Supplementary Materials for

The role of multiple global change factors in driving soil functions and microbial biodiversity

Matthias C. Rillig[1,2,3*], Masahiro Ryo[1,2,3], Anika Lehmann[1,2], Carlos A. Aguilar-Trigueros[1,2], Sabine Buchert[1,2], Anja Wulf[1,2], Aiko Iwasaki[1,2], Julien Roy[1,2], Gaowen Yang[1,2]

[1] Freie Universität Berlin, Institute of Biology, 14195 Berlin, Germany
[2] Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), 14195 Berlin, Germany
[3] These authors contributed equally to this work

Correspondence to: rillig@zedat.fu-berlin.de

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Materials and Methods

Literature synthesis.

We conducted a literature search on the 15th of September 2018 in Web of KnowledgeTM with the following search string: (warming OR temperature OR precipitation OR drought OR water* OR CO2 OR "carbon dioxide" OR O3 OR ozone OR N OR nitrogen OR sal* OR "global change" OR "climate change" OR "land use chang*" OR (plant NEAR/3 invas*) OR ((invas* OR alien) AND species) OR fungicid* OR bacteriocid* OR herbicid* OR pesticide* OR habitat loss* OR agricult*expans* OR "land use chang*" OR "land convers*" OR phosphor* OR P OR fertiliz* OR N OR excess nitrogen* OR excess phosph* OR nutrient pollution OR nitr* pollut* OR phosph* pollut*) AND (soil). Our search was restricted to articles written in English language and published between 1945 and 2017. We included no citation indices and focused only on the Web of Science category "Ecology". We retrieved 4,202 hits.

In order to be included in our data set, articles had to present data on soil systems derived from experimental studies (no observations or simulations were allowed); thus, factors had to be applied by experimenters. If samples from field experiments were used in an additional laboratory assay and both field and lab data were presented in the same study, we focused on the field experiment only.

Following the search string, studies had to present data on at least one of the following nine global change drivers as treatment factors: temperature (ambient and increase/decrease), water (drought or irrigation/precipitation manipulation), CO_2 (ambient and elevated), O_3 (ambient and elevated), fertilization (including inorganic and organic forms of N and P fertilizer), land use change (i.e. conversion from natural to cultivated land; reforestation/renaturation of cultivated land), species invasion (plant, animal or microbes), agrochemicals (including pesticides, soil conditioners and surfactants), salt (ambient and elevated). Furthermore, in rare cases, we were not able to gain access to articles. This was the case for articles published before 1990 (concerning <10 articles) and those published in the Journal of Soil and Water Conservation; hence, these were not considered in our screening.

A total of 1228 articles matched our criteria and were incorporated in our data synthesis. For this, we collected data on publication year, publication name, experimental context (agricultural or natural system), setting (lab or field), response variable focus (community or process measure or both) and tested factors of global change (Data S1). For data visualization, we used the R package "ggplot2" and its extensions "ggpubr" and "geomnet" in R v.3.4.3 (*18, 19*); the latter package was used for generating the network graph.

We collected data on how many articles were published in the Web of Science database in the category "Ecology" for each year for which we found articles matching our inclusion criteria (1968-2017). These data were incorporated to account for the overall increase in publications over the examined time span.

We additionally investigated all studies reporting effects for three and four global change drivers, asking in how many such studies and response variables higher order interactions were found. To do this, we screened model summary tables and counted for how many tested response variables per study a significant ($p<0.05$) three- or four-way interaction term was reported.

Microcosms and experimental design.

We used 50 mL conical tubes (Corning propylene centrifuge tubes) with screw-top caps as experimental units. Caps contained a septum through which a temperature sensor was inserted (or a temperature sensor dummy made from the same material). The lid of each tube was modified to contain a septum that allowed sampling of CO_2 (see below). Tubes were placed in beakers filled with sand to provide insulation from neighboring units and placed in a fully randomized fashion inside a controlled environment chamber (incubated in the dark, 60% relative humidity, ambient temperature 16°C). Tubes were filled with 25.0 g of freshly collected soil (sieved to 2 mm, with all coarse organic material removed, ground to a powder and added back to soil per experimental unit) from a local grassland (52° 33' 09.53''N, 12° 40' 07.86'' E). Soil properties (LUFA, Rostock) were: 86.6% sand, 10.8% silt, 2.6% clay; pH ($CaCl_2$) 4.1; 2.8 mg 100 g-1 P; 2.0 mg 100 g-1 K. To this, 5.0 g of previously sterilized (autoclaving, 121°C, 1h) 'loading' soil was mixed that contained the appropriate dose of the chemical treatments, so that all tubes contained 30.0 g of soil in the end. We used this 'loading' soil to achieve more effective mixing of chemical agents into soil; it was sterilized to avoid any exaggerated effects on the soil community.
The treatments represented 10 factors of global change, all applied individually to each experimental unit, and representing abiotic factors, resource availability, chemical toxicants and compounds (inorganic and synthetic

organic), and physically-acting agents (microplastic). The factors were: temperature (+ 5°C increase), nitrogen enrichment (NH_4NO_3 added to the equivalent of 100 kg N ha^{-1} yr^{-1}), drought (30% of water holding capacity compared to 60%), salinity (4.0 dS m^{-1} with NaCl), microplastic (1.0 g polyester microplastic fibers kg^{-1} soil), insecticide (50 ng g^{-1} imidacloprid), herbicide (50 mg kg^{-1} glyphosate formulated as 'Roundup'®), antibiotic (3.05 mg kg^{-1} oxytetracycline), antifungal agent (6.0 mg kg^{-1} carbendazim), and heavy metal (100 mg Cu kg^{-1} as $CuSO_4$). For full details and justifications for these factors please see below. Since some compounds (imidacloprid, oxytetracycline) were dissolved in DMSO, we insured the same amount of DMSO and water was added to all experimental units. We carried out a test where we added DMSO at the target concentration to soil under control conditions (n=10 each for DMSO addition and no-addition); there were no effects of DMSO on the key variable of soil aggregation (Fig. S6). All experimental units also received the exact same amount of handling and mixing time.

We had the following levels of replication: control (n=20), individual factors (n=8 each), and factor richness levels (n=10 each), for a total of 140 experimental units. We used random sampling from the pool of 10 factors for the factor richness levels to create replicates. We did this because our objective is to determine how an increasing number (not factor identity or particular factor combinations) of global change factors influence soil ecosystem processes. This means that our resolution does not allow statements on specific, individual factor interactions. Addressing such interactions would result in an experimental design encompassing all factor combinations with 1,024 unique treatments, which, applying our level of replication, would mean 10,230 experimental units.

The temperature treatment was applied using a heating cable wrapped around the individual tubes (PT2011, Exo Terra, Germany) with a separate controller per experimental unit (ETC-902, VOLTCRAFT, Germany).

At the start of the experiment, we added 4.20 mL of water (equivalent to 60% water holding capacity) to each experimental unit (except for drought, which received half this amount). The experiment ran for six weeks, with soil respiration measured after 3 weeks, and was then harvested.

Experimental treatments.

We here present the rationale for the 10 tested factors of global change.

1) Warming. We used an increment of 5.0°C over an ambient temperature of 16.0°C to simulate warmer spring temperatures. Temperatures were recorded (tempmate B2 data logger, imec Messtechnik GmbH, Heilbronn, Germany) in soils in a set of experimental units (n = 3 each, for warming and control) to verify treatment application. A level of +5.0°C is frequently used in studies of warming effects on soil systems, and this level corresponds to climate scenarios predicted for the next 100 years (*20*).

2) Nitrogen enrichment. Nitrogen enrichment as a consequence of human activity has long been recognized as a factor of global change (*21*). We added ammonium nitrate (>-98%, p.A., ACS. Roth GmbH, Karlsruhe, D. article K299.1) to the experimental units in dissolved form. We used a one-time addition to reflect annual accumulation rates on a ha^{-1} yr^{-1} basis as is common in the literature (*22*), and we assumed 10 cm soil depth for conversion to addition rates in the experiment. We added the equivalent of 100 kg N ha^{-1} yr^{-1} to reflect a high N enrichment rate.

3) Drought. Increased occurrence of drought is a well-established aspect of climate change (*23*) also with direct relevance to Europe (e.g. (*24*)) and the Brandenburg region in Germany, near Berlin, from which the soils were taken. We represented drought by adding half of the amount of water at the beginning of the experiment, compared to control water levels that were at 60% of water holding capacity. This level of drought is often employed in experimental studies (e.g. (*25*)). Our experiment thus simulated a drought episode; however, without the transition from sufficient water availability to the drought situation.

4) Heavy metal. Heavy metal contamination represents pollution with persistent (i.e. non-degradable) inorganic compounds. We chose copper, which is of high relevance, because of mining, atmospheric deposition, and also use in organic agriculture; copper is an important soil pollutant in Europe (*26*). We added copper (ii)-sulphate - pentahydrate (BioChemica. AppliChem GmbH, Darmstadt, Germany) to the soil in dissolved form to a final concentration of 100 mg Cu kg^{-1} to simulate a hotspot of copper contamination (Cu concentrations in German soils range typically from 2 to 50 mg kg^{-1}). Such concentrations can occur in Cu polluted sites (e.g. mining, agriculture).

5) Microplastic. Microplastics have been found in many ecosystem compartments, including many soils worldwide, and can be regarded as a factor of global change (*27, 28*). Microplastic comes in many forms, and we previously showed that microplastic fibers exert effects of primarily physical nature; we thus used polyester fibers (Glorex Inspirations, 'Bastelwatte', 100% polyester, item number: 6252105; cut by hand) at a concentration of 0.1% (w/w) of soil, similar to levels used previously (*29*). Procedures were as described before, e.g. the cutting and mixing, and we microwaved the fibers for 3 min to reduce the risk of introducing microbes with the fibers.

6) Salinity. The Millenium Ecosystem Assessment (*30*) recognizes soil salinization as a major human-induced driver; salinization also has a strong European relevance (*31*). Soils with an electrical conductivity between 4.0

and 8.0 dS m^{-1} are regarded as moderately saline. We chose the lower end of this spectrum here, that is 4.0 dS m^{-1}, which we achieved by adding NaCl to the soil (that is, our treatment also reflects sodicity).
Synthetic organic chemicals have recently been argued to be regarded as factors of global change (*32*). We used several substances of different chemical classes, target organisms, and also representing different uses as separate factors in our experiment: an herbicide, an antibiotic, an insecticide and an antifungal agent.
7) Herbicide. Glyphosate is one of the most commonly used herbicides worldwide. We used the commercial product Roundup® PowerFlex (Monsanto Agrar Deutschland, Düsseldorf), which contains 480 g L^{-1} glyphosate as active ingredient (a.i.), but also other additives, such as surfactants. We chose to use the commercial product together with its formulation rather than just the active ingredient, since the surfactant may contribute to effects. This followed the protocol of Ratcliff et al. (*33*) who used an application rate of 50 mg kg^{-1} a.i. to simulate the approximate soil glyphosate concentration following a single application at the recommended field rate of 5 kg a.i. ha^{-1}.
8) Antibiotics. There has been growing concern about antibiotics in the environment due to the spread of antibiotic resistance worldwide (*34*). There are many different classes of antibiotics, including tetracyclines; these are used on humans and in a veterinary context and residues have been documented in soils (*35*). Oxytetracycline has been found at soil concentrations of 305.000 ng kg^{-1}, and it has been shown to persist and accumulate in the soil environment (*35*). We used the equivalent of ten times this amount (added as oxytetracycline-dihydrate; Sigma-Aldrich, MO, USA, catalog #PHR1537) to simulate a temporary hotspot, resulting for example from a fresh deposition of manure on an agricultural field.
9) Insecticide. Neonicotinoid pesticides are now the most widely used class of insecticides in the world (*36*), and are discussed in terms of posing risks to non-target organisms. We here used imidacloprid, which is one of the three most widely used agricultural neonicotinoids (along with clothianidin and thiamethoxam). Observed values of imidacloprid in German and UK agricultural soils ranged from 1.6 up to 50 ng g^{-1} (*36*); we thus used 50 ng g^{-1} imidacloprid (PESTANAL ® analytical standard, Sigma-Aldrich, MO, USA, catalog #37894) in our experiment.
10) Antifungal agent. Azoles are the most commonly used class of fungicides (*37*), and are being used on animals (including humans) and on plants in agricultural fields. Carbendazim, a benzimidazole, has been used previously in soil research, using a recommended field application rate of approx. 6.0 mg kg^{-1} and 20 and 40-fold this concentration (*38*), other studies have used 1 mg kg^{-1} to 100 mg kg^{-1} of carbendazim (*39*); we here used 6.0 mg kg^{-1} (PESTANAL ® analytical standard, Sigma-Aldrich, MO, USA, catalog #45368).

Response variables.

Here we describe the response variables investigated in our test systems. These variables are chosen to represent important biological processes (soil respiration, decomposition), physical properties (water-stable soil aggregates, water repellency), and fungal biodiversity (community composition, dispersion, nestedness). Soil fungi are known to play an important role in mediating these processes and the physical properties examined.
1) Soil respiration. We measured soil respiration as CO_2 concentration (ppm) after three weeks of the experiment. We sampled 3 ml of air from the headspace of each tube and injected this sample into an infrared gas analyzer (LiCOR 6400xt) following Bradford et al. (*40*). At the beginning of the experiment, we flushed each of the tubes with CO2-free air for five minutes to standardize among experimental units.
2) Decomposition. As an indicator of decomposition, we inserted a pre-weighed cellulose filter paper square (approx. 30 mg; Testo AG, Germany; item #0554.0308) into the soil, which was retrieved at the end of the experiment. The corrected weight of this filter paper was used as an indication of decomposition (percent decomposition).
3) Water-stable soil aggregates. We followed a modified protocol by Kemper and Rosenau (*41*). Briefly, the percentage of water stable aggregates was determined by placing samples (4.0 g) on small sieves with a mesh size of 0.25 mm. We used capillary re-wetting with deionized water and inserted samples into a sieving machine (Agrisearch Equipment, Eijkelkamp, Giesbeek, Netherlands). Calculations of the percentage of water-stable aggregates (%WSA) per sample were according to: %WSA = (water stable fraction-coarse matter)/(4.0 g-coarse matter).
4) Soil water repellency. We measured soil water repellency with the water drop penetration time (wdpt (*42*)) method, where a droplet (8 µl) of deionized water is placed onto the soil surface, and the time in seconds is counted until the droplet soaks in (carried out in triplicate per sample).
5) Soil fungal community analysis. We used amplicon sequence variant (ASV) richness. DNA was extracted from 0.25 g soil using the PowerSoil DNA isolation kit (MoBio Laboratories Inc., Carlsbad, CA, USA), following the manufacturer's instructions. The fungal ITS2 genomic region was amplified by PCR using the fITS7 (5′- GTGARTCATCGAATCTTTG- 3′) and the ITS4 primers (5′- TCCTCCGCTTATTGATATGC- 3′)(*43*), respectively extended in 5′ with the p5 and p7 Illumina sequencing adaptors. The amplicon library was sequenced on an Illumina MiSeq 2000 platform (Illumina Inc., San Diego, CA, USA) at the Berlin Center for Genomics in Biodiversity Research (BeGenDiv, Berlin, Germany) using 2x300-bp paired- end sequencing. We used DADA2 (*44*) to obtain denoised, chimera-free, non-singleton

fungal ASVs. For full detail of soil fungal molecular analyses, bioinformatics and statistics see section "Analysis of soil fungal diversity and description of fungal sequence dataset" below. All statistical analyses for fungal data processing were conducted in R v.3.4.3 (*18*) using "vegan" (*45*) and base packages.

6) Soil fungal community composition. Community compositional shifts were assessed using Bray-Curtis dissimilarity. Unconstrained multivariate ordination (NMDS) of the Bray-Curtis sample pairwise dissimilarities indicated that community compositions were mainly differentiated along the first axis. Therefore, we use the first axis as a relative measure of community compositional shifts. The second axis mainly represents community dispersion (as explained in the following paragraph).

7) Soil fungal community dispersion. To represent the within-treatment variability in community composition among replicates (i.e. community dispersion), we calculated the mean distance to the group centroid of all replicates on the NMDS ordination space for each treatment. The longer the mean distance, the more dispersed the community.

8) Soil fungal community nestedness. To test whether ASV-poor communities were a subset of the ASV-rich communities (i.e. nestedness), the temperature metric of the presence/absence community matrix was calculated (*46*). Non-randomness of nestedness was assessed using 999 null matrices (row and column permutation) and statistical robustness was assessed using a two-sided test.

Statistical analyses of treatment effects.

To test if the effects in multi-factor experiments are predictable from the effects measured in single treatments, we quantified the effect size of each treatment in single factor experiments, combined them, and compared them to the observations in multi-factor experiments. We considered the recent argument on null hypothesis significance testing and the scientific meaningfulness of effect sizes (e.g. (*47-50*)). Following them, we neither use the term "statistically significant" nor evaluate results dichotomously by setting a significance level (e.g. α = 0.05); and, we basically assess effect size by comparing the means and 95% confidence intervals (CIs).

An effect size and CIs of each single factor was estimated using a non-parametric bootstrap (*51*), because of unknown probability distributions and high flexibility (Fig. S5). We define effect size as the raw difference in mean between control and treatment (i.e., non-standardized absolute effect size). We are aware of other approaches to estimate or adjust CIs (e.g. BCa (*52*)), but they differ little in practice (*53*). The algorithm conducts sub-sampling with replacement (10,000 iterations).

Then, we used plausible null assumptions for combining effect size of single factors following Schäfer and Piggott (*54*), namely additive, dominative, and multiplicative assumptions. Under the additive assumption, each factor has a unique effect; in the dominative case, the strongest factor dominates (overrides) the others (in case of positive and negative effects, this means picking the strongest absolute value); for the multiplicative case, proportional effect changes are considered and mathematically combined as if effects acted consecutively. Importantly, none of the assumptions can take factor interactions such as synergism and antagonism into account. Our intention to apply these three assumptions was not to figure out the most preferable one. Rather, we test if none of these assumptions were met (i.e. the joint effect sizes predicted in these assumptions do not include the actual joint effect size), so that we can regard the joint effects of multiple factors as fully unpredictable due to synergistic higher-order interactions. Although all assumptions are not necessarily reasonable for all response variables, having multiple (null) expectations is a recommended practice in the recent statistics literature (*50, 55, 56*), and we avoided subjectively removing some assumptions depending on response variable attributes.

For the additive assumption, in each level of factor richness, the effect sizes of the corresponding single factors are simply summed up. For each replicate m (= 1, 2, …, 10),

$$\sum_{i \in K_m} ES_i$$

was calculated, where ES_i is effect size of a single factor i , K_m is the unique subset of factors randomly chosen from the 10 factors for the replicate m: e.g. K_1 = [Temperature, Copper], K_2 = [Glyphosate, Copper], …, K_{10} = [Microplastic, Salinity] at the number of factors level of 2. Since each set of K_m has 8 replicates for each single factor (e.g. K_1 has 8 replicates for temperature and 8 replicates for copper), we applied a bootstrap procedure (1,000 iterations; see Fig. S5). Using this, each Km has 1,000 iterated effect size predictions, and therefore 10,000 effect size predictions were made when all replicates were considered. Finally, the mean value and 95%CI were calculated from the distribution.

The same bootstrap procedures were used in multiplicative and dominative assumptions. In the multiplicative assumption, ES_i was divided by the mean of the control and then multiplied as follows (modified from Thompson et al. (*57*)):

$$CT \prod_{i \in K_m} (1 + \frac{ES_i}{CT}) - CT$$

The formula can consider both positive and negative effect size. However, note that the multiplicative assumption becomes unstable when the control value is very close to zero and relative change values are extremely high. Under the dominative assumption, the ES_i having the highest absolute size within K_m was selected. Therefore, it assumes that only the maximum effect size of single factor determines the joint effect size, regardless of other weaker factors.

Furthermore, we investigated if the responses have a consistent directional change along the number of factors, how much the number of factors alone explains variability in the responses, and how much knowing factor identity and effect size information increases predictability. To represent possible nonlinear changes and factor interactions, we applied a random forest machine learning algorithm (*58*). To test if the change is directional along the number of factors, each response variable was modeled using the number of factors. Then, factor identity (i.e. whether a factor was present or not; binary coding for each factor) was added to the model as explanatory variables. Moreover, instead of adding factor identity, the effect sizes estimated based on the three assumptions described above (additive, multiplicative, and dominative) were added to the model as explanatory variables (i.e. an expected effect size value for each assumption). After constructing these three models (i.e. number, identity, and including effect size information), we evaluated how well each model explains the variability (R^2, %). The algorithm hyperparameters were set as follows, after confirming the performance stability: the number of trees = 1000 and the random feature selection = 4. Bootstrap resampling was applied to estimate the 95%CIs of R^2 and fitted curve (10,000 iterations).

For the entire processes we created an R script, available at github ("https://github.com/masahiroryo/joint-ES-estimate"). For the visualization, we used R v. 3.4.3 and its packages "ggplot2" (*19*), "ggridges" (*59*), "ggepi" (https://github.com/lwjohnst86/ggepi; v0.0.1.9000), and "patchwork" (https://github.com/thomasp85/patchwork). For the random forest analysis, we used the packages "party" (*60*) and "caret" (*61*).

Analysis of soil fungal diversity

DNA extraction, ITS2 PCR amplification and Illumina sequencing preparation. Soil DNA was extracted from 0.25 g using the PowerSoil DNA isolation kit (MoBio Laboratories Inc., Carlsbad, CA, USA), following the manufacturer's instructions. The fungal ITS2 genomic region was amplified by PCR using the fITS7 (5′-GTGARTCATCGAATCTTTG-3′) and the ITS4 primers (5′- TCCTCCGCTTA TTGATATGC-3′) (*43*), respectively extended in 5′ with the p5 and p7 Illumina sequencing adaptors.
The PCR cycles were as follows: a denaturation step for 3 min at 98°C, 30 cycles of denaturation for 20 s at 98°C, annealing for 20 s at 50°C, and elongation for 30 s at 72°C, and a final elongation step of 5 min at 72°C. PCRs were performed in a 25 µl volume containing 0.5 U of KAPA HiFi polymerase (Kapa Biosystems, Woburn, MA, USA), 1x KAPA HiFi buffer, 0.2 mM of each dNTP, 0.3 µM of each primer, and 2 µl of DNA template. PCR amplification was performed in duplicate and duplicate PCR products of each sample were subsequently pooled. The PCR products were purified using magnetic beads in a 0.8:1 v:v (GC Biotech, Alphen aan den Rijn, The Netherlands). The purified products were used in a second PCR step reduced to 10 cycles with similar cycle set up, with primers containing the sequencing adaptors and a 8 nt long index sequence for multiplex sequencing using 4 µl of DNA. PCR products were again purified with magnetic beads in a 0.8:1 v:v (GC Biotech, Alphen aan den Rijn, The Netherlands). DNA quantification was performed using PicoGreen technology (Invitrogen, Carlsbad, CA, USA) and the final PCR products were pooled on an equimolar basis. The amplicon library was sequenced on an Illumina MiSeq 2000 platform (Illumina Inc., San Diego, CA, USA) at the Berlin Center for Genomics in Biodiversity Research (BeGenDiv, Berlin, Germany) using 2x300-bp paired-end sequencing.

Bioinformatics. Raw reads were demultiplexed allowing no error in the index sequence for sample assignment. We used DADA2 (*44*) to obtain denoised, chimera-free, non-singleton fungal amplicon sequence variants (ASVs). Raw reads containing any ambiguous bases were removed. Primers were removed using cutadapt including the reverse complement sequence of the reverse and forward primer sequence in the forward and reverse reads, respectively. Reads with more than 2 and 5 maximum expected number of errors for forward and reverse reads were excluded. Non-singleton ASVs were inferred on a sample basis. Chimera were identified de novo (sequences that corresponded to subsets of two more abundant sequences) and removed. Taxonomic annotation of ASVs was performed using the Naive Bayesian Classifier (*62*) against UNITE (*63*). ASVs were considered fungal if they were annotated at least at the phylum level at an 80% confidence threshold. We inferred a total 854 fungal ASVs in the 140 samples. After random read subsampling to a common sequencing depth (150 reads), 346 fungal ASVs in 139 samples were included in the analysis to test the effect of increasing

treatment levels on soil fungal community composition and diversity. A complete description of the dataset can be found below.

Statistical analyses of soil fungal diversity. Community compositional change was measured using Bray-Curtis dissimilarity. The effect of factor richness and treatment identity was tested using a permutational multivariate analysis of variance on pairwise Bray-Curtis dissimilarities (*64*). The statistical robustness was assessed using 999 Monte-Carlo permutations where sample assignment to treatments was randomized. The community variability within treatments (single and richness factor experiments) was calculated using multivariate homogeneity of group variance (*65*) and statistical robustness assessed using 999 Monte-Carlo permutations. A linear model was fitted to regress community variability with increasing factor richness. Pairwise Bray-Curtis dissimilarities were visualized using a non-metric multidimensional scaling (NMDS) ordination.

Community taxa richness was measured as the number of ASVs in samples. A linear model was fitted to regress ASV richness with increasing factor richness. We tested whether ASV-poor communities were a subset of the ASV-rich communities, i.e. nestedness. Nestedness was measured using the temperature metric of the presence/absence community matrix, which calculates the number of "surprises" of the absence or presence of ASVs between the observed community matrix (arranged by decreasing sample richness and sample occurrence) and a perfectly nested matrix (*46*). To test the non-randomness of nestedness, we used a null model that pemutes the rows (samples) and columns (ASVs) of the matrix to generate null matrices while maintaining rows and columns totals. 999 null matrices were generated and statistical robustness was assessed using a two-sided test. Sample pairwise nestedness was further calculated and nestedness within and among treatment levels was tested using ANOVA. The difference in sample ranks of the different treatment levels in the lowest temperature matrix was also tested with a Kruskal-Wallis test, followed by a one-sided Wilcoxon test to test whether the rank of samples between two specific treatment levels differ.

All statistical analyses were conducted in R (*18*) using "vegan" (*45*) and base packages. Data visualization was performed using the graphical R packages "gplots" (*66*) and "ggplot2" and its extensions "ggpubr" (*19*).

Supplementary Text

Description of the fungal sequence dataset

From 2813975 raw reads, 854 denoised non-singleton non-chimeric fungal ASVs were inferred, totalling 1546914 reads (55% of the reads retained; per sample: 57% ± 10% mean ± sd) for 140 samples. The major loss of reads happened at the quality filtering step (70% of the reads retained; per sample: 68% ± 8.5%, compared to >85% for any other steps including denoising, paired-end read merging, chimera and non-fungal ESV removal).

Ten phyla were retrieved. ASVs were annotated to 33 classes, 60 orders, 113 families, 156 genera and 139 species, but the proportion of ASVs to be annotated at any level strongly decreased at lower levels. The most strongly represented phylum was Ascomycota (604 ASVs, 68% of the ASVs), then Basidiomycota (149, 17%), followed by Mortierellomycota (32, 3.6%), Glomeromycota (28, 3.2%), Rozellomycota (25, 2.8%), Mucoromycota (21, 2.3%), Chytridiomycota (12, 1.4%), Zoopagomycota (5, <1%) and Olpidiomycota (2, <1%) and Monoblepharomycota (2, <1%). The most abundant phylum is Ascomycota (47.5% of the reads), followed by Mucoromycota (31.6%), Basidiomycota (18.8%) and Mortierellomycota (1.7%). All other phyla were represented by less than 0.1% of the reads. ASV sequence length showed a unimodal distribution that peaks at ~ 250 bp and ranging from 165 bp to 454 bp (265 bp ± 35 bp mean ± sd). Although taxonomy is correlated to sequence length, our coverage of sequence length suggests that there is no taxonomic bias in the dataset.

Accumulation curve of ASVs with addition of samples revealed no saturation, indicating that unique ASVs were found in each sample. Those ASVs were neither the most numerous nor abundant, totalling 0.5% of the reads. Most ASVs were found in two samples, but again totalling 0.5% of the reads. In contrast, there were a few highly abundant ASVs, and they tended to occur in most samples. No ASV was found to occur in all samples but one was found in 139 samples (coded "esv1", absent from sample 130).

The final number of fungal reads per sample was 7649.39± 5201.451 (min:13 in sample 69, max:199530 in sample 72). Sequencing depth and richness were slightly correlated (Pearson's R=0.51). The number of fungal ASVs per sample was 70.5 ± 38 (min: 2 in sample 69, max: 245 in sample 72).

After removing sample 69 (failure to amplify), and randomly resampling to 150 reads per sample to account for sequencing depth, 346 ASVs belonging to 8 phyla (loss of Olpidiomycota and Monoblepharomycota) were

retained and the ASV richness was 29.4 ± 10.4 (min: 6 in sample 137, max: 51 in samples 113 and 52). ASV richness per sample before and after random resampling were well correlated (Pearson's R=0.72). Similarly, Bray-Curtis dissimilarity between samples before and after random resampling were also well correlated (Mantel R=0.66).

We inferred 346 fungal ASVs in 139 samples, after random read subsampling to a common sequencing depth, that we have included in the analysis to test the effect of increasing treatment levels on soil fungal community composition and diversity. Centroid location in a multivariate space clearly differed among treatment levels (PERMANOVA, df=1, F=20.2, R^2=0.11, p-value=0.001) and additionally among treatments (PERMANOVA, df=13, F=3.5, R^2=0.24, p-value=0.001). This indicates that community composition consistently differed with increasing treatment levels and therefore some degree of predictability. However, community composition variability increased with increasing treatment level (df=1, t=5.049, F=25.5, R^2=0.15, p-value=1.39e-06). This indicates that even though increasing treatment levels select for particular phylotypes that differ from the lower treatment levels (notably the single treatments), the abundance and occurrence of those phylotypes tend to vary more when increasing treatment levels, indicating lower predictability and increasing dispersion. Moreover, community variability clearly differed between treatments (df=14, F=5.5, p-value=0.001) but this was less clear among treatments level (df=5, F=2.5, p-value=0.036,) likely because single treatment effect had opposing effect on community composition, at the ASV level and at the phylum level. These results suggest that the effect of multiple factors on soil fungal community composition will not be easily predicted from the effect of single treatment factors.

We also observed that ASV richness strongly decreased with increasing treatment level (df=1, t=-7.311, F=53.4, R^2=0.27, p=2.01e-11). The losses of ASVs were non-random, which is in agreement with the selection of specific phylotypes: notably, ASV-poor communities were constituted of a subset of the ASVs in ASV-rich communities (temperature=7.3, SES=-8.3, p=0.001). Specifically, Basidiomycota are lost with an increasing number of factors while Ascomycota are apparently more stress-tolerant generalists (Fig. S2 and S3). Additional sample pairwise comparisons revealed higher nestedness between samples of different treatment levels than of similar ones (ANOVA F=425.74, df=1, p-value < 2.2e-16), with higher level treatments being nested within lower treatment levels, especially single factor treatments.

Tests on sample ranks of the matrix with the lowest temperature further confirm that ranks of different treatment levels tend to differ (Kruskal-Wallis chi-squared = 11.477, df = 5, p-value = 0.0427) with treatment level 10 being consistently at higher ranks (i.e. nested within other treatment levels) than lower treatment levels (Level 10 versus all others level treatments: Wilcoxon W = 989, p-value = 0.002555).

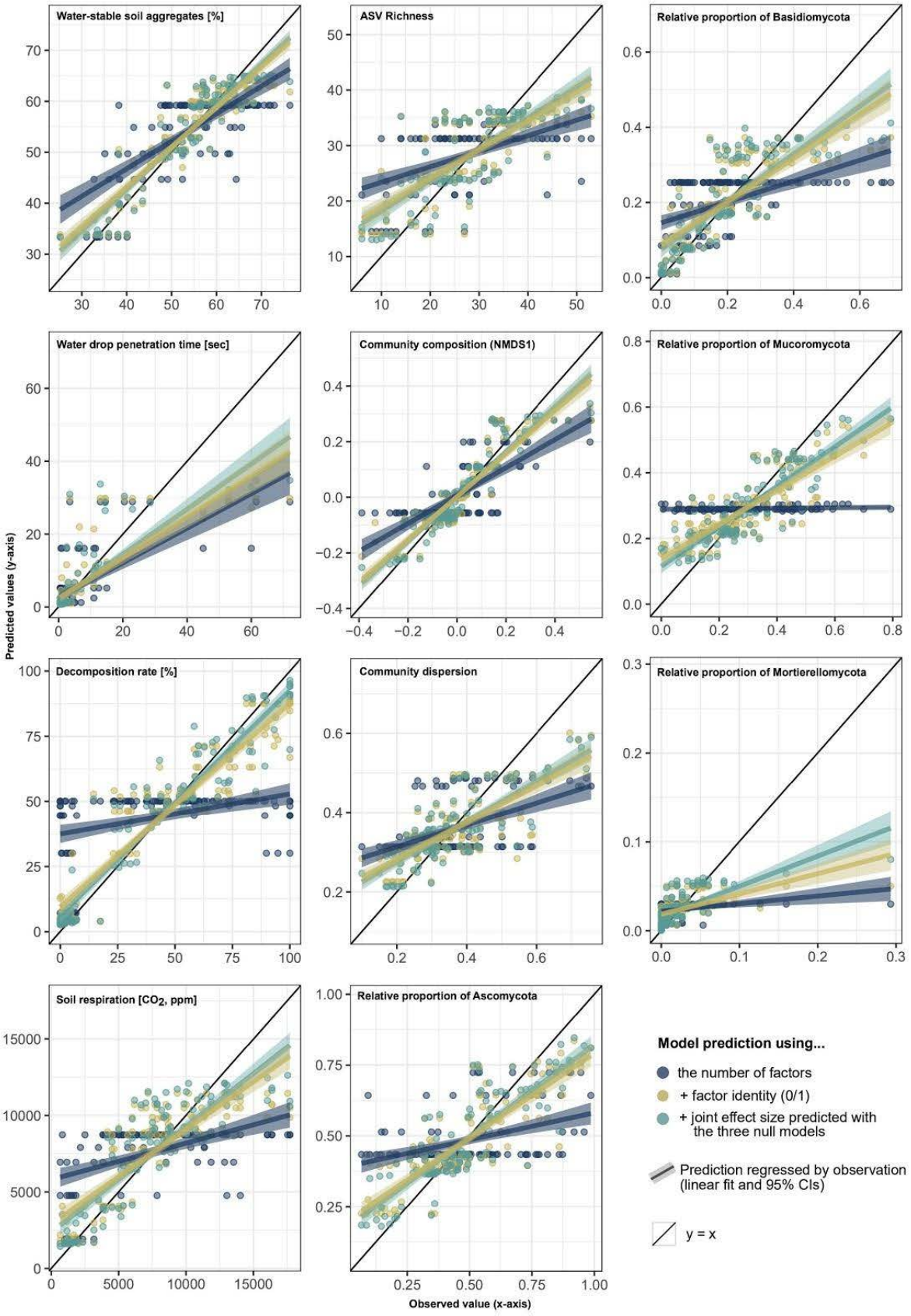
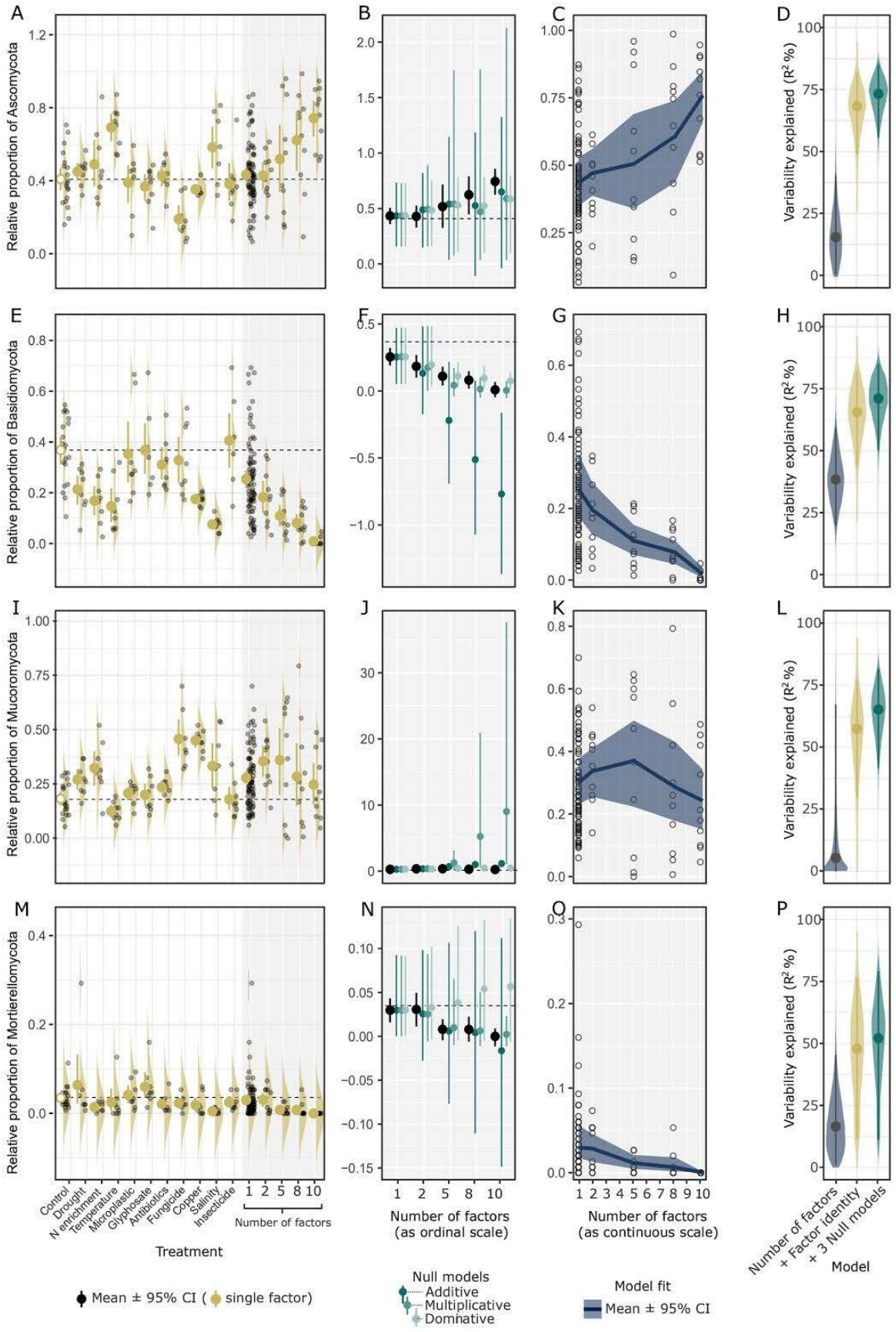


Fig. S1.
Observation vs. prediction. Predictions using a random forest machine learning algorithm based on three different models. Predictions based only on the number of factors (blue) generally show a correlation with observations but the slopes are shallower than predictions with more information (yellow and green). This indicates that general trends can be predicted solely by the number of factors, but additional information (i.e. factor identity or effect size information) is needed for better predicting the severity of the changes.

436



437
438 **Fig. S2.**

439 Effects of different global change factors applied singly and applied in different multiples (2, 5, 8, 10 interacting
440 factors) on the soil fungal community. Relative abundance of fungal taxa at the phylum level. See the caption of
441 Fig. 3 and 4 for explanation of symbols.
442

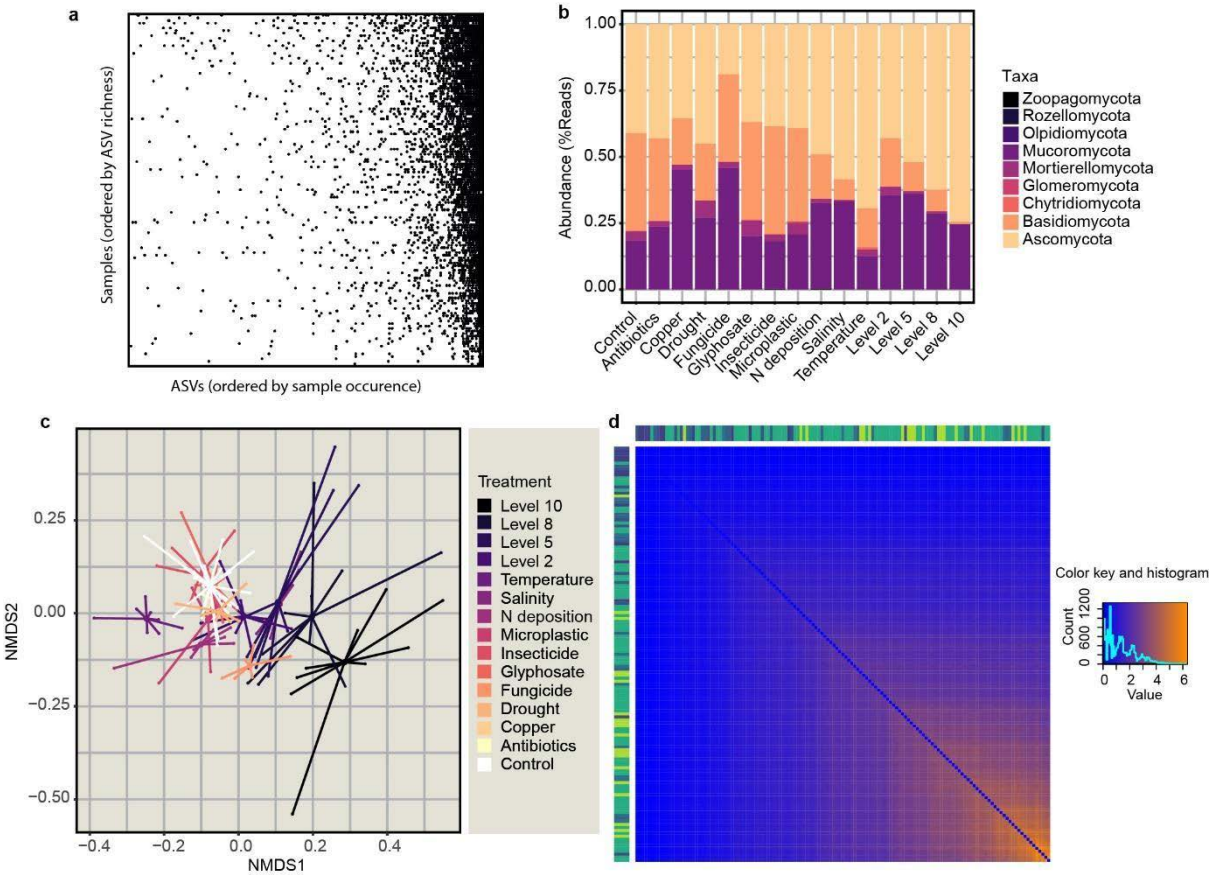


Fig. S3.

Characteristics of the soil fungal community in response to multiple drivers of global change. A. Arranged presence/absence matrix of ASVs in samples; rows from top to bottom are samples arranged by decreasing ASV richness; columns from left to right are ASVs arranged by increasing sample occurrence. B. Relative abundance of fungal taxa (at phylum level) across the single factor treatments and factor richness experiment.C. Non-metric multidimensional scaling (NMDS) of Bray-Curtis sample pairwise dissimilarities. Color represents the treatment applied to samples. Samples are connected to their group centroid by lines. In the main analysis, community composition and dispersion were represented by the scores on the 1st axis of the NMDS plot and the mean value of Euclidean distances from all replicates to the group median on the NMDS plot, respectively. D. Heatmap of nestedness between samples. The samples in the square matrix are ordered by increasing ASV richness. The colors in rows and columns represent number of factors applied to samples. The colors within the heatmap represent nestedness from low (orange) to high (blue).

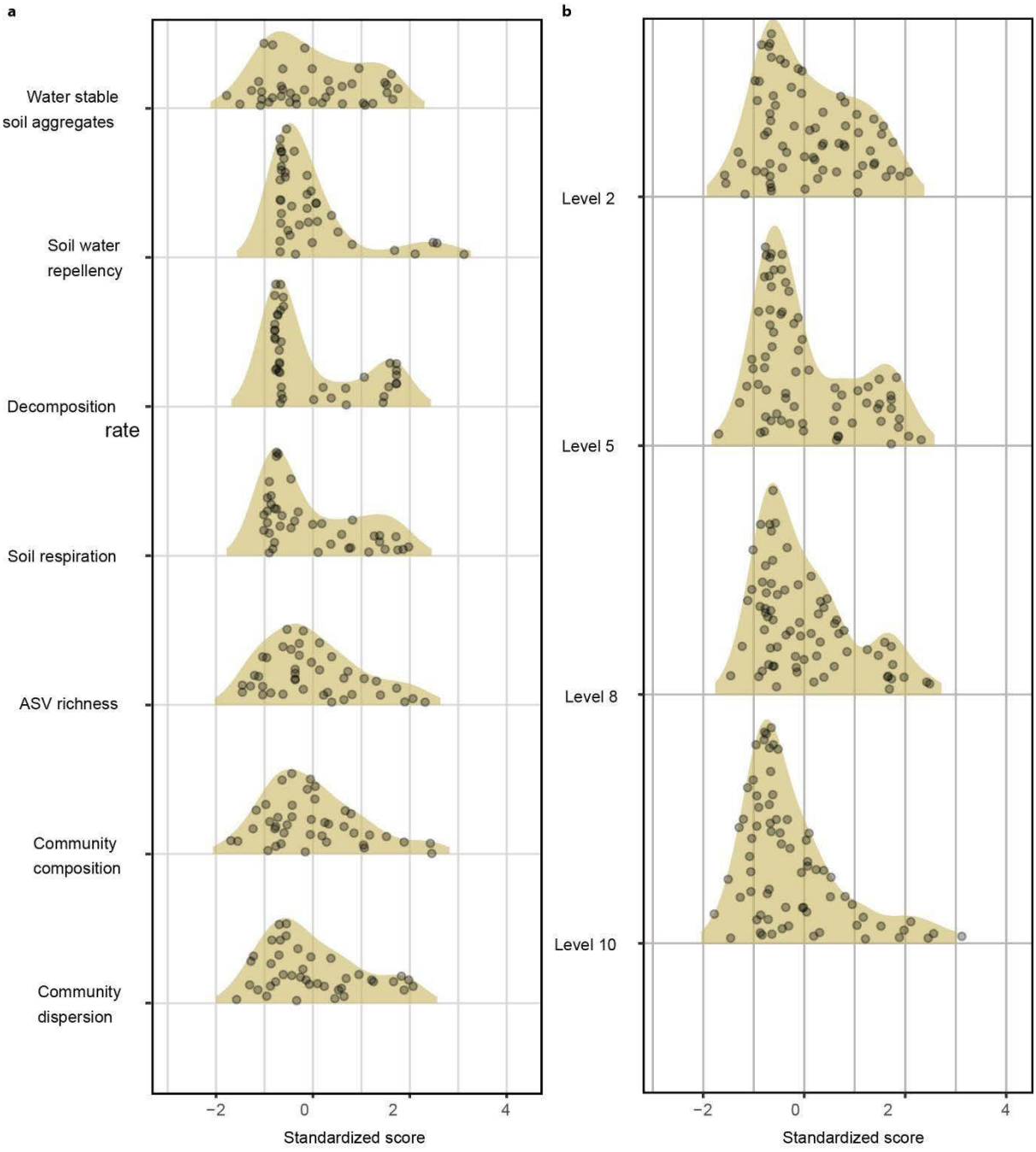


Fig. S4.
Bimodal-like distributions of the standardized measured response variables. These might indicate regime shifts caused by the interaction of multiple factors: the distribution for each variable (and all level data are combined) in the left panel, and the distribution for each level (and all variables are combined) in the right panel. Observing a bimodality together with shift in an internal system state (e.g. community composition and dispersion; Fig. 4) is often thought to be a reasonable indication of regime shifts. Yet, we did not directly demonstrate such a regime shift, as our experiment was not designed for this (e.g. hysteresis and positive feedback are considered direct support for the existence of a bistability state).

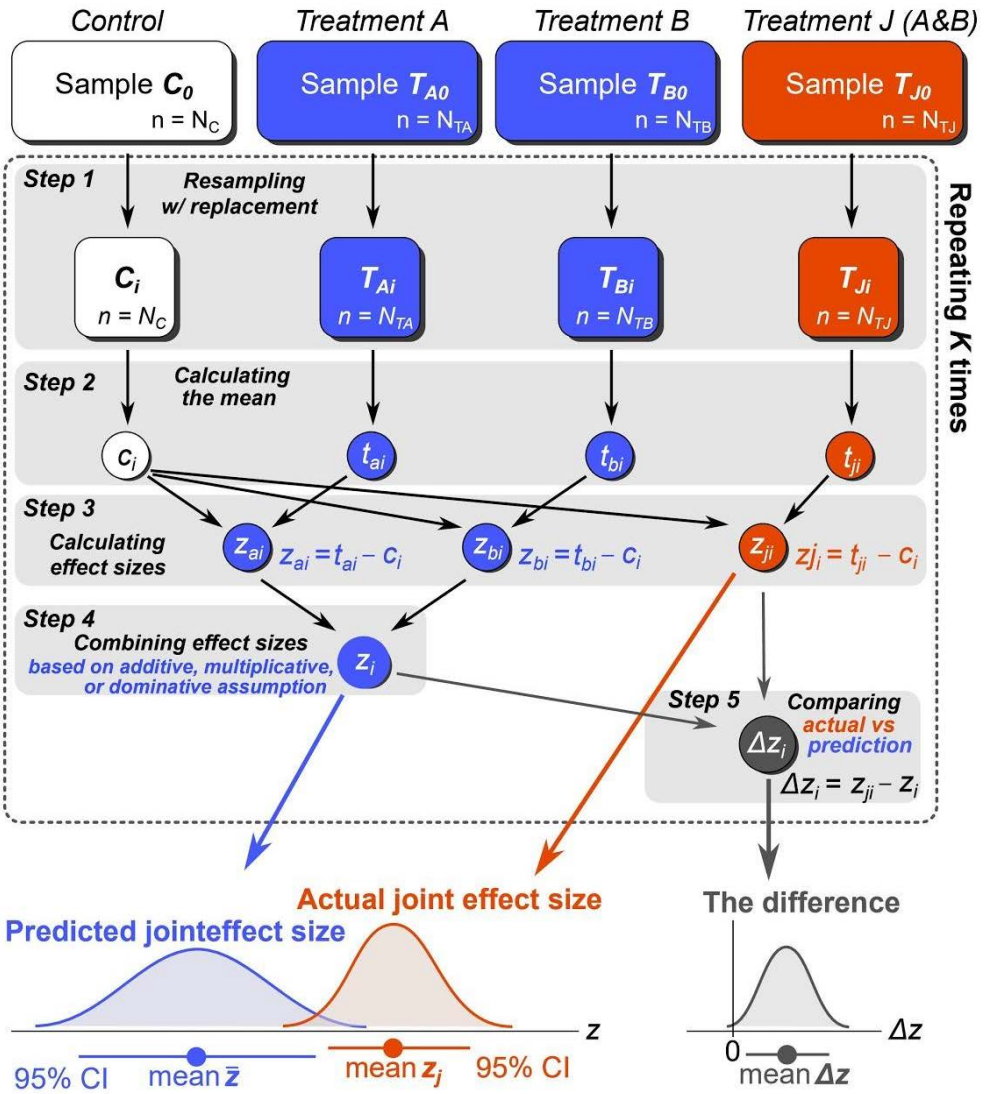


Fig. S5.

Bootstrap resampling scheme for estimating the joint effect size at each level of factor number. For brevity, as an example, only factors A and B are shown to represent a replicate in level 2, without losing generality. In Step 1, for each of control, treatments A, B, and J (J: actual joint effect of A and B), it performs sub-sampling with replacement to generate C_i, T_{Ai}, T_{Bi}, and T_{Ji}, respectively, where i represents i-th iteration (i = 1, 2, …, K). In Step 2, for each, it calculates the mean (c_i, t_{ai}, t_{bi}, and t_{ji}). In Step 3, it calculates the effect size of each treatment (z_{ai}, z_{bi}, z_{ji}) as an absolute difference from the mean of the control. In Step 4, the effect sizes of single factors are combined to generate a predicted effect size z_i: how to combine them depends on assumption type (additive, multiplicative, or dominative). In Step 5, the difference between the predicted effect size z_i and the actual joint effect size z_{ji} is calculated as Δz_i. Steps 1–5 are repeated K times to draw the bootstrap distributions of the actual joint effect size, the predicted joint effect size, and the difference. Finally, the mean and 95% confidence intervals (CIs) were taken from these distributions, and a p-value was calculated from the difference distribution Δz, estimating the probability that the difference between the actual joint effect size and the predicted joint effect size crosses over zero by chance.

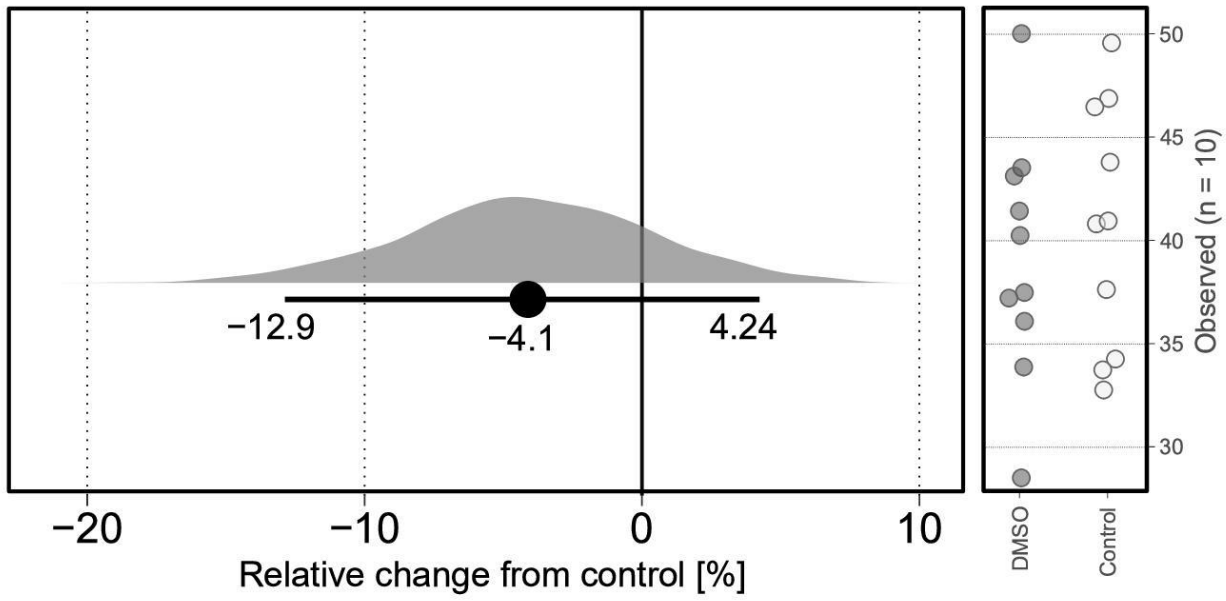


Fig. S6.
Results of separate experiment to test for effects of DMSO on soil aggregation (water stable soil aggregates). Shown are bootstrapped difference in mean and 95% confidence interval. Raw data (n = 10) are in the right panel: there was no difference between control and DMSO treatment. DMSO is used as a solvent and was added to all experimental units at the same rate.

Table S1.

The effect size estimates and their predictability tests along factor richness. For each response variable in each level of factor richness, a p-value was estimated under each assumption, quantifying the probability that the mean of the actual effect size differs from the assumption by chance.

Variables	Target	Level 2				Level 5				Level 8				Level 10			
		CI 2.5%	Mean	CI 97.5%	p value	CI 2.5%	Mean	CI 97.5%	p value	CI 2.5%	Mean	CI 97.5%	p value	CI 2.5%	Mean	CI 97.5%	p value
Water Stable Aggregates	Actual	-10.0	-4.7	0.6		-16.3	-9.9	-3.3		-20.5	-15.0	-8.8		-29.6	-26.2	-22.9	
	Additive	-13.8	-1.3	12.4	0.3800	-23.6	-2.7	17.2	0.2884	-23.4	0.7	24.4	0.1053	-27.2	-2.2	21.6	0.0307
	Multiplicative	-13.0	-1.3	13.1	0.3988	-20.7	-2.8	18.1	0.3052	-20.9	0.0	27.1	0.1066	-23.6	-3.3	21.7	0.0125
	Dominative	-12.3	-1.9	9.3	0.4582	-12.8	-2.1	10.0	0.2272	-13.2	-2.9	10.4	0.0412	-13.6	-4.8	10.6	< 0.0001
Water drop penetration time	Actual	-0.2	1.3	3.5		1.0	4.1	7.7		4.0	14.2	27.2		16.8	29.6	44.3	
	Additive	-0.9	0.5	3.4	0.2318	-2.0	1.1	4.2	0.1083	-2.3	1.8	4.9	0.0076	-1.8	2.2	5.2	< 0.0001
	Multiplicative	-0.8	0.4	3.2	0.2119	-1.1	1.2	8.0	0.1296	-1.2	2.9	18.8	0.0711	-1.2	4.3	27.8	0.0253
	Dominative	-0.5	0.5	3.4	0.2437	-0.6	1.4	3.8	0.1227	-0.5	2.1	3.9	0.0054	1.7	2.8	4.0	< 0.0001
Decomposition rate	Actual	-28.2	-5.0	18.6		-33.5	-4.6	25.1		-45.2	-19.6	9.5		-54.3	-45.8	-37.4	
	Additive	-61.5	-7.4	36.0	0.4796	-83.9	4.0	96.8	0.4590	-88.7	6.5	114.3	0.3269	-84.9	4.0	92.8	0.1386
	Multiplicative	-53.0	-11.6	40.3	0.4356	-55.2	-0.4	168.4	0.3938	-56.4	-9.9	220.0	0.3045	-57.4	-42.2	-6.7	0.4569
	Dominative	-53.5	-8.8	43.6	0.4799	-54.1	-6.3	49.1	0.4514	-54.9	-7.0	50.5	0.4598	-55.6	-23.4	50.7	0.4994
CO2 concentration at third week	Actual	-4968.3	-2741.9	-609.5		-6597	-3394	-142		-8342	-5423	-2021		-9721	-8572	-7465	
	Additive	-8498.5	-3736.1	1373.6	0.3637	-18406	-6882	4927	0.3026	-24391	-11635	2586	0.2031	-26473	-14269	-2425	0.1816
	Multiplicative	-7661.5	-3826.0	1080.3	0.3352	-9898	-5319	4902	0.2570	-10505	-7326	876	0.2145	-10776	-8653	-5100	0.4208
	Dominative	-7465.0	-3563.3	3456.9	0.3463	-7711	-4181	4171	0.3121	-7828	-5205	3958	0.4570	-7974	-6441	-5114	0.0131
ASV Richness	Actual	-6.2	-0.1	5.9		-18.6	-12.0	-4.5		-20.6	-13.7	-6.2		-26.9	-21.8	-16.5	
	Additive	-21.8	-6.1	9.2	0.2686	-54.3	-22.8	4.3	0.2583	-67.3	-34.0	-0.7	0.1241	-78.5	-40.7	-3.3	0.1664
	Multiplicative	-19.0	-5.7	9.5	0.2651	-31.2	-17.7	2.0	0.2462	-33.6	-23.3	-5.5	0.1114	-35.2	-25.4	-8.4	0.2551
	Dominative	-16.1	-5.2	7.6	0.2699	-17.9	-11.7	5.4	0.4898	-18.2	-13.0	-8.0	0.4652	-18.6	-13.3	-8.2	0.0134
Community composition	Actual	0.0	0.1	0.1		0.099	0.184	0.270		0.187	0.273	0.374		0.187	0.273	0.374	
	Additive	-0.2	0.1	0.2	0.4190	-0.3	0.1	0.4	0.3146	-0.2	0.2	0.6	0.3136	-0.2	0.2	0.6	0.3136
	Multiplicative	-0.2	0.1	0.3	0.4814	-0.6	0.0	0.7	0.1014	-0.2	0.1	0.4	0.0421	-0.2	0.1	0.4	0.0421
	Dominative	-0.2	0.0	0.2	0.4052	-0.2	0.0	0.2	0.2465	-0.2	0.0	0.2	0.0287	-0.2	0.0	0.2	0.0287
Community dispersion	Actual	-0.1	-0.1	0.0		-0.051	0.067	0.196		-0.016	0.073	0.161		0.018	0.096	0.179	
	Additive	-0.3	-0.2	0.0	0.1080	-0.7	-0.4	-0.2	0.0006	-1.0	-0.7	-0.3	< 0.0001	-1.3	-0.9	-0.5	< 0.0001
	Multiplicative	-0.3	-0.2	0.0	0.1108	-0.4	-0.3	-0.1	0.0007	-0.4	-0.3	-0.3	< 0.0001	-0.4	-0.4	-0.3	< 0.0001
	Dominative	-0.2	-0.1	0.0	0.1406	-0.2	-0.2	-0.1	0.0033	-0.2	-0.2	-0.1	0.0002	-0.2	-0.2	-0.2	< 0.0001

Table S2.
The way forward. We here outline some key next steps for future experiments building on the approach used here, providing an explanation for why these steps are important, and we explain how our study specifically set the stage and could be built upon.

Next steps	Explanation	How our study set the stage
Higher degree of realism and system complexity	Carry out field experiments (potentially in various ecosystem types) that also include other organism groups, such as plants and larger soil fauna (e.g. earthworms)	We show here that patterns and trajectories emerge quite clearly with a relatively low number of replicates (140 total); this is a clear advantage, since field studies are in principle logistically feasible with this level of replication (compared to many hundred plots)
Increased mechanistic resolution	Our study design was not optimized for detecting specific factor interactions (which is what factorial designs are for)	The design used here could be combined with a factorial design. The factorial design would include a few factors (perhaps only) of special interest. This factorial design could then be repeated on the background of all remaining factors either being toggled on or off. An equivalent idea could be use also for fractional factorial designs
Dealing with multiple levels of one factor	We here used one level for each factor (e.g. 50% of water for drought), informed by literature values or scenarios, as is typical of much global change biology research.	The design we used could be combined with a design that tests multiple levels of a factor, for one or a few selected factors. For example, other factors that are not in the focus of such tests could be collectively toggled on or off.
Detecting important interactions	Using machine learning, we could not specify which combinations of factors are important.	Advanced machine learning methods might be able to identify important higher order interactions in studies such as ours, once they have been tested on small datasets (*67*, *68*)

Data S1 (separate file)
Data for literature synthesis presented in Fig. 1 (panel A - C is covered by the dataset "multifactorialty" and panel D by "network analysis")