

Time series

Data

The oak_seeds.csv dataset shows the number of seeds of *Quercus crispula* oak collected annually (1980-2017) by 16 traps located in a stand of this species in Japan.

```
seed <- read.csv("../donnees/oak_seeds.csv")
head(seed)
```

```
##   year trap seeds
## 1 1980    1    13
## 2 1980    2   131
## 3 1980    3    44
## 4 1980    4    44
## 5 1980    5    47
## 6 1980    6    27
```

The oak_weather.csv file contains annual weather data for that same site:

```
weather <- read.csv("../donnees/oak_weather.csv")
head(weather)
```

```
##   year temp_fl temp_gr rain_fl rain_gr
## 1 1980   14.9   15.2     75    437
## 2 1981    9.3   15.4     40    766
## 3 1982   11.5   15.8    109    487
## 4 1983   11.5   15.9     49    657
## 5 1984   13.4   17.1     49    622
## 6 1985   11.5   16.9     63    501
```

- *temp_fl*: Mean temperature (°C) during the flowering period of the tree.
- *temp_gr*: Mean temperature (°C) during the growing season.
- *rain_fl*: Total amount of rain (in mm) during the flowering period of the tree.
- *rain_gr*: Total amount of rain (in mm) during the growing season.

These data come from the following study:

Shibata, M., Masaki, T., Yagihashi, T., Shimada, T., & Saitoh, T. (2019). Data from: Decadal changes in masting behaviour of oak trees with rising temperature. Dryad Digital Repository. <https://doi.org/10.5061/dryad.v6wwpzgrb>

1. Site-level time series

- (a) Calculate the total number of seeds collected per year (all traps combined) and apply a square root transformation to the result. Convert the result into a temporal data frame (*tsibble*) and view the resulting time series.

Note: Since we will be using linear rather than generalized models in this exercise, the square root transformation is intended to stabilize the variance of the count data.

- (b) Visualize the temporal correlations for this series. What type of ARIMA model (AR and/or MA, as well as their order) might be appropriate here?
- (c) Fit an ARIMA model by letting the function automatically choose the type and order of the model. What do the estimated coefficients mean?
- (d) Join the `weather` dataset and fit an ARIMA model with the four weather variables as external predictors. Do these variables help to better predict the number of seeds produced per year?
- (e) What type of model is chosen by `ARIMA()` if you consider only the sub-series starting in the year 2000, without an external predictor? Explain this choice from the graph in (a) and the temporal correlations for this subseries.
- (f) Calculate the forecasts from the models in (c) and (e) for the next five years. How do these forecasts differ?

2. Trap-level time series

- (a) Go back to the original table showing the number of seeds per year and trap, then apply the square root transformation to the number of seeds. Then use the `lme` function from the *nlme* package to fit a linear mixed model including: the fixed effect of weather variables, the random effect of the trap and the temporal correlations from one year to another.

Here is an example of how to specify a random effect of a `GROUP` variable on the intercept of a `lme` model, as well as an ARMA correlation between successive elements of the same `GROUP`:

```
library(nlme)
mod_lme <- lme(..., data = ...,
              random = list(GROUP = ~1),
              correlation = corARMA(p = ..., q = ..., form = ~ 1 | GROUP))
```

- (b) Compare the accuracy of the fixed effects in this model in (a) to the model in 1(d). What is the reason for this difference?