

Supplemental Information

Table S1, S2 and S3 provide descriptive statistics for the biodiversity datasets used in this study, including the number of observations, species richness, and spatial distribution for plants, butterflies, and birds. These values highlight the differences in sampling intensity between taxa and justify the use of non-parametric richness estimators to correct for unequal survey effort.

Table S1: Databases used for modeling the diversity of vascular plants (Tracheophyta) (n = 4,279,291).

Database name	Number of observations
PlantNet automatically identified occurrences	477,497
Divers jeux de données floristiques produits par le réseau des botanistes amateurs dans les départements du 06, 11, 13, 30, 34, 48, 66, 83, 84	434,947
Données flores issues de la base de données de Jean-Marc Lewin	419,632
Divers jeux de données floristiques produits par le CBNMED sur son territoire d'agrément	364,440
Relevés floristiques du protocole de l'Inventaire National Forestier de l'IGN	355,155
Divers jeux de données floristiques produits par le CEN du Languedoc-Roussillon dans les départements du 06, 11, 13, 30, 34, 48, 66, 83, 84	158,601
Observations opportunistes Flore, Fonge et Bryophytes des salariés et bénévoles de l'association Nature en Occitanie (ex-Nature en Midi-Pyrénées) en Midi-Pyrénées	130,001
Données d'observations flore du réseau Tela Botanica (hors programmes spécifiques)	129,947
Observations opportunistes Flore, Fonge et Bryophytes du CRPF en Midi-Pyrénées	110,756
Divers jeux de données floristiques produits par la Fédération France Orchidées (FFO)	107,440
Pl@ntNet observations	97,433
Jeux de données floristiques produits par les Réserves Naturelles Catalanes	89,977
Observations opportunistes Flore, Fonge et Bryophytes des salariés et bénévoles du CBNPMP	82,179
Carnet en Ligne	76,200
Divers jeux de données flore en provenance du bureau d'études BIOTOPE dans les départements du 06,11,13,30,34,48,66,83,84	58,106
Inventaire floristique vasculaire et lichens	57,154
Divers jeux de données floristiques produits par la Fédération Aude Claire	52,653
Divers jeux de données issues de la bibliographie dépouillées par le CBNMED	46,915
Inventaire général de la flore vasculaire de Midi-Pyrénées par le CBNPMP	43,574
Observation.org, Nature data from around the World	41,719
Divers jeux de données floristiques produits par l'AMBHHC dans les départements du 06,11,13,30,34,48,66,83,84	41,471
Application INPN Espèces: Observations naturalistes, participatives et opportunistes, fondées sur des photographies	37,694
Observations opportunistes Flore, Fonge et Bryophytes des salariés et bénévoles de l'ANA	37,135

Divers jeux de données floristiques produits par les Ecologistes de l'Euzière dans les départements du 06,11,13,30,34,48,66,83,84	35,863
iNaturalist Research-grade Observations	35,075
Divers jeux de données floristiques produits par la Société d'Horticulture et d'Histoire Naturelle du 34 dans les dép. du 06,11,13,30,34,48,66,83,84	34,031
Observations Flore, Fonge et Bryophytes du CEN Midi-Pyrénées	26,217
Données acquises dans le cadre de l'activité du bureau d'étude BIOTOPE (1994 - 2013)	24,863
Observations Flore, Fonge et Bryophytes de l'ONF en Midi-Pyrénées	24,630
Appel à Projet Connaissances en Occitanie sur les départements 11,30,34,48 et 66	23,824
Observations occasionnelles Parc national des Pyrénées	23,461
Divers jeux de données flore en provenance du bureau d'études ECOMED dans les départements du 06, 11, 13, 30, 34, 48, 66, 83, 84	23,285
Divers jeux de données flore en provenance du bureau d'études Cabinet Barbanson Environnement dans les départements du 06,11,13,30,34,48,66,83,84	21,461
Inventaires réalisés par le CBNPMP pour sa mission d'appui à Natura 2000 en Midi-Pyrénées	20,429
Divers jeux de données flore produits par le CBNMed sur l'ex région LR dans le cadre d'inventaires habitats naturels financés par la DREAL Occitanie	19,119
Contact aléatoire tous règnes confondus _ obs_occ	18,511
Divers jeux de données du CBNMED produites dans le cadre de programmes scientifiques	16,867
Relevés phytosociologiques produits par le CBNMed dans le Gard et l'Hérault dans le cadre du programme CarHab	16,668
Production de données flore dans le cadre de l'animation par le CBNMed du réseau des botanistes amateurs de l'ex région LR	15,525
Divers jeux de données floristiques produits par le Bureau d'études Nymphalis dans les départements du 06, 11, 13, 30, 34, 48, 66, 83, 84	15,006
Divers données flores historiques produites par les anciens CEN LR et Lozères et absentes de la base SICEN Occitanie.	14,650
Relevés de végétations produits par le CBNMed dans le cadre du programme Trame Turquoise	14,039
Contact aléatoire tous règnes confondus	12,688
Inventaires réalisés pour le projet de PNR d'Ariège	12,644
Divers jeux de données floristiques produits par le Parc naturel régional de la Narbonnaise	12,440
Observations Flore, Fonge et Bryophytes du Parc National des Pyrénées	11,886
ONF - Observations opportunistes et inventaires	11,236
Flore - Inventaire non protocolé	11,177
[ABC] - PNP - Inventaire Faune-Flore-Fonge	10,925
Données flore produites par le CBNMed dans l'ex région LR dans la cadre l'inventaire en continu des ZNIEFF	10,051
Données Vigie-Flore	9,455
Données d'observation flore issues du programme Sauvages de ma rue (MNHN, Tela Botanica)	9,299
Recherche et suivi de la flore rare et menacée par le CBNPMP en Occitanie	8,662
Divers jeux de données floristiques produits par le Bureau d'études Naturalia Environnement dans les départements du 06, 11, 13, 30, 34, 48, 66, 83, 84	8,290

Données anciennes produites par le Parc national des Cévennes et saisies par le CBNMed	7,880
Données flores produites par le CBNMed dans cadre du programme Messicoles de Cerdagne	7,714
[ABC] - Amis du PNP - Inventaire Faune-Flore-Fonge	7,394
PRA Euphorbia Peplis porté par le CBNMed dans le cadre de 6 projets de mesures compensatoires	7,049
Données flore produites par le CBNMed dans l'ex région LR dans le cadre de mesure compensatoire	6,975
Divers jeux de données floristiques produits en ex LR par la Société Mycologique et Botanique de Catalogne Nord	6,872
Données d'observation saisies via le projet Pl@ntNet (Tela Botanica)	6,550
Inventaires du CBNPMP dans le cadre de CARHAB	5,867
Divers jeux de données flore sur l'ex région LR en provenance du bureau d'études CRB Environnement	5,803
Divers jeux de données floristiques produits par le monde associatif dans l'ex région LR	5,419
Programmes messicoles dans les départements du 11, 30, 34, 48, 66.	5,280
naturgucker	4,850
Divers jeux de données flore sur l'ex région LR en provenance du Bureau d'études Jean Laurent Hentz	4,786
Inventaires du CBNPMP sur les pelouses sèches en Midi-Pyrénées	4,655
Données naturalistes de PARADIS Anne-Hélène	4,614
Relevés phytosociologiques pour la conception de la typologie et de la cartographie du site Natura 2000 'Causses du Larzac	4,516
Inventaires réalisés par le CBNPMP dans la RNR du Pibeste-Aoulhet	4,504
Données Prune Pellet	4,460
Divers jeux de données du CBNMED produites dans le cadre d'expertises	4,446
Données flore produites par le CBNMed dans l'ex région LR dans le cadre de concours prairies fleuries	4,370
Divers jeux de données flore en provenance de divers Bureaux d'études dans les départements du 06, 11, 13, 30, 34, 48, 66, 83, 84	3,986
Cartographie des Leguminosae (Fabaceae) en France	3,787
Atlas de la Biodiversité Communale sur le territoire du Parc National des Pyrénées - Données du CEN MP	3,715
ENS du Gard - plans de gestion et inventaires (bureaux d'étude)	3,713
Divers jeux de données floristiques produits en ex LR par l'association Barmy Botanists	3,651
Production de données flore dans le cadre du programme européen FLORAPYR EFA 100-15 dans l'Aude et les PO	3,635
Divers jeux de données flore en provenance du Bureau d'études Nicolas Borel Consultant dans les départements du 06,11,13,30,34,48,66,83,84	3,621
Divers jeux de données floristiques produits en ex LR par l'association Lozérienne pour l'Etude et la Protection de l'Environnement	3,613
Inventaires du CBNPMP dans le cadre du Concours Prairies fleuries en Midi-Pyrénées	3,607
Divers jeux de données floristiques produits par l'association Gard Nature dans les	3,590

départements du 06,11,13,30,34,48,66,83,84	
Inventaire et suivis d'espèces dans le Parc national des Cévennes	3,579
Divers jeux de données floristiques produits par la Tour du Valat dans les départements du 06,11,13,30,34,48,66,83,84	3,489
[HORS_PROTOCOL] - Contributions à la connaissance sur la biodiversité du Parc National des Pyrénées	3,427
Divers jeux de données sur les EVEC du CPIE Haut Languedoc	3,364
fundiv france database	3,257
Jeux de données produits par le CBNMed lors de programmes sur les habitats naturels dans l'ex région LR	3,256
Données naturalistes de D'ONOFRIO Patrice	3,244
Jeu de données flores produit dans le cadre du PRA Lagunes temporaires	3,167
Spatially explicit database of tree related microhabitats (TreMs). Version 1.2. Integrate+ project	3,147
Jeu de données personnel de Bruncvik Roman	3,054
Divers jeux de données flore sur l'ex région LR en provenance du bureau d'études CERA Environnement	3,044
Divers jeux de données floristiques produits en ex LR par le Syndicat mixte du bassin de l'Or	3,034
Observations opportunistes Flore, Fonge et Bryophytes des salariés et bénévoles de l'association Nature Comminges en Midi-Pyrénées	2,857
Inventaires réalisées par le CBNPMP dans le cadre de l'appui à la MAET HERB07 en Midi-Pyrénées	2,769
Divers jeux de données floristiques produits par le Bureau d'études ECOSPHERE dans les départements du 06,11,13,30,34,48,66,83,84	2,588
Divers jeux de données floristiques fournis par le CBN Alpin sur le TAG du CBNMed	2,410
Données d'inventaires et suivis des messicoles recueillies par NMP en Midi-Pyrénées	2,398
Divers jeux de données issues de la saisie des herbiers par le CBNMED	2,386
Divers jeux de données floristiques produits par le Syndicat des Hautes Vallées Cévenoles	2,377
PASS_Flore_CORIEAULYS	2,363
Divers jeux de données floristiques produits dans le cadre du programme RESEDA-Flore	2,321
Divers jeux de données flore sur l'ex région LR en provenance du bureau d'études Carex	2,191
Flore SULMONT	2,180
Divers jeux de données floristiques de l'agence nat. de sécurité sanitaire, de l'environnement et du travail - Labo Santé des Vgtx	2,022
Données flore produites dans le cadre de l'ABC Camargue gardoise	2,014
Relevés phytosociologiques produits par le CBNMed pour la conception de la typologie des sites Natura 2000 'Etangs Palavasiens' et 'Etang de l'Or	1,993
Inventaires réalisés par le CBNPMP dans la RNR d'Aulon	1,981
Divers jeux de données flore sur l'ex région LR en provenance du Bureau d'études Ecotone	1,947
Inventaires réalisés par les opérateurs Natura 2000 dans le cadre des DOCOB en Midi-Pyrénées	1,932
Diverses données floristiques issues de structures publiques dans les départements du 06, 11, 13, 30, 34, 48, 66, 83, 84	1,914

Données d'inventaires et suivis des messicoles recueillies par le PNR HL en Midi-Pyrénées	1,888
Données naturalistes de Paul ROUVEYROL	1,883
Inventaire des végétations et habitats péri-aquatiques dans le secteur de plaine de Midi-Pyrénées par le CBNPMP	1,848
Import données Baznat flore	1,799
Inventaires floristiques produits par le CBNMED sur la RN du Bagnas	1,783
Dépouillements bibliographiques réalisées par le CBNPMP	1,767
CS Le Trabet	1,688
Vegetation surveys in mountainous area (Pyrenees, France)	1,554
Divers jeux de données flore sur l'ex région LR en provenance du bureau d'études NATURAE Groupe LAMO	1,471
Observations faunistiques et floristiques - Inventaires 2020	1,451
Flore - Inventaire non standardisé.	1,440
Inventaires Aéro Biodiversité - DCM - Observation opportuniste tous règnes confondus	1,410
Etude prairies fleuries	1,392
Inventaires Aéro Biodiversité - TLS - Observation opportuniste tous règnes confondus	1,375
Données d'inventaires et suivis des messicoles recueillies par le CEN MP en Midi-Pyrénées	1,373
Inventaires Aéro Biodiversité - PGF - Observation opportuniste tous règnes confondus	1,358
Relevés phytosociologiques ciblés sur les végétations aquatiques et riveraines	1,354
Observations d'Ambroisie à feuilles d'armoise (<i>Ambrosia artemisiifolia</i>) et d'Ambroisie trifide (<i>Ambrosia trifida</i>)	1,343
Contact aléatoire tout règnes confondus, jdd accessible par tous	1,332
Inventaires du CBNPMP sur les vieilles forêts pyrénéennes	1,302
Données naturalistes de DUSOULIER François	1,280
ENS Département du Gard - Données opportunistes agents de terrain	1,233
Données flore produites dans le cadre de l'ABC de Pouzols	1,194
Observations NEO (saisies avec l'outil géonature OccTax)	1,185
Recherche de taxons rares ou menacées de plantes vasculaires de la sous-trame milieux humides de Midi-Pyrénées	1,185
Système d'évaluation de l'état des Eaux (SEEE) - Macrophytes	1,177
Observations de plantes exotiques envahissantes réalisées par le CBNPMP	1,165
Flore - Inventaire non standardisé	1,139
Divers jeux de données floristiques produits par l'Office national des forêts dans les départements du 06, 11, 13, 30, 34, 48, 66, 83, 84	1,111
Divers jeux de données floristiques produits par le Syndicat mixte de la Camargue gardoise	1,094
ABC Argeles-Gazost	1,089
Divers jeux de données flore en provenance du bureau d'études MTDA dans les départements du 06,11,13,30,34,48,66,83,84	1,079
Divers jeux de données floristiques produits en ex LR par la Société d'Etude des Sciences de l'Aude	1,019

Table S2: Databases used for modeling the diversity of butterflies (Rhopalocera) (n = 1,106,770).

Database name	Number of observations
Données CEN L-R Papillons/Libellule	310,072
Données opportunistes du CEN Midi-pyrenees	186,760
Versement collectif Faune Occitanie 2020 amphibiens, reptiles, mammifères	119,679
Données des Lépidoptères de France issues du Portail Artemisiae	106,594
Import données Baznat invert	70,212
Observation.org, Nature Data from around the world	61,860
Données de l'Atlas écologique régional des papillons de jour et zygènes	46,559
Suivi Temporel des Rhopalocères de France (STERF)	22,862
INaturalist Reasearch-grade Observations	22,038
données EE pour l'atlas biodiv'occitanie	19,191
Données Observatoire des Papillons des Jardins	18,941
Données opportunistes du CEN Languedoc-Roussilon	13,854
Suivi Photographique des Insectes Pollinisateurs (SPIPOLL)	7,407
Application INPN Espèces: Observations naturalistes, participatives et opportunistes, fondées sur des photographies	7,207
Données acquises dans le cadre de l'activité du bureau d'étude BIOTOPE (1994 - 2013)	6,535
Données opportunistes du CEN Occitanie	6,412
Données invertébrés produites par l'OPIE	6,073
Observations occasionnelles Parc national des Pyrénées	5,217
Données ANA - CEN Ariège sur diverses études	5,024
Données invertébrés OPIE dans le cadre de la participation au réseau N2000	4,905
ONF - Observations opportunistes et inventaires	4,304
Contact aléatoire tout règnes confondus, jdd accessible par tous	3,999
Atlas de la Biodiversité Communale sur le territoire du Parc National des Pyrénées - Données du CEN MP	3,880
Revue oreina [2008-]	3,878
Données photographiques validées de la galerie du forum 'Le Monde des Insectes	3,726
Données zones humides CEN M-P	3,649
Données CEN L-R départements	3,505
Observations NEO (saisies avec l'outil géonature OccTax)	3,329
Données CEN M-P Milieux secs	3,063
Données ANA - CEN Ariège opportunistes	2,936
Insectes - Inventaire non protocolé	2,923
Données Protocole Papillons Gestionnaires (PROPAGE)	2,368

observado	2,189
Jeu de données personnel de Lock Jude	2,117
Données CEN M-P appui réserves naturelles	2,100
Données invertébrés dans le cadre de la participation au réseau N2000	1,984
Jeu de données personnel de Urrustoy Etienne	1,417
Données CEN M-P - EDF	1,306
Données appui collectivités territoriales CEN L-R	1,232
Données invertébrés sites CEN MP	1,227
Données invertébrés sites CEN-LR	1,110
Natural History Museum Rotterdam - Specimens	1,106
Données OPIE partenariats PNR	1,011
Données OPIE Compensation collectivités	1,009

Table S3: Databases used for modeling bird diversity (Aves) (n = 4,765,000)

Database name	Number of observations
Données opportunistes OISEAUX de Faune-Occitanie	3,333,951
Données Oiseaux des Jardins	480,715
Données STOC-EPS	235,438
EOD - eBird Observation Dataset	214,361
Import données Baznat oiseaux	150,436
Observation.org, Nature data from around the World	52,356
Données SHOC	29,577
Données Birdlab	23,570
Données Oiseaux dans le cadre de la participation au réseau N2000	20,429
Données oiseaux dans le cadre des politiques publiques de départements	15,878
Oiseaux - Inventaire non protocolé	14,384
Données naturalistes de SWIFT Olivier	13,965
Données Oiseaux issues du programme régional de gestion de la sous-trame Milieux ouverts et semi-ouverts dans sa composante « semi-naturelle »	13,391
Données acquises dans le cadre de l'activité du bureau d'étude BIOTOPE (1994 - 2013)	12,859
Données Oiseaux issues de la gestion de site ou d'appui à la gestion de site	10,977
iNaturalist Research-grade Observations	10,881
Observations occasionnelles Parc national des Pyrénées	10,080
Données Oiseaux issues des mesures compensatoires de projets portés par les collectivités territoriales	8,781
Contact aléatoire tout règnes confondus, jdd accessible par tous	7,576
ONF - Observations opportunistes et inventaires	7,191
Données d'observations 'espèce' réalisées par des agents de l'OFB jusqu'au 31 décembre	7,046

2018.	
Données Oiseaux issues de l'appui au Parcs Naturels Régionaux	6,114
Application INPN Espèces: Observations naturalistes, participatives et opportunistes, fondées sur des photographies	6,070
Observations NEO (saisies avec l'outil géonature OccTax)	5,804
[HORS_PROTOCOL] - Contributions à la connaissance sur la biodiversité du Parc National des Pyrénées	5,762
Inventaire et suivis d'espèces dans le Parc national des Cévennes	5,570
Données Oiseaux de Nîmes Métropole versées au SINP	5,460
Données Oiseaux issues de la mise en oeuvre des suivis compensatoires de propriétaires privés ou partenariat public-privé	5,030
Xeno-canto - Bird sounds from around the world	4,372
naturgucker	3,781
Données Oiseaux dans le cadre de l'appui au réserves naturelles	3,481
JDD faune - Projet TEREKA REVA - Renouveau Villariès Albi	3,302
[OISEAUX] - STOC	3,199
[ABC] - PNP - Inventaire Faune-Flore-Fonge	2,180
Plan National d'Actions Ourde canepetière	2,105
Visiolittoral : surveillance naturaliste des sites du Conservatoire du littoral	1,942
Données Oiseaux issues des mesures compensatoires routes départementales	1,714
Suivi post implantation Truc de l'Homme 2018-2019	1,613
Programme FLASH : Répartition Merle noir 2012 - 2016 Coordonnées GPS Lambert 93	1,539
Jeu de données personnel de Bergès Christophe	1,530
Données Oiseaux issues des mesures compensatoires voies ferrées	1,521
Données naturalistes de Benoit FONTAINE	1,400
Oiseaux - Inventaire protocolé	1,357
Suivi LIFE+ Envoll	1,354
Données naturalistes de Julien BIRARD	1,298
Programme FLASH : Répartition Etourneau sansonnet 2012 - 2016 Coordonnées GPS Lambert 93	1,247
Données brutes faune flore avifaune chiroptères	1,233
ABC Argeles-Gazost	1,214
[OISEAUX] - Oiseaux alpins communs - Suivi	1,211
Programme FLASH : Répartition Grive draine 2012 - 2016 Coordonnées GPS Lambert 93	1,195
Données brutes de biodiversité	1,135
STMH_données opportunistes	1,125
Programme FLASH : Répartition Tourterelle turque 2012 - 2016 Coordonnées GPS Lambert 93	1,107
Données oiseaux issues d'études d'impacts	1,079
Données brutes Avifaune - Ferme éolienne du Puech	1,042

Table S4: Moran indices and robust Lagrange multipliers (SAR, SEM) for Tracheophyta (vascular plants), Rhopalocera (butterflies), Aves (birds) for the regional models with overall geodiversity. The lines highlighted in yellow correspond to the selected model (lowest AIC). The rows highlighted in yellow correspond to the models with the lowest Akaike Information Criterion (AIC) values, meaning they offer the best trade-off between explanatory power and model parsimony. Selecting these models ensures that the biodiversity–geodiversity relationships are interpreted from statistically robust frameworks that account for spatial dependence while avoiding overfitting.

Taxon	Scale (km)	Test	Statistical	AIC	p-value
<i>Tracheophyta</i>	5	Moran I	0.576		0.00E+00
		Lagrange robuste SAR	27.942	18248.31	1.25E-07
		Lagrange robuste SEM	252.407	18124.78	0.00E+00
	7.5	Moran I	0.577		0.00E+00
		Lagrange robuste SAR	34.181	12238.70	5.01E-09
		Lagrange robuste SEM	107.428	12174.37	0.00E+00
	10	Moran I	0.572		0.00E+00
		Lagrange robuste SAR	40.368	7755.65	2.10E-10
		Lagrange robuste SEM	52.310	7714.31	4.73E-13
	5	Moran I	0.343		0.00E+00
		Lagrange robuste SAR	39.667	14901.82	3.01E-10
		Lagrange robuste SEM	28.524	14941.46	9.25E-08
<i>Rhopalocera</i>	7.5	Moran I	0.417		0.00E+00
		Lagrange robuste SAR	36.930	8322.68	1.22E-09
		Lagrange robuste SEM	18.076	8334.92	2.12E-05
	10	Moran I	0.435		0.00E+00

<i>Aves</i>	5	Lagrange robuste SAR	47.129	4941.58	6.64E-12
		Lagrange robuste SEM	30.458	4953.54	3.411E-08
		Moran I	0.491		0.00E+00
	7.5	Lagrange robuste SAR	5.769	23338.66	0.016307
		Lagrange robuste SEM	32.102	23244.23	1.462E-08
		Moran I	0.501		0.00E+00
	10	Lagrange robuste SAR	3.219	10447.40	0.0727
		Lagrange robuste SEM	32.101	10387.37	1.463E-08
		Moran I	0.502		0.00E+00
	10	Lagrange robuste SAR	3.404	5749.96	0.06503
		Lagrange robuste SEM	20.565	5728.77	5.76E-06

Table S5. Moran indices and robust Lagrange multipliers (SAR, SEM) for Tracheophyta (vascular plants), Rhopalocera (butterflies), Aves (birds) for the regional models with **the four components of geodiversity**. The lines highlighted in yellow correspond to the selected model (lowest AIC).

Taxon	Scale (km)	Test	Statistical	AIC	p-value
<i>Tracheophyta</i>	5	Moran I	0.576		0.00E+00
		Lagrange robuste SAR	28.420	18233.07	9.763E-08
		Lagrange robuste SEM	248.412	18114.87	0.00E+00
	7.5	Moran I	0.577		0.00E+00
		Lagrange robuste SAR	33.171	12230.58	8.40E-089
		Lagrange robuste SEM	110.740	12172.73	0.00E+00
	10	Moran I	0.572		0.00E+00
		Lagrange robuste SAR	42.987	7738.51	5.50E-11
		Lagrange robuste SEM	56.246	7703.21	6.39E-14

		SEM			
<i>Rhopalocera</i>	5	Moran I		0.338	0.00E+00
		Lagrange SAR	robuste	42.998	14897.57
		Lagrange SEM	robuste	25.896	14939.95
	7.5	Moran I		0.404	0.00E+00
		Lagrange SAR	robuste	45.533	8321.45
		Lagrange SEM	robuste	13.891	8338.24
	10	Moran I		0.418	0.00E+00
		Lagrange SAR	robuste	53.364	4938.19
		Lagrange SEM	robuste	25.450	4956.81
<i>Aves</i>	5	Moran I		0.456	0.00E+00
		Lagrange SAR	robuste	56.317	23160.20
		Lagrange SEM	robuste	63.499	23117.32
	7.5	Moran I		0.482	0.00E+00
		Lagrange SAR	robuste	28.373	10342.27
		Lagrange SEM	robuste	46.893	10293.38
	10	Moran I		0.496	0.00E+00
		Lagrange SAR	robuste	15.556	5680.47
		Lagrange SEM	robuste	32.167	5648.38

Table S6: Comparisons of the quality of fit of the spatial models (SAR, SEM) and the linear model for Tracheophyta (land plants), Rhopalocera (moth) and Aves (birds) in the model with all the components of geodiversity.

For each taxon and scale, spatial models consistently achieve much higher adjusted R^2 values than linear models, showing that explicitly accounting for spatial dependence greatly improves explanatory power. The Akaike Information Criterion (AIC) values are also markedly lower for spatial models, confirming their superior fit while avoiding overfitting. For vascular plants (Tracheophyta), SEM models capture over 80% of the variance at 10 km scale, compared to only 42% for linear models, illustrating the robustness gained by correcting for spatial structure. For butterflies (Rhopalocera), SAR models perform best, with

adjusted R^2 increasing from 0.29 (lm) to 0.54 (SAR) at 10 km, suggesting that species distributions are strongly spatially autocorrelated. For birds (Aves), SEM models provide the best fit, improving adjusted R^2 values by a factor of 3–4 relative to linear models. These results demonstrate that failing to account for spatial autocorrelation would lead to substantial underestimation of geodiversity–biodiversity relationships and highlight the necessity of using spatially explicit approaches in such analyses.

Taxon	Scale (km)	Test	Coefficient autoregression	p-value	Adjusted R^2	Adjusted R^2 (lm)	AIC	AIC (lm)
<i>Tracheophyta</i>	5	SEM	0.765	<0.001	0.663	0.204	18114.871	18978.264
	7.5		0.869	<0.001	0.774	0.346	12172.733	12910.690
	10		0.894	<0.001	0.825	0.424	7703.217	8227.000
<i>Rhopalocera</i>	5	SAR	0.523	<0.001	0.481	0.288	14897.577	15307.878
	7.5		0.663	<0.001	0.661	0.449	8321.450	8705.031
	10		0.707	<0.001	0.746	0.535	4938.194	5237.043
<i>Aves</i>	5	SEM	0.785	<0.001	0.590	0.139	23117.328	24384.374
	7.5		0.841	<0.001	0.648	0.159	10293.382	10978.452
	10		0.861	<0.001	0.682	0.175	5648.383	6057.333

Fig. S1 : Schematic representation of the principles of geographically weighted regression, modified by Feuillet et al (2019)

In GWR, the relationship between variables is estimated locally by applying a spatial weighting function that decreases with distance. The weight w_{ij} signed to each observation j depends on its distance d_{ij} from the focal location i , scaled by a neighborhood window or bandwidth h . As shown in the figure, nearby points contribute more strongly to the local regression than distant points, which receive lower weights. This approach allows model coefficients to vary across space, thereby capturing non-stationary relationships that cannot be identified with global models. In the context of our study, GWR enables us to detect spatial heterogeneity in geodiversity–biodiversity associations, highlighting areas where these relationships are stronger, weaker, or even negative.

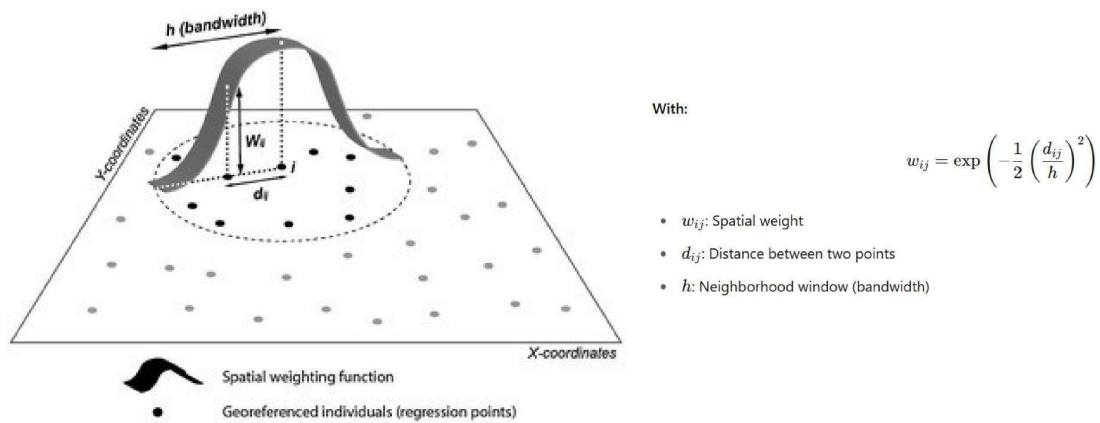
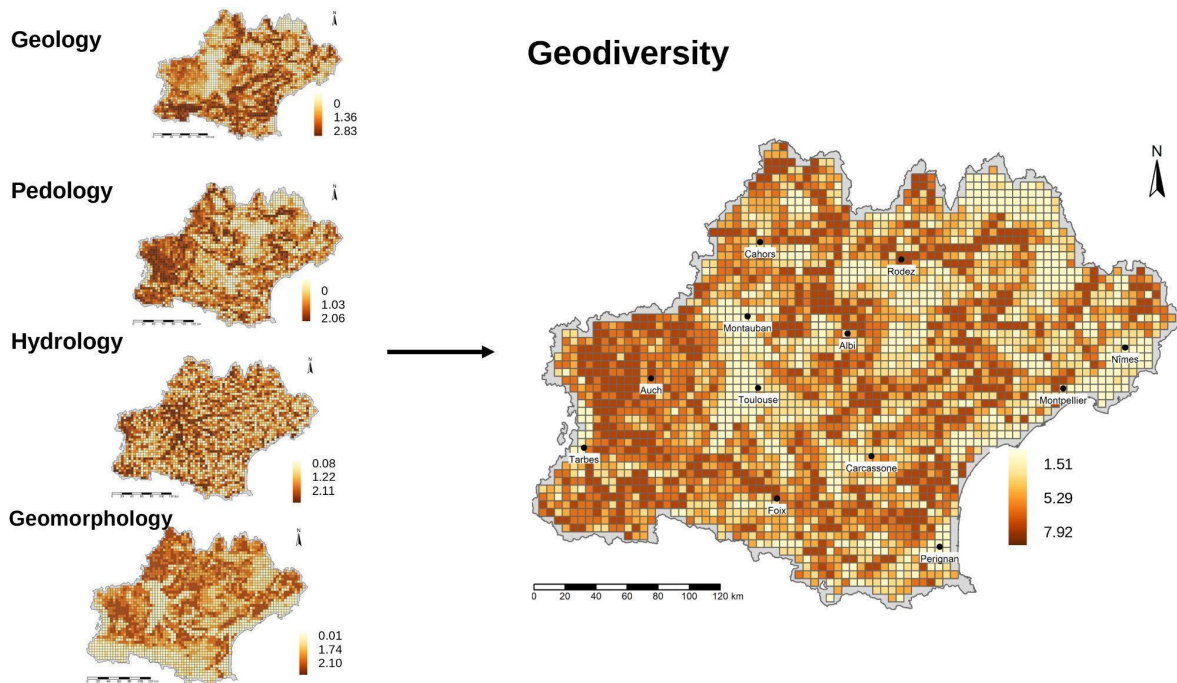


Fig. S2. Maps of geological, pedological, hydrological, and geomorphological diversity and resulting overall geodiversity at the three spatial scales considered: 5 (A), 7.5 (B), and 10 km (C) grid resolution.

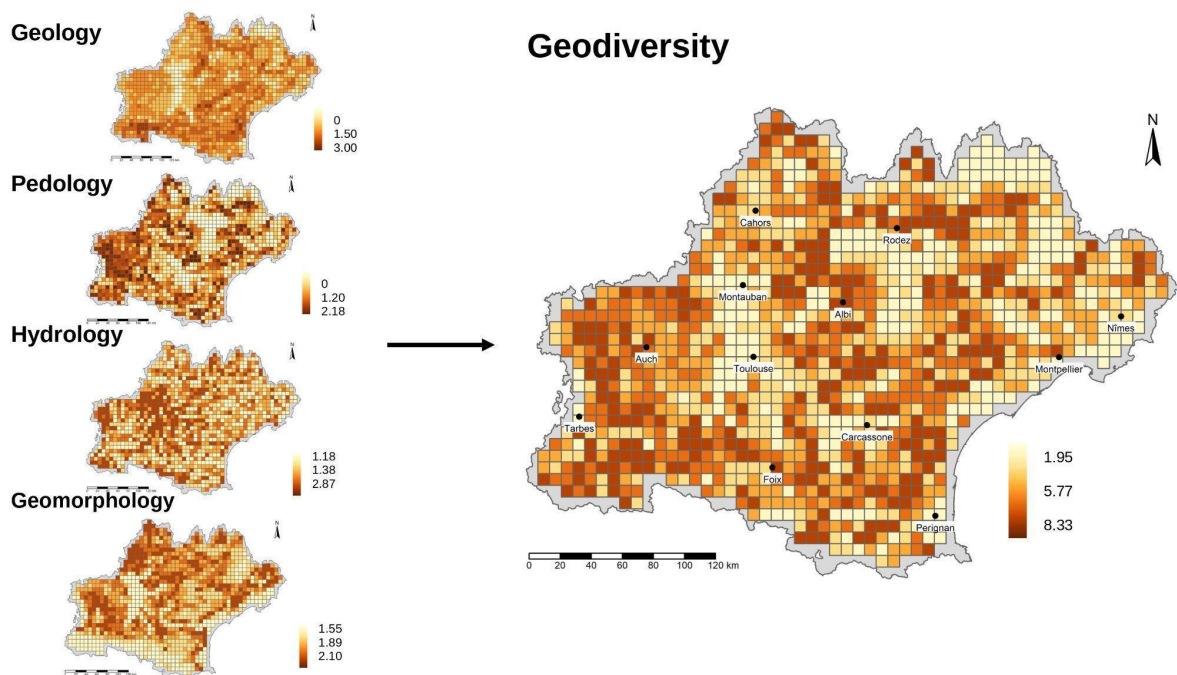
At finer resolution (5 km), local heterogeneity is more visible, with strong contrasts between mountainous and lowland areas. As grid size increases (7.5 km and 10 km), patterns become smoother and some local variations are diluted, but broad-scale gradients remain consistent. Across all resolutions, higher geodiversity values are concentrated in mountainous areas: in the north, near the Massif Central around Rodez and Cahors; and in the south, along the Pyrenees near Tarbes and Perpignan. By contrast, lower geodiversity values dominate coastal and lowland regions, particularly in the east near Montpellier, Nîmes, and Narbonne, and in the west near Auch and Toulouse. The central part of the region, around Albi and Carcassonne, shows intermediate values, reflecting its role as a transition zone between mountains and plains. These maps highlight both the robustness of overall spatial patterns

and the influence of resolution on the detection of fine-scale variability, which is central to the Modifiable Areal Unit Problem (MAUP) discussed in the manuscript

A- Grid of 5 km



D- Grid of 7.5 km



C- Grid of 10 km

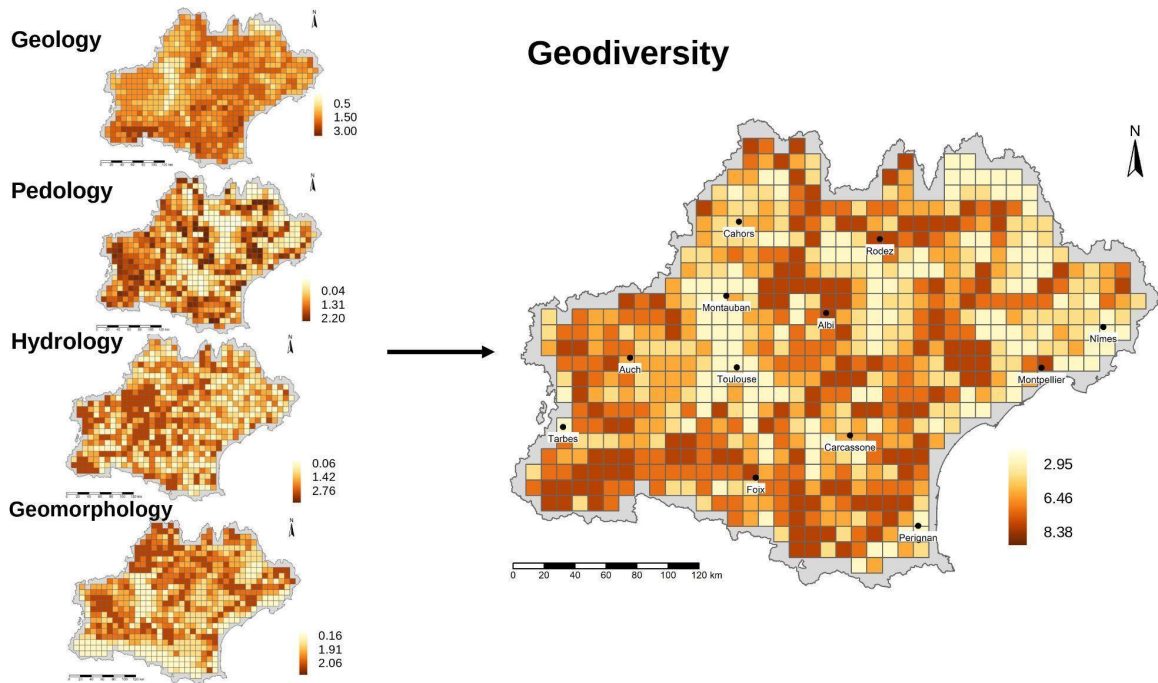
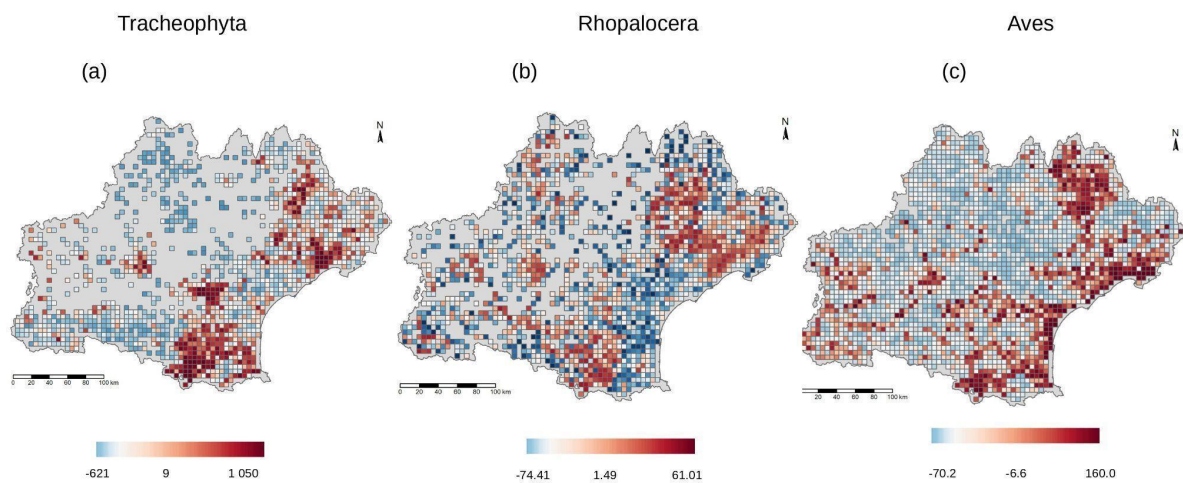


Fig.S3 Maps of the residuals of the linear model for the three taxa at the three spatial scales considered: 5 (A), 7.5 (B), and 10 km (C) grid resolution.

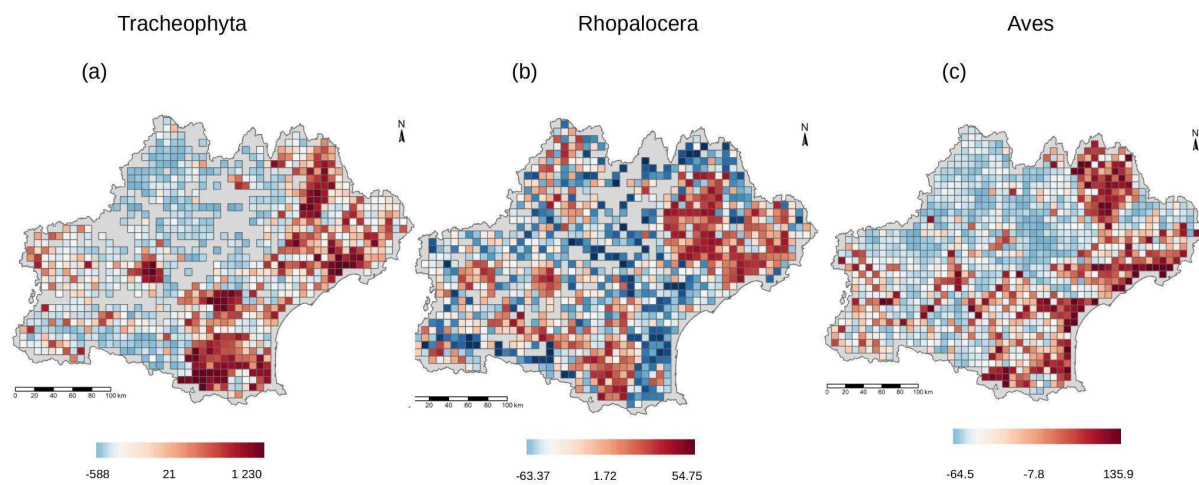
In statistical modeling, residuals correspond to the difference between observed values of species richness and the values predicted by the model. Positive residuals (red cells) indicate that the model underestimates species richness in those locations, while negative residuals (blue cells) indicate overestimation. Grey cells with mesh represent non-significant relationships, while fully grey areas correspond to excluded cells due to insufficient biodiversity exhaustivity (<50% singletons, see Methods 2.2).

At all scales, residuals are not randomly distributed but show strong spatial clustering. In particular, mountainous regions such as the Pyrenees (south, near Tarbes and Perpignan) and the Massif Central (north, near Rodez and Cahors) tend to have large positive residuals, showing that linear models systematically underestimate biodiversity in these areas. Conversely, lowland and coastal areas, especially around Montpellier, Nîmes, and Toulouse, often display negative residuals, suggesting overestimation of richness. The persistence of such structured residuals reveals that linear models fail to capture spatial processes underlying biodiversity patterns. This provides a strong justification for the use of spatially explicit models (SAR/SEM), which are better suited to account for spatial autocorrelation and improve explanatory power.

A- Grid of 5 km



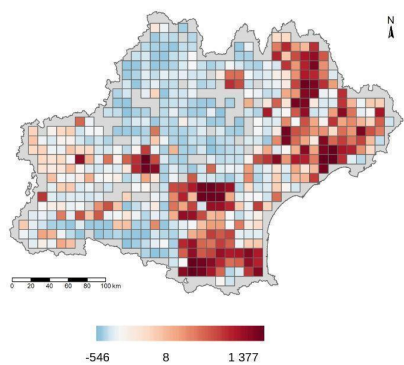
B- Grid of 7.5 km



C- Grid of 10 km

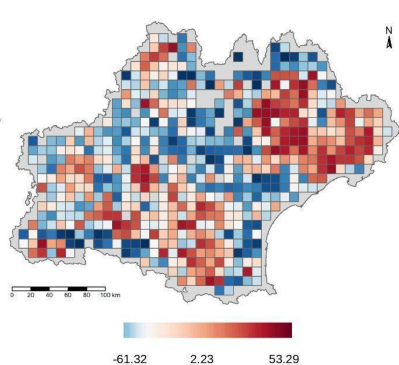
Tracheophyta

(a)



Rhopalocera

(b)



Aves

(c)

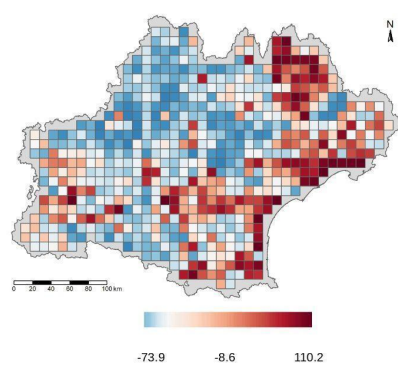
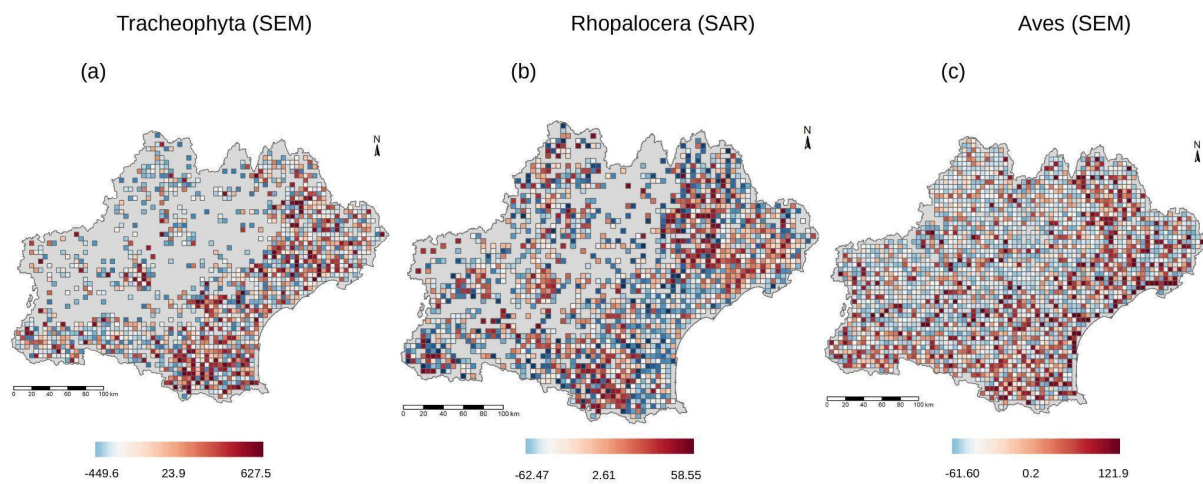


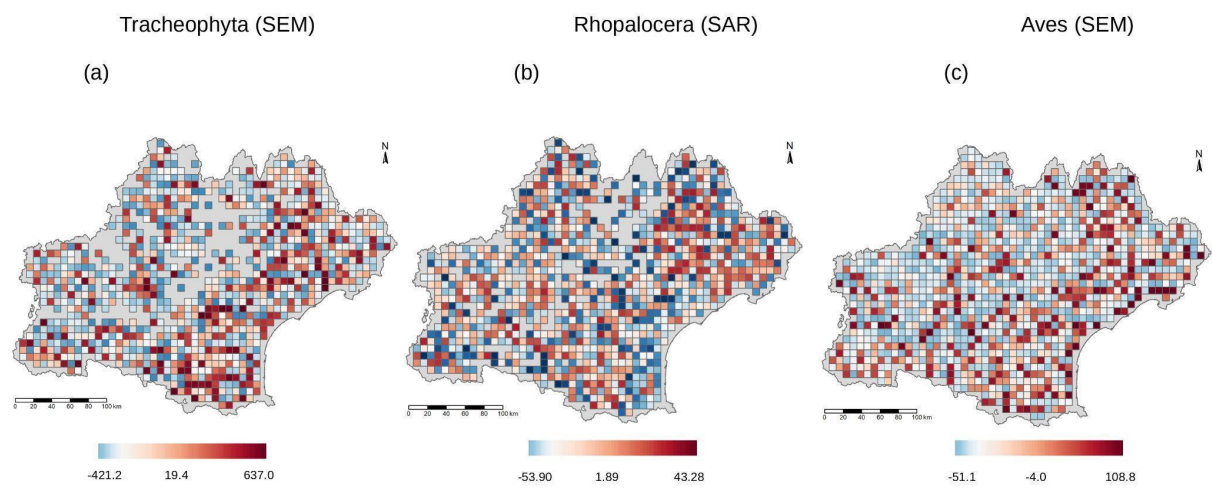
Fig. S4: Maps of the residuals of the best spatially corrected model (SAR or SEM) for the three taxa at the three spatial scales considered: 5 (A), 7.5 (B), and 10 km (C) grid resolution. As in Figure S3, residuals represent the difference between observed and predicted species richness: red cells correspond to underestimation, and blue cells to overestimation by the model.

In contrast with the linear models (Fig. S3), residuals from the spatial models show no clear clustering or strong spatial patterns. Instead, red and blue cells are scattered in a seemingly random manner across the study area, with no major geographic concentration in mountains, lowlands, or coastal regions. This more homogeneous distribution indicates that spatial autocorrelation has been effectively corrected for, meaning that SAR and SEM models successfully account for spatial dependencies in the data. The absence of residual clustering confirms that the spatially explicit models provide a more robust and unbiased framework for analyzing biodiversity–geodiversity relationships across taxa and scales.

A- Grid of 5 km



B- Grid of 7.5 km



B- Grid of 10 km

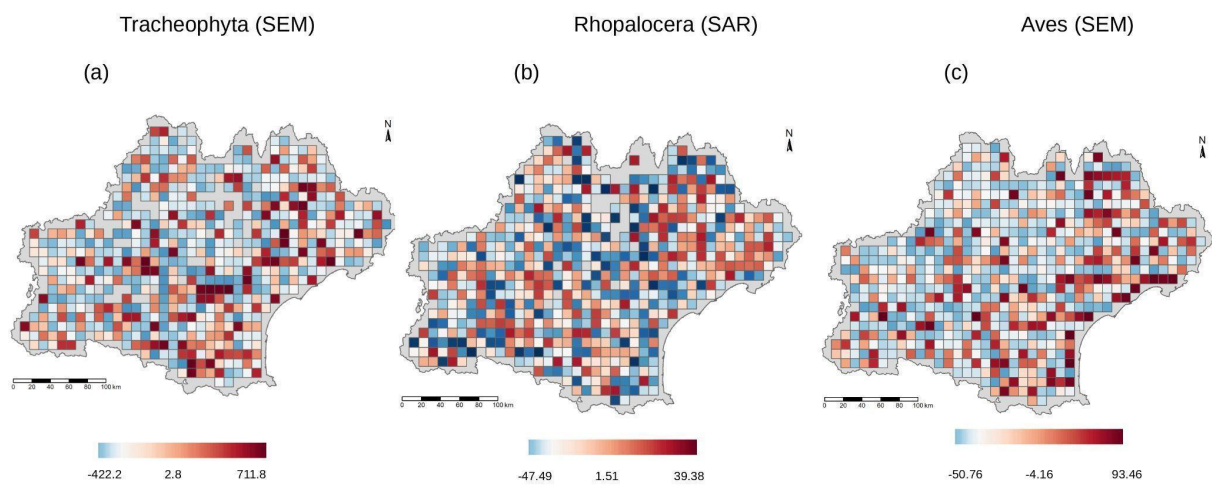


Figure S5 presents the spatial variation of abiotic coefficients estimated with geographically weighted regression (GWR) for vascular plants (A), butterflies (B), and birds (C) at the three spatial resolutions considered (5 km, 7.5 km, and 10 km). In GWR, coefficients are calculated locally rather than globally, allowing relationships between biodiversity and abiotic predictors to vary across space. Red clusters indicate areas where the predictor has a strong positive effect on species richness, while blue clusters highlight negative effects.

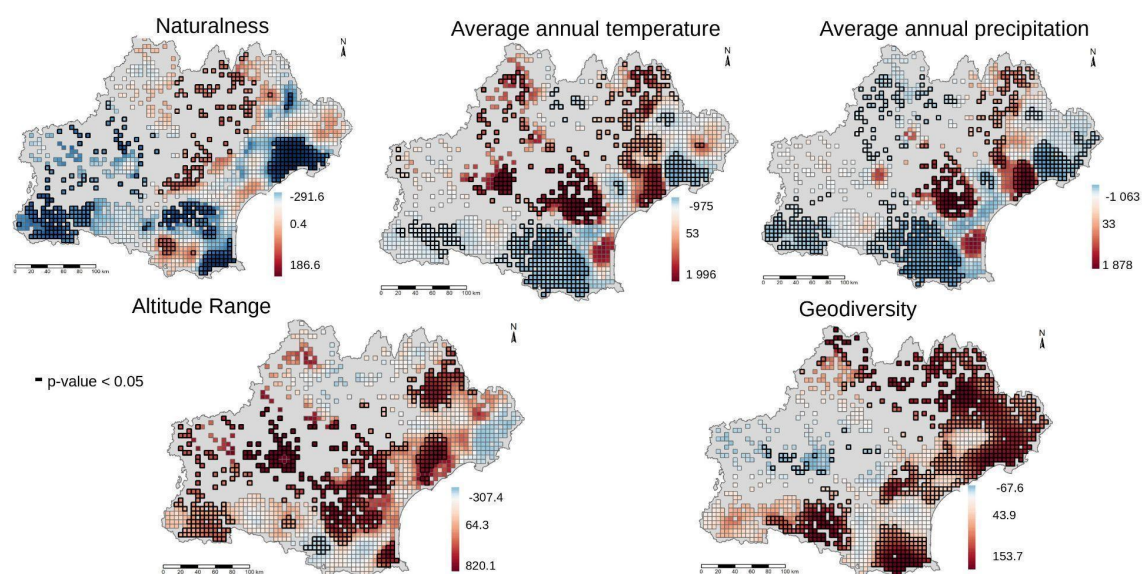
Across all taxa and resolutions, clear spatial clusters emerge, showing that the influence of abiotic variables such as geodiversity, naturalness, climate, and altitude range is not uniform across the region. For example, mountainous areas (Pyrenees and Massif Central) frequently display strong positive effects, while lowland or coastal regions often show weaker or even negative associations. These patterns illustrate the phenomenon of spatial non-stationarity, meaning that the strength and direction of biodiversity–environment relationships depend heavily on local ecological and anthropogenic conditions.

Importantly, the persistence of both red and blue clusters across different taxa and scales indicates that geodiversity and other abiotic drivers can act as biodiversity promoters in some contexts but may have negligible or adverse effects in others. This highlights the need for context-specific conservation strategies: extrapolating global effects would mask important local nuances, whereas GWR provides a tool to identify geographic areas where geodiversity-based conservation is most relevant.

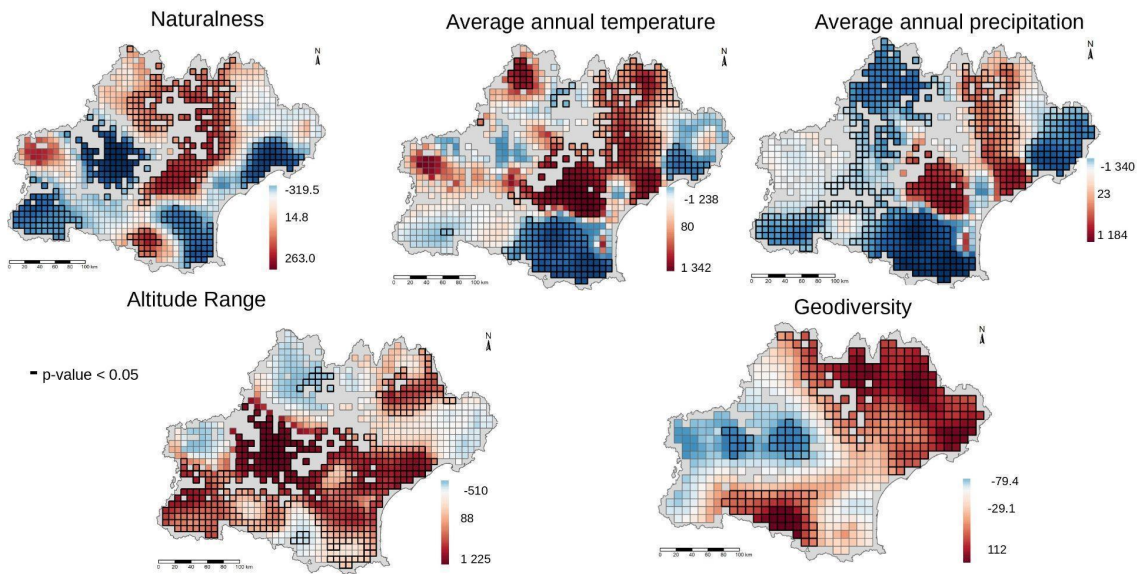
Fig.S5: Maps of the value of abiotic coefficients following GWR for vascular plants (A) butterflies (B) and birds (C) at the three spatial scales considered: 5, 7.5, and 10 km grid resolution.

A- Plants

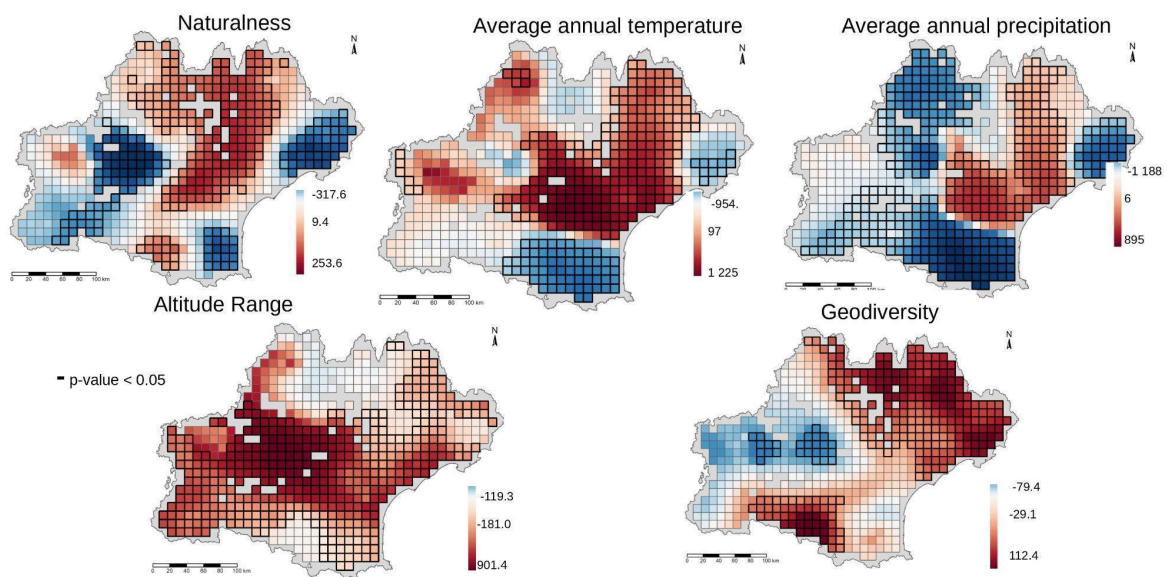
Grid of 5 km



Grid of 7.5 km

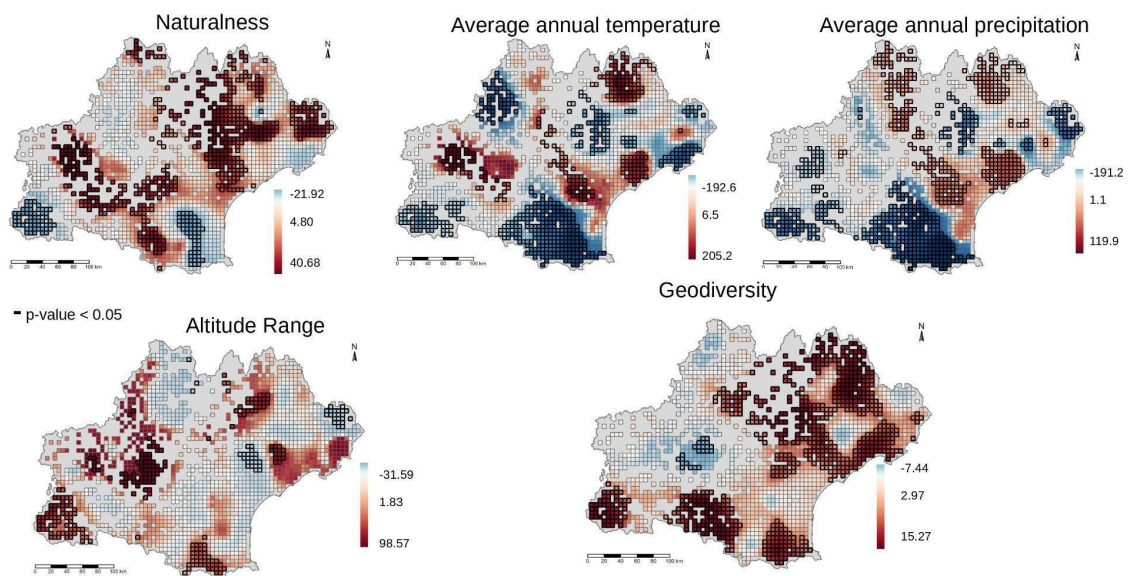


Grid of 10 km

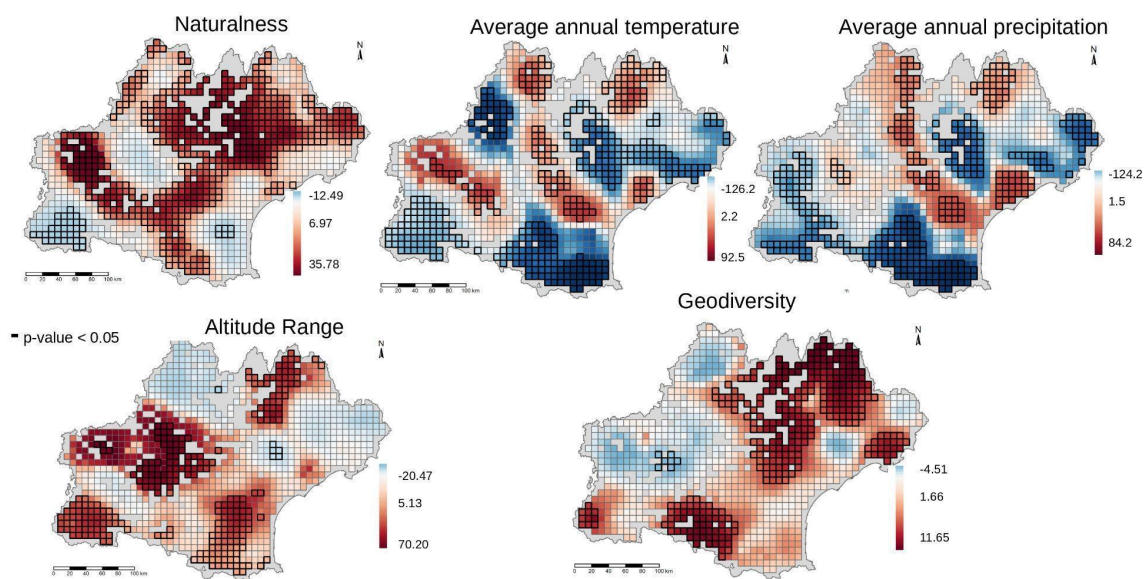


B- Butterflies

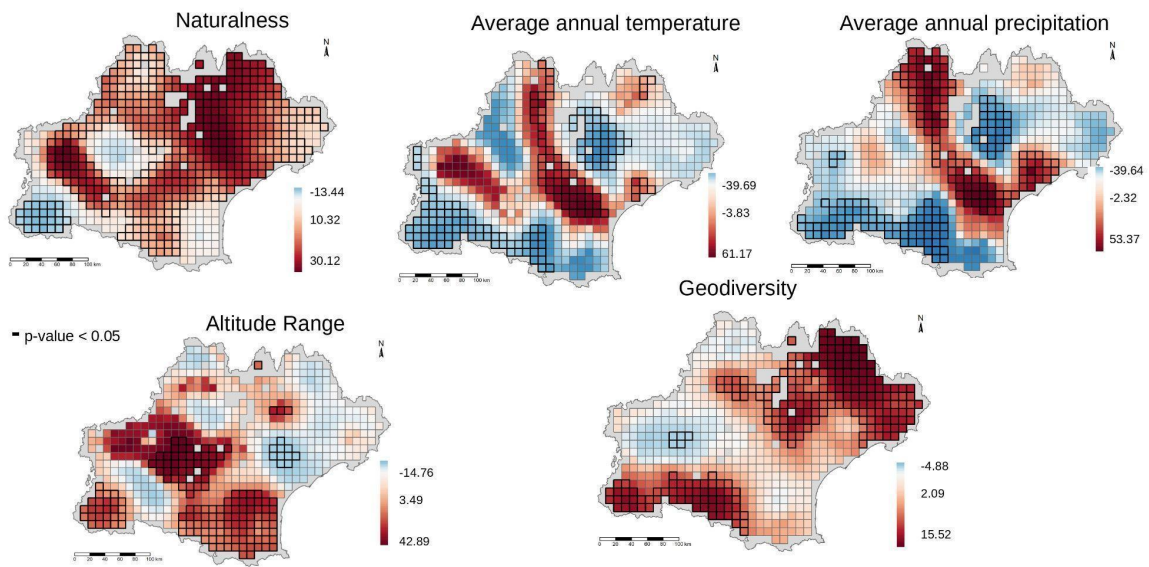
Grid of 5 km



Grid of 7.5 km

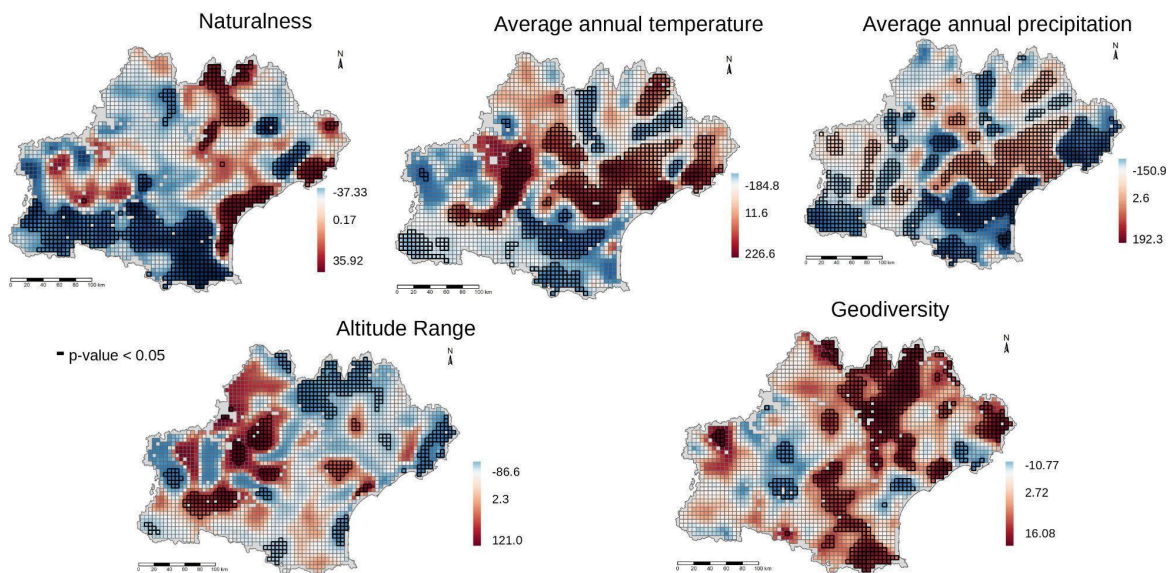


Grid of 10 km

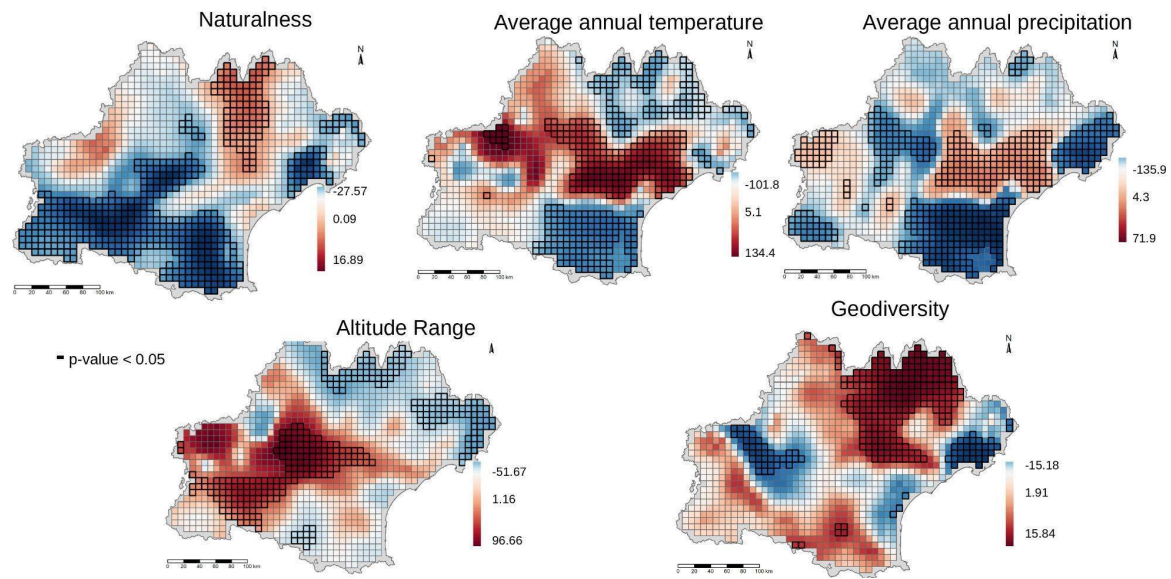


C-Birds

Grid of 5 km



Grid of 7.5 km



Grid of 10 km

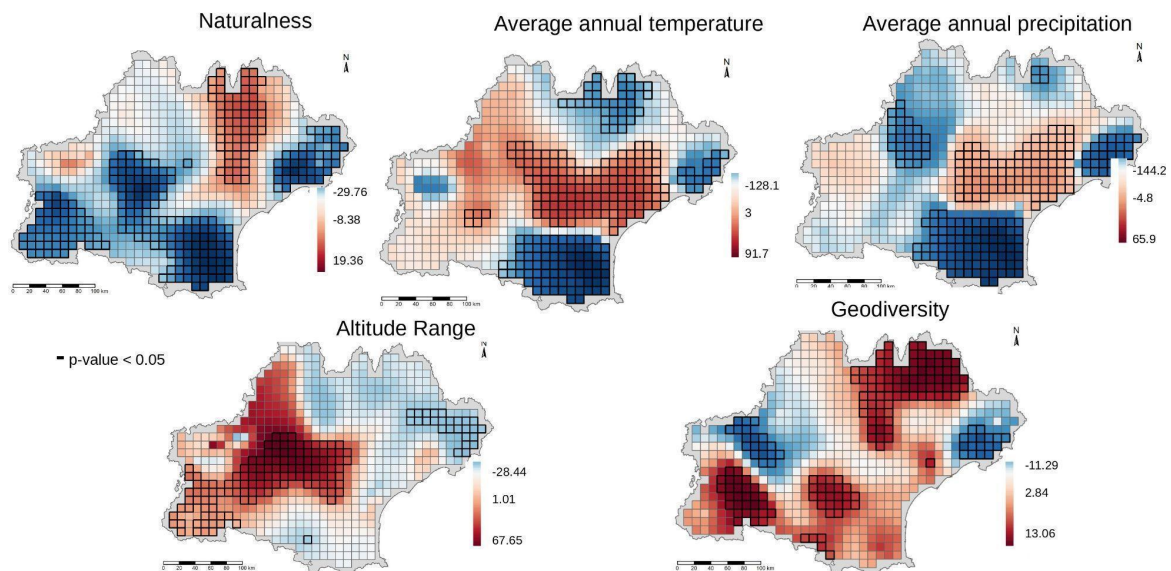


Fig. S6: Intersect of significant positive GWR correlations common between taxa. Number of grid cells showing a significant correlation for the three taxa is represented for each predictive environmental variable and spatial scale (5, 7.5 and 10km).

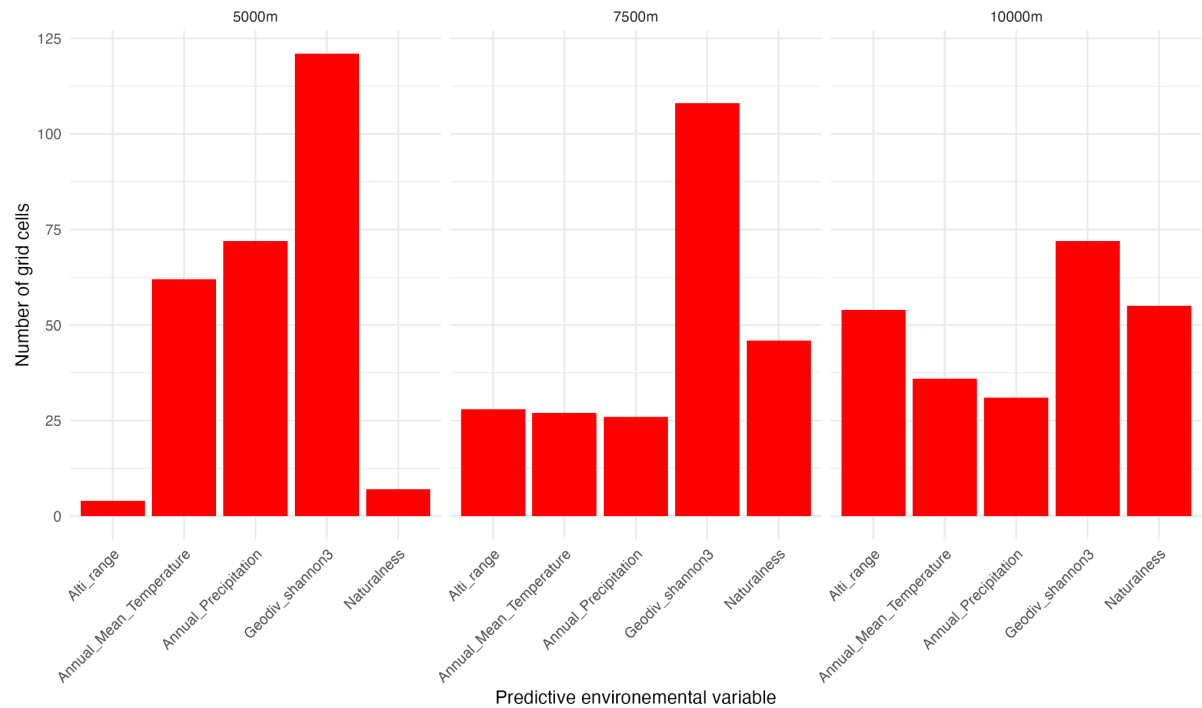
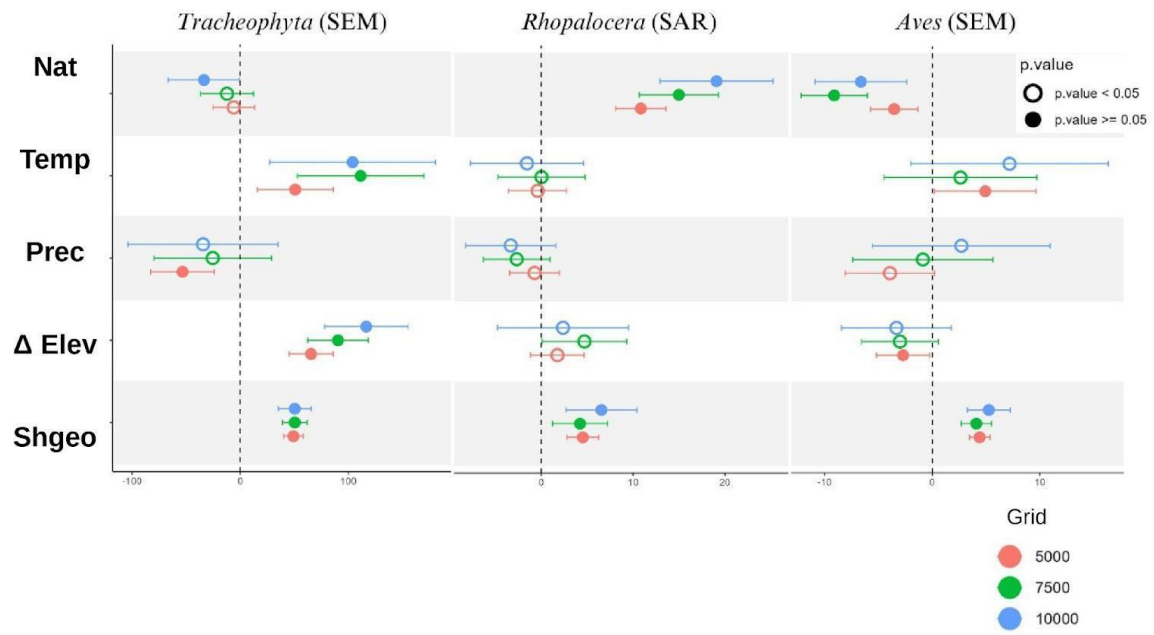


Fig S.7: Standardized coefficients ($\pm$ 95% CI) from the best-fitting spatial models (SAR and SEM), showing the effects of overall geodiversity (without geomorphological diversity) and environmental covariates on regional species richness. Tracheophyta: vascular plants; Rhopalocera: butterflies; Aves: birds; Nat: naturalness; Temp: annual mean temperature; Prec: annual mean precipitation; Δ Elev: altitude range; Shgeo: Shannon geodiversity.



Appendix A

In our initial modeling approach, which used total overall geodiversity, human footprint naturalness, annual mean temperature, annual mean precipitation, and elevation range as linear predictors of species richness, no evidence of problematic multicollinearity (all VIFs < 4), heteroscedasticity, or non-normality was detected in any of the models, regardless of the taxonomic group or spatial resolution considered. But in all our models, Moran's I index was positive and significant (p-value < 0.05), indicating positive autocorrelation and therefore spatial dependence in the residuals (Fig. A1, Table S4). This spatial dependence can be observed using the Moran scatterplot (Fig. A1a), which showed the positive linear relationship between the model's residuals and the spatially lagged residuals (mean of neighboring residuals). Another way to check this condition was through a spatial analysis of the residuals on a map (Fig. A1b), allowing us to visualize the structure of the spatial autocorrelation. For example, for birds at a resolution of 5 km, the model seems to underestimate their richness along the Mediterranean coastline. All residual maps and Moran's I values are presented in Fig.S3 and Table S4.

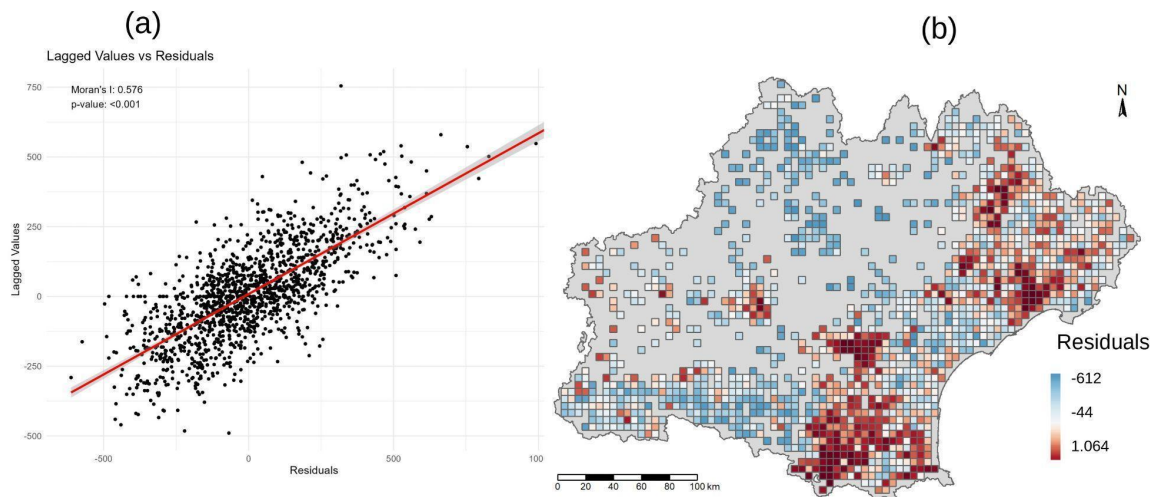


Fig. A1: Analysis of the spatial autocorrelation of the residuals of the linear bird model. (a) Moran diagram. (b) Mapping of linear model residuals (mesh = 5 km). The color gradient (b) represents the residual autocorrelation values. It goes from a negative residual autocorrelation value (blue) to a positive one (dark red).

The violation of spatial independence led us to reject the use of the linear model and to select the appropriate spatial model using Lagrange Multiplier tests. Based on these tests, the SAR

model was selected (lowest AIC) for all spatial resolutions in the case of butterflies, whereas the SEM model was preferred for vascular plants and birds (Table S4). This suggests that butterfly species richness in a given cell is directly influenced by species richness in neighboring cells, while for plants and birds, the observed spatial autocorrelation in the residuals is more likely due to spatially structured factors not accounted for in our modeling framework. To ensure that the issue of spatial dependence was resolved by our spatial models, an analysis of the residual distribution from the new models was conducted again (Fig. S3). Following the construction of the SAR and SEM models, no residual autocorrelation was detected (p-value Moran's $I > 0.05$), and the spatial structures of the residuals previously observed in Fig. 3 also disappeared (e.g., Fig. A2 for birds at 5 km resolution; and Fig. S4 for all taxa and spatial scales).

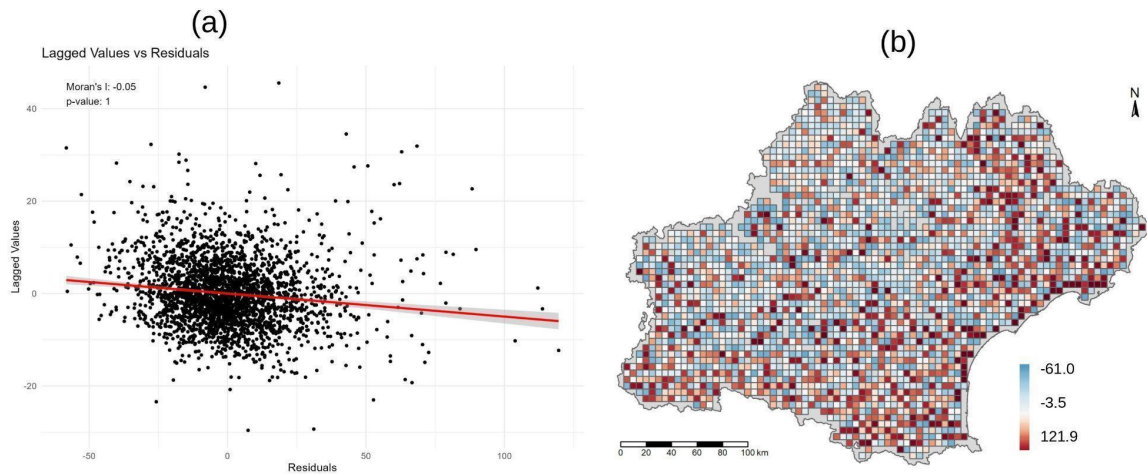


Fig. A2: Analysis of the autocorrelation of the residuals of the bird SEM model. (a) Moran diagram (b) Mapping of residuals from the SEM model (mesh = 5 km). The color gradient (b) represents the residual autocorrelation values. It goes from a negative residual autocorrelation value (blue) to a positive one (dark red).