

Robust regression - Graded lab

This work must be submitted before **Friday February 17th** at 5 p.m. on Moodle.

Numbers in parentheses indicate the number of points for each question.

Total: 10 points.

Data

We will use the dailyTWD.csv table containing data on Relative Extractable Water in the soil (REW) and Tree Water Deficit (TWD) for 50 trees, 4 sites, 3 species (EPN, PET and PIG) and 2 soil types (clay and sand).

```
data = read.csv("../donnees/dailyTWD.csv")
data[,1] = as.factor(data[,1])
data[,2] = as.factor(data[,2])
data[,3] = as.factor(data[,3])
data[,4] = as.factor(data[,4])
str(data)
```

```
## 'data.frame': 7800 obs. of 7 variables:
## $ tree : Factor w/ 50 levels "Dendro_argile_authier_2021.SEVolt_Avg.1.",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ site : Factor w/ 4 levels "Dendro_argile_authier_2021",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ species : Factor w/ 3 levels "EPN","PET","PIG": 1 1 1 1 1 1 1 1 1 1 ...
## $ type : Factor w/ 2 levels "clay","sand": 1 1 1 1 1 1 1 1 1 1 ...
## $ day : chr "2021-05-26" "2021-05-27" "2021-05-28" "2021-05-29" ...
## $ REW20mean: num 0.8 0.868 0.866 0.851 0.826 ...
## $ TWDmaxstd: num 0.0776 0.2692 0.7661 1.0844 1.4131 ...
```

REW is a metric of available water in the soil for plant growth (0 to 1). See

Granier, A., Bréda, N., Biron, P., & Villette, S. (1999). A lumped water balance model to evaluate duration and intensity of drought constraints in forest stands. *Ecological Modelling*, 116(2), 269–283. doi: 10.1016/S0304-3800(98)00205-1

TWD is computed from half hourly dendrometer data recording stem radius variations (here the data are standardized relative to a reference period; TWD >1 is greater than reference). See

Knüsel, S., Peters, R. L., Haeni, M., Wilhelm, M., & Zweifel, R. (2021). Processing and Extraction of Seasonal Tree Physiological Parameters from Stem Radius Time Series. *Forests*, Vol. 12, p. 765. doi: 10.3390/f12060765

All data are daily averages

1. The relationship between REW and TWD

- a. Data formatting. Convert day information on an object facilitating manipulation of dates and times (use the `as.POSIXct` function; we will see later in the course the different options to handle time series). Then, subset the data to only retain days \geq “2021-07-15” and \leq “2021-09-12”. You can overwrite the old matrix “data”. **(1)**
- b. Data exploration. With `ggplot`, plot REW (REW20mean) against time (day), using one color per site (line chart). **(1)**
- c. Data exploration. Group the data by site, day and species, then compute the mean “TWDmaxstd” for each group. In this way, you average the data of different trees. Subsequently, plot the time series of TWD only for the PIG species using different colours per different sites as in b. **(1)**
- d. Interpret previous time series (in b and c) and describe with a short text the patterns and relations you see. **(1)**
- e. Plot a scatter plot (dots) using the full “data” matrix (result of a) showing REW in the x-axis and TWD in the y-axis. Only plot data for the PIG species. Use the option `theme(aspect.ratio = 1)` of `ggplot`. **(1)**
- f. Describe the relations you see in (e) and define which robust regression technique may be appropriated for these data and why. **(1)**
- g. Run a quantile regression of TWD (only the PIG species) as a function of REW using the 1st and 9th deciles (0.1 and 0.9), the 1st and 3rd quartiles (0.25 and 0.75) and the median. Then, interpret the regression coefficients for each quantile using the summary, the plot of the summary and a short text you will write. **(1)**
- h. Run three different quantile regressions of TWD as a function of REW. One for EPN, one for PET and one for PIG. Only consider the 9th decile. Using the `summary` function, compare the regression coefficients and their confidence intervals among species, writing main interpretations. **(1)**
- i. Plot a scatter plot (dots) using the full “data” matrix (result of a) showing REW in the x-axis and TWD in the y-axis. Here, contrary to what you have done in e, you will plot all three species together with different colours. Add the result of quantile regressions (1st and 9th deciles) for each species in the same plot. **(2)**