

Occurrences_Human&Crops

Eléna Manfrini

2025-01-24

J'ai analysé l'impact des données d'occurrences d'*Hermetia illucens* sur le comportement de nos modèles.

Initialement, Quentin avait choisi de n'inclure que les données d'occurrences issues d'observations humaines et de machines. Pour l'espèce *Hermetia illucens*, cela correspondait à 17 668 occurrences.

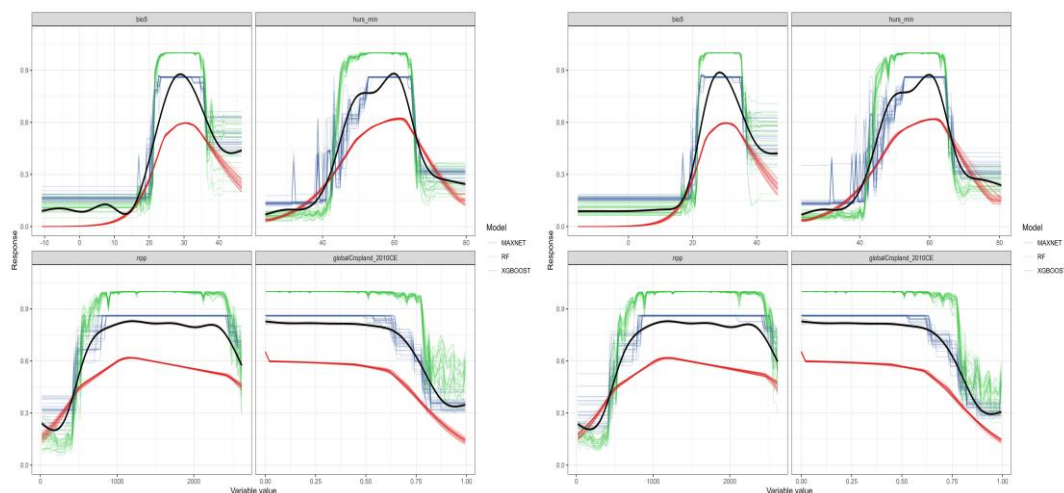
Cependant, en examinant plus attentivement les types de données d'occurrences disponibles, je me suis aperçue que leurs définitions n'étaient pas toujours très claires (<https://gbif.github.io/parsers/apidocs/org/gbif/api/vocabulary/BasisOfRecord.html>). Par conséquent, il me semble préférable de maximiser le nombre de données utilisées plutôt que de les restreindre trop. Cela dit, deux catégories apparaissent évidentes à exclure : "UNKNOWN" et "FOSSIL_SPECIMEN", afin de garantir une plus forte probabilité d'avoir des occurrences d'individus trouvés vivants.

En retenant tous les autres types d'occurrences et en excluant uniquement ces deux catégories, nous obtenons 18 688 occurrences pour *Hermetia illucens*.

Examinons maintenant l'effet de ces différences sur nos modèles.

Bio5, Hurs_min, Npp et Cropland.

Concernant les réponses de nos modèles sur les variables bio5, hurs_min, npp et cropland.

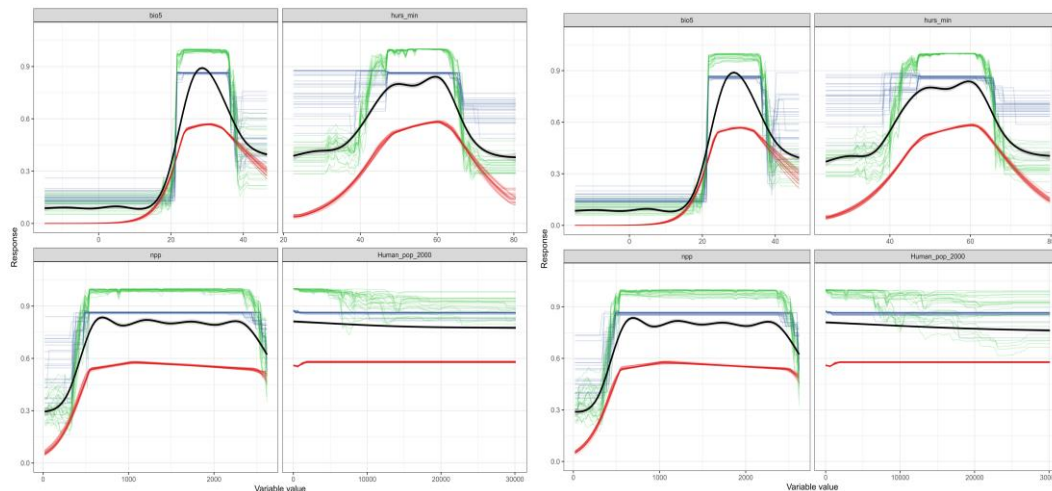


Human and Machine observation All observations (-Fossil & Unknown)

La différence des réponses des modèles par rapport aux nombres initiaux d'occurrences n'est pas flagrante. En effet, les réponses restent globalement similaires.

Bio5, Hurs_min, Npp et Human population

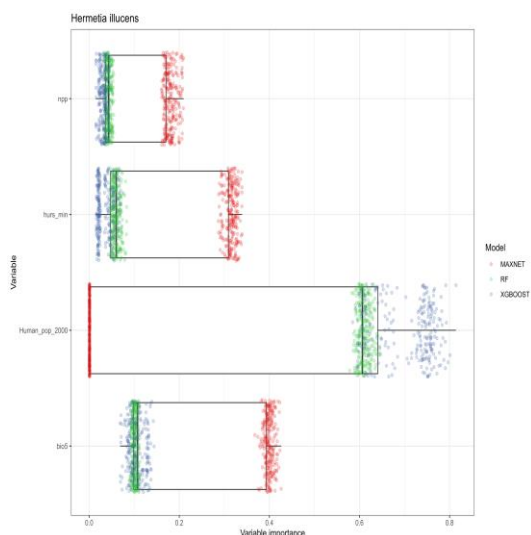
Concernant les réponses de nos modèles sur les variables bio5, hurs_min, npp et human population.



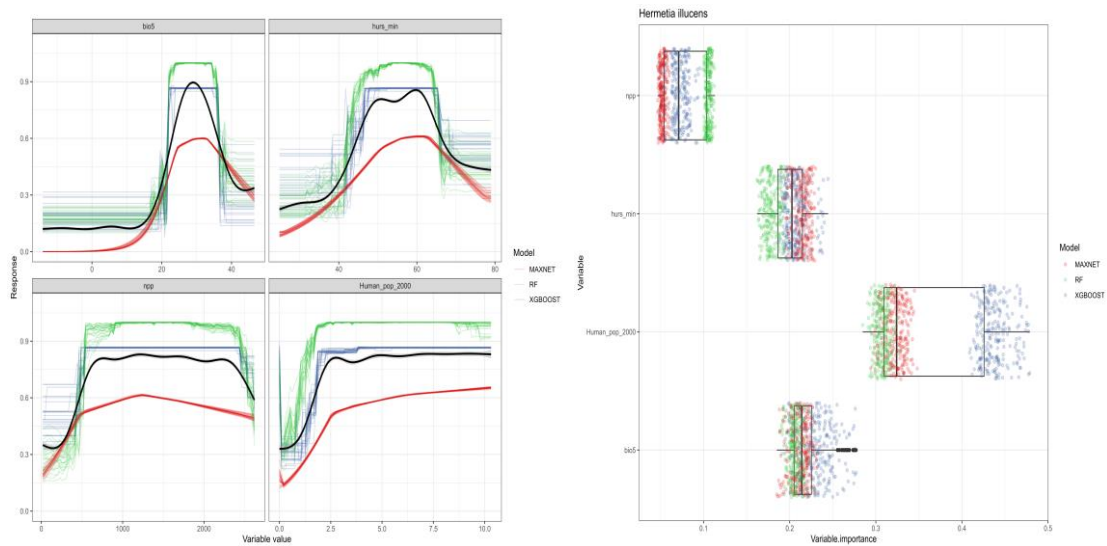
Human and Machine observation

All observations (-Fossil & Unknown)

De même que pour cropland, les relations ne changent pas avec ou sans le rajout d'occurrences. Les modèles ne semblent pas réagir à la variable Hum pop dans les deux cas. Cela pourrait-il être dû à une trop grande disparité dans l'échelle des valeurs ? Est-ce qu'il serait pertinent d'essayer de mettre sous format log le raster de population afin d'homogénéiser les données et faciliter l'analyses pour les modèles ?



Mise en échelle log du raster Human pop + toutes les occurences (- Fossil and Unknown)



Les modèles ont l'air d'avoir trouvé une relation entre la densité de population humaine et les occurrences pour *Hermetia illucens* grâce à une mise sous format log. La transformation log permet une homogénéisation des valeurs et donc une réduction de la disparité des valeurs extrêmes ce qui permet sûrement aux modèles de mieux voir la relation.

La transformation log de la densité de population humaine a même fait passer cette variable comme étant la plus importante par nos modèles.