

Modèles additifs généralisés - Exercices en classe

Données

Le fichier `dendro_wa082.csv`, basé sur le jeu de données *wa082* inclus dans le package *dplR*, contient des séries dendrochronologiques de 23 sapins gracieux (*Abies amabilis*) échantillonnés dans l'état du Washington (nord-ouest américain). La première colonne représente l'année et chaque autre colonne représente la croissance annuelle de la surface terrière d'un arbre, telle que déterminée par les cernes de croissance. Les valeurs manquantes NA représentent des années antérieures à la formation du premier cerne de l'arbre.

```
wa <- read.csv("../donnees/dendro_wa082.csv")
wa[1:5, 1:8]
```

##	year	X712011	X712012	X712021	X712022	X712031	X712032	X712041
## 1	1698	NA	NA	NA	NA	NA	NA	NA
## 2	1699	NA	NA	NA	NA	NA	NA	NA
## 3	1700	NA	NA	NA	NA	NA	NA	NA
## 4	1701	NA	NA	NA	NA	NA	NA	NA
## 5	1702	NA	NA	NA	NA	NA	NA	NA

1. Préparation des données

- Utilisez la fonction `pivot_longer` du package *tidyr* pour transformer `wa` en un tableau avec trois colonnes: l'année, l'arbre et la croissance.
- Ajoutez des colonnes pour représenter l'âge et la surface terrière (croissance cumulative) correspondant à chaque paire arbre-année. Avec le package *dplyr*, vous pouvez trier les données par arbre et année, grouper les données par arbre, puis calculer l'âge avec `row_number` et la surface terrière avec `cumsum` (somme cumulative).
- Illustrez les séries de croissance de chaque arbre. Il est recommandé de représenter le logarithme de la croissance en surface terrière.

2. Croissance en fonction de l'âge et de la surface terrière

La croissance annuelle d'un arbre dépend de facteurs intrinsèques (ex.: âge et taille actuelle de l'arbre) et extrinsèques, notamment les conditions climatiques. Afin d'isoler l'effet du climat sur la croissance, il est donc nécessaire de retirer des séries de croissance la tendance due à l'âge et la taille de chaque arbre. Nous utiliserons ici un GAM pour estimer cette tendance, avec une forme du modèle semblable à celle proposée dans l'étude de Girardin et al. (2016).

- Ajustez un modèle additif avec la formule $\log(cst) \sim \log(st) + s(age)$, où *cst* est la croissance en surface terrière de l'année et *st* est la surface terrière. Assurez-vous que le paramètre *k* de la spline est assez élevé. Comment interprétez-vous le coefficient de $\log(st)$? Comment décrivez-vous (brièvement) l'effet de l'âge?
- Ajoutez maintenant un effet aléatoire de l'arbre sur l'ordonnée à l'origine du modèle en (a). Vérifiez l'ajustement du modèle, incluant la normalité des effets aléatoires. Quelle est la fraction de la variance de $\log(cst)$ expliquée par ce modèle?
- Comparez deux façons d'inclure la variation de la croissance d'une année à l'autre au modèle en (b): (1) une spline selon l'année (considérée comme variable numérique) ou (2) un effet aléatoire de l'année (considérée comme facteur) sur l'ordonnée à l'origine du modèle. Quelles sont les différences entre les suppositions des deux versions du modèle?

- (d) Avec la fonction `predict(..., type = "terms")`, nous pouvons obtenir la contribution de chaque terme du modèle à la réponse prédite. Utilisez cette méthode pour illustrer les effets aléatoires de l'année estimés pour le deuxième modèle en (c).

Références

Girardin, M.P. et al. (2016) No growth stimulation of Canada's boreal forest under half-century of combined warming and CO2 fertilization. PNAS 113, E8406-E8414.