OIKOS

Research article

Relationships between ecosystem functions vary among years and plots and are driven by plant species richness

Laura Argens[1], Wolfgang W. Weisser[1], Anne Ebeling[2], Nico Eisenhauer[3,4], Markus Lange[5], Yvonne Oelmann[6], Christiane Roscher[4,7], Holger Schielzeth[2], Bernhard Schmid[8], Wolfgang Wilcke[9] and Sebastian T. Meyer[1]

[1]Department for Life Science Systems, School of Life Sciences, Technical University of Munich, Terrestrial Ecology Research Group, Freising, Germany
[2]Population Ecology Group, Institute of Ecology and Evolution, Friedrich Schiller University, Jena, Germany
[3]Institute of Biology, Leipzig University, Leipzig, Germany
[4]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
[5]Max Planck Institute for Biogeochemistry, Jena, Germany
[6]Geoecology, University of Tübingen, Tübingen, Germany
[7]Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany
[8]University of Zurich, Zurich, Switzerland
[9]Institute of Geography and Geoecology, Karlsruhe Institute of Technology, Karlsruhe, Germany

Correspondence: Laura Argens (laura.argens@tum.de)

Ecosystem management aims at providing many ecosystem services simultaneously. Such ecosystem service multifunctionality can be limited by tradeoffs and increased by synergies among the underlying ecosystem functions (EF), which need to be understood to develop targeted management. Previous studies found differences in the correlation between EFs. We hypothesised that correlations between EFs are variable even under the controlled conditions of a field experiment and that seasonal and annual variation, plant species richness, and plot identity (identity effects of plots, such as the presence and proportion of functional groups) are drivers of these correlations. We used data on 31 EFs related to plants, consumers, and physical soil properties that were measured over 5 to 19 years, up to three times per year, in a temperate grassland experiment with 80 different plots, constituting six sown plant species richness levels (1, 2, 4, 8, 16, 60 species). We found that correlations between pairs of EFs were variable, and correlations between two particular EFs could range from weak to strong or negative to positive correlations among the repeated measurements. To determine the drivers of pairwise EF correlations, the covariance between EFs was partitioned into contributions from species richness, plot identity, and time (including years and seasons). We found that most of the covariance for synergies was explained by species richness (26.5%), whereas for tradeoffs, most covariance was explained by plot identity (29.5%). Additionally, some EF pairs were more affected by differences among years and seasons, showing a higher temporal variation. Therefore, correlations between two EFs from single measurements are insufficient to draw conclusions on tradeoffs and synergies. Consequently, pairs of EFs need to

www.oikosjournal.org

be measured repeatedly under different conditions to describe their relationships with more certainty and be able to derive recommendations for the management of grasslands.

Introduction

Land management and policy aim to improve human well-being by providing multiple ecosystem services, i.e. ecosystem service multifunctionality (Dade et al. 2019, Manning et al. 2018). The Millennium Ecosystem Assessment (2005) defined ecosystem services (ES) as the 'benefits people obtain from ecosystems, e.g. food, water, timber and cultural values. Ecosystem services derive from ecosystem functions (EF) (Balvanera et al. 2006, Costanza et al. 2017). Ecosystem functions describe the biogeochemical processes influenced by the organisms and their traits to sustain an ecosystem (Millennium ecosystem assessment 2005, Reiss et al. 2009). Ecosystem functions include a set of ecological processes and attributes, which can be characterised by various ecosystem variables, processes or attributes, allowing for detailed inquiries into the underlying function of an ecosystem (Weisser et al. 2017). The flow and exchange of materials and energy in ecosystems, i.e. the ecosystem functions, can be measured directly (Naeem 1998) or indirectly via ecosystem properties, such as storage and retention of water or nutrients (Costanza et al. 2017). In the last decades, the average global crop yields have been rising due to more intensive management practices in agriculture (Foley et al. 2011). These management practices had negative side effects on the environment, such as declines in native pollinators, increases in pests and diseases, and degrading land and water (Gordon et al. 2008, Foley et al. 2011). On the other hand, one important aim of nature conservation is to protect areas in order to preserve important ES, such as carbon sequestration and climate regulation, and to avoid widespread biodiversity declines (Watson and Venter 2017). While ES multifunctionality may be an implicit or explicit management aim, current management strategies often focus on providing single ecosystem services, e.g. maximising productivity or the value of nature conservation. For example, in agriculture, conventional farming practices may prioritise high crop yields or pest control while disregarding other ecological services like the biodiversity of pollinators or soil health (Mondelaers et al. 2009). ES multifunctionality requires EF multifunctionality (Manning et al. 2018). Since many ecosystem functions improve with increasing plant species richness (Scherber et al. 2010, Weisser et al. 2017), diversifying ecosystems have been proposed as an alternative management target, and studies have found a generally positive relationship between plant species richness and EF multifunctionality (Cardinale et al. 2006, Gamfeldt et al. 2008, Pasari et al. 2013, Dooley et al. 2015, Finney and Kaye 2017, Hautier et al. 2018, Meyer et al. 2018).

One challenge of promoting EF multifunctionality is that the simultaneous enhancement of all EFs is likely impossible because there are tradeoffs between EFs (Rodríguez et al. 2006, Manning et al. 2018, Meyer et al. 2018). Such tradeoffs occur when the provisioning of one EF improves at the expense of another EF. For example, under conventional management of single crops, high productivity often is associated with soil degradation (Kleinman et al. 2011, Pereira et al. 2023). In contrast, synergies among EFs occur when EFs are co-varying in the same direction (Rodríguez et al. 2006). For example, high below-ground biomass production is related to high below-ground carbon storage (Hanisch et al. 2020).

Two mechanisms can cause correlations between EFs. The first mechanism consists of common drivers affecting multiple EFs (Bennett et al. 2009), referred to as the common-driver mechanism in the following. Environmental conditions can improve one EF while deteriorating another EF (Bradford et al. 2014), thereby causing a tradeoff between the two EFs or a synergy if both EF would improve or deteriorate in the same way in response to the environmental condition. For example, Maestre et al. (2012) found that an increase in temperature decreased EF multifunctionality, which could indicate that either individual EFs are negatively affected by increasing temperature or that higher temperature can cause weaker synergies and/or stronger tradeoffs among EFs. The second mechanism consists of physiological or ecological constraints among EFs (Bennett et al. 2009), referred to as ecological-constraints-mechanism in the following. As resources are limited within an ecosystem, not all EFs can be improved simultaneously, independent of external drivers. Carbon sequestration, for example, can be enhanced by afforestation, but during tree growth, evapotranspiration is increased, and water availability deteriorates (Engel et al. 2005). Management strategies cannot easily overcome ecological constraints. Consequently, correlations among EFs need to be understood to mitigate tradeoffs and enhance synergies (Shen et al. 2020). One decision strategy for ecosystem management could be to consider the occurring species traits to avoid potential tradeoffs, as species traits link EFs with each other (Hanisch et al. 2020). An attempt to consider species traits is to maximise the number of species present, as each species possesses a large number of traits, or to consider functional groups, classifying groups of plant species according to plant traits, which seem more likely to influence EFs (Tilman 2001, Roscher et al. 2004). Consequently, correlations among EFs and the underlying drivers need to be understood to mitigate tradeoffs and enhance synergies (Shen et al. 2020), which is essential to manage ecosystems for ES multifunctionality.

Previous syntheses on the relationships between EFs reported inconsistent results, i.e. EFs are provided at unstable levels throughout time (Cardinale et al. 2012), indicating

unstable EF-relationships over time. Inconsistent results were also shown when ES were compared between studies in a meta-analysis (Lee and Lautenbach 2016). For example, for the ES 'Nutrition biomass' and 'Life cycle maintenance, habitat and gene pool protection', as many as 50–75% of the case studies reported a tradeoff, whereas 25–50% reported a synergy or no relationship between these ES. For the ESs 'Intellectual and representative interactions' and 'Physical and experiential interactions', 50–75% of the studies reported synergies, while 25–50% reported the opposite or no relationship. While Lee and Lautenbach (2016) investigated the variability of ES relationships, the results also hold relevance for relationships between EFs and underline that the causes of these conflicting results are still subject to debate (Dade et al. 2019).

There are several possibilities why the relationship between two particular EFs or ESs could differ among studies. First, the relationship between EFs or ESs can change based on the scale or land system considered, e.g. urban area versus agricultural area (Adhikari and Hartemink 2016, Lee and Lautenbach 2016). Second, most studies investigated EF- or ES relationships based on single measurements. However, ecological drivers, such as diversity or nutrient availability, can change over time and cause variations in relationships between EFs (Crouzat et al. 2015, Torralba et al. 2018, Zheng et al. 2019). Third, differences in the ecosystem investigated, or in abiotic conditions among sites, can cause variation regarding EF relationships among studies. Land-use type (Li et al. 2018), management intensity (Rodríguez et al. 2006), and environmental factors like climate and soil pH have been shown to strongly affect individual EFs (Wang et al. 2021), and the correlations between EFs (Spake et al. 2017). If these drivers affect EFs differently, a change in the driver will change the relationship between these EFs. One example would be EFs dependent on water availability, such as shoot length and root length, being positively related within a year of high precipitation (Pérez-Ramos et al. 2012) and showing a weaker relationship at low precipitation when plants invest more in roots than shoots (Mokany et al. 2006). In addition, previous studies have found that drivers of individual EFs are of different importance at different places and time points (Isbell et al. 2011, Crouzat et al. 2015, Torralba et al. 2018, Zheng et al. 2019, Martin et al. 2020, Shen et al. 2020, Willemen 2020). This implies that also the variability in EF relationships may differ among places and time points as these drivers can influence EF relationships directly by changing the ecological dependency of the two EFs or indirectly by affecting EFs individually and, therefore, causing a change in their covariance. Finally, also differences in the statistical methods used to evaluate relationships between ESs can bias results (Lee and Lautenbach 2016). For example, no-effect relationships were more likely to be found when correlation coefficients were used, whereas descriptive methods such as GIS analyses, which quantify and describe ES relationships based on the cooccurrence of ESs at the same location, showed a higher probability of identifying tradeoffs (Lee and Lautenbach 2016). In summary, there are several

reasons why relationships between different EFs and ESs may vary. Whereas a few studies recorded the variation of individual EFs (van der Plas et al. 2020) and their drivers over time (Gaglio et al. 2020, van der Plas et al. 2020), such studies are lacking for EF relationships.

To understand whether EF relationships are inherently variable or whether meta-analyses detected variability because of differences among studies, studies investigating EF relationships repeatedly under comparable conditions are needed. Furthermore, the drivers of EF relationships need to be investigated to understand what might cause variability in EF relationships. Drivers and variability of EF relationships might depend on the individual EFs or their proxies investigated. For example, it was shown that plant diversity has particularly strong effects on lower trophic levels and effects dampen with increasing trophic levels (Scherber et al. 2010). Consequently, it can be expected that EFs depending on different components of the ecosystem (e.g. plant productivity and soil microbes) show different EF relationships or a higher variability of EF relationships. Furthermore, we expect to see similar EF relationships between EFs depending on the same components of the ecosystem, e.g. between EFs representing plant productivity and EFs representing invasion resistance.

Here we used data from 31 EFs repeatedly measured during 5–19 years in a large-scale temperate grassland biodiversity experiment, i.e. the Jena Experiment (Roscher et al. 2004, Weisser et al. 2017). The 31 EFs covered different components of the ecosystem related to plant productivity, plant nutrients, soil microbes, consumers, invasion resistance, soil properties, and soil nitrogen and carbon concentrations, which are called classes of EFs hereafter. Our study aimed to systematically investigate the variability in the pairwise relationships between EFs and the underlying drivers of variability. Specifically, we addressed the following questions:

1. How variable are EF relationships over replicated measurements under the controlled conditions of a field experiment?
2. What drives the relationship among EFs? How much do years, seasons, species richness and the identity of the plots contribute to these relationships by affecting pairs of EFs in similar or opposing ways?
3. Are synergies and tradeoffs driven differently by years, seasons, plant species richness and the identity of the studied plots?

Material and methods

Study site

In 2002, the Jena Experiment, a biodiversity experiment with 82 plots, was established at a former arable field near the city of Jena (Germany) (Roscher et al. 2004, Weisser et al. 2017). The plots were sown in May 2002 with species richness (SR) of 1, 2, 4, 8, 16 and 60 grassland plant species, with 16, 16, 16, 16, 14 and 4 replicates, respectively (each replicate was a

unique species composition, i.e. community, except for the highest richness level where all replicates had the same species composition). Plot identity ('plotID') represents unique plots containing different plant communities with diverse compositional features (Jochum et al. 2020). The experiment reduces stochastic variation by standardising abiotic conditions across all plots. However, the lack of replicates at the species composition level prevents us from determining if the effects are solely due to species composition. Additionally, species composition may have changed over time, along with abiotic factors, contributing to the net differences observed in plotID (Wagg et al. 2022). Plant species for communities with 1–16 species were randomly chosen from a pool of 60 plant species typical for *Arrhenatherum* grasslands with restrictions to create different levels of functional-group richness within each level of species richness. We distinguished three functional groups, namely grasses, herbs (small herbs and tall herbs combined), and legumes, based on ecologically relevant attributes (Roscher et al. 2004). Species richness and functional group richness (FGR, number of functional groups per community) were varied as independently as possible (Roscher et al. 2004). All plots were mown twice a year, did not receive any fertiliser, and were weeded two to

three times a year (Roscher et al. 2004). The chosen mowing regime corresponds to the region's typical management of extensively used hay meadows (Weisser et al. 2017). Two monocultures were given up due to the weak establishment of the target species in the first years, resulting in 80 plots used for this analysis.

Dataset

We based this analysis on 31 EFs measured during 5–19 years in the Jena Experiment (full description in the Supporting information). These EFs are indicative of eight classes of EFs: plant productivity, plant nutrients, soil microbes, consumers, invasion resistance, soil carbon, soil nitrogen and soil properties (Table 1). The EFs within one class of EFs are often related.

The data were categorised into spring (March, April, May), summer (June, July, August), and autumn (September, October, November) according to the meteorological seasons of the Northern Hemisphere (while in winter, no measurements were taken). In the case of multiple measurements of the same EF per season and year, the raw data were averaged per plot, year, and season. The EFs were always measured

Table 1. List of all Ecosystem functions (EF), the classes of EFs they represent, the abbreviations for the EFs used in the following, and if EFs were inverted to represent a valuable function according to humans' perspective.

Class of EF	Ecosystem function (EF)	Abbreviation for EF	Inverted
Consumer	Herbivory damage in %	Herbivory	no
	Predation as bite-marks on dummy caterpillars in%	Predation	no
Invasion resistance	Invasion resistance as biomass on weeded subplots (inverse of biomass of weeds)	$\text{WeedBM}^{(-1)}$	yes
	Invasion resistance as cover (inverse of weed cover)	$\text{WeedCover}^{(-1)}$	yes
	Invasion resistance as species number on not weeded subplots (inverse of the number of weed species)	$\text{WeedSPr}_{nw}^{(-1)}$	yes
	Invasion resistance as species number on weeded subplots (inverse of the number of weed species)	$\text{WeedSPr}_{w}^{(-1)}$	yes
Plant productivity	Standing root biomass	RootBM	no
	Biomass of target plants	ShootBM	no
	Total cover (inverse of bare ground)	$\text{BareGround}^{(-1)}$	yes
	Cover of target plants	PlantCover	no
	Plant height	PlantHeight	no
	Leaf area index	LAI	no
Plant nutrients	Plant tissue carbon (C) concentration	PlantC	no
	Plant tissue calcium (Ca) concentration	PlantCa	no
	Plant tissue potassium (K) concentration	PlantK	no
	Plant tissue magnesium (Mg) concentration	PlantMg	no
	Plant tissue nitrogen (N) concentration	PlantN	no
	Plant tissue sodium (Na) concentration	PlantNa	no
	Plant tissue phosphorus (P) concentration	PlantP	no
Soil carbon	Inorganic carbon concentration	SoilC_{inorg}	no
	Organic carbon concentration	SoilC_{org}	no
	Dissolved inorganic carbon concentration in soil solution	SoilDIC	no
	Dissolved organic carbon concentration in soil solution	SoilDOC	no
Soil microbes	Biomass of microbes	MicrobeBM	no
	Basal respiration	SoilResp	no
Soil nitrogen	Total nitrogen concentration in soil	SoilN	no
	NH_4-N concentration in soil	SoilNH_4	no
	NO_3-N concentrations in soil	SoilNO_3	no
	Mineral nitrogen concentration (sum of the NH_4-N and NO_3-N concentrations) in soil	SoilN_{min}	no
Soil properties	Looseness of soil (Inverse of soil bulk density)	$\text{SoilDensity}^{(-1)}$	yes
	Water content in the soil	SoilH_2O	no

on all plots but in different numbers of years and seasons. The number of years ranged from 5 to 19, and most EFs were measured once or twice a year. A dataset comprising all plots is referred to as a measurement in the following. The number of measurements ranged from a minimum of five (SoilDensity) to a maximum of 36 (PlantHeight). The inverse of some EFs was used to represent a valuable function according to humans' perspective, enabling to identify synergies and tradeoffs (Table 1).

Data preparation

All data manipulations and later analyses were conducted using the free software R ver. 3.6.2 (www.r-project.org) and Rstudio ver. 1.1.442 (RStudio Team 2016). To approximate normal distributions, all raw data (EF_{raw}) were boxcox-transformed (EF_{boxcox}) using two lambda-values (λ and λ_2) estimated with the package 'geoR' (Ribeiro 2020):

$$EF_{boxcox} = boxcox\rangle transformed(EF) = \frac{EF_{raw} + \lambda_2^{\lambda} - 1}{\lambda}$$

To scale all EFs to a comparable range of 0–1, the EF_{boxcox} were minmax-transformed (EF_{minmax}):

$$EF_{minmax} = minmax\rangle transformed(EF)$$

$$= \frac{EF_{boxcox} - min(EF_{boxcox})}{(EF_{boxcox}) - min(EF_{boxcox})}$$

Variation in individual EFs

The variation of individual EFs was quantified as a standard deviation over all data points (individual measures on plots). The individual EFs are not consistently measured at the same time. However, there is sufficient overlap in the timing of measurements (Supporting information). Variation of individual EFs is expected to be comparable and not biased by the identity of years and seasons measurements were taken. However, we tested whether the variation of individual EFs depended on the number of repeated measures, meaning how often in time EFs were measured (number of years × number of seasons). Therefore, a model with the standard deviation per individual EF depending on the explanatory variable 'number of repeated measures' (number of years × number of seasons an individual EFs was measured) was run.

The drivers of the variation in individual EFs (EF_{minmax}) were tested in a linear model with the explanatory terms 'block' (factor with four levels), 'SR' (initial number of species planted, log-transformed continuous variable), 'plotID' (factor with 80 levels), 'season' (factor with three levels, as no measurements were done in winter), 'year' (continuous variable), and their interactions (model 1). The plot identity (plotID) effect mainly accounts for differences among the initially planted communities. This set of terms is referred to

as 'drivers' in the following. The same model was conducted for EFs measured only once per year, excluding 'season' and the respective interaction terms (model 2).

$$model(1) \quad X \sim block + log(SR) + plotID + season + year$$
$$+ season:year + log(SR):(season + year + season:year)$$
$$+ plotID:(season + year + season:year)$$
$$model(2) \quad X \sim block + log(SR) + plotID + year$$
$$+ log(SR):year) + plotID:year$$

(note that here, following conventions of R, we use the colon instead of a multiplication sign as an interaction operator).

To analyse whether different classes of EFs were affected differently by drivers, the variance in individual EFs explained by individual drivers was calculated by dividing the sum of squares explained by the driver by the total sum of squares in the respective model of the individual EFs explained above. In a subsequent model, the explained variation per EF and per driver was used as meta-data. The variation was tested against the classes of EFs (with the different classes of EFs as levels) and the drivers (with the levels 'block', 'SR', 'plotID', 'season', 'year' and their interactions) as independent variables.

Relationships between pairs of EFs

Relationships between EF pairs were statistically investigated using covariances and correlations. In correlations, the relationship between two EFs was standardised by the variation of the individual EFs (product of their standard deviations), enabling us to compare relationships between different EF pairs. To calculate correlation coefficients, we used the R-package 'Hmisc' ver. 4.4-2 (www.r-project.org, Harrell and Dupont 2020). We used the non-standardised relationships (covariances) to analyse the influence of drivers on relationships among EFs.

Variation in EF correlations
To quantify the general strength and variation of EF correlations, we calculated the mean and the standard deviation of Fisher's Z-transformed correlation coefficients for each EF pair. Correlation coefficients were calculated among measurements on all plots at a particular time point and then averaged across time points. Hence we refer to this correlation as the mean correlation. It includes the effects of species richness and plot identity. In order to plot the EF relationships as correlation coefficients on a scale of −1 (perfect negative) to 1 (perfect positive correlation), the mean correlation coefficients were back-transformed from Z-scale. Mean correlations were defined as no correlation ($-0.1 > r < 0.1$), weak ($|r| > 0.1$ and $|r| < 0.3$), moderate ($|r| > 0.3$ and $|r| < 0.5$), and strong ($|r| > 0.5$) correlations after Cohen (2013).

The standard deviation of the individual correlations at the different time points quantifies the temporal

variation (among seasons and years) of correlations among EFs. However, using all time points to calculate the temporal variation might be influenced by the number of time points and the identity of time points (deviating years or seasons). Therefore, first, we checked whether this temporal variation, based on all time points, depended on the number of time points. We analysed the temporal variation of the correlations per EF pair as a function of the number of time points that an EF pair was measured (number of years × number of seasons). The number of repeated measures for pairwise EFs, meaning the number of times two EFs were measured at the same time (same year and same season), ranged from 0 to 36 times (the Supporting information contains an overview of the individual EFs and at what time (years and seasons) they were measured). Second, we checked whether the variation of correlations per EF pair depended on the identity of the time point that the EF pair was measured. Therefore, for each EF pair, we randomly chose four time points to calculate a standard deviation of the respective correlation coefficients. For each EF pair, this was done 20 times. The range of these 20 standard deviations per EF pair was used to check whether the standard deviation for that EF pair was stable (small range indicating no identity effect of years and seasons) or not (large range indicating strong identity effects of years or seasons).

Drivers of the covariance between EF pairs

To analyse whether years, seasons, species richness, and plot identity affect EF relationships by driving individual EFs in similar or opposing ways, we partitioned overall covariances into contributions of the different explanatory terms. Here, plot identity was further decomposed in the effects of functional group richness and the presence of the functional groups legumes, herbs (tall and short herbs combined), or grasses. This decomposition of covariances was based on an additive partitioning of sums of products (SPs) in the same way as additive partitioning of the sum of squares (SS) is used in a decomposition of variances in an analysis of variance (ANOVA). This type of covariance analysis has previously been used to investigate, for example, the influence of explanatory terms on trait–trait relationships (He et al. 2009) and is frequently used in quantitative genetic and phylogenetic approaches (Kempthorne 1957, Bell 1989).

The decomposition of covariances includes the following steps:

Preservation of the sum of squares (SS) for each EF pair
The SS were obtained from general linear models, implemented with the *lm()* function in R (Mangiafico 2015). We run the decomposition of covariances twice: Once to estimate the general effect of species richness (SR), plot identity (plotID), season, year and their interactions on EF-relationships (model 1 and model 2). Model 1 encompassed all explanatory terms, while model 2 excluded 'season' and its interaction terms, specifically targeting EFs measured once per year (similar to the analysis on the individual EFs). Second, we run the decomposition of covariances additionally to estimate the effect of the functional groups

(grasses, herbs and legumes; FG_i), inserted as sown proportions per plotID, and the functional group richness (FGR) (model 3 and model 4). To be able to investigate all pairs of EFs, model 3 encompassed all explanatory terms, while model 4 excluded 'season' and its interaction terms. Model 3 and model 4 were executed individually for each functional group, meaning that the following steps were carried out separately for each functional group.

$$\text{model} (3) \quad X \sim \text{block} + \log(\text{SR}) + \text{FGR} + \text{FG}_i$$
$$+ \text{plotID} + \text{season} + \text{year} + \text{season:year}$$
$$+ \log(\text{SR}) : (\text{season} + \text{year} + \text{season:year})$$
$$+ \text{FGR} : (\text{season} + \text{year} + \text{season:year})$$
$$+ \text{FG}_i : (\text{season} + \text{year} + \text{season:year})$$
$$+ \text{plotID} : (\text{season} + \text{year} + \text{season:year})$$

$$\text{model} (4) \quad X \sim \text{block} + \log(\text{SR}) + \text{FGR}$$
$$+ \text{FG}_i + \text{plotID} + \text{year}$$
$$+ \log(\text{SR}) : \text{year}) + \text{FGR:year}$$
$$+ \text{FG}_i : \text{year} + \text{plotID:year}$$

(note that here, following conventions of R, we use the colon instead of a multiplication sign as an interaction operator).

To summarise, for the first decomposition analysis, we used model 1 and model 2 to examine the drivers (SR, plotID, season, year and their interactions), for the second decomposition analysis, we used model 3 and model 4 for each functional group, to investigate the drivers and additionally the effects of FGR and the proportions of grasses, herbs and legumes present in the plant communities. The following steps were done for each of the decompositions:

For each EF pair, the linear models were run three times: one for each of the individual EFs (X and Y) and one for the sum of the two EFs (X + Y), based on the measurements from different time points of $\text{EF}_{\text{minmax}}$. For EF-pairs, which were just present in one season, the same models were run, reduced by the term 'season' and its interactions (model 2 or model 4).

Calculating sums of products for each EF pair
The sums of products, which are equivalent to covariances, were obtained per EF pair using the following formula:

$$\text{SP}(X,Y) = \frac{\text{SS}(X+Y) - \text{SS}(X) - \text{SS}(Y)}{2}$$

where X and Y are the EFs of interest, and $X + Y$ is the sum of the two EFs.

Like in ANOVA, SPs were divided by their degrees of freedom to obtain mean SPs (MSPs), which were divided by residual MSP to calculate F-ratios and significances. Because there are nested effects, not all terms could be tested against 'Residuals'. 'Block', 'log(SR)', and 'FGR' had to be tested at the level of variation between plots with different species compositions (plotID). Similarly, the interaction terms 'log(SR):(season + year + season:year)', 'FGR:(season + year + season:year)', and 'FG$_l$:(season + year + season:year)' had to be tested for the same reason against 'plotID:(season + year + season:year)'. All other terms were tested against 'Residuals' (Supporting information). It has been shown that for balanced experimental designs such as the Jena Experiment, this method is comparable to linear mixed-model analysis using restricted maximum likelihood methods (Schmid et al. 2017).

Calculating the percentage per driver per EF pair and separating EF pairs into synergies and drivers

Because SPs are additive, we can express the influence of each driver on EF covariation (i.e. the relationship between the EFs) by calculating the percentage of the (absolute) total sum of products explained, similar to a percentage variance explained (He et al. 2009). However, unlike variances, covariances are either positive, indicating a positive relationship between two variables, or negative, indicating an inverse (i.e. tradeoff) relationship between two variables. The sign of SPs for each explanatory term informs us about whether covariances are positive or negative. This means that we could deduce whether the individual drivers affected the EFs in a pair in an antagonistic (negative covariance) or synergistic (positive covariance) way. Therefore, we show 'signed percentages' of covariance in the results by multiplying the absolute percentages with the sign of the respective covariance.

Sensitivity analysis

Per combination of EF-class, one EF-pair is drawn for representation, and the results of the decomposition of covariances were stored. This results in 27 EF-pairs, representing each one combination of EF classes (e.g. plant productivity and soil nutrients). From this subset, the results of the decomposition for these 27 EF-pairs were used to calculate average values per driver and per correlation type (synergy versus tradeoff). By repeating this process 100 times, an average effect per driver and correlation type could be estimated, while the range of minimum and maximum values indicated the influence of the identity of EF-pairs and, consequently, the influence of the number of EF-pairs within each combination of EF-classes on the results of the decomposition.

Results

Variation in individual EFs

First, we compared the variation of individual EFs and EF classes. The average standard deviation, calculated by averaging all standard deviations of all EFs, was 0.17. While some EFs varied strongly in time among replicated measures, other EFs showed a low variation (Table 2; the minimum standard deviation was 0.07 for plant carbon, and the maximum standard deviation was 0.38 for plant sodium). The variation of individual EFs did not depend on the number of times (number of years × number of seasons) they were measured ($F_{1,29} = 0.753$, $p = 0.393$) (Supporting information). Classes of EFs did not differ significantly in their variation ($F_{7,23} = 0.76$, $p = 0.63$; Table 2).

Second, we tested if measures of individual EFs differed among years, seasons, species richness levels, and plot identities (Fig. 1). Considering all EFs, year explained on average 4.7% of the variation of EFs, and season explained on average 4.1%. Additionally, species richness explained 8.6%, while plot identity explained 21.3% of the variation of individual EFs. These differences explained about one-fourth of the variation of the individual EFs. One-third of the variation in individual EFs was unexplained: 34.1% for EFs, which were measured in several seasons, and 48.3% for EFs measured in just one season. All tested interaction terms explained only a small part of the total variation of the EFs (Fig. 1).

Different variables explained the variation of EFs in different classes of EFs. For example, for invasion resistance and plant productivity, species richness explained a large proportion of the variation (on average, 23.1 and 21.0%; Fig. 1, green). For consumer-related functions, year explained a large proportion (on average 11.9%; Fig. 1, red). For plant nutrients, plot identity explained, on average, 38% of the variation (Fig. 1, blue). This means that classes of EFs were differently affected by biological and environmental conditions (classes of EF: $F_{7,245} = 4.2$, $p = < 0.01$; Driver $F_{12,245} = 38.35$, $p < 0.01$; classes of EF: Driver: $F_{60,245} = 5.4$, $p < 0.01$, where 'driver' represents the year, season, SR, plotID and their interactions).

Variation in EF correlations

Positive correlations (indicating synergies) and negative correlations (indicating tradeoffs) were observed across all measures (Fig. 2, upper triangle). For instance, plant height and shoot biomass showed a synergy, while soil-dissolved carbon and plant height showed a tradeoff in their mean correlations. The strength of these correlations (defined after Cohen 2013) differed among pairs of EFs, with some EF pairs showed no correlation, while others showed weak, moderate, or strong correlations. We observed no strong negative correlations. All EFs showed positive correlations to some and negative correlations to other EFs (according to their mean correlation) (Fig. 2, upper triangle), with EFs in some classes showing predominantly positive correlations (plant productivity and invasion resistance) and others mostly negative correlations (plant productivity and plant nutrients). The EF correlations were robust against the method of calculating correlations, i.e. whether we used mean correlations (correlation coefficient averaged across time points), grand-total correlations (one correlation coefficient using all data from all time points), or between-group correlations (one correlation

Table 2. Variation in ecosystem functions (EF) expressed as standard deviation. The standard deviation over all measurements and the average standard deviation per class of EF correlations are listed. The table also lists the number of years and seasons the EFs were measured.

Classes of EF	EF	n years	n seasons	Standard deviation	Average standard deviation per class of EF
Consumer	Herbivory	5	2	0.12	0.2
	Predation	7	2	0.27	
Invasion resistance	WeedBM$^{(-1)}$	16	2	0.18	0.18
	WeedCover$^{(-1)}$	18	2	0.19	
	WeedSPr$_{nw}$$^{(-1)}$	5	2	0.16	
	WeedSPr$_{w}$$^{(-1)}$	5	2	0.19	
Plant productivity	RootBM	7	2	0.17	0.17
	ShootBM	18	2	0.16	
	BareGround$^{(-1)}$	18	2	0.25	
	PlantCover	18	2	0.14	
	PlantHeight	18	2	0.15	
	LAI	18	2	0.13	
Plant nutrients	PlantC	10	2	0.07	0.16
	PlantCa	5	2	0.09	
	PlantK	5	2	0.16	
	PlantMg	5	2	0.14	
	PlantN	10	2	0.14	
	PlantNa	5	2	0.38	
	PlantP	5	2	0.12	
Soil carbon	SoilC$_{inorg}$	5	1	0.28	0.23
	SoilC$_{org}$	5	1	0.18	
	SoilDIC	4	1	0.28	
	SoilDOC	4	1	0.17	
Soil microbes	MicrobeBM	17	1	0.13	0.14
	SoilResp	17	1	0.14	
Soil nitrogen	SoilN	5	1	0.20	0.12
	SoilNH$_4$	16	1	0.17	
	SoilN$_{min}$	16	1	0.12	
	SoilNO3	16	1	0.14	
Soil properties	SoilDensity$^{(-1)}$	5	1	0.17	0.16
	SoilH$_2$O	15	1	0.07	

coefficient calculated with data points averaged across time points) (Supporting information). As expected, EF correlations tended to be stronger when the variation of individual measurements across time points was removed, i.e. for the between-group correlations (Supporting information).

The variation per EF pair was quantified by the standard deviation of correlation coefficients, which were calculated for every time point when the two EFs were measured in the same year and season. Overall, there was considerable variation in EF correlations (mean SD = 0.16, mean correlation coefficient = 0.14) (Fig. 2, lower triangle). We tested if the variation in the EF correlations depended on the correlation's average strength. EF pairs generally showed a higher variation in their correlation when they showed a stronger correlation irrespective of this correlation being positive or negative ($F_{5,281} = 9.4$, $p < 0.001$, Supporting information). Additionally, we tested if the variation in the EF correlations depended on the number of times the EF-pair was measured (number of years × number of seasons). EF pairs generally showed an increasing variation in their correlations with a higher number of times the EF-pair was measured ($F_{1,572} = 120.91$, $p < 0.001$) (Supporting information). Lastly, we checked whether the variation in correlations among EFs depended on the identity of time points

they were measured. These ranges of temporal variation were rather small, on average showing a SD ± 0.08 (Supporting information). Furthermore, the range of temporal variation of correlations was different for the individual EF-pairs, some EF-pairs showed a strong identity effect of time points (e.g. SoilNH4__SoilN$_{min}$, PlantCover__WeedCover) and some a weak identity effect of time points (e.g. ShootBM__SoilN, PlantC__SoilC$_{org}$) (Supporting information).

Drivers of the covariance of EF pairs

To test if the drivers year or season, species richness, and plot identity affect relationships among EFs, we quantified the covariance between all pairs of EFs and the contribution of each driver to these covariances in percentage. These percentages were signed because the drivers can contribute to the EF covariances by affecting the underlying EFs synergistically (signed positive) or antagonistically (signed negative). Importantly, the contribution of individual drivers can have antagonistic (more negative covariance) or synergistic (more positive covariance) effects irrespective of the overall relationship between the respective EF being a synergy or a tradeoff.

All tested drivers (year, season, SR and plot ID) affected the covariances between EFs. The largest fraction of covariance
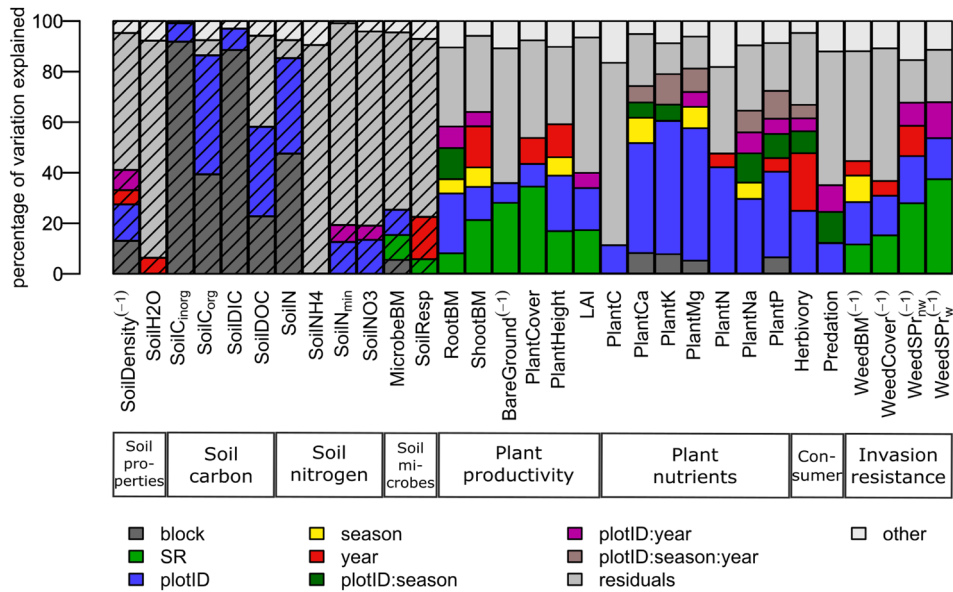
Figure 1. Percentage of variation in individual EFs that was explained by year, season, species richness (SR), plot identity (plotID), and the interaction among these variables. The influence of the explanatory terms is plotted in% of the total sum of squares, corresponding to increments in multiple $R^2 \times 100$. Explanatory terms are plotted for individual effects > 5%. All effects less < 5% were summarised as 'other', e.g. various interaction effects. Hatched barplots represent a simpler model, including only year, SR, plotID and their interactions for EF, which were measured in only one season. Non-hatched barplots represent full models, including all terms. The graph corresponds to a hierarchical partitioning of type one, but because explanatory terms were not correlated, there was no need to average across different fitting sequences.
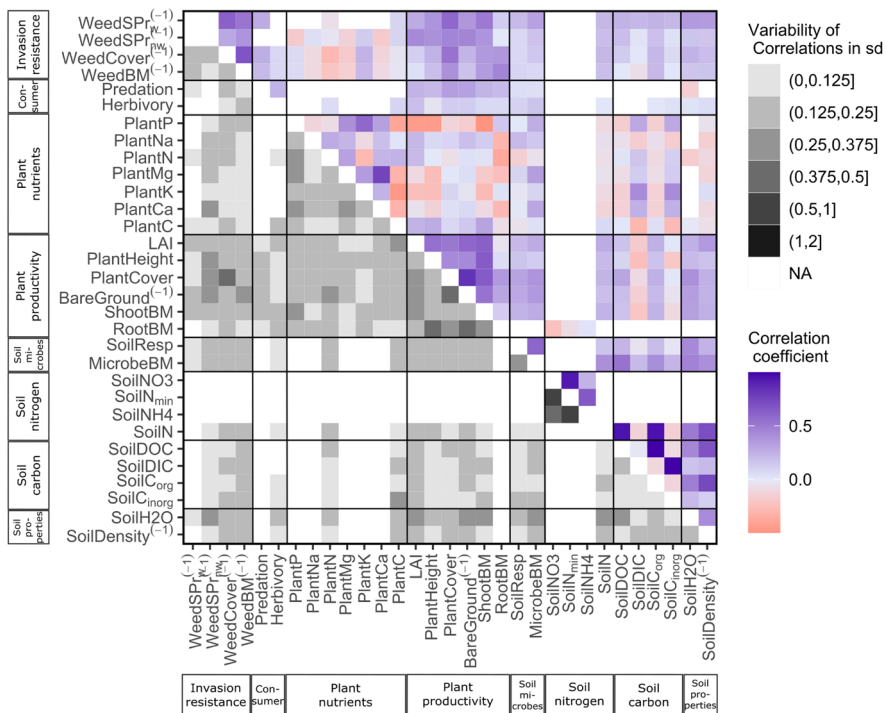

Figure 2. Variation in the correlations between pairs of ecosystem functions (EFs) (lower triangle) and average correlation between these EFs (upper triangle). The different EFs (list in Table 1) were grouped into classes. Mean correlations were calculated using Fisher's Z transformation of EF correlations per season and year that were averaged over time. The standard deviation of the EF correlations per year and season was calculated to estimate the variation of EF correlations. When no average correlation is shown, the respective EF was not measured in the same season and year. A missing standard deviation for an EF pair shown to have a correlation coefficient represents cases where a correlation coefficient could only be calculated for a single time point.

among EFs was explained by SR and plotID. However, effects differed between EF pairs in synergies and tradeoffs (defined by the sign of the mean correlation, Fig. 2, upper triangle). For synergies, most of the covariance was explained by SR (26.8%). In contrast, for tradeoffs, most of the covariance was explained by plot identity (–18.3%, Fig. 3), with the negative value indicating that the individual EFs were driven antagonistically, causing a tradeoff. When further investigating plotID, the proportion of herbs and legumes already explained half of the effect of plot ID in tradeoffs (Supporting information). In more detail, the proportion of legumes and herbs seemed to have a positive effect on many tradeoffs. However, there are strong negative effects on a few EF pairs, resulting in a negative mean (legumes: mean −5.83% median −0.7%, herbs: mean –8.21% median –6.05%). For synergies, plot ID had intermediate positive effects (12.9%), partially due to the proportion of grasses and legumes. Also, here, it seems that the effect of grasses and legumes is positive for the majority of EF pairs. However, there are a few strong negative effects, causing a negative trend in EF relationships (grasses: mean 1.43% median 0.13%, legumes: mean 2.35% median 0.27%). Year and

season caused both positive and negative covariances so that the average percentages explained by year and season were low (3.1 and –0.6%). For tradeoffs, the average percentage of explained covariance by SR was low (–1.6%), contributing positively and negatively to covariance. Season contributed an additional –12.1% to covariance, while year explained very little (–3.7%). Interactions between drivers explained very little covariance (Supporting information). Unexplained residual covariance was, on average, |1.9%| of the covariance (for synergies 2.1% and for tradeoffs –1.6%), suggesting a low amount of random covariation between EFs.

To test whether the pattern in the drivers depend on the identity of the EF-pairs included, we did a sensitivity analysis resulting in an average effect per driver and per correlation type, and a range of minimum- and maximum values per driver and correlation type, both indicating how much the results of the decomposition were influenced by the identity of the EF-pairs and therefore the number of EF-pairs within one combination of EF-classes. The results of the sensitivity analysis showed that the average effect per driver in the sensitivity analysis (Supporting information) is comparable to the average effect per driver shown in the results of the decomposition on all EF pairs (Fig. 3). Synergies were less influenced by the identity of EF-pairs than the tradeoffs, indicated by the range of effect per driver (Supporting information). For the tradeoffs, the explained covariance in% showed a higher variability per driver, especially the drivers plotID and season depended more on the identity of the EF-pairs present (Supporting information). However, differences and patterns between the drivers are strong for both synergies and tradeoffs.

Discussion

We investigated the variation in the correlations between different EFs and the drivers of these relationships. We found that correlations were variable, and correlations between two particular EFs could range from weak to strong or from negative to positive among the repeated measurements. Overall, EF pairs generally showed an increasing variation in their correlations with a higher number of times the EF-pair was measured. The correlations among pairwise EFs were differently affected by the identity of time points (years and seasons). That means that some EF-pairs showed more stable correlations throughout time, whereas other EF-pairs were more affected by differences in years and seasons and therefore showed a higher temporal variation. Species richness and plot identity (including the proportions of legumes, grasses, and herbs) explained the largest fraction of covariance among EFs, while the effects of time (year, season and their interaction) explained little covariance. We found that most of the covariance for synergies was explained by species richness (26.8%), whereas for tradeoffs, most covariance was explained by plot identity (−18.3%). Time explained −15.8% of covariance for tradeoffs but little for synergies (|3.7%|). Correlations among EFs and the drivers of these correlations varied over time. These results indicate the
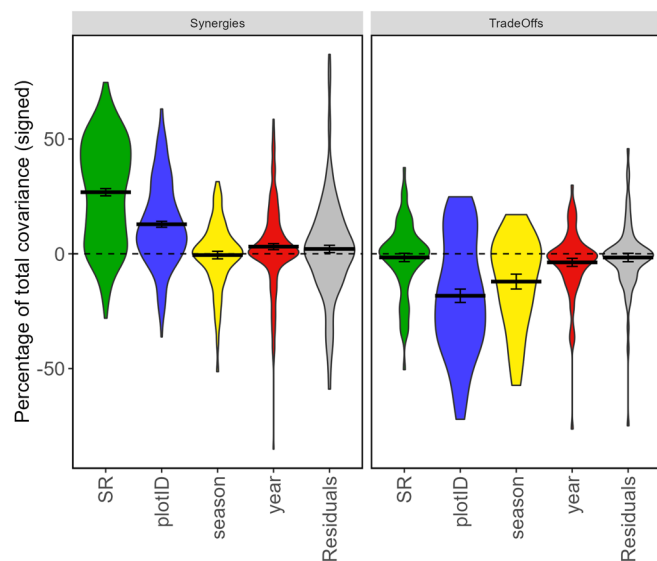

Figure 3. The contribution of SR, plotID, season and year to total covariance between the 116 pairs of EFs, separated for EF pairs showing positive relationships (synergies, 78 pairs) and negative relationships (tradeoffs, 38 EF pairs) according to their mean correlation. The violin plots show for each driver the mean (solid line), the standard error, and the distribution of contributions to the covariance of EF pairs. Positive contributions indicate that the driver causes positive covariances between pairs of EFs, synergistically driving the two individual EFs. Negative contributions indicate that the driver causes negative covariances between pairs of EFs, driving the two individual EFs antagonistically. Results are derived by partitioning overall covariances into contributions of the different drivers; see the method section for the explanation. In this graph, only effects are shown, which on average, explain > 5% of covariance. In the Supporting information, the same graph with all variables (including FGR and the proportion of grasses, legumes and herbs) is shown.

importance of repeated measurements of ecosystem functions (EFs) over time to avoid spurious conclusions. Furthermore, they suggest that land management practices that promote biodiversity and reduce negative identity effects can enhance multifunctionality in grasslands. Our finding that the proportion of legumes and herbs contributed to identity effects driving tradeoffs suggests the potential presence of unfavourable plant species or combinations of plant species that could be managed to enhance multifunctionality in grasslands.

We found that even under the controlled conditions of our experiment, correlations among EFs were variable. High temporal variation of individual EFs had been documented before (Carpenter et al. 2009, Cardinale et al. 2012, Gaglio et al. 2020, Qiu et al. 2020, van der Plas et al. 2020). However, until now, inconsistent correlations between EFs or ESs have only been found when different studies were compared (Lee and Lautenbach 2016). Although all functions were measured with a consistent methodology at a single field site, tradeoffs were as variable as synergies (Supporting information), and relationships for many pairs of functions could range from synergy to tradeoff when correlations were calculated for different time points, which confirms the previous study of Lee and Lautenbach (2016), investigating ES relationships. Lee and Lautenbach (2016) found that the agreement on the type of relationship for a particular pair of ES, i.e. synergy, tradeoff, or no-effect relationship, decreased the more often the relationships were measured. Similarly, we found that the variability of EF relationships increased with the number of measurements (Supplementary information), which indicates that single measurements can be misleading when EF relationships are identified. Furthermore, we showed that not the identity of time points (years and seasons) but the identity of EF pairs were associated with a high variation in EF relationships (Supporting information). That means that it depends on the particular EF pair whether their correlation was highly variable because of differences between years or seasons. One explanation could be that ecosystem processes vary caused by a change or adjustment of biotic assemblages as a response to their environmental conditions (Turner and Chapin 2005), leading to changes in EF relationships or -multifunctionality with changing environmental conditions (Zirbel et al. 2019). In our study, the variation in relationships between EFs originated from the temporal variation in EF drivers (possible reasons could be inter-annual variation in rainfall, temperature or other cyclic patterns such as boom and bust cycles of herbivory), while in Lee and Lautenbach (2016), the variation in the relationships among ESs was introduced by different studies, and therefore additional site-dependent contexts.

Regarding the identified EF relationships (mean correlations among all the different EFs), we found both synergies and tradeoffs that can be explained by biological processes and therefore confirm other studies investigating the individual EFs (Jarrell and Beverly 1981, Allan et al. 2013). For example, EFs of the classes plant nutrients, and plant productivity often showed a tradeoff, indicating a dilution effect. i.e. when plant growth improved, plant nutrient concentrations decreased in the plant tissue (Jarrell and Beverly 1981). However, the carbon concentration of plants (PlantC) showed mainly synergies with EFs of the class plant productivity. One reason could be that high biomass reflects a high nutrient efficiency and thus comparatively low nutrient concentrations and correspondingly high C concentrations (Allan et al. 2013). Furthermore, the organic carbon in the soil (SoilDOC, SoilCorg) was positively related to plant productivity (Fig. 2). This is consistent with studies showing that high biomass production leads to an accumulation of dead plant material in the soil (Post and Kwon 2000) or root exudation of plants (Raich and Tufekciogul 2000). As the Jena Experiment was established on depleted arable soil, a higher carbon concentration in the soil occurred faster with higher biomass production, but in the end, the carbon concentration might be the same on all plots due to accelerated litter decomposition (Weisser et al. 2017). EFs of the class Invasion resistance showed synergies with EFs of the class plant productivity. This is consistent with former studies, showing that high biomass of the native species suppressed invasive species (Yannelli et al. 2020, Rojas-Botero et al. 2022), often due to complete use of available resources (Hector et al. 2001, Roscher et al. 2009). Summarising these examples, the relationships identified here for the classes of EFs are consistent with the underlying biological processes.

We found that synergies and tradeoffs have different drivers. Species richness is a known driver of many EFs (Gamfeldt and Roger 2017, Weisser et al. 2017, Craven et al. 2018) and ESs (van der Plas 2019). When two EFs improve with higher SR, a positive covariance is introduced, strengthening their relationship.

We showed that an increase in SR had a predominantly positive impact on the investigated EFs (Fig. 3). This suggests that the positive effects of SR on the covariance between EFs were widespread, with synergistic EFs being more strongly influenced by SR compared to EFs involved in tradeoffs. However, the relationship between two EFs weakens when SR has contrasting effects on the two EFs, as indicated by a few pairs of EFs for which we showed SR to cause negative covariance. In our study, plot identity represents all differences among plots. This encompasses differences arising from diverse plant communities, such as FGR, the proportion of different functional groups (Supporting information), and the presence of other groups of organisms (e.g. microbes, insects) associated with particular plant communities. While the experimental design ensures that stochastic variation is reduced by standardising abiotic conditions, including initial conditions and management practices, across all plots, abiotic factors cannot fully be excluded. Within the long-term experiment, species composition may have changed over time (Wagg et al. 2022). For instance, we observed significant variations among plots with identical compositions, such as the four plots of the 60-species mixture. In addition to changing plant communities, abiotic factors could change due to feedback at the plot level. Several long-term studies have shown that abiotic parameters such as pH and nitrogen mineralisation can change over time, even at the plot/ treatment level (Weisser et al. 2017, Cusser et al. 2021). As a result,

we interpret plot identity as the distinctions attributed to the different plant communities and changed abiotic conditions induced through feedback loops present in the study. These identity effects were mostly positive for EFs in synergies and mostly negative for EFs in tradeoffs. This can be explained by selection effects, which have been documented repeatedly by comparing the performance, such as biomass production, of plant communities (Marquard et al. 2009). A high performance of a plant community may be associated with a high abundance of certain species (Loreau and Hector 2001) and, therefore, with many simultaneously occurring EFs (many synergies). Our finding that the proportion of legumes can drive synergies is one indicator of the positive selection effect. Low performance is associated with a negative selection effect (Loreau and Hector 2001) and could be related to the occurrence of just a few EFs, as they are restricted by tradeoffs. Our finding that the proportion of herbs and legumes drive tradeoffs is an indicator of the negative selection effect. Outside an experimental setup, the equivalent of SR and plotID would be the different communities associated with landscape patches. As a consequence of different biotic communities and abiotic conditions, different levels of individual EFs would occur in these patches, and different correlations among EFs could be identified. While the effect of different aspects of SR and plotID (such as FGR, or proportions of functional groups) on individual EFs is frequently investigated, further research is needed to identify how different SR/ plotID impacts the relationships between EF.

Time (year, season and their interaction) explained some, albeit little, covariance among EFs and affected synergies and tradeoffs differently. Time can become a driver of EF relationships when environmental conditions vary over time, e.g. temperature and extreme events, affecting the biological activity of organisms. Tradeoffs were more affected by temporal effects than synergies. The season was among the main drivers of tradeoffs, reflecting the pronounced change in abiotic conditions among seasons in the temperate zone. As an example, soil microbial activity strongly responds to climatic conditions, affecting carbon and nutrient cycling (Frey et al. 2013). Further, competition strongly affects tradeoffs, which can be changed by altering the environmental conditions, such as the availability of water or light that fluctuate with time. Consequently, temporal variability in these drivers can induce variability in EF correlations underlying the importance of repeated measurements to identify the true relationships between EF.

When investigating the drivers of EF correlations, we found that some covariance among EFs could not be explained by any of the drivers tested in our study. We interpret this unexplained covariance as EF pairs affected by the ecological-constraints mechanism. Plants have access to a limited pool of resources they can invest in, e.g. in growth or defence against natural enemies, resulting in a growth-defence tradeoff (Karasov et al. 2017). Because providing unlimited resources within a local patch is impossible, ecological tradeoffs resulting from resource limitation are inevitable. Further, the simultaneous provision of EFs can be limited by competition. For example, in our study, improved plant productivity

was associated with higher invasion resistance (considered good), likely due to intensified competition for space and light between the resident plant community and potential invading plant individuals in our plots. Furthermore, the higher the root biomass was, the lower the soil nutrient concentrations, implying competition among plant species for available nutrients. Resource limitations and competition may limit biological activities, leading to tradeoffs between EFs. These tradeoffs can be weakened when competition in diverse communities is reduced by complementarity between species (Weisser et al. 2017). Understanding how ecological constraints affect relationships between EFs is an important topic for further investigation.

Our results have implications for land management aiming at promoting ES multifunctionality. Relationships among EFs affect EF multifunctionality since they can either promote (synergies) or limit (tradeoffs) EF multifunctionality. Analysing the drivers of relationships between EFs, we showed that species richness can promote synergies among EFs, resulting in increased EF multifunctionality. Consequently, promoting diversity is a means to foster EF multifunctionality, confirming previous empirical biodiversity EF multifunctionality relationships (Isbell et al. 2011, Lefcheck et al. 2015, Meyer et al. 2018). Further, we showed that plot identity effects, including functional group richness and the proportion of individual functional groups, were important drivers for tradeoffs between EFs. While we tested for identity effects of plots, there are likely individual plant species that cause these tradeoffs by maximising some EFs at the expense of other EFs. When future research can identify such plant species with strong effects on tradeoffs, land management can target low densities of these disadvantageous species to reduce tradeoffs between EF and promote EF multifunctionality. Nevertheless, competition for resources and the resulting ecological tradeoffs between EFs are challenging to resolve.

Alternative viewpoints

There are two topics authors disagree on in the interpretation: first, there is an ongoing debate regarding the interpretation of plotID in our study. Some argue that in our controlled experiment with similar starting conditions for each plot, plotID primarily reflects plant community identities, which can influence abiotic conditions. Consequently, this interpretation indicates that any abiotic differences among plots were caused by the plant communities and their changes over time. Others contend that it encompasses both composition effects and unexplained environmental variations. However, it is important to acknowledge that certain factors, such as environmental variations in the field site, are beyond our control, even with the inclusion of experimental blocks.

Second, it should be noted that there is an uneven representation of different EFs within each EF class. This disparity in representation has the potential to influence the identification of important drivers for EF relationships. Despite conducting a sensitivity analysis on the drivers, it is possible that with a different

set of EFs, certain drivers may assume greater influence on EF relationships or lose their effects. However, it is important to acknowledge that our study is limited by the available data and the resources to measure EFs simultaneously in the same year.

Conclusions

Our study showed that even under the controlled conditions of a single experimental field site, correlations among EFs were variable over time. Consequently, repeated measurements of EF are needed to avoid spurious and non-generalisable conclusions about relationships among EFs.

Moreover, our results show the potential for land management to promote EF multifunctionality during establishment and management by incorporating two principles. First, maintaining or increasing the biodiversity of grasslands to increase synergies among EFs; second, reducing the proportion of disadvantageous species or functional groups and assembling communities to reduce tradeoffs, to promote EF multifunctionality.

Future studies should continue to investigate the drivers of EF relationships to identify common drivers causing tradeoffs and separate common drivers from potential ecological constraints. Importantly, these studies should also address how environmental conditions can change these relationships and identify influential species enabling recommendations on how to adapt management for maximising EF multifunctionality.

Author contributions

Laura Argens: Conceptualization (equal); Formal analysis (lead); Methodology (lead); Project administration (lead); Visualization (lead); Writing – original draft (lead). **Wolfgang W. Weisser**: Conceptualization (supporting); Funding acquisition (equal); Investigation (equal); Supervision (supporting); Writing – review and editing (equal). **Anne Ebeling**: Investigation (equal); Writing – review and editing (equal). **Nico Eisenhauer**: Investigation (equal);, Writing – review and editing (equal). **Markus Lange**: Investigation (equal);

Writing – review and editing (equal). **Yvonne Oelmann**: Investigation (equal); Writing – review and editing (equal). **Christiane Roscher**: Investigation (equal);, Writing – review and editing (equal). **Holger Schielzeth**: Formal analysis (supporting); Investigation (equal); Methodology (equal); Writing – review and editing (equal). **Bernhard Schmid**: Formal analysis (supporting); Methodology (equal);, Writing – review and editing (equal). **Wolfgang Wilcke**: Investigation (equal);, Writing – review and editing (equal). **Sebastian T. Meyer**: Conceptualization (equal);, Formal analysis (supporting); Funding acquisition (equal);, Investigation (equal);, Methodology (equal);, Supervision (lead); Visualization (supporting); Writing – original draft (supporting); Writing – review and editing (lead).

Data availability statement

This work is based on data elaborated by multiple projects of the Jena Experiment, which is funded by the Deutsche Forschungsgemeinschaft (For 5000). The datasets are publicly available in the Jena Experiment database (https://jexis.idiv.de/), or the data publisher PANGEAE (https://www.pangaea.de/). In the Supporting information, the IDs ans DOIs of the datasets are listed. However, to give data owners and collectors time to perform their analysis the Jena Experiment' data and publication policy includes by default an embargo period of three years from the end of data collection/data assembly which applies to some datasets. These datasets will be made publicly available via the same data repository.

Supporting information

The Supporting information associated with this article is available with the online version.

References

Adhikari, K. and Hartemink, A. E. 2016. Linking soils to ecosystem services – a global review. – Geoderma 262: 101–111.

Allan, E. et al. 2013. A comparison of the strength of biodiversity effects across multiple functions. – Oecologia 173: 223–237.

Balvanera, P., Pfisterer, A. B., Buchmann, N., He, J. S., Nakashizuka, T., Raffaelli, D. and Schmid, B. 2006. Quantifying the evidence for biodiversity effects on ecosystem functioning and services. – Ecol. Lett. 9: 1146–1156.

Bell, G. 1989. A comparative method. – Am. Nat. 133: 553–571.

Bennett, E. M., Peterson, G. D. and Gordon, L. J. 2009. Understanding relationships among multiple ecosystem services. – Ecol. Lett. 12: 1394–1404.

Bradford, M. A., Wood, S. A., Bardgett, R. D., Black, H. I. J., Bonkowski, M., Eggers, T., Grayston, S. J., Kandeler, E., Manning, P., Setälä, H. and Jones, T. H. 2014. Discontinuity in the responses of ecosystem processes and multifunctionality to altered soil community composition. – Proc. Natl Acad. Sci. USA 111: 14478.

Cardinale, B. J., Srivastava, D. S., Duffy, J. E., Wright, J. P., Downing, A. L., Sankaran, M. and Jouseau, C. 2006. Effects of bio-

diversity on the functioning of trophic groups and ecosystems. – Nature 443: 989–992.

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., Mace, G. M., Tilman, D., Wardle, D. A., Kinzig, A. P., Daily, G. C., Loreau, M., Grace, J. B., Larigauderie, A., Srivastava, D. S. and Naeem, S. 2012. Biodiversity loss and its impact on humanity. – Nature 486: 59–67.

Carpenter, S. R., Mooney, H. A., Agard, J., Capistrano, D., DeFries, R. S., Díaz, S., Dietz, T., Duraiappah, A. K., Oteng-Yeboah, A. and Pereira, H. M. 2009. Science for managing ecosystem services: beyond the millennium ecosystem assessment. – Proc. Natl Acad. Sci. USA 106: 1305–1312.

Cohen, J. 2013. Statistical power analysis for the behavioral sciences. – Routledge.

Costanza, R., de Groot, R., Braat, L., Kubiszewski, I., Fioramonti, L., Sutton, P., Farber, S. and Grasso, M. 2017. Twenty years of ecosystem services: how far have we come and how far do we still need to go? – Ecosyst. Serv. 28: 1–16.

Craven, D. et al. 2018. Multiple facets of biodiversity drive the diversity–stability relationship. – Nat. Ecol. Evol. 2: 1579–1587.

Crouzat, E., Mouchet, M., Turkelboom, F., Byczek, C., Meersmans, J., Berger, F., Verkerk, P. J. and Lavorel, S. 2015. Assessing bundles of ecosystem services from regional to landscape scale: insights from the French Alps. – J. Appl. Ecol. 52: 1145–1155.

Cusser, S., Helms, J. IV, Bahlai, C. A. and Haddad, N. M. 2021. How long do population level field experiments need to be? Utilising data from the 40-year-old LTER network. – Ecol. Lett. 24: 1103–1111.

Dade, M. C., Mitchell, M. G. E., McAlpine, C. A. and Rhodes, J. R. 2019. Assessing ecosystem service tradeoffs and synergies: the need for a more mechanistic approach. – Ambio 48: 1116–1128.

Dooley, Á., Isbell, F., Kirwan, L., Connolly, J., Finn, J. A., Brophy, C. and Buckley, Y. 2015. Testing the effects of diversity on ecosystem multifunctionality using a multivariate model. – Ecol. Lett. 18: 1242–1251.

Engel, V., Jobbágy, E. G., Stieglitz, M., Williams, M. and Jackson, R. B. 2005. Hydrological consequences of *Eucalyptus* afforestation in the Argentine Pampas. – Water Resour. Res. 41: W10409.

Finney, D. M. and Kaye, J. P. 2017. Functional diversity in cover crop polycultures increases multifunctionality of an agricultural system. – J. Appl. Ecol. 54: 509–517.

Foley, J. A. et al. 2011. Solutions for a cultivated planet. – Nature 478: 337–342.

Frey, S. D., Lee, J., Melillo, J. M. and Six, J. 2013. The temperature response of soil microbial efficiency and its feedback to climate. – Nat. Clim. Change 3: 395–398.

Gaglio, M., Aschonitis, V., Castaldelli, G. and Fano, E. A. 2020. Land use intensification rather than land cover change affects regulating services in the mountainous Adige river basin (Italy). – Ecosyst. Serv. 45: 101158.

Gamfeldt, L. and Roger, F. 2017. Revisiting the biodiversity–ecosystem multifunctionality relationship. – Nat. Ecol. Evol. 1: 168.

Gamfeldt, L., Hillebrand, H. and Jonsson, P. R. 2008. Multiple functions increase the importance of biodiversity for overall ecosystem functioning. – Ecology 89: 1223–1231.

Gordon, L. J., Peterson, G. D. and Bennett, E. M. 2008. Agricultural modifications of hydrological flows create ecological surprises. – Trends Ecol. Evol. 23: 211–219.

Hanisch, M., Schweiger, O., Cord, A. F., Volk, M. and Knapp, S. 2020. Plant functional traits shape multiple ecosystem services, their tradeoffs and synergies in grasslands. – J. Appl. Ecol. 57: 1535–1550.

Harrell, F. E. Jr, and Dupont, C. 2020. Hmisc: harrell miscellaneous. – R package ver. 4.4-2., https://hbiostat.org/R/Hmisc/.

Hautier, Y. et al. 2018. Local loss and spatial homogenization of plant diversity reduce ecosystem multifunctionality. – Nat. Ecol. Evol. 2: 50–56.

He, J. S., Wang, X., Flynn, D. F., Wang, L., Schmid, B. and Fang, J. 2009. Taxonomic, phylogenetic, and environmental tradeoffs between leaf productivity and persistence. – Ecology 90: 2779–2791.

Hector, A., Dobson, K., Minns, A., Bazeley-White, E. and Hartley Lawton, J. H. 2001. Community diversity and invasion resistance: an experimental test in a grassland ecosystem and a review of comparable studies. – Ecol. Res. 16: 819–831.

Isbell, F., Calcagno, V., Hector, A., Connolly, J., Harpole, W. S., Reich, P. B., Scherer-Lorenzen, M., Schmid, B., Tilman, D., van Ruijven, J., Weigelt, A., Wilsey, B. J., Zavaleta, E. S. and Loreau, M. 2011. High plant diversity is needed to maintain ecosystem services. – Nature 477: 199–202.

Jarrell, W. M. and Beverly, R. B. 1981. The dilution effect in plant nutrition studies. – Adv. Agron. 34: 197–224.

Jochum, M. et al. 2020. The results of biodiversity–ecosystem functioning experiments are realistic. – Nat. Ecol. Evol. 4: 1485–1494.

Karasov, T. L., Chae, E., Herman, J. J. and Bergelson, J. 2017. Mechanisms to mitigate the tradeoff between growth and defense. – Plant Cell 29: 666–680.

Kempthorne, O. 1957. An introduction to genetic statistics. – Wiley.

Kleinman, P., Sharpley, A., Buda, A., McDowell, R. and Allen, A. 2011. Soil controls of phosphorus in runoff: management barriers and opportunities. – Can. J. Soil Sci. 91: 329–338.

Lee, H. and Lautenbach, S. 2016. A quantitative review of relationships between ecosystem services. – Ecol. Indic. 66: 340–351.

Lefcheck, J. S., Byrnes, J. E. K., Isbell, F., Gamfeldt, L., Griffin, J. N., Eisenhauer, N., Hensel, M. J. S., Hector, A., Cardinale, B. J. and Duffy, J. E. 2015. Biodiversity enhances ecosystem multifunctionality across trophic levels and habitats. – Nat. Commun. 6: 6936.

Li, B., Chen, N., Wang, Y. and Wang, W. 2018. Spatio-temporal quantification of the tradeoffs and synergies among ecosystem services based on grid-cells: a case study of Guanzhong Basin, NW China. – Ecol. Indic. 94: 246–253.

Loreau, M. and Hector, A. 2001. Partitioning selection and complementarity in biodiversity experiments. – Nature 412: 72–76.

Maestre, F. T. et al. 2012. Plant species richness and ecosystem multifunctionality in global drylands. – Science 335: 214–218.

Mangiafico, S. 2015. An R companion for the handbook of biological statistics. – www.rcompanion.org/rcompanion/.

Manning, P., van der Plas, F., Soliveres, S., Allan, E., Maestre, F. T., Mace, G., Whittingham, M. J. and Fischer, M. 2018. Redefining ecosystem multifunctionality. – Nat. Ecol. Evol. 2: 427–436.

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

Martin, G., Durand, J.-L., Duru, M., Gastal, F., Julier, B., Litrico, I., Louarn, G., Médiène, S., Moreau, D., Valentin-Morison, M., Novak, S., Parnaudeau, V., Paschalidou, F., Vertès, F., Voisin, A.-S., Cellier, P. and Jeuffroy, M.-H. 2020. Role of ley pastures in tomorrow's cropping systems. A review. – Agron. Sustain. Dev. 40: 17.

Meyer, S. T. et al. 2018. Biodiversity-multifunctionality relationships depend on identity and number of measured functions. – Nat. Ecol. Evol. 2: 44–49.

Millennium Ecosystem Assessment 2005. Ecosystems and human well-being. – Island Press.

Mokany, K., Raison, R. J. and Prokushkin, A. S. 2006. Critical analysis of root: shoot ratios in terrestrial biomes. – Global Change Biol. 12: 84–96.

Mondelaers, K., Aertsens, J. and Van Huylenbroeck, G. 2009. A meta-analysis of the differences in environmental impacts between organic and conventional farming. – Brit. Food J. 111: 1098–1119.

Naeem, S. 1998. Species redundancy and ecosystem reliability. – Conserv. Biol. 12: 39–45.

Pasari, J. R., Levi, T., Zavaleta, E. S. and Tilman, D. 2013. Several scales of biodiversity affect ecosystem multifunctionality. – Proc. Natl Acad. Sci. USA 110: 10219–10222.

Pereira, P., Bogunovic, I., Inacio, M., Zhao, W. and Barcelo, D. 2023. Agriculture intensification impacts on soil and water ecosystem services, Copernicus Meetings, EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-1423.

Pérez-Ramos, I. M., Roumet, C., Cruz, P., Blanchard, A., Autran, P. and Garnier, E. 2012. Evidence for a 'plant community economics spectrum' driven by nutrient and water limitations in a Mediterranean rangeland of southern France. – J. Ecol. 100: 1315–1327.

Post, W. M. and Kwon, K. C. 2000. Soil carbon sequestration and land-use change: processes and potential. – Global Change Biol. 6: 317–327.

Qiu, J., Carpenter, S. R., Booth, E. G., Motew, M. and Kucharik, C. J. 2020. Spatial and temporal variability of future ecosystem services in an agricultural landscape. – Landscape Ecol. 35: 2569–2586.

Raich, J. W. and Tufekciogul, A. 2000. Vegetation and soil respiration: correlations and controls. – Biogeochemistry 48: 71–90.

Reiss, J., Bridle, J. R., Montoya, J. M. and Woodward, G. 2009. Emerging horizons in biodiversity and ecosystem functioning research. – Trends Ecol. Evol. 24: 505–514.

Ribeiro, P. J. D. Jr., Peter, J., Schlather, M., Bivand, R. and Ripley, B. 2020. geoR: analysis of geostatistical data. – R package., http://www.leg.ufpr.br/geoR/.

Rodríguez, J. P., Beard, T. D. Jr, Bennett, E. M., Cumming, G. S., Cork, S. J., Agard, J., Dobson, A. P. and Peterson, G. D. 2006. Tradeoffs across space, time and ecosystem services. – Ecol. Soc. 11.

Rojas-Botero, S., Kollmann, J. and Teixeira, L. H. 2022. Competitive trait hierarchies of native communities and invasive propagule pressure consistently predict invasion success during grassland establishment. – Biol. Invas. 24: 107–122.

Roscher, C., Schumacher, J., Baade, J., Wilcke, W., Gleixner, G., Weisser, W. W., Schmid, B. and Schulze, E.-D. 2004. The role of biodiversity for element cycling and trophic interactions: an experimental approach in a grassland community. – Basic Appl. Ecol. 5: 107–121.

Roscher, C., Beßler, H., Oelmann, Y., Engels, C., Wilcke, W. and Schulze, E.-D. 2009. Resources, recruitment limitation and invader species identity determine pattern of spontaneous invasion in experimental grasslands. – J. Ecol. 97: 32–47.

RStudio Team. 2016. RStudio: integrated development environment for R. – RStudio, Inc.

Scherber, C. et al. 2010. Bottom-up effects of plant diversity on multitrophic interactions in a biodiversity experiment. – Nature 468: 553–556.

Schmid, B., Baruffol, M., Wang, Z. and Niklaus, P. A. 2017. A guide to analyzing biodiversity experiments. – J. Plant Ecol. 10: 91–110.

Shen, J., Li, S., Liang, Z., Liu, L., Li, D. and Wu, S. 2020. Exploring the heterogeneity and nonlinearity of trade-offs and synergies among ecosystem services bundles in the Beijing-Tianjin-Hebei urban agglomeration. – Ecosyst. Serv. 43: 101103.

Spake, R., Lasseur, R., Crouzat, E., Bullock, J. M., Lavorel, S., Parks, K. E., Schaafsma, M., Bennett, E. M., Maes, J., Mulligan, M., Mouchet, M., Peterson, G. D., Schulp, C. J. E., Thuiller, W., Turner, M. G., Verburg, P. H. and Eigenbrod, F. 2017. Unpacking ecosystem service bundles: towards predictive mapping of synergies and tradeoffs between ecosystem services. – Global Environ. Change 47: 37–50.

Tilman, D. 2001. Functional diversity. – Encyclopedia Biodiver. 3: 109–120.

Torralba, M., Fagerholm, N., Hartel, T., Moreno, G. and Plieninger, T. 2018. A social-ecological analysis of ecosystem services supply and tradeoffs in European wood-pastures. – Sci. Adv. 4: eaar2176.

Turner, M. G. and Chapin, F. S. 2005. Causes and consequences of spatial heterogeneity in ecosystem function. – In: Lovett, G. M., Turner, M. G., Jones, C. G. and Weathers, K. C. (eds), Ecosystem function in heterogeneous landscapes. Springer, pp. 9–30.

van der Plas, F. 2019. Biodiversity and ecosystem functioning in naturally assembled communities. – Biol. Rev. Camb. Phil. Soc. 94: 1220–1245.

van der Plas, F. et al. 2020. Plant traits alone are poor predictors of ecosystem properties and long-term ecosystem functioning. – Nat. Ecol. Evol 4: 1602–1611.

Wagg, C., Roscher, C., Weigelt, A., Vogel, A., Ebeling, A., de Luca, E., Roeder, A., Kleinspehn, C., Temperton, V. M., Meyer, S. T., Scherer-Lorenzen, M., Buchmann, N., Fischer, M., Weisser, W. W., Eisenhauer, N. and Schmid, B. 2022. Biodiversity–stability relationships strengthen over time in a long-term grassland experiment. – Nat. Commun. 13: 7752.

Wang, S., Liu, Z., Chen, Y. and Fang, C. 2021. Factors influencing ecosystem services in the Pearl River Delta, China: spatiotemporal differentiation and varying importance. – Resour. Conserv. Recy. 168: 105477.

Watson, J. E. M. and Venter, O. 2017. Ecology: a global plan for nature conservation. – Nature 550: 48–49.

Weisser, W. W. et al. 2017. Biodiversity effects on ecosystem functioning in a 15-year grassland experiment: patterns, mechanisms, and open questions. – Basic Appl. Ecol. 23: 1–73.

Willemen, L. 2020. It's about time: advancing spatial analyses of ecosystem services and their application. – Ecosyst. Serv. 44: 101125.

Yannelli, F. A., MacLaren, C. and Kollmann, J. 2020. Moving away from limiting similarity during restoration: timing of arrival and native biomass are better proxies of invasion suppression in grassland communities. – Front. Ecol. Evol. 8: 238.

Zheng, H., Wang, L. and Wu, T. 2019. Coordinating ecosystem service tradeoffs to achieve win-win outcomes: a review of the approaches. – J. Environ. Sci. (China) 82: 103–112.

Zirbel, C. R., Grman, E., Bassett, T. and Brudvig, L. A. 2019. Landscape context explains ecosystem multifunctionality in restored grasslands better than plant diversity. – Ecology 100: e02634.