Brief methological explanation and preliminary results

Here we summarise all the critical steps conducted to analise bee preferences.

Filtering criteria

1) Filter records above 1985 and 1987 (this is decided based on the GIS data available) for Europe and USA respectively.
2) Filter by unique capture event.
3) Filter by minimum of 3 decimals on coordinates.
4) Filter number of levels per species (minimum N=50).
5) Filter by wide geographical distribution.

Land use classification

1) Europe

Link with land European use classes

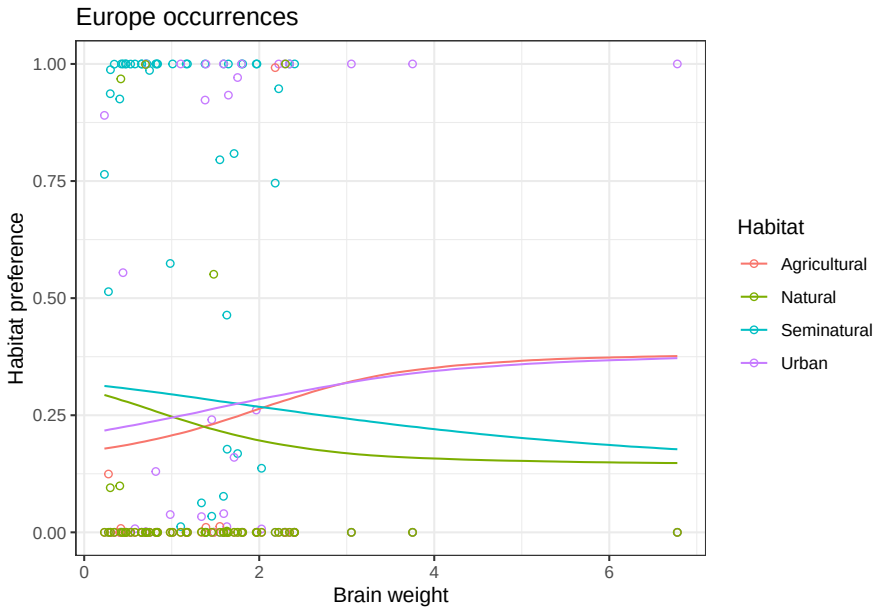
https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/html/index-clc-512.html

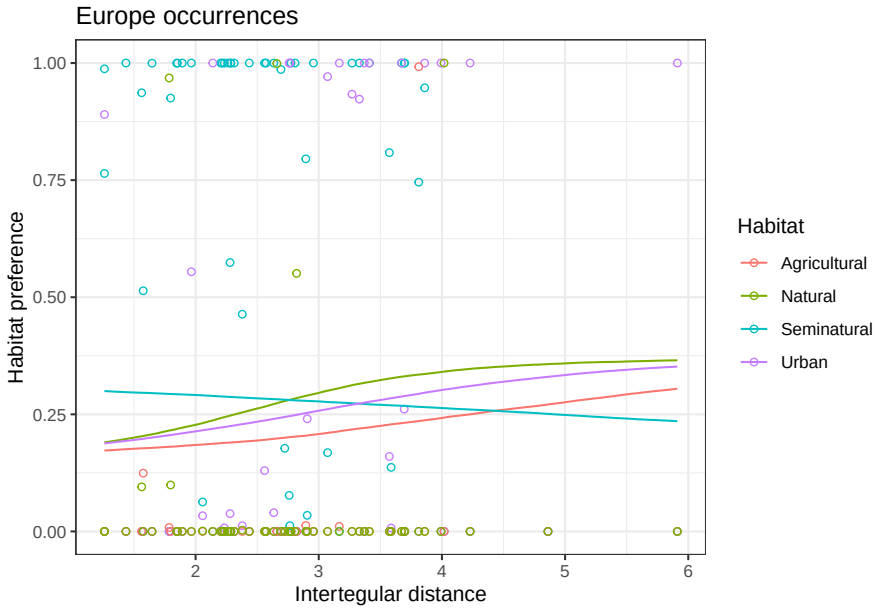
Land_use	n_rows	Cover_names
Industrial or commercial units	121485	Urban
Permanently irrigated land	66733	Seminatural
Annual crops associated with permanent crops	46655	Agricultural
Land principally occupied by agriculture, with significant areas of natural vegetation	26878	Seminatural
Coniferous forest	16835	Natural
Agro-forestry areas	13265	Seminatural
Non-irrigated arable land	10128	Agricultural
Road and rail networks and associated land	10023	Seminatural
Mixed forest	7948	Natural
Natural grasslands	7639	Natural
Sport and leisure facilities	6280	Seminatural
Moors and heathland	2838	Natural
Port areas	2512	Urban
Sclerophyllous vegetation	2420	Natural
Fruit trees and berry plantations	2168	Agricultural
Peat bogs	2149	Natural
Discontinuous urban fabric	2015	Seminatural
Coastal lagoons	1998	Discard
Beaches, dunes, sands	1438	Discard
Dump sites	1391	Discard
Salt marshes	1155	Natural
Olive groves	1036	Agricultural
Bare rocks	698	Discard
Water bodies	689	Discard
Water courses	631	Discard
Sea and ocean	522	Discard
Mineral extraction sites	457	Discard
Airports	393	Urban
Construction sites	305	Urban
Salines	208	Seminatural
Green urban areas	192	Urban
Burnt areas	35	Discard
Estuaries	22	Discard
Sparsely vegetated areas	10	Natural

Check final levels per category:

Cover_names	Levels
Seminatural	123387
Urban	126902
Agricultural	59987
Natural	40994

Then, we calculate preferences and model the data.





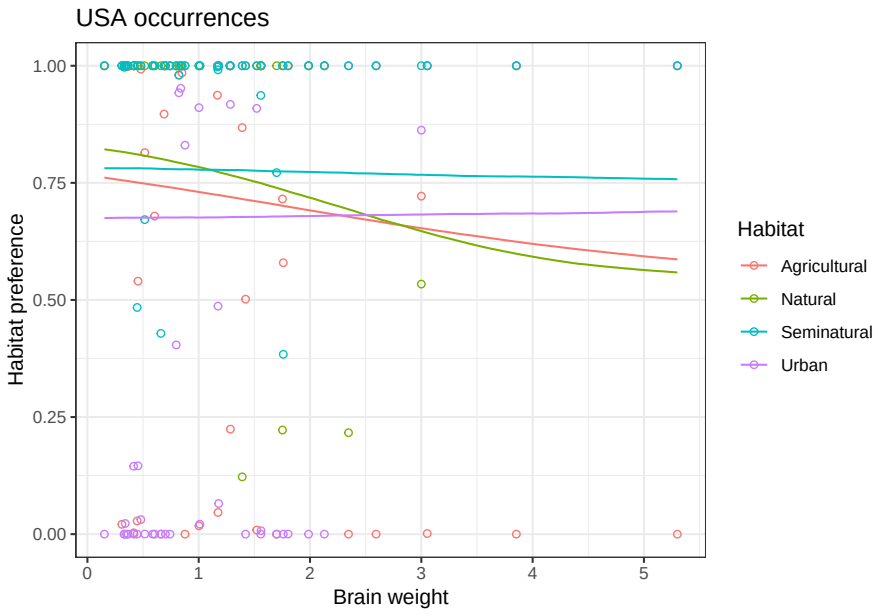
2) USA

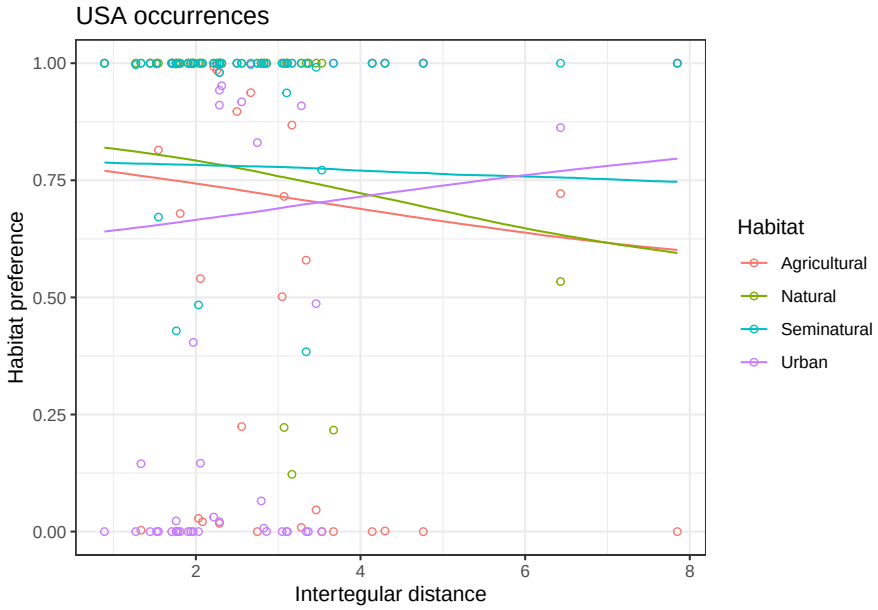
Land_use	n_rows	Cover_names
Developed, Open Space	9713	Seminatural
Developed, Low Intensity	8029	Seminatural
Developed, Medium Intensity	7774	Urban
Deciduous Forest	6573	Natural
Hay/Pasture	5970	Agricultural
Developed, High Intensity	3472	Urban
Mixed Forest	2592	Natural
Cultivated Crops	2240	Agricultural
Woody Wetlands	2192	Natural
Shrub/Scrub	1127	Natural
Evergreen Forest	999	Natural
Herbaceous	926	Natural
Open Water	744	Discard
Emergent Herbaceous Wetlands	588	Natural
Barren Land	264	Discard

Check final levels per category:

cover.names	n_rows
Agricultural	8210
Natural	14997
Urban	11246
Seminatural	17742

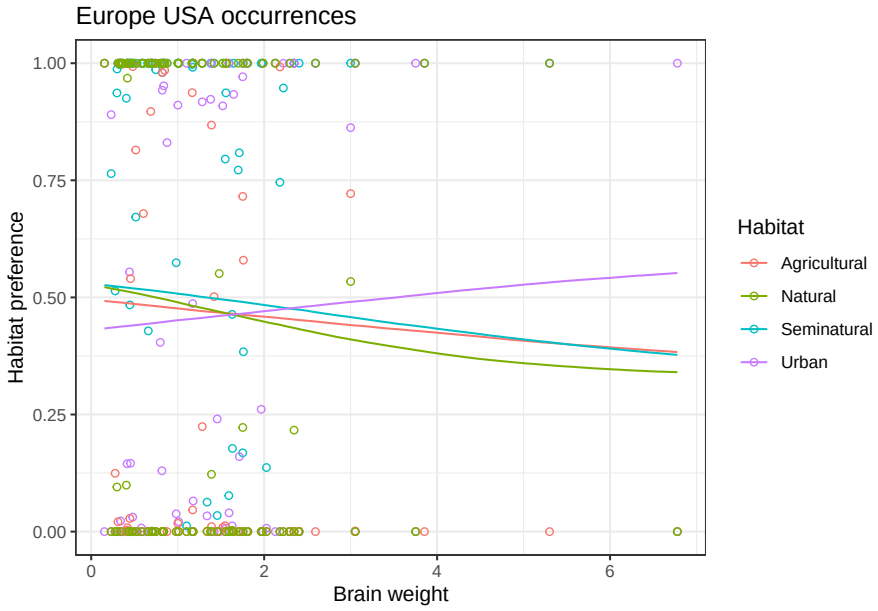
Then, we calculate preferences and model the data.

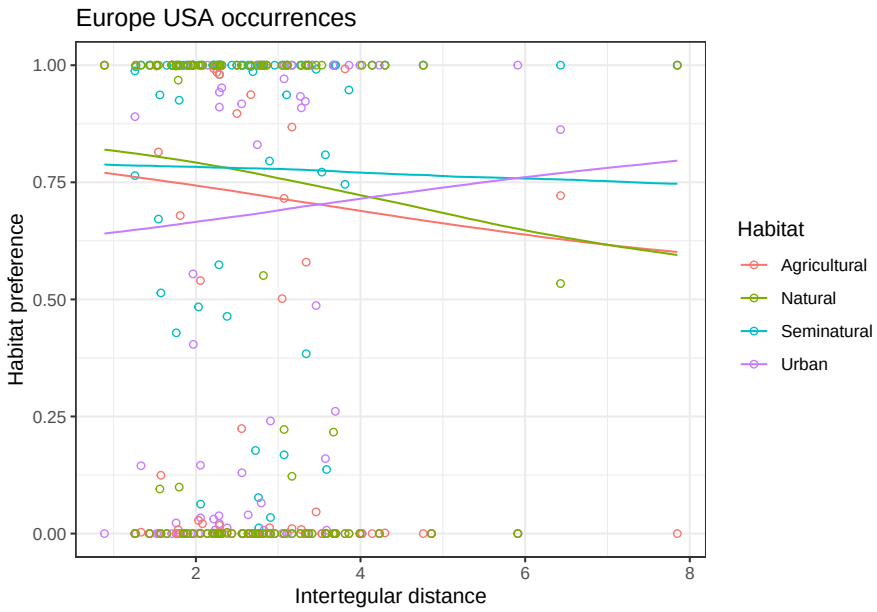




Note: The models have been run with non-informative priors from the brms function but some work on priors can be done... I have explored a bit and the results were quite similar overall but reviewers seem to don't like this approach as one of the goals of Bayesian is to use the prior knowledge of our variables to improve our predictions.

2) Finally, we consider Europe and USA together




Europe USA occurrences
Habitat preference
1.00
0.75
0.50
0.25
0.00
2
4
6
8
Intertegular distance
Habitat
Agricultural
Natural
Seminatural
Urban