Time series - Lab 2

Contents

Data

For this exercise we will use data from a flux tower located in a black spruce forest near Chibougamau.

> **Reference**: 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.

Flux towers measure net ecosystem exchange or the amount of gas that is exchanged between the atmosphere and the ecosystem using eddy covariance technique.

> **Weblink**: https://www.neonscience.org/3d-interactive-flux-tower

> **Weblink**: https://www.youtube.com/watch?v=CR4Anc8Mkas

> **Weblink**: https://www.neonscience.org/data-collection/meteorology

We will start loading the required packages and the data.

```r
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
```

The columns are:

- *Year* is the year of the observation
- *Day* is the day of the year of the observation (1-365)
- *GapFilled_NEP* is the daily net ecosystem productivity (*umol C m-2 of stand s-1*)
- *GapFilled_R* is the daily ecosystem respiration (*umol C m-2 of stand s-1*)
- *GapFilled_GEP* is the daily gross ecosystem productivity (*umol C m-2 of stand s-1*)

- *TimeSteps* is an integer saying how many half hourly data composed the daily aggregates

1. Process and explore NEP time series

(1a) Create a temporal data frame (*tsibble*). As a first step, you must add a column containing the date using the information in *Year* and *Day*. Consult the following website to understand how to deal with date/time data in *R*: https://www.stat.berkeley.edu/~s133/dates.html (1 point)

(1b) One of the problems working with daily data is to deal with leap years. In this case we load data with constant 365 days per year. This is a common solution to simplify the data processing, especially in modelling. In order to add one more day per each leap year we can use the functions *fill_gaps* (https://www.rdocumentation.org/packages/tsibble/versions/1.0.0/topics/fill_gaps) and *tidyr::fill* (https://www.rdocumentation.org/packages/tidyr/versions/1.1.3/topics/fill). We can specify that the added rows have *Day* equal to 366 and *GapFilled_NEP*, *GapFilled_R*, *GapFilled_GEP*, and *Year* equal to the value of the preceding row. (1 point)

(1c) Obtain a new temporal data frame (*tsibble*) containing mean monthly values of *GapFilled_NEP*. Plot the obtained time series and comment it. How does the time series vary over time? What do negative values mean? (1 point)

(1d) Plot the 3 time-series of daily values (*GapFilled_NEP*, *GapFilled_R*, *GapFilled_GEP*), the annual seasonality of *GapFilled_NEP* (use the daily dataset as well as the monthly dataset providing two distinct plots), and the trend of *GapFilled_NEP* data for each month over time (use the monthly dataset). When does the growing season start and end at the study site? When does the peak of photosynthesis occur? Is there any evident trend in the mean monthly values? (1 point)

(1e) Extract the several components of the *GapFilled_NEP* daily time series (trend, seasonality, and residuals). What is the components' relative importance? What does it mean? Finally, store the components into a new temporal data frame (*tsibble*). (1 point)

(1f) Analyze the autocorrelation and the partial autocorrelation of the *GapFilled_NEP* daily time series and of its residual component extracted in 1e. What do you deduce from these plots? (1 point)

2. GAM model for NEP

(2a) Fit a Generalized Additive Model (GAM) on *GapFilled_NEP* using Day of Year (DOY) as a smooth term. Plot the estimated smooth function. Evaluate whether the model residuals meet key assumptions: Homoscedasticity (residuals vs fitted), Normality (QQ-plot), Absence of temporal autocorrelation (time series plot and ACF). (1 point)

(2b) Extend the previous GAM by incorporating temporal autocorrelation in the residuals using an AR correlation structure (p = 1, q = 0). Which model performs best (with or without temporal autocorrelation)? Why? Do the residuals of the best model meet the model's assumptions (homoscedasticity, normality, no remaining temporal autocorrelation)? (1 point)

3. GAM model for NEP with external predictors

We will start loading meteorological data for the flux tower site.

```
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
```

The columns are:

- *Year* is the year of the observation
- *Day* is the day of the year of the observation (1-365)
- *Tmax* is the daily maximum temperature ($°C$)
- *Tmin* is the daily minimum temperature ($°C$)
- *Precip* is the daily precipitation sum (cm)
- *CO2* is daily CO2 concentration (ppm)

Once you have loaded the meteorological data you must create a temporal data frame with these data (same procedure than 1a) and gap fill these data (same procedure than 1b).

```
My_meteo = mutate(My_meteo, Date = as.Date(paste(My_meteo$Year,My_meteo$Day), format='%Y %j'))
My_meteo = as_tsibble(My_meteo, index = Date)
My_meteo = My_meteo %>%
  fill_gaps(Day = 366) %>%
  tidyr::fill(Tmax, .direction = "down") %>%
  tidyr::fill(Tmin, .direction = "down") %>%
  tidyr::fill(Precip, .direction = "down") %>%
  tidyr::fill(CO2, .direction = "down") %>%
  tidyr::fill(Year, .direction = "down")
```

(3a) Find the meteorological or environmental variable (*Tmax*, *Tmin*, *Precip* or *CO2*) that correlates the most with the *GapFilled_NEP* daily time series. (1 point)

(3b) Join the flux and meteorological tables (*inner_join*) and plot the relationship (scatterplot) between the variable found in 3a (x-axis) and *GapFilled_NEP* (y-axis). Comment on this relationship. (1 point)

(3c) Extend the previous GAM model (2b) including a quadratic term for the variable found in 2a. Inspect and interpret the output of the new GAM model (summary function). Compare the AIC of this new model to the models from 2a and 2b. Visualize the model predictions (fitted values) against the observed data, using: a scatterplot of GapFilled_NEP vs the used climate variable; a scatterplot of GapFilled_NEP vs DOY (Day of Year). (2 point).