

# Laboratoire évalué : séries temporelles

## Contents

Données . . . . .	1
1. Modèle ARIMA pour le NEP . . . . .	2
2. Modèle ARIMA pour le NEP avec prédicteurs externes . . . . .	2

Ce laboratoire devra être soumis sur Moodle le 21 avril.

## Données

Pour cet exercice, nous utiliserons les données d'une tour à flux située dans une forêt d'épinettes noires près de Chibougamau.

**Référence bibliographique:** Bergeron, Margolis, Black, Coursolle, Dunn, Barr, & Wofsy. (2007). Comparison of carbon dioxide fluxes over three boreal black spruce forests in Canada. *Global Change Biology*, 13(1), 89–107. <https://doi.org/10.1111/j.1365-2486.2006.01281.x>.

Les tours à flux mesurent la productivité nette d'un écosystème et la quantité de gaz (CO<sub>2</sub>, H<sub>2</sub>O et CH<sub>4</sub>) échangée entre l'atmosphère et l'écosystème à l'aide de la méthode de covariance des tourbillons (*eddy covariance*).

**Weblink:** <https://www.battelleecology.org/data-collection/flux-tower-measurements>

Nous commencerons par charger les packages requis et les données.

```
library(fpp3)
library(dplyr)
library(ggplot2)
library(cowplot)
EOBS_fluxnet <- read.csv("../donnees/EOBS_fluxnet2.csv")
head(EOBS_fluxnet)
```

```
##   Year Day GapFilled_NEP GapFilled_R GapFilled_GEP TimeSteps
## 1 2004   1    -0.3702583    0.3702583           0         48
## 2 2004   2    -0.3226569    0.3226569           0         48
## 3 2004   3    -0.3143513    0.3143513           0         48
## 4 2004   4    -0.3108769    0.3108769           0         48
## 5 2004   5    -0.3105173    0.3105173           0         48
## 6 2004   6    -0.3069446    0.3069446           0         48
```

Les colonnes sont :

- *Year* est l'année de l'observation
- *Day* est le jour de l'année (1-365)
- *GapFilled\_NEP* est la productivité nette quotidienne de l'écosystème (*umol C m-2 of stand s-1*)
- *GapFilled\_R* est la respiration quotidienne de l'écosystème (*umol C m-2 of stand s-1*)
- *GapFilled\_GEP* est la productivité brute quotidienne de l'écosystème (*umol C m-2 of stand s-1*)
- *TimeSteps* est un entier indiquant le nombre de données demi-horaires qui composent les moyennes quotidiennes

## 1. Modèle ARIMA pour le NEP

(1a) Créez un tableau de données temporelles (*tsibble*). Dans un premier temps, vous devez ajouter une colonne contenant la date en utilisant les informations dans les colonnes *Year* et *Day*. Consultez le site Web suivant pour comprendre comment gérer les données temporelles dans *R* : <https://www.stat.berkeley.edu/~s133/dates.html>

(1b) L'un des problèmes de l'utilisation des données journalières est la gestion des années bissextiles. Dans ce cas, nous avons des données avec un nombre de jours constant par année (365 jours par an). Il s'agit d'une solution courante pour simplifier le traitement des données, notamment en modélisation. Afin d'ajouter un jour de plus pour les années bissextiles, nous pouvons utiliser les fonctions *fill\_gaps* ([https://www.rdocumentation.org/packages/tsibble/versions/1.0.0/topics/fill\\_gaps](https://www.rdocumentation.org/packages/tsibble/versions/1.0.0/topics/fill_gaps)) et *tidyr::fill* (<https://www.rdocumentation.org/packages/tidyr/versions/1.1.3/topics/fill>). Nous pouvons spécifier que les lignes ajoutées ont *Day* égal à 366 et *GapFilled\_NEP*, *GapFilled\_R*, *GapFilled\_GEP* et *Year* égaux à la valeur de la ligne précédente.

(1c) Obtenez un nouveau tableau de données temporelles (*tsibble*) contenant les valeurs mensuelles moyennes de *GapFilled\_NEP*. Tracez la série chronologique obtenue et commentez-la. Comment la série chronologique varie-t-elle dans le temps ? Que signifient les valeurs négatives ?

(1d) Tracez les 3 séries chronologiques des valeurs quotidiennes (*GapFilled\_NEP*, *GapFilled\_R*, *GapFilled\_GEP*), la saisonnalité annuelle de *GapFilled\_NEP* (utilisez le jeu de données journalières ainsi que le jeu de données mensuelles en produisant deux graphiques distincts), et la tendance des données *GapFilled\_NEP* pour chaque mois sur la période d'observation (utilisez le jeu de données mensuelles). Quand la saison de croissance commence-t-elle et se termine-t-elle au site d'étude ? Quand se produit le pic de photosynthèse ? Y a-t-il une tendance évidente dans les valeurs mensuelles moyennes ?

(1e) Extraire les différentes composantes de la série chronologique journalière de *GapFilled\_NEP* (tendance, saisonnalité et résidus). Quelle est l'importance relative des composantes ? Qu'est-ce que ça veut dire ? Enfin, stockez les composants dans un nouveau tableau de données temporelles (*tsibble*).

(1f) Analyser l'autocorrélation et l'autocorrélation partielle de la série chronologique journalière de *GapFilled\_NEP* et de la composante résiduelle extraite en 1e. Que déduisez-vous de ces graphiques ?

(1g) Ajustez un modèle ARIMA à la série chronologique journalière de *GapFilled\_NEP*. Laissez le modèle choisir le modèle ARIMA approprié. Quel type de modèle ARIMA est automatiquement sélectionné ? Les résidus du modèle ARIMA respectent-ils les hypothèses du modèle ?

(1h) Effectuez des prévisions et tracez une année supplémentaire de données journalières de *GapFilled\_NEP* en utilisant le modèle sélectionné à l'étape précédente.

## 2. Modèle ARIMA pour le NEP avec prédicteurs externes

Nous commencerons par charger des données météorologiques pour le site de la tour à flux.

```
My_meteo=read.delim("../donnees/EOBS_fluxnet_inmet2.txt",skip=1,header=F)
names(My_meteo)= c("Year","Day","Tmax","Tmin","Precip","CO2")
head(My_meteo)
```

```
##   Year Day      Tmax      Tmin Precip      CO2
## 1 2004   1 -14.280000 -20.70000  0.306 384.4005
## 2 2004   2 -13.340000 -17.90000  0.203 384.3611
## 3 2004   3  -1.859991 -13.34000  0.401 384.3216
## 4 2004   4  -8.299994 -24.65999  0.000 384.2822
## 5 2004   5 -18.000010 -28.14001  0.157 384.2427
## 6 2004   6 -17.900000 -20.32000  0.109 384.2033
```

Les colonnes sont :

- *Year* est l'année de l'observation
- *Day* est le jour de l'année (1-365)
- *Tmax* est la température maximale journalière (°C)
- *Tmin* est la température minimale journalière (°C)
- *Precip* est la somme des précipitations journalières (cm)
- *CO2* est la concentration en CO2 de l'atmosphère (ppm)

Une fois que vous avez chargé les données météorologiques, vous devez créer un tableau de données temporelles avec ces données (même procédure qu'au point 1a) et combler les années bissextiles en ajoutant un jour (même procédure qu'au point 1b).

(2a) Trouvez la variable météorologique ou environnementale (*Tmax*, *Tmin*, *Precip* ou *CO2*) qui corrèle le plus à la série chronologique journalière de *GapFilled\_NEP*.

(2b) Joignez les données de flux et météorologiques (*inner\_join*) et tracez la relation (nuage de points) entre la variable trouvée en 2a (axe des x) et *GapFilled\_NEP* (axe des y). Commentez cette relation.

(2c) Appliquez un modèle linéaire sans aucun terme ARIMA incluant un terme quadratique pour la variable trouvée en 2a. Ensuite, tracez les prévisions du modèle sur le graphique du point 2b. Enfin, tracez les résidus de ce modèle en fonction du temps. Le modèle représente-t-il correctement les données ? Essayez d'expliquer pourquoi les résidus semblent être principalement positif au début de la saison de croissance.

(2d) Ajouter des termes ARIMA au modèle du point 2c et comparer l'AIC de ce modèle avec l'AIC du modèle en 1g. Comparez les termes ARIMA sélectionnés ici et en 1g.