

# Régression robuste aux valeurs extrêmes - Laboratoire évalué

Ce travail doit être remis avant **vendredi 17 février** à 17h sur Moodle.

**Les chiffres entre parenthèses indiquent le nombre de points pour chaque question.**

**Total : 10 points.**

## Données

Nous utiliserons le tableau dailyTWD.csv contenant des données sur la teneur en eau relative du sol (REW) et le déficit en eau des arbres (TWD) pour 50 arbres, 4 sites, 3 espèces (EPN, PET et PIG) et 2 types de sol (argile et sable).

```
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 1
## $ site : Factor w/ 4 levels "Dendro_argile_authier_2021",...: 1 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 1 ...
## $ type : Factor w/ 2 levels "clay","sand": 1 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 est une mesure de l'eau disponible dans le sol pour la croissance des plantes (0 à 1). Voir:

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

Le TWD est calculé à partir des données demi-horaires de dendromètres enregistrant les variations du rayon des tiges (ici, les données sont normalisées par rapport à une période de référence ;  $TWD > 1$  est supérieur à la référence). Voir

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

Toutes les données sont des moyennes journalières

## 1. La relation entre REW et TWD

- a. Préparation des données. Convertir les informations temporelles (“day”) en objets facilitant la manipulation des dates et des heures (utilisez la fonction `as.POSIXct` ; nous verrons plus loin dans le cours les différentes options pour gérer les séries temporelles). Ensuite, créer un sous-ensemble de données en conservant que les jours  $\geq$  “2021-07-15” et  $\leq$  “2021-09-12”. Vous pouvez écraser l’ancienne matrice “data”. **(1)**
- b. Exploration des données. Avec `ggplot`, tracer le REW (REW20mean) en fonction du temps (day), en utilisant une couleur par site (graphique en ligne). **(1)**
- c. Exploration des données. Regrouper les données par site, jour et espèce, puis calculez la moyenne de “TWDmaxstd” pour chaque groupe. De cette façon, vous faites la moyenne des données de différents arbres. Par la suite, tracer les séries temporelles de TWD uniquement pour l’espèce `PIG` en utilisant différentes couleurs pour différents sites comme en b. **(1)**
- d. Interpréter les séries temporelles précédentes (en b et c) et décrire avec un court texte les patrons et les relations que vous voyez. **(1)**
- e. Tracer un nuage de points en utilisant la matrice “data” complète (résultat de a) montrant REW sur l’axe des x et TWD sur l’axe des y. Uniquement tracer les données pour l’espèce `PIG`. Utiliser l’option `theme(aspect.ratio = 1)` de `ggplot`. **(1)**
- f. Décrire les relations en (e) et définir quelle technique de régression robuste peut être appropriée pour ces données et pourquoi. **(1)**
- g. Exécuter une régression quantile de TWD (uniquement l’espèce `PIG`) en fonction de REW en utilisant les 1er et 9e déciles (0,1 et 0,9), les 1er et 3e quartiles (0,25 et 0,75) et la médiane. Ensuite, interpréter les coefficients de régression pour chaque quantile à l’aide de `summary`, du graphique de `summary` et d’un court texte que vous rédigerez. **(1)**
- h. Exécuter trois régressions quantiles différentes de TWD en fonction de REW. Une pour `EPN`, une pour `PET` et une pour `PIG`. Ne considérer que le 9e décile. À l’aide de la fonction “summary”, comparez les coefficients de régression et leurs intervalles de confiance entre les espèces, en écrivant les principales interprétations. **(1)**
- i. Tracer un nuage de points en utilisant la matrice “données” complète (résultat de a) montrant REW sur l’axe des x et TWD sur l’axe des y. Ici, contrairement à ce que vous avez fait en e, vous tracerez les trois espèces ensemble avec des couleurs différentes. Ajouter, ensuite, le résultat des régressions quantiles (1er et 9e déciles) pour chaque espèce dans la même graphique. **(2)**