Metrics based on habitat area and condition are poor proxies for invertebrate biodiversity

Authors: **Duffus, Natalie E. [1*^], Atkins, Thomas B. [1^],** zu Ermgassen, Sophus O.S.E. [1], Grenyer, Richard.[2], Bull, Joseph W. [1], Castell, Dan A. [2], Stone, Ben. [2], Tooher, Niamh. [2], Milner-Gulland, E.J. [1], Lewis, Owen T.[1]

* Corresponding author: natalie.duffus@biology.ox.ac.uk

^ Equal first author

Affiliations:

1. Department of Biology, University of Oxford
2. School of Geography and the Environment, University of Oxford

Abstract

There is increasing demand for standardised, easy-to-use metrics to assess progress towards achieving biodiversity targets and the effectiveness of ecological compensation schemes. Biodiversity metrics based on combining habitat area and habitat condition scores are proliferating rapidly, but there is limited evidence on how they relate to ecological outcomes. Here, we test the relationship between the statutory biodiversity metric used for Biodiversity Net Gain (BNG) in England — and as the basis for new biodiversity credit systems around the world — and invertebrate richness, abundance, and community composition. We find that the combined area-condition BNG metric does not capture the value of arable farmland and grassland sites for invertebrate biodiversity: invertebrate communities were highly variable across sites that had the same type and condition under the BNG metric. There was no significant relationship between invertebrate abundance or species richness and metric scores. Our results highlight the need to incorporate factors beyond habitat type and condition into site evaluations, and to complement metric use with species-based surveys.

Introduction

Increased conservation efforts are required to prevent further declines in biodiversity [1,2]. Alongside efforts to promote nature recovery, there has been a proliferation of ecological compensation initiatives seeking to enable infrastructure development while compensating for ecological harms [4]. These range from regulations mandating a No Net Loss (NNL) or Net Gain in biodiversity, to voluntary commitments to improve biodiversity by purchasing biodiversity credits [4,5]. Many countries are also developing strategies for financing biodiversity conservation to achieve the aspirations of the Kunming-Montreal agreement and other policy commitments [6]

Such efforts and mechanisms require robust biodiversity metrics to set goals, track progress, estimate risk, and evidence success. Biodiversity metrics based on combining the area and condition of habitats are proliferating (Table 1) and have the advantage of being relatively simple to calculate [5,7].

Table 1. Examples of combined area-condition biodiversity metrics including metrics derived from the England BNG metric.

Metric	**Calculation**	**Application / Use**
Habitat Hectares [8]	Habitat extent * habitat condition (assessed using a condition benchmark)	For Victoria's Native Vegetation Framework in Australia.
SD VISta Nature Framework [9]	Habitat extent * habitat condition	To evidence a positive biodiversity outcome to generate a Nature Credit
Composite Biotope Value [10]	Biotype + structural attributes + Red List attributes + management attributes – threats	To measure the progress of restoration activities against a benchmark in Austria
Biometric [11]	habitat condition * restoration potential * habitat extent	As part of the Biodiversity Banking assessment methodology in New South Wales, Australia
Taskforce for Nature-related Financial Disclosures Metrics [12]	Includes core and additional metrics on ecosystem extent and ecological condition	For organisations to measure impacts on nature and identify opportunities to improve it

Quality Hectares [13]	Habitat extent * habitat quality (context-specific markers of quality used)	To measure progress towards a net positive impact on biodiversity in Madagascar, and influencing multiple other projects
Statutory Biodiversity Metric [14]	Habitat extent * habitat distinctiveness * habitat condition	To evidence mandatory 10% Biodiversity Net Gain (BNG) in England's planning system
Combined area-condition metrics which are direct derivatives of the England BNG metric		
CLIMB metric [15]	Used to measure baseline biodiversity value, estimate compensation requirements and report on biodiversity enhancements in Sweden	
Singapore Biodiversity Accounting Metric [16]	Used to demonstrate biodiversity value of sites pre- and post- development and calculate biodiversity gains and losses in Singapore	
Wallacea Trust Biodiversity Credit Methodology [17]	Allows user to choose a basket of metrics, one of which is the England Statutory Biodiversity Metric. Used to generate evidence of positive biodiversity outcomes which translate into biodiversity credits	
Americas Biodiversity Metric [18]	A decision-making tool that quantifies biodiversity and can be used to evidence No Net Loss or Net Gain. Currently applies to North-Eastern USA but it is anticipated that it well be applied to all US and Canadian habitats	
Scotland Biodiversity Metric [19]	There is ongoing work to adapt the England metric for use in Scotland planning system to evidence biodiversity enhancements from development	

Condition-area metrics are therefore becoming fundamental to how society and nature markets measure biodiversity. Unless these metrics are reliable proxies of wider biodiversity, policy and nature markets will fail to improve it, or may even be detrimental. For example, in Victoria (Australia), the biodiversity offsetting system delivers limited additionality, with real-world biodiversity declining even in situations where the area-condition metric used to measure it increased [20].

Despite widespread concern that area-condition metrics may not provide good proxies for biodiversity, risking sub-optimal ecological outcomes [20–26], studies investigating the relationship between combined area-condition metric scores and biodiversity estimated using primary data across multiple taxa are lacking. Here, we test whether scores from a combined area-condition metric correlate with measures of ground

invertebrate biodiversity based on samples collected using pitfall traps. Ground invertebrates are taxonomically and functionally diverse, with short generation times [27,28] and limited mobility compared to more frequently-studied groups such as birds, mammals or butterflies. These factors make them an especially useful focal group, as their measured richness, abundance and species composition will respond at the fine spatial scales relevant to the application of area-condition metrics. We focus on the metric for Biodiversity Net Gain (BNG) in England, which is used to evidence whether developments provide a mandatory 10% BNG [14,29]. The England BNG metric calculates biodiversity 'units' before and after development by multiplying site size by a biodiversity score that is itself derived by multiplying a score for habitat distinctiveness by a score for habitat condition. Not only does the BNG Metric in England underpin a major national policy goal, but it is highly influential worldwide, with derivatives emerging globally (Table 1). Therefore, its performance has direct relevance to the broader question of how effective combined area-condition metrics are at representing biodiversity.

Results

*Invertebrate abundance and distinctiveness*condition score*

We found no evidence for a relationship between a habitat parcel's distinctiveness*condition score calculated using the BNG metric and the abundance of ground invertebrates, for grassland and arable habitat parcels surveyed in August and September 2022 (Fig 1a. $z_{1,11}$= -0.86, p=0.39). In 2023, using a more efficient pitfall trap design and surveying a different set of habitat parcels encompassing a wider range of grassland and arable habitat types, there was similarly no significant relationship between distinctiveness*condition score and mean total ground invertebrate abundance (May: Fig 1b. $z_{1,16}$= 1.86, p=0.06, June: Fig 1c. $z_{1,18}$=0.99, p=0.32, August: Fig 1d. $z_{1,16}$=0.16, p=0.88). There was also no significant relationship in either year between the abundance of the dominant groups in the samples (carabids and spiders) and distinctiveness*condition scores (p>0.1).

Figure 1. Mean ground invertebrate abundance and distinctiveness*condition scores over four weeks in 2022 (A), and three one-week periods across three months in 2023 (B,C,D).

*Species richness and distinctiveness*condition score*

In 2022 there was no significant relationship between rarefied carabid beetle richness and distinctiveness*condition score (Fig 2a. $z_{1,11}$= -1.28, p=0.20). In 2023, there was no significant relationship between total rarefied ground invertebrate species richness quantified by DNA metabarcoding and distinctiveness*condition score (Fig 2b. $z_{1,15}$= 0.53, p=0.59). There was also no significant relationship between total rarefied Operational Taxonomic Unit (OTU) richness and distinctiveness*condition score ($z_{1,15}$=0.54, p=0.60).

Focusing on individual taxonomic groups, there were no evidence of relationships between distinctiveness*condition scores and species richness quantified using DNA metabarcoding for spiders (Fig 2c. $z_{1,17}$=0.82, p=0.41), carabid beetles (Fig 2d. $z_{1,17}$=1.19, p=0.23), rove beetles (Fig 2e. $z_{1,17}$=0.15, p=0.88), Collembola (Fig 2f. $z_{1,17}$ =-0.57, p=0.56) or Stylommatophora ($z_{1,17}$= 0.39, p=0.69).

Figure 2. Ground invertebrate species richness and distinctiveness*condition scores for carabid beetles (A), total ground invertebrate species richness (B), spiders (C), carabids (D), staphylinids (E) and Collembola (F). Data for (A) are from 2022 sampling, and (B to D) from 2023.

Community composition

A 3D Nonmetric multidimensional scaling (NMDS) analysis of the number of family-level OTUs derived from DNA metabarcoding in each habitat parcel shows the clustering of invertebrate communities by habitat type in May, June, and August 2023 (Fig 3a-c). The ordination revealed almost complete overlap of communities drawn from habitats with differing distinctiveness scores under the BNG metric. In August 2023, there is some evidence of separation between lower and higher distinctiveness habitats, although not between habitat types within those distinctiveness groupings.

Figure 3. Nonmetric multidimensional scaling of family-level Operational Taxonomic Units (OTUs) from 2023. The least stress solution was 0.09 for May (A), 0.12 for June (B), 0.13 for August (C).

Species of conservation concern

We recorded 13 species which are nationally scarce in the United Kingdom, including Biodiversity Action Plan (UKBAP) species. These were found uniformly across a range of habitat parcels, from ‘low’ to ‘very high distinctiveness’ and ‘poor’ to ‘good condition’ (Supplementary material 1).

Discussion

Combined area-condition biodiversity metrics are widely applied proxies for evidencing improvements to biodiversity, and their use is proliferating rapidly as demand for easy-to-use biodiversity quantification methods rises to satisfy many policy and business objectives (Table 1). Here we fail to find evidence that a prominent combined area-condition metric (the BNG metric) is related to the abundance and species richness of ground invertebrates. If use of these metrics result in trades of large, low-scoring sites for smaller, high-scoring sites (as observed in practice: Rampling et al. 2023) then wider biodiversity benefits may not be achieved. In England, harm to high-scoring sites such as lowland meadows is strongly disincentivised, and where it does occur it requires bespoke compensation to be agreed with local authorities. This is not the case for lower-scoring sites, and if the metric is the only measure of biodiversity value considered, our results therefore imply regular undervaluing of sites of high value for invertebrates, which is likely also to be an issue for other components of biodiversity.

If, as we found here, combined area-condition metrics are not reliable proxies for wider biodiversity value, sites with high faunal biodiversity value but low metric scores will be undervalued in offsetting calculations. In this scenario, a low baseline metric score would mean fewer offsets would be required than would be necessary to balance the environmental damage caused by a development requiring ecological compensation. For voluntary biodiversity credit initiatives, an inaccurate proxy metric would lead to investment in credits with uncertain or even overinflated biodiversity benefits. If such biodiversity credits become central to how businesses claim to achieve ‘nature-positive’ objectives (i.e. claiming they deliver more improvements to biodiversity than damage) or address their nature-related risks under prevailing Corporate Sustainability Reporting frameworks and policies (e.g. the Taskforce for Nature-relate Financial Disclosures or the EU Corporate Sustainability Reporting Directive), using a

metric that doesn't reliably capture faunal diversity risks creating the illusion of biodiversity gains [30]. Overestimating the biodiversity benefits delivered in these systems could have negative repercussions. In the case of the voluntary carbon market, evidence that carbon offset projects were not delivering as much gain as promised [31] led to a 56% reduction in the volume of the voluntary carbon market the following year (Forest Trends Ecosystem Marketplace 2024). Robust biodiversity metrics are essential to avoid overpromising and underdelivering, to prevent investors losing confidence in funding biodiversity conservation through biodiversity credit systems [33].

Our results are also concerning given evidence of declining populations of ground invertebrate species [34–36], their contribution to important ecosystem functions such as nutrient cycling, pest control, and decomposition [37–39], and their role as food resources for taxa such as farmland birds [40]. Recent analyses suggest that the BNG metric is also a poor proxy for birds and other vertebrates of conservation concern [23,25], although the higher mobility and larger territories and home ranges of most vertebrates make it less likely that patterns in their abundance and diversity align with BNG metric scores in fine-grained, heterogeneous landscapes such as much of lowland England, where parcel sizes or habitat units for which BNG scores are calculated are small. Given the importance of the ecosystem functions fulfilled by ground invertebrates and their significance in food webs, optimising combined area-condition metrics to better reflect their habitat requirements could trigger better habitat protection and compensation for other taxa. Indeed, certain groups of ground invertebrates, such as ground beetles, are good proxies for the diversity of other invertebrates, vertebrates, and plants [41]. Conversely, optimisation of metrics for vertebrate requirements may possibly result in outcomes antithetical to invertebrate biodiversity, where conservation planning assumes vertebrates will have an ‘umbrella effect’, without accounting for patterns of invertebrate richness and distribution [42].

Like many other combined area-condition metrics, the BNG metric takes a habitat-based approach [43]. However, our results indicate that the habitat type classification used in the metric does not predict invertebrate biodiversity, with high variability between sites of the same habitat type. The BNG metric is underpinned by the UK Habitat Classification (UKHab) system, and each habitat type within the metric

constitutes an umbrella for a range of different sub-communities [44]. This could explain some of the variation we observed between sites classified as the same habitat type but does not account for the higher invertebrate species richness and abundance observed in some 'low' and 'medium distinctiveness' grassland sites compared to 'very high distinctiveness' sites.

The condition assessment applied to grasslands also risks exacerbating the mismatch between metric scores and invertebrate biodiversity by devaluing sites with habitat features of value to invertebrates [26]. Sites with more than 5% cover of scattered scrub, bare ground, and plant species 'indicative of sub-optimal condition' are assigned a lower score [14]. This simplified scoring does not recognise the value of bare ground and scrub for creating heterogenous sites which support unique ground invertebrate assemblages [45,46]. While vegetation-based features simplify monitoring and capture some useful elements of biodiversity, it is important that the features chosen are relevant to a wider range of flora and fauna.

Refinement of the BNG metric habitat classification and condition assessment may still not capture important dimensions explaining variations in biodiversity among sites. Landscape-level factors are important determinants of a site's biodiversity, not only for invertebrates, but also for groups such as insect pollinators, bats, and birds [47]. Landscape-level factors such as connectivity, which is particularly important for population persistence of less mobile, more specialised taxa [48], are not currently captured by the metric. Site age is also a consideration, given the long time-scales over which some invertebrates take to colonise new sites [49]. Developing a way to incorporate site age and connectivity could ensure that older, well-connected sites achieve higher scores, reflecting a higher likelihood of well-established invertebrate communities. For example, the BNG metric uses a measure of 'strategic significance' which assigns a higher score to post-development habitats which contribute to the delivery of local priorities for nature recovery[50]. However, this does not recognise the connectivity value of sites pre-development, such as a low scoring grassland that connects high scoring grasslands.

Combined area-condition metrics could be refined to better represent faunal diversity by: using a robust appropriately fine-grained habitat classification system; selecting

indicators of site quality that reflect the habitat requirements of a range of faunal species; and incorporating landscape-scale metrics such as connectivity. However, in the absence of evidence that combined area-condition metrics are robust proxies for faunal diversity, metrics should be supplemented with on-the-ground species data collection to understand the biodiversity present on sites under consideration for either development or conservation action. This is especially pertinent where conservation evaluations or prioritisation are occurring. We also found large differences in community composition in different months within a single season (Fig 3a-c), demonstrating the need for repeated data collection to account for seasonal variation. Using species data alongside combined area-condition metrics would help identify sites with low metric score which nonetheless support populations of species of conservation concern, or highly diverse or unique communities. This would enable more nuanced decision-making which reflects the true biodiversity value of sites.

Amid high demand for easy-to-use biodiversity proxies, combined area-condition biodiversity metrics have become widely adopted tools for a range of biodiversity monitoring applications including offsetting schemes and biodiversity credit calculations. Once a particular metric becomes established, it is likely to be emulated by others (as is happening with the England BNG metric; Table 1). This is good, if lessons from the implementation of a metric are used to improve it and to inform the development of other metrics. Our data demonstrate that area-condition metrics can be a poor proxy for invertebrate diversity sampled using pitfall traps, in terms of abundance, richness, community composition, and the occurrence of species of conservation concern. Combined area-condition metrics could be improved by refining indicators of ecological quality to better reflect habitat features required by a range of species, including invertebrates. However, without accompanying data on the occurrence of these species, uncertainty will remain about the full biodiversity benefits of interventions designed using combined area-condition metrics. Ultimately, this leads to the risk that development will occur on sites of high biodiversity value, that biodiversity losses will not be fully offset, or that the biodiversity benefits of biodiversity credits will be overestimated.

Author Contribution Statement

O.T.L., E.J.M.G., R.G., S.O.S.E.z.E., N.E.D., and T.B.A. conceptualised the study. N.E.D., T.B.A., D.C., B.S., and N.T., collected the data. N.E.D. and T.B.A. analysed the data and produced the figures. N.E.D. and T.B.A. wrote the initial manuscript. O.T.L., R.G., S.O.S.E.z.E., J.W.B., E.J.M.G., N.E.D., and T.B.A., contributed to subsequent iterations of the manuscript and all authors approved the final version for submission.

Acknowledgements

The authors would like to thank the Agile sprint team for their support on this project, including Prue Addison and Julia Baker for their advice on fieldwork and the use of the metric, and Isobel Hawkins, Dom Meeks, and Lucy Jones for their assistance with data collection. We also thank Community Ecology Research Oxford (CERO) for their support, including Dave Hemprich-Bennett and Chris Terry for their valuable advice on the data analysis. Finally, we would like to thank all the landowners for allowing us the use of their sites for this study.

Funding

N.E.D., T.B.A., R.G., J.B., E.J.M.G., and O.T.L. were supported by the Natural Environment Research Council (NERC) [grant number NE/W004976/1] as part of the Agile Initiative at the Oxford Martin School. N.E.D. is funded by the Natural Environment Research Council NE/S007474/1 Oxford-NERC Doctoral Training Partnership in Environmental Research and an Oxford Reuben Scholarship. S.O.S.E.z.E. and J.W.B. are supported by EU Horizon 2020 project SUPERB (Systemic solutions for upscaling of urgent ecosystem restoration for forest-related biodiversity and ecosystem services; Ref.: GA-101036849). D.C. was supported by funding from the Oxford-NERC Doctoral Training Partnership in Environmental Research. N.T. was supported by funding from Jesus College, University of Oxford. This publication also arises from research funded by the John Fell Oxford University Press Research Fund under grant 2023/13555 to R.G. and N. E. D. and funding from the Leverhulme Trust funded Leverhulme Centre for Nature Recovery, University of Oxford.

Methods

To test the correlation between BNG metric score and measures of invertebrate biodiversity we sampled a range of habitat types representing combinations of habitat distinctiveness and condition scores. We focused on grassland habitat parcels in agricultural landholdings, as well as areas managed for wildlife conservation. This focus was motivated by a study of early BNG-adopting councils, which found that grasslands, in particular 'modified grasslands' and 'other neutral grasslands', are the most common habitats in BNG metric calculations [51].

Site selection and survey

In 2022, 14 habitat parcels (with habitat parcel defined as an area of the same habitat type and condition) were sampled from five landholdings in Oxfordshire and Buckinghamshire. Under the UK Habitat Classification system (UKHab) [44], these were assessed to be five 'modified grassland' parcels, seven 'other neutral grassland' parcels, and two 'arable field margin' parcels. The 2022 parcels ranged from 0.6 ha to 12.3ha, with a mean parcel size of 4.0ha. In 2023, a new set of 22 habitat parcels were investigated from five landholdings in Oxfordshire, Gloucestershire, and Buckinghamshire. These were assessed under UKHab to be two 'temporary grass and clover ley', seven 'modified grassland', seven 'other neutral grassland', four 'lowland meadow', and three 'wood pasture and parkland' habitat parcels. In 2023, habitat parcels ranged from 0.2ha to 8.7ha with a mean area of 2.4ha.

In both years, statutory biodiversity metric scores were generated for each habitat parcel using standard protocols for calculating BNG metric scores: (1) desk study to identify any statutory or non-statutory site designations using the DEFRA Magic Map [52]; (2) area delineation in the field; (3) floral survey; (4) habitat classification using UKHab 2.0 [44]; (5) metric condition assessment appropriate for habitat type [14]; (6) calculation of biodiversity units using the statutory biodiversity metric [14]. For the floral survey, between 5-10 1 m^2 quadrats were used, depending on the size and heterogeneity of the habitat parcel. These were placed haphazardly whilst ensuring that the quadrat configuration captured any site heterogeneity (e.g., soil moisture and slope). Condition was assessed using the quadrat data and by extensive site walkovers to identify the extent of features that reduce habitat condition (e.g., scrub encroachment).

Invertebrate surveys

Pitfall traps used a twin plastic cup (width: 94mm, height: 135mm) containing 90ml of water, 10ml of propylene glycol, and <1ml of detergent. Galvanised chicken wire (mesh size: 35mm) was placed above the trap and flush with the ground surface to prevent vertebrate bycatch, and a black corrugated plastic rain cover (diameter: 150mm) was positioned 30mm above each trap, secured with tent pegs. Collected specimens were transferred to ethanol. In 2023, the trap design was modified to increase its effectiveness by installing X-shaped rigid foam guidance barriers 650mm in length and 70mm in height [53]. In each habitat parcel, the number of traps per parcel varied between three and nine, increasing with size and heterogeneity of the habitat parcel. Trap placement was subjective and designed to represent the parcel and its habitat features.

In 2022, the traps ran for 4 weeks during August and September. Pitfall traps were active for approximately two weeks and then emptied and reset for another two weeks. In 2023, the traps were set for 1 week of each month in May, June, and August. The shorter timescale was appropriate because of higher catches arising from using guidance barriers and to prevent DNA degradation. Samples that were interfered with by vertebrates or otherwise damaged were removed from the dataset. In 2022, a total of 84 pitfall catches were collected across all sites, with each catch representing 28 trap-days. In 2023, 252 pitfall catches were collected across all sites, with each catch representing 7 trap-days.

Invertebrate sorting and identification

In both years, the mean total invertebrate abundance of each parcel was taken by counting the number of individuals across major taxonomic groups: beetles (Coleoptera), spiders (Araneae), woodlice (Oniscidea), centipedes (Chilopoda), millipedes (Diplopoda), worms (Annelida), slugs and snails (Stylommatophora), and harvestmen (Opiliones). In 2022, ground beetles (Carabidae) were identified to species using Luff (2007). In 2023, to examine invertebrate species richness trends in more depth, samples were subjected to DNA metabarcoding by a commercial facility (NatureMetrics Ltd., Guildford, UK) using standard protocols (Supplementary Material

2). Due to resource constraints, the pitfall traps from each habitat parcel for each individual replicate were pooled. This decision was taken rather than pooling the temporal replicates for each individual pitfall trap so that seasonal variation could be examined. This meant that variation in DNA metabarcoding outcome within each habitat parcel between pitfall traps could not be examined, but we were able to retain this resolution for the abundance data.

Data analysis

Data analysis and plotting was conducted in R using the packages ggplot2, vegan, dplyr, tidyverse, MASS, lme4, DHARMa, scatterplot3d, and iNEXT [55–64].

2022 and 2023 – Invertebrate abundance data

Generalised Linear Mixed Models (GLMMs) were fitted to test the relationship between distinctiveness*condition score of habitat parcels and invertebrate abundance. Landholding was included as a random effect in all models to account for the spatial structure of the habitat parcels in the dataset. Invertebrate abundance counts were modelled specifying a negative binomial error distribution, which greatly improved model diagnostics relative to Poisson errors, with the ratio of residual deviance to degrees of freedom close to 1 in all cases.

2022 – Carabid species richness data

Carabid species richness was rarefied using coverage-based rarefaction to account for variation in sampling intensity and inventory completeness among parcels [56]. Richness was rarefied to the minimum sample coverage of all area habitats (0.8575). The mean and confidence intervals were generated using 50,000 bootstraps. As above, GLMMs were fitted to test the relationship between rarefied carabid species richness and distinctiveness*condition scoring of habitat parcels using a negative binomial error distribution and landholding as a random effect.

2023 – Species richness data

DNA metabarcoding generated 689 Operational Taxonomic Units (OTUs) across all samples. OTUs from taxonomic groups not explicitly targeted by pitfall traps (non-ant Hymenoptera, Orthoptera, Lepidoptera, Mecoptera, Diptera, Nematoda, Plecoptera, Siphonaptera and Thysanoptera) were removed from the dataset, resulting in 417

target OTUs. For analyses of OTU richness, OTUs which occurred only once in the dataset were removed, leaving 263 OTUs. However, only 50.5% of OTUs were resolved to a species-level identification. Given that multiple OTUs can have the same species-level identification, we also analysed species richness alongside OTU richness. To do this, we removed OTUs that did not have a species-level identification and amalgamated multiple OTUs that had the same species-level identification. This left 145 OTUs which were used to analyse total ground invertebrate species richness and the species richness of the main taxonomic groups.

Both OTU and species richness were rarefied to account for differences in sampling intensity using iNEXT [56]. Richness was rarefied to a minimum sampling coverage of 55% and samples which fell below this threshold were excluded from the analysis. GLMMs were used to test the relationship between total rarefied invertebrate OTU and species richness and distinctiveness*condition scores, with a negative binomial error distribution and including landholding as a random effect.

Due to zero inflation in the data and small datasets for individual taxonomic groups, sample coverage rarefaction was not possible for the species richness of individual taxonomic groups. Instead, the raw species richness value was used, excluding samples with <8 pitfall traps in total, and the number of pitfalls per parcel per season (between 8 and 15) was used as a covariate in the models to account for differences in sampling intensity. This was done for carabid beetle (Carabidae), rove beetle (Staphylinidae), spider (Araneae), slug and snail (Stylommatophora), true bug (Hemiptera), woodlice (Oniscidea), and springtail (Collembola) species richness. Sampling intensity was not a significant predictor of species richness for any group. GLMMs were fitted using a negative binomial error distribution and landholding as a random effect.

2023 – Community Composition

To compare community composition across sites we used the OTUs from the DNA metabarcoding dataset which were identified to the family-level within our target pitfall taxa for May, June, and August [65]. This included 381 OTUs from 78 families. The species-level dataset was too sparse to produce a meaningful output from this analysis. Using numbers of family-level OTUs reduced the zero-inflation of the data

and allowed clustering analysis. We used the number of family-level OTUs per sample and calculated a distance matrix using Bray-Curtis in the R package 'vegan' with 1000 permutations on 3 dimensions until the least-stress solution had been calculated [55].

Species of conservation concern

We deemed species to be of conservation concern if they were assessed to be threatened or scarce in Great Britain or if they were listed on a relevant policy list. Our dataset included species of principal importance in England (England biodiversity list) and species assessed as nationally scarce from relevant taxonomic assessments [66–68].

References

1. Díaz, S. *et al.* Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science* **366**, eaax3100 (2019).
2. Leclère, D. *et al.* Bending the curve of terrestrial biodiversity needs an integrated strategy. *Nature* **585**, 551–556 (2020).
3. Convention on Biological Diversity. Kunming-Montreal Global Biodiversity Framework. https://www.cbd.int/gbf/ (2023).
4. zu Ermgassen, S. O. S. E. *et al.* The ecological outcomes of biodiversity offsets under "no net loss" policies: A global review. *CONSERVATION LETTERS* **12**, (2019).
5. Wunder, S. *et al.* Biodiversity credits: learning lessons from other approaches to incentivise conservation. Preprint at https://doi.org/10.31219/osf.io/qgwfc (2024).
6. zu Ermgassen, S. zu *et al.* The current state, opportunities and challenges for upscaling private investment in biodiversity in Europe. Preprint at https://doi.org/10.31235/osf.io/2u6ky (2024).

7. Marshall, E., Wintle, B. A., Southwell, D. & Kujala, H. What are we measuring? A review of metrics used to describe biodiversity in offsets exchanges. *Biological Conservation* **241**, 108250 (2020).
8. Parkes, D., Newell, G. & Cheal, D. Assessing the quality of native vegetation: The 'habitat hectares' approach. *Ecological Management & Restoration* **4**, S29–S38 (2003).
9. Verra. *SDVISta Nature Framework*. https://verra.org/wp-content/uploads/2023/09/SD-VISta-Nature-Framework-v0.1-for-Public-Consultation.pdf (2023).
10. Pöll, C. E., Willner, W. & Wrbka, T. Challenging the practice of biodiversity offsets: ecological restoration success evaluation of a large-scale railway project. *Landscape Ecol Eng* **12**, 85–97 (2016).
11. Gibbons, P. *et al.* An operational method to assess impacts of land clearing on terrestrial biodiversity. *Ecological Indicators* **9**, 26–40 (2009).
12. TNFD. *The TNFD Nature-Related Risk and Opportunity Management and Disclosure Framework. Beta v0.4 Annex 4.3 Disclosure Metrics Annexes*. https://tnfd.global/wp-content/uploads/2023/07/23-23882-TNFD_v0.4_Annex_4.3_v3-1-1.pdf?v=1690527800 (2023).
13. Temple, H. *et al. Forecasting the Path towards a Net Positive Impact on Biodiversity for Rio Tinto QMM*. (2012).
14. DEFRA. Statutory biodiversity metric tools and guides. *GOV.UK* https://www.gov.uk/government/publications/statutory-biodiversity-metric-tools-and-guides (2024).
15. Ecogain. CLIMB - Changing land use impact on biodiversity. http://climb.ecogain.se (2023).

16. AECOM. SG Bio Accounting Metric - Singapore Biodiversity Accounting Metric. https://anz.planengage.com/singapore-bio-metric/page/SG-Bio-Accounting-Metric-10 (2023).
17. Wallacea Trust. *Methodology for Quantifiying Units of Biodiversity Gain*. https://wallaceatrust.org/wp-content/uploads/2022/12/Biodiversity-credit-methodology-V3.pdf (2023).
18. Ramboll. Americas Biodiversity Metric 1.0 - Ramboll Group. https://www.ramboll.com/en-us/americas-biodiversity-metric (2024).
19. NatureScot. A Biodiversity Metric for Scotland's Planning System - Key Issues consultation | NatureScot. https://www.nature.scot/doc/biodiversity-metric-scotlands-planning-system-key-issues-consultation (2024).
20. zu Ermgassen, S. O. S. E. *et al.* Evaluating the impact of biodiversity offsetting on native vegetation. *Global Change Biology* **29**, 4397–4411 (2023).
21. Hanford, J. K., Crowther, M. S. & Hochuli, D. F. Effectiveness of vegetation-based biodiversity offset metrics as surrogates for ants. *Conservation Biology* **31**, 161–171 (2017).
22. Thorn, S., Hobbs, R. J. & Valentine, L. E. Effectiveness of biodiversity offsets: An assessment of a controversial offset in Perth, Western Australia. *Biological Conservation* **228**, 291–300 (2018).
23. Hawkins, I., zu Ermgassen, S. O. S. E., Grub, H., Treweek & Milner-Gulland, E. J. No Consistent Relationship Found Between Biodiversity Metric Habitat Scores and the Presence of Species of Conservation Priority. *Inpractice* 16–20 (2022).

24. Rampling, E. E., zu Ermgassen, S. O. S. E., Hawkins, I. & Bull, J. W. Achieving biodiversity net gain by addressing governance gaps underpinning ecological compensation policies. *Conservation Biology* **n/a**, (2023).
25. Marshall, C. A. M. *et al.* England's statutory biodiversity metric enhances plant, but not bird nor butterfly, biodiversity. *Journal of Applied Ecology* **n/a**, (2024).
26. Duffus, N. E. *et al.* Leveraging Biodiversity Net Gain to address invertebrate declines in England. *EcoEvoRxiv* (2024) doi:10.32942/X2TP5V.
27. Perner, J. & Malt, S. Assessment of changing agricultural land use: response of vegetation, ground-dwelling spiders and beetles to the conversion of arable land into grassland. *Agriculture, Ecosystems & Environment* **98**, 169–181 (2003).
28. Gerlach, J., Samways, M. & Pryke, J. Terrestrial invertebrates as bioindicators: an overview of available taxonomic groups. *J Insect Conserv* **17**, 831–850 (2013).
29. DEFRA. Biodiversity net gain. *GOV.UK* https://www.gov.uk/government/collections/biodiversity-net-gain (2024).
30. Maron, M. *et al.* 'Nature positive' must incorporate, not undermine, the mitigation hierarchy. *Nat Ecol Evol* **8**, 14–17 (2024).
31. West, T. A. P. *et al.* Action needed to make carbon offsets from forest conservation work for climate change mitigation. *Science* **381**, 873–877 (2023).
32. Forest Trends' Ecosystem Marketplace. *State of the Voluntary Carbon Market 2024: On the Path of Maturity*. https://3298623.fs1.hubspotusercontent-na1.net/hubfs/3298623/SOVCM%202024/State_of_the_Voluntary_Carbon_Markets_20240529%201.pdf (2024).

519 33. Swinfield, T., Shrikanth, S., Bull, J. W., Madhavapeddy, A. & zu Ermgassen, S. O. S. E.
520 Nature-based credit markets at a crossroads. *Nat Sustain* 1–4 (2024)
521 doi:10.1038/s41893-024-01403-w.
522 34. Brooks, D. R. *et al.* Large carabid beetle declines in a United Kingdom monitoring
523 network increases evidence for a widespread loss in insect biodiversity. *Journal of*
524 *Applied Ecology* **49**, 1009–1019 (2012).
525 35. Mancini, F. *et al.* Invertebrate biodiversity continues to decline in cropland. *Proceedings*
526 *of the Royal Society B: Biological Sciences* **290**, 20230897 (2023).
527 36. Ewald, J. A. *et al.* Fifty years of monitoring changes in the abundance of invertebrates in
528 the cereal ecosystem of the Sussex Downs, England. *Insect Conservation and Diversity*
529 **n/a**, (2024).
530 37. Nichols, E. *et al.* Ecological functions and ecosystem services provided by Scarabaeinae
531 dung beetles. *Biological Conservation* **141**, 1461–1474 (2008).
532 38. Dainese, M. *et al.* A global synthesis reveals biodiversity-mediated benefits for crop
533 production. *Science Advances* **5**, eaax0121 (2019).
534 39. Seibold, S. *et al.* The contribution of insects to global forest deadwood decomposition.
535 *Nature* **597**, 77–81 (2021).
536 40. Holland, J. m., Hutchison, M. a. s., Smith, B. & Aebischer, N. j. A review of invertebrates
537 and seed-bearing plants as food for farmland birds in Europe. *Annals of Applied Biology*
538 **148**, 49–71 (2006).
539 41. Schuldt, A. & Assmann, T. Invertebrate diversity and national responsibility for species
540 conservation across Europe – A multi-taxon approach. *Biological Conservation* **143**,
541 2747–2756 (2010).

42. Oberprieler, S. K., Andersen, A. N., Gillespie, G. R. & Einoder, L. D. Vertebrates are poor umbrellas for invertebrates: cross-taxon congruence in an Australian tropical savanna. *Ecosphere* **10**, e02755 (2019).
43. Treweek, J., Butcher, B. & Temple, H. Biodiversity offsets: possible methods for measuring biodiversity losses and gains for use in the UK. *In Practice* **69**, 29–32 (2010).
44. UKHab. ukhab – UK Habitat Classification. *UKHab* https://ukhab.org/ (2023).
45. Key, R. Bare ground and the conservation of invertebrates. *British Wildlife* **11**, 183–191 (2000).
46. Lyons, A., Ashton, P. A., Powell, I. & Oxbrough, A. Epigeal spider assemblage responses to vegetation structure under contrasting grazing management in upland calcareous grasslands. *Insect Conservation and Diversity* **11**, 383–395 (2018).
47. Le Provost, G. *et al.* Contrasting responses of above- and belowground diversity to multiple components of land-use intensity. *Nat Commun* **12**, 3918 (2021).
48. Wamser, S., Diekötter, T., Boldt, L., Wolters, V. & Dauber, J. Trait-specific effects of habitat isolation on carabid species richness and community composition in managed grasslands. *Insect Conservation and Diversity* **5**, 9–18 (2012).
49. Woodcock, B. A. & McDonald, A. W. What goes wrong? Why the restoration of beetle assemblages lags behind plants during the restoration of a species rich flood-plain meadow. *Fritillary* (2010).
50. Department for Environment, Food & Rural Affairs. Local nature recovery strategies. *GOV.UK* https://www.gov.uk/government/publications/local-nature-recovery-strategies/local-nature-recovery-strategies (2023).

51. zu Ermgassen, S. O. S. E. *et al.* Exploring the ecological outcomes of mandatory biodiversity net gain using evidence from early-adopter jurisdictions in England. *CONSERVATION LETTERS* **14**, (2021).
52. Natural England. MAGIC. *MAGIC Map* https://magic.defra.gov.uk/About_MAGIC.htm (2023).
53. Boetzl, F. A., Ries, E., Schneider, G. & Krauss, J. It's a matter of design—how pitfall trap design affects trap samples and possible predictions. *PeerJ* **6**, e5078 (2018).
54. Luff, M. L. *Handbooks for the Identification of British Insects. Vol. 4, Pt. 2: The Carabidae (Ground Beetles) of Britain and Ireland*. vol. 4 (Royal Entomological Society, St. Albans, 2007).
55. Dixon, P. VEGAN, a package of R functions for community ecology - Dixon - 2003 - Journal of Vegetation Science - Wiley Online Library. *Journal of Vegetation Science* **14**, 927–930 (2003).
56. Hsieh, T. C., Ma, K. H. & Chao, A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution* **7**, 1451–1456 (2016).
57. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. https://ggplot2.tidyverse.org/authors.html#citation (2016).
58. Wickham, H. *et al.* dplyr: A Grammar of Data Manipulation. (2023).
59. R Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (2024).
60. Hartig, F. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.4.6. (2022).

61. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S. Fourth Edition.* (Springer, New York, 2002).

62. Ligges, U. & Maechler, M. scatterplot3d - An R Package for Visualizing Multivariate Data. *Journal of Statistical Software* **8**, 1–20 (2003).

63. Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 1–48 (2015).

64. Wickham, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* **4**, 1686 (2019).

65. Hemprich-Bennett, D. R. *et al.* Selective Logging Shows No Impact on the Dietary Breadth of a Generalist Bat Species: The Fawn Leaf-Nosed Bat (Hipposideros cervinus). *Front. Ecol. Evol.* **9**, (2021).

66. Natural England. *A Review of the Beetles of Great Britain: The Darkling Beetles and Their Allies. Species Status No.18 (NECR148)*. https://publications.naturalengland.org.uk/publication/5862553594888192 (2014).

67. Natural England. *A Review of the Beetles of Great Britain: Ground Beetles (Carabidae). Species Status No.25. NECR189*. https://publications.naturalengland.org.uk/publication/6270849377107968 (2016).

68. DEFRA. *List of Habitats and Species of Principal Importance in England*. https://www.gov.uk/government/publications/habitats-and-species-of-principal-importance-in-england (2022).

69. Harvey, P. *et al. A Review of the Scarce and Threatened Spiders (Araneae) of Great Britain: Species Status No. 22. NRW Evidence Report No: 11*. (2017).

Figures

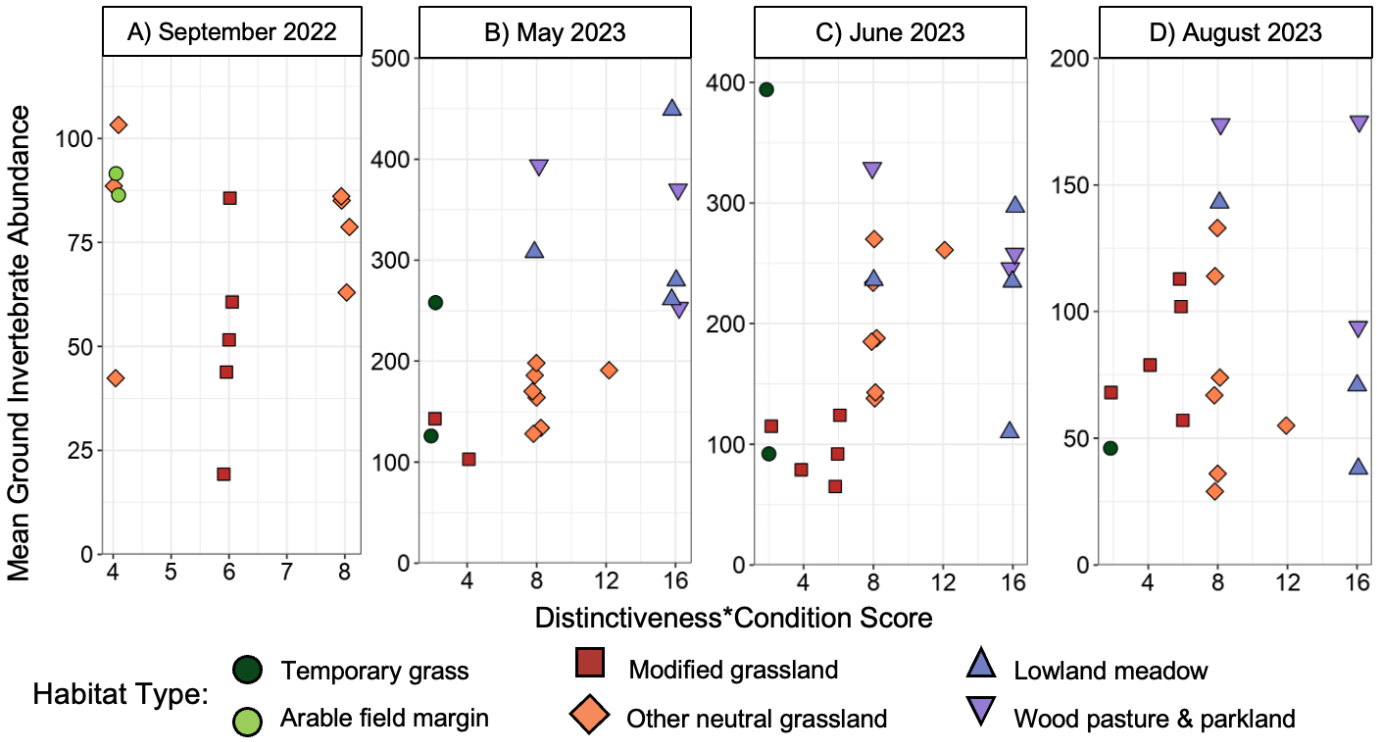

Figure 1. Mean ground invertebrate abundance and distinctiveness*condition scores over four weeks in 2022 (A), and three one-week periods across three months in 2023 (B,C,D).

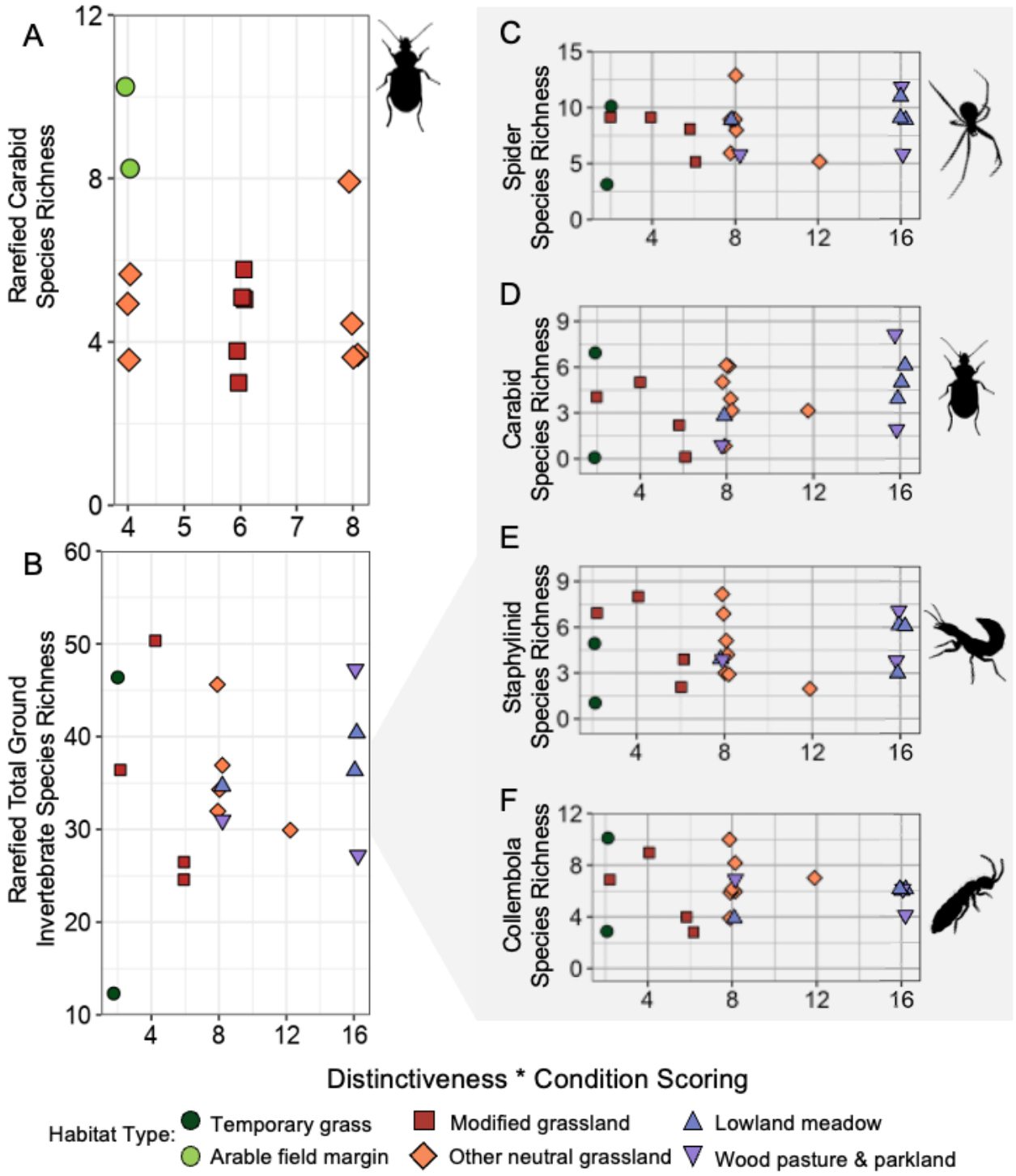

Figure 2. Ground invertebrate species richness and distinctiveness*condition scores for carabid beetles (A), total ground invertebrate species richness (B), spiders (C), carabids (D), staphylinids (E) and Collembola (F). Data for (A) are from 2022 sampling, and (B to D) from 2023.

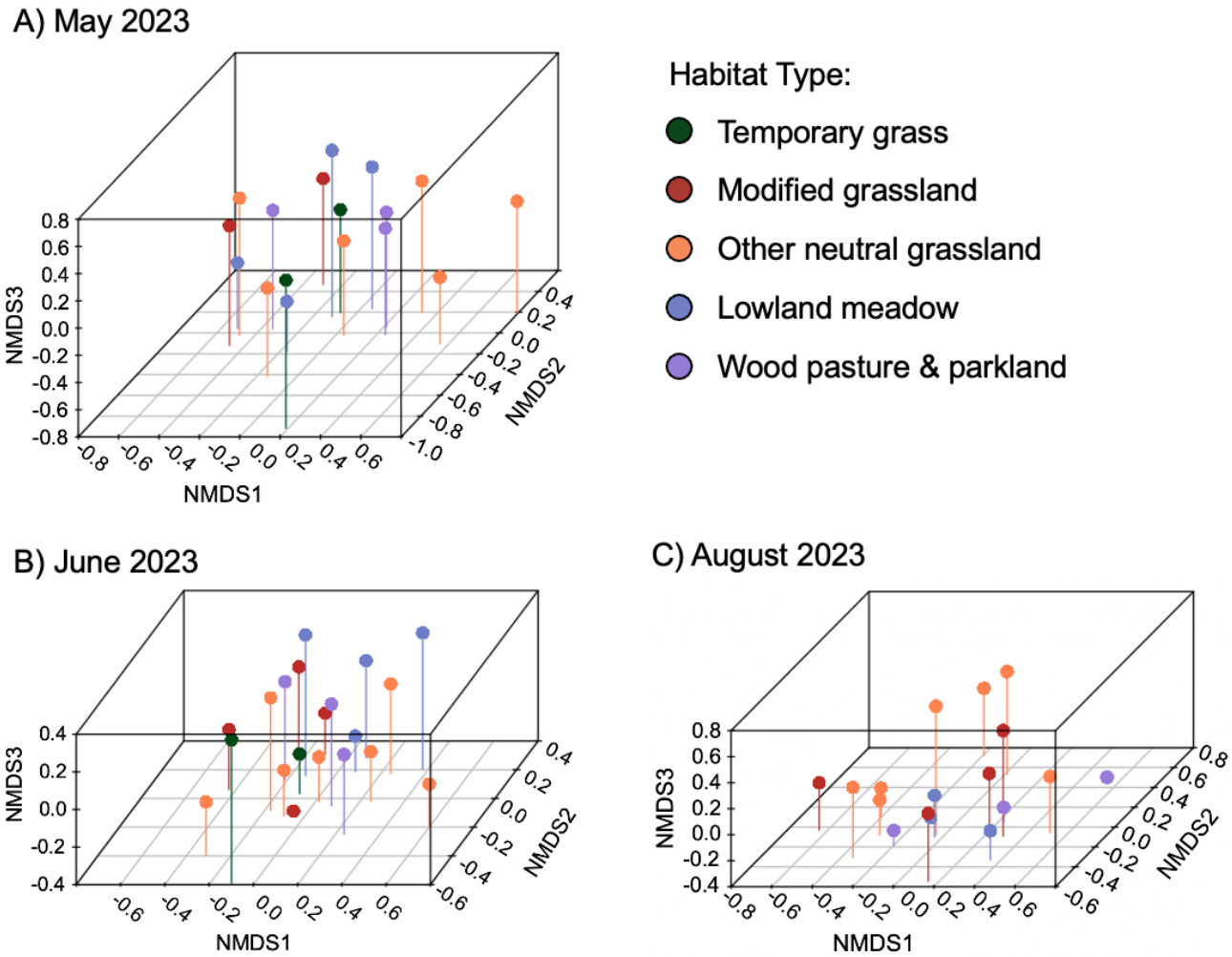

Figure 3. Nonmetric multidimensional scaling of family-level Operational Taxonomic Units (OTUs) from 2023. The least stress solution was 0.09 for May (A), 0.12 for June (B), 0.13 for August (C).