



**HAL**  
open science

## Thresholds for ecological responses to global change do not emerge from empirical data

Helmut Hillebrand, Ian Donohue, W. Stanley Stanley Harpole, Dorothee Hodapp, Michal Kucera, Aleksandra M Lewandowska, Julian Merder, José Montoya, Jan A Freund

### ► To cite this version:

Helmut Hillebrand, Ian Donohue, W. Stanley Stanley Harpole, Dorothee Hodapp, Michal Kucera, et al.. Thresholds for ecological responses to global change do not emerge from empirical data. *Nature Ecology & Evolution*, 2020, 4 (11), pp.1502-1509. 10.1038/s41559-020-1256-9 . hal-03008956

**HAL Id: hal-03008956**

**<https://hal.science/hal-03008956v1>**

Submitted on 17 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Thresholds for ecological responses to global change do not emerge from**  
2 **empirical data**

3 **Authors:** Helmut Hillebrand<sup>1,2,3\*</sup>, Ian Donohue<sup>4</sup>, W. Stanley Harpole<sup>5,6,7</sup>, Dorothee  
4 Hodapp<sup>2,3</sup>, Michal Kucera<sup>8</sup>, Aleksandra M. Lewandowska<sup>9</sup>, Julian Merder<sup>10</sup>, Jose M.  
5 Montoya<sup>11</sup>, Jan A. Freund<sup>12</sup>

6 **Affiliations:**

7 <sup>1</sup>Plankton Ecology Lab, Institute for Chemistry and Biology of the Marine Environment, Carl  
8 von Ossietzky University Oldenburg, Schleusenstrasse 1, 26382 Wilhelmshaven, Germany.

9 <sup>2</sup>Helmholtz-Institute for Functional Marine Biodiversity at the University of Oldenburg  
10 [HIFMB], Ammerländer Heerstrasse 231, 26129 Oldenburg.

11 <sup>3</sup>Alfred-Wegener-Institute, Helmholtz-Centre for Polar and Marine Research, Bremerhaven,  
12 Germany.

13 <sup>4</sup>School of Natural Sciences, Department of Zoology, Trinity College Dublin, Dublin 2,  
14 Ireland.

15 <sup>5</sup>Helmholtz Center for Environmental Research – UFZ, Department of Physiological  
16 Diversity, Permoserstrasse 15, 04318 Leipzig, Germany.

17 <sup>6</sup>German Centre for Integrative Biodiversity Research (iDiv), Deutscher Platz 5e, 04103  
18 Leipzig, Germany.

19 <sup>7</sup>Martin Luther University Halle-Wittenberg, Am Kirchtor 1, 06108 Halle (Saale), Germany.

20 <sup>8</sup>MARUM – Center for Marine Environmental Sciences, University of Bremen, Leobener  
21 Strasse 8, 28359 Bremen, Germany.

22 <sup>9</sup>Tvärminne Zoological Station, University of Helsinki, J.A. Palménin tie 260, 10900 Hanko,  
23 Finland.

24 <sup>10</sup> Marine Geochemistry, Institute for Chemistry and Biology of the Marine Environment,  
25 Carl von Ossietzky University Oldenburg, Carl von Ossietzky Straße 9-11, D-26129  
26 Oldenburg, Germany.

27 <sup>11</sup>Centre for Biodiversity Theory and Modelling, Theoretical and Empirical Ecology Station,  
28 CNRS and Paul Sabatier University, Moulis, France.

29 <sup>12</sup>Theoretical Physics / Complex Systems, Institute for Chemistry and Biology of the Marine  
30 Environment, Carl von Ossietzky University Oldenburg, Carl von Ossietzky Straße 9-11, D-  
31 26129 Oldenburg, Germany.

32  
33 \*Correspondence to: [helmut.hillebrand@uni-oldenburg.de](mailto:helmut.hillebrand@uni-oldenburg.de)

34 **In order to understand ecosystem responses to anthropogenic global change, a**  
35 **prevailing framework is the definition of threshold levels of pressure, above which**  
36 **response magnitudes and their variances increase disproportionately. However, we lack**  
37 **systematic quantitative evidence as to whether empirical data allow definition of such**  
38 **thresholds. Here, we summarize 36 meta-analyses measuring more than 4600 global**  
39 **change impacts on natural communities. We find that threshold transgressions were**  
40 **rarely detectable, either within or across meta-analyses. Instead, ecological responses**  
41 **were characterized mostly by progressively increasing magnitude and variance when**  
42 **pressure increased. Sensitivity analyses with modelled data revealed that minor**  
43 **variances in the response are sufficient to preclude the detection of thresholds from**  
44 **data, even if they are present. The simulations reinforced our contention that global**  
45 **change biology needs to abandon the general expectation that system properties allow**  
46 **defining thresholds as a way to manage nature under global change. Rather, highly**  
47 **variable responses, even under weak pressures, suggest that ‘safe-operating spaces’ are**  
48 **unlikely to be quantifiable.**

49 Concepts of thresholds, tipping points and regime shifts dominate current ecological  
50 frameworks aiming to understand ecosystem responses to anthropogenic global change<sup>1-4</sup>. A  
51 threshold corresponds to a level of environmental pressure that creates a discontinuity in the  
52 ecosystem response to this pressure. Thresholds and tipping points pervade environmental  
53 policy documents<sup>5,6</sup> as they allow definition of levels of pressure below which ecosystem  
54 responses remain within “safe ecological limits”<sup>6</sup>, and above which response magnitudes and  
55 their variances increase disproportionately<sup>7,8</sup>. Anticipating when and under what conditions  
56 such threshold transgression might occur is important for sustainable environmental  
57 management.

58 Threshold-related concepts and their implementation in policy hinge upon the  
59 assumption that the presence of thresholds can be detected in data or – even better – predicted.

60 Testing this assumption requires knowledge of the ecosystem response to an environmental  
61 pressure for present-day and potential future pressure magnitudes. Ecological meta-analysis  
62 has led to the publication of thousands of effect sizes in response to *in-situ* trends or  
63 experimental manipulations of key pressures of global change such as eutrophication,  
64 warming, land-use change, fisheries, and ocean acidification. Each study in a meta-analysis  
65 quantifies the magnitude of the response of an ecosystem variable to the strength of an  
66 applied environmental pressure (Fig. 1a). The entire set of studies in the meta-analysis then  
67 represents a wide range of pressure strengths, which often exceed the conditions observed in  
68 nature, but might be expected in future ecosystems. We capitalize on this richness of data by  
69 combining available information from 36 meta-analyses, providing 4601 effect sizes across  
70 ecosystems and pressures of global change into multiple tests of whether these data sets –  
71 individually or aggregated – reveal a response pattern that indicates transgression of a  
72 threshold (Fig. 1b). We first tested whether and how ecosystems respond to increased  
73 environmental pressures by simply exploring whether ecosystems show a directional change  
74 in response to a pressure, regardless of the presence of a threshold (Fig. 1c). Second, we  
75 quantified discontinuities in the variance of responses, which would be a way to define the  
76 existence of a threshold. Finally, we tested for existence of multimodality of responses, which  
77 would be stronger evidence for alternative states under different environmental pressures.

78

## 79 **Results**

80 To test for general changes of systems along gradients of environmental pressures, we  
81 used an averaged Kullback-Leibler (KL) divergence method (see *Methods*) to quantify the  
82 overall deviation between the response distribution for a given stressor value and the marginal  
83 response distribution, that is, the response distribution when collapsing all response data onto  
84 a single axis ignoring the magnitude of the stressor variable. Most meta-analyses (23 of 36)  
85 showed changes in the response magnitude along the gradient of pressure strengths (KL,

86 Table S1). This provides strong evidence that direction and increasing magnitude of global  
87 environmental pressures have significant effects on ecosystem variables. While necessary,  
88 this evidence is not sufficient to support the general prevalence of threshold-type responses  
89 across ecosystems.

90 If thresholds are common, then we expect to see increased variance in response  
91 variables as the pressure strength crosses the threshold value<sup>7,8</sup> (as sketched in Fig. 1c). To  
92 test for discontinuities in the variance of effect size responses, we used a weighted quantile  
93 ratio (QR) of interquantile range (95%-5%) to quantify substantial inhomogeneity in the  
94 width of the response distribution across the range of observed stressors (see *Methods*).  
95 Significant changes in the variance of effect sizes were present in only 8 out of 36 cases (QR,  
96 Table S1), challenging the widespread expectation of rising variance as a signal of threshold  
97 transgression. Moreover, in those cases with a significant QR test, the increase in variance  
98 occurred frequently only at the most extreme pressure level observed in the respective meta-  
99 analysis (see below for further details).

100 Stronger evidence for threshold-type ecosystem responses to increasing environmental  
101 pressure would be provided by the existence of multimodal distributional patterns, reflecting a  
102 state transition. We used Hartigan's dip test method (HD; see *Methods*) to assess the  
103 multimodality of effect sizes<sup>9</sup>, which provides a narrow test for the case of bi-(multi)-stability  
104 of responses. We found no support for widespread existence of alternative states in ecological  
105 responses to increasing pressure intensities. None of the 36 meta-analyses revealed any sign  
106 of bimodality in the frequency distribution of effect sizes (HD,  $p > 0.3$  in every case, Table  
107 S1).

108 Comparing these empirical results (Table S1) to model data (Fig. 2, Extended Data  
109 S1) with known presence or absence of thresholds shows that our three approaches are  
110 suitable to detect threshold transgression. For idealized data, the three tests provide a clear  
111 differentiation between gradual and threshold-associated disproportional changes in response

112 magnitudes. However, empirical observations will be affected by different sources of  
113 variance, both systematic (cases with different locations of thresholds and magnitudes of  
114 response shift) and stochastic. With increasing noise to signal ratios, thresholds – although  
115 present – quickly become undetectable, as the power of QR and HD declines rapidly. The  
116 exponential decline in detection probability for QR shows that thresholds can only be  
117 identified reliably for nearly ideal data without random variation around the response  
118 magnitude (scenarios g-i in Fig. 2), with the exception of the unlikely case that all systems are  
119 characterized by the same threshold (scenario f in Fig. 2). For HD, the power collapses  
120 completely with only moderate noise levels (Fig. 2). Only KL is still able to detect changes in  
121 response magnitude with increasing pressure with increasing variance, either around gradual  
122 shifts in response magnitude (scenarios c-d in Fig. 2) or around thresholds (scenarios e-i). The  
123 simulations corroborate our general empirical finding across the 36 datasets that thresholds  
124 are rarely detectable in data even if using statistical methods developed for threshold  
125 detection.

126         Even when thresholds were empirically detected, limited inference can be made as  
127 shown by highlighting several individual meta-analysis datasets to illustrate specific  
128 ecosystem responses to particular environmental pressures. The first meta-analysis in our data  
129 set (MA1.1) exemplifies the general results. The overall response of biomass production to  
130 biodiversity loss tended to be negative, and became more negative for larger proportions of  
131 species lost without changes in the variational range of effect sizes (Fig. 3). This gradual  
132 response type was also found in the analysis of fertilization effects on biomass production  
133 (MA2.1), and in soil responses to changes in precipitation (MA8) and land-use change (MA9)  
134 as well as prey responses to predator loss (MA 16.1). Ten additional examples of this type of  
135 response involving other drivers of environmental change are provided in the supporting  
136 material (Extended Data S2, Table S1). In all of these cases, the magnitude of the  
137 environmental change altered the magnitude of the response – as expected – but the variance

138 around this relationship did not indicate the emergence of a “novel” ecosystem response  
139 beyond a pressure threshold. Eight cases showed significant QR tests, of which three showed  
140 an increase in response variance only at highest pressure strength and two a reduction in  
141 response variance with increasing pressure. Thus, only three out of 36 cases showed a shifting  
142 distribution of effect sizes with increasing pressure that was consistent with the emergence of  
143 new types of responses above a threshold. These comprise land-use change effects on  
144 mammal abundance (MA6.5), warming effects on corals (MA10), and fertilization effects on  
145 microbial respiration (MA17.2, all Extended Data S2). By contrast, in 12 of the 36 meta-  
146 analyses, neither KL nor QR were significant (exemplified by MA23.1 in Fig. 3, for others  
147 see Extended Data S2), indicating that no increases in response magnitudes or threshold  
148 transgressions were observed.

149         The above results are relevant for across-system analyses of single pressure gradients,  
150 but in many cases management might not have *a priori* knowledge of which pressure gradient  
151 leads to transgressions. In order to analyze this situation, we further aggregated our analysis  
152 across drivers, organism groups and ecosystems, by standardizing and normalizing the  
153 pressure gradient to a median of 0 and a range of -1 to 1 (Fig. 4). The range of responses was  
154 impressive, the effect sizes in cases indicated more than 200-fold increase or decrease in the  
155 measured ecosystem variable (Fig. 4a). Both KL and QR tests were highly significant for the  
156 aggregated data, indicating a strong impact of pressure intensity on the strength and variance  
157 of the ecological response (Table S1). However, this increase in the variance of effect sizes  
158 was found for studies with normalized pressures greater than 0.5, which comprised the top  
159 3.5% of the manipulated range of potential impacts (Fig. 4b). This observation resembles a  
160 “sledgehammer effect”, that is, system transformation by huge impact, which is a trivial  
161 consequence of the large pressure magnitude and the complete transformation of the system.

162         As the sign of the effect size depends upon the specific association of driver and effect  
163 in each meta-analysis, we also analyzed the absolute magnitude of response (|LRR|)

164 independent of sign for the aggregated data set (Fig. 4c). We found that the median |LRR|  
165 increased with increasing environmental pressure, as did the variance, particularly so at the  
166 highest pressure magnitudes (significant KL and QR tests, Table S1). The median |LRR|  
167 corresponded to 1.5-2-fold increases or decreases in process rates or properties, whereas the  
168 range of responses (i.e., the 5-95% quantiles of |LRR|) exceeded 5-fold changes even at the  
169 smallest pressure strengths. Thus, even at very small pressures, very large responses can  
170 occur.

## 171 **Discussion**

172 Analysis of the 4601 experiments that we assembled here, potentially the most  
173 comprehensive data available, did not enable us to estimate where thresholds might have been  
174 crossed. Instead, the data suggest that the ecosystem impacts of human-induced changes in  
175 environmental drivers are better characterized by gradual shifts in response magnitudes with  
176 increasing pressure coupled with broad variations around this trend. While our analyses do  
177 not rule out the existence of tipping points, they bring into question the utility of threshold-  
178 based concepts in management and policy if we cannot detect thresholds in nature<sup>10,11</sup>.  
179 Expectation of threshold responses ultimately leads to an underestimation of the large  
180 consequences of small environmental pressures<sup>12</sup>. Moreover, it marginalizes the importance  
181 of other, more complex non-linear dynamics under global change, which may underlie the  
182 considerable variance around gradually increasing response magnitudes.

183 Our use of field and semi-natural experiments has the advantage that these often  
184 involve pressures that are larger than observed environmental conditions, as they commonly  
185 incorporate future scenarios of severe environmental change<sup>13</sup>. This counters the argument  
186 that thresholds exist but have not yet been reached. Still, some caveats to our approach need  
187 to be acknowledged. First, the absence of evidence is obviously not the evidence of absence:  
188 as shown by our explicit analysis of test power, the existence of thresholds can be masked by  
189 high inter-study variance (especially for HD). However, this also questions the usefulness of



190 thresholds if their occurrence is dependent on the complex interaction of multiple pressures  
191 and their detection is only possible under very high signal-to-noise ratios. Without *a priori*  
192 knowledge across specific systems of when thresholds might appear, any definition of  
193 thresholds – even if precautionary principles are used – must remain arbitrary. Second, we  
194 focused on functional, not compositional aspects of ecosystems, and do not make conclusions  
195 about threshold pressures for changes in composition. However, compositional and  
196 functional stability often show interdependencies<sup>14</sup> because compensatory dynamics between  
197 species may dampen the response in ecosystem functions<sup>15</sup> or allow for rapid recovery from a  
198 phase shift<sup>16-18</sup>. Given that the functions addressed here often are aggregate properties of the  
199 communities investigated, we thus consider it unlikely that thresholds are more prevalent for  
200 compositional responses. Third, the temporal extent of the experimental studies in our data  
201 base is limited; it rarely exceeds the scale of tens of generations of organisms. However, there  
202 is no strong support to why threshold transgressions should increase through time. Threshold-  
203 related concepts thus would be untestable in ecology, as their absence could always be  
204 ascribed to insufficiently long observation periods.

205         The lack of clearly-defined and generally applicable thresholds distinguishing between  
206 tolerable and non-tolerable responses has obvious implications for environmental policies.  
207 The use of thresholds has been critically discussed in ecosystem management, conservation  
208 and restoration<sup>19-21</sup> to establish precautionary principles for environmental policy. Using such  
209 threshold arguments in a world where changes are too case-specific and variable to allow  
210 prediction of tipping points undermines this precautionary argument. It leads to the  
211 anticipation of major system transformation as thresholds are passed, whereas the majority of  
212 observed responses to environmental change represent progressively shifting baselines on  
213 time-scales of human perceptions<sup>22,23</sup>. Consequently, environmental concerns might appear  
214 overstated if thresholds are taken for the general case but critical transitions associated with  
215 transgressing thresholds are not observed<sup>24,25</sup>. The frequently major and highly variable

216 responses we observed even at low pressure magnitudes indicate that safe-operating spaces  
217 are unlikely to be definable from data. The data resonate well with the fact that conceptually  
218 thresholds occur under special and limiting conditions. Our results thus question the pervasive  
219 presence of threshold concepts in management and policy.

- 221 1 Scheffer, M., Carpenter, S., Foley, J. A., Folke, C. & Walker, B. Catastrophic shifts in ecosystems.  
222 *Nature* **413**, 591-596 (2001).
- 223 2 Scheffer, M. *Critical Transitions in Nature and Society*. (Princeton University Press 2009).
- 224 3 Rockström, J. *et al.* A safe operating space for humanity. *Nature* **461**, 472-475, doi:10.1038/461472a  
225 (2009).
- 226 4 Folke, C. *et al.* Regime shifts, resilience and biodiversity in ecosystem management. *Annual Review of*  
227 *Ecology, Evolution, and Systematics* **35**, 557-581 (2004).
- 228 5 Donohue, I. *et al.* Navigating the complexity of ecological stability. *Ecology Letters* **19**, 1172-1185,  
229 doi:10.1111/ele.12648 (2016).
- 230 6 UN. United Nations. Aichi biodiversity targets (<https://www.cbd.int/sp/targets/>). Report No.  
231 <https://www.cbd.int/sp/targets/>, (2010).
- 232 7 Carpenter, S. R. & Brock, W. A. Rising variance: a leading indicator of ecological transition. *Ecology*  
233 *Letters* **9**, 308-315 (2006).
- 234 8 Scheffer, M. *et al.* Early-warning signals for critical transitions. *Nature* **461**, 53-59 (2009).
- 235 9 Hartigan, J. A. & Hartigan, P. M. The Dip Test of Unimodality. *Ann. Statist.* **13**, 70-84,  
236 doi:10.1214/aos/1176346577 (1985).
- 237 10 Montoya, J. M., Donohue, I. & Pimm, S. L. Planetary Boundaries for Biodiversity: Implausible  
238 Science, Pernicious Policies. *Trends Ecol. Evol.* **33**, 71-73, doi:10.1016/j.tree.2017.10.004 (2018).
- 239 11 Pimm, S. L., Donohue, I., Montoya, J. M. & Loreau, M. Measuring resilience is essential to understand  
240 it. *Nature Sustainability* **2**, 895-897, doi:10.1038/s41893-019-0399-7 (2019).
- 241 12 Clark, C. M. & Tilman, D. Loss of plant species after chronic low-level nitrogen deposition to prairie  
242 grasslands. *Nature* **451**, 712-715 (2008).
- 243 13 Korell, L., Auge, H., Chase, J. M., Harpole, W. S. & Knight, T. M. We need more realistic climate  
244 change experiments for understanding ecosystems of the future. *Global Change Biology* **0**,  
245 doi:10.1111/gcb.14797 (2019).
- 246 14 Hillebrand, H. *et al.* Decomposing multiple dimensions of stability in global change experiments.  
247 *Ecology Letters* **21**, 21-30, doi:10.1111/ele.12867 (2018).
- 248 15 Connell, S. D. & Ghedini, G. Resisting regime-shifts: the stabilising effect of compensatory processes.  
249 *Trends Ecol. Evol.* **30**, 513-515, doi:10.1016/j.tree.2015.06.014 (2015).
- 250 16 Bruno, J. F., Sweatman, H., Precht, W. F., Selig, E. R. & Schutte, V. G. W. Assessing evidence of  
251 phase shifts from coral to macroalgal dominance on coral reefs. *Ecology* **90**, 1478-1484 (2009).
- 252 17 Diaz-Pulido, G. *et al.* Doom and Boom on a Resilient Reef: Climate Change, Algal Overgrowth and  
253 Coral Recovery. *PLOS ONE* **4**, e5239 (2009).
- 254 18 Carpenter, S. R. *et al.* Early Warnings of Regime Shifts: A Whole-Ecosystem Experiment. *Science* **332**,  
255 1079-1082, doi:10.1126/science.1203672 (2011).
- 256 19 Suding, K. N. & Hobbs, R. J. Threshold models in restoration and conservation: a developing  
257 framework. *Trends Ecol. Evol.* **24**, 271-279, doi:<https://doi.org/10.1016/j.tree.2008.11.012> (2009).
- 258 20 Vaquer-Sunyer, R. & Duarte, C. M. Thresholds of hypoxia for marine biodiversity. *Proceedings of the*  
259 *National Academy of Sciences* **105**, 15452-15457, doi:10.1073/pnas.0803833105 (2008).
- 260 21 Groffman, P. M. *et al.* Ecological Thresholds: The Key to Successful Environmental Management or an  
261 Important Concept with No Practical Application? *Ecosystems* **9**, 1-13, doi:10.1007/s10021-003-0142-z  
262 (2006).
- 263 22 Hughes, T. P., Carpenter, S., Rockstrom, J., Scheffer, M. & Walker, B. Multiscale regime shifts and  
264 planetary boundaries. *Trends Ecol. Evol.* **28**, 389-395, doi:10.1016/j.tree.2013.05.019 (2013).
- 265 23 Papworth, S. K., Rist, J., Coad, L. & Milner-Gulland, E. J. Evidence for shifting baseline syndrome in  
266 conservation. *Conservation Letters* **2**, 93-100, doi:10.1111/j.1755-263X.2009.00049.x (2009).
- 267 24 Schlesinger, W. H. Thresholds risk prolonged degradation. *Nature Reports Climate Change* **3**,  
268 doi:10.1038/climate.2009.1093 (2009).
- 269 25 Duarte, C. M. *et al.* Reconsidering Ocean Calamities. *BioScience* **65**, 130-139,  
270 doi:10.1093/biosci/biu198 (2015).
- 271 26 Sheather, S. J. & Jones, M. C. A Reliable Data-Based Bandwidth Selection Method for Kernel Density  
272 Estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* **53**, 683-690 (1991).
- 273 27 Chase, J. M. & Knight, T. M. Scale-dependent effect sizes of ecological drivers on biodiversity: why  
274 standardised sampling is not enough. *Ecology Letters* **16**, 17-26, doi:10.1111/ele.12112 (2013).
- 275 28 Cardinale, B. J. *et al.* Effects of biodiversity on the functioning of trophic groups and ecosystems.  
276 *Nature* **443**, 989-992 (2006).
- 277 29 Gruner, D. S. *et al.* A cross-system synthesis of consumer and nutrient resource control on producer  
278 biomass. *Ecology Letters* **11**, 740-755, doi:10.1111/j.1461-0248.2008.01192.x (2008).

279 30 Elser, J. J. *et al.* Global analysis of nitrogen and phosphorus limitation of primary producers in  
280 freshwater, marine and terrestrial ecosystems. *Ecology Letters* **10**, 1135-1142 (2007).

281 31 Lin, D., Xia, J. & Wan, S. Climate warming and biomass accumulation of terrestrial plants: a meta-  
282 analysis. *New Phytologist* **188**, 187-198, doi:10.1111/j.1469-8137.2010.03347.x (2010).

283 32 Treseder, K. K. Nitrogen additions and microbial biomass: a meta-analysis of ecosystem studies.  
284 *Ecology Letters* **11**, 1111-1120, doi:10.1111/j.1461-0248.2008.01230.x (2008).

285 33 Akiyama, H., Yan, X. & Yagi, K. Evaluation of effectiveness of enhanced-efficiency fertilizers as  
286 mitigation options for N<sub>2</sub>O and NO emissions from agricultural soils: meta-analysis. *Global Change*  
287 *Biology* **16**, 1837-1846, doi:10.1111/j.1365-2486.2009.02031.x (2010).

288 34 Gibson, L. *et al.* Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature* **478**, 378-  
289 381, doi:10.1038/nature10425 (2011).

290 35 Liang, J. Y., Qi, X., Souza, L. & Luo, Y. Q. Processes regulating progressive nitrogen limitation under  
291 elevated carbon dioxide: a meta-analysis. *Biogeosciences* **13**, 2689-2699, doi:10.5194/bg-13-2689-2016  
292 (2016).

293 36 Liu, L. L. *et al.* A cross-biome synthesis of soil respiration and its determinants under simulated  
294 precipitation changes. *Global Change Biology* **22**, 1394-1405, doi:10.1111/gcb.13156 (2016).

295 37 van Lent, J., Hergoualc'h, K. & Verchot, L. V. Reviews and syntheses: Soil N<sub>2</sub>O and NO emissions  
296 from land use and land-use change in the tropics and subtropics: a meta-analysis. *Biogeosciences* **12**,  
297 7299-7313, doi:10.5194/bg-12-7299-2015 (2015).

298 38 Ateweberhan, M. & McClanahan, T. R. Relationship between historical sea-surface temperature  
299 variability and climate change-induced coral mortality in the western Indian Ocean. *Marine Pollution*  
300 *Bulletin* **60**, 964-970, doi:10.1016/j.marpolbul.2010.03.033 (2010).

301 39 Gärtner, M. *et al.* Invasive plants as drivers of regime shifts: identifying high-priority invaders that alter  
302 feedback relationships. *Diversity and Distributions* **20**, 733-744, doi:10.1111/ddi.12182 (2014).

303 40 Dooley, S. R. & Treseder, K. K. The effect of fire on microbial biomass: a meta-analysis of field  
304 studies. *Biogeochemistry* **109**, 49-61, doi:10.1007/s10533-011-9633-8 (2012).

305 41 Dijkstra, F. A. & Adams, M. A. Fire Eases Imbalances of Nitrogen and Phosphorus in Woody Plants.  
306 *Ecosystems* **18**, 769-779, doi:10.1007/s10021-015-9861-1 (2015).

307 42 Lu, M. *et al.* Responses of ecosystem carbon cycle to experimental warming: a meta-analysis. *Ecology*  
308 **94**, 726-738 (2013).

309 43 Griffin, J. N., Byrnes, J. E. K. & Cardinale, B. J. Effects of predator richness on prey suppression: a  
310 meta-analysis. *Ecology* **94**, 2180-2187, doi:10.1890/13-0179.1 (2013).

311 44 Srivastava, D. S. *et al.* Diversity has stronger top-down than bottom-up effects on decomposition.  
312 *Ecology* **90**, 1073-1083 (2009).

313 45 Östman, Ö. *et al.* Top-down control as important as nutrient enrichment for eutrophication effects in  
314 North Atlantic coastal ecosystems. *J. Appl. Ecol.* **53**, 1138-1147, doi:10.1111/1365-2664.12654 (2016).

315 46 Katano, I., Doi, H., Eriksson, B. K. & Hillebrand, H. A cross-system meta-analysis reveals coupled  
316 predation effects on prey biomass and diversity. *Oikos* **124**, 1427-1435, doi:10.1111/oik.02430 (2015).

317 47 Borer, E. T. *et al.* What determines the strength of a trophic cascade? *Ecology* **86**, 528-537,  
318 doi:10.1890/03-0816 (2005).

319 48 Hodapp, D. & Hillebrand, H. Effect of consumer loss on resource removal depends on species-specific  
320 traits. *Ecosphere* **8**, e01742-n/a, doi:10.1002/ecs2.1742 (2017).

321 49 Liu, L. L. & Greaver, T. L. A global perspective on belowground carbon dynamics under nitrogen  
322 enrichment. *Ecology Letters* **13**, 819-828, doi:10.1111/j.1461-0248.2010.01482.x (2010).

323 50 Martinson, H. M. & Fagan, W. F. Trophic disruption: a meta-analysis of how habitat fragmentation  
324 affects resource consumption in terrestrial arthropod systems. *Ecology Letters* **17**, 1178-1189,  
325 doi:10.1111/ele.12305 (2014).

326 51 Holden, S. & Treseder, K. A meta-analysis of soil microbial biomass responses to forest disturbances.  
327 *Frontiers in Microbiology* **4**, doi:10.3389/fmicb.2013.00163 (2013).

328 52 Nagelkerken, I. & Connell, S. D. Global alteration of ocean ecosystem functioning due to increasing  
329 human CO<sub>2</sub> emissions. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13272-13277,  
330 doi:10.1073/pnas.1510856112 (2015).

331 53 Kaiser, M. J. *et al.* Global analysis of response and recovery of benthic biota to fishing. *Mar. Ecol.*  
332 *Prog. Ser.* **311**, 1-14 (2006).

333 54 Gill, D. A. *et al.* Capacity shortfalls hinder the performance of marine protected areas globally. *Nature*  
334 **534**, 665-669, doi:10.1038/nature21708 (2017).

335 55 Gallardo, B., Clavero, M., Sánchez, M. I. & Vilà, M. Global ecological impacts of invasive species in  
336 aquatic ecosystems. *Global Change Biology* **22**, 151-163, doi:10.1111/gcb.13004 (2016).

337 56 Vila, M. *et al.* Ecological impacts of invasive alien plants: a meta-analysis of their effects on species,  
338 communities and ecosystems. *Ecology Letters* **14**, 702-708, doi:10.1111/j.1461-0248.2011.01628.x  
339 (2011).

340 57 Scheffer, M. & Carpenter, S. R. Catastrophic regime shifts in ecosystems: linking theory to observation.  
341 *Trends Ecol. Evol.* **18**, 648-656, doi:<http://dx.doi.org/10.1016/j.tree.2003.09.002> (2003).  
342 58 Andersen, T., Carstensen, J., Hernandez-Garcia, E. & Duarte, C. M. Ecological thresholds and regime  
343 shifts: approaches to identification. *Trends Ecol. Evol.* **24**, 49-57 (2009).  
344 59 Fasiolo, M., Goude, Y., Nedellec, R. & N. Wood, S. Fast calibrated additive quantile regression.  
345 *Journal of the American Statistical Association*, doi:<https://doi.org/10.1080/01621459.2020.1725521>  
346 (2020).

347

## 348 **Acknowledgments**

349 The data reported in this paper are presented and derived from 36 different meta-analyses.  
350 They are archived and available from each of these as indicated in the Supplementary Text.  
351 The concept of this paper has emerged during scientific discussions with Thorsten Blenckner  
352 at Stockholm University, at an UK NERC/BESS Tansley Working Group on ecological  
353 stability and the TippingPond EU Biodiversa project. The actual work has been funded by the  
354 Lower Saxony Ministry of Science and Culture through the MARBAS project to HH and the  
355 HIFMB, a collaboration between the Alfred-Wegener-Institute, Helmholtz-Center for Polar  
356 and Marine Research, and the Carl-von-Ossietzky University Oldenburg, initially funded by  
357 the Ministry for Science and Culture of Lower Saxony and the Volkswagen Foundation  
358 through the “Niedersächsisches Vorab” grant program (grant number ZN3285). The work was  
359 finalized with support by Deutsche Forschungsgemeinschaft, grant no HI848/26-1. Liv  
360 Toaspern helped gathering data from invasion meta-analyses. Peter Ruckdeschel helped with  
361 the statistical approach. Montserrat Vila provided additional information on their published  
362 meta-analyses. We acknowledge the comments by Ulrike Feudel, Gabriele Gerlach, and the  
363 members of the Plankton Ecology Lab on the manuscript that helped detailing our  
364 argumentation.

## 365 **Author contribution**

366 HH designed the analysis and discussed the framework with ID and JMM, JF and JM  
367 developed the statistical approach with input by HH and DH. HH assembled the effect size  
368 information, JF and JM performed the statistical analysis. MK, AL & WSH provided input on  
369 paleoecological and experimental constraints, respectively. HH wrote the manuscript together  
370 with WSH and JMM as well as input from all co-authors.

371 The authors declare no competing interests.

372

373 **Figure Legends**

374

375 **Fig. 1. Detecting thresholds in response to environmental change.** (a) Classically, the  
376 approach to detecting thresholds is to address the discontinuity of responses to an  
377 environmental driver over time. Instead of a temporal axis, our analyses use the multitude of  
378 experiments or observations testing the same driver in independent studies. Each meta-  
379 analysis summarizes the results of multiple experiments characterized by different magnitudes  
380 of the same pressure and response magnitudes  $\pm$  sampling variance. The basis of each meta-  
381 analysis is represented by single experiments (or observational studies) measuring the  
382 response in a variable of interest in control and disturbed environments (insert). The distance  
383 in the environmental variable (e.g., temperature in warming experiments) between control and  
384 treatment gives the intensity of the pressure, the log response ratios (LRR) measure the  
385 relative change in the response variable (e.g., plant biomass) based on treatment and control  
386 means, whereas the pooled standard deviations result in an estimate of sampling variance per  
387 study (varLRR). (b) If the response shows discontinuity, we expect a tendency towards a new  
388 category of responses (red cases reflecting critical transitions) at higher pressure strengths. (c)  
389 We developed two robust non-parametric test statistics and assessed their statistical  
390 significance using permutation tests: Kullback-Leibler (KL) divergence to test for general  
391 changes in the response magnitude along the pressure gradient and the weighted quantile ratio  
392 (QR) of interquantile (5%-95%) ranges to test for changes in the variability of effect sizes.  
393 We tested for multimodal frequency distribution of effect sizes, reflecting alternative  
394 responses to a common driver using Hartigan's dip test (HD). To visualize the KL approach,  
395 we indicated a potential realized distribution of responses by a red area, compared to a  
396 randomized distribution (blue area, see Methods). The significant deviation between realized  
397 and randomized responses can occur if there is gradual increase in response with increasing  
398 pressure (orange line) or if shifts in the response (red solid line) occur at a threshold (vertical  
399 dashed line).

400

401 **Fig. 2: Detection probability for thresholds in global change experiments using kernel**  
402 **density estimation.** We analyzed the test power for 9 scenarios of responses to pressure in  
403 meta-analyses, the derivation of each scenario is described in the supplementary online  
404 material, Extended Data S3. Scenarios **a-d** do not comprise a threshold, where scenario **a**  
405 is the null model without an effect of the pressure on the response. Scenarios **e-i** do comprise a  
406 threshold, for the latter two combined with intermediate responses. For the three statistical  
407 test used in our analyses, the expected outcome is colour-coded, with green representing that  
408 the test should be significant. We then tested the proportion of 1000 simulated data sets for  
409 which the tests were significant with a probability  $p = 0.05$  (black) and  $p = 0.01$  (blue). We  
410 did for increasing noise variance (= inverse signal-to-noise ratio). Bandwidth selection was  
411 based on the "solve- the-equation" method of Sheater & Jones<sup>26</sup>. The estimated bandwidth was  
412 adjusted by a factor of 2.5 in each case because this optimized test power for all cases. The  
413 three tests together allow perfect detection of thresholds in the absence of noise (scenarios **e-**  
414 **h**), only if threshold-type and gradual responses are mixed (scenario **i**), the analysis of  
415 multimodality (HD) is no longer able to pick up the threshold embedded in the data, as the  
416 simultaneous increase in mean and variance of the response (as in scenario **d**) masks modes in

417 the response distribution. With increasing noise variance, however, the detection probability  
418 for thresholds via HD and QR rapidly decreases.

419

420 **Fig. 3: Example meta-analyses testing for changes in the response magnitude along with**  
421 **increasing pressure intensity.** Red and blue shaded regions indicate the (5%-95%)  
422 interquantile ranges for, respectively, the bivariate data (including the pressure gradient) and  
423 the marginal distribution of LRR (integrating out the pressure gradient). Dashed red and blue  
424 thick lines trace the related median (50% quantile). Overlain are the data points and, at the  
425 bottom, the yellow shaded area indicates the distribution of stressor variables resulting from a  
426 weighted kernel density estimation (Extended Data S2). Color codes for habitat (dark blue:  
427 freshwater; aquamarine: marine; green: terrestrial), circle size reflects statistical weight.  
428 Please note that the suggestive break in the responses in MA8 is induced by a lack of data  
429 covering intermediate pressure magnitude.

430

431 **Fig. 4. Analysis of aggregate data across meta-analyses. (a)** Log response ratios (LRR) of  
432 ecological processes across a gradient of environmental change, where the different pressures  
433 were normalized to a median of 0 and a range of -1 to 1. Color codes for habitat (dark blue:  
434 freshwater; aquamarine: marine; green: terrestrial), circle size reflects statistical weight.  
435 Shaded regions indicate the interquantile (5%-95%) ranges for the marginal distribution  
436 (blue) and the bivariate distribution (red). Density of values along the stressor and the  
437 response axis are given below (yellow) or at the right margin (green), respectively. **(b)** Same  
438 as (a), but without single effect sizes, focusing on the distribution of response magnitudes  
439 over the normalized pressure gradient. **(c)** Same as (b), but for absolute response magnitudes.  
440 Note the change in scale of the Y-axis in the three panels.

441

442 **Methods**

443 Data

444 We searched the ISI® Web of Science (WoS) using a search string targeted towards  
445 detecting meta-analyses in a global change context (Topic: ["metaanalysis" or "meta-analysis"  
446 or "metaanalyses" or "meta-analyses"] AND Topic: ["global change" or "fertili\*" or "land-  
447 use" or "acidification" or "warming" or "temperature" or "eutrophication" or "disturbance" or  
448 "invasion" or "extinction" or "drought" or "ultraviolet"] AND Topic: [chang\* or  
449 manipulation\* or experim\* or treatm\*]). We refined the results by focusing on the WoS  
450 research area "Environmental sciences and ecology". This search (done September 11, 2016)  
451 yielded 979 studies, from which the majority did not fit all of our inclusion criteria (upon  
452 request, we provide a list of all studies with the study-specific criteria to include or exclude),  
453 which were:

- 454 • The paper provided a formal meta-analysis with effect sizes, which quantified the  
455 responses to a factor that represented a global change impact. The factor was either an  
456 experimental treatment or an in-situ change. This excluded numerous studies that either  
457 were verbal/vote-counting reviews or provided effect sizes as a response to non-global-  
458 change factors (e.g. mitigation efforts).
- 459 • The response was measured at the level of ecological communities or ecosystems. This  
460 excluded studies where responses were measured at the level of single species, as these  
461 were deemed inappropriate to detect regime shifts, or at the level of human societies (e.g.,  
462 health aspects, economy). We also excluded fossil data as not being affected by  
463 anthropogenic global change and non-biological response variables (e.g., the effect of CO<sub>2</sub>-  
464 enrichment on water pH).
- 465 • Given that effect sizes on species richness have recently been criticized strongly for being  
466 statistically biased<sup>27</sup>, we decided not to use biodiversity response variables but only  
467 functional processes or properties at the community or ecosystem level (details see below).



468 As we explicitly address the statistical distribution of effect sizes (see below), this  
469 statistical bias was considered to be potentially misleading in the context of our analysis.  
470 However, we used cases where biodiversity loss was the manipulated component of global  
471 change and a functional response was measured.

472 From the remaining 162 meta-analyses that fulfilled these criteria, we extracted the  
473 information needed to perform our analyses. This included a measure of the magnitude of the  
474 stressor (impact, driver) and the effect size as well as its sampling variance or weight  
475 (response). When the information was not given in an online appendix or associated data  
476 table, we contacted the authors to ask for data access. Still, we had to exclude further meta-  
477 analyses, as they

- 478 • did not quantify the stressor magnitude. This was especially common in meta-analyses  
479 addressing the response to invasive species
- 480 • did not contain enough cases to perform analysis. We set the critical number of effect sizes  
481 to 35 as a minimum to detect variance shifts
- 482 • overlapped with other meta-analyses on the same subject. This was especially found for  
483 analyses on eutrophication and biodiversity loss, where we always opted for the most  
484 consistent and information-dense alternative.
- 485 • did not provide available data.

486 The final database contained 24 meta-analyses (information derived from 29 papers<sup>28-</sup>  
487 <sup>56</sup>), which were divided into a total of 36 cases (Table S1). Subsetting multiple cases from a  
488 meta-analysis was done if different drivers were tested or different response categories were  
489 used in a single meta-analysis. We followed the authors in defining response categories and  
490 stressor variables. We excluded laboratory experiments and focused our study solely on field  
491 experiments and observational studies. The resulting dataset reflects ecological responses in  
492 the form of ecosystem processes (primary or secondary production, feeding rates, element  
493 fluxes) to the most pervasive anthropogenic alterations of our planet (Table S1).

494

495 *Statistical approach*

496 For each meta-analysis data set containing a measure of the stressor magnitude (X),  
497 the response variable (log response ratio, LRR) and its sampling variance (var.LRR), we  
498 assessed whether the data set reflects any statistically significant influence of the stressor  
499 variable on the response. As the data basis of each meta-analysis, and thus the sources of  
500 variation of LRR within each data set, is unknown to us, we devised three robust non-  
501 parametric test statistics and assessed their statistical significance by permutation tests.

502 An averaged Kullback-Leibler (KL) divergence quantified the overall deviation  
503 between the response distribution for a given stressor value and the marginal response  
504 distribution, that is, the response distribution when collapsing all response data onto a single  
505 axis ignoring the stressor variable. Second, a weighted quantile ratio (QR) of interquantile  
506 range (95%-5%) was then used to quantify substantial variability of the response distribution  
507 width across the range of observed stressors. Finally, we used Hartigan's dip (HD) test to  
508 assess the multimodality of effect sizes<sup>9</sup>. Based on simulation-based p-values, HD provides a  
509 narrow test for the case of bi-(multi)-stability of responses, analogous to the bimodality test  
510 proposed by Scheffer & Carpenter<sup>57</sup>. A significant HD indicates that the responses along the  
511 pressure gradient fall into two (or more) clearly separated categories, which indicates the  
512 presence of two (or more) alternative ecosystem states. Essentially, strict bimodality across a  
513 wide range of studies is a rather narrow expectation, but we include this test as the bifurcation  
514 case is the one most often discussed in considerations of thresholds, tipping points and regime  
515 shifts<sup>57,58</sup>.

516 For both KL and QR, the assessment of statistical significance was done by a  
517 permutation test: the null-hypothesis (NH) that the response distribution is unrelated to the  
518 stressor is simulated by breaking up paired variables (X, LRR, var.LRR) and recombining  
519 them in the form (X', LRR, var.LRR), where X' is a permutation of recorded stressor values.

520 If the NH were valid, this permutation should induce no substantial difference. Computing the  
521 two test statistics (KL, QR) for the permuted data set ( $X'$ , LRR, var.LRR) and repeating these  
522 steps 10,000 times generates the distribution of the test statistics under validity of the NH and  
523 allows extraction of a p-value as the fraction of permutations that yielded a similar or larger  
524 value for the test statistic (KL or QR) as the original data set ( $X$ , LRR, var.LRR).

525 In comparison to alternative approaches, our methods are robust and non-parametric -  
526 they do not rely on functional assumptions and use only the supposed smoothness of a  
527 possible connection between stressor and response. Reconstructing the NH by simulating  
528 surrogate data guarantees perfect control of errors of the first kind (false positive statements)  
529 and even would handle a constant bias of estimators. Given the breadth of underlying meta-  
530 analyses, we also consider our analysis highly conservative with regard to publication bias  
531 and study selection. Finally, using a weighted approach downgrades the influence of studies  
532 with very high internal variance, and thus decreases the chance of missing threshold-like  
533 responses because of too noisy data (false negative statements).

534 It should be noted that neither the single experiments summarized in each meta-  
535 analysis nor the meta-analyses themselves, were designed to detect thresholds. The inclusion  
536 of studies not necessarily looking for thresholds actually reduces the risk of publication bias  
537 towards positive results. However, even if the underlying experiments were not planned to  
538 detect thresholds, our statistical approach should reveal these if they fall into the covered  
539 range of stressors, which can be expected as this range encompasses stressor magnitudes not  
540 yet experienced under realistic conditions.

541

#### 542 *Statistical analyses*

543 For each effect size in each meta-analysis, a statistical weight is assigned to each data  
544 point as the log-transformed inverse sampling variance of the effect size

545 
$$\log \left( 1 + \frac{1}{var.LRR} \right) \tag{1}$$

546 As described above, surrogate data sets (reflecting the NH) are created by permuting the list  
 547 of stressor values in X (yielding X' = X shuffled). From the list of stressor values, a smooth  
 548 probability distribution  $p_X(gx)$  is computed via weighted [with statistical weights calculated  
 549 following equation (1)] kernel density estimation (with a Gaussian kernel and an optimized  
 550 bandwidth, cf. *Simulations* below) for grid points  $gx$  that span the range of observed stressor  
 551 values (Extended Data S4). A smooth density surface over the grid ( $gx, gy$ ) in the (X, LRR)  
 552 plane is computed from the data set (and the surrogates) via a two-dimensional weighted  
 553 ([with statistical weights calculated following equation (1)] kernel density estimation  
 554 (bivariate Gaussian D-class kernel with optimized bandwidth) (Extended Data S5). For each  
 555 grid point  $gx$ , the density profile along  $gy$  is converted to a conditional probability distribution  
 556  $p_{LRR|X}(gy|gx)$  by normalization (Extended Data S6, with results for the original data and the  
 557 surrogate data). Based on the conditional cumulative distribution function,

558 
$$F_{LRR|X}(gy|gx) = \sum_{gy' \leq gy} p_{LRR|X} \sum p_{LRR|X}(gy'|gx) \tag{2}$$

559 (Extended Data S7), the 5%, 50% (median), and 95% quantiles can be extracted for each grid  
 560 point  $gx$  (Extended Data S8). The test statistics that we devised are:

561

562 (i) the average Kullback-Leibler divergence

563 
$$KL = \sum_{gx} p_X(gx) \sum_{gy} p_{LRR|X}(gy|gx) \log \frac{p_{LRR|X}(gy|gx)}{p_{LRR}(gy)} \tag{3}$$

564 that shares the useful property of being non-negative and that vanishes if, and only if,  
 565  $p_{LRR|X} \equiv p_{LRR}$  (almost everywhere). Pronounced differences between the two empirical  
 566 distributions are thus condensed in values substantially larger than zero.

567

568 (ii) the ratio of interquantile (5%-95%) ranges

569 
$$QR = \frac{IQR_5^{95}(99\%)}{IQR_5^{95}(1\%)} \quad (4)$$

570 where  $IQR_5^{95}(qx)$  denotes the  $qx$ -quantile of the 5%-95% interquantile range of the  
571 conditional probability distribution  $p_{LRR|X}(gy|gx)$ , and the subsequent percentage in brackets  
572 indicates the related weighted quantile across the stressor grid points. We choose this latter  
573 definition for robustness, rather than the max/min ratio which may be prone to distortions by  
574 extremes. This measure was devised to indicate substantial changes of the LRR variance  
575 along the stressor axis.

576

577 (iii) the Hartigan's dip (HD) test statistic tests for multimodality, which, if significant,  
578 indicates that a frequency distribution has more than one mode.

579

580 Values of all test statistics obtained for the original data set were assessed for  
581 statistical significance. This was done by excessively repeating the permutation strategy to  
582 create surrogate data in accordance with the null hypothesis of a non-existent connection  
583 between stressor X and response LRR. p-values for both test statistics (KL and QR) were  
584 obtained as fractions of 10,000 surrogate sets (in case of HD, 2000 permutations), leading to  
585 test statistics exceeding related values of the original data set.

586 In addition to the employed kernel density estimates generating cumulative  
587 distribution functions and derived quantiles, we used a nonlinear quantile regression supplied  
588 by the R package “qgam”<sup>59</sup>. This package is based on general additive models (GAMs) and  
589 returns quantiles instead of standard mean response. With “qgam” we estimated the following  
590 quantiles: 0.001, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.999. Because these quantile curves  
591 were computed sequentially, independently resultant lines could intersect. To resolve this  
592 problem, we used the R package “cobs” to perform a penalized B-spline regression of  
593 obtained quantiles (separately for every grid point  $gx$ ), bound to the constraint of a monotonic

594 increase, thus yielding a smooth cumulative distribution function. As for the kernel density  
595 estimation, exemplary 5%, 50% (median), and 95% quantiles are shown in Extended Data S8.

596 Under default settings, the “qgam” routine was very time consuming and, in  
597 comparison with the bandwidth optimized kernel density method, had inferior test  
598 power (Figs. 2 & S1). This may be due to the fact that, because of excessive  
599 run time an optimization of qgam parameters was not feasible. We therefore constrain  
600 reporting of our results to these obtained with the optimized kernel density method.

601

## 602 *Simulations*

603 We examined the performance of our tests by simulating artificial data sets that combined  
604 nine deterministic backbone structures with additive noise (normally distributed random  
605 fluctuations) of controlled intensity. The deterministic backbone structures were chosen to  
606 reflect a broad range of scenarios. The noise intensity is quantified via the inverse signal-to-  
607 noise ratio (ISNR), i.e. the size ratio of fluctuations and backbone structure. The nine cases  
608 are depicted in Extended Data S3, each for small (ISNR: 0.05) and large (ISNR: 0.95) noise  
609 intensity. Fig. 2 and S1, we list the expected outcome of the three designed tests for the noise-  
610 free case. To assess the performance of the tests under various noise conditions, we simulated,  
611 for each isnr value (in the range [0-1]), 1000 artificial data sets and collected related test  
612 decisions (for two decision criteria  $p$ : 0.05 and 0.01 and all three tests). In case of an expected  
613 positive test, the fraction of positive test decisions thus estimates the test power (1-error of the  
614 second kind). We note that simulations of the test power were also underlying the  
615 optimization of the kernel bandwidth, where Bandwidth selection was based on the “solve- the-  
616 equation” method of Sheater & Jones<sup>26</sup>.

617 In all simulated cases with small to moderate noise (ISNR < 0.5), threshold structures  
618 in simulated response ~ stressor relations could be detected with high reliability (at least for  
619 the KL and QR test). Of course, for strong noise (ISNR  $\geq 1$ ), thresholds may be masked by

620 random fluctuations reflecting natural variability. In such situations, the underlying threshold  
621 structure, though present, will no longer be ecologically relevant because it is overridden by  
622 natural variability.

623

624 Data Availability: All data and code are available at

625 <https://zenodo.org/record/3828869#.XsI4ZmgzaUk> (<http://doi.org/10.5281/zenodo.3828869>).





627 **Legends Extended Data**

628

629

630 **Extended Data S1: Test power as in Fig. 2, but for the “qgam” approach.** Fractions of  
631 positive test results (equals test power when test should be positive) for simulated test cases.  
632 We analyzed the test power for 9 scenarios of responses to pressure in meta-analyses, the  
633 derivation of each scenario is described in the supplementary online material, Extended Data  
634 S7. Scenarios a–d do not comprise a threshold, where scenario a is the null model without an  
635 effect of the pressure on the response. Scenarios e–i do comprise a threshold, for the latter two  
636 combined with intermediate responses. For the three statistical test used in our analyses, the  
637 expected outcome is colour-coded, with green representing that the test should be significant.  
638 We then tested the proportion of 1000 simulated data sets for which the tests were significant  
639 with a probability  $p = 0.05$  (black) and  $p = 0.01$  (blue). We did for increasing noise variance  
640 (= inverse signal-to-noise ratio). The three tests together allow perfect detection of thresholds  
641 at the absence of noise (scenarios e-h), only if threshold-type and gradual responses are mixed  
642 (scenario i), the analysis of multimodality (HD) fails, giving the same output as a gradual  
643 increase in mean and variance of the response (scenario d). With increasing noise variance,  
644 however, the detection probability for thresholds via HD and QR rapidly decreases. We used  
645 default settings for the “qgam” approach due to high runtimes and computational effort, thus  
646 settings are not optimized as for the test power calculations based on kernel density  
647 estimation. Note: HD is equal to kernel method, because it is not based on different quantile  
648 estimations.

649

650 **Extended Data S2:** Further meta-analyses testing for changes in the response magnitude  
651 along increasing pressure strength. Red and blue shaded regions indicate the (5%-95%)  
652 interquantile ranges for the bivariate data (including the pressure gradient) and the univariate  
653 LRR data (ignoring the pressure gradient = homogeneous marginal probability), respectively.  
654 Solid red and dashed blue thick lines trace the related median (50% quantile). Overlain are the  
655 data points and at the bottom the yellow shaded area indicates the distribution  $p_x(g_x)$  resultant  
656 from a weighted kernel density estimation (Extended Data S2). Color codes for habitat  
657 (darkblue: freshwater, aquamarine: marine, green: terrestrial), circle size reflect statistical  
658 weight.

659

660 **Extended Data S3: Test cases at different noise levels.** In order to assess the power of our  
661 statistical tests, we simulated artificial meta-analyses combining prototypical  
662 response~stressor relationships with (normally distributed) random fluctuations reflecting  
663 natural variability, and compared related statistical test results with expectations. Stressor  
664 range (along horizontal range) and deterministic effect sizes (along vertical axis) are  
665 normalized to  $[-0.5, 0.5] \times [0.5, 0.5]$ . Stressor values are normally distributed with mean zero.  
666 The relative intensity of random fluctuations is quantified by inverse signal to-noise ratio  
667 (isnr). A grey background indicates absence of thresholds, yellow background threshold  
668 presence.

669

670 **(a)** (neutral -simple-): Here pressure strength has no impact on the response, which falls into a  
671 single response. Thus, we assume that across all “studies” in this “meta-analysis”, there is one  
672 main response type and no threshold.

673 **(b)** (neutral -bimodal-): Here pressure strength has no impact on the response, which falls into  
674 either of two alternative attractors: a weak and a strong response. Thus, we assume that across  
675 all “studies” in this “meta-analysis”, there are two main response types and no threshold.

676 **(c)** (plain trend, proportionate response): A gradual response with no change in variability  
677 revealing a trend but no threshold.

678 **(d)** (gradual, no threshold): A nonlinear but smooth increase with smoothly increasing  
679 variability. Here we assume that the responses increase with some normally distributed error  
680 with the pressure without transgressing any threshold.

681 **(e)** (saddle-node bifurcation): A widely discussed model situation in the context of  
682 ‘tipping points’ and ‘catastrophic regime shifts’.

683 **(f)** (strict threshold): Here we assume that across all studies in a meta-analysis, the  
684 response switches from weak to strong (as defined in case a) at exactly the same threshold for  
685 each study. This assumption is very unrealistic (see below) but makes the case when there are  
686 two main response types and a global threshold holding for any single study in the meta-  
687 analysis.

688 **(g)** (variable threshold): Here we assume that all studies in a meta-analysis potentially  
689 transgress a threshold, but the position of the threshold differs. Thus, the probability that the  
690 response switches from weak to strong increases with increasing pressure. Response  
691 similar to Case a.

692 **(h)** (variable threshold with intermediates): Here we assumed that not all studies in a  
693 meta-analysis potentially transgresses a threshold, but some of the studies show gradual  
694 responses. As in Case f, the position of the threshold differs between studies and the  
695 probability that the response switches from weak to strong increases with increasing pressure.  
696 As for cases a,b,e and f, we assume there are two main response types. This scenario can be  
697 distinguished from case d by the abrupt change in variance along the pressure gradient.

698 **(i)** (variable threshold and variable effect sizes below and above threshold): Here we  
699 assumed that the position of the threshold differs between studies (as in Case f) and any  
700 experiment in the study had a 50% chance that the threshold was crossed, independent of the  
701 pressure magnitude. By contrast to cases a, b and e-h, we relax the assumption that there are  
702 two main response types, but transgressing the thresholds leads to an increase in effect size,  
703 which depended on the position on the pressure gradient. Thus, if a study with a large  
704 pressure magnitude transgressed the threshold, the increase in response magnitude was larger  
705 than if a study with an overall small pressure did so.

706 **Extended Data S4. Permutation example.** An example data set **(a)** together with a surrogate  
707 data set based on permuted X values **(b)**; as in Fig. 2 of the main text, color codes habitat  
708 (blue: marine, green: terrestrial), circle size reflects statistical weight, and the yellow shaded  
709 area indicates the distribution  $p_X(gx)$  resultant from a weighted kernel density estimation.

710 **Extended Data S5: Two-dimensional probability densities.** Densities are calculated over a  
711 grid  $(gx,gy)$  for the original data set **(a)** and the surrogate data set **(b)**

712

713 **Extended Data S6: Conditional probability distribution example.** The conditional  
714 probability distribution  $p_{LRRVX}(gy \vee gx)$  for each grid point  $gx$  together with the marginal  
715 distribution  $p_{LRR}(gy)$  (thick black line). **(a)** original data set, **(b)** surrogate data set.

716

717 **Extended Data S7: Cumulative distribution example.** The cumulative distribution  
718 functions  $F_{LRRVX}(gy \vee gx)$  and  $F_{LRR}(gy)$  (thick black line) for the probability profiles shown  
719 in Fig.S3. **(a)** original, **(b)** surrogate.

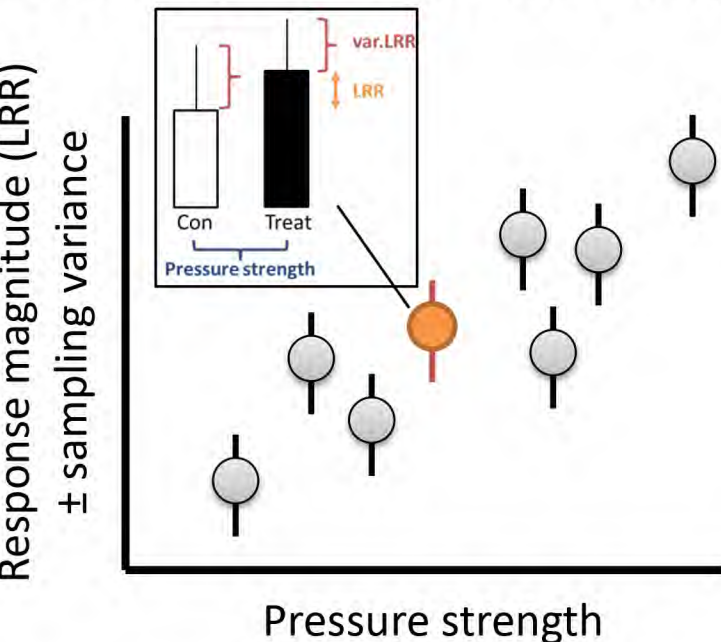
720

721

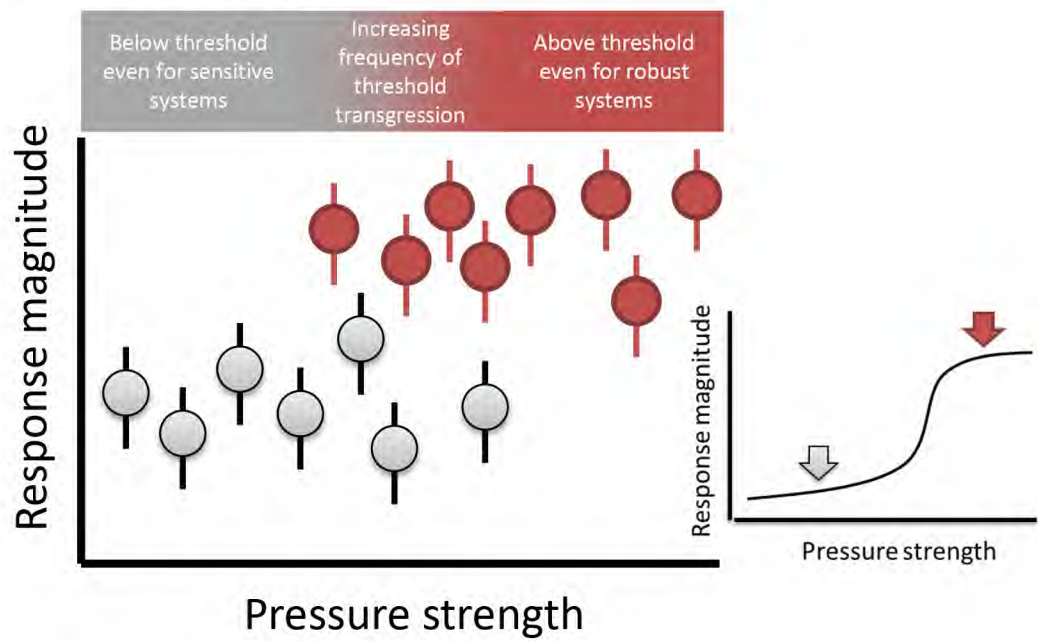
722 **Extended Data S8: Comparison of kernel density estimation and “qgam”.** Images of the  
723 reconstructed statistical structures for an original data set (MA1.1) and one of its surrogate  
724 data sets. **(a)** Quantiles estimated by optimized kernel density estimation; **(b)** Quantiles  
725 estimated by “qgam”.

726

**a Primary data from meta-analysis**

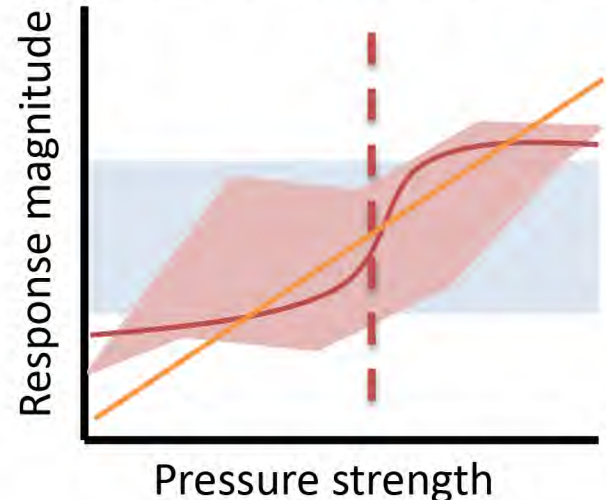


**b Rationale**

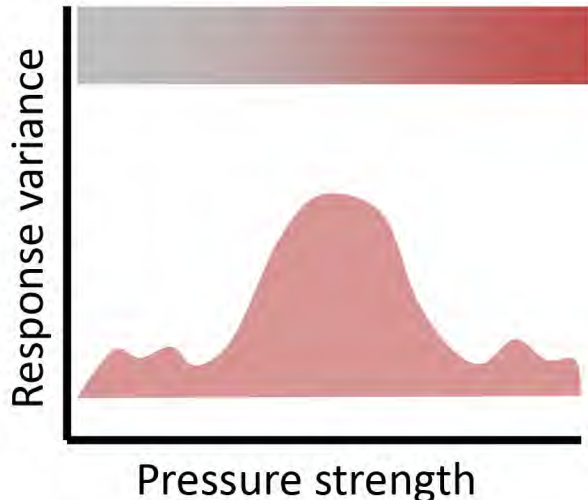


**c Statistical tests**

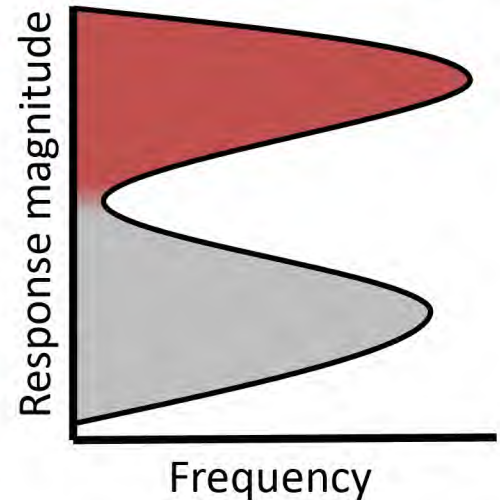
**KL:** impact of pressure strength

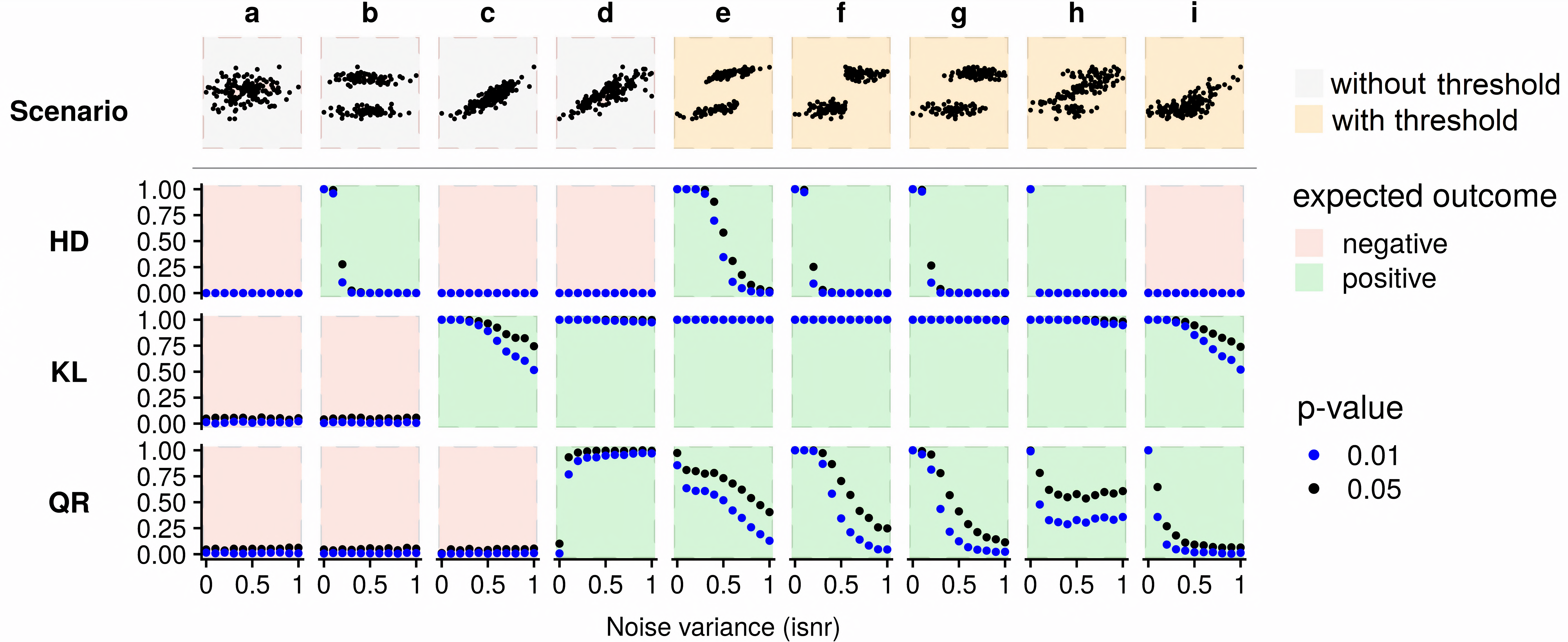


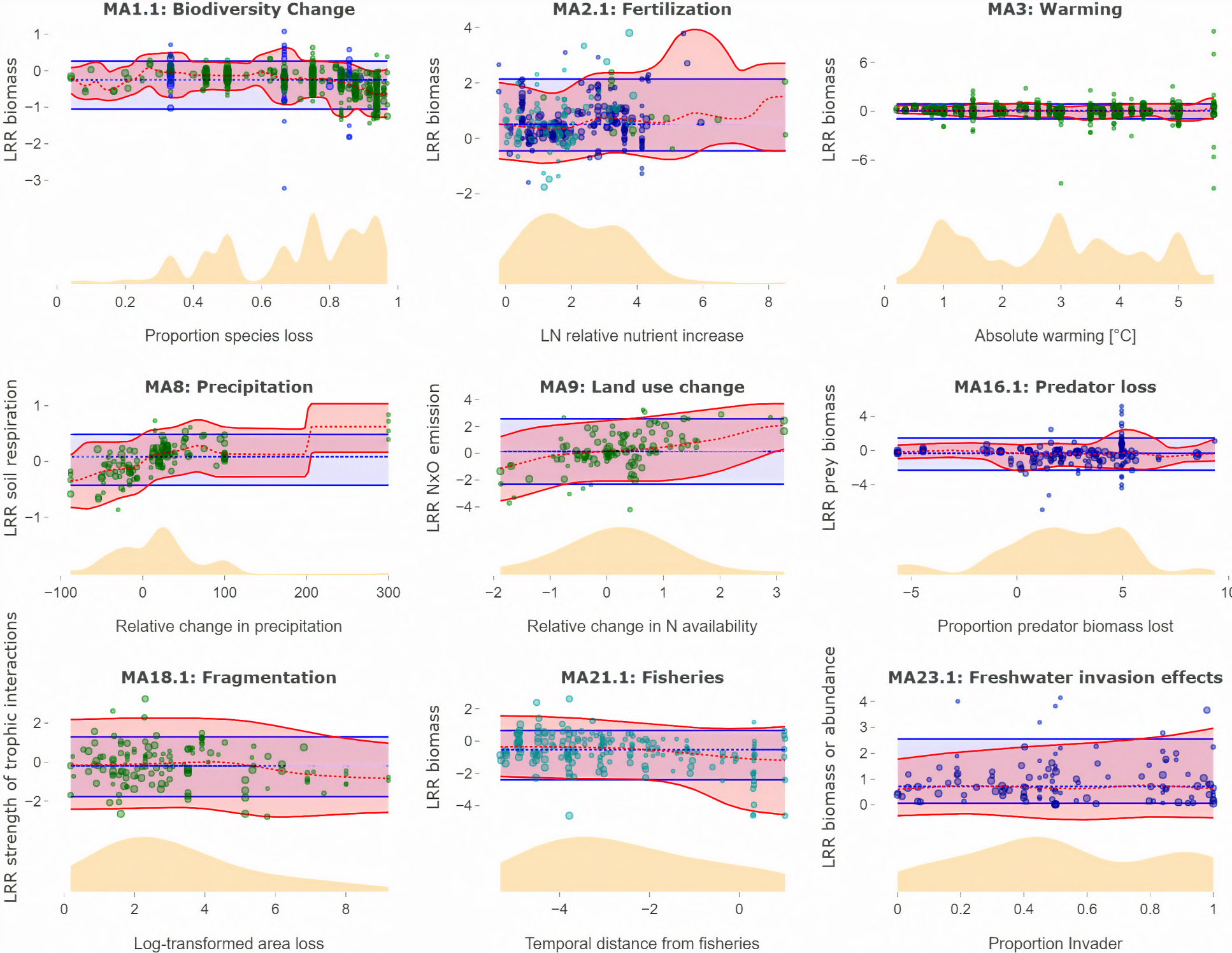
**QR:** test for variance increase



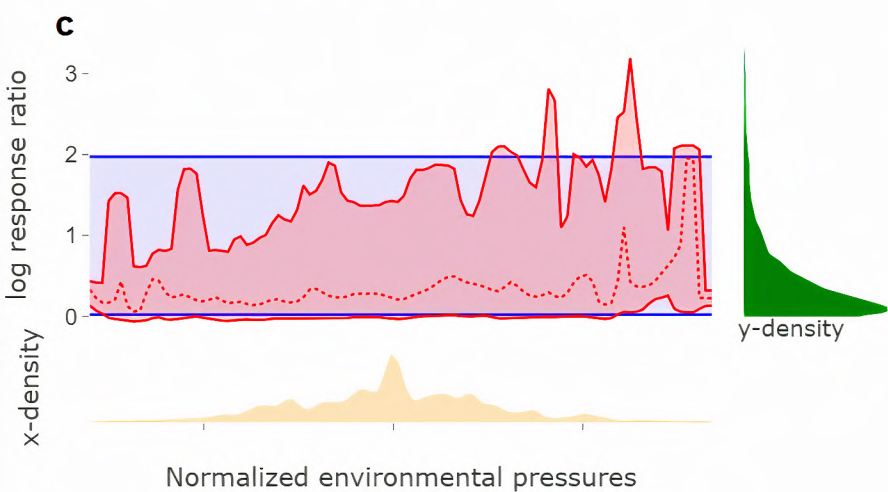
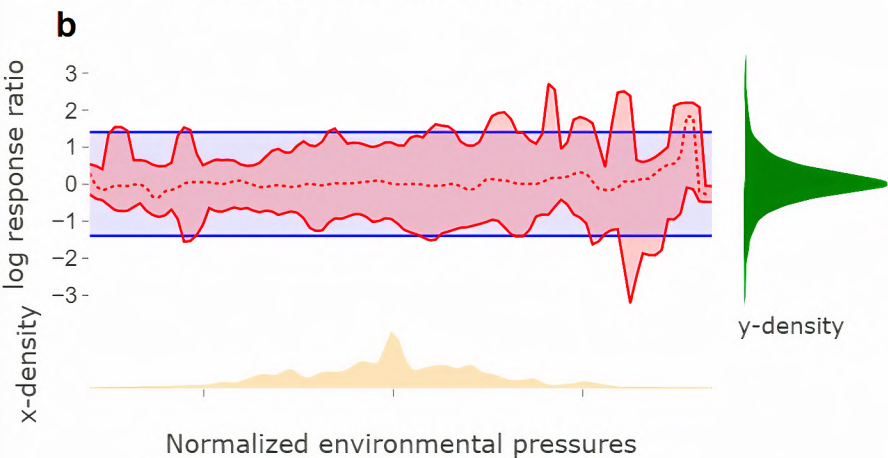
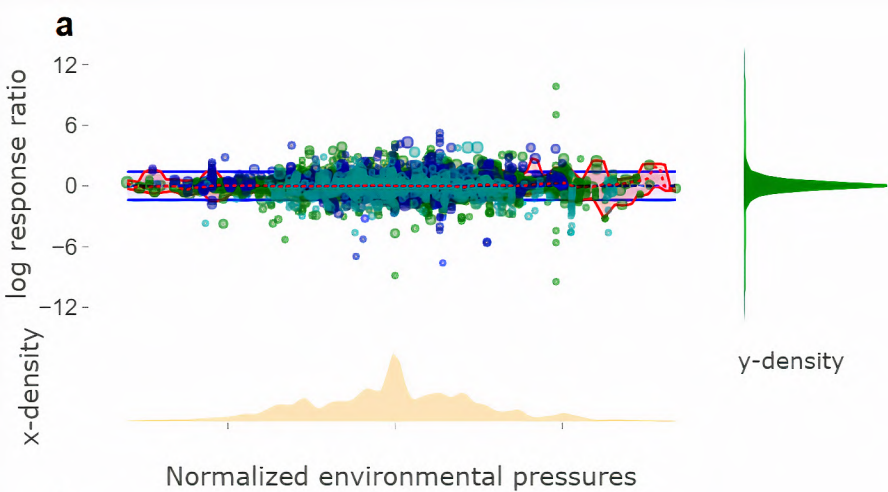
**HD:** test for multimodality



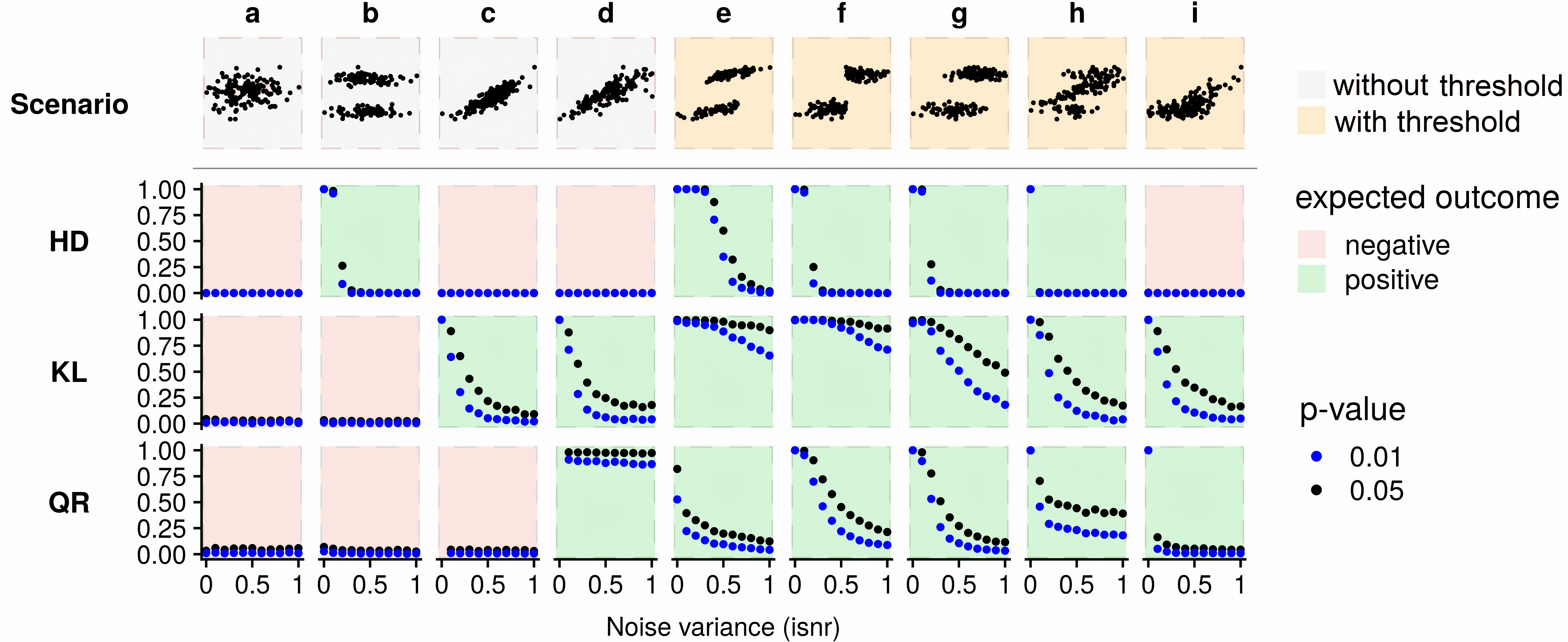




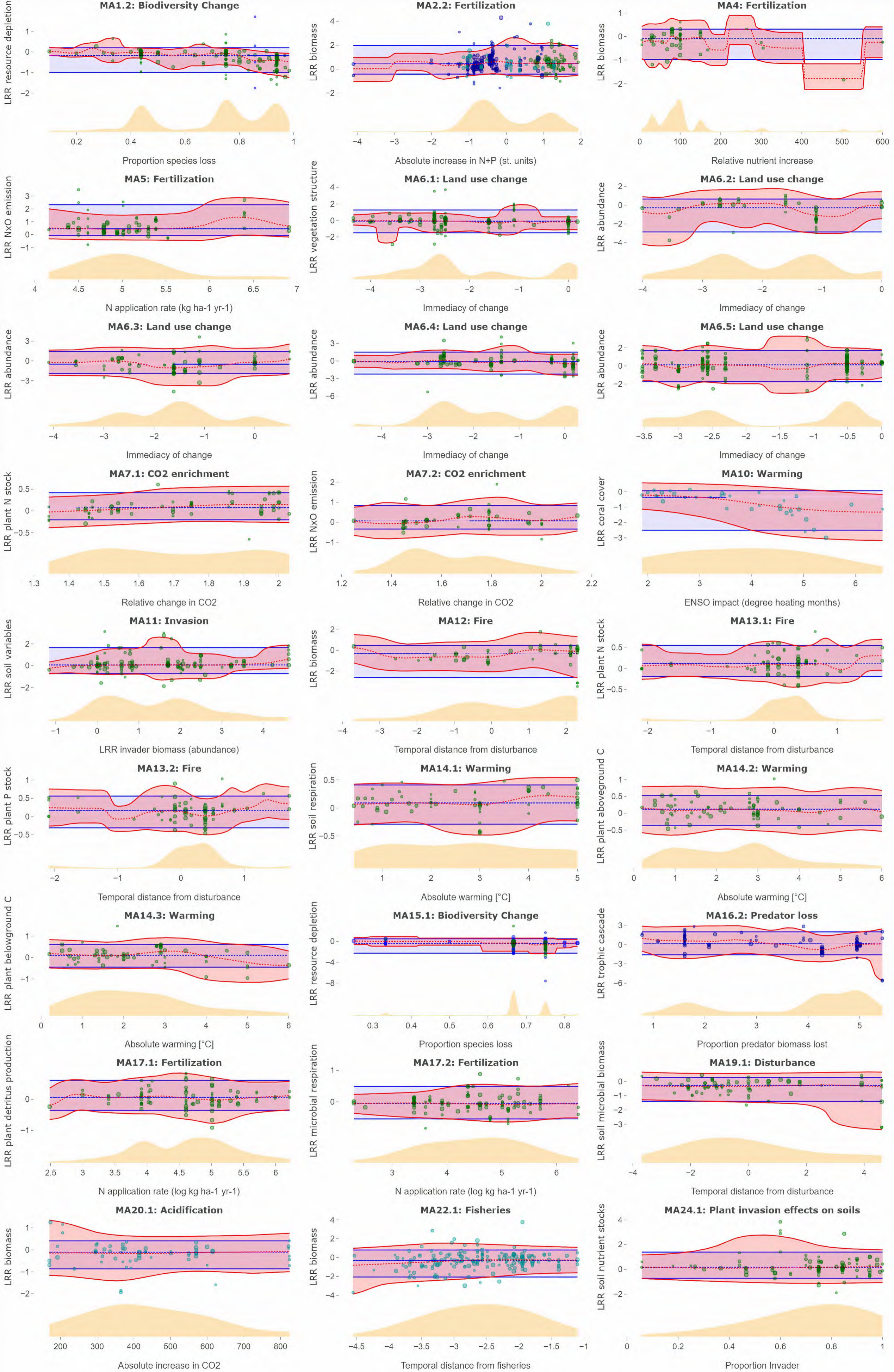
— lower quantile marginal    — lower quantile    ⋯ median marginal    ● aquatic    ● freshwater    — density  
— upper quantile marginal    — upper quantile    ⋯ median    ● terrestrial    ● marine

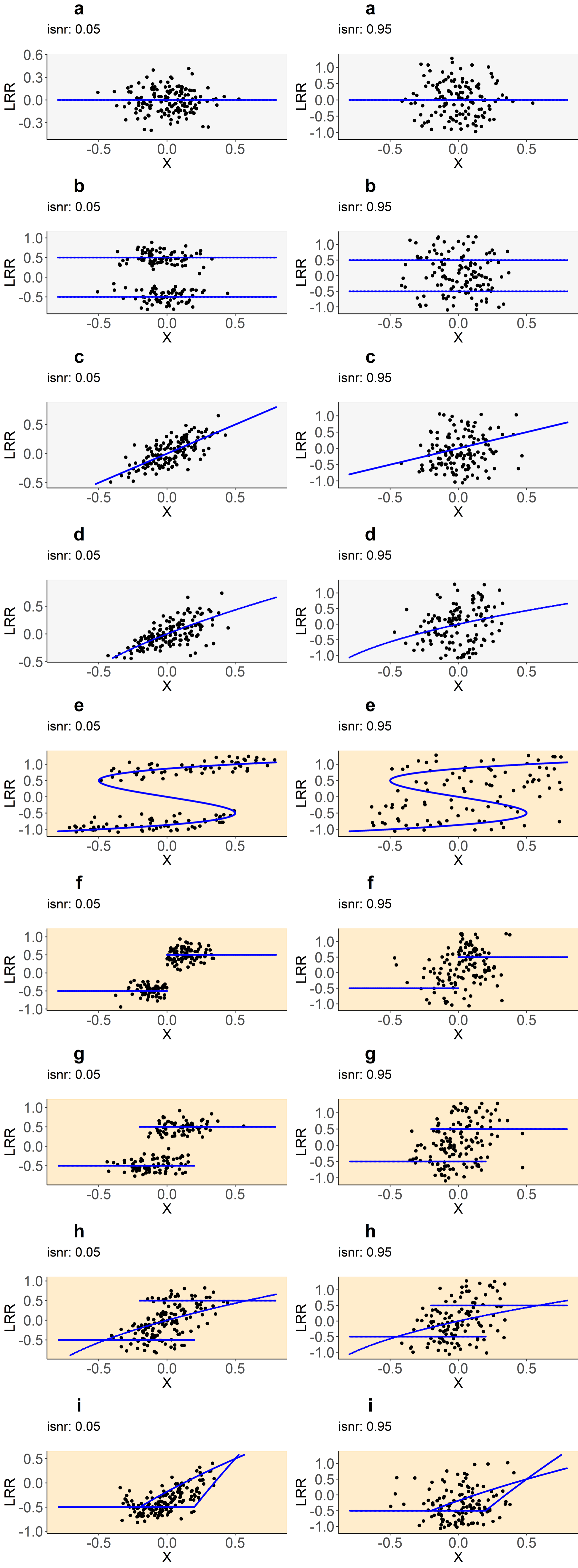


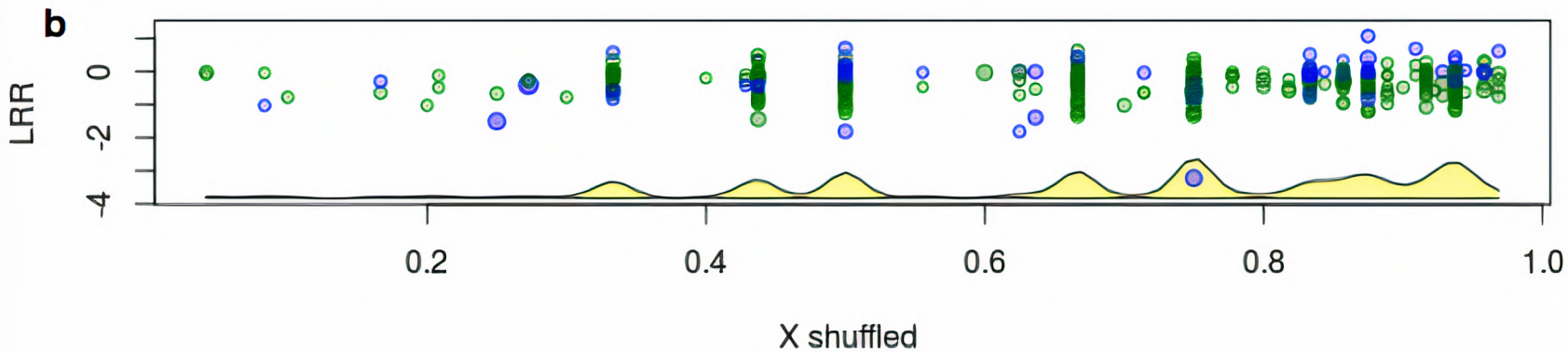
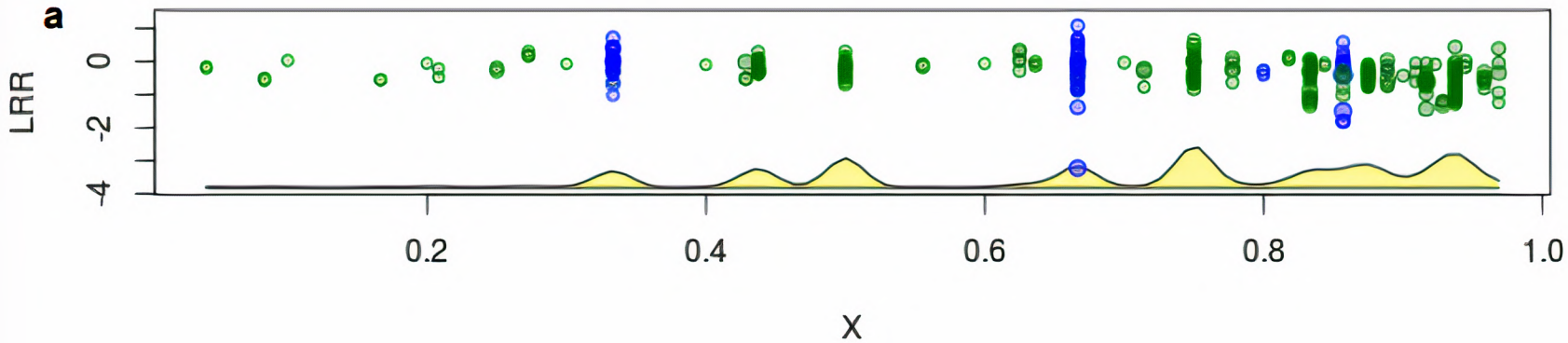
- lower quantile marginal
- upper quantile marginal
- median marginal
- ⋯ median
- freshwater
- marine
- x - density
- y - density
- lower quantile
- upper quantile
- aquatic
- terrestrial

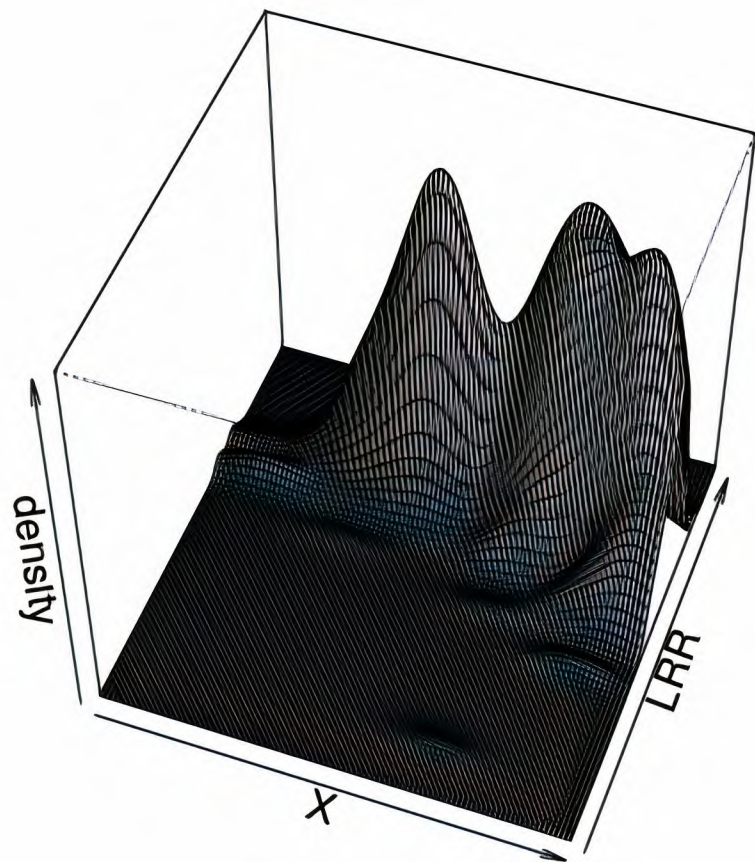
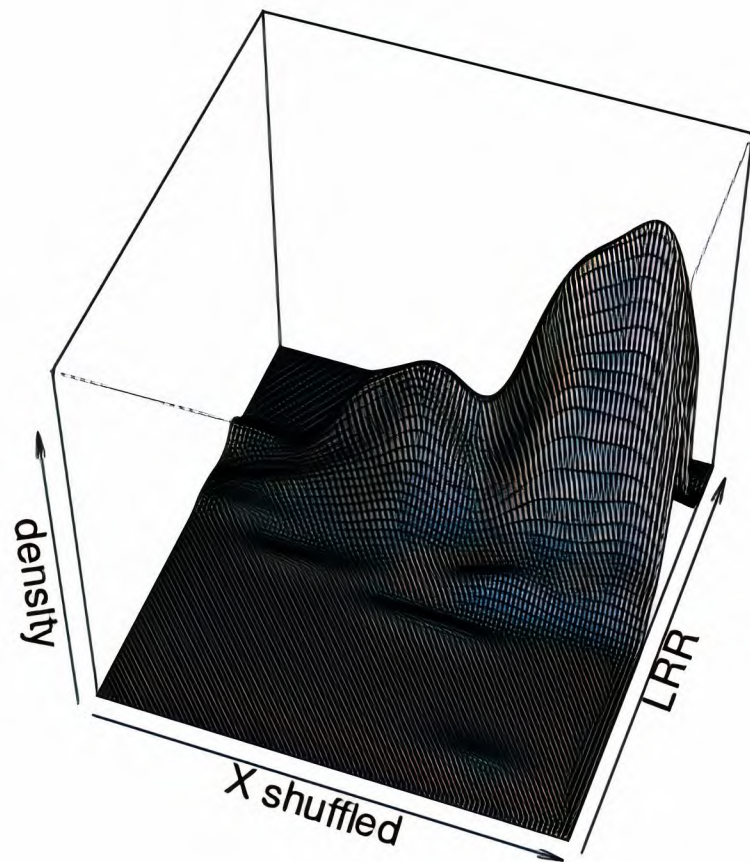






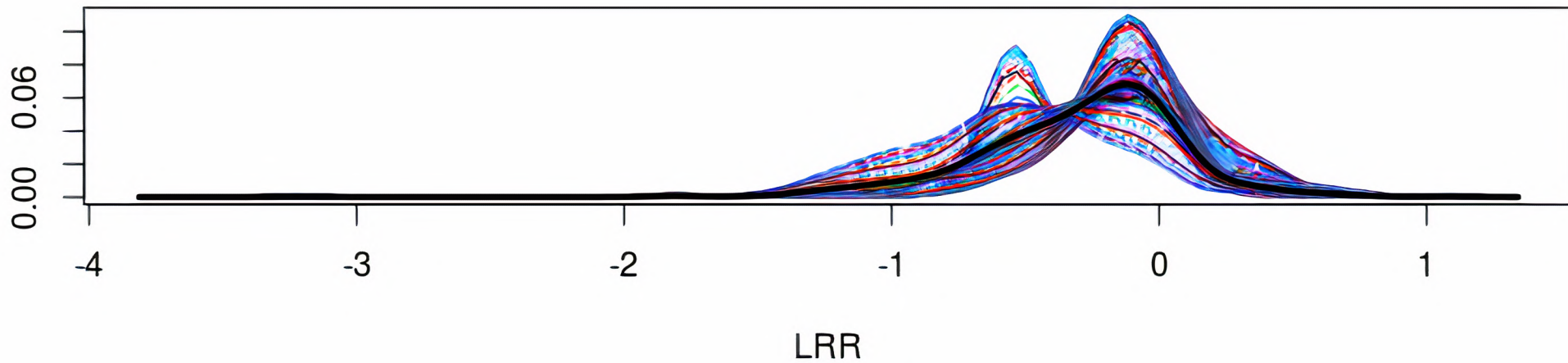




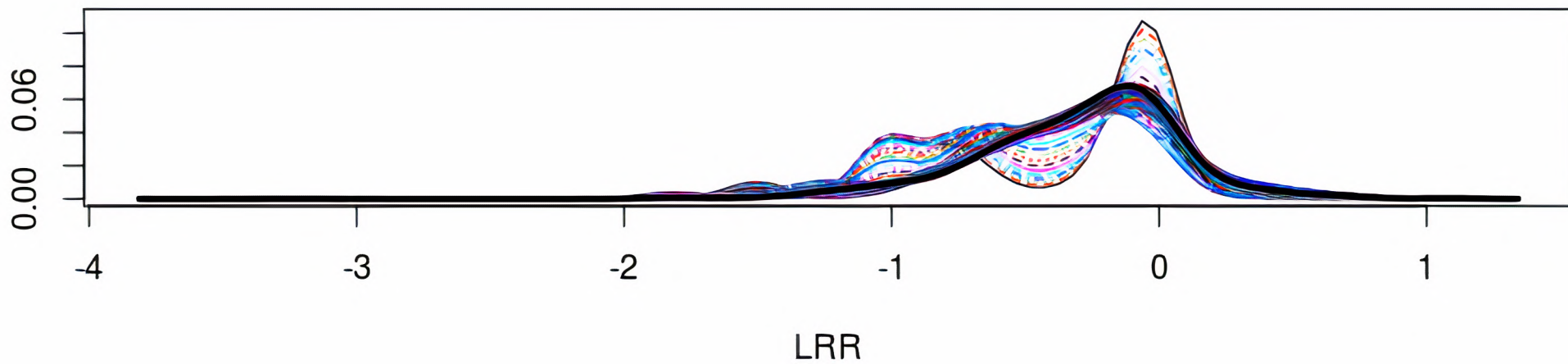
**a****b**

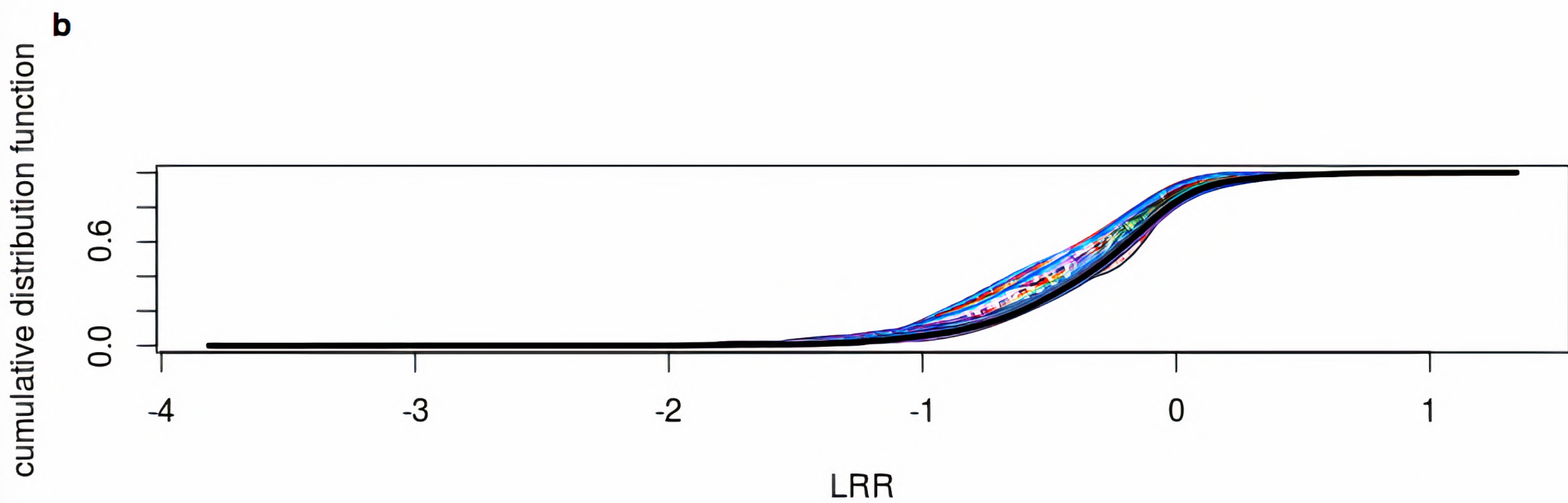
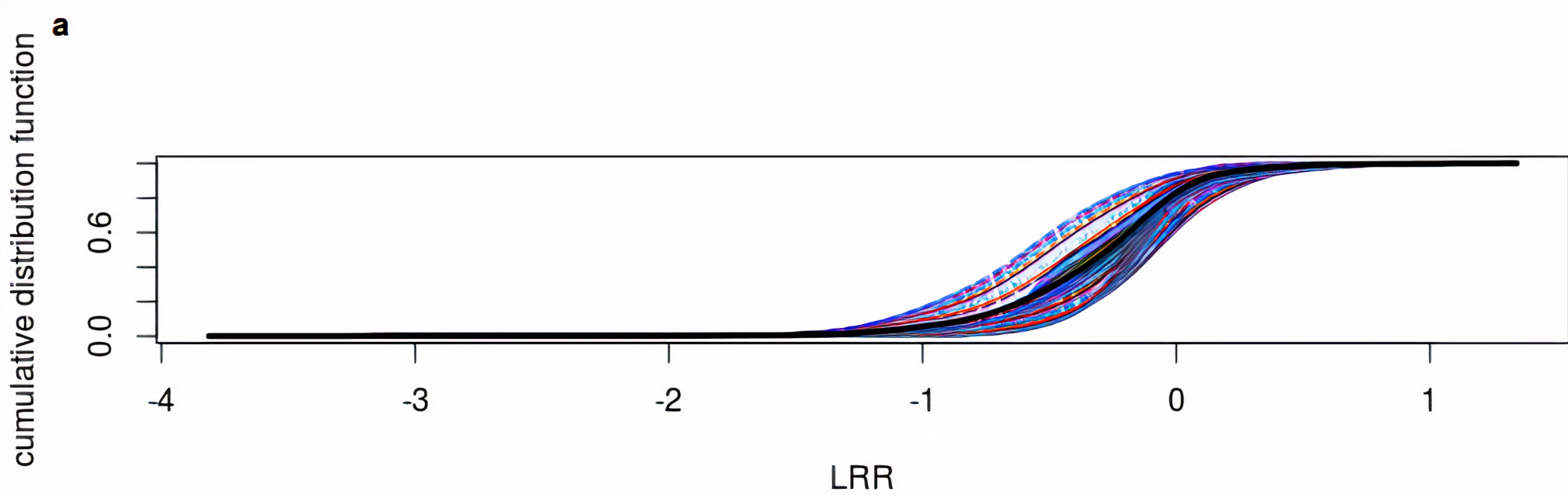
**a**

probability distributions

**b**

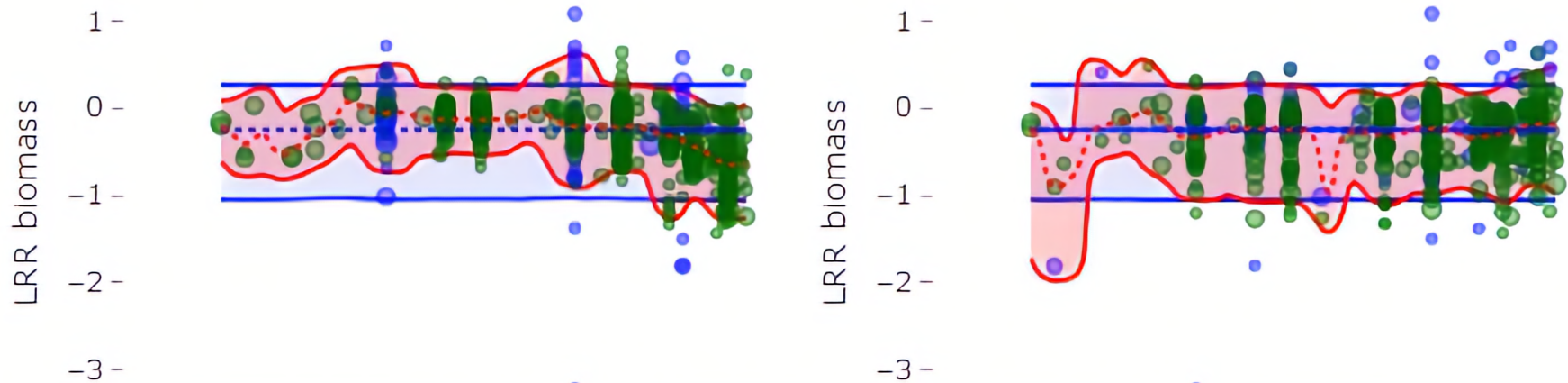
probability distributions



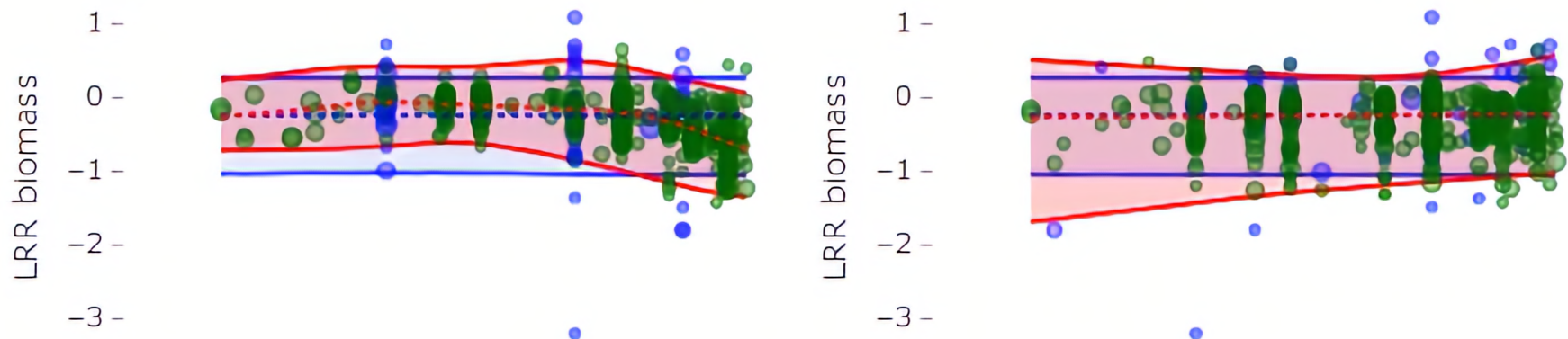


# MA1.1: Biodiversity Change

**a**



**b**



Proportion species loss (original)

Proportion species loss (shuffled)

