Searching for the best-performing multi-species grassland mixtures: evolving research questions, and experimental design and analysis

Brophy C.[1], Suter M.[2], Thivierge M.-N.[3], Vishwakarma R.[1], Komainda M.[4] and Finn J.A.[5]

[1]School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland; [2]Agroscope, Forage Production and Grassland Systems, 8046 Zurich, Switzerland; [3]Agriculture and Agri-Food Canada, Quebec Research and Development Centre, Quebec City, QC, Canada; [4] Georg-August-University Göttingen, Department of Crop Sciences, Grassland Science, Von-Siebold-Str. 8, 37075 Göttingen; [5] Teagasc, Environment Research Centre, Johnstown Castle, Wexford, Ireland

Abstract

The design and analysis of biodiversity and ecosystem function (BEF) studies have both evolved in recent decades to permit the addressing of increasingly complex research questions. Simple linear regression techniques can address how average patterns in ecosystem function are affected by species richness, while ANOVA approaches permit the comparison of specific communities contained in the design. The Diversity–Interactions modelling approach can address both of these research questions and, furthermore, it can assess the relative performance of individual species, assess if and how species interact, and can identify species diversity regions (characterised by species richness, composition and proportions) that lead to the best-performing mixtures. In this paper, we discuss experimental designs for varying research questions. We also show the added benefits from using the Diversity–Interactions modelling approach over traditional approaches using simulated BEF data and data from a productive grassland experiment. Being able to identify the best-performing mixtures can provide valuable knowledge for the management of both nutrient-rich productive grasslands where there is a degree of control over mixture design at sowing, and nutrient-poor semi-natural grasslands where management decisions can be taken to influence species dynamics.

Keywords: biodiversity and ecosystem function, Diversity–Interactions modelling, forage yields, methodology, optimal design of mixtures

Introduction

Within ecological systems, a distinct discipline of Biodiversity and Ecosystem Function (BEF) has been investigating whether more diverse assemblages perform better than less diverse ones (Hooper *et al.*, 2005; Tilman *et al.*, 2014). Examples of biological systems that have been studied in this way include bacteria, protists, annelids with or without insects, trees, and grasslands (e.g., Hooper *et al.*, 2005; Weisser *et al.*, 2017; Isbell *et al.*, 2017). What unites many of these studies is a concern about the effect of declining biodiversity on ecosystem function, given the ongoing extinction crisis. Given their economic and ecological importance, global distribution and relative ease of manipulation for experiments, grasslands have long been a prominent focus of BEF science. In addition to investigations on declining biodiversity by manipulating species richness, grassland BEF studies have explored in various ways the effects of mixing species through experimentally manipulating species identities and/or the proportions in which they are combined.

The ability to quantify the effects of mixing species on a given ecosystem function relies on the statistical methodology used, and the better the methodology, the better the ability to identify a range of best-performing communities. Here, our primary aim is to outline advances in the design and analysis of BEF studies to better quantify how changes in species diversity affect ecosystem function and identify a range of best-performing multi-species mixtures. We are especially interested in methods that can quantify how species richness, species identities/community composition, and species proportions/relative abundance jointly affect ecosystem function. When quantifying the effect of species diversity on an ecosystem function, the nature of the research questions that are asked can have far-reaching implications for the selection of the most appropriate experimental design and statistical analysis. We discuss the pros and cons of different methodologies in unravelling the underlying complexity of the effects of changing diversity in grassland communities on ecosystem function and illustrate them using analyses of simulated and empirical data.

Semi-natural and productive grassland BEF studies

Grasslands BEF research can differentiate between nutrient-poor to mesotrophic semi-natural grassland systems that typically have high species richness, and nutrient-rich productive grassland systems that are often intensively managed to optimise forage yield production. Species-rich grasslands BEF studies have contributed a rich theoretical and empirical literature (Tilman *et al.*, 2001; Roscher *et al.*, 2005; Marquard *et al.*, 2009; Craven *et al.*, 2016); however, it cannot be simply assumed that this theory and methodology can be applied to productive grasslands (Lüscher *et al.*, 2022; Finn *et al.*, 2024). In the following, we discuss overlaps between these two grassland types in concepts and principles.

Within BEF studies from semi-natural grasslands, a strong motivating theme has been the effect of species extinctions, which readily leads to questions about the effect of changes in species richness on ecosystem functions. Inevitably, however, manipulations of species richness also lead to concomitant changes in species composition and relative abundance. For example, consider a four species pool (species 1, 2, 3 and 4) and sown proportions in communities: (A) 0.5:0.5:0:0, (B) 0:0:0.5:0.5, (C) 0.25:0.25:0.25:0.25, (D) 0.7:0.1:0.1:0.1. Across these communities, sown richness varies (communities A and B vs. C and D), and at a given level of richness, composition/species identities can vary (community A vs. B) or sown proportions can vary (community C vs. D). The latter is often referred to as a manipulation of evenness; where evenness is a community level quantification of how varied the sown species proportions are (e.g., Wilsey and Polley, 2004). Whenever species richness changes, then species identities and proportions must also vary for at least some of the species (e.g. community A versus C). Furthermore, species' extinctions are rarely random (Schläpfer *et al.*, 2005), which implies a range of possible trajectories from higher richness to lower, with some trajectories more likely to occur than others (an effect of species composition). Also, the extinction of any species is preceded by a period when it experiences a declining proportion before eventually it goes extinct. Designs of BEF studies from semi-natural grasslands manipulating species richness thus face the challenge of appropriately reflecting natural communities with largely different species proportions and their dynamics over time, where the ratio of the largest to smallest species' biomass in a mixture can become more than 100 (Roscher *et al.*, 2005).

In comparison, within nutrient-rich productive grasslands that are typically sown on a regular basis with agronomic cultivars of high-yielding forage species, landowners have a very high degree of control over the richness, composition and proportions of species at the time of sowing. Agronomic experiments that investigate the design of multi-species mixtures typically manipulate some combination of the number of species (richness), identities (community composition) and sown species proportions (linked to community evenness) across experimental units, and measure a response with the ultimate aim of identifying high- or highest-performing communities. The nature of the research questions will help with selecting the most appropriate experimental design and statistical analysis.

Thus, despite considerable differences in the motivations and practical applications, both semi-natural and productive grassland systems share a common need for a quantitative analytical framework that can incorporate these three dimensions of species diversity: species richness, composition, and proportions.

Diversity–Interactions modelling: a statistical tool for analysing data from BEF studies

In previous BEF studies, simple linear regression models have commonly been fit to explain the relationship between grassland yield and sown species richness, often with richness log or otherwise transformed (e.g. Hector *et al.*, 1999; Temperton *et al.*, 2007); these models will collectively be referred to as the 'richness model' hereafter. Another common approach is to use ANOVA analyses, where each unique manipulated community is treated as a level of a factor (e.g. Bélanger *et al.*, 2020; Abalos *et al.*, 2021), or, occasionally, ANOVA approaches that have been refined to account for functional group categorisations or legume presence (e.g. Marquard *et al.*, 2009). Here, we illustrate a modelling framework called Diversity–Interactions modelling (Kirwan *et al.*, 2009; Connolly *et al.*, 2013; Moral *et al.*, 2023) that can be used for quantifying the effects of multiple components of initial species diversity (richness, composition, and proportions) on a subsequently measured ecosystem function within a single analysis.

Diversity–Interactions models follow a regression approach with initial species proportions and their interactions as predictors, where initial species proportions may be sown proportions or realised proportions from a preceding time. The inclusion of each species proportion as a predictor captures species identities effects. The interaction terms can take many forms, and varying their formulations allows the testing of a range of biological hypotheses. For example, different specification of these terms allows testing of whether interactions being driven by individual pairs of species, whether species functional groupings dictate how species mix, or whether all pairs of species interact in the same way. Moreover, a non-linear exponent can be included on a pairwise interaction and its value can identify the shape of the BEF relationship as richness increases; under certain conditions, setting this exponent to 0.87 equates to using 'log richness' in the richness model (Moral *et al.*, 2023). As with any regression model, the range of predictor combinations that lead to maximal (or minimal) responses can be identified through optimisation methods, i.e. regions of initial species proportions that lead to the best responses can be identified.

BEF studies, design and research questions

The research questions of interest should motivate the design of any BEF experiment. For example, there are substantial differences between the following questions: what is the difference between a lower-diversity community and a higher-diversity community?; what is the effect of a change in species richness from one to two to four to eight and more species?; what is the effect of changing the species proportions in a community?, and; from a given set of species, which combination of species and their proportions defines the community or range of communities that are best-performing?

If the average trend with increasing/decreasing richness is primarily of interest, random assemblages of communities across all levels of richness will be appropriate (e.g. Hector *et al.*, 1999), although some stratification at each level of richness may be employed (e.g. Bell *et al.*, 2005), and a richness model can be applied. Often in these designs, two replicates for each unique community are used; this can be helpful for testing for lack of fit but is not required for the richness model. This methodology has been heavily employed when analysing data from semi-natural grassland BEF experiments. If the comparison of specific communities of interest is the main research question, ANOVA can be used, but multiple replicates of each unique community are required in the design. This methodology has been used frequently when analysing data from BEF studies, especially on productive grasslands (e.g. Bélanger *et al.*, 2020; Abalos *et al.*, 2021).

If the research question of interest relates to identifying the range of multi-species mixtures that will lead to the best ecosystem function outcomes, the richness model and ANOVA will not be conclusive. The richness model can identify the best richness level on average, and the ANOVA model can identify the best out of the design communities only. For the richness model, variation due to composition and changing proportions at a given level of richness are assumed part of the model residual error variation, but a large amount of this variation can be attributed to composition and proportions (Moral *et al.*, 2023; Finn *et al.*, 2024). Among others, an important benefit of Diversity–Interactions modelling is its ability to predict and make inference beyond the specific communities in the experimental design (but within the continuous design space). This is important because the best-performing community/communities may be different to those that are in the experimental design; for example, a 50:50 binary community might be in the design, but 75:25 community may outperform it. Thus, a Diversity–Interactions model can identify the range of communities, characterised by richness, composition and species proportions, that will lead to the best outcomes. This approach can further answer questions on the relative performance of species both in monoculture and in their contributions to mixtures, and can quantify species interactions that may lead to benefits of mixing species.

If designing a BEF experiment for use with Diversity–Interactions modelling, it is important (but not imperative) that there is wide coverage of the design space. It is required to have varying proportions for each species across the mixtures, and the wider the range the better; this naturally happens when richness is manipulated in a design. Furthermore, it is recommended (but not required) to include replicated monocultures of each species and to have one or more replicates of two-species communities containing each pair of species to improve power of the model fit. It is also desirable to have both equi-proportional and unbalanced communities to provide better coverage of the design space across the richness levels of interest. However, it is not crucial to have monocultures of all species or to have unbalanced communities in the design for the Diversity–Interactions approach to be useful. For example, if a given species was absent in monoculture, that species identity effect could still be estimated to determine the species' contribution to mixtures, but predictions and inference for that species should only be between the proportions for which it was present in the design, and not in monoculture. For a completed experiment that was not designed optimally for fitting a Diversity–Interactions model, frequently the approach will still have value. The only hard requirement is that there must be a range of initial species proportions for each species in the pool. Finally, as with all regression approaches, predictions can be interpolated to any point within the design space (here the species proportions).

In the next two sections, we show two examples, a first dataset suited to a richness model, and a second suited to ANOVA, to show the additional knowledge that can be gained when using Diversity–Interactions modelling. Analyses and visualisations were implemented using R statistical software, including the packages DImodels (Moral *et al.*, 2023), PieGlyph (Vishwakarma *et al.*, 2023) and DImodelsVis (Vishwakarma *et al.*, 2024).

Added value of Diversity–Interactions modelling: going beyond richness

Here we use simulated data assuming a BEF experiment where the diversity of four species was manipulated. The design included monocultures of all species and two-, three- and four-species mixtures that varied in composition and proportions, and a yield response was simulated. Firstly, a log richness model was fitted to the data, and it captured the increasing but saturating pattern of the yield response with richness (Figure 1a). Secondly, a range of Diversity–Interactions models were fitted (see Table 1 in Moral *et al.*, 2023 for details on the range of models) and the best model selected had the four species identity effects, a single interaction coefficient that captured the interaction strength between any pair of species (and its estimate was positive), and the non-linear interaction exponent parameter was included. Predicting from the Diversity–Interactions model illustrates the effects of changing species composition and proportions, as well as capturing the increasing but saturating relationship with richness (Figures 1b and 1c). At a given level of richness, communities dominated by species 3 and 4 led to higher yields than communities dominated by species 1 and 2 (e.g. compare multiple predictions at richness = 2 in Figures 1b and 1c). Species richness was not the sole driver of this BEF relationship; there were two-species communities that outperformed four-species communities (e.g. species 1 and 4 50:50 compared to the community 0.9:0.033:0.033:0.033; Figure 1b). Whilst this is an extreme example, Figures 1b and 1c illustrate the importance of species identity and proportion in identifying regions of sown proportions that lead to the best-performing communities. Thus, with Diversity–Interactions modelling, we can retain the benefits of richness modelling by predicting on average across richness levels, and gain information by capturing the additional effects of species identity and proportions at any richness level. This provides insight on the effects of the multiple components of species diversity on ecosystem function.

Added value of Diversity–Interactions modelling: going beyond ANOVA

Here, we use data from a five-species productive grassland experiment (Komainda *et al.*, 2020). The species used in this mesocosm experiment were the grasses *Lolium perenne* (Lp) and *Dactylis glomerata* (Dg), the herbs *Plantago lanceolata* (Pl) and *Taraxacum officinale* (To), and the legume *Trifolium repens* (Tr). Across mesocosms, species diversity was manipulated to include monocultures of each species, all possible three-species equi-proportional communities, and the five-species equi-proportional community. The species diversity treatment was crossed with a multi-phase drought stress/water-fed control treatment. We modelled yield data from only phase three of the drought stressed treatment using both ANOVA and Diversity–Interactions modelling to compare the two approaches.

From the ANOVA analysis, the predicted yield varied considerably across the communities in the design (Figure 2a). Generally, communities that included the legume Tr performed well, while those without Tr had the lowest yields. Pairwise comparisons can identify the best-performing communities, out of those in the design. For example, the Tr monoculture out-performed the five-species ($P = 0.03$)

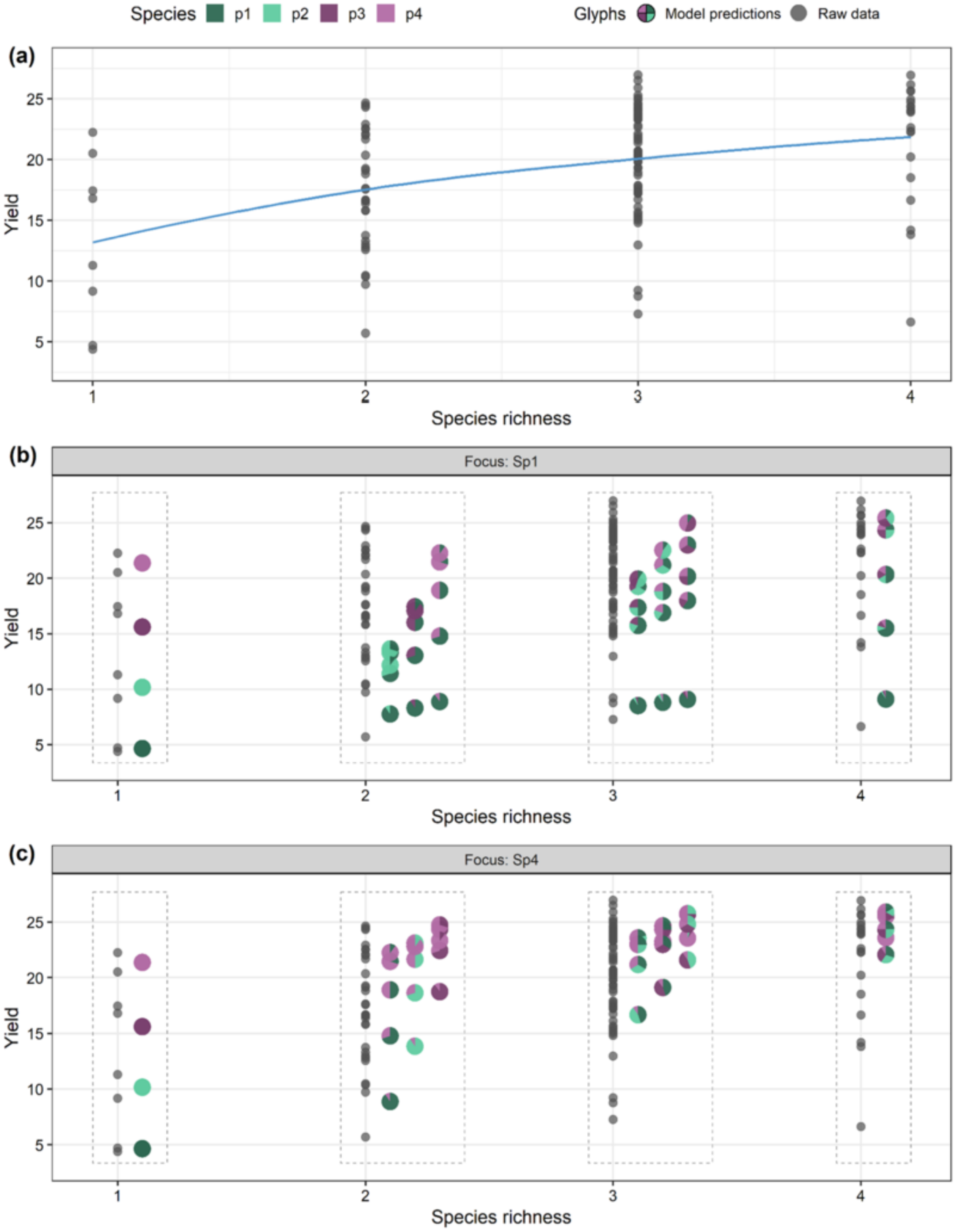

Figure 1. Illustration of simulated data typically arising from a grassland study that manipulated species diversity. (a) Predicted yield (blue line) from a richness model with log richness as a predictor, with raw data (grey circles). (b and c) Predicted yield (pie-glyphs) from the 'Average pairwise interaction' DI model fitted to the simulated data (grey dots). Within each vertical stack, the pie-glyphs represent quantitative predictions for communities containing species Sp1 (panel b) or Sp4 (panel c) at proportions ranging from 0.1 to 0.9 with the remaining proportions split equally among the other species in the mixture. Panels (b) and (c) show how predicted yield can be affected by multiple dimensions of species diversity, illustrated by richness (*x*-axis), composition (changes by moving horizontally right across the stacks of pie-glyphs at each level of richness), and species proportions (changes by moving up or down in each stack of pie-glyphs at each level of richness).

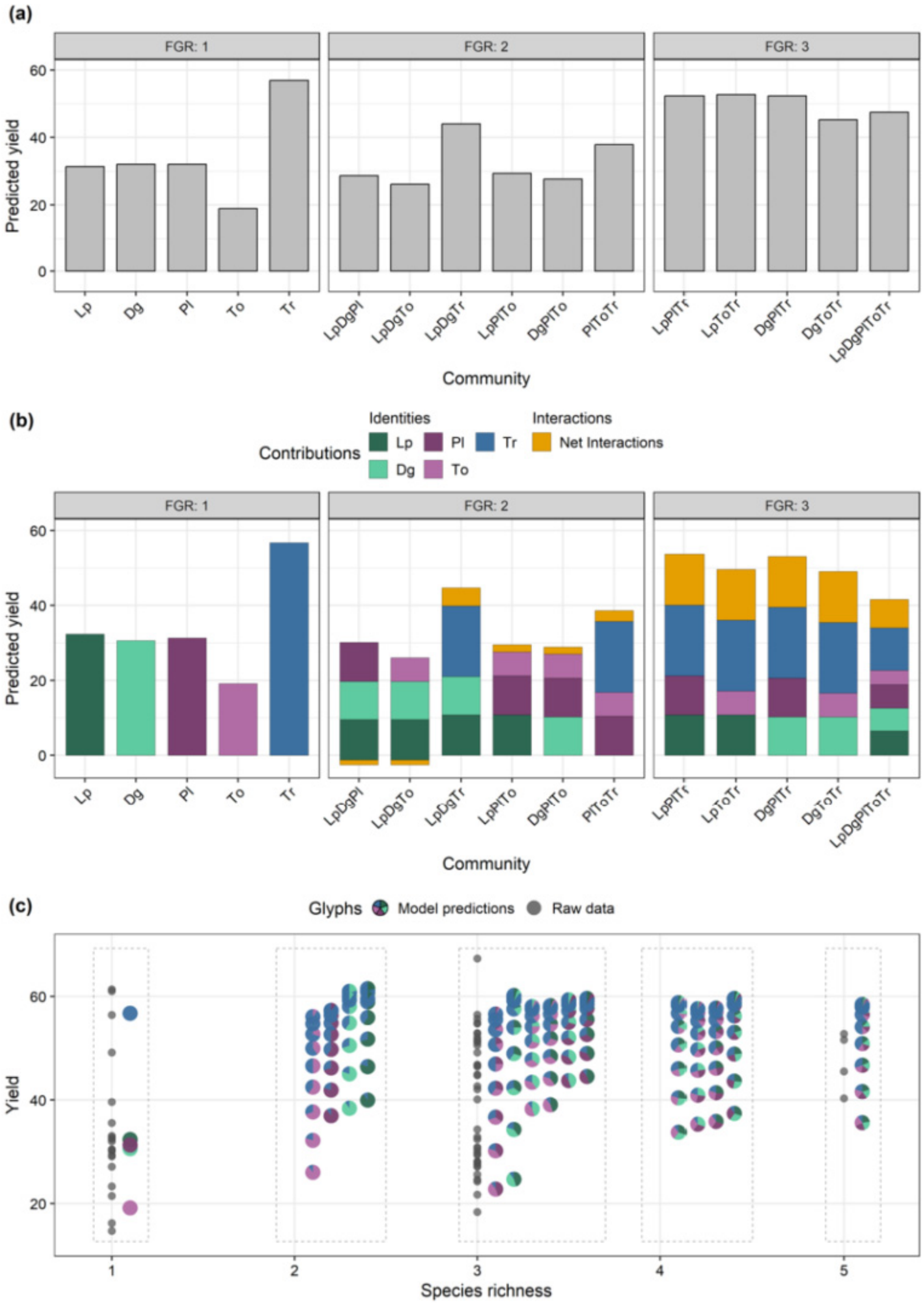

Figure 2. Illustration of the benefits of Diversity–Interactions models over ANOVA. (a) Predicted yield from the ANOVA model. (b) Predicted yield from the 'functional group' Diversity–Interactions model, with the contributions of species identities and net interactions shown in the bar segments. The bars are grouped by functional group richness (FGR). (c) Raw data (grey dots) and predicted yield (pie-glyphs) for communities with varying composition and proportions at each level of species richness. The pie-glyphs are arranged horizontally according to species composition, and within each vertical stack the proportion of *T. repens* increases from 0.1 to 0.9, with the remaining proportions split equally among the other species in the mixture.

Table 1. Parameter estimates from the 'functional group' Diversity–Interactions model.

Term	Explanation	Estimate	SE	P-value
β_1	*Lolium perenne* (Lp)	32.3	3.000	
β_2	*Dactylis glomerata* (Dg)	30.6	3.000	
β_3	*Plantago lanceolata* (Pl)	31.3	3.000	
β_4	*Taraxacum officinale* (To)	19.1	3.000	
β_5	*Trifolium repens* (Tr)	56.7	3.207	
δ_{GH}	Grass-herb	30.0	10.668	0.007
δ_{GL}	Grass-legume	57.5	15.155	<0.001
δ_{HL}	Herb-legume	34.5	15.155	0.027
δ_{GG}	Grass-grass	−71.5	23.539	0.004
δ_{HH}	Herb-herb	−42.9	23.539	0.074

The first five terms are the identity effects (coefficient of the sown proportion of each species). The interaction structure selected captures a unique interaction for any pair of species from differing functional groups (grass-herb, grass-legume, herb-legume) and one for the two grasses (grass-grass) and one for the two herbs (herb-herb). The equation of the final model is:

$$y = \sum_{i=1}^{5} \beta_i p_i + \delta_{\mathrm{GH}} \sum_{\substack{i \in \{1,2\} \\ j \in \{3,4\}}} p_i p_j + \delta_{\mathrm{GL}}(p_1 p_5 + p_2 p_5) + \delta_{\mathrm{HL}}(p_3 p_5 + p_4 p_5) + \delta_{\mathrm{GG}} p_1 p_2 + \delta_{\mathrm{HH}} p_3 p_4 + \varepsilon$$

where the error term is assumed normally distributed with mean 0 and constant variance.

and the three-species Dg-Pl-To ($P < 0.001$) mixtures, while it performed comparably to the three-species Lp-Pl-Tr mixture ($P = 0.3$).

The best Diversity–Interactions model selected was the 'functional group' model (Table 1). This included a term for each of the five sown species proportions (identity effects) and five interaction terms that capture grass-legume, grass-herb, herb-legume, grass-grass and herb-herb interactions. For example, the coefficient for grass-legume quantifies the interaction between either of the grasses (Lp or Dg) and the legume (Tr), while the grass-grass coefficient quantifies the interaction between the two grass species (Lp and Dg).

Predictions for the design communities from the fitted Diversity–Interactions model gave similar results to the ANOVA analysis (Figure 2a vs. 2b). However, the Diversity–Interactions model can also show the contributions of the species identity and interaction effects (Figure 2b, segments in each bar). Including the legume Tr in any mixture boosted its performance due to its strong identity effect (Table 1: Tr coefficient is 56.7) and its ability to interact with grass (Table 1: grass-legume coefficient is 57.5) and herb species (Table 1: herb-legume coefficient is 34.5). Mixtures that included only grasses and herbs did not result in strong interaction effects (Figure 2b, small net interactions for those communities) because the positive grass-herb interaction effect was counteracted by either the negative grass-grass or herb-herb effects. This added knowledge the Diversity-Interactions model provides explanation as to *why* some communities differ, not only *if* they differ. For example, the Lp-Dg-Tr mixture yielded higher than the Lp-Dg-To mixture (the two only differ by Tr and To) and we can see that this is due to the high identity effect of Tr compared to To (Table 1, β_5 vs. β_4) combined with the higher grass-legume compared to grass-herb interaction (Table 1, δ_{GL} vs. δ_{GH}).

Not only can the fitted Diversity–Interactions model predict for the design points in the experiment, but it can also predict for any combination of proportions of the five species that are not explicitly in the design. This is because a Diversity–Interactions model is a regression model with continuous predictors (the sown proportions of each species) and thus we can interpolate to any point in the design

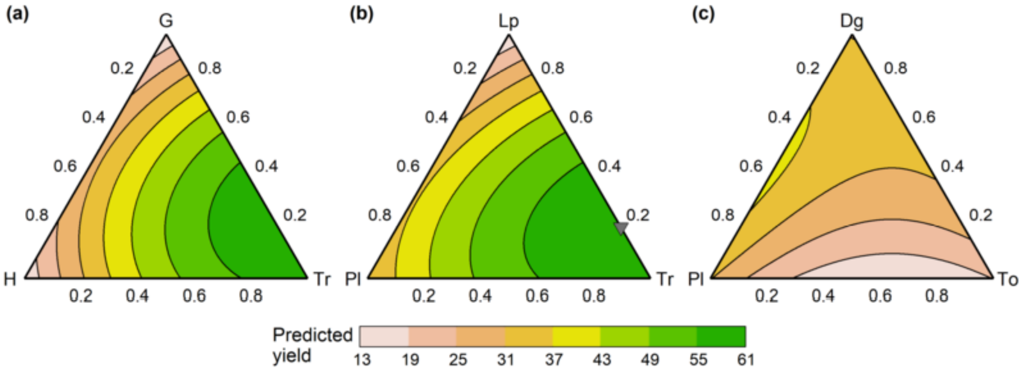

Figure 3. Predicted yield across different regions of the five-dimensional simplex space shown as contours within ternary diagrams. (a) The functional group ternary space with the total grass (G) and herb (H) proportion split equally between *L. perenne* (Lp) and *D. glomerata* (Dg), and between *P. lanceolata* (Pl) and *T. officinale* (To), respectively. For example, the vertex labelled G contains 100% of grasses and corresponds to a binary mixture containing 50% each of Lp and Dg. (b) All one- to three-species communities containing *L. perenne*, *P. lanceolata*, and *T. repens* (Tr). (c) All one- to three-species communities containing Dg, Pl and To. The best-performing mixture across the five sown species proportions falls within ternary (b) and is highlighted with a grey inverted triangle.

space. This means that if the 'best' mixture or mixtures lie outside of exact sown design communities, they can still be identified by Diversity–Interactions modelling. For example, we can predict for two- or four-species mixtures even though they were not in the design (Figure 2c) and obtain inference for those predictions. The 10:90 To:Tr, for instance, significantly outperforms the 90:10 To:Tr mixture ($p < 0.001$).

To illustrate the ability of the Diversity–Interactions model to predict around the design space, we focus on predicting for communities which include the legume species *T. repens* (given its strong performance) in Figure 2c, as well as for a range of five- and three species mixtures in a ternary diagram in Figure 3. In Figure 2c, for each vertical stack of pie-glyphs at each level of species richness, the proportion of Tr increases systematically as one moves upwards, showing that the higher the proportion of Tr sown, the higher the predicted yield. Any community containing over 60% of Tr has a predicted yield above 50 g mesocosm^{-1}. Moreover, when combining Tr with either of the grasses, this high performance is maintained, even for mixtures containing up to 60% grass (Figure 3a). As long as a mixture contains Tr, high yields can be consistently achieved regardless of the other species (Figures 3a and 3b). In contrast, mixtures not containing Tr had a lower yield up to approx. 80% relative to those containing high proportions of Tr (Figure 3c). The best-performing mixture contains 80% of *T. repens* and 20% *L. perenne* (grey point; Figure 3b). However, the predicted yield is rather consistent in a range around this point, indicating flexibility in selecting an optimal combination of sown species proportions for maximising performance. Thus, Diversity–Interactions modelling offers insights beyond those available from ANOVA by enabling predictions and inference for communities not present in the design. This offers more nuanced insights into the BEF relationship and helps identify best-performing ranges of species proportions as opposed to one or multiple best-performing communities solely from the *a priori* selected design communities.

Limitations and practical implications

There are some limitations with Diversity–Interactions models. A valid concern is whether they can be applied to model systems with lots of species. The approach has been applied successfully to data with nine (Connolly *et al.*, 2013), 16 (McKenna *et al.*, 2019) and 72 (Moral *et al.*, 2023) species, but it is important to balance parsimony with model usefulness when modelling many species (Brophy *et al.*, 2017). A second concern is the effect of the time gap between the assessment of species proportions and the measure of the response of interest. In this paper, we have continually referred to initial or sown species proportions. If modelling a response from a grassland study where the plots were established several months or years prior to recording the response, it is possible that realised proportions at some stage following sowing would be more useful predictors than sown proportions. When using realised proportions, weed proportion can also be included as a predictor. An important consideration is that the realised proportions must be recorded at a time point prior to recording the response; if they are recorded at the same time, then they are all responses and alternative approaches are required. The concern regarding changed predictors due to a time gap between establishment and recording the response refers not only to Diversity–Interactions models but is a general one. For example, realised species richness at a later time can differ greatly from the sown species richness, and realised species richness could therefore be a more appropriate predictor than sown richness in a richness model.

A major advantage of the Diversity–Interactions modelling approach is the ability to predict around the design space of species proportions. It is this aspect that allows us to identify regions that lead to the best-performing mixtures. In the literature (e.g. Lüscher *et al.*, 2022, 2025 this volume; Finn *et al.*, 2024), several practical lessons have been identified for the selection of the best-performing productive grassland mixtures. These include 1) select species that deliver satisfactory agronomic performance, e.g. high-yielding, highly digestible, tolerant of frequent defoliation, and have ability to acquire and use plant-available nitrogen; 2) target species that maximise complementarity for the response(s) of interest; and 3) carefully consider the species proportions, although sometimes quite a wide range of proportions of species and functional groups can lead to high-performing mixtures (Finn *et al.*, 2013; Grange *et al.*, 2021).

Through our discussion and examples, we have demonstrated the added benefits of fitting Diversity–Interactions models to BEF data. We conclude that Diversity–Interactions modelling can be a valuable tool for analysing data from BEF studies and can address more complex research questions than traditional approaches.

Acknowledgements

JAF and RV were supported by Multi4More (project ref. 21R456) funded by the Department of Agriculture, Food and the Marine funded by the Department of Agriculture, Food and the Marine and the Department of Agriculture, Environment and Rural Affairs. CB received funding from the European Union's Horizon 2021 doctoral network programme under the Marie Skłodowska-Curie grant agreement No. 101072579 (LegumeLegacy).

References

Abalos D., De Deyn G.B., Philippot L., Oram N.J., Oudová B., Pantelis I. … and van Groenigen J.W. (2021) Manipulating plant community composition to steer efficient N-cycling in intensively managed grasslands. *Journal of Applied Ecology* 58, 167–180.

Bélanger G., Tremblay G.F., Seguin P., Lajeunesse J., Bittman S. and Hunt D. (2020) Cutting management of alfalfa-based mixtures in contrasting agroclimatic regions. *Agronomy Journal* 112, 1160–1175.

Bell T., Newman J.A., Silverman B.W., Turner S.L. and Lilley A.K. (2005). The contribution of species richness and composition to bacterial services. *Nature* 436, 1157–1160.

Brophy C., Dooley Á., Kirwan L., Finn J. A., McDonnell J., Bell T. … and Connolly J. (2017) Biodiversity and ecosystem function: making sense of numerous species interactions in multi-species communities. *Ecology* 98, 1771–1778.

Craven D., Isbell F., Manning P., Connolly J., Bruelheide H., Ebeling A. … and Eisenhauer N. (2016) Plant diversity effects on grassland productivity are robust to both nutrient enrichment and drought. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371, 20150277.

Connolly J., Bell T., Bolger T., Brophy C., Carnus T., Finn J.A. … and Weigelt A. (2013) An improved model to predict the effects of changing biodiversity levels on ecosystem function. *Journal of Ecology* 101, 344–355.

Finn J.A., Kirwan L., Connolly J., Sebastià M.-T., Helgadottir A., Baadshaug O.H. … and Lüscher A. (2013) Ecosystem function enhanced by combining four functional types of plant species in intensively managed grassland mixtures: A 3-year continental-scale field experiment. *Journal of Applied Ecology* 50, 365–375.

Finn J.A., Suter M., Vishwakarma R., Oram N.J., Lüscher A. and Brophy C. (2024) Design principles for multi-species productive grasslands: Quantifying effects of diversity beyond richness. *Journal of Ecology* 112, 2471–2479.

Grange G., Finn J.A. and Brophy C. (2021) Plant diversity enhanced yield and mitigated drought impacts in intensively managed grassland communities. *Journal of Applied Ecology* 58, 1864–1875.

Hector A., Schmid B., Beierkuhnlein C., Caldeira M.C., Diemer M., Dimitrakopoulos P.G. … and Lawton J.H. (1999) Plant diversity and productivity experiments in European grasslands. *Science* 286, 1123–1127.

Hooper D.U., Chapin III F.S., Ewel J.J., Hector A., Inchausti P., Lavorel S. … and Schmid B. (2005) Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecological Monographs* 75, 3–35.

Isbell F., Adler P.R., Eisenhauer N., Fornara D., Kimmel K., Kremen C. … and Scherer-Lorenzen M. (2017) Benefits of increasing plant diversity in sustainable agroecosystems. *Journal of Ecology* 105, 871–879.

Kirwan L., Connolly J., Finn J.A., Brophy C., Lüscher A., Nyfeler D. and Sebastià M.-T. (2009) Diversity–interaction modeling: Estimating contributions of species identities and interactions to ecosystem function. *Ecology* 90, 2032–2038.

Komainda M., Küchenmeister F., Küchenmeister K., Kayser M., Wrage-Mönnig N. and Isselstein J. (2020) Drought tolerance is determined by species identity and functional group diversity rather than by species diversity within multi-species swards. *European Journal of Agronomy* 119, 126116.

Lüscher A., Barkaoui K., Finn J.A., Suter D., Suter M. and Volaire F. (2022) Using plant diversity to reduce vulnerability and increase drought resilience of permanent and sown productive grasslands. *Grass and Forage Science* 77, 235–246.

Marquard E., Weigelt A., Temperton V.M., Roscher C., Schumacher J., Buchmann N. … and Schmid B. (2009) Plant species richness and functional composition drive overyielding in a six-year grassland experiment. *Ecology* 90, 3290–3302.

McKenna T.P., McDonnell J., Yurkonis K.A. and Brophy C. (2019) *Helianthus maximiliani* and species fine-scale spatial pattern affect diversity interactions in reconstructed tallgrass prairies. *Ecology and Evolution* 9, 12171–12181.

Moral R.A., Vishwakarma R., Connolly J., Byrne L., Hurley C., Finn J.A. and Brophy C. (2023) Going beyond richness: Modelling the BEF relationship using species identity, evenness, richness and species interactions via the DImodels R package. *Methods in Ecology and Evolution* 14, 2250–2258.

Roscher C., Temperton V.M., Scherer-Lorenzen M., Schmitz M., Schumacher J., Schmid B. … and Schulze E.-D. (2005) Overyielding in experimental grassland communities – irrespective of species pool or spatial scale. *Ecology Letters* 8, 419–429.

Schläpfer F., Pfisterer A.B. and Schmid B. (2005) Non-random species extinction and plant production: implications for ecosystem functioning. *Journal of Applied Ecology* 42, 13–24.

Temperton V.M., Mwangi P.N., Scherer-Lorenzen M., Schmid B. and Buchmann N. (2007) Positive interactions between nitrogen-fixing legumes and four different neighbouring species in a biodiversity experiment. *Oecologia* 151, 190–205.

Tilman D., Isbell F. and Cowles J.M. (2014) Biodiversity and ecosystem functioning. *Annual Review of Ecology, Evolution, and Systematics* 45, 471–493.

Tilman D., Reich P.B., Knops J., Wedin D., Mielke T. and Lehman C. (2001) Diversity and productivity in a long-term grassland experiment. *Science* 294, 843–845.

Vishwakarma R., Brophy C. and Hurley C. (2023) PieGlyph: Axis Invariant Scatter Pie Plots (0.1.0). Available online at https://cran.r-project.org/web/packages/PieGlyph/index.html

Vishwakarma R., Brophy C., Byrne L. and Hurley C. (2024) DImodelsVis: Visualising and interpreting statistical models fit to compositional data. R package (1.0.1). Available online at https://cran.r-project.org/web/packages/DImodelsVis/index.html

Weisser W.W., Roscher C., Meyer S.T., Ebeling A., Luo G., Allan E. … and Eisenhauer N. (2017) Biodiversity effects on ecosystem functioning in a 15-year grassland experiment: Patterns, mechanisms, and open questions. *Basic and Applied Ecology* 23, 1–73.

Wilsey B.J. and Polley H.W. (2004) Realistically low species evenness does not alter grassland species-richness–productivity relationships. *Ecology* 85, 2693–2700.