


REVIEWS AND SYNTHESSES

Meta-analysis on pulse disturbances reveals differences in functional and compositional recovery across ecosystems

Helmut Hillebrand^{1,2,3*}  and Charlotte Kunze^{1,2,3}

The peer review history for this article is available at <https://publons.com/publon/10.1111/ele.13457>

Abstract

Most ecosystems are affected by anthropogenic or natural pulse disturbances, which alter the community composition and functioning for a limited period of time. Whether and how quickly communities recover from such pulses is central to our understanding of biodiversity dynamics and ecosystem organisation, but also to nature conservation and management. Here, we present a meta-analysis of 508 (semi-)natural field experiments globally distributed across marine, terrestrial and freshwater ecosystems. We found recovery to be significant yet incomplete. At the end of the experiments, disturbed treatments resembled controls again when considering abundance (94%), biomass (82%), and univariate diversity measures (88%). Most disturbed treatments did not further depart from control after the pulse, indicating that few studies showed novel trajectories induced by the pulse. Only multivariate community composition on average showed little recovery: disturbed species composition remained dissimilar to the control throughout most experiments. Still, when experiments revealed a higher compositional stability, they tended to also show higher functional stability. Recovery was more complete when systems had high resistance, whereas resilience and resistance were negatively correlated. The overall results were highly consistent across studies, but significant differences between ecosystems and organism groups appeared. Future research on disturbances should aim to understand these differences, but also fill obvious gaps in the empirical assessments for regions (especially the tropics), ecosystems and organisms. In summary, we provide general evidence that (semi-)natural communities can recover from pulse disturbances, but compositional aspects are more vulnerable to long-lasting effects of pulse disturbance than the emergent functions associated to them.

Keywords

Biodiversity, disturbance, experiments, recovery, stability.

Ecology Letters (2020) 23: 575–585

INTRODUCTION

The recovery of ecosystems from pulse disturbances has long been recognised as a central aspect of community organisation (Levin & Paine 1974; Hastings 1980; Menge & Sutherland 1987). Pulse disturbances are defined as sudden, temporally constrained changes in the environment that alter biomass – and potentially also composition – of ecological communities (Jentsch & White 2019). Pulse disturbances are ubiquitous in natural ecosystems, in the form of physical forces (storms, waves, heat, drought, or floods), chemical inputs (pollutants, nutrients, heavy rainfall) or biological processes (pest outbreaks, bioturbation). Understanding the response to pulse disturbances becomes more important as human interventions into Earth's climate and biogeochemistry alter the frequency and extent of naturally occurring pulses and creates novel disturbance types, which often are related to land use (ploughing, tilling, etc) or the extraction of resources (mining, fisheries). This human interference with ecosystem dynamics

has triggered renewed scientific interest in how stable ecological communities are when facing disturbances (Pimm 1984; Ives & Carpenter 2007; Donohue *et al.* 2013; Donohue *et al.* 2016).

The intensity and frequency of pulse disturbances is a central component of classical ecological concepts such as the intermediate disturbance hypothesis (IDH, Connell 1978), dynamic equilibrium theory (Huston 1979), and the stress gradient framework of community organisation (Menge & Sutherland 1976). Even though the underlying assumptions of these concepts are strongly contested (Chesson & Huntly 1997; Fox 2013), the empirical analysis of disturbance responses created a diverse body of scientific literature comprising highly controlled microcosms, semi-natural field experiments and observational studies, all of which compare disturbances to a non-disturbed reference. Another approach analyses temporal community dynamics (before and) after a disturbance event within a single site, as secondary successions often start from disturbance events (Connell & Slatyer 1977).

¹Institute for Chemistry and Biology of Marine Environments [ICBM], Carl-von-Ossietzky University Oldenburg, Schleusenstrasse 1,26382, Wilhelmshaven, Germany

²Helmholtz-Institute for Functional Marine Biodiversity at the University of Oldenburg [HIFMB], Ammerländer Heerstrasse 231, 26129, Oldenburg, Germany

³Alfred Wegener Institute, Helmholtz-Centre for Polar and Marine Research [AWI], Bremerhaven, Germany

*Correspondence: E-mail: helmut.hillebrand@uni-oldenburg.de

In a comparable approach, chronosequences allow addressing recovery when a spatial gradient reflects the time course after the disturbance, and this allows substituting space for time (Peterson *et al.* 2002).

Multiple meta-analyses have summarised this broad empirical literature on disturbances (Mackey & Currie 2001; Jones & Schmitz 2009; Murphy & Romanuk 2012; Murphy & Romanuk 2014). Mackey & Currie (2001) amend the theoretical criticism of the IDH by showing that also the empirical evidence shows inconsistent support of peaked diversity at intermediate levels or frequencies of disturbances. Jones & Schmitz (2009) were most interested in the question whether or not an ecosystem shows signs of recovery after human interventions, providing a vote-counting effort across 240 studies. They concluded that – in principle – recovery is possible in most cases. By contrast, Murphy & Romanuk (2012, 2014) were less focused on recovery, but quantified the predominantly negative impact of the most common types of anthropogenic disturbances on abundance and diversity and especially the predictability of this impact (using the variance among replicates in the response to the disturbance). They found that species richness tended to decline following disturbance (Murphy & Romanuk 2014), whereas different disturbance types had contrasting effects on the predictability of responses (Murphy & Romanuk 2012). This differentiation in predictability may also require different management strategies to enable recovery in disturbed ecosystems. Several additional analyses have quantified the responses to a certain type of disturbance (Cameron *et al.* 2016; Clarke *et al.* 2017) or for a certain ecosystem (Winfree *et al.* 2009; Holden & Treseder 2013).

Despite the overarching picture provided by these and other synthesis efforts, none of them has actually quantified the degree of recovery in different ecosystems and elaborated on which covariates enhance or inhibit recovery from pulse disturbances. Such an analysis is timely given the recent advances in understanding the multiple dimensions of stability composing the response to pulse (and other) disturbances (Kéfi *et al.* 2019). Donohue *et al.* (2013) provided an approach that allows quantifying the response to a disturbance by addressing multiple dimensions of stability. After a pulse disturbance, the response can be driven by immediate resistance (the ability to withstand the disturbance), resilience (the rate of approaching a non-disturbed control), recovery (the degree to which a non-disturbed state is finally reached) and the temporal invariability of the recovery trend. Depending on how correlated these aspects of stability are, the dimensionality of the entire response can range between 1 (all stability aspects strongly correlated) and n , where n is the number of measured aspects, which then are completely independent. In a follow-up analysis of the existing literature on stability, it became evident that the vast majority of studies analysed only 1 aspect of stability at a time (Donohue *et al.* 2016; Kéfi *et al.* 2019). Moreover most studies focused on either functional or compositional responses, but few analysed how functional and compositional recovery are linked each other (but see Baert *et al.* 2016; Gülzow *et al.* 2017).

Using two aquatic global change experiments as test cases, Hillebrand *et al.* (2018b) used a multidimensional approach for both functional and compositional stability and found that

recovery of a function became more likely if the composition recovered as well, whereas lack of compositional recovery could result in higher or lower functionality of the system – or even the same. Radchuk *et al.* (2019) used the same approach to compare the outcome of different models (and model types) and concluded that the dimensionality of stability depends on type and extent of the disturbance. Arnoldi *et al.* (2019) made a similar point by comparing diversity-stability relationships in light of different disturbance types and their impact on different species in the community. They concluded that species abundance relates to systems variability and it is therefore necessary to include all measures of stability to understand and predict ecosystem responses to perturbations. This opens the question of how common and complete functional and compositional recovery is, and whether these are linked to each other and affected by the same variables reflecting disturbance type, ecosystem, organism group, or experimental protocols.

To answer this question, we performed a meta-analysis of pulse disturbance experiments performed in (semi-)natural communities in the field or in large mesocosms. Thus, we neither used press disturbance treatments nor experiments in the lab or on single species. We also included only studies that had a simultaneously running control to which the disturbed treatment could be compared, which excluded studies substituting space for time (e.g. chronosequences) and temporal analyses of post-disturbance succession without a non-disturbed reference location or control. In total, 508 field and in-situ-mesocosm experiments fulfilled these criteria. On average each experiment contributed roughly 10 unique comparisons of disturbed treatments to undisturbed controls, which were different time points or variables measured, such that the overall database consisted of 5152 observations of treatment to control. This data set was used to obtain unique effect sizes for each experiment and dimension of stability (resistance, resilience, recovery, and temporal invariability), if possible for both functional and compositional aspects.

Most studies used biomass, abundance or related variables (such as percent cover, area-specific productivity) to quantify the functional response to the disturbance, which we accordingly used to calculate functional stability. As compositional aspects, many studies provided single diversity metrics such as richness or Shannon diversity estimates, which have been heavily criticised for huge potential artefacts inflicted by their sensitivity to species pool size, relative abundances and sampling effort (Chase & Knight 2013). Therefore, we used whenever possible reported species occurrences (or proportions) or ordinations as more informative measures of dissimilarity between treatment and control. Moreover as a substantial subset of studies provided information on both univariate and multivariate measures of composition, we were able to test how well changes in composition were reflected by univariate diversity indices. Thus, we had four response categories in our meta-analyses, number-based (abundance) or biomass-based functionality, and compositional responses reflected by indices or dissimilarity measures.

We used this data set to test the following hypotheses:

Recovery is the norm and composition as well as functioning at the end of the experiments does not significantly differ

anymore from the control (hypothesis H1a). This recovery can be reached by high resistance or high resilience in communities initially more affected (less resistant) (H1b). The dimensionality of stability depends on the strength of correlations between the different metrics (H1c): We expect negative correlations between resistance and resilience (as following low resistance a system has a higher probability to exhibit faster recovery dynamics, i.e. resilience), but positive correlations between resistance or resilience and recovery. Temporal invariability is expected to be unrelated to the other three aspects of stability.

Compositional recovery and functional recovery show different degree of completion, but are generally linked such that functional recovery is more complete if composition recovers as well (H2). Thus, the dimensionality of stability depends on the strength of correlation between functional and compositional resistance, resilience, recovery and temporal stability, respectively.

We expect substantial variation in the magnitude (and perhaps also sign) of the stability effect sizes and hypothesise that these can be partially explained by systematic differences between ecosystem types (marine, freshwater, terrestrial), organism groups, type of disturbance category, and geographic location (H3). One mechanism for system differences in recovery could be different degree of isolation (freshwater) vs. connectivity (marine, terrestrial), where more continuous space allows for higher spatial recovery. Therefore we also contrasted experiments open and closed to dispersal, and more confined mesocosms vs. field studies. Smaller organisms or warmer regions were expected to exhibit higher resilience and recovery due to higher growth rate potential. Some disturbances could potentially result in longer-lasting effects (e.g. by remaining presence of chemical pulses) than others (e.g. mechanical biomass removal).

MATERIAL AND METHODS

Studies were obtained based on a search at Web of Science (www.webofknowledge.com/WOS, assessed April 3rd, 2018) using the search term '(experiment* or manipul* or mesocosm* or microcosm*) AND recover* AND (disturb* or perturb* or pulse) AND (communit* or composit* or diversit* or assembl*)'. From the 943 studies obtained, titles and abstracts were scanned to retrieve experiments which fulfilled the following three selection criteria: (1) The study comprised a disturbance treatment and an undisturbed control that were run simultaneously and independent of each other. (2) The disturbance was a pulse treatment inflicted upon a community of naturally co-occurring species in the field or in outdoor mesocosms. (3) A community wide metric of function (abundance, biomass) or composition (indices, dissimilarity) was available at least for the end of the experiment (recovery), but preferably for multiple time points following the disturbance. We amended this database by additional studies cited in previous meta-analyses (Jones & Schmitz 2009; Murphy & Romanuk 2012; Murphy & Romanuk 2014).

These criteria led to a database comprising 508 experiments from 110 publications (Appendix S1, Table S1). The data were globally distributed across all ecosystem types, despite a

preponderance of studies in temperate regions (Appendix S2). For each time point we obtained means and standard deviations for the available univariate response variable (abundance, biomass, index) for control and treatment. Biomass and abundance were roughly equally used (254 and 216 experiments, respectively) as functional variables. For diversity indices, 237 experiments provided richness estimates, 18 a common index (mostly Shannon-Wiener). 102 experiments allowed calculating diversity metrics ourselves by providing composition data, for which we followed recommendations for using the effective number of species (ENS) as a metric robust against sampling issues (for details, see Chase & Knight 2013). For all these univariate metrics, we compared treatment and control for every time point using log response ratios (LRR) as effect size and their sampling variance (var.LRR) as a measure of precision (Hillebrand & Gurevitch 2016). Dissimilarity metrics were available for 282 experiments, where the vast majority (275) used an abundance-weighted compositional distance between treatment and control (most commonly Bray-Curtis, $n = 217$). Only a small proportion of studies (7) used presence-absence dissimilarity metrics such as Jaccard.

The original dissimilarities (independent of metric) ranged between 0 and 1, with higher dissimilarity being closer to 1. As all univariate response variables tended to be negatively affected by disturbances ($LRR < 0$ when abundance, biomass or diversity declines), we also expressed dissimilarity in negative numbers with 0 = lowest dissimilarity and -1 being highest dissimilarity. Thereby, positive slopes after the disturbance always reflected increasing recovery over time.

These responses were measured on single trophic levels, and if multiple levels were measured, they were analysed separately. In addition to the responses we obtained information on potential co-variables, for which we created meaningful categories (Appendix S1, Table S1). This includes the system (marine, terrestrial, freshwater), the approach (field or mesocosm), latitude and longitude of the site, whether the experiment was open to dispersal (open or closed), the type of disturbance (physical, chemical, biological, physico-chemical), the duration of the experiment in days, and the group of organisms (plants, invertebrates, vertebrates, microalgae, microbes [bacteria, protists]).

All calculations were performed using R (R Development Core Team 2018; RStudio Team 2018), and meta-analyses were performed using the 'metafor' package (Viechtbauer 2010). To assess the different dimensions of stability, we used the approach outlined by Hillebrand *et al.* (2018b) with minor modifications reflecting that we analysed responses across rather than within studies. Resistance was quantified by the deviation from control (as LRR or dissimilarity) on the first sampling date after the disturbance period. Across studies, this was on average at 8.5% (interquartile 1.6–10.1%) of the experiment duration, and acknowledges the expert knowledge of the authors of the primary study on when to measure disturbance impact. Strong deviations from 0 are congruent with low resistance, and the var.LRR for this time point reflected the sampling variance. Analogously, final recovery was assessed by LRR or dissimilarity compared to the control at the final date of sampling. When reported – we also quantified

the pre-disturbance distance between treatment and control as a baseline for comparison (start). The latter was especially important for dissimilarity, which rarely is 0 and the recovery cannot be measured by returning to a dissimilarity of 0.

We changed the approach by Hillebrand *et al.* (2018b) for resilience and temporal stability, as we replaced the linear regression of LRR to time by a weighted meta-analysis regression using the metafor package. Thus, as resilience metric we obtained the slope of a weighted regression using LRR over log-transformed time, with the inverse of var.LRR as weights. Given the huge span in organism groups and experimental duration, we transformed the duration of the experiment to a scale of 0 to 1, with 1 being the last sampling date. Thereby, we made slopes comparable by acknowledging that the authors of the primary studies have selected a time frame that reflects the recovery potential of the organisms involved. Using absolute duration would make slopes consistently smaller for larger, more long-lived organisms and thus longer-lasting experiments (see Appendix S3 for a comparison of normalised and absolute resilience estimates as well as a discussion of alternative normalisations). As described previously (Hillebrand *et al.* 2018b), using log-transformed relative duration linearised the recovery dynamics and allowed fitting a linear slope. The standard deviation of the residuals around this weighted regression was then obtained to calculate temporal stability as ln-transformed (1/sd). Sampling variance for the slope was its squared standard error, whereas the temporal stability was used unweighted.

To test hypothesis 1, we used a weighted random effect meta-analysis to estimate grand means (\pm 95% confidence intervals) for each of the stability aspects (resistance, resilience, and recovery) across all experiments. We calculated these separately for each of the four response categories (abundance, biomass, index, composition) by including the unique identifier for each experiment as random variable. The same approach was used for temporal stability, which however lacked sampling variance resulting in an unweighted meta-analysis. We used the 95% confidence intervals (CI) to test hypothesis H1a, which was falsified if the overall effect size for recovery differed from zero (i.e. when the CI did not include 0) or – in case of dissimilarity – from the pre-disturbance dissimilarity. We complement these analyses by also showing the density distribution of effect sizes for each response category and stability aspect in the supplementary material, using the 'ggplot2' (Wickham 2016) and 'ggridges' (Wilke 2018) packages.

We further tested for relationships between the different aspects of stability (H1b and H1c), using weighted correlations in the 'weights' package in R (Pasek 2018). As weights we used a combination of the weights of both correlated aspects, for example, resistance and resilience, by obtaining the log-transformed mean of the inverse sampling variances of both correlated aspects. Rare cases with missing sampling variances were used with a weight of 1.

We used the same approach to assess the correlations between functional and compositional aspects of the response, testing H2. For each aspect of stability (resistance, resilience, recovery, temporal), we assessed the weighted correlation between any combination of functional (biomass, abundance)

and compositional (index, composition) response. The lack of significant correlations would falsify hypothesis H2.

Finally, we tested H3 by amending the above mentioned random effect meta-analysis for resistance, resilience, temporal stability and recovery, by moderators potentially influencing the stability. System, response category, organisms group, openness of the experiment to dispersal, type of experiment (mesocosm vs. field), and disturbance type were categorical moderators, absolute latitude a continuous predictor (Table S1).

RESULTS

Before the onset of the disturbance (Fig. 1, start), control and treatment plots showed non-significant differences in abundance, biomass, composition, and standing diversity ($P > 0.1$ for all four response categories). Only for composition, the mode of the density distribution was non-zero given that any two assemblages will show dissimilarity in species inventory and abundance (Appendix S4, Fig. S4-1). These differences between control and treatment significantly increased due to the disturbance (Fig. 1, resistance), with all response categories showing on average negative responses to the disturbance. Average effect sizes for resistance corresponds to a remaining biomass of 40%, remaining abundance of 53%, and remaining standing diversity of 59% (all significantly different from zero, $P < 0.0001$). Dissimilarity almost doubled on average from pre- to post-disturbance.

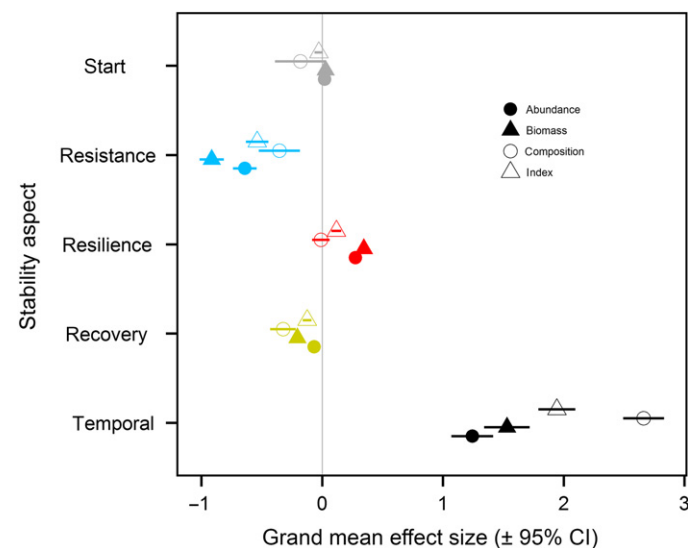


Figure 1 Grand mean effect sizes and their 95% confidence intervals for different dimensions of stability following a pulse disturbance. The four dimensions differ in their metric (resistance and resilience are log response ratios, resilience is a weighted slope through log response ratios over time, temporal stability is the inverse of residual variance around this trend), the overall effect sizes are unitless grand means from weighted meta-analyses. These are compared to the pre-disturbance difference between control and treatment (start). The effect sizes are separated for the different response categories abundance, biomass, composition and index. For interpretation, a vertical line at 0 indicates highest possible resistance and recovery (no deviation to control), but lowest resilience (no trend) and temporal stability

Across studies, grand mean estimates for resilience were significantly positive for all univariate measures (abundance, biomass, diversity indices), and thus opposite in sign to the resistance effect size indicating a general recovery trend (Fig. 1, resilience, Appendix S4, Fig. S4-1). However, the grand mean resilience for composition was not significantly different from zero ($P = 0.78$). When comparing the magnitudes of the resilience grand means, compositional response categories (composition, indices) showed significantly lower resilience as the 95% CIs did not overlap with the grand mean resilience in functional variables (abundance, biomass).

For final recovery, the grand mean effect size for all response categories remained significantly different from zero, but the magnitude of this difference was in most cases smaller than after the disturbance, indicating significant yet incomplete recovery (Fig. 1, recovery, Fig. S4-1). Recovery of abundance was most complete as the grand mean corresponded to 94% recovery (still significantly < 0 , $P = 0.002$), compared to 82% for biomass ($P < 0.0001$). Whereas the grand mean for univariate metrics of diversity also indicated a substantial recovery (88%, $P < 0.0001$), the multivariate assessment of composition remained highly different from the control. In fact, the grand mean effect for recovery in composition (average = -0.32 , CI = -0.43 to -0.22) was highly similar to the estimate for resistance (average = -0.36 , CI = -0.52 to -0.18). Temporal stability was higher in compositional than functional response categories (Fig. 1, temporal).

As some of the 508 experiments reported both ordination and index related compositional variables, we had 561 unique combinations to test multiple dimensions of stability in functional and compositional space. From these, 131 experiments showed a further deviation from the control after the disturbance, compared to 430 with a recovering trend. However, only 11 of these experiments with departing trajectories showed resilience estimates significantly different from zero at a study-wide P -level of 0.1. Thus 23% of all studies showed some sign of lacking recovery, but significant novel trajectories after the pulse were rare ($< 2\%$ of all 508 experiments). By contrast, 171 of the 430 experiments with a recovery trend had study-wide P -levels < 0.1 (30% of all experiments).

Resilience and resistance were strongly negatively correlated (Fig. 2a), i.e., systems with low resistance showed high resilience irrespective of whether low resistance was reflecting a loss in biomass, abundance or diversity or an increase in dissimilarity (Appendix S4, Fig. S4-2). By contrast recovery did only weakly correlate with resilience (Fig. 2b), only for composition, a strong positive correlation between resilience and recovery appeared (Appendix S4, Fig. S4-2). Resistance was correlated to recovery, as the most resistant systems also showed little deviation between control and treatment at the end of the experiment, independent of the response category (functional or compositional) addressed (Fig. 2c, Fig. S4-2). When comparing start and recovery LRR (Fig. 2d), it becomes obvious that recovery remained incomplete, indicated

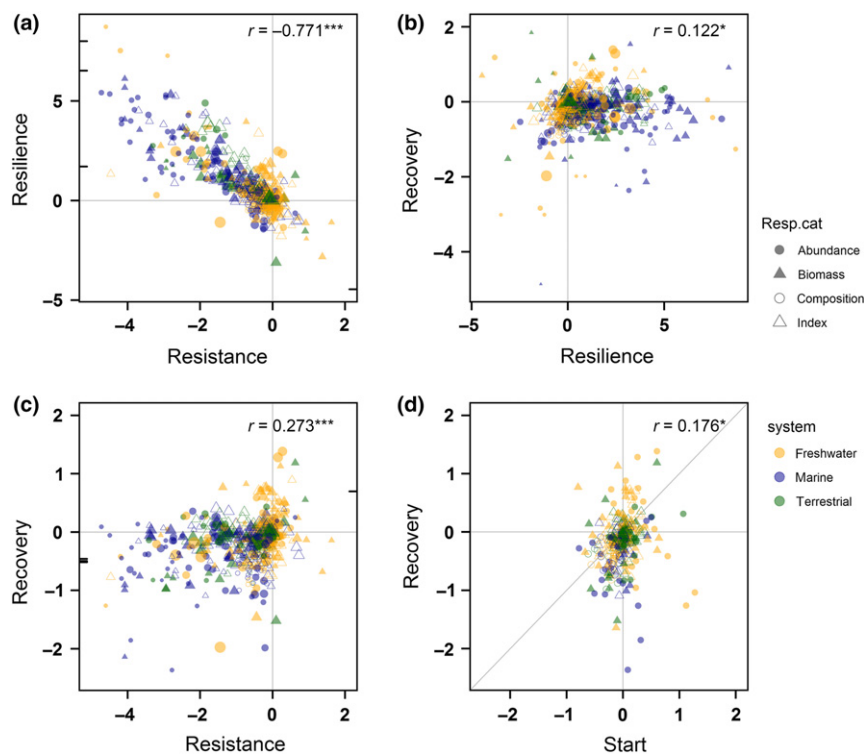


Figure 2 Correlations between different aspects of stability in response to a pulse disturbance. Symbols denote response categories (resp.cat), colours denote the ecosystem type. (a) Resistance vs. resilience. (b) Resilience vs. recovery. (c) Resistance vs. recovery. (d) Recovery vs. start differences between control and treatment. Symbol size corresponds to the weight of the experiment, weighted correlation coefficients are given for each panel including significance levels ($***P < 0.001$, $*P < 0.05$, n.s. = not significant). Please note: Vertical and horizontal lines at 0 have different meanings for the different stability aspects, reflecting highest stability for resistance and recovery (no deviation to control), but lowest stability for resilience (no trend). In panel D, the diagonal line is the 1:1 relationship. See Fig. S4-2 for a display separating the response categories

by stronger vertical than horizontal deviation of the cases from the 1:1 line.

Temporal stability was correlated significantly with the majority of compositional and functional dimensions of stability (Appendix S4, Fig. S4-4). High temporal stability was always associated to high resistance and recovery. However, high resilience tended to be associated to low temporal stability, indicating that fast recovery was related to high variance in the response categories. Index- and dissimilarity-based assessments of compositional stability were also overall significantly correlated (Appendix S4 Fig. S4-S5; $r > 0.3$, $P < 0.01$ for all four stability aspects). However, especially for recovery the bivariate assessment reinforced the conclusion that full recovery of univariate diversity metrics (log response ratio = 0) did not coincide with a full recovery of community composition.

Compositional and functional levels of stability were positively correlated across studies for each of the four aspects of stability studied (Fig. 3). High resistance in functions was coupled to high compositional resistance, especially when the latter was based on univariate indices of composition. The same correlation appeared for resilience and temporal stability. For recovery, the correlation was weaker, but still a full recovery of the composition mostly coincided with a high probability of full functional recovery.

Few predictors had strong significant impacts across the different aspects of stability and the models explained only a

moderate fraction of the heterogeneity in the effect sizes (pseudo- R^2 between 0 and 35%, Appendix 5, Table S1). Only the models for temporal stability showed higher explanatory power. Still, for each of the four stability aspects significant moderators altered the response to the pulse disturbance. For resistance (Fig. 4a), freshwater ecosystems showed higher compositional resistance than terrestrial and higher functional resistance than marine ecosystems. Invertebrates had lower functional resistance than most other organisms (except plants) and lower compositional resistance than vertebrates, which were most resistant overall. Resistance generally increased with latitude, and was higher in mesocosms than field experiments for composition.

Resilience often showed opposite moderator effects (Fig. 4b), with higher resilience in terrestrial systems than in freshwater, lower functional resilience in most organisms groups (except plants) compared to invertebrates (and lowest in vertebrates), as well as lower resilience in mesocosms. Additionally, compositional stability measured by dissimilarity showed a significantly reduced resilience compared to index-based measures of dissimilarity, which resulted also in lower recovery (Fig. 4c). Recovery was higher in freshwater than in both other ecosystem types for both composition and function. Recovery was highly similar between organism groups (as resistance and resilience effects cancelled each other out), with the exception of lower functional recovery in vertebrates, and higher compositional recovery in microbes.

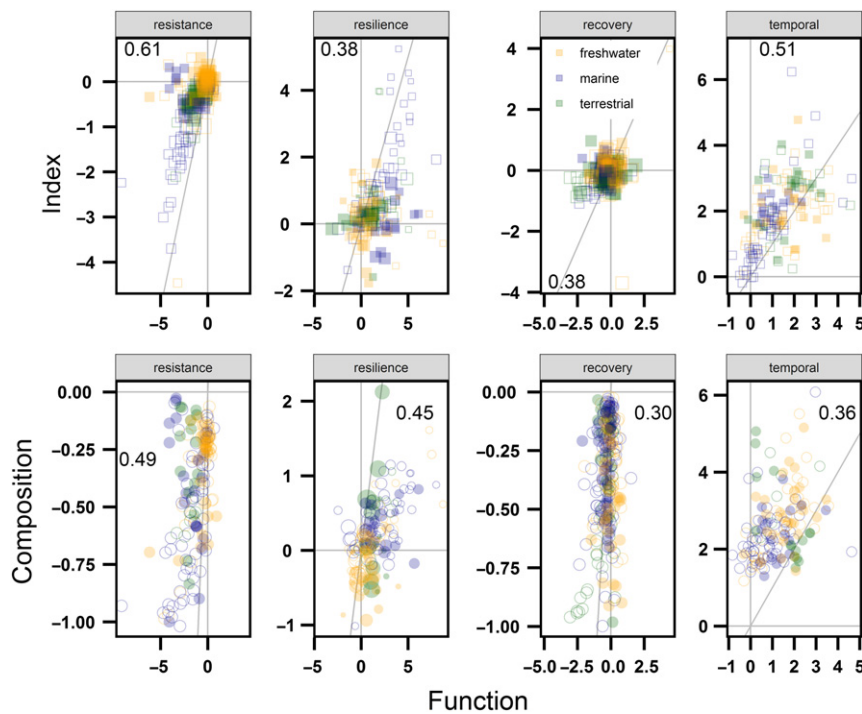


Figure 3 Correlation between compositional and functional stability in response to a pulse disturbance. The correlation is separated for different aspects of stability (resistance, resilience, recovery, temporal stability) and for experiments measuring compositional stability by univariate indices or by assessing compositional dissimilarity. Colours denote the ecosystem type, experiments either used biomass (filled symbols) or abundance (open symbols) as functional response category. Symbol size corresponds to the weight of the experiment, weighted correlation coefficients are given for each panel (all significant at $P < 0.001$, only for recovery both correlations at $P < 0.01$). Vertical and horizontal lines at 0 have different meanings for the different stability aspects, reflecting highest stability for resistance and recovery (no deviation to control), but lowest stability regarding resilience (no trend) and temporal stability (high variance). The diagonal line is the 1:1 relationship

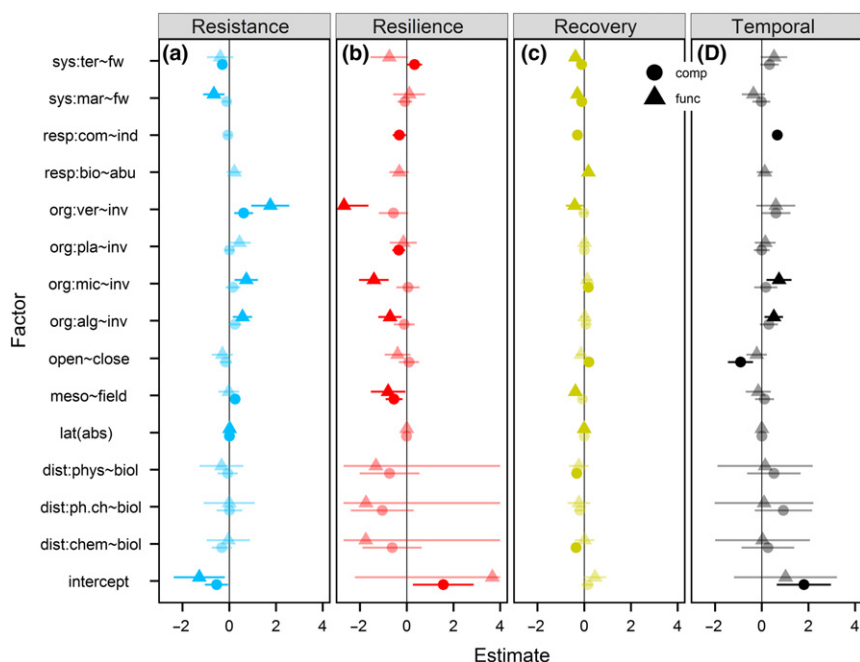


Figure 4 Moderators of the random effect meta-analysis of (a) resistance, (b) resilience, (c) recovery and (d) temporal stability. For each moderator, significant deviations from the intercept are given in bold symbols, non-significant in light symbols. Error bars denote the standard error of the parameter estimate, which has been truncated if larger than the range (–2 to + 4) plotted here. A full list of estimates can be found in Table S2. The factors analysed are abbreviates as **system**, **response category**, **organism group**, and **disturbance type**, the contrasts within the factors are specified with ~ and comprise **terrestrial** or **marine** ~ freshwater, **composition** ~ **index**, **biomass** ~ **abundance**, **vertebrate** or **plant** or **microbes** or **microalgae** ~ **invertebrate**, **open** ~ **close** dispersal, **mesocosm** ~ **field** studies, **physical** or **physico-chemical** or **chemical** ~ **biological** disturbances. The slope with absolute latitude **lat** (**abs**) and intercept are given as well

Compositional recovery was higher when systems were open to dispersal and disturbance was biological. Functional recovery was higher at low latitudes and rather for biomass than abundance as response category. Higher temporal stability was associated to unicellular organisms (microbes, microalgae) compared to invertebrates for functional stability (Fig. 4d). For composition, closed systems and experiments using dissimilarity showed higher temporal stability.

DISCUSSION

Summarising the results of 508 field and mesocosm experiments, our meta-analysis on recovery converged on three main conclusions: First, recovery is the norm but remains incomplete at least for the duration of the studies included here (hypothesis H1a). The degree of recovery was higher if resistance was high or – less pervasive – if resilience was high, whereas between resistance and resilience a strong negative correlation emerged (H1b and H1c). Second, compositional recovery was lower than functional recovery, especially when dissimilarity- or ordination-based methods were used to address composition. Still, functional and compositional stability aspects were generally positively correlated such that experiments with high stability in composition were also functionally more stable (H2). Third, stability aspects differed between systems, organisms and experimental approaches, although the unexplained heterogeneity in effect sizes remained high (H3). Some of these moderator effects were

highly congruent across aspects of stability, whereas others were specific for single aspects and only compositional or functional stability. These results have strong relevance for our ecological understanding of disturbances and stability, and also for ecosystem management. Before addressing these, we start with discussing some methodological issues and caveats.

We chose a weighted meta-analysis of log response ratios, and thus focused on the relative response to the disturbance. The weighting reflected differences in replication and setup between experiments by including the statistical reliability of stability estimates. It has been argued that such weighting downgrades more natural experiments with low replication compared to more artificial experiments allowing high replication, but (1) this general bias seems to be non-existent (Hillebrand & Gurevitch 2014) and (2) the present analysis does not include artificial lab experiments. Thus, we are highly confident that our analysis is robust and provides a substantial overview of our current evidence regarding stability following pulse disturbances. Our decision to normalise resilience slopes to the duration of the experiment is extensively discussed in the appendix (Appendix S3), but comparison to results without this normalisation reveals minor changes in the outcome of the meta-analysis.

As in all meta-analyses, our conclusions can only be as robust as the primary literature is. As in many empirical syntheses (Hillebrand 2009), the primary studies on disturbance are biased towards the temperate zones of Europe and North

America, which reflects an existing lack of publications from other regions, especially the tropics. Even though the slopes were small, our meta-analysis showed a latitudinal gradient in resistance (increasing) and recovery (decreasing) (H3). However, this does not necessarily mean that tropic systems recover faster as there are very few studies at latitudes below 23°N or S. We thus strongly advocate to experimentally analyse the role of pulse disturbance and the recovery potential in tropical regions. This does not only allow filling a gap, as tropical areas may show different responses as they, for example, are characterised by higher species richness and spatial beta-diversity as well as higher temperature and potentially lower generation times. A recent global analysis of grasslands showed that the temporal dynamics of communities strongly depends on the spatial diversity allowing colonisation (Hodapp *et al.* 2018). Mapped to recovery, this suggests that disturbance studies in the field would strongly profit from a metacommunity approach taking the spatial component of diversity outside the study plots into consideration, as has been shown in some microcosm experiments (Gülzow *et al.* 2017). Other potential gaps in the present scientific literature are vertebrate studies, polar marine studies as well as deep sea studies, all of which are strongly underrepresented. Despite these caveats, however, we would like to stress that – for the evidence available – the grand mean effect sizes were highly consistent and statistically highly significant, allowing us to provide conclusion for the existing large set of experiments.

The majority of studies showed clear signs of recovery and the overall weighted effect sizes clearly indicated that especially functional aspects of the disturbed communities approached the non-disturbed controls. Additionally, very few experiments (<2%) continued to deviate more from the control after the disturbance. With the exception of compositional dissimilarity (see below), it is valid to conclude that disturbances rarely pushed the communities to novel trajectories for functioning or univariate diversity. The discussion whether or not disturbances can push ecological communities to alternative states has been active in the ecological literature (Didham *et al.* 2005; Jiang & Patel 2008; Fukami & Nakajima 2011; Graham *et al.* 2015). At least from the set of experiments analysed here, we conclude that this risk is low, given the preponderance of recovery dynamics. Switching to an alternative state requires multiple basins of attraction for community dynamics (May 1977) and the crossing of an environmental threshold leading to a non-transient change (Knowlton 1992). Despite the large magnitude of change induced by some of the pulses (see mean and variance for resistance estimates in Fig. 1 & Fig. S4-1), such thresholds were rarely crossed.

Thus, recovery seems to be the norm for the functional aspects of communities and for composition when collapsed to a univariate biodiversity metric ignoring species identities (Fig. 1). However, for multivariate assessments of composition resilience was close to 0 and the grand mean recovery effect size remained highly different compared to the start (Fig. 1), reinforcing the statement that changes in species composition are not well reflected by changes in richness or other measures (Chase & Knight 2013; Chase *et al.* 2018;

Hillebrand *et al.* 2018a). A third of all studies measuring multivariate compositional resilience (57 out of 162) showed further divergence of composition after the disturbance, including seven of the eight studies with significantly increasing deviations. The potential reasons for such a long-lasting effect on compositions are manifold. In isolated systems, species not surviving the disturbance cannot re-immigrate, and thus full recovery of composition remains impossible. At least at the level of detail we can infer in our analysis, the higher compositional recovery in experiments open to dispersal is in line with such an explanation. However, even with dispersal, community composition may remain altered if composition depends on the history of community assembly, for example, through priority effects (Fukami 2015; Ojima & Jiang 2017), or disturbance induced changes prevent the re-immigration of sensitive species even after the disturbance ceases.

Further, our results indicate – as has been previously described – that functions can be restored even in the absence of compositional recovery (Hoover *et al.* 2014), as functional redundancy can allow a different community to perform nearly the same functions (Allison & Martiny 2008). Models (Yachi & Loreau 1999; Cottingham *et al.* 2001) and empirical assessments of biodiversity – invariability relationships (Allan *et al.* 2011; Hautier *et al.* 2014) converged on the conclusions that it actually is the change in composition that allows maintaining a stable ecosystem function under changing environmental conditions. A recent microcosm study on ciliates found also overall ecosystem stability (based on a multifunctionality approach) to depend on biodiversity, but in a complex non-linear way as different functions may scale differently to the species present (Pennekamp *et al.* 2018). Thus, pulse disturbance may select for different assemblages, which can only maintain functional stability if there is functional redundancy.

Whereas this result highlights the ability of maintaining functionality under altered composition, the converse conclusion is true as well: if composition recovers, there is a high chance of functional recovery as well, reflected by the positive correlation between functional and compositional levels of all aspects of stability (resistance, resilience, recovery, temporal stability). Functional recovery varies widely for experiments which lack compositional recovery, but becomes predictable and more complete if the recovery of composition was high. Similar results have previously been obtained from single lab (Gülzow *et al.* 2017) and field (Hillebrand *et al.* 2018b) experiments.

Our results may thus have important consequences for management of ecosystems affected by acute changes in conditions, which is an important aspect of restoration ecology (Geist & Hawkins 2016). The obvious gap between fast recovery of functional variables and the slow (or even absent) recovery of the concise species composition indicates that restoration measures for conservation may not yield short-term effects (Borja *et al.* 2010). At the same time, the success of such measures cannot be monitored by functional variables alone.

The correlation structure between metrics of post-disturbance stability and their functional vs. compositional magnitude reflect previous conclusions that stability is a

multi-dimensional construct, but with a dimensionality smaller than the number of aspects measured (Hillebrand *et al.* 2018b; Radchuk *et al.* 2019). In addition to the correlation between resistance and resilience, which reflects the way the stability response is decomposed, we found an ecologically obvious relationship between resistance and recovery, as a system withstanding disturbance will show minimal deviances between control and treatment at the end. The weaker, but significant correlation between resilience and recovery also indicated that faster trends led to higher recovery. Additionally, we would like to emphasise that temporal stability around the recovery trend shows a very distinct relationship to all three other metrics, reflecting that temporal stability is high if the community is not strongly affected (resistance maximal close to 0) or does not change afterwards (resilience minimal close to 0).

In summary, acknowledging again the potential gaps in the scientific literature, we found very consistent grand mean effects across studies. Resilience and recovery were highly predictable for functions and emergent properties of the composition, but not for the concise composition of species. Variance around these grand mean effects was comparably small, but still systematic differences in stability aspects between studies emerged. In addition to latitude and system openness (see above), main differences were found between ecosystem types and between organism groups. The level of information from the original studies does not allow disentangling actual mechanisms for these differences, as obviously traits of the ecosystems (connectivity, size) and the organisms (size, mobility, trophic position, life history) will alter the susceptibility to and recovery from disturbances (Cardillo *et al.* 2005; Starzomski & Srivastava 2007; Sunday *et al.* 2012; Pinsky *et al.* 2019). Thus, the observed differences open up novel questions for future research: does the higher recovery and resistance in freshwater systems reflect that these undergo more frequent disturbances based on their relative small size and isolation compared to land masses and marine realms? Which traits of organisms allow predicting tolerance to or recovery from disturbance, and thus explain the difference between organism groups involved? We found little evidence for different types of disturbances affecting the different stability metrics, in contrast to previous meta-analyses on disturbance effects (Murphy & Romanuk 2012; Murphy & Romanuk 2014). However, because of the multivariate assessment of predictors, our categorisations had to be rather coarse, which does not preclude the existence of differences at finer scales between different anthropogenic disturbance types as reported by Murphy & Romanuk (2012, 2014).

Some clear recommendations for future research emerge from the results above. Tropical disturbance studies (and other unrepresented regions and organisms groups) are needed, as well as meta-community studies with an explicit assessment of how spatial heterogeneity affects temporal dynamics of species composition and functions after disturbance. Cross-system approaches would help to understand different stability in different ecosystem types. Given that multiple dimensions of stability prevail, we especially would recommend studies that allow disentangling which species

contributes how much to the different aspects of stability following a disturbance. Thereby, traits could be mapped for species being resistant, resilient or temporally invariant and eventually be transformed into a predicted stability, which could be compared to observed stability.

ACKNOWLEDGEMENT

HH acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG HI 848/26-1), whereas CK was funded by HIFMB, a collaboration between the Alfred-Wegener-Institute, Helmholtz-Center for Polar and Marine Research, and the Carl-von-Ossietzky University Oldenburg, initially funded by the Ministry for Science and Culture of Lower Saxony and the Volkswagen Foundation through the 'Niedersächsisches Vorab' grant program (grant number ZN3285). We thank Friso Muijers, Andrea Mentges and Svenja Mintenig, who helped with literature search, and Barbara Bauer for comments on the paper.

AUTHORSHIP

HH designed the study. HH and CK conducted the literature search and extracted the information from the primary studies. HH performed the statistical analyses and both authors wrote the manuscript.

DATA AVAILABILITY STATEMENT

Primary data are from published studies, the full data set with all extracted effect sizes has been stored at DataDryad <https://doi.org/10.5061/dryad.cz8w9gj09>.

REFERENCES

- Allan, E., Weisser, W., Weigelt, A., Roscher, C., Fischer, M. & Hillebrand, H. (2011). More diverse plant communities have higher functioning over time due to turnover in complementary dominant species. *Proc. Natl Acad. Sci.*, 108, 17034–17039.
- Allison, S.D. & Martiny, J.B.H. (2008). Resistance, resilience, and redundancy in microbial communities. *Proc. Natl Acad. Sci. USA*, 105, 11512–11519.
- Arnoldi, J.-F., Loreau, M. & Haegeman, B. (2019). The inherent multidimensionality of temporal variability: how common and rare species shape stability patterns. *Ecol. Lett.*, 22, 1557–1567.
- Baert, J.M., De Laender, F., Sabbe, K. & Janssen, C.R. (2016). Biodiversity increases functional and compositional resistance, but decreases resilience in phytoplankton communities. *Ecology*, 97, 3433–3440.
- Borja, Á., Dauer, D.M., Elliott, M. & Simenstad, C.A. (2010). Medium- and long-term recovery of estuarine and coastal ecosystems: patterns, rates and restoration effectiveness. *Estuaries Coasts*, 33, 1249–1260.
- Cameron, E.K., Vila, M. & Cabeza, M. (2016). Global meta-analysis of the impacts of terrestrial invertebrate invaders on species, communities and ecosystems. *Glob. Ecol. Biogeogr.*, 25, 596–606.
- Cardillo, M., Mace, G.M., Jones, K.E., Bielby, J., Bininda-Emonds, O.R.P., Sechrest, W. *et al.* (2005). Multiple causes of high extinction risk in large mammal species. *Science*, 309, 1239–1241.
- Chase, J.M. & Knight, T.M. (2013). Scale-dependent effect sizes of ecological drivers on biodiversity: why standardised sampling is not enough. *Ecol. Lett.*, 16, 17–26.
- Chase, J.M., McGill, B.J., McGlinn, D.J., May, F., Blowes, S.A., Xiao, X. *et al.* (2018). Embracing scale-dependence to achieve a deeper

- understanding of biodiversity and its change across communities. *Ecol. Lett.*, 21, 1737–1751.
- Chesson, P. & Huntly, N. (1997). The roles of harsh and fluctuating conditions in the dynamics of ecological communities. *Am. Nat.*, 150, 519–553.
- Clarke, L.J., Hughes, K.M., Esteves, L.S., Herbert, R.J.H. & Stillman, R.A. (2017). Intertidal invertebrate harvesting: a meta-analysis of impacts and recovery in an important waterbird prey resource. *Mar. Ecol. Prog. Ser.*, 584, 229–244.
- Connell, J.H. (1978). Diversity in tropical rain forests and coral reefs. *Science*, 199, 1302–1310.
- Connell, J.H. & Slatyer, R.O. (1977). Mechanisms of succession in natural communities and their role in community stability and organization. *Am. Nat.*, 111, 1119–1144.
- Cottingham, K.L., Brown, B.L. & Lennon, J.T. (2001). Biodiversity may regulate the temporal variability of ecological systems. *Ecol. Lett.*, 4, 72–85.
- Didham, R.K., Watts, C.H. & Norton, D.A. (2005). Are systems with strong underlying abiotic regimes more likely to exhibit alternative stable states? *Oikos*, 110, 409–416.
- Donohue, I., Petchey, O.L., Montoya, J.M., Jackson, A.L., McNally, L., Viana, M. *et al.* (2013). On the dimensionality of ecological stability. *Ecol. Lett.*, 16, 421–429.
- Donohue, I., Hillebrand, H., Montoya, J.M., Petchey, O.L., Pimm, S.L., Fowler, M.S. *et al.* (2016). Navigating the complexity of ecological stability. *Ecol. Lett.*, 19, 1172–1185.
- Fox, J.W. (2013). The intermediate disturbance hypothesis should be abandoned. *Trends Ecol. Evol.*, 28, 86–92.
- Fukami, T. (2015). Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annu. Rev. Ecol. Syst.*, 46, 1–23.
- Fukami, T. & Nakajima, M. (2011). Community assembly: alternative stable states or alternative transient states? *Ecol. Lett.*, 14, 973–984.
- Geist, J. & Hawkins, S.J. (2016). Habitat recovery and restoration in aquatic ecosystems: current progress and future challenges. *Aquat. Conserv.*, 26, 942–962.
- Graham, N.A.J., Jennings, S., MacNeil, M.A., Mouillot, D. & Wilson, S.K. (2015). Predicting climate-driven regime shifts versus rebound potential in coral reefs. *Nature*, 518, 94–+.
- Gülzow, N., Muijsers, F., Ptacnik, R. & Hillebrand, H. (2017). Functional and structural stability are linked in phytoplankton metacommunities of different connectivity. *Ecography*, 40, 719–732.
- Hastings, A. (1980). Disturbance, coexistence, history and competition for species. *Theor. Pop. Biol.*, 18, 363–373.
- Hautier, Y., Seabloom, E.W., Borer, E.T., Adler, P.B., Harpole, W.S., Hillebrand, H. *et al.* (2014). Eutrophication weakens stabilizing effects of diversity in natural grasslands. *Nature*, 508, 521–525.
- Hillebrand, H. (2009). Meta-analysis of grazer control of periphyton biomass across aquatic ecosystems. *J. Phycol.*, 45, 798–806.
- Hillebrand, H. & Gurevitch, J. (2014). Meta-analysis results are unlikely to be biased by differences in variance and replication between ecological lab and field studies. *Oikos*, 123, 794–799.
- Hillebrand, H. & Gurevitch, J. (2016). Meta-Analysis and Systematic Reviews in Ecology. In: eLS. John Wiley & Sons Ltd.
- Hillebrand, H., Blasius, B., Borer, E.T., Chase, J.M., Downing, J.A., Eriksson, B.K. *et al.* (2018a). Biodiversity change is uncoupled from species richness trends: consequences for conservation and monitoring. *J. Appl. Ecol.*, 55, 169–184.
- Hillebrand, H., Langenheder, S., Lebet, K., Lindström, E., Östman, Ö. & Striebel, M. (2018b). Decomposing multiple dimensions of stability in global change experiments. *Ecol. Lett.*, 21, 21–30.
- Hodapp, D., Borer, Elizabeth T., Harpole, W.S., Lind, Eric M., Seabloom, Eric W., Adler, Peter B. *et al.* (2018). Spatial heterogeneity in species composition constrains plant community responses to herbivory and fertilisation. *Ecol. Lett.*, 21, 1364–1371.
- Holden, S. & Treseder, K. (2013). A meta-analysis of soil microbial biomass responses to forest disturbances. *Frontiers in Microbiology*, 4, 1–17.
- Hoover, D.L., Knapp, A.K. & Smith, M.D. (2014). Resistance and resilience of a grassland ecosystem to climate extremes. *Ecology*, 95, 2646–2656.
- Huston, M. (1979). A general hypothesis of species diversity. *Am. Nat.*, 113, 81–101.
- Ives, A.R. & Carpenter, S.R. (2007). Stability and diversity of ecosystems. *Science*, 317, 58–62.
- Jentsch, A. & White, P. (2019). A theory of pulse dynamics and disturbance in ecology. *Ecology*, 100, e02734.
- Jiang, L. & Patel, S.N. (2008). Community assembly in the presence of disturbance: A microcosm experiment. *Ecology*, 89, 1931–1940.
- Jones, H.P. & Schmitz, O.J. (2009). Rapid recovery of damaged ecosystems. *PLoS ONE*, 4, e5653.
- Kéfi, S., Domínguez-García, V., Donohue, I., Fontaine, C., Thébault, E. & Dakos, V. (2019). Advancing our understanding of ecological stability. *Ecol. Lett.*, 22, 1349–1356.
- Knowlton, N. (1992). Thresholds and multiple stable states in coral reef community dynamics. *Am. Zool.*, 32, 674–682.
- Levin, S.A. & Paine, R.T. (1974). Disturbance, patch formation, and community structure. *Proc. Natl Acad. Sci. USA*, 71, 2744–2747.
- Mackey, R.L. & Currie, D.J. (2001). The diversity-disturbance relationship: is it generally strong and peaked? *Ecology*, 82, 3479–3492.
- May, R.M. (1977). Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature*, 269, 471–477.
- Menge, B.A. & Sutherland, J.P. (1976). Species diversity gradients: synthesis of the roles of predation, competition, and temporal heterogeneity. *Am. Nat.*, 110, 351–369.
- Menge, B.A. & Sutherland, J.P. (1987). Community regulation: variation in disturbance, competition, and predation in relation to environmental stress and recruitment. *Am. Nat.*, 130, 730–757.
- Murphy, G.E.P. & Romanuk, T.N. (2012). A meta-analysis of community response predictability to anthropogenic disturbances. *Am. Nat.*, 180, 316–327.
- Murphy, G.E.P. & Romanuk, T.N. (2014). A meta-analysis of declines in local species richness from human disturbances. *Ecol. Evol.*, 4, 91–103.
- Ojima, M.N. & Jiang, L. (2017). Interactive effects of disturbance and dispersal on community assembly. *Oikos*, 126, 682–691.
- Pasek, J. (2018). Package "weights". In: <https://CRAN.R-project.org/package=weights>.
- Pennekamp, F., Pontarp, M., Tabi, A., Altermatt, F., Alther, R., Choffat, Y. *et al.* (2018). Biodiversity increases and decreases ecosystem stability. *Nature*. 563(7729), 109–112
- Peterson, B.J., Rose, C.D., Rutten, L.M. & Fourqurean, J.W. (2002). Disturbance and recovery following catastrophic grazing: studies of a successional chronosequence in a seagrass bed. *Oikos*, 97, 361–370.
- Pimm, S.L. (1984). The complexity and stability of ecosystems. *Nature*, 307, 321–326.
- Pinsky, M.L., Eikeset, A.M., McCauley, D.J., Payne, J.L. & Sunday, J.M. (2019). Greater vulnerability to warming of marine versus terrestrial ectotherms. *Nature*.
- Radchuk, V., Laender, F.D., Cabral, J.S., Boulangeat, I., Crawford, M., Bohn, F. *et al.* (2019). The dimensionality of stability depends on disturbance type. *Ecol. Lett.*, 22, 674–684.
- R Development Core Team 2018. *R: A language and environment for statistical computing (version 3.5.1)*. Vienna, Austria: R Foundation for Statistical Computing.
- R Studio Team (2018). RStudio: Integrated Development for R (Version 1.1.423). In: RStudio Inc. Boston, MA.
- Starzomski, B.M. & Srivastava, D.S. (2007). Landscape geometry determines community response to disturbance. *Oikos*, 116, 690–699.
- Sunday, J.M., Bates, A.E. & Dulvy, N.K. (2012). Thermal tolerance and the global redistribution of animals. *Nat. Clim. Chang.*, 2, 686.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.*, 36, 1–48.

- Wickham, H. (2016). *ggplot2 - Elegant Graphics for Data Analysis*, 2nd edn. Springer, Dordrecht, NL.
- Wilke, C.O. (2018). Ridgeline Plots in 'ggplot2'. In, <https://CRAN.R-project.org/package=ggridges>.
- Winfrey, R., Aguilar, R., Vazquez, D.P., LeBuhn, G. & Aizen, M.A. (2009). A meta-analysis of bees' responses to anthropogenic disturbance. *Ecology*, 90, 2068–2076.
- Yachi, S. & Loreau, M. (1999). Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. *Proc. Natl Acad. Sci. USA*, 96, 1463–1468.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Editor, Serguei Saavedra

Manuscript received 6 September 2019

First decision made 13 October 2019

Manuscript accepted 18 December 2019