

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

4
 5 Patrick. J. C. White, Mark A. Lee, Dave J. Roberts & Lorna J. Cole

6
 7 **SECTION A**

8
 9 **Table S1. Estimates of dry matter (DM) yield in kg per ha, metabolisable energy (ME) per kg**
 10 **and resultant ME per ha for productive land-uses considered in this study.**

ESTIMATE	LAND-USE				
	Silage	Wheat	Barley	Oilseed rape	Fodder beet
YIELD (kg DM ha ⁻¹)	6,800	6,900	6,500	3,600	9,600
ME (MJ kg ⁻¹ DM)	10.8	13.8	13.2	19.1	12.0
ME (MJ ha ⁻¹) x 10 ³	73	95	86	69	115

11
 12 **Table S2. Parameter estimates (± standard error) for all confidence set models given in**
 13 **Table 1 in the main text.**

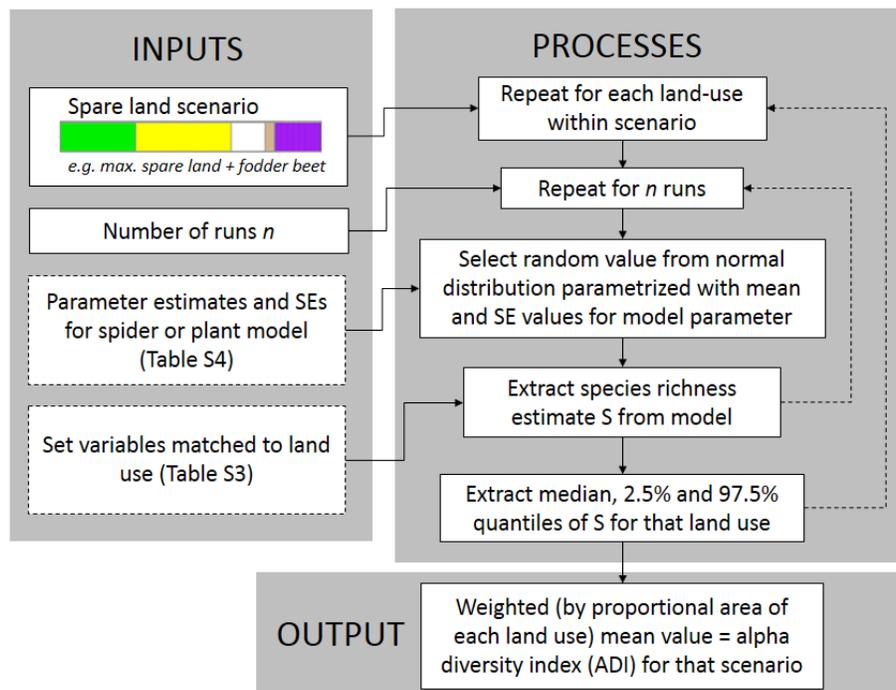
DIVERSITY INDEX	TAXA	MODEL	Intercept	1 st variable	2 nd variable	Interaction
Alpha (Δ_{AD})	Epigeal spiders	RICH*SPARE	-0.253 (±0.103)	-0.000 (±0.003)	-0.225 (±0.009)	0.010 (±0.000)
	Vascular plants	RICH*SPARE	0.651 (±0.146)	-0.003 (±0.002)	-0.083 (±0.014)	0.010 (±0.000)
Beta (Δ_{BD})	Epigeal spiders	DCA*EVEN	1.02 (±0.33)	-0.90 (±0.19)	-1.35 (±0.41)	1.29 (±0.23)
		DCA*HETER	0.99 (±0.35)	-0.87 (±0.20)	-0.70 (±0.26)	0.69 (±0.14)
	Plants	HETER	-2.28 (±0.38)	2.02 (±0.30)	-	-
		EVEN	-2.46 (±0.51)	3.56 (±0.65)	-	-
		DCA+HETER	-2.30 (±0.41)	0.00 (±0.11)	2.03 (0.35)	-
DCA+EVEN	-2.89 (±0.53)	0.22 (±0.11)	3.42 (0.62)	-		

16

17 **SECTION B – Methodology for production of alpha and beta diversity indices for each**
18 **scenario**

19 **Alpha diversity index**

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



38

39 **FIGURE S1. Workflow for producing an alpha diversity index (ADI) for each scenario.**
 40 **This fits into the overall workflow at Figure 2a in the main document.**

41 **Table S3. Fixed variables used for species richness models for vascular plants (Wilson et al.**
 42 **2003) and epigeal spiders (Downie et al. 1999) according to land-use. Variables were set at**
 43 **the mean values for the land-use component from the original data set, except for altitude**
 44 **and annual cutting (marked '*') which were specified by us. Fixed values were taken as the**
 45 **mean values for that variable across all sites of a given land-use within the data (Yelloly**
 46 **1999). Dummy variables indicating the presence/absence of a hedge (spider model) or**
 47 **boundary wall (plant model) were fixed at zero.**

Model	Variable	Land-use component					
		Silage	Barley	Wheat	Oilseed	Fodder beet	Extensive grass.
Plant	Sward height (cm)	18.9	72.2	74.5	99.8	56.6	22.0
Plant	% bare-ground	2.0	67.1	61.7	63.7	24.4	3.1
Plant	Stocking density (livestock units ha ⁻¹)	0	0	0	0	0	0
Plant	% bryophyte cover	3.3	2.3	2.7	0.1	0	26.6
Spider	Altitude (m)	100	100	100	100	100	100
Spider	Number of vegetation layers	2.4	2.4	2.7	2.5	1.7	1.6
Spider	Soil organic content (% loss on ignition)	13.4	5.3	6.2	8.0	30.3	32.9
Spider	Annual cutting [FACTOR LEVELS]*	≥2 cuts	1 cut	1 cut	1 cut	1 cut	0 cuts
Spider	NVC category [FACTOR LEVELS]	2	1	1	1	1	7

48
 49
 50 We obtained the model parameter standard errors (SEs) (Table S4) and used them to
 51 account for uncertainty in the published models (Downie et al., 1999; Wilson et al., 2003).
 52 For individual land-uses, models were run 10³ times and in each run a random parameter
 53 value was drawn from a normal distribution based on that parameter's estimate and SE.
 54 Over all runs, we took median S, plus 2.5% and 97.5% quantiles. The original models were
 55 developed to estimate S for a single land-use, rather than a mixed land-use scenario. It was
 56 therefore not possible to directly obtain a whole-scenario cumulative (gamma) S estimate.
 57 Therefore the ADI was calculated as the mean S across land-uses for each scenario,
 58 weighted by the proportional area of each land-use.

59
 60
 61
 62

63
64
65
66
67
68
69

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).

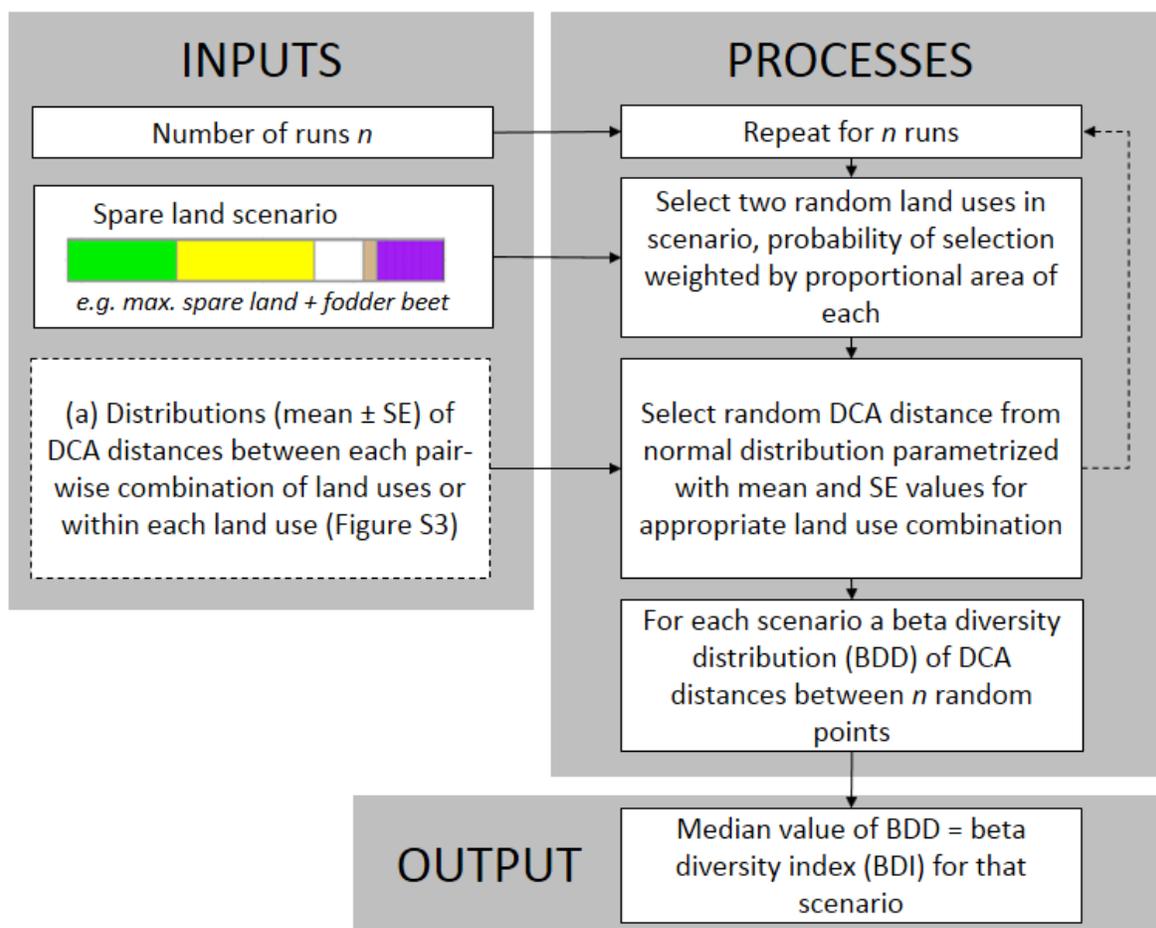
Parameter	Model	
	Plant	Spider
Intercept	3.068 ± 0.128	2.368 ± 0.154
Sward height (m)	-0.009 ± 0.002	-
Presence of wall*	0.061 ± 0.103	-
ln (% Bare-ground + 1)	-0.150 ± 0.036	-
Stocking density (LU ha ⁻¹)	-0.173 ± 0.041	-
Ln (% Bryophyte cover + 1)	0.370 ± 0.067	-
Presence of land use 3 or 9*	0.883 ±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

70
71

72 A beta-diversity index

73 Figures S2 and S3 show the workflow for the estimation of the beta diversity index (BDI)
74 for each scenario, which fits into the overall workflow at Figure 2b in the main document.
75 The accompanying R code for this analysis is provided in Section D of the Supporting
76 Information. Our BDI was based on Euclidean distances on a detrended correspondence
77 analysis (DCA) bi-plot, based on the data from Downie et al. (1999) and Wilson et al.

78 (2003), published in (Murphy et al. 1998). Note that while alpha-diversity models were for
 79 vascular plants only (Wilson et al. 2003), beta-diversity DCA data were for vascular plants
 80 and bryophytes combined (Murphy et al. 1998) although they were derived from the same
 81 study. DCA axes provide an estimate of species turnover, with two sites ≥ 4 units apart
 82 considered to have a very low probability of sharing species (Leps & Smilauer 2003).
 83 Euclidean distances between sites on a DCA bi-plot can therefore be considered an index of
 84 pairwise beta-diversity, with sites that are further apart having a greater complementarity
 85 in the species they support (Smol et al. 2005).

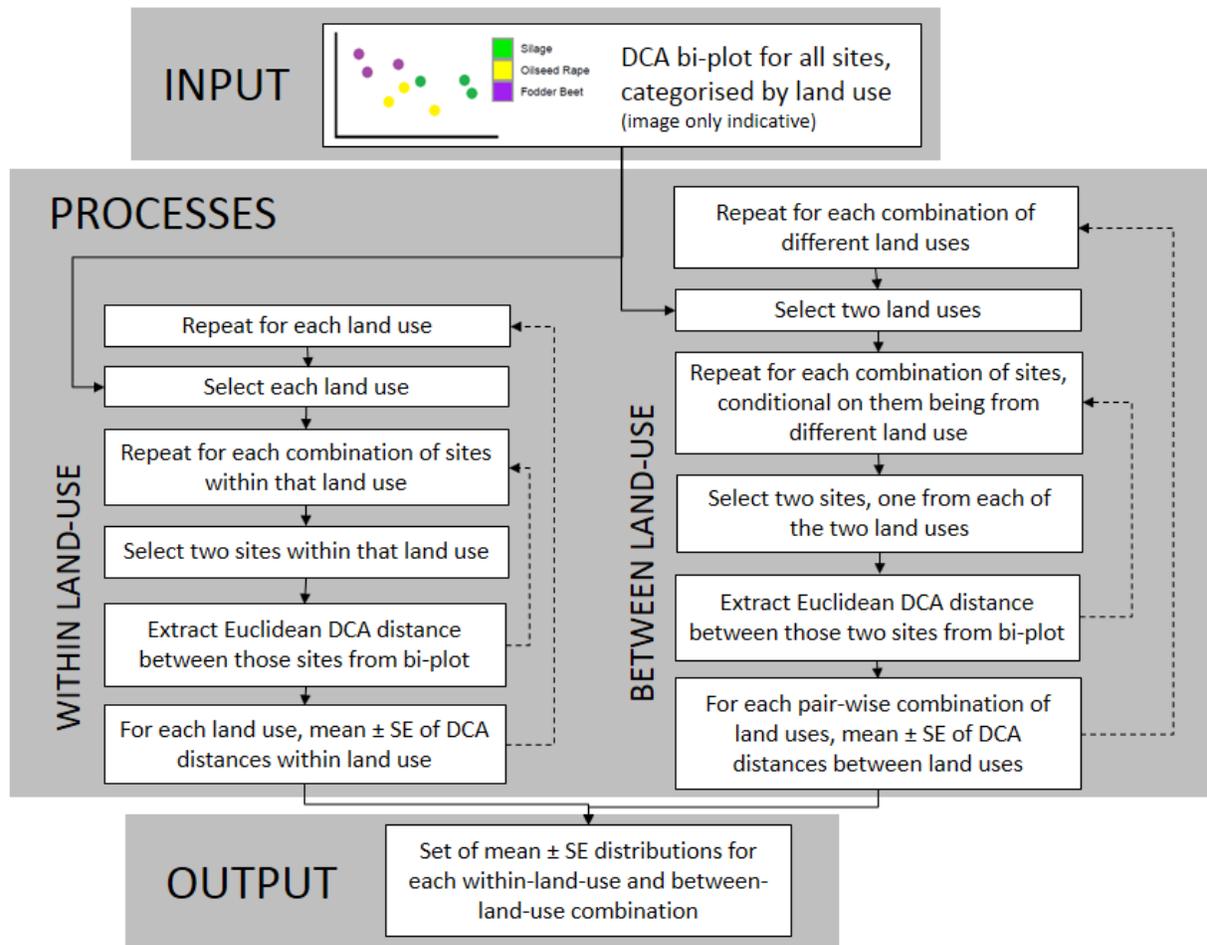


86
 87 **FIGURE S2. Workflow for producing a beta diversity index (BDI) for each scenario.**
 88 **This fits into the overall workflow at Figure 2b in the main document. The sub-**
 89 **workflow indicated by (a) is shown in Figure S3.**

90 For each unique pair of different land-uses (15 for spiders, 10 for plants where wheat and
 91 barley were amalgamated) we calculated the Euclidean distance on the DCA bi-plot
 92 (hereafter 'DCA distance') between each possible combination of sites with a specific land
 93 use pairing, conditional on the pair having a different land-use (nm combinations, where n
 94 = number of sites of the first land-use and m = number of sites of the second). Mean and SE
 95 values gave a distribution of beta-diversity between each land-use pair. Additionally, for
 96 each land-use we calculated the DCA distance between each possible pair of sites within a

97 land-use ($n!/[n-2]!2$ combinations, where n = number of sites). Mean and SE values gave a
 98 distribution of beta-diversity within each land-use. For plants, there were insufficient
 99 wheat sites with DCA data available ($n = 3$) to complete robust analyses, so we combined
 100 wheat and barley into a single category 'cereal'. Within the three wheat sites the mean DCA
 101 distance was 0.70, compared to 0.49 between wheat and barley, 0.38 within barley, and a
 102 mean of 1.93 across all crop combinations, justifying this amalgamation.

103



104

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

109 For each taxa and scenario a weighted beta-diversity distribution (BDD) was defined as the
 110 distribution of predicted DCA distances between any two random locations within that
 111 scenario, given their respective land-use types. To do so, we generated 10^4 pairs of
 112 locations, with each location within each pair being randomly assigned a land-use from that
 113 scenario, weighting this assignment by the relative area occupied by that land-use in that
 114 scenario. For each pairing, we then generated a beta-diversity value which was extracted

115 by random sampling from a normal distribution parameterised with the mean and SE of
116 the DCA distance between sites for that land-use combination (either between-land-use or
117 within-land-use). The beta-diversity index (BDI) for a scenario was taken as the median
118 value across all pairs.

119

120 **SECTION C - Alpha diversity model code**

121

122 Data files referred to are available from the Edinburgh Napier University Repository. DOI:
123 10.17869/ENU.2018.1141073 (White et al., 2018).

124

125 **Species richness models**

126

127 The below model represents regression Model B from Appendix 2 in Downie et al. (1999).
128 This is saved as SPIDERBrnorm.modified.txt. Each model parameter is presented with its
129 estimate (taken from Downie et al. 1999) and standard error (obtained from the authors).
130 For stochasticity in the model, each time this is run one of the parameters is selected at
131 random. It then randomly selects from a normal distribution parameterised with that
132 parameter's estimate and standard error, while holding all other parameters at their fixed
133 estimates. The final output SspiB gives an estimate of epigeal spider species richness based
134 on the parameter estimates and the values of each variable that are presented to the model
135 either as fixed variables, or land-use based variables (explained below).

```
136 > n.pars<-8  
137 > r.par<-sample(1:n.pars,1)  
138 > intercept<-ifelse(r.par==1,rnorm(1,2.368,0.1536),2.368)  
139 > field<-ifelse(SP.FIELD==7,ifelse(r.par==2,rnorm(1,0.4885,0.1870),0.4885),if  
140 else(SP.FIELD==8,ifelse(r.par==2,rnorm(1,0.2638,0.09125),0.2638),0))  
141 > nvc<-ifelse(NVC==3|NVC==4,ifelse(r.par==3,rnorm(1,0.3638,0.07257),0.3638),i  
142 felse(NVC==6|NVC==8|NVC==10,  
143 + ifelse(r.par==3,rnorm(1,0.2939,0.1025),0.2939),ifelse(NVC==7|NVC==9,ifelse(  
144 r.par==3,rnorm(1,0.7162,0.1050),0.7162),0)))  
145 > layers<-ifelse(r.par==4,rnorm(1,0.1614,0.04127),0.1614)*LAYERS  
146 > hedge<-ifelse(HEDGE==1,ifelse(r.par==5,rnorm(1,-0.2736,0.08651),-0.2736),0)  
147 > organicp<-ifelse(r.par==6,rnorm(1,-0.04840,0.01775),-0.04840)*sqrt(ORGANICP  
148 +0.5)  
149 > altitude.ln<-ifelse(r.par==7,rnorm(1,0.1022,0.03041),0.1022)*log(ALTITUDE+1  
150 )  
151 > cutting<-ifelse(CUTTING==3,ifelse(r.par==8,rnorm(1,-0.3055,0.1179),-0.3055)  
152 ,0)  
153 > SspiB<-exp(intercept+field+nvc+layers+hedge+organicp+altitude.ln+cutting)-1
```

154 Similarly, the below model represents the vascular plant regression model from Table 6 in
155 Wilson et al. (2003). This operates in the same stochastic manner as the spider model,

156 above. This is saved as VASCULARrnorm.modified.txt. The final output Svas gives an
 157 estimate of vascular plant species richness based on the parameter estimates and the
 158 values of each variable that are presented to the model either as fixed variables, or land-use
 159 based variables (explained below).

```

160 > n.pars<-8
161 > r.par<-sample(1:n.pars,1)
162 > intercept<-ifelse(r.par==1,rnorm(1,3.068,0.1280),3.068)
163 > field<-ifelse(PL.FIELD==3|PL.FIELD==9,ifelse(r.par==2,rnorm(1,0.8829,0.1669
164 ),0.8829),0)
165 > swardh<-ifelse(r.par==3,rnorm(1,-0.008743,0.001835),-0.008743)*SWARDH
166 > wall<-ifelse(WALL==1,ifelse(r.par==4,rnorm(1,0.06072,0.1026),0.06072),0)
167 > bareground.ln<-ifelse(r.par==5,rnorm(1,-0.1497,0.03587),-0.1497)*log(BAREGR
168 OUND+1)
169 > cattle<-ifelse(r.par==6,rnorm(1,-0.1728,0.04091),-0.1728)*CATTLE
170 > moss<-ifelse(r.par==7,rnorm(1,-0.02741,0.005695),-0.02741)*MOSS
171 > moss.ln<-ifelse(r.par==8,rnorm(1,0.3695,0.06679),0.3695)*log(MOSS+1)
172 > Svas<-exp(intercept+field+swardh+wall+bareground.ln+cattle+moss+moss.ln)-1

```

173

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

175

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

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

```

189 > SETVARS<-read.csv(file=file.choose())

```

	TYPE	PL.FIELD	SP.FIELD	LAYERS	ORGANICP	AVAILP	LITTER	MOSS	SWARDH	BIOMASS	
190	1	BARL	1	2	2.38	5.31	65.00	7.4	2.3	72.2	30.3
191	2	PAST	2	5	2.41	13.37	33.26	3.9	3.3	18.9	5.3
192	3	OILS	3	2	2.50	8.00	202.50	19.5	0.1	99.8	12.2
193	4	SETA	4	4	4.00	6.35	73.50	32.2	1.9	5.7	3.5
194	5	WHEA	5	1	2.67	6.20	70.67	13.6	2.7	74.5	34.2
195	6	GORS	6	7	3.50	9.80	14.50	5.9	12.1	26.7	16.2
196	7	HEAT	7	8	1.57	33.03	12.11	17.2	30.5	16.6	7.0
197											

198	8	EXTG	8	6	1.55	32.90	8.60	7.0	26.6	22.0	14.1
199	9	ROOT	9	3	1.67	30.30	17.07	9.7	0.0	55.6	2.4
200		BAREGROUND									
201	1				67.1						
202	2				2.0						
203	3				63.7						
204	4				1.1						
205	5				61.7						
206	6				0.0						
207	7				0.0						
208	8				3.1						
209	9				24.4						

210 The other means by which fixed variables are entered into the model is within the scenario
211 specifications themselves, discussed in the following section.

212

213 Loading scenarios

214

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

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

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

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

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

	SCENARIO	COMPONENT	AREA	TYPE	NVC	HEDGE	ALTITUDE	CUTTING	WALL	CATTLE
229	1	INDIVIDUAL	10.0	PAST	2	0	100	3	0	0
230	2	INDIVIDUAL	10.0	BARL	1	0	100	1	0	0
231	3	INDIVIDUAL	10.0	WHEA	1	0	100	1	0	0
232	4	INDIVIDUAL	10.0	OILS	1	0	100	1	0	0
233	5	INDIVIDUAL	10.0	ROOT	1	0	100	1	0	0
234	6	INDIVIDUAL	10.0	EXTG	7	0	100	1	0	0
235	7	INDIVIDUAL	10.0	SETA	1	0	100	1	0	0
236	8	BL	25.6	PAST	2	0	100	3	0	0
237	9	BL	11.9	BARL	1	0	100	1	0	0
238	10	BL	15.4	WHEA	1	0	100	1	0	0
239	11	BL	50.0	OILS	1	0	100	1	0	0

240	12	BL	5	0.9	ROOT	1	0	100	1	0	0
241	13	PC1	1	28.4	PAST	2	0	100	3	0	0
242	14	PC1	2	3.7	BARL	1	0	100	1	0	0
243	15	PC1	3	15.8	WHEA	1	0	100	1	0	0
244	16	PC1	4	52.0	OILS	1	0	100	1	0	0
245	17	PC1	5	3.1	ROOT	1	0	100	1	0	0
246	18	PC2	1	28.4	PAST	2	0	100	3	0	0

247

248 Specifications for the stochastic alpha-diversity model

249

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

```
256 > scenarioID<-"PC5.EXTG"
```

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

```
258 > RUNS<-1000
```

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

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

264

265 Preliminary code before running the stochastic model

266

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

```
270 > scenario<-SCENARIO.SET[SCENARIO.SET$SCENARIO==scenarioID,];scenario
271   SCENARIO COMPONENT AREA TYPE NVC HEDGE ALTITUDE CUTTING WALL CATTLE
272 151 PC5.EXTG      1 31.6 PAST  2    0     100      3  0    0
273 152 PC5.EXTG      2  9.3 BARL  1    0     100      1  0    0
274 153 PC5.EXTG      3  9.5 WHEA  1    0     100      1  0    0
275 154 PC5.EXTG      4 36.7 OILS  1    0     100      1  0    0
276 155 PC5.EXTG      5  3.1 ROOT  1    0     100      1  0    0
277 156 PC5.EXTG      6 13.7 EXTG  7    0     100      1  0    0
```

278 The stochastic alpha diversity model runs a series of loops and this preliminary code
279 creates intermediate dataframes to store intermediate results prior to summary, below.

```
280 > SPIDER.RUN<-data.frame(rep(0,nrow(scenario)+1))  
281 > VASC.RUN<-data.frame(rep(0,nrow(scenario)+1))
```

282

283 The stochastic alpha-diversity model

284

285 The stochastic alpha diversity model runs two loops. The inner loop estimates a stochastic
286 species richness estimate for epigeal spiders and vascular plants for land-use *i* within the
287 scenario and populates them into column *h* of the intermediate dataframes SPIDER.RUN and
288 PLANT.RUN. The outer loop repeats this for *h* runs, specified by RUNS and generates a
289 weighted mean species richness estimate for each run, weighted by the relative area each
290 land use occupies.

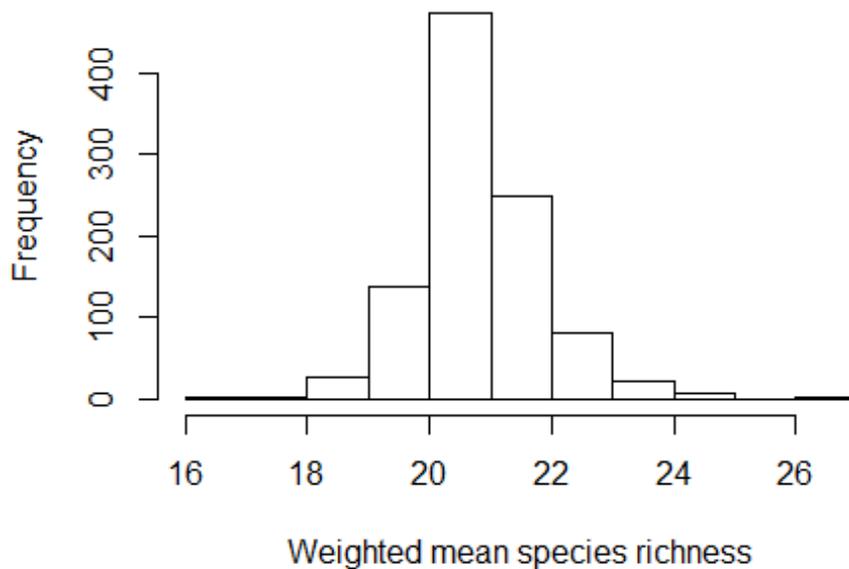
```
291 > setwd(dir=DIRECTORY)  
292 >  
293 > # START OF LOOP FOR RUN h WITHIN ALL RUNS  
294 > for(h in 1:RUNS){  
295 + #print(c(h," / ",RUNS),quote=F) #THIS OPTION ALLOWS TRACKING OF PROGRESS OF  
296 RUNS - IT IS SUPPRESSED HERE  
297 +  
298 +  
299 + # START OF LOOP FOR LAND-USE TYPE i WITHIN SPECIFIED SCENARIO  
300 + for(i in 1:nrow(scenario)){  
301 + source("COMPONENTVARS.txt")  
302 +  
303 + # READ IN SOURCE CODE FOR VASCULAR PLANT MODEL AND EPIGEAL SPIDER MODEL A  
304 ND ESTIMATE SPECIES RICHNESS FOR THE LAND-USE  
305 + # NOTE THAT FOR EACH i, A DIFFERENT SPECIES RICHNESS IS ESTIMATED STOCHAS  
306 TICALLY BASED ON STANDARD ERROR ESTIMATES OF MODELS  
307 + source("VASCULARrnorm.modified.txt");VASC.RUN[i,h]<-round(Svas,2)  
308 + source("SPIDERBrnorm.modified.txt");SPIDER.RUN[i,h]<-round(SspiB,2)  
309 +  
310 + }# END OF LOOP FOR LAND-USE TYPE i WITHIN SPECIFIED SCENARIO  
311 +  
312 + # FOR EACH RUN, A WEIGHTED MEAN SPECIES RICHNESS ESTIMATE IS MADE, WEIGHT  
313 ED BY RELATIVE AREA OF EACH LAND-USE  
314 + p.area<-scenario$AREA/sum(scenario$AREA)  
315 + SPIDER.RUN[nrow(scenario)+1,h]<-sum(SPIDER.RUN[1:nrow(SPIDER.RUN)-1,h]*p.  
316 area)  
317 + VASC.RUN[nrow(scenario)+1,h]<-sum(VASC.RUN[1:nrow(VASC.RUN)-1,h]*p.area)  
318 +  
319 + }# END OF LOOP FOR RUN h WITHIN ALL RUNS
```

320 The output is a dataframe of h columns, each of which is a run, and i+1 rows where i is the
321 number of land-uses in the scenario. The additional, final row is the weighted mean species
322 richness for that land-use. The first 6 columns only are shown.

```
323 > SPIDER.RUN[,1:6]
324   rep.0..nrow.scenario...1.      V2      V3      V4      V5      V6
325 1      14.53000 12.11000 14.53000 15.01000 15.38000 14.49000
326 2      20.74000 26.45000 21.36000 22.68000 19.53000 21.36000
327 3      22.23000 20.19000 20.93000 23.09000 22.23000 23.26000
328 4      17.69000 21.24000 21.24000 21.24000 22.19000 20.91000
329 5      16.13000 16.13000 16.01000 16.13000 16.13000 16.13000
330 6      25.45000 38.74000 33.00000 33.53000 33.00000 32.68000
331 7      18.39371 20.98859 20.57623 21.11133 21.12895 20.62192
```

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

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



336
337 **Figure S4. Histogram of weighted mean species richness estimations for 1,000 runs**
338 **of the stochastic alpha-diversity model for the scenario 15%. spare land + extensive**
339 **grassland.**

340

341 Note that as the distribution shows strong central tendency, the mean and median for the
342 weighted mean species richness are very similar.

```
343 > mean(as.numeric(paste(SPIDER.RUN[nrow(scenario)+1,])))  
344 [1] 20.81855
```

```
345 > median(as.numeric(paste(SPIDER.RUN[nrow(scenario)+1,])))  
346 [1] 20.73984
```

347

348 **Populating a final output dataframe**

349

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

```
354 > LANDUSE<-c(as.character(scenario$TYPE),"ALL")  
355 > AREA<-c(scenario$AREA,sum(scenario$AREA))  
356 > SPIDER.MEAN<-rep(0,nrow(scenario)+1);SPIDER.95LO<-rep(0,nrow(scenario)+1);S  
357 PIDER.95HI<-rep(0,nrow(scenario)+1)  
358 > VASC.MEAN<-rep(0,nrow(scenario)+1);VASC.95LO<-rep(0,nrow(scenario)+1);VASC.  
359 95HI<-rep(0,nrow(scenario)+1)  
360 > output<-data.frame(LANDUSE,AREA,SPIDER.MEAN,SPIDER.95LO,SPIDER.95HI,VASC.ME  
361 AN,VASC.95LO,VASC.95HI)  
362 > p.area<-output$AREA[1:nrow(output)-1]/output$AREA[nrow(output)]  
363 > #output$SPIDER.MEAN[nrow(output)]<-sum(output$SPIDER.MEAN[1:nrow(output)-1]  
364 *p.area)  
365 > #output$VASC.MEAN[nrow(output)]<-sum(output$VASC.MEAN[1:nrow(output)-1]*p.a  
366 rea)  
367 > output  
368   LANDUSE  AREA SPIDER.MEAN SPIDER.95LO SPIDER.95HI VASC.MEAN VASC.95LO  
369 1   PAST   31.6           0           0           0           0           0  
370 2   BARL    9.3           0           0           0           0           0  
371 3   WHEA    9.5           0           0           0           0           0  
372 4   OILS   36.7           0           0           0           0           0  
373 5   ROOT    3.1           0           0           0           0           0  
374 6   EXTG   13.7           0           0           0           0           0  
375 7    ALL  103.9           0           0           0           0           0  
376   VASC.95HI  
377 1           0  
378 2           0  
379 3           0  
380 4           0  
381 5           0  
382 6           0  
383 7           0
```

384 Finally, the output dataframe is populated using the mean, lower 95% quantile and upper
 385 95% quantile of species richness for each species for each land-use and for the weighted
 386 mean across all scenarios.

```

387 > output$SPIDER.MEAN<-c(rowMeans(SPIDER.RUN))
388 > output$VASC.MEAN<-c(rowMeans(VASC.RUN))
389 > QLO<-function(x) {quantile(x,probs=0.025)} # BESPOKE LOWER 95% QUANTILE FU
390 NCTION
391 > QHI<-function(x) {quantile(x,probs=0.975)} # BESPOKE UPPER 95% QUANTILE FU
392 NCTION
393 > output$SPIDER.95LO<-c(apply(SPIDER.RUN,1,QLO))
394 > output$VASC.95LO<-c(apply(VASC.RUN,1,QLO))
395 > output$SPIDER.95HI<-c(apply(SPIDER.RUN,1,QHI))
396 > output$VASC.95HI<-c(apply(VASC.RUN,1,QHI))
397 >
398 > output # DISPLAYS THE OUTPUT DATFRAME
399 LANDUSE AREA SPIDER.MEAN SPIDER.95LO SPIDER.95HI VASC.MEAN VASC.95LO
400 1 PAST 31.6 14.60506 11.32950 18.43200 23.15478 19.92575
401 2 BARL 9.3 21.48863 17.38925 26.35250 7.89850 6.17975
402 3 WHEA 9.5 22.30467 18.18000 27.41150 8.12309 6.45975
403 4 OILS 36.7 21.40539 17.60825 26.62025 11.10441 8.30700
404 5 ROOT 3.1 16.15268 12.95900 19.68075 18.83240 14.82975
405 6 EXTG 13.7 33.14873 26.25850 41.01075 22.67104 17.72975
406 7 ALL 103.9 20.81855 18.88218 23.08025 15.96556 14.55723
407 VASC.95HI
408 1 26.00025
409 2 9.92025
410 3 10.19025
411 4 13.96400
412 5 23.37275
413 6 29.11400
414 7 17.41407

```

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

417

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

419

420 Data files referred to are available from the Edinburgh Napier University Repository. DOI:
 421 10.17869/ENU.2018.1141073 (White et al., 2018).

422

423 The beta-diversity model uses the same scenario dataset structure as the alpha-diversity
 424 model, since we tested the same set of scenarios, although many of the fixed variable fields
 425 in this are not used to assess beta diversity. Rather, we also imported a dataset of
 426 detrended correspondence analysis (DCA) axes 1 and 2 coordinates (described in main

427 text) for a series of sites throughout Scotland, each of which is assigned a different land-
428 use, which match against those in our scenarios. These DCA values are saved in the file DCA
429 VALUES.csv and imported to a dataframe object DCA.VALUES.

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

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

```
435 > DCA.VALUES<-DCA.VALUES[DCA.VALUES$TAXA=="SPIDER",];head(DCA.VALUES,10)
436   TAXA  SITE  AXIS1  AXIS2  CROP
437 1 SPIDER  MM1  5.60  2.01  BARL
438 2 SPIDER  RH2  5.15  1.85  BARL
439 3 SPIDER  RH6  5.31  1.84  BARL
440 4 SPIDER  RH7  5.21  1.93  BARL
441 5 SPIDER  SAMS1 4.91  1.68  BARL
442 6 SPIDER  SAVF1 4.83  1.68  BARL
443 7 SPIDER  KPLS1 2.41  0.55  EXTG
444 8 SPIDER  KPLS3 3.73  1.73  EXTG
445 9 SPIDER  KPLS4 2.56  0.04  EXTG
446 10 SPIDER KPLS5 3.63  1.70  EXTG
```

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

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

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

```
456 > distd<-as.matrix(dist(DCA.VALUES[,3:4]));distd[1:6,1:6]
457   1      2      3      4      5      6
458 1 0.000000 0.4775982 0.3361547 0.3981206 0.7648529 0.8377350
459 2 0.4775982 0.0000000 0.1603122 0.1000000 0.2941088 0.3623534
460 3 0.3361547 0.1603122 0.0000000 0.1345362 0.4308132 0.5059644
461 4 0.3981206 0.1000000 0.1345362 0.0000000 0.3905125 0.4548626
462 5 0.7648529 0.2941088 0.4308132 0.3905125 0.0000000 0.0800000
463 6 0.8377350 0.3623534 0.5059644 0.4548626 0.0800000 0.0000000
```

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

```
466 > types<-as.factor(levels(DCA.VALUES$CROP))
467 > combos<-expand.grid(types,types);head(combos,10)
468   Var1 Var2
```

```

469 1 BARL BARL
470 2 EXTG BARL
471 3 GORS BARL
472 4 HEAT BARL
473 5 OILS BARL
474 6 PAST BARL
475 7 ROOT BARL
476 8 SETA BARL
477 9 WHEA BARL
478 10 BARL EXTG

```

479 A loop is required to estimate the mean, standard error and sample size (n) of DCA
480 distances for all pairs of sample plots for each pair-wise land-use combination. Initially, and
481 output dataframe comparisons is generated, which will be populated by the loop. Only the
482 first 10 lines are shown.

```

483 > means<-rep(0,nrow(combos));se<-rep(0,nrow(combos));count<-rep(0,nrow(combos
484 ))
485 > comparisons<-cbind(combos,means,se,count);head(comparisons,10)
486   Var1 Var2 means se count
487 1 BARL BARL    0  0     0
488 2 EXTG BARL    0  0     0
489 3 GORS BARL    0  0     0
490 4 HEAT BARL    0  0     0
491 5 OILS BARL    0  0     0
492 6 PAST BARL    0  0     0
493 7 ROOT BARL    0  0     0
494 8 SETA BARL    0  0     0
495 9 WHEA BARL    0  0     0
496 10 BARL EXTG    0  0     0

```

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

```

499 > for (i in 1:nrow(combos)){
500 +
501 +   (rows<-which(DCA.VALUES$CROP==combos[i,1]))
502 +   (cols<-which(DCA.VALUES$CROP==combos[i,2]))
503 +
504 +   ifelse(sum(rows-cols)!=0,
505 +     subdists<-dstd[rows,cols],
506 +     subdists<-dstd[rows,cols][lower.tri(dstd[rows,cols])])
507 +   comparisons$means[i]<-round(mean(subdists),2)
508 +   comparisons$se[i]<-round(sd(subdists)/sqrt(length(subdists)),2)
509 +   comparisons$count[i]<-length(subdists)}
510 > head(comparisons,10)
511   Var1 Var2 means se count
512 1 BARL BARL  0.38 0.06   15
513 2 EXTG BARL  2.73 0.14   48
514 3 GORS BARL  3.55 0.12    6

```

```

515 4 HEAT BARL 3.99 0.08 42
516 5 OILS BARL 0.53 0.07 18
517 6 PAST BARL 0.95 0.08 108
518 7 ROOT BARL 1.65 0.11 18
519 8 SETA BARL 0.39 0.12 6
520 9 WHEA BARL 0.44 0.07 18
521 10 BARL EXTG 2.73 0.14 48

```

522 As discussed, we use the same scenario set (SCENARIO.SET), which was imported for the
523 alpha diversity analyses, earlier. As with the DCA.values, for plant diversity we combine
524 wheat and barley to a single 'cereal' category, using the below code. This is not used for
525 spider diversity analyses which we are demonstrating here.

```

526 > levels(SCENARIO.SET$TYPE)
527 [1] "BARL" "EXTG" "OILS" "PAST" "ROOT" "SETA" "WHEA"

528 > levels(SCENARIO.SET$TYPE)<-c("CERE","EXTG","OILS","PAST","ROOT","SETA","CER
529 E")

```

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

```

535 > scenarioID<-"PC5.EXTG"
536 > scenario<-SCENARIO.SET[SCENARIO.SET$SCENARIO==scenarioID,];scenario
537   SCENARIO COMPONENT AREA TYPE NVC HEDGE ALTITUDE CUTTING WALL CATTLE
538 151 PC5.EXTG      1 31.6 PAST  2    0    100      3    0    0
539 152 PC5.EXTG      2  9.3 BARL  1    0    100      1    0    0
540 153 PC5.EXTG      3  9.5 WHEA  1    0    100      1    0    0
541 154 PC5.EXTG      4 36.7 OILS  1    0    100      1    0    0
542 155 PC5.EXTG      5  3.1 ROOT  1    0    100      1    0    0
543 156 PC5.EXTG      6 13.7 EXTG  7    0    100      1    0    0

```

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

```

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

```

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

```

552 > runs<-10000
553 > DCA.each.run<-rep(0,runs)

```

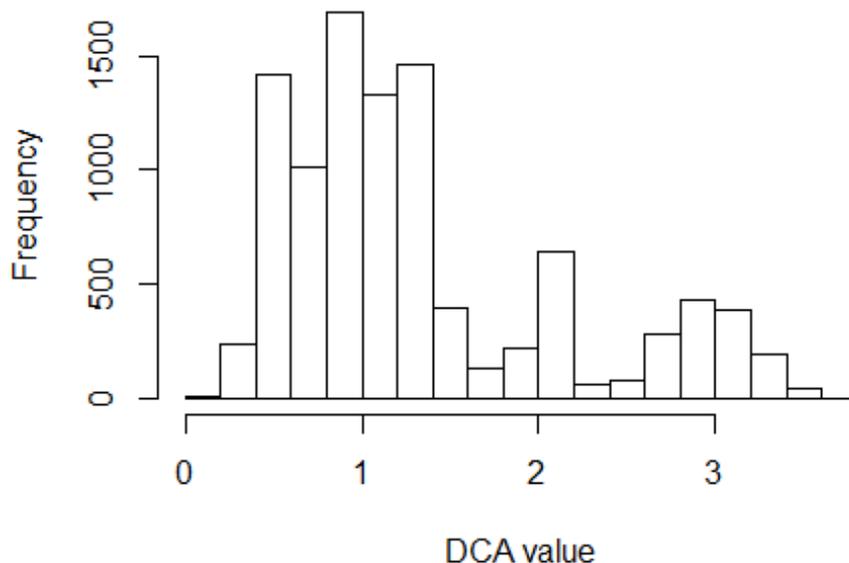
554 Finally, we run a loop that randomly selects two land uses weighted by their relative
555 proportions and then randomly selects a DCA distance based on the mean and standard

556 error of the DCA distances between the plots of those two land-uses (including within land-
557 uses).

```
558 > for (i in 1:runs){  
559 + habitat1<-as.character(sampleDist(scenario$TYPE,scenario$AREA))  
560 + habitat2<-as.character(sampleDist(scenario$TYPE,scenario$AREA))  
561 + mean<-comparisons[comparisons$Var1==habitat1 & comparisons$Var2==habitat2,3  
562 ]  
563 + se<-comparisons[comparisons$Var1==habitat1 & comparisons$Var2==habitat2,4]  
564 + DCA.each.run[i]<-rnorm(1,mean,se)  
565 + #print(c(as.character(habitat1),as.character(habitat2))) # CAN BE TURNED ON  
566 TO TRACK PROGRESS OF ANALYSIS - IT IS SUPPRESSED HERE  
567 + }
```

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

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



571

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

574

575 The median and 95% quantiles are then taken for this scenario. By running this for each
576 taxa for each scenario, we were able to estimate changes in the beta diversity index for
577 each scenario relative to that of the baseline scenario.

```
578 > median(DCA.each.run,na.rm=T)
579 [1] 1.100794
```

```
580 > QLO(DCA.each.run)
581     2.5%
582 0.404121
```

```
583 > QHI(DCA.each.run)
584     97.5%
585 3.188257
```

586

587 **Supporting Information references**

588 Downie, I.S., Wilson, W.L., Abernethy, V.J., McCracken, D.I., Foster, G.N., Ribera, I., et al.
589 (1999). The impact of different agricultural land-uses on epigeal spider diversity in
590 Scotland. *J. Insect Conserv.*, 3, 273-286

591 Murphy, K.J., McCracken, D.I., Foster, G.N., Waterhouse, A., Furness, R., Abernethy, V.J.,
592 Downie, I.S., Wilson, W.L., Adam, A. & Ribera, I. (1998) Functional Analysis of Plant-
593 Invertebrate-Bird Biodiversity on Scottish Agricultural Land. Final Report to SOAEFD No:
594 UGW/814/94, 51. University of Glasgow, Glasgow.

595 Smol, J.P., Wolfe, A.P., Birks, H.J.B., Douglas, M.S. V, Jones, V.J., Korhola, A., Pienitz, R.,
596 Rühland, K., Sorvari, S., Antoniades, D., Brooks, S.J., Fallu, M.-A., Hughes, M., Keatley, B.E.,
597 Laing, T.E., Michelutti, N., Nazarova, L., Nyman, M., Paterson, A.M., Perren, B., Quinlan, R.,
598 Rautio, M., Saulnier-Talbot, E., Siitonen, S., Solovieva, N. & Weckström, J. (2005) Climate-
599 driven regime shifts in the biological communities of arctic lakes. *Proceedings of the*
600 *National Academy of Sciences of the United States of America*, 102, 4397–402.

601 White, P.J.C., Lee, M.A., Roberts, D.J. & Cole, L.J. (2018) Data from: Routes to achieving
602 sustainable intensification in simulated dairy farms - the importance of production
603 efficiency and complimentary land uses. Edinburgh Napier University Repository.
604 <https://doi.org/10.17869/ENU.2018.1141073>

605 Wilson, W., Abernethy, V., Murphy, K., Adam, A., McCracken, D., Downie, I., et al. (2003).
606 Prediction of plant diversity response to land-use change on Scottish agricultural land.
607 *Agric. Ecosyst. Environ.*, 94, 249-263

608 Yellooly, J. M. (1999). Biodiversity user interface. Unpublished MSc thesis, University of
609 Glasgow.