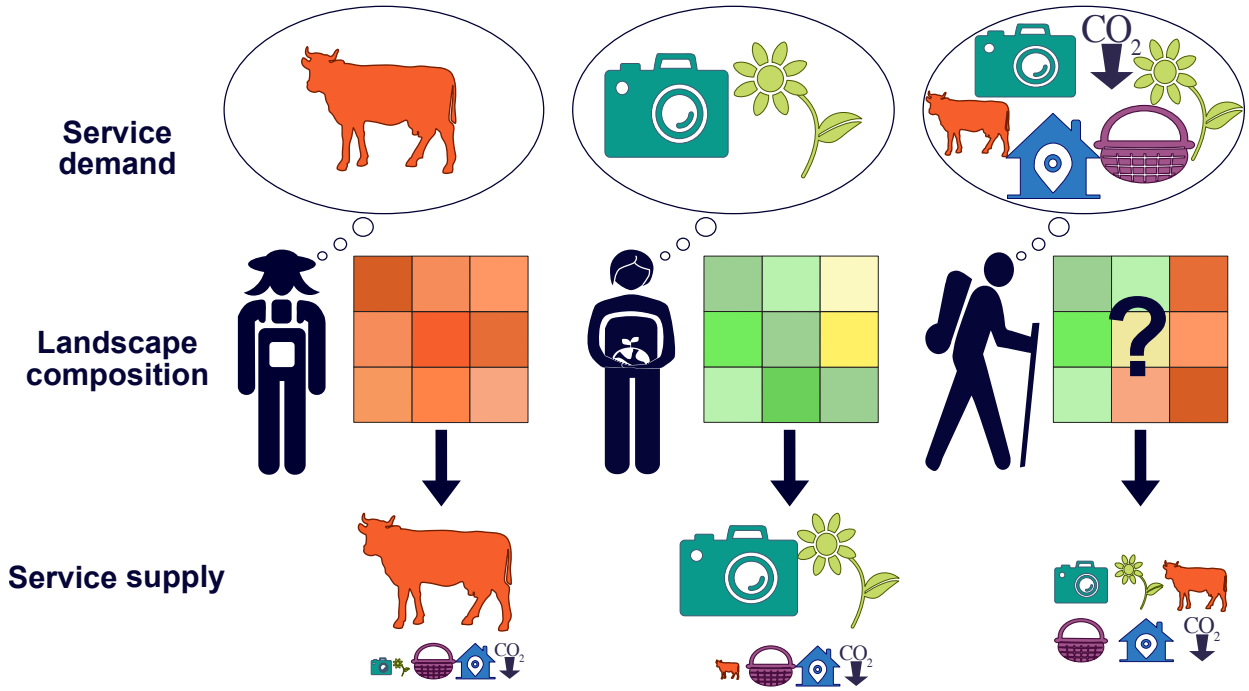
Landscape management for grassland multifunctionality

Neyret M., Fischer M., Allan E., Hölzel N., Klaus V. H., Kleinebecker T., Krauss J., Le Provost G., Peter. S, Schenk N., Simons N.K., van der Plas F., Binkenstein J., Börshig C., Jung K., Prati D., Schäfer D., Schäfer M., Schöning I., Schrumpf M., Tschapka M., Westphal C. & Manning P.

Highlights

- An online tool identifies optimal landscape compositions for desired ecosystem services
- When the desired services are synergic, the optimum is their common best landscape composition
- When the desired services trade-off, a mix of grassland intensity is most multifunctional
- Such tools could support decision-making processes and aid conflict resolution

Graphical abstract



Abstract

Land-use intensification has contrasting effects on different ecosystem services, often leading to land-use conflicts. While multiple studies have demonstrated how landscape-scale strategies can minimise the trade-off between agricultural production and biodiversity conservation, little is known about which land use strategies maximise the landscape level supply of multiple ecosystem services (landscape multifunctionality), a common goal of stakeholder communities.

We combine comprehensive data collected from 150 German grassland sites with a simulation approach to identify landscape compositions, with differing proportions of low-, medium-, and high-intensity grasslands, that minimise trade-offs between the six main grassland ecosystem services prioritised by local stakeholders: biodiversity conservation, aesthetic value, productivity, carbon storage, foraging, and regional identity. Results are made accessible through an online tool that provides information on which compositions best meet any combination of user-defined priorities (https://neyret.shinyapps.io/landscape_composition_for_multifunctionality/).

Results show that an optimal landscape composition can be identified for any pattern of ecosystem service priorities. However, multifunctionality was similar and low for all landscape compositions in cases where there are strong trade-offs between services (e.g. aesthetic value and fodder production), where many services were prioritised, and where drivers other than land use played an important role. We also found that if moderate service levels are deemed acceptable, then strategies in which both high and low intensity grasslands are present can deliver landscape multifunctionality. The tool presented can aid informed decision-making by predicting the impact of future changes in landscape composition, and by allowing for the relative roles of stakeholder priorities and biophysical trade-offs to be understood by scientists and practitioners alike.

Keywords

Land management, ecosystem services, multifunctionality, grasslands, land use, simulation modelling

1. Introduction

Habitat conversion and land-use intensification are driving biodiversity loss and changes to ecosystem service supply across the world (IPBES 2019). While high land-use intensification promotes a small number of ecosystem services related to food production, it is often detrimental to biodiversity conservation (Bennett et al., 2009; Lavorel et al., 2011; Raudsepp-Hearne et al.,

2010) and other regulating or cultural ecosystem services that depend on biodiversity for their delivery (Allan et al., 2015; Clec'h et al., 2019). Such contrasting responses of different ecosystem services to ecosystem drivers often make it impossible to achieve high levels of all desired services at a local scale (van der Plas et al., 2019). This has led to land-use conflicts, which are becoming increasingly common across the globe (Goldstein et al., 2012).

To date, much of the work on minimising trade-offs between ecosystem services within landscapes has compared a 'land sparing' strategy, in which semi-natural high-biodiversity areas and intensive farmland are spatially segregated, and a 'land sharing' strategy in which biodiversity conservation and commodity production are co-delivered in a landscape of intermediate intensity, and where these different land uses form a mosaic (Green, 2005). Within this field, most studies have found that land sparing is the best way to achieve high levels of both biodiversity conservation and commodity production (Phalan et al., 2011; Simons & Weisser, 2017). However, multiple studies have also stressed the limitations of the land sharing versus land sparing concept. The framework focuses on just two extreme strategies, and on only two services - commodity production and biodiversity conservation (Bennett, 2017; Fischer et al., 2014), while in reality, most landscapes are expected to provide multiple services, even within a single ecosystem type. This is the case for semi-natural grasslands (*sensu* Bullock et al. 2011), which supply a wide range of highly prioritised ecosystem services including water provision, climate regulation (carbon storage) and recreation services, in addition to food production and biodiversity conservation (Bengtsson et al., 2019). Accounting for these additional ecosystem services could significantly affect which land-use strategy is considered optimal, meaning the best strategy for achieving high levels of multiple services within grassland landscapes remains unknown.

One way of measuring how land-use affects the overall supply of multiple services is the ecosystem service multifunctionality approach (Manning et al., 2018). Ecosystem service multifunctionality is defined as the simultaneous supply of multiple prioritised ecosystem services, relative to their human demand (Linders et al. 2021). It builds upon the metrics used in biodiversity-ecosystem functioning research (Allan et al., 2015; Barnes et al., 2017; Byrnes et al 2014) by combining ecological and biophysical data describing the supply of multiple ecosystem services with social data that quantifies the relative priority given by stakeholder groups to each service. The approach also advances on existing methods such as the identification of ecosystem service bundles (Frei et al., 2018; Raymond et al., 2009) by measuring the overall supply of ecosystem services relative to their demand. The resulting multifunctionality metrics can therefore be seen as summarising the overall benefit provided by a system to stakeholders.

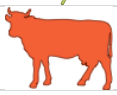
Here, we combine the multifunctionality approach with data simulation methods to identify the optimal landscape composition for multiple ecosystem services. This approach involves varying the proportion of land under different intensities in data simulations and measuring the outcome on ecosystem service multifunctionality. We also investigate how the relative priority of services to land users affects the optimal strategy. The analysis was achieved by using ecosystem service data collected at 150 grassland sites that vary in their intensity, found in the three regions of the large-scale and long-term Biodiversity Exploratories project, in Germany. This was utilised in simulations in which artificial 'landscapes' of varying composition, in terms of land use intensity, were assembled from site-level data (Fig. 1). We base our metrics of multifunctionality on six services which are directly linked to final benefits (*sensu* the cascade model, Mace et al., 2012; Fisher & Turner, 2008): fodder production, biodiversity conservation, climate change mitigation, aesthetic value, foraging opportunities and regional identity, covering all the services provided by grasslands that were demanded by the main stakeholder groups, as identified in a social survey (Supplementary Fig. A1). We hypothesised that heterogenous landscapes composed of both high- and low-intensity sites would have the highest multifunctionality, when considering fodder production and biodiversity conservation (van der Plas et al., 2019).

2. Material and methods

2.1 Study design

We used data from 150 grassland plots (hereafter sites) studied within the large-scale and long-term Biodiversity Exploratories project in Germany (https://www.biodiversity-exploratories.de/). The sites were located in three regions: the UNESCO Biosphere Area Schwäbische Alb (South-West region), in and around the National Park Hainich (Central region; both are hilly regions with calcareous bedrock), and the UNESCO Biosphere Reserve Schorfheide-Chorin (North of Germany: flat, with a mixture of sandy and organic soils, see Fischer et al. (2010) for details). The regions were selected to be broadly representative of the three main landscape types of Southern, Central and Northern Germany in terms of environmental conditions and management regimes. Sites measured 50 x 50m and were selected to be representative of the whole field they were in. Field size ranged from 3 to 148 ha. The 50 sites of each region spanned the full range of land-use intensity within the region, while minimising variation in potentially confounding environmental factors.

Table 1 Estimation of ecosystem services from site-scale ecosystem service indicators.

Ecosystem service		Site-scale ecosystem service indicator	Landscape-scale* ecosystem service indicator
	Biodiversity conservation	Number of plant species (alpha diversity)	Number of plant species (gamma diversity)
	Fodder production	Estimated biomass production (as per Simons & Weisser 2017) x plant nitrogen concentration x 6.25 (Lee, 2018)	Sum of protein production of all sites in the landscape, i.e. material that can be transformed into livestock production (Lee 2018)
	Aesthetic value	Butterfly abundance Flower cover Bird richness	Average butterfly abundance in the landscape Average flower cover in the landscape Number of bird families (gamma diversity) in the landscape (Hedblom *et al.*, 2014)
	Climate change mitigation (carbon storage)	C stock at 0-10cm depth	Sum of soil C stocks in the landscape
	Regional identity	Cover of plants / richness of birds associated with German identity, historical habitats (*Juniperus* grasslands)	Sum of charismatic plants cover, gamma diversity of charismatic birds, number of grasslands with *Juniperus*
	Foraging	Cover of edible plants	Sum of edible plants cover

* All landscape-scale indicators were weighted equally within each service

It has been shown that the shape of the response of ecosystem services to their drivers can affect which landscape strategies are identified as optimal (e.g. Phalan et al. 2011), and in our

study, the response of each service to land use intensity differs strongly across regions, even after correcting for the effects of environmental covariates. For instance, biodiversity in the North region is inherently and uniformly low, hence an almost 'flat' biodiversity response to LUI is observed, while diversity declines strongly between low and high intensity in the other two regions (see Results, Fig. 2a). As a result, the optimal strategy is likely to be region-specific and identifying optimal strategies based on 'averaged' responses could give dangerously misleading results. Therefore, we chose to run all analyses at the regional level.

2.2 Land-use intensity

Data on site management was collected annually since 2007 from site owners using a questionnaire. We quantified grazing intensity as the number of livestock units × the number of days of grazing (cattle younger than 1 year corresponded to 0.3 livestock units (LU), cattle 2 years to 0.6 LU, cattle older than 2 years to 1 LU, sheep and goats younger than 1 year to 0.05 LU, sheep and goats older than 1 year to 0.1 LU, horses younger than 3 years to 0.7 LU, and horses older than 3 years to 1.1 LU; Fischer et al. 2010). Fertilisation intensity was defined as the amount of nitrogen addition, excluding dung inputs from grazing animals (kg N $ha^{-1}y^{-1}$), and mowing frequency is the annual number of mowing events. For each site these three land-use intensity (LUI) components were standardised, square-root transformed, summed, and then averaged between 2007 and 2012 to obtain an overall LUI value (Blüthgen et al., 2012). We then classified all sites as low-, medium- or high-intensity based on whether their LUI index (Fig. 1 step 1a) belonged to the lower, middle or top third (0-33%, 33-66%, 66-100% quantiles) of all LUI indices within the considered region (which resulted in a classification equivalent to classifying based on all regions altogether, Fisher test: $p < 10^{-10}$ for all regions). Confidence intervals for grazing, mowing and fertilisation intensities for each LUI class in the three regions are presented in Table 2. The intensity gradient was mostly driven by fertilisation and cutting frequency in the South-West and Central regions, and by grazing intensity and fertilisation in the North (Fig. 2b), and all three regions span a similar range of LUI values (Table 2).

Table 2 Description of the variations of land-use intensity components. Confidence intervals for fertilisation, mowing and grazing intensities in each region, for each land-use intensity (LUI) class. 95% confidence intervals were calculated based on management of individual sites on the period 2007-2012.

	LUI class	South-West	Central	North
LUI index	Low	1 (0.9-1.2)	1 (0.9-1.1)	1.2 (1.1-1.3)
	Medium	1.7 (1.6-1.7)	1.7 (1.6-1.8)	1.5 (1.4-1.5)
	High	2.2 (2-2.4)	2.2 (2.1-2.4)	2.3 (2.1-2.5)
Grazing intensity (Livestock units. days.ha^{-1})	Low	82.2 (49.2-115.3)	86.5 (64.4-108.6)	103.4 (52.2-154.6)
	Medium	97.6 (34.5-160.7)	102.5 (48.3-156.6)	239.5 (140.9-338.1)
	High	156.7 (24.6-288.8)	160.7 (39.6-281.8)	215.9 (80.4-351.4)
Mowing intensity (Cut.yr^{-1})	Low	0.5 (0.1-0.8)	0.3 (0-0.6)	0.8 (0.6-1.1)
	Medium	1.4 (0.9-1.8)	1.2 (0.9-1.5)	0.8 (0.4-1.2)
	High	1.8 (1.3-2.3)	1.6 (1.2-2)	1.2 (0.8-1.7)
Fertilisation (kg.N.ha^{-1})	Low	1.1 (-0.6-2.8)	1.4 (-1.6-4.4)	0.4 (-0.4-1.2)
	Medium	38 (23.4-52.7)	34.9 (18-51.8)	0.6 (-0.7-2)
	High	95 (67.4-122.6)	91.1 (65.4-116.9)	42.8 (25.8-59.7)

2.3 Ecosystem services priority

Three expert-workshops were conducted in autumn 2018 in the three Exploratories, with representatives of numerous pre-selected interest groups. These led to the identification of a full set of 14 interest groups and a list of landscape-level ecosystem services that are demanded regionally. We then restricted the list to services with direct links to final benefits, thus excluding regulating services (e.g. pollination) which underpin the supply of other services (e.g. food production) but do not directly benefit humans. We also excluded water-based services and the production of energy from technology which we weren't able to cover and both had relatively low weighting. The final list consisted of 12 ecosystem services (Fig. A1). We then conducted a survey across all stakeholder groups in 2019, in which 321 respondents were requested to distribute a maximum of 20 points across all services to quantify their personal priorities. At the beginning of the survey, written consent was requested for the collection and processing of the

anonymous personal data; all participants to the survey and workshops could withdraw at any time.

Next, the survey results were subsetted to services provided by grasslands (e.g. removing timber and food crop production), meaning six services were retained: biodiversity conservation, livestock production, aesthetic value, carbon storage, regional identity, and foraging. We then selected, as examples, four stakeholder groups who assigned contrasting priorities to these different services: locals, nature conservation associations, farmers and the tourism sector (126 respondents in total). Priority scores of for each service were then normalised by the total number of points attributed to grassland services by each respondent. The priority scores for each group did not vary significantly across regions so we used overall scores.

2.4 Ecosystem services supply

We estimated ecosystem service supply from several indicators (Table 1), measured at each of the 150 sites. Biodiversity conservation was based on total plant species richness as plant alpha-diversity at the site level (a good proxy for diversity across multiple trophic levels at these sites, Manning et al., 2015), and gamma-diversity at the landscape level. Fodder production was calculated as total fodder protein production, a common agronomical indicator used in livestock production models, and which better represents livestock production potential than biomass, a large part of which can be indigestible (e.g. Waghorn and Clark, 2004; Herrero, 2015; Lee, 2018). We calculated this indicator from direct measures of grassland aboveground biomass production and shoot protein content. Climate change mitigation was quantified as soil organic carbon stocks in the top 10 cm, as deeper stocks are unlikely to be affected strongly by management actions. For the aesthetic value measure, we integrated direct measures of flower cover, the number of bird families observed and the abundance of butterflies. Regional identity included the species richness (resp. cover) of birds (resp. plants) associated with German identity (see Table B1 and B2), and the foraging service was measured based on the cover of edible plants (see Table B3). Details on the measurements of these different indicators and their aggregation from site- to landscape- level can be found in the supplementary methods (appendix B).

Before estimating landscape-level services, we imputed missing values for individual indicators using predictive mean matching on the dataset comprising all services (98 out of 1500 values, R mice package v3.13.0; van Buuren & Groothuis-Oudshoorn, 2011). The missing values were mostly for flower cover, and some for butterfly abundance, and were equally distributed among regions and land-use intensities. We also corrected site-level indicator values for environmental

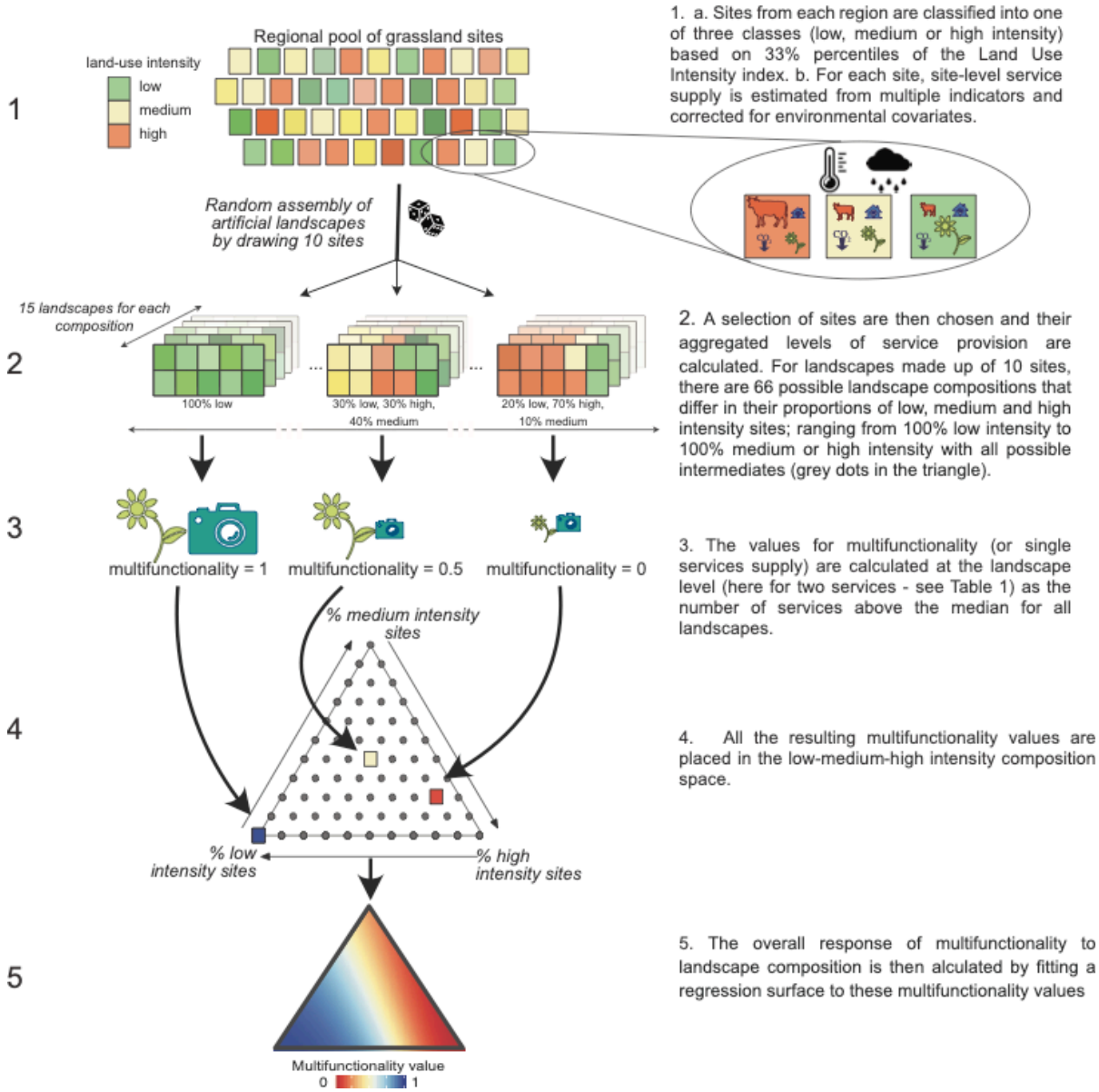
covariates (Figure 1, step 1b), thus ensuring differences were caused by management and not covarying factors. Corrected values were the residuals from linear models in which the ecosystem service indicators were the response variable and predictors were: the region, pH, soil depth, sand and clay content, mean annual temperature, mean annual rainfall (see Allan et al. (2015) and Hijmans et al. (2005) for details on these measurements), and a topographic wetness index (see Le Provost et al. 2021 and supplementary methods). To account for a site's surroundings, we also used the proportion of grassland in a 1km radius as a predictor, as surrounding grassland habitat may act as a source of colonization for local biodiversity (e.g. Henckel et al., 2015; Le Provost et al., 2017; Tscharntke et al., 2012). Surrounding grassland cover was calculated from 2008 land-cover data that were mapped in QGIS v 3.6 and classified into five broad categories: croplands, grasslands, forests, water bodies, roads and urban areas.

All indicators except gamma diversities were corrected at the site level. This was done to avoid biasing landscape estimates with sites from environmental extremes. These would not be detected when averaging at the landscape level. However, plant and bird species gamma diversity can only be calculated at the landscape level, so were corrected at this scale using the landscape-level averages of the same environmental factors as described above, and also a landscape level environmental heterogeneity variable. Landscape heterogeneity was calculated as the volume of the convex hull of the selected sites in a PCA that included all environmental variables.

2.5 Site-level analyses

We first analysed the relationship between all site-level service indicators and land-use intensity class. Within each region, we scaled the services between 0 and 1 and fitted ANOVAs with land-use class as the explanatory variable; followed by a pairwise mean comparison.

Figure 1 Steps of the analysis



2.6 Landscape simulations

An overview of the simulation process is presented in Fig. 1. We conducted the simulations separately for each region, as the relationships between land use and ecosystem services differed strongly between them (Fig. 2).

Landscapes were simulated by randomly drawing sites from the pool of 50 sites from each region (Fig. 1, step 2). Each artificial landscape was composed of ten sites to avoid the high similarity of landscapes composed of more sites. For instance, as there were approximately 16 sites of each intensity class in each region, the most extreme strategies (e.g. 100% low intensity) have only 1 possible landscape composition for landscapes of 16 sites, but 16 possible compositions, which share 90% of their sites for landscapes of 15 sites. For landscapes made up of 10 sites, there are 66 possible landscape compositions that differ in their proportions of low, medium and high intensity sites; ranging from 100% low intensity to 100% medium or high intensity with all possible intermediates (grey dots in the middle triangle of Fig. 1). For each of these compositions, we generated 15 unique artificial landscapes by randomly drawing sites from the regional pool, resulting in 15 x 66 = 990 landscapes. For each simulated landscape, we then calculated landscape-scale ecosystem service indicators, as described above. Simulated landscapes had a combined area of 10 x 50m x 50m sites and so 2.5 ha, but we note that each site is representative of a field ranging from 3 to 148 ha, meaning the results are representative of much larger 'farm-scale' landscapes.

Finally, we calculated landscape-scale ecosystem service supply and multifunctionality (Fig. 1, step 3) as described below. We fitted binomial linear models with multifunctionality as the response variable and with a second-degree polynom of the proportions of low and high land-use intensity as explanatory variables (Fig. 1, step 4-5).

2.7 Landscape-level ecosystem multifunctionality

The response of all combinations of the six main ecosystem services (i.e. single services, all combinations of 2 to 5 services, and all six services) to landscape composition were investigated. Because trade-offs between services mean that it is unlikely that all services can be maintained at high levels (Bennett et al., 2009; Raudsepp-Hearne et al., 2010; van der Plas et al 2019), managers are often faced with hard choices. To simulate the potential compromises that can be made we therefore generated two contrasting multifunctionality metrics (hereafter multifunctionality scenarios - representing different governance decisions), both of which are multifunctionality measures in which the supply is multiplied by stakeholder priority (Manning et al. 2018). In the first, governors choose to provide a small number of services at high levels, e.g. to meet the needs of a single or few groups to the exclusion of others (hereafter 'threshold scenario'). In the second, governors opt for a compromise situation in which all services are provided at moderate levels but without any guarantee of them being high (hereafter 'compromise scenario'). In the threshold scenario, multifunctionality (ranging from 0 to 1) is the proportion of prioritised services that pass a given threshold, and this threshold is the median

value of the service for all landscapes within the considered region. For the compromise scenario we scored multifunctionality as 1 if all services exceeded a 25% threshold (i.e. above the 25% quantile of the service distribution in all landscapes within the region), and 0 otherwise. In addition to these two theoretical scenarios, we also calculated stakeholder-specific multifunctionality values. This was done by scoring each service 1 or 0 depending on whether it passed the 50% threshold (as done in the 'threshold' scenario) and averaging the resulting scores, weighted by the relative priority of each service to the stakeholder group (Manning et al. 2018). Priority scores were obtained from points allocation of the social survey (see section 2.3 on Ecosystem services priority).

2.8 Dependence of multifunctionality range to the number of services included and to environmental covariates

The response of multifunctionality to landscape composition became increasingly complex as more services were considered (see Results). To understand this complexity, and to identify general rules underlying why optimal landscapes could be identified in some cases but not others, we formed hypotheses and conducted additional analyses. As the objective of these analyses was to find generalities in what affects the responsiveness of multifunctionality to drivers, these were performed upon combined data from all three regions.

Multifunctionality responsiveness is the degree by which multifunctionality responds to landscape composition and it was calculated as the range (2.5% to 97.5% quantiles) of the fitted values of the binomial GLM models described above. Visually, this corresponds to the strength of the colour gradient in the triangle plots presented (Fig. 4 and Fig. 5). While the potential range of multifunctionality is always 1 (i.e. varying from none to all services above the threshold within a landscape), in reality the range of fitted values depends on model fit and the degree of response to landscape composition, or other drivers.

To investigate a hypothesised relationship between multifunctionality responsiveness and the number of ecosystem services included in its calculation, we regressed multifunctionality responsiveness upon the number of ecosystem services included in the landscape-scale assessment (ranging from 1 for individual services to 6 for multifunctionality including all services).

Multifunctionality responsiveness was also hypothesised to depend on the degree of contrast in responses to land-use intensity of the different services included in the assessment. To test this, we estimated slope coefficients of linear regressions between each service and land-use

intensity at the site level, and calculated the 'service response variance' as the variance of the slope coefficients. We expected that multifunctionality would decrease with the service response variance (van der Plas et al., 2019), but only when services had a strong response to land use intensity. Thus, we also calculated the 'service response mean' as the mean of the absolute values of the slope coefficients. A high service response mean indicates that all services respond strongly to LUI, irrespective of the direction of the response. This service response mean was then classified as 'high' (resp. 'low') if higher (resp. lower) than the median. We then fitted a linear model in which with the range of multifunctionality values was regressed upon the service response variance, the service response mean as a qualitative variable (high or low), and their interaction.

Finally, we tested the hypothesis that multifunctionality was more responsive to land use intensity when land use played a large role in driving the component functions, relative to other drivers. To do this, we examined the linear relationship between the response of multifunctionality and the relative strength of the LUI effect compared to other environmental covariates. For each single ecosystem service and each region, we quantified the relative strength of the effect of land-use intensity (RS_{LUI}) as:

$$RS_{LUI} = \left[\frac{corr(ES, LUI)}{max_j(\, corr(ES, EC_j)}\right]$$

Where *corr* is the correlation, *ES* the ecosystem service (uncorrected for the environment), *LUI* the value of land-use intensity, and EC_j the environmental covariates.

2.9 Online tool

Given the sensitivity of the results to the choice of services prioritised and the multifunctionality scenario (see results) we developed an online tool to allow users to investigate which landscape composition best delivers multifunctionality, for any given set of ecosystem service priorities (https://neyret.shinyapps.io/landscape_composition_for_multifunctionality/). This Shiny App (R package shiny v. 1.6, Chang et al. 2021) is based on landscape-scale service supply values pre-calculated for a large range of parameters (see supplementary methods for details on sensitivity analyses) and on a R script similar to the one used for the analyses presented in this manuscript. It allows users to select any combination of parameters (number of sites included, choice of multifunctionality metric, etc.) and service priority scores and to investigate the resulting variations in multifunctionality.

3. Results

3.1 Relationships between land-use intensity and ecosystem services

At the single-site scale, the optimal land-use intensity for individual services can be easily identified. Across all regions, fodder production consistently increases with land-use intensity while conservation and aesthetic values respond negatively to land-use intensity (Fig. 2). Carbon stocks do not vary with land-use intensity, because environmental variables such as soil texture and mineralogy play a larger role than those of land use within the study regions (Herold et al. 2014). Foraging opportunities and regional identity did not respond consistently to land-use intensity. The trade-offs and synergies between services at the landscape scale (Fig. 3) are consistent with these site-scale results (Fig. 2). Conservation value is synergic with aesthetic value (Pearson's $r = 0.28$ for all regions, $P < 0.001$) but both display a trade-off with fodder production (respectively $r = - 0.21$ and $r = -0.41$, $P < 0.001$). The other services do not show any consistent correlations.

Figure 2 Relationship between ecosystem service supply and land-use intensity across the study regions. a. Variation of ecosystem services supply with land-use intensity. Values shown are calculated at the site level as the average of their component indicators (see Table 1 and supplementary figures). Values were scaled between 0 and 1. Symbols indicate significant differences (ANOVA and pairwise comparisons; **** $p < 10^{-4}$; *** $p < 10^{-3}$; ** $p < 10^{-2}$; * $p < 0.05$; ° $p < 0.1$). b. Characterisation of land-use intensity based on mowing, grazing and fertilisation levels in the different regions. The width of the symbols is proportional to average values for the corresponding intensity.

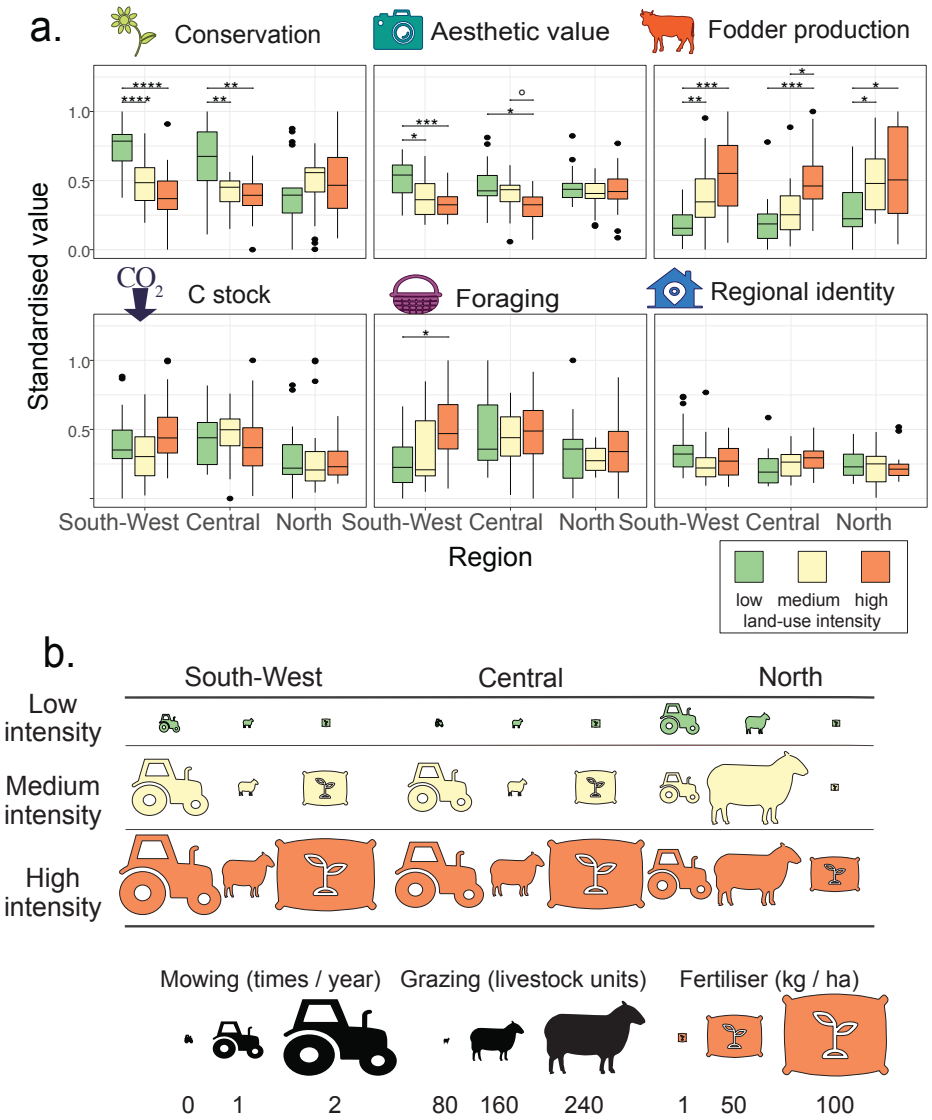
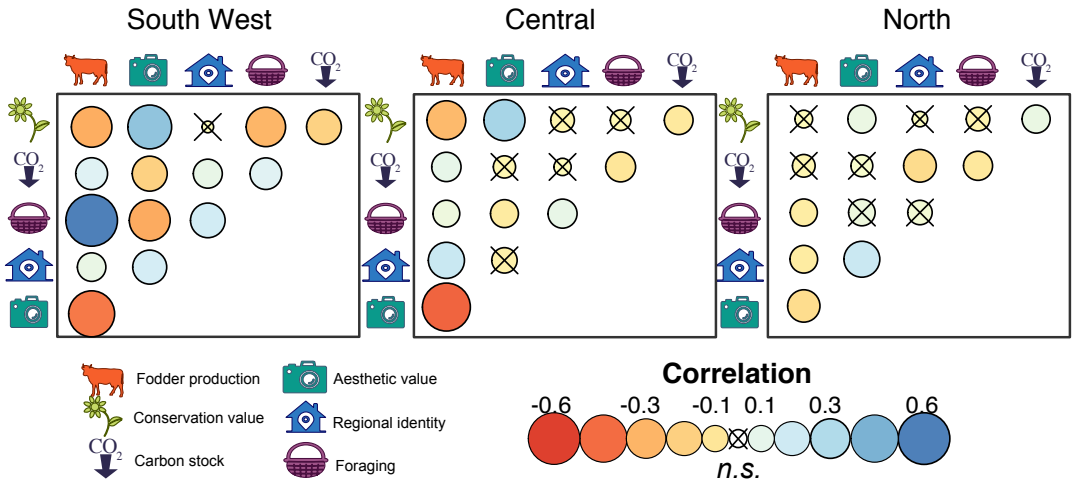


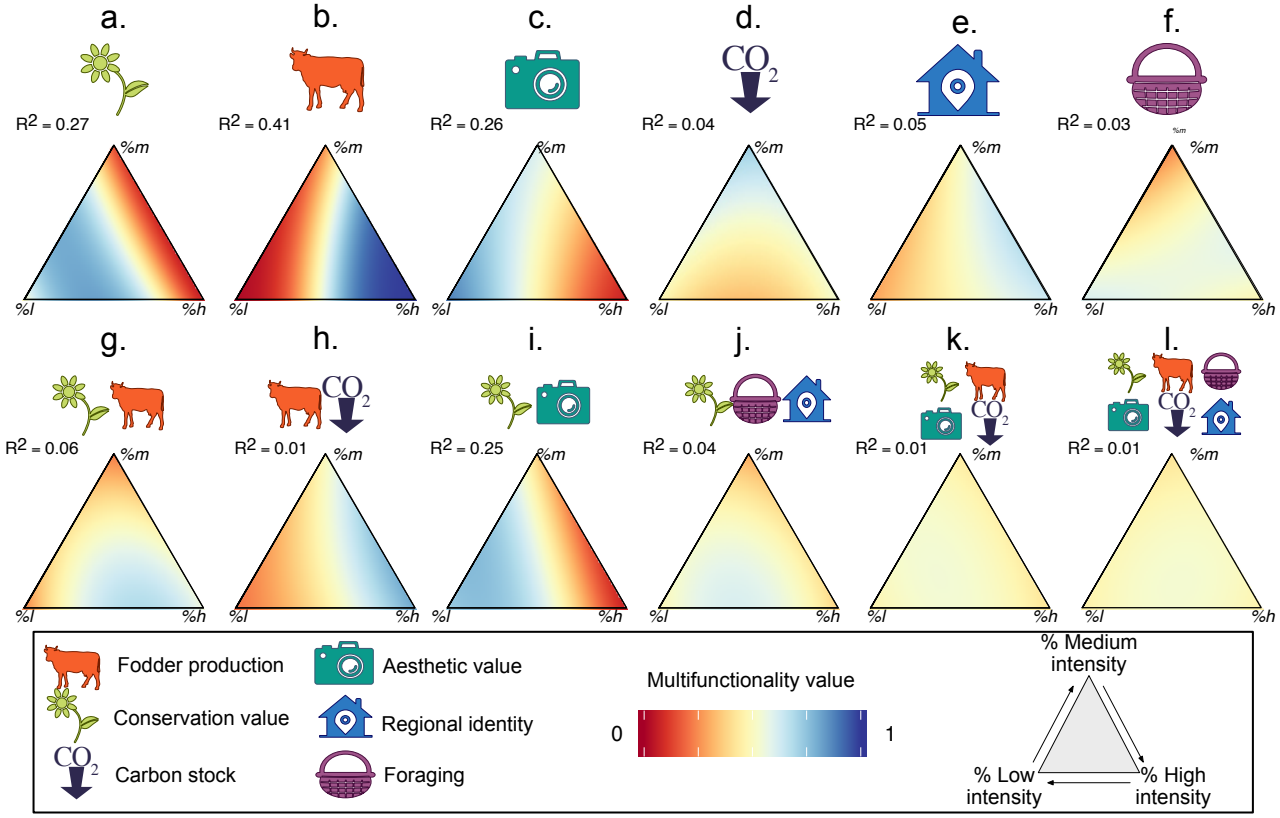
Figure 3 Trade-offs between landscape-scale ecosystem services. The colour and size of the circles denote the strength of the correlation between pairs of variables, within each region. Crosses indicate no significant correlations at 5% (Holm correction for multiple testing).



3.2 Optimal land-use allocation at the landscape scale

At the landscape scale, effective landscape strategies can be identified where only a few services are desired, but optimisation becomes increasingly difficult as more services are considered. The optimal land-use allocation pattern also depends strongly on whether achieving moderate levels of all services, or high levels of a few, is the priority. This sensitivity is best explored in the online tool (https://neyret.shinyapps.io/landscape_composition_for_multifunctionality/), which allows users to investigate the best management strategy given any set of ecosystem service priorities and the impact of land use changes from current conditions. In the text below we highlight a few of the possible combinations of this parameter space, and demonstrate the sensitivity of multifunctionality to multiple factors. We then illustrate our results using data collected from four of the main stakeholder groups of the three study regions: farmers, conservationists, locals and the tourism sector.

Figure 4 Dependency of overall ecosystem service supply on the services demanded and landscape composition. Landscape composition is presented as proportions of low, medium and high-intensity sites, for selected combinations of ecosystem services in the Central region of the Exploratories. For single ecosystem services (top row), the service supply corresponds to the probability of the given service being above the median. For combinations of multiple services (middle and bottom rows), multifunctionality is the proportion of services above the median. A broader set of service combinations in all regions can be found in Figure S3. R^2 values were calculated from generalised linear models (see Methods).



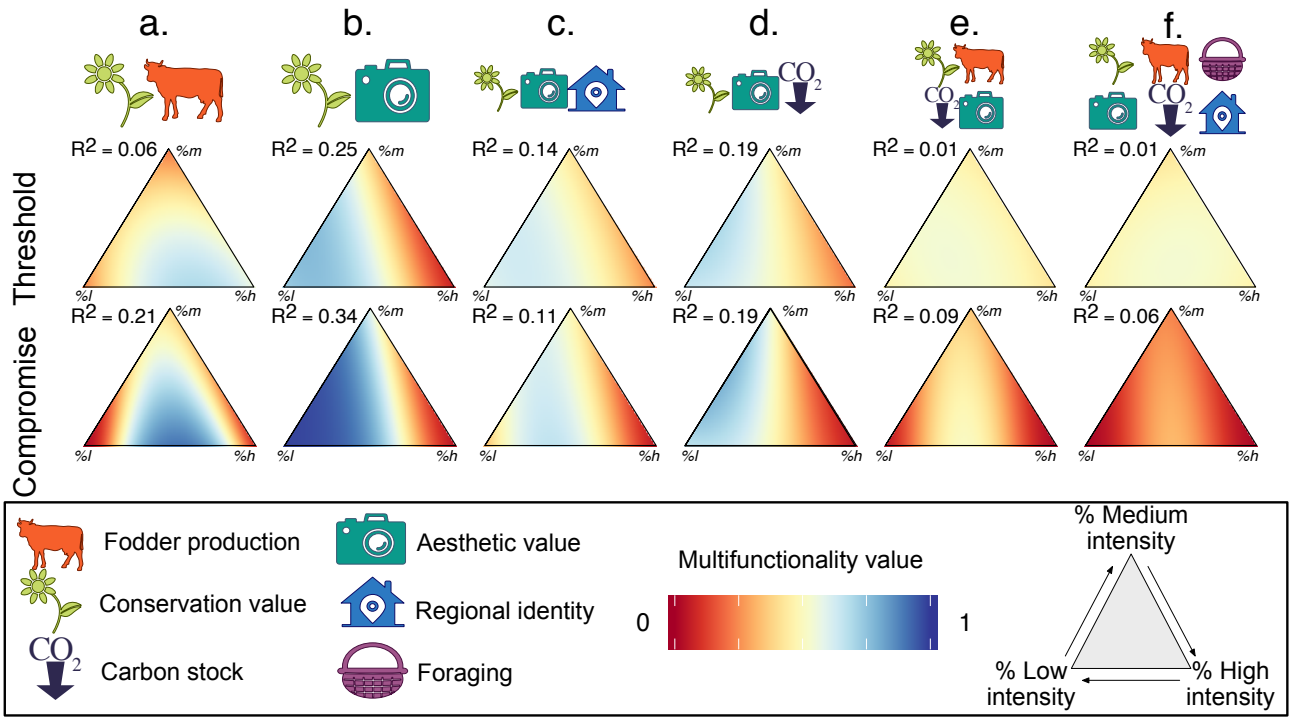
In the first set of examples, we present the 'threshold' scenario in which land governors consider a landscape that provide high levels of some services, potentially to the exclusion of some others. Here we find that for individual services, the optimal landscape composition is predictable and consistent with the site-level results, i.e. that when services are significantly affected by land use intensity, the highest service values are found in homogeneous landscapes composed of sites with land-use intensities favouring that particular service (Fig. 4. a-c). The optimal landscape composition when two ecosystem services are considered depends on whether these services have consistent or contrasting responses to land-use intensity. When the two services are synergic, they behave as a single service and optimal landscape composition is found at the common optimum of the two services. For example, a clear optimum can be found for

conservation and aesthetic value (Fig. 4i). In contrast, if the two services respond contrastingly to land-use intensity, then whether an optimum could be found depends on the form and strength of their relationship with land-use intensity. For example, a common objective of landscape management is to combine food production with biodiversity conservation (Phalan et al. 2011). As there is a strong trade-off between these services (Fig. 3), only a partial optimum with high levels of both services can be found (Fig. 4g), with the landscape composition delivering this depending on regional differences in the response of services to land-use intensity (see Fig. A2 for details), and the relative responsiveness of the services considered to land use intensity. For three or more services (Fig. 4j-l) the identification of a clear optimal land-use strategy becomes even more challenging. In these cases, multifunctionality varies very little across the full range of landscape composition (maximum R^2 17%, and often < 10%, Fig. A2), with relatively uniform multifunctionality values of about 50%, regardless of the landscape composition.

In the 'compromise' scenario, land governors consider a landscape multifunctional if it balances the demands of different stakeholder groups, ensuring moderate, but not necessarily high, level of all services. This gives very different results in comparison to the 'threshold' scenario (for selected service combinations see Fig. 5). The two scenarios give similar results when services are synergic (e.g. Fig. 5b), but when the considered services display a trade-off it is easier to identify successful land-use strategies in the 'compromise' scenario, and especially when there are only two services (Fig. 5a). In this case 'compromise' multifunctionality is highest in landscapes composed of both high- and low-intensity sites, and with few medium-intensity sites, i.e. – broadly similar to a land-sparing strategy, and consistent with our original hypothesis. When multiple services are considered (Fig. 5d-f), variation in multifunctionality across the different

strategies also tends to be higher in the 'compromise' than the 'threshold' scenario, thus indicating than an optimum can be identified.

Figure 5 Dependency of multifunctionality on stakeholder demand patterns, as represented by 'threshold' and 'compromise' scenarios. Values also depend on landscape composition (in proportions of low, medium and high-intensity sites). In the threshold scenario, multifunctionality is calculated as the number of services above the median (top row, repeated from Fig. 4). In the compromise scenario, multifunctionality equals 1 if all services are above the 25th quantile, and 0 otherwise (bottom row). R^2 values were calculated from generalised linear models (see Methods). Only data from the Central region and certain service combinations are presented, other service combinations and regions can be found in Figure S2 and S9.



Optimal land-use for contrasting stakeholder groups

Different stakeholder groups had contrasting service demand, but all groups allocated points to all services (Fig. A1). This demand for multiple services resulted in relatively weak responses of overall multifunctionality to landscape composition (Fig. A3). Consistent with the results described above, the optimal landscape also depended on the relative priority of different services. For instance, farmers reported a clear priority for productivity (40% of the total priority score for grassland services) and so find their optimum in landscapes composed of mostly high-intensity sites. In contrast, nature conservation associations prioritised biodiversity (39%) and so found their optimum in landscapes dominated by low-intensity sites. Tourists and locals weighted

services more equally (e.g. 26% conservation, 22% aesthetic value, 20% regional identity for tourists) meaning their optimum landscape compositions were less easy to identify, although an optimum was usually found for intermediate to high proportions of low-intensity sites.

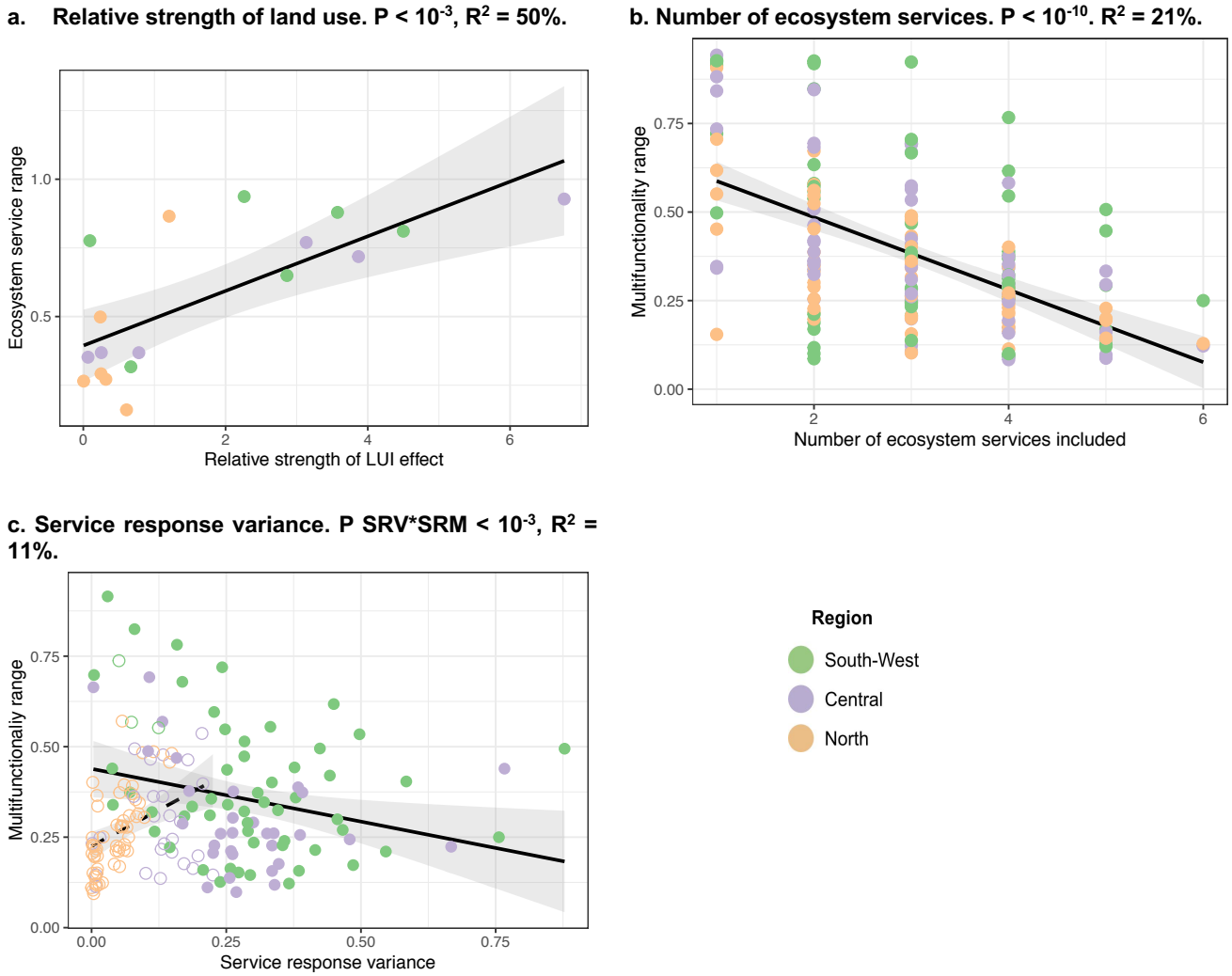
3.3 Identifying the drivers of multifunctionality

To explore why optimal landscape strategies cannot always be identified when multiple services are prioritised, we tested several hypotheses. Hypothesis 1, some services are primarily driven by environmental drivers (e.g. climate and underlying geology) and thus respond weakly to changes in landscape composition. Hypothesis 2, if services respond contrastingly to land use intensity, and so trade-off, then their respective contributions to multifunctionality counteract each other, resulting in a flat response of multifunctionality to changes in landscape composition. Hypothesis 3, increasing the number of services demanded increases the chance that service responses will contrast, and also aggregates increasing amounts of variation, thus weakening the response of multifunctionality to landscape composition).

Hypothesis 1 was supported: in the threshold scenario, the responsiveness of multifunctionality to landscape composition (defined as the range of fitted multifunctionality values, visualised as "colourfulness" in Fig. 4-5; see Methods) increased when land-use intensity had a relatively large effect on the services compared to other environmental drivers (Fig. 6a, $P < 10^{-3}$, $R^2 = 50\%$). Hypotheses 2 and 3 were also supported; variation in multifunctionality decreased with increasing variance in the response of services to land-use intensity, at least when the service response were strong (high SRM) (Fig. 6c, $P < 10^{-3}$, $R^2 = 11\%$), and with increases in the numbers of services included in the analysis (Fig. 6b, $P < 10^{-3}$. $R^2 = 21\%$). Most of these relationships were also found in the compromise scenario (Table C1).

3.4 Additional analyses

Figure 6 Factors explaining the sensitivity of multifunctionality to landscape composition. Figures show the responsiveness of multifunctionality (range between 5% and 95% quantiles of the predicted values) to landscape composition depending on (a) the strength of each individual ecosystem service's response to landscape composition relative to the effects of land use and environmental covariates (b) the number of ecosystem services included in the calculation of multifunctionality (all possible combinations, of 1 to 4 services) and (c) the service-response variance among the included ecosystem services (all possible combinations); P SRV*SRM indicates p-value for the interaction between service response variance and the service response mean. Each dot represents individual services (a), or one combination of services (b, c), per region. The lines show the prediction of a linear model, with multifunctionality range as the response and the considered factor as the explanatory variable. In panel c, filled dots and solid line show high mean response; and empty dots with dashed line low mean response.



In addition to the main cases presented here we identified several other sensitivities including additional metrics for multifunctionality calculation at the landscape level, the number of sites included in each landscape, and the use of raw data instead of that corrected for environmental variation. We encourage readers to explore these sensitivities in the app, although the corresponding figures and specificities are also presented extensively in the supplementary information (Supplementary Fig. C3 to C15). In addition, we also investigated the response of

multifunctionality to the mean and coefficient of variation of land-use intensity at the landscape level. The results were similar to those described here in that unless the services were synergic, no clear optimum could be found when several services were demanded (Fig. C17).

5. Discussion

While the land-sharing or -sparing debate has aided our understanding of the trade-offs between commodity production and conservation (Phalan, 2018) we show that strategies that are broadly comparable to these cannot provide high multifunctionality in grassland landscapes, if high levels of multiple ecosystem services are desired. In particular, while our approach allows stakeholders to identify an optimal landscape composition for any given set of priorities, a demand for multiple ecosystem services typically led to low to moderate multifunctionality that hardly varied with landscape composition. This indicates that there is no landscape composition in which supply can meet the demand for all services. We predict that this difficulty in achieving high multifunctionality is general to many ecosystems and landscapes, as the presence of other drivers and weak and negative correlations between services are commonplace (Bennett et al., 2009). Various studies have advocated for the consideration of more complex strategies for balancing commodity production with conservation (Bennett, 2017; Butsic & Kuemmerle, 2015; Chan et al. 2006; Fischer et al., 2014; Phalan et al., 2011, Simons and Weisser 2017). By employing a rigorous approach based on direct, in-field measurements of ecosystem service indicators, we further show that considering not only trade-offs and synergies between ecosystem services, but also information describing the ecosystem service priorities of stakeholders, helps identify land management options that have greater precision and relevance to land users. The approach presented also allows the potential causes of land-use conflicts to be identified, as it can assess whether low multifunctionality is caused by trade-offs in the supply of ecosystem services, or unrealistic and incompatible demands on the ecosystem by stakeholders. Furthermore, the impact of any given land use change, e.g. deintensification of 50% of medium intensity grasslands, on stakeholders with different priorities can be estimated and accounted for in decision making.

In our study system, ecosystem services showed contrasting responses to land-use intensity, including the commonly observed trade-off between production and biodiversity or cultural services (Allan et al., 2015; Cordingley et al., 2016; Lavorel et al., 2011; Raudsepp-Hearne et al., 2010). Understanding contrasting responses of ecosystem services to land management is fundamental to identifying landscape-level strategies. Here, we show that strong management-driven trade-offs preclude multifunctionality when high levels of services are required. As a result, even complex landscape strategies may fail to deliver high levels of multiple ecosystem services

(Allan et al., 2015) and landscape management is likely to require “hard choices” (Cordingley et al., 2016; Slade et al., 2017) regarding which services to prioritise, and which are secondary. Social survey of stakeholders in our study system showed that all stakeholders demanded multifunctionality (supporting previous results, e.g. Hölting et al. 2020). Although it is possible to find a landscape optimum when stakeholders clearly prioritise one service (e.g., farmers and livestock production), the multifunctionality demanded makes landscape optimisation difficult even for a single stakeholder group, as multiple services that do not synergise were often given equally high priority.

Although high levels of all services may be unattainable, we show that it is possible to provide limited levels of multiple services by combining sites at low and high intensities, a strategy broadly similar to land-sparing. In this respect, our results show that the optimal strategy depends heavily the levels of service provision landscape managers are willing to accept. While different stakeholders favour different sets of services, landscape-level governors are faced with a difficult choice: create a landscape with a few services at high value, which will create winners and losers among stakeholder groups, or one that minimises the trade-offs among services so that all are present at moderate levels, meaning that all stakeholder groups must accept sub-optimal levels of ecosystem services.

While advancing on previous studies by incorporating multiple services, we acknowledge that our approach to identifying optimal landscape strategies is simple and ignores much of the complexity found in natural systems. First, land use decisions are also affected by the relationship between ecosystem service supply and the amount of benefit provided: while we considered the supply-benefit relationship for all services as behaving in a threshold manner, this is clearly a simplification. More accurate and service specific representation of the relationships between supply and benefit (e.g. Linders et al. 2021) could provide additional insights into which landscape compositions best deliver multifunctionality. For instance, at very low intensity fodder production might not profitable enough for farmers, resulting in no benefit despite some supply, and with a risk of pasture abandonment. While tailored supply-benefit relationships have been rarely used in multifunctionality metrics due to the difficulty in defining their shape, their integration could greatly improve model predictions (Manning et al. 2018, Linders et al. 2020). Both the prioritisation of services by stakeholders, and supply-benefit relationships, are also likely to change in time and space (Boesing et al. 2020), e.g. as market forces and society change. Incorporating this temporal dimension remains a challenge for future multifunctionality research.

An aspect highlighted by our study is the difficulty in identifying land management strategies when ecosystem services respond to multiple drivers. Additional drivers can be either anthropogenic (e.g. other land-use changes, overexploitation, Carpenter et al., (2009)) or environmental (e.g. soil (Adhikari & Hartemink, 2016), climate, or elevation (Lavorel et al., 2011)). Failing to account for these drivers can obscure the relationship between land-use composition and multifunctionality. In this study, the use of real-plot data at least partly integrates this inherent variability, which also explains why some trends are not as clear-cut as those that could be obtained from models using fixed values. Environmental drivers will differ in their effect on different services, and so can modify their trade-offs (Clec'h et al., 2019). Therefore, the development of strategies to achieve landscape multifunctionality also needs to be informed by regional knowledge (Butsic & Kuemmerle, 2015; Clec'h et al., 2019). For instance, in our analysis the Northern region responded very differently to the other two regions. This was due to regional specificities, such as its uniformly low plant diversity and the association of low-intensity sites with organic soils, which shifted the optimal landscape compositions to different regions of the triangular space compared to the other regions (Fig. A2). Despite this context-dependency, we expect that our study areas are representative of their regions both in terms of environmental conditions and management, making our results broadly applicable to Germany.

In addition to local drivers, the delivery of many ecosystem services depends on the movement of matter or organisms among landscape units (Mitchell et al., 2014). For instance, pollination, water quality, or pest and disease control are affected by landscape complexity, fragmentation and surrounding land uses (Duarte et al., 2018). Accordingly, we advocate the incorporation of spatial interactions between landscape units (Lindborg et al., 2017) into future models, elements which may modify and expand upon the conclusions presented here.

Our system consists of only one land-use type and does not include unmanaged land. This makes it only broadly comparable to the land-sparing and -sharing strategies, which typically integrate more diverse land-use types and also aspects relating to their spatial organisation in a landscape (e.g., land sharing strategies may lead to a patchwork of intensively farmed and semi-natural landscape units, in which overall service provision may be influenced by their configuration). However, we argue that the approach presented here could be extended to many different land-use and management regimes, provided that appropriate data on services and drivers is available. Steps must also be taken to ensure that insights from such studies are in a format that can be communicated effectively to land managers. For instance, we argue that presenting proportions of land in a number of land-use categories is more easily transferable than indices of land-use intensity heterogeneity. Strategies for knowledge transfer also need to be developed. We suggest that online tools like the one presented here provide a useful

demonstration tool for communicating land-use options to land managers and policy makers, as they could be used to explore options, understand the causes of conflicts and trigger discussions, thus helping to support decision-making among different groups of stakeholders. However, full application of findings such as those presented here also requires the existence of structures that aim to identify landscape strategies and operationalise them at a community level, such as the 'landscape approach' (DeFries & Rosenzweig, 2010; Sayer et al., 2013). This aims to balance competing land-use priorities to promote environmental conservation and human well-being based on a participatory approach (e.g. the African Forest Landscape Restoration Initiative). Government and corporate policies can also implement such strategies, e.g. via agri-environment schemes that may guide the allocation of different land-use types or land-use intensities to different parts of the landscape (Whittingham, 2011). We suggest that demonstrating the impacts of different management options via apps such as that presented here, can foster understanding and aid decision making in both of these settings.

Overall, this study shows that landscape strategies are highly sensitive to the identity of the services desired and the type of multifunctionality demanded by stakeholders, making participatory approaches to the development of land management strategies essential. When high levels of all services are required, we show that optimising landscape composition is usually possible for two services. However, when there are strong trade-offs among services or significant effects of other environmental drivers, winning management options become increasingly hard to identify unless stakeholders are willing to accept moderate service levels, which can be delivered by strategies akin to land sparing. Across the world, landscapes are increasingly required to provide a wide range of services. This study stresses the need for both theoretical studies and applied social and ecological research into which services are required, at what scale, and how they are affected by environmental drivers. Such knowledge is essential if we are to identify land-use strategies that minimise conflict between stakeholders, and promote the sustainable use of all ecosystem services.

5. Acknowledgements

We thank Steffen Boch for his help in designing the list of edible and culturally important plants. We thank the managers of the three Exploratories Konstans Wells, Swen Renner, Kirsten Reichel-Jung, Sonja Gockel, Kerstin Wiesner, Katrin Lorenzen, Andreas Hemp, Martin Gorke and all former managers for their work in maintaining the plot and project infrastructure; Simone Pfeiffer, Maren Gleisberg and Christiane Fischer for giving support through the central office, Jens Nieschulze and Michael Owonibi for managing the central data base, and Markus Fischer, Eduard Linsenmair, Dominik Hessenmöller, Daniel Prati, Ingo Schöning, François

Buscot, Ernst-Detlef Schulze, Wolfgang W. Weisser and the late Elisabeth Kalko for their role in setting up the Biodiversity Exploratories project.

The work has been funded by the DFG Priority Program 1374 “Infrastructure-Biodiversity-Exploratories”. Field work permits were issued by the responsible state environmental offices of Baden-Württemberg, Thüringen, and Brandenburg.

6. Data accessibility

The data used in this paper and full code to reproduce the results will be made available via GitHub/FigShare and Zenodo.

7. References

Adhikari, K., & Hartemink, A.E. (2016). Linking soils to ecosystem services—A global review. *Geoderma*, *262*, 101–111. http://dx.doi.org/10.1016/j.geoderma.2015.08.009

Allan, E. et al. (2015). Land use intensification alters ecosystem multifunctionality via loss of biodiversity and changes to functional composition. *Ecology Letters*, *18*(8), 834–843. https://doi.org/10.1111/ele.12469

Barnes, A.D., et al. (2017). Direct and cascading impacts of tropical land-use change on multi-trophic biodiversity. *Nat Ecol Evol* 1, 1511–1519. https ://doi.org/10.1038/s41559-017-0275-7

Bengtsson, J. et al. (2019). Grasslands—More important for ecosystem services than you might think. *Ecosphere*, *10*(2), e02582. https://doi.org/10.1002/ecs2.2582

Bennett, E.M. (2017). Changing the agriculture and environment conversation. *Nature Ecology and Evolution*, *1*(1), 1–2. DOI: 10.1038/s41559-016-0018

Bennett, E.M. et al. (2009). Understanding relationships among multiple ecosystem services. *Ecology Letters*, *12*(12), 1394–1404. doi: 10.1111/j.1461-0248.2009.01387.x

Blüthgen, N., et al. (2012). A quantitative index of land-use intensity in grasslands: Integrating mowing, grazing and fertilization. *Basic and Applied Ecology*, *13*(3), 207–220. http://dx.doi.org/10.1016/j.baae.2012.04.001

Boesing et al. (2020). Ecosystem services at risk: integrating spatiotemporal dynamics of supply and demand to promote long-term provision. *One Earth*, 3(6) 704-713. https://doi.org/10.1016/j.oneear.2020.11.003

Butsic, V., & Kuemmerle, T. (2015). Using optimization methods to align food production and biodiversity conservation beyond land sharing and land sparing. *Ecological Applications*, *25*(3), 589–595. DOI: 10.1890/14-1927.1

Byrnes, J.E.K., et al. (2014). Investigating the relationship between biodiversity and ecosystem multifunctionality: challenges and solutions. *Methods Ecol Evol* 5, 111–124. https://doi.org/10.1111/2041-210X.12143

Carpenter, S.R., et al. (2009). Science for managing ecosystem services: Beyond the Millennium Ecosystem Assessment. *Proceedings of the National Academy of Sciences*, *106*(5), 1305–1312. https://doi.org/10.1073/pnas.0808772106

Chan, K.M.A., Shaw, M.R., Cameron, D.R., Underwood, E.C., Daily, G.C., 2006. Conservation Planning for Ecosystem Services. PloS Biol 4, e379. https://doi.org/10.1371/journal.pbio.0040379

Chang, W. et al. (2021). shiny: Web Application Framework for R. R package version 1.6.0. https://CRAN.R-project.org/package=shiny

Clec'h, L., et al. (2019). Assessment of spatial variability of multiple ecosystem services in grasslands of different intensities. *Journal of Environmental Management*. https://doi.org/10.1016/j.jenvman.2019.109372

Cordingley, J.E., et al. (2016). Can landscape-scale approaches to conservation management resolve biodiversity–ecosystem service trade-offs? *Journal of Applied Ecology*, 96–105. doi: 10.1111/1365-2664.12545

DeFries, R., & Rosenzweig, C. (2010). Toward a whole-landscape approach for sustainable land use in the tropics. *Proceedings of the National Academy of Sciences*, *107*(46), 19627–19632. https://doi.org/10.1073/pnas.1011163107

Duarte, G.T., et al. (2018). The effects of landscape patterns on ecosystem services: Meta-analyses of landscape services. *Landscape Ecology*, *33*(8), 1247–1257. https://doi.org/10.1007/s10980-018-0673-5

Fischer, J., et al. (2014). Land sparing versus land sharing: Moving forward. *Conservation Letters*, *7*(3), 149–157. doi: 10.1111/conl.12084

Fischer, M., et al. (2010). Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. *Basic and Applied Ecology*, *11*(6), 473–485. https://doi.org/10.1016/j.baae.2010.07.009

Fisher, B., & Turner, R.K. (2008). Ecosystem services: Classification for valuation. *Biological Conservation*, *141*(5), 1167–1169. https://doi.org/10.1016/j.biocon.2008.02.019

Frei, B., et al. (201. Bright spots in agricultural landscapes: Identifying areas exceeding expectations for multifunctionality and biodiversity. *J Appl Ecol* 55, 2731–2743. https ://doi.org/10.1111/1365-2664.13191

Goldstein, J.H., et al. (2012). Integrating ecosystem-service trade-offs into land-use decisions. *Proceedings of the National Academy of Sciences*, *109*(19), 7565–7570. https://doi.org/10.1073/pnas.1201040109

Green, R.E. (2005). Farming and the Fate of Wild Nature. *Science*, *307*(5709), 550–555. https ://doi.org/10.1126/science.1106049

Haines-Young, R. and M.B. Potschin (2017). Common International Classification of Ecosystem Services (CICES) V5.1 and Guidance on the Application of the Revised Structure.

Hedblom, M., et al. (2014). Bird song diversity influences young people's appreciation of urban landscapes. *Urban Forestry & Urban Greening*, *13*(3), 469–474. https://doi.org/10.1016/j.ufug.2014.04.002

Henckel, L., et al. (2015). Organic fields sustain weed metacommunity dynamics in farmland landscapes. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1808), 20150002–20150002. http://dx.doi.org/10.1098/rspb.2015.0002

Herold, N., Schöning, I., Michalzik, B. *et al.* Controls on soil carbon storage and turnover in German landscapes. *Biogeochemistry* **119,** 435–451 (2014). https://doi.org/10.1007/s10533-014-9978-x

Herrero, M., Havlik, P., Valin, H., Notenbaert, A., Rufino, M.C., Thornton, P.K., Blummel, M., Weiss, F., Grace, D., Obersteiner, M., 2013. Biomass use, production, feed efficiencies, and greenhouse gas emissions from global livestock systems. Proceedings of the National Academy of Sciences 110, 20888–20893. https://doi.org/10.1073/pnas.1308149110

Hijmans, R.J., et al. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

IPBES (2019): Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Díaz, S., et al. IPBES secretariat, Bonn, Germany. 56 pages. https ://doi.org/10.5281/zenodo.3553579

Knoke, T., et al. (2020). Accounting for multiple ecosystem services in a simulation of land-use decisions: Does it reduce tropical deforestation? Glob Change Biol 26, 2403–2420. https://doi.org/10.1111/gcb.15003

Lavorel, S., et al (2011). Using plant functional traits to understand the landscape distribution of multiple ecosystem services. *Journal of Ecology*, *99*(1), 135–147. https://doi.org/10.1111/j.1365-2745.2010.01753.x

Le Provost, G., et al. (2017). Trait-matching and mass effect determine the functional response of herbivore communities to land-use intensification. *Functional Ecology*, *31*(8), 1600–1611. https://doi.org/10.1111/1365-2435.12849

Le Provost et al. (2021) Contrasting responses of above- and belowground diversity to multiple components of land-use intensity. *Nature Communications.* (in press)

Lee, M.A. (2018). A global comparison of the nutritive values of forage plants grown in contrasting environments. *Journal of Plant Research*, *131*(4), 641–654. https://doi.org/10.1007/s10265-018-1024-y

Linders, T. et al. (2021). Stakeholder priorities determine the impact of an alien tree invasion on ecosystem multifunctionality*. People & Nature*. https://doi.org/10.1002/pan3.10197

Lindborg, R., et al. (2017). How spatial scale shapes the generation and management of multiple ecosystem services. *Ecosphere*, *8*(4), e01741. https://doi.org/10.1002/ecs2.1741

Mace, G.M., et al. (2012). Biodiversity and ecosystem services: a multilayered relationship. *Trends in Ecology & Evolution*, *27*(1), 19–26. doi:10.1016/j.tree.2011.08.006

Manning, P., et al. (2015). Grassland management intensification weakens the associations among the diversities of multiple plant and animal taxa. *Ecology*, *96*(6), 1492–1501. https://doi.org/10.1890/14-1307.1

Manning, P., et al. (2018). Redefining ecosystem multifunctionality. *Nature Ecology & Evolution*, *2*(3), 427–436. https://doi.org/10.1038/s41559-017-0461-7

Mitchell, M.G.E., et al. (2014). Forest fragments modulate the provision of multiple ecosystem services. *Journal of Applied Ecology*, *51*(4), 909–918. https://doi.org/10.1111/1365-2664.12241

Newbold, T., et al. (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, *520*(7545), 45–50. https://doi.org/10.1038/nature14324

Phalan, B. (2018). What have we learned from the land sparing-sharing model? *Sustainability*, *10*(6), 1760. doi:10.3390/su10061760

Phalan, B. et al. (2011). Reconciling food production and biodiversity conservation: Land sharing and land sparing compared. *Science*, *333*(6047), 1289–1291. DOI: 10.1126/science.1208742

Raymond, C.M., et al. (2009). Mapping community values for natural capital and ecosystem services. Ecological Economics 68, 1301–1315. https://doi.org/10.1016/j.ecolecon.2008.12.006

Raudsepp-Hearne, C., et al. (2010). Ecosystem service bundles for analysing tradeoffs in diverse landscapes. *Proceedings of the National Academy of Sciences*, *107*(11), 5242–5247. https://doi.org/10.1073/pnas.0907284107

Sayer, J., et al. (2013). Ten principles for a landscape approach to reconciling agriculture, conservation, and other competing land uses. *Proceedings of the National Academy of Sciences*, *110*(21), 8349–8356. https://doi.org/10.1073/pnas.1210595110

Simons, N.K., & eisser, W. W. (2017). Agricultural intensification without biodiversity loss is possible in grassland landscapes. *Nature Ecology & Evolution*. https://doi.org/10.1038/s41559-017-0227-2

Slade, E.M., et al. (2017). The importance of species identity and interactions for multifunctionality depends on how ecosystem functions are valued. *Ecology*, *98*(10), 2626–2639. DOI: 10.1002/ecy.1954

Soliveres, S., et al. (2016). Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality. *Nature*, *536*. https://doi.org/doi:10.1038/nature19092

Tscharntke, T., et al. (2012). Landscape moderation of biodiversity patterns and processes—Eight hypotheses. *Biological Reviews*, *87*(3), 661–685. https://doi.org/10.1111/j.1469-185X.2011.00216.x

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. In *Journal of Statistical Software* (Vol. 45, Issue 3, pp. 1–67). https://www.jstatsoft.org/v45/i03/

van der Plas, F., et al. (2019). Towards the development of general rules describing landscape heterogeneity-multifunctionality relationships. *Journal of Applied Ecology*, *56*(1), 168–179. https://doi.org/10.1111/1365-2664.13260

Waghorn, G., Clark, D., 2004. Feeding value of pastures for ruminants. New Zealand Veterinary Journal 52, 320–331. https://doi.org/10.1080/00480169.2004.36448

Whittingham, M.J. (2011). The future of agri-environment schemes: Biodiversity gains and ecosystem service delivery?: Editorial. *Journal of Applied Ecology*, *48*(3), 509–513. https://doi.org/10.1111/j.1365-2664.2011.01987.

Landscape management for grassland multifunctionality

Supplementary material

Appendix A. Additional information

Table A1 Variation in the ecosystem service measures and their component service indicators, with the land-use intensity class in each region. Values were first corrected for the environment (see Methods). Different letters indicate differences significant at 5%.

Ecosystem services or *service indicator*	Land-use intensity	South-West	Central	North
Livestock production	low	-66.45 +/- 18.44 [a]	-37.2 +/- 14.36 [a]	-34.97 +/- 14.71 [a]
	medium	9.9 +/- 19 [b]	-10.13 +/- 14.8 [a]	11.09 +/- 15.17 [ab]
	high	57.12 +/- 18.44 [b]	46.73 +/- 14.36 [b]	24.53 +/- 14.71 [b]
Biomass production	low	-33.44 +/- 7.04 [a]	-20.06 +/- 5.84 [a]	-15.8 +/- 5.16 [a]
	medium	5.69 +/- 7.25 [b]	-4.54 +/- 6.02 [a]	5.56 +/- 5.31 [b]
	high	28.09 +/- 7.04 [b]	24.33 +/- 5.84 [b]	10.57 +/- 5.16 [b]
N content	low	-0.01 +/- 0.08 [a]	-0.1 +/- 0.07 [a]	0.07 +/- 0.13 [a]
	medium	0.01 +/- 0.08 [a]	0.02 +/- 0.08 [a]	-0.03 +/- 0.13 [a]
	high	-0.01 +/- 0.08 [a]	-0.02 +/- 0.07 [a]	0.07 +/- 0.13 [a]
Aesthetic	low	0.51 +/- 0.03 [b]	0.49 +/- 0.04 [b]	0.43 +/- 0.02 [a]
	medium	0.37 +/- 0.03 [a]	0.41 +/- 0.04 [ab]	0.4 +/- 0.03 [a]
	high	0.33 +/- 0.03 [a]	0.33 +/- 0.04 [a]	0.41 +/- 0.02 [a]
Flower cover (sqrt)	low	0.17 +/- 0.2 [a]	0.29 +/- 0.24 [a]	-0.06 +/- 0.15 [a]
	medium	-0.25 +/- 0.21 [a]	-0.02 +/- 0.25 [a]	0 +/- 0.16 [a]
	high	0.06 +/- 0.2 [a]	-0.27 +/- 0.24 [a]	0.07 +/- 0.15 [a]
Butterfly abundance (sqrt)	low	2.75 +/- 0.73 [b]	1.32 +/- 0.77 [b]	0.56 +/- 0.46 [a]
	medium	-0.48 +/- 0.76 [a]	0.28 +/- 0.79 [ab]	-0.58 +/- 0.47 [a]
	high	-2.7 +/- 0.73 [a]	-1.36 +/- 0.77 [a]	0.16 +/- 0.46 [a]
Bird family richness	low	0.95 +/- 0.54 [a]	0.84 +/- 0.72 [a]	0.49 +/- 0.5 [a]
	medium	-0.24 +/- 0.56 [a]	0.28 +/- 0.74 [a]	-0.03 +/- 0.52 [a]
	high	-0.88 +/- 0.54 [a]	-0.99 +/- 0.72 [a]	-0.42 +/- 0.5 [a]
Conservation *(plant richness)*	low	9.24 +/- 1.97 [b]	10.13 +/- 2.64 [b]	-0.71 +/- 1 [a]
	medium	-3.04 +/- 2.03 [a]	-3.77 +/- 2.72 [a]	-0.05 +/- 1.03 [a]
	high	-6.38 +/- 1.97 [a]	-6.58 +/- 2.64 [a]	0.76 +/- 1 [a]
C stock	low	-0.04 +/- 0.22 [a]	-0.02 +/- 0.24 [a]	-0.02 +/- 0.56 [a]
	medium	-0.27 +/- 0.22 [a]	0.11 +/- 0.25 [a]	0.27 +/- 0.58 [a]
	high	0.3 +/- 0.22 [a]	-0.08 +/- 0.24 [a]	-0.23 +/- 0.56 [a]
Foraging *(cover edible plants)*	low	-25.63 +/- 12.12 [a]	4.06 +/- 12.12 [a]	11.4 +/- 22.52 [a]
	medium	-2.57 +/- 12.5 [ab]	-3.59 +/- 12.49 [a]	-18.51 +/- 23.22 [a]
	high	23.98 +/- 12.12 [b]	2.21 +/- 12.12 [a]	7.2 +/- 22.52 [a]
Regional ID	low	0.36 +/- 0.04 [a]	0.24 +/- 0.03 [a]	0.26 +/- 0.03 [a]
	medium	0.25 +/- 0.04 [a]	0.27 +/- 0.03 [a]	0.25 +/- 0.03 [a]
	high	0.27 +/- 0.04 [a]	0.31 +/- 0.03 [a]	0.24 +/- 0.03 [a]
Juniperus grasslands	low	0.24 +/- 0.07 [a]	0	0
	medium	0.06 +/- 0.07 [a]	0	0
	high	0 +/- 0.07 [a]	0	0
Cover cultural plants	low	-13.61 +/- 14.88 [a]	-23.4 +/- 15.2 [a]	-11.39 +/- 18.84 [a]
	medium	-17.34 +/- 15.34 [a]	-0.68 +/- 15.67 [ab]	-1.86 +/- 19.42 [a]
	high	27.06 +/- 14.88 [a]	34.08 +/- 15.2 [b]	5.98 +/- 18.84 [a]
Richness cultural birds	low	0.51 +/- 0.31 [a]	0.18 +/- 0.39 [a]	0.17 +/- 0.27 [a]
	medium	-0.15 +/- 0.32 [a]	0.16 +/- 0.4 [a]	-0.23 +/- 0.28 [a]
	high	-0.47 +/- 0.31 [a]	0.1 +/- 0.39 [a]	-0.28 +/- 0.27 [a]

Figure A 1. Stakeholder groups' ecosystem preferences. a. Mean demand score for all considered services, per stakeholder group (coloured dots) or all groups considered (crosses). We then retained for the analysis only the six main services that can be delivered by grasslands (red box). Four stakeholder groups are presented as an example in this paper; they are shown with full dots while the other groups are shown with empty circles. b. Mean demand scores for the services included in the analysis of the current paper, for the selected stakeholder group; the demand scores were normalised by the number of points given by responders to only those six services.

a.

b.

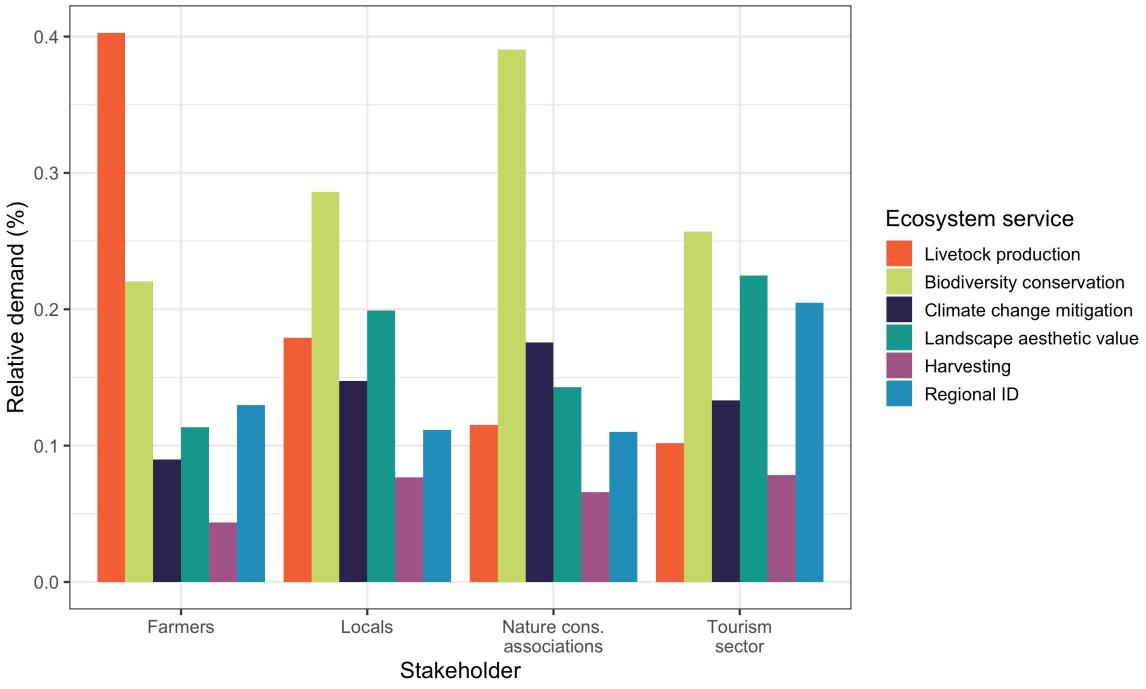


Figure A2 Estimated multifunctionality values depending on landscape composition (in proportions of low, medium and high-intensity sites). **This figure shows a wider selection of the service combinations for the 'threshold' approach shown in Fig. 4.**

For single ecosystem services (top row), the value presented corresponds to the probability of the given service to be above the median. For combinations of multiple services (middle and bottom rows), multifunctionality is the expected proportion of services above the median. Blue indicates higher multifunctionality values, orange lower.

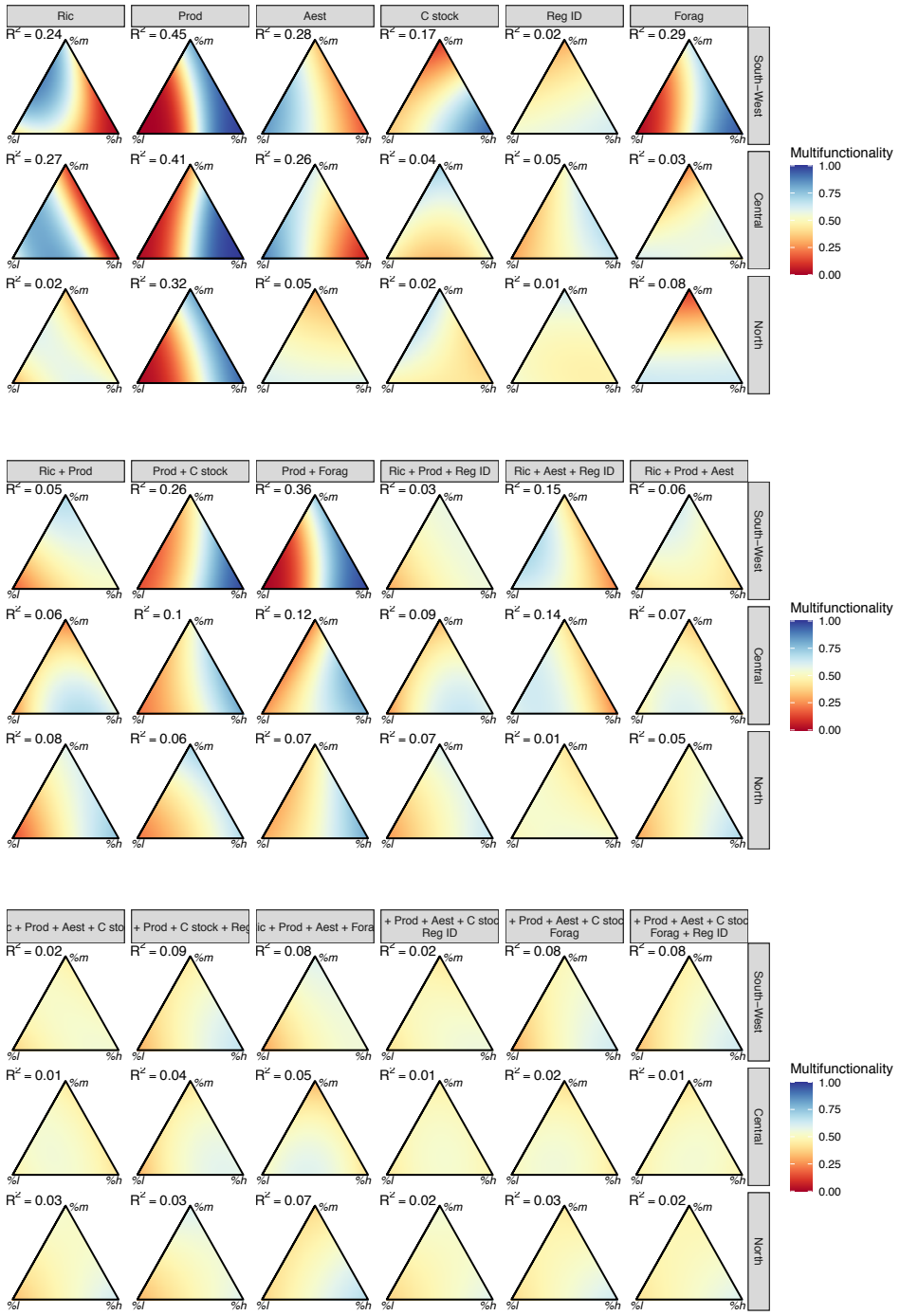
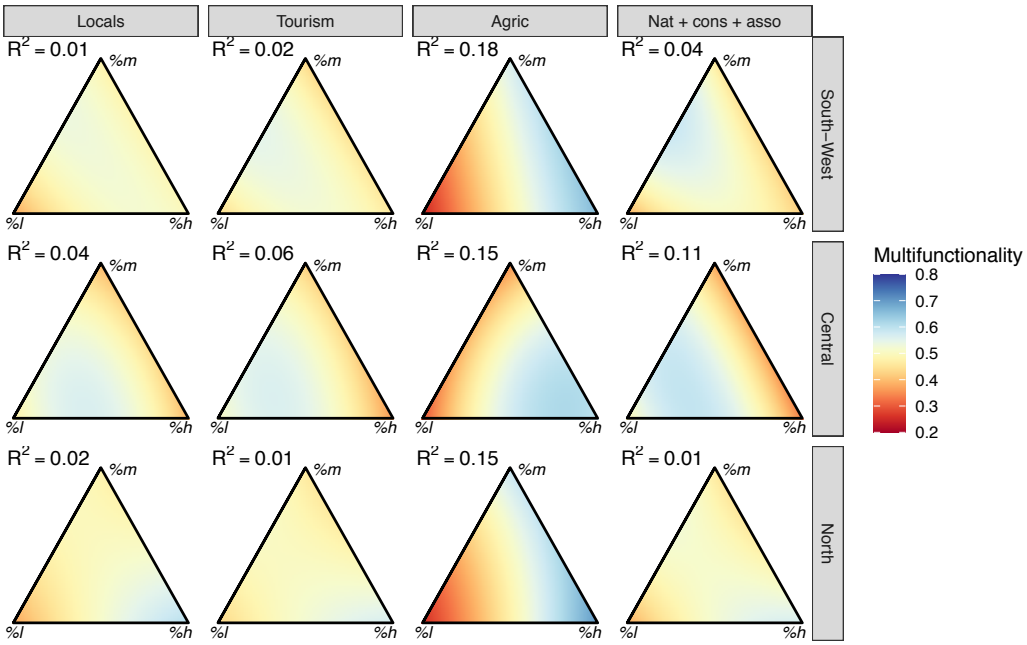


Figure A3. Estimated multifunctionality values for different stakeholder groups (column), depending on landscape composition (in proportions of low, medium and high-intensity sites) for the three regions (rows). Multifunctionality values are calculated by multiplying each service's threshold value (i.e., 1 if the service is above the median, 0 otherwise), weighted by its relative priority to the stakeholder group.

Note that the scale had been expanded (ranging from 0.3 to 0.8 instead of 0 to 1) compared to the other figures due to low multifunctionality ranges.

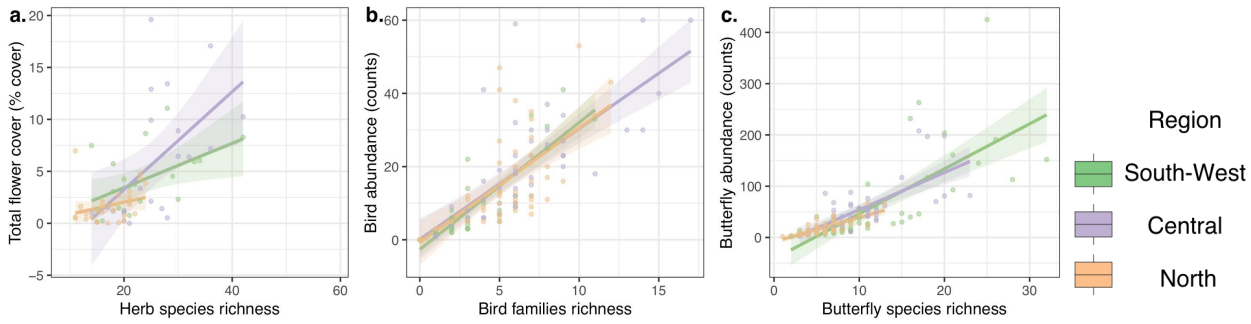


Appendix B. Supplementary methods

Measurement of site-level indicators and aggregation to landscape level

The 'biodiversity conservation' service at the site-level was based on total plant species richness as plant alpha-diversity. Plant species richness has been shown to be a good proxy for diversity at multiple trophic levels at these sites (correlation of 0.67 and 0.68 between the whole ecosystem multidiversity index (Allan et al 2014) and the richness of asterids and rosids respectively, for instance (Manning et al., 2015)). We chose not to include other taxa to prevent co-linearity with the other service measures (see below). This indicator was calculated at the landscape level as gamma diversity.

Figure B1 Correlations among potential service indicators



The fodder production service was calculated as total fodder protein production, a common agronomical indicator (Lee, 2018) that we calculated based on grassland aboveground biomass production and shoot protein content. Between mid-May and mid-June each year, aboveground biomass was harvested by clipping the vegetation 2 – 3 cm above ground in four randomly placed

quadrats of 0.5 m × 0.5 m in each subplot. The plant biomass was dried at 80°C for 48 hours, weighed and summed over the four quadrats. Biomass was then averaged between 2008 and 2012. In order to convert this one-time biomass measurements into estimates of annual field productivity, we used the information on the number of cuts and the number of livestock units in a site to estimate the total biomass production used by farming activities, i.e. converted into fodder or consumed directly by livestock. Details of this estimation process can be found in Simons & Weisser (2017). We then multiplied this productivity by plant shoot protein levels, a common indicator of forage quality (Lee, 2018). Total nitrogen concentrations in ground samples of aboveground biomass were determined using an elemental auto-analyser (NA1500, CarloErba, Milan, Italy), and multiplied by 6.25 to obtain protein content (Lee, 2018). The landscape-scale protein production was then calculated as the sum of the production of all individual sites in the landscape.

Climate change mitigation was quantified as soil organic carbon stocks in the top 10 cm, as deeper stocks are unlikely to be affected strongly by management actions. We sampled composite samples for each site, prepared by mixing 14 mineral surface soil samples per site. Soil samples were taken along two 18 m transects in each site using a split tube auger, 40 cm long and 5 cm wide (Eijkelkamp, Giesbeek, The Netherlands). Composite samples were weighed, homogenized, air-dried and sieved (<2 mm). We then measured total carbon (TC) contents by dry combustion in a CN analyser “Vario Max” (Elementar Analysensysteme GmbH, Hanau, Germany) on ground subsamples. We determined inorganic carbon (IC) contents after combustion of organic carbon in a muffle furnace (450°C for 16 h). We then calculated the soil organic carbon (SOC) content as the difference between TC and IC, and the SOC concentration based on the weight of the dry fine-earth (105°C) and its volume. SOC concentration was then multiplied by soil bulk density to obtain site-level carbon stock values. The landscape-scale soil carbon stock was calculated as the sum of the soil carbon stock of all individual sites.

The aesthetic value measure integrated flower cover, the number of bird families and abundance of butterflies. The choice of these indicators was led by studies showing people’s preference for bird richness over abundance (Cox & Gaston, 2015), including song diversity (Hedblom et al., 2014); and for flower-rich landscapes (Graves et al., 2017). Flowering units were counted between May and September 2009 for all flowering plant species (excluding grasses and sedges) on transects along the four edges of each site, in a total area of 600m^2. For abundant species, the number of flowering units was extrapolated to the whole site from a smaller area of 112 m^2. The total flower cover was calculated at the site scale as the sum of the individual flower cover of all plant species (see Binkenstein et al. (2013) for details). Butterfly and day-active moths (hereafter termed as Lepidoptera) abundance was measured in 2008 and averaged

among sites within each landscape (Börschig et al., 2013). We conducted surveys of Lepidoptera from early May to mid-August. We sampled Lepidoptera during 3 surveys, each along one fixed 300m transect of 30min in each site. Each transect was divided in 50m sections of 5min intervals and we recorded all Lepidoptera within a 5 m corridor. Birds were surveyed by standardized audio-visual point-counts and all birds exhibiting territorial displays (singing and calling) were recorded. We used fixed-radius point counts and recorded all males of each bird species during a five-minute interval per site. Each site was visited five times between 15 March and 15 June each year. The data was then aggregated by family. Landscape-scale bird richness was calculated as the total number of bird families found in the landscape (i.e. in at least one site and one year) between 2009 and 2012. These three indicators were then scaled and averaged to estimate landscape-scale aesthetic value. Richness and abundance were usually highly correlated for the three groups (correlation of 0.75 for butterflies, 0.72 for birds, and 0.52 for plants), and the number of families for birds was highly correlated (0.96) to species richness; thus, other selections of indicators would have led to similar results (Fig. B1).

The regional identity service was estimated from the richness of culturally important birds and the cover of culturally important plants. The list of culturally important bird species was obtained from an online survey among 57 German respondents who were asked to say whether each species observed on the sites (see above) was unimportant (0 points), slightly important (1 point) or very important (5 points) species, based on whether the species “were part of German cultural identity or that of the study regions, e.g. by regularly appearing in German folklore, iconography, or popular entertainment”. We then retained the 25% species with highest average scores (Table B1), and used the survey data described above to calculate species richness. The list of culturally important plant species also followed the same definition (“part of German cultural identity or that of the study regions, e.g. by regularly appearing in German folklore, iconography, or popular entertainment”) but was created by two German experts with botanical knowledge of the species in the study regions (Table B2). We then used the survey data described above to calculate total cover of the species they listed.

Finally, the foraging service was calculated as the total cover of edible plants. The list of edible plants was obtained by the same botanical experts as the culturally important species and is presented in Table B3.

Table B1. List of culturally important birds

Alauda arvensis	*Falco tinnunculus*	*Milvus milvus*
Buteo buteo	*Fringilla coelebs*	*Parus major*
Ciconia ciconia	*Garrulus glandarius*	*Pica pica*

Corvus corax	*Grus grus*	*Sturnus vulgaris*
Cuculus canorus	*Hirundo rustica*	*Troglodytes troglodytes*
Cyanistes caeruleus		*Turdus merula*

Table B2. List of culturally important plants

Acer sp.	*Colchicum autumnale*	*Prunus avium*	*Trifolium pratense*
Bellis perennis	*Equisetum arvense*	*Pseudolysimachion spicatum*	*Trifolium repens*
Betula pendula	*Fraxinus excelsior*	*Pulsatilla vulgaris*	*Urtica dioica*
Briza media	*Gentiana verna*	*Quercus robur*	*Veronica arvensis*
Campanula glomerata	*Gentianella germanica*	*Rosa canina aggr.*	*Veronica chamaedrys*
Campanula patula	*Geum urbanum*	*Taraxacum sp*	*Veronica filiformis*
Campanula persicifolia	*Juniperus communis*	*Thymus pulegioides*	*Veronica hederifolia*
Campanula rapunculoides	*Leucanthemum vulgare aggr.*	*Tilia sp.*	*Veronica officinalis*
Campanula rotundifolia	*Myosotis arvensis*	*Trifolium alpestre*	*Veronica persica*
Campanula sp.	*Myosotis discolor*	*Trifolium arvense*	*Veronica serpyllifolia*
Carlina acaulis	*Myosotis ramosissima*	*Trifolium campestre*	*Veronica teucrium*
Carpinus betulus	*Origanum vulgare*	*Trifolium dubium*	*Viola arvensis*
Centaurium erythraea	*Phragmites australis*	*Trifolium medium*	*Viola hirta*
Cichorium intybus	*Primula elatior veris agg*	*Trifolium montanum*	*Viola sp.*

Table B3. List of edible plants

Achillea millefolium aggr.	*Carum carvi*	*Lamium album*	*Rumex acetosella*
Aegopodium podagraria	*Chenopodium album*	*Lamium maculatum*	*Rumex thyrsiflorus*
Allium cf oleraceum	*Cichorium intybus*	*Leontodon hispidus*	*Salvia pratensis*
Angelica sylvestris	*Cirsium oleraceum*	*Lepidium campestre*	*Sanguisorba minor*
Anthriscus sylvestris	*Crataegus sp.*	*Origanum vulgare*	*Silene dioica*
Aphanes arvensis	*Daucus carota*	*Pastinaca sativa*	*Silene vulgaris*
Arctium lappa	*Elymus repens*	*Phyteuma orbiculare*	*Sinapis alba*
Arctium minus	*Erigeron acris*	*Phyteuma spicatum*	*Sinapis arvensis*
Arctium sp.	*Fallopia convolvulus*	*Prunus avium*	*Sisymbrium officinale*
Arctium tomentosum	*Filipendula ulmaria*	*Prunus sp.*	*Stellaria media*
Astragalus glycyphyllos	*Fragaria vesca*	*Prunus spinosa*	*Symphytum officinale*
Bellis perennis	*Fragaria viridis*	*Ranunculus ficaria*	*Taraxacum sp.*
Bistorta officinalis	*Glechoma hederacea*	*Rosa canina aggr.*	*Thlaspi arvense*
Brassica rapa	*Heracleum sphondylium*	*Rubus caesius*	*Thymus pulegioides*
Bunium bulbocastanum	*Hypericum perforatum*	*Rubus sp.*	*Tragopogon pratensis*
Cardamine pratensis	*Juniperus communis*	*Rumex acetosa*	

Topographical Wetness Index calculation

We calculated the Topographic Wetness Index (TWI) of each site, defined as $\ln(a/\tan B)$ where a is the specific catchment area (cumulative upslope area which drains through a Digital Elevation Model (DEM, http://www.bkg.bund.de) cell, divided by per unit contour length) and

tan*B* is the slope gradient in radians calculated over a local region surrounding the cell of interest (Gessler et al. 1995; Sørensen et al. 2006). TWI therefore combines both upslope contributing area (determining the amount of water received from upslope areas) and slope (determining the loss of water from the site to downslope areas). TWI was calculated from raster DEM data with a cell size of 25 m for all sites, using ArcGIS tools (flow direction and flow accumulation tools of the hydrology toolset and raster calculator). The TWI measure used was the average value for a 4 × 4 window centred on the site, i.e. 16 DEM cells corresponding to an area of 100 m ×100 m. See Le Provost et al. 2021 for further details.

Supplementary references

Binkenstein, J. et al. (2013). Increasing land-use intensity decreases floral colour diversity of plant communities in temperate grasslands. *Oecologia*, *173*(2), 461–471. https://doi.org/10.1007/s00442-013-2627-6

Börschig, C. et al. (2013). Traits of butterfly communities change from specialist to generalist characteristics with increasing land-use intensity. *Basic and Applied Ecology*, 547–554. https://doi.org/10.1016/j.baae.2013.09.002

Breunig, T. & Demuth, S. (1999). *Rote Liste der Farn—Und Samenpflanzen Baden—Württembergs* (Landesanstalt für Umweltschutz Baden-Württemberg. Naturschutz-Praxis, No. 2; Artenschutz, p. 247).

Cox, D.T.C. & Gaston, K. J. (2015). Likeability of Garden Birds: Importance of Species Knowledge & Richness in Connecting People to Nature. *PLOS ONE*, *10*(11), e0141505. https://doi.org/10.1371/journal.pone.0141505

Gessler P.E. et al. (2007). Soil-landscape modelling and spatial prediction of soil attributes. Int J Geogr Inf Syst 9:421–432. https://doi.org/10.1080/02693799508902047

Graves, R. A. et al. (2017). Species richness alone does not predict cultural ecosystem service value. *Proceedings of the National Academy of Sciences*, *114*(14), 3774–3779. https://doi.org/10.1073/pnas.1701370114

Korsch, H., & Westhus, W. (2001). Rote Liste der Farn- und Blütenpflanzen (Pteridophyta et Spermatophyta) Thüringens. *Naturschutzreport*, *18*, 273–296.

Ristow, M. et al. (2006). Rote Liste der etablierten Gefäßpflanzen Brandenburgs. *Naturschutz Und Landschaftspflege in Branden- Burg 15*, *15*(4), 12.

Sørensen R. et al. (2006) On the calculation of the topographic wetness index: evaluation of different methods based on field observations. Hydrol Earth Syst Sci 10:101–112. https://doi.org/10.5194/hess-10-101-2006

Herrero, M. et al. (2013) Biomass use, production, feed efficiencies, and greenhouse gas emissions from global livestock systems. Proceedings of the National Academy of Sciences 110, 20888–20893. https://doi.org/10.1073/pnas.1308149110

Waghorn, G. & Clark, D. (2004) Feeding value of pastures for ruminants. New Zealand Veterinary Journal 52, 320–331. https://doi.org/10.1080/00480169.2004.36448

Appendix C. Details on sensitivity analyses

We explored how multiple choices in the simulations and calculation of multifunctionality could affect the results of our analyses. In the first part of this appendix, we first list all the different sensitivities that we identified and detail how their implementation modified the methodology detailed in the main text. In the second part, we consider in turn each of the main results presented in the paper and describe the potential variations highlighted by the analyses. We describe only analyses departing from the main analysis by one parameter – further combinations can be explored in the online tool.

1. Methods

The sensitivities we identified were the following:

a. Analysis on non-environmentally corrected data.

The main analysis was conducted on service indicator values that were first corrected at the site level by environmental covariates (see Methods). In these sensitivity analyses we also conduct the analyses using raw data.

b. Classification into classes of low, medium and high land-use intensity.

In the main analyses, the sites are classified in the three intensity categories based on 33% quantiles of the intensity values (e.g. lowest, medium and highest thirds) within the regions. In these sensitivity analyses we remove the “intermediate”, potentially confounding sites by using sites within the lowest, medium and highest 20% quantiles (e.g. lowest, medium and highest fifths) of the intensities within each region. This reduced the number of sites available during the landscape simulations, and to avoid running the analysis on identical landscapes we adapted the analyses by lowering the number of landscape replicates for each combination (*).

c. Number of sites per landscape.

In the main analysis, each landscape was composed of 10 sites. We also run similar analyses with 7 and 13 sites per landscapes. This also affected the number of possible landscape combinations, and we consequently adapted the number of landscape replicates for each combination (*).

d. Calculation of landscape-scale service indicator values.

In the main analysis, landscape-scale service indicator values consisted either of gamma diversity (plant and bird diversity) or of the sum of the services provided by all sites in the landscape (other services). We also considered a situation in which high level of each service is expected in only part of the landscape, and calculated the landscape indicator value as the maximum of the indicator in all sites of the landscape.

e. Calculation of landscape-scale multifunctionality

In the main analysis, landscape scale multifunctionality was calculated as the number of services above the median or as 1 if all services were over 25%, 0 otherwise. We also calculated multifunctionality as the number of services above the 40th or 60th quantiles of the distribution of the values in all landscapes (measured as 95th quantile to avoid outliers).

We also measured multifunctionality as the average of the (scaled) values for all considered services.

(*) In the main analysis, there were 66 possible landscape compositions (from 0 to 10 sites of each intensity), and 15 random landscape replicates per composition, hence 990 different landscapes. Changing the number of sites per landscapes changed the number of possible combinations (36 possible combinations for 7 sites, 105 for 13 sites) and to keep the total number of simulated landscapes approximately similar, we used 1000/(number of combinations) landscape replicates per combination (i.e. 10 for 13 sites, 28 for 7 sites).

When using only the 20% lowest, medium and highest intensity sites decreased the size of the regional pool from which to build landscapes. In that case we used only 7 sites per landscape and the extreme compositions (e.g. 100% of one intensity) were represented by slightly less landscape replicates.

2. Results

a. Site-level correlations among indicators and variation of service indicator values with land-use intensity and.

When correcting for the environment, there were positive correlations between flower cover, butterfly abundance and plant species richness. These indicators were usually positively correlated with bird richness and negatively correlated with biomass production. Most of these relationships were similar when considering raw, non environmentally-corrected data (Table C1). As shown in Table C1, the variation of site-level service indicators with land-use intensity was not strongly affected by using raw data instead of environmentally corrected residuals, except that the response of some services (e.g. plant richness) to intensity was more linear (no apparent “threshold”) when it was corrected for the environment.

b. Landscape-level correlations among ecosystem services.

Correlations among landscape-scale services was mostly similar when considering services calculated as the maximum value of each service (instead of the sum) in the landscape (Fig. C3), when changing the number of sites per landscape (Fig. C4, Fig. C5) or using data not corrected for the environment (Fig. C6).

c. Multifunctionality response to landscape composition

The following figures present the multifunctionality response to landscape composition as affected by the parameters of the model.

Calculating the landscape-scale services as the maximum instead of the sum (Fig. C8) did not change the direction of the response. It led to weaker responses of single services (top row) and marginally stronger variability when including multiple services.

Decreasing (Fig. C9) or increasing (Fig. C10) the number of sites per landscape slightly weakened (resp. strengthened) the response of single services to landscape composition, possibly because including more sites made for a higher chance to select sites with very high or low values, especially in the extreme compositions (e.g. 100% low intensity, 100% high intensity). When considering multiple services, including more sites did not change the response of multifunctionality.

For multifunctionality counted as the proportion of services above a given threshold, changing the threshold from the median to the 40th or 60th quantile of the distribution slightly switched the multifunctionality to higher or, respectively, lower values (Fig. C11 and Fig. C12) but the general form of the response was not affected.

The same was observed in the “compromise” scenario, when multifunctionality was calculated as 1 when all services were above a threshold, and 0 otherwise. Changing the threshold from the 25th quantile of the distribution to the 15th or 35th quantiles (Fig. C13 and Fig. C14) made it easier (respectively, more difficult) to provide all services at the required level, resulting in overall higher (resp. lower) multifunctionality values but without affecting the form of the response.

Calculating multifunctionality as the average of all services gave similar results as the main thresholding approach (Fig. C15).

d. Effect of other drivers on the responsivity of multifunctionality to landscape composition

Table C2 presents the result of the model with the responsivity (i.e. range) of multifunctionality over all landscape compositions as a response, and the ratio of the effect of land-use intensity and other environmental variables; the number of services included; or the service response variance as explanatory variables.

The relative effect of land-use intensity compared to other environmental variables was calculated as the ratio between the slope coefficient between individual services and land-use intensity over the maximum slope coefficient between the service and all other environmental covariates. It was significantly positive regardless of the model parameters, except when landscape-scale services were calculated based on the maximum of the sites.

The multifunctionality range decreased with the number of services considered in 20 out of the 22 scenarios considered, supporting our main conclusions. However, it did not vary with the number of services when multifunctionality was calculated as 1 if all the services were above the

15th percentile. This is because for low thresholds such as these, multifunctionality is expected to be high everywhere if too few services are considered.

Finally, the responsiveness of multifunctionality significantly decreased with the service response variance in 18 of the 22 considered scenarios. There was no positive relationship.

Table C1 Site-level correlation among ecosystem services. a. corrected for the environment, and b. raw service values. * P < 0.05, ** P < 0.01, *** P < 0.001.

a. Indicators corrected for the environment	Biomass production	Plant nitrogen content	Flower cover (sqrt)	Butterfly abundance (sqrt)	Bird family richness	Plant richness	Soil C stock	Cover edible plants	*Juniperus* grasslands	Cover cultural plants
Plant nitrogen content	0.16									
Flower cover (sqrt)	0.02	-0.06								
Butterfly abundance (sqrt)	-0.48 ***	-0.16	0.17 *							
Bird family richness	-0.25 **	-0.07	-0.03	0.28 ***						
Plant richness	-0.5 ***	-0.07	0.41 ***	0.47 ***	0.35 ***					
Soil C stock	0.02	0.11	-0.02	-0.02	0.04	-0.09				
Cover edible plants	0.17 *	0.08	-0.04	-0.16	-0.01	-0.15	-0.12			
Juniperus grasslands	-0.28 *	-0.01	-0.04	0.31 **	0.29 *	0.24	0.07	-0.09		
Cover cultural plants	0.28 ***	0.01	0.16 *	-0.19 *	-0.03	-0.14	-0.17 *	0.38 ***	-0.1	
Richness cultural birds	-0.18 *	-0.05	0.02	0.23 **	0.79 ***	0.27 ***	0.05	-0.09	0.29 ***	0.04
b. Raw indicator values	Biomass production	Plant nitrogen content	Flower cover (sqrt)	Butterfly abundance (sqrt)	Bird family richness	Plant richness	Soil C stock	Cover edible plants	*Juniperus* grasslands	Cover cultural plants
Plant nitrogen content	0.19 *									
Flower cover (sqrt)	-0.11	-0.12								
Butterfly abundance (sqrt)	-0.53 ***	-0.2 *	0.36 ***							
Bird family richness	-0.29 ***	-0.11	-0.03	0.26 **						
Plant richness	-0.53 ***	-0.17 *	0.62 ***	0.6 ***	0.31 ***					
Soil C stock	0.14	0.15	-0.31 ***	-0.3 ***	-0.04	-0.43 ***				
Cover edible plants	0.21 **	0.07	-0.19 *	-0.25 **	-0.01	-0.27 **	0.12			
Juniperus grasslands	-0.24 **	-0.07	0.04	0.38 ***	0.22 **	0.24 **	-0.09	-0.07		
Cover cultural plants	0.24 **	0.02	0.25 **	-0.05	-0.08	0.01	-0.3 ***	0.22 **	-0.04	
Richness cultural birds	-0.24 **	-0.12	0.05	0.27 ***	0.82 ***	0.3 ***	-0.13	-0.1	0.26 **	0.03

Table C2 Variation in ecosystem services and their component indicators with the land-use intensity class in each region (mean ± sd). Indicators were not corrected for the environment (see Methods). Different letters indicate differences significant at 5%.

Ecosystem service or *service indicator*	Land-use intensity	South-West	Central	North
Livestock production	low	52.07 +/- 18.45 [a]	50.91 +/- 13.94 [a]	105.9 +/- 15.08 [a]
	medium	167.19 +/- 19.02 [b]	98.04 +/- 14.37 [a]	141.2 +/- 15.54 [ab]
	high	189.16 +/- 18.45 [b]	163.26 +/- 13.94 [b]	169.04 +/- 15.08 [b]
Biomass production	low	23.66 +/- 6.92 [a]	25.82 +/- 5.99 [a]	48 +/- 5.47 [a]
	medium	73.01 +/- 7.13 [b]	46.36 +/- 6.18 [a]	64.78 +/- 5.64 [ab]
	high	88.68 +/- 6.92 [b]	78.39 +/- 5.99 [b]	72.74 +/- 5.47 [b]
N content	low	2.12 +/- 0.08 [a]	1.95 +/- 0.07 [a]	2.24 +/- 0.13 [a]
	medium	2.18 +/- 0.08 [a]	2.1 +/- 0.07 [a]	2.18 +/- 0.14 [a]
	high	2.12 +/- 0.08 [a]	2.08 +/- 0.07 [a]	2.27 +/- 0.13 [a]
Aesthetic	low	0.49 +/- 0.03 [b]	0.51 +/- 0.03 [b]	0.26 +/- 0.02 [a]
	medium	0.29 +/- 0.03 [a]	0.43 +/- 0.04 [ab]	0.28 +/- 0.02 [a]
	high	0.3 +/- 0.03 [a]	0.32 +/- 0.03 [a]	0.25 +/- 0.02 [a]
Flower cover (sqrt)	low	2.56 +/- 0.21 [b]	2.53 +/- 0.26 [b]	1 +/- 0.16 [a]
	medium	1.68 +/- 0.21 [a]	2.13 +/- 0.26 [ab]	1.36 +/- 0.17 [a]
	high	2.28 +/- 0.21 [ab]	1.63 +/- 0.26 [a]	1.18 +/- 0.16 [a]
Butterfly abundance (sqrt)	low	10.32 +/- 0.81 [b]	8.47 +/- 0.81 [b]	4.32 +/- 0.43 [a]
	medium	5.49 +/- 0.84 [a]	6.84 +/- 0.84 [ab]	4.24 +/- 0.44 [a]
	high	4.12 +/- 0.81 [a]	5.14 +/- 0.81 [a]	4.2 +/- 0.43 [a]
Bird family richness	low	4.59 +/- 0.56 [b]	6.53 +/- 0.75 [a]	5.06 +/- 0.51 [a]
	medium	2.88 +/- 0.58 [ab]	6.25 +/- 0.77 [a]	4.56 +/- 0.52 [a]
	high	2.47 +/- 0.56 [a]	4.65 +/- 0.75 [a]	4.06 +/- 0.51 [a]
Conservation *(plant richness)*	low	54.88 +/- 2.01 [b]	59.88 +/- 2.87 [b]	27 +/- 1.27 [a]
	medium	36.5 +/- 2.07 [a]	44.69 +/- 2.96 [a]	31 +/- 1.31 [a]
	high	37.35 +/- 2.01 [a]	38 +/- 2.87 [a]	29.24 +/- 1.27 [a]
C stock	low	3.75 +/- 0.18 [a]	3.1 +/- 0.2 [a]	6.49 +/- 0.79 [a]
	medium	4.36 +/- 0.19 [ab]	3.57 +/- 0.21 [a]	4.79 +/- 0.82 [a]
	high	4.57 +/- 0.18 [b]	3.5 +/- 0.2 [a]	5.32 +/- 0.79 [a]
Foraging *(Cover edible plants)*	low	60.64 +/- 12.13 [a]	87.25 +/- 12.28 [a]	148.6 +/- 23.66 [a]
	medium	105.07 +/- 12.51 [b]	90.03 +/- 12.66 [a]	81.67 +/- 24.39 [a]
	high	122.56 +/- 12.13 [b]	95.84 +/- 12.28 [a]	130.83 +/- 23.66 [a]
Regional ID	low	0.33 +/- 0.04 [b]	0.23 +/- 0.03 [a]	0.19 +/- 0.03 [a]
	medium	0.19 +/- 0.04 [a]	0.26 +/- 0.03 [a]	0.22 +/- 0.03 [a]
	high	0.22 +/- 0.04 [ab]	0.28 +/- 0.03 [a]	0.18 +/- 0.03 [a]
Juniperus grasslands	low	0.24 +/- 0.07 [a]	0 +/- 0 [a]	0 +/- 0 [a]
	medium	0.06 +/- 0.07 [a]	0 +/- 0 [a]	0 +/- 0 [a]
	high	0 +/- 0.07 [a]	0 +/- 0 [a]	0 +/- 0 [a]
Cover cultural plants	low	114.42 +/- 15.43 [a]	77.58 +/- 14.72 [a]	69.17 +/- 20.6 [a]
	medium	93.27 +/- 15.91 [a]	98.6 +/- 15.17 [a]	113.05 +/- 21.23 [a]
	high	145.01 +/- 15.43 [a]	126.26 +/- 14.72 [a]	100.92 +/- 20.6 [a]
Richness cultural birds	low	2.29 +/- 0.32 [b]	2.65 +/- 0.41 [a]	2 +/- 0.26 [a]
	medium	1.12 +/- 0.33 [a]	2.69 +/- 0.42 [a]	1.75 +/- 0.27 [a]
	high	1.06 +/- 0.32 [a]	2.41 +/- 0.41 [a]	1.47 +/- 0.26 [a]

Figure C3 Trade-offs between landscape-scale ecosystem service measures. This figure differs from Fig. 3 as the landscape-scale services were calculated based on the maximum (instead of sum) of the services provided by all sites in the landscape.

The colour and size of the circles denote the strength of the correlation between pairs of variables, within each region. Crosses indicate no significant correlations at 5% (Holm correction for multiple testing).

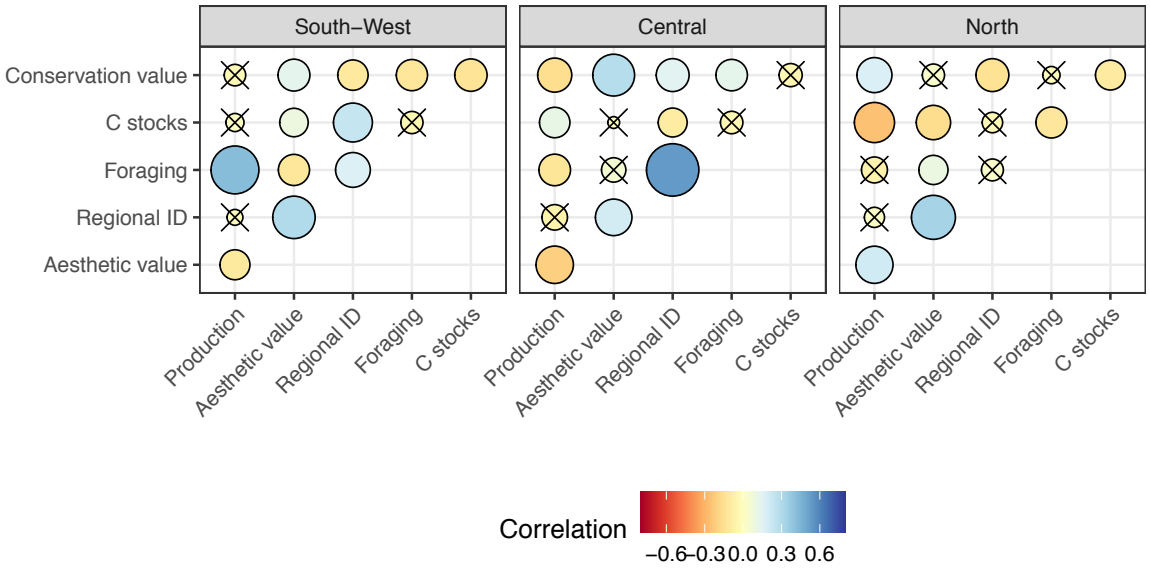


Figure C4 Trade-offs between landscape-scale ecosystem service measures. This figure differs from Fig. 3 as the landscapes included 7 (instead of 10) sites.

The colour and size of the circles denote the strength of the correlation between pairs of variables, within each region. Crosses indicate no significant correlations at 5% (Holm correction for multiple testing).

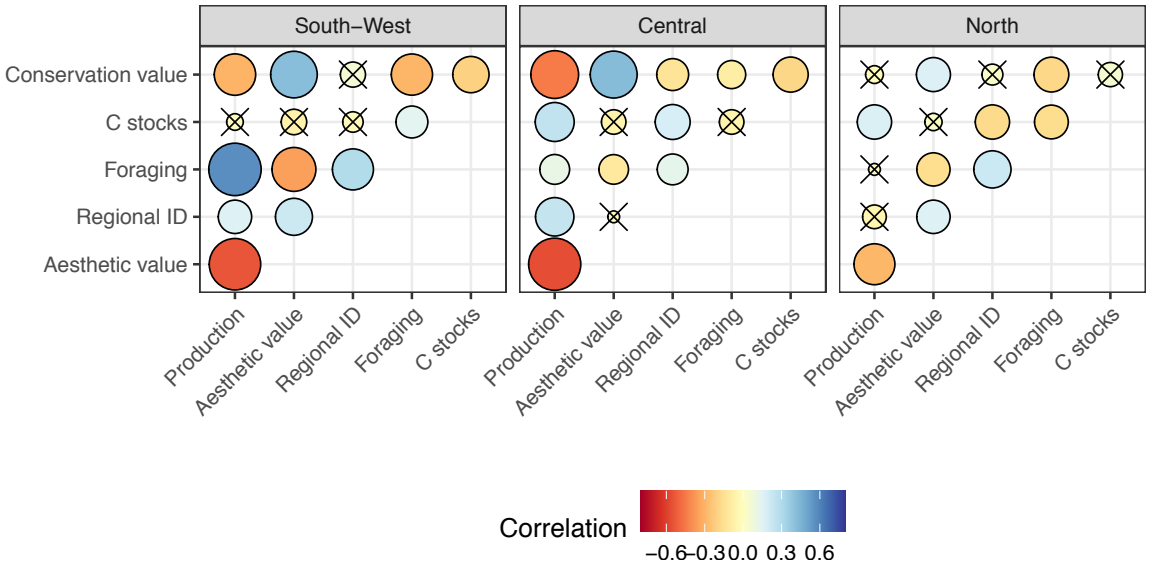


Figure C5 Trade-offs between landscape-scale ecosystem service measures. This figure differs from Fig. 3 as the landscapes included 13 (instead of 10) sites.

The colour and size of the circles denote the strength of the correlation between pairs of variables, within each region. Crosses indicate no significant correlations at 5% (Holm correction for multiple testing).

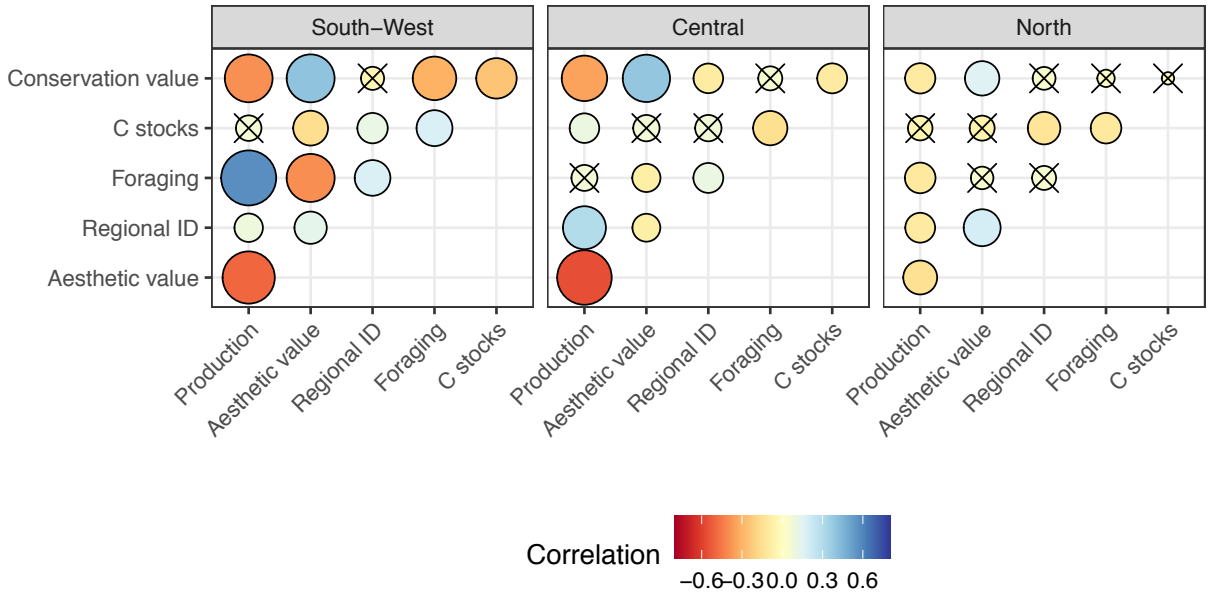


Figure C6 Trade-offs between landscape-scale ecosystem service measures. This figure differs from Fig. 3 as the ecosystem service indicators were not corrected for the environment before analysis.

The colour and size of the circles denote the strength of the correlation between pairs of variables, within each region. Crosses indicate no significant correlations at 5% (Holm correction for multiple testing).

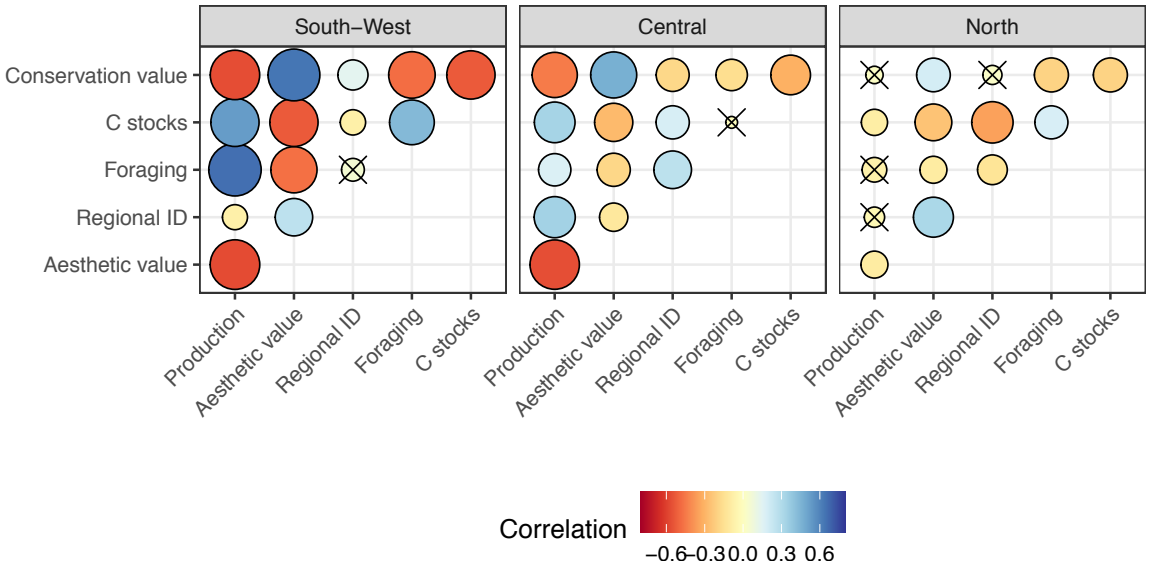


Figure C7 Estimated multifunctionality values depending on landscape composition (in proportions of low, medium and high-intensity sites). This figure shows all the service combinations for the 'compromise' approach, partly shown in Fig. 5.

For single ecosystem services (top row), the value presented corresponds to the probability of the given service to be above the median. For combinations of multiple services (middle and bottom rows), multifunctionality is the expected proportion of services above the median. Blue indicates higher multifunctionality values, orange lower.

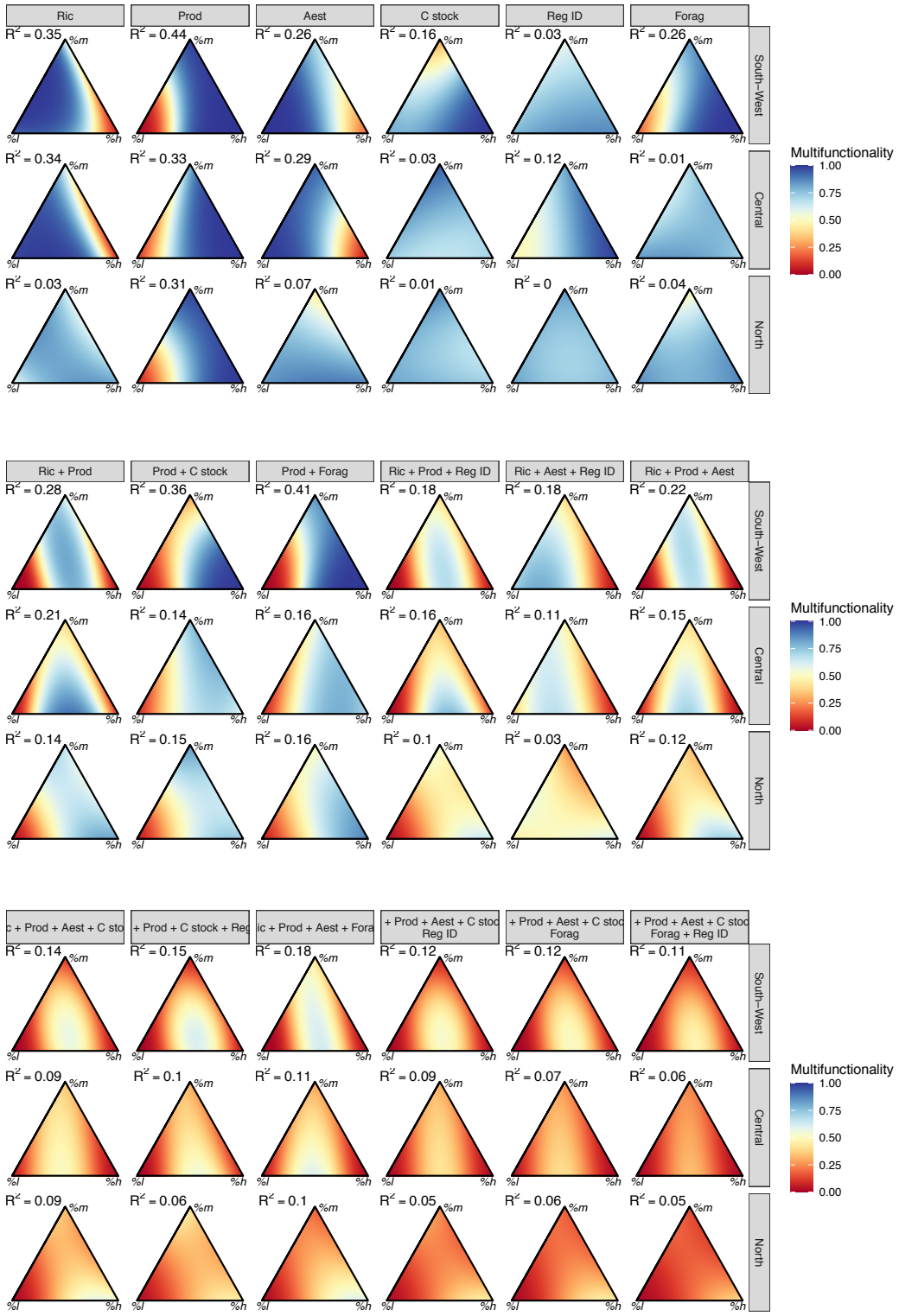


Figure C8 Estimated multifunctionality values depending on landscape composition (in proportions of low, medium and high-intensity sites). This figure differs from Fig. 4 in that landscape-scale ecosystem service values were calculated as the maximum, not the sum, of site-level ecosystem services.

For single ecosystem services (top row), the value presented corresponds to the probability of the given service to be above the median. For combinations of multiple services (middle and bottom rows), multifunctionality is the expected proportion of services above the median. Blue indicates higher multifunctionality values, orange lower.

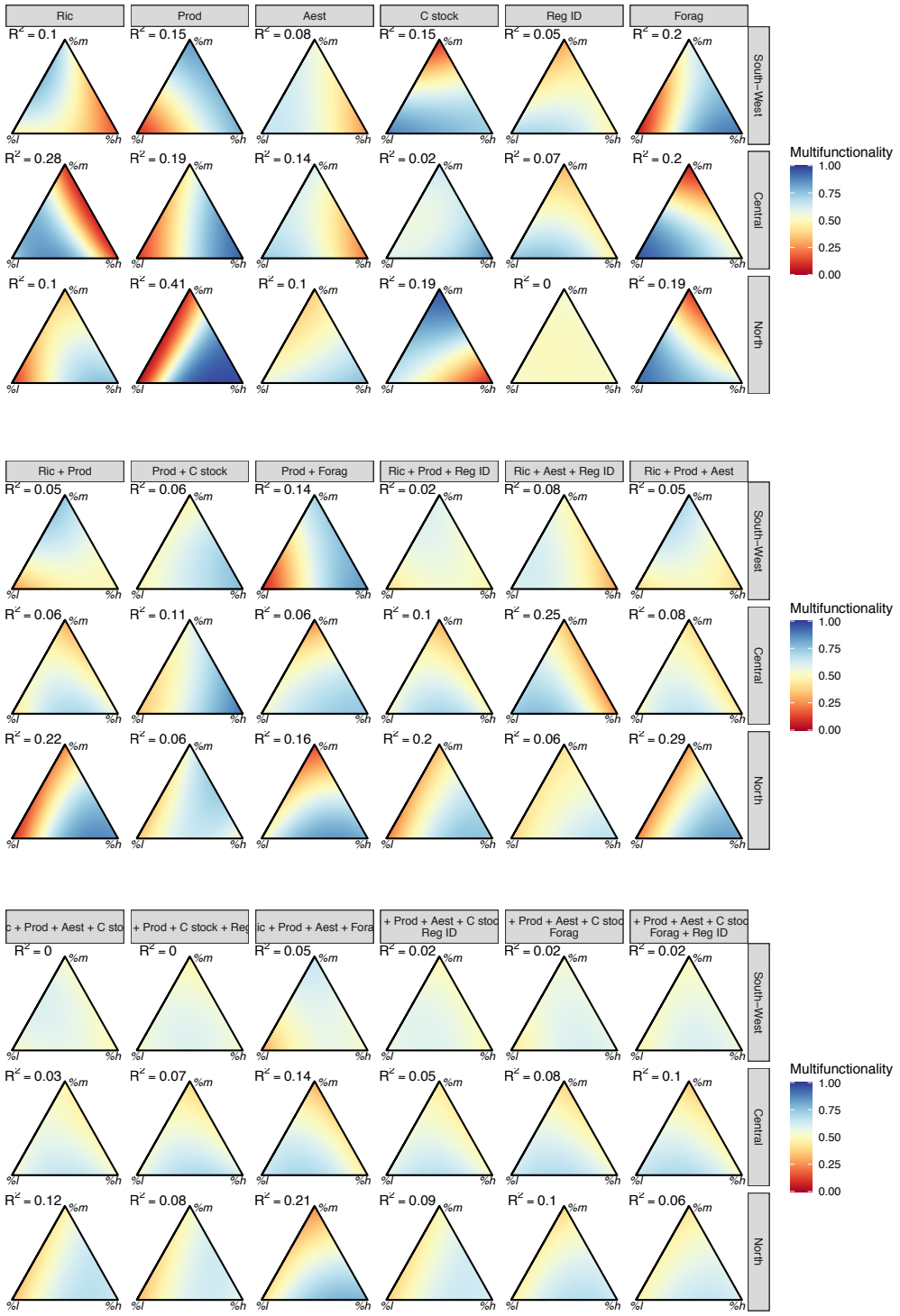


Figure C9 Estimated multifunctionality values depending on landscape composition (in proportions of low, medium and high-intensity sites). This figure differs from Fig. 4 in that landscapes were composed of 7 sites, instead of 10.

For single ecosystem services (top row), the value presented corresponds to the probability of the given service to be above the median. For combinations of multiple services (middle and bottom rows), multifunctionality is the expected proportion of services above the median. Blue indicates higher multifunctionality values, orange lower.

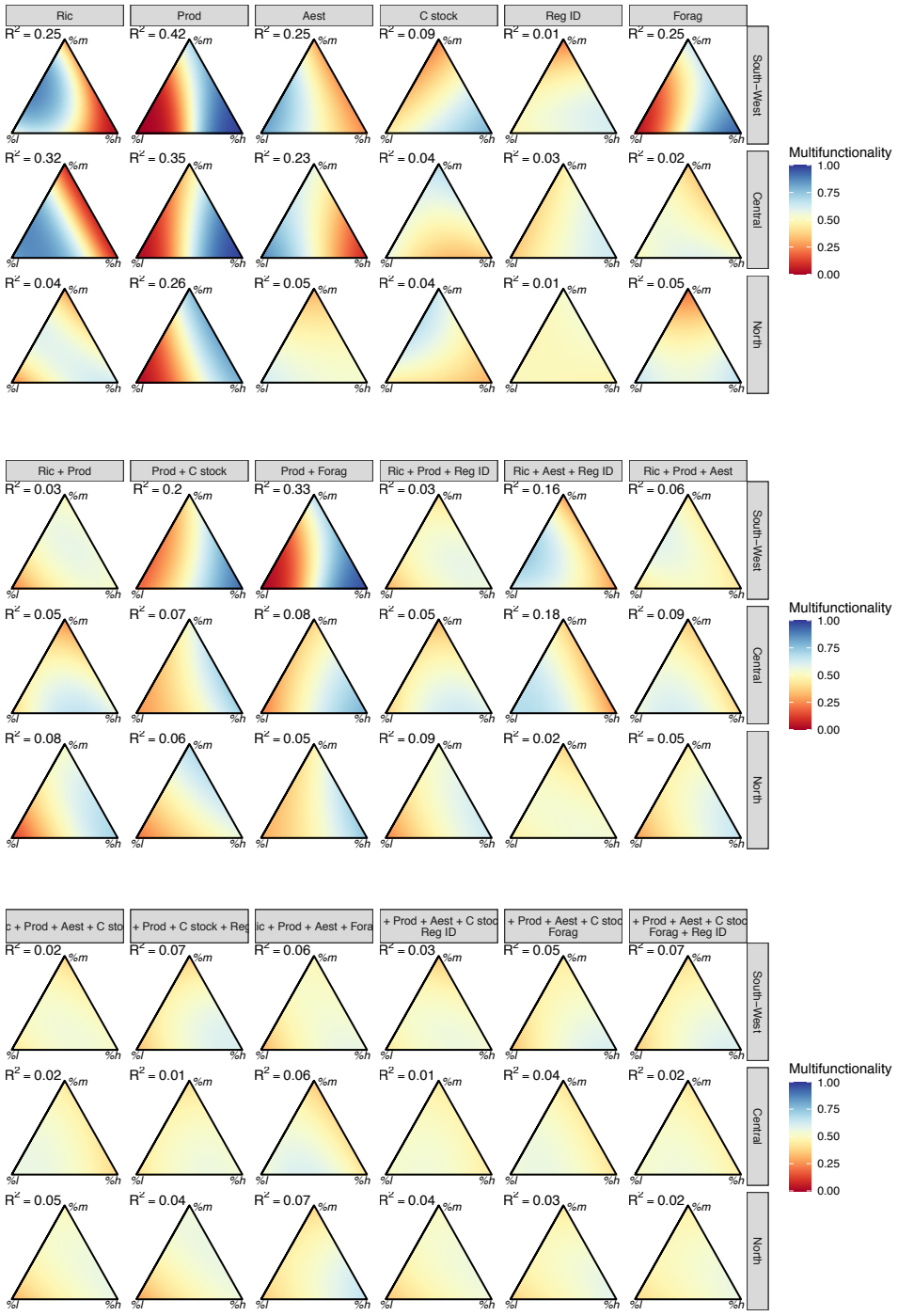


Figure C10 Estimated multifunctionality values depending on landscape composition (in proportions of low, medium and high-intensity sites). This figure differs from Fig. 4 in that landscapes were composed of 13 sites, instead of 10.

For single ecosystem services (top row), the value presented corresponds to the probability of the given service to be above the median. For combinations of multiple services (middle and bottom rows), multifunctionality is the expected proportion of services above the median. Blue indicates higher multifunctionality values, orange lower.

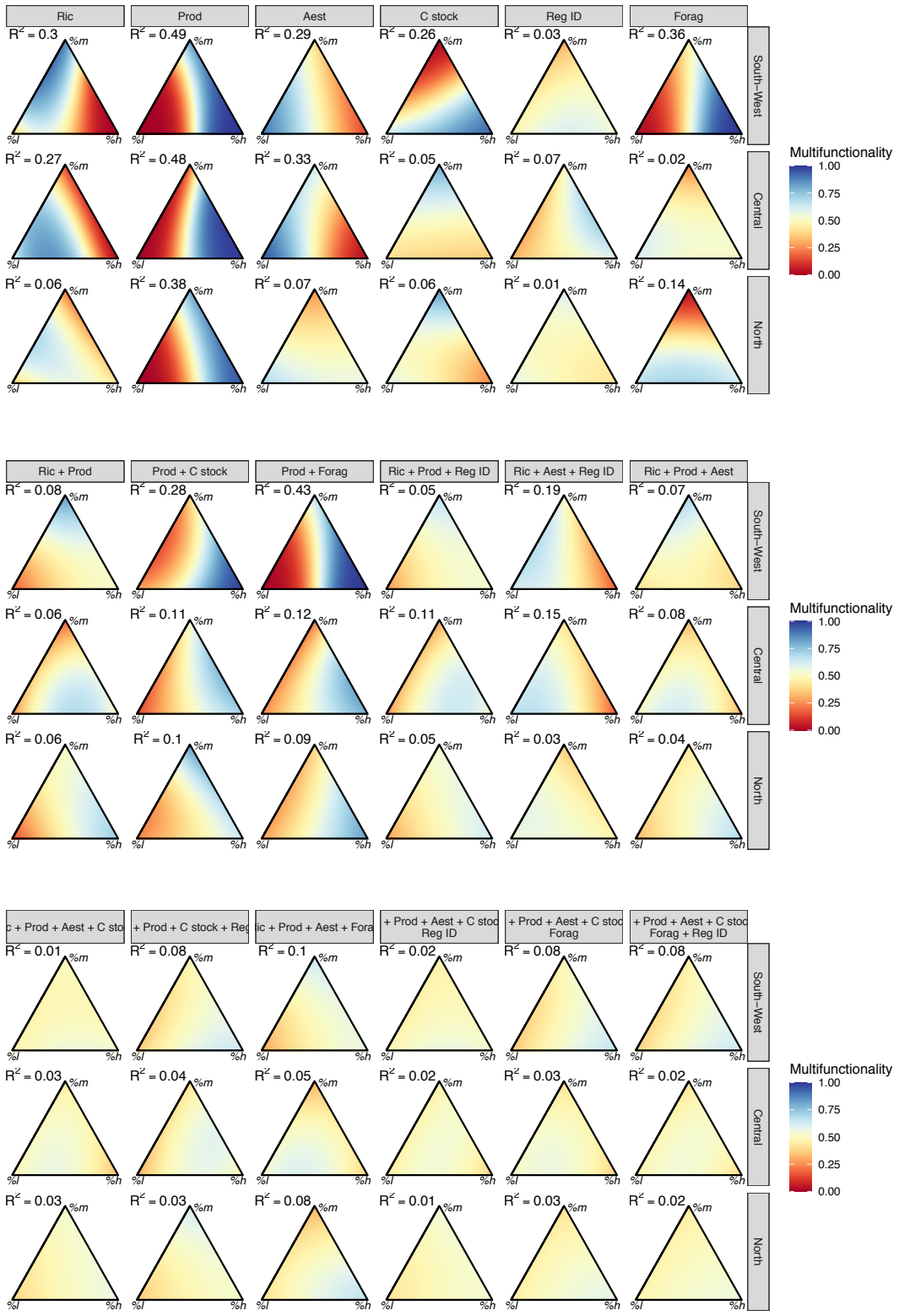


Figure C11 Estimated multifunctionality values depending on landscape composition (in proportions of low, medium and high-intensity sites). This figure differs from Fig. 4 in that the threshold was set to the 40th percentile instead of the median.

For single ecosystem services (top row), the value presented corresponds to the probability of the given service to be above the threshold. For combinations of multiple services (middle and bottom rows), multifunctionality is the expected proportion of services above the threshold. Blue indicates higher multifunctionality values, orange lower.

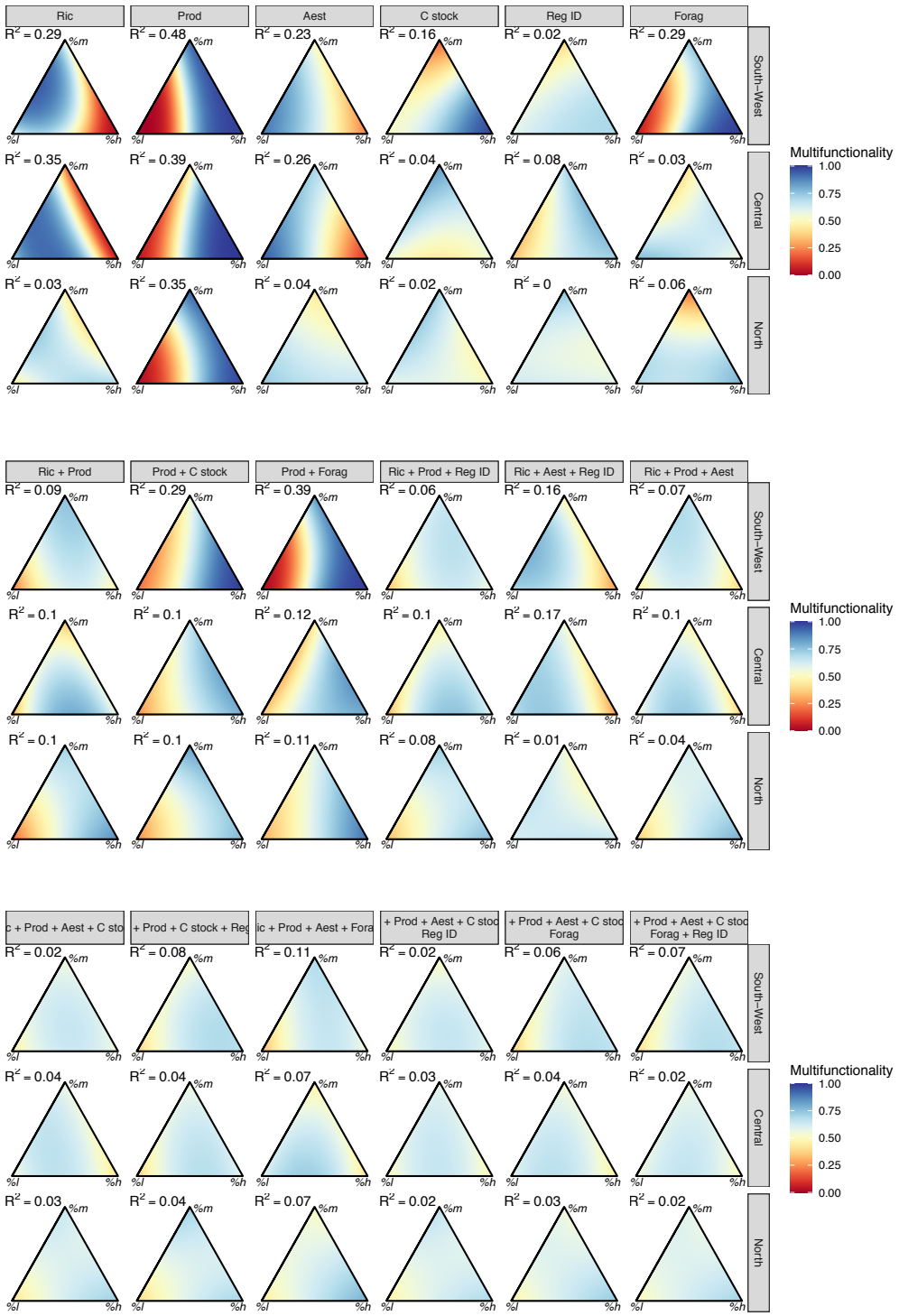


Figure C12 Estimated multifunctionality values depending on landscape composition (in proportions of low, medium and high-intensity sites). This figure differs from Fig. 4 in that the threshold was set to the 60th percentile instead of the median.

For single ecosystem services (top row), the value presented corresponds to the probability of the given service to be above the threshold. For combinations of multiple services (middle and bottom rows), multifunctionality is the expected proportion of services above the threshold. Blue indicates higher multifunctionality values, orange lower.

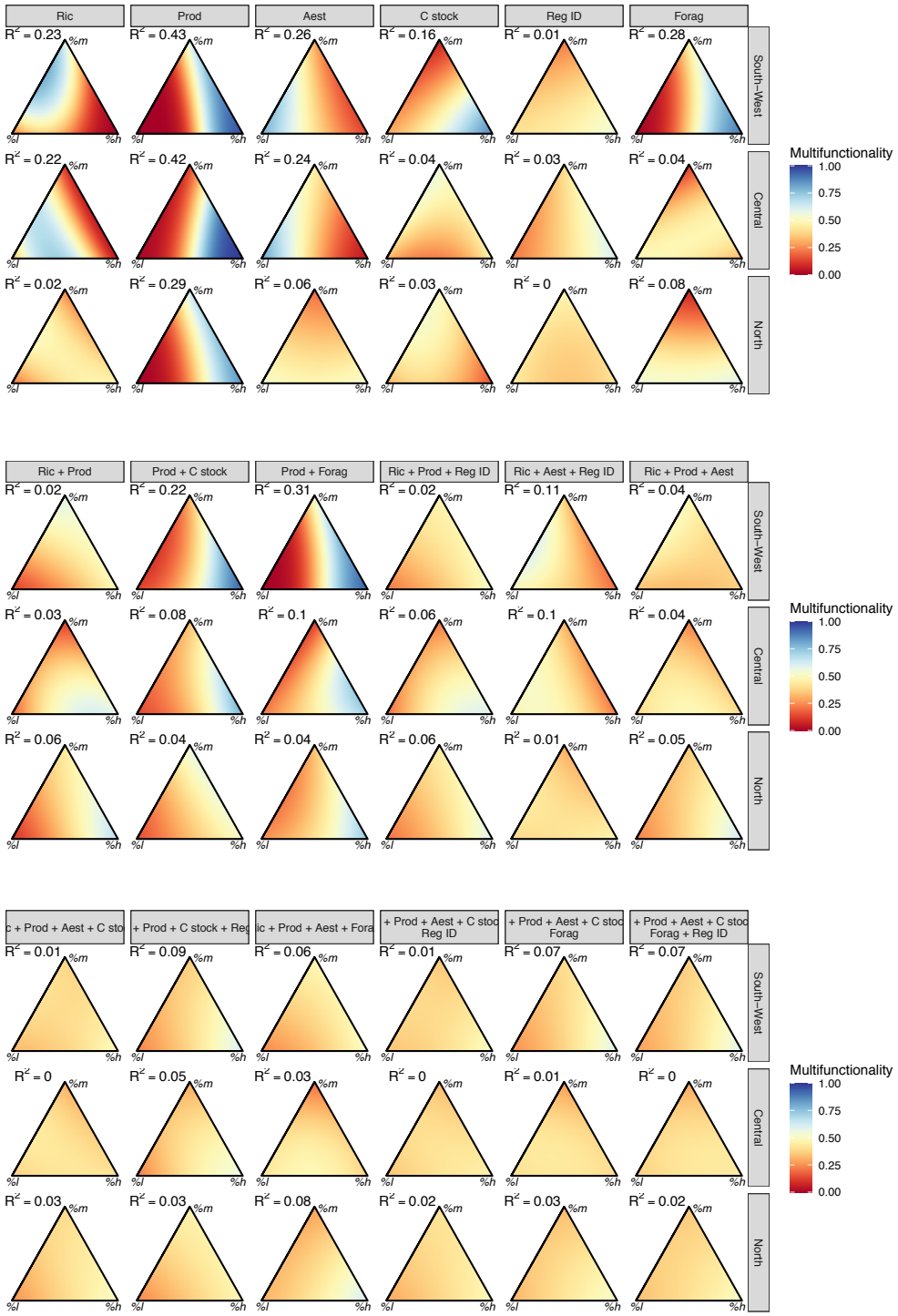


Figure C13 Estimated multifunctionality values depending on landscape composition (in proportions of low, medium and high-intensity sites). This figure differs from Fig. 5 in that the multifunctionality was calculated as 1 if all the services were above a 15th percentile threshold, and 0 otherwise (instead of a 25th percentile threshold).

The value presented corresponds to the probability that all given services are above the threshold. Blue indicates higher multifunctionality values, orange lower.

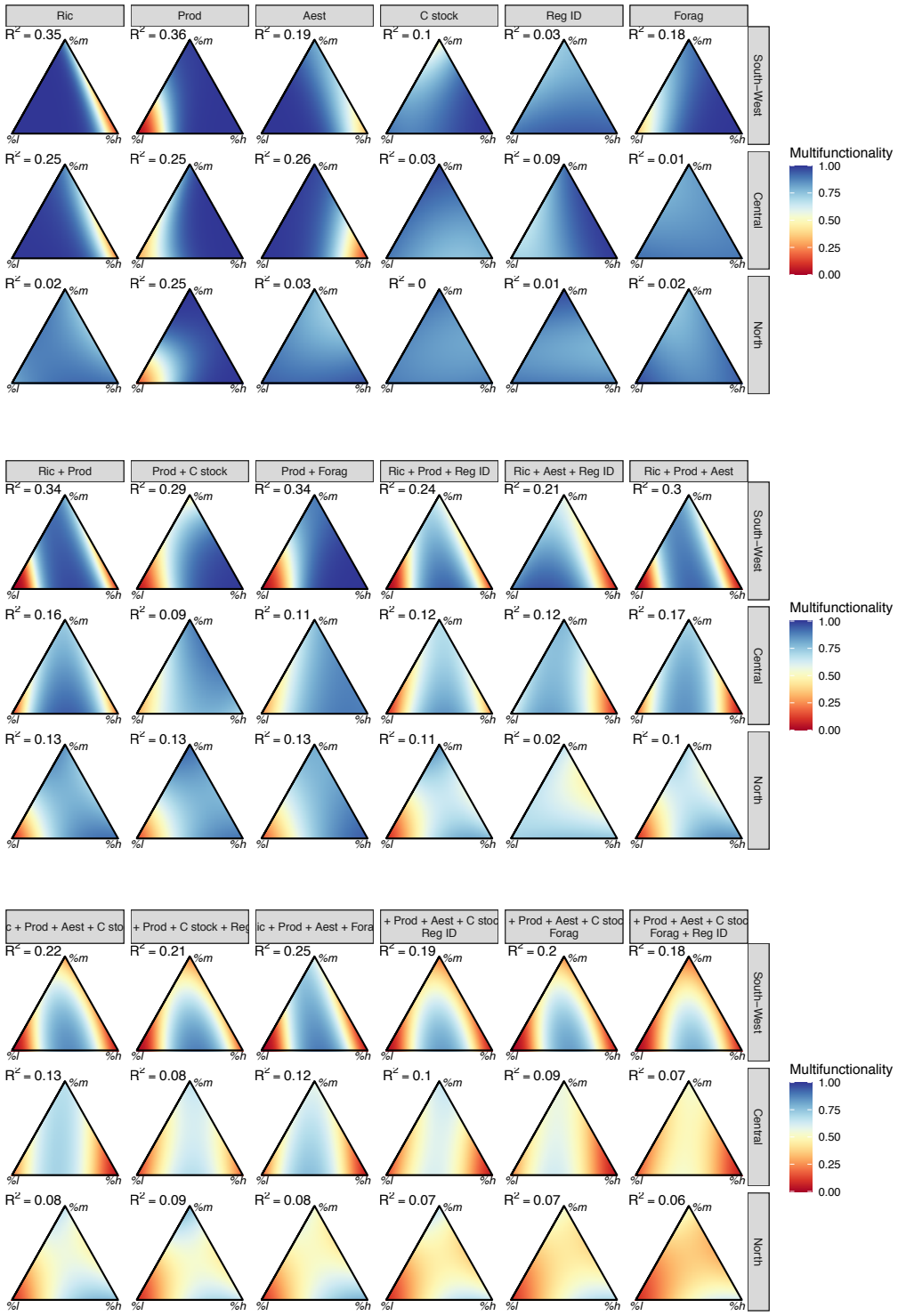


Figure C14 Estimated multifunctionality values depending on landscape composition (in proportions of low, medium and high-intensity sites). This figure differs from Fig. 5 in that the multifunctionality was calculated as 1 if all the services were above a 35th percentile threshold, and 0 otherwise (instead of a 25th percentile threshold).

The value presented corresponds to the probability that all given services are above the threshold. Blue indicates higher multifunctionality values, orange lower.

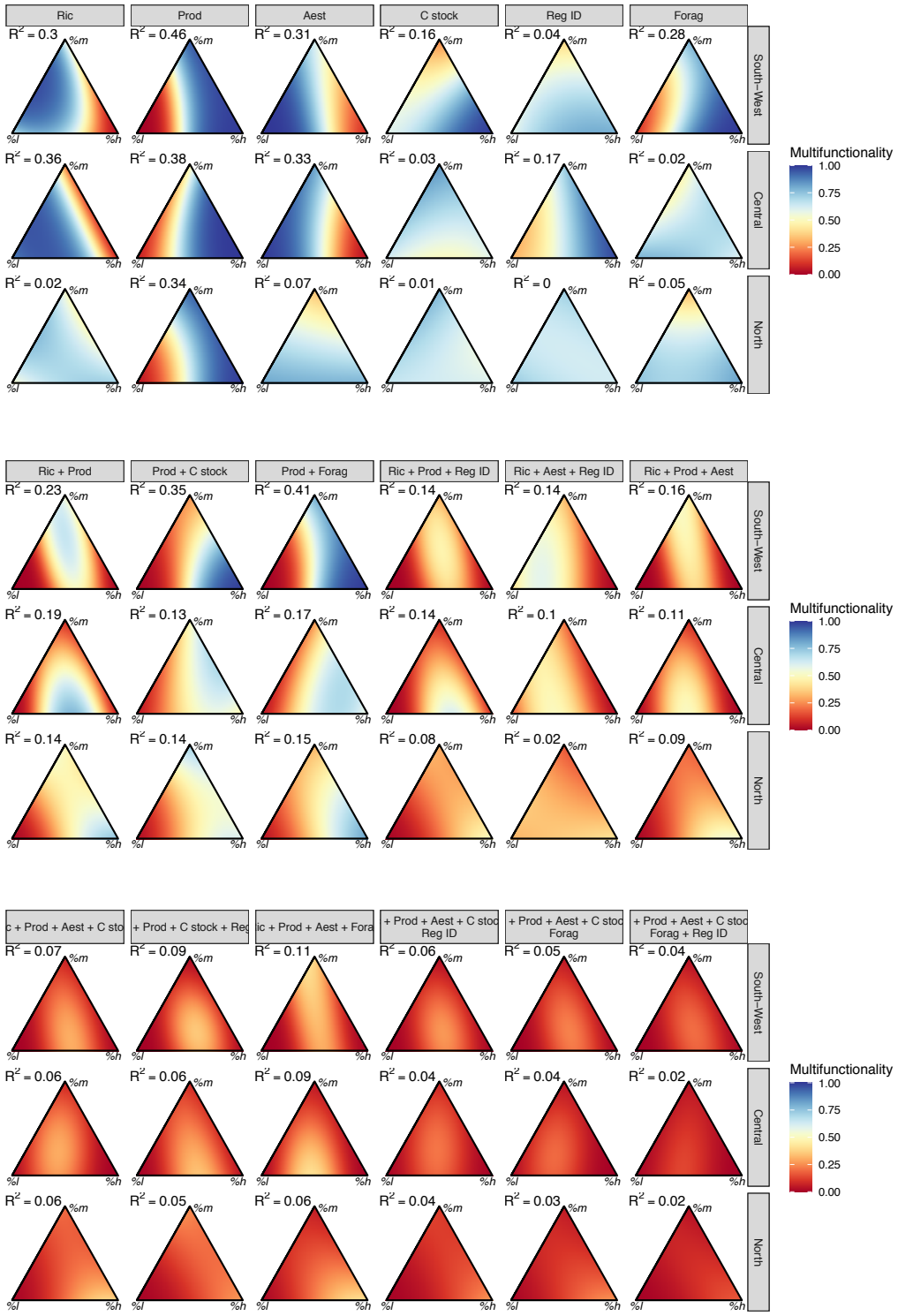


Figure C15 Estimated multifunctionality values depending on landscape composition (in proportions of low, medium and high-intensity sites). This figure differs from Fig. 4 in that landscapes multifunctionality was calculated as the average of the (scaled) values of all considered services, instead of the number of services above a threshold.

Blue indicates higher multifunctionality values, orange lower.

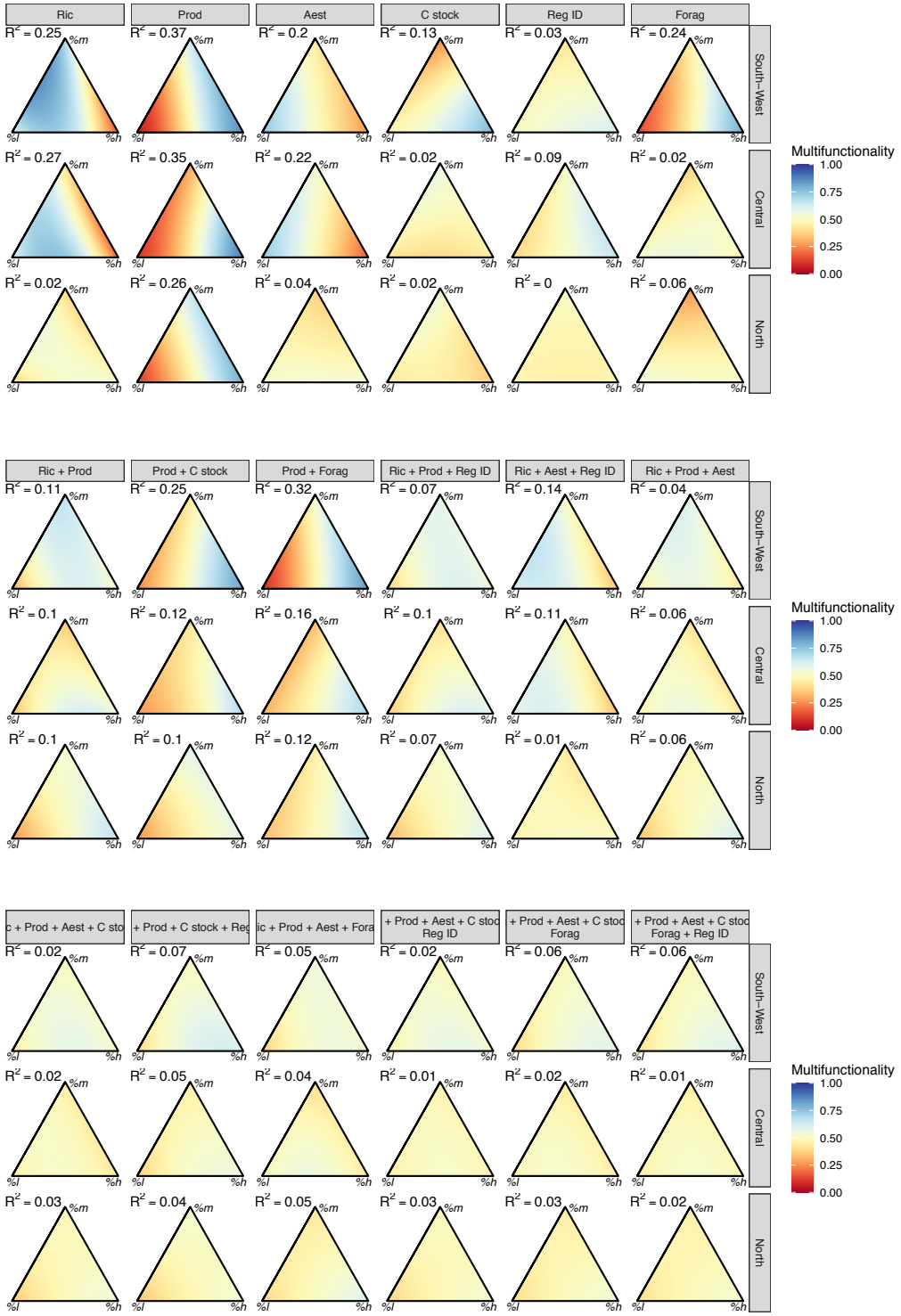


Figure C16 Estimated multifunctionality values depending on the mean (x-axis) and coefficient of variation (y-axis) of the land-use intensity in the landscape. The area outside the coloured represent combinations of intensity mean and variation that were not observed within the region.

Blue indicates higher multifunctionality values, orange lower.

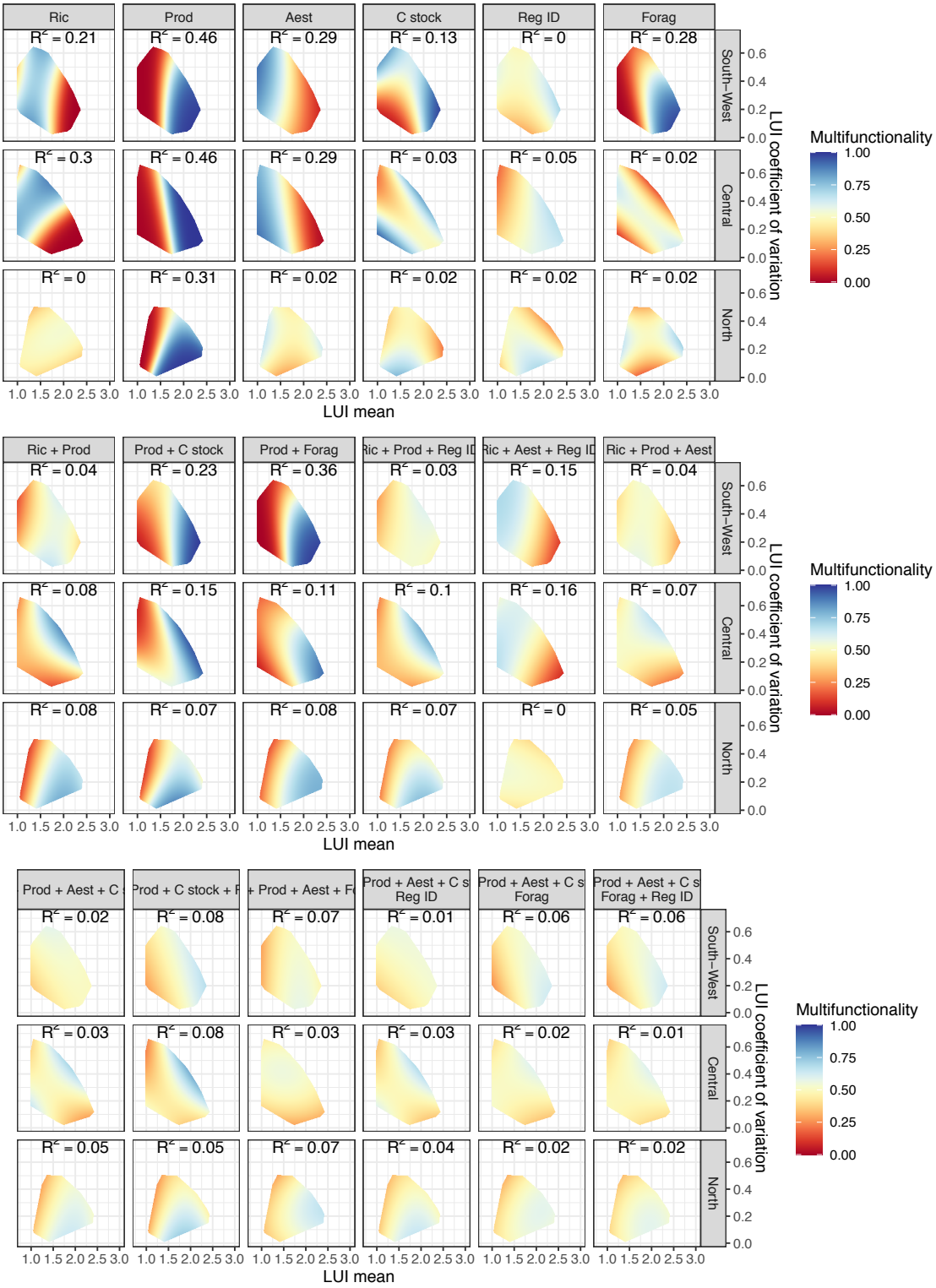


Table C2 Variability of multifunctionality responsiveness with the ratio of the effect of intensity and other environmental variables ("Ratio"), the service response variance ("SRV"), and the number of services included in the analysis, depending on model parameters. Model results for ratio and number of services are presented as slope of the relationship ± standard deviation (* P> 0.05, ** P < 0.01, *** P< 0.001). For conciseness, model results for SRV include slope for the 'high SRM' group (service variance mean > median); the p-value is the significance of the interaction SRV*SRM. Sensitivity analyses include whether sites were classified into low-, medium- and high intensity s using all intensity values (quantile:30%) or only the lowest, middle and highest fifths (quantile: 20%); the number of sites per landscape (7, 10 or 13); the calculation of services at the landscape scale (either sum or max of the values observed at the site-level). Are also considered the different methodology to measure multifunctionality (number of services above a given threshold, either based on quantiles of the distribution of percentage of the maximum observed vales; average of service values; or as 1 if all services are above a threshold, 0 otherwise). Note that for the ratio of land use effect to other environmental variables, only models for service values not corrected for the environment are relevant and presented here.

Env. correction	Intensity classes	N°of sites	Landscape-scale ES	Multifunctionality	Threshold	Ratio	SRV (high SRM)	N° of services
no	20% quantile	7	mean	methodThreshold0.5	0.5	0,22+/-0,05***	-0,85+/-0,1***	-0,09+/-0,01***
no	30% quantile	7	mean	methodThreshold0.5	0.5	0,25+/-0,05***	-0,79+/-0,09***	-0,1+/-0,01***
no	30% quantile	10	max	methodThreshold0.5	0.5	0,13+/-0,06	-0,49+/-0,08***	-0,06+/-0,01***
no	30% quantile	10	mean	methodAverage	NA	0,22+/-0,04***	-0,53+/-0,07***	-0,07+/-0,01***
no	30% quantile	10	mean	methodminimise	0.15	0,28+/-0,05***	0,11+/-0,06***	0,01+/-0,01
no	30% quantile	10	mean	methodminimise	0.25	0,23+/-0,05***	-0,08+/-0,08***	-0,06+/-0,01***
no	30% quantile	10	mean	methodminimise	0.35	0,22+/-0,05***	-0,31+/-0,11***	-0,13+/-0,01***
no	30% quantile	10	mean	methodThreshold0.4	0.4	0,24+/-0,06***	-0,64+/-0,09***	-0,09+/-0,01***
no	30% quantile	10	mean	methodThreshold0.5	0.5	0,24+/-0,05***	-0,81+/-0,1***	-0,1+/-0,01***
no	30% quantile	10	mean	methodThreshold0.6	0.6	0,23+/-0,05***	-0,7+/-0,09***	-0,1+/-0,01***
no	30% quantile	13	mean	methodThreshold0.5	0.5	0,22+/-0,06**	-0,77+/-0,11***	-0,11+/-0,01***
yes	20% quantile	7	mean	methodThreshold0.5	0.5	NA	-0,53+/-0,11*	-0,08+/-0,01***
yes	30% quantile	7	mean	methodThreshold0.5	0.5	NA	-0,4+/-0,09***	-0,07+/-0,01***
yes	30% quantile	10	max	methodThreshold0.5	0.5	NA	-0,15+/-0,09***	-0,08+/-0,01***
yes	30% quantile	10	mean	methodAverageNA	NA	NA	-0,23+/-0,07***	-0,05+/-0,01***
yes	30% quantile	10	mean	methodminimise	0.15	NA	0,14+/-0,09***	0,02+/-0,01
yes	30% quantile	10	mean	methodminimise	0.25	NA	-0,05+/-0,09***	-0,05+/-0,01***
yes	30% quantile	10	mean	methodminimise	0.35	NA	-0,22+/-0,12***	-0,1+/-0,01***
yes	30% quantile	10	mean	methodThreshold0.4	0.4	NA	-0,26+/-0,1***	-0,08+/-0,01***
yes	30% quantile	10	mean	methodThreshold0.5	0.5	NA	-0,29+/-0,1***	-0,08+/-0,01***
yes	30% quantile	10	mean	methodThreshold0.6	0.6	NA	-0,22+/-0,1***	-0,07+/-0,01***
yes	30% quantile	13	mean	methodThreshold0.5	0.5	NA	-0,29+/-0,11***	-0,1+/-0,01***