Graded lab: Time series

Contents

This lab should be submitted to Moodle on April 22nd.

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.battelleecology.org/data-collection/flux-tower-measurements

We will start loading the required packages and the data.

```
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. ARIMA model for NEP

(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

(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.

(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?

(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?

(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*).

(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?

(1g) Adjust an ARIMA model to the *GapFilled_NEP* daily time series. Let the model choose the appropriate ARIMA model. What kind of ARIMA model is automatically selected? Do the the ARIMA residuals meet the model's assumptions?

(1h) Forecast and plot one additional year of *GapFilled_NEP* daily data based on the selected model in the previous step.

2. ARIMA 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).

(2a) Find the meteorological or environmental variable (*Tmax*, *Tmin*, *Precip* or *CO2*) that correlates the most with the *GapFilled_NEP* daily time series.

(2b) Join the flux and meteorological tables (*inner_join*) and plot the relationship (scatterplot) between the variable found in 2a (x-axis) and *GapFilled_NEP* (y-axis). Comment on this relationship.

(2c) Apply a linear model without any ARIMA terms including a quadratic term for the variable found in 2a. Then, plot the model predictions on the plot of 2b. Finally, plot the residuals of this model as a function of time. Does the model represent the data correctly? Try to explain why the residuals seem to be mostly positive at the beginning of the growing season.

(2d) Add ARIMA terms to the model of 2c and compare the AIC of this model with the AIC of the model in 1g. Compare the ARIMA terms selected here and in 1g.