

Supplementary Information 1: Detailed Methods

Net effects of multiple stressors in freshwater ecosystems: a meta-analysis

Jackson, M.C.^{1*}, Loewen, C.J.G.², Vinebrooke, R.D.², Chimimba, C.T.¹ *mjackson@zoology.up.ac.za

1. Centre for Invasion Biology, Department of Zoology and Entomology, University of Pretoria, South Africa
2. Department of Biological Sciences, University of Alberta, Canada

Data search and selection criteria

To identify papers reporting an influence of multiple stressors on freshwater systems we searched the ISI Web of Knowledge database (<http://apps.isiknowledge.com>) between July 2013 and March 2014 for the key word combination (“synerg*” or “antagon*” or “additive” or “combined” or “multiple” or “factorial” or “experiment”) and (“freshwater” or “river” or “stream” or “lake”) and (“impact” or “effect”). In total, approximately 11,000 papers were screened for our study. We only considered those papers that investigated quantifiable effects of stressors independently and in combination compared to a stressor-free control, in nature or under experimental conditions. We only used papers that reported on the effects of extreme or novel manipulations of environmental conditions (i.e., stressors) rather than natural disturbances or interactions (e.g., competition or predation) between native species. Reported stressors included acidification, higher temperatures, exposure to ultraviolet radiation (UVR), contamination (xenobiotics or salinity), eutrophication, habitat alteration (physical manipulation, sedimentation, altered flow regime or drought) and invasive species. Data were acquired directly from text and tables, from figures using Data Thief software (Tummers *et al.*, 2010) or by contacting the corresponding authors.

We use the term ‘observation’ to refer to individual responses used in our analyses, and the term ‘paper’ to refer to their source documents. In several cases, multiple observations were extracted from individual papers where multiple experiments were conducted (i.e., using different sets of species, study locations or stressor combinations) or where multiple organism groups were measured (e.g., producers, invertebrates or vertebrates). If not appropriately summarized at the organism group level, multiple observations were extracted

where measurements were taken on distinct functional groups (e.g., phytoplankton and periphyton). In these cases, multiple observations were considered independent for the purposes of our initial comparisons. Crain *et al.*, (2008) only considered direct stressor impacts in their marine meta-analysis; however, here we include and merge all impacts because it is often difficult to distinguish between direct and indirect effects of stressors in freshwater ecosystems and many studies do not state if the impact is considered indirect (e.g., food-web mediated). Further, indirect effects may be just as important, or more important, than direct effects in terms of ecological relevance. If a study examined the impact of multiple stressors experimentally over time, we only used the longest time period (before complete mortality). When recovery to the stressors was included in the study we used the final sampling point before the recovery period. If a study examined the effect of different degrees of stress (e.g., low and high nutrient enrichment) we only considered the most severe scenario (i.e., high nutrient enrichment) with the exception of studies where the most extreme scenario caused 100% mortality in response organisms; in these cases, we used the second most severe scenario. Several ecotoxicology studies tested the effects of a toxicant across a wide range of temperatures; in these cases, we compared ambient temperatures (usually defined in the papers) to warming of 4 – 10 °C, based on warming predictions by the Intergovernmental Panel on Climate Change (IPCC, 2013). Experiments consisted of laboratory/microcosm studies (taking place inside), mesocosm studies (taking place outside) and *in situ* studies (taking place in a natural system in an enclosure or mesocosm).

We considered observations measured using any of the following response metrics: (i) survival, (ii) growth/size, (iii) condition, (iv) reproductivity, (v) behaviour, (vi) total biomass/abundance, (vii) diversity and (viii) leaf decomposition (Table S1). We included leaf decomposition as it is an important aspect of freshwater ecosystem functioning and was considered in several of our selected papers. Where multiple response metrics were used to

report impact on a single organism group in a single experiment (e.g., both plant biomass and plant diversity), we treated each response metric as an independent observation for inclusion in our ‘full dataset’, which was used only for our full global random effects meta-analysis and our detailed and pooled response metric mixed effects analyses (see ‘Weighted meta-analyses’ below for model details). Our full dataset initially contained 318 observations; however, we then excluded 32 observations where the calculated predicted net effects were deemed impossible (see ‘Interaction effect size calculations’ below for details). Thus our full dataset was reduced to 286 observations.

For the remainder of our comparisons, we excluded all diversity metrics ($n = 31$) and reduced our dataset to include only the most inclusive ‘functional performance’ response metrics per experiment for each organism group. For experiments where multiple response metrics were reported, the single most inclusive response metric was selected on a preferential basis where community responses were preferred over population or organism-level responses and metrics were selected in favour of biomass/abundance over survival, survival over growth/size, growth/size over condition, condition over reproductivity and reproductivity over behaviour. However, if the same experiment measured impact separately on multiple organism groups (e.g., producers and invertebrates), each observation was retained. This ‘most inclusive response metric dataset’ ($n = 230$) was used for the majority of our meta-analyses (i.e., those not specifically comparing response metrics) to minimize data non-independence. See Table S1 for a list of all observations included in our datasets.

Effect size calculation and interpretation

For each observation, we extracted mean, standard deviation and sample size values for each treatment combination (stressor A; stressor B; stressor A and B; no stressor control). Interaction effect sizes were then calculated for each observation in our dataset using Hedges

d , an estimate of the standardised mean difference not biased by small sample sizes (Gurevitch & Hedges 2001). The interaction effect sizes were calculated by comparing the predicted additive effect to the actual observed effect of both stressors. Each interaction effect size was therefore based on the absolute difference between the observed net impact of paired stressors against a hypothetical additive outcome based on the sum of their single independent effects. For each mean response variable (X) from the four treatment groups (i.e., control (u), stressor A (A), stressor B (B) and both stressors A and B (AB)), the predicted additive effect (X_p) was calculated as:

$$X_p = (X_A - X_u) + (X_B - X_u) + X_u$$

At this stage, we removed any studies from our dataset ($n = 32$) where the predicted effect was deemed impossible (e.g., survival greater than 100% or biomass less than zero). Hedge's d effect sizes were then calculated for each observation by comparing the predicted additive effect with the actual observed effect of both stressors applied in combination (AB, o):

$$\frac{X_o - X_p}{S} j$$

where j is a weighting factor based on the number of replicates (n), calculated as:

$$1 - \frac{3}{4(n_o + n_p - 2) - 1}$$

S is the pooled standard deviation, calculated as:

$$\sqrt{\frac{(n_o - 1)\sigma_o^2 + (n_p - 1)\sigma_p^2}{n_o + n_p - 2}}$$

and the predicted standard deviation (σ_p^2) was calculated by pooling σ_A and σ_B and the pooled sample size (n_p) was calculated as $n_A + n_B$. Finally, the variance (Vd) around each interaction effect size was calculated as:

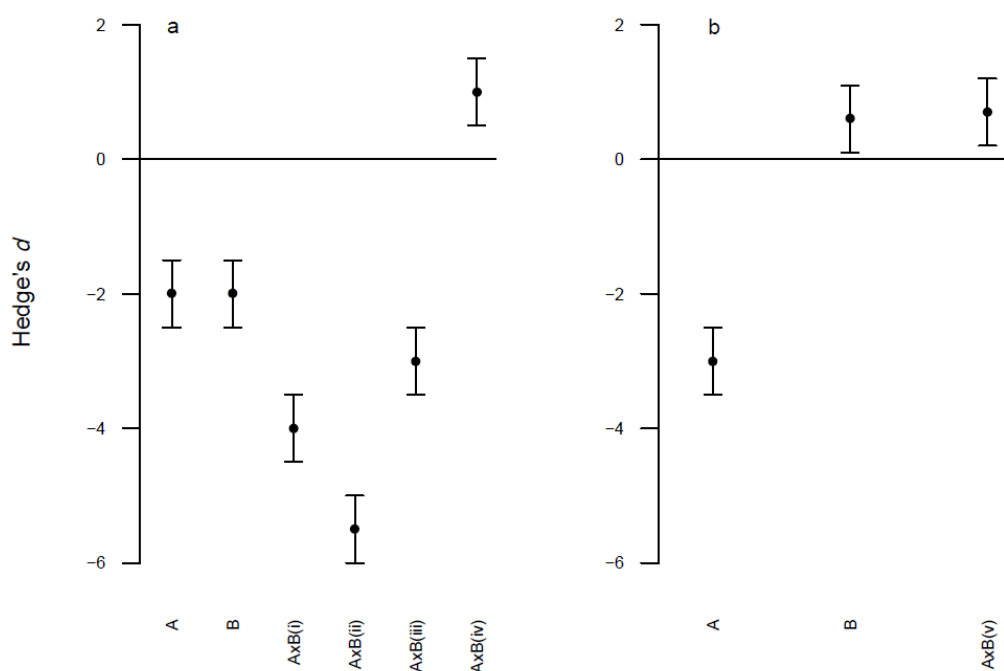
$$\frac{n_o + n_p}{n_o n_p} + \frac{d^2}{2(n_o + n_p)}$$

Because stressors may impart either negative or positive effects on biological receptors, we inverted the response direction/sign (-/+) of our calculated interaction effect sizes where the predicted additive effects (X_p) were negative (i.e., where both independent effects were negative, or if in opposing directions, where the negative effect had the higher absolute value). This was done so that our interaction effect sizes would be directly comparable, regardless of their directionality (Piggot *et al.*, 2015). As a result, an interaction effect size (d) of zero represents an exact additive effect (i.e., a combined effect equal the sum of their independent effects), a positive d denotes a synergistic interaction (i.e., a combined effect greater than the sum of their independent effects) and a negative d reflects either antagonism or a reversal interaction (i.e., a combined effect less than the sum of their independent effects). To distinguish between antagonistic and reversal interactions, we compared the direction (negative or positive, relative to the control) of the observed response to both stressors applied in combination (X_o) with the direction of their predicted additive response (X_p), and assigned reversals where they were opposite. Interaction significance was assessed using 95% confidence intervals calculated around each effect size (from a t -distribution), such that any interactions with intervals crossing zero were deemed additive.

In addition to interaction effect sizes (used in the meta-analytic models described below), we also calculated individual effect sizes for each independent and combined stressor response to assess their directionality. For each mean response variable (X) from the four treatment groups (i.e., control (u), stressor A (A), stressor B (B) and both stressors A and B (AB)), the individual effect size d of each stressor combination (i ; A or B or AB) was calculated using the same Hedge's d equation above where the predicted responses (p) was replaced with the control (u) and the observed response (o) was replaced with the response in the presence of the stressor/s (i). Individual effect sizes reflect responses in the presence of the stressor/s compared to a control and therefore a positive d reflects an increase in survival,

growth, size, condition, biomass, abundance, diversity or decomposition compared to the control; and a negative d indicates a decrease (Fig. S1). Individual effect sizes are useful for illustrating reversal interactions, where the combined effects of stressors are in the opposite direction than that of the strongest individual effect and thus their predicted additive effect, relative to the experimental control.

Fig. S1 The theoretical independent and combined effects of stressors A and B expressed as individual effect sizes (Hedge's d), relative to an experimental control ($= 0$). Negative and positive values represent declines and increases in the measured response variables (e.g., abundance, survival and diversity), respectively. **(a)** When both stressors occur simultaneously, the magnitude of combined effects can be either additive ($=A+B$; i), synergistic ($>A+B$; ii), antagonistic ($<A+B$; iii) or reversed (opposite direction of $A+B$, relative to the control; iv). **(b)** A reversal (v) may also occur where stressors A and B have opposing independent effects if their combined impact is in the opposite direction than that of the larger independent effect.



Weighted meta-analyses

Mean interaction effect sizes across studies were estimated from weighted meta-analyses. In each analysis, ‘Observation ID’ was treated as a random effect to account for the random component of effect size variation among observations, and inverse unconditional variance weights (w) were calculated for each interaction effect size as:

$$\frac{1}{Vd + \hat{\sigma}_{pooled}^2}$$

where $\hat{\sigma}_{pooled}^2$ is the pooled within-class variance estimated from the meta-analytic model (see Gurevitch & Hedges 2001 for further details of random/mixed model analysis). In addition to using random effects meta-analyses to assess the global mean interaction effect sizes across all observations included in our ‘full’ and ‘most inclusive response metric’ datasets, we conducted a series of mixed effects meta-analyses where selected categorical moderators were treated as fixed effects to assess mean interactions at each level of each category (where $n \geq 8$).

Using our full dataset, we conducted a detailed response metric analysis to assess the sensitivity of different response metrics to multiple stressor impacts. We followed this with a pooled response metric analysis, where responses metrics were reassigned as reflecting either ‘diversity’ (plant diversity or animal diversity) or ‘functional performance’ (all other response metrics considered), to assess the sensitivities of these broader response metric categories. We then used our reduced most inclusive response metric dataset to estimate mean effect sizes for different response levels, organism groups, and stressor pairs. Percentile bootstrapped 95% confidence intervals (represented by the lowest and highest 2.5% of the bootstrapped values) were calculated around each mean interaction effect size to assess significance. Similar to the assessment of interaction effect sizes for single observations, a positive mean effect reflects synergy, a negative mean effect reflects antagonism (reversals could not be distinguished with this method) and cases where the confidences intervals

crossed zero were deemed additive. Details of meta-analytical models are as follows (results provided in Table S2; however, significance was only evaluated for groups/levels where $n \geq 8$):

- Model 1 (full global random effects meta-analysis) used the full dataset ($n = 286$), treating ‘Observation ID’ as a random effect;
- Model 2 (detailed response metric mixed effects meta-analysis) used the full dataset ($n = 286$), treating ‘Observation ID’ as a random effect and the moderator ‘detailed response metric’ as a fixed effect (13 levels, including animal biomass/abundance, animal condition, animal diversity, animal growth/size, animal survival, behaviour, decomposition, other biomass/abundance, other survival, plant biomass/abundance, plant diversity, plant growth/size and reproductivity);
- Model 3 (pooled response metric mixed effects meta-analysis) used the full dataset ($n = 286$), treating ‘Observation ID’ as a random effect and the moderator ‘pooled response metric’ as a fixed effect (two levels, including diversity and functional performance);
- Model 4 (reduced global random effects meta-analysis) used the most inclusive response metric dataset ($n = 230$), treating ‘Observation ID’ as a random effect;
- Model 5 (response level mixed effects meta-analysis) used the most inclusive response metric dataset ($n = 230$), treating ‘Observation ID’ as a random effect and the moderator ‘level of biological organization’ as a fixed effect (three levels, including community, organism and population);
- Model 6 (organism group mixed effects meta-analysis) used the most inclusive response metric dataset ($n = 230$), treating ‘Observation ID’ as a random effect and the moderator ‘organism group’ as a fixed effect (six levels, including bacteria, fungi, invertebrate, producer, vertebrate and virus);

- Model 7 (stressor pair mixed effects meta-analysis) used the most inclusive response metric dataset' (n = 229; one observation was dropped from the model because the stressor pair was not replicated), treating 'Observation ID' as a random effect and the moderator 'stressor pair' as a fixed effect (20 levels, including acidification x contamination, acidification x habitat alteration, acidification x UVR, acidification x warming, contamination x contamination, contamination x habitat alteration, contamination x invasion, contamination x nitrification, contamination x UVR, contamination x warming, habitat alteration x habitat alteration, habitat alteration x invasion, habitat alteration x nitrification, habitat alteration x warming, invasion x invasion, invasion x nitrification, invasion x warming, nitrification x UVR, nitrification x warming and warming x UVR).

Vote-counting methods

A vote counting approach was used to complement our weighted meta-analyses, as reversal interactions could not be distinguished from antagonistic interactions using mean interaction effect sizes. Additionally, this method considers the frequency of interaction types among individual studies and therefore considers the occurrence of interactions at a more local scale. Randomisation tests of independence (Monte Carlo approximation using 9,999 permutations) were used to assess whether the frequencies of interaction types differed significantly among levels of each categorical moderator where $n \geq 8$. Randomisation tests, rather than standard chi-squared tests of independence, were used because of their utility for assessing data with small within group sample sizes and expected frequencies (Roff & Bentzen, 1989). Interaction frequencies should be interpreted as the commonality of interaction types among individual studies, as opposed to the mean interaction effect sizes obtained from our weighted meta-analytic models, which reflect the combined responses

across studies. Therefore, smaller interaction effect sizes, which may frequently be assigned as additive in individual studies, may be revealed as non-additive by mean interaction effect sizes owing to greater statistical power.

References

Crain CM, Kroeker K, Halpern BS (2008) Interactive and cumulative effects of multiple human stressors in marine systems. *Ecology Letters*, 11, 1304–1315.

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

Gurevitch J, Hedges LV (2001) Meta-analysis: combining the results of independent experiments. In: *Design and Analysis of Ecological Experiments, Second Edition* (eds Scheiner SM, Gurevitch J), pp. 347–369. Oxford University Press, New York, USA.

IPCC (2013) Working Group I contribution to the IPCC fifth assessment report. *Climate change 2013: the physical sciences basis summary for policymakers*, Cambridge, UK, Cambridge University Press.

Roff DA, Bentzen P (1989) The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. *Molecular Biology and Evolution*, 6, 539–545.

Tummers B, van der Laan J, Huyser K (2010) Data Thief III: Version 1.6. URL: <http://www.datathief.org>.