Accounting for dispersal and biotic interactions in order to disentangle the drivers of species distributions and their abundances

Isabelle BOULANGEAT, Dominique GRAVEL, Wilfried THUILLER

Supporting Information

The supplementary information provides additional results and details concerning either the data or the method. It includes six sections with tables, figures and equations:

1) Details on the datasets

Table S1: Species observation data. The following table gives the species number of observation in each abundance class, for each modelled species, over the 8160 sampling plots.

Abundance classes:	0	1	2	3	4	5	6
Abies alba	7427	236	156	144	94	74	29
Alnus glutinosa	8078	26	11	18	13	9	5
Arnica montana	7935	137	55	31	2	0	0
Bromus erectus	7191	224	196	193	186	126	44
Buxus sempervirens	7571	129	105	154	122	63	16
Cacalia alliariae	7813	127	87	73	40	18	2
Carex ferruginea	8052	20	25	21	31	9	2
Dactylis glomerata	7043	568	298	179	61	10	1
Dryas octopetala	7939	67	49	58	31	14	2
Euphorbia cyparissias	7355	587	182	34	2	0	0
Festuca paniculata	7888	93	47	51	51	28	2
Geranium sylvaticum	7386	364	236	136	35	3	0
Kobresia myosuroides	7959	60	42	42	51	6	0
Larix decidua	7839	110	59	45	61	27	19
Phragmites australis	7949	67	45	46	25	9	19
Plantago alpina	7433	199	280	174	57	15	2
Polygonum viviparum	7308	351	353	135	12	1	0
Ranunculus glacialis	7949	130	63	18	0	0	0
Rhododendron ferrugineum	7738	148	83	78	67	33	13
Urtica dioica	7755	204	109	62	22	4	4
Vaccinium myrtillus	7290	204	221	224	159	47	15

Table S2: Dispersal parameters. The following table gives the dispersal parameters used in the dispersal model. They have been attributed according to their dispersal class following Vittoz *et. al* (2007).

Species	Dispersal class	d99 (m)	ldd (m)
Abies alba	4	150	1000
Alnus glutinosa	4	150	1000
Arnica montana	3	15	1000
Bromus erectus	4	150	1000
Buxus sempervirens	3	15	1000
Cacalia alliariae	5	500	5000
Carex ferruginea	6	1500	5000
Dactylis glomerata	7	5000	10000
Dryas octopetala	3	15	1000
Euphorbia cyparissias	7	5000	10000
Festuca paniculata	3	15	1000
Geranium sylvaticum	6	1500	5000
Kobresia myosuroides	6	1500	5000
Larix decidua	6	1500	5000
Phragmites australis	5	500	5000
Plantago alpina	6	1500	5000
Polygonum viviparum	1	1	1000
Ranunculus glacialis	4	150	1000
Rhododendron ferrugineum	1	1	1000
Urtica dioica	7	5000	10000
Vaccinium myrtillus	7	5000	10000

Table S3: Abiotic variables' description. We provide a detailed description of each abiotic variable and precise the source.

Name	**Description**	**Source**
Temperature	Annual mean daily monthly mean temperature(°C)	Meteo France[1] 100x100m resolution
Precipitations	Winter precipitations (January to March)(mm)	
Moisture	Moisture index for the growing season (June to August) measured as the balance between gains from precipitation and losses from potential evapo-transpiration ($mm.d^{-1}$)	
Radiation	Potential yearly global radiation, expressing the potential amount of direct and diffuse solar irradiation ($kJ.m^{-2}.d^{-1}$)	
Soil carbon	Percentage of carbon in the bedrock	Soil European Database http://eusoils.jrc.ec.europa.eu/data.html 1kmx1km resolution
Available water capacity	Available water that can be stored in soil and be available for use by plants	
Slope	Slope angle (°)	French Digital elevation model at resolution 50x50m (IGN-France)
Topography	Difference between the average elevation in a circular moving window and the centre cell of the window	
Wetness index	Topographic wetness index (TWI) [2]	
Corine land cover	Corine Land Cover simplified to 7 classes: artificial surfaces (1**), agricultural areas (2**), forest (31*), scrub (32*), open spaces (33*), wetlands (4**) and water bodies (5**)	European Topic Centre on Land Use and Spatial Information. 100mx100m resolution Date of delivery: 14 Sep 2009

[1] Bénichou, P. and O. Le Breton. 1987. Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques. La Météorologie 7:23-34.

[2] TWI=upslope contributing area / tan(slope angle). TWI is an hydrological index developed by Beven and Kirkby 1979. This method calculates the amount of water that runs through a pixel and corrects this for slope in order to have a measure of potential soil humidity. Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrol. Sci. Bull., 24, 43–69, 1979.

Fig.S1 Vegetation datasets. Left: study area with the 8160 community plots where species are modelled. Right : study area with all additional presence-only data, used to build the different indices.

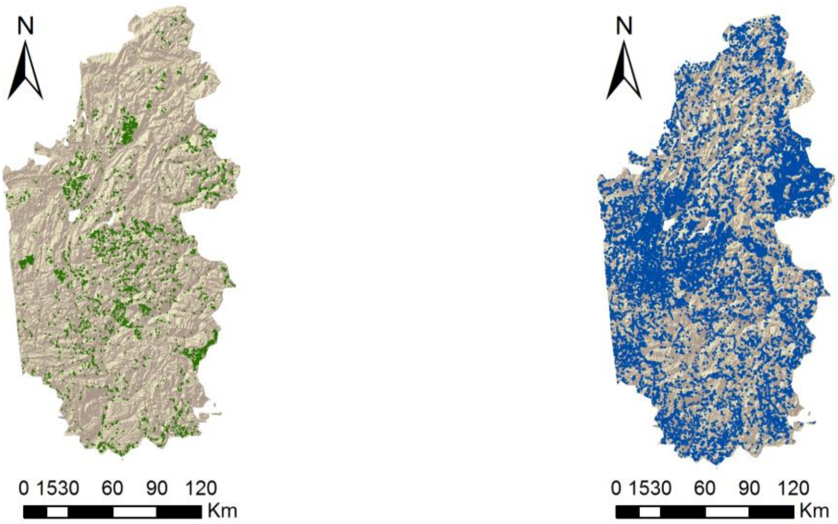
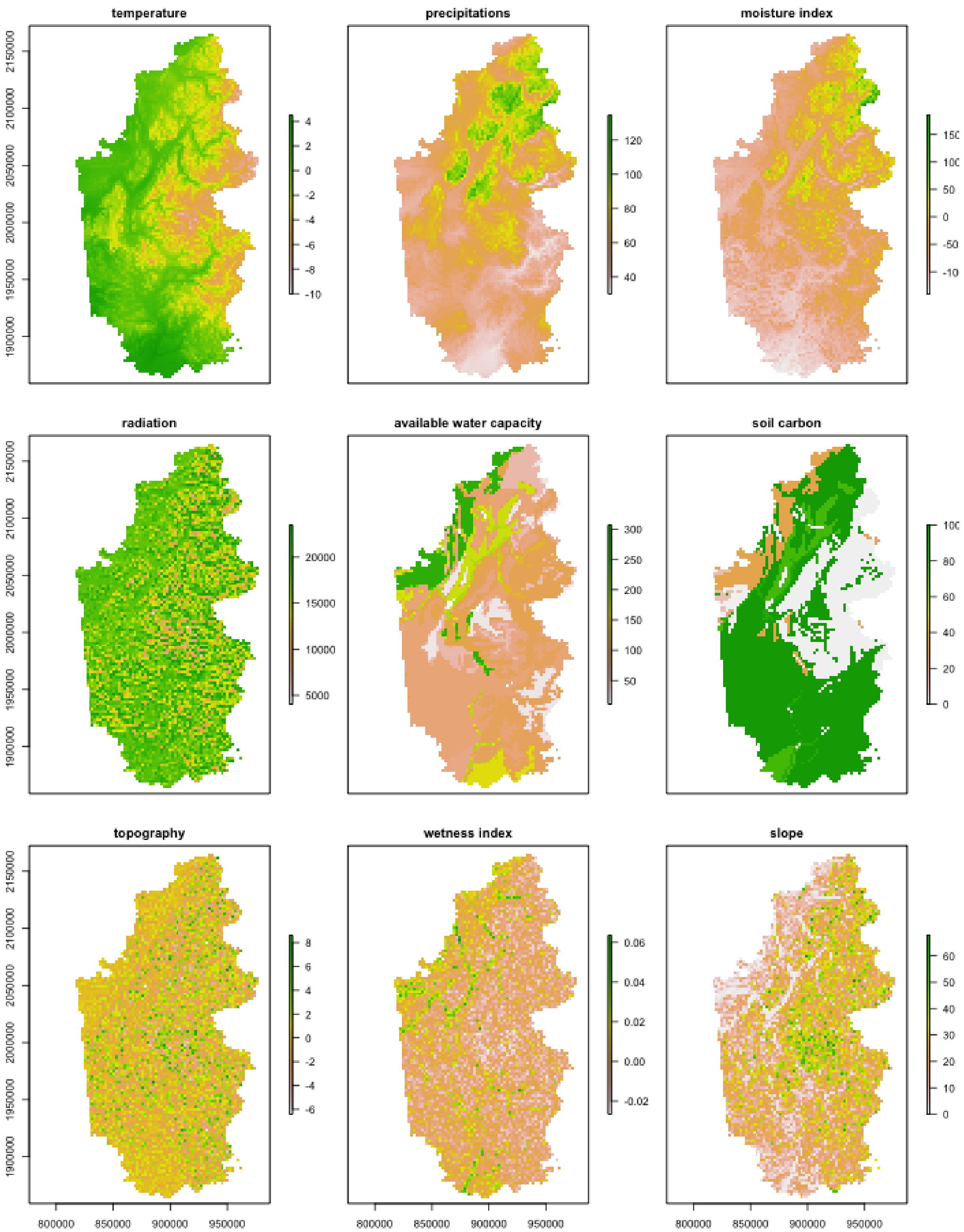
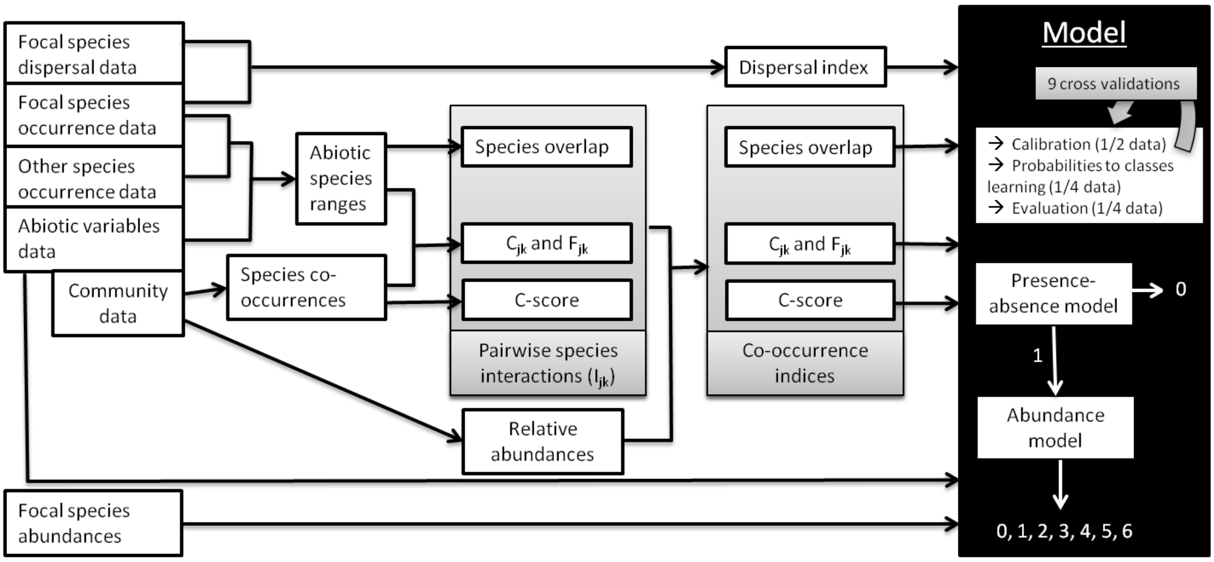


Fig.S2 Topo-climatic variables distributions. The following maps show variable spatial distributions. The geographic scale is in meters. The average temperature is in °C, the winter precipitation is in mm, the moisture index of the growing season is in mm.d^{-1}, the potential yearly global radiation is in kJ.m^{-2}.d^{-1}, the available water capacity is in mm and the carbon in the bedrock is in percentage, the topography is in meters, the topographic wetness index has no units, and the slope angle is in degree.



2) Details on the methods

Fig. S3. Model workflow. From vegetation databases abiotic variables and community composition, we derived the different co-occurrence indices. The resulting indices were then used as predictor variables in the main model. This model consists in two steps. The first step is a classical presence-absence distribution model and the second step models abundance classes for the sites where presence has been predicted by the first step. We used a cross-validation approach, dividing the dataset into three parts in each repetition, for calibration (50%), transformation into classes (25%) and evaluation (25%). For presence-absence, the binary transformation into presence-absence was based on the threshold optimizing the True Skill Statistic. For abundance classes, it was based on weights optimizing the Hanssen-Kuipers score (eq. S1).



Eq. S1 Calculation of the Hanssen-Kuipers discriminant score.

$$HK = \frac{\frac{1}{N}\sum_i n(P_i, O_i) - \frac{1}{N^2}\sum_i N(P_i).N(O_i)}{1 - \frac{1}{N^2}\sum_i (N(O_i))^2}$$

where N is the total number of sites, $n(P_i, O_i)$ is the number of predictions that match with obervations for the class i, $N(O_i)$ is the number of observations in the class i and $N(P_i)$ is the number of predictions in the class i.

Eq. S2 Calculation of variable importance.

The variable importance for the variable x is calculated using the Out-Of-Bag (OOB) sample which is a subsample of the data that has been put aside during the calibration and is used as independant data to evaluate the model. The variable importance is the mean difference in accuracy between normal predictions and predictions with a randomly permuted variable x. For each permuation, the variable importance VI is equal to :

$$VI_x = \frac{1}{N}\sum_i n(P_i, O_i) - \frac{1}{N}\sum_i n(P_i^x, O_i)$$

where N is the number of plots in the OOB sample, $n(P_i, O_i)$ is the number of predictions that match with observations for the class i and P^x is the prediction vector obtained after randomly permuting the predictor variable x.

3) Intermediate results

Table S4: Correlations between proposed indices and topo-climatic variables. The following table gives the Pearson correlations in sampled sites between the proposed indices and all topo-climatic variables, for the 21 focal species. Highest correlations (above 0,6) are highlighted. T°: average temperature, Precip: winter precipitations, Moist:moisture index for the growing season, Rad: potential yearly radiation, WC: soil available water capacity, Carbon: percentage of carbon in the bedrock, topo: topography, wetness: topographic wetness index, slope.

	T°	Precip.	Moist.	Rad.	WC	Carbon	topo	wetness	slope
Abies alba									
species pool	0.14	-0.01	-0.15	0.00	-0.09	0.10	-0.08	-0.02	-0.04
repulsion	-0.11	0.08	0.07	0.05	-0.09	-0.01	0.01	-0.06	0.03
attraction	0.09	0.09	0.05	-0.07	0.08	0.04	-0.01	-0.01	0.05
Cscore	-0.23	-0.01	0.07	0.06	-0.11	-0.03	0.09	-0.07	-0.02
dispersal	0.12	0.22	0.14	-0.17	-0.02	0.15	0.04	-0.14	0.14
Alnus glutinosa									
species pool	0.48	-0.30	-0.52	0.15	0.06	0.20	-0.12	0.14	-0.19
repulsion	**0.67**	-0.26	-0.51	0.20	0.25	0.26	-0.06	0.11	-0.08
attraction	0.33	-0.22	-0.28	0.07	0.13	0.09	-0.04	0.33	-0.20
Cscore	-0.17	0.10	0.14	-0.04	-0.10	0.01	0.11	-0.17	0.09
dispersal	0.27	-0.11	-0.17	0.05	0.28	-0.06	-0.06	0.26	-0.20
Arnica montana									
species pool	-0.28	0.17	0.18	-0.12	-0.30	-0.01	-0.05	-0.19	0.13
repulsion	-0.57	0.42	0.56	-0.24	-0.29	-0.10	0.13	-0.35	0.27
attraction	-0.22	0.02	0.10	0.00	-0.13	0.00	0.03	-0.03	-0.04
Cscore	0.01	0.01	0.00	-0.02	-0.02	0.06	0.07	-0.10	0.07
dispersal	-0.30	0.23	0.32	-0.04	-0.05	-0.24	0.02	-0.11	0.10
Bromus erectus									
species pool	0.16	-0.04	-0.18	0.01	-0.08	0.11	-0.09	-0.01	-0.05
repulsion	0.56	-0.15	-0.35	0.07	0.22	0.24	-0.03	0.11	-0.04
attraction	0.03	0.09	0.04	0.00	0.03	-0.02	-0.05	0.03	-0.04
Cscore	-0.23	0.17	0.23	-0.13	-0.11	-0.02	0.11	-0.15	0.10
dispersal	0.36	-0.29	-0.43	0.26	0.04	0.17	-0.03	0.05	-0.11
Buxus sempervirens									
species pool	0.45	-0.28	-0.49	0.14	0.04	0.19	-0.12	0.13	-0.17
repulsion	**0.72**	-0.24	-0.53	0.20	0.27	0.22	-0.13	0.25	-0.17
attraction	0.39	-0.21	-0.29	0.09	0.18	0.10	-0.04	0.28	-0.15
Cscore	-0.29	0.16	0.24	-0.09	-0.14	-0.05	0.09	-0.11	0.04
dispersal	0.57	-0.30	-0.47	0.15	0.17	0.33	0.05	-0.01	-0.08
Cacalia alliariae									
species pool	-0.09	0.07	0.02	-0.06	-0.22	0.04	-0.07	-0.12	0.06
repulsion	-0.41	0.35	0.43	-0.16	-0.22	-0.05	0.12	-0.29	0.22
attraction	-0.30	0.27	0.33	-0.09	-0.13	-0.04	0.09	-0.16	0.15
Cscore	-0.14	-0.01	0.03	0.04	-0.08	0.02	0.10	-0.09	0.01
dispersal	-0.03	0.41	0.37	-0.30	-0.02	0.02	0.08	-0.17	0.18
Carex ferruginea									
species pool	0.06	0.05	-0.07	-0.03	-0.12	0.08	-0.07	-0.06	0.00
repulsion	-0.48	0.41	0.52	-0.25	-0.24	-0.08	0.13	-0.31	0.25
attraction	-0.04	0.10	0.09	-0.01	-0.02	0.00	0.02	-0.03	0.00
Cscore	-0.07	0.04	0.05	-0.03	-0.05	0.03	0.09	-0.11	0.05
dispersal	-0.12	0.52	0.45	-0.12	-0.02	0.05	0.13	-0.18	0.15
Dactylis glomerata									
species pool	0.14	-0.02	-0.16	0.00	-0.09	0.10	-0.09	-0.02	-0.04
repulsion	0.31	-0.02	-0.14	-0.01	0.10	0.22	0.07	-0.09	0.11
attraction	0.11	-0.10	-0.16	0.18	-0.01	0.06	0.03	-0.06	0.05
Cscore	-0.17	0.10	0.16	-0.08	-0.09	0.01	0.13	-0.17	0.12
dispersal	0.39	0.00	-0.20	0.15	0.19	0.00	0.00	0.09	-0.08
Dryas octopetala									
species pool	0.09	0.02	-0.11	-0.02	-0.11	0.09	-0.08	-0.04	-0.02
repulsion	-0.52	0.39	0.51	-0.21	-0.27	-0.10	0.12	-0.32	0.23
attraction	0.03	-0.03	-0.04	0.02	0.01	-0.01	-0.01	0.01	-0.03
Cscore	0.00	0.05	0.03	-0.03	-0.02	0.05	0.06	-0.08	0.03
dispersal	-0.42	0.11	0.31	-0.19	-0.10	-0.22	0.00	-0.18	0.20

Euphorbia cyparissias									
species pool	0.21	-0.11	-0.26	0.04	-0.09	0.13	-0.10	0.01	-0.06
repulsion	-0.13	0.15	0.18	-0.12	-0.10	0.08	0.12	-0.25	0.15
attraction	0.01	0.07	0.04	0.02	0.02	-0.03	-0.05	0.04	-0.06
Cscore	-0.12	0.10	0.13	-0.08	-0.06	0.02	0.09	-0.11	0.05
dispersal	0.13	-0.29	-0.29	0.11	-0.03	0.03	-0.03	-0.04	0.10
Festuca paniculata									
species pool	-0.27	0.16	0.17	-0.12	-0.30	-0.01	-0.06	-0.18	0.12
repulsion	-0.51	0.38	0.51	-0.24	-0.28	-0.06	0.12	-0.32	0.24
attraction	-0.03	-0.02	0.00	0.00	-0.01	-0.02	-0.04	-0.01	0.03
Cscore	-0.03	0.04	0.04	-0.04	-0.03	0.06	0.08	-0.11	0.06
dispersal	-0.35	-0.06	0.10	-0.01	-0.14	-0.31	-0.07	-0.09	0.15
Geranium sylvaticum									
species pool	0.09	-0.04	-0.15	0.00	-0.14	0.10	-0.09	-0.04	-0.02
repulsion	-0.51	0.32	0.45	-0.17	-0.27	-0.06	0.14	-0.35	0.25
attraction	-0.02	0.18	0.15	-0.09	0.03	-0.03	0.00	-0.04	0.02
Cscore	-0.06	0.01	0.03	0.00	-0.05	0.05	0.11	-0.12	0.06
dispersal	-0.20	0.36	0.37	-0.19	-0.06	-0.21	0.00	-0.17	0.22
Kobresia myosuroides									
species pool	0.09	0.01	-0.12	-0.02	-0.11	0.09	-0.08	-0.04	-0.02
repulsion	**-0.70**	0.37	0.57	-0.19	-0.33	-0.19	0.12	-0.30	0.21
attraction	-0.11	-0.03	0.03	-0.01	-0.11	0.04	-0.02	0.00	0.00
Cscore	0.04	0.05	0.01	-0.02	0.00	0.05	0.06	-0.08	0.05
dispersal	-0.57	0.06	0.33	-0.12	-0.14	-0.28	0.04	-0.13	0.08
Larix decidua									
species pool	0.13	-0.01	-0.14	-0.01	-0.09	0.10	-0.08	-0.03	-0.03
repulsion	-0.34	0.27	0.33	-0.12	-0.20	-0.01	0.12	-0.30	0.19
attraction	-0.15	0.08	0.07	0.09	-0.09	-0.08	0.08	-0.16	0.15
Cscore	-0.08	0.05	0.06	-0.02	-0.06	0.04	0.09	-0.10	0.03
dispersal	-0.20	-0.22	0.03	-0.19	-0.10	-0.16	-0.12	-0.06	0.18
Phragmites australis									
species pool	0.15	-0.03	-0.18	0.01	-0.08	0.11	-0.09	-0.01	-0.05
repulsion	0.37	-0.12	-0.26	0.10	0.11	0.20	0.04	-0.03	0.00
attraction	0.03	-0.03	-0.04	0.01	0.01	0.00	-0.02	0.04	-0.03
Cscore	-0.16	0.11	0.14	-0.05	-0.10	0.02	0.11	-0.17	0.11
dispersal	0.36	-0.16	-0.27	0.08	0.31	0.00	-0.10	0.41	-0.33
Plantago alpina									
species pool	0.08	0.02	-0.10	-0.02	-0.11	0.08	-0.08	-0.05	-0.01
repulsion	**-0.67**	0.43	**0.61**	-0.25	-0.31	-0.22	0.10	-0.37	0.30
attraction	-0.22	0.16	0.19	0.03	-0.09	-0.01	0.10	-0.12	0.06
Cscore	0.13	0.01	-0.04	-0.02	0.05	0.07	0.06	-0.08	0.08
dispersal	-0.57	0.08	0.31	-0.08	-0.35	-0.10	0.01	-0.11	0.06
Polygonum viviparum									
species pool	0.11	0.00	-0.13	-0.01	-0.10	0.09	-0.08	-0.04	-0.03
repulsion	**-0.70**	0.40	0.59	-0.22	-0.32	-0.21	0.14	-0.40	0.29
attraction	-0.05	0.19	0.19	-0.14	0.01	-0.03	0.01	-0.07	0.06
Cscore	0.12	0.02	-0.04	-0.01	0.04	0.08	0.06	-0.08	0.07
dispersal	**-0.67**	0.10	0.43	-0.24	-0.29	-0.23	-0.09	-0.08	0.08
Ranunculus glacialis									
species pool	0.10	-0.01	-0.13	-0.01	-0.11	0.09	-0.08	-0.04	-0.03
repulsion	**-0.71**	0.32	0.54	-0.16	-0.33	-0.19	0.12	-0.27	0.17
attraction	-0.02	0.01	0.01	-0.02	0.00	-0.02	0.02	-0.02	0.01
Cscore	-0.01	0.07	0.04	-0.02	-0.03	0.05	0.07	-0.10	0.06
dispersal	-0.35	0.07	0.21	-0.06	-0.11	-0.12	0.11	-0.12	0.06
Rhododendron ferrugineum									
species pool	-0.21	0.14	0.13	-0.11	-0.27	0.01	-0.06	-0.16	0.11
repulsion	-0.50	0.40	0.51	-0.21	-0.28	-0.05	0.10	-0.31	0.24
attraction	-0.08	0.23	0.22	-0.13	0.01	-0.03	0.03	-0.09	0.07
Cscore	0.00	-0.02	-0.03	0.02	-0.04	0.08	0.07	-0.08	0.02
dispersal	-0.35	0.32	0.44	-0.30	-0.03	-0.45	-0.01	-0.19	0.27
Urtica dioica									
species pool	0.14	-0.02	-0.16	0.00	-0.09	0.10	-0.08	-0.02	-0.04
repulsion	-0.01	0.09	0.05	0.00	-0.05	0.11	0.11	-0.22	0.13
attraction	-0.01	0.05	0.05	-0.06	-0.01	0.01	-0.01	-0.03	0.05
Cscore	-0.14	0.05	0.09	0.00	-0.08	0.02	0.12	-0.13	0.05
dispersal	0.28	0.12	-0.02	-0.05	0.19	-0.03	0.04	0.00	0.00
Vaccinium myrtillus									
species pool	0.13	-0.06	-0.19	0.01	-0.12	0.11	-0.09	-0.02	-0.03
repulsion	-0.51	0.25	0.40	-0.12	-0.29	-0.06	0.09	-0.28	0.18
attraction	0.01	0.01	0.00	0.01	0.02	-0.03	0.04	-0.03	0.01
Cscore	-0.03	-0.07	-0.06	0.06	-0.05	0.07	0.06	-0.06	0.00

dispersal	-0.17	0.47	0.40	-0.16	-0.03	-0.31	0.10	-0.23	0.22

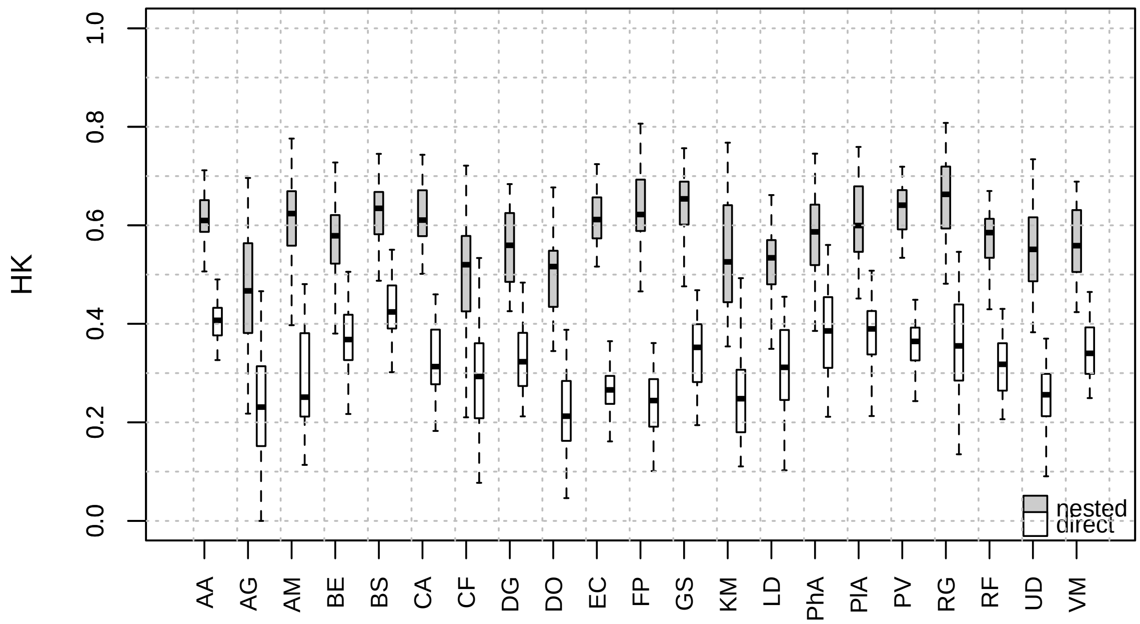
Table S5: Sensitivity to the dispersal distance class. The table shows Pearson correlations between various estimates of the dispersal index for *Geranium sempervirens*. Each time, we estimated the dispersal index using distance parameters from a different dispersal class (1, 3, 4, 5, 6, and 7).

Dispersal class	1	3	4	5	6	7
1	1.00	1.00	0.73	0.84	0.95	0.98
3		1.00	0.73	0.84	0.95	0.98
4			1.00	0.97	0.88	0.77
5				1.00	0.96	0.86
6					1.00	0.95
7						1.00

Fig. S4. Species distributions for the dispersal model. These distributions are the result of the random forest model used to build the dispersal index, for the 21 focal species. They show the spatial configuration of realized species distributions.

4) Effects of the nested modelling method

Fig.S5 Effect of the nested modelling method on model evaluations. The following figures show the variation in model evaluation across four models and nine repetitions, for each modelled species. The following abbreviations are used to name the species: AA=*Abies alba,* AG= *Alnus glutinosa,* AM=*Arnica montana,* BE=*Bromus erectus,* BS=*Buxus sempervirens,* CA=*Cacalia alliariae,* CF=*Carex ferruginae,* DG=*Dactylis glomerata,* DO=*Dryas octopetala,* EC=*Euphorbia cyparissias,* FP=*Festuca paniculata,* GS=*Geranium sempervirens,* KM=*Kobresia myosuroides,* LD=*Larix decidua,* PhA=*Phragmites australis,* PlA=*Plantago alpina,* PV=*Polygonum viviparum,* RG=*Ranunculus glacialis,* RF=*Rhododendron ferrugineum,* UD=*Urtica dioica,* VM=*Vaccinium myrtillus.* The white boxplots are evaluations of direct modelling, which is a direct fit of abundance classes. The grey boxplots are evaluations of nested modelling, where abundance classes are fitted only in sites where presence is predicted. The evaluation method is the Hanssen-Kuipers discriminant (HK), which varies from 0 to 1 for perfect fit.



5) Variable importance

Fig. S6: Detailed variable importance. The following figures show the variable importance for each species, measured as the average change in model accuracy across repetitions when the focal variable is randomized. The segments show confidence intervals (mean+/-1.96*sd) across repetitions. A variable was considered not significant when its importance confidence interval included zero.

Fig. S6a: Variable importance for presence-absence step

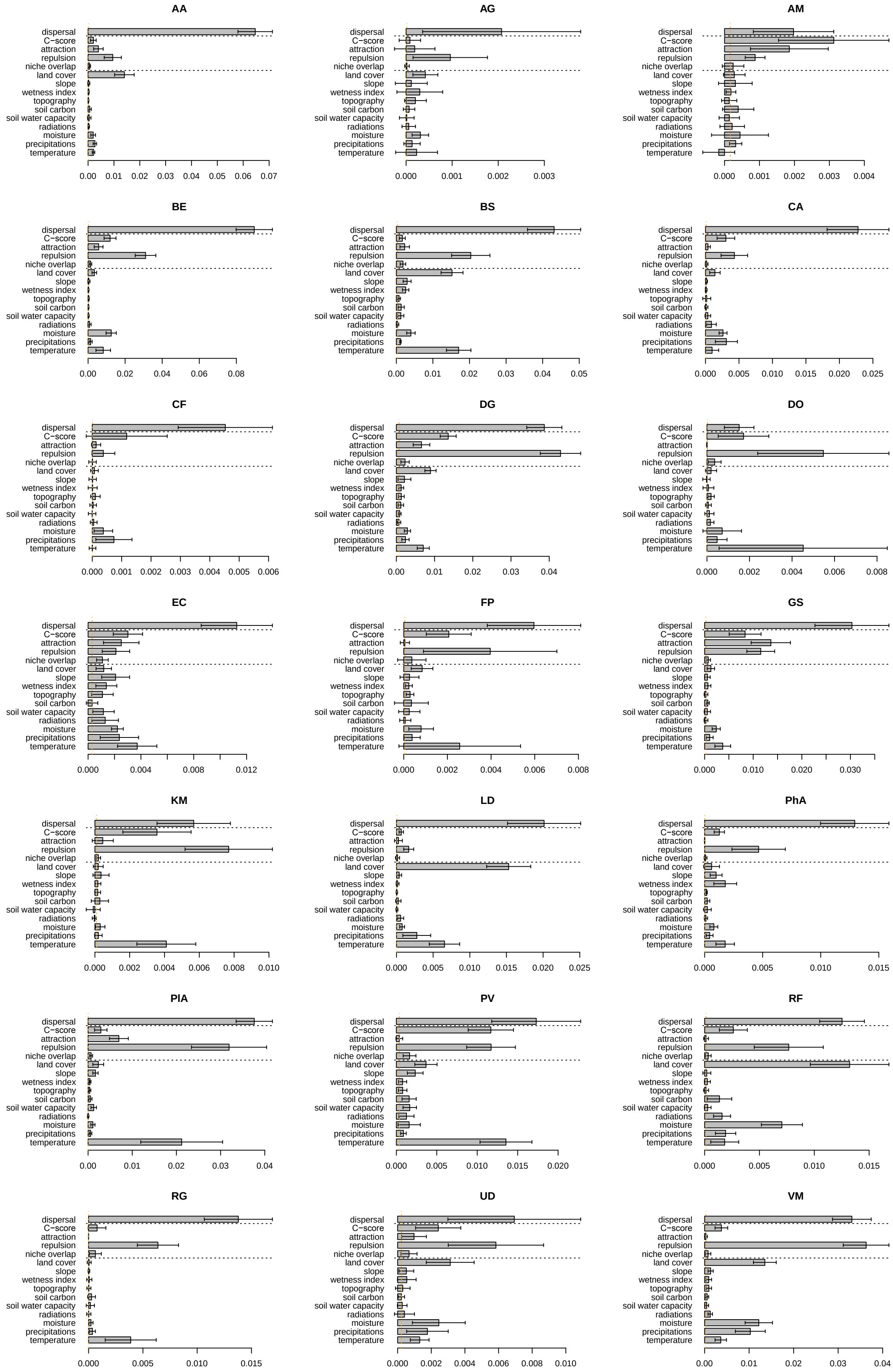
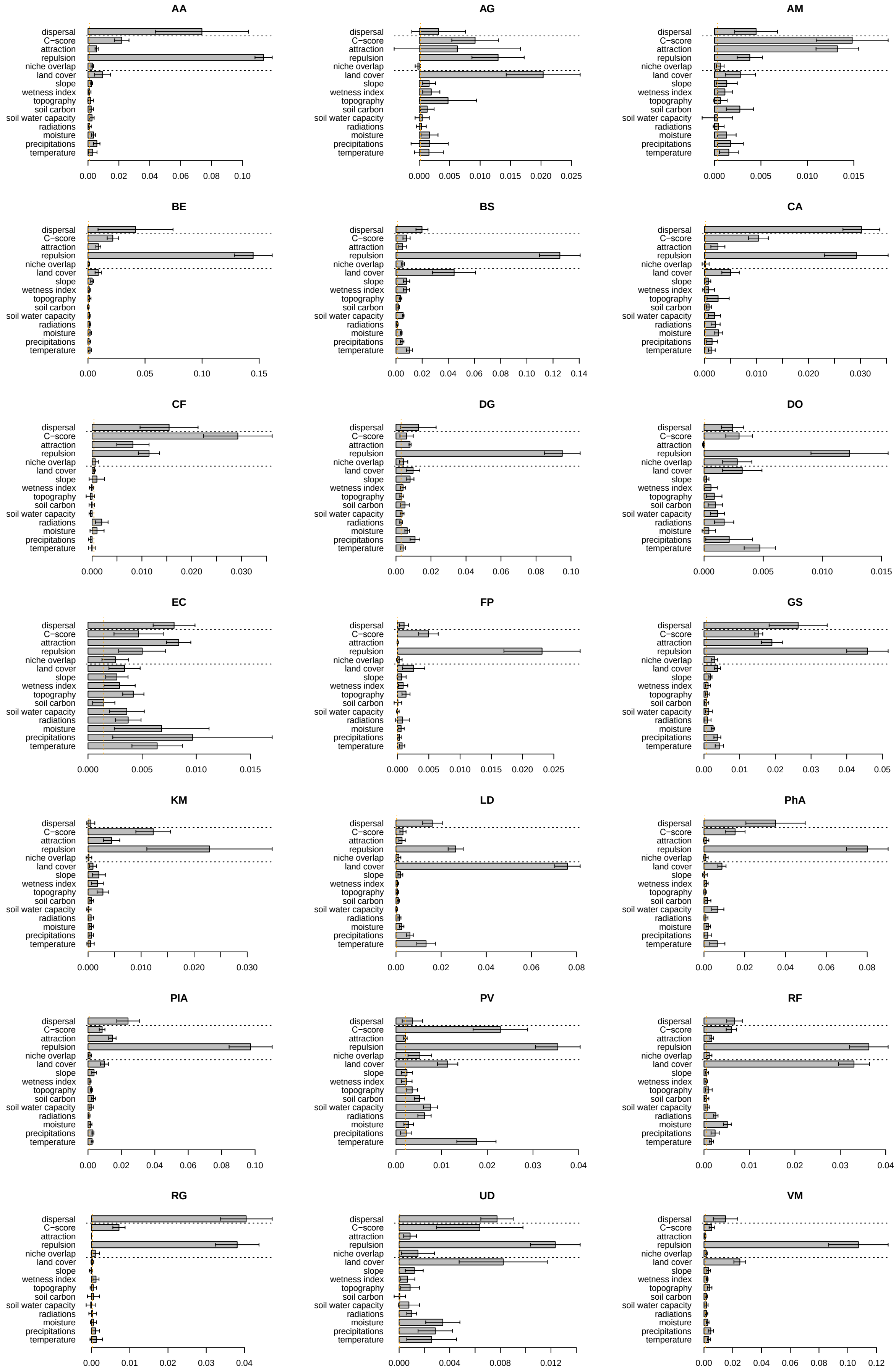


Fig. S6b: Variable importance for abundance variation (2nd modelling step).



6) Effects of the different drivers on the abiotic niche

Fig. S7. Effect of dispersal and biotic interactions on the abiotic niche for 3 other species. A: *Dactylis glomerata*. **B:** *Vaccinium myrtillus*. **C:** *Plantago alpina.* The abiotic niche space is represented by the first two axes (53% of inertia) of a PCA of the abiotic variables. **(a)** Realized niche. Predictions of model ABD are presented in this figure. Left: density of predicted presences normalized by the number of sample plots within each grid cell. Right: third quartile of predicted abundance class within each grid cell. Low: <5% cover; Medium: 5% to 25% cover; High:>25% cover. **(b)** Left/right: Proportion of sources/sinks among predicted presences. Middle: abundances in source and sink plots. **(c)** Effect of biotic interactions. Left: density of predicted presences with co-occurrence indices equalling zero, normalized by the number of sample plots within each grid cell. Right: negative and positive effects of the biotic interactions.

Fig. S7A *Dactylis glomerata*

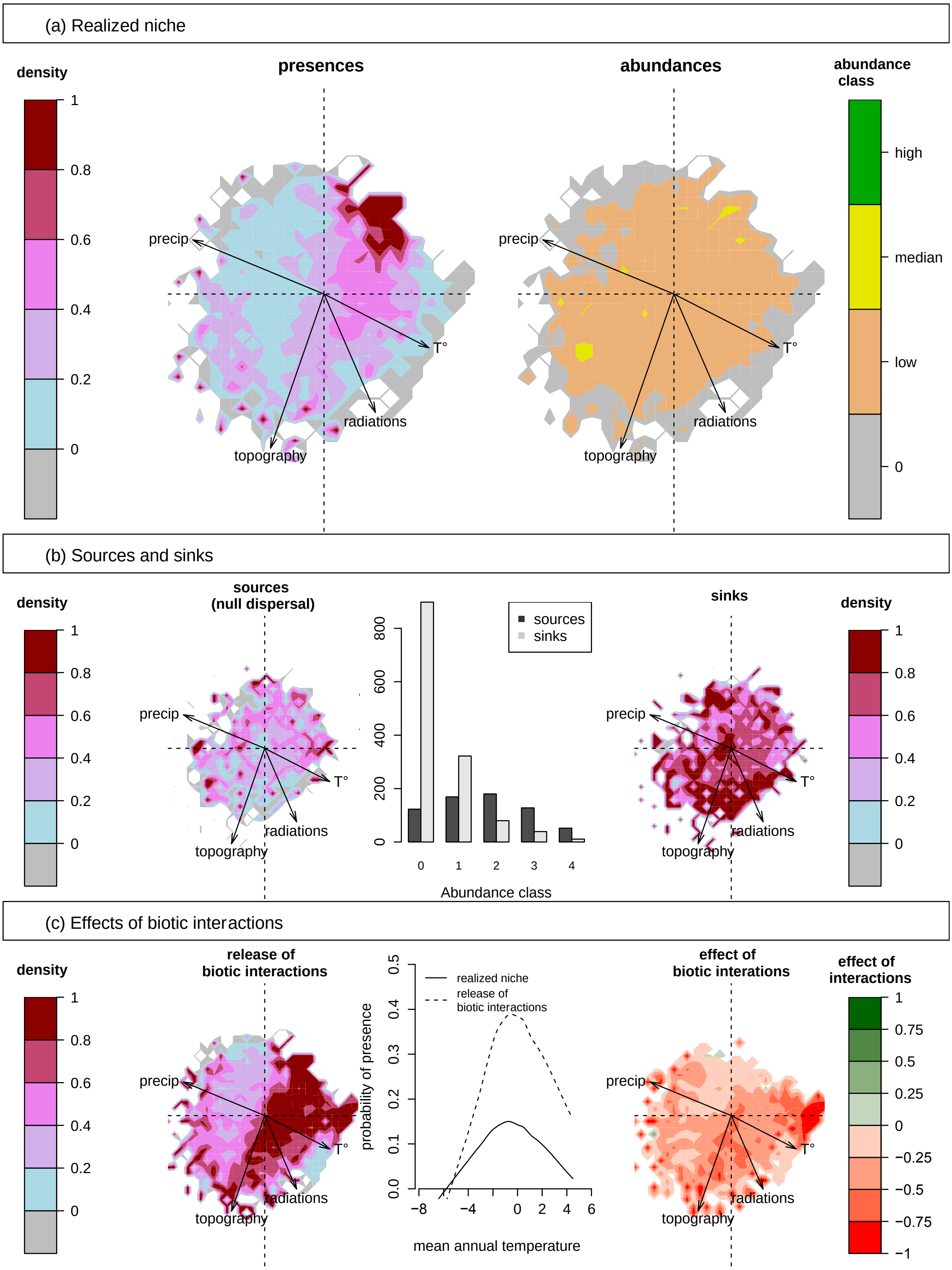


Fig. S7B *Vaccinium myrtillus*

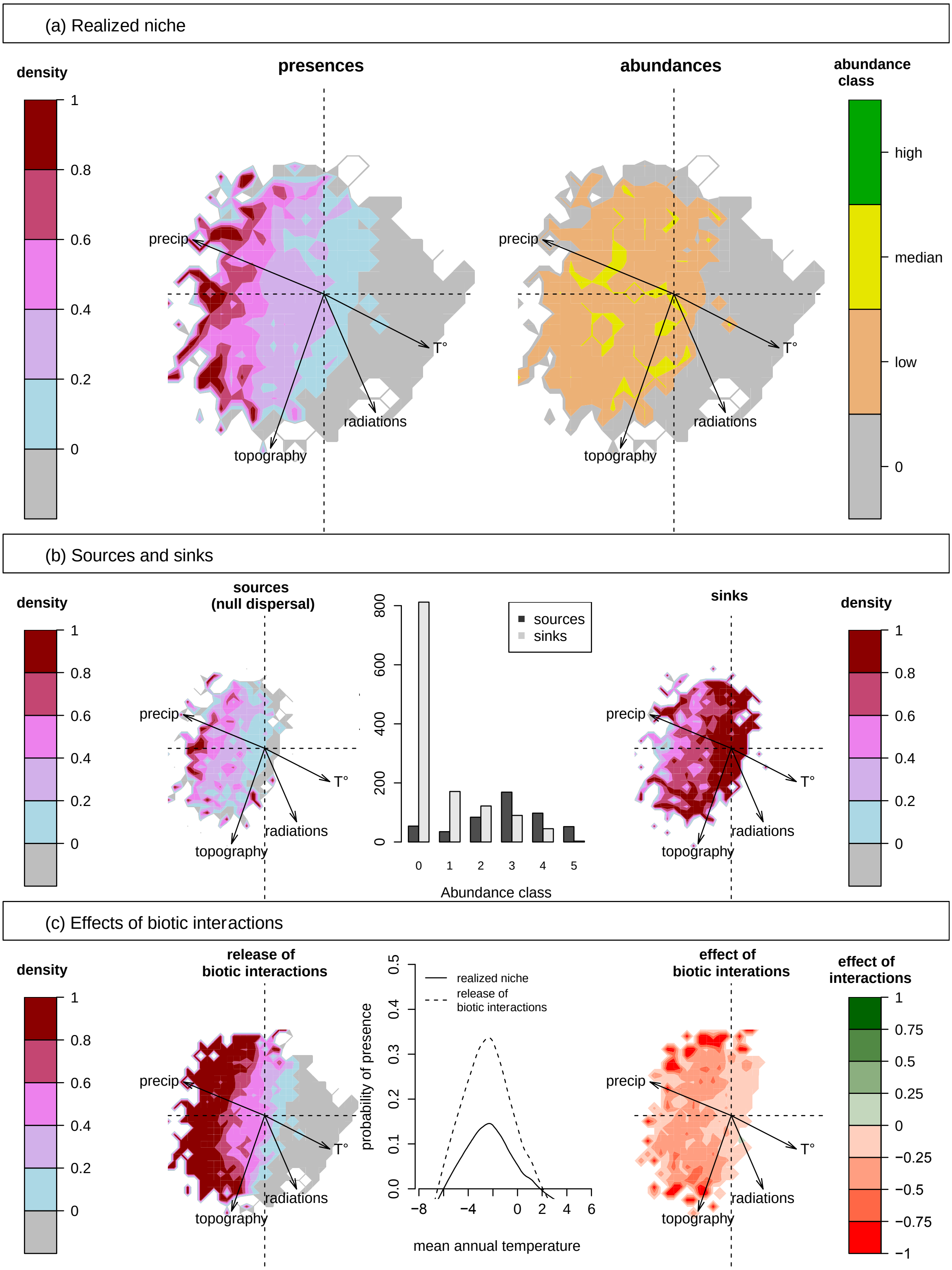


Fig. S7C *Plantago alpina*