Supporting Information for "Routes to achieving sustainable intensification in simulated dairy farms - the importance of production efficiency and complimentary land uses"

Patrick. J. C. White, Mark A. Lee, Dave J. Roberts & Lorna J. Cole

SECTION A

Table S1. Estimates of dry matter (DM) yield in kg per ha, metabolisable energy (ME) per kg and resultant ME per ha for productive land-uses considered in this study.

<table>
<thead>
<tr>
<th>LAND-USE</th>
<th>Silage</th>
<th>Wheat</th>
<th>Barley</th>
<th>Oilseed rape</th>
<th>Fodder beet</th>
</tr>
</thead>
<tbody>
<tr>
<td>YIELD (kg DM ha⁻¹)</td>
<td>6,800</td>
<td>6,900</td>
<td>6,500</td>
<td>3,600</td>
<td>9,600</td>
</tr>
<tr>
<td>ME (MJ kg⁻¹ DM)</td>
<td>10.8</td>
<td>13.8</td>
<td>13.2</td>
<td>19.1</td>
<td>12.0</td>
</tr>
<tr>
<td>ME (MJ ha⁻¹) x 10³</td>
<td>73</td>
<td>95</td>
<td>85</td>
<td>69</td>
<td>115</td>
</tr>
</tbody>
</table>

Table S2. Parameter estimates (± standard error) for all confidence set models given in Table 1 in the main text.

<table>
<thead>
<tr>
<th>DIVERSITY INDEX</th>
<th>TAXA</th>
<th>MODEL</th>
<th>Intercept</th>
<th>1st variable</th>
<th>2nd variable</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha (Δα00)</td>
<td>Epigonal spiders</td>
<td>RICH*SPARE</td>
<td>-0.253 (±0.105)</td>
<td>-0.000 (±0.003)</td>
<td>-0.225 (±0.009)</td>
<td>0.010 (±0.000)</td>
</tr>
<tr>
<td></td>
<td>Vascular plants</td>
<td>RICH*SPARE</td>
<td>0.651 (±0.146)</td>
<td>0.003 (±0.002)</td>
<td>-0.083 (±0.014)</td>
<td>0.010 (±0.000)</td>
</tr>
<tr>
<td>Beta (Δβ00)</td>
<td>Epigonal spiders</td>
<td>DCA*EVEN</td>
<td>1.02 (±0.39)</td>
<td>-0.90 (±0.19)</td>
<td>-1.35 (±0.41)</td>
<td>1.29 (±0.23)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DCA*HETER</td>
<td>0.99 (±0.35)</td>
<td>-0.87 (±0.20)</td>
<td>-0.70 (±0.25)</td>
<td>0.69 (±0.14)</td>
</tr>
<tr>
<td></td>
<td>Plants</td>
<td>HETER</td>
<td>-2.28 (±0.38)</td>
<td>2.02 (±0.30)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EVEN</td>
<td>-2.46 (±0.51)</td>
<td>3.56 (±0.65)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DCA*HETER</td>
<td>-2.30 (±0.41)</td>
<td>0.00 (±0.11)</td>
<td>2.03 (±0.35)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DCA*EVEN</td>
<td>-2.89 (±0.53)</td>
<td>0.22 (±0.11)</td>
<td>3.42 (±0.52)</td>
<td>-</td>
</tr>
</tbody>
</table>
SECTION B – Methodology for production of alpha and beta diversity indices for each scenario

Alpha diversity index

Figure S1 shows the workflow for estimating the alpha diversity index (ADI) for each scenario, which fits into the overall workflow at Figure 2a in the main document. The accompanying R code for this analysis is provided in Section C of the Supporting Information. As stated in the main document, to generate an alpha diversity estimates we used species richness ($S$) models from Downie et al. (1999) and Wilson et al. (2003). These were developed to explore drivers of field-scale $S$ from a range of variables describing sites’ land-use and management. We used these models to provide estimates of $S$ for our six land-use components (i.e. the crops included in feed rations plus extensive grassland), by setting values of continuous variables (sward height, % bare ground, %bryophyte cover, stocking density, number of vegetation layers, and soil organic content) equal to their mean values recorded across each study site, for each land-use separately (Table S3). An annual cutting factor was set as the typical value for each land-use (0 cuts, 1 annual cut, or ≥ 2 annual cuts), while NVC categories were taken as the most common NVC category for that land-use in the study (Table S3). Altitude was selected as 100 m, approximating the mean (94 m) across land-uses used in the original study and an altitude appropriate to that of lowland farmland in the UK. Therefore models predicted plant and spider species richness respectively by varying land-use category, and modifying the values of other variables accordingly.
FIGURE S1. Workflow for producing an alpha diversity index (ADI) for each scenario. This fits into the overall workflow at Figure 2a in the main document.

Table S3. Fixed variables used for species richness models for vascular plants (Wilson et al. 2003) and epigeal spiders (Downie et al. 1999) according to land-use. Variables were set at the mean values for the land-use component from the original data set, except for altitude and annual cutting (marked "*") which were specified by us. Fixed values were taken as the mean values for that variable across all sites of a given land-use within the data (Yelloly 1999). Dummy variables indicating the presence/absence of a hedge (spider model) or boundary wall (plant model) were fixed at zero.

<table>
<thead>
<tr>
<th>Model</th>
<th>Variable</th>
<th>Silage</th>
<th>Barley</th>
<th>Wheat</th>
<th>Oilled</th>
<th>Fodder beet</th>
<th>Extensive grass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant</td>
<td>Sward height (cm)</td>
<td>18.9</td>
<td>72.2</td>
<td>74.5</td>
<td>99.8</td>
<td>56.5</td>
<td>12.0</td>
</tr>
<tr>
<td>Plant</td>
<td>% bare-ground</td>
<td>2.0</td>
<td>67.1</td>
<td>61.7</td>
<td>63.7</td>
<td>24.4</td>
<td>3.1</td>
</tr>
<tr>
<td>Plant</td>
<td>Stocking density (livestock units ha⁻¹)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Plant</td>
<td>% bryophyte cover</td>
<td>3.3</td>
<td>2.3</td>
<td>2.7</td>
<td>0.1</td>
<td>0</td>
<td>26.6</td>
</tr>
<tr>
<td>Spider</td>
<td>Altitude (m)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Spider</td>
<td>Number of vegetation layers</td>
<td>2.4</td>
<td>2.4</td>
<td>2.7</td>
<td>2.5</td>
<td>1.7</td>
<td>1.6</td>
</tr>
<tr>
<td>Spider</td>
<td>Soil organic content (% loss on ignition)</td>
<td>13.4</td>
<td>5.3</td>
<td>6.2</td>
<td>0.0</td>
<td>30.3</td>
<td>32.9</td>
</tr>
<tr>
<td>Spider</td>
<td>Annual cutting [FACTOR LEVELS]*</td>
<td>≥2 cuts</td>
<td>1 cut</td>
<td>1 cut</td>
<td>1 cut</td>
<td>1 cut</td>
<td>0 cuts</td>
</tr>
<tr>
<td>Spider</td>
<td>NYC category [FACTOR LEVELS]</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

We obtained the model parameter standard errors (SEs) (Table S4) and used them to account for uncertainty in the published models (Downie et al., 1999; Wilson et al., 2003). For individual land-uses, models were run $10^3$ times and in each run a random parameter value was drawn from a normal distribution based on that parameter's estimate and SE. Over all runs, we took median S, plus 2.5% and 97.5% quantiles. The original models were developed to estimate S for a single land-use, rather than a mixed land-use scenario. It was therefore not possible to directly obtain a whole-scenario cumulative (gamma) S estimate. Therefore the ADI was calculated as the mean S across land-uses for each scenario, weighted by the proportional area of each land-use.
Table S4. Parameter values and their standard errors for the vascular plant (Wilson et al. 2003) and epigeal spider (Downie et al. 1999) species richness models. Species richness estimates for a given taxa are formed by multiplying the variable values by the parameter estimates, summing together with the intercept, taking the exponential and subtracting 1. Parameters marked with an asterisk take dummy values (0 = no, 1 = yes).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Plant</th>
<th>Spider</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.068 ± 0.128</td>
<td>2.368 ± 0.154</td>
</tr>
<tr>
<td>Sward height (m)</td>
<td>-0.009 ± 0.002</td>
<td>-</td>
</tr>
<tr>
<td>Presence of wall*</td>
<td>0.061 ± 0.103</td>
<td>-</td>
</tr>
<tr>
<td>ln (% bare-ground + 1)</td>
<td>-0.150 ± 0.036</td>
<td>-</td>
</tr>
<tr>
<td>Stocking density (LU ha⁻¹)</td>
<td>-0.173 ± 0.041</td>
<td>-</td>
</tr>
<tr>
<td>Ln (% Bryophyte cover + 1)</td>
<td>0.370 ± 0.067</td>
<td>-</td>
</tr>
<tr>
<td>Presence of land use 3 or 9*</td>
<td>0.883 ± 0.167</td>
<td>-</td>
</tr>
<tr>
<td>[a] (Altitude (m) + 1)</td>
<td>-</td>
<td>0.102 ± 0.030</td>
</tr>
<tr>
<td>Number of vegetation layers</td>
<td>-</td>
<td>0.161 ± 0.041</td>
</tr>
<tr>
<td>V (Organic content [units])</td>
<td>-</td>
<td>-0.048 ± 0.018</td>
</tr>
<tr>
<td>Presence of hedge*</td>
<td>-</td>
<td>-0.274 ± 0.087</td>
</tr>
<tr>
<td>Presence of ≥ 2 cuts*</td>
<td>-</td>
<td>-0.306 ± 0.118</td>
</tr>
<tr>
<td>Presence of NVC 3 or 4*</td>
<td>-</td>
<td>0.364 ± 0.073</td>
</tr>
<tr>
<td>Presence of NVC 6, 8 or 10*</td>
<td>-</td>
<td>0.294 ± 0.103</td>
</tr>
<tr>
<td>Presence of NVC 7 or 9*</td>
<td>-</td>
<td>0.716 ± 0.105</td>
</tr>
<tr>
<td>Presence of land use 7*</td>
<td>-</td>
<td>0.489 ± 0.187</td>
</tr>
<tr>
<td>Presence of land-use 8*</td>
<td>-</td>
<td>0.264 ± 0.091</td>
</tr>
</tbody>
</table>

A beta-diversity index

Figures S2 and S3 show the workflow for the estimation of the beta diversity index (BDI) for each scenario, which fits into the overall workflow at Figure 2b in the main document. The accompanying R code for this analysis is provided in Section D of the Supporting Information. Our BDI was based on Euclidean distances on a detrended correspondence analysis (DCA) bi-plot, based on the data from Downie et al. (1999) and Wilson et al.
Note that while alpha-diversity models were for vascular plants only (Wilson et al. 2003), beta-diversity DCA data were for vascular plants and bryophytes combined (Murphy et al. 1998) although they were derived from the same study. DCA axes provide an estimate of species turnover, with two sites ≥4 units apart considered to have a very low probability of sharing species (Leps & Smilauer 2003). Euclidean distances between sites on a DCA bi-plot can therefore be considered an index of pairwise beta-diversity, with sites that are further apart having a greater complementarity in the species they support (Smol et al. 2005).

**FIGURE S2.** Workflow for producing a beta diversity index (BDI) for each scenario. This fits into the overall workflow at Figure 2b in the main document. The sub-workflow indicated by (a) is shown in Figure S3.

For each unique pair of different land-uses (15 for spiders, 10 for plants where wheat and barley were amalgamated) we calculated the Euclidean distance on the DCA bi-plot (hereafter 'DCA distance') between each possible combination of sites with a specific land use pairing, conditional on the pair having a different land-use (nm combinations, where n = number of sites of the first land-use and m = number of sites of the second). Mean and SE values gave a distribution of beta-diversity between each land-use pair. Additionally, for each land-use we calculated the DCA distance between each possible pair of sites within a
land-use \( (n/[n-2]!2 \) combinations, where \( n \) = number of sites). Mean and SE values gave a distribution of beta-diversity within each land-use. For plants, there were insufficient wheat sites with DCA data available (\( n = 3 \)) to complete robust analyses, so we combined wheat and barley into a single category ‘cereal’. Within the three wheat sites the mean DCA distance was 0.70, compared to 0.49 between wheat and barley, 0.38 within barley, and a mean of 1.93 across all crop combinations, justifying this amalgamation.

**FIGURE S3.** Workflow for producing distributions (mean ± SE) of DCA distances between each combination of land uses or within land use, across all sites in the original studies of Downie et al. (1999) and Wilson et al. (2003). This fits into the workflow at Figure S2a.

For each taxa and scenario a weighted beta-diversity distribution (BDD) was defined as the distribution of predicted DCA distances between any two random locations within that scenario, given their respective land-use types. To do so, we generated \( 10^4 \) pairs of locations, with each location within each pair being randomly assigned a land-use from that scenario, weighting this assignment by the relative area occupied by that land-use in that scenario. For each pairing, we then generated a beta-diversity value which was extracted.
by random sampling from a normal distribution parameterised with the mean and SE of
the DCA distance between sites for that land-use combination (either between-land-use or
within-land-use). The beta-diversity index (BDI) for a scenario was taken as the median
value across all pairs.

**SECTION C - Alpha diversity model code**

Data files referred to are available from the Edinburgh Napier University Repository. DOI:
10.17869/ENU.2018.1141073 (White et al., 2018).

**Species richness models**

The below model represents regression Model B from Appendix 2 in Downie et al. (1999).
This is saved as SPIDERBrnorm.modified.txt. Each model parameter is presented with its
estimate (taken from Downie et al. 1999) and standard error (obtained from the authors).

For stochasticity in the model, each time this is run one of the parameters is selected at
random. It then randomly selects from a normal distribution parameterised with that
parameter's estimate and standard error, while holding all other parameters at their fixed
estimates. The final output SspiB gives an estimate of epigeal spider species richness based
on the parameter estimates and the values of each variable that are presented to the model
either as fixed variables, or land-use based variables (explained below).

```r
> n.pars<-8
> r.par<-sample(1:n.pars,1)
> intercept<-ifelse(r.par==1,rnorm(1,2.368,0.1536),2.368)
> field<-ifelse(SP.FIELD==7,ifelse(r.par==2,rnorm(1,0.4885,0.1870),0.4885),ifelse(SP.FIELD==8,rnorm(1,0.2638,0.09125),0.2638))
> nvc<-ifelse(NVC==3|NVC==4,ifelse(r.par==3,rnorm(1,0.3638,0.07257),0.3638),ifelse(NVC==6|NVC==8|NVC==10,ifelse(r.par==3,rnorm(1,0.2939,0.1025),0.2939),ifelse(NVC==7|NVC==9,ifelse(r.par==3,rnorm(1,0.7162,0.1050),0.7162),0))
> layers<-ifelse(r.par==4,rnorm(1,0.1614,0.04127),0.1614)*LAYERS
> hedge<-ifelse(HEDGE==1,ifelse(r.par==5,rnorm(1,-0.2736,0.08651),-0.2736),0)
> organicp<-ifelse(r.par==6,rnorm(1,-0.04840,0.01775),-0.04840)*sqrt(ORGANICP+0.5)
> altitude.ln<-ifelse(r.par==7,rnorm(1,0.1022,0.03041),0.1022)*log(ALTITUDE+1)
> cutting<-ifelse(CUTTING==3,ifelse(r.par==8,rnorm(1,-0.3055,0.1179),-0.3055),0)
> SspiB<-exp(intercept+field+nvc+layers+hedge+organicp+altitude.ln+cutting)-1
```

Similarly, the below model represents the vascular plant regression model from Table 6 in
Wilson et al. (2003). This operates in the same stochastic manner as the spider model,
above. This is saved as VASCULARrnorm.modified.txt. The final output $S_{vas}$ gives an estimate of vascular plant species richness based on the parameter estimates and the values of each variable that are presented to the model either as fixed variables, or land-use based variables (explained below).

```r
> n.pars<-8
> r.par<sample(1:n.pars,1)
> intercept<ifelse(r.par==1,rnorm(1,3.068,0.1280),3.068)
> field<ifelse(PL.FIELD==3|PL.FIELD==9,ifelse(r.par==2,rnorm(1,0.8829,0.1669),0.8829),0)
> swardh<ifelse(r.par==3,rnorm(1,-0.008743,0.001835),-0.008743)*SWARDH
> wall<ifelse(WALL==1,ifelse(r.par==4,rnorm(1,0.06072,0.1026),0.06072),0)
> bareground.ln<ifelse(r.par==5,rnorm(1,-0.1497,0.03587),-0.1497)*log(BAREGROUND+1)
> cattle<ifelse(r.par==6,rnorm(1,-0.1728,0.04091),-0.1728)*CATTLE
> moss<ifelse(r.par==7,rnorm(1,-0.02741,0.005695),-0.02741)*MOSS
> moss.ln<ifelse(r.par==8,rnorm(1,0.3695,0.06679),0.3695)*log(MOSS+1)
> Svas<-exp(intercept+field+swardh+wall+bareground.ln+cattle+moss+moss.ln)-1
```

**Variables that are fixed for each land-use**

As an artifact of the way the model was developed, 'fixed' variables that are set against land-use type are introduced in two ways. The first way is in a dataset called SETVARIABLES.csv which is loaded to a dataframe SETVARS which contains a separate numerical land-use coding for vascular plants (PL.FIELD) and spiders (SP.FIELD) as a legacy of the coding systems used by Downie et al. (1999) and Wilson et al. (2003). The values of fixed variables were provided in Table S3.

These variables could have been hard-coded into VASCULARrnorm.modified.txt and SPIDERrnorm.modified.txt (above) but the current system allows functionality for modifying these in future applications. It also includes some habitat types or land-uses not used in the home-grown feeds system scenarios we used (e.g. GORS = gorse, HEAT = heather moor, SETA = setaside). Land-uses represented in our scenarios are BARL = barley, PAST = silage, OILS = oilseed rape, WHEA = wheat, EXTG = extensive grassland, ROOT = fodder beet.

```r
> SETVARS<read.csv(file=file.choose())
```

<table>
<thead>
<tr>
<th>TYPE</th>
<th>PL.FIELD</th>
<th>SP.FIELD</th>
<th>LAYERS</th>
<th>ORGANICP</th>
<th>AVAILP</th>
<th>LITTER</th>
<th>MOSS</th>
<th>SWARDH</th>
<th>BIOMASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BARL</td>
<td>1</td>
<td>2</td>
<td>2.38</td>
<td>5.31</td>
<td>65.00</td>
<td>7.4</td>
<td>2.3</td>
<td>72.2</td>
<td>30.3</td>
</tr>
<tr>
<td>PAST</td>
<td>2</td>
<td>5</td>
<td>2.41</td>
<td>13.37</td>
<td>33.26</td>
<td>3.9</td>
<td>3.3</td>
<td>18.9</td>
<td>5.3</td>
</tr>
<tr>
<td>OILS</td>
<td>3</td>
<td>2</td>
<td>2.50</td>
<td>8.00</td>
<td>202.50</td>
<td>19.5</td>
<td>0.1</td>
<td>99.8</td>
<td>12.2</td>
</tr>
<tr>
<td>SETA</td>
<td>4</td>
<td>4</td>
<td>4.00</td>
<td>6.35</td>
<td>73.50</td>
<td>32.2</td>
<td>1.9</td>
<td>5.7</td>
<td>3.5</td>
</tr>
<tr>
<td>WHEA</td>
<td>5</td>
<td>1</td>
<td>2.67</td>
<td>6.20</td>
<td>70.67</td>
<td>13.6</td>
<td>2.7</td>
<td>74.5</td>
<td>34.2</td>
</tr>
<tr>
<td>GORS</td>
<td>6</td>
<td>7</td>
<td>3.50</td>
<td>9.80</td>
<td>14.50</td>
<td>5.9</td>
<td>12.1</td>
<td>26.7</td>
<td>16.2</td>
</tr>
<tr>
<td>HEAT</td>
<td>7</td>
<td>8</td>
<td>1.57</td>
<td>33.03</td>
<td>12.11</td>
<td>17.2</td>
<td>30.5</td>
<td>16.6</td>
<td>7.0</td>
</tr>
</tbody>
</table>
The other means by which fixed variables are entered into the model is within the scenario specifications themselves, discussed in the following section.

### Loading scenarios

Scenarios are loaded in the file FULL_SCENARIOS.csv to an object SCENARIO.SET.

```r
> SCENARIO.SET <- read.csv(file = file.choose())
```

Each scenario is named in SCENARIO.SET$SCENARIO, comprising of various land uses (SCENARIO.SET$TYPE) of different area (SCENARIO.SET$AREA). The component numbering (SCENARIO.SET$COMPONENT) facilitates later loops in the model. Some of the variables that are required in the original models by Downie et al. (1999) and Wilson et al. (2003) are provided here, which are fixed for each land-use type. As discussed, this is a legacy of model creation, but it does allow increased functionality of the model as these can be varied for each scenario.

The first 18 rows are shown here. The INDIVIDUAL scenario is simply there to allow estimate of alpha diversity of each component, while BL (baseline), PC1 (1% spare land), PC2 (6% spare land) etc. are the first three scenarios for assessment.

```r
> head(SCENARIO.SET, 18)
```

<table>
<thead>
<tr>
<th>SCENARIO COMPONENT</th>
<th>AREA</th>
<th>TYPE</th>
<th>NVC</th>
<th>HEDGE</th>
<th>ALTITUDE</th>
<th>CUTTING</th>
<th>WALL</th>
<th>CATTLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INDIVIDUAL</td>
<td>10.0</td>
<td>PAST</td>
<td>2</td>
<td>0</td>
<td>100</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INDIVIDUAL</td>
<td>10.0</td>
<td>BARL</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INDIVIDUAL</td>
<td>10.0</td>
<td>WHEA</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INDIVIDUAL</td>
<td>10.0</td>
<td>OILS</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INDIVIDUAL</td>
<td>10.0</td>
<td>ROOT</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INDIVIDUAL</td>
<td>10.0</td>
<td>EXTG</td>
<td>7</td>
<td>0</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INDIVIDUAL</td>
<td>10.0</td>
<td>SETA</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BL</td>
<td>15.6</td>
<td>PAST</td>
<td>2</td>
<td>0</td>
<td>100</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BL</td>
<td>11.9</td>
<td>BARL</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BL</td>
<td>15.4</td>
<td>WHEA</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BL</td>
<td>50.0</td>
<td>OILS</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Specifications for the stochastic alpha-diversity model

When the fixed variables and scenario specifications have been loaded and called SETVARS and SCENARIO.SET respectively, the model is specified with two settings. Firstly, the scenario identity is set as an object scenarioID. For example, here we set the scenario identity as "PC5.EXTG" which means the fifth test-system (i.e. 15% spare land PC5; see Figure 1 in the main text) with extensive grassland (EXTG) added to that spare land component.

> scenarioID <- "PC5.EXTG"

Secondly, we specify the number of model runs we want. We used 1,000:

> RUNS <- 1000

Finally, we need to save a directory pathway to an object DIRECTORY from which models are sourced. This directory should contain the vascular plant model (VASCULARrnorm.modified.txt) and epigeal spider model (SPIDERBrnorm.modified.txt) described above. As an example:

> DIRECTORY <- "C:/My Documents/Diversity-production modelling"

Preliminary code before running the stochastic model

First we select the scenario details for the scenario for which we wish to generate an alpha diversity index, which we specified above. This contains the land-uses in this scenario and their area (ha), along with fixed values for some parameters.

> scenario <- SCENARIO.SET[SCENARIO.SET$SCENARIO == scenarioID,]
The stochastic alpha diversity model runs two loops. The inner loop estimates a stochastic species richness estimate for epigeal spiders and vascular plants for land-use $i$ within the scenario and populates them into column $h$ of the intermediate dataframes SPIDER.RUN and PLANT.RUN. The outer loop repeats this for $h$ runs, specified by RUNS and generates a weighted mean species richness estimate for each run, weighted by the relative area each land use occupies.

```r
> setwd(dir=DIRECTORY)
> # START OF LOOP FOR RUN $h$ WITHIN ALL RUNS
> for(h in 1:RUNS){
+ #print(c(h," / ",RUNS),quote=F) #THIS OPTION ALLOWS TRACKING OF PROGRESS OF RUNS - IT IS SUPPRESSED HERE
+ # FOR EACH RUN, A WEIGHTED MEAN SPECIES RICHNESS ESTIMATE IS MADE, WEIGHTED BY RELATIVE AREA OF EACH LAND-USE
+ p.area<-scenario$AREA/sum(scenario$AREA)
+ SPIDER.RUN[nrow(scenario)+1,h]<-sum(SPIDER.RUN[1:nrow(SPIDER.RUN)-1,h]*p.area)
+ VASC.RUN[nrow(scenario)+1,h]<-sum(VASC.RUN[1:nrow(VASC.RUN)-1,h]*p.area)
+ }
```
The output is a dataframe of $h$ columns, each of which is a run, and $i+1$ rows where $i$ is the number of land-uses in the scenario. The additional, final row is the weighted mean species richness for that land-use. The first 6 columns only are shown.

```r
> SPIDER.RUN[,1:6]
rep.0..nrow.scenario....1.    V2     V3     V4     V5     V6
3    22.23000 20.19000 20.93000 23.09000 22.23000 23.26000
6    25.45000 38.74000 33.00000 33.53000 33.00000 32.68000
```

The distribution of the weighted mean species richness across the 1,000 runs is in Figure S4 (note the precise distribution will vary due to stochasticity in the model).

```r
> hist(as.numeric(paste(SPIDER.RUN[nrow(scenario)+1,])),main=NULL,xlab="Weighted mean species richness")
```

![Histogram](image)

**Figure S4.** Histogram of weighted mean species richness estimations for 1,000 runs of the stochastic alpha-diversity model for the scenario 15% spare land + extensive grassland.
Note that as the distribution shows strong central tendency, the mean and median for the weighted mean species richness are very similar.

```r
> mean(as.numeric(paste(SPIDER.RUN[nrow(scenario)+1,])))
[1] 20.81855
> median(as.numeric(paste(SPIDER.RUN[nrow(scenario)+1,])))
[1] 20.73984
```

### Populating a final output dataframe

For ease of data extraction, a final output dataframe can be generated for the given scenario which is initially populated with the land-uses and their areas (ha) alone. This provides spaces for mean and 95% quantiles for each land-use and the weighted mean species richness.

```r
> LANDUSE<-c(as.character(scenario$TYPE),"ALL")
> AREA<-c(scenario$AREA,sum(scenario$AREA))
> SPIDER.MEAN<-rep(0,nrow(scenario)+1);SPIDER.95LO<-rep(0,nrow(scenario)+1);SPIDER.95HI<-rep(0,nrow(scenario)+1)
> VASC.MEAN<-rep(0,nrow(scenario)+1);VASC.95LO<-rep(0,nrow(scenario)+1);VASC.95HI<-rep(0,nrow(scenario)+1)
> output<-data.frame(LANDUSE,AREA,SPIDER.MEAN,SPIDER.95LO,SPIDER.95HI,VASC.MEAN,VASC.95LO,VASC.95HI)

> p.area<-output$AREA[1:nrow(output)-1]/output$AREA[nrow(output)]
> #output$SPIDER.MEAN[nrow(output)]<-sum(output$SPIDER.MEAN[1:nrow(output)-1]*p.area)
> #output$VASC.MEAN[nrow(output)]<-sum(output$VASC.MEAN[1:nrow(output)-1]*p.area)
> output

<table>
<thead>
<tr>
<th>LANDUSE</th>
<th>AREA</th>
<th>SPIDER.MEAN</th>
<th>SPIDER.95LO</th>
<th>SPIDER.95HI</th>
<th>VASC.MEAN</th>
<th>VASC.95LO</th>
<th>VASC.95HI</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAST</td>
<td>31.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BARL</td>
<td>9.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WHEA</td>
<td>9.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OILS</td>
<td>36.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ROOT</td>
<td>3.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EXTG</td>
<td>13.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ALL</td>
<td>103.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
```
Finally, the output dataframe is populated using the mean, lower 95% quantile and upper 95% quantile of species richness for each species for each land-use and for the weighted mean across all scenarios.

```r
> output$SPIDER.MEAN <- c(rowMeans(SPIDER.RUN))
> output$VASC.MEAN <- c(rowMeans(VASC.RUN))
> QLO <- function(x) {quantile(x, probs=0.025)}  # BESPOKE LOWER 95% QUANTILE FUNCTION
> QHI <- function(x) {quantile(x, probs=0.975)}  # BESPOKE UPPER 95% QUANTILE FUNCTION
> output$SPIDER.95LO <- c(apply(SPIDER.RUN, 1, QLO))
> output$VASC.95LO <- c(apply(VASC.RUN, 1, QLO))
> output$SPIDER.95HI <- c(apply(SPIDER.RUN, 1, QHI))
> output$VASC.95HI <- c(apply(VASC.RUN, 1, QHI))
```

By running this for each scenario, we were able to estimate changes in the alpha diversity index for each scenario relative to that of the baseline scenario.

**SECTION D - Beta-diversity model code**

Data files referred to are available from the Edinburgh Napier University Repository. DOI: 10.17869/ENU.2018.1141073 (White et al., 2018).

The beta-diversity model uses the same scenario dataset structure as the alpha-diversity model, since we tested the same set of scenarios, although many of the fixed variable fields in this are not used to assess beta diversity. Rather, we also imported a dataset of detrended correspondence analysis (DCA) axes 1 and 2 coordinates (described in main
text) for a series of sites throughout Scotland, each of which is assigned a different land-use, which match against those in our scenarios. These DCA values are saved in the file DCA VALUES.csv and imported to a dataframe object DCA.VALUES.

```
> DCA.VALUES<-read.csv(file.choose())

Unlike the alpha diversity model which estimates plant and spider alpha diversity indices simultaneously, here we need to run the model separately for spiders and plants, for each scenario to be tested. To do so, we select rows using the DCA.VALUES$TAXA variable, which has values "SPIDER" or "PLANT". The first 10 rows are shown.

```
> DCA.VALUES<-DCA.VALUES[DCA.VALUES$TAXA=="SPIDER",];head(DCA.VALUES,10)

<table>
<thead>
<tr>
<th>TAXA</th>
<th>SITE</th>
<th>AXIS1</th>
<th>AXIS2</th>
<th>CROP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPIDER</td>
<td>MM1</td>
<td>5.60</td>
<td>2.01</td>
<td>BARL</td>
</tr>
<tr>
<td>SPIDER</td>
<td>RH2</td>
<td>5.15</td>
<td>1.85</td>
<td>BARL</td>
</tr>
<tr>
<td>SPIDER</td>
<td>RH6</td>
<td>5.31</td>
<td>1.84</td>
<td>BARL</td>
</tr>
<tr>
<td>SPIDER</td>
<td>RH7</td>
<td>5.21</td>
<td>1.93</td>
<td>BARL</td>
</tr>
<tr>
<td>SPIDER</td>
<td>SAMS1</td>
<td>4.91</td>
<td>1.68</td>
<td>BARL</td>
</tr>
<tr>
<td>SPIDER</td>
<td>SAVF1</td>
<td>4.83</td>
<td>1.68</td>
<td>BARL</td>
</tr>
<tr>
<td>SPIDER</td>
<td>KPLS1</td>
<td>2.41</td>
<td>0.55</td>
<td>EXTG</td>
</tr>
<tr>
<td>SPIDER</td>
<td>KPLS3</td>
<td>3.73</td>
<td>1.73</td>
<td>EXTG</td>
</tr>
<tr>
<td>SPIDER</td>
<td>KPLS4</td>
<td>2.56</td>
<td>0.04</td>
<td>EXTG</td>
</tr>
<tr>
<td>SPIDER</td>
<td>KPLS5</td>
<td>3.63</td>
<td>1.70</td>
<td>EXTG</td>
</tr>
</tbody>
</table>
```

For plants, we would have to combine wheat and barley together to a 'cereal' land-use category (see above). For this, we would run the below code.

```
> levels(DCA.VALUES$CROP)
> levels(DCA.VALUES$CROP)<-c("CERE","EXTG","GORS","HEAT","OILS","PAST","ROOT","SETA","CERE")

We create a Euclidean distance matrix of all rows in `DCA.VALUES' against all others. The main diagonal demonstrates zero distances, as expected, since a sample site cannot differ from itself, and the matrix is symmetrical about this main diagonal. Only the first six rows and columns are shown.

```
> distd<-as.matrix(dist(DCA.VALUES[,3:4]));distd[1:6,1:6]

```

```
1 0.0000000 0.4775982 0.3361547 0.3981206 0.7648529 0.8377350
2 0.4775982 0.0000000 0.1603122 0.1000000 0.2941088 0.3623534
3 0.3361547 0.1603122 0.0000000 0.1345362 0.4308132 0.5059644
4 0.3981206 0.1000000 0.1345362 0.0000000 0.3905125 0.4548626
5 0.7648529 0.2941088 0.4308132 0.3905125 0.0000000 0.0800000
6 0.8377350 0.3623534 0.5059644 0.4548626 0.0800000 0.0000000
```

Then all possible combinations of land-uses are generated. Only the first 10 lines are shown.

```
> types<-as.factor(levels(DCA.VALUES$CROP))
> combos<-expand.grid(types,types);head(combos,10)
Var1 Var2
```
A loop is required to estimate the mean, standard error and sample size (n) of DCA
distances for all pairs of sample plots for each pair-wise land-use combination. Initially, and
output dataframe comparisons is generated, which will be populated by the loop. Only the
first 10 lines are shown.

```r
> means<-rep(0,nrow(combos));se<-rep(0,nrow(combos));count<-rep(0,nrow(combos))
> comparisons<-cbind(combos,means,se,count);head(comparisons,10)
      Var1 Var2 means  se count
   1 BARL BARL   0 0.00   0
   2 EXTG BARL   0 0.00   0
   3 GORS BARL   0 0.00   0
   4 HEAT BARL   0 0.00   0
   5 OILS BARL   0 0.00   0
   6 PAST BARL   0 0.00   0
   7 ROOT BARL   0 0.00   0
   8 SETA BARL   0 0.00   0
   9 WHEA BARL   0 0.00   0
  10 BARL EXTG   0 0.00   0
```

The loop is below, with the first 10 lines of the populated comparisons dataframe now
shown.

```r
> for (i in 1:nrow(combos)){
+   (rows<-which(DCA.VALUES$CROP==combos[i,1]))
+   (cols<-which(DCA.VALUES$CROP==combos[i,2]))
+   ifelse(sum(rows-cols)!=0,
+   subdists<-distd[rows,cols],
+   subdists<-distd[rows,cols][lower.tri(distd[rows,cols])])
+   comparisons$means[i]<-round(mean(subdists),2)
+   comparisons$se[i]<-round(sd(subdists)/sqrt(length(subdists)),2)
+   comparisons$count[i]<-length(subdists)}
> head(comparisons,10)
      Var1 Var2 means  se count
   1 BARL BARL 0.38 0.06   15
   2 EXTG BARL 2.73 0.14   48
   3 GORS BARL 3.55 0.12   6
```
As discussed, we use the same scenario set (SCENARIO_SET), which was imported for the alpha diversity analyses, earlier. As with the DCA.values, for plant diversity we combine wheat and barley to a single 'cereal' category, using the below code. This is not used for spider diversity analyses which we are demonstrating here.

```r
> levels(SCENARIO_SET$TYPE)
[1] "BARL" "EXTG" "OILS" "PAST" "ROOT" "SETA" "WHEA"

> levels(SCENARIO_SET$TYPE)<-c("CERE","EXTG","OILS","PAST","ROOT","SETA","CERE")
```

As with the alpha-diversity model, the scenario identity is set as an object scenarioID. For example, here, again, we set the scenario identity as "PC5.EXTG" which means the fifth test-system (i.e. 15% spare land; see Figure 1 in the main text) with extensive grassland (EXTG) added to that spare land component, and then we filter SCENARIO_SET to a new object scenario. Note this is the version for spiders since here wheat and barley are separate.

```r
> scenarioID<-'PC5.EXTG'
> scenario<-SCENARIO_SET[SCENARIO_SET$SCENARIO==scenarioID,]
```

Three short bespoke functions are required within the stochastic beta diversity model, below.

```r
> sampleDist<-function(habitats,p.areas) {sample(x=habitats,size=1,replace=T,prob=p.areas)}
> QLO<-function(x) {quantile(x,probs=0.025,na.rm=T)}
> QHI<-function(x) {quantile(x,probs=0.975,na.rm=T)}
```

We also specify the number of random pairwise points (runs) we want, and generate a blank vector DCA.each.run to populate with these. Here we have specified 10,000 runs.

```r
> runs<-10000
> DCA.each.run<-rep(0,runs)
```

Finally, we run a loop that randomly selects two land uses weighted by their relative proportions and then randomly selects a DCA distance based on the mean and standard
error of the DCA distances between the plots of those two land-uses (including within land-uses).

```r
> for (i in 1:runs){
+ habitat1<-as.character(sampleDist(scenario$TYPE,scenario$AREA))
+ habitat2<-as.character(sampleDist(scenario$TYPE,scenario$AREA))
+ mean<-comparisons[comparisons$Var1==habitat1 & comparisons$Var2==habitat2,3]
+ se<-comparisons[comparisons$Var1==habitat1 & comparisons$Var2==habitat2,4]
+ DCA.each.run[i]<-rnorm(1,mean,se)
+ } #print(c(as.character(habitat1),as.character(habitat2))) # CAN BE TURNED ON TO TRACK PROGRESS OF ANALYSIS - IT IS SUPPRESSED HERE
```

It can be seen in Figure S5 that the distributions of DCA values are multi-modal, reflecting the fact that these are sampled from a limited discrete set of distributions.

```r
> hist(DCA.each.run,main=NULL,xlab="DCA value")
```

**Figure S5.** Histogram of DCA distances for 10,000 runs of the stochastic beta diversity model for the scenario 15% spare land + extensive grassland.

The median and 95% quantiles are then taken for this scenario. By running this for each taxa for each scenario, we were able to estimate changes in the beta diversity index for each scenario relative to that of the baseline scenario.
> median(DCA.each.run, na.rm=T)
[1] 1.100794

> QLO(DCA.each.run)
2.5%
0.404121

> QHI(DCA.each.run)
97.5%
3.188257

Supporting Information references


