Generalized additive models - Solutions

Data

The dendro_wa082.csv dataset, based on the *wa082* dataset included in the *dplR* package, contains dendrochronological series of 23 Pacific silver firs (*Abies amabilis*) sampled in the state of Washington (northwestern USA). The first column represents the year and each other column represents the annual basal area increment (BAI) of a tree, as determined by tree rings. Missing values (`NA`) represent years prior to the formation of the first tree ring.

```
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. Data preparation

(a) Use the `pivot_longer` function of the *tidyr* package to transform `wa` in a table with three columns: year, tree and basal area increment (bai).

(b) Add columns to represent the age and basal area (cumulative growth) for each tree-year pair. With the *dplyr* package, you can sort the data by tree and year, group the data by tree, then calculate the age with `row_number` and the basal area with `cumsum` (cumulative sum).

(c) Illustrate the growth series of each tree. It is recommended to show the logarithm of the basal area increment.

2. Growth according to age and basal area

The annual growth of a tree depends on both intrinsic (e.g. age and current size of the tree) and extrinsic factors, including climatic conditions. In order to isolate the effect of climate on growth, it is therefore necessary to remove from the growth series the trend due to the age and size of each tree. Here we will use a GAM to estimate this trend, with a form of the model similar to that proposed in the study by Girardin et al. (2016).

(a) Fit an additive model with the formula `log(bai) ~ log(ba) + s(age)`, where *bai* is the basal area increment of the year and *ba* is the basal area. Make sure that the k parameter of the spline is high enough. How do you interpret the coefficient of `log(ba)`? How do you (briefly) describe the effect of age?

(b) Now add a random effect of the tree on the intercept of the model in (a). Check the fit of the model, including the normality of the random effects. What is the fraction of the variance of `log(bai)` explained by this model?

(c) Compare two ways of including interannual growth variation in the model in (b): (1) a spline by year (with year as a numerical variable) or (2) a random effect of year (with year as a factor) on the intercept. What are the differences between the assumptions of the two versions of the model?

(d) With the function `predict(..., type = "terms")`, we can obtain the contribution of each term in the model to the predicted response. Use this method to illustrate the random year effects estimated for the second model in (c).

References

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.