Supporting information for :

Jointly estimating ecological niches and spatial sampling effort from multiple species occurrences

Christophe Botella[1,2,3,4], Alexis Joly[1], Pierre Bonnet[3,5], François Munoz[6], and Pascal Monestiez[4]

[1]INRIA Sophia-Antipolis - ZENITH team, LIRMM - UMR 5506 - CC 477, 161 rue Ada, 34095 Montpellier Cedex 5, France.

[2]INRA, UMR AMAP, F-34398 Montpellier, France.

[3]Univ Montpellier, UMR AMAP, Montpellier, France.

[4]INRA, BioSP, Site Agroparc, 84914 Avignon, France.

[5]CIRAD, UMR AMAP, F-34398 Montpellier, France.

[6]Université Grenoble Alpes, 621 avenue Centrale, 38400 Saint-Martin-d'Hères, France.

1 Appendix A: Approximation of model likelihood through a weighted sum over quadrature points

We detail the approximation of equation (3) (in the main manuscript) For any $i \in [|1, N|]$. (i) For the Monte Carlo approximation of the integral, we uniformly draw $Z_q = \{z_1^q, ..., z_Q^q\}$ points over $\cup_{j=1}^{C} c_j$ marked for species i and decompose our sum approximation as follows:

$$\begin{aligned} \int_D s(z)\lambda_i(z)dz &\approx \sum_{z \in Z_i \cup Z_q} w(z)s(z)\lambda_i(z) \\ &= \frac{1}{100}\sum_{z \in Z_i} \frac{|D|}{n_i} s(z)\lambda_i(z) + \frac{99}{100}\sum_{z \in Z_q} \frac{|D|}{Q} s(z)\lambda_i(z) \end{aligned}$$

For the real dataset of occurrences, we use an alternative strategy to insure that all the sampling cells have background points and that they capture the environmental variability of each cell. We uniformly draw a fixed number (6) of background points uniformly in each sampling cell. As each sampling cell has the same size in the present case, we can keep the same weighting scheme as previously, and the procedures weighted sum will also converge to the target integral. We can show this by decomposing the integral into a sum of integrals over each sampling cell multiplied by the inverse of the total number of cells and then using the Monte Carlo (because points are uniformly drawn inside cells).

2 Appendix B: Environmental variables tables

Name	Description	Values	Resolution (m)
CHBIO_1	Annual Mean Temperature	[-10.6,18.4]	1000
CHBIO_5	Max Temperature of Warmest Month	[36.4,6.2]	1000
CHBIO_12	Annual Precipitation	[318,2543]	1000
etp	Potential Evapo Transpiration	[133,1176]	1000
alti	Elevation	[-188,4672]	90
slope	Absolute elevation gradient	[0,13457]	90
awc_top	Topsoil available water capacity	$\{0, 120, 165, 210\}$	1000
bs_top	Base saturation of the topsoil	$\{35, 62, 85\}$	1000
spht	Aggregated land cover	{culti.,for.,past.,urb.,other}	100

Table 1: Table of environmental variables used in this study.

CLC category description	spht category name	Raster code
Non-irrigated arable land	cultivated	12
Permanently irrigated land	cultivated	13
Vineyards	cultivated	15
Fruit trees and berry plantations	cultivated	16
Complex cultivation patterns	cultivated	20
Land principally occupied by agriculture, with significant areas of natural vegetation	cultivated	21
Agro-forestry areas	cultivated	22
Pastures	grasslands	18
Natural grasslands	grasslands	26
Moors and heathland	grasslands	27
Sclerophyllous vegetation	grasslands	28
Broad-leaved forest	forest	23
Coniferous forest	forest	24
Mixed forest	forest	25
Transitional woodland-shrub	forest	29
Continuous urban fabric	urban	1
Discontinuous urban fabric	urban	2
Industrial or commercial units	urban	3
Road and rail networks and associated land	urban	4
Airports	urban	6
Green urban areas	urban	10
Sport and leisure facilities	urban	11
Port areas	other	5
Mineral extraction sites	other	7
Dump sites	other	8
Construction sites	other	9
Rice fields	other	14
Olive groves	other	17
Annual crops associated with permanent crops	other	19
Beaches, dunes, sands	other	30
Bare rocks	other	31
Sparsely vegetated areas	other	32
Burnt areas	other	33
Glaciers and perpetual snow	other	34
Inland marshes	other	35
Peat bogs	other	36
Salt marshes	other	37
Salines	other	38
Intertidal flats	other	39
Water courses	other	40
Water bodies	other	41
Coastal lagoons	other	42
Estuaries	other	43
Sea and ocean	other	44
No data	other	48
Unclassified land surface	other	49
Unclassified water bodies	other	50

Table 2: spht (Aggregated land cover) categories correspondance with Corine Land Cover 2012.