

# Scénario: Exploration d'une série chronologique de températures

## Résumé

*Scénario d'analyse d'un jeu de données : la série des températures centrales en Grande-Bretagne. Description et classification de courbes de température. Utilisation des différentes techniques descriptives uni, bi et multidimensionnelles : ACP, k-means, CAH.*

## 1 Objectif

Le thème principal de ce projet est d'explorer une question très actuelle : le réchauffement climatique. L'objectif est d'illustrer cette question par des techniques exploratoires appliquées à la plus ancienne séries d'observations climatologiques. En moyenne, les dernières années apparaissent comme plus chaudes. L'idée est d'essayer de caractériser plus précisément ce qui explique ce réchauffement moyen : les températures d'hiver ? celles de l'été ? plus ou moins de valeurs extrêmes dans un sens ou dans l'autre ?

## 2 Données

La série CET des moyennes mensuelles des températures centrales en Angleterre débute en 1659 C'est la plus longue des séries de températures enregistrées disponibles pour des études climatiques. Elle est régulièrement mise à jour par le Bureau Météorologique du Royaume Uni. Elle représente une moyenne calculée sur plusieurs stations du centre de l'Angleterre car cela permet, entre autres, de suppléer à des valeurs manquantes.

Les données sont issues du site `hadobs.metoffice.com/hadcet/` ; elles sont accessibles dans le fichier `cetm.dat` du répertoire "data". Cette série est celle des moyennes mensuelles mais les données journalières sont également disponibles. Beaucoup d'autres approches peuvent donc être abordées plus que l'étude de la moyenne mensuelle, celle des valeurs sup, des inf, des écarts ou encore des dispersions.

## 3 Travail à réaliser

Différents outils sont disponibles pour répondre à la question posée, il s'agit de les mettre en œuvre afin d'y répondre de façon pertinente. Trois phases seront considérées. Celles-ci sont décrites dans leurs grandes lignes mais les initiatives personnelles sont bienvenues.

### 3.1 Description élémentaire de la série

- Représenter simultanément les distributions des moyennes mensuelles (boxplot), commenter.
- Extraire la série chronologique des moyennes annuelles, la représenter conjointement avec une courbe d'approximation par lissage (fonction `loess`)
- Même chose avec la série des moyennes mensuelles ainsi que celle des anomalies climatiques : températures moyennes mensuelles auxquelles moyenne sur la période a été retranchée.
- Comparer la tendance linéaire calculer sur une moyenne des mois d'hiver avec la tendance linéaire des mois d'été.
- Tracer quelques courbes de températures pour différentes années. Celles-ci apparaissent nettement plus irrégulières que les données précédemment étudiées décrivant les températures des villes en France. En effet, ces dernières sont des moyennes sur dix ans des moyennes mensuelles ; elles sont en conséquence beaucoup plus lisses. Un lissage spline va être utilisé sur la série anglaise afin de gommer le "bruit" parasite.
- Utiliser la fonction en annexe pour tracer simultanément une courbe d'observations avec les lissages splines pour différentes valeurs du coefficient de pénalisation.

### 3.2 Analyse en composantes principales

L'étude précédente fait clairement apparaître un comportement particulier des, disons, 30 dernières années. L'objectif de cette section va être d'en préciser les caractéristiques par une analyse en composantes principales.

- Calculer l'ACP centrée non réduite des données brutes et représenter les graphiques associés : décroissance des valeurs propres, boîtes parallèles des composantes, biplot, fonctions propres (vecteurs propres en fonction du temps).

- Lisser toutes les lignes de la matrice (courbes de températures mensuelles des données) avec une valeur du paramètre de lissage puis recalculer l'ACP.
- Observer le comportement des valeurs propres pour différentes valeurs de ce paramètre. Choisir une valeur permettant une bonne séparation des deux premières valeurs propres, c'est-à-dire un bon rapport "signal sur bruit" en deux dimensions. Commenter les graphes de l'ACP dans ce dernier cas. Interpréter l'axe 1 et l'axe 2.
- Construire plus précisément un graphique permettant de mettre en évidence les 30 dernières années. Interpréter leur position dans le plan, par rapport aux deux axes, pour apporter des éléments de réponse à la question posée.

### 3.3 Classification

- Opérer une [classification hiérarchique](#) des courbes lissées précédentes en testant deux distances : celle euclidienne usuelle et celle basée sur les corrélations. L'objectif de cette dernière est de s'attacher à rapprocher des formes de courbes plutôt que leurs valeurs respectives.
- Représenter les classes dans le plan principal de l'ACP, du MDS ou encore de l'AFD ainsi que les groupes des courbes superposées par classe avec leur barycentre.

## Annexe : quelques outils en R

```
# prise en compte des données
cetmA=read.table("cetm.dat")
moyan=cetm[,13]
cetm=cetmA[,-13]
boxplot(cetm)

# lissages variés
ts.plot(cetA)
lines(lowess(cetA))
ceth=cetm[,c(1,2,11,12)]
cete=cetm[,c(5,6,7,8)]
```

```
cetem=apply(cete,1,mean)
ts.plot(cetem)
lines(lowess(cetem))
cethm=apply(ceth,1,mean)
ts.plot(cethm)
lines(lowess(cethm))

# anomalies climatiques
cet_anomalie=t(t(cetem)-apply(cetem,2,mean))
anomalie.ts=ts(as.vector(cet_anomalie))
ts.plot(anomalie.ts,col=5)
lines(lowess(anomalie.ts,f=1/5),col=1)
cet.ts=as.ts(as.matrix(t(as.matrix(cetem))))

# fonction de lissage spline des lignes d'une matrice
lsm=
function(y, spar)
{
  n <- nrow(y)
  p <- ncol(y)
  ychap <- y
  for(i in 1:n) {
    ychap[i, ] <- smooth.spline(1:12,y[i, ],
      spar=spar)$y
  }
  ychap
}

# Utilisation et visualisation de courbes lissées
sm04.cet=lsm(cetm,0.4)
ts.plot(t(rbind(cetm[100,],sm04.cet[100,])))
plot.ts(t(sm04.cet[100:107,]))

# ACP en reliant les 18 dernières années
acp=princomp(lsm(cetm,spar=0.5))
```

```
plot(acp$scores[,1:2])
lines(acp$scores[330:348,1:2])
abline(h=0)
abline(v=0)

# les vecteurs propres associés aux 10
# premières valeurs propres
plot.ts(acp$loadings[,1:10])
```