

## Régression Non Linéaire avec Application sous R

*GM5 – Feuille TP 6***Estimation par backfitting  
dans les modèles additifs non linéaires**

L'objet de ce TP est d'estimer avec l'algorithme du backfitting et la fonction `gam` du package `mgcv`, le modèle additif non linéaire introduit dans le cours précédent, pour expliquer le maximum de la concentration en ozone du jour (en  $\mu g/m^3$ ), à partir des 3 variables explicatives `temp` (maximum de la température du jour, en  $^{\circ}C$ ), `vent` (maximum de vitesse de vent du jour, entre 14h et 18h, en  $m/s$ ) et `mozon` (moyenne spatiale de la concentration en ozone de la veille sur la région parisienne, en  $\mu g/m^3$ ).

1. Récupérer les données à l'URL

<http://lmi2.insa-rouen.fr/~bportier/Data/p13.txt>

2. Construire un échantillon d'apprentissage (80% des données) pour mettre au point le modèle et un échantillon test (20% des données) pour évaluer les performances du modèle. On pourra s'inspirer de la procédure décrite dans un TP proposé par P. Besse, que l'on trouvera à l'URL

[http://www.lsp.ups-tlse.fr/Besse/pub/TP/appr\\_se/tp4\\_cancer\\_lec\\_extract\\_R.pdf](http://www.lsp.ups-tlse.fr/Besse/pub/TP/appr_se/tp4_cancer_lec_extract_R.pdf)  
et dont voici le code R :

```
> set.seed(111) # initialisation du générateur
# Extraction des échantillons
> test.ratio=.2 # part de l'échantillon test
> npop=nrow(data) # nombre de lignes dans les données
> nvar=ncol(data) # nombre de colonnes
> ntest=ceiling(npop*test.ratio) # taille de l'échantillon test
> testi=sample(1:npop,ntest) # indices de l'échantillon test
> appri=setdiff(1:npop,testi) # indices complémentaires de l'échant. d'apprentis
# Construction des échantillons avec les variables explicatives .
> datapq=data[appri,] # construction de l'échantillon d'apprentissage
> datestq=data[testi,] # construction de l'échantillon test
> summary(datapq) # vérifications
> summary(datestq)
```

Ce code doit bien évidemment être adapté à votre jeu de données.

3. Estimer le modèle à l'aide de l'échantillon d'apprentissage. Commenter les résultats obtenus. On discutera notamment sur les effets estimés obtenus pour chaque variable explicative.

4. A l'aide de l'échantillon test, étudier le pouvoir prédictif du modèle. On pourra étudier les résidus et les différents critères de performances fournis par les fonctions suivantes :

```
source(Perfo.R)
source(TabDep.R)
Perfo(ozon, ozonprev)
TabDep(ozon, ozonprev, 130,180,130)
```

où ozonprev désigne l'ozone prévu par la modèle que vous aurez construit.

5. Comparer les résultats obtenus avec ceux que l'on obtiendrait si l'on modélisait le maximum de la concentration en ozone à l'aide d'un modèle de régression linéaire en les 3 variables explicatives à disposition.