eLTER Training
Time series analysis for ecologists
Wed 1st of Dec 2021 14–16 CET

CHAT questions and answers:

- **The data itself or the residuals should be normal?**
  - both, by inspecting the residuals we can see if the data were correctly modelled considering normailty

- **what is the difference between log and log10?**
  - log is a natural logarithm in R. log --> log(e). I think
  - natural logarithm (ln) is the logarithm with the Euler number as base (e). log10 has the base 10 ;-)

- **In which case do you use multiplicative or additive TS?**
  - It is the data itself that defines it, not that the user choses
    - This could be a good reference for multiplicative time series example: https://www.abs.gov.au/websitedbs/d3310114.nsf/home/time+series+analysis:+the+basics

- **What happens after you find a type of autocorrelation? You use other ways of trend analysis?**
  - once you find autocorrelation you have to include it in your model (e.g. use gls(dependent variable ~time, correlation =corAR1())

- **Could you please explain again the acf principle**
  - correlate the time series with itself, then you start to shift it by lag, if subsequent observations influence themselves, you observe the autocorrelation. Iteratively you change the lag distance and see how far this influence can reach (=how data auto-correllate = acf!)

- **is "rcompanion" the good package name? I don't find it**
  - you need to update your R, I had the same issue
  - it is only available for the newest R versions
  - UO: you can find on Cran also the older versions of a package that work with your R-version

- **When looking for autocorrelation in the residuals: is it important to always look at the acf plot, or can I just check the pacf?**
  - invest those 10 seconds to see the ACF (minor effort), because both ACF and PCF are like twins and the more you understand your data, the better it is

- **don't we have to use - type- normalize?**
  - normalize has to be used *only!!!* when you use gls!!!!!!, when you do normal lm, use normal residuals

- **how can I put into a graph arima regression, with IC?**
  - this is shown on L279 of the skript

- **If you have a time series for several variables (e.g. tree growth, temperature and rainfall) and you want to explain one variable with the other ones, do you need to look at acf for each variable or only the variable you want to explain?**
  - acf for dependent variable, I think… and acf for residuals of the model
  - UO: you specify your model (let`s say: tree.growth ~temperature + rainfall) and then you inspect the residuals, when you find autocorrelations, you have to start thinking which variable is responsible for this; some lagged effect of rain? Try to include a correlation structure, ecc and inspect the residuals

- **Hi! A breakpoint could be a change in a land management?**
  - UO: yes

- **what other change point tests do you recommend?**
  - On L387 of the script there is a link to a page that does a good job of comparing methods

- **Can I test if there are breakpoints in the timeseries and where they are? E.g. that would give info where statistically significant jumps occured. Can breakpoints do it too?**
  - Yes, have a look at the script from L424 downwards

- **What should be the minimum length of the time series to be able to detect auto-correlation? Min 10 years, or 20 years, 50 years? Any ideas?**
  - Length of a timeseries is not important in this context because if your data do not have it, they do not have it

- **If you have several breakpoints, do we still include an overall ACF structure for the model, or do we need to break the series into several sections?**
  - One ACF plot is enough for inspecting the whole model

- **could you recommend us any good R tools for generating/editing time series data from other formats (e.g. tables with date and data)?**
  - Data harmonisation is tricky, I do not have any recommandation

- **@detecting break points: does not this depend also on the detection threshold? the smaller it is chosen, the more break points are detected, the bigger it is the less will be detected. So the question would be what the given threshold in used script is, isn't it**
  - Yes, that is true and it depends on the user on what he/she wants to put the most weight (ecological knowledge is also important)

- **Don't most changepoint detection methods also assume that the data is Independent (so not autocorrelated)?**
  - This assumption is often assumed but in some models as shown you can introduce the autocorrelation part

- **I don't understand why k = 12 because of 12 months. k is how "wiggly" the smooth may be. that should be independent of the months per year.**
  - K is not about the wigglyness, k determines the knots

- **I would like to know if there is a method to estimate the % importance of each variable in a GAM model that contains several variables? Thank you**
  - ○ I never came across this question/modelling but interesting question; have a look at the two links in the skript

- **Why did we need to switch from GAM to GAMM?**
  - ○ The switching was needed to be able to model residual autocorrelation