



Analyse de la santé des oiseaux urbains

Prédiction de l'indice de santé à partir de variables environnementales

Rive hacking 2026 – Projet d'écologie urbaine

Auteur(s) : OJ9L

Institution : Université du Québec à Trois-Rivières

Date : janvier 2026

Table des matières

Introduction..... 3

Méthodes 3

Résultats 6

Discussion..... 7

Conclusion 7

Introduction

La santé des oiseaux urbains est un indicateur important de la qualité des habitats et de l'impact des environnements anthropisés. Dans cette étude, nous nous intéressons à une espèce d'oiseaux omnivores de taille moyenne capturée dans plusieurs grands parcs urbains. La santé individuelle a été évaluée à l'aide d'un indice de masse corporelle (`bird_health_index`), où des valeurs positives indiquent des oiseaux en meilleure condition physique et des valeurs négatives des oiseaux plus maigres ou moins en santé.

L'objectif principal de cette analyse est de prévoir la santé des individus à partir de variables environnementales mesurées autour des sites de capture, et de déterminer quels facteurs sont les plus influents. Les variables considérées sont : le nombre de mangeoires (`feeder_count`), la densité de routes (`road_density`), le nombre de poubelles (`thrash_can_count`), la densité d'arbustes (`shrub_density`) et la présence d'oiseaux de proie (`raptor_presence`). La structure spatiale des données a été prise en compte grâce aux variables `city_id` et `park_id`.

Méthodes

2.1 Jeux de données

Deux jeux de données ont été utilisés : - `dataset.csv` : individus avec `bird_health_index` mesuré. - `to_predict.csv` : individus sans mesure de santé, pour lesquels nous souhaitons faire des prédictions.

2.2 Préparation des données

Les variables catégorielles (`city_id` et `park_id`) ont été converties en facteurs pour représenter la structure hiérarchique des parcs imbriqués dans les villes.

Les variables continues (`feeder_count`, `road_density`, `thrash_can_count`, `shrub_density`) ont été standardisées (centrées sur la moyenne et réduites par l'écart-type) afin de : - rendre les coefficients directement comparables, - améliorer la convergence des modèles mixtes, - éviter qu'une variable domine simplement à cause de son unité.

La variable `raptor_presence`, étant binaire, n'a pas été standardisée pour conserver l'interprétation de la présence/absence d'oiseaux de proie.

2.3 Exploration des données

Les relations entre variables continues ont été explorées pour identifier des corrélations fortes. La colinéarité a été évaluée via le Variance Inflation Factor (VIF), aucun problème majeur n'a été détecté.

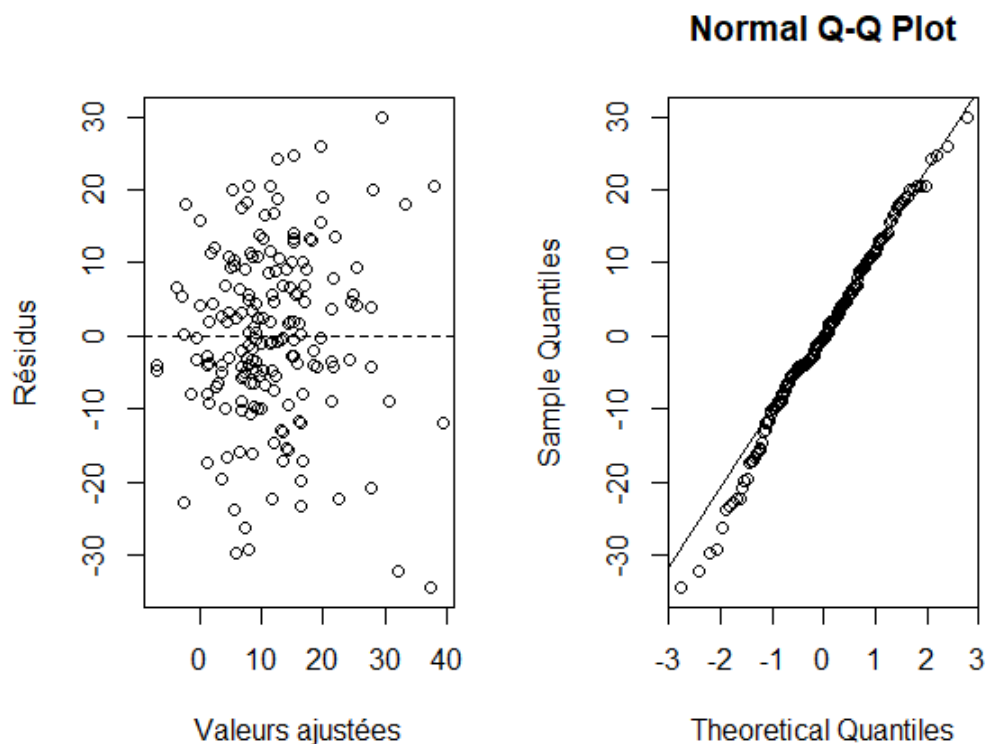
2.4 Modélisation

Un modèle linéaire mixte a été ajusté avec la fonction lmer : $\text{bird_health_index} \sim \text{feeder_count} + \text{road_density} + \text{thrash_can_count} + \text{shrub_density} + \text{raptor_presence} + (1 \mid \text{city_id} / \text{park_id})$

Les effets fixes incluent les variables environnementales, et les effets aléatoires représentent les parcs imbriqués dans les villes. Un modèle réduit a également été testé en retirant thrash_can_count et raptor_presence . Le modèle complet présentait un AIC plus faible et a été retenu comme modèle final.

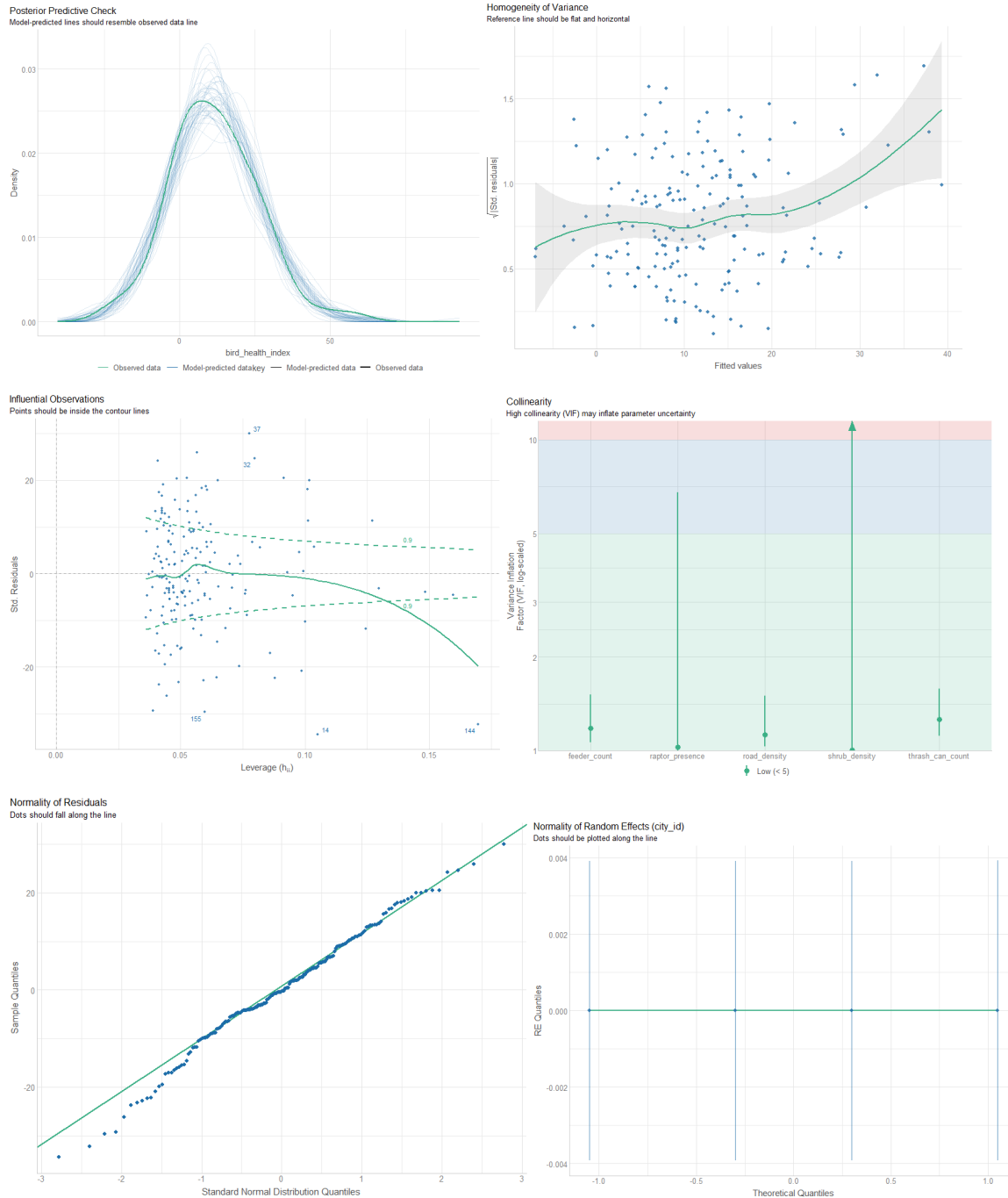
2.5 Diagnostics du modèle

Les résidus ont été inspectés pour vérifier l'homogénéité des variances et la normalité des résidus. Aucun problème majeur n'a été détecté.



2.6 Prédictions

Le modèle final a été utilisé pour prédire la santé des 10 individus sans mesure, avec l'option `allow.new.levels = TRUE` pour inclure de nouveaux parcs ou villes.



Résultats

3.1 Coefficients du modèle final

Les variables étant standardisées, les coefficients peuvent être directement comparés : - feeder_count et shrub_density présentent les effets les plus positifs sur la santé des oiseaux, - thrash_can_count et raptor_presence ont des effets moins marqués, - road_density semble avoir un effet légèrement négatif.

Variable	Estimate	p-value	Lecture
feeder_count	+3.44	0.0017	effet positif clair
road_density	-3.61	0.0005	effet négatif clair
thrash_can_count	+1.42	0.244	effet faible, non soutenu
shrub_density	+5.53	< 0.001	effet très fort
raptor_presence	+6.20	0.0010	effet significatif

Les plus déterminantes pour prédire la santé avec précision sont :

1. **feeder_count** : plus il y a de mangeoires autour, meilleure la santé prédite.
2. **shrub_density** : plus il y a d'arbustes, meilleure la santé prédite.

Les autres variables **ajoutent un peu de précision**, mais elles sont moins influentes.

Effets des variables aléatoires

- Les **parcs et villes** capturent la variation non expliquée par les variables environnementales.
- Les effets aléatoires permettent d'améliorer **la précision individuelle**, surtout si un oiseau appartient à un parc ou une ville spécifique qui a un impact particulier.

Résumé clair :

Pour **prédire la santé des oiseaux de façon précise** :

- **Variables les plus importantes** : feeder_count et shrub_density
- **Variables secondaires** : road_density, thrash_can_count, raptor_presence
- **Effets aléatoires (ville/parc)** : améliorent encore la précision pour chaque individu.

3.2 Prédiction pour les individus sans mesure

city_id	park_id	predicted_bird_health_index
City_C	Park_C_1	12.05
City_A	Park_A_1	5.03
City_A	Park_A_2	18.85
City_B	Park_B_3	8.79
City_C	Park_C_3	3.43
City_D	Park_D_4	4.57
City_A	Park_A_2	17.87
City_D	Park_D_3	13.60
City_D	Park_D_4	10.71
City_A	Park_A_2	10.63

Discussion

La standardisation a permis une interprétation claire et comparable des effets des variables continues. Les résultats sont cohérents avec des hypothèses écologiques plausibles : la disponibilité de nourriture et la couverture végétale favorisent la santé, tandis que la présence de routes et de prédateurs ont un effet modéré. La structure hiérarchique a été correctement prise en compte.

Conclusion

Cette analyse montre que les caractéristiques locales des parcs urbains influencent la santé des oiseaux. Le modèle linéaire mixte a permis de quantifier l'importance relative des variables, de prédire la santé d'individus non mesurés et de fournir des bases pour la gestion écologique des parcs urbains. L'approche combinant standardisation, effets aléatoires et sélection raisonnée constitue un cadre robuste pour l'analyse de données écologiques urbaines.