GeoLIFECLEF : Synthèse de l’Overview

Location-based species recommendation

Goal is to automatically predict the **list of species that are the most likely to be observed** at a given location.

Data : Large scaled spatialized environmental data

Large training set of species occurrences, each one being associated to a **multi-channel image** characterizing the local environment.

=> Predict the distribution

Problem : usually not possible to learn a species distribution model directly from spatial positions because of the limited number of occurrences and the sampling bias.

From a machine learn ing point of view, the challenge will thus be treatable as a **multi-channel image classification task**.

Environmental data : possibles problems

* Bias in the sampling and lack of occurrences at certains locations.
* Image channels do not have the same ‘scope’ of resolution : 1km, 100m, 90m, 12.5m
* Reporting dependencies (?)

A data occurrence is composed this way :

-Locations coordinates - punctual environmental values (extracted from the rasters) - environmental variables tensor (64x64 pixels x33)

Questions ?

Why the perfect row in Table 1 only has a score of 0.5593

What are fusion models (like CNN + Random forest)

Not 1 model per species?

What could cause the difference in score, for the same algorithm type of the same team, between different runs? Like ST\_17 ans ST\_10.