A Web of Science search was performed on 15th October 2018 using a combination of search terms that captured groups of soil organisms AND metrics of biodiversity AND global change drivers. Within the global change drivers, the six groups were represented (OR).

**The combination of search terms was as follows:**

TS=( (((soil OR below$ground OR below-ground) AND (biota OR fauna OR micro$fauna OR macro$fauna OR meso$fauna OR animal\* OR arthropod\* OR invert\* OR “inverte\* decomposer\*” OR detritivore\* OR macroarthropod\* OR rotifer\* OR mite\* OR acari\* OR protozoa\* OR tardigrad\* OR isopod\* OR protist\* OR micro-arthropod\* OR microarthropod\* OR ciliat\* OR termit\*)) OR nematod\* OR oligochaet\* OR annelid\* OR collembol\* OR springtail\* OR earthworm\* OR enchytrae\* OR lumbricid\* OR “soil biodiversity” OR “below$ground biodiversity” OR “soil divers\*” OR “below$ground divers\*”))

AND

TS=(“species richness” OR richness OR “number of species” OR “number of taxa” OR diversity OR Shannon\* OR evenness OR abundance OR density OR communit\* OR biomass)

AND

(TS=(invad\* OR exotic OR alien OR invas\* OR non$native OR peregrine OR introduc\* OR non$indigenous)

OR

TS=(“nitrogen deposition” OR “nutrient deposition” OR “atmospheric deposition” OR \*eutroph\* OR fertili\* OR “nutrient\* enrichment” OR “nutrient pollut\*”)

OR

TS=(pollut\* OR contamin\* OR toxi\* OR metal\* OR asbestos OR radionuclide\* OR radioactiv\* OR pharmaceutic\* OR “emerging contamin\*” OR “synthetic organic chem\*” OR “personal care product\*” OR plastic\* OR “polycyclic aromatic hydrocarbon\*” OR pesticide\* or herbicide\* or fungicide\* or molluscide\* or nematicide\* or insecticide\* OR agrochemical\* OR “oil spill” OR “brine spill” OR “petrol spill” OR mining OR smelting OR “industrial activit\*” OR “waste disposal” OR wastewater OR sludge OR sewage)

OR

TS = (“climat\* change” OR drought OR temperature\* OR warming OR heat\* OR precipitation\* OR rain\* OR flood\* OR irrigation OR moisture OR watering OR fire OR “carbon dioxide” OR CO2)

OR

TS=(“habitat loss” OR “habitat fragment\*” OR “edge effect\*” OR fragment\*)

OR

TS=(land-use OR landuse OR land use OR agricultural intensi\* OR forest\* OR agricultur\* OR grassland\* OR pasture\* OR meadow\* OR agroforest\* OR plantation\* OR urban\* OR farm\* OR abandon\* OR fallow\* OR graz\* OR arable OR till\* OR ploug\* OR habitat degrad\* OR habitat destruct\* OR logg\* OR deforest\* OR ((land use OR landuse OR land-use OR cropland OR agricultur\*) AND (intensi\* OR expansion)) )

OR

TS=(“global change” OR “environmental change” OR disturbance\* OR stress\* ) )

**Abstract screening**

The search retrieved 25,591 records, that were further filtered to remove obviously irrelevant papers using the Web of Science categories (MARINE FRESHWATER BIOLOGY OR OCEANOGRAPHY OR LIMNOLOGY were excluded) and non-primary research articles (REVIEW OR MEETING ABSTRACT OR NOTE OR NEWS ITEM OR RETRACTED PUBLICATION OR BIBLIOGRAPHY OR DATABASE REVIEW). Resulting in 24,979 records considered for screening.

Titles and abstracts were screened for suitability. Papers were suitable if they (1) investigated the effect of at least one GCD on at least one group of soil fauna and (2) contained at least one reference, undisturbed, or control site, and one site impacted by the GCD.

We used a machine learning algorithm in the program Abstrackr to assist the screening process (Wallace et al. 2012). Abstrackr assigns confidence scores to all papers based on the screened abstracts and titles. This is dynamic process through the screening stage. We manually screened 9,535 abstracts (6,143 were irrelevant and 3,389 were included), until the Abstrackr confidence score was 0.58 or under for the remaining 15,444 articles. The cutoff value of 0.58 was chosen based on a quality control procedure in which we randomly sampled 5% of the records within each 0.1 band of confidence scores, and screened their titles to check that they ‘may be’ suitable or were “definitely not” suitable. The cut-off confidence score was then based on the point where the number of ‘definitely not’ suitable papers was the majority of the titles within each 0.01 band.

**Full-text screening**

The full text of 3,389 papers were manually screened. In order to be suitable the article needed to have (1) measured at least one soil fauna group (i.e., earthworms, macrofauna, oribatid mites), (2) captured the impact of one or several GCDs according to our GCD-specific inclusion criteria (see below) and (3) present the data (mean values, variance etc.) allowing us to calculate an effect size for the meta-analysis.

1. Soil fauna groups

We used the sampling methodology as criteria for defining soil/below-ground fauna. Suitable sampling methods could include soil cores, excavated soil blocks, or mustard extraction. Methods that only sampled the litter fauna were considered not suitable, i.e. studies reporting aboveground invertebrate sampled from pitfall traps only. If the pitfall traps were associated to another method targeting the soil (soil core, pit and quadrat), they were included.

In addition, studies where the soil fauna group were manipulated, i.e., removal or additional of species, were excluded.

1. Global change

**Land-use intensification** studies report soil biodiversity in sites with increasing intensity of use (mechanical, tillage, agricultural or forestry practices, etc.) within one land cover system (i.e. forest, grassland, cropland etc.). Changes in fertilization rates and pesticides are however classified in other GCD categories (nutrient enrichment and pollution).

**A land-use change** study had to contrast sites with different land cover/use. For example, a grazed grassland versus a forest versus an urban site. It could also report changes in soil biodiversity before versus after a change in land use (e.g. deforestation studies). We collected this information but currently not plan to integrate those studies to the global meta-analysis.

**Habitat fragmentation and loss** studies had sites in habitats (of any system) that change in their size, isolation or disturbance. This covered changes in habitat amount across sites, changes in distance to the edge of the habitat across sites, and changes in the connectivity of the habitat to other habitats.

**Chemical pollution** studies compared sites with varying levels of chemical concentrations (as documented/quantified in the paper itself or in a related cited publication). Importantly, studies investigating pollution by nutrient only (fertilization, nitrogen deposition etc.) were classified in the nutrient enrichment category.

**Nutrient enrichment** studies compared sites presenting contrasting levels of either organic (manure, mulch) or inorganic (N, P, K, any mixture of them, Ca, S).

**Climate change** studies covered a range of treatments (gas (N2O, CO2, CH4), fire, precipitation, temperature, any mixture of them as well as extreme climate events). Changes in the intensity, frequency or amount of the treatment were all considered.

**Invasive species** studies investigated the impacts of invasive plants, animals (both above- and belowground), and pathogens by comparing sites at different invasion stages.

**Multiple drivers** whenstudies combined several of these drivers (either separately or in a full- factorial design crossing all the drivers), they were classified into all of the drivers investigated in the paper.

1. Presence of data

Data could be presented within tables, figures or the text. Studies containing means, or ways to calculate means, and variance (standard deviations, standard errors, confidence intervals etc.) were suitable. Studies reporting only multivariate statistics (PCA etc.) were excluded.

At the full-text screening step, we also collected information about each included paper, classifying the studies according to global change drivers and taxonomic groups, as well as documenting the country and the community metric measured (i.e., richness, abundance an/or biomass).

**References**

Byron C. Wallace, Kevin Small, Carla E. Brodley, Joseph Lau and Thomas A. Trikalinos. **Deploying an interactive machine learning system in an evidence-based practice center: abstrackr**. In *Proc. of the ACM International Health Informatics Symposium (IHI)*, p.819--824. 2012.