson et al., 2021). Furthermore, different stressors are capable of affecting the same response metrics, albeit through contrasting ecological mechanisms. For example, sertraline (a pharmaceutical product) may increase the mortality rate of species X. In contrast, cadmium may have no direct effect on species X, but instead greatly increases the mortality of species Y, which species X predates upon. Overall, both stressors result in declines in species X's abundance either directly through an increased mortality rate, or indirectly through an absence of available food resources. As such, what is the expectation for the combined effect of these stressors on species X if they act simultaneously? Under the additive null model, we would expect the overall effect on the abundance of species X to be equal to the sum of the effects of the individual stressors in isolation. However, this may not necessarily be the case. It could be that sertraline reduces the abundance of species X, which in turn means that there are fewer individuals competing for fewer food resources (i.e., reduced intra-specific competition). In such a scenario, sertraline may lessen the effect of cadmium on species X's abundance, resulting in an antagonistic interaction occurring. As such, species interactions have been viewed as having a crucial role in determining the effects of stressors, and their interactions, within communities (Bruder et al. 2019; Seibold et al., 2018).

The ecological phenomena outlined here are merely examples of the wide-range of processes that occur within any ecosystem; however, they illustrate the complexity of these systems and the plethora of direct, and indirect, effects that a stressor is capable of having. Further complexity arises from the effects of stressors at the cellular level (Sinclair et al., 2009; Sokolova, 2021), with there being calls for toxicology to be better incorporated within multiple stressor ecology (Orr et al., 2020). However, cellular approaches to understanding stressor interactions requires detailed knowledge of the physiological mechanisms which stressors affect; although for the vast majority of stressors and organisms, this information is simply not known (Spears et al., 2021).

1.2 Current approaches in multiple stressor ecology

Of the empirical, meta-analytical, and theoretical approaches to investigate multiple stressor interactions, the latter perhaps represents the most underutilised approach within freshwater ecosystems. Theoretical approaches to understanding multiple stressor impacts have instead been more widely implemented in general (e.g., Fronhofer et al., 2018; Sentis et al., 2021; Thompson et al., 2018a) or marine environments (e.g., Beauchesne et al., 2021; Brown et al., 2013; Fu et al., 2018). To date, most theoretical freshwater studies consider single stressor effects (e.g., Harvey et al., 2018; O'Gorman et al., 2017; 2019), with relatively few freshwater studies implementing theoretical ecology to understand the impacts of multiple stressors. Furthermore, those freshwater theoretical studies which do consider multiple stressors often implement disparate models and address wide-ranging hypotheses making direct comparisons between them difficult. For example, Belarde & Railsback (2016) use ecological theory to investigate the effects of habitat alteration and a non-native species on the growth and survival of a freshwater piscivorous fish. Overall, Belarde & Railsback (2016) conclude that while stressor interactions predominately matched null model predictions, environmental factors had a substantial impact on the prevalence of antagonistic interactions. In contrast, Galic et al. (2018) implement a theoretical model of freshwater amphipods and leaf litter (see Galic et al., 2017) to determine how co-occurring stressors interact to affect various ecosystem properties. Within these simulations, the observed interaction was found to be dependent on both the response variable and the modes of action of both stressors (Galic et al., 2018). As such, there is limited scope to draw comparisons across theoretical, freshwater, multiple stressor studies. However, there is substantial potential for this freshwater theoretical ecology to expand and fill an obvious gap in the multiple stressor literature.

To date, empirical multiple stressor research in freshwater systems has predominately focussed on microcosm, or mesocosm, studies (e.g., Archer et al., 2020; Cambronero et al., 2018; Richardson et al., 2019), although there are empirical studies considering multiple stressors effects at greater geographic scales (e.g., Birk et al., 2020; Ryo et al., 2018). Such an approach builds upon the use of mesocosms in single stressor experiments (see Stewart et al., 2013); but within a multiple stressor context, experiments are usually conducted in a factorial design, with treatments for i) control conditions, ii) only the first stressor present, iii) only the second stressor present, and iv) both stressors present. The benefit of this approach is that the effect of co-occurring stressors can be compared to a null model prediction (e.g., the additive null model) given that a response metric (e.g., biomass or survival) can be measured under all four treatment conditions. However, this experimental approach is not without limitations, and may be resource intensive (Boyd et al., 2018). Indeed, decisions regarding the number of stressors and the number of replicates per treatment are required. For example, a fully factorial design with two stressors and four replicates would require 16 experimental units (e.g., mesocosms). However, if a fully factorial design is implemented with three or four stressors (each with four replicates) then 32 or 64 experimental units would be required respectively. Such increases in the required numbers of experimental units are even more stark if multiple intensities of each stressor are required. As such, as found by several meta-analyses (e.g., Gomez Isaza et al., 2020; Seifert et al., 2020) the majority of multiple stressor experiments frequently have low (i.e., less than five) numbers of replicates per treatment, potentially to navigate some of the limitations surrounding experimental design. Accordingly, resource considerations may limit the scope of any experiments, particularly where large, or specialist, experimental units are required. However, as with the results of theoretical studies, the findings of multiple stressor experiments are likewise divergent. Conclusions may differ between experiments (or potentially even within a single experiment) as functionally, or taxonomically, similar organisms may respond to co-occurring stressors in disparate ways (e.g., Christensen et al., 2006; Piggott et al., 2012). As such, simple qualitative comparisons between experiments are often difficult to reliably make. Instead, other analytical approaches are needed to draw general conclusions from across multiple stressor experiments (e.g., meta-analyses).

Meta-analyses are an area of multiple stressor research which are becoming increasingly prevalent, particularly those exclusively, or partially, considering stressor interactions in freshwater environments (Figure 1.2). In brief, meta-analyses collate data from multiple experiments which are then aggregated and analysed to determine a single aggregated metric (see Gurevitch et al., 2018). For multiple stressor meta-analyses, null models (e.g., the additive null model) can again be implemented, being used to classify interactions both for individual experiments, and the overall aggregated metric. In total, there have been six meta-analyses (published before 2021) exclusively considering multiple stressors in freshwater environments; although Lange et al. (2018) expands upon the dataset first used by Matthaei & Lange (2016). The conclusions of these meta-analyses can greatly differ, with no generalities apparent for the frequency of interaction classifications for individual experiments or the overall interaction classification assigned to the aggregated metric. For individual experiments, antagonistic (Jackson et al., 2016) and null (Lange et al., 2018) interactions have separately been found to dominate, while roughly equal frequencies of antagonistic, synergistic, and null interactions have also been reported (Gomez Isaza et al., 2020). For the interaction classification of the aggregated metric, meta-analyses have likewise diverged and reported an overall interaction class of null (Egea-Serrano et al., 2012), antagonistic (Jackson et al., 2016; Lange et al., 2018), or synergistic (Bancroft et al., 2008) interactions. The disparate results across multiple stressor meta-analyses are unexpected, with it anticipated that the results should align to a far greater degree. At present, the absence of generalities across multiple stressor meta-analyses (for all realms) is unexplained, with ecological rationale failing to explain these disparities (Côté et al., 2016). Indeed, ecological variables such as response metric, organisation level, or feeding group have all been suggested as being important, or unimportant, covariables in explaining multiple stressor interactions, with there being no consistency across meta-analyses (Côté et al., 2016). As such, an explanation for the absence of any generalities in multiple stressor meta-analyses remains undetermined but is urgently required.

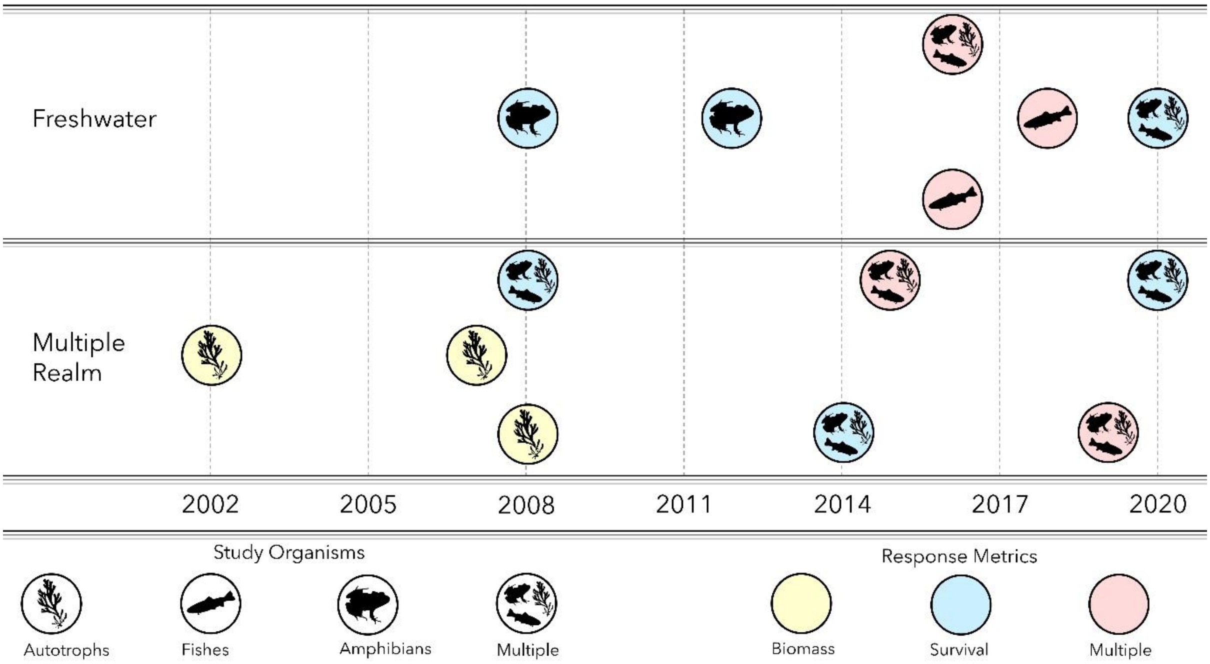

Figure 1.2: Timeline of different meta-analyses (published before 2021) which solely, or partially, consider the effects of multiple stressors in the freshwater realm, labelled 'Freshwater' and 'Multiple Realm' respectively. Key denotes the different study organisms and response metrics each meta-analysis considers. Silhouettes from phylopic.org

1.3 Furthering understanding of stressor interactions

Despite being a relatively new field, there has been a surge in the number of multiple stressor papers (both overall and those considering freshwater environments) recently published (Figure 1.3). However, questions remain as to what is required in order to better understand multiple stressor interactions. From one perspective the answer is obvious, with a greater quantity of empirical data being required. For instance, this can take the form of testing new combinations of stressors against previously unconsidered ecological communities or the replication of previous experiments. However, alongside an increased quantity of data, other developments are nonetheless required. For example, multiple stressor ecology requires an increased understanding, and appreciation, of the statistical methods commonly implemented, alongside a greater emphasis on the development of ecological theory. These necessary developments, alongside others, are discussed below.

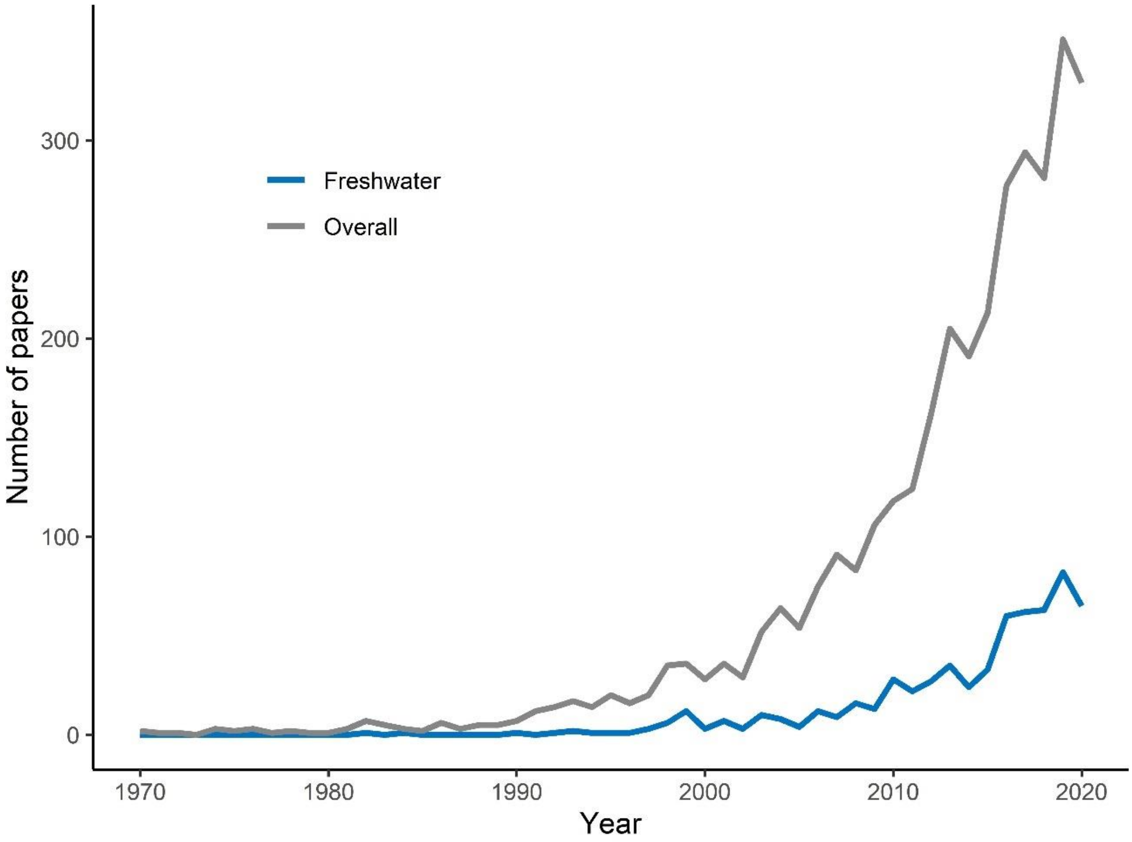

Figure 1.3: Number of studies published by year (up to, and including, 2020) which consider the effects of multiple stressors. Searches were conducted in Web of Science, with separate searches conducted for all multiple stressor papers (including freshwater studies) (grey line), and those which were specific to the freshwater realm (blue line). Search terms were based on those of Orr et al. (2020) (Appendix One).

Within multiple stressor ecology, the additive and multiplicative null models (see Folt et al., 1999; Sih et al., 1998; Soluk & Collins, 1988), are the most commonly used approaches for classifying stressor interactions. Indeed, although these two null models are widely implemented, there is relatively little knowledge of their statistical properties. For instance, both Folt et al. (1999) and Sih et al. (1998) describe (using verbal arguments) how the additive null model may be biased towards reporting antagonistic interactions given that it can make ecologically unfeasible predictions (e.g., a mortality rate of over 100%). However, there have been no attempts to determine whether other sources of bias are occurring within either the additive or multiplicative null models. As such, it may be that these null models are being widely-implemented and interpreted without an appreciation of the nuances or limitations of these approaches. Furthermore, the results of studies, or meta-analyses, which implement these contrasting null models are often compared (e.g., Côté et al., 2016), yet these null models employ differing underlying assumptions and as such may not be directly comparable. Indeed, there is definitely a need to determine the conditions under which we can expect the results of these two null models to align or differ, with only Stephens et al. (2013) implementing both null models on the same dataset. For example, an interaction may be assigned an antagonistic interaction classification by the multiplicative null model, but a synergistic interaction classification by the additive null model. Ultimately, it is important to determine whether there is a statistical, in addition to ecological, rationale for implementing one null model over the other.

As described above, the additive and multiplicative null models are the most widely adopted approaches when considering stressor interactions. Many studies determine, and classify, interactions through the calculation of effect sizes which correspond to the additive (e.g., Gurevitch et al., 2000) and multiplicative (e.g., Lajeunesse, 2011) null models (approaches that are subsequently built on throughout this thesis). Furthermore, while not necessarily widely appreciated (Schäfer & Piggott, 2018), these null models may also be implemented through analysis of variance or generalised linear models (e.g., Birk et al., 2020). However, alongside the widely-implemented additive and multiplicative null models, an increasing number of null models are being developed. Recently presented null models include the compositional null model (Thompson et al., 2018b) and the Rescaled Bliss Independence Model (Tekin et al., 2020). However, it is currently unclear when, or whether, such novel approaches should be implemented. Furthermore, there are currently no, or limited, comparisons to the established null model approaches. For instance, Tekin et al. (2020) compare the Rescaled Bliss Independence Model to analysis of variance, but not more commonplace additive or multiplicative null model approaches. Similarly, Thompson et al. (2018b) compare the compositional null model to versions of both the additive and multiplicative null models, though potential issues with this novel null model have been highlighted (see Orr et al., 2021a). Ultimately, it is unclear whether these new methods represent an upgrade on the existing null model approaches. As outlined above, a thorough understanding of the existing null model approaches (including their benefits and limitations) is required. Such an understanding would hence allow novel null models to be benchmarked against the established methods.

As with most ecological disciplines, multiple stressor ecology faces data limitations. Indeed, there are limitations surrounding both the number of studies considering a specific question and the quantity of data generated by any given experiment. As discussed above, most multiple stressor experiments have limited numbers of replicates per treatment as a way of mitigating resource limitations. However, there is little (if any) knowledge of the statistical implications of limited sample sizes upon the ability of the null models to detect non-null interactions (i.e., ecological surprises). For instance, should we expect an experiment with ten replicates per treatment to be better able to detect an interaction than an experiment with three replicates per treatment? If so, what is the relative benefit in increasing sample sizes from three to ten replicates per treatment? Or, alternatively, what sample size provides the best trade-off between statistical power and resource management? In many cases, decisions on sample sizes are simply based on heuristic rationale, with little consideration of the statistical implications of such a decision. Indeed, it may be that current experimental designs are only able to determine whether a non-null interaction is occurring if the interaction is exceptionally strong. As such, understanding the implications of an experimental design, in relation to the implemented null model, will be required in order for multiple stressor ecology to fully interpret existing results.

As described above, one of the pressing questions facing multiple stressor research is: *Why do multiple stressor meta-analyses fail to determine consistent generalities?* Indeed, this question applies to all multiple stressor meta-analyses not just those solely considering freshwater stressor interactions. As discussed previously, explanations for the absence of any consistent generalities have almost entirely focussed on ecological rationale for this disparity (see Côté et al., 2016). However, given this has so far failed to provide any explanation, it may be that other factors are instead responsible. One such factor could be the different methods employed by meta-analyses (see Hungate et al., 2009). Some methodological differences, such as whether a meta-analysis implements an additive or multiplicative null model, are easily apparent. However, there may also be a broader suite of subtler methodological differences which are driving these disparities yet are currently overlooked. For instance, there are various different iterations of both the additive (e.g., Darling & Côté, 2008;

Hillebrand, 2002; Siviter et al., 2021; Zhou et al., 2016) and multiplicative (e.g., Gomez Isaza et al., 2020; Harvey et al., 2013) null models, though how these different null model variants may affect any meta-analytical results is currently unknown.

Within multiple stressor ecology, theoretical and empirical analyses are seldom conducted in tandem. However, it is possible for theoretical ecology to generate expectations, regarding the effects of stressor interactions, which could then be sought to be proved, or disproved, by empirical analyses. For instance, theoretical ecology has generated predictions regarding how stressor interactions, in a given system, are dependent on stressor modes of action and measured response variables (Galic et al., 2018) or how conservation effectiveness can be determined by stressor interactions (Brown et al., 2013). Such examples illustrate the potential for theory to help inform and direct empirical analyses; although, inversely, empirical analyses could also generate predictions which are then tested by ecological theory. Either way, a greater emphasis on a combined empirical and theoretical approach may further multiple stressor ecology by allowing ecological observations to be contrasted against a potentially mechanistic understanding of stressor interactions.

1.4 Thesis aims

The research outlined within this thesis centres on several core aims. These aims are stated below, before being described in the following section.

i. Determine whether ecological surprises (i.e., antagonistic, reversal, and synergistic interactions) are more prevalent than currently expected.
ii. Determine whether ecology, or methodology, are responsible for the conclusions of previous multiple stressor research.
iii. Determine the causes responsible for the absence of consistent generalities across multiple stressor meta-analyses.
iv. Determine what covariables (e.g., feeding group or response metric) are important in determining how stressors interact in aquatic ecosystems.

1.5 Thesis outline

In this thesis, I focus upon null model approaches to understand the interactions of stressors. While alternative approaches have been suggested (e.g., De Laender, 2018; Spears et al., 2021), these null models remain the most popular choice for classifying the interactions of multiple stressors with there similarly being established statistical methods for their implementation (e.g., Gurevitch et al., 2000; Lajeunesse, 2011). However, despite their popularity, these null models are not necessarily well understood, an issue which I attempt to remedy.

In Chapter Two, I combine theoretical and meta-analytical approaches to better understand how multiple stressors can interact to affect freshwater ecosystems at higher levels of ecological organisation. Simulations of food chains are used to determine how both the frequency of individual interaction classifications and overall meta-analytical metrics vary under ecologically relevant conditions. These findings are then compared to the results of a large-scale meta-analysis focussing on the effects of multiple stressors, on biomasses or densities, in freshwater environments. In Chapter Three, I address the limited understanding of the additive and multiplicative null models. Here, I build on the methods of Chapter Two, determining how these null models respond to ecologically relevant levels of observation error and sample sizes, and determining the conditions under which the results of these null models align or differ. I similarly determine whether some results, frequently reported by multiple stressor studies, are indeed due to ecology or are simply artefacts of these statistical tools. Penultimately, in Chapter Four, I address the issue identified by Côté et al. (2016) of why multiple stressor meta-analyses have so far failed to determine consistent generalities. Here, I focus on methodological explanations for these disparities, building upon the findings of the previous chapters. In doing so I collate, correct, and reanalyse datasets from seven published multiple stressor meta-analyses using a single consistent statistical framework. I determine whether the contrasting implementation of null models, across meta-analyses, is responsible for these differences that have often been attributed to underlying ecological variation. Finally, in Chapter Five, I collate the findings from each of the above chapters and examine the key messages from across this thesis. In doing so, I summarize current issues within the field of multiple stressor ecology and outline potential avenues for future research.

Chapter Two – Classifying ecosystem stressor interactions: Theory highlights the data limitations of the additive null model and the difficulty in revealing ecological surprises

2.1 Abstract

Understanding how multiple co-occurring environmental stressors combine to affect biodiversity and ecosystem services is an on-going grand challenge for ecology. Currently, progress has been made through accumulating large numbers of smaller-scale empirical studies that are then investigated by meta-analyses to detect general patterns. There is particular interest in detecting, understanding, and predicting ‘ecological surprises’ where stressors interact in a non-additive (e.g., antagonistic or synergistic) manner, but so far few general results have emerged. However, the ability of the statistical tools to recover non-additive interactions in the face of data uncertainty is unstudied, so crucially, we do not know how well the empirical results reflect the true stressor interactions. Here, we investigate the performance of the commonly implemented additive null model. A meta-analysis of a large (545 interactions) empirical dataset for the effects of pairs of stressors on freshwater communities reveals additive interactions dominate individual studies, whereas pooling the data leads to an antagonistic summary interaction class. However, analyses of simulated data from food chain models, where the underlying interactions are known, suggest both sets of results may be due to observation error within the data. Specifically, we show that the additive null model is highly sensitive to observation error, with non-additive interactions being reliably detected at only unrealistically low levels of data uncertainty. Similarly, plausible levels of observation error lead to meta-analyses reporting antagonistic summary interaction classifications even when synergies co-dominate. Therefore, while our empirical results broadly agree with those of previous freshwater meta-analyses, we conclude these patterns may be driven by statistical sampling rather than any ecological mechanisms. Further investigation of candidate null models used to define stressor-pair interactions is essential, and once any artefacts are accounted for, the so-called ‘ecological surprises’ may be more frequent than was previously assumed.

2.2 Introduction

Ecological communities are being subjected to a wide variety of external stressors (Halpern et al., 2015) that act across terrestrial, freshwater, and marine biomes and threaten ecosystems and their services (Scheffers et al., 2016). These stressors, also termed drivers, factors, or perturbations (Orr et al., 2020), are frequently anthropogenic in origin (Vörösmarty et al., 2010; Geldmann et al., 2014), but are capable of being abiotic or biotic (Przeslawski et al., 2015), and are able to act at the local to global scales (Ban et al., 2014; França et al., 2020). While individual stressors (e.g., climate change, habitat alteration, or pollution) are themselves capable of inducing changes in biodiversity or ecosystems and their services (Dirzo et al., 2014; Tittensor et al., 2014; Newbold et al., 2015), ecosystems are frequently, if not predominately, acted upon by multiple stressors simultaneously (Crain et al., 2008). Despite the negative connotations surrounding the term *stressor*, stressors are capable of inducing effects that are either beneficial or detrimental to the affected ecosystem (Kroeker et al., 2017). One of the grand challenges facing ecologists is to be able to detect, understand, and predict how these different types of ecosystem stressors interact to affect biodiversity and ecosystem services (Hodgson & Halpern, 2019); although the challenge is more difficult since the observed interactions can substantially deviate from what is anticipated (Christensen et al., 2006). Ultimately, knowledge of how stressors interact is important in guiding conservation and management initiatives, and in helping to prevent remediation measures from being ineffective, or even potentially harming those ecosystems they are intended to preserve (Brown et al., 2013; Côté et al., 2016).

Aquatic ecosystems and communities are particularly threatened by multiple stressors (Birk et al., 2020). For instance, Halpern et al. (2008) describe how every marine area is subjected to human influence, with 41% of these areas being impacted by multiple stressors. Moreover, freshwaters represent some of the most at-risk ecosystems and are frequently exposed to a wide range of stressors (Hecky et al., 2010; Ormerod et al., 2010; Woodward et al., 2010; He et al., 2019), with freshwater biodiversity declining at rates exceeding even those of the most impacted terrestrial ecosystems (Sala et al., 2000), and potentially endangering vital ecosystem services (Malaj et al., 2014). While stressors often interact to impact freshwater ecosystems (Birk et al., 2020), their presence in freshwater ecosystems is not a new phenomenon, with some freshwater bodies having been subjected to stressors for several centuries (Dudgeon et al., 2006). However, the stressors that freshwater systems are currently facing has expanded, with the introduction of novel stressors, such as nanomaterials, while existing ones are continuing to have severe impacts (Reid et al., 2019). The cumulative impact of multiple stressors has been identified as one of the most pressing and emerging threats to freshwater biodiversity, but despite this, our current understanding of both how stressors interact, and the severity of their effects, is poor (Reid et al., 2019).

The term *ecological surprise* (*sensu* Paine et al., 1998) is often used to describe the changes in a biological response variable that contrast those anticipated when multiple stressors interact (e.g., Christensen et al., 2006; Jackson et al., 2016). Although an ecological surprise may be defined as an interaction that is either greater than, or less than, the expected magnitude from a null model, particular focus has been on interactions of stressors which interact synergistically; i.e., where the combined effect is greater than the sum of the individual effects. Synergistic interactions of multiple stressors are important to document, firstly due to their potential to have a dramatic effect on ecological communities, and secondly because the presence of a synergistic interaction means management strategies can potentially have a large effect by mitigating against just one of the interacting stressors (Brown et al., 2013; Côté et al., 2016; Haller-Bull & Bode, 2019). Because of their potential impact there has been a great deal of effort in recording the frequency of synergy in stressors across different ecosystems and communities (Côté et al., 2016). However, there is always a danger that an emphasis on their importance could lead to over-estimating the frequency of synergisms or other forms of ecological surprise (e.g., antagonisms) within the multiple stressor literature and, as highlighted by Côté et al. (2016), there is little evidence to suggest that stressors predominately interact in a synergistic manner. A pertinent question which has yet to be addressed is whether ecological surprises should be expected, or whether the prevalence of these interactions are skewed in some way by reporting biases, statistical sampling, or both.

There is relatively little ecological theory that generates expectations of when and how often the cumulative effects of pairs of stressors should be synergistic, or indeed any other type of interaction. This is in contrast to other ecological interactions, such as the effects of multiple predators on prey density and biomass, where a much richer body of theoretical knowledge has been used to generate a number of hypotheses for testing (Sih et al., 1998; Schmitz, 2007). Instead, progress on ecosystem stressor interactions has been made largely by meta-analyses across a number of experiments, realms, trophic levels, measured traits, taxonomic groups, and stressor types (e.g., Crain et al., 2008; Darling and Côté, 2008; Wu et al., 2011; Jackson et al., 2016). Within ecosystem stressor research, the most popular approach is to use the additive null model where the stressor interaction is predicted to be simply the sum of their individual effects (e.g., Crain et al., 2008; Darling & Côté, 2008; Strain et al., 2014; Jackson et al., 2016), although the multiplicative null model, the log-transformed version of the additive model, is also relatively common (e.g., Bancroft et al., 2008; Gruner et al., 2008; Harvey et al., 2013; Rosenblatt & Schmitz, 2014). These null models classify interactions as either being null (i.e., the additive or multiplicative effect of interacting stressors), synergisms (i.e., greater than the null), or antagonisms (i.e., less than the null). While distinctions are increasingly being made for various forms of antagonistic interactions in this simple scheme (e.g., Jackson et al., 2016), there exists a range of other classification schemes (Orr et al., 2020), and these have been implemented across a number of studies (e.g., Travers-Trolet et al., 2014; Piggott et al., 2015). The profusion of null models can make it difficult to generalise results across different studies. A ‘synergistic’ or ‘antagonistic’ interaction may have contrasting definitions depending on the scheme being used leading to the same interactions being labelled differently under contrasting schemes; hence the biological and statistical interpretation is therefore dependent on the null model being applied. One way round this issue is to pool published data together to harness increased statistical power and conduct a meta-analysis to search for generalities under a particular null model (examples listed in Côté et al., 2016). However, despite their potential, these meta-analyses have to date not identified any general covariates capable of explaining the broad patterns of multiple stressor interactions, meaning we still lack general predictions of the consequences of multiple stressors (Côté et al., 2016).

Given the lack of consistent generalities from empirical studies, the development of ecological theory within multiple stressor research may represent an approach capable of providing novel insights. Some theory has been developed for particular case studies (e.g., Brown et al., 2013; Galic et al., 2018), but only a few studies (e.g., Haller-Bull & Bode, 2019) have so far investigated more general insights. Of primary interest is the generation of theory which can provide a mechanistic underpinning to the field, and potentially allow for an increased understanding of multiple stressor interactions, compared to that provided solely by a null model approach (De Laender, 2018). However, theory could also be used to better understand the results obtained from the null model approach to empirical classification of stressor interactions. In particular, we know of no study that has investigated how robust the null models are to noisy data (i.e., sampling uncertainty and/or process variation); yet understanding this is important before we can draw strong conclusions from the empirical analyses. This knowledge is also important for evaluating the relative performances of the profusion of null models, and is therefore something which may help guide the end-user to decide which null model may be both appropriate and likely to yield important results in the face of what is often noisy and/or limited data.

Here, we begin to close these gaps in understanding by testing for the prevalence of non-additive effects of co-occurring pairs of stressors in freshwater ecosystems. We first develop classical community ecology models based on Lotka-Volterra consumer-resource dynamics in order to simulate data from biologically simple food webs impacted by pairs of stressors. This provides us with ’data’ where we know the underlying stressor-pair interactions. We then use this simulated data to investigate the ability of the additive null model to recover interactions under a range of different levels of data uncertainty which we model as observation error. With a better understanding of the statistical null model we then review the experimental literature to compile and analyse the largest (in terms of the number of interactions) dataset for the effects of co-occurring stressor interactions on the biomasses and densities of freshwater organisms. In particular, we ask whether ecological surprises are common in freshwater stressor interactions. The simulation experiments allow considerable insights into our empirical analyses and help prevent over-interpretation of our results.

2.3 Methods

Theoretical models

In order to provide a better understanding of the empirical results that follow, we built food chain models based on the classical Lotka-Volterra consumer resource equations (Heath et al., 2014). We chose these models since we believe stressors may act directly on population- and trophic-level patterns, but also indirectly via trophic cascades (e.g., a species may be indirectly affected if its primary resource is directly affected by a stressor). This approach is also broadly in line with our empirical data which focuses on population and community-level metrics as the responses to stressor treatments (see below). To increase the robustness of our conclusions we considered two forms of the model; one where (within trophic level) density dependence affects the death rates of each trophic level, and the other where consumer uptake is density regulated (Table 2.1). Both these scenarios were analysed by Heath et al. (2014) to investigate the roles of different types of density dependence on trophic cascades (see details therein). In both models the basal level of the chain describes the dynamics of a key nutrient that limits the productivity of the food chain, and we assumed nutrients are added at a constant rate, ω (Heath et al., 2014). Each subsequent equation then describes a different type of consumer. The first level is wholly dependent on the nutrients and may represent a primary producer such as an algal species that requires a key mineral such as silica. The second level consumes the first trophic level and is in turn consumed by a third trophic level, and so on until the apex consumer is reached. In the density dependence model (Equation 2.1; Table 2.1), the consumer (trophic level *i)* exploits the resource (trophic level $i-1$) with a constant consumption/attack rate, α_i, and the conversion efficiency parameter, ε_i, determines the proportion of the consumed resource that is converted into new consumers (Heath et al., 2014). Under density dependence, the consumer is self-regulated by the intraspecific density dependence parameter λ_i, which leads to an increase in death rate as the consumer density increases (Heath et al., 2014). In contrast, the consumer uptake regulation model (Equation 2.2; Table 2.1), assumes the effect of increasing consumers is to slow down the consumption of the resource, perhaps due to increased interference (Heath et al., 2014). In this case, the parameter ν_i, determines the consumer density at which the maximum per capita uptake rate is halved, defined as the density $x_i = 1/\nu_i$ (Heath et al., 2014).

Table 2.1: Equations used to establish theoretical food-chains. The equations, sets, and a brief description of the equivalent ecological trophic level are shown.

	Equation Type	**Equation**	**Description**
2.1.1	*Density Dependence*	$\frac{dx_n}{dt} = \alpha_n \varepsilon_n x_{n-1} x_n - \delta_n x_n - \lambda_n x_n^2$	*Change in density of Apex Consumer (x_n)*
2.1.2	*Density Dependence*	$\frac{dx_i}{dt} = \alpha_i \varepsilon_i x_{i-1} x_i - \alpha_{i+1} x_i x_{i+1} - \delta_i x_i - \lambda_i x_i^2$	*Change in density of Non-Apex Consumer (x_i)*
2.1.3	*Density Dependence*	$\frac{dx_0}{dt} = \omega - \alpha_1 x_0 x_1$	*Change in density of Nutrients (x_0)*
2.2.1	*Consumer Uptake Regulation*	$\frac{dx_n}{dt} = \frac{\alpha_n \varepsilon_n x_{n-1} x_n}{1 + \nu_n x_n} - \delta_n x_n$	*Change in density of Apex Consumer (x_n)*
2.2.2	*Consumer Uptake Regulation*	$\frac{dx_i}{dt} = \frac{\alpha_i \varepsilon_i x_{i-1} x_i}{1 + \nu_i x_i} - \frac{\alpha_{i+1} x_i x_{i+1}}{1 + \nu_{i+1} x_{i+1}} - \delta_i x_i$	*Change in density of Non-Apex Consumer (x_i)*
2.2.3	*Consumer Uptake Regulation*	$\frac{dx_0}{dt} = \omega - \frac{\alpha_1 x_0 x_1}{1 + \nu_1 x_1}$	*Change in density of Nutrients (x_0)*

Using these equations, we established food-chains comprising either three, four, or five trophic levels, and the equation for each trophic level models how the biomass or density changed over time. For simplicity we assumed all key parameters (nutrient input ω; consumption rates α_i; conversion efficiencies ε_i; uptake regulators ν_i; density independent δ_i, and dependent death rates λ_i, for trophic level *i*) do not vary over time, and we investigated the effect of stressors on equilibrium biomasses/densities. The models also ignore spatial structure in the community, which also remain closed to immigration from outside apart from the constant input of the nutrient. Hence these models represent the simplest form of community dynamics that can be used to investigate the effects of multiple stressors as well as the manner in which they interact.

Stressors to the food chains were modelled by changing the values for parameters and comparing the resultant equilibrium densities across all trophic levels to the equilibria for a set of baseline parameter values. Equations 2.1 and 2.2 are not mechanistic models for specific stressors (e.g., pollution, temperature) but instead capture the net effect of stressors on the ecological processes of the food web species. For simplicity, we assumed each stressor had either a positive or negative effect on one model parameter (i.e., ω, α_i, ε_i, or δ_i), and we investigated how pairs of stressors interact to affect community densities. The baseline parameters were drawn from uniform distributions with ranges given in Table 2.2. Therefore, for a given food chain, the baseline parameters for all trophic levels were independently sampled from the distribution of values given in Table 2.2. Similarly, the processes (parameters) affected by each stressor were randomly selected from the possible candidates, and the intensity of its effect on the baseline rate was drawn from a uniform distribution with the ranges shown in Table 2.2. The baseline parameter set therefore represented the control community, and as in experimental studies that employ the factorial design approach (e.g., Matthaei et al., 2010; Davis et al., 2018), we manipulated our model communities by investigating the effect of each stressor acting alone, as well as the stressors acting in combination. From these cases, we then computed the type of stressor interaction and how they combined to alter the community densities (see below for definitions of how stressor interactions are computed). We therefore chose one trophic level at random from the entire food chain, excluding the nutrient level. We focused on this population/trophic level and mirrored it in our selection of empirical data (see below). This also means the species or trophic levels under scrutiny were not always directly affected by the stressor but could be affected solely due to a trophic cascade effect. It is also important to note that a stressor could have led to either an increase or a decrease in parameter value relative to the baseline; and that multiple stressors could have acted on the same, or a different trophic level. We chose to model the scenario where each stressor affected only one parameter (and therefore one biological process); hence within our model communities, stressors did not interact at the parameter level. However, relaxing this assumption to allow two stressors to affect a single process (parameter) did not alter our results (Appendix Two).

Table 2.2: Explanation of the different parameters within Equations 2.1 and 2.2, with the mechanism they reflect, alongside the minimum and maximum values for the ranges of baseline parameter values. Parameter values were drawn from a uniform distribution U~(a, b) with lower limit, a, and upper limit, b, with the limits differing between the baseline and stressed parameters. The method for determining stressed parameter values is detailed in Appendix Two.

Parameter	**Ecological Mechanism**	**Baseline Value Range**	**Stressed Value Range**
α	The rate at which a trophic level predates upon the trophic level directly below.	α_b = U~(0.25, 0.75)	U~(0.01, 0.99)
ε	The efficiency at which a trophic level can transform consumed matter into new individuals.	ε_b = U~(0.25, 0.75)	U~(0.01, 0.99)
δ	The density independent mortality rate of a trophic level.	δ_b = U~(0.25, 0.75)	U~(0.01, 0.99)
ω	The constant rate at which a resource (x_0) is input into the food chain.	ω_b = U~(25, 75)	U~(1, 99)
λ	The density dependent mortality rate of a trophic level.	λ_b = U~(0.00625, 0.025)	The parameter was not under selection for alteration by a stressor
ν	A limit to the uptake rate of a consumer through a trait-mediated response, that may be behavioural or otherwise.	ν_b = U~(0.05, 0.15)	The parameter was not under selection for alteration by a stressor

Overall, 1,320,000 different combinations, of equations, food-chain lengths, stressor pairs, and randomly selected baseline values were generated. Equilibrium densities, for each of these combinations, were calculated using Mathematica 10.4 (Wolfram Research, Inc., 2016). We only considered cases where the equilibria were all stable, and feasible (i.e., all densities were positive), and only equilibrium densities for trophic levels x_1 and above were included in the stressor interaction results (i.e., we excluded the nutrient level from our stressor interaction analyses). Stability was assessed by determining the Jacobian matrix for each community and calculating the corresponding eigenvalues. For every community, all eigenvalues had a negative real part with the equilibria being point attractors.

Across all 1,320,000 combinations, 79.9% of the parameter sets resulted in the determination of equilibrium densities that were both stable and feasible, with the discarded 20.1% of parameter sets resulting in at least one biologically unfeasible density. From the full set of stable and feasible communities, we randomly selected 360,000 theoretical interactions, and for each community we randomly selected a single trophic level for the focus of our estimation of the stressor interaction. All subsequent analyses of the theoretical data were performed on this group of 360,000 theoretical interactions. This subsetting was required as there was a negative relationship between the number of trophic levels and the likelihood of the community being both stable and feasible, which biased the full dataset towards communities with only three trophic levels. The final 360,000 stressor interactions were selected with weighted probabilities to ensure approximately one third (i.e., ~120,000) were from each of the three food chain lengths, and that each model (Table 2.1) was also approximately equally represented.

Unlike the empirical studies used in the meta-analyses below, the food chain models are purely deterministic, meaning that there are no random fluctuations around the equilibrium densities. In effect, for any given pair of stressors, there is no uncertainty (observation error) in the theoretical data. Clearly, this differs from the empirical data where observation error leads to an estimate of the densities/biomasses under investigation in the control and treatment replicate communities, and this observation error may lead to some stressor interactions being misclassified. For a better comparison with the empirical data, and to test the robustness of the additive null model to observation error, we modelled observation error by taking the 360,000 theoretical interactions from our original analyses and then multiplying the density of each trophic level by a random number drawn from a Gaussian distribution with a mean of 1.00 and standard deviation of σ. This process was repeated between three and six times for each treatment, analogous to the number of replicates per treatment found in our empirical data (see below). Thus, larger values for σ led to larger deviations around the equilibrium biomasses, and therefore a larger observation error, with an increased likelihood that the stressor interaction was misclassified. Standard deviations, σ, were from one of 50 different levels, ranging from 1×10^{-6} to 5×10^{-1}, in consistent logarithmic increments (e.g., 8×10^{-6}, 9×10^{-6}, 1×10^{-5}, 2×10^{-5}, etc.). The interpretation of σ is straightforward, as we would expect 99.7% of all observations to fall within 3σ's of the 'true' stressor effect (i.e., the biomass/density in the absence of any observation error). Appendix Two details a complete overview of how observation error was incorporated into the theoretical data.

Collation of empirical data

Through Web of Science we searched the primary scientific literature, for papers published before 1st January 2019, which investigated the impacts of multiple stressors on freshwater communities. In order to be incorporated, papers needed to report results where there was a factorial design, namely: (i) a control (without stressors); (ii) each stressor acting individually; and (iii) the stressors acting simultaneously. We required papers to report the mean value of the response, the number of replicates, and standard deviation or standard error for each treatment in the factorial design; failure to report any of this information led to the study being excluded from our analysis. Additionally, papers were required to report at least one of the following untransformed metrics: biomass, abundance, density, or chlorophyll-a of one or more groups of organisms within the stressed community. Hence, and in line with our trophic models, the focus of our effort was directed towards studies that report the effects of stressors acting at the population and community levels. Papers often reported the impacts of stressors on multiple different groups of organisms within a community; when this occurred, the responses of all different groups of organisms were included within the overall dataset. The different groups of organisms could comprise: populations of a single species (e.g., *Daphnia pulex*); a group of organisms within the same feeding guild (e.g., herbivores); a group of taxonomically similar organisms (e.g., taxa within the genera *Ephemeroptera*, *Plecoptera, and Trichoptera*); or a group of similar organisms (e.g., macroinvertebrates or algae). To be included within our dataset, papers had to investigate communities comprising a minimum of two different groups of organisms. Studies investigated a wide range of different stressors, although these were subsequently grouped into broader stressor categories, such as temperature, contamination, and habitat alteration.

Previous analyses have frequently focused on collating data for only the greatest single intensity of a stressor (e.g., Jackson et al., 2016). In contrast, where studies reported the responses of communities to multiple intensities of different stressors, data for all of the different intensities were collated. All interactions considering the different intensities of stressors were included in the overall dataset, although covariation in data due to repeated experiments across different stressor intensities was accounted for in the final meta-analyses (see section below).

Some studies reported multiple different response metrics for the same group of organisms, included the same species within multiple different groups, or reported data for the same experiment over multiple different time points. Accordingly, in order to reduce the correlation/covariance within the overall dataset, these interactions were removed from our analyses. For instance, where the effects of stressor interactions on multiple different traits were reported, those considering density as the biological response metric were prioritised over abundances, which were in turn prioritised over biomasses, or those considering chlorophyll-a, respectively. Similarly, where papers reported data for stressor interactions over multiple different time points, only the final time point was used as this best matched our equilibrium assumption for the theoretical models.

Appendix Two gives a complete overview of the different search terms used to find studies, the method used to determine whether the data for a study could be collated, the processes for extracting and collating the data, and the process for removing interactions to prevent covariance.

The determination of effect sizes and the classification of interactions

Across both the theoretical and empirical datasets, we used the same method to determine the classification of an interaction, using the factorial form of the effect size metric, Hedges' *d* (Gurevitch et al., 2000). Hedges' *d* is frequently used to investigate the impacts of multiple stressors as it estimates the standardised mean difference between the means of stressed and control samples and is unbiased by small sample sizes (Hedges & Olkin, 1985). It is calculated by comparing the effect of the interaction on ecological communities to the sum of effects of the stressors acting individually; namely, an additive null model. In line with current methods, we inverted the sign of the interactions when the expected effect of the additive null model was negative (Jackson et al., 2016; Piggott et al., 2015). This method allowed for interaction effect sizes to be compared regardless of their directionality. We therefore focused on the classification of the interaction as opposed to the absolute magnitude/polarity of the effects. Appendix Two gives a complete breakdown of the equations used for calculating Hedges' *d*.

Once Hedges' *d* for a given interaction of stressors was calculated, we then classified the interaction into one of four types as illustrated in Figure 2.1 and following the convention of Jackson et al. (2016). In brief, the four interaction classifications were: (i) *Additive*, where the effect of the additive null model was statistically indistinguishable from the effect of observed interaction; (ii) *Synergistic*, where the observed interaction effect was greater than the effect of the additive null model; (iii) *Antagonistic*, where the observed interaction effect was less than the effect of the additive null model, but both effects had the same polarity; and (iv) *Reversal,* where the observed interaction effect was less than the effect of the additive null model, but the observed and expected effects had contrasting polarities. The distinction between antagonistic and reversal interactions is relatively recent (e.g., Travers-Trolet et al., 2014; Jackson et al., 2016), with most research still continuing to use the appellation of antagonistic to refer to both antagonistic and reversal interactions (e.g., Velasco et al., 2019; Gomez Isaza et al., 2020).

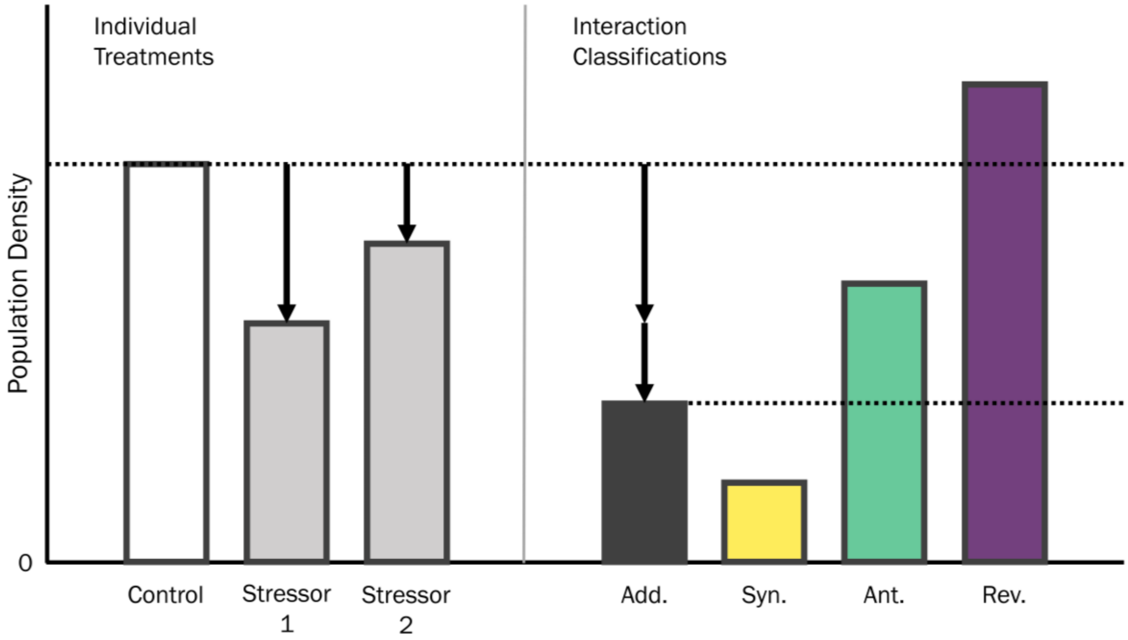

Figure 2.1: Graphical depiction of interaction types using population density as a response metric. White and grey bars denote densities under control and single stressors, respectively. The black bar denotes the additive (Add.) interaction classification (i.e., the sum of the effects for the individual stressors shown by the black arrows). The yellow bar denotes a synergistic (Syn.) interaction classification (i.e., a decrease in population density greater than the additive effect). The green bar denotes an antagonistic (Ant.) interaction classification (i.e., a decrease in population density less than the additive effect). The purple bar denotes a reversal (Rev.) interaction classification (i.e., a change in population density in the opposite direction to that of the additive effect).

In our method, if Hedges' *d* was positive, the interaction was classed as synergistic, and if negative, the interaction was classed as either an antagonistic or reversal interaction, although this could only be determined by comparing the effect of the additive null model to the observed effect (as outlined above). Each value of Hedges' *d* had corresponding 95% confidence intervals; if these confidence intervals incorporated 0, then an interaction was deemed to be additive. The classification scheme outlined above is one of a number of possible choices (e.g., Crain et al., 2008; Jackson et al., 2016), and Appendix Two details a comparison of how these different schemes contrast each other.

Vote-counting

Following the classification of all interactions, we implemented a vote-counting method to determine the relative proportions of the interaction classes across both the theoretical and empirical datasets. To consider the effect of different strengths of observation error on the ability to detect the 'true' stressor interaction in the modelled data, we computed the frequency of interaction types for each level of observation error investigated.

Summary effect sizes

Alongside the vote-counting method, we calculated summary effect sizes across both the theoretical and empirical datasets. The calculation of summary effect sizes represents one of the key components that defines a formal meta-analysis (Koricheva & Gurevitch, 2014), allowing for the collation of the individual effect sizes of multiple independent experiments or studies and determining a single summary effect. Pooling the data in this way increases the statistical power of our analyses, and therefore leads to a greater probability of correctly rejecting the null hypothesis that stressor interactions are additive. Meta-analyses and summary effect sizes are both useful and well-established within the field of multiple stressors (e.g., Crain et al., 2008; Jackson et al., 2016; Seifert et al., 2020), and give higher weightings to individual effect sizes with lower uncertainties (i.e., lower variances) which lead to more precise estimates of the overall summary effect size (Koricheva & Gurevitch, 2014).

For the empirical analysis, summary effect sizes were determined by using a weighted random effect model and implemented in the *metafor* package (Viechtbauer, 2010) in R. Random effects were specified as being the identity (ID) of the study group of organisms nested within the ID for study. The random effects were specified in order to account for both within- and between-study variation contained within the empirical dataset. Additionally, some empirical studies considered multiple intensities of one or more stressors, and as such, calculations of the interaction class for each intensity of stressor used the same control. To account for any covariance between the different intensities of a single stressor, we incorporated covariance-variance matrices within the meta-analytical models. For the empirical dataset, mixed effect models were also conducted with the fixed effects of stressor pair or organism group, alongside the previously described random effects (see Appendix Two). The summary effect size for the theoretical dataset was also determined using a similar process. However, due to computational limitations caused by the number of interactions under analysis (360,000 interactions at each level of observation error), fixed effect models for the theoretical data were fitted using the *lm* function. The models applied to both the theoretical and empirical datasets are explained in further detail within Appendix Two. While we detail the results of both the vote-counting and summary effect size methods, our results primarily focus on summary effect sizes, in line with recommendations for meta-analyses (Gurevitch et al., 2018).

The overall effect from a meta-analysis needs to be checked for consistency among effect sizes, termed as heterogeneity (Nakagawa et al., 2017). We used the I^2 statistic, which is bounded between 0% and 100%, with 25%, 50%, and 75% being suggested as levels for, low, medium, and high heterogeneity, respectively (Higgins et al., 2003). Ecological meta-analyses often report high levels of heterogeneity (Senior et al., 2016), perhaps due to the variation in study organisms common to the questions being addressed, and we may have expected a high value here due to both range of study organism and range of stressor type. To explore the potential causes of heterogeneity within the empirical meta-analysis, we conducted separate meta-analyses on a sub-group of the dataset, a similar process to running a meta-regression (Nakagawa et al., 2017), using organism group (i.e., producer or consumer) as the categorical moderator to explore heterogeneity (see Appendix Two). We also considered publication bias (see Appendix Two); although it should be noted that common tests for publication bias within meta-analyses can be limited by high heterogeneity (Nakagawa et al., 2017).

2.4 Results

Stressor interactions within theoretical data

We found no strong difference between the classification of stressor interactions from either form of food chain model (Table 2.1), or between the different lengths (three, four, five trophic levels) of food chains (see Appendix Two), showing that the frequencies of interaction classifications were robust to these details of the models. For the entire theoretical dataset of 360,000 stressor interactions (comprising both Consumer Uptake Regulation and Density

Dependence Equations, and across food chains of three, four and five levels) without observation error, antagonistic and synergistic interactions were the most frequently assigned (0.483 and 0.480, respectively), followed by reversal (0.0288), and finally additive interactions (0.00856). These frequencies represent the 'true' interactions as classified by the additive null model, under no data uncertainty. However, the ability of the additive null model to recover these interaction frequencies is very sensitive to observation error. Increasing observation error led to more interactions being classified as additive (the null model) and at higher levels additive interactions were clearly dominant (Figure 2.2a). This pattern could be generated if our theoretical interactions only weakly deviated from additivity, but checks confirmed that this was not the case, and that over 50% of interactions deviated from additivity by more than 5% (see Appendix Two).

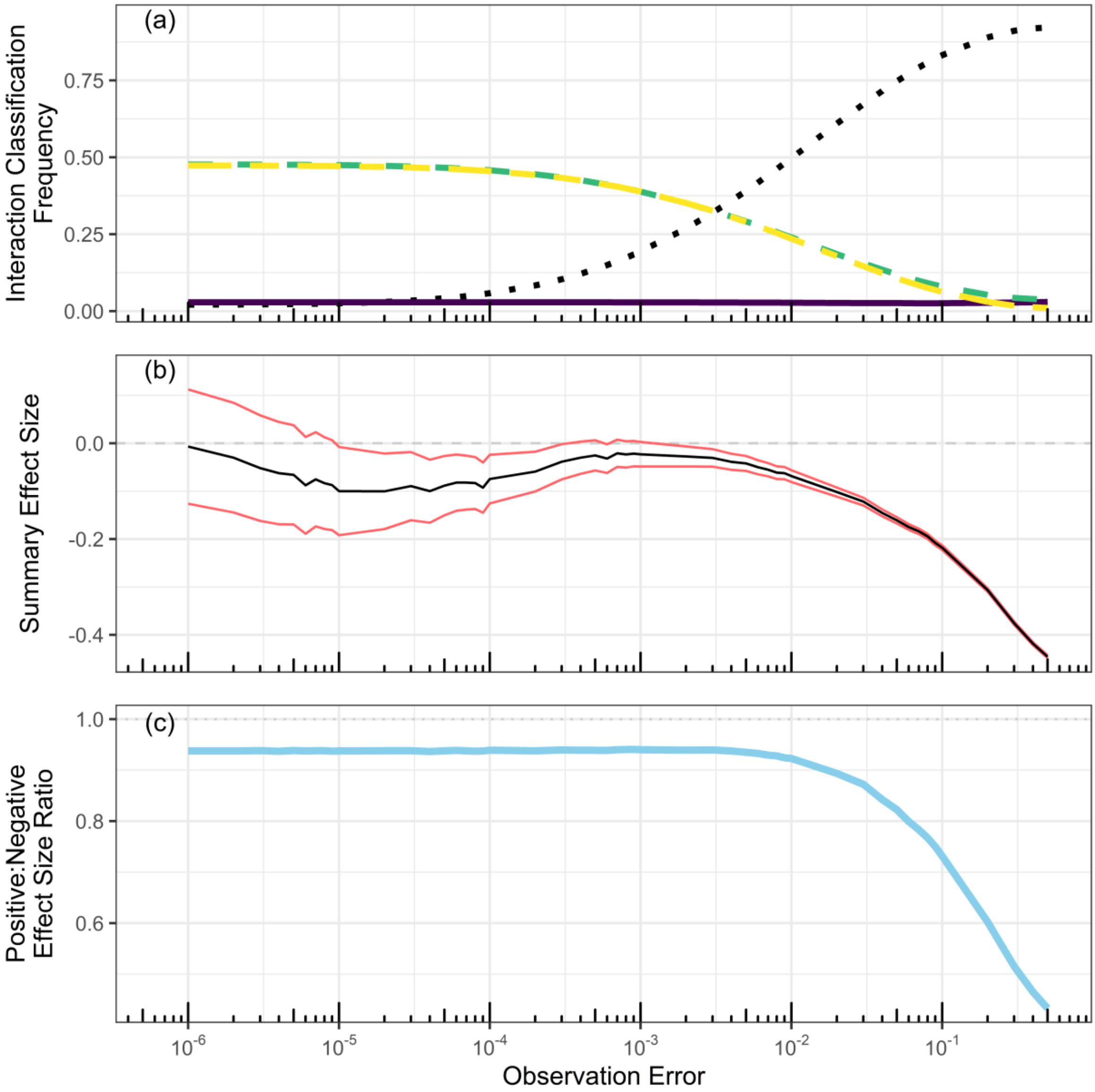

Figure 2.2: The effect of observation error (σ) on the stressor interaction categorisation, and summary meta-analytic effect sizes in the theoretical data. a) Frequency of the different interaction classes for the 360,000 theoretical stressor interactions at each level of observation error: Dotted black line denotes additive interactions; green short-dashed line indicates antagonistic interactions; yellow long-dashed line denotes synergistic interactions; and purple line indicates reversal interactions. b) Summary effect sizes for the 360,000 theoretical stressor interactions, at each level of observation error - Black line denotes summary effect sizes; and red lines denote 95% confidence intervals. c) The ratio of positive to negative summary effect sizes at each level of observation error.

The summary effect size, and summary interaction class as generated from the meta-analytical framework also showed high sensitivity to observation error, although in these analyses the outcome was rather different (Figure 2.2b). For low levels of observation error, the 95% confidence intervals of the summary effect size overlapped zero, indicative of an additive summary interaction class. This occurred because the frequency and magnitudes of synergistic (positive effect size) and antagonistic/reversal (negative effect sizes) interactions were approximately equal for low observation error (Figure 2.2a), effectively cancelling one another out, and the large confidence intervals were caused by the underlying large variance in effect sizes (See Appendix Two). However, with increasing observation error, the summary effect sizes became increasingly more negative, and confidence intervals for these summary effect sizes did not overlap zero, indicating an antagonistic/reversal summary interaction class. This result may seem surprising since, similar to the case of individual interactions (Figure 2.2a), we may expect increased observation error to lead to summary effects with larger confidence intervals that overlapped zero. Instead, we found that, although the ‘true’ stressor interactions (i.e., in the absence of observation error) were roughly equally divided between synergy and antagonism, the summary effect became increasingly more negative as observation error increased, indicating observation error affected synergistic and antagonistic interactions asymmetrically. Further inspection showed an increase in the proportion of negative effect sizes as observation error increased (Figure 2.2c), with this being mirrored by a decreasing summary effect size (Figure 2.2b). Although not so obvious due to the dominance of additive interactions, a similar trend could be observed in the frequencies of interaction classes at higher observation errors, with synergistic interactions heading towards 0 frequency faster than antagonistic interactions (Figure 2.2a). Hence, analyses of our model results with varying levels of observation error suggested synergies in pairs of ecosystem stressors may be under-reported in many empirical studies.

Theoretical expectations

In summary, our theoretical analyses led us to conclude that at biologically plausible levels of observation error (i.e., >0.01), we should expect (i) the empirical data to be dominated by additive interactions for individual interactions (Figure 2.2a), but (ii) in contrast the summary effect sizes computed across a large body of such studies may indicate a dominant role for antagonistic, or reversal, interactions. Both of these results may occur even if, as in our simulated data, synergies are common for the ‘true’ interaction classifications.

Stressor interactions within freshwater empirical data

Our literature search within the Web of Science yielded 1805 papers that met our search criteria, 58 of which met our criteria for inclusion. They included 545 interactions summarised in Figure 2.3 to illustrate the frequency of different interaction classifications and the overall summary effect sizes and interaction classes. Additive interactions were the most frequent (0.829), followed by antagonistic (0.0991), reversal (0.0477), and finally synergistic (0.0239) interactions (Figure 2.3a). Additionally, the summary effect size for the entire dataset was negative with 95% confidence intervals that did not overlap zero (-0.632 ± 0.260), indicative of an antagonistic/reversal summary interaction class (Figure 2.3b).

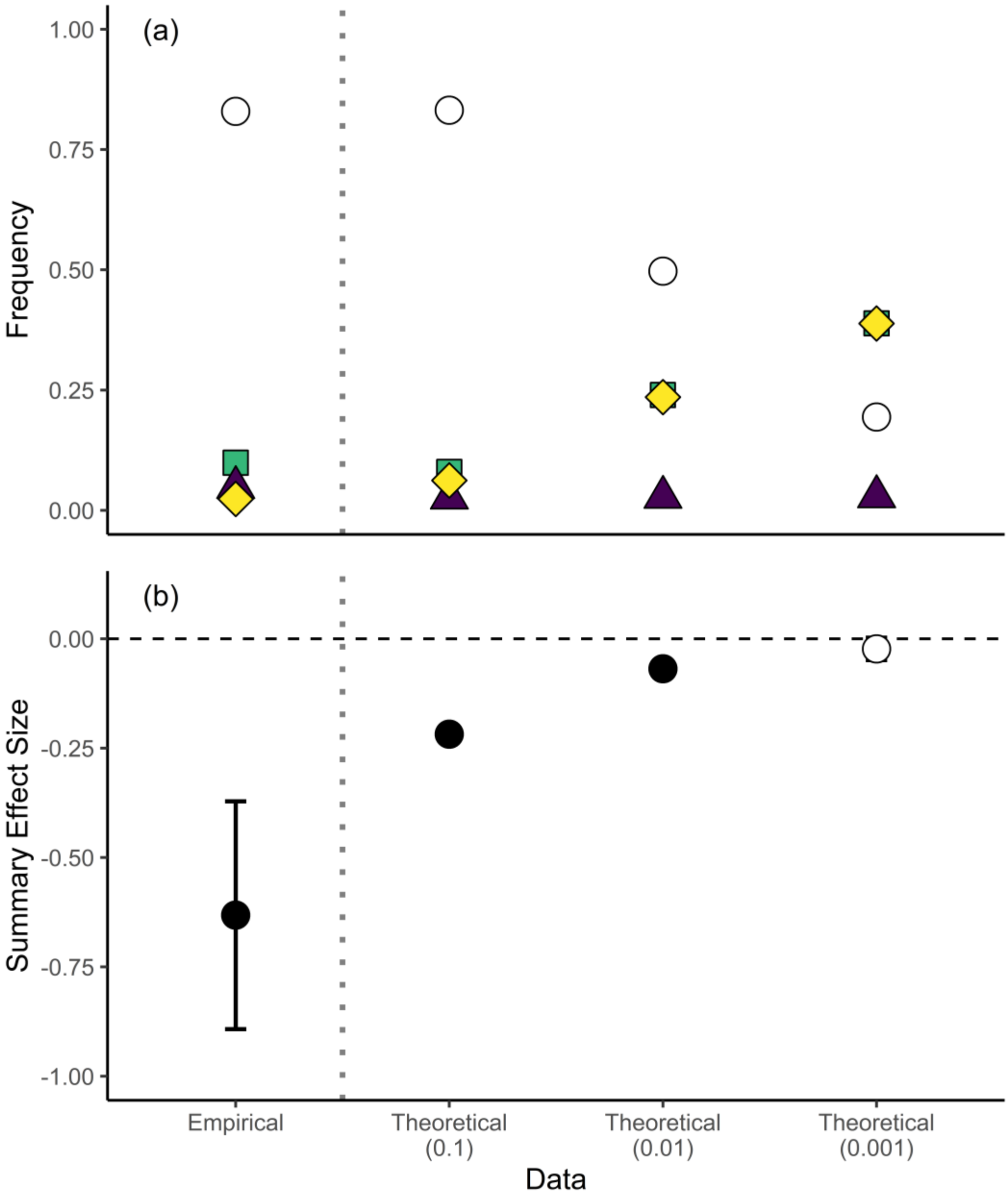

Figure 2.3: Comparisons of the analyses of the freshwater stressor interaction dataset (545 interactions) with the full theoretical stressor interaction dataset for different levels of observation error, σ (given in parentheses on the x-axis). Comparisons are for: a) proportions of the different interaction classes; and b) summary effect sizes for the empirical and theoretical dataset. Figure 2.3a; white circles denote additive interactions, green squares denote antagonistic interactions, yellow diamonds denote synergistic interactions, purple triangles denote reversal interactions. Figure 2.3b; closed circles denote significant summary effect sizes (i.e., 95% confidence intervals do not overlap zero), and open circles denote non-significant summary effect sizes (i.e., 95% confidence intervals overlap zero).

Our meta-analysis showed medium-level heterogeneity (I^2 = 48.5%) although this was considerably lower than the mean heterogeneity (I^2 = 91.7%) found in an analysis of previous ecological meta-analyses (Senior et al., 2016). Furthermore, two additional meta-analyses were conducted on sub-groups of the empirical dataset, with the categorical moderator of organism group used to explore this heterogeneity (Nakagawa et al., 2017). However, these additional meta-analyses failed to uncover any source of this heterogeneity, with the meta-analysis for consumers reporting medium-level heterogeneity (I^2 = 42.5%) and the producer meta-analysis reporting high-level heterogeneity (I^2 = 67.7%) (see Appendix Two).

Comparison of empirical and theoretical interaction classifications

Overall, we found close agreement between our theoretical models with biologically reasonable levels of observation error and the freshwater empirical data (Figure 2.3). Vote-counting results highlight how individual interactions tended to return an additive classification in the empirical dataset, and that this is expected in the theoretical data when estimates of metrics used to classify the interactions are mostly within 10% of the true value (Figure 2.3a). Similarly, summary effect sizes were negative in the simulated data under even a very small level of observation error (Figure 2.3b) despite synergies and antagonisms co-dominating the ‘true’ interactions. This implies the summary effect size reported in the empirical data for freshwater communities (Figure 2.3b) may not necessarily be representative of the underlying ‘true’ stressor interactions.

2.5 Discussion

There has been much interest in understanding and cataloguing the joint effects of stressors on ecological communities and ecosystems (Côté et al., 2016; Schäfer & Piggott, 2018; Thompson et al., 2018b), but to date there has been relatively little guidance from ecological theory. Similarly, we know of no demonstration of the abilities of the statistical tools used for classifying interactions to recover known interactions in the face of data uncertainty. Here, our aim was to test the statistical tools used to define stressor interactions using data simulated from ecological theory in order to gain a deeper understanding of a freshwater dataset. Our empirical analyses generate two main results: (1) vote-counting analyses suggest additive interactions to be by far the most dominant stressor interaction types in freshwater community experiments (Figure 2.3a); but (2) our meta-analysis shows antagonism to be the summary interaction class (Figure 2.3b). However, the analyses of the simulated data suggest both results should be expected under plausible levels of observation error (i.e., >0.01) in the data (Figure 2.3), and that only under unrealistically low levels of precision should we expect to recover the 'true' stressor-pair interactions in either individual studies or meta-analyses (Figure 2.2). We believe that once these statistical aspects are considered, the so-called 'ecological surprises' (*sensu* Paine et al., 1998) may in fact be more prevalent in both our freshwater dataset, and more widely.

Null model sensitivity to observation error

The choice of the null model is hotly debated within ecological stressor research (Schäfer & Piggott, 2018), and it has been argued that null models should be able to accurately predict the combined effects of stressors (Orr et al., 2020). Our results (Figure 2.2) are the first attempt to quantify the degree of accuracy for the most commonly used null model, and we conclude that for all but the very lowest levels of observation error it is difficult to correctly reject the additive null interaction (Figure 2.2a). In other words, we find weak statistical power to recover the underlying stressor-pair interactions. On this basis, and given that most experiments have low sample sizes (we report a mean of 3.83 with a maximum of 16 per treatment in our empirical data), we consider it premature to conclude that most stressor interactions are *truly* additive in the freshwater data we collected. Instead, we should be careful to conclude that in the majority of cases we do not have sufficient evidence to reject the null (additive) interaction. However, it means that we should take notice whenever a non-null interaction is returned by the additive null model, since only strong non-additive effects are likely to be detected (see Appendix Two).

Perhaps more surprising is our finding that meta-analyses using the additive null model report antagonism as the summary interaction classification when observation error is non-negligible, despite synergies co-dominating in our simulation data (Figure 2.2b). A naïve expectation would be for increased observation error to lead to summary effect sizes centred around zero with large confidence intervals making it difficult to rule out an additive summary interaction in our simulation data. Smaller confidence intervals at higher levels of observation error are easily explained by the effect sizes becoming more similar due to the high variances of the response metrics (Appendix Two). Additionally, it is clear that observation error has an asymmetric effect on antagonisms and synergies, with this leading to a shift towards negative effect sizes dominating the distribution of simulated effect sizes (Figure 2.2c). Hence, although the return of an antagonistic summary interaction for our empirical dataset is mirrored in previous analyses of freshwater stressor experiments (Jackson et al., 2016; Lange et al., 2018), we cannot conclude that this is strong evidence for the dominance of antagonism in freshwater ecosystems. The simulation data therefore adds valuable interpretation of our empirical data that would otherwise be missed, and in so-doing highlights the importance of benchmarking statistical tools against data with known attributes.

The high sensitivity to estimation uncertainty may be key reasons why stressor synergies are not as often reported as may be expected (Darling and Côté, 2008; Côté et al., 2016), although other reasons may also contribute, and we can also not rule out that the empirical results do truly reflect the underlying interactions. However, we believe our finding of high sensitivity to observation error in the null model is more general than either our theoretical results, or our freshwater dataset, and we suggest future studies should investigate other null models for their robustness to observation error and sample sizes. Such analyses would build on previous descriptions of the null models (e.g., Sih et al., 1998; Folt et al., 1999; Sih et al., 2004) and would be particularly useful if analyses considered the effect of sample size on statistical power, as this will help guide future empirical studies to improve the detection rate of non-null stressor interactions.

Theoretical expectations for interaction frequencies

Our food chain models imply that, given adequate sample sizes (see above), we should expect synergistic and antagonistic interactions to co-dominate at the population and trophic levels, whereas additive interactions and reversals should be relatively rare. It may well be the case that our models are not good descriptors of the data we analyse; certainly, we ignore much important detail that is likely a feature in the data, such as spatial structure and temporal variation in parameters caused by external perturbations not linked to the stressors, and more complex food web structure involving omnivory or parasitism. Unfortunately, the null model sensitivity to observation error implies we do not yet have the tools with which to discern the relative abilities of different theoretical models to capture the empirical data. However, our key theoretical finding for the relative rarity of additive interactions appears to be echoed in the few other theoretical studies on stressor interactions in ecological communities (e.g., Travers-Trolet et al., 2014; Thompson et al., 2018a; Haller-Bull & Bode, 2019). This agreement is despite a variety of key differences in the model assumptions. In particular, Haller-Bull and Bode (2019) focused on populations rather than multispecies communities, but found dominant roles for synergistic and antagonistic interactions, with additive interactions occurring most frequently for stressors affecting the carrying capacity. Similar to our model, Thompson et al. (2018a) also focused on multispecies communities, but they assumed biological interactions were constant, whereas we allow interactions (consumption and conversion rates) to be modified by stressors, an assumption that seems likely to be met on a regular basis. For example, stressors have been shown to influence resource competition (Kroeker et al., 2013); susceptibility to parasitism in oysters (Lenihan et al., 1999); and modify the flow of energy through aquatic food webs by inducing changes in trophic links (Schrama et al., 2017). Despite this difference, Thompson et al. (2018a) found additive interactions were most prominent when species facilitated each other (i.e., positive species interactions), but that synergy or antagonism in combined stressor effects on species richness or community biomass were more common when species interactions are negative (competition or resource use).

The apparent rarity of additive interactions in all of these models may appear at odds with the possible interpretation that two stressors acting on different species within a community could lead to an additive joint effect (Jackson et al., 2016). However, feedbacks in the food web, like those found in our models, mean that even if a species is unaffected directly by a stressor, it is highly likely that top-down or bottom-up effects will lead to indirect interactions for many species, and as a result, additive interactions are extremely unlikely in the absence of uncertainty (e.g., observation error). Indeed, we anticipate that additive interactions may only truly occur in scenarios where species in different and very weakly interacting sub-communities are affected by different stressors, or, as found by Thompson et al. (2018a), where species interactions are predominantly positive. We believe there will be an increasing role of theory in generating hypotheses for the ways in which stressors interact (De Laender, 2018), and the most progress will be made when the theory is developed so it can be directly compared against empirical data, much as we have done here.

Mechanistic understanding of multiple stressors

In this study, we sought to address the question of *how* multiple stressors interact. This approach, when applied across both theoretical and empirical datasets can allow us to discern what may be expected across the interactions of multiple stressors. Future research may seek to address the question of *why* multiple stressors interact in the manner that they do. Undoubtedly, these two questions are entwinned, with the answers to each of these questions highly likely to be dependent on the other. However, while the use of null models is essential in determining the combined effect of multiple stressors (Thompson et al., 2018b), the adoption of a mechanistic approach to investigating multiple stressors may provide novel insights which address these joint questions (De Laender, 2018; Schäfer & Piggott, 2018). For instance, a mechanistic understanding may allow for responses such as co-tolerance or co-susceptibility (Todgham & Stillman, 2013) to stressors to be more thoroughly understood from an ecological perspective. Ultimately, as our results imply, such an understanding is likely to require a large amount of empirical data to fully understand; however, there is ample scope for theoretical ecology to help fill this gap in our collective understanding of multiple stressors, and to generate specific hypotheses to be tested. Similarly, a mechanistic understanding of multiple stressor interactions would prove invaluable when mitigating the effects of stressors or implementing conservation initiatives.

Future developments

Our analysis represents a novel approach combining both theoretical and empirical methods. While this analysis provides a solid foundation, there are several aspects that could be adjusted in future research. Firstly, there is a clear need to better understand the limitations and data requirements of the null models (e.g., Gurevitch et al., 2000; Lajeunesse, 2011;

Thompson et al., 2018b) that are used to classify stressor interactions. Such knowledge would be very useful in guiding experimental design that would maximise the probability of uncovering non-null stressor interactions and would therefore provide a better understanding of their true prevalence. Knowing how many data points are required before we can realistically hope to detect a particular type of pattern, in this case a stressor interaction type of a given strength, is a critical component of experimental design. Moreover, our work has also uncovered some hitherto undescribed biases that lead to meta-analyses potentially over-emphasising antagonisms, and it is important to investigate other null models for this feature as well as looking for methods to reduce this bias. Secondly, the theoretical communities manipulated here combine multiple populations each on a separate trophic level. While this builds upon similar research conducted on a single population (Brown et al., 2013; Haller-Bull & Bode, 2019), there is scope for this approach to be expanded to consider more complex communities, for instance with multiple populations on a single trophic level (e.g., Thompson et al., 2018a). Finally, the manner in which stressors interact at the parameter or process level can occur in numerous ways, for instance either additively or multiplicatively (Haller-Bull & Bode, 2019). However, whether a process or parameter is impacted in an additive or multiplicative manner, will cause a stressed parameter value to change by differing degrees, with this in turn potentially resulting in contrasting frequencies of interaction classifications at the population level. Accordingly, the manner in which a process or parameter (e.g., feeding rate, mortality) is impacted may be determined by the individual stressors; for instance, if two simultaneously acting stressors are entirely independent of one another then their effect on an ecological process may be additive (Haller-Bull & Bode, 2019). Consequently, allowing stressors to impact the same process undoubtedly represents an area for expansion, particularly when considering how impacts at the parameter level affect population level properties.

Conclusions

Determining the ways multiple stressors interact is key when attempting to mitigate their effects, with the class of the observed interaction potentially outlining whether the removal of a stressor will have a beneficial, limited, or detrimental impact to an ecosystem (Brown et al., 2013; Côté et al., 2016). Our results show the value of developing a theoretical framework which can aid in the interpretation of environmental stressor interactions, and we hope more general theory that makes specific predictions based on ecological mechanisms (e.g., De Laender, 2018; Fu et al., 2018; Thompson et al., 2018a) will be developed *and* tested in future. However, our results also highlight the urgent need to better understand the strengths and limitations of the null models that are used to classify the cumulative effects of community stressors, and we also believe a unified approach to the meta-analyses of individual studies will increase our understanding of how environmental stressors combine.

Chapter Three – Multiple stressor null models frequently fail to detect interactions due to low statistical power

3.1 Abstract

Ecosystems across the globe are being impacted by multiple anthropogenic drivers. One pressing question facing ecologists is understanding how these multiple stressors interact to impact ecosystems. Predominately, studies have investigated the interactions of stressors using null models, with the additive and multiplicative null models being those most widely applied. Such approaches classify interactions as being synergistic, antagonistic, reversal, or null. Despite their wide-spread use, there has been no thorough analysis of these null models, nor a systematic test of the robustness of their results to sample size or sampling error in the estimates of stressor effects. Using simulated food web models, we demonstrate that the additive and multiplicative null models are not directly comparable, illustrated by the null models assigning over a third of all interactions different classifications. We highlight that both null models have weak power to correctly classify interactions at commonly implemented sample sizes (i.e., ≤6 replicates), unless data uncertainty is unrealistically low; hence the majority of interactions are assigned a null classification given they are indistinguishable from the implemented null model. Using mathematical approximations and simulations, we demonstrate that increasing sample size increases the power to detect the true interactions; however, power only slowly increases with sample size. We demonstrate that for common experimental sample sizes, only exceptionally large effect sizes are able to be assigned a non-null classification. Ultimately, our results may aid researchers in the design of their experiments, and the subsequent interpretation of their results. Overall, our results show no clear statistical advantage of using one null model over the other, although we conclude that it is not possible nor even meaningful to compare interaction types under different null models. Furthermore, the low statistical power of commonly used null models means we are likely missing many synergistic and antagonistic stressor interactions.

3.2 Introduction

Globally, ecosystems are being impacted by a plethora of external anthropogenic stressors (*sensu* disturbances, drivers, factors, or pressures) (Blowes et al., 2019; Christensen et al., 2006), with such stressors encompassing a wide range of environmental or biotic changes including land-use change, invasive species, climate change, and pollution (Brook et al., 2008; Dirzo et al., 2014; Hillebrand et al., 2020; Jackson et al., 2020). As such, an individual stressor is capable of having impacts upon biodiversity, species abundances, and ecosystem services (Newbold et al., 2015; Sala et al., 2020; Tittensor et al., 2014; Vinebrooke et al., 2004), with ecosystems within freshwater, marine, and terrestrial realms all at risk (Beaumelle et al., 2021; O'Hara et al., 2021; Reid et al., 2019). Although ecosystems are frequently subjected to multiple stressors, understanding how these concurrently acting stressors interact is difficult, with the combined effects of these stressors frequently unknown (Hodgson & Halpern, 2019). Accordingly, understanding, predicting, and mitigating the effects of multiple interacting stressors upon various ecosystem properties represents one of the major, yet urgent, challenges to be confronted by ecologists and conservationists (Côté et al., 2016; Jackson et al., 2021; Lindenmayer et al., 2020).

At present, the inferred effects of multiple interacting stressors are predominately determined through the implementation of null models (De Laender, 2018), where the observed response is compared to an expectation that the stressors are non-interacting. Of these null models, the additive null model (Gurevitch et al., 2000) is the most widely applied (e.g., Crain et al., 2008; Jackson et al., 2016) and suggests that the overall effect of the multiple interacting stressors is equal to the sum of the effects of the stressors acting individually. While it has been previously assumed that the majority of interactions are able to be explained by the additive null model, there is growing evidence that substantial numbers of stressor interactions have effects different to those predicted by the additive null model (Côté et al., 2016). Increasingly, studies are illustrating the prevalence of interactions with effects that are: i) greater than anticipated by the additive null model (*synergistic interactions*); ii) less than predicted by the additive null model (*antagonistic interactions*); iii) opposite to that suggested by the additive null model (*reversal interactions*). Frequently, these non-additive interactions are considered to be *ecological surprises* (*sensu* Paine et al., 1998), namely that they deviate from the expectation of the additive null model. Comparisons of the different interaction classes are illustrated by Figure 3.1.

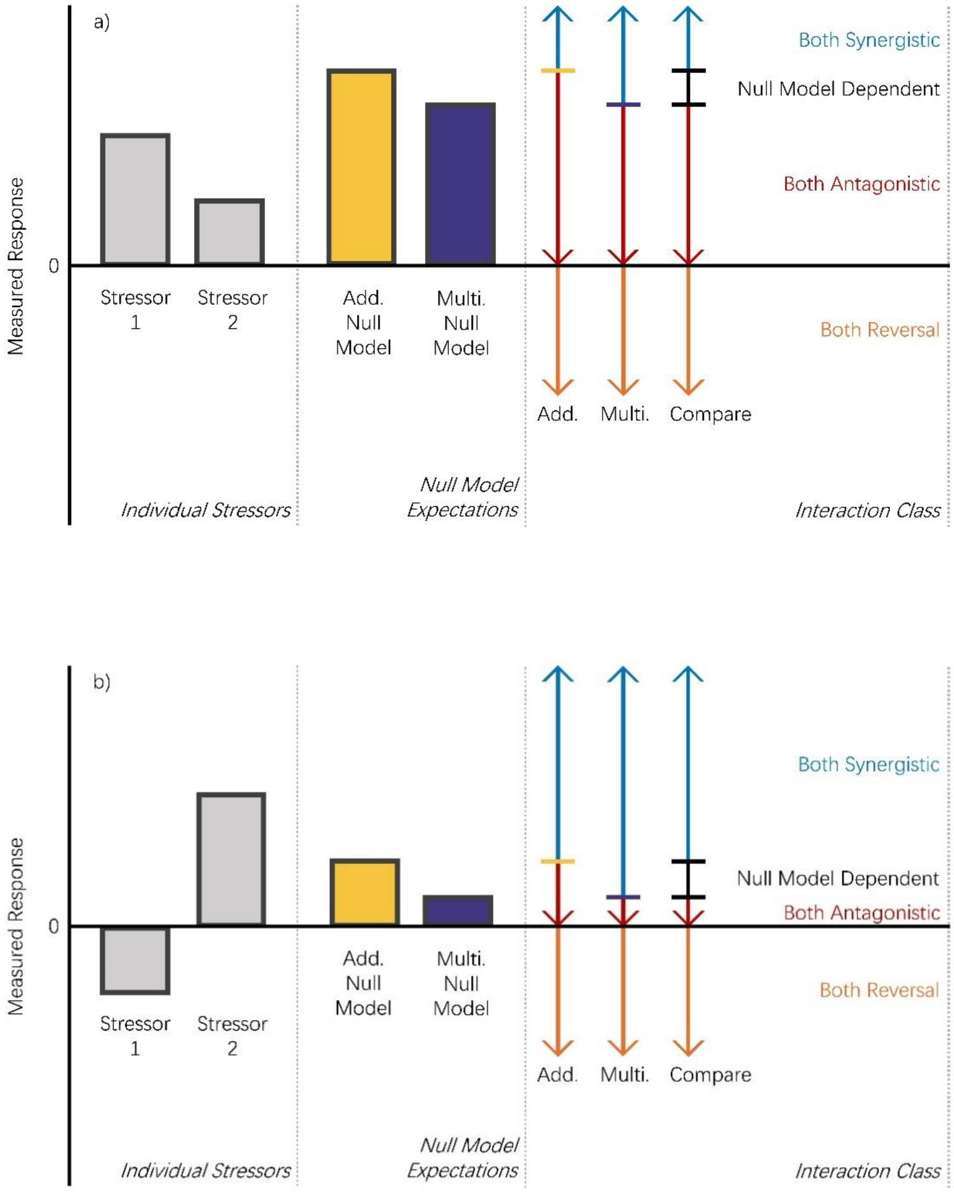

Figure 3.1: Comparison of the additive and multiplicative null models, illustrating the conditions under which the different interaction classifications are observed. Individual Stressors: Grey bars represent the change in a measured response due to a given stressor (Stressor 1 or Stressor 2). Null Model Expectations: Yellow bar represents the expected response under an additive null model; Purple bar represents the expected response under a multiplicative null model. Interaction Class: Arrows denote how an interaction would be classified for a given observed response. Blue arrows denote synergistic

interactions; Pink arrows denote antagonistic interactions; Orange arrows denote reversal interactions. For the additive and multiplicative null models, a null interaction is shown by the purple and yellow lines respectively. When comparing between the additive and multiplicative null models, the black line denotes the range of observed responses where the additive and multiplicative null models classify interactions differently (i.e., the interaction class is null model dependent). Whether an interaction class varies between null models, or not, is inherently explainable by the underlying algebra of the null models (see Appendix Three). a) Both Stressor 1 and Stressor 2 act upon the response in the same direction. b) Stressor 1 and Stressor 2 act upon the response in contrasting directions.

While the majority of studies currently apply the additive null model there is a growing body of literature suggesting alternative null models (e.g., Tekin et al., 2020; Thompson et al., 2018b) that may potentially be more appropriate when investigating the impacts of multiple stressors (Dey & Koops, 2021; Schäfer & Piggott, 2018). Frequently, this argument centres on the parsimonious nature of the additive null model which limits, or even prevents, any mechanistic insight into *how* stressors interact; with there similarly being calls for multiple stressor research to focus on the mechanisms underpinning stressor interactions (Orr et al., 2020; Schäfer & Piggott, 2018). Accordingly, the multiplicative null model (Hawkes & Sullivan, 2001; Lajeunesse, 2011) is the alternative which is most widely discussed (see Folt et al., 1999; Morris et al., 2007; Sih et al., 1998) and implemented (e.g., Gomez Isaza et al., 2020; Harvey et al., 2013). The simplest description of the multiplicative null model is that it represents the logarithmic form of the additive null model. However, this numeric transformation results in differences between the two null models including the fundamental assumptions underpinning each approach (Schäfer & Piggott, 2018). Given these assumptions, it has been suggested that the multiplicative null model may better reflect biological and ecological systems (Kerkhoff & Enquist, 2009), with this being particularly pertinent when certain responses are considered (e.g., mortality or survival) (Fournier et al., 2006). Furthermore, while interactions under the multiplicative null model are capable of varying from the expected effect in a similar manner to the additive null model, the classification assigned to any given interaction (i.e., synergistic, antagonistic, or reversal interactions) may also differ between the two null models (Figure 3.1). For instance, an interaction classed as being antagonistic by the additive null model may be deemed as being a synergistic interaction by the multiplicative null model.

Ultimately, the differences between the two null models, outlined above, result in several important ramifications. Firstly, when individual studies implement differing null models, direct comparisons of these studies are decidedly difficult; indeed, as a result of the contrasting assumptions underpinning each null model, the hypotheses they test are different. Secondly, where the conservation or management of a particular ecosystem is based on the results of a specific null model a different course of action may have been adopted if an alternative null model was used. Furthermore, there are potential consequences to incorrectly determining the interaction classification with conservation strategies being rendered ineffective or even detrimental (Brown et al., 2013; Côté et al., 2016), and it is therefore imperative to understand how the additive and multiplicative null models inherently relate to one another.

When planning an experiment, one fundamental criterion that ecologists should consider is whether an experimental design is capable of detecting a biologically important effect (Steidl et al., 1997). For the above null models, such prospective power analyses necessitate knowledge of the sample sizes required to detect an effect of a given strength against a backdrop of data uncertainty. In the particular case of stressor interactions, the aim of all null model tests is to uncover an effect that shows a departure from the null hypothesis (i.e., observing an antagonistic, reversal, or synergistic interaction). Accordingly, there is a need to understand how each of the three constituent components of the null model calculations (treatment means, treatment uncertainty, and sample sizes) can influence the results of these null models. However, we are not aware of any previous analysis which attempts to understand the role of each of these components, with this meaning that there is currently minimal guidance for ecological experiments. Indeed, there is currently a distinct lack of knowledge regarding how experimental design, or ecologically relevant levels of uncertainty can impact the results of these null models, or even the conditions under which results differ between the additive and multiplicative null models. Without knowledge on the attributes or statistical power of these null models it can be very hard to make robust conclusions from the results of model tests, or design experiments that are likely to yield results that reflect the underlying effects of the stressors on the biological system of concern. As examples, two recent multiple stressor meta-analyses (Gomez Isaza et al., 2020; Seifert et al., 2020) included no experiments with more than six replicates per treatment, while a third (Burgess et al.,

2021) found <1% of experiments used more than eight replicates per treatment. Currently there is no understanding of how these samples sizes might impact the power of either null model.

Here, we focus on closing these gaps in knowledge about the additive and multiplicative null models for stressor interactions, outlining the role that treatment means, treatment uncertainty, and sample sizes play in determining model results. To this aim, we simulate data using an ecological model that has been shown to produce plausible distributions of interaction types (Burgess et al., 2021), whilst allowing for full control over the sample sizes and level of observation (measurement) error in the data. Using this simulated data, where we know the 'true' interaction type (i.e., that inferred in the absence of observation or sampling error), we are able to address the following questions:

Question one: What are the conditions under which the additive and multiplicative null models agree, and disagree, and can we define these conditions from the model formulae?

Question two: To what degree does treatment uncertainty (in the form of observation error) reduce the ability of both the additive and multiplicative null models to correctly infer the 'true' interaction classification?

Question three: What is the relationship between the power to correctly reject the null models and sample size, and are current ranges of sample sizes adequate to detect most non-null interactions?

3.3 Methods

Additive null model

Within multiple stressor research the form of the additive null model most commonly implemented is the factorial iteration of the null model, (namely *Hedges' d*), outlined by Gurevitch et al., (2000). *Hedges' d*, (henceforth referred to as the additive null model), is an estimate of the standardised mean difference between the means of the control and treatment samples, but also has the benefit of being unbiased by small sample sizes (Hedges & Olkin, 1985). The additive null model is calculated by comparing the effect of the stressors acting separately to the effect of the stressors acting simultaneously. Within our analysis we also implement this form of the null model (Equation 3.1). The calculation of the additive null model depends on three variables, (namely the mean response value (X_X), the number of replicate measurements (N_X), and the standard deviation around the mean (SD_X)), with each of these variables being taken under the four different treatments, (Control (C), Stressor A only (A), Stressor B only (B), Stressors A & B (I)). The calculation of the additive effect size, (ES_{Add}), is shown by Equation 3.1.1, with the algebraic notation used as described above. The pooled sampling standard deviation is denoted by s, (Equation 3.1.2); while $J(m)$ (Equation 3.1.3) is the small sample bias correction factor, often used where there is a small number of replicates per interaction (Borenstein et al., 2009). For each additive effect size a corresponding variance, (V_{Add}), standard error, (SE_{Add}), and confidence intervals, (CI_{Add}), are calculated by Equations 3.1.4 – 3.1.6. Here, we set the significance level to 0.05 ($Z_{\alpha/2}$ = 1.96), hence 95% confidence intervals are calculated.

3.1.1)
$$ES_{Add} = \frac{X_I - X_A - X_B + X_C}{s} \cdot J(m)$$

3.1.2)
$$s = \sqrt{\frac{(N_I - 1) \cdot (SD_I)^2 + (N_A - 1) \cdot (SD_A)^2 + (N_B - 1) \cdot (SD_B)^2 + (N_C - 1) \cdot (SD_C)^2}{N_I + N_A + N_B + N_C - 4}}$$

3.1.3)
$$J(m) = 1 - \frac{3}{4 \cdot (N_I + N_A + N_B + N_C - 4) - 1}$$

3.1.4)
$$V_{Add} = (J(m))^2 \cdot (\frac{1}{N_I} + \frac{1}{N_A} + \frac{1}{N_B} + \frac{1}{N_C} + \frac{(ES_{Add})^2}{2(N_I + N_A + N_B + N_C)})$$

3.1.5)
$$SE_{Add} = \sqrt{V_{Add}}$$

3.1.6)
$$CI_{Add} = Z_{\alpha/2} \cdot SE_{Add}$$

Multiplicative null model

We implement the form of the multiplicative null model detailed by Lajeunesse (2011). Despite a somewhat similar appearance to the additive null model, the multiplicative null model applies a logarithmic transformation to the measured response values (Equation 3.2.1) and again incorporates measurements of the same variables for all four treatments used in the additive null model (notation used is the same as above). The multiplicative effect size, (ES_{Mul}), is calculated by Equation 3.2.1; while the corresponding variance, (V_{Mul}), standard error, (SE_{Mul}), and confidence intervals, (CI_{Mul}), are determined using 3.2.2 – 3.2.4. As above, we set the significance level to 0.05 ($Z_{\alpha/2}$ = 1.96), hence 95% confidence intervals are calculated.

3.2.1) $$ES_{Mul} = \ln(X_I) - \ln(X_A) - \ln(X_B) + \ln(X_C)$$

3.2.2) $$V_{Mul} = \frac{(SD_I)^2}{(X_I)^2 \cdot N_I} + \frac{(SD_A)^2}{(X_A)^2 \cdot N_A} + \frac{(SD_B)^2}{(X_B)^2 \cdot N_B} + \frac{(SD_C)^2}{(X_C)^2 \cdot N_C}$$

3.2.3) $$SE_{Mul} = \sqrt{V_{Mul}}$$

3.2.4) $$CI_{Mul} = Z_{\alpha/2} \cdot SE_{Mul}$$

Interaction classifications

When using either the additive or multiplicative null models, individual interactions are able to be classified into one of four classes depending upon both effect sizes (Equations 3.1.1, 3.2.1) and corresponding confidence intervals (Equations 3.1.6, 3.2.4). Using the naming conventions of (Orr et al., 2020), the classifications are: null, antagonistic, synergistic, or reversal interactions (Figure 3.1). An interaction was classed as null, regardless of whether the effect size was positive or negative, if the 95% confidence intervals for that interaction overlapped zero. Interactions were classed as synergistic if the effect size was positive and the 95% confidence intervals for that interaction did not overlap zero. Interactions were classed as antagonistic or reversal interactions, if the effect size was negative, and the 95% confidence intervals did not overlap zero. Furthermore, for an interaction to be classed as antagonistic, the expected and observed interaction effects had to both act in the same direction (e.g., for an additive interaction: $X_I - X_C > 0$ & $X_A + X_B - 2X_C > 0$). In contrast, for an interaction to be classed as reversal, the expected and observed interaction effects had to both act in differing directions (e.g., for an additive interaction: $X_I - X_C < 0$ & $X_A + X_B - 2X_C > 0$). Finally, it is important to note that a null interaction classification does not necessarily mean that stressors interact in an additive or multiplicative manner. A null classification denotes interactions where it was not possible to reject the null hypothesis (i.e., null model), not that the additive, or multiplicative, null model was accepted. The method for classifying interactions is explored more in Appendix Three.

Theoretical framework

Described below is an overview of the framework used to simulate interaction data for our analyses. For full details, including formulae and details of distributions, refer to Appendix Three.

Throughout our analysis we implement a theoretical framework (Burgess et al., 2021) to generate population densities analogous to those from empirical experiments. The theoretical framework is based upon Lotka-Volterra consumer-resource equations (Heath et al., 2014), enabling us to establish food chains of three, four, or five trophic levels and determine the densities of populations at equilibrium. Regardless of the length of food chain considered, the models contain three distinct types of trophic level. These levels are: i) basal trophic level, equivalent to a pool of a limiting nutrient; ii) consumer trophic level, which only predates the trophic level directly below, and is only predated upon by the trophic level directly above; iii) apex consumer trophic level, which only predates on the trophic level directly below, but is not predated upon itself. Both Heath et al., (2014) and Burgess et al., (2021) outline these models to a greater degree with additional detailing found there.

In brief, the food-chain models (and therefore population densities) are governed by the densities of the trophic levels and five key parameters; namely attack rate, α; conversion efficiency, ε; density independent mortality rate, δ; density dependent mortality rate λ; and the basal (i.e., nutrient) input rate ω. The attack rate, α_i, denotes the rate at which trophic level *i* feeds upon a resource (trophic level *i-1*). The conversion efficiency, ε_i, represents the proportion of the consumed resource that is converted into new consumers within trophic level *i*. The density independent mortality rate, δ_i, represents the background rate at which the density of trophic level *i* is reduced with each model time-step. The density dependent mortality rate, λ_i, increases the mortality rate of a trophic level with increasing population density. Finally, the basal input rate, ω, represents the constant rate at which the basal trophic level increases in density.

We aim to investigate how stressors interact to impact the equilibrium densities of trophic levels within a theoretical food chain. As with empirical studies of multiple stressor interactions, our theoretical framework similarly employs a factorial experimental design. For the control treatment, the values of each of the above parameters (for each trophic level) are drawn from pre-determined uniform distributions. Each of these parameters were randomly drawn across all trophic levels; hence for example, the value for the conversion efficiency parameter at trophic level *i* is unlikely to be the same as that for trophic level *j*. Following the method used by Burgess et al., (2021) we assume that all the previously described key parameters do not vary over time, with the intention of this framework to explore how stressors impact equilibrium densities.

Subsequently, within our theoretical framework, food chains are subjected to stressors by altering the value for a single parameter (e.g., α, ε, δ, or ω) for a single trophic level, and keeping all other parameters constant under control values. Within these simulations, all parameters are equally likely to be selected to be stressed, although a single parameter is only ever impacted by one stressor (i.e., two stressors cannot act on the same parameter). For the interaction treatment of our experimental design, all parameters were under control conditions, except for those which were stressed in the individual treatments which were instead assigned their stressed values. Accordingly, for each of the four treatments we calculate the equilibrium densities of each trophic level within the food chain, and we use these densities as our treatment means to classify the stressor interactions using the null models (Equations 3.1 and 3.2, *Interaction classifications* section). When determining how stressors interact to impact populations, we randomly chose one trophic level from the entire theoretical food chain, (with the exception of the basal/nutrient trophic level which was never selected). Accordingly, our subsequent analysis is centred on individual trophic levels rather than entire communities; however, this approach means that the trophic level investigated was not always directly impacted by a given stressor but was able to be indirectly impacted through cascading effects.

Implementing the above approach, a dataset comprising of 100,000 interactions was collated. These interactions represented approximately equal numbers of equilibrium densities for populations from the three different lengths of food chain. The calculation of the equilibrium densities for each food chain, across each of the four treatments, was conducted using Mathematica v10.4 (Wolfram Research Inc., 2016). In line with the method of Burgess et al., (2021), the equilibrium densities of all 100,000 interactions were both stable and greater than zero. These 100,000 interactions form the basis for each of subsequent analyses, detailed below.

Treatment means, treatment uncertainty, and sample sizes

Question one centres around understanding the additive and multiplicative null models under idealised conditions (e.g., in the complete absence of uncertainty, or with infinite sample sizes). Such an analysis provides the clearest indication of how the null models differ from one another, based solely upon the 100,000 treatment means described above, allowing the 'true' classification of any interaction to be determined. Accordingly, a focus upon treatment means allows the conditions under which the results of the null models agree or disagree to be better understood.

Given that the food chain models are solved to a given equilibria, there is an absence of uncertainty at this stage. To mirror empirical experiments, we modelled treatment uncertainty, in the form of observation error for the equilibrium densities recorded at each treatment in each of the 100,000 interactions. In brief, we follow the method of Burgess et al., (2021), with each equilibrium density being multiplied by a number drawn from a Gaussian distribution with a mean of 1.00 and a standard deviation σ, with this process being repeated four times per treatment; with the number of replicates used here mirroring those frequently used within empirical experiments. Smaller values of σ will result in a tighter distribution of observed mean densities around the true population mean compared to larger values of σ. Our approach allows for an intuitive understanding of σ, as we can expect 99.7% of all observed treatment densities to fall within 3σ of the true value. We use 250 levels of σ, hence referred to as levels of observation error, ranging from $1.0\text{x}10^{-3}$ to $2.5\text{x}10^{-1}$ (see Appendix Three for more details). We then compute the probability of assigning the true interaction classification to each interaction under every level of observation error, thereby investigating the ability of each null model to correctly classify interactions in the face of uncertainty (Question two).

To address Question three and investigate the role which sample size plays in the determination of interaction classifications, a similar method was employed as for treatment uncertainty (see Appendix Three). For three levels of observation error (0.001, 0.01, 0.1) the number of replicates per treatment was varied between three and 100, and for each combination of sample size and observation error, we computed the frequency of interactions that were assigned the 'true' interaction classification for each null model (see *Treatment means*). As such, this analysis provides a form of power analysis quantifying how increases in sample size may improve the ability of empirical experiments to detect significant interactions.

3.4 Results

Question one: What are the conditions under which the additive and multiplicative null models agree, and disagree, and can we define these conditions from the model formulae?

Despite testing differing hypotheses, both the additive and multiplicative null models report similar frequencies of the different interaction classifications in the absence of any variation. Both null models report that antagonistic interactions are the most prevalent (additive 0.493; multiplicative 0.519), followed by a slightly reduced number of synergistic interactions (0.485; 0.462), minimal frequencies of reversal interactions (0.023; 0.020), and a complete absence of null interactions (0.000; 0.000).

The frequencies of the interactions reported suggest a high degree of alignment between both null models. However, when considering how individual interactions are classified by both null models, it is evident that substantial differences exist (Table 3.1). While the majority of interactions are assigned the same classification, over one third of interactions (33.6%) are assigned classifications which are null model dependent. This difference is predominately explained by interactions being assigned a synergistic classification by one null model and an antagonistic class by the other.

Table 3.1: Proportions of interactions classed differently, or the same, across additive and multiplicative null models. Shaded boxes indicated the proportion of interactions where both null models agree.

		Additive Null Model			
		Synergistic	Antagonistic	Reversal	Null
Multiplicative Null Model	Synergistic	0.299	0.158	0.005	0.000
	Antagonistic	0.181	0.328	0.010	0.000
	Reversal	0.005	0.007	0.007	0.000
	Null	0.000	0.000	0.000	0.000

Question two: To what degree does treatment uncertainty (in the form of observation error) reduce the ability of both the additive and multiplicative null models to correctly infer the 'true' interaction classification?

As shown by Figure 3.2, both null models are sensitive to data uncertainty in the form of observation error. The frequency of null interactions rapidly increases accounting for over 50% of all interactions for both null models by σ~0.005 (i.e., when nearly all sampled treatment means are within 1.5% of the true values). Hence even at small levels of observation error both the additive and multiplicative null models are frequently unable to be correctly rejected. Accordingly, under the levels of observation error considered here, it is expected that the majority of interactions would be assigned a null interaction class. Indeed, as shown for both the additive (Figure 3.2a) and multiplicative (Figure 3.2b) null models, the frequencies of synergistic and antagonistic interactions rapidly decline with increasing levels of observation error. However, the frequency of reversal interactions is relatively constant, regardless of the level of observation error.

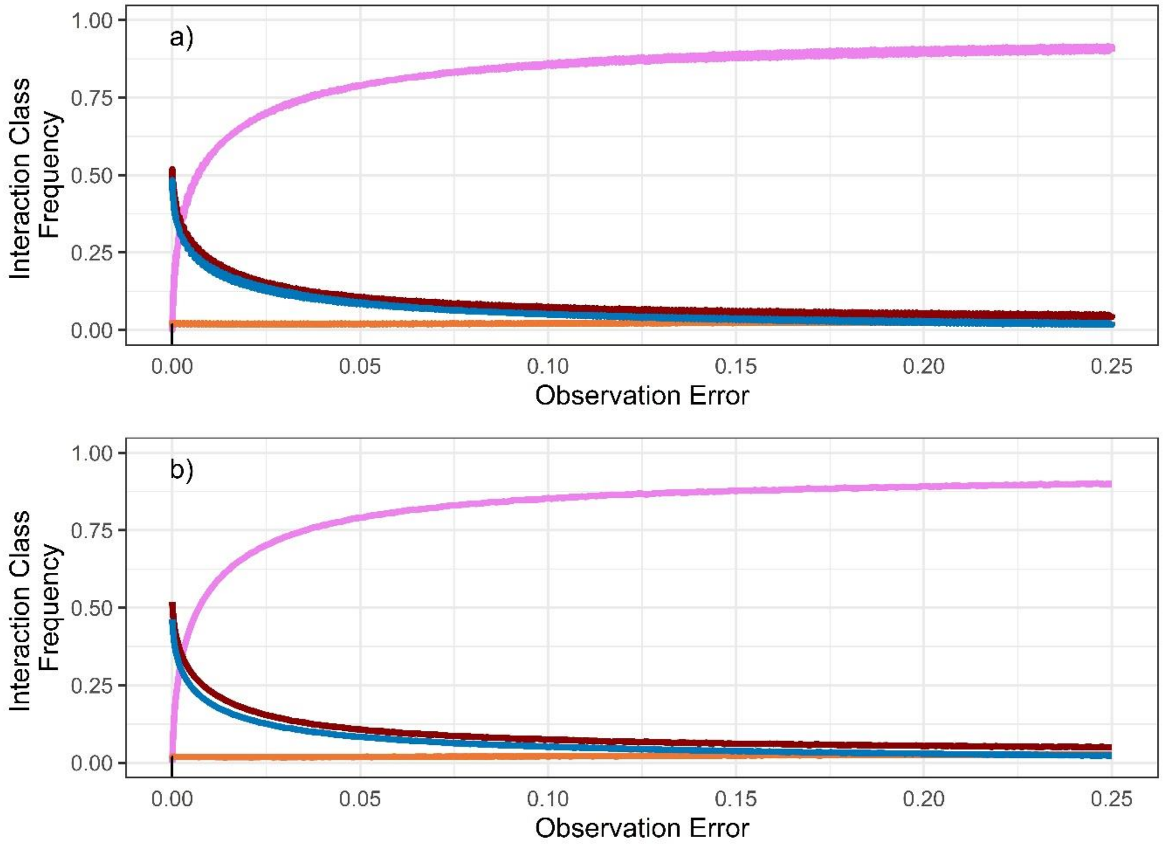

Figure 3.2: The effect of observation error on the frequency of interaction classifications, showing how the proportions differed from those in the absence of observation error, using an additive null model (a) and multiplicative null model (b). Proportions of the different interaction classes are for the 100,000 simulated interactions at each level of observation error. Pink line indicates null interactions. Orange line denotes reversal interactions. Blue line indicates synergistic interactions. Red line denotes antagonistic interactions.

Question three: What is the relationship between the power to correctly reject the null models and sample size, and are current ranges of sample sizes adequate to detect most non-null interactions?

As shown by Figure 3.3, increasing sample size leads to a higher rate of correctly rejecting both the additive and multiplicative null models (i.e., an increase in the true positive rate). However, even for large sample sizes, it can be difficult to detect the ‘true’ interaction if observation error is anything other than minimal. Moreover, at lower levels of observation error (i.e., 0.001 or 0.01) the multiplicative null model is correctly rejected more frequently than the additive null model; however, the additive null model is correctly rejected, more frequently, for the highest level (0.1) of observation error. Our results suggest that increasing those sample sizes frequently observed in multiple stressor studies, will result either null model being correctly rejected more frequently.

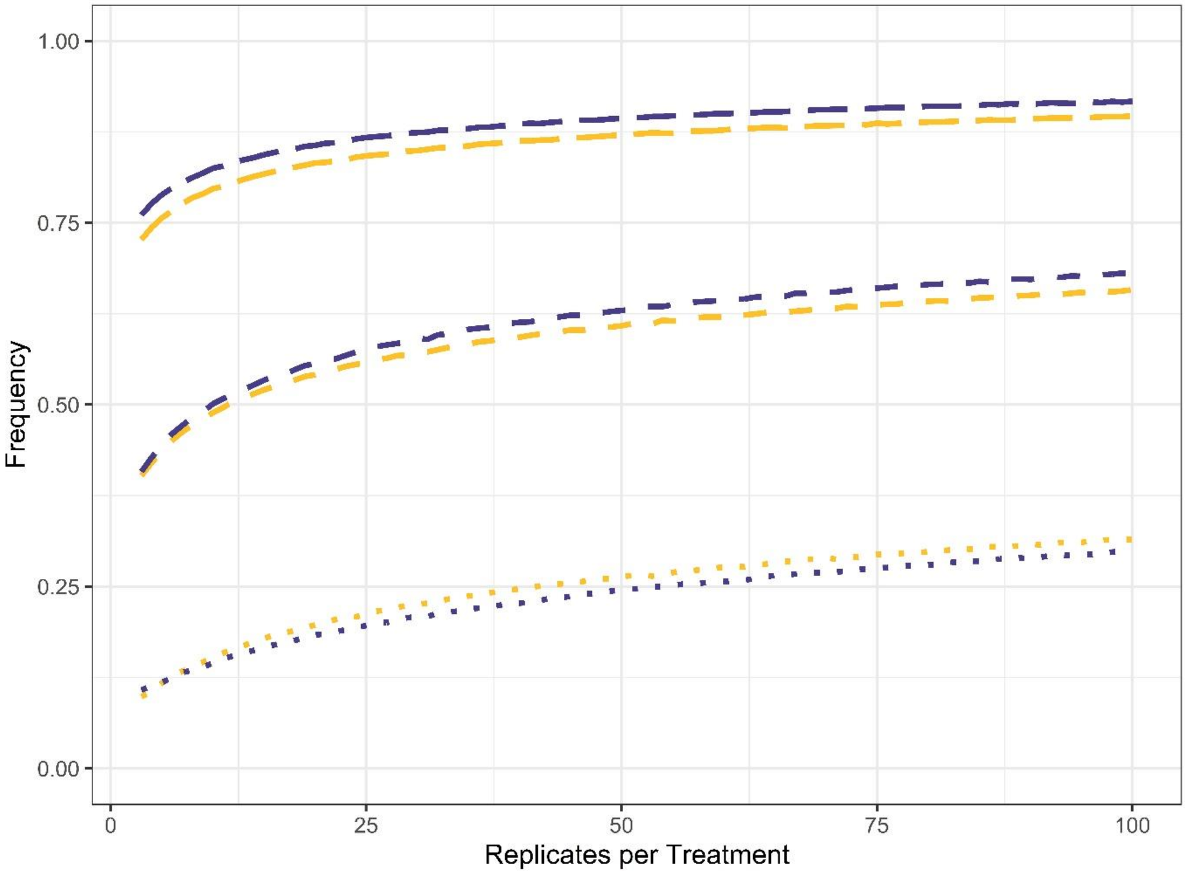

Figure 3.3: Frequency of interactions that are correctly classified (i.e., true positive rate) for a range of different treatment replicates. Three different levels of observation error are considered (0.001, 0.01, 0.1), with these being indicated by longdashed, dashed, and dotted lines respectively. Yellow lines denote use of an additive null model, purple lines denote use of a multiplicative null model.

By combining and rearranging Equations 3.1.1 – 3.1.6, it is possible to express the additive effect size as a function of sample size (Inequality 3.1). This inequality shows that for a given sample size, there is a minimum additive effect size value which must be exceeded in order for the interaction to be significantly different to the null model (i.e., critical effect size; see Lakens, 2021), and therefore be correctly classified as being non-null.

$$|ES_{Add}| > \sqrt{\left(\frac{1}{N_C}+\frac{1}{N_A}+\frac{1}{N_B}+\frac{1}{N_I}\right)\cdot\left(\frac{2\cdot(Z_{\alpha/2})^2\cdot(N_C+N_A+N_B+N_I)\cdot\left(J(m)\right)^2}{2\cdot(N_C+N_A+N_B+N_I)-(Z_{\alpha/2})^2\cdot\left(J(m)\right)^2}\right)}$$

(Inequality 3.1)

Unfortunately, the multiplicative null model does not yield such an inequality for calculating the critical effect size, due to a more complex structure in the calculation of the variance. However, we can instead estimate the relationship between the critical effect size and sample size using our simulated data. Figure 3.4 illustrates the relationship between effect sizes and samples sizes for both the additive (Figure 3.4a) and multiplicative (Figure 3.4b) null models using our simulated data. For the additive null model, there is a clear threshold (Inequality 3.1) with points falling within this threshold being assigned a null classification, with those which fall outside this threshold given a synergistic, antagonistic, or reversal classification. In contrast, for the multiplicative null model it is apparent that a similar, general, relationship exists between sample size and whether an interaction is assigned a null classification (Figure 3.4b). However, the lack of a sharp threshold of critical effect sizes for the multiplicative null model (Figure 3.4b) indicates that the relationship of this threshold to sample size is more complex than for the additive null model (Figure 3.4a; Inequality 3.1).

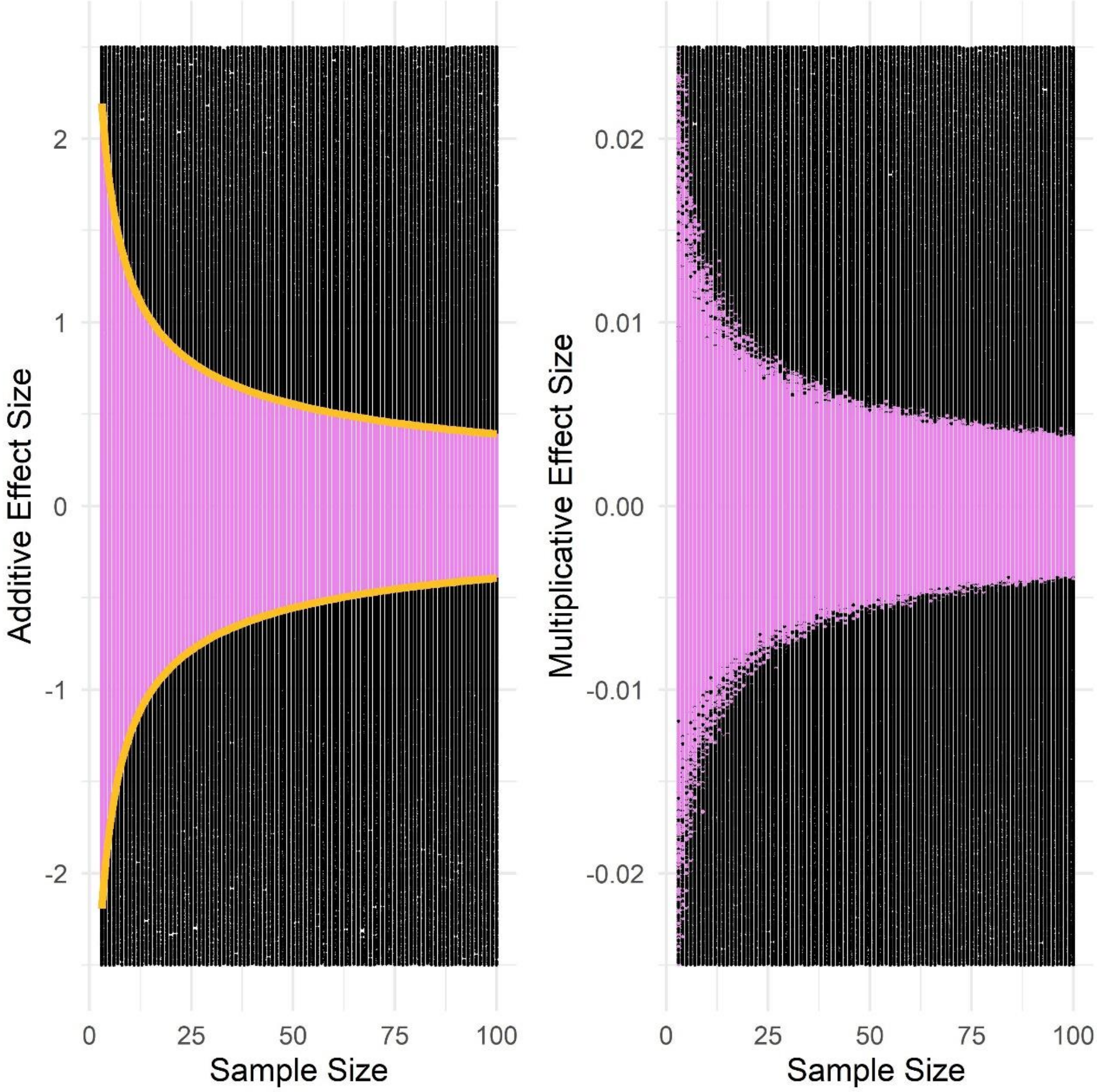

Figure 3.4: The relationship between treatment sample size and effect size for the additive (a) and multiplicative (b) null models. For each treatment sample size (3-100), points are plotted from simulations where an observation error of 0.01 was used. Pink points illustrate interactions that are classified as null (i.e., effect sizes not statistically different to zero), while black points denote interactions that are not classified as being null (i.e., effect size are statistically different to zero and a classification of antagonistic, reversal, or synergistic). The yellow line (Figure 3.4a) denotes the critical effect size threshold (Inequality 3.1) that interaction effect sizes must exceed in order for an interaction to be significantly different to zero (and hence classified as being non-null) under the additive null model. As in the methods, 95% confidence intervals were calculated (i.e., for Inequality 3.1, $Z_{\alpha/2} = 1.96$).

3.4 Discussion

While differences between null models have been discussed, in the context of multiple stressors, for many years (Folt et al., 1999; Soluk & Collins, 1988), we still lack a strong understanding of how they relate to one another or how robust they are to variation in sample measurements, or sample size. Here, we help rectify this by comparing the widely implemented additive and multiplicative null models and their abilities to infer the correct stressor interaction types in the face of multiple sources of uncertainty. Overall, our results illustrate the following points. Firstly, despite apparent similarities in interaction classifications, there are considerable differences between the additive and multiplicative null models (including both their results and the hypotheses they test), and as such it is difficult to draw direct comparisons between them. Secondly, both models are very sensitive to uncertainty in the estimation of the stressor effects, and it can be very hard to reject the null interaction even with very small measurement errors. This may mean many non-null interactions (especially synergistic and antagonistic interactions) may have been missed in previous experiments, though the null models appear to be robust in their ability to detect reversal interactions regardless of treatment uncertainty. Thirdly, our results show the key role that sample size plays in determining whether an interaction is significantly different to either null model. Indeed, it is clear that larger sample sizes allow smaller significant effects to be observed. However, for sample sizes most commonly implemented within ecological experiments it is evident that only those interactions with uncommonly large effects will ever be classified as being different to the null model. Overall, we conclude that from an analytical perspective there is not an obviously *better* null model, and that null model choice can therefore be based upon biological arguments without undue concerns over loss of statistical power.

Empirical implications

Given the threat that multiple stressors pose to natural systems (Lindenmayer et al., 2020), there is a pressing need to consider how stressors interact when implementing conservation initiatives (Didham et al., 2007), as this can help inform decisions on when, where and how to act (Brown et al., 2013; Côté et al., 2016). However, it is equally (if not more) important to accurately identify the classification of any interaction that is occurring. Incorrectly assuming how stressors interact can lead to conservation actions either being ineffective or, in some situations, detrimental to the systems that they aim to preserve (Brown et al., 2013). Our results imply that synergistic, antagonistic, and reversal interactions may be frequently missed due to even low levels of uncertainty or small sample sizes, with this having consequences for management strategies. Ultimately, treatment means, uncertainty, and sample sizes all impact the calculation of effect sizes, variances (and hence the classification of interactions) in differing ways. Accordingly, an understanding of these relationships can both aid the interpretation of results and allow experimental approaches to be adjusted to ensure they are capable of answering the study question.

As shown by the equations for both the additive (Equations 3.1.1 – 3.1.3) and multiplicative (Equation 3.2.1) null models, treatment means are a crucial component in the calculation of effect sizes, and hence the classifying of interactions. However, the two null models we have focussed on are clearly different, hence it should be no surprise that they lead to different interaction classifications. This difference in the results of null models is well-documented within fields such as pharmacology (Tang et al., 2015), where the class of an observed interaction may be prefixed by the null model implemented (e.g., Loewe antagonism) (Greco et al., 1992). However, Tang et al., (2015) note the difficulties that arise when interactions are classified differently across null models, a phenomena we similarly show with the frequency of null model dependent interactions. Currently, for multiple stressor ecology, there is no equivalent framework for interpreting, or reporting, results. This discrepancy should be remedied, given that the ambiguous usage of terms (such as synergistic or antagonistic interactions) is potentially hindering interpretation of multiple stressor interactions, both within and across disciplines (Dey & Koops, 2021; Orr et al., 2020). Additionally, given the contrasting nature of these null models, we have derived algebraic inequalities (Inequalities S1, S2) that can explain differences in interaction classifications as a function of treatment means. These inequalities provide useful insights regarding the values of treatment means under which interactions are classified differently between null models, although these are only for the scenario of no treatment uncertainty. Ultimately, we echo previous calls (Schäfer & Piggott, 2018) to thoroughly consider the choice of null model to be implemented within a study, especially given that it is difficult to compare studies which implement different null models (see *Null Model Selection*).

Treatment uncertainty (in the form of treatment standard deviations) impacts the calculation of additive effect sizes (Equations 3.1.1 – 3.1.3) and variances (Equation 3.1.4) as well as the variance of multiplicative effect sizes (Equation 3.2.2). As shown by Figure 3.2, under ecologically plausible levels of observation error, we would expect the majority of interactions to be assigned a null interaction classification (i.e., we are unable to reject the null model). Indeed, the dominance of null interactions has previously been observed in several meta-analyses (e.g., Burgess et al., 2021; Lange et al., 2018; Stephens et al., 2013), illustrating that the results we present here are likely consistent across theoretical and empirical analyses. It is important to note that while uncertainty may arise due to technical variation (e.g., observation error), forms of biological variation such as heterogenous environmental factors or genetic variation may also contribute to any uncertainty (Higdon, 2013).

While treatment means and uncertainty are unknowns that an experimenter is trying to estimate, sample size is the one factor that can, usually, be controlled and built into the experimental design. This arguably means that sample size is the most important of the three components to consider when attempting to increase the statistical power of any experiment. As our results show, increases in statistical power can be made by increasing the number of experimental replicates, but, within our analyses, the improvements are not dramatic (Figure 3.3). These results corroborate earlier work that investigated the effect of sample size on models to detect interactions in epidemiological studies and found that statistical power can be very low under commonly encountered situations, and that the power converged very slowly with sample size (Greenland, 1983).

Here, we provide a closed form solution (Inequality 3.1) that gives the critical effect size value required to be able to correctly reject the additive null model for a given sample size. As verified by our simulations, additive effect sizes with magnitudes greater than this threshold will be assigned a non-null (i.e., antagonistic, reversal, synergistic) interaction class, while any effect size with a magnitude less than the threshold will be assigned a null classification. The derivation of this additive critical effect size threshold has major implications for the design of multiple stressor experiments. Inequality 3.1 allows researchers to conduct an a-priori analysis to determine the critical (i.e., minimum) additive effect size required to reject the null hypothesis (i.e., null model) for their experimental design. Indeed, such a finding allows researchers to determine whether their experimental design will only enable them to statistically distinguish exceptionally large effects or achieve sufficient power to determine the minimum effect size a researcher wishes to detect (Lakens, 2021). As previously highlighted, the empirical experiments included in three recent meta-analyses (Burgess et al., 2021; Gomez Isaza et al., 2020; Seifert et al., 2020) predominately have eight or fewer replicates per treatment, with this meaning that only exceptionally large effects are able to be assigned a non-null interaction classification. Unfortunately, there is no closed form solution that we could find for the multiplicative null model, but our simulations suggest a similar pattern of effect size and sample size, albeit with effect sizes being on a different scale (Figure 3.4).

While some researchers may wish for a heuristic guide, or ‘rule of thumb’, regarding sample sizes, such guidance may actually be unhelpful or even hinder experiments (Lakens, 2021). Overall, the minimum sample sizes required by a researcher vary with the question asked or hypothesis investigated. For instance, it might be important to measure even a statistically small deviation from a null model; hence a high statistical power is required. Alternatively, it might only be important to detect when large deviations from a null model occur; here, a lower statistical power will suffice. In both scenarios, decisions are made regarding the minimum effect size of ecological importance, which dictates the scope of any experiment. Resource limitations aside, sample size is an attribute which researchers have explicit control over; therefore, it is important that researchers determine, a-priori, whether their experimental design is capable of addressing their research question (Lakens, 2021; Mudge et al., 2012). However, there may be instances where an experiment has low statistical power, sample sizes cannot be increased, but an insight into the effects of interacting stressors is still required. In such a situation, statistical power may be increased by raising the level of significance (i.e., >0.05), increasing the likelihood of the null hypothesis (i.e., null model) being rejected, though the likelihood of Type-I errors will similarly be increased. As shown by Figure S3.2 (Supplementary Material), increasing the significance level results in a greater frequency of interactions being assigned their ‘true’ interaction classification for both null models, across all levels of observation error and number of replicates. Crucially, the altered level of significance must be justified, and not arbitrarily chosen (Lakens, 2021). As noted by Lakens

(2021), such justification may be provided by either a cost-benefit analysis (see Field et al., 2004) or a compromise power analysis (see Lakens, 2021).

Null model selection

Ultimately, the choice of null model is at the discretion of the individual researcher (Tang et al., 2015), but knowledge on statistical properties can be an important criterion in deciding which null model to use. Given our results indicate a similar level of sensitivity to both observation error and sample size we do not argue that either the additive or multiplicative null model is obviously *better* than the other, meaning the user can select a null model based upon biologically apposite criteria (see Schäfer & Piggott, 2018).

While null models are capable of providing insights into the combined effects of interacting stressors, it has been suggested that the underlying assumptions of these statistical approaches limit their usefulness from an ecological perspective (De Laender, 2018; Schäfer & Piggott, 2018). One of the greatest criticisms of the additive null model is that it is capable of exceeding a feasible range of effects such that, for example, survival probability could drop below 0 or rise above 1 (Folt et al., 1999). This is particularly pertinent when considering responses that are bounded within fixed limits (e.g., mortality) (Fournier et al., 2006); as such, the use of the multiplicative null model in this scenario can negate the issue of unfeasible expectations (Darling & Côté, 2008; Fournier et al., 2006). However, rather than simply addressing a weakness of the additive null model, the multiplicative null model is explicitly useful in its own right. As noted by Kerkhoff & Enquist (2009), many biological processes (e.g., growth and metabolism), are inherently multiplicative; hence, the application of the multiplicative null model is perhaps a more ecologically appropriate approach.

Questions remain as to whether an alternative null model would be more appropriate in any given situation, as it may better reflect the varied mechanisms through which stressors induce effects. Of the alternatives, the most often discussed is the dominance null model, where the predicted effect of multiple simultaneous stressors is simply assumed to be equal to that of the individual stressor with the greatest effect (Folt et al., 1999). Here, the dominance null model may reflect a protective mechanism, namely cross-tolerance, where the exposure to one stressor results in reduced susceptibility to an additional stressor (Todgham & Stillman, 2013; Velasco et al., 2019). Furthermore, reviews (Orr et al., 2020; Schäfer & Piggott, 2018) have suggested that future research should prioritise null models capable of providing mechanistic insights into multiple stressor interactions, for instance by considering the many physiological processes which stressors may affect (see Sokolova, 2021). Accordingly, the allied fields of toxicology and ecotoxicology may potentially provide methods that may be adopted in part (Schäfer & Piggott, 2018). Following this, Tekin et al., (2020) illustrated how an adapted toxicological model (Bliss Independence Model; see Bliss, 1939) could be applied to multiple stressors in an ecological context. However, it is important to consider that while other methods may be more established in their respective fields (e.g., Tekin et al., 2020; see Bliss, 1939; Segre et al., 2005; Yeh et al., 2006), within multiple stressor ecology they perhaps represent a departure from the more widely implemented null models.

Finally, understanding how the null models relate to one another is important for drawing conclusions based upon results across multiple studies, and also perhaps for null model choice, but both of these depend upon an understanding of their statistical power. For example, it is clear that the Bliss Independence Model generates different predictions to other null models (Tekin et al., 2020), but are the results more robust to changes in sample size and measurement error? Two models could report a similar type of interaction for different reasons (e.g., a lack of data, or because it is the true interaction), but in most cases we do not have any understanding of the data requirements; this represents a fundamental issue which we hope future work will address.

Limitations

Broadly, our results can be grouped into two categories; those that provide general insights, and those that are dataset specific. The results based upon algebraic relationships (e.g., Inequality 3.1; Figure 3.4) are general and can be applied to any dataset, whereas results that depend upon data simulation might be dependent upon the distribution of the effect sizes generated by the underlying food chain model. Whilst the food chain model used here has been shown to generate ecologically plausible results (Burgess et al., 2021), the relationships shown here may differ with other datasets (for instance, where a response other than density is considered) which contain differing distributions of effect sizes. Ultimately, we anticipate that many trends will remain across different datasets (e.g., the dominance of null interactions at commonly implemented sample sizes), though others (e.g., the frequency of antagonistic, reversal, and synergistic interactions in the absence of uncertainty) may vary between datasets.

Conclusions

‘Will my study answer my research question?’ is the most fundamental question a researcher can ask when designing a study (Johnson et al., 2015), and its answer requires an understanding of the statistical power of the tools at hand. Despite a large amount of effort going into determining the ways ecosystem stressors interact, and the development of a number of alternative models, there has been little attempt to quantify the statistical power of the null models used to classify the interactions. Here we have shown that the commonly used additive and multiplicative null models are both very sensitive to measurement error in the estimated effects of the stressors, and that it is difficult to reject the null model interaction for even large sample sizes. This means there is likely a large number of stressor interactions that are truly synergistic or antagonistic, but which are being incorrectly assigned, and therefore remain hidden from ecologists. This has deep implications for management strategies that might be based on these results, and it also makes it difficult to draw strong conclusions from reviews of the experimental literature. There have been calls (Orr et al., 2020; Schäfer & Piggott, 2018) to move to a more mechanistic understanding of stressor interactions. Accordingly, this may mean a shift in the statistical tools that are implemented; however, there is a pressing need to understand both the benefits, data requirements, and limitations of any analytical approach so that the maximum ecological insight can be gained.

Chapter Four – Methodological variation obscures generalities across multiple stressor meta-analyses

4.1 Abstract

Multiple co-occurring stressors affect the majority of freshwater and marine ecosystems. However, how these stressors interact to affect ecosystems is not well understood; with aquatic meta-analyses revealing few (if any) consistent generalities. Some meta-analyses conclude that stressors interact synergistically (greater effect than expected) whilst others suggest that multiple stressors combine antagonistically (lesser effect than expected). To date, there are no proven explanations for the differences between these meta-analyses. As such, we determine whether methodological variation (a currently overlooked facet) is responsible for the disparate conclusions between multiple stressor meta-analyses. We collate datasets from seven aquatic multiple stressor meta-analyses, which are then analysed using a consistent multiplicative null model framework. Our results illustrate that once methodological variation is removed, consistent generalities are apparent across all seven datasets, with antagonisms dominating and a distinct absence of synergisms. Furthermore, by aggregating these individual datasets, we collate the single largest multiple stressor meta-analytical dataset. Analysis of this aggregated dataset highlights that response metric is an ecologically important covariable in determining stressor interactions; with multiple stressors affecting density in a more strongly antagonistic manner than either biomass or survival. Similarly, multiple stressor interactions are shown to significantly differ between feeding groups for the response metric of density, but not biomass. Additionally, we do not observe any differences in interactions where the combined effect of stressors is predicted to have a positive or negative effect. Our results highlight that methodological variation is likely preventing multiple stressor meta-analyses from reaching consistent generalities, indicating that researchers must consider the methods used in future analyses. Furthermore, ecological variation is likewise undoubtedly important to consider, with our results also highlighting where data deficiencies currently exist in multiple stressor research. Finally, our results provide important insights for conservation practices which seek to remediate the effects of interacting stressors in aquatic ecosystems.

4.2 Introduction

Whilst providing numerous ecosystem services (Sala et al., 2021; Smale et al., 2019; Sterner et al., 2020), such as the provision of food resources (Bernhardt & O'Connor, 2021; Costello et al., 2020), aquatic ecosystems are threatened by a multiplicity of stressors (Barbarossa et al., 2021; O'Hara et al., 2021; Su et al., 2021). Climate change (Trisos et al., 2020), habitat alteration (Best, 2019), invasive species (Diagne et al., 2021), and pollution (Johansen et al., 2017) are examples of well-documented stressors which are capable of impacting individual organisms (Sokolova, 2021) through to entire ecosystems (Marshall et al., 2017), as well as their functions (Arrigo et al., 2020) and services (Andrello et al., 2017). However, novel stressors are also continuing to emerge (Reid et al., 2019).

Despite a focus on single stressor experiments (Jackson et al., 2021), freshwater and marine ecosystems are frequently subjected to multiple co-occurring, and potentially interacting, stressors (EEA, 2018; Halpern et al., 2015; Ormerod et al., 2010). However, predicting how stressors combine to affect these ecosystems is challenging (Hodgson & Halpern, 2019), with the combined effect of stressors often based upon the effects of individual stressors. Accordingly, individual stressor effects are estimated by measuring one or more biological responses (e.g., population densities or individual growth rates), and comparing this to a control scenario without the stressor present (e.g., Griffith et al., 2012). Null models, such as the additive or multiplicative null models (see Folt et al., 1999; Sih et al., 1998; Soluk & Collins, 1988), are often then used to determine an expected combined effect of stressors based upon single stressor impacts (Burgess et al., 2021). However, the observed effect of interacting stressors is capable of deviating from the null model expectation (sometimes referred to as an 'ecological surprise'; sensu Paine et al., 1998) leading to pairs of stressors interacting in one of four different ways (Figure 4.1). Specifically, the observed effect of interacting stressors can (i) exceed that predicted by the null model (synergistic interaction); (ii) equal the predicted effect (null interaction); or be less than predicted, with either (iii) combined effects having the same polarity as the individual effects (antagonistic interaction), or (iv) the or the combined effects having different polarities to the individual effects (reversal interaction). Determining how stressors interact represents one of the grand challenges facing ecologists (Fleishman et al., 2011; Hodgson & Halpern, 2019). However, addressing this grand challenge is potentially complicated by stressors interacting in contrasting ways across different response metrics. Indeed, it has been shown that, within a theoretical freshwater system, the same combination of stressors can result in additive interactions for body size and population biomass, synergistic interactions for abundance and resource processing, and antagonistic interactions for reproduction rates (Galic et al., 2018).

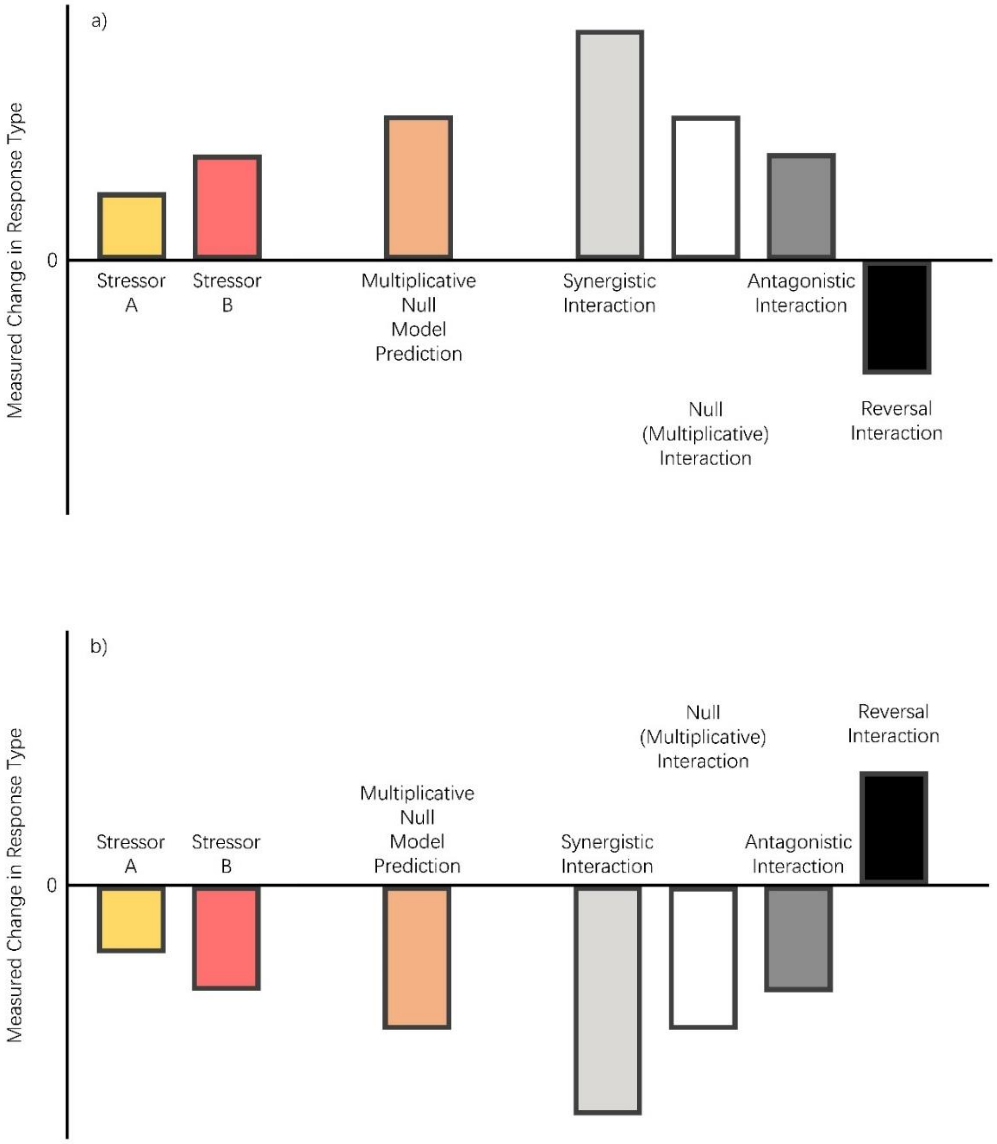

Figure 4.1: Graphical illustration of interaction types for a given response metric. Yellow and red bars denote the change in a response metric as a consequence of individual stressors, while the expected interactive effect of these stressors (under a multiplicative null model) is shown by the orange bar. Compared to this multiplicative expectation, examples of synergistic (light grey), null (white), antagonistic (dark grey), and reversal (black) interactions are shown. Differences in directionality of stressor effects are shown between Figure 4.1a (positive) and Figure 4.1b (negative).

Whilst ecosystems across realms are affected by co-occurring stressors (Vinebrooke et al., 2004), there is growing evidence that aquatic ecosystems are particularly at risk to the effects of interacting stressors (e.g., Crain et al., 2008; Jackson et al., 2016). Indeed, freshwater species have experienced an average decline in population abundances of 84% since 1970 (WWF, 2020), with these declines greater than those observed for terrestrial species (Dudgeon, 2019). Such declines are severe given that rivers, reservoirs, and lakes cover 2.3% of the Earth's surface (Lehner & Döll, 2004) yet freshwater ecosystems contain 9.5% of the animal species found on Earth (Balian et al., 2008). Furthermore, 32% of all freshwater species are threatened with extinction, compared to the 24% of terrestrial species under threat (Collen et al., 2014). Similarly, multiple interacting stressors have been identified as an intensifying threat which endangers freshwater ecosystems (Reid et al., 2019). Likewise, despite covering a much greater surface area than freshwater environments, marine ecosystems are similarly at risk from multiple stressors (Halpern et al., 2019); indeed, between 1970 and 2012, marine species have undergone average declines in population abundances of 36% (WWF, 2016). Only a small minority of the ocean (2.3%) is not affected by multiple stressors (Halpern et al., 2015), with at-risk species facing negative effects of stressors across an average of 57% of their range (O'Hara et al., 2021). Furthermore, the negative impacts of stressors are intensifying across many at-risk species' ranges (O'Hara et al., 2021) while some marine environments are subjected to over fifteen co-occurring stressors (Côté et al., 2016).

Given the sheer number of stressors and their interactions, the plethora of potentially impacted species and communities, and the myriad of potential response metrics, testing every potentially important combination of anthropogenic and ecological variables in aquatic ecosystems is an unfeasible proposition (Côté et al., 2016). Nevertheless, empirical studies investigating multiple stressor interactions can provide important insights into whether conservation measures will be effective for a given ecosystem (Brown et al., 2013). However, whether the results of these empirical studies are applicable to different ecosystems or alternative combinations of stressors is not necessarily clear. As such, an alternative approach to understand multiple stressor interactions is to search for generalities which are applicable to broad categories of stressors, ecosystems, or response metrics (Côté et al., 2016). To date, meta-analyses have been the primary tool used to search for these generalities (Côté et al., 2016). Such meta-analyses vary in the stressor combinations (e.g., Bancroft et al., 2008; Gomez Isaza et al., 2020), realms (e.g., Jackson et al., 2016; Wu et al., 2011), taxonomic groups (e.g., Egea-Serrano et al., 2012; Seifert et al., 2020), or response metrics (e.g., Burgess et al., 2021; Darling & Côté, 2008) that they consider, with all collating data from multiple empirical studies.

However, while it may be anticipated that multiple meta-analyses will often reach the same conclusions, this has not occurred to date. As shown by Table 4.1, several meta-analyses for aquatic ecosystems report overall synergistic effects (e.g., Bancroft et al., 2008; Crain et al., 2008), whilst others report overall antagonistic effects (e.g., Burgess et al., 2021; Jackson et al., 2016). Similar patterns are also observed for the frequency of individual interaction classifications, with some meta-analyses reporting that null interactions dominate (e.g., Burgess et al., 2021; Lange et al., 2018), whilst others report that antagonisms (e.g., Jackson et al., 2016) or synergisms (e.g., Przeslawski et al., 2015) are the most prevalent. Alternatively, other meta-analyses report approximately equal frequencies of interaction classifications (e.g., Crain et al., 2008; Darling & Côté 2008; Gomez Isaza et al., 2020). Ultimately, the differences between meta-analyses are not currently explainable by a consistent ecological criterion; as such, few generalities have emerged from multiple stressor meta-analyses (Côté et al., 2016). For instance, no congruous patterns have been observed for meta-analyses conducted on marine data, though several freshwater meta-analyses have reported antagonistic summary interaction classifications (Table 4.1).

*Table 4.1: Comparisons of meta-data, methods, and main results for selected freshwater (light blue), marine (dark blue), or mixed-realm (black) meta-analyses. * denotes that reversal interactions were not distinguished from antagonistic interactions.*

Meta-analysis	Study Group (Realm)	Stressor combination	Response metric	Organisation level	Feeding group	Overall Interaction Class	Individual interactions % Null | Syn. | Ant. | Rev.	Null model	Used in analysis
Bancroft et al., (2008)	Amphibians (Freshwater)	UV-B and second stressor	Survival	Population	Heterotroph	Synergistic	Not reported	Multi.	No
Crain et al., (2008)	Multiple (Marine)	Multiple	Density, growth, survival, etc.	Community, population, organism	Autotroph, heterotroph	Synergistic	26 | 36 | 38 | *	Add.	Yes
Darling & Côté, (2008)	Multiple (Multiple)	Multiple	Survival	Population	Heterotroph	Not reported	23 | 35 | 42 | *	Add.	No
Harvey et al., (2013)	Multiple (Marine)	Temperature & CO_2	Calcification, growth, survival, etc.	Community, population	Autotroph, heterotroph	Not reported	Not reported	Multi.	Yes
Przeslawski et al., (2015)	Embryo / Larvae (Marine)	Multiple	Development, growth, survival, etc.	Population, organism	Autotroph, heterotroph	Not reported	17 | 65 | 17 | *	Add.	Yes
Jackson et al., (2016)	Multiple (Freshwater)	Multiple	Biomass, density, survival, etc.	Community, population, organism	Autotroph, heterotroph	Antagonistic	17 | 28 | 41 | 15	Add.	Yes
Lange et al., (2018)	Fishes (Freshwater)	Multiple	Biomass, growth, survival, etc.	Population, organism	Heterotroph	Antagonistic	80 | 3 | 17 | *	Add.	Yes
Gomez Isaza et al., (2020)	Multiple (Freshwater)	Nitrate and second stressor	Survival	Population	Heterotroph	Not reported	33 | 25 | 42 | *	Multi.	Yes
Seifert et al., (2020)	Phytoplankton (Marine)	Multiple	Growth	Population	Autotroph	Not reported	Not reported	Add.	No
Stockbridge et al., (2020)	Seagrasses (Marine)	Multiple	Biomass, mortality, survival, etc.	Population	Autotroph	Not reported	57 | 36 | 6 | *	Add.	No
Burgess et al., (2021)	Multiple (Freshwater)	Multiple	Biomass, density	Community, population	Autotroph, heterotroph	Antagonistic	82 | 2 | 10 | 5	Add.	Yes

These disparate results may be attributed to variation in meta-analyses that is based upon either (i) data or (ii) methodology. Data-based variation includes meta-analyses analysing contrasting ecological data (e.g., differing response metrics, taxonomic groups, or stressor combinations) and is often the most commonly considered source of variation between meta-analyses. However, methodological variation between meta-analyses represents a previously overlooked facet. Some methodological differences are easily apparent, such as whether the meta-analysis employs an additive or multiplicative null model (see Table 4.1); with these null models having previously been shown to differ from one another in their underlying assumptions and results (Chapter Three). Similarly, alternative frameworks (e.g., Jackson et al., 2016; Piggott et al., 2015) are sometimes used to classify interactions. As such, despite being oft compared, it should not be expected that meta-analyses employing different null models, or interaction classification frameworks, should return the same results.

A final less evident source of methodological variation is that meta-analyses may employ differing versions of the same null model. For instance, Darling & Côté (2008), Hillebrand (2002), Siviter et al. (2021) and Yue et al. (2017) all employ an additive null model; however, each meta-analysis calculates the additive null model using different equations. For instance, Hillebrand (2002), Siviter et al. (2021), and Yue et al. (2017), all use variants of the additive null model outlined by Gurevitch et al. (2000) (*Hedges' d*); whilst Darling & Côté (2008), implement an entirely different effect size metric. Similarly, contrasting versions of the multiplicative null model are also utilised in different meta-analyses (e.g., Darling & Côté 2008; Dey & Koops, 2021; Hawkes & Sullivan, 2001; Rosenblatt & Schmitz, 2014). Accordingly, the extent to which the methodological variation between meta-analyses impacts the results and conclusions of these studies is currently unknown. Here, we attempt to determine whether methodological variation is one of the primary drivers of these differences by testing the below hypothesis.

Hypothesis one: Elimination of methodological variation results in minimal differences in both interaction classification frequencies and summary effect sizes between meta-analytical datasets.

In order to test this hypothesis, seven different meta-analyses, all considering the impacts of multiple stressors in aquatic systems (see Table 4.1), were collated and individually analysed using the same consistent framework utilising the multiplicative null model. Based upon a previous analysis (Hungate et al., 2009), it was our expectation that methodological variation would be responsible for some, but not all, of the differences between meta-analyses.

Additionally, the aggregation of these datasets provides an opportunity to analyse the single largest meta-analytical dataset on the effects of multiple stressors on aquatic systems. Accordingly, this large-scale dataset was then used to test three simple hypotheses, based on ecological rationale, regarding differences in interacting stressor effects between response metrics (Hypothesis two), feeding groups (Hypothesis three) and the expected direction of effects (Hypothesis four).

Stressors are capable of affecting multiple different response metrics (e.g., density, growth, or survival) (Kroeker et al., 2013); however, given that stressors affect disparate physiological pathways, response metrics can be impacted in contrasting ways (Segner et al., 2014). For example, stressors may swiftly affect physiological responses, though response metrics such as growth may only be substantially impacted by chronic stressors (Lange et al., 2018). Similarly, compensatory dynamics are also capable of moderating stressor effects (Jackson et al., 2016). For instance, at the community level, a reduction in the biomass of one species may be offset by an increase in the biomass of another (Ives & Cardinale, 2004); however, such compensatory dynamics may not necessarily occur where response metrics are measured at lower levels of ecological organisation (e.g., survival). Whilst these processes all modify the effects of single stressors, they likewise influence how stressor interactions affect different response metrics. Indeed, stressor interactions have previously been shown to vary between physiological and survival metrics in freshwater fishes (Lange et al., 2018).

Accordingly, multiple stressor interactions are potentially capable of varying across response metrics, with such differences being either contrasting interaction classifications (e.g., antagonisms or synergisms) or differing interaction strengths between response metrics. Based upon the above ecological processes, our expectation is that multiple stressor interactions will vary across response metrics; however, in line with recent analyses (Chapter Three) we anticipate that these interactions are unlikely to be synergistic. As such, the following hypothesis was tested.

Hypothesis two: The effects of multiple stressors significantly differ across response metrics.

Multiple covariables have been purported as being important in determining stressor interactions, though few have been found to be consistently important across meta-analyses (Côté et al., 2016). However, feeding group (e.g., autotrophs or heterotrophs) is one such covariate which is often reported as moderating stressor interactions; although the exact nature of this relationship may vary (e.g., Crain et al., 2008; Jackson et al., 2016). While differences between feeding groups are readily apparent, such physiological disparities between autotrophs and heterotrophs may result in differing biological mechanisms being affected by stressors (Bracewell et al., 2019). Given differences in physiology and stressor modes of action, it is likely that multiple stressors will affect autotrophs and heterotrophs in a divergent manner. Indeed, as before, such effects could be differences in either observed interaction classifications, or the relative strength of any interaction.

To date, meta-analyses have primarily investigated the role of feeding group in isolation, without consideration of other potentially important covariates. Instead, we investigate whether interactions differ between feeding groups across response traits. To our knowledge, such an analysis has not previously been conducted. However, given that stressors can affect differing physiological processes (Bracewell et al., 2019), our expectation is that stressor interactions will differ between feeding groups for a given response metric. Accordingly, we test the following hypothesis.

Hypothesis three: For a given response metric, the effect of multiple stressors significantly differs between autotrophic and heterotrophic feeding groups.

Despite some differences in nomenclature (Orr et al., 2020), it is widely known that stressors are capable of inducing either positive, or negative, effects upon an ecosystem (Kroeker et al., 2017). Indeed, a stressor which has a negative effect on one species may have a directly, or indirectly, positive effect on another species (Piggott et al., 2015). As such, the combined effect of two stressors can also have either a positive or negative impact, with it possible to calculate an expectation for the combined effect of two interacting stressors based upon their individual effects (see Section 4.3).

However, to our knowledge, there has been limited research to determine whether the effect of interacting stressors differs depending upon whether a positive or negative combined effect is expected. Should we expect interactions, where a positive combined effect is predicted, to be synergistic? Or should we anticipate interactions, where a negative combined effect is predicted, to be antagonistic? To date, there is no guidance as to whether we should expect interactions to vary with the predicted direction of combined stressor effects.

Our naïve expectation was that interactions would not vary with the predicted direction of combined effects. Instead, we anticipate that response metric and feeding group may instead be more important in determining stressor interactions; though the lack of previous experiments or theoretical underpinnings provide no support, or counterbalance, to our expectation:

Hypothesis four: For a given response metric, the effects of multiple stressors will not significantly differ between stressor interactions which have an expected negative, or positive, effect.

4.3 Methods

Data collation

Seven multiple stressor meta-analyses were selected, with three and four datasets from the marine (Crain et al., 2008; Harvey et al., 2013; Przeslawski et al., 2015) and freshwater realms (Burgess et al., 2021; Gomez Isaza et al., 2020; Jackson et al., 2016; Lange et al., 2018) respectively. Each of these meta-analyses consider different combinations of covariates (Table 4.1), such as stressor combination and response metric. If the datasets for each meta-analysis were not publicly available, the corresponding author for each study was contacted. An additional marine dataset on phytoplankton growth (Seifert et al., 2020) was initially considered for inclusion within this analysis. However, given the logarithmic transformation applied to this data, and the corresponding challenges in inferring null model results from logarithmically transformed data (Griffen et al., 2016), the inclusion of this dataset was decided against.

For hypothesis one, each of the seven meta-analytical datasets were kept as close to those used in the original analysis for each meta-analysis as possible. However, the data for each interaction was compared to original source for the data with any numerical issues that were identified being rectified (see Table 4.2).

Table 4.2: Examples of the data quality issues encountered within the datasets, alongside their impacts upon results and remedies.

Issue	Impact	Remedy
Numeric values for means, standard deviations, or sample sizes are missing.	Effect sizes, variances, and confidence intervals cannot be calculated. The data will not be able to be analysed.	Determine whether the missing data can be reextracted from the original data source.
Reported standard deviations actually represent other error metrics (e.g. standard error or 95% confidence intervals).	Effect size variances and confidence intervals are incorrect. Potential to alter the reported interaction classification and impact the summary effect size.	Determine whether standard deviations can be reextracted from the original data source.
Reported means actually represent other metrics (e.g. medians, or model coefficients).	Effect sizes, variances, and confidence intervals are incorrect. The reported interaction classification may differ, the summary effect size will be impacted.	Determine whether means can be reextracted from the original data source.
Controls, single stressor, and combined stressor treatments are provided in the wrong order.	Effect sizes are incorrect. Potential to alter the reported interaction classification and impact the summary effect size.	Compare the data for each treatment to the original data source.
Technical, as opposed to biological, replicates are reported (i.e., pseudo-replication is occurring).	Effect size variances and confidence intervals are incorrect. Potential to alter the reported interaction classification and impact the summary effect size.	Compare the number of replicates for each treatment to the original data source. Methods should also be checked to ensure that experimental replication occurs.
Numeric values are incorrect.	Effect sizes, variances, and confidence intervals are incorrect. The reported interaction classification may differ, the summary effect size will be impacted.	Compare the numeric values for each treatment to the original data source.
Reported data has been transformed (e.g., logarithmic transformation).	Effect sizes, variances, and confidence intervals are incorrect. The null model assumptions are not met.	The data may be able to be used if the transformation can be reversed.

Data aggregation

So as to address hypotheses two - four, the seven individual meta-analytical datasets were combined into a single dataset (hence referred to as the aggregated dataset). While the aggregation of multiple meta-analyses has previously been conducted for ecological systems (e.g., Hillebrand et al., 2020) such an approach has not previously been conducted for multiple stressor meta-analyses.

Across each of the seven datasets, multiple categorical variables were collated, namely: realm, response metric, stressor combination, feeding group, and organisation level. As different meta-analyses may classify categorical variables in contrasting ways, a single consistent framework was used to classify these variables which involved consulting the original source of the data (see Appendix Four). Despite data for multiple categorical variables being collated, these variables were not subsequently analysed. This is due to the analysis being limited to a few, core, hypotheses as recommended to reduce any potential issues with multiplicity (Bender et al., 2008; O'Dea et al., 2021).

Overall, once duplicated interactions were removed, the aggregated dataset comprised 821 interactions, accounting for twelve different response metrics. For hypotheses two – four, our analyses solely focussed on the response metrics of biomass, density, and survival. Overall, these three response metrics account for the majority (76.0%) of all interactions (see Results; Figure 4.3).

The analysis for hypothesis one uses the meta-analytical datasets as close to their original forms as possible. However, for hypotheses two - four, interactions were additionally excluded if the following conditions occurred. Firstly, where non-independence of data occurred (i.e., multiple metrics were reported for the same study group), data was excluded so that density data was prioritised over biomass data, which was in turn prioritised over survival data. Secondly, data was excluded if the stressors were not external pressures (e.g., interactions where sampling time or stocking density was a stressor were excluded). Finally, data was excluded if the means and standard deviations were not reported in a standardised manner. Survival data was required to be expressed as proportion survival, whilst biomass and density data were required to be untransformed data (e.g., individuals l^{-1} or mg m^{-2}).

For hypotheses three and four, the aggregated dataset was then subset, according to each of the response metrics, with these subsets hence termed the aggregated biomass, aggregated density, and aggregated survival datasets.

Multiplicative null model

The multiplicative null model was selected for this analysis, as it is thought to be more ecologically realistic than the additive null model (Schäfer & Piggott, 2018), and has been implemented in previous multiple stressor meta-analyses (e.g., Harvey et al., 2013, Rosenblatt & Schmitz, 2014). Similarly, the multiplicative null model has a corresponding effect size (i.e., factorial response ratio) which is easily understandable, statistically rigorous, and mathematically detailed (see Lajeunesse, 2011). Here, the multiplicative null model represents the factorial form of the response ratio, comparing the observed effect of multiple stressors against that which is expected based upon the effects of individual stressors. The calculation of the multiplicative null model is based upon three variables: the mean response metric value (X_X), the standard deviation around the mean (SD_X), and the number of replicate measurements (N_X). Each of these variables are required for each of the four treatments: control conditions (*C*), stressor A only (*A*), stressor B only (*B*), and stressors A and B together (*I*).

The equation for calculating the multiplicative effect size is shown by Equation 4.1, using the notation shown above. For each effect size, corresponding variances (Equation 4.2), standard errors (Equation 4.3), and confidence intervals (Equation 4.4) are calculated. Together these values are used to classify individual interactions and calculate summary effect sizes (see below).

Prior to interactions being assigned classifications, or the calculation of summary effect sizes, directionality (see Figure 4.1) was removed from each interaction. Following a widely used method (e.g., Jackson et al., 2016; Siviter et al., 2021), where the expected effect of the individual stressors for a given interaction (Equation 4.5) was negative, the sign of the effect size for that interaction (Equation 4.1) was inverted. This allows interaction effect sizes to be comparable to one another regardless of their directionality. The observed effect of interacting stressors was likewise calculated (Equation 4.6).

4.1) $$ES = \ln(X_I) - \ln(X_A) - \ln(X_B) + \ln(X_C)$$

4.2) $$V = \frac{(SD_I)^2}{(X_I)^2 \cdot N_I} + \frac{(SD_A)^2}{(X_A)^2 \cdot N_A} + \frac{(SD_B)^2}{(X_B)^2 \cdot N_B} + \frac{(SD_C)^2}{(X_C)^2 \cdot N_C}$$

4.3) $$SE = \sqrt{V}$$

4.4) $$CI = Z_{\alpha/2} \cdot SE$$

4.5) $$Exp. = \ln(X_A) + \ln(X_B) - 2 \cdot \ln(X_C)$$

4.6) $$Obs. = \ln(X_I) - \ln(X_C)$$

Throughout this analysis, α (i.e., the level of significance) was set at 0.05, hence $Z_{\alpha/2} = 1.96$.

Classification of interactions

Using the multiplicative null model, a given interaction can be assigned one of four interaction classifications based upon its effect size and confidence intervals. By implementing a widely used classification system (e.g., Burgess et al., 2021; Jackson et al., 2016) an interaction can be assigned either an *antagonistic*, *null*, *reversal*, or *synergistic* classification (Figure 4.1).

The classification system uses the following framework to assign each individual interaction a classification.

i) If the 95% confidence intervals overlap zero (i.e., the multiplicative null model cannot be rejected), the interaction was assigned a null interaction classification.

ii) If the multiplicative null model can be rejected (i.e., 95% confidence intervals do not overlap zero) and the effect size is positive, then the interaction was assigned a synergistic classification.

iii) If the null model can be rejected, the interaction effect size is negative, and both the expected (Equation 4.5) and observed (Equation 4.6) effects have the same polarity, the interaction is assigned an antagonistic classification.

iv) If the null model can be rejected, the interaction effect size is negative, and both the expected (Equation 4.5) and observed (Equation 4.6) effects have differing polarities, the interaction is assigned a reversal classification.

Vote-counting

Once all interactions were assigned a classification, a vote-counting method was conducted to determine whether the relative proportions of the interaction classes vary between the different meta-analytical datasets (hypothesis one). For hypotheses two - four the results of this analysis are shown in Appendix Four.

Meta-analyses

In addition to the vote-counting methods, formal meta-analyses were conducted. Here, meta-analyses allow for a single summary effect size to be obtained for each of the individual meta-analytical datasets (hypothesis one), and each of the considered response metrics from the aggregated dataset (hypothesis two). Similarly, summary effect sizes can be obtained for the sub-group analyses across a single response metric (hypotheses three and four).

Meta-analytical models were implemented using the metafor package (v2.1.0) (Viechtbauer, 2010), with either random- (hypothesis one) or mixed-effect (hypotheses two - four) models being specified. Across all meta-analytical models, we accounted for between- and within-study variation using random effects. Here, random-effects were specified using a hierarchical structure where a unique identifier for the study group of organisms which was nested within a unique identifier for each individual study. Similarly, within mixed-effect models, fixed-effects were specified as a categorical covariate (i.e., response metric, feeding group, or expected direction of effects). Across all meta-analytical models, weighting of the individual effect sizes occurs with those with lower uncertainty (i.e., lower variances) being given a higher weighting within the model.

For each summary effect size, it was possible to assign a summary interaction classification, following a similar procedure to that for individual interactions (see *Classification of interactions*). While synergistic and null summary interaction classifications were assigned in the same manner as for individual interactions, it was not possible to differentiate between antagonistic and reversal summary interaction classifications. Accordingly, whenever a summary effect size was significantly negative, this was assigned an antagonistic summary interaction class. As a post-hoc analysis we implemented Wald-type χ^2 tests to investigate statistically significant differences in summary effects (hypotheses two – four) between pairs of factors of interest.

Several of the meta-analytical datasets included data for multiple intensities of one or more stressors. As such, calculations of effect sizes sometimes used the same control resulting in covariance across these effect sizes. Accordingly, covariance-variance matrices were induced within meta-analytical models as a means of accounting for any potential covariance (see Appendix Four).

Finally, for each of the meta-analytical models, model diagnostics were conducted. Heterogeneity was assessed using the I^2 statistic, with the boundaries of 25%, 50%, and 75% being suggested as indicating low, moderate, or high heterogeneity respectively (Higgins et al., 2003). Due to variation in study organisms, environments, and other variables, ecological meta-analyses often report high heterogeneity (Senior et al., 2016). Additionally, potentially influential results were assessed using $\hat{R}$, Cook's Distance, and DFBETA. Where potentially influential effect sizes were identified, the data was reassessed for errors (see Table 4.2), whilst the original source of the data was assessed to determine whether the effect sizes should be included from an ecological perspective (e.g., whether a factorial design was used, whether external stressors were implemented). Heterogeneity and model diagnostics are reported in Appendix Four.

4.4 Results

Hypothesis one: Elimination of methodological variation results in minimal differences in both interaction classification frequencies and summary effect sizes between meta-analytical datasets.

Across all seven meta-analytical datasets, null interactions are the most prevalent individual interaction class, accounting for between 58.5% and 81.2% of all interactions (Figure 4.2a). However, there is no consistent pattern in the frequency of the antagonistic, reversal, and synergistic interaction classes across the datasets.

All seven meta-analytical datasets report negative summary effect sizes (Figure 4.2b), with the effect sizes for five datasets (Burgess et al., 2021; Harvey et al., 2013; Jackson et al., 2016; Lange et al., 2018; Przeslawski et al., 2015) being assigned an antagonistic summary interaction class, while the effects sizes for the remaining datasets (Crain et al., 2008; Gomez Isaza et al., 2020) were assigned a null summary interaction class. Additionally, there is substantial variation between the different summary effect sizes.

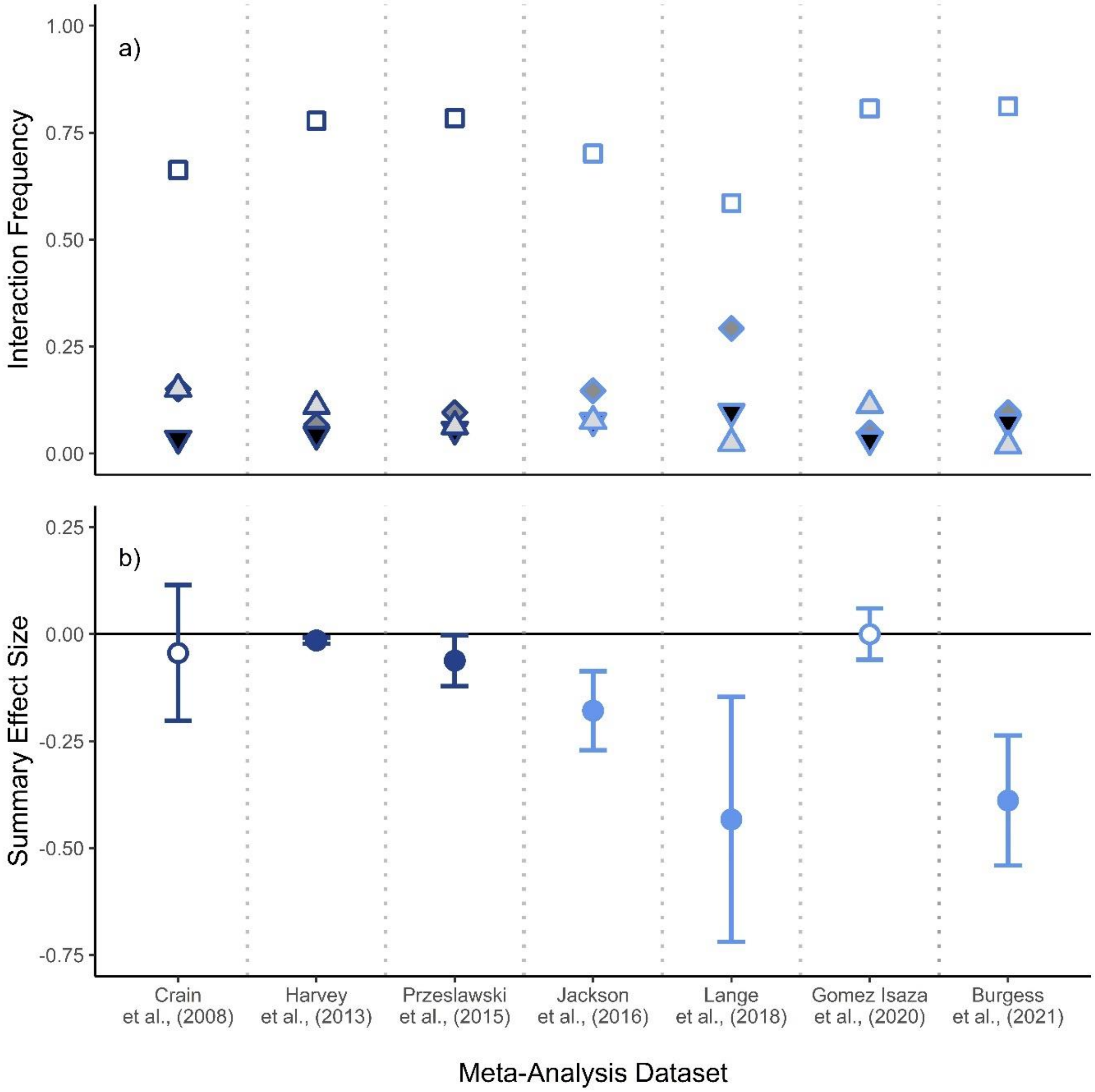

Figure 4.2: Results of the multiplicative null model across the seven meta-analytical datasets that were compiled. Figure 4.2a: the frequencies of null (white squares), synergistic (light-grey upward triangles), antagonistic (dark grey diamonds), and reversal interactions (black downward triangles) are shown. Figure 4.2b: filled points denote interactions where 95% confidence intervals do not overlap zero, hollow points denote interaction interactions where 95% confidence intervals overlap zero. Points outlined in dark blue or light blue denote datasets based upon the marine and freshwater realms respectively. Confidence intervals of the summary effect sizes are denoted by solid and dashed lines respectively.

Aggregated dataset

Once the individual datasets have been aggregated into a single dataset (with duplicate interactions being removed), it is evident that some response metrics are more frequently considered than others, with the response metrics of biomass, density, and survival accounting for 76.0% of all interactions (Figure 4.3a). However, response metrics vary in whether they are predominately measured within the marine, or freshwater realms. Similarly, it is apparent that the vast majority of interactions within this dataset have sample sizes of four, or fewer, replicates per treatment with no interaction having more than twelve replicates per treatment (Figure 4.3b).

For the aggregated biomass, density, and survival datasets, up to 24 unique stressor combinations are reported (Figure 4.3c). Similarly, for these datasets, all data is measured at higher levels of ecological organisation with the majority of biomass and density data being measured at the community level, while all survival data is measured at the population level (Figure 4.3d). Finally, it is evident that there are differences in the feeding groups, and body sizes, for organisms across the different aggregated response metric datasets (Figure 4.3e). As such, the aggregated survival dataset overwhelmingly considers heterotrophic macroorganisms, the aggregated biomass dataset predominately represents autotrophic microorganisms, whilst the aggregated density dataset is a more even combination of feeding groups and body sizes.

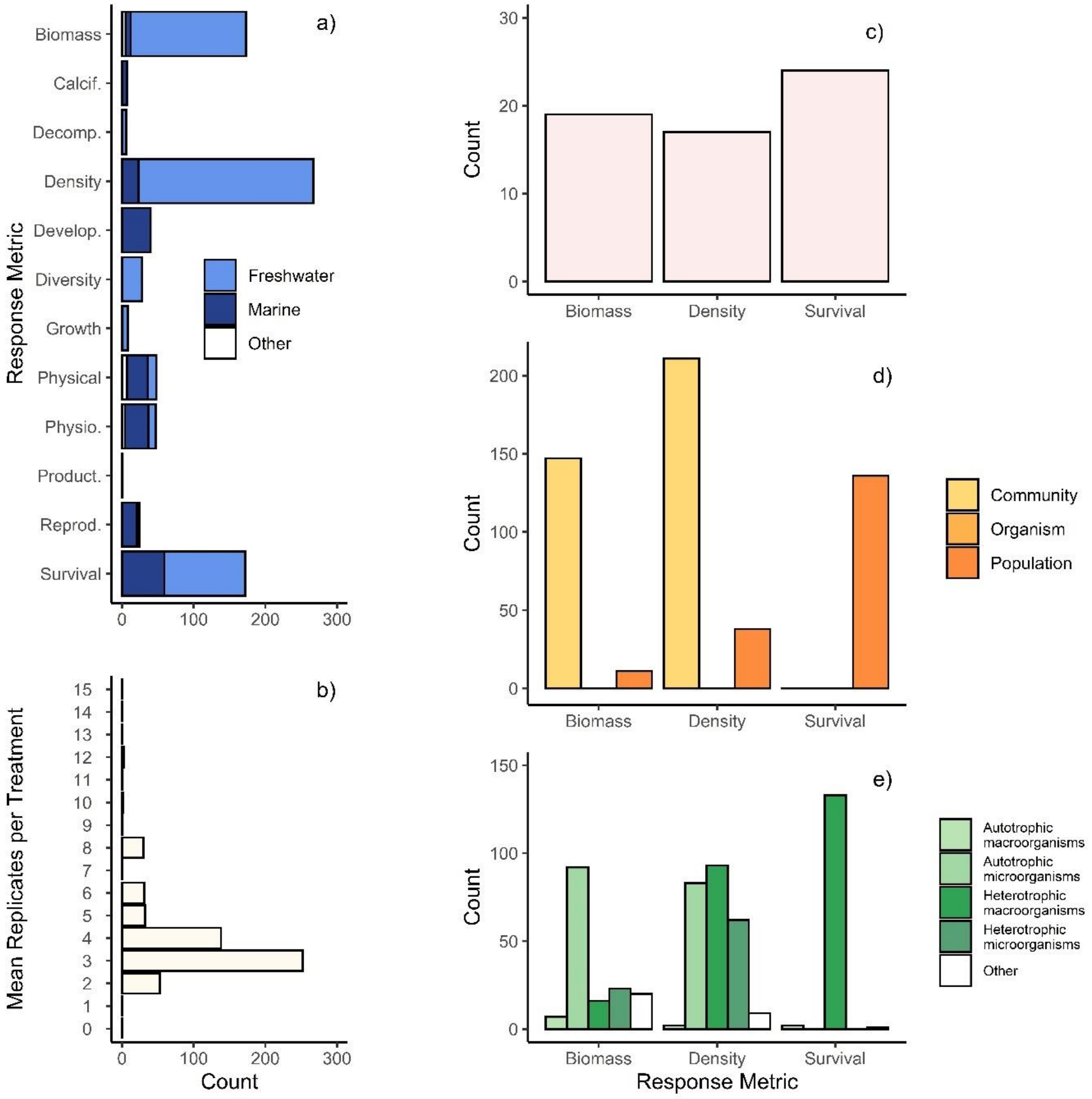

Figure 4.3: Summary of the aggregated dataset, though with removal of duplicate interactions, (Figure 4.3ab) and the aggregated biomass, density, and survival datasets (Figure 4.3cde). Figure 4.3a: Number of interactions for each response metric in the aggregated dataset, separated by realm. Figure 4.3b: Number of interactions by mean number of replicates per treatment. Where the mean number of replicates is a non-integer, it is rounded to the nearest integer. Figure 4.3c: Number of unique stressor combinations for the aggregated biomass, density, and survival datasets. Figure 4.3d: Number of interactions, across levels of ecological organisation, for the aggregated biomass, density, and survival datasets. Figure 4.3e: Number of interactions, across feeding group and body size, for the aggregated biomass, density, and survival datasets.

Hypothesis two: The effects of multiple stressors significantly differ across response metrics.

For biomass, density, and survival data, all summary effect sizes are negative, with those for biomass and density being significantly different to zero, indicating an antagonistic interaction; however, the effect size for survival is not significantly different to zero, indicating a null interaction (Figure 4.4). No significant difference was observed between the effect sizes for biomass and survival ($\chi^2 = 0.1892$, $p = 0.6635$), though significant differences were observed between the effect sizes of biomass and density ($\chi^2 = 6.6564$, $p = 0.0099$) as well as density and survival ($\chi^2 = 7.9844$, $p = 0.0047$).

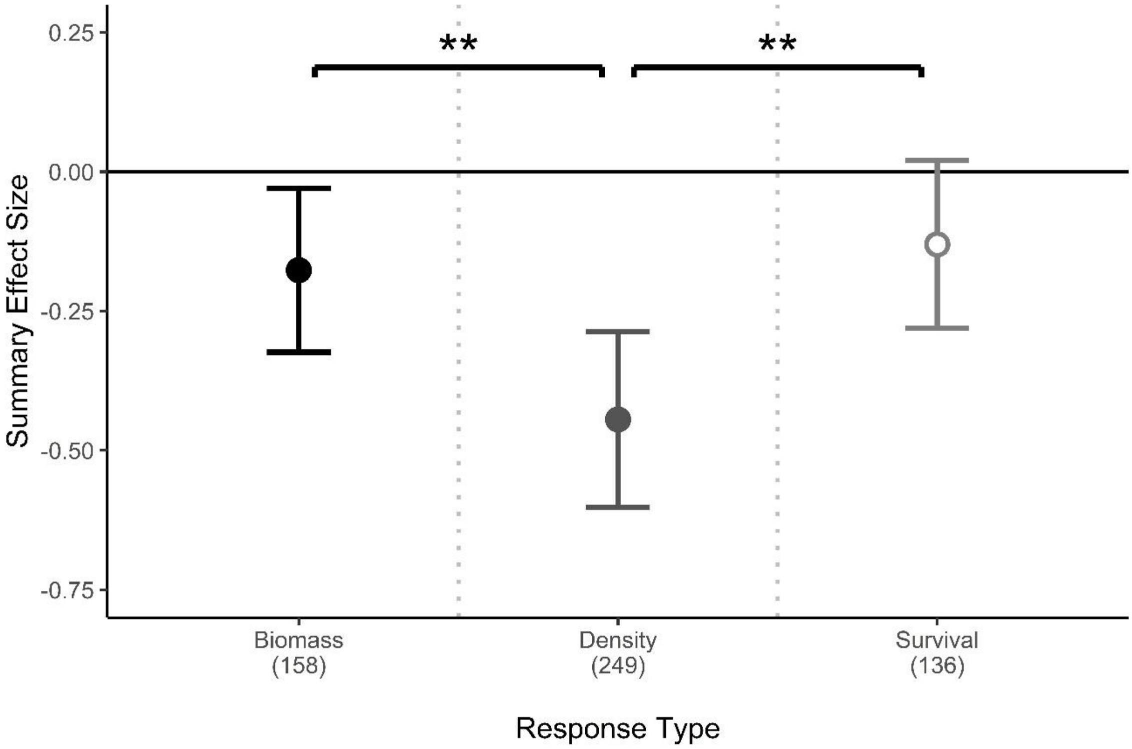

Figure 4.4: Summary effect sizes for the biomass, density, and survival response metrics. Filled points denote significant summary effect sizes (i.e., 95% confidence intervals do not overlap zero), hollow points denote non-significant summary effect sizes (i.e., 95% confidence intervals overlap zero). Differences in the summary effect sizes for each response metric, between levels of a fixed effect, are tested using a Wald-type chi squared test. Levels of significance are denoted by the following annotation: p value < 0.05 (*), p value < 0.01 (**), p value < 0.001 (***). Numbers in parenthesises denote the total number of interactions within each group.

Hypothesis three: For a given response metric, the effect of multiple stressors significantly differs between autotrophic and heterotrophic feeding groups.

Comparing summary effect sizes between autotrophic and heterotrophic feeding groups yields contrasting results across response metrics (Figure 4.5ab). For density data (Figure 4.5b), no significant difference was observed between the summary effect sizes for each feeding group ($\chi^2 = 0.9607$, $p = 0.3270$), with both being assigned an antagonistic summary interaction classification.

In contrast, for the aggregated biomass dataset (Figure 4.5a), the summary effect size for the heterotrophic feeding group is negative and significantly different to zero (again indicating an antagonistic interaction), whilst the autotroph summary effect size is non-significantly negative (indicating a null interaction between stressors). Additionally, a significant difference was reported between the biomass summary effect sizes for the different feeding groups ($\chi^2 = 6.7921$, $p = 0.0092$).

While results of this analysis are shown for both the biomass and density aggregated datasets, the data within the survival aggregated dataset was overwhelmingly heterotrophic (Figure 4.3e). Accordingly, analysis comparing summary effect sizes between autotrophic and heterotrophic feeding groups for survival data was not shown as no meaningful comparison can be made.

Hypothesis four: For a given response metric, the effects of multiple stressors will not significantly differ between stressor interactions which have an expected negative, or positive, effect.

Comparisons of the summary effect sizes between the expected direction of interaction effects yields similar results across all response metrics (Figure 4.5cde). For all response metrics and expected direction of effects, the summary effect sizes were negative and significantly different to zero, indicating antagonistic interactions. The single exception was for biomass data, where the direction of the expected effect of interacting stressors was negative; here, the summary effect was negative but not significantly different to zero, indicating a null interaction.

Finally, no significant differences in effect sizes were observed between the contrasting directions of expected stressor effects across any of the response metrics (biomass: χ^2 = 1.2134, p = 0.2707; density: χ^2 = 0.0017, p = 0.9668; survival: χ^2 = 2.2777, p = 0.1312).

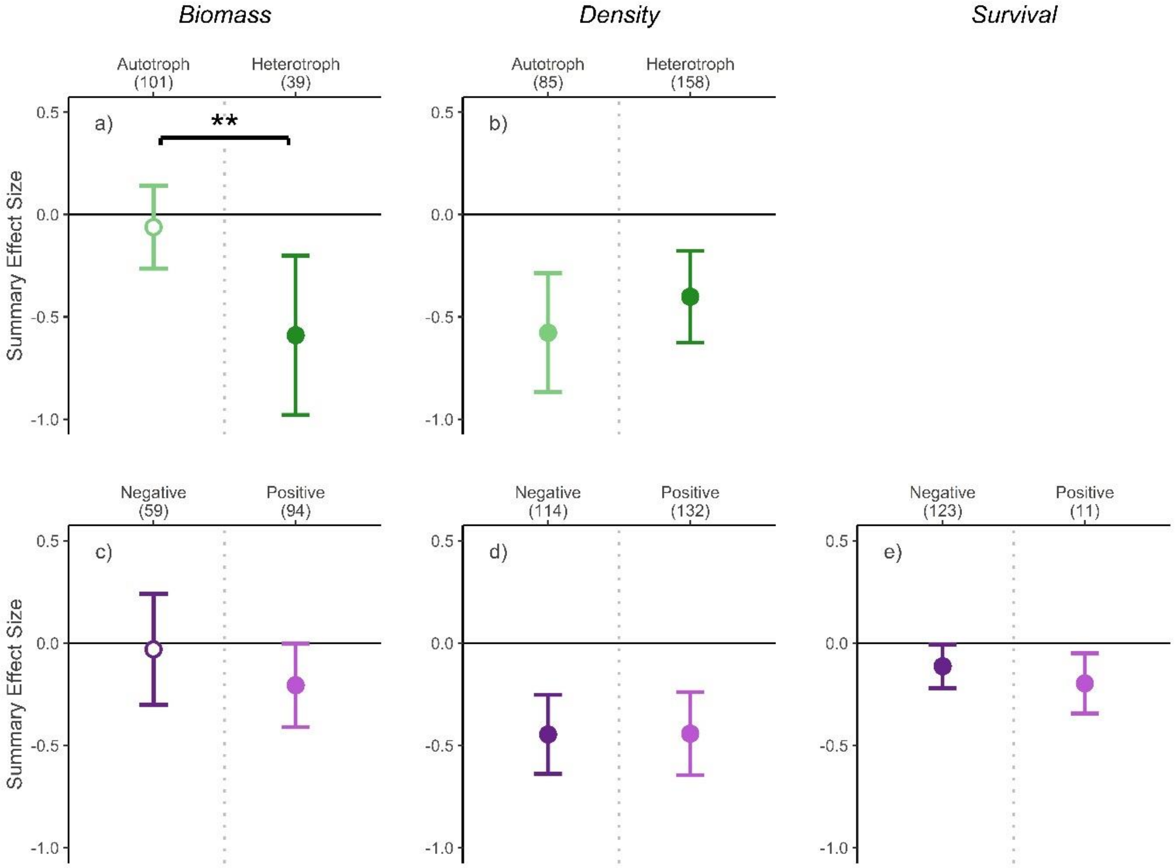

Figure 4.5: Summary effect sizes for each response metric across the fixed effects of feeding group (Figure 4.5ab) and expected direction of effects (Figure 4.5cde). Filled points denote interactions where 95% confidence intervals do not overlap zero, hollow points denote interaction interactions where 95% confidence intervals overlap zero. Differences in the summary effect sizes for each response metric, between levels of a fixed effect, are tested using a Wald-type chi squared test. Levels of significance are denoted by the following annotation: p value < 0.05 (*), p value < 0.01 (**), p value < 0.001 (***). Numbers in parenthesises denote the total number of interactions within each group.

4.5 Discussion

Meta-analyses are capable of providing general insights into how multiple stressors are likely to interact. However, despite the potential to highlight wide-ranging generalities, agreement between multiple stressor meta-analyses has to date been limited (Côté et al., 2016). By removing methodological variation and using the same statistical framework across a large number of stressor-pair interactions we have shown that strong generalities emerge (Figures 4.3 – 4.5). Firstly, synergistic interactions are rare. Summary effect sizes across all analyses are negative and assigned an antagonistic, or null, summary interaction classification; with our results supporting suggestions that the prevalence of synergisms is overstated (Côté et al., 2016). Secondly, individual interactions are overwhelmingly assigned null classifications. However, rather than being attributed to ecological phenomena, previous analyses have shown that small sample sizes limit our ability to reject a null model (Chapter Three). Thirdly, some differences in interactions are observed between response metrics and feeding groups. However, no difference in interactions were found between stressor interactions where positive or negative effects were expected.

Methodological variation

Our results highlight that methodological variation is partially responsible for the wide-ranging results of previous meta-analyses. Following the removal of methodological variation, all meta-analytical datasets report that individual interactions are predominately assigned a null classification (Figure 4.2a), whilst all summary effect sizes are negative and assigned either an antagonistic, or null, summary interaction classification (Figure 4.2b). Our results contrast many of the findings of these original analyses. However, methodological variation partially explains the absence of any generalities across meta-analyses (Côté et al., 2016), and may explain some disparities between these studies. Nevertheless, following the removal of methodological variation from our analysis, differences between the meta-analytical datasets remain (e.g., differences in the strength of antagonistic interactions across meta-analytical datasets). These differences can hence be attributed to ecological variation within, and between, the datasets (e.g., contrasting response metrics). Understanding the sources of this ecological variation is crucial as it may lead to both the discovery of important covariates which drive stressor interactions, as well as widely-applicable generalities which have implications for future research and the mitigation of interacting stressors. Whilst we explore potentially important covariates (Figures 4.4 – 4.5), one generality evident from the analysis of methodological variation is the distinct lack of synergisms, with this result mirroring with the conclusions of Côté et al. (2016). Additionally, while meta-analyses attempt to identify ecologically important covariables, it is also important to determine where results are driven by factors other than ecology. Previous analysis has shown that the prevalence of null classifications for individual interactions is driven by sample sizes, not ecological factors (Chapter Three). Indeed, within our aggregated dataset, the majority of interactions have small sample sizes (Figure 4.3b), which in turn means that null models are unlikely to be rejected unless exceptionally strong signals are present (Chapter Three). Accordingly, in keeping with recommendations (Gurevitch et al., 2018), meta-analyses should predominately focus on summary effect sizes as opposed to solely drawing, potentially erroneous, ecological inferences from the vote-counting of individual interaction classifications.

Multiple versions of the additive and multiplicative null models have previously been used, with the seven meta-analyses considered here all using differing null models or classification frameworks. As such, it is not entirely clear how these differences alter the results of any given meta-analysis. However, one example which is easily interpretable, relates to the frequency of the different interaction classifications. Whilst the calculation of the standard error is shown by Equation 4.3, it may be incorrectly assumed that the square root of the effect size variance is the standard deviation, not the standard error. Hence, Equation 4.7 may be incorrectly used to calculate effect size standard errors.

4.7)
$$SE = \frac{\sqrt{V}}{\sqrt{N_C + N_A + N_B + N_I}}$$

Accordingly, the use of Equation 4.7 (as opposed to Equation 4.6) leads to 95% confidence intervals being incorrectly reduced by a factor of $\sqrt{N_C + N_A + N_B + N_I}$. Based upon the data within the aggregated dataset this would lead to confidence intervals being reduced by a minimum of 64.6%, though in some more extreme cases confidence intervals would be reduced by 85.6%. Such a reduction greatly reduces the likelihood of the confidence intervals for individual interactions overlapping zero, hence leading to a greater prevalence of synergistic, antagonistic, and reversal interactions than might otherwise be expected.

Our findings represent the first illustration of the major impacts of methodological variation upon multiple stressor meta-analyses. Furthermore, our results echo those of a previous analysis (Hungate et al., 2009) which found that the substantial divergence in the conclusions of four highly similar soil carbon meta-analyses principally occurred due to methodological variation. Accordingly, our results highlight that methodological differences represent a substantial source of variation between multiple stressor meta-analyses.

The effect size calculation we implement here is the factorial log response ratio which is widely implemented and mathematically detailed (Lajeunesse, 2011); although we note that if alternative effect sizes are used, alternative conclusions may be reached. As such it is imperative that future meta-analyses explicitly consider the potential impacts of a particular methodology from both an ecological *and* statistical perspective. Indeed, the choice of statistical procedure can have substantial implications for both the meta-analytical results and the subsequent ecological interpretation (Hungate et al., 2009).

Finally, there is a need for model diagnostics to be more widely used within multiple stressor meta-analyses. While visual assessment of funnel plots is a commonly implemented diagnostic measure, it has been shown to be a poor method for determining whether publication bias is present (Simmonds, 2015). Furthermore, common tests for publication bias are often limited by high levels of heterogeneity (Nagakawa et al., 2017) which are frequently reported within ecological meta-analyses (Senior et al., 2016). In contrast, alternative diagnostic measures such as Cooks' Distance, DFBETAS, and $\hat{R}$ should be implemented. These metrics are capable of identifying potentially influential data points and determining whether the results of a meta-analysis are being unduly driven by a small number of interactions. Furthermore, use of these metrics may also prevent any issues with data quality (see Table 4.2) from occurring, as any potential issues may be highlighted during these post-hoc analyses.

Response metrics

Hypothesis two is based upon the expectation that stressors affect different biological mechanisms, and hence impact response metrics (e.g., density, survival) in contrasting ways (Segner et al., 2014). Our results support this hypothesis (Figure 4.4), with density being affected by multiple stressors in a significantly more antagonistic manner than either biomass or survival. However, whether the observed differences between response metrics occur due to disparate physiological pathways is unclear (see *Physiological pathways*). Alternatively, compensatory effects may be responsible for the antagonistic interactions reported for biomass and density data. At the community level, a reduction in the biomass or density of one species may be offset by increases in others (Ives & Cardinale, 2004). For instance, a single stress-tolerant species may only become dominant within a community in the presence of stressors, as has been shown for an herbivorous zooplankton species (Christensen et al., 2006). As such, this phenomena may lead to an antagonistic interaction occurring, as any changes in community biomass or density would potentially be less than that predicted by the null model. However, compensatory effects do not explain the significant difference between biomass and density summary effect sizes. Accordingly, such varying responses between metrics could be due to differences in the effects of stressors. For instance, a stressor could lead to an increase in the density of species but lead to smaller organisms due to increased competition for limited resources (e.g., Thaker et al., 2018). Alternatively, a stressor could lead to a reduction in species density but a positive effect on biomass as a consequence of compensatory effects where individuals that survive become larger, potentially due to a greater availability of resources (Galic et al., 2018).

Whilst all response metrics report antagonistic (or null) interactions, there are potential implications for any conservation measures which aim to mitigate the effect of stressors and their interactions (Côté et al., 2016). Indeed, where antagonistic interactions occur, the removal of a stressor may have a limited, or even detrimental, effect upon an affected system (Brown et al., 2013). Similarly, the relative strength of an interaction should also be considered. Given that interacting stressors have a significantly stronger antagonistic effect upon density compared to either biomass or survival, the effects of any remediation measures may be asymmetrical across response metrics.

To date, few meta-analyses have investigated effect sizes across response metrics. Of those that have, Lange et al. (2018) found substantial differences between the effect sizes for biomass and survival; although, the results of other analyses are less clear (e.g., Jackson et al., 2016). It is also important to note that all previous meta-analyses which compare across response metrics implement an additive null model. However, the additive null model is not necessarily reflective of biological systems (Schäfer & Piggott, 2018), and can even lead to unfeasible predictions (e.g., a predicted survival of less than zero) (Folt et al., 1999). Where a bounded response metric is considered, interaction data is likely skewed in favour of antagonisms, given that synergisms or truly null interactions are unlikely, or unable, to be reported by the additive null model (Lange et al., 2018). As such, the lack of clarity in the effects of multiple stressors across response metrics may in part be due to the use of a potentially ecologically inappropriate null model.

Feeding groups

We anticipated that the effects of multiple stressors would differ between autotrophs and heterotrophs, potentially as a consequence of differing stressor modes of action (Bracewell et al., 2019). However, no consistent pattern was observed across response metrics, with no significant difference between feeding groups for the aggregated density data (Figure 4.5b), but a significant difference between the summary effect sizes for the autotrophic and heterotrophic biomass data (Figure 4.5a).

Accordingly, for the biomass data, it may be that our original assumptions regarding differing physiological mechanisms are met. However, it is important to note that the overwhelming majority of the autotrophic biomass data represents microorganism communities (e.g., broad groups of phytoplankton) whereas the heterotrophic biomass data represents a more even ratio of microorganisms and macroorganisms (Figure 4.3e). As such, rather than reflecting differences between heterotrophic and autotrophic organisms, the observed effects may instead relate to differences in body sizes. Furthermore, body size is shown to be positively correlated to generation time (Gillooly, 2000), with generation time also likely to affect a species' response to stressors (Brook et al., 2008; Jackson et al., 2021).

Whilst autotrophic and heterotrophic density data is likewise split between microorganisms and macroorganisms in a similar manner to the biomass data (Figure 4.3e), biomass and density are not necessarily equivalent metrics, or affected by interacting stressors in the same manner (see Galic et al., 2018). As such, these observed differences between response metrics may occur due to biomass and density being directly, and indirectly, affected by ecological or physiological processes in differing ways. Ultimately, the differences between feeding groups across response metrics (with particular emphasis on autotrophic, or microorganism, biomass) undoubtedly warrants further investigation.

Expected direction of effects

Stressors may have either positive or negative effects (Kroeker et al., 2017; Orr et al., 2020), and therefore the combined impact of multiple stressors is similarly capable of eliciting either a positive or negative response. However, we are not aware of any meta-analyses which have attempted to determine whether the class, or strength, of an interaction varies depending upon whether it has an expected positive or negative effect. Here we attempt to address this gap in the literature. As our results show (Figure 4.5cde), across all response metrics, either antagonistic or null interaction classes are reported where positive or negative effects of multiple stressors are expected. Additionally, there were no statistically significant differences between the effect sizes where positive or negative effects were expected for any response metric. Our results support our hypothesis, suggesting that the effects of multiple stressors are consistent (both in strength and interaction class) regardless of whether a positive or negative effect of stressors is expected. However, our analysis is not capable of elucidating the physiological pathways through which positive, or negative, effects are induced. While no significant differences, based upon expected stressor effects, were observed within this meta-analysis, it may be important to consider predicted stressor effects in the context of organisation level. Accordingly, at the level of the individual, differences may be apparent depending upon whether stressors affect the same, or disparate, mechanisms (see *Physiological pathways*). In contrast, at higher levels of organisation (e.g., communities) compensatory dynamics may lead to any differences being minimized. However, the aggregated datasets considered here are predominately represented by higher levels of ecological organisation; as such, analysis of organisation level was beyond the remit of our meta-analysis. Nevertheless, this analysis represents the first attempt to understand the effects of stressor directionality; accordingly, our understanding of this phenomena would undoubtedly be improved by future experimental or theoretical analyses. Our analysis also suggests that, given no significant differences between interactions with contrasting directionalities, there may be minimal ramifications for management practices. Accordingly, conservation actions may instead focus on alternative stressor properties (e.g., stressor modes of action) that may be more relevant for the remediation of a given system.

Physiological pathways

Meta-analyses are capable of revealing how stressor interactions affect ecosystems; however, such insights are broadly limited to wide-ranging generalities. Accordingly, to compliment these general insights, there is a need to consider the specific physiological mechanisms underpinning the effects of stressors and their interactions (Schäfer & Piggott, 2018). Furthermore, stressor modes of action may help explain the results of a given null model (Townsend et al., 2008), with null interactions potentially occurring where stressors impact disparate physiological mechanisms (Gunderson et al., 2016). Similarly, a given stressor has the potential to modulate the effect of an additional stressor, either exacerbating (*cross-susceptibility*) or mitigating (*cross-tolerance*) its impact on a physiological response (Todgham & Stillman, 2013). As such, enhanced knowledge of these phenomena, in conjunction with physiological pathways, and the timing of stressor interactions (Gunderson et al., 2016; Jackson et al., 2021), may be crucial when transitioning multiple stressor ecology into practical conservation measures (Rodgers & Gomez Isaza, 2021).

An understanding of stressor modes of action may also shape the scope of multiple stressor meta-analyses and the hypotheses that are addressed. For instance, many meta-analyses group stressors into categories (e.g., contaminants, habitat alteration) based on their qualitative properties. However, large-scale meta-analyses often find no generalities across stressor combinations (e.g., Burgess et al., 2021; Jackson et al., 2016). Given the broad nature of this meta-analysis, coupled with the large number of unique stressor combinations for each response metric (Figure 4.3c), we do not attempt to explore the role of stressor combination in determining stressor interactions.

Finally, there is evidence that qualitatively dissimilar stressors may act through the same physiological pathways, with both temperature and chlorpyrifos being shown to impact thyroid hormones in *Acanthurus triostegus*, (Besson et al., 2020). Such physiological effects are also capable of transitioning to the community level through the alteration of predator-prey relationships (Besson et al., 2020), a phenomena observed for a range of aquatic stressors and taxa (e.g., Shears & Ross, 2010). As such, this suggests that precedence should be given to a stressor's mode of action, as opposed to its qualitative properties; however, such mechanistic insights are seldom reported and often inadequately understood (Spears et al., 2021). Accordingly, increased consideration of physiological pathways may lead to the discovery of more generalities for multiple stressor interactions (across all levels of organisation), which in turn may aid understanding of both stressor interactions and modes of action. Ultimately, given the potential benefits, greater overlap of these complimentary approaches should be viewed as a priority for multiple stressor ecology.

Future directions

One of the main aims of meta-analyses is to determine whether ecologically important covariables are capable of explaining sources of heterogeneity within a given dataset. As such, within this multiple stressor meta-analysis, we have sought to determine whether response metric, feeding group, or expected direction of effects are important covariables in determining how stressors interact in aquatic ecosystems. However, while there is an ecological rationale underpinning each of these covariables, there is a plethora of other covariables which future multiple stressor analyses (meta-analytical or otherwise) may wish to investigate. Indeed, there are numerous calls for multiple stressor ecology to place a greater emphasis on understanding the mechanisms underpinning stressor effects (e.g., Kroeker et al., 2017; Schäfer & Piggott, 2018; see *Physiological pathways*). However, alternative covariables to consider include the level of ecological organisation at which the effect of multiple stressors is measured. As such, future multiple stressor research may wish to focus on higher levels of ecological organisation given that a species' sensitivity to stressors is partially defined by its relationships to other species in a community (Beauchesne et al., 2021), as well as conservation actions primarily being enacted at these higher levels of ecological organisation (Orr et al., 2020). To date, meta-analyses have failed to determine generalities regarding whether certain organisation levels are more likely to be affected by particular interaction classifications (Côté et al., 2016; Orr et al., 2020); though as we have illustrated here, such disparities may arise as a consequence of differences in methodology, not ecology. However, there is growing body of ecological theory investigating stressor interactions across levels of organisation (Orr et al., 2020). This theory has shown the importance of interactions between species in determining whether community level response metrics are affected by synergisms or antagonisms (De Laender, 2018; Thompson et al., 2018a). Furthermore, theoretical freshwater systems have shown that interaction classifications may differ, or align, across different organisation levels (Galic et al., 2018).

However, within this analysis, the response metrics we consider are skewed towards either community or population organisation levels (Figure 4.3d); as such, analysis of organisation level was not possible. Accordingly, future multiple stressor meta-analyses may wish to investigate the importance of organisation level as a covariable in determining stressor interactions.

A second potential explanatory variable, which future analyses may wish to consider, is whether interactions differ between realms (i.e., freshwater and marine ecosystems). Whilst combining datasets from different realms, we do not test its effect within our analyses, for several reasons. Firstly, of the seven datasets considered here, the four freshwater datasets are more recently published, as such they are able to collate data from a wider pool of the scientific literature than the three marine studies. Secondly, the freshwater datasets primarily consider biomass, density, or survival response metrics; in contrast, the marine datasets consider a much broader range of response metrics, with each metric often having a smaller number of interactions (Figure 4.3a). Together, these attributes mean that our aggregated dataset is weighted towards freshwater interactions, with the ramification that comparisons across realms are not necessarily meaningful. As such, if future research wishes to explore the effect of realm, then further generalities may be uncovered. However, it may be necessary to consider more specific ecosystem types than simply freshwater or marine realms (e.g., coastal, estuarine, or wetland environments), as stressors may interact differently in lacustrine or riverine ecosystems (Birk et al., 2020).

Ultimately, one of the greatest limitations faced by multiple stressor meta-analyses is the availability of data that is able to be analysed. As shown by Figure 4.3a, particular response metrics (e.g., diversity, growth, and reproduction) are under-represented in meta-analyses. Similarly, as discussed above, data limitations mean that determining the importance of covariables such as realm, organisation level, or even feeding group remain difficult. Accordingly, to overcome this issue, future empirical studies may wish to focus on those areas which we have highlighted has being potentially data deficient (see Figure 4.3). Similarly, there is likewise ample scope for future multiple stressor meta-analyses to collate new datasets, building upon those already compiled, and incorporating more recently published studies.

Our analysis has revealed that many of the disparities between multiple stressor meta-analyses have arisen due to methodological, not ecological, variation (see *Methodological variation*). Here, we have standardised methods across all analyses by using a single consistent version of the multiplicative null model and interaction classification framework. While the multiplicative null model is generally considered to be more ecologically realistic than the additive null model (Schäfer & Piggott, 2018), there are an increasing number of alternatives which may be appropriate under particular circumstances (e.g., Tekin et al., 2020; Thompson et al., 2018b). Given the methodological variation outlined by our analysis, future analyses should explicitly describe all equations used within null model calculations; such details would hence allow simple comparisons between studies which potentially highlight methodological rationale for why the results of two studies differ. Similarly, future analyses should explicitly justify why a particular null model was implemented. Given that versions of the additive and multiplicative null models have been shown to be comparable in terms of statistical power (Chapter Three), researchers can hence select a null model based upon ecological criteria, which should be described.

Conclusions

Both freshwater and marine ecosystems are being subjected to multiple co-occurring stressors (Halpern et al., 2015; Reid et al., 2019). Furthermore, understanding how stressors interact is imperative when enacting conservation measures upon an aquatic ecosystem (Brown et al., 2013). However, no generalities across multiple stressor meta-analyses have previously been reported (Côté et al., 2016) as such hindering our collective understanding of stressor interactions in freshwater and marine ecosystems. Our results illustrate that methodological variation is responsible for the absence of both generalities (e.g., absence of synergisms), and ecologically important covariates (e.g., response metric) across aquatic multiple stressor meta-analyses. These findings illustrate the need to thoroughly consider both the statistical tools used to classify stressor interactions and any comparisons that are made across meta-analyses. Our results likewise have implications for the management of aquatic ecosystems given that response metric and feeding group are identified as potentially important covariates for determining how stressors interact. Importantly, our analysis also highlights where data deficiencies in aquatic meta-analyses currently exist, and hence limit the ability of these analyses to determine whether particular covariates (e.g., realm, organisation level, certain response metrics) are important in driving stressor interactions in aquatic ecosystems. Ultimately, future research should seek to address the issues we spotlight here, so that a greater understanding of multiple stressor interactions can transition into tangible benefits for the conservation and remediation of affected freshwater and marine ecosystems.

Chapter Five – Discussion

Nearly all ecosystems are subjected to multiple stressors (Halpern et al., 2015). Indeed, understanding how co-occurring stressors interact to affect ecosystems is one of the most pressing questions currently facing ecologists (Fleishman et al., 2011; Rudd, 2014). However, multiple stressor ecology has to date been hindered by the absence of generalities applicable across studies. The results of such studies may frequently differ with various interaction classifications reported (e.g., antagonistic, null, or synergistic interactions). Furthermore, similar studies may likewise reach divergent conclusions as to how stressors interact (e.g., Jackson et al., 2016; Lange et al., 2018). This absence of any generalities ultimately harms multiple stressor ecology as a field, limiting the insights it can provide to sister disciplines (e.g., conservation science) and obscuring future directions for research. To date, the disparate conclusions of much multiple stressor research has often been attributed to underlying variation from ecological or environmental sources. However, attempts to identify ecological covariables capable of explaining this variation have so far failed to provide any consistent insights (Côté et al., 2016). While ecological, or environmental, factors are undoubtedly important in determining how stressors interact, the current failure to identify these covariables suggests that other factors may be influencing results. Accordingly, variation in statistical methods (i.e., null models) is one such factor which is potentially capable of explaining differences between studies. Indeed, there is an increasing appreciation that there are inherent differences between null models (i.e., additive and multiplicative null models) including the hypotheses that they test (Schäfer & Piggott, 2018). However, there is a limited understanding of the statistical methods through which these null models are applied. Importantly, there may be instances where results arise due to the statistical properties of these methods, yet they are erroneously ascribed ecological importance. Accordingly, an understanding of the statistical tools frequently employed in multiple stressor ecology is currently missing, yet crucial to the progression of the field.

Within this thesis, the research I have detailed has attempted to understand multiple stressor interactions from various different, often novel, perspectives. Broadly, I have utilised theoretical (Chapters Two and Three) or meta-analytical (Chapters Two and Four) approaches to better understand the statistical tools used to analyse, and classify, the interactions of multiple stressors. While the use of null models is widespread in multiple stressor ecology (e.g., Folt et al., 1999; Galic et al., 2018; Siviter et al., 2021) a detailed understanding of these approaches was previously missing. As such, it was unclear whether these null models encompass any biases (aside from those previously described; see Sih et al., 1998) which may skew any interpretation of how stressors interact to affect all levels of ecological organisation. Accordingly, this thesis has addressed these limitations. In Chapter Two, I have shown that the dominance of antagonistic summary interaction classifications, under the additive null model, may be an artefact of the statistical tool. Indeed, a theoretical dataset with an approximately equal frequency of 'true' synergistic and antagonistic interactions, will report an antagonistic summary interaction classification under ecologically relevant sample sizes and observation errors. Such a finding is mirrored in the results of the largest published multiple stressor meta-analysis (Chapter Two; Burgess et al., 2021). In Chapter Three, I build on this analysis illustrating that some patterns (e.g., frequency of individual interaction classifications) are predominately driven by sample sizes and other statistical factors, with this occurring regardless of whether the additive or multiplicative null models are used. As such, ecological interpretations of these results are likely inaccurate and instead highlighting that methodological choices are often limiting our ability to determine how stressors truly interact. In Chapter Four, I build on the previous chapters reanalysing seven multiple stressor meta-analyses determining why the results of these studies are so disparate. The results of this chapter illustrate that methodological differences are one of the major causes of these contrasting conclusions, with the removal of these differences rendering the results more consistent across all seven meta-analyses. Furthermore, I highlight that multiple stressor interactions vary in classification, or strength, across response metrics, a previously disputed finding (Côté et al., 2016).

As such, the analysis, results, and conclusions of this thesis build on the existing multiple stressor literature, offering explanations for several of the more prominent issues within the field of multiple stressor ecology. Accordingly, the analysis of Chapters Two, Three, and Four has met the following aims, first outlined in Chapter One.

i. Determine whether ecological surprises (i.e., antagonistic, reversal, and synergistic interactions) are more prevalent than currently expected.

ii. Determine whether ecology, or methodology, are responsible for the conclusions of previous multiple stressor research.

iii. Determine the causes responsible for the absence of consistent generalities across multiple stressor meta-analyses.

iv. Determine what covariables (e.g., feeding group or response metric) are important in determining how stressors interact in aquatic ecosystems.

Within this chapter, I explore the contributions of Chapters Two, Three, and Four to the field of multiple stressor ecology. Firstly, I outline the findings of each chapter, highlighting overarching conclusions, and the subsequent implications for the collective understanding of interacting stressors. Secondly, building on the findings of this thesis, I explore future directions for multiple stressor ecology. In doing so, I spotlight where methodological, empirical, and theoretical advances could be made. Finally, I provide a broader commentary on multiple stressor ecology, outlining those areas (outside the remit of this thesis) which require further research and postulating on our current understanding of how interacting stressors affect ecosystems.

5.1 Implications for the understanding of multiple stressor interactions

Multiple stressor ecology is often primarily concerned with answering the question: *How do co-occurring stressors interact?* Indeed, within this thesis, this is a question I have often attempted to answer. However, it may be argued that the majority of studies view obtaining an answer to this question as more important than the process through which the answer is gained. While this in itself is not necessarily an issue, it undoubtedly becomes one if the process is subsequently shown to be inappropriate, or a valid process is incorrectly applied. Within this thesis, I have attempted to gain a greater understanding of the methods used to determine how stressors interact. In doing so, I have illustrated potential issues regarding both the null models and experimental design (Chapters Two and Three) alongside instances where, unfortunately, statistical procedures have sometimes been erroneously conducted (Chapter Four). Below, I outline the conclusions of each chapter, describing the implications both for past research and how these findings can be utilised by future multiple stressor studies.

To date, there has been relatively little research which seeks to combine theoretical and empirical data within multiple stressor ecology. However, in Chapter Two, I help fill this literature gap by using both theoretical and empirical approaches to better understand the interactions of multiple stressors. Here, the empirical approach comprises one of the largest published multiple stressor meta-analysis (Burgess et al., 2021) which considers freshwater density data. In contrast, the theoretical approach considers food chains established using fundamental ecology (Heath et al., 2014) which are then subjected to stressors by altering parameters corresponding to ecological processes (e.g., mortality rates). These theoretical simulations revealed that under increasing levels of uncertainty, individual interactions are increasingly likely to be assigned a null classification. Furthermore, at lower levels of uncertainty, there is an approximately equal split of negative to positive effect sizes for individual interactions, with summary effect sizes assigned a null classification. However, under increasing uncertainty, individual interactions are more likely to have a negative (than positive) effect size; summary effect sizes are likewise skewed towards being assigned antagonistic interaction classifications. The results of the freshwater meta-analysis likewise mirror these findings, with null classifications dominating for individual interactions, but an antagonistic summary effect size is reported. The findings of Chapter Two have implications for past multiple stressor research. Both the theoretical and empirical approaches report that the majority of individual interactions are assigned a null classification, but summary effect sizes are classified as antagonistic. However, these results likewise align with previous freshwater research (e.g., Lange et al., 2018) and contrast other studies (e.g., Jackson et al., 2016). As such, these results indicate that there may be ecological processes (or statistical attributes) that are currently unconsidered but instrumental in driving reported patterns. Accordingly, the results of Chapter Two raise further questions regarding potential artefacts in the statistical tools (or modelling approaches) or currently unconsidered ecological processes which may be instrumental in driving the reported patterns across this chapter and previous research. Ultimately, the results of Chapter Two have directed and shaped the analyses of the subsequent chapters of this thesis.

In Chapter Three, I expanded upon the theoretical methods used in Chapter Two, to better understand the null model approaches often utilised within multiple stressor ecology. The research presented within this chapter explored both the additive and multiplicative null models from a statistical perspective, as opposed to the ecological viewpoint from which these tools are predominately viewed. The analyses of these null models spotlight several important points. Firstly, the results of the additive and multiplicative null models cannot be directly compared. This is due to the fact that these null models implement differing assumptions and hence test different hypotheses (namely whether stressors interact in an additive or multiplicative manner), with interactions often being assigned contrasting classifications by the different null models. Accordingly, comparisons of studies which implement differing null models (e.g., Côté et al., 2016) will be inherently difficult. This analysis has also revealed that, statistically, neither null model is obviously better than the other, meaning that the choice of null model to be implemented by a study can be based on ecological criteria. However, it has been shown that both sample sizes and the uncertainty surrounding treatment means may have significant effects on null model conclusions. For instance, experiments which utilise small sample sizes (i.e., those common in multiple stressor ecology) frequently lack the necessary statistical power to determine whether an interaction between stressors is occurring. These findings have implications for the interpretation of previous multiple stressor research. While previous multiple stressor meta-analyses (e.g., Lange et al., 2018) have concluded that the majority of stressor interactions are assigned a null (i.e., additive) interaction classification, this may simply reflect an absence of statistical power (due to limited sample sizes) as opposed to any ecological mechanisms. Accordingly, it is important that researchers do not assign ecological importance to results which are simply statistical artefacts.

Previous research has highlighted the absence of generalities from multiple stressor meta-analyses (Côté et al., 2016). In Chapter Four, I collate, rectify, and reanalyse datasets from seven freshwater and marine multiple stressor meta-analyses. In doing so I implement the multiplicative null model under a single consistent framework for all seven datasets, with this analysis revealing that the removal of methodological differences results in consistent conclusions across all multiple stressor meta-analyses. Furthermore, some of the previously disparate results have been shown to be due to erroneous methods. The results of these meta-analyses align with those of Chapters Two and Three, with all meta-analyses reporting that most individual interactions are assigned null classifications (likely due to small sample sizes within multiple stressor experiments) while all summary effect sizes are assigned either an antagonistic or null interaction classification. However, whether the prevalence of antagonistic summary interaction classifications is due to ecology or artefacts within the methods (Chapter Two) remains to be determined. The formation of a single dataset (by amalgamating data from each of the seven meta-analyses) also reveals that response metric (e.g., biomass, density, survival) and feeding group (e.g., autotrophs, or heterotrophs) may be important variables in determining stressor interactions. As such, it is likely that methodological differences (or errors) have prevented consistent generalities across multiple stressor meta-analyses from being identified. Overall, this may potentially have severe consequences with multiple stressor research being diverted to search for ecological explanations for methodological differences.

5.2 Thesis contribution to future analyses

Overall, this thesis makes several important findings, which in turn should be considered, and ideally implemented, within any future multiple stressor analyses. Firstly, there must be a greater appreciation of the null models used in multiple stressor ecology, particularly regarding the implementation and subsequent interpretation of these statistical tools. For instance, it is important to consider that null models test whether a particular hypothesis (e.g., an additivity of stressor effects) can be rejected. Accordingly, a failure to reject the null model does not equate to the null model being accepted. While this may appear to be a trivial difference in conclusions, it is nonetheless important to acknowledge, and respect, this difference. As such, a failure to reject a null model may be due to a number of factors. For instance, there may be an ecological basis for why a null model was rejected, namely that stressors truly interact in an additive (or multiplicative) manner hence no deviation from the null model occurs. Alternatively, such factors may be statistical, perhaps revolving around high levels of variation, or low sample sizes. Indeed, if there are high levels of variation it may be difficult to reject the null model even if an interaction between stressors is occurring (Chapter Two). Similarly, if an experiment uses low sample sizes per treatment, then that experiment will only have the necessary statistical power to detect exceptionally strong signals (e.g., strong synergistic or antagonistic interactions) (Chapter Three). As such, there is a need for researchers to consider the statistical properties (e.g., power) of their experiment a-priori. Indeed, calculations of critical effect sizes (Chapter Three) for an experiment will indicate whether it has the necessary statistical power to detect an effect of a desired strength, or only exceptionally strong interactions. Ideally, the calculation of any effect size (e.g., Gurevitch et al., 2000; Lajeunesse, 2011) should be accompanied by the critical effect size (where appropriate) so that readers are able to determine the statistical power of any experiment and how this may affect the interpretation of any results. Indeed, a greater appreciation of the power of any experiment may lead to an increased consideration of the relative strength of stressor interactions, alongside their classification. For instance, if an experiment with low statistical power reveals a synergistic interaction, then such interaction could be described as being a 'strong synergism'. Alternatively, if the same experiment instead returns a null interaction classification, then this may be described as 'a strong interaction was not found to occur' as a weaker interaction may still be occurring which the experiment failed to detect.

There have also been suggestions that multiple stressor experiments should consider a greater range of stressor intensities, given that this may help reveal how stressors interact from a mechanistic perspective (Griffen et al., 2016; Schäfer & Piggott, 2018). However, given that resource limitations are commonplace in multiple stressor ecology it may be that considering additional stressor intensities may further reduce the already limited statistical power of any experiments. As such, if resource limitations are an issue, it may be prudent to adopt a pragmatic approach and focus on fewer stressor intensities each with more replicates. Alternatively, if the consideration of multiple stressor intensities is required, then different statistical analyses may be necessary. Other approaches have been suggested (e.g., Spears et al., 2021) though these tools may not be as widely used as the approaches implemented here with their benefits or limitations yet to be fully outlined. Furthermore, where new null models (e.g., Tekin et al., 2020; Thompson et al., 2018b) are presented, these must be thoroughly understood, and benchmarked against existing widely used methods, before they are widely adopted. Indeed, such an understanding may reveal ecological or statistical rationale which encourage (or potentially discourage) transitioning from one method to another (e.g., Orr et al., 2021a).

Meta-analyses represent some of the most influential (and highly-cited) research within multiple stressor ecology. However, as previously illustrated (Chapter Four), methodological differences have prevented these meta-analyses from revealing consistent generalities, which ultimately has held back multiple stressor ecology as a field. So as to prevent methodological differences (or errors) from occurring in the future, it is important that (as a minimum) multiple stressor meta-analyses outline all equations that have been used and ensure that raw data is openly available. Indeed, such actions would perhaps prevent methodological issues from arising so easily, given that equations could be compared to their original sources and raw data able to be reanalysed. Additionally, there is ample scope for future multiple stressor meta-analyses. Indeed, the discovery of generalities is an important aspect of multiple stressor ecology, given the potential for these generalities to shape management actions. However, our knowledge of these generalities is currently poor. As such, future meta-analyses may wish to determine whether a given co-variable is important in determining how stressors interact. Such meta-analyses may wish to expand on previous studies or focus on those variables which are currently data deficient (Chapter Four) or have been the subject of contrasting conclusions to date (see Côté et al., 2016).

Using the existing datasets of seven aquatic meta-analyses (Chapter Four), current gaps in the multiple stressor experimental literature have been identified. As such, future aquatic experimental studies have the potential to rectify these omissions from the scientific literature. However, it is important that future experiments consider environmental conditions or stressor combinations that reflect those likely to be experienced by ecosystems (Korell et al., 2020). For instance, there is little to be gained (from a conservation standpoint) in determining how an ecosystem responds to a pharmaceutical contaminant, if that ecosystem is unlikely to be subjected to said contaminant. To date, empirical studies within the freshwater or marine realm have often prioritised response metrics such as abundance, density, or survival. However, there is the potential for future experiments to consider other community level metrics such as stability (e.g., Polazzo & Rico, 2021) to gain a perspective of how stressors may interact to affect alternative ecosystem properties. Similarly, future research may likewise wish to analyse long term datasets (e.g., RivFishTIME; Comte et al., 2021a) in an attempt to understand how co-occurring stressors affect ecosystems in ‘real world’ conditions (e.g., Comte et al., 2021b; Erős et al., 2020). Indeed, such analyses may reveal how metrics such as population densities, species interactions, or stability vary in response to multiple stressors across time-scales. Furthermore, there are increasing discussions on the need to consider stressor time-scales (relative to a species' generation time) within multiple stressor ecology (Jackson et al., 2021).

As previously identified (Chapters One - Three), ecological theory has not been widely utilised within multiple stressor ecology. However, the incorporation of ecological theory into multiple stressor ecology is essential, given that it may be used to make predictions that can be tested with empirical data (e.g., Chapter Two); similarly, such theory may also be used to test the findings of experimental analyses. Indeed, this absence of theoretical ecology is particularly acute when considering freshwater environments, though this likewise provides ample opportunity for future studies to build on the limited research in this realm. Accordingly, one potential avenue for expansion is the use of fundamental theoretical ecology, within multiple stressor research, in novel ways (e.g., Galic et al., 2018). Alternatively, existing general ecosystem models (e.g., Christensen & Walters, 2004; Harfoot et al., 2014; Purves et al., 2013) could be more widely expanded (e.g., Bartlett et al., 2016; Fu et al., 2018) to consider how multiple stressors affect various levels of ecological organisation.

5.3 Broad overview of multiple stressor ecology

Multiple stressor ecology is a relatively broad discipline and (as previously discussed) encompasses all levels of ecological organisation and overlaps with a range of other fields (e.g., ecotoxicology, conservation science). However, there are various commentaries on where attention should be focussed within multiple stressor ecology. Indeed, many studies (e.g., Griffen et al., 2016; Schäfer & Piggott, 2018) suggest that multiple stressor ecology should initially focus on gaining a mechanistic understanding of stressors, most likely at the organismal level given that stressors primarily act on individuals (Maltby, 1999). Alternatively, Orr et al., (2020) describe how multiple stressor research should prioritise understanding how interacting stressors affect higher levels of biological organisation, given that conservation actions are primarily enacted on communities or ecosystems. As has been repeatedly noted, stressors do not affect different organisation levels in isolation, instead stressor effects may transition between these different levels (e.g., Besson et al., 2020; Shears & Ross, 2010). In a similar manner, insights gained from multiple stressor research conducted at one organisation level will likewise percolate, providing benefits to our understanding of stressor effects at other levels of organisation (Griffen et al., 2016). Accordingly, focussing future research on one organisation level will likely lead to both benefits and ramifications for the field as a whole. However, discussions regarding the future of multiple stressor ecology are undoubtedly important, although it would be remiss not to frame these arguments in regard to the relative benefits which focussing on these different levels of organisation can provide.

The need for a mechanistic understanding of stressor effects has been outlined by numerous studies (e.g., Griffen et al., 2016; Orr et al., 2021b; Schäfer & Piggott, 2018). However, despite being oft discussed, a mechanistic understanding of stressors and their interactions is by no means easy to achieve. Indeed, a large amount of data is required for a mechanistic understanding, although currently this data is often limited or absent (Spears et al., 2021). A limited mechanistic understanding is further compounded by the sheer number of stressors (and their intensities) coupled with a plethora of organisms and myriad of physiological pathways which are potentially able to be affected (Côté et al., 2016). As such, given the multiplicity of stressor and organism combinations, prioritisation of particular stressors or organisms of concern is likely to be necessary. However, given the currently limited data on stressor mechanisms, there is ample scope for novel experimental studies to fill this literature gap. Similarly, it is likely that experiments conducted at the individual level will be less resource intensive than those at higher levels of organisation. As such, for these experiments, there is the potential to implement much larger sample sizes than those typically used for community level experiments. As such, these larger sample sizes would in turn allow much subtler stressor effects to be determined given the increased statistical power of these experiments (Chapter Three). Indeed, this greater statistical power may be crucial in determining whether stressors interact to perturb physiological mechanisms which may subsequently affect other response metrics in different ways (e.g., survival).

The importance of stressor interactions within management planning is often discussed within multiple stressor ecology (e.g., Brown et al., 2014; Falkenburg et al., 2013; Kath et al., 2018); indeed, it has been shown that a failure to account for stressor actions can lead to conservation actions failing (Brown et al., 2013). Given that such actions are primarily focused on communities and ecosystems, Orr et al. (2020) have called for a prioritisation of these organisation levels within multiple stressor ecology. Indeed, community or ecosystem level studies may reveal how species interactions, community stability and even ecosystem services are affected by interacting stressors. However, interpreting any changes in these response metrics may rely on understanding population level effects (e.g., changes in feeding or survival rates); although these may in turn require knowledge of stressor effects at the individual level to fully understand (see above). Indeed, an improved ability to manage impacted ecosystems requires knowledge of stressor effects across all levels of organisation (De Laender et al., 2016; Polazzo & Rico, 2021). However, an increased consideration of the effects of stressors at an ecosystem or community level will undoubtedly be important. Although it is equally important (if not more so) that multiple stressor ecology collaborates to a greater degree with conservation science. To date, multiple stressor ecology is primarily academic in its outlook. However, to have the maximum benefits to science (and society) there must be shift towards prioritising the tangible, or practical, benefits from multiple stressor ecology. Indeed, there have been few (if any) instances where an understanding of stressor interactions has resulted in a successful management plan, with none apparent for freshwater ecosystems (Spears et al., 2021). However, the absence of management decisions based on multiple stressor ecology may be due in part to the lack of clear generalities which the field has provided (Côté et al., 2016). Overall, while an increased prioritisation of multiple stressor effects at the community or ecosystem levels is perhaps welcome, researchers must consider the direct (i.e., practical) implications of their work and how conservation or management actions may be aided.

Finally, based upon previous results (Chapters Two and Four) it may be that stressors generally interact in an antagonistic manner. As such, if the majority of stressor interactions are truly antagonistic then what are the ecological implications? Indeed, is this evidence that ecosystems inherently push back against perturbations? If such mitigation does occur, it is unlikely to be a single process which may alleviate any stressor effects, but instead an amalgamation of processes across organismal (e.g., physiological; Sokolova, 2021), population (e.g., genetic; Orr et al., 2021b), and community (e.g., stability; Polazzo & Rico, 2021) levels. Importantly, it may also be that these mitigation processes fundamentally change the properties of an ecosystem. As such, if stressors are removed from an ecosystem, it will not necessarily revert to the same state it inhabited prior to the stressors. Accordingly, if a population evolves to adapt to the presence of stressors, then the removal of these stressors may have unforeseen, negative, consequences for this population and others in the community (Orr et al., 2021b). Indeed, such mechanisms may likewise make understanding how ecosystems respond to the removal of interacting stressors even more challenging. As such, this further highlights the need for stressor mechanisms to be better understood and for multiple stressor ecology to become more entwinned with conservation science.

5.4 Concluding remarks

Various anthropogenic actions are having significant impacts on the natural world (Best, 2019; Dudgeon, 2019). Indeed, the effects of many of these stressors (e.g., climate change or pollution) are increasingly understood when considered in isolation (Reid et al., 2019). However, there remains much uncertainty as to how these stressors may combine to affect ecosystems (Côté et al., 2016). Individual experiments and meta-analyses have all reported that antagonistic, null, and synergistic interactions should be expected, though there has been little consistency in these conclusions. Accordingly, the research within this thesis has helped illustrate why such generalities are so far absent. For instance, multiple stressor experiments may only have the necessary statistical power to detect exceptionally strong interactions, meaning that many weaker (though potentially still important) interactions between stressors are not reported (Chapter Three). For multiple stressor meta-analyses, disparate (and sometimes erroneous) methodologies are likely responsible for the absence of any generalities (Chapter Four).

Although grounded in established science (e.g., Soluk & Collins, 1988), multiple stressor ecology is still relatively immature as a field of research. Indeed, this relative infancy may be partially responsible for the absence of generalities, as there is minimal guidance of how decisions (with statistical or analytical consequences) may affect ecological interpretations. As such, while future aims have previously been suggested for multiple stressor ecology (e.g., Griffen et al., 2016; Orr et al., 2020; Schäfer & Piggott, 2018) it may be prudent to first consider the statistical aspects of the field (e.g., sample sizes and analytical techniques) which may be preventing these aims from being met. It is only once these statistical attributes are better understood that multiple stressor ecology can make greater progress in gaining mechanistic insights or overlapping with conservation science.

Bibliography

Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences*, 113(42), 11770-11775.

Anderson, D. M. (2009). Approaches to monitoring, control and management of harmful algal blooms (HABs). *Ocean & Coastal Management*, 52(7), 342-347.

Andrello, M., Guilhaumon, F., Albouy, C., Parravicini, V., Scholtens, J., Verley, P., ... & Mouillot, D. (2017). Global mismatch between fishing dependency and larval supply from marine reserves. *Nature Communications*, 8(1), 1-9.

Archer, L. C., Hutton, S. A., Harman, L., McCormick, S. D., O'Grady, M. N., Kerry, J. P., ... & Reed, T. E. (2020). Food and temperature stressors have opposing effects in determining flexible migration decisions in brown trout (Salmo trutta). *Global Change Biology*, 26(5), 2878-2896.

Arrigo, K. R., van Dijken, G. L., Cameron, M. A., van der Grient, J., Wedding, L. M., Hazen, L., ... & Bailey, R. M. (2020). Synergistic interactions among growing stressors increase risk to an Arctic ecosystem. *Nature Communications*, 11(1), 1-8.

Balian, E. V., Segers, H., Lévèque, C. & Martens, K. (2008). The freshwater animal diversity assessment: an overview of the results. *Hydrobiologia*, 595, 627–637.

Ban, S. S., Graham, N. A., & Connolly, S. R. (2014). Evidence for multiple stressor interactions and effects on coral reefs. *Global Change Biology*, 20(3), 681-697.

Bancroft, B. A., Baker, N. J., & Blaustein, A. R. (2008). A meta-analysis of the effects of ultraviolet B radiation and its synergistic interactions with pH, contaminants, and disease on amphibian survival. *Conservation Biology*, 22(4), 987-996.

Barbarossa, V., Bosmans, J., Wanders, N., King, H., Bierkens, M. F., Huijbregts, M. A., & Schipper, A. M. (2021). Threats of global warming to the world's freshwater fishes. *Nature Communications*, 12(1), 1-10.

Bartlett, L. J., Newbold, T., Purves, D. W., Tittensor, D. P., & Harfoot, M. B. (2016). Synergistic impacts of habitat loss and fragmentation on model ecosystems. *Proceedings of the Royal Society B: Biological Sciences*, 283(1839), 20161027.

Beauchesne, D., Cazelles, K., Archambault, P., Dee, L.E. & Gravel, D. (2021) On the sensitivity of food webs to multiple stressors. *Ecology Letters*, 00, 1 – 19

Beaumelle, L., Thouvenot, L., Hines, J., Jochum, M., Eisenhauer, N., & Phillips, H. R. (2021). Soil fauna diversity and chemical stressors: a review of knowledge gaps and roadmap for future research. *Ecography*, 44(6), 845-859.

Belarde, T. A., & Railsback, S. F. (2016). New predictions from old theory: Emergent effects of multiple stressors in a model of piscivorous fish. *Ecological Modelling*, 326, 54-62.

Bender, R., Bunce, C., Clarke, M., Gates, S., Lange, S., Pace, N. L., & Thorlund, K. (2008). Attention should be given to multiplicity issues in systematic reviews. *Journal of Clinical Epidemiology*, 61(9), 857-865.

Bernhardt, J. R., & O'Connor, M. I. (2021). Aquatic biodiversity enhances multiple nutritional benefits to humans. *Proceedings of the National Academy of Sciences*, 118(15).

Besson, M., Feeney, W. E., Moniz, I., François, L., Brooker, R. M., Holzer, G., ... & Lecchini, D. (2020). Anthropogenic stressors impact fish sensory development and survival via thyroid disruption. *Nature Communications*, 11(1), 1-10.

Best, J. (2019). Anthropogenic stresses on the world's big rivers. *Nature Geoscience*, 12(1), 7-21.

Birk, S., Chapman, D., Carvalho, L., Spears, B. M., Andersen, H. E., Argillier, C., ... & Hering, D. (2020). Impacts of multiple stressors on freshwater biota across spatial scales and ecosystems. *Nature Ecology & Evolution*, 4(8), 1060-1068.

Bliss, C. I. (1939). The toxicity of poisons applied jointly 1. *Annals of Applied Biology*, 26(3), 585-615.

Blowes, S. A., Supp, S. R., Antão, L. H., Bates, A., Bruelheide, H., Chase, J. M., ... & Dornelas, M. (2019). The geography of biodiversity change in marine and terrestrial assemblages. *Science*, 366(6463), 339-345.

Borenstein, M., Cooper, H., Hedges, L. and Valentine, J. (2009). Effect sizes for continuous data. *The Handbook of Research Synthesis and Meta-Analysis*, 2, 221-235.

Boyd, P. W., Collins, S., Dupont, S., Fabricius, K., Gattuso, J. P., Havenhand, J., ... & Pörtner, H. O. (2018). Experimental strategies to assess the biological ramifications of multiple drivers of global ocean change—a review. *Global Change Biology*, 24(6), 2239-2261.

Bracewell, S., Verdonschot, R. C., Schäfer, R. B., Bush, A., Lapen, D. R., & Van den Brink, P. J. (2019). Qualifying the effects of single and multiple stressors on the food web structure of Dutch drainage ditches using a literature review and conceptual models. *Science of the Total Environment*, 684, 727-740.

Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., ... & Zhang, J. (2018). Declining oxygen in the global ocean and coastal waters. *Science*, 359(6371).

Brook, B. W., Sodhi, N. S., & Bradshaw, C. J. (2008). Synergies among extinction drivers under global change. *Trends in Ecology & Evolution*, 23(8), 453-460.

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., & West, G. B. (2004). Toward a metabolic theory of ecology. *Ecology*, 85(7), 1771-1789.

Brown, C. J., Saunders, M. I., Possingham, H. P., & Richardson, A. J. (2013). Managing for interactions between local and global stressors of ecosystems. *PLoS One*, 8(6), e65765.

Brown, C. J., Saunders, M. I., Possingham, H. P., & Richardson, A. J. (2014). Interactions between global and local stressors of ecosystems determine management effectiveness in cumulative impact mapping. *Diversity and Distributions*, 20(5), 538-546.

Bruder, A., Frainer, A., Rota, T., & Primicerio, R. (2019). The importance of ecological networks in multiple-stressor research and management. *Frontiers in Environmental Science*, 7, 59.

Burgess, B. J., Purves, D., Mace, G., & Murrell, D. J. (2021). Classifying ecosystem stressor interactions: Theory highlights the data limitations of the additive null model and the difficulty in revealing ecological surprises. *Global Change Biology*.

Cambronero, M. C., Marshall, H., De Meester, L., Davidson, T. A., Beckerman, A. P., & Orsini, L. (2018). Predictability of the impact of multiple stressors on the keystone species Daphnia. *Scientific Reports*, 8(1), 1-11.

Christensen, V., & Walters, C. J. (2004). Ecopath with Ecosim: methods, capabilities and limitations. *Ecological Modelling*, 172(2-4), 109-139.

Christensen, M. R., Graham, M. D., Vinebrooke, R. D., Findlay, D. L., Paterson, M. J., & Turner, M. A. (2006). Multiple anthropogenic stressors cause ecological surprises in boreal lakes. *Global Change Biology*, 12(12), 2316-2322.

Collen, B., Whitton, F., Dyer, E. E., Baillie, J. E., Cumberlidge, N., Darwall, W. R., ... & Böhm, M. (2014). Global patterns of freshwater species diversity, threat and endemism. *Global Ecology and Biogeography*, 23(1), 40-51.

Comte, L., Carvajal-Quintero, J., Tedesco, P. A., Giam, X., Brose, U., Erős, T., ... & Olden, J. D. (2021a). RivFishTIME: A global database of fish time-series to study global change ecology in riverine systems. *Global Ecology and Biogeography*, 30(1), 38-50.

Comte, L., Olden, J. D., Tedesco, P. A., Ruhi, A., & Giam, X. (2021b). Climate and land-use changes interact to drive long-term reorganization of riverine fish communities globally. *Proceedings of the National Academy of Sciences*, 118(27).

Costello, C., Cao, L., Gelcich, S., Cisneros-Mata, M. Á., Free, C. M., Froehlich, H. E., ... & Lubchenco, J. (2020). The future of food from the sea. *Nature*, 588(7836), 95-100.

Côté, I. M., Darling, E. S., & Brown, C. J. (2016). Interactions among ecosystem stressors and their importance in conservation. *Proceedings of the Royal Society B: Biological Sciences*, 283(1824), 20152592.

Crain, C. M., Kroeker, K., & Halpern, B. S. (2008). Interactive and cumulative effects of multiple human stressors in marine systems. *Ecology Letters*, 11(12), 1304-1315.

Darling, E. S., & Côté, I. M. (2008). Quantifying the evidence for ecological synergies. *Ecology Letters*, 11(12), 1278-1286.

Davis, S. J., Mellander, P. E., Kelly, A. M., Matthaei, C. D., Piggott, J. J., & Kelly-Quinn, M. (2018). Multiple-stressor effects of sediment, phosphorus and nitrogen on stream macroinvertebrate communities. *Science of the Total Environment*, 637, 577-587.

De Laender, F., Rohr, J. R., Ashauer, R., Baird, D. J., Berger, U., Eisenhauer, N., ... & Van den Brink, P. J. (2016). Reintroducing environmental change drivers in biodiversity–ecosystem functioning research. *Trends in Ecology & Evolution*, 31(12), 905-915.

De Laender, F. (2018). Community-and ecosystem-level effects of multiple environmental change drivers: Beyond null model testing. *Global Change Biology*, 24(11), 5021-5030.

Dey, C. J., & Koops, M. A. (2021). The consequences of null model selection for predicting mortality from multiple stressors. *Proceedings of the Royal Society B*, 288(1948), 20203126.

Diagne, C., Leroy, B., Vaissière, A. C., Gozlan, R. E., Roiz, D., Jarić, I., ... & Courchamp, F. (2021). High and rising economic costs of biological invasions worldwide. *Nature*, 592(7855), 571-576.

Didham, R. K., Tylianakis, J. M., Gemmell, N. J., Rand, T. A., & Ewers, R. M. (2007). Interactive effects of habitat modification and species invasion on native species decline. *Trends in Ecology & Evolution*, 22(9), 489-496.

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J., & Collen, B. (2014). Defaunation in the Anthropocene. *Science*, 345(6195), 401-406.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z. I., Knowler, D. J., Lévêque, C., ... & Sullivan, C. A. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews*, 81(2), 163-182.

Dudgeon, D. (2019). Multiple threats imperil freshwater biodiversity in the Anthropocene. *Current Biology*, 29(19), R960-R967.

EEA, (2018). *European waters Assessment of status and pressures 2018*. EEA Report, No 7/2018, 1–90.

Egea-Serrano, A., Relyea, R. A., Tejedo, M., & Torralva, M. (2012). Understanding of the impact of chemicals on amphibians: a meta-analytic review. *Ecology and Evolution*, 2(7), 1382-1397.

Enochs, I. C., Manzello, D. P., Donham, E. M., Kolodziej, G., Okano, R., Johnston, L., ... & Price, N. N. (2015). Shift from coral to macroalgae dominance on a volcanically acidified reef. *Nature Climate Change*, 5(12), 1083-1088.

Erős, T., Comte, L., Filipe, A. F., Ruhi, A., Tedesco, P. A., Brose, U., ... & Olden, J. D. (2020). Effects of nonnative species on the stability of riverine fish communities. *Ecography*, 43(8), 1156-1166.

Falkenberg, L. J., Connell, S. D., & Russell, B. D. (2013). Disrupting the effects of synergies between stressors: improved water quality dampens the effects of future CO 2 on a marine habitat. *Journal of Applied Ecology*, 50(1), 51-58.

Fey, S. B., Siepielski, A. M., Nusslé, S., Cervantes-Yoshida, K., Hwan, J. L., Huber, E. R., ... & Carlson, S. M. (2015). Recent shifts in the occurrence, cause, and magnitude of animal mass mortality events. *Proceedings of the National Academy of Sciences*, 112(4), 1083-1088.

Field, S. A., Tyre, A. J., Jonzén, N., Rhodes, J. R., & Possingham, H. P. (2004). Minimizing the cost of environmental management decisions by optimizing statistical thresholds. *Ecology Letters*, 7(8), 669-675.

Fleishman, E., Blockstein, D. E., Hall, J. A., Mascia, M. B., Rudd, M. A., Scott, J. M., ... & Vedder, A. (2011). Top 40 priorities for science to inform US conservation and management policy. *BioScience*, 61(4), 290-300.

Fisher, M. C., & Garner, T. W. (2020). Chytrid fungi and global amphibian declines. *Nature Reviews Microbiology*, 18(6), 332-343.

Folt, C. L., Chen, C. Y., Moore, M. V., & Burnaford, J. (1999). Synergism and antagonism among multiple stressors. *Limnology and Oceanography*, 44(3part2), 864-877.

Fournier, V., Rosenheim, J. A., Brodeur, J., Diez, J. M., & Johnson, M. W. (2006). Multiple plant exploiters on a shared host: testing for nonadditive effects on plant performance. *Ecological Applications*, 16(6), 2382-2398.

França, F. M., Benkwitt, C. E., Peralta, G., Robinson, J. P., Graham, N. A., Tylianakis, J. M., ... & Barlow, J. (2020). Climatic and local stressor interactions threaten tropical forests and coral reefs. *Philosophical Transactions of the Royal Society B*, 375(1794), 20190116.

Fronhofer, E. A., Legrand, D., Altermatt, F., Ansart, A., Blanchet, S., Bonte, D., ... & Cote, J. (2018). Bottom-up and top-down control of dispersal across major organismal groups. *Nature Ecology & Evolution*, 2(12), 1859-1863.

Fu, C., Travers-Trolet, M., Velez, L., Grüss, A., Bundy, A., Shannon, L. J., ... & Shin, Y. J. (2018). Risky business: the combined effects of fishing and changes in primary productivity on fish communities. *Ecological Modelling*, 368, 265-276.

Galic, N., Grimm, V., & Forbes, V. E. (2017). Impaired ecosystem process despite little effects on populations: modeling combined effects of warming and toxicants. *Global Change Biology*, 23(8), 2973-2989.

Galic, N., Sullivan, L. L., Grimm, V., & Forbes, V. E. (2018). When things don't add up: quantifying impacts of multiple stressors from individual metabolism to ecosystem processing. *Ecology Letters*, 21(4), 568-577.

Gallardo, B., Clavero, M., Sánchez, M. I., & Vilà, M. (2016). Global ecological impacts of invasive species in aquatic ecosystems. *Global Change Biology*, 22(1), 151-163.

Geldmann, J., Joppa, L. N., & Burgess, N. D. (2014). Mapping change in human pressure globally on land and within protected areas. *Conservation Biology*, 28(6), 1604-1616.

Gillooly, J. F. (2000). Effect of body size and temperature on generation time in zooplankton. *Journal of Plankton Research*, 22(2), 241-251.

Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M., & Charnov, E. L. (2001). Effects of size and temperature on metabolic rate. *Science*, 293(5538), 2248-2251.

Gomez Isaza, D. F., Cramp, R. L., & Franklin, C. E. (2020). Living in polluted waters: A meta-analysis of the effects of nitrate and interactions with other environmental stressors on freshwater taxa. *Environmental Pollution*, 261, 114091.

Greco, W., Unkelbach, H. D., Pöch, G., Sühnel, J., Kundi, M., & Bödeker, W. (1992). Consensus on concepts and terminology for combined-action assessment: the Saariselkä agreement. *Archives of Complex Environmental Studies*, 4(3), 65-69.

Greenland, S. (1983). Tests for interaction in epidemiologic studies: a review and a study of power. *Statistics in Medicine*, 2(2), 243-251.

Griffen, B. D., Belgrad, B. A., Cannizzo, Z. J., Knotts, E. R., & Hancock, E. R. (2016). Rethinking our approach to multiple stressor studies in marine environments. *Marine Ecology Progress Series*, 543, 273-281.

Griffith, G. P., Fulton, E. A., Gorton, R., & Richardson, A. J. (2012). Predicting interactions among fishing, ocean warming, and ocean acidification in a marine system with whole-ecosystem models. *Conservation Biology*, 26(6), 1145-1152.

Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., ... & Zarfl, C. (2019). Mapping the world's free-flowing rivers. *Nature*, 569(7755), 215-221.

Gruner, D. S., Smith, J. E., Seabloom, E. W., Sandin, S. A., Ngai, J. T., Hillebrand, H., ... & Bolker, B. M. (2008). A cross-system synthesis of consumer and nutrient resource control on producer biomass. *Ecology Letters*, 11(7), 740-755.

Gunderson, A. R., Armstrong, E. J., & Stillman, J. H. (2016). Multiple stressors in a changing world: the need for an improved perspective on physiological responses to the dynamic marine environment. *Annual Review of Marine Science*, 8, 357-378.

Gurevitch, J., Morrison, J. A., & Hedges, L. V. (2000). The interaction between competition and predation: a meta-analysis of field experiments. *The American Naturalist*, 155(4), 435-453.

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, 555(7695), 175-182.

Haller-Bull, V., & Bode, M. (2019). Superadditive and subadditive dynamics are not inherent to the types of interacting threat. *PLoS One*, 14(8), e0211444.

Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'Agrosa, C., ... & Watson, R. (2008). A global map of human impact on marine ecosystems. *Science*, 319(5865), 948-952.

Halpern, B. S., Frazier, M., Potapenko, J., Casey, K. S., Koenig, K., Longo, C., ... & Walbridge, S. (2015). Spatial and temporal changes in cumulative human impacts on the world's ocean. *Nature Communications*, 6(1), 1-7.

Halpern, B. S., Frazier, M., Afflerbach, J., Lowndes, J. S., Micheli, F., O'Hara, C., ... & Selkoe, K. A. (2019). Recent pace of change in human impact on the world's ocean. *Scientific Reports*, 9(1), 1-8.

Harfoot, M. B., Newbold, T., Tittensor, D. P., Emmott, S., Hutton, J., Lyutsarev, V., ... & Purves, D. W. (2014). Emergent global patterns of ecosystem structure and function from a mechanistic general ecosystem model. *PLoS Biology*, 12(4), e1001841.

Harvey, B. P., Gwynn-Jones, D., & Moore, P. J. (2013). Meta-analysis reveals complex marine biological responses to the interactive effects of ocean acidification and warming. *Ecology and Evolution*, 3(4), 1016-1030.

Harvey, E., Gounand, I., Fronhofer, E. A., & Altermatt, F. (2018). Disturbance reverses classic biodiversity predictions in river-like landscapes. *Proceedings of the Royal Society B*, 285(1893), 20182441.

Hawkes, C. V., & Sullivan, J. J. (2001). The impact of herbivory on plants in different resource conditions: a meta-analysis. *Ecology*, 82(7), 2045-2058.

He, F., Zarfl, C., Bremerich, V., David, J. N., Hogan, Z., Kalinkat, G., ... & Jähnig, S. C. (2019). The global decline of freshwater megafauna. *Global Change Biology*, 25(11), 3883-3892.

Heath, M. R., Speirs, D. C., & Steele, J. H. (2014). Understanding patterns and processes in models of trophic cascades. *Ecology Letters*, 17(1), 101-114.

Hecky, R. E., Mugidde, R., Ramlal, P. S., Talbot, M. R., & Kling, G. W. (2010). Multiple stressors cause rapid ecosystem change in Lake Victoria. *Freshwater Biology*, 55, 19-42.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. *Academic press, New York*.

Heino, J., Virkkala, R., & Toivonen, H. (2009). Climate change and freshwater biodiversity: detected patterns, future trends and adaptations in northern regions. *Biological Reviews*, 84(1), 39-54.

Higdon R. (2013) Experimental Design, Variability. In: Dubitzky W., Wolkenhauer O., Cho KH., Yokota H. (eds) Encyclopedia of Systems Biology. *Springer*, New York, NY.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557-560.

Hillebrand, H. (2002). Top-down versus bottom-up control of autotrophic biomass—a meta-analysis on experiments with periphyton. *Journal of the North American Benthological Society*, 21(3), 349-369.

Hillebrand, H., Donohue, I., Harpole, W. S., Hodapp, D., Kucera, M., Lewandowska, A. M., ... & Freund, J. A. (2020). Thresholds for ecological responses to global change do not emerge from empirical data. *Nature Ecology & Evolution*, 4(11), 1502-1509.

Hodgson, E. E., Essington, T. E., & Halpern, B. S. (2017). Density dependence governs when population responses to multiple stressors are magnified or mitigated. *Ecology*, 98(10), 2673-2683.

Hodgson, E. E., & Halpern, B. S. (2019). Investigating cumulative effects across ecological scales. *Conservation Biology*, 33(1), 22-32.

Hungate, B. A., Van Groenigen, K. J., Six, J., Jastrow, J. D., Luo, Y., De Graaff, M. A., ... & Osenberg, C. W. (2009). Assessing the effect of elevated carbon dioxide on soil carbon: a comparison of four meta-analyses. *Global Change Biology*, 15(8), 2020-2034.

Ives, A. R., & Cardinale, B. J. (2004). Food-web interactions govern the resistance of communities after non-random extinctions. *Nature*, 429(6988), 174-177.

Jackson, M. C., Loewen, C. J., Vinebrooke, R. D., & Chimimba, C. T. (2016). Net effects of multiple stressors in freshwater ecosystems: a meta-analysis. *Global Change Biology*, 22(1), 180-189.

Jackson, M. C., Fourie, H. E., Dalu, T., Woodford, D. J., Wasserman, R. J., Zengeya, T. A., ... & Weyl, O. L. (2020). Food web properties vary with climate and land use in South African streams. *Functional Ecology*, 34(8), 1653-1665.

Jackson, M. C., Pawar, S., & Woodward, G. (2021). The temporal dynamics of multiple stressor effects: From individuals to ecosystems. *Trends in Ecology & Evolution*.

Johansen, J. L., Allan, B. J., Rummer, J. L., & Esbaugh, A. J. (2017). Oil exposure disrupts early life-history stages of coral reef fishes via behavioural impairments. *Nature Ecology & Evolution*, 1(8), 1146-1152.

Johnson, P. T., & Paull, S. H. (2011). The ecology and emergence of diseases in fresh waters. *Freshwater Biology*, 56(4), 638-657.

Johnson, P. C., Barry, S. J., Ferguson, H. M., & Müller, P. (2015). Power analysis for generalized linear mixed models in ecology and evolution. *Methods in Ecology and Evolution*, 6(2), 133-142.

Juvigny-Khenafou, N. P., Piggott, J. J., Atkinson, D., Zhang, Y., Macaulay, S. J., Wu, N., & Matthaei, C. D. (2021). Impacts of multiple anthropogenic stressors on stream macroinvertebrate community composition and functional diversity. *Ecology and Evolution*, 11(1), 133-152.

Kath, J., Thomson, J. R., Thompson, R. M., Kefford, B. J., Dyer, F. J., & Mac Nally, R. (2018). Interactions among stressors may be weak: implications for management of freshwater macroinvertebrate communities. *Diversity and Distributions*, 24(7), 939-950.

Kerkhoff, A. J., & Enquist, B. J. (2009). Multiplicative by nature: why logarithmic transformation is necessary in allometry. *Journal of Theoretical Biology*, 257(3), 519-521.

Knouft, J. H., & Ficklin, D. L. (2017). The potential impacts of climate change on biodiversity in flowing freshwater systems. *Annual Review of Ecology, Evolution, and Systematics*, 48, 111-133.

Korell, L., Auge, H., Chase, J. M., Harpole, S., & Knight, T. M. (2020). We need more realistic climate change experiments for understanding ecosystems of the future. *Global Change Biology*, 26(2), 325-327.

Kratina, P., Greig, H. S., Thompson, P. L., Carvalho-Pereira, T. S., & Shurin, J. B. (2012). Warming modifies trophic cascades and eutrophication in experimental freshwater communities. *Ecology*, 93(6), 1421-1430.

Kroeker, K. J., Kordas, R. L., Crim, R., Hendriks, I. E., Ramajo, L., Singh, G. S., ... & Gattuso, J. P. (2013). Impacts of ocean acidification on marine organisms: quantifying sensitivities and interaction with warming. *Global Change Biology*, 19(6), 1884-1896.

Kroeker, K. J., Kordas, R. L., & Harley, C. D. (2017). Embracing interactions in ocean acidification research: confronting multiple stressor scenarios and context dependence. *Biology Letters*, 13(3), 20160802.

Koricheva, J., & Gurevitch, J. (2014). Uses and misuses of meta-analysis in plant ecology. *Journal of Ecology*, 102(4), 828-844.

Lajeunesse, M. J. (2011). On the meta-analysis of response ratios for studies with correlated and multi-group designs. *Ecology*, 92(11), 2049-2055.

Lakens, D. (2021). Sample size justification. *PsyArXiv*.

Lange, K., Bruder, A., Matthaei, C. D., Brodersen, J., & Paterson, R. A. (2018). Multiple-stressor effects on freshwater fish: Importance of taxonomy and life stage. *Fish and Fisheries*, 19(6), 974-983.

Lehner, B., & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, 296(1-4), 1-22.

Lenihan, H. S., Micheli, F., Shelton, S. W., & Peterson, C. H. (1999). The influence of multiple environmental stressors on susceptibility to parasites: an experimental determination with oysters. *Limnology and Oceanography*, 44(3part2), 910-924.

Lewis, S. L., & Maslin, M. A. (2015). Defining the anthropocene. *Nature*, 519(7542), 171-180.

Li, F., Altermatt, F., Yang, J., An, S., Li, A., & Zhang, X. (2020). Human activities' fingerprint on multitrophic biodiversity and ecosystem functions across a major river catchment in China. *Global Change Biology*, 26(12), 6867-6879.

Lindenmayer, D. B., Foster, C. N., Westgate, M. J., Scheele, B. C., & Blanchard, W. (2020). Managing interacting disturbances: Lessons from a case study in Australian forests. *Journal of Applied Ecology*, 57(9), 1711-1716.

Mace, G. M. (2014). Whose conservation? *Science*, 345(6204), 1558-1560.

Malaj, E., Peter, C., Grote, M., Kühne, R., Mondy, C. P., Usseglio-Polatera, P., ... & Schäfer, R. B. (2014). Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. *Proceedings of the National Academy of Sciences*, 111(26), 9549-9554.

Maltby, L. (1999). Studying stress: The importance of organism-level responses. *Ecological Applications*, 9(2), 431-440.

Marshall, K. N., Kaplan, I. C., Hodgson, E. E., Hermann, A., Busch, D. S., McElhany, P., ... & Fulton, E. A. (2017). Risks of ocean acidification in the California Current food web and fisheries: ecosystem model projections. *Global Change Biology*, 23(4), 1525-1539.

Matthaei, C. D., Piggott, J. J., & Townsend, C. R. (2010). Multiple stressors in agricultural streams: interactions among sediment addition, nutrient enrichment and water abstraction. *Journal of Applied Ecology*, 47(3), 639-649.

Matthaei, C. D., & Lange, K. (2016). Multiple-stressor effects on freshwater fish: a review and meta-analysis. *Conservation of Freshwater Fishes*, 178-214.

Mebane, C. A., Schmidt, T. S., & Balistrieri, L. S. (2017). Larval aquatic insect responses to cadmium and zinc in experimental streams. *Environmental Toxicology and Chemistry*, 36(3), 749-762.

Morris, W. F., Hufbauer, R. A., Agrawal, A. A., Bever, J. D., Borowicz, V. A., Gilbert, G. S., ... & Vázquez, D. P. (2007). Direct and interactive effects of enemies and mutualists on plant performance: a meta-analysis. *Ecology*, 88(4), 1021-1029.

Mudge, J. F., Baker, L. F., Edge, C. B., & Houlahan, J. E. (2012). Setting an optimal α that minimizes errors in null hypothesis significance tests. *PLoS One*, 7(2), e32734.

Muyssen, B. T., Messiaen, M., & Janssen, C. R. (2010). Combined cadmium and temperature acclimation in Daphnia magna: physiological and sub-cellular effects. *Ecotoxicology and Environmental Safety*, 73(5), 735-742.

Nakagawa, S., Noble, D. W., Senior, A. M., & Lagisz, M. (2017). Meta-evaluation of meta-analysis: ten appraisal questions for biologists. *BMC Biology*, 15(1), 1-14.

Newbold, T., Hudson, L. N., Hill, S. L., Contu, S., Lysenko, I., Senior, R. A., ... & Purvis, A. (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, 520(7545), 45-50.

O'Brien, A. L., Dafforn, K. A., Chariton, A. A., Johnston, E. L., & Mayer-Pinto, M. (2019). After decades of stressor research in urban estuarine ecosystems the focus is still on single stressors: a systematic literature review and meta-analysis. *Science of the Total Environment*, 684, 753-764.

O'Dea, R. E., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W., Parker, T. H., ... & Nakagawa, S. (2021). Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. *Biological Reviews*.

O’Gorman, E. J., Zhao, L., Pichler, D. E., Adams, G., Friberg, N., Rall, B. C., ... & Woodward, G. (2017). Unexpected changes in community size structure in a natural warming experiment. *Nature Climate Change*, 7(9), 659-663.

O’Gorman, E. J., Petchey, O. L., Faulkner, K. J., Gallo, B., Gordon, T. A., Neto-Cerejeira, J., ... & Woodward, G. (2019). A simple model predicts how warming simplifies wild food webs. *Nature Climate Change*, 9(8), 611-616.

O’Hara, C. C., Frazier, M., & Halpern, B. S. (2021). At-risk marine biodiversity faces extensive, expanding, and intensifying human impacts. *Science*, 372(6537), 84-87.

Ormerod, S. J., Dobson, M., Hildrew, A. G., & Townsend, C. (2010). Multiple stressors in freshwater ecosystems. *Freshwater Biology*, 55, 1-4.

Orr, J. A., Vinebrooke, R. D., Jackson, M. C., Kroeker, K. J., Kordas, R. L., Mantyka-Pringle, C., ... & Piggott, J. J. (2020). Towards a unified study of multiple stressors: divisions and common goals across research disciplines. *Proceedings of the Royal Society B*, 287(1926), 20200421.

Orr, J. A., Piggott, J. J., Jackson, A. L., & Arnoldi, J. F. (2021a). Scaling up uncertain predictions to higher levels of organisation tends to underestimate change. *Methods in Ecology and Evolution*.

Orr, J. A., Luijckx, P., Arnoldi, J. F., Jackson, A. L., & Piggott, J. J. (2021b). Rapid evolution generates synergism between multiple stressors: Linking theory and an evolution experiment. *Global Change Biology*.

Paine, R. T., Tegner, M. J., & Johnson, E. A. (1998). Compounded perturbations yield ecological surprises. *Ecosystems*, 1(6), 535-545.

Piggott, J. J., Lange, K., Townsend, C. R., & Matthaei, C. D. (2012). Multiple stressors in agricultural streams: a mesocosm study of interactions among raised water temperature, sediment addition and nutrient enrichment. *PLoS One*, 7(11), e49873.

Piggott, J. J., Townsend, C. R., & Matthaei, C. D. (2015). Reconceptualizing synergism and antagonism among multiple stressors. *Ecology and Evolution*, 5(7), 1538-1547.

Polazzo, F., & Rico, A. (2021). Effects of multiple stressors on the dimensionality of ecological stability. *Ecology Letters*.

Przeslawski, R., Byrne, M., & Mellin, C. (2015). A review and meta-analysis of the effects of multiple abiotic stressors on marine embryos and larvae. *Global Change Biology*, 21(6), 2122-2140.

Purves, D., Scharlemann, J. P., Harfoot, M., Newbold, T., Tittensor, D. P., Hutton, J., & Emmott, S. (2013). Time to model all life on Earth. *Nature*, 493(7432), 295-297.

Rader, R., Bartomeus, I., Garibaldi, L. A., Garratt, M. P., Howlett, B. G., Winfree, R., ... & Woyciechowski, M. (2016). Non-bee insects are important contributors to global crop pollination. *Proceedings of the National Academy of Sciences*, 113(1), 146-151.

Reeves, M. K., Jensen, P., Dolph, C. L., Holyoak, M., & Trust, K. A. (2010). Multiple stressors and the cause of amphibian abnormalities. *Ecological Monographs*, 80(3), 423-440.

Reid, A. J., Carlson, A. K., Creed, I. F., Eliason, E. J., Gell, P. A., Johnson, P. T., ... & Cooke, S. J. (2019). Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews*, 94(3), 849-873.

Reidy Liermann, C., Nilsson, C., Robertson, J., & Ng, R. Y. (2012). Implications of dam obstruction for global freshwater fish diversity. *BioScience*, 62(6), 539-548.

Richardson, J., Feuchtmayr, H., Miller, C., Hunter, P. D., Maberly, S. C., & Carvalho, L. (2019). Response of cyanobacteria and phytoplankton abundance to warming, extreme rainfall events and nutrient enrichment. *Global Change Biology*, 25(10), 3365-3380.

Roccuzzo, S., Beckerman, A. P., & Trögl, J. (2021). New perspectives on the bioremediation of endocrine disrupting compounds from wastewater using algae-, bacteria-and fungi-based technologies. *International Journal of Environmental Science and Technology*, 18(1), 89-106.

Rodgers, E. M., & Gomez Isaza, D. F. (2021). Harnessing the potential of cross-protection stressor interactions for conservation: a review. *Conservation Physiology*, 9(1), coab037.

Rosenblatt, A. E., & Schmitz, O. J. (2014). Interactive effects of multiple climate change variables on trophic interactions: a meta-analysis. *Climate Change Responses*, 1(1), 1-10.

Rudd, M. A. (2014). Scientists' perspectives on global ocean research priorities. *Frontiers in Marine Science*, 1, 36.

Ryo, M., Harvey, E., Robinson, C. T., & Altermatt, F. (2018). Nonlinear higher order abiotic interactions explain riverine biodiversity. *Journal of Biogeography*, 45(3), 628-639.

Sala, O. E., Chapin, F. S., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., ... & Wall, D. H. (2000). Global biodiversity scenarios for the year 2100. *Science*, 287(5459), 1770-1774.

Sala, E., Mayorga, J., Bradley, D., Cabral, R. B., Atwood, T. B., Auber, A., ... & Lubchenco, J. (2021). Protecting the global ocean for biodiversity, food and climate. *Nature*, 592(7854), 397-402.

Schäfer, R. B., & Piggott, J. J. (2018). Advancing understanding and prediction in multiple stressor research through a mechanistic basis for null models. *Global Change Biology*, 24(5), 1817-1826.

Schrama, M., Barmentlo, S. H., Hunting, E. R., van Logtestijn, R. S., Vijver, M. G., & van Bodegom, P. M. (2017). Pressure-induced shifts in trophic linkages in a simplified aquatic food web. *Frontiers in Environmental Science*, 5, 75.

Scheffers, B. R., De Meester, L., Bridge, T. C., Hoffmann, A. A., Pandolfi, J. M., Corlett, R. T., ... & Watson, J. E. (2016). The broad footprint of climate change from genes to biomes to people. *Science*, 354(6313).

Schmitz, O. J. (2007). Predator diversity and trophic interactions. *Ecology*, 88(10), 2415-2426.

Segner, H., Schmitt-Jansen, M., & Sabater, S. (2014). Assessing the impact of multiple stressors on aquatic biota: The receptor’s side matters. *Environmental Science & Technology*, 48, 7690–7696.

Segre, D., DeLuna, A., Church, G. M., & Kishony, R. (2005). Modular epistasis in yeast metabolism. *Nature Genetics*, 37(1), 77-83.

Seibold, S., Cadotte, M. W., MacIvor, J. S., Thorn, S., & Müller, J. (2018). The necessity of multitrophic approaches in community ecology. *Trends in Ecology & Evolution*, 33(10), 754-764.

Seifert, M., Rost, B., Trimborn, S., & Hauck, J. (2020). Meta-analysis of multiple driver effects on marine phytoplankton highlights modulating role of pCO2. *Global Change Biology*, 26(12), 6787-6804.

Senior, A. M., Grueber, C. E., Kamiya, T., Lagisz, M., O'Dwyer, K., Santos, E. S., & Nakagawa, S. (2016). Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. *Ecology*, 97(12), 3293-3299.

Sentis, A., Montoya, J. M., & Lurgi, M. (2021). Warming indirectly increases invasion success in food webs. *Proceedings of the Royal Society B*, 288(1947), 20202622.

Shears, N. T., & Ross, P. M. (2010). Toxic cascades: multiple anthropogenic stressors have complex and unanticipated interactive effects on temperate reefs. *Ecology Letters*, 13(9), 1149-1159.

Shrimpton, J. M., Zydlewski, J. D., & Heath, J. W. (2007). Effect of daily oscillation in temperature and increased suspended sediment on growth and smolting in juvenile chinook salmon, Oncorhynchus tshawytscha. *Aquaculture*, 273(2-3), 269-276.

Sih, A., Englund, G., & Wooster, D. (1998). Emergent impacts of multiple predators on prey. *Trends in Ecology & Evolution*, 13(9), 350-355.

Sih, A., Bell, A. M., & Kerby, J. L. (2004). Two stressors are far deadlier than one. *Trends in Ecology & Evolution*, 19(6), 274-276.

Simmonds, M. (2015). Quantifying the risk of error when interpreting funnel plots. *Systematic Reviews*, 4(1), 1-7.

Sinclair, B. J., Ferguson, L. V., Salehipour-Shirazi, G., & MacMillan, H. A. (2013). Cross-tolerance and cross-talk in the cold: relating low temperatures to desiccation and immune stress in insects. *Integrative and Comparative Biology*, 53(4), 545-556.

Siviter, H., Bailes, E. J., Martin, C. D., Oliver, T. R., Koricheva, J., Leadbeater, E., & Brown, M. J. (2021). Agrochemicals interact synergistically to increase bee mortality. *Nature*, 1-4.

Smale, D. A., Wernberg, T., Oliver, E. C., Thomsen, M., Harvey, B. P., Straub, S. C., ... & Moore, P. J. (2019). Marine heatwaves threaten global biodiversity and the provision of ecosystem services. *Nature Climate Change*, 9(4), 306-312.

Sokolova, I. (2021). Bioenergetics in environmental adaptation and stress tolerance of aquatic ectotherms: linking physiology and ecology in a multi-stressor landscape. *Journal of Experimental Biology*, 224(Suppl_1), jeb236802.

Soluk, D. A., & Collins, N. C. (1988). Synergistic interactions between fish and stoneflies: facilitation and interference among stream predators. *Oikos*, 94-100.

Spears, B. M., Chapman, D. S., Carvalho, L., Feld, C. K., Gessner, M. O., Piggott, J. J., ... & Birk, S. (2021). Making Waves. Bridging theory and practice towards multiple stressor management in freshwater ecosystems. *Water Research*, 116981.

Steidl, R. J., Hayes, J. P., & Schauber, E. (1997). Statistical power analysis in wildlife research. The *Journal of Wildlife Management*, 270-279.

Stephens, A. E., Srivastava, D. S., & Myers, J. H. (2013). Strength in numbers? Effects of multiple natural enemy species on plant performance. *Proceedings of the Royal Society B: Biological Sciences*, 280(1760), 20122756.

Sterner, R. W., Keeler, B., Polasky, S., Poudel, R., Rhude, K., & Rogers, M. (2020). Ecosystem services of Earth’s largest freshwater lakes. *Ecosystem Services*, 41, 101046.

Stewart, R. I., Dossena, M., Bohan, D. A., Jeppesen, E., Kordas, R. L., Ledger, M. E., ... & Woodward, G. (2013). Mesocosm experiments as a tool for ecological climate-change research. *Advances in Ecological Research*, 48, 71-181.

Stockbridge, J., Jones, A. R., & Gillanders, B. M. (2020). A meta-analysis of multiple stressors on seagrasses in the context of marine spatial cumulative impacts assessment. *Scientific Reports*, 10(1), 1-11.

Strain, E. M., Thomson, R. J., Micheli, F., Mancuso, F. P., & Airoldi, L. (2014). Identifying the interacting roles of stressors in driving the global loss of canopy-forming to mat-forming algae in marine ecosystems. *Global Change Biology*, 20(11), 3300-3312.

Su, G., Logez, M., Xu, J., Tao, S., Villéger, S., & Brosse, S. (2021). Human impacts on global freshwater fish biodiversity. *Science*, 371(6531), 835-838.

Tang, J., Wennerberg, K., & Aittokallio, T. (2015). What is synergy? The Saariselkä agreement revisited. *Frontiers in Pharmacology*, 6, 181.

Tekin, E., Diamant, E. S., Cruz-Loya, M., Enriquez, V., Singh, N., Savage, V. M., & Yeh, P. J. (2020). Using a newly introduced framework to measure ecological stressor interactions. *Ecology Letters*, 23(9), 1391-1403.

Thompson, P. L., MacLennan, M. M., & Vinebrooke, R. D. (2018a). Species interactions cause non-additive effects of multiple environmental stressors on communities. *Ecosphere*, 9(11), e02518.

Thompson, P. L., MacLennan, M. M., & Vinebrooke, R. D. (2018b). An improved null model for assessing the net effects of multiple stressors on communities. *Global Change Biology*, 24(1), 517-525.

Tickner, D., Opperman, J. J., Abell, R., Acreman, M., Arthington, A. H., Bunn, S. E., ... & Young, L. (2020). Bending the curve of global freshwater biodiversity loss: an emergency recovery plan. *BioScience*, 70(4), 330-342.

Tittensor, D. P., Walpole, M., Hill, S. L., Boyce, D. G., Britten, G. L., Burgess, N. D., ... & Ye, Y. (2014). A mid-term analysis of progress toward international biodiversity targets. *Science*, 346(6206), 241-244.

Todgham, A. E., & Stillman, J. H. (2013). Physiological responses to shifts in multiple environmental stressors: relevance in a changing world. *Integrative and Comparative Biology*, 53(4), 539-544.

Townsend, C. R., Uhlmann, S. S., & Matthaei, C. D. (2008). Individual and combined responses of stream ecosystems to multiple stressors. *Journal of Applied Ecology*, 45(6), 1810-1819.

Travers-Trolet, M., Shin, Y. J., Shannon, L. J., Moloney, C. L., & Field, J. G. (2014). Combined fishing and climate forcing in the southern Benguela upwelling ecosystem: an end-to-end modelling approach reveals dampened effects. *PLoS One*, 9(4), e94286.

Trisos, C. H., Merow, C., & Pigot, A. L. (2020). The projected timing of abrupt ecological disruption from climate change. *Nature*, 580(7804), 496-501.

Velasco, J., Gutiérrez-Cánovas, C., Botella-Cruz, M., Sánchez-Fernández, D., Arribas, P., Carbonell, J. A., ... & Pallarés, S. (2019). Effects of salinity changes on aquatic organisms in a multiple stressor context. *Philosophical Transactions of the Royal Society B*, 374(1764), 20180011.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48.

Vinebrooke, R. D., Cottingham, K. L., Norberg, J., Scheffer, M., Dodson, S. I., Maberly, S. C., & Sommer, U. (2004). Impacts of multiple stressors on biodiversity and ecosystem functioning: The role of species co-tolerance. *Oikos*, 104(3), 451-457.

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., ... & Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature*, 467(7315), 555-561.

Wolfram Research. Inc. (2016). Mathematica (Version 10.4). *Wolfram Research, Inc., Champaign, Illinois.*

Woodward, G., Perkins, D. M., & Brown, L. E. (2010). Climate change and freshwater ecosystems: impacts across multiple levels of organization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1549), 2093-2106.

Woodward, G., Gessner, M. O., Giller, P. S., Gulis, V., Hladyz, S., Lecerf, A., ... & Chauvet, E. (2012). Continental-scale effects of nutrient pollution on stream ecosystem functioning. *Science*, 336(6087), 1438-1440.

Wu, Z., Dijkstra, P., Koch, G. W., Peñuelas, J., & Hungate, B. A. (2011). Responses of terrestrial ecosystems to temperature and precipitation change: A meta-analysis of experimental manipulation. *Global Change Biology*, 17(2), 927-942.

Wurtsbaugh, W. A., Paerl, H. W., & Dodds, W. K. (2019). Nutrients, eutrophication and harmful algal blooms along the freshwater to marine continuum. *Wiley Interdisciplinary Reviews: Water*, 6(5), e1373.

WWF. (2016). *Living Planet Report 2016 – Risk and resilience in a new era*. WWF International, Gland, Switzerland.

WWF. (2020). *Living Planet Report 2020 – Bending the curve of biodiversity loss*. Almond, R.E.A., Grooten M. and Petersen, T. (Eds). WWF, Gland, Switzerland.

Yeh, P., Tschumi, A. I., & Kishony, R. (2006). Functional classification of drugs by properties of their pairwise interactions. *Nature Genetics*, 38(4), 489-494.

Yue, K., Fornara, D. A., Yang, W., Peng, Y., Li, Z., Wu, F., & Peng, C. (2017). Effects of three global change drivers on terrestrial C: N: P stoichiometry: a global synthesis. *Global Change Biology*, 23(6), 2450-2463.

Yvon-Durocher, G., Montoya, J. M., Trimmer, M., & Woodward, G. (2011). Warming alters the size spectrum and shifts the distribution of biomass in freshwater ecosystems. *Global Change Biology*, 17(4), 1681-1694.

Zhao, S. Y., Sun, Y. P., Lin, Q. Q., & Han, B. P. (2013). Effects of silver carp (Hypophthalmichthys molitrix) and nutrients on the plankton community of a deep, tropical reservoir: an enclosure experiment. *Freshwater Biology*, 58(1), 100-113.

Zhou, L., Zhou, X., Shao, J., Nie, Y., He, Y., Jiang, L., ... & Hosseini Bai, S. (2016). Interactive effects of global change factors on soil respiration and its components: a meta-analysis. *Global Change Biology*, 22(9), 3157-3169.

Appendices

The following pages detail the appendices for Chapters One, Two, Three, and Four. For Appendices Two, Three, and Four, a table of contents is provided. Page numbers for each of these appendices are outlined below.

Appendix One

Web of Science was searched (22/06/2021) to determine the number of multiple stressor papers published per year which were either general or freshwater specific. The following search terms were used and based upon those described by Orr et al. (2020).

Overall number of multiple stressor papers

TI = (“multiple stress*” OR “stressors” OR “global change factors” OR “environmental factors” OR “global change drivers” OR “multiple drivers” OR “synerg*” OR “amplif*” OR “antagon*” OR “dampen*” OR “additive” OR “multiplicative” OR “reversal” OR “interactive effects” OR “multifactor” OR “nitrogen and phosphorus limitation” OR “multiple limiting resources” OR “nutrient co-limitation” OR “global change experiments” OR “counteract*”)

AND

TS = (“multiple stress*” OR “stressors” OR “global change factors” OR “global change drivers” OR “multifactor” OR “cumulative effect*” OR “net effect*” OR “combined effect*” OR “interacting” OR “nitrogen and phosphorus limitation” OR “multiple limiting resources” OR “nutrient co-limitation” OR “global change experiments”)

AND

WC = (“Ecology” OR “Toxicology” OR “Environmental Sciences” OR “Plant Sciences” OR “Zoology” OR “Marine Freshwater Biology” OR “Limnology” OR “Oceanography” OR “Multidisciplinary Sciences” OR “Plant Sciences” OR “Microbiology” OR “Biodiversity Conservation”)

AND

PY = (1900-2020)

Number of multiple stressor papers considering the freshwater realm

TI = (“multiple stress*” OR “stressors” OR “global change factors” OR “environmental factors” OR “global change drivers” OR “multiple drivers” OR “synerg*” OR “amplif*” OR “antagon*” OR “dampen*” OR “additive” OR “multiplicative” OR “reversal” OR “interactive effects” OR “multifactor” OR “nitrogen and phosphorus limitation” OR “multiple limiting resources” OR “nutrient co-limitation” OR “global change experiments” OR “counteract*”)

AND

TS = (“multiple stress*” OR “stressors” OR “global change factors” OR “global change drivers” OR “multifactor” OR “cumulative effect*” OR “net effect*” OR “combined effect*” OR “interacting” OR “nitrogen and phosphorus limitation” OR “multiple limiting resources” OR “nutrient co-limitation” OR “global change experiments”)

AND

WC = (“Ecology” OR “Toxicology” OR “Environmental Sciences” OR “Plant Sciences” OR “Zoology” OR “Marine Freshwater Biology” OR “Limnology” OR “Oceanography” OR “Multidisciplinary Sciences” OR “Plant Sciences” OR “Microbiology” OR “Biodiversity Conservation”)

AND

PY = (1900-2020)

AND

TS = (“Freshwater” OR “river” OR “stream” OR “pond” or “lake” or “catchment” or “riparian”)

References

Orr, J. A., Vinebrooke, R. D., Jackson, M. C., Kroeker, K. J., Kordas, R. L., Mantyka-Pringle, C., ... & Piggott, J. J. (2020). Towards a unified study of multiple stressors: divisions and common goals across research disciplines. *Proceedings of the Royal Society B*, 287(1926), 20200421.

Appendix Two

Supplementary material for Chapter Two.

A table of contents is provided below.

Theoretical simulations

The following section details the method used to generate data for the theoretical simulations.

Equations

The equations (Equations 2.1 & 2.2) used throughout this analysis have been taken from Heath et al., (2014); though the equations have been minorly adapted to include density independent mortality at all trophic levels (except for the bottom trophic level) as opposed to only the apex trophic level. The two sets of equations comprise differing forms of dynamics and regulation. Equation 2.1 (DD-Equations) comprises a term for density-dependent mortality, whereas in Equation 2.2 (CR-Equations) consumer uptake is regulated by density. These two sets of equations have been explored in detail, and compared to alternate forms of classic ecological theory, by Heath et al., (2014).

Generation of Theoretical Data

Theoretical stressor interaction data was generated using the following method (expanded from that in Section 2.3).

1) Parameter combinations were randomly determined for 1,320,000 food chains, where each combination of equation type (DD-Equations or CR-Equations) and food chain length (three, four, or five trophic levels) accounts for 220,000 parameter combinations. Each parameter was randomly assigned a value from a predetermined range, with ranges differing depending upon the parameter in question (Table 2.2). The parameter values were drawn from a uniform distribution (i.e., any value in the range is as likely to be selected as any other). These values are considered to be *baseline* parameter values.

2) Two parameters were randomly chosen to be subjected to stress with all candidate parameters being chosen with equal probability. These two parameters have different values to their baseline equivalents, but all other parameters will maintain the same

values. Stressed parameter values were chosen from the ranges given in Table 2.2, using a uniform distribution. Only the parameters of attack rate (α), conversion efficiency (ε), density independent mortality (δ), or nutrient input rate (ω) were able to be stressed. The parameters selected to be stressed could be on the same trophic level (e.g. α_2 and ε_2) the same biological process on different trophic levels (e.g. ε_2 and ε_1) or different biological processes on different trophic levels (e.g. ε_1 and α_2). Additionally, α_1 was not stressed in any of the simulations. This is due to the fact that altering α_1 only had any impact upon the nutrient trophic level, and no impact upon any other trophic levels in the simulations (unpublished data). Given that this analysis was primarily concerned about *non-nutrient trophic levels*, α_1 was not stressed in any analysis.

Stressor A parameter values are the same as the baseline values except that the parameter value for the first stressed parameter is changed to the stressed value.

Stressor B parameter values are the same as the baseline values except that the parameter value for the second stressed parameter is changed to the stressed value.

Interaction parameter values are the same as the baseline values except that the parameter value for both stressed parameters are changed to the corresponding stressed values.

Overall, there are 5,280,000 parameter combinations equating to 1,320,000 parameter groups for individual food chains under baseline, stressor A, stressor B, and interaction conditions.

When determining the value of the stressed parameters, these were again selected using a uniform distribution. When α, ε, or δ were selected to be stressed, the stressed value was determined from a uniform distribution (U~(0.01, 0.99)), which was bounded at 0.01 and 0.99 for the lower and upper limits respectively. The single caveat placed upon this uniform distribution was that stressed parameter values were unable to have a value within the range of the baseline value ± 0.025, this was so as to ensure that any stressed parameter value was suitably different to the baseline parameter value. The same approach was adopted when ω was selected to be stressed, though the limits of the uniform distribution (U~(1, 99)) were 1 and 99, with the excluded range being the baseline parameter value ± 2.5.

3) For each of the 1,320,000 parameter groups, the equilibrium densities of all trophic levels in the food chain, under all levels of stress, were determined using the *NSolve* function in Mathematica 10.4 (Wolfram Research Inc., 2016). Equilibrium densities are referred to as the densities of each trophic level in the food chain that result in there being no net change in the density of any other trophic level. Furthermore, additional constraints were placed on the equilibrium densities in that they needed to be feasible (i.e. the equilibrium densities of all trophic levels in the food chain had to be positive and non-zero).

   For some parameter groups, equilibrium densities were able to be calculated for the baseline parameter values but not for either stressor A, stressor B, or the interaction parameter values. When this occurred the equilibrium densities for the entire parameter group were discarded from any further analysis.

   Overall, feasible equilibrium densities could be found, under all four levels of stress, for 1,054,272 (79.9%) parameter groups.

4) From the 1,054,272 parameter groups, 360,000 were randomly selected (with each combination of food chain length and equations accounting for ~16.667% of interactions). From each of these 360,000 parameter groups, the equilibrium density of a single trophic level (consistent across each level of stress) was randomly selected and collated into a single dataset for analysis, hence referred to as the *analysis dataset*. All subsequent analysis was performed on this analysis dataset. For this analysis, the number of interactions (360,000) was large enough to ensure that repeating the analysis would not cause substantially different results to occur, whilst also being computationally feasible. Furthermore, this also allowed for significant numbers of interactions to be included from i) both types of equations, ii) all lengths of food chain, iii) all considered trophic levels.

Simulation of observation error

For each parameter group, when calculating the equilibrium densities, there was no innate variation included within the equations. Hence, using the same parameter values will return the same equilibrium densities each time. However, this is not necessarily reflective of ecological systems. To account for this, observation error was incorporated into the analysis dataset through the following method.

5) Firstly, the level of observation error, σ, to be included was chosen, where σ refers to the standard deviation of a normal distribution with a mean of 1 (i.e. $N\sim(1, \sigma)$). Within this analysis, observation error was included with equal spacing on a logarithmic scale (between $1\text{x}10^{-6}$ and $5\text{x}10^{-1}$) with there being 50 different levels of observation error investigated (alongside the absence of observation error). Through the rest of this method we use a level of observation error of σ.

6) For each of the 360,000 parameter groups in the analysis, a number of replicates (*n*) (either three, four, five, or six), was randomly selected using a uniform distribution. The number of replicates was consistent across the different treatments for each parameter group (i.e. baseline, stressor A, stressor B, and interaction). The range of *n* was chosen to mirror the ranges of replicates reported in the majority of empirical studies that made up our freshwater stressor analysis.

7) For each treatment within each parameter group, *n* values from a Gaussian distribution with a mean of 1 and standard deviation of σ were taken. These values were then multiplied by the equilibrium density for that treatment; with the mean and standard deviation of these new *n* equilibrium densities being calculated.

   This was conducted for each treatment within each parameter group. Additionally, if any of the *n* values from the Gaussian distribution were equal to, or greater than, 2 or less than, or equal, to 0, then the *n* values from the Gaussian distribution were re-drawn.

Accordingly, by having the mean, standard deviation, and number of replicates for each treatment, it is possible to calculate effect size metrics (Hedges' d) and determine interaction classifications.

Additional Results

Overall, the qualitative pattern for the interaction classification frequencies varied with observation error. Figure S2.1.1 shows how the frequency of the different interaction classes varies with observation error for the overall dataset of 360,000 interactions, as well as this dataset when subset for food chains using either CR-Equations or DD-Equations. Similarly, Figure S2.1.2 shows how the frequency of interaction classes varies when the overall dataset is subset by the length of food chain. Figure S2.1.3 shows how interactions classed as being reversal, antagonistic, or synergistic under no observation error, are reclassified when observation error is incorporated at different levels. Figure S2.1.4 illustrates the distribution of effect sizes under different levels of observation error. Penultimately, Figure S2.1.5 shows the frequency of the interaction classifications, summary effect sizes, and ratio of positive to negative effect sizes under different levels of observation error when two stressors are capable of acting upon the same single parameter. The method for this section is near identical to that outlined above, with the only difference being that the same parameter can be impacted by two individual stressors, with the interaction stressed parameter value being the additive effect of the individual stressors. The interaction stressed parameter value still had to be within the same limits as before, though if it exceeded these limits the individual stressed parameter values were redrawn. As shown, the results of this analysis (Figure S2.1.5) are highly similar to when two stressors only ever act on different parameters (Figure 2.2).

Finally, Figure S2.1.6 illustrates the frequency of interactions that are assigned the 'true' interaction classification (across all levels of observation error) for different levels of deviation. Here, deviation is calculated as $\frac{D_I - D_A - D_B + D_C}{D_A + D_B - 2D_C} \cdot 100\%$, with this metric denoting the percentage difference between the observed and additive null model predicted interaction effects. For Figure S2.1.6, interactions have been grouped, according to their deviation, into bins of 5% intervals (i.e. 0% ≤ deviation < 5%; 5% ≤ deviation < 10%; etc.). We found no evidence for interactions only weakly deviating from additivity, with the majority (53.1%) of interactions have a deviation greater than ±5%, including numerous interactions having a deviation of greater than ±50% (10.1%) or greater than ±100% (5.2%). Under each level of observation error, the frequency of interactions that are correctly classified has been determined. As expected, correct classification of the most minorly deviating interactions (i.e., $0\% \leq$ deviation $< 5\%$) declines quite quickly with increasing observation error (Figure S2.1.6). However, the correct classification of the more strongly deviating interactions declines much more slowly. If we compare Figure 2.2 to Figure 2.1.6 we can see that by an observation level of around 1×10^{-2} most correctly assigned interactions are in the larger deviation classes, but that as we observe in Figure 2.2, there is still a high total proportion of interactions that can be correctly detected.

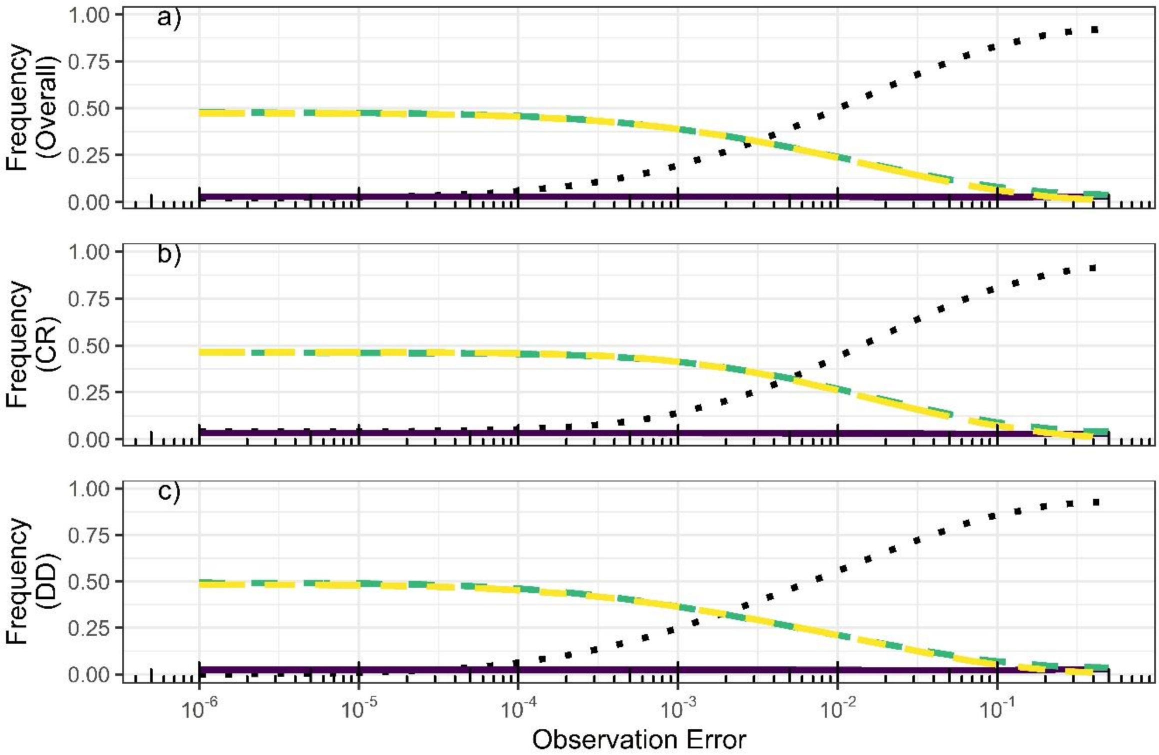

Figure S2.1.1: Prevalence of the different classifications of interactions for the theoretical data, when observation error (σ) was implemented, separated for data from Consumer Uptake Regulation Equations, Density Dependence Equations, and the combined dataset. a) Overall dataset, b) only food chains with CR dynamics, c) only food chains with DD dynamics. Dotted black line denotes additive interactions. Dashed green line indicates antagonistic interactions. Long-dashed yellow line denotes synergistic interactions. Solid purple line indicates reversal interactions.

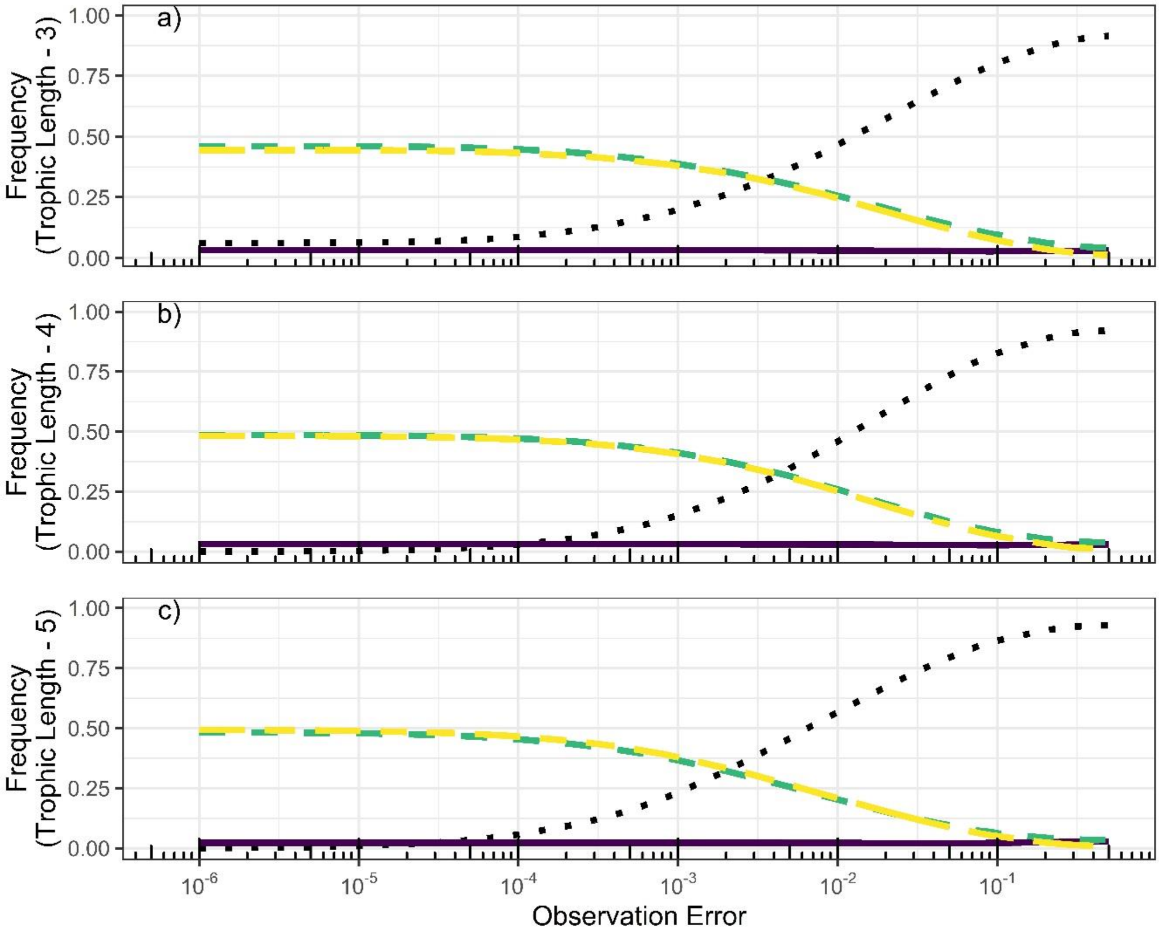

Figure S2.1.2: Prevalence of the different classifications of interactions for the theoretical data, when observation error (σ) was implemented, separated for data from the three different lengths of food chain. a) Communities comprising three trophic levels, b) Communities comprising four trophic levels, c) Communities comprising five trophic levels. Dotted black line denotes additive interactions. Dashed green line indicates antagonistic interactions. Long-dashed yellow line denotes synergistic interactions. Solid purple line indicates reversal interactions.

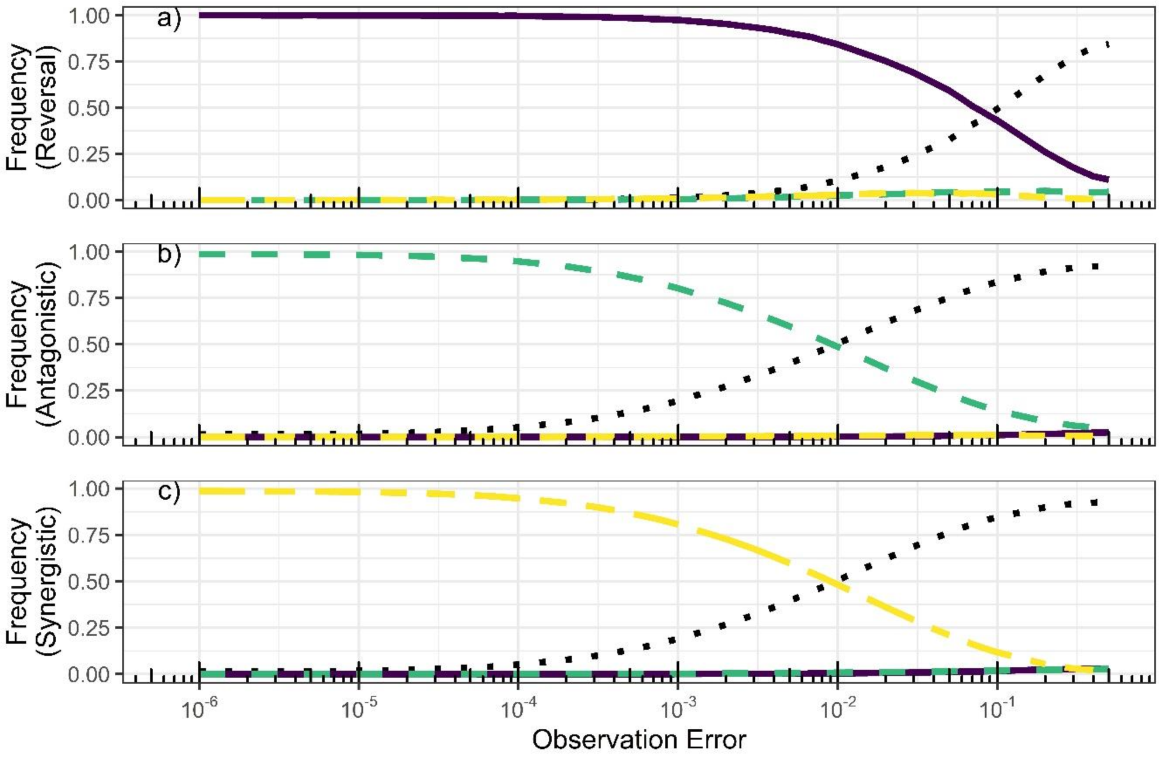

Figure S2.1.3: For a given 'true' interaction classification (i.e. when observation error is not considered) the proportions of this interaction classification that are reassigned to a different class when observation error (σ) was implemented. The data shown is for the combined dataset of containing simulations using either set of equations, comprising 360,000 interactions. a) Reversal interactions, b) Antagonistic interactions, c) Synergistic Interactions. Dotted black line denotes additive interactions. Dashed green line indicates antagonistic interactions. Long-dashed yellow line denotes synergistic interactions. Solid purple line indicates reversal interactions.

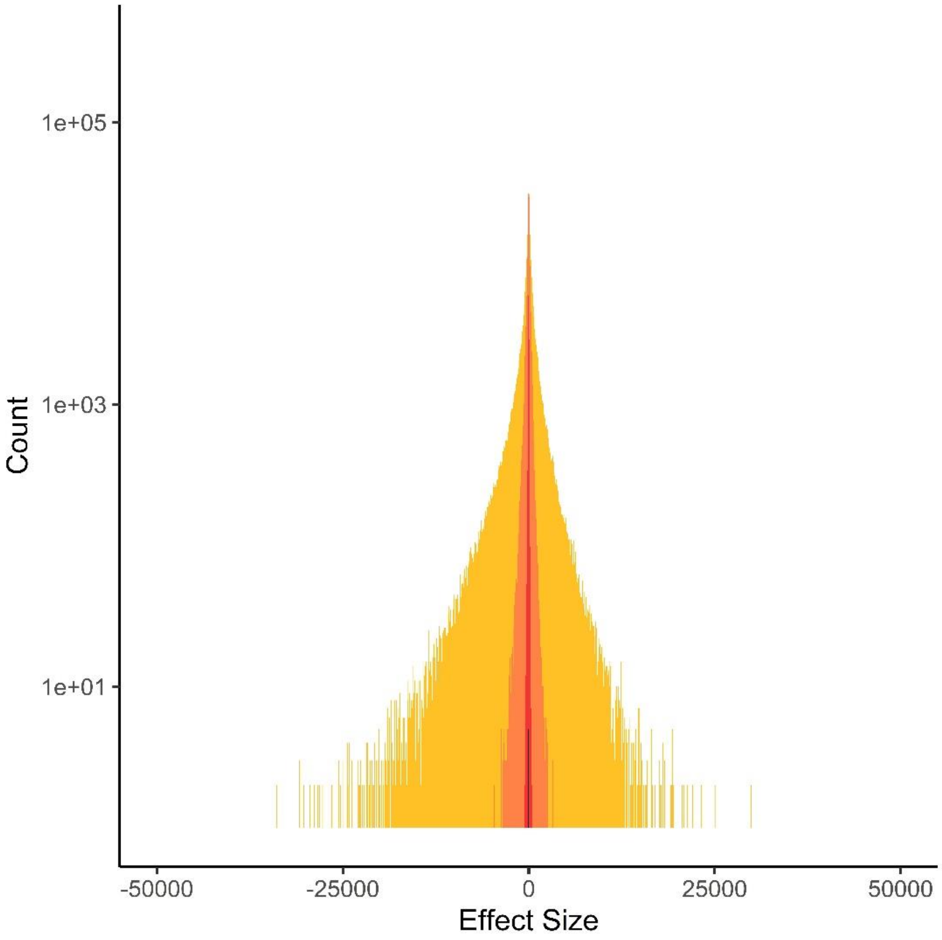

Figure S2.1.4: Distribution of effect sizes for the theoretical data, under multiple levels of observation error. Yellow indicates the effect size distribution for an observation error (σ) of $1x10^{-4}$*; Orange indicates the effect size distribution for an observation error of* $1x10^{-3}$*; Red indicates the effect size distribution for an observation error of* $1x10^{-2}$*; Black indicates the effect size distribution for an observation error of* $1x10^{-1}$*.*

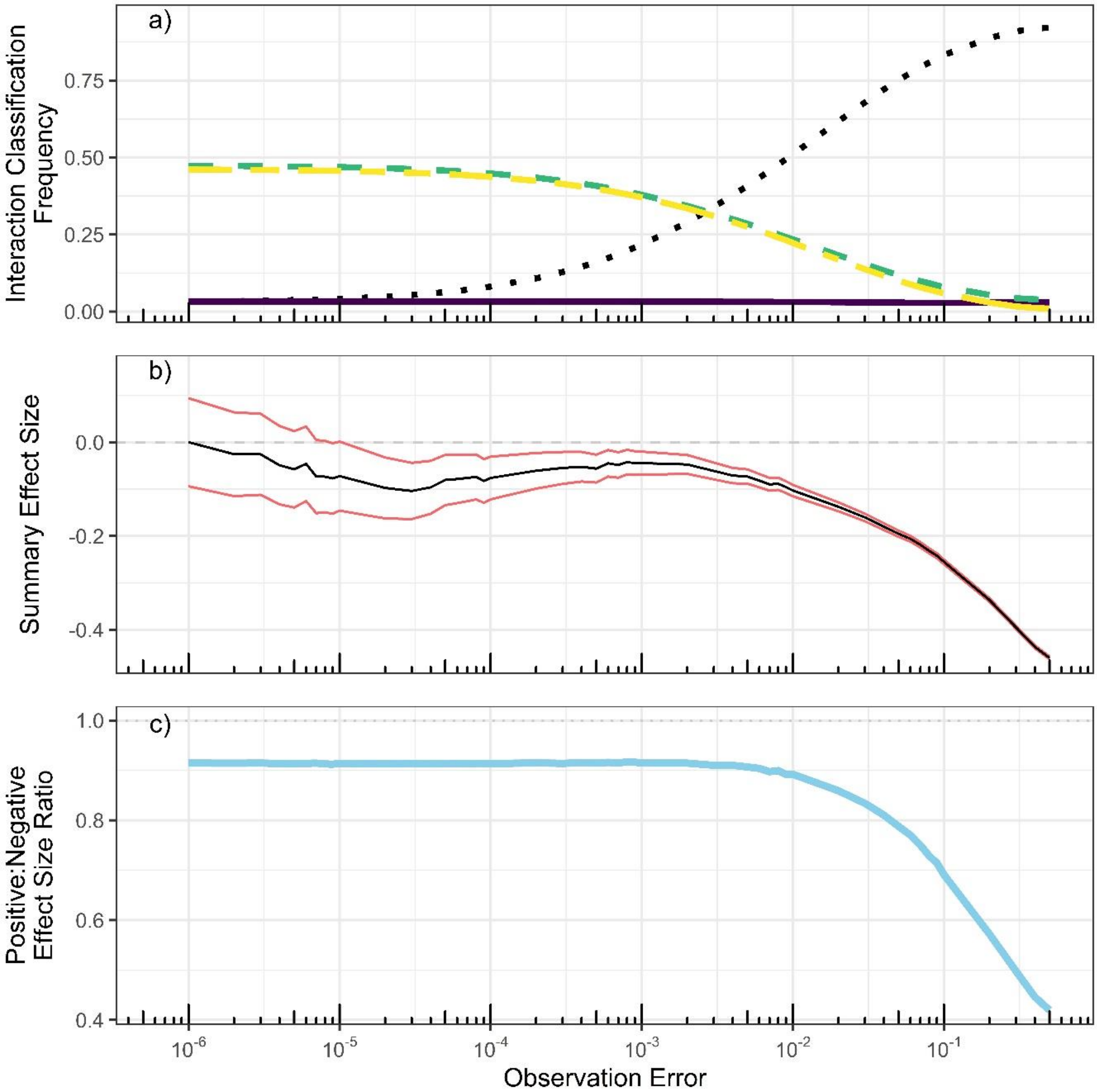

Figure S2.1.5: The effect of observation error (σ) on the stressor interaction categorisation, and summary meta-analytic effect sizes in the theoretical data where two stressors are able to act on the same parameter. a) Frequency of the different interaction classes for the 360,000 theoretical interactions at each level of observation error. Dotted black line denotes additive interactions. Green short-dashed line indicates antagonistic interactions. Yellow long-dashed line denotes synergistic interactions. Purple line indicates reversal interactions. b) Summary effect sizes for the 360,000 theoretical interactions, at each level of observation error. Black lines denote summary effect sizes, and red lines denote 95% confidence intervals. c) Ratio of positive to negative summary effect sizes at each level of observation error.

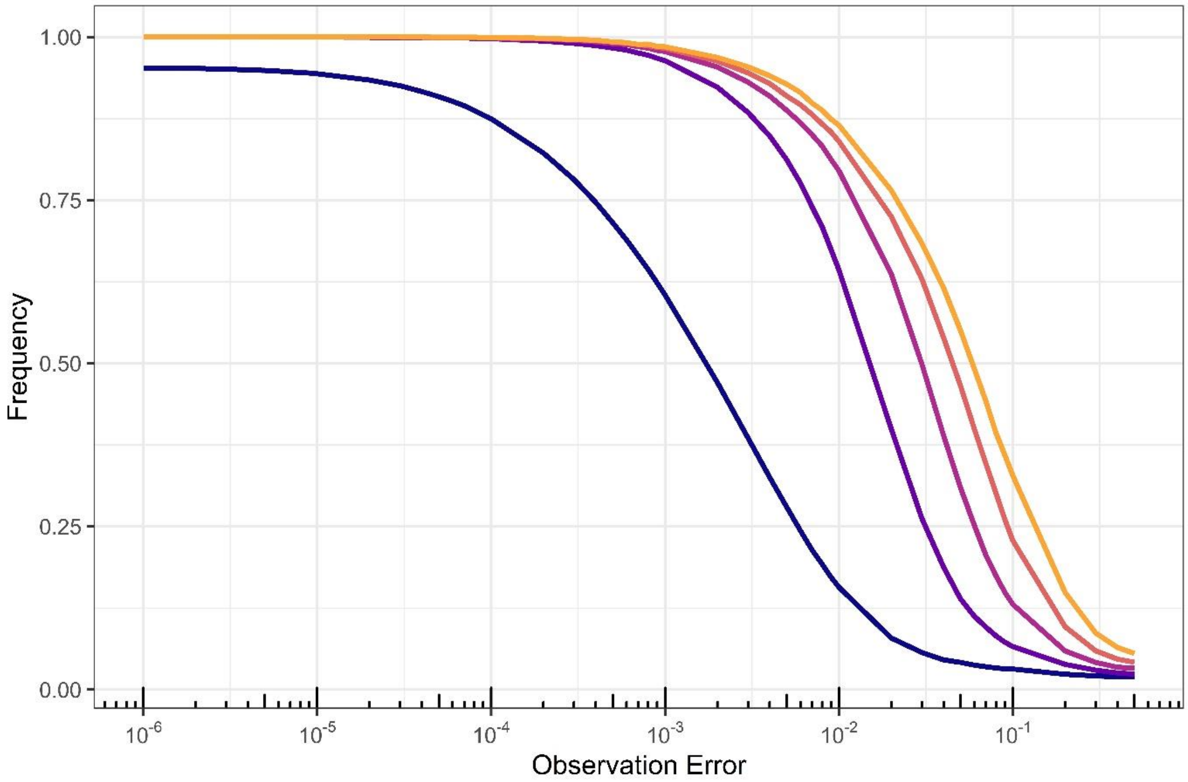

Figure S2.1.6: Frequency of interactions that are assigned the correct classification in the presence of observation error (σ). Lines refer to different levels of deviation. Dark blue - 0% ≤ deviation < 5%; purple - 5% ≤ deviation < 10%; pink - 10% ≤ deviation < 15%; orange - 15% ≤ deviation < 20%; yellow - 20% ≤ deviation < 25%.

References

Heath, M. R., Speirs, D. C., & Steele, J. H. (2014). Understanding patterns and processes in models of trophic cascades. *Ecology Letters*, 17(1), 101-114.

Wolfram Research. Inc. (2016). Mathematica (Version 10.4). *Wolfram Research, Inc., Champaign, Illinois.*

Literature search

The following section details the literature search for empirical data to be used within the multiple stressor meta-analysis.

A search for literature was conducted in Web of Science (26/02/2019) using the following search criteria, for papers published up to, and including, 31st December 2018:

TS = ("synerg*" OR "antagon*" OR "additive" OR "dampen*" OR "combined" OR "multiple" OR "factorial" OR "experiment" OR "stressor" OR "multiple stressor") AND TS = ("freshwater" OR "river" OR "stream" OR "lake" OR "catchment" OR "pond" OR "riparian") AND TS = ("impact" OR "effect") AND TS = ("communit*") AND TS = ("biomas*" OR "abundan*" OR "densit*") AND SU=("Biodiversity & Conservation" OR "Environmental Sciences & Ecology" OR "Fisheries" OR "Marine & Freshwater Biology" OR "Mathematical & Computational Biology" OR "Microbiology" OR "Plant Sciences" OR "Zoology")

The search terms used here were adapted from those used by Jackson et al., (2016).

Overall, 1805 entries were returned.

To be included within this dataset, studies had to investigate the impacts of two stressors, upon a freshwater community. There was no limit placed upon the types of stressor that could be considered, so long as the stressor was a manipulation of environmental, or novel ecological, conditions as opposed to measuring the response of native communities to normal, or expected, events or disturbances.

Only data referring to the abundance, biomass, density (or in some cases chlorophyll-a, or other chemical concentrations from which biomass could be estimated) was collated. For each group of organisms, the study had to report, mean values, standard deviation or standard error, and the number of replicates for the following:

i) Data for the group of organisms under control conditions (i.e. no stressors)
ii) Data for the group of organisms under Stressor A (a single stressor)
iii) Data for the group of organisms under Stressor B (a single stressor)

iv) Data for the group of organisms under both Stressors A and B (both stressors acting simultaneously)

Each study could report data for multiple different groups of organisms, all of which was collated. Additionally, some studies reported data for freshwater communities exposed to multiple different intensities of a stressor, all of these were collated, so long as the above factorial design was followed. Where studies did not report mean values, standard deviation or standard error, or the number of replicates, the data was not incorporated within the dataset.

Each paper was individually screened to determine whether it was suitable for inclusion within our study. Firstly, titles and abstracts were assessed to determine whether the paper was potentially relevant to our study. If a paper was potentially relevant the methods were read to determine whether the study design was factorial and met the above criteria for inclusion. If the paper met the criteria, then the results and supplementary information were considered to determine whether all required data (above) was presented. Overall, data extraction was attempted for 124 papers. However, at this stage 66 papers were deemed unsuitable for data extraction, due to issues such as missing data, or having figures that were too unclear for data extraction.

To be included within this dataset, studies had to investigate freshwater communities. Those experiments which were conducted upon other such aquatic environments, e.g. marsh, brackish, estuarine, or coastal environments were not considered. From each study, data was extracted from tables or from figures using WebPlotDigitizer software (Rohatgi, 2017).

Overall, 58 papers were found, which were able to contribute data to this dataset, contributing a total of 859 interactions. However, to reduce covariance within the dataset between these interactions, the following protocol was adopted.

i) Where studies reported data for multiple response metrics, preference was given in the following order: Density, Abundance, Biomass, and Chlorophyll-a (and other chemical proxies for biomass).

ii) Where studies reported data for multiple overlapping groups of organisms, preference was given to subsets of groups of organisms as opposed to coarser groupings. For instance, if a study reported densities of i) overall periphyton, alongside data for constituent ii) algae and iii) bacteria; only ii and iii would be included in the dataset.

iii) Where studies reported data for the same group of organism, but in different locations within the study environment, only one location was used, with preference being given to the location where a more complete set of data was able to be extracted.

iv) Where studies reported data for multiple time points, preference was given to the time point measured latest in the experiment.

After following this protocol, the number of interactions was reduced to 545.

Within the dataset, each stressor is assigned into one of seven different categories depending upon qualitative properties of the stressor. The seven different categories are: community composition, contamination, habitat alteration, light, nutrients, salinity, and temperature, with explanations and examples given in Table S2.2.1.

Table S2.2.1: The seven broad categories of stressor considered within this analysis, with examples and definitions provided.

Stressor Category	Explanation	Example
Community Composition	The addition or removal of a group of organisms from the community.	e.g. Invasive species
Contamination	The addition of a chemical (with the exception of nutrients) at concentrations not normally found in the system.	e.g. Pesticides
Habitat Alteration	Changing the physical properties of a system from those normally seen	e.g. Sedimentation
Light	Change in the levels of light that a system receives	e.g. UV
Nutrients	Addition of nutrients to a system at levels exceeding those normally found	e.g. Nitrogen
Salinity	An increase in the salinity of the freshwater study system	e.g. Salt addition
Temperature	An increase in the temperature of the freshwater study system	e.g. Temperature increase

References

Jackson, M. C., Loewen, C. J., Vinebrooke, R. D., & Chimimba, C. T. (2016). Net effects of multiple stressors in freshwater ecosystems: a meta-analysis. *Global Change Biology*, 22(1), 180-189.

Rohatgi, A. (2017) WebPlotDigitalizer: HTML5 based online tool to extract numerical data from plot images. Version 4.1. URL <http://arohatgi.info/WebPlotDigitizer/app/>

Calculation of Hedges' d

This section details the calculation of Hedges' d, used to implement the additive null model.

For each interaction the effect size metric of Hedges' d was calculated. Hedges' d is an estimate of the standardised mean difference that is not biased by small sample sizes (Hedges and Olkin, 1985). Here we use the method outlined in Gurevitch et al., (2000) to calculate Hedges' d, which uses an additive null model to compare the predicted effect of two stressors upon a system (i.e. the sum of their effects) to the effect that is actually observed. In order to calculate this, experiments are required to use a factorial design with four treatments: control conditions with no stressors (C), only stressor A present (A), only stressor B present (B), and both stressors present (I). For each of these treatments, the mean density/abundance/biomass of the response group of organisms (D_X) the standard deviation of the mean response (s_x), and the number of replicates (N_x) is required.

Overall, Hedges' d (d_{int}) was calculated using the following equations:

$$d_{int} = \frac{D_I - D_A - D_B + D_C}{s}$$

Where s is the pooled standard deviation of d_{int}:

$$s = \sqrt{\frac{(N_I - 1)(s_I)^2 + (N_A - 1)(s_A)^2 + (N_B - 1)(s_B)^2 + (N_C - 1)(s_C)^2}{(N_I + N_A + N_B + N_C - 4)}}$$

The variance of d_{int} can be calculated as:

$$v_{int} = \left(\frac{1}{N_I} + \frac{1}{N_A} + \frac{1}{N_B} + \frac{1}{N_C} + \frac{d_{int}^2}{2(N_I + N_A + N_B + N_C)}\right)$$

Where the standard error of d_{int} can be calculated (Borenstein 2009) as:

$$SE_{int} = \sqrt{v_{int}}$$

Where 95% confidence intervals are calculated as:

$$CI_{95\%} = 1.96 * SE_{int}$$

For each interaction, directionality was removed using the method stated in Section 2.3 and outlined in Piggott et al., (2015). This involved inverting the response direction of the calculated Hedges' d where the predictive additive effects were negative.

For each interaction, it is possible to use Hedges' d to determine an interaction classification (namely, additive, antagonistic, synergistic, or reversal). Accordingly, an interaction was classified as being additive if, $d_{int} \pm CI_{95\%}$, overlaps zero. If $d_{int} \pm CI_{95\%}$ does not overlap zero, then an interaction was classified under the following conditions:

i) $d_{int} > 0$, the interaction is classed as being synergistic

ii) $d_{int} < 0$ and

$D_A + D_B - 2D_C$ and $D_I - D_C$ have the same polarity, the interaction is classed as being antagonistic

iii) $d_{int} < 0$ and

$D_A + D_B - 2D_C$ and $D_I - D_C$have the different polarities, the interaction is classed as being reversal

The method was used for all empirical interactions, and for theoretical interactions where observation error was included.

Theoretical interactions without observation error

When no observation error was present in the theoretical experiments, d_{int} was instead classified using the following equation:

$$d_{int} = D_I - D_A - D_B + D_C$$

This is due to the fact that with no observation error, the was no variation in the densities of trophic levels, hence effect sizes could be determined exactly with no confidence intervals.

Accordingly, under this set of scenarios, if:

iv) $d_{int} > 0$, the interaction is classed as being synergistic

v) $d_{int} < 0$ and
$D_A + D_B - 2D_C$ and $D_I - D_C$ have the same polarity, the interaction is classed as being antagonistic

vi) $d_{int} < 0$ and
$D_A + D_B - 2D_C$ and $D_I - D_C$ have the different polarities, the interaction is classed as being reversal

vii) $d_{int} = 0$, the interaction is classed as being additive

References

Borenstein, M., Cooper, H., Hedges, L. & Valentine, J. (2009). Effect sizes for continuous data. *The Handbook of Research Synthesis and Meta-Analysis*, 2, 221-235.

Gurevitch, J., Morrison, J. A., & Hedges, L. V. (2000). The interaction between competition and predation: a meta-analysis of field experiments. *The American Naturalist*, 155(4), 435-453.

Hedges, L. V., & I. Olkin. (1985). Statistical methods for meta-analysis. *Academic Press, New York*.

Piggott, J.J., Townsend, C.R. & Matthaei, C.D. (2015). Reconceptualizing synergism and antagonism among multiple stressors. *Ecology and Evolution*, *5*(7), 1538-1547.

Naming conventions

Within this research, we have adopted the naming convention for interaction classes, first established by Jackson et al., (2016), and additionally outlined by Orr et al., (2020). Table S2.4.1 outlines how the naming of interaction classes within this study corresponds to those used within previous studies.

Table S2.4.1: Comparisons of the naming of interaction classifications within this study, and other studies.

This research	Travers-Trolet et al., (2014) / Fu et al., (2018)	Common Nomenclature
Additive	Additive	Additive
Synergistic	Synergistic	Synergistic
Antagonistic	*Dampened*	*Antagonistic*
Reversal	*Antagonistic*	*Antagonistic*

The term *antagonistic* is frequently used to describe different classes of interactions depending upon the study. Frequently, studies do not differentiate between *antagonistic* and *reversal* interactions simply classifying both as being *antagonistic*.

References

Fu, C., Travers-Trolet, M., Velez, L., Grüss, A., Bundy, A., Shannon, L. J., ... & Shin, Y. J. (2018). Risky business: the combined effects of fishing and changes in primary productivity on fish communities. *Ecological Modelling*, 368, 265-276.

Jackson, M. C., Loewen, C. J., Vinebrooke, R. D., & Chimimba, C. T. (2016). Net effects of multiple stressors in freshwater ecosystems: a meta-analysis. *Global Change Biology*, 22(1), 180-189.

Orr, J. A., Vinebrooke, R. D., Jackson, M. C., Kroeker, K. J., Kordas, R. L., Mantyka-Pringle, C., ... & Piggott, J. J. (2020). Towards a unified study of multiple stressors: divisions and common goals across research disciplines. *Proceedings of the Royal Society B*, 287(1926), 20200421.

Travers-Trolet, M., Shin, Y. J., Shannon, L. J., Moloney, C. L., & Field, J. G. (2014). Combined fishing and climate forcing in the southern Benguela upwelling ecosystem: an end-to-end modelling approach reveals dampened effects. *PLoS One*, 9(4), e94286.

Weighted meta-analyses

Meta-analyses were conducted using the *metafor* package (Viechtbauer, 2010) in R, using the function *rma.mv*. The meta-analyses conducted within our research were weighted. For every meta-analyses, each interaction was weighted, with the weighting given to each interaction by determined by the *rma.mv()* function.

Many of the included empirical studies conducted experiments where multiple treatments of the same stressor were investigated, or the same control was used for multiple factorial experiments. In order to account for covariance between these interactions, covariance-variance matrices were used within the empirical meta-analyses. Covariance-variance matrices take the form of the below example (Equation S5.1). In this example the variance of the four interactions are along the major diagonal (v_i) whilst interactions 1 and 2 share a control, hence the covariance of these two interactions ($c_{1,2}$) is incorporated in the corresponding off-diagonal positions. Where two interactions do not share a control (there is no covariance) the corresponding off-diagonal position is filled with a 0.

$$\begin{pmatrix} v_1 & c_{1,2} & 0 & 0 \\ c_{1,2} & v_2 & 0 & 0 \\ 0 & 0 & v_3 & 0 \\ 0 & 0 & 0 & v_4 \end{pmatrix}$$

Equation S5.1, an example of a covariance-variance matrix.

For the empirical meta-analyses, random effects were specified as being:

i) The study the interaction was published in,
ii) The study group of organisms.

Within the meta-analytical model, ii) was nested within i). Accordingly, i) allows us to account for between study variance, whilst ii) allows us to account for within study variance.

For the meta-analytical results within the main text of the paper, there were no fixed effects specified. However, two additional meta-analytical models were conducted, with the categorical moderators of *organism group* and *stressor combination*, to test whether our results were robust across subsets of the empirical data (Figures S2.5.1 – S2.5.2). The results from these meta-analytical models are in line with the main findings of this study. All sub-groups reported summary effect sizes that were negative, with confidence intervals that did not overlap zero (i.e. antagonistic/reversal interactions) or had confidence intervals that overlapped zero (additive interactions). However, it should be noted that many of these sub-groups had small numbers of interactions, and as such care should be taken when inferring conclusions from these results.

For the theoretical meta-analytical models, a slightly different approach was used, with this due to computational constraints. Here, weighted fixed effect models were conducted using the *lm* function in R. For models using the theoretical data, the weighting given to each interaction was specified as being the reciprocal of the interaction variance. Discrepancies between the use of *metafor*, *lm*, *nlme*, and *lme4* have been previously documented (see <http://www.metafor-project.org/doku.php/tips:i2_multilevel_multivariate>). Ultimately, it has been shown that the summary effect size between the two methods remains the same, but that the standard errors differ. While we acknowledge the differences between the two methods, due to the computational constraints it was unavoidable.

Figures S2.5.3 – S2.5.4 are variants of Figure 2.3 (main text) but allow for comparisons between the results of the empirical meta-analysis and subsets of the theoretical analyses where only interactions from either the DD- or CR-Equations are considered. As in Figure 2.3, three different levels of observation error (0.1, 0.01, 0.001) are considered.

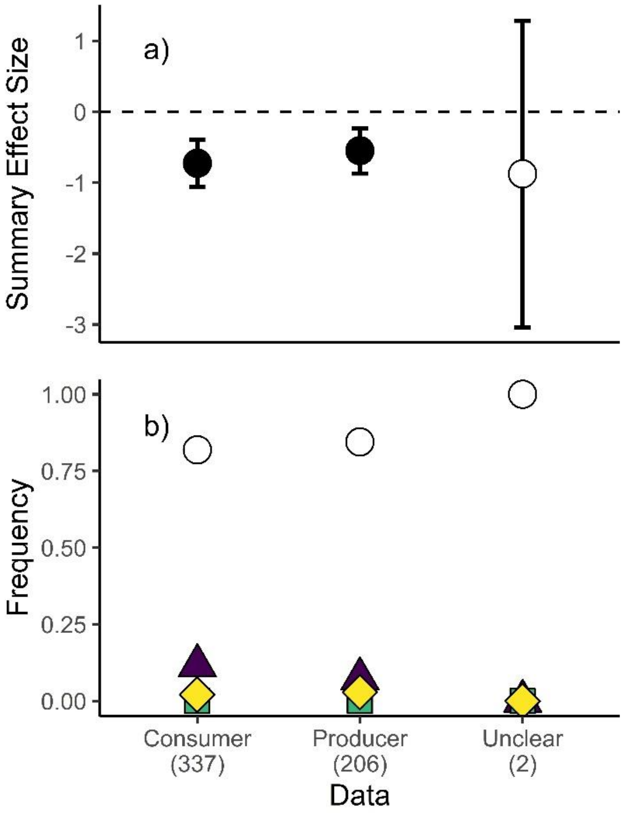

Figure S2.5.1: a) Results of meta-analytical model with the inclusion of organism group (i.e. producer, consumer, or unclear) as a fixed effect. For each organism group, the summary effect size, with accompanying 95% confidence intervals are shown. Filled circles indicate that the 95% confidence intervals do not overlap zero, while unfilled circles indicate that the 95% confidence intervals overlap zero. b) The frequency of the different interaction classes, for each organism group, legend as for Figure 2.3. For x-axis labels, numbers in brackets indicate the number of interactions for organism group.

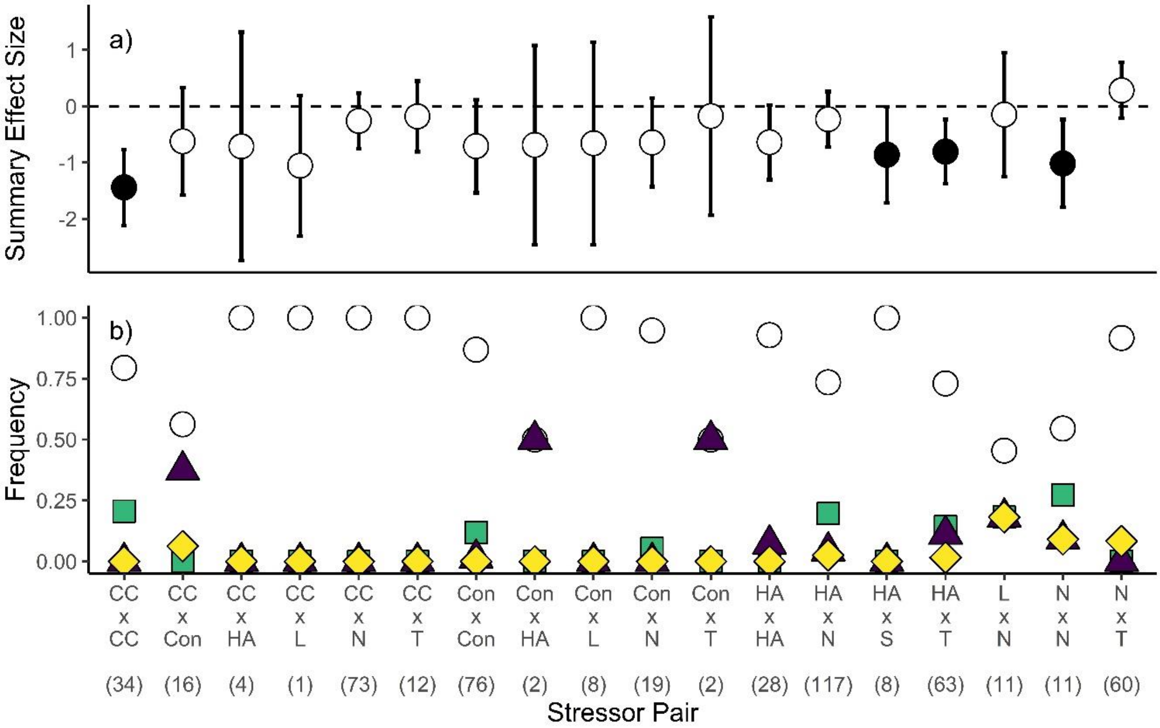

Figure S2.5.2: a) Results of meta-analytical model with the inclusion of stressor combination as a fixed effect. For each stressor combination, the summary effect size, with accompanying 95% confidence intervals are shown. Filled circles indicate that the 95% confidence intervals do not overlap zero, while unfilled circles indicate that the 95% confidence intervals overlap zero. b) The frequency of the different interaction classes, for each stressor combination, legend as for Figure 2.3. For x-axis labels, numbers in brackets indicate the number of interactions for stressor combination. The following notation denotes each group of stressors: CC – community composition; Con – contamination; HA – habitat alteration; L – light; N – nutrients; S – salinity; T – temperature.

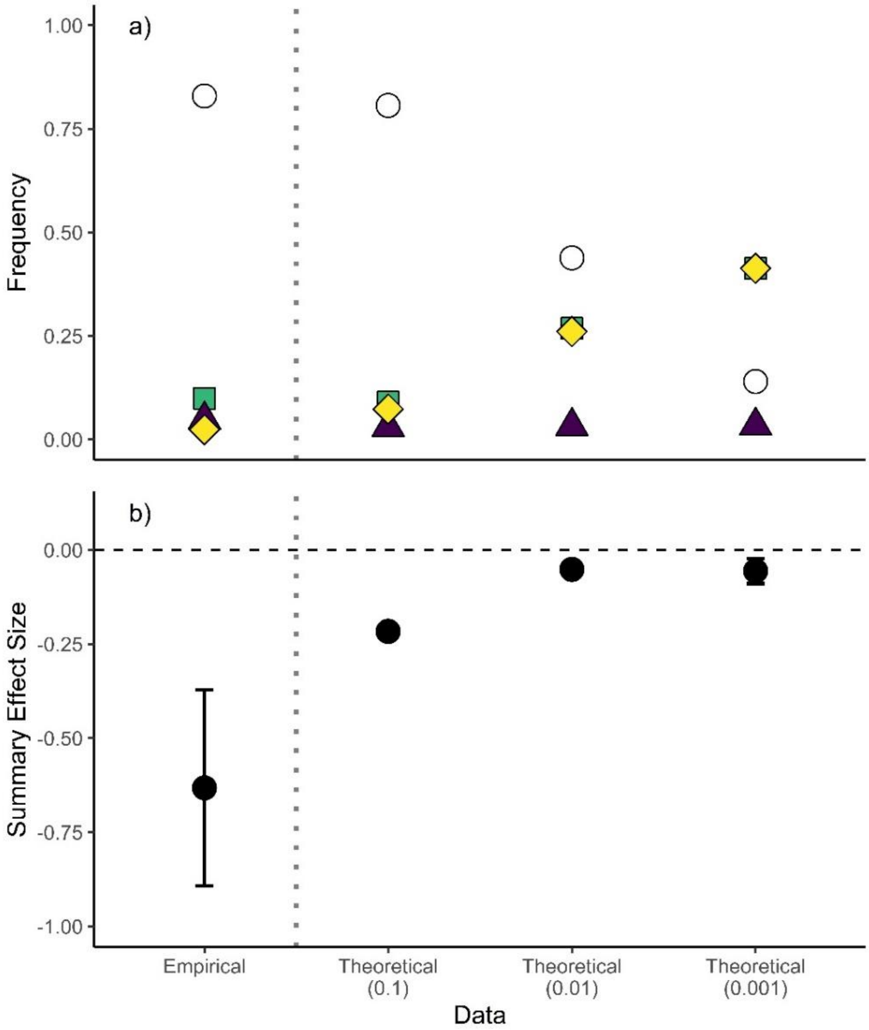

Figure S2.5.3: *Comparisons of the analyses of the freshwater interaction dataset with the theoretical interaction dataset for different levels of observation error, σ (given in parentheses on the x-axis). For the theoretical dataset only effect sizes for interactions determined using DD Equations (Table 1 main text) are shown. Comparisons are for a) proportions of the different interaction classes, and b) summary effect sizes for the empirical and theoretical dataset. The empirical dataset comprised 545 interactions, while the theoretical dataset comprised all simulated interactions at the level of observation error stated within the parentheses. For Figure S2.5.3a) White circles denote additive interactions. Green squares denote antagonistic interactions. Yellow diamonds denote synergistic interactions. Purple triangles denote reversal interactions. For Figure S2.5.3b) Closed circles denote significant summary effect sizes (i.e. 95% confidence intervals do not overlap zero); open circles denote non-significant summary effect sizes (i.e. 95% confidence intervals overlap zero).*

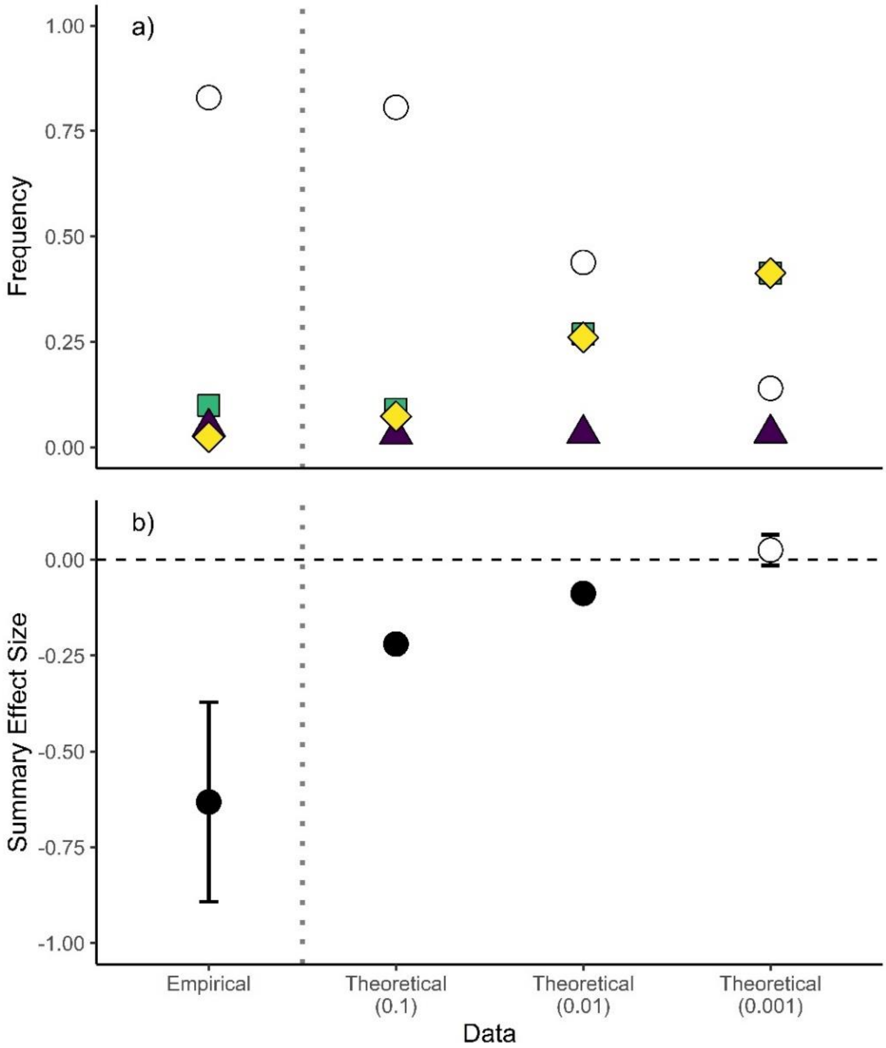

Figure S2.5.4: *Comparisons of the analyses of the freshwater interaction dataset with the theoretical interaction dataset for different levels of observation error, σ (given in parentheses on the x-axis). For the theoretical dataset only effect sizes for interactions determined using CR Equations (Table 1 main text) are shown. Comparisons are for a) proportions of the different interaction classes, and b) summary effect sizes for the empirical and theoretical dataset. The empirical dataset comprised 545 interactions, while the theoretical dataset comprised all simulated interactions at the level of observation error stated within the parentheses. For Figure S2.5.4a) White circles denote additive interactions. Green squares denote antagonistic interactions. Yellow diamonds denote synergistic interactions. Purple triangles denote reversal interactions. For Figure S2.5.4b) Closed circles denote significant summary effect sizes (i.e. 95% confidence intervals do not overlap zero); open circles denote non-significant summary effect sizes (i.e. 95% confidence intervals overlap zero).*

References

The metafor Package, I^2 for Multilevel and Multivariate Models, 2019, viewed 22/04/2020, < http://www.metafor-project.org/doku.php/tips:i2_multilevel_multivariate>

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48. URL: http://www.jstatsoft.org/v36/i03/

Meta-analysis heterogeneity and asymmetry

This section discusses methods used to assess potential heterogeneity and asymmetry within meta-analyses.

Heterogeneity

In order to quantify the effects of heterogeneity within the meta-analytical models, I^2 statistics were calculated. For the overall meta-analytical model an I^2 statistic of 48.5%, just below the benchmark of high heterogeneity from Higgins et al., (2003); though the usefulness of these benchmarks in ecological meta-analyses is questioned (Nakagawa et al., 2017). Overall, this suggests that heterogeneity within our meta-analysis was lower than either the mean (91.69%) or median (84.67%) I^2 found in an analysis of previous ecological meta-analyses (Senior et al., 2016).

Subsequently, in order to explore some of the sources of heterogeneity we conducted two additional meta-analyses on subgroups of the empirical dataset using the ecologically relevant categorical moderator of organism group (namely Producer or Consumer) to subset the data. Overall, there were 206 interactions for producers, and 337 interactions for consumers (while a further two interactions were classed as *unclear* – namely where the response of dinoflagellates to interacting stressors was measured – and excluded from further analysis). Overall, running meta-analytical models following the same method as outlined in the methods, the meta-analytical model for producers returned an effect size of -0.571 ±0.377, corresponding to an antagonistic/reversal summary interaction class; while the meta-analytical model for consumers returned an effect size of -0.614 ±0.333, also corresponding to a summary interaction class of antagonistic/reversal. For both of these sub-group meta-analyses I^2 statistics were calculated. For the producer meta-analytical model, I^2 was equal to 67.7%, while for the consumer meta-analytical model, I^2 was equal to 42.5%. Overall, the sub-group analysis revealed that even within these sub-groups, heterogeneity was either moderate or high.

Publication Bias

Within the overall meta-analysis, moderate-high levels of heterogeneity were observed. However, this can have ramifications for the traditional methods of assessing publication bias within meta-analyses (Ioannidis & Trikalinos, 2007; Nakagawa et al., 2017). Publication bias can manifest through a lack of studies where there are small, or insignificant, effects being published, and instead studies reporting large, or significant effects, are predominately published. As has been previously noted, ecological meta-analyses frequently report high levels of heterogeneity (Senior et al., 2016), as such assessing publication bias can be challenging, as indications of publication bias may instead reflect high levels of heterogeneity.

For our analysis we report the Fail-Safe Number (using the Rosenberg approach) which indicates the number of studies averaging null results that would need to be added to the analysis to reduce the significance level to a target alpha level (0.05). For our analysis the reported Fail-Safe Number is 33674.

Additionally, we include two funnel plots illustrating the relationship between effect size and sample size (Figure S2.6.1a) and effect size and the inverse of the interaction variance (S2.6.1b). As shown by Figure S2.6.1a, there are consistent bands of sample sizes which correspond to sample sizes exactly devisable by four (i.e. 8, 12, 16, 20, 32), which correspond to equal numbers of replicates for each of the four treatments within each interaction. Similarly, these bands are also visible within Figure S2.6.1b. Interpretation of these plots is difficult for several reasons. Firstly, there is correlation between both sample sizes and effect sizes, and the variance and the effect sizes. This is because the sample size is used in the calculation of the effect size and the effect size is used in the calculation of the variance (see *Calculation of Hedges' d*). Further complicating this is the complex structure of our data, and the inability of the plots to accurately reflect this. While the meta-analytical models account for both covariance between interactions where interactions share a control or control and treatment (see *Weighted meta-analyses*), or the specified random effects (see *Weighted meta-analyses*), this is not able to be accounted for within these plots. Similarly, we implement Egger's test (modified to account for multivariate/multi-level meta-analyses), which is a measure of funnel plot symmetry, using the inverse of interaction variance as the specified moderator (p-value < 0.001) which indicated significant asymmetry within the funnel plot. However, again we note both the potential difficulties in interpreting funnel plots, and the difficulties which heterogeneity poses to identifying publication bias.

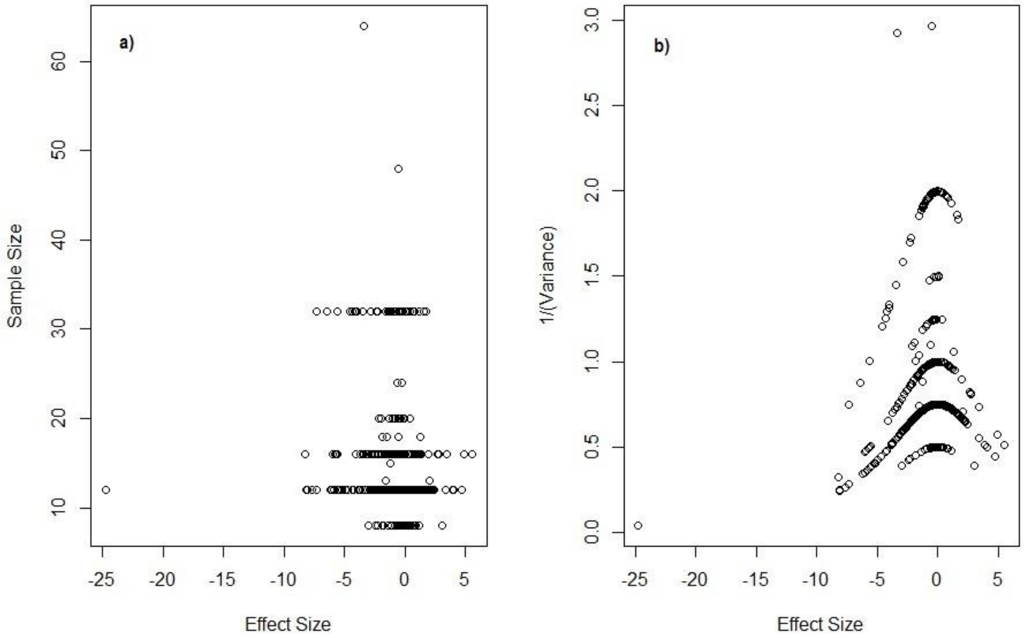

Figure S2.6.1: a – Interaction effect size against sample size for the 545 interactions within the empirical dataset. b – Interaction effect size against the inverse of interaction variance for the 545 interactions within the empirical dataset.

References

Ioannidis, J.P. & Trikalinos, T.A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *Cmaj*, 176(8), 1091-1096.

Nakagawa, S., Noble, D.W., Senior, A.M. & Lagisz, M. (2017). Meta-evaluation of meta-analysis: ten appraisal questions for biologists. *BMC Biology*, 15(1), 1-14.

Senior, A. M., Grueber, C. E., Kamiya, T., Lagisz, M., O'Dwyer, K., Santos, E. S., & Nakagawa, S. (2016). Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. *Ecology*, 97(12), 3293-3299.

Empirical data reference list

Below is the reference list for studies from which empirical data was extracted (listed in alphabetical order).

Alexander, A. C., Culp, J. M., Baird, D. J., & Cessna, A. J. (2016). Nutrient–insecticide interactions decouple density-dependent predation pressure in aquatic insects. *Freshwater Biology*, 61(12), 2090-2101.

Aristi, I., Casellas, M., Elosegi, A., Insa, S., Petrovic, M., Sabater, S., & Acuña, V. (2016). Nutrients versus emerging contaminants–or a dynamic match between subsidy and stress effects on stream biofilms. *Environmental Pollution*, 212, 208-215.

Beermann, A. J., Elbrecht, V., Karnatz, S., Ma, L., Matthaei, C. D., Piggott, J. J., & Leese, F. (2018). Multiple-stressor effects on stream macroinvertebrate communities: A mesocosm experiment manipulating salinity, fine sediment and flow velocity. *Science of the Total Environment*, 610, 961-971.

Brittain, J. T., & Strecker, A. L. (2018). The interactive effect of multiple stressors on crustacean zooplankton communities in montane lakes. *Water Resources Research*, 54(2), 939-954.

Cadmus, P., Guasch, H., Herdrich, A. T., Bonet, B., Urrea, G., & Clements, W. H. (2018). Structural and functional responses of periphyton and macroinvertebrate communities to ferric Fe, Cu, and Zn in stream mesocosms. *Environmental Toxicology and Chemistry*, 37(5), 1320-1329.

Camacho, A., & de Wit, R. (2003). Effect of nitrogen and phosphorus additions on a benthic microbial mat from a hypersaline lake. *Aquatic Microbial Ecology*, 32(3), 261-273.

Cheever, B. M., & Simon, K. S. (2009). Seasonal influence of brook trout and mottled sculpin on lower trophic levels in an Appalachian stream. *Freshwater Biology*, 54(3), 524-535.

Christensen, M. R., Graham, M. D., Vinebrooke, R. D., Findlay, D. L., Paterson, M. J., & Turner, M. A. (2006). Multiple anthropogenic stressors cause ecological surprises in boreal lakes. *Global Change Biology*, 12(12), 2316-2322.

Cottingham, K. L., Glaholt, S., & Brown, A. C. (2004). Zooplankton community structure affects how phytoplankton respond to nutrient pulses. *Ecology*, 85(1), 158-171..

Dahl, J. (1998). The impact of vertebrate and invertebrate predators on a stream benthic community. *Oecologia*, 117(1), 217-226.

Dalton, R. L., Boutin, C., & Pick, F. R. (2015). Determining in situ periphyton community responses to nutrient and atrazine gradients via pigment analysis. *Science of the Total Environment*, 515, 70-82.

Davis, S. J., Mellander, P. E., Kelly, A. M., Matthaei, C. D., Piggott, J. J., & Kelly-Quinn, M. (2018). Multiple-stressor effects of sediment, phosphorus and nitrogen on stream macroinvertebrate communities. *Science of the Total Environment*, 637, 577-587.

De Stefano, L. G., Gattás, F., Vinocur, A., Cristos, D., Rojas, D., Cataldo, D., & Pizarro, H. (2018). Comparative impact of two glyphosate-based formulations in interaction with Limnoperna fortunei on freshwater phytoplankton. *Ecological Indicators*, 85, 575-584.

Duarte, S., Pascoal, C., Alves, A., Correia, A., & Cássio, F. (2008). Copper and zinc mixtures induce shifts in microbial communities and reduce leaf litter decomposition in streams. *Freshwater Biology*, 53(1), 91-101.

Fernández-Aláez, M., Fernández-Aláez, C., Bécares, E., Valentín, M., Goma, J., & Castrillo, P. (2004). A 2-year experimental study on nutrient and predator influences on food web constituents in a shallow lake of north-west Spain. *Freshwater Biology*, 49(12), 1574-1592.

Ferragut, C., & de Campos Bicudo, D. (2012). Effect of N and P enrichment on periphytic algal community succession in a tropical oligotrophic reservoir. *Limnology*, 13(1), 131-141.

Gardeström, J., Ermold, M., Goedkoop, W., & McKie, B. G. (2016). Disturbance history influences stressor impacts: effects of a fungicide and nutrients on microbial diversity and litter decomposition. *Freshwater Biology*, 61(12), 2171-2184.

Gattás, F., De Stefano, L. G., Vinocur, A., Bordet, F., Espinosa, M. S., Pizarro, H., & Cataldo, D. (2018). Impact of interaction between Limnoperna fortunei and Roundup Max® on freshwater phytoplankton: An in situ approach in Salto Grande reservoir (Argentina). *Chemosphere*, 209, 748-757.

Geyer, R. L., Smith, G. R., & Rettig, J. E. (2016). Effects of Roundup formulations, nutrient addition, and Western mosquitofish (Gambusia affinis) on aquatic communities. *Environmental Science and Pollution Research*, 23(12), 11729-11739.

Graham, M. D., & Vinebrooke, R. D. (2009). Extreme weather events alter planktonic communities in boreal lakes. *Limnology and Oceanography*, 54(6part2), 2481-2492.

Gustina, G. W., & Hoffmann, J. P. (2000). Periphyton dynamics in a subalpine mountain stream during winter. *Arctic, Antarctic, and Alpine Research*, 32(2), 127-134.

Haglund, A. L., & Hillebrand, H. (2005). The effect of grazing and nutrient supply on periphyton associated bacteria. *FEMS Microbiology Ecology*, 52(1), 31-41.

Ho, B. S., & Dudgeon, D. (2016). Are high densities of fishes and shrimp associated with top-down control of tropical benthic communities? A test in three Hong Kong streams. *Freshwater Biology*, 61(1), 57-68.

Jackson, M. C., Jones, T., Milligan, M., Sheath, D., Taylor, J., Ellis, A., ... & Grey, J. (2014). Niche differentiation among invasive crayfish and their impacts on ecosystem structure and functioning. *Freshwater Biology*, 59(6), 1123-1135.

Jackson, M. C., Grey, J., Miller, K., Britton, J. R., & Donohue, I. (2016). Dietary niche constriction when invaders meet natives: evidence from freshwater decapods. *Journal of Animal Ecology*, 85(4), 1098-1107.

Kashian, D. R., Zuellig, R. E., Mitchell, K. A., & Clements, W. H. (2007). The cost of tolerance: Sensitivity of stream benthic communities to UV-B and metals. *Ecological Applications*, 17(2), 365-375.

Kiffney, P. M., & Richardson, J. S. (2001). Interactions among nutrients, periphyton, and invertebrate and vertebrate (Ascaphus truei) grazers in experimental channels. *Copeia*, 2001(2), 422-429.

Korbee, N., Carrillo, P., Mata, M. T., Rosillo, S., Medina-Sánchez, J. M., & Figueroa, F. L. (2012). Effects of ultraviolet radiation and nutrients on the structure–function of phytoplankton in a high mountain lake. *Photochemical & Photobiological Sciences*, 11(6), 1087-1098.

Liess, A., Lange, K., Schulz, F., Piggott, J. J., Matthaei, C. D., & Townsend, C. R. (2009). Light, nutrients and grazing interact to determine diatom species richness via changes to productivity, nutrient state and grazer activity. *Journal of Ecology*, 97(2), 326-336.

Louhi, P., Richardson, J. S., & Muotka, T. (2017). Sediment addition reduces the importance of predation on ecosystem functions in experimental stream channels. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(1), 32-40.

Lozano, V. L., Vinocur, A., y García, C. S., Allende, L., Cristos, D. S., Rojas, D., ... & Pizarro, H. (2018). Effects of glyphosate and 2, 4-D mixture on freshwater phytoplankton and periphyton communities: a microcosms approach. *Ecotoxicology and Environmental Safety*, 148, 1010-1019.

Mackey, K. R., Hunter, D., Fischer, E. V., Jiang, Y., Allen, B., Chen, Y., ... & Paytan, A. (2013). Aerosol-nutrient-induced picoplankton growth in Lake Tahoe. *Journal of Geophysical Research: Biogeosciences*, 118(3), 1054-1067.

MacLennan, M. M., Dings-Avery, C., & Vinebrooke, R. D. (2015). Invasive trout increase the climatic sensitivity of zooplankton communities in naturally fishless lakes. *Freshwater Biology*, 60(8), 1502-1513.

Matthaei, C. D., Piggott, J. J., & Townsend, C. R. (2010). Multiple stressors in agricultural streams: interactions among sediment addition, nutrient enrichment and water abstraction. *Journal of Applied Ecology*, 47(3), 639-649.

McKee, D., Atkinson, D., Collings, S. E., Eaton, J. W., Gill, A. B., Harvey, I., ... & Moss, B. (2003). Response of freshwater microcosm communities to nutrients, fish, and elevated temperature during winter and summer. *Limnology and Oceanography*, 48(2), 707-722.

Mebane, C. A., Schmidt, T. S., & Balistrieri, L. S. (2017). Larval aquatic insect responses to cadmium and zinc in experimental streams. *Environmental Toxicology and Chemistry*, 36(3), 749-762.

Molinos, J. G., & Donohue, I. (2010). Interactions among temporal patterns determine the effects of multiple stressors. *Ecological Applications*, 20(7), 1794-1800.

Muñoz, I., Real, M., Guasch, H., Navarro, E., & Sabater, S. (2001). Effects of atrazine on periphyton under grazing pressure. *Aquatic Toxicology*, 55(3-4), 239-249.

Murdock, J. N., Shields, F. D., & Lizotte, R. E. (2013). Periphyton responses to nutrient and atrazine mixtures introduced through agricultural runoff. *Ecotoxicology*, 22(2), 215-230.

Mustonen, K. R., Mykrä, H., Louhi, P., Markkola, A., Tolkkinen, M., Huusko, A., ... & Muotka, T. (2016). Sediments and flow have mainly independent effects on multitrophic stream communities and ecosystem functions. *Ecological Applications*, 26(7), 2116-2129.

Pagano, M., Koffi, M. A., Cecchi, P., Corbin, D., Champalbert, G., & Saint-Jean, L. (2003). An experimental study of the effects of nutrient supply and Chaoborus predation on zooplankton communities of a shallow tropical reservoir (Lake Brobo, Côte d'Ivoire). *Freshwater Biology*, 48(8), 1379-1395.

Pannard, A., Bormans, M., Lefebvre, S., Claquin, P., & Lagadeuc, Y. (2007). Phytoplankton size distribution and community structure: influence of nutrient input and sedimentary loss. *Journal of Plankton Research*, 29(7), 583-598.

Perez-Fuentetaja, A., McQueen, D. J., & Demers, E. (1996). Stability of oligotrophic and eutrophic planktonic communities after disturbance by fish. *Oikos*, 98-110.

Pérez, M. T., & Sommaruga, R. (2007). Interactive effects of solar radiation and dissolved organic matter on bacterial activity and community structure. *Environmental Microbiology*, 9(9), 2200-2210.

Piggott, J. J., Lange, K., Townsend, C. R., & Matthaei, C. D. (2012). Multiple stressors in agricultural streams: a mesocosm study of interactions among raised water temperature, sediment addition and nutrient enrichment. *PLoS One*, 7(11), e49873.

Piggott, J. J., Townsend, C. R., & Matthaei, C. D. (2015). Climate warming and agricultural stressors interact to determine stream macroinvertebrate community dynamics. *Global Change Biology*, 21(5), 1887-1906.

Pilati, A., Vanni, M. J., González, M. J., & Gaulke, A. K. (2009). Effects of agricultural subsidies of nutrients and detritus on fish and plankton of shallow-reservoir ecosystems. *Ecological Applications*, 19(4), 942-960.

Pizarro, H., Di Fiori, E., Sinistro, R., Ramírez, M., Rodríguez, P., Vinocur, A., & Cataldo, D. (2016). Impact of multiple anthropogenic stressors on freshwater: how do glyphosate and the invasive mussel Limnoperna fortunei affect microbial communities and water quality?. *Ecotoxicology*, 25(1), 56-68.

Portinho, J. L., Nielsen, D. L., Daré, L., Henry, R., Oliveira, R. C., & Branco, C. C. (2018). Mixture of commercial herbicides based on 2, 4-D and glyphosate mixture can suppress the emergence of zooplankton from sediments. *Chemosphere*, 203, 151-159.

Pringle, C. M., & Hamazaki, T. (1998). The role of omnivory in a neotropical stream: separating diurnal and nocturnal effects. *Ecology*, 79(1), 269-280.

Rohr, J. R., & Crumrine, P. W. (2005). Effects of an herbicide and an insecticide on pond community structure and processes. *Ecological Applications*, 15(4), 1135-1147.

Sanches, L. F., Guariento, R. D., Caliman, A., Bozelli, R. L., & Esteves, F. A. (2011). Effects of nutrients and light on periphytic biomass and nutrient stoichiometry in a tropical black-water aquatic ecosystem. *Hydrobiologia*, 669(1), 35-44.

Scott, S. E., Pray, C. L., Nowlin, W. H., & Zhang, Y. (2012). Effects of native and invasive species on stream ecosystem functioning. *Aquatic Sciences*, 74(4), 793-808.

Smith, K. F., & Lester, P. J. (2006). Cyanobacterial blooms appear to be driven by top-down rather than bottom-up effects in the Lower Karori Reservoir (Wellington, New Zealand). *New Zealand Journal of Marine and Freshwater Research*, 40(1), 53-63.

Wendt-Rasch, L., Pirzadeh, P., & Woin, P. (2003). Effects of metsulfuron methyl and cypermethrin exposure on freshwater model ecosystems. *Aquatic Toxicology*, 63(3), 243-256.

Wickham, S. A. (1998). The direct and indirect impact of Daphnia and Cyclops on a freshwater microbial food web. *Journal of Plankton Research*, 20(4), 739-755.

Zhao, S. Y., Sun, Y. P., Lin, Q. Q., & Han, B. P. (2013). Effects of silver carp (Hypophthalmichthys molitrix) and nutrients on the plankton community of a deep, tropical reservoir: an enclosure experiment. *Freshwater Biology*, 58(1), 100-113.

Zingel, P., Cremona, F., Nõges, T., Cao, Y., Neif, É. M., Coppens, J., ... & Jeppesen, E. (2018). Effects of warming and nutrients on the microbial food web in shallow lake mesocosms. *European Journal of Protistology*, 64, 1-12.

Summary figures for empirical data

Additional figures detailing the geographic distribution of papers from which empirical data was extracted (Figure S2.8.1) and the number of papers per year that were included in this analysis, and the number of papers per year that were revealed in the systemic literature search (Figure S2.8.2).

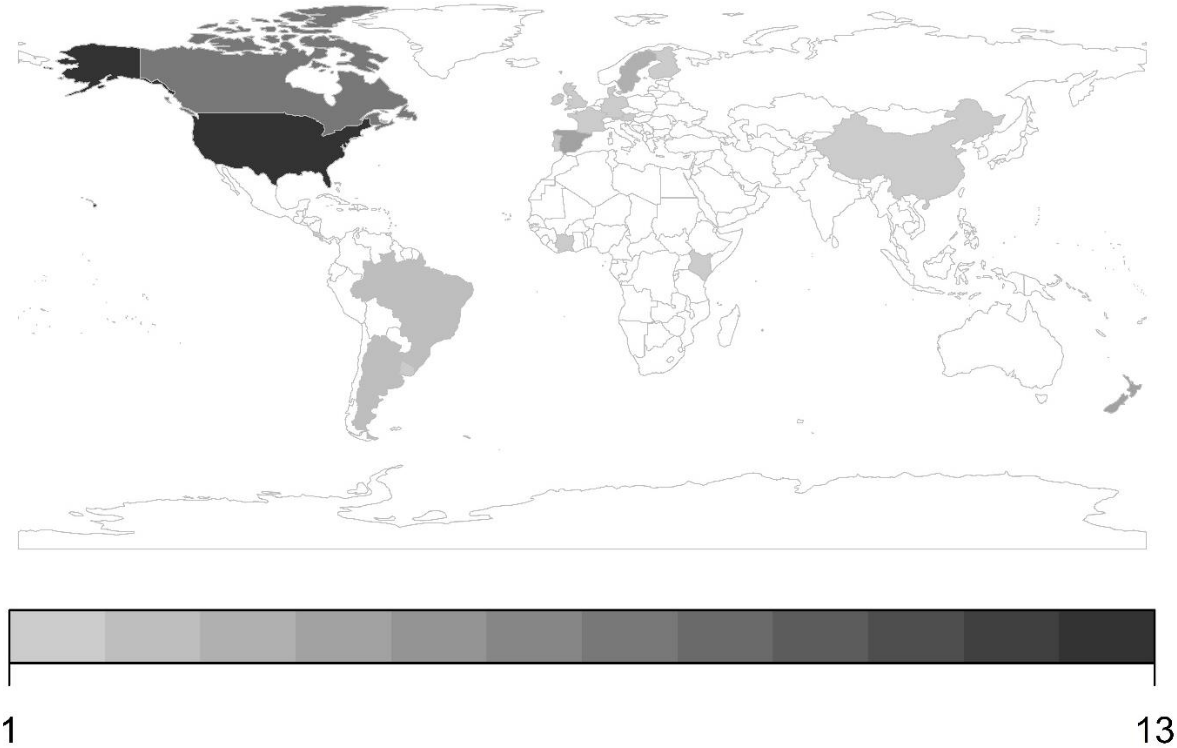

Figure S2.8.1: Number of papers, from which data was extracted, that were included in our analysis shown by country of origin. Darker colours indicate more papers originating from the country in question. Countries coloured white indicate no papers originating from the country in question.

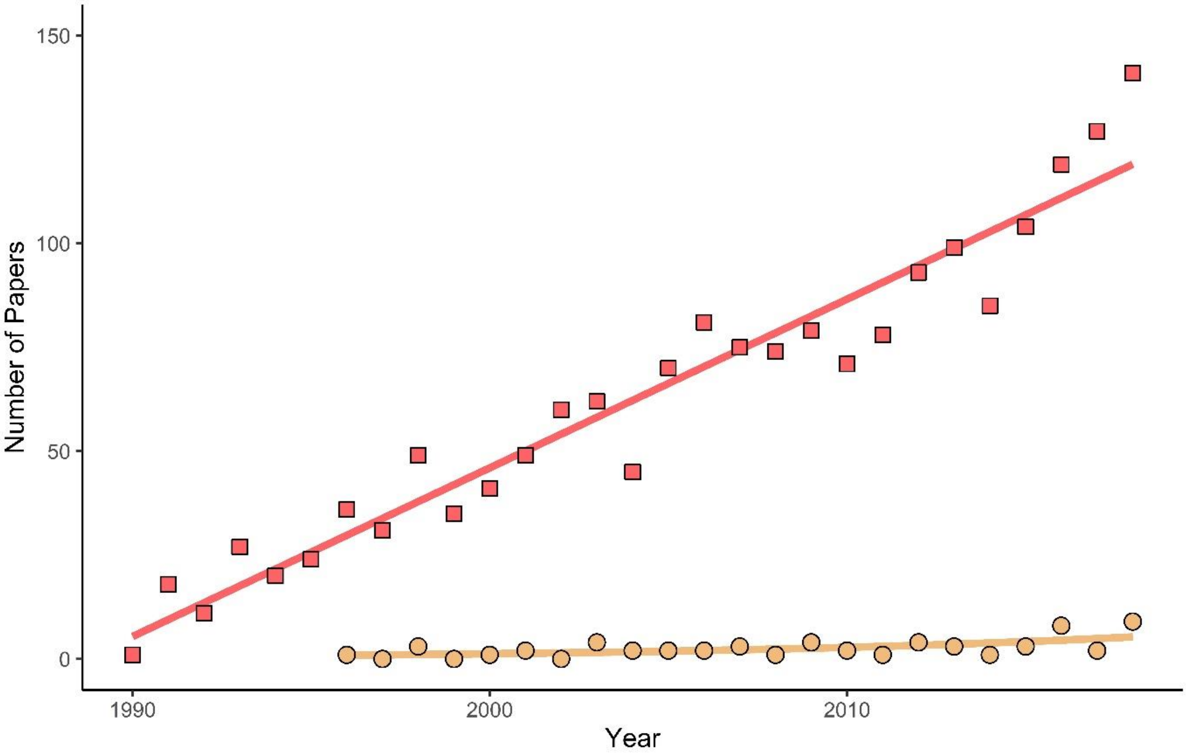

Figure S2.8.2: Number of papers by year that were revealed by the systematic literature review (red squares) and those that were included in our study (orange circles). Lines indicate generalised linear models fit to the data, with Year as the explanatory variable and Number of Papers as the response variable, both Gaussian and Poisson distributions were used to fit the models, with AIC values used to select the best fitting model. For the systematic literature review, a Gaussian distribution with intercept = -8074.1961, p<0.001; and coefficient =4.0601, p<0.001 was the best fitting model. For the papers included within our analysis, a Poisson distribution with intercept = -160.25750, p<0.001; and coefficient = 0.08024, p<0.001 was the best fitting model.

Appendix Three

Supplementary material for Chapter Three.

A table of contents is provided below.

Theoretical framework

Outlined below is the detailed method for the generation of data using a food chain model, alongside its subsequent analysis. The theoretical framework used here is based on that implemented by Burgess et al., (2021).

Food chain model

The food chain model we use within our analysis is based on equations from Heath et al., (2014) (Equation S3.1). These equations are used to calculate the change over time in density for each trophic level and contain a form of self-regulation in density-dependent mortality (i.e. the trophic level experiences a greater rate of mortality under higher densities, compared to lower densities). As described in Section 3.3, the equations comprise four different parameters which represent: i) attack rate (α); ii) conversion efficiency (ε); iii) density independent mortality (δ); iv) density dependent mortality (λ). The food chains modelled within our analysis were able to comprise either three, four, or five different trophic levels; although regardless of the length of food chain modelled, the same equations (Equation S3.1) were used. However, within a given food chain, each of the different trophic levels were modelled using different equations. The change in density of the bottom trophic level (x_0) - representing the pool of an essential element, mineral, or nutrient - was modelled using Equation S3.1.1. Here, the change in x_0 is modelled as the constant input of said nutrient minus the uptake of the nutrient by the trophic level directly above the nutrient pool (i.e. x_1). Accordingly, the change in density of all trophic levels (x_i) aside from the bottom (x_0) and top (x_n) trophic levels is modelled using Equation S3.1.2. Here, the change in x_i is modelled as the increase in density due to conversion of predated resources (x_{i-1}), minus the decrease in density due to predation by the above trophic level (x_{i+1}), density independent mortality, and density dependent mortality. Finally, the change in density of the top (x_n) trophic level is modelled using Equation S3.1.3. Here, the change in x_n is modelled as the increase in density as a result of predating on the below trophic level (x_{n-1}) minus the change in density due to both density dependent and density independent mortality. For the top trophic level there is no form of predation from an above trophic level.

$$Equation\ S3.1.1: \qquad \frac{dx_0}{dt} = \omega - \alpha_1 x_0 x_1$$

$$Equation\ S3.1.2: \qquad \frac{dx_i}{dt} = \alpha_i \varepsilon_i x_{i-1} x_i - \alpha_{i+1} x_i x_{i+1} - \delta_i x_i - \lambda_i x_i^2$$

$$Equation\ S3.1.3: \qquad \frac{dx_n}{dt} = \alpha_n \varepsilon_n x_{n-1} x_n - \delta_n x_n - \lambda_n x_n^2$$

Data generation

Building on the description given in Section 3.3, the theoretical framework generated data in the following manner. Firstly, for each length of food chain, 220,000 parameter combinations were randomly determined (i.e., a total of 660,000 parameter combinations). For each combination, each parameter was assigned a random value from a predetermined uniform distribution, though these distributions differed for each parameter (Table S3.1.1). These combinations and parameter values are hence referred to as the *control* parameter values.

Table S3.1.1: Uniform distributions from which the control values of each parameter could be randomly drawn.

Parameter	**Control parameter distribution**
α	U ~ (0.25, 0.75)
ε	U ~ (0.25, 0.75)
δ	U ~ (0.25, 0.75)
ω	U ~ (25, 75)
λ	U ~ (0.00625, 0.025)

Secondly, for each combination of parameters, two were randomly selected (with all parameters equally likely to be chosen) to be subjected to stress and have their values altered. Accordingly, all other parameter values remain the same as their control values, though the values of the stressed parameters differ. The values of stressed parameters were randomly drawn from the uniform distributions shown by Table S3.1.2; following the method of Burgess et al. (2021), only the parameters of attack rate, conversion efficiency, density independent mortality, or nutrient input rate were randomly selected to be stressed. The single restriction that was placed on the stressed parameter values, was that they were unable to be in the range of the control parameter value ± 0.025 (except for ω where the excluded range was the control parameter value ± 2.5). The rationale for this limitation was to make sure that the stressed parameter values were meaningfully different to the control parameter values. The parameters which were randomly selected to be stressed could be the same processes across differing trophic levels, differing processes on the same trophic level, or differing processes across differing trophic levels. Furthermore, α_1 was never able to be selected to be stressed. This follows the method of Burgess et al. (2021), which describes how the alteration of α_1 only impacted the bottom trophic level, not any other trophic level. Given that our method follows that of Burgess et al. (2021) in investigating only non-basal trophic levels, α_1 was not stressed in our analysis.

Table S3.1.2: Uniform distributions from which the stressed values of each parameter could be randomly drawn.

Parameter	**Stressed parameter distribution**
α	U ~ (0.01, 0.99)
ε	U ~ (0.01, 0.99)
δ	U ~ (0.01, 0.99)
ω	U ~ (1, 99)

As described in the main text, the theoretical framework implements a factorial design similar to those found within experimental studies. Accordingly, corresponding treatments are: i) *Control parameter values*; *Stressor 1 parameter values* (i.e. all parameters have the same values as the control parameters with the exception that the first stressed parameter has its parameter value changed to the stressed value); *Stressor 2 parameter values* (i.e. all parameters have the same values as the control parameters with the exception that the second stressed parameter has its parameter value changed to the stressed value); *Interaction parameter values* (i.e. all parameters have the same values as the control parameters with the exception that both stressed parameters have their parameter values changed to the stressed values). In totality, we investigated 2,640,000 different parameter combinations for individual food chains, with each treatment having a corresponding 660,000 parameter combinations.

Thirdly, as described in the main text, the focus of this analysis was on equilibrium densities. Here, equilibrium densities refer to the density of each trophic level in a given food chain so that there is no net change in the density of any other trophic level. Hence, for each of the 2,640,000 parameter combinations, the equilibrium densities were determined using Mathematica v10.4 (Wolfram Research Inc., 2016). We placed additional constraints on the equilibrium densities, ensuring that they were greater than zero (representing biologically feasible densities). Where a feasible density was unable to be determined (for any of the four treatments within the factorial design), the entire factorial treatment design (i.e. control, stressor 1, stressor 2, and interaction equilibrium densities) was excluded from any subsequent analysis. Finally, it was possible to determine equilibrium densities for every treatment for 516,349 (78.23%) of all factorial treatments. Accordingly, from these 516,349 factorial treatments, 100,000 were randomly selected (with approximately one third of interactions coming from each length of food chain). For each of these 100,000 factorial treatments, a single trophic level was randomly selected with the equilibrium densities of this trophic level, under each of the factorial treatments, being collated into a single dataset, upon which all further analysis was conducted. As shown by Burgess et al. (2021), there are no significant differences in the responses of the various food chain lengths, hence allowing interactions from different food chain lengths to be pooled.

Data analysis – Treatment means

As described in Chapter Three, question one centres on understanding both null models in the absence of uncertainty. Accordingly, using the exact equilibrium densities (i.e. treatment means) provides the best opportunity to determine the 'true' interaction classification under each null model and the conditions under which the null models agree or disagree in their classifications. Accordingly, the above dataset was used to address question one.

Data analysis – Treatment uncertainty

Following the method of Burgess et al., (2021), uncertainty was incorporated into the theoretical framework in the form of observation error. Firstly, observation error was modelled using a normal distribution, with a mean of 1.00 and standard deviation of σ. Here, σ represents the level of observation error to be included. Within this analysis, 250 levels of observation error were used ranging from 0.001 to 0.25, in increments of 0.001. Secondly, each of the 100,000 factorial treatments was randomly assigned a number of replicates (*n*) (for this analysis, *n* was assigned a constant value of four replicates across all treatments). Finally, for each treatment within a factorial treatment design, *n* values from a normal distribution with a standard deviation of σ and a mean of 1.00 were drawn. The single limit placed upon these values was that they could not be greater than or equal to 2.00, nor less than or equal to 0.00. The equilibrium density for a given treatment were then multiplied by these values, with the mean and standard deviation of these new *n* equilibrium densities being determined. Accordingly, each factorial treatment design has corresponding values for the mean, standard deviation, and sample size for every treatment. As such, the additive and multiplicative null models are then able to be applied to each factorial treatment design.

Data analysis – Sample size

In order to understand the role of sample size in the results of both null models (i.e. question three), a similar method to that used to explore treatment uncertainty was implemented. Indeed, the implemented method, differs only to that used for treatment uncertainty in two fundamental ways. Firstly, rather than consider 250 levels of observation error, three levels were selected (0.001, 0.01, and 0.1). Secondly, rather than randomly drawing the sample size from a predetermined range for each factorial treatment design, the assigned sample size was systematically varied between three and 100 replicates (in steps of a single replicate).

As for treatment uncertainty, for each treatment within a factorial treatment design, *n* values (i.e. iteratively varied between 3 and 100) from a normal distribution with a standard deviation of σ (i.e. varied between 0.001, 0.01, and 0.1) and a mean of 1.00 were drawn. As before, the single limit placed upon these values was that they could not be greater than or equal to 2.00, nor less than or equal to 0.00. The equilibrium density for a given treatment was then multiplied by these values, with the mean and standard deviation of these new *n* equilibrium densities being determined.

As before, each factorial treatment design has corresponding values for the mean, standard deviation, and sample size for every treatment. Therefore, the additive and multiplicative null models are then able to be applied to each factorial treatment design.

References

Burgess, B. J., Purves, D., Mace, G., & Murrell, D. J. (2021). Classifying ecosystem stressor interactions: Theory highlights the data limitations of the additive null model and the difficulty in revealing ecological surprises. *Global Change Biology*.

Heath, M. R., Speirs, D. C., & Steele, J. H. (2014). Understanding patterns and processes in models of trophic cascades. *Ecology Letters*, 17(1), 101-114.

Wolfram Research. Inc. (2016). Mathematica (Version 10.4). *Wolfram Research, Inc., Champaign, Illinois.*

Interaction classifications

As described in Section 3.3, an interaction can be classified as a synergistic, antagonistic, reversal, or null interaction using either the additive (Gurevitch et al., 2000) or multiplicative (Lajeunesse, 2011) null models. As detailed in Section 3.3, the classification of an interaction is dependent upon the effect size (ES), associated confidence intervals ($CI_{95\%}$), the observed ($Obs.$) and predicted ($Pred.$) interactive effect of stressors.

For the additive null model, the $Pred.$ can be calculated as $X_A + X_B - 2X_C$; while for the multiplicative null model, $Pred.$ can be calculated as $\ln\left(\frac{X_A X_B}{X_C^2}\right)$. Similarly, for the additive null model, $Obs.$ can be calculated as $X_I - X_C$; while for the multiplicative null model, $Obs.$ Can be calculated as $\ln\left(\frac{X_I}{X_C}\right)$.

Using this scheme, interactions were classified as follows:

When $Pred. \geq 0$

1) If $ES \pm CI_{95\%}$ overlaps zero, then the interaction was assigned a null classification.
2) If $ES > 0$, the interaction was assigned a synergistic classification.
3) If $ES < 0$, and if $Obs.$ and $Pred.$ both have the same polarity, the interaction was assigned an antagonistic classification.
4) If $ES < 0$, and if $Obs.$ and $Pred.$ have contrasting polarities, the interaction was assigned a reversal classification.

When $Pred. < 0$

1) If $ES \pm CI_{95\%}$ overlaps zero, then the interaction was assigned a null classification.
2) If $ES < 0$, the interaction was assigned a synergistic classification.
3) If $ES > 0$, and if $Obs.$ and $Pred.$ both have the same polarity, the interaction was assigned an antagonistic classification.
4) If $ES > 0$, and if $Obs.$ and $Pred.$ have contrasting polarities, the interaction was assigned a reversal classification.

Question one was addressed in the absence of any treatment uncertainty, hence there are no standard deviations, or effect sizes associated with any treatments (i.e. only means). Accordingly, for the analysis centred on this question, slightly different steps to classify interactions were used:

When $Pred. \geq 0$

1) If $ES = 0$, then the interaction was assigned a null classification.
2) If $ES > 0$, the interaction was assigned a synergistic classification.
3) If $ES < 0$, and if $Obs.$ and $Pred.$ both have the same polarity, the interaction was assigned an antagonistic classification.
4) If $ES < 0$, and if $Obs.$ and $Pred.$ have contrasting polarities, the interaction was assigned a reversal classification.

When $Pred. < 0$

1) If $ES = 0$, then the interaction was assigned a null classification.
2) If $ES < 0$, the interaction was assigned a synergistic classification.
3) If $ES > 0$, and if $Obs.$ and $Pred.$ both have the same polarity, the interaction was assigned an antagonistic classification.
4) If $ES > 0$, and if $Obs.$ and $Pred.$ have contrasting polarities, the interaction was assigned a reversal classification.

References

Gurevitch, J., Morrison, J.A. and Hedges, L.V. (2000). The interaction between competition and predation: a meta-analysis of field experiments. *The American Naturalist*, 155(4), 435-453.

Lajeunesse, M.J. (2011). On the meta-analysis of response ratios for studies with correlated and multi-group designs. *Ecology*, 92(11), 2049-2055.

Treatment means

As described in Section 3.4 (see *Treatment means*), the classification of interactions may vary between null models. Indeed, within our analyses the different null models assigned contrasting classifications to over one third of all interactions. However, the differences (and indeed similarities) between the classifications of the different null models can be attributed to their underlying algebraic relationships. Accordingly, by combining and rearranging both null model equations, it is possible to derive two inequalities which, when met, detail the conditions under which the additive and multiplicative null models report differing interaction classifications.

$$Inequality\ S3.3.1: \qquad 0 \leq \frac{X_I X_C - X_A X_C - X_B X_C + X_C^2}{X_A X_B - X_A X_C - X_B X_C + X_C^2} \leq 1$$

$$Inequality\ S3.3.2: \qquad 0 < \frac{2X_C^2 - X_A X_C - X_B X_C}{X_A X_B - X_A X_C - X_B X_C + X_C^2} < 1$$

X_C, X_A, X_B, and X_I represent mean response values under control, the first stressor, the second stressor, and both stressors respectively. If either inequality is met, then the interaction will be classified differently between the additive and multiplicative null models.

Interpretation of the inequalities can be explained as follows, illustrated by Figure S3.3.1. If only Inequality S3.3.1 is met, an interaction will be classified differently between the additive and multiplicative null models, though neither null model classifies an interaction as being a reversal interaction (blue points, Figure S3.3.1). If only Inequality S3.3.2 is met, an interaction will be classified differently between the null models, with one null model reporting an antagonistic interaction while the other reports a reversal interaction (red points, Figure S3.3.1). If both Inequalities S3.3.1 and S3.3.2 are met, an interaction will be classified differently between the null models, with one null model reporting a reversal interaction and the other reporting either an antagonistic, null, or synergistic interaction (purple points, Figure S3.3.1).

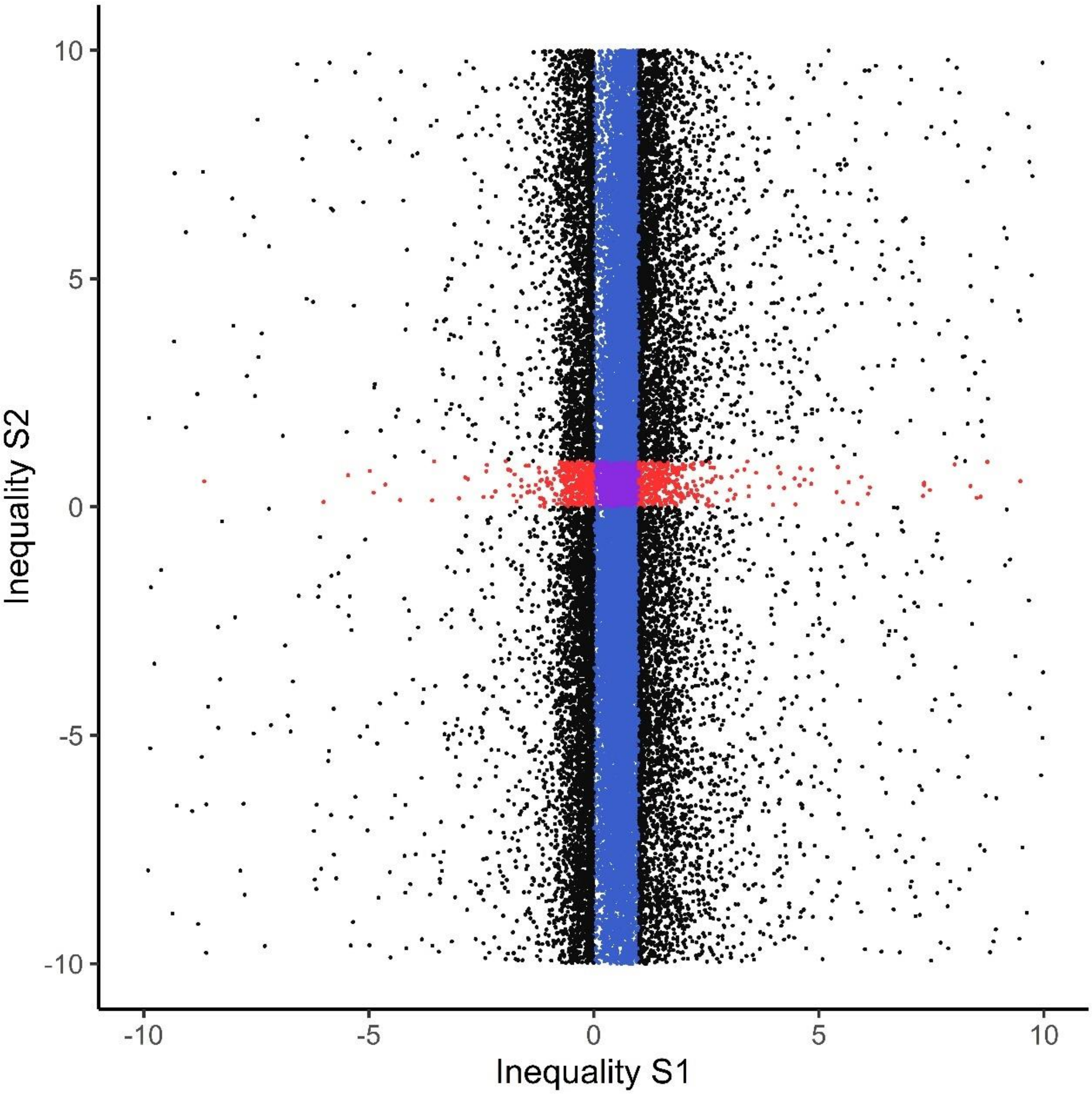

Figure S3.3.1: Values of Inequality S3.3.1 and Inequality S3.3.2 for each interaction in the absence of observation error. Points in black denote interactions were neither inequality was met, hence the interaction was assigned the same classification by both the additive and multiplicative null models. Blue points denote interactions where Inequality S3.3.1 was met, but Inequality S3.3.2 was not. Red points indicate interactions where Inequality S3.3.2 was met, but Inequality S3.3.1 was not. Purple points indicate interactions where both inequalities were met. Refer to the above explanation of each inequality, and the interactions assigned by each null model.

Treatment uncertainty

The reason that observation error leads to the dominance of null interactions (e.g., Figure 3.2) is as follows. For the multiplicative null model, the effect size variance is proportional to the standard deviation (Equation 3.2.2). Hence higher levels of observation error result in larger variances whilst the effect size itself remains unaffected. For the additive null model higher levels of observation error reduce both the magnitude of the effect size (Equations 3.1.1 – 3.1.2), and the magnitude of the effect size relative to the variance of the effect size (Equation 3.1.4). Accordingly, if all other variables remain the same, higher levels of observation error would lead to larger confidence intervals surrounding the multiplicative effect size (increasing the likelihood that the interaction effect size will be non-significant). Alternatively, higher levels of observation error would lead to greater reduction in the magnitude of the additive effect size compared to the reduction in the associated confidence intervals (hence increasing the likelihood that the interaction effect size will be non-significant).

Level of significance

As outlined in the discussion, the choice of significance level may be justifiably altered, to increase statistical power of an experiment, when alternatives (e.g., increasing sample sizes) are not possible. The potential impacts of increased significance levels upon the frequency of correctly classified interactions (across sample sizes and levels of uncertainty) are shown by Figure S3.4.2. Three different levels of significance (0.05, 0.10,0.20) are used.

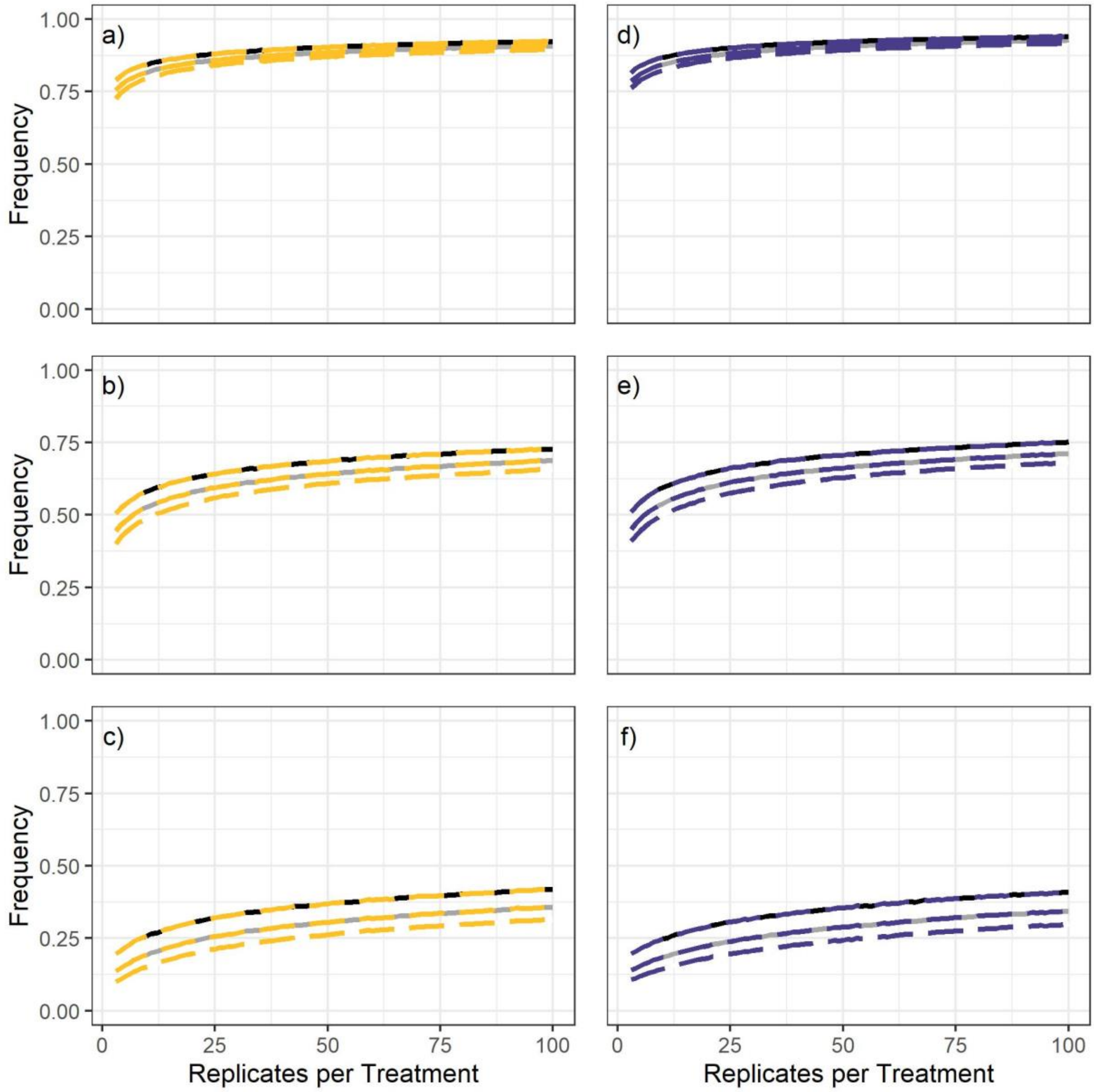

Figure S3.4.2: Frequency of interactions that are correctly classified (i.e., true positive rate) for a range of different treatment replicates. The frequency of interactions correctly classified by the additive null model (a,b,c) are shown with yellow lines, the frequency of interactions correctly classified by the multiplicative (d,e,f) are shown with purple lines. Three different levels of observation error are considered: 0.001 (a,d), 0.01 (b,e), and 0.1 (c,d). The level of significance is denoted by the secondary colour of the dashed lines: α=0.05 – white dashed lines; α=0.10 – grey dashed lines; α=0.20 – black dashed lines.

Appendix Four

Supplementary material for Chapter Four.

A table of contents is provided below.

Categorial data

Multiple stressor experiments are wide-ranging, considering a broad range of stressors, response metrics, and other covariables. As such, following the methods of previous multiple stressor meta-analyses (e.g., Crain et al., 2008; Jackson et al., 2016), we devise categories, which these covariables are subsequently grouped into. Here, we form categories for stressors, realm, response metrics, and organisation level. These categories (alongside examples where necessary) are outlined below.

Stressor categories

Multiple stressor experiments consider a diverse range of stressors. Accordingly, within this analysis, stressors have been grouped into the following nine distinct categories:

i)	Temperature	(e.g., temperature increase)
ii)	Light	(e.g., UVB radiation)
iii)	pH	(e.g., CO_2 increase or acidification)
iv)	Salinity	(e.g., increased salinity or increased ionic load)
v)	Community alteration	(e.g., predator removal or addition of invasive species)
vi)	Habitat alteration	(e.g., dredging or sedimentation)
vii)	Disease	(e.g., *Batrachochytrium dendrobatidis*)
viii)	Nutrients	(e.g., phosphate or ammonium nitrate)
ix)	Contamination	(e.g., sertraline or cadmium)

Accordingly, when stressor combinations are determined, they are classed as a combination of two of the above categories (e.g., Contamination x Temperature, or Disease x Nutrients).

Realm

Within this analysis, data has been collated from seven different aquatic multiple stressor meta-analyses. Of these, four meta-analyses focus on freshwater environments, while three focus on marine ecosystems. As such, the vast majority of the data is assigned a realm of either *freshwater* or *marine*. Such classifications reflect experiments that were conducted in, or on organisms from, either freshwater or marine environments (e.g., coastal environments, coral reefs, ponds, rivers, or lakes). A small minority of interactions were assigned a category of *Other*. This category reflects experiments conducted in, or on organisms from, environments which are exposed to both fresh and saline waters; examples of such environments include coastal lagoons or marshes.

Response metrics

The meta-analyses from which our data is collated, consider a wide-range of different response metrics. Here, we group these into twelve categories (shown below). Of these, some (e.g., biomass, density, and survival) are more prevalent than others. As outlined in the Section 4.3, within our analyses, survival data was expressed as proportions while biomass and density were required to be expressed in untransformed units (e.g., individuals per litre).

i)	Biomass	(e.g., biomass or chlorophyll-a)
ii)	Calcification	(e.g., calcification rate)
iii)	Decomposition	(e.g., decomposition rate)
iv)	Density	(e.g., abundance or density)
v)	Development	(e.g., germination rate)
vi)	Diversity	(e.g., species richness)
vii)	Growth	(e.g., growth rate)
viii)	Physical	(e.g., length or height)
ix)	Physiological	(e.g., photosynthetic rate or hormone concentration)
x)	Productivity	(e.g., production rate)
xi)	Reproduction	(e.g., reproduction rate)
xii)	Survival	(e.g., survival rate)

Organisation level

Multiple stressor experiments are capable of being conducted at a range of levels of ecological organisation. Accordingly, we categorise interactions based upon the organisation level they consider. As such, interactions were categorized as being either *organism* (i.e., individual organisms were measured), *population* (i.e., an attribute of an entire population was measured), or *community* (i.e., an attribute of more than one population, such as zooplankton, was measured).

References

Crain, C. M., Kroeker, K., & Halpern, B. S. (2008). Interactive and cumulative effects of multiple human stressors in marine systems. *Ecology Letters*, 11(12), 1304-1315.

Jackson, M. C., Loewen, C. J., Vinebrooke, R. D., & Chimimba, C. T. (2016). Net effects of multiple stressors in freshwater ecosystems: a meta-analysis. *Global Change Biology*, 22(1), 180-189.

Vote-counting

The frequency at which different interaction classifications are assigned to individual interactions (i.e., vote-counting results) for hypotheses two (Figure S4.2.1), three, and four (Figure S4.2.2) are shown below.

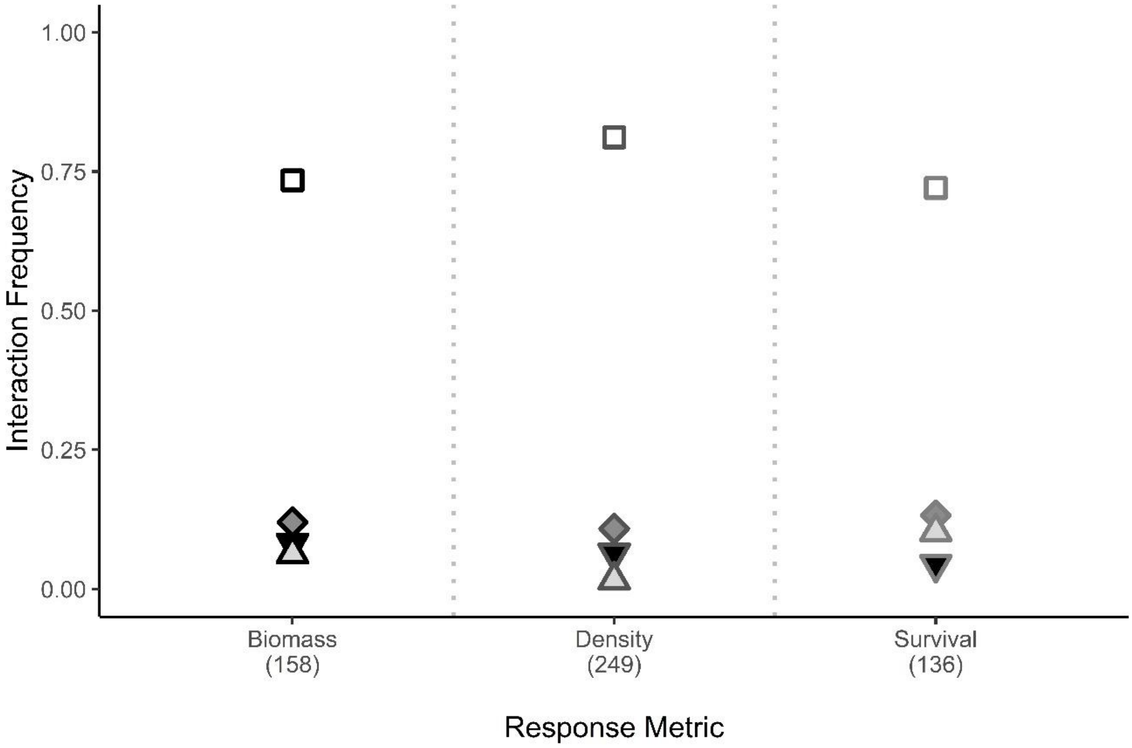

Figure S4.2.1: Vote-counting results from the analyses for hypothesis two. The frequencies of null (white squares), synergistic (light-grey upward triangles), antagonistic (dark grey diamonds), and reversal interactions (black downward triangles) are shown. Numbers in brackets refer to the number of interactions for each response metric.

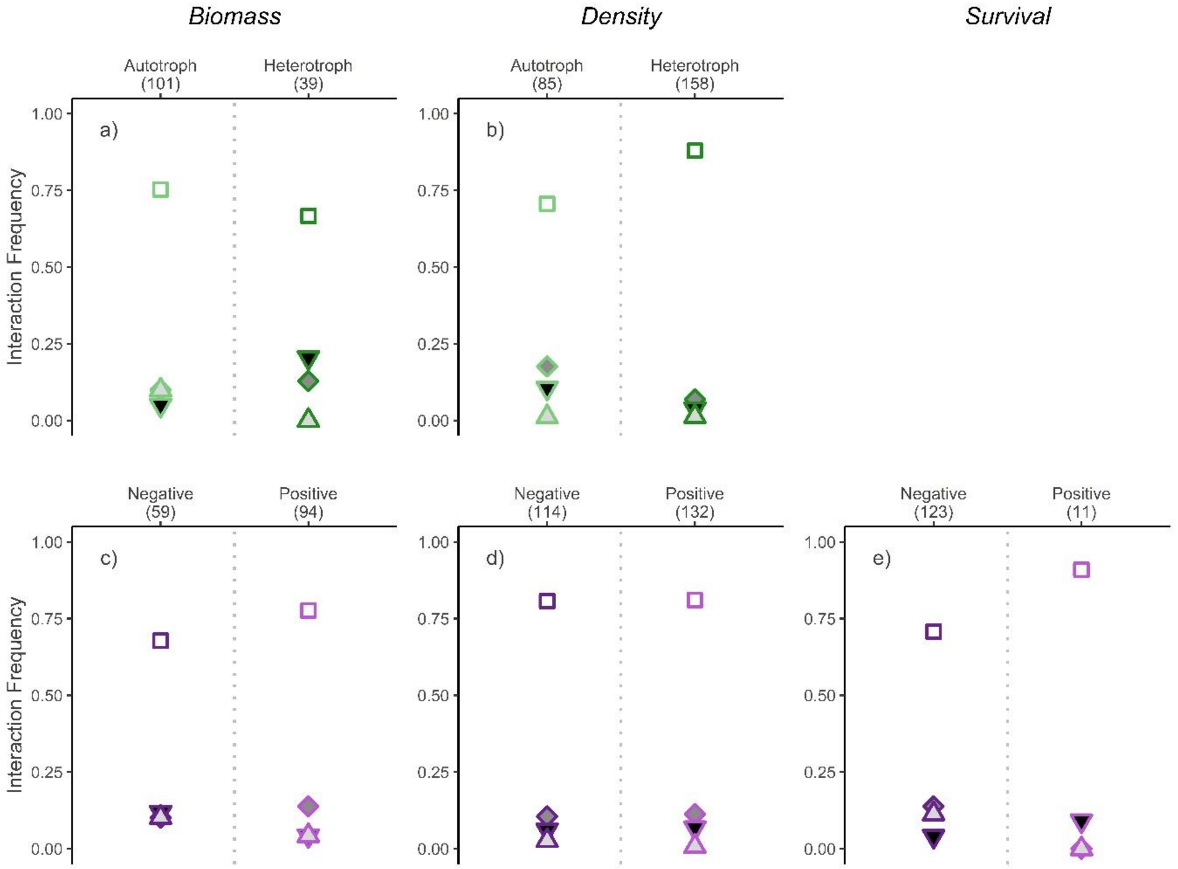

Figure S4.2.2: Vote-counting results from the analyses for hypotheses three (a-b) and four (c-e). The frequencies of null (white squares), synergistic (light-grey upward triangles), antagonistic (dark grey diamonds), and reversal interactions (black downward triangles) are shown. Numbers in brackets refer to the number of interactions for each group.

Model diagnostics

As outlined in Section 4.3, we conduct model diagnostics through the use of $\hat{R}$ (leverage), Cook's Distance, and DFBETA to assess for potentially influential results which may significantly affect the conclusions of any analysis. The results of these analyses are shown for hypothesis two (Figures S4.2.3 – S4.2.5) and hypotheses three and four (Figures S4.2.6 – S4.2.8). This analysis revealed three influential data points (all for biomass data in the analysis for hypothesis four). As such, these data points were removed for this (and only this) analysis (Figure 4.5c). The data was compared to the original sources with no issues with the data apparent. Likewise, potential ecological rationale for these influential data points could not be found.

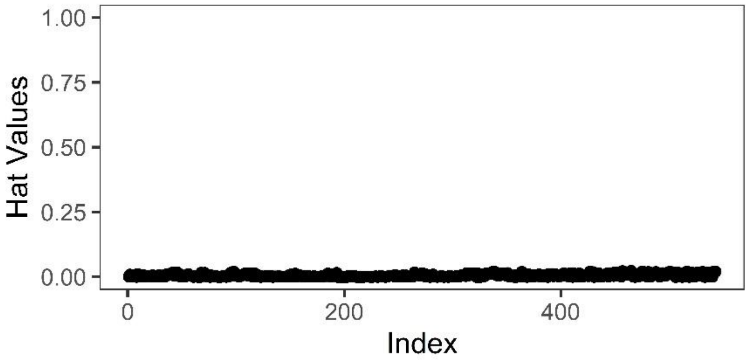

Figure S4.2.3: $\hat{R}$ values for the meta-analytical analysis of hypothesis two (Figure 4.4). X-axis is interaction number (i.e., no bearing on $\hat{R}$ values).

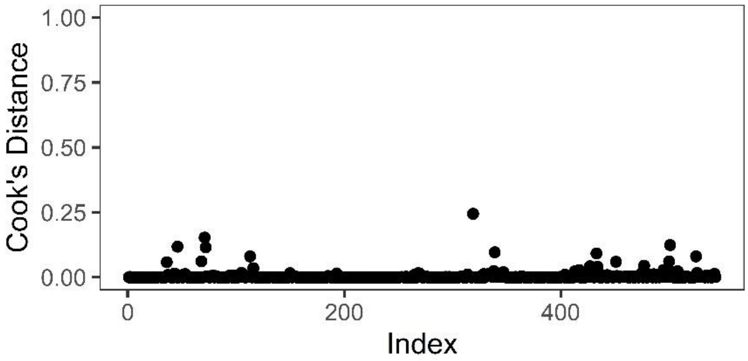

Figure S4.2.4: Cook's Distance values for the meta-analytical analysis of hypothesis two (Figure 4.4). X-axis is interaction number (i.e., no bearing on Cook's Distance values).

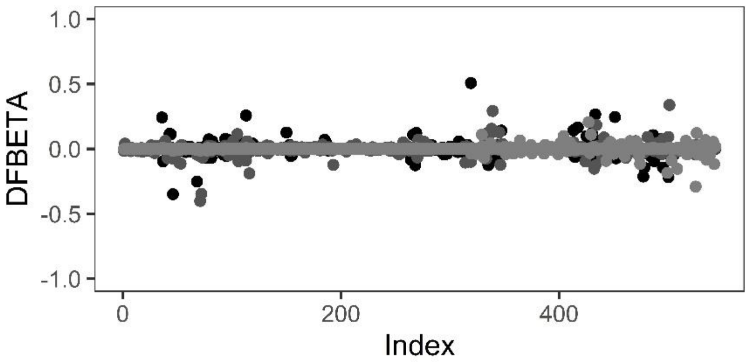

Figure S4.2.5: DFBETA values for the meta-analytical analysis of hypothesis two (Figure 4.4). X-axis is interaction number (i.e., no bearing on DFBETA values). Colour scheme same as in Figure 4.4.

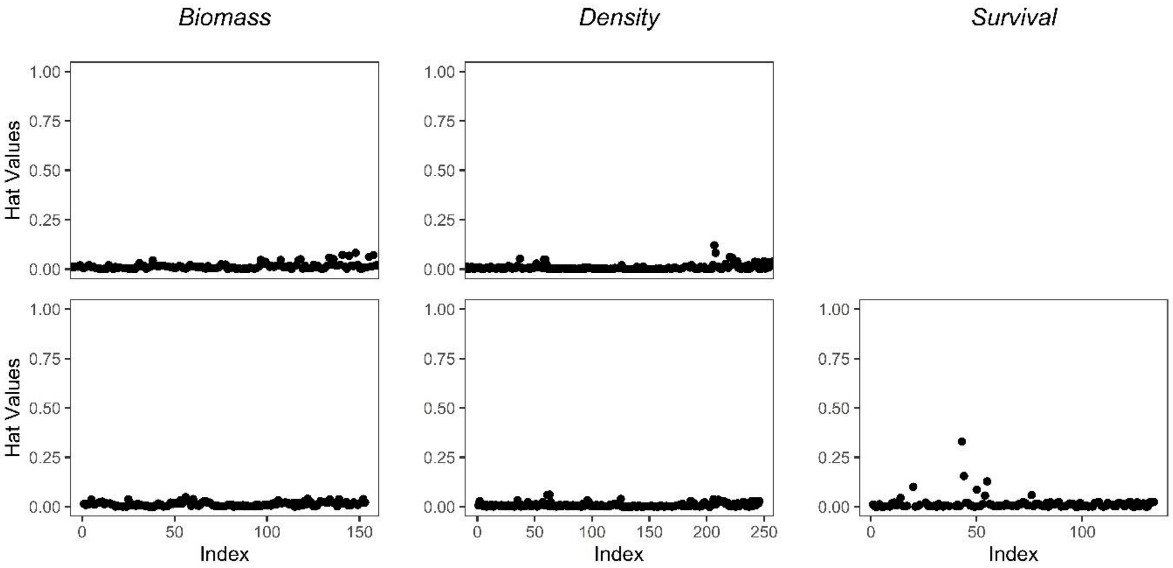

Figure S4.2.6: $\hat{R}$ values for the meta-analytical analysis of hypotheses three (Figure 4.5ab) and four (Figure 4.5cde). X-axis is interaction number (i.e., no bearing on $\hat{R}$ values).

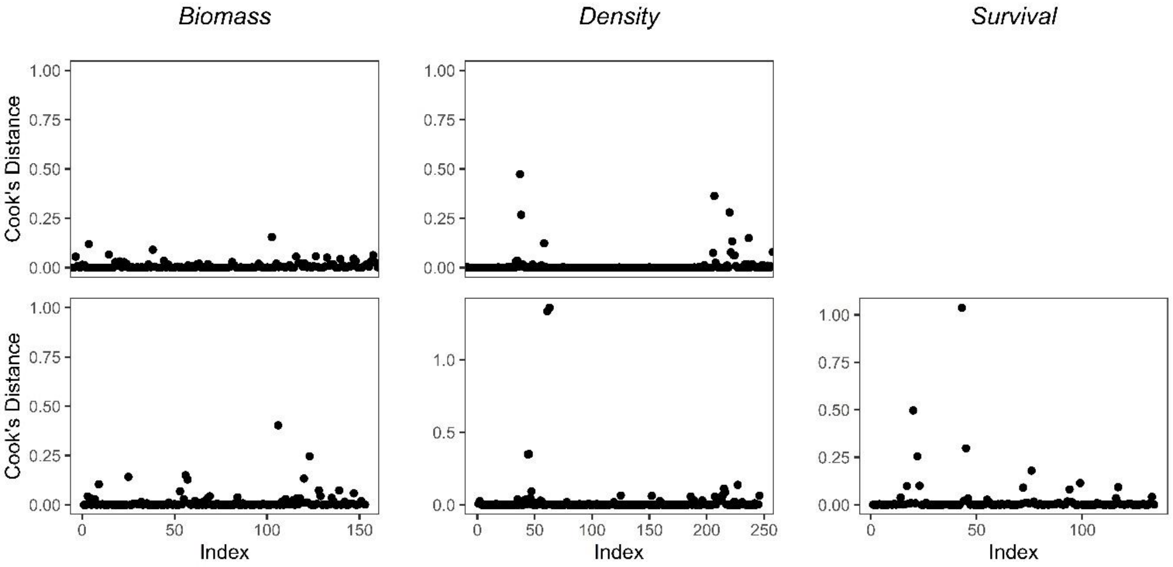

Figure S4.2.7: Cook's Distance values for the meta-analytical analysis of hypotheses three (Figure 4.5ab) and four (Figure 4.5cde). X-axis is interaction number (i.e., no bearing on Cook's Distance values).

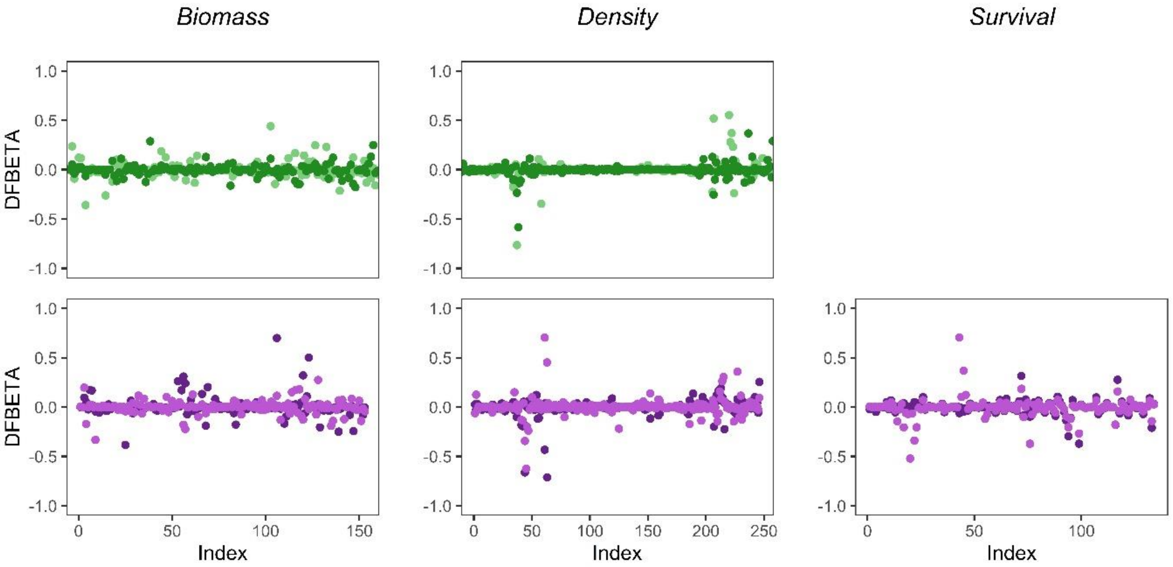

Figure S4.2.8: DFBETA values for the meta-analytical analysis of hypotheses three (Figure 4.5ab) and four (Figure 4.5cde). X-axis is interaction number (i.e., no bearing on DFBETA values). Colour scheme same as in Figure 4.5.

Heterogeneity

I^2 values were calculated as a measure of heterogeneity in meta-analytical models. The I^2 values are classed as high using the values suggested by Higgins et al., (2003); however, within ecological meta-analyses I^2 values are frequently in excess of 90% (Senior et al., 2016). I^2 values are shown in Table S4.3.1.

Table S4.3.1: I^2 values for the meta-analytical models conducted on aggregated datasets.

Dataset	Fixed effect	I^2
Aggregated dataset	Response metric	95.8%
Aggregated biomass dataset	Feeding group	86.1%
Aggregated density dataset	Feeding group	74.8%
Aggregated biomass dataset	Expected direction of effects	87.4%
Aggregated density dataset	Expected direction of effects	82.7%
Aggregated survival dataset	Expected direction of effects	96.8%

References

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557-560.

Senior, A. M., Grueber, C. E., Kamiya, T., Lagisz, M., O'Dwyer, K., Santos, E. S., & Nakagawa, S. (2016). Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. *Ecology*, 97(12), 3293-3299.

Covariance-variance matrices

All meta-analytical models were conducted using the *rma.mv* function from the *metafor* package (v2.1.0) (Viechtbauer, 2010) in R (v3.6.1). All of the datasets considered here included data from empirical studies which shared the same control. For instance, the same baseline may be used to compare multiple different combinations of stressors, or many intensities of the same stressors. Accordingly, covariance exists between these interactions where a control (or a control and a single treatment) are common across multiple interactions. As such, within these models we implement covariance-variance matrices, which allow the meta-analytical models to account for covariance between effect sizes. An example of a covariance-variance matrix is shown below (Equation S4.4.1)

$$\begin{pmatrix} var_a & 0 & 0 & 0 & 0 & 0 \\ 0 & var_b & cov_{b,c} & 0 & 0 & 0 \\ 0 & cov_{b,c} & var_c & 0 & 0 & 0 \\ 0 & 0 & 0 & var_d & cov_{d,e} & cov_{d,f} \\ 0 & 0 & 0 & cov_{d,e} & var_e & cov_{e,f} \\ 0 & 0 & 0 & cov_{d,f} & cov_{e,f} & var_f \end{pmatrix}$$

Equation S4.4.1: Example of a covariance-variance matrix. The major diagonal contains the variances (var_x) for effect sizes *a – f.* Off-diagonals are either $cov_{x,y}$, indicating that effect sizes *x* and *y* share the same control, or *0*, indicating no covariance between effect sizes. In this example, effect size *a* does not share a control with any other effect size. Effect sizes *b* and *c* share the same control as one another. Similarly, effect sizes *d*, *e*, and *f* all share the same control indicated by the relevant off-diagonal having a non-zero value.

References

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48.

Significance level

Throughout our analyses, we set α = 0.05. While there is a discussion about the implications of multiple-testing (i.e., multiple comparisons) and Type I errors in meta-analyses, there is no consensus in how to account for this issue (Borenstein, 2009). As noted by Borenstein (2009), suggestions include: not adjusting α, but evaluating results in context; using a lower (yet still arbitrary significance criteria e.g. α = 0.01); or lowering α using a procedure such as the Benjamini-Hochberg correction. However, it is important to note that for meta-analyses, procedures such as the Bonferroni correction are unlikely to be an adequate solution to addressing issues with multiple comparisons (Bender et al., 2008).

A recent review of meta-analyses in education and psychology (Polanin & Pigott, 2014) found that over 80% of studies did not implement a correction procedure to adjust α, while 93% of studies did not consider the implications of Type I error for their results. Similarly, Polanin & Pigott (2014) illustrated that correction techniques may reduce the instance of Type I errors occurring, but that the effects upon different statistical tests may vary.

In this analysis, we follow the recommendation of O'Dea et al., (2021) who advise that all analyses are justified and determined prior to determining results. Here we formulate four hypotheses, each relating to biological or ecological mechanisms. As noted in Section 4.3, we have collated categorical data for several variables (stressor combination, organisation level, realm) that were not tested further. Given the potential ramifications of α adjustment techniques, we do not implement such a procedure here.

Within ecology, many meta-analyses test all covariables in their dataset, under the premise that each of these could be ecologically relevant (Rothstein et al., 2013). However, this can lead to an increase in the instance of Type I error and the incorrect inference of an important explanatory variable (Rothstein et al., 2013). By taking the steps outlined above, we have attempted to limit the effects of Type I error upon our analysis.

References

Bender, R., Bunce, C., Clarke, M., Gates, S., Lange, S., Pace, N. L., & Thorlund, K. (2008). Attention should be given to multiplicity issues in systematic reviews. *Journal of Clinical Epidemiology*, 61(9), 857-865.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

O'Dea, R. E., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W., Parker, T. H., ... & Nakagawa, S. (2021). Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. *Biological Reviews*.

Polanin, J. R., & Pigott, T. D. (2015). The use of meta-analytic statistical significance testing. *Research Synthesis Methods*, 6(1), 63-73.

Rothstein, H. R., Lortie, C. J., Stewart, G. B., Koricheva, J., & Gurevitch, J. (2013). 20. Quality Standards for Research Syntheses. In *Handbook of meta-analysis in ecology and evolution* (pp. 323-338). Princeton University Press.

Multiplicative null model derivation

The multiplicative null model represents an approach to determining the expected effects due to multiple stressors. Outlined by Sih et al. (1998) and Folt et al. (1999), the multiplicative null model can be calculated using Equation S4.7.1, with A, B, and C referring to response metric values under Stressor A, Stressor B, and control conditions respectively.

$$S4.7.1) \qquad Expected\ Difference\ Under\ the\ Multiplicative\ Null\ Model = \frac{A \cdot B}{C} - C$$

While Equation S4.7.1 represents the difference from the control value, the absolute value expected under the multiplicative null model can be calculated by Equation S4.7.2.

$$S4.7.2) \qquad Expected\ Absolute\ Value\ Under\ the\ Multiplicative\ Null\ Model = \frac{A \cdot B}{C}$$

Example 1:

Condition value (C) = 100

Stressor A value (A) = 70

Stressor B value (B) = 80

Expected difference (to the control) under the multiplicative null model is -44

Expected absolute value under the multiplicative null model is 56.

Example 2:

Condition value (C) = 100

Stressor A value (A) = 40

Stressor B value (B) = 30

Expected difference (to the control) under the multiplicative null model is -88

Expected absolute value under the multiplicative null model is 12

In some situations, the multiplicative null model is described as Equations S4.7.3 or S4.7.4. A form of this equation is presented by Sih et al. (1998), where it is used the determine the proportion of a population which survive two stressors.

$$S4.7.3) \qquad Exp.\ Proport.\ Diff.\ Under\ the\ Multi.Null\ Model = P_A P_B - P_A - P_B$$

In the example used by Sih et al. (1998), P_A and P_B correspond to the proportion of prey killed by stressor A (i.e. $\frac{C-A}{C}$) and stressor B (i.e. $\frac{C-B}{C}$) respectively. As such, Equation S4.7.4 represents the difference from the control value, the proportional value expected under this form of the multiplicative null model.

$$S4.7.4) \qquad Exp.Proport.Value\ Under\ the\ Multi.Null\ Model = 1 + P_A P_B - P_A - P_B$$

Examples using these forms of the multiplicative null model are shown by Examples 3 and 4. Importantly, this form of the multiplicative null model provides the same results as the form of the multiplicative null model discussed above (i.e. compare the results of Examples 1 and 3, and Examples 2 and 4).

Example 3:

Condition value (C) = 100

Stressor A value (A) = 70

Stressor B value (B) = 80

Expected proportional difference (to the control) under the multiplicative null model is -0.44

Expected proportional value under the multiplicative null model is 0.56.

Example 4:

Condition value (C) = 100

Stressor A value (A) = 40

Stressor B value (B) = 30

Expected difference (to the control) under the multiplicative null model is -0.88

Expected absolute value under the multiplicative null model is 0.12

Finally, it is important to note that the form of the multiplicative null models provided by Equations S4.7.3 and S4.7.4 only applies where P_A and P_B correspond to the proportion of prey killed by stressor A (i.e. $\frac{C-A}{C}$) and stressor B (i.e. $\frac{C-B}{C}$) respectively. It does ***not*** apply where P_A and P_B correspond to the proportion of prey which survive stressor A (i.e. $\frac{A}{C}$) and stressor B (i.e. $\frac{B}{C}$) respectively. If Equations S4.7.3 and S4.7.4 are used where P_A and P_B correspond to the proportions survived by Stressors A and B, then incorrect results will be returned by the null models.

References

Folt, C. L., Chen, C. Y., Moore, M. V., & Burnaford, J. (1999). Synergism and antagonism among multiple stressors. *Limnology and Oceanography*, 44(3part2), 864-877.

Sih, A., Englund, G., & Wooster, D. (1998). Emergent impacts of multiple predators on prey. *Trends in Ecology & Evolution*, 13(9), 350-355.

Meta-analytical dataset reference list

In total 209 published studies provided data for the analyses in Chapter Four. A full reference list for these papers is provided below.

References

Ade, C. M., Boone, M. D., & Puglis, H. J. (2010). Effects of an insecticide and potential predators on green frogs and northern cricket frogs. *Journal of Herpetology*, 44(4), 591-600.

Alexander, A. C., Luis, A. T., Culp, J. M., Baird, D. J., & Cessna, A. J. (2013). Can nutrients mask community responses to insecticide mixtures?. *Ecotoxicology*, 22(7), 1085-1100.

Alexander, A. C., Culp, J. M., Baird, D. J., & Cessna, A. J. (2016). Nutrient–insecticide interactions decouple density-dependent predation pressure in aquatic insects. *Freshwater Biology*, 61(12), 2090-2101.

Alutoin, S., Boberg, J., Nyström, M., & Tedengren, M. (2001). Effects of the multiple stressors copper and reduced salinity on the metabolism of the hermatypic coral Porites lutea. *Marine Environmental Research*, 52(3), 289-299.

Anlauf, H., D'Croz, L., & O'Dea, A. (2011). A corrosive concoction: the combined effects of ocean warming and acidification on the early growth of a stony coral are multiplicative. *Journal of Experimental Marine Biology and Ecology*, 397(1), 13-20.

Aristi, I., Casellas, M., Elosegi, A., Insa, S., Petrovic, M., Sabater, S., & Acuña, V. (2016). Nutrients versus emerging contaminants–or a dynamic match between subsidy and stress effects on stream biofilms. *Environmental Pollution*, 212, 208-215.

Arnberg, M., Calosi, P., Spicer, J. I., Tandberg, A. H. S., Nilsen, M., Westerlund, S., & Bechmann, R. K. (2013). Elevated temperature elicits greater effects than decreased pH on the development, feeding and metabolism of northern shrimp (Pandalus borealis) larvae. *Marine Biology*, 160(8), 2037-2048.

Babarro, J. M., & de Zwaan, A. (2002). Influence of abiotic factors on bacterial proliferation and anoxic survival of the sea mussel Mytilus edulis L. *Journal of Experimental Marine Biology and Ecology*, 273(1), 33-49.

Baldwin, A. H., & Mendelssohn, I. A. (1998). Effects of salinity and water level on coastal marshes: an experimental test of disturbance as a catalyst for vegetation change. *Aquatic Botany*, 61(4), 255-268.

Banks, J. L. (1994). Raceway density and water flow as factors affecting spring chinook salmon (Oncorhynchus tshawytscha) during rearing and after release. *Aquaculture*, 119(2-3), 201-217.

Baxter, C. V., Fausch, K. D., Murakami, M., & Chapman, P. L. (2004). Fish invasion restructures stream and forest food webs by interrupting reciprocal prey subsidies. *Ecology*, 85(10), 2656-2663.

Beermann, A. J., Elbrecht, V., Karnatz, S., Ma, L., Matthaei, C. D., Piggott, J. J., & Leese, F. (2018). Multiple-stressor effects on stream macroinvertebrate communities: A mesocosm experiment manipulating salinity, fine sediment and flow velocity. *Science of the Total Environment*, 610, 961-971.

Boone, M. D., & Bridges-Britton, C. M. (2006). Examining multiple sublethal contaminants on the gray treefrog (Hyla versicolor): effects of an insecticide, herbicide, and fertilizer. *Environmental Toxicology and Chemistry: An International Journal*, 25(12), 3261-3265.

Boone, M. D., Semlitsch, R. D., Little, E. E., & Doyle, M. C. (2007). Multiple stressors in amphibian communities: effects of chemical contamination, bullfrogs, and fish. *Ecological Applications*, 17(1), 291-301.

Brittain, J. T., & Strecker, A. L. (2018). The interactive effect of multiple stressors on crustacean zooplankton communities in montane lakes. *Water Resources Research*, 54(2), 939-954.

Brown, J. R., Miiller, T., & Kerby, J. L. (2013). The interactive effect of an emerging infectious disease and an emerging contaminant on Woodhouse's toad (Anaxyrus woodhousii) tadpoles. *Environmental Toxicology and Chemistry*, 32(9), 2003-2008.

Bruno, J. F., Petes, L. E., Drew Harvell, C., & Hettinger, A. (2003). Nutrient enrichment can increase the severity of coral diseases. *Ecology Letters*, 6(12), 1056-1061.

Buck, J. C., Scheessele, E. A., Relyea, R. A., & Blaustein, A. R. (2012). The effects of multiple stressors on wetland communities: pesticides, pathogens and competing amphibians. *Freshwater Biology*, 57(1), 61-73.

Byrne, M., Soars, N. A., Ho, M. A., Wong, E., McElroy, D., Selvakumaraswamy, P., ... & Davis, A. R. (2010). Fertilization in a suite of coastal marine invertebrates from SE Australia is robust to near-future ocean warming and acidification. *Marine Biology*, 157(9), 2061-2069.

Byrne, M., Soars, N., Selvakumaraswamy, P., Dworjanyn, S. A., & Davis, A. R. (2010). Sea urchin fertilization in a warm, acidified and high pCO2 ocean across a range of sperm densities. *Marine Environmental Research*, 69(4), 234-239.

Byrne, M., Ho, M., Wong, E., Soars, N. A., Selvakumaraswamy, P., Shepard-Brennand, H., ... & Davis, A. R. (2011). Unshelled abalone and corrupted urchins: development of marine calcifiers in a changing ocean. *Proceedings of the Royal Society B: Biological Sciences*, 278(1716), 2376-2383.

Byrne, M., Foo, S., Soars, N. A., Wolfe, K. D., Nguyen, H. D., Hardy, N., & Dworjanyn, S. A. (2013). Ocean warming will mitigate the effects of acidification on calcifying sea urchin larvae (Heliocidaris tuberculata) from the Australian global warming hot spot. *Journal of Experimental Marine Biology and Ecology*, 448, 250-257.

Byrne, M., Gonzalez-Bernat, M., Doo, S., Foo, S., Soars, N., & Lamare, M. (2013). Effects of ocean warming and acidification on embryos and non-calcifying larvae of the invasive sea star Patiriella regularis. *Marine Ecology Progress Series*, 473, 235-246.

Byrne, M., Ho, M. A., Koleits, L., Price, C., King, C. K., Virtue, P., ... & Lamare, M. (2013). Vulnerability of the calcifying larval stage of the Antarctic sea urchin Sterechinus neumayeri to near-future ocean acidification and warming. *Global Change Biology*, 19(7), 2264-2275.

Cadmus, P., Guasch, H., Herdrich, A. T., Bonet, B., Urrea, G., & Clements, W. H. (2018). Structural and functional responses of periphyton and macroinvertebrate communities to ferric Fe, Cu, and Zn in stream mesocosms. *Environmental Toxicology and Chemistry*, 37(5), 1320-1329.

Camacho, A., & de Wit, R. (2003). Effect of nitrogen and phosphorus additions on a benthic microbial mat from a hypersaline lake. *Aquatic Microbial Ecology*, 32(3), 261-273.

Cancino, J., Gallardo, J., & Torres, F. (2003). Combined effects of dissolved oxygen concentration and water temperature on embryonic development and larval shell secretion in the marine snail Chorus giganteus (Gastropoda: Muricidae). *Marine Biology*, 142(1), 133-139.

Carvalho, L. M., Correia, P. M., Caçador, I., & Martins-Loução, M. A. (2003). Effects of salinity and flooding on the infectivity of salt marsh arbuscular mycorrhizal fungi in Aster tripolium L. *Biology and Fertility of Soils*, 38(3), 137-143.

Charpentier, A., Mesléard, F., & Grillas, P. (1998). The role of water level and salinity in the regulation of Juncus gerardi populations in former ricefields in southern France. *Journal of Vegetation Science*, 9(3), 361-370.

Cheever, B. M., & Simon, K. S. (2009). Seasonal influence of brook trout and mottled sculpin on lower trophic levels in an Appalachian stream. *Freshwater Biology*, 54(3), 524-535.

Chen, C. Y., Hathaway, K. M., & Folt, C. L. (2004). Multiple stress effects of Vision® herbicide, pH, and food on zooplankton and larval amphibian species from forest wetlands. *Environmental Toxicology and Chemistry: An International Journal*, 23(4), 823-831.

Christensen, M. R., Graham, M. D., Vinebrooke, R. D., Findlay, D. L., Paterson, M. J., & Turner, M. A. (2006). Multiple anthropogenic stressors cause ecological surprises in boreal lakes. *Global Change Biology*, 12(12), 2316-2322.

Chua, C. M., Leggat, W., Moya, A., & Baird, A. H. (2013). Temperature affects the early life history stages of corals more than near future ocean acidification. *Marine Ecology Progress Series*, 475, 85-92.

Clements, W. H., Cadmus, P., & Brinkman, S. F. (2013). Responses of aquatic insects to Cu and Zn in stream microcosms: understanding differences between single species tests and field responses. *Environmental Science & Technology*, 47(13), 7506-7513.

Cottingham, K. L., Glaholt, S., & Brown, A. C. (2004). Zooplankton community structure affects how phytoplankton respond to nutrient pulses. Ecology, 85(1), 158-171.Dahl, J., 1998. The impact of vertebrate and invertebrate predators on a stream benthic community. *Oecologia*, 117(1-2), pp.217-226.

Cumbo, V. R., Fan, T. Y., & Edmunds, P. J. (2013). Effects of exposure duration on the response of Pocillopora damicornis larvae to elevated temperature and high pCO2. *Journal of Experimental Marine Biology and Ecology*, 439, 100-107.

Dalton, R. L., Boutin, C., & Pick, F. R. (2015). Determining in situ periphyton community responses to nutrient and atrazine gradients via pigment analysis. *Science of the Total Environment*, 515, 70-82.

Day, R. H., Doyle, T. W., & Draugelis-Dale, R. O. (2006). Interactive effects of substrate, hydroperiod, and nutrients on seedling growth of Salix nigra and Taxodium distichum. *Environmental and Experimental Botany*, 55(1-2), 163-174.

De Stefano, L. G., Gattás, F., Vinocur, A., Cristos, D., Rojas, D., Cataldo, D., & Pizarro, H. (2018). Comparative impact of two glyphosate-based formulations in interaction with Limnoperna fortunei on freshwater phytoplankton. *Ecological Indicators*, 85, 575-584.

Devakie, M. N., & Ali, A. B. (2000). Salinity–temperature and nutritional effects on the setting rate of larvae of the tropical oyster, Crassostrea iredalei (Faustino). *Aquaculture*, 184(1-2), 105-114.

Domaizon, I., Lepère, C., Debroas, D., Bouvy, M., Ghiglione, J. F., Jacquet, S., ... & Bouvier, T. (2012). Short-term responses of unicellular planktonic eukaryotes to increases in temperature and UVB radiation. *BMC Microbiology*, 12(1), 1-13.

Dove, M. C., & O'Connor, W. A. (2007). Salinity and temperature tolerance of Sydney rock oysters Saccostrea glomerata during early ontogeny. *Journal of Shellfish Research*, 26(4), 939-947.

Duarte, S., Pascoal, C., Alves, A., Correia, A., & Cássio, F. (2008). Copper and zinc mixtures induce shifts in microbial communities and reduce leaf litter decomposition in streams. *Freshwater Biology*, 53(1), 91-101.

Egilsdottir, H., Spicer, J. I., & Rundle, S. D. (2009). The effect of CO2 acidified sea water and reduced salinity on aspects of the embryonic development of the amphipod Echinogammarus marinus (Leach). *Marine Pollution Bulletin*, 58(8), 1187-1191.

Elfwing, T., & Tedengren, M. (2002). Effects of copper and reduced salinity on grazing activity and macroalgae production: a short-term study on a mollusc grazer, Trochus maculatus, and two species of macroalgae in the inner Gulf of Thailand. *Marine Biology*, 140(5), 913-919.

Feng, Y., Warner, M. E., Zhang, Y., Sun, J., Fu, F. X., Rose, J. M., & Hutchins, D. A. (2008). Interactive effects of increased pCO2, temperature and irradiance on the marine coccolithophore Emiliania huxleyi (Prymnesiophyceae). *European Journal of Phycology*, 43(1), 87-98.

Feng, Y., Hare, C. E., Leblanc, K., Rose, J. M., Zhang, Y., DiTullio, G. R., ... & Hutchins, D. A. (2009). Effects of increased pCO2 and temperature on the North Atlantic spring bloom. I. The phytoplankton community and biogeochemical response. *Marine Ecology Progress Series*, 388, 13-25.

Fernandes, I., Duarte, S., Cássio, F., & Pascoal, C. (2009). Mixtures of zinc and phosphate affect leaf litter decomposition by aquatic fungi in streams. *Science of the Total Environment*, 407(14), 4283-4288.

Fernández-Aláez, M., Fernández-Aláez, C., Bécares, E., Valentín, M., Goma, J., & Castrillo, P. (2004). A 2-year experimental study on nutrient and predator influences on food web constituents in a shallow lake of north-west Spain. *Freshwater Biology*, 49(12), 1574-1592.

Ferragut, C., & de Campos Bicudo, D. (2012). Effect of N and P enrichment on periphytic algal community succession in a tropical oligotrophic reservoir. *Limnology*, 13(1), 131-141.

Ferrier-Pages, C., Richard, C., Forcioli, D., Allemand, D., Pichon, M., & Shick, J. M. (2007). Effects of temperature and UV radiation increases on the photosynthetic efficiency in four scleractinian coral species. *The Biological Bulletin*, 213(1), 76-87.

Findlay, H. S., Kendall, M. A., Spicer, J. I., & Widdicombe, S. (2010). Post-larval development of two intertidal barnacles at elevated CO 2 and temperature. *Marine Biology*, 157(4), 725-735.

Fockedey, N., Ghekiere, A., Bruwiere, S., Janssen, C. R., & Vincx, M. (2006). Effect of salinity and temperature on the intra-marsupial development of the brackish water mysid Neomysis integer (Crustacea: Mysidacea). *Marine Biology*, 148(6), 1339-1356.

Foo, S. A., Dworjanyn, S. A., Poore, A. G., & Byrne, M. (2012). Adaptive capacity of the habitat modifying sea urchin Centrostephanus rodgersii to ocean warming and ocean acidification: performance of early embryos. *PLoS One*, 7(8), e42497

Fu, F. X., Warner, M. E., Zhang, Y., Feng, Y., & Hutchins, D. A. (2007). Effects of Increased temperature and CO2 on photosynthesis, growth, and elemental ratios in marine Synechococcus and Prochlorococcus (cyanobacteria). *Journal of Phycology*, 43(3), 485-496.

Fu, F. X., Zhang, Y., Warner, M. E., Feng, Y., Sun, J., & Hutchins, D. A. (2008). A comparison of future increased CO2 and temperature effects on sympatric Heterosigma akashiwo and Prorocentrum minimum. *Harmful Algae*, 7(1), 76-90.

Funck, J. A., Clivot, H., Felten, V., Rousselle, P., Guérold, F., & Danger, M. (2013). Phosphorus availability modulates the toxic effect of silver on aquatic fungi and leaf litter decomposition. *Aquatic Toxicology*, 144, 199-207.

Gahl, M. K., Pauli, B. D., & Houlahan, J. E. (2011). Effects of chytrid fungus and a glyphosate-based herbicide on survival and growth of wood frogs (Lithobates sylvaticus). *Ecological Applications*, 21(7), 2521-2529.

Gama-Flores, J. L., Sarma, S. S. S., & Nandini, S. (2005). Interaction among copper toxicity, temperature and salinity on the population dynamics of Brachionus rotundiformis (Rotifera). *Hydrobiologia*, 546(1), 559-568.

Gardeström, J., Ermold, M., Goedkoop, W., & McKie, B. G. (2016). Disturbance history influences stressor impacts: effects of a fungicide and nutrients on microbial diversity and litter decomposition. *Freshwater Biology*, 61(12), 2171-2184.

Gattás, F., De Stefano, L. G., Vinocur, A., Bordet, F., Espinosa, M. S., Pizarro, H., & Cataldo, D. (2018). Impact of interaction between Limnoperna fortunei and Roundup Max® on freshwater phytoplankton: An in situ approach in Salto Grande reservoir (Argentina). *Chemosphere*, 209, 748-757.

Geyer, R. L., Smith, G. R., & Rettig, J. E. (2016). Effects of Roundup formulations, nutrient addition, and Western mosquitofish (Gambusia affinis) on aquatic communities. *Environmental Science and Pollution Research*, 23(12), 11729-11739.

Gomez Isaza, D. F., Cramp, R. L., & Franklin, C. E. (2018). Negative impacts of elevated nitrate on physiological performance are not exacerbated by low pH. *Aquatic Toxicology*, 200, 217-225.

Gonzalez, M. J., Knoll, L. B., & Vanni, M. J. (2010). Differential effects of elevated nutrient and sediment inputs on survival, growth and biomass of a common larval fish species (Dorosoma cepedianum). *Freshwater Biology*, 55(3), 654-669.

Gorgula, S. K., & Connell, S. D. (2004). Expansive covers of turf-forming algae on human-dominated coast: the relative effects of increasing nutrient and sediment loads. *Marine Biology*, 145(3), 613-619.

Graham, M. D., & Vinebrooke, R. D. (2009). Extreme weather events alter planktonic communities in boreal lakes. *Limnology and Oceanography*, 54(6part2), 2481-2492.

Gustina, G. W., & Hoffmann, J. P. (2000). Periphyton dynamics in a subalpine mountain stream during winter. *Arctic, Antarctic, and Alpine Research*, 32(2), 127-134.

Hagerthey, S. E., Defew, E. C., & Paterson, D. M. (2002). Influence of Corophium volutator and Hydrobia ulvae on intertidal benthic diatom assemblages under different nutrient and temperature regimes. *Marine Ecology Progress Series*, 245, 47-59.

Haglund, A. L., & Hillebrand, H. (2005). The effect of grazing and nutrient supply on periphyton associated bacteria. *FEMS Microbiology Ecology*, 52(1), 31-41.

Häkkinen, J., Vehniäinen, E., & Oikari, A. (2004). High sensitivity of northern pike larvae to UV-B but no UV-photoinduced toxicity of retene. *Aquatic Toxicology*, 66(4), 393-404.

Hassell, K. L., Coutin, P. C., & Nugegoda, D. (2008). Hypoxia, low salinity and lowered temperature reduce embryo survival and hatch rates in black bream Acanthopagrus butcheri (Munro, 1949). *Journal of Fish Biology*, 72(7), 1623-1636.

Hatch, A. C., & Blaustein, A. R. (2000). Combined effects of UV-B, nitrate, and low pH reduce the survival and activity level of larval cascades frogs (Rana cascadae). *Archives of Environmental Contamination and Toxicology*, 39(4), 494-499.

Hellings, S. E., & Gallagher, J. L. (1992). The effects of salinity and flooding on Phragmites australis. *Journal of Applied Ecology*, 41-49.

Heugens, E. H., Tokkie, L. T., Kraak, M. H., Hendriks, A. J., Van Straalen, N. M., & Admiraal, W. (2006). Population growth of Daphnia magna under multiple stress conditions: joint effects of temperature, food, and cadmium. *Environmental Toxicology and Chemistry: An International Journal*, 25(5), 1399-1407.

Ho, B. S., & Dudgeon, D. (2016). Are high densities of fishes and shrimp associated with top-down control of tropical benthic communities? A test in three Hong Kong streams. *Freshwater Biology*, 61(1), 57-68.

Hoffman, J. R., Hansen, L. J., & Klinger, T. (2003). Interactions between UV radiation and temperature limit inferences from single-factor experiments. *Journal of Phycology*, 39(2), 268-272.

Holzapfel, A. M., & Vinebrooke, R. D. (2005). Environmental warming increases invasion potential of alpine lake communities by imported species. *Global Change Biology*, 11(11), 2009-2015.

Howard, R. J., & Mendelssohn, I. A. (2000). Structure and composition of oligohaline marsh plant communities exposed to salinity pulses. *Aquatic Botany*, 68(2), 143-164.

Howard, R. J., & Rafferty, P. S. (2006). Clonal variation in response to salinity and flooding stress in four marsh macrophytes of the northern gulf of Mexico, USA. *Environmental and Experimental Botany*, 56(3), 301-313.

Jackson, M. C., Jones, T., Milligan, M., Sheath, D., Taylor, J., Ellis, A., ... & Grey, J. (2014). Niche differentiation among invasive crayfish and their impacts on ecosystem structure and functioning. *Freshwater Biology*, 59(6), 1123-1135.

Jackson, M. C., Grey, J., Miller, K., Britton, J. R., & Donohue, I. (2016). Dietary niche constriction when invaders meet natives: evidence from freshwater decapods. *Journal of Animal Ecology*, 85(4), 1098-1107.

Kahn, A. E., & Durako, M. J. (2006). Thalassia testudinum seedling responses to changes in salinity and nitrogen levels. *Journal of Experimental Marine Biology and Ecology*, 335(1), 1-12.

Kamer, K., & Fong, P. (2001). Nitrogen enrichment ameliorates the negative effects of reduced salinity on the green macroalga Enteromorpha intestinalis. *Marine Ecology Progress Series*, 218, 87-93.

Karsten, U., Dummermuth, A., Hoyer, K., & Wiencke, C. (2003). Interactive effects of ultraviolet radiation and salinity on the ecophysiology of two Arctic red algae from shallow waters. *Polar Biology*, 26(4), 249-258.

Kashian, D. R., Zuellig, R. E., Mitchell, K. A., & Clements, W. H. (2007). The cost of tolerance: Sensitivity of stream benthic communities to UV-B and metals. *Ecological Applications*, 17(2), 365-375.

Kiffney, P. M., & Richardson, J. S. (2001). Interactions among nutrients, periphyton, and invertebrate and vertebrate (Ascaphus truei) grazers in experimental channels. *Copeia*, 2001(2), 422-429.

Kimberly, D. A., & Salice, C. J. (2014). Complex interactions between climate change and toxicants: evidence that temperature variability increases sensitivity to cadmium. *Ecotoxicology*, 23(5), 809-817.

Knillmann, S., Stampfli, N. C., Noskov, Y. A., Beketov, M. A., & Liess, M. (2013). Elevated temperature prolongs long-term effects of a pesticide on Daphnia spp. due to altered competition in zooplankton communities. *Global Change Biology*, 19(5), 1598-1609.

Koch, M. S., & Erskine, J. M. (2001). Sulfide as a phytotoxin to the tropical seagrass Thalassia testudinum: interactions with light, salinity and temperature. *Journal of Experimental Marine Biology and Ecology*, 266(1), 81-95.

Kratina, P., Greig, H. S., Thompson, P. L., Carvalho-Pereira, T. S., & Shurin, J. B. (2012). Warming modifies trophic cascades and eutrophication in experimental freshwater communities. *Ecology*, 93(6), 1421-1430.

Krishnamurthy, S. V., & Smith, G. R. (2010). Growth, abnormalities, and mortality of tadpoles of American toad exposed to combinations of malathion and nitrate. *Environmental Toxicology and Chemistry*, 29(12), 2777-2782.

Kuhns, L. A., & Berg, M. B. (1999). Benthic invertebrate community responses to round goby (Neogobius melanostomus) and zebra mussel (Dreissena polymorpha) invasion in southern Lake Michigan. *Journal of Great Lakes Research*, 25(4), 910-917.

Kumlu, M. E. T. İ. N., Eroldogan, O. T., & Aktas, M. (2000). Effects of temperature and salinity on larval growth, survival and development of Penaeus semisulcatus. *Aquaculture*, 188(1-2), 167-173.

Laughlin, R. B., & Neff, J. M. (1979). Interactive effects of salinity, temperature and polycyclic aromatic hydrocarbons on the survival and development rate of larvae of the mud crab Rhithropanopeus harrisii. *Marine Biology*, 53(3), 281-291.

Lenihan, H. S., Peterson, C. H., Kim, S. L., Conlan, K. E., Fairey, R., McDonald, C., ... & Oliver, J. S. (2003). Variation in marine benthic community composition allows discrimination of multiple stressors. *Marine Ecology Progress Series*, 261, 63-73.

Lenssen, G. M., Van Duin, W. E., Jak, P., & Rozema, J. (1995). The response of Aster tripolium and Puccinellia maritima to atmospheric carbon dioxide enrichment and their interactions with flooding and salinity. *Aquatic Botany*, 50(2), 181-192.

Li, R., & Brawley, S. H. (2004). Improved survival under heat stress in intertidal embryos (Fucus spp.) simultaneously exposed to hypersalinity and the effect of parental thermal history. *Marine Biology*, 144(2), 205-213.

Liess, A., Lange, K., Schulz, F., Piggott, J. J., Matthaei, C. D., & Townsend, C. R. (2009). Light, nutrients and grazing interact to determine diatom species richness via changes to productivity, nutrient state and grazer activity. *Journal of Ecology*, 97(2), 326-336.

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

Long, L. E., Saylor, L. S., & Soule, M. E. (1995). A pH/UV-B synergism in amphibians. *Conservation Biology*, 9(5), 1301-1303.

Longhi, M. L., Ferreyra, G., Schloss, I., & Roy, S. (2006). Variable phytoplankton response to enhanced UV-B and nitrate addition in mesocosm experiments at three latitudes (Canada, Brazil and Argentina). *Marine Ecology Progress Series*, 313, 57-72.

Louhi, P., Richardson, J. S., & Muotka, T. (2017). Sediment addition reduces the importance of predation on ecosystem functions in experimental stream channels. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(1), 32-40.

Lozano, V. L., Vinocur, A., y García, C. S., Allende, L., Cristos, D. S., Rojas, D., ... & Pizarro, H. (2018). Effects of glyphosate and 2, 4-D mixture on freshwater phytoplankton and periphyton communities: a microcosms approach. *Ecotoxicology and Environmental Safety*, 148, 1010-1019.

Mackey, K. R., Hunter, D., Fischer, E. V., Jiang, Y., Allen, B., Chen, Y., ... & Paytan, A. (2013). Aerosol-nutrient-induced picoplankton growth in Lake Tahoe. *Journal of Geophysical Research: Biogeosciences*, 118(3), 1054-1067.

MacLennan, M. M., Dings-Avery, C., & Vinebrooke, R. D. (2015). Invasive trout increase the climatic sensitivity of zooplankton communities in naturally fishless lakes. *Freshwater Biology*, 60(8), 1502-1513.

Magbanua, F. S., Townsend, C. R., Hageman, K. J., Lange, K., Lear, G., Lewis, G. D., & Matthaei, C. D. (2013). Understanding the combined influence of fine sediment and glyphosate herbicide on stream periphyton communities. *Water Research*, 47(14), 5110-5120.

Matthaei, C. D., Piggott, J. J., & Townsend, C. R. (2010). Multiple stressors in agricultural streams: interactions among sediment addition, nutrient enrichment and water abstraction. *Journal of Applied Ecology*, 47(3), 639-649.

McKee, D., Atkinson, D., Collings, S. E., Eaton, J. W., Gill, A. B., Harvey, I., ... & Moss, B. (2003). Response of freshwater microcosm communities to nutrients, fish, and elevated temperature during winter and summer. *Limnology and Oceanography*, 48(2), 707-722.

Mebane, C. A., Schmidt, T. S., & Balistrieri, L. S. (2017). Larval aquatic insect responses to cadmium and zinc in experimental streams. *Environmental Toxicology and Chemistry*, 36(3), 749-762.

Molinos, J. G., & Donohue, I. (2010). Interactions among temporal patterns determine the effects of multiple stressors. *Ecological Applications*, 20(7), 1794-1800.

Moss, B., McKee, D., Atkinson, D., Collings, S. E., Eaton, J. W., Gill, A. B., ... & Wilson, D. (2003). How important is climate? Effects of warming, nutrient addition and fish on phytoplankton in shallow lake microcosms. *Journal of Applied Ecology*, 40(5), 782-792.

Muñoz, I., Real, M., Guasch, H., Navarro, E., & Sabater, S. (2001). Effects of atrazine on periphyton under grazing pressure. *Aquatic Toxicology*, 55(3-4), 239-249.

Murdock, J. N., Shields, F. D., & Lizotte, R. E. (2013). Periphyton responses to nutrient and atrazine mixtures introduced through agricultural runoff. *Ecotoxicology*, 22(2), 215-230.

Mustonen, K. R., Mykrä, H., Louhi, P., Markkola, A., Tolkkinen, M., Huusko, A., ... & Muotka, T. (2016). Sediments and flow have mainly independent effects on multitrophic stream communities and ecosystem functions. *Ecological Applications*, 26(7), 2116-2129.

Muyssen, B. T., Messiaen, M., & Janssen, C. R. (2010). Combined cadmium and temperature acclimation in Daphnia magna: physiological and sub-cellular effects. *Ecotoxicology and Environmental Safety*, 73(5), 735-742.

Naidoo, G., & Kift, J. (2006). Responses of the saltmarsh rush Juncus kraussii to salinity and waterlogging. *Aquatic Botany*, 84(3), 217-225.

Nasrolahi, A., Pansch, C., Lenz, M., & Wahl, M. (2012). Being young in a changing world: how temperature and salinity changes interactively modify the performance of larval stages of the barnacle Amphibalanus improvisus. *Marine Biology*, 159(2), 331-340.

Nguyen, H. D., Doo, S. S., Soars, N. A., & Byrne, M. (2012). Noncalcifying larvae in a changing ocean: warming, not acidification/hypercapnia, is the dominant stressor on development of the sea star M eridiastra calcar. *Global Change Biology*, 18(8), 2466-2476.

Nielsen, M. V., & Tønseth, C. P. (1991). Temperature and salinity effect on growth and chemical composition of Gyrodinium aureolum Hulburt in culture. *Journal of Plankton Research*, 13(2), 389-398.

O'Connor, W. A., & Lawler, N. F. (2004). Salinity and temperature tolerance of embryos and juveniles of the pearl oyster, Pinctada imbricata Röding. *Aquaculture*, 229(1-4), 493-506.

Ortiz-Santaliestra, M. E., & Sparling, D. W. (2007). Alteration of larval development and metamorphosis by nitrate and perchlorate in southern leopard frogs (Rana sphenocephala). *Archives of Environmental Contamination and Toxicology*, 53(4), 639-646.

Ortiz-Santaliestra, M. E., Fernández-Benéitez, M. J., Lizana, M., & Marco, A. (2010). Adaptation to osmotic stress provides protection against ammonium nitrate in Pelophylax perezi embryos. *Environmental Pollution*, 158(3), 934-940.

Ortiz-Santaliestra, M. E., Fernández-Benéitez, M. J., Lizana, M., & Marco, A. (2011). Influence of a combination of agricultural chemicals on embryos of the endangered gold-striped salamander (Chioglossa lusitanica). *Archives of Environmental Contamination and Toxicology*, 60(4), 672-680.

Ortiz-Santaliestra, M. E., Fernández-Benéitez, M. J., Lizana, M., & Marco, A. (2011). Responses of toad tadpoles to ammonium nitrate fertilizer and predatory stress: differences between populations on a local scale. *Environmental Toxicology and Chemistry*, 30(6), 1440-1446.

Ortiz-Santaliestra, M. E., & Marco, A. (2015). Influence of dissolved oxygen conditions on toxicity of ammonium nitrate to larval natterjack toads. *Archives of Environmental Contamination and Toxicology*, 69(1), 95-103.

Pagano, M., Koffi, M. A., Cecchi, P., Corbin, D., Champalbert, G., & Saint-Jean, L. (2003). An experimental study of the effects of nutrient supply and Chaoborus predation on zooplankton communities of a shallow tropical reservoir (Lake Brobo, Côte d'Ivoire). *Freshwater Biology*, 48(8), 1379-1395.

Pahkala, M., Räsänen, K., Laurila, A., Johanson, U., Björn, L. O., & Merilä, J. (2002). Lethal and sublethal effects of UV-B/pH synergism on common frog embryos. *Conservation Biology*, 16(4), 1063-1073.

Pandey, R. B., Adams, G. L., & Warren, L. W. (2011). Survival and precopulatory guarding behavior of Hyalella azteca (Amphipoda) exposed to nitrate in the presence of atrazine. *Environmental Toxicology and Chemistry*, 30(5), 1170-1177.

Pannard, A., Bormans, M., Lefebvre, S., Claquin, P., & Lagadeuc, Y. (2007). Phytoplankton size distribution and community structure: influence of nutrient input and sedimentary loss. *Journal of Plankton Research*, 29(7), 583-598.

Pansch, C., Nasrolahi, A., Appelhans, Y. S., & Wahl, M. (2012). Impacts of ocean warming and acidification on the larval development of the barnacle Amphibalanus improvisus. *Journal of Experimental Marine Biology and Ecology*, 420, 48-55.

Paula, J., Mendes, R. N., Paci, S., McLaughlin, P., Gherardi, F., & Emmerson, W. (2001). Combined effects of temperature and salinity on the larval development of the estuarine mud prawn Upogebia africana (Crustacea, Thalassinidea). *Hydrobiologia*, 449, 141-148.

Pechenik, J. A., Marsden, I. D., & Pechenik, O. (2003). Effects of temperature, salinity, and air exposure on development of the estuarine pulmonate gastropod Amphibola crenata. *Journal of Experimental Marine Biology and Ecology*, 292(2), 159-176.

Pelletier, M. C., Burgess, R. M., Ho, K. T., Kuhn, A., McKinney, R. A., & Ryba, S. A. (1997). Phototoxicity of individual polycyclic aromatic hydrocarbons and petroleum to marine invertebrate larvae and juveniles. *Environmental Toxicology and Chemistry: An International Journal*, 16(10), 2190-2199.

Perez-Fuentetaja, A., McQueen, D. J., & Demers, E. (1996). Stability of oligotrophic and eutrophic planktonic communities after disturbance by fish. *Oikos*, 98-110.

Pérez, M. T., & Sommaruga, R. (2007). Interactive effects of solar radiation and dissolved organic matter on bacterial activity and community structure. *Environmental Microbiology*, 9(9), 2200-2210.

Piggott, J. J., Lange, K., Townsend, C. R., & Matthaei, C. D. (2012). Multiple stressors in agricultural streams: a mesocosm study of interactions among raised water temperature, sediment addition and nutrient enrichment. *PLoS One*, 7(11), e49873.

Piggott, J. J., Townsend, C. R., & Matthaei, C. D. (2015). Climate warming and agricultural stressors interact to determine stream macroinvertebrate community dynamics. *Global Change Biology*, 21(5), 1887-1906.

Pilati, A., Vanni, M. J., González, M. J., & Gaulke, A. K. (2009). Effects of agricultural subsidies of nutrients and detritus on fish and plankton of shallow-reservoir ecosystems. *Ecological Applications*, 19(4), 942-960.

Pizarro, H., Di Fiori, E., Sinistro, R., Ramírez, M., Rodríguez, P., Vinocur, A., & Cataldo, D. (2016). Impact of multiple anthropogenic stressors on freshwater: how do glyphosate and the invasive mussel Limnoperna fortunei affect microbial communities and water quality?. *Ecotoxicology*, 25(1), 56-68.

Piazzi, L., Balata, D., Ceccherelli, G., & Cinelli, F. (2005). Interactive effect of sedimentation and Caulerpa racemosa var. cylindracea invasion on macroalgal assemblages in the Mediterranean Sea. *Estuarine, Coastal and Shelf Science*, 64(2-3), 467-474.

Portinho, J. L., Nielsen, D. L., Daré, L., Henry, R., Oliveira, R. C., & Branco, C. C. (2018). Mixture of commercial herbicides based on 2, 4-D and glyphosate mixture can suppress the emergence of zooplankton from sediments. *Chemosphere*, 203, 151-159.

Preston, D. L., Henderson, J. S., & Johnson, P. T. (2012). Community ecology of invasions: direct and indirect effects of multiple invasive species on aquatic communities. *Ecology*, 93(6), 1254-1261.

Pringle, C. M., & Hamazaki, T. (1998). The role of omnivory in a neotropical stream: separating diurnal and nocturnal effects. *Ecology*, 79(1), 269-280.

Proia, L., Vilches, C., Boninneau, C., Kantiani, L., Farré, M., Romaní, A. M., ... & Guasch, H. (2013). Drought episode modulates the response of river biofilms to triclosan. *Aquatic Toxicology*, 127, 36-45.

Przeslawski, R. (2005). Combined effects of solar radiation and desiccation on the mortality and development of encapsulated embryos of rocky shore gastropods. *Marine Ecology Progress Series*, 298, 169-177.

Przeslawski, R., Davis, A. R., & Benkendorff, K. (2005). Synergistic effects associated with climate change and the development of rocky shore molluscs. *Global Change Biology*, 11(3), 515-522.

Puglis, H. J., & Boone, M. D. (2011). Effects of technical-grade active ingredient vs. commercial formulation of seven pesticides in the presence or absence of UV radiation on survival of green frog tadpoles. *Archives of Environmental Contamination and Toxicology*, 60(1), 145-155.

Putnam, H. M., Mayfield, A. B., Fan, T. Y., Chen, C. S., & Gates, R. D. (2013). The physiological and molecular responses of larvae from the reef-building coral Pocillopora damicornis exposed to near-future increases in temperature and p CO 2. *Marine Biology*, 160(8), 2157-2173.

Qiu, J. W., & Qian, P. Y. (1997). Combined effects of salinity, temperature and food on early development of the polychaete Hydroides elegans. *Marine Ecology Progress Series*, 152, 79-88.

Qiu, J. W., & Qian, P. Y. (1998). Combined effects of salinity and temperature on juvenile survival, growth and maturation in the polychaete Hydroides elegans. *Marine Ecology Progress Series*, 168, 127-134.

Rejmánková, E., & Komárková, J. (2005). Response of cyanobacterial mats to nutrient and salinity changes. *Aquatic Botany*, 83(2), 87-107.

Relyea, R. A. (2006). The effects of pesticides, pH, and predatory stress on amphibians under mesocosm conditions. *Ecotoxicology*, 15(6), 503-511.

Rogell, B., Hofman, M., Eklund, M., Laurila, A., & Höglund, J. (2009). The interaction of multiple environmental stressors affects adaptation to a novel habitat in the natterjack toad Bufo calamita. *Journal of Evolutionary Biology*, 22(11), 2267-2277.

Rohr, J. R., Elskus, A. A., Shepherd, B. S., Crowley, P. H., McCarthy, T. M., Niedzwiecki, J. H., ... & Palmer, B. D. (2004). Multiple stressors and salamanders: effects of an herbicide, food limitation, and hydroperiod. *Ecological Applications*, 1028-1040.

Rohr, J. R., & Crumrine, P. W. (2005). Effects of an herbicide and an insecticide on pond community structure and processes. *Ecological Applications*, 15(4), 1135-1147.

Roller, R. A., & Stickle, W. B. (1993). Effects of temperature and salinity acclimation of adults on larval survival, physiology, and early development of Lytechinus variegatus (Echinodermata: Echinoidea). *Marine Biology*, 116(4), 583-591.

Romansic, J. M., Diez, K. A., Higashi, E. M., & Blaustein, A. R. (2006). Effects of nitrate and the pathogenic water mold Saprolegnia on survival of amphibian larvae. *Diseases of Aquatic Organisms*, 68(3), 235-243.

Rosenberg, R., & Costlow, J. D. (1976). Synergistic effects of cadmium and salinity combined with constant and cycling temperatures on the larval development of two estuarine crab species. *Marine Biology*, 38(4), 291-303.

Rotter, S., Heilmeier, H., Altenburger, R., & Schmitt-Jansen, M. (2013). Multiple stressors in periphyton–comparison of observed and predicted tolerance responses to high ionic loads and herbicide exposure. *Journal of Applied Ecology*, 50(6), 1459-1468.

Russell, J., & Phillips, N. (2009). Species-specific vulnerability of benthic marine embryos of congeneric snails (Haminoea spp.) to ultraviolet radiation and other intertidal stressors. *The Biological Bulletin*, 217(1), 65-72.

Russell, J., & Phillips, N. E. (2009). Synergistic effects of ultraviolet radiation and conditions at low tide on egg masses of limpets (Benhamina obliquata and Siphonaria australis) in New Zealand. *Marine Biology*, 156(4), 579-587.

Sandifer, P. A. (1973). Effects of temperature and salinity on larval development of grass shrimp, Palaemonetes vulgaris (Decapoda, Caridea). *Fishery Bulletin*, 71(1), 115.

Sanches, L. F., Guariento, R. D., Caliman, A., Bozelli, R. L., & Esteves, F. A. (2011). Effects of nutrients and light on periphytic biomass and nutrient stoichiometry in a tropical black-water aquatic ecosystem. *Hydrobiologia*, 669(1), 35-44.

Sargian, P., Pelletier, É., Mostajir, B., Ferreyra, G. A., & Demers, S. (2005). TBT toxicity on a natural planktonic assemblage exposed to enhanced ultraviolet-B radiation. *Aquatic Toxicology*, 73(3), 299-314.

Sargian, P., Mas, S., Pelletier, E., & Demers, S. (2007). Multiple stressors on an Antarctic microplankton assemblage: water soluble crude oil and enhanced UVBR level at Ushuaia (Argentina). *Polar Biology*, 30(7), 829-841.

Schlöder, C., & D'Croz, L. (2004). Responses of massive and branching coral species to the combined effects of water temperature and nitrate enrichment. *Journal of Experimental Marine Biology and Ecology*, 313(2), 255-268.

Scott, S. E., Pray, C. L., Nowlin, W. H., & Zhang, Y. (2012). Effects of native and invasive species on stream ecosystem functioning. *Aquatic Sciences*, 74(4), 793-808.

Seeland, A., Albrand, J., Oehlmann, J., & Müller, R. (2013). Life stage-specific effects of the fungicide pyrimethanil and temperature on the snail Physella acuta (Draparnaud, 1805) disclose the pitfalls for the aquatic risk assessment under global climate change. *Environmental Pollution*, 174, 1-9.

Sheppard Brennand, H., Soars, N., Dworjanyn, S. A., Davis, A. R., & Byrne, M. (2010). Impact of ocean warming and ocean acidification on larval development and calcification in the sea urchin Tripneustes gratilla. *PLoS One*, 5(6), e11372.

Shurin, J. B., Clasen, J. L., Greig, H. S., Kratina, P., & Thompson, P. L. (2012). Warming shifts top-down and bottom-up control of pond food web structure and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1605), 3008-3017.

Smith, K. F., & Lester, P. J. (2006). Cyanobacterial blooms appear to be driven by top-down rather than bottom-up effects in the Lower Karori Reservoir (Wellington, New Zealand). New Zealand *Journal of Marine and Freshwater Research*, 40(1), 53-63.

Smith, G. R., Krishnamurthy, S. V., Burger, A. C., & Mills, L. B. (2011). Differential effects of malathion and nitrate exposure on American toad and wood frog tadpoles. *Archives of Environmental Contamination and Toxicology*, 60(2), 327-335.

Smith, G. R., & Dibble, C. J. (2012). Effects of an invasive fish (Gambusia affinis) and anthropogenic nutrient enrichment on American toad (Anaxyrus americanus) tadpoles. *Journal of Herpetology*, 46(2), 198-202.

Smith, G. R., Dibble, C. J., Terlecky, A. J., Dayer, C. B., Burner, A. B., & Ogle, M. E. (2013). Effects of invasive western mosquitofish and ammonium nitrate on green frog tadpoles. *Copeia*, 2013(2), 248-253.

Sousa, A. I., Martins, I., Lillebø, A. I., Flindt, M. R., & Pardal, M. A. (2007). Influence of salinity, nutrients and light on the germination and growth of Enteromorpha sp. spores. *Journal of Experimental Marine Biology and Ecology*, 341(1), 142-150.

Southerland, H. A., & Lewitus, A. J. (2004). Physiological responses of estuarine phytoplankton to ultraviolet light-induced fluoranthene toxicity. *Journal of Experimental Marine Biology and Ecology*, 298(2), 303-322.

Steevens, J. A., Slattery, M., Schlenk, D., Aryl, A., & Benson, W. H. (1999). Effects of ultraviolet-B light and polyaromatic hydrocarbon exposure on sea urchin development and bacterial bioluminescence. *Marine Environmental Research*, 48(4-5), 439-457.

Sullivan, K. B., & Spence, K. M. (2003). Effects of sublethal concentrations of atrazine and nitrate on metamorphosis of the African clawed frog. *Environmental Toxicology and Chemistry: An International Journal*, 22(3), 627-635.

Sundbäck, K., Petersen, D. G., Dahllöf, I., & Larson, F. (2007). Combined nutrient–toxicant effects on a shallow-water marine sediment system: sensitivity and resilience of ecosystem functions. *Marine Ecology Progress Series*, 330, 13-30.

Suomalainen, L. R., Tiirola, M. A., & Valtonen, E. T. (2005). Influence of rearing conditions on Flavobacterium columnare infection of rainbow trout, Oncorhynchus mykiss (Walbaum). *Journal of Fish Diseases*, 28(5), 271-277.

Talmage, S. C., & Gobler, C. J. (2011). Effects of elevated temperature and carbon dioxide on the growth and survival of larvae and juveniles of three species of Northwest Atlantic bivalves. *PLoS One*, 6(10), e26941.

Tasmin, R., Shimasaki, Y., Tsuyama, M., Qiu, X., Khalil, F., Okino, N., ... & Oshima, Y. (2014). Elevated water temperature reduces the acute toxicity of the widely used herbicide diuron to a green alga, Pseudokirchneriella subcapitata. *Environmental Science and Pollution Research*, 21(2), 1064-1070.

Thiyagarajan, V., Harder, T., & Qian, P. Y. (2003). Combined effects of temperature and salinity on larval development and attachment of the subtidal barnacle Balanus trigonus Darwin. *Journal of Experimental Marine Biology and Ecology*, 287(2), 223-236.

Thiyagarajan, V., & Ko, G. W. K. (2012). Larval growth response of the Portuguese oyster (Crassostrea angulata) to multiple climate change stressors. *Aquaculture*, 370, 90-95.

Thompson, P. L., St-Jacques, M. C., & Vinebrooke, R. D. (2008). Impacts of climate warming and nitrogen deposition on alpine plankton in lake and pond habitats: an in vitro experiment. *Arctic, Antarctic, and Alpine Research*, 40(1), 192-198.

Thompson, P. L., & Shurin, J. B. (2012). Regional zooplankton biodiversity provides limited buffering of pond ecosystems against climate change. *Journal of Animal Ecology*, 81(1), 251-259.

Torquemada, Y. F., Durako, M. J., & Lizaso, J. L. S. (2005). Effects of salinity and possible interactions with temperature and pH on growth and photosynthesis of Halophila johnsonii Eiseman. *Marine Biology*, 148(2), 251-260.

Valenti Jr, T. W., Perez-Hurtado, P., Chambliss, C. K., & Brooks, B. W. (2009). Aquatic toxicity of sertraline to Pimephales promelas at environmentally relevant surface water pH. *Environmental Toxicology and Chemistry*, 28(12), 2685-2694.

Valenti Jr, T. W., James, S. V., Lahousse, M. J., Schug, K. A., Roelke, D. L., Grover, J. P., & Brooks, B. W. (2010). A mechanistic explanation for pH-dependent ambient aquatic toxicity of Prymnesium parvum Carter. *Toxicon*, 55(5), 990-998.

Van de Staaij, J. W. M., Lenssen, G. M., Stroetenga, M., & Rozema, J. (1993). The combined effects of elevated CO 2 levels and UV-B radiation on growth characteristics of Elymus athericus (= E. pycnanathus). *Vegetatio*, 104(1), 433-439.

Van Katwijk, M. M., Schmitz, G. H. W., Gasseling, A. P., & Van Avesaath, P. H. (1999). Effects of salinity and nutrient load and their interaction on Zostera marina. *Marine Ecology Progress Series*, 190, 155-165.

Vázquez, M. G., Ituarte, R. B., Bas, C. C., & Spivak, E. D. (2013). Effects of temperature and salinity on the ovarian cycle and the embryonic development of the invasive shrimp Palaemon macrodactylus. *Journal of Crustacean Biology*, 33(2), 218-223.

Vernberg, W. B., DeCoursey, P. J., & Padgett, W. J. (1973). Synergistic effects of environmental variables on larvae of Uca pugilator. *Marine Biology*, 22(4), 307-312.

Verween, A., Vincx, M., & Degraer, S. (2007). The effect of temperature and salinity on the survival of Mytilopsis leucophaeata larvae (Mollusca, Bivalvia): The search for environmental limits. *Journal of Experimental Marine Biology and Ecology*, 348(1-2), 111-120.

Villeneuve, A., Montuelle, B., & Bouchez, A. (2010). Influence of slight differences in environmental conditions (light, hydrodynamics) on the structure and function of periphyton. *Aquatic Sciences*, 72(1), 33-44.

Wacksman, M. N., Maul, J. D., & Lydy, M. J. (2006). Impact of atrazine on chlorpyrifos toxicity in four aquatic vertebrates. *Archives of Environmental Contamination and Toxicology*, 51(4), 681-689.

Wagenhoff, A., Townsend, C. R., & Matthaei, C. D. (2012). Macroinvertebrate responses along broad stressor gradients of deposited fine sediment and dissolved nutrients: a stream mesocosm experiment. *Journal of Applied Ecology*, 49(4), 892-902.

Wagenhoff, A., Lange, K., Townsend, C. R., & Matthaei, C. D. (2013). Patterns of benthic algae and cyanobacteria along twin-stressor gradients of nutrients and fine sediment: a stream mesocosm experiment. *Freshwater Biology*, 58(9), 1849-1863.

Walsh, W. A., Swanson, C., & Lee, C. S. (1991). Combined effects of temperature and salinity on embryonic development and hatching of striped mullet, Mugil cephalus. *Aquaculture*, 97(2-3), 281-289.

Wang, H., Zhu, X., Wang, Y., Luo, M., & Liu, Z. (2012). Determination of optimum temperature and salinity for fertilization and hatching in the Chinese pearl oyster Pinctada martensii (Dunker). *Aquaculture*, 358, 292-297.

Webb, E. C., & Mendelssohn, I. A. (1996). Factors affecting vegetation dieback of an oligohaline marsh in coastal Louisiana: field manipulation of salinity and submergence. *American Journal of Botany*, 83(11), 1429-1434.

Weisse, T., Laufenstein, N., & Weithoff, G. (2013). Multiple environmental stressors confine the ecological niche of the rotifer Cephalodella acidophila. *Freshwater Biology*, 58(5), 1008-1015.

Wendt-Rasch, L., Pirzadeh, P., & Woin, P. (2003). Effects of metsulfuron methyl and cypermethrin exposure on freshwater model ecosystems. *Aquatic Toxicology*, 63(3), 243-256.

Wickham, S. A. (1998). The direct and indirect impact of Daphnia and Cyclops on a freshwater microbial food web. *Journal of Plankton Research*, 20(4), 739-755.

Wulff, A., Wängberg, S. Å. K., Sundbäck, K., Nilsson, C., & Underwood, G. J. (2000). Effects of UVB radiation on a marine microphytobenthic community growing on a sand-substratum under different nutrient conditions. *Limnology and Oceanography*, 45(5), 1144-1152.

Zacharia, S., & Kakati, V. S. (2004). Optimal salinity and temperature for early developmental stages of Penaeus merguiensis De man. *Aquaculture*, 232(1-4), 373-382.

Zhao, S.Y., Sun, Y.P., Lin, Q.Q. and Han, B.P. (2013). Effects of silver carp (Hypophthalmichthys molitrix) and nutrients on the plankton community of a deep, tropical reservoir: an enclosure experiment. *Freshwater Biology*, 58(1), 100-113.

Zhu, G., Cao, T., Zhang, M., Ni, L., & Zhang, X. (2014). Fertile sediment and ammonium enrichment decrease the growth and biomechanical strength of submersed macrophyte Myriophyllum spicatum in an experiment. *Hydrobiologia*, 727(1), 109-120.

Zingel, P., Cremona, F., Nõges, T., Cao, Y., Neif, É. M., Coppens, J., ... & Jeppesen, E. (2018). Effects of warming and nutrients on the microbial food web in shallow lake mesocosms. *European Journal of Protistology*, 64, 1-12.