

Ajustement du modèle de DE RÉAUMUR (1735) aux données publiées (1735-1740) et manuscrites (1734-1756) de DE RÉAUMUR

Pr Jean R. Lobry

Estimation par régression non-linéaire des paramètres du modèle de DE RÉAUMUR. Intérêt des régions de confiance pour les paramètres dans la mise en évidence de la sur-paramétrisation du modèle. Illustrations des déboires possibles en régression non-linéaire avec une forte corrélation structurelle entre les paramètres, une mesa et un minimum local.

Contents

1	Introduction	3
1.1	Le modèle de DE RÉAUMUR	3
1.2	Implémentation du modèle	3
1.3	La somme des carrés des écarts	4
1.4	Estimation des paramètres	4
1.5	Région de confiance pour les paramètres	4
1.6	Les données de température	4
2	Données publiées (n = 6)	4
2.1	Les données	4
2.2	Estimation ponctuelle des paramètres	5
2.3	Région de confiance des paramètres	6
2.3.1	Corrélation structurelle entre les paramètres.	7
2.3.2	Présence de mesa	8
2.3.3	Surparamétrisation du modèle	9
2.3.4	Minimum local	10
3	Données manuscrites (n = 23)	12
3.1	Les données	12
3.2	Estimation ponctuelle des paramètres	12
3.3	Région de confiance	13
3.4	Artillerie lourde (n = 208)	14

4	Annexes	18
4.1	Données phénologiques	18
4.1.1	Origine des données	18
4.1.2	Positionnement des données d'intérêt	19
4.1.3	Conversion en jour de l'année	21
4.1.4	Sauvegarde post-prétraitement	21
4.2	Données météorologiques publiées	21
4.2.1	Origine des données	21
4.2.2	Notation des degrés négatifs	23
4.2.3	Heures d'observation	24
4.2.4	Géolocalisation	26
4.2.5	Les degrés de température	27
4.2.6	Sauvegarde post-prétraitement	30
4.3	Données météorologiques manuscrites	30
4.3.1	Origine des données	30
4.3.2	Données manquantes	31
4.3.3	Le nombre d'observations	32
4.3.4	Sauvegarde post-prétraitement	32
4.4	Données météorologiques de Rousseau 2024	32
4.4.1	Origine des données	32
4.4.2	Données manquantes	32
4.4.3	Sauvegarde post-prétraitement	33

1 Introduction

1.1 Le modèle de de Réaumur

LE modèle de DE RÉAUMUR [9] cherche à prédire, pour une année donnée, la date d'un stade phénologique, par exemple la floraison, à partir des données de température journalière. Les données en entrée, T_d , sont les températures moyennes journalières, mais on censure les températures négatives (d'où le terme $\max(T_d, 0)$ dans l'équation ci-après). Il consiste à faire la somme au cours du temps des degrés positifs jusqu'à ce qu'un seuil critique soit atteint. Il ne comporte que deux paramètres : la date t_0 à partir de laquelle on commence à sommer et le seuil critique $F^* \geq 0$ qu'il faut dépasser. La date prédite, t_d , tout comme t_0 sont exprimées classiquement pour ce type de modèle par le rang du jour dans l'année.

$$t_d = \min(t^*) \mid \sum_{d=t_0}^{t^*} \max(T_d, 0) > F^*$$

1.2 Implémentation du modèle

LA fonction `dddR1()` définie ci-dessous calcule la date prédite par le modèle de DE RÉAUMUR pour une série de valeur de température donnée et un jeu de paramètre donné. Pour éviter d'avoir une fonction objective « en escalier » on fait une interpolation avec une fonction de lissage « serrée » des données météorologiques de façon à pouvoir travailler avec des t_0 et t_d pas forcément entiers.

```
# Entrées :
# d : vecteur des jours consécutifs de l'année
# Td : vecteur des températures moyennes correspondant
# param[1] : t0 : début de l'intégration
# param[2] : Seuil F* à dépasser
dddR1 <- function(d, Td, param, warn = FALSE){
  d <- as.integer(d)
  # Check parameter values
  if(length(d) != length(Td)) stop("Different length for d and Td")
  if(!all(diff(d) == 1L)) stop("Consecutive values expected for d")
  if(length(param) != 2) stop("2 parameters are expected")

  t0 <- param[1]
  # On force t0 à rester dans les jours observés
  if(t0 < d[1]) t0 <- d[1]
  if(t0 > d[length(d)]) t0 <- d[length(d)]

  Fstar <- param[2]

  # On ne conserve que les degrés positifs
  Td[Td < 0] <- 0.0

  # On calcule les degrés-jours
  dj <- cumsum(Td)


  if(dj[length(dj)] < Fstar){
    if(warn) warning("F* unreachable")
    return(d[length(d)])
  }
  lo <- loess(dj~d, span = 0.1) # Lissage serré
  f <- function(x) predict(lo, x) - Fstar - predict(lo, t0)
  tryres <- try(uniroot(f, c(d[1], d[length(d)]))$root, silent = TRUE)
  if(inherits(tryres, "try-error")) return(d[length(d)])
  return(tryres)
}
```

1.3 La somme des carrés des écarts

CETTE fonction renvoie la somme des carrés des écarts et calcule les valeurs prédites dans la colonne `théo` de la table `Reauphe` :

```
SSRR1 <- function(param, dmin = 1, dmax = 365){
  dspan <- dmin:dmax
  for(i in seq_len(nrow(Reauphe))){
    the_year <- Reauphe[i, "Year"]
    # indice du premier janvier de l'année :
    imétéo <- with(Reaumet, which(Year == the_year & doy == 1))
    Reauphe[i, "théo"] <-<- with(Reaumet,
      dddR1(dspan, Tmean[(imétéo + dmin - 1):(imétéo + dmax - 1)], param))
  }
  return(with(Reauphe, sum((Day_of_Year - théo)^2)))
}
```

1.4 Estimation des paramètres

ON utilisera ici la fonction de base  [17] `nlm()` pour trouver la valeur des paramètres qui minimise la somme des carrés des écarts. Voir la fiche de TD¹ « régression non linéaire » pour plus de détails sur les sorties de cette fonction.

1.5 Région de confiance pour les paramètres

UNE région de confiance pour la valeur des paramètres avec un risque de première espèce α est donnée [2] par l'ensemble des valeurs des paramètres telles que la somme des carrés des résidus n'excède pas un seuil donné,

$$\theta \mid S(\theta) \leq S(\hat{\theta}) \left(1 + \frac{p}{n-p} F_{p;n-p}^{\alpha}\right)$$


où p est le nombre de paramètres du modèle, n le nombre de points disponibles dans le jeu de données, et $\hat{\theta}$ le vecteur des valeurs des paramètres tel que la somme des carrés des écarts, $S(\hat{\theta})$, soit minimale.

1.6 Les données de température

IL S'AGIT de séries de températures moyennes journalières. Attention, on utilise ici les degrés utilisés par DE RÉAUMUR. Pour les convertir en degrés CELSIUS il faut [20] les multiplier par un facteur $\frac{5}{9} \times 2.066 \approx 1.148$. Cela ne change rien ici pour l'estimation de t_0 , mais l'estimation de F^* doit être multipliée par ce facteur si on la souhaite en degrés CELSIUS.

2 Données publiées (n = 6)

2.1 Les données

IL s'agit des données de température publiées par DE RÉAUMUR (*cf.* annexe page 21) et des dates de vendange (*cf.* annexe page 18) pour les années 1735 à 1740 incluses pour la région Île-de-France. Les importer dans  avec :

¹tldr46 : <https://esb.univ-lyon1.fr/pdf/tldr46.pdf>

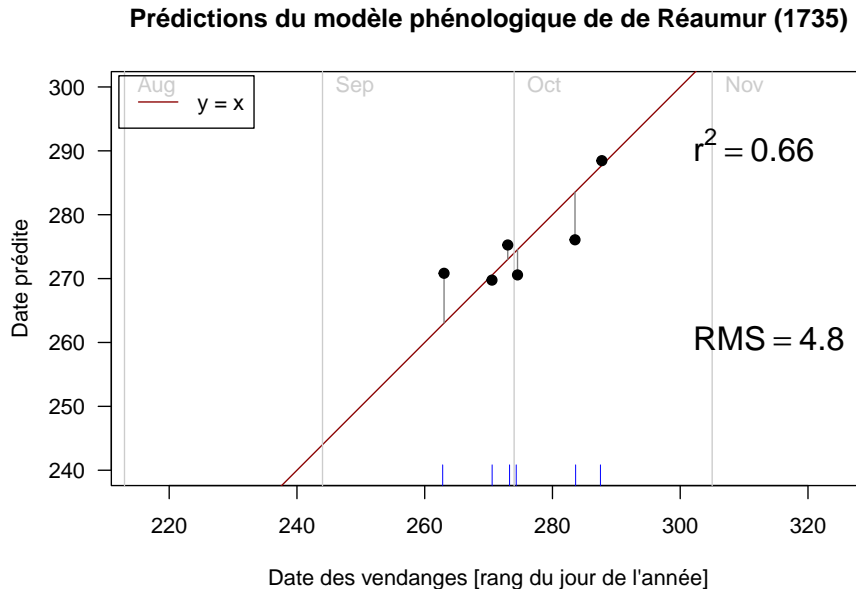
```
chmin <- "https://esb.univ-lyon1.fr/donnees/tdr4R/"
load(url(paste0(chmin, "Reaumet.Rda")))
load(url(paste0(chmin, "Reauphe.Rda")))
Reauphe <- subset(Reauphe, Year %in% 1735:1740)
```

2.2 Estimation ponctuelle des paramètres

```
(resnlm <- nlm(SSRR1, c(100, 2300)))
$minimum
[1] 138.3634
$estimate
[1] 141.4557 2006.4153
$gradient
[1] 2.852705e-08 2.144219e-09
$code
[1] 1
$iterations
[1] 28
```

ON obtient une somme des carrés des écarts minimum de 138.4 pour $t_0 = 141.5$ (soit autour du 21 mai) et $F^* = 2006.4$. La valeur de la somme des carrés des écarts n'est pas très parlante, on peut calculer l'écart quadratique moyen (23.1) et prendre sa racine carrée pour revenir aux unités de départ (4.8). Cela signifie que l'on a une précision de l'ordre de 5 jours ici pour la prédiction de la date des vendanges avec le modèle de DE RÉAUMUR. Graphiquement :

```
plotpred <- function(x, y, lims = c(240, 300), ...){
  require(lubridate)
  plot(x, y, pch = ".", las = 1, xlim = lims, ylim = lims,
       asp = 1, bg = rgb(0, 0, 0, 0.5),
       main = "Prédictions du modèle phénologique de de Réaumur (1735)",
       ylab = "Date prédite",
       xlab = "Date des vendanges [rang du jour de l'année]", ...)
  abline(c(0, 1), col = "darkred")
  segments(x, x, x, y, col = grey(0.5))
  mois <- as.Date(sprintf("2017-%0.2d-01", 1:12))
  abline(v = yday(mois), col = grey(0.8))
  text(yday(mois), rep(300, 12), month(mois, label = TRUE, abbr = T), pos = 4, col = grey(0.8))
  legend("topleft", inset = 0.01, legend = "y = x", lty = 1, col = "darkred")
  r2 <- signif(cor(x, y)^2, 2) ; if(is.na(r2)) r2 <- 0
  text(300, 290, bquote(r^2 == .(r2)), cex = 1.5, pos = 4)
  RMS <- signif(sqrt(mean((x - y)^2)), 2)
  text(300, 260, bquote(RMS == .(RMS)), cex = 1.5, pos = 4)
  points(x, y, pch = 19)
  rug(jitter(x), col = "blue", ticksize = 0.05)
}
invisible(SSRR1(resnlm$estimate)) # pour calculer les valeurs prédites
y <- Reauphe$théo ; x <- Reauphe$Day_of_Year
plotpred(x, y)
```



TOUT ceci pourrait sembler *a priori* satisfaisant, mais l'examen de la région de confiance pour la valeur des paramètres va nous montrer que nous sommes bien loin d'une situation idéale.

2.3 Région de confiance des paramètres

Nous allons explorer la somme des carrés des écarts avec une grille sur le plan des paramètres, comme les calculs peuvent être assez long, on les effectue en arrière-plan et on sauvegarde les résultats dans l'objet `matSSR.Rda` :

```
scemin <- resnlm$minimum
n <- nrow(Reauphe)
p <- 2 ; alpha <- 0.05
seuil <- scemin*( 1 + (p*qt(p = 1 - alpha, df1 = p, df2 = n - p))/(n - p) )
npts <- 128
p1seq <- seq(90, 300, le = npts)
p2seq <- seq(0, 2500, le = npts)
matSSR <- matrix(NA, nrow = length(p1seq), ncol = length(p2seq))
for(i in seq_len(nrow(matSSR))){
  for(j in seq_len(ncol(matSSR))){
    matSSR[i, j] <- SSRR1(c(p1seq[i], p2seq[j]))
  }
}
save(p1seq, p2seq, matSSR, resnlm, seuil, file = "matSSR.Rda")
```

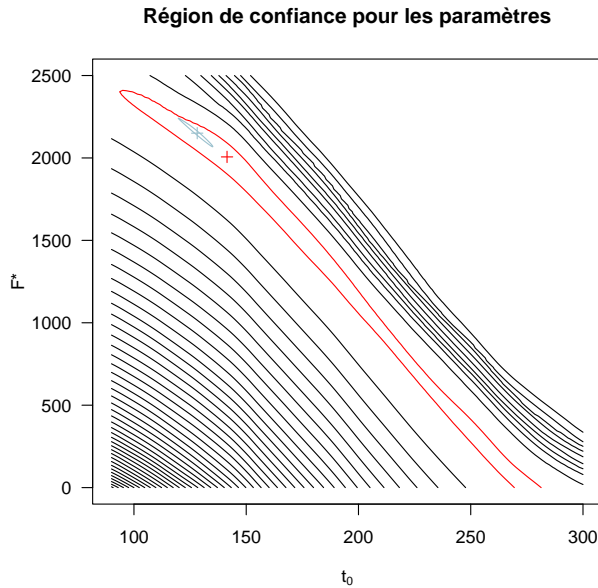
La fonction `R` de base `contour()` permet de visualiser simplement avec des courbes de niveau l'allure de la somme des carrés des écarts. La croix rouge représente notre estimation ponctuelle, $\hat{\theta}$, de la valeur des paramètres. La courbe en rouge donne la région de confiance pour les paramètres. La croix bleue donne la « vraie » valeur des paramètres obtenue avec un jeu de données bien plus conséquent ($n = 208$ section 3.4 page 14) et la courbe en bleu la région de confiance associée.

```
load(url(paste0(chmin, "matSSR.Rda")))
contour(p1seq, p2seq, matSSR, nlevels = 50, drawlabels = FALSE,
        xlab = expression(t[0]), ylab = "F*", las = 1,
```

```

main = "Région de confiance pour les paramètres")
contour(p1seq, p2seq, matSSR, levels = seuil, drawlabels = FALSE,
        add = TRUE, col = "red")
points(resnlm$estimate[1], resnlm$estimate[2], pch = 3, col = "red")
load(url(paste0(chmin, "matSSR4.Rda")))
contour(p1seq, p2seq, matSSR4, levels = seuil, drawlabels = FALSE,
        add = TRUE, col = "lightblue3")
points(resnlm3$estimate[1], resnlm3$estimate[2], pch = 3, col = "lightblue3")

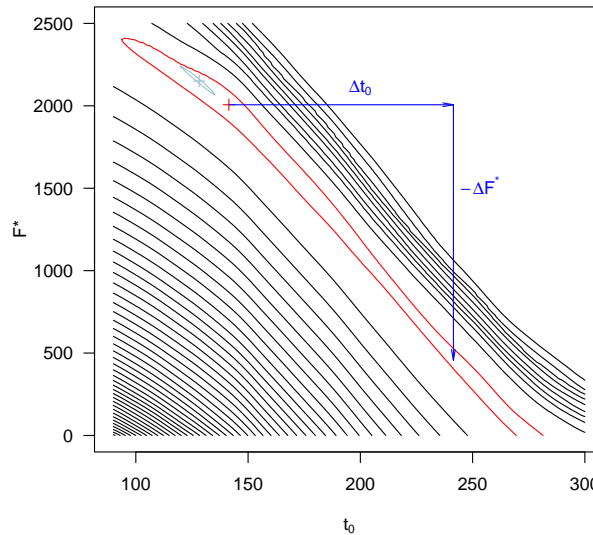
```



2.3.1 Corrélation structurelle entre les paramètres.

ON CONSTATE que la région de confiance est extrêmement étirée selon une direction parallèle à la seconde bissectrice : il y a une corrélation structurelle négative entre les paramètres du modèle. Si par la pensée on part de notre estimation $\hat{\theta}$ et que l'on se déplace horizontalement d'un incrément Δt_0 suffisant pour sortir de la région de confiance, on voit que l'on peut compenser en se déplaçant verticalement d'un incrément $-\Delta F^*$ pour revenir dans la région de confiance :

Région de confiance pour les paramètres



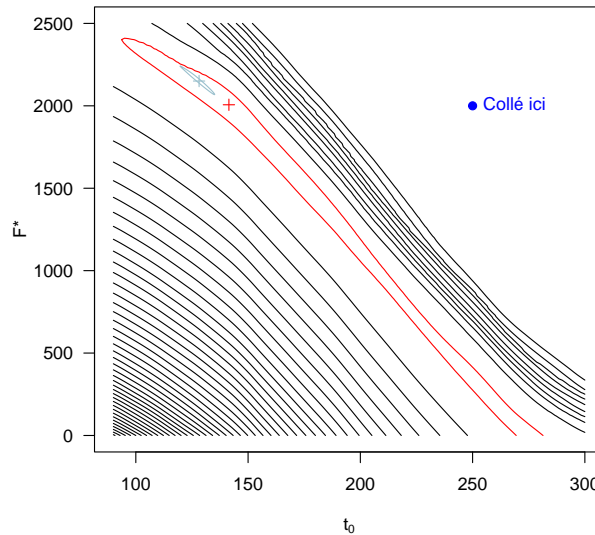
AUTREMENT DIT, si je décide de commencer à sommer les degrés plus tard, je peux compenser en abaissant le seuil F^* à atteindre pour déclencher la date des vendanges. C'est une propriété intrinsèque du modèle, c'est pourquoi on parle de corrélation *structurelle* entre les paramètres. Ce n'est pas une propriété désirable parce qu'elle complique l'estimation des paramètres avec des méthodes de type gradient et augmente l'étendue des intervalles de confiance marginaux pour les paramètres.

2.3.2 Présence de mesa

DANS le cadran supérieur droit de la représentation de la somme des carrés des écarts on constate la présence d'une mesa, c'est à dire d'un haut plateau complètement plat. Comme la pente y est nulle, les algorithmes de type gradient restent piégés sur cette hauteur :

```
resnlmMesa <- nlm(SSRR1, c(250, 2000))
resnlmMesa$estimate
[1] 250 2000
```


Région de confiance pour les paramètres

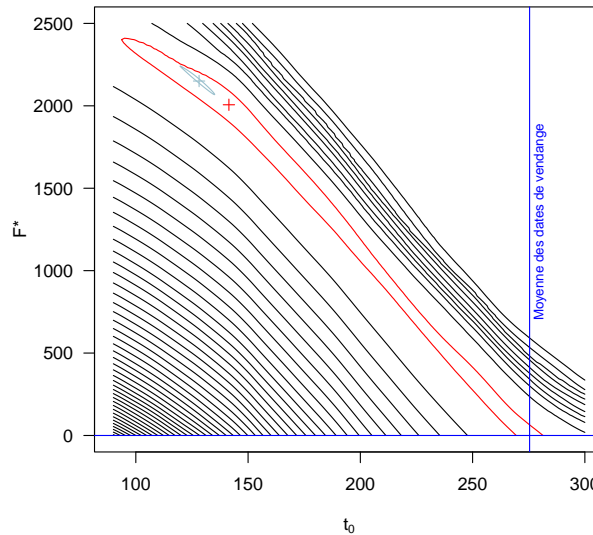


CECI illustre tout l'intérêt d'avoir une bonne estimation initiale des paramètres quand on fait de la régression non-linéaire. L'origine de cette mesa est que lorsque l'on prend des seuils F^* trop élevés, il devient alors impossible de les atteindre en sommant tous les degrés disponibles. Dans ce cas de figure la fonction `dddR1()` renvoie pour t_d la plus grande date disponible dans le jeu de données, la somme des carrés des écarts est alors insensible aux variations des paramètres.

2.3.3 Surparamétrisation du modèle

ON REMARQUE que dans la région de confiance il y a des valeurs nulles pour le seuil ($F^* = 0$). Le modèle de DE RÉAUMUR se réduit dans ce cas à $t_d = t_0$ puisque que le seuil est immédiatement atteint. La valeur de t_0 qui minimise la somme des carrés des écarts n'est rien d'autre que la moyenne des dates de vendange observées :

Région de confiance pour les paramètres



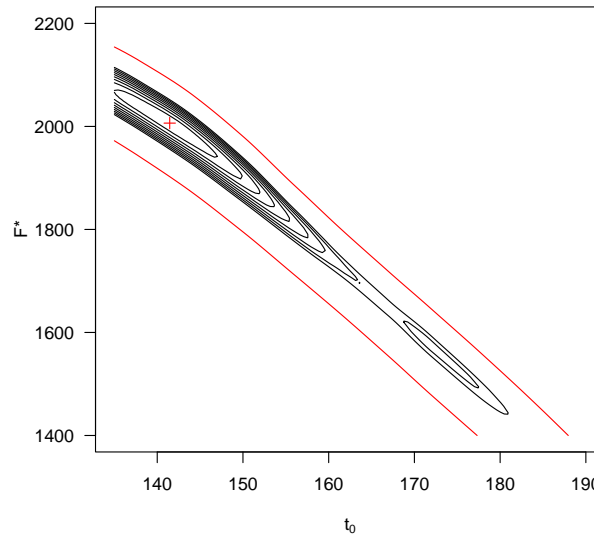
LE modèle de DE RÉAUMUR à 2 paramètres ne fait donc pas significativement mieux que le modèle à 1 paramètre qui résume les données par leur moyenne. On dit que le modèle est sur-paramétré : avec 5 points observés il n'y a pas suffisamment de données pour estimer de façon fiable les paramètres du modèle.

2.3.4 Minimum local

SI L'ON ZOOOME la surface de réponse pour la somme des carrés des écarts dans la région en bas à droite du minimum, on se rend compte qu'il existe un minimum local à environ $t_0 = 171.7$ et $F^* = 1579$:

```
npts <- 128
p1seq <- seq(135, 190, le = npts)
p2seq <- seq(1400, 2200, le = npts)
matSSR <- matrix(NA, nrow = length(p1seq), ncol = length(p2seq))
for(i in seq_len(nrow(matSSR))){
  for(j in seq_len(ncol(matSSR))){
    matSSR[i, j] <- SSRR1(c(p1seq[i], p2seq[j]))
  }
}
save(p1seq, p2seq, matSSR, resnlm, seuil, file = "minloc.Rda")
```

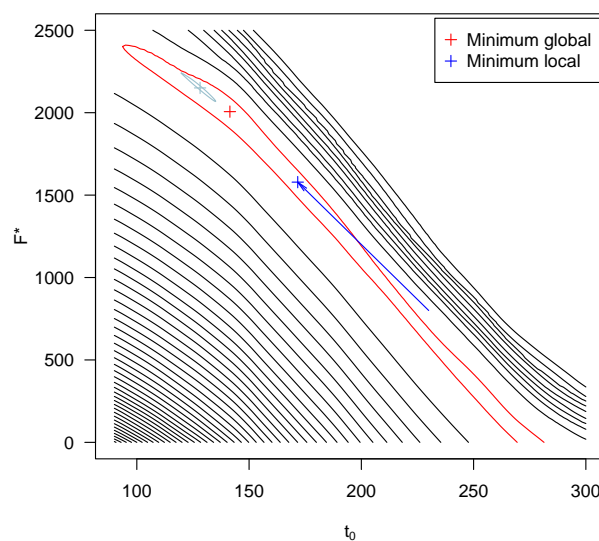
Région de confiance pour les paramètres




La conséquence directe de ceci est que si l'on part avec une mauvaise estimation initiale pour la valeur des paramètres (par exemple $t_0 = 230$ et $F^* = 800$), un algorithme utilisant une estimation locale du gradient va rester piégé dans ce minimum local :

```
badguess <- c(230, 800)
resnlmMinLoc <- nlm(SSRR1, badguess)
resnlmMinLoc$estimate
[1] 171.6747 1578.7346
```

Région de confiance pour les paramètres




C'EST une situation très pénible parce que sauf à faire une exploration systématique dans l'espace des paramètres, comme on l'a fait ici, il n'y a pas moyen de savoir si le minimum trouvé par un algorithme du type gradient est un minimum local ou bien global. Pour pallier ce problème on peut tenter d'utiliser des algorithmes de type « recuit simulé » (*simulated annealing*) avec par exemple la fonction `GenSA()` du paquet  éponyme [23] mais les temps de calcul sont alors beaucoup beaucoup *beaucoup* plus longs².

```
library("GenSA")
system.time(resGenSA <- GenSA(badguess, SSRR1, lower = c(0, 0), upper = c(280, 10^4)))
save(resGenSA, file = "resGenSA.Rda")

load(url(paste0(chmin, "resGenSA.Rda")))
resGenSA$par
[1] 141.4556 2006.4160
```

3 Données manuscrites (n = 23)

3.1 Les données

IL s'agit des données de température décryptées par Daniel ROUSSEAU [20] à partir des notes manuscrites de DE RÉAUMUR (*cf.* annexe page 30) et des dates de vendange (*cf.* annexe page 18) pour les années 1734 à 1756 incluses pour la région Île-de-France. Les importer dans  avec :

```
load(url(paste0(chmin, "Reaumet2.Rda")))
load(url(paste0(chmin, "Reauphe.Rda")))
Reauphe <- Reauphe[Reauphe$Year %in% 1734:1756, ]
```

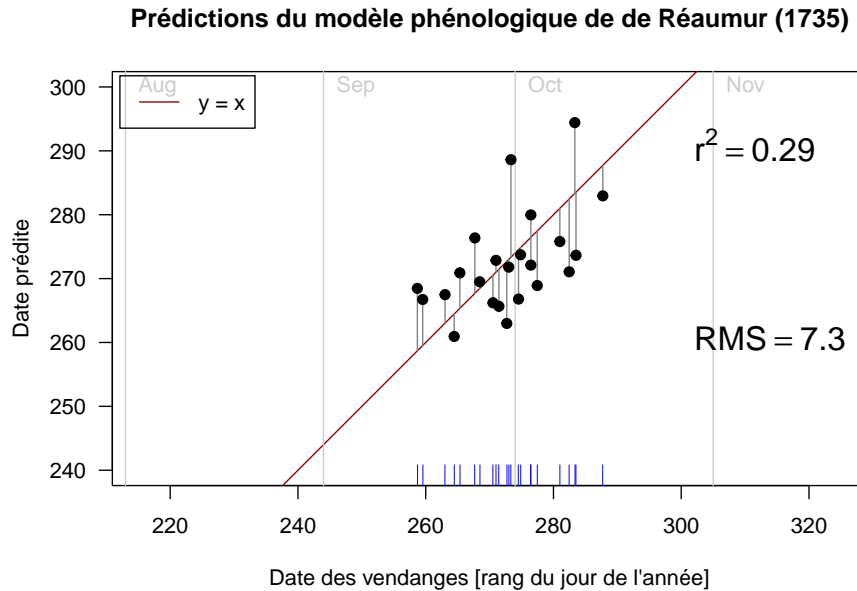
3.2 Estimation ponctuelle des paramètres

```
resnlm2 <- nlm(SSRR1, c(100, 2100))
resnlm2$estimate
[1] 139.5434 1991.1174
```

LES valeurs estimées pour les paramètres sont très proches de celles obtenues avec le jeu des données publiées. On a un $t_0 \approx 140$ soit le 20 mai au lieu du 21 mai précédemment. Les prédictions du modèle sont un peu moins bonnes :

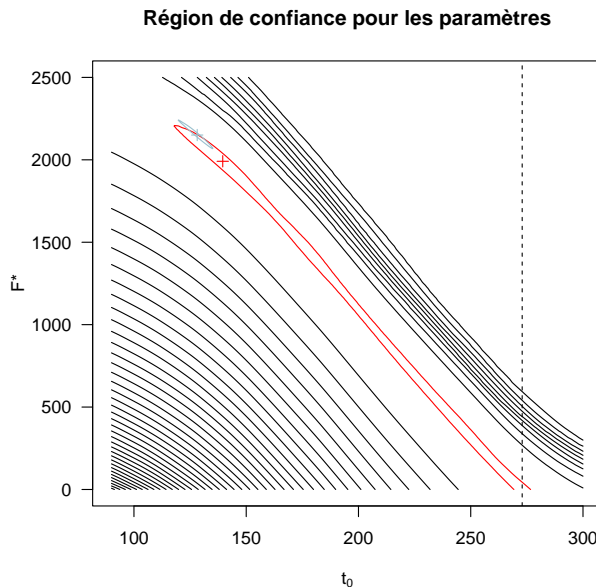
```
invisible(SSRR1(resnlm2$estimate)) # pour calculer les valeurs prédites
y <- Reauphe$théo ; x <- Reauphe$Day_of_Year
plotpred(x, y)
```

²De l'ordre d'une heure sur mon MacBook Pro.



3.3 Région de confiance

```
load(url(paste0(chmin, "matSSR2.Rda")))
contour(p1seq, p2seq, matSSR2, nlevels = 50, drawlabels = FALSE,
        xlab = expression(t[0]), ylab = "F*", las = 1,
        main = "Région de confiance pour les paramètres")
contour(p1seq, p2seq, matSSR2, levels = seuil, drawlabels = FALSE,
        add = TRUE, col = "red")
points(resnlm2$estimate[1], resnlm2$estimate[2], pch = 3, col = "red")
abline(v = mean(Reauphe$Day_of_Year), lty = 2)
load(url(paste0(chmin, "matSSR4.Rda")))
contour(p1seq, p2seq, matSSR4, levels = seuil, drawlabels = FALSE,
        add = TRUE, col = "lightblue3")
points(resnlm3$estimate[1], resnlm3$estimate[2], pch = 3, col = "lightblue3")
```



La région de confiance pour les paramètres est plus étroite qu’avec les données publiées (section 2.3 page 6) mais nous avons toujours ce « canyon vers la moyenne » qui nous dit que le modèle ne fait pas significativement mieux que la simple moyenne des valeurs observées. Il n’y a donc pas moyen d’estimer de façon fiable la valeur des paramètres du modèle de DE RÉAUMUR à partir des données de DE RÉAUMUR. Notons cependant que l’estimation ponctuelle n’est pas complètement délirante puisqu’assez proche de la « vraie » valeur des paramètres.

3.4 Artillerie lourde (n = 208)

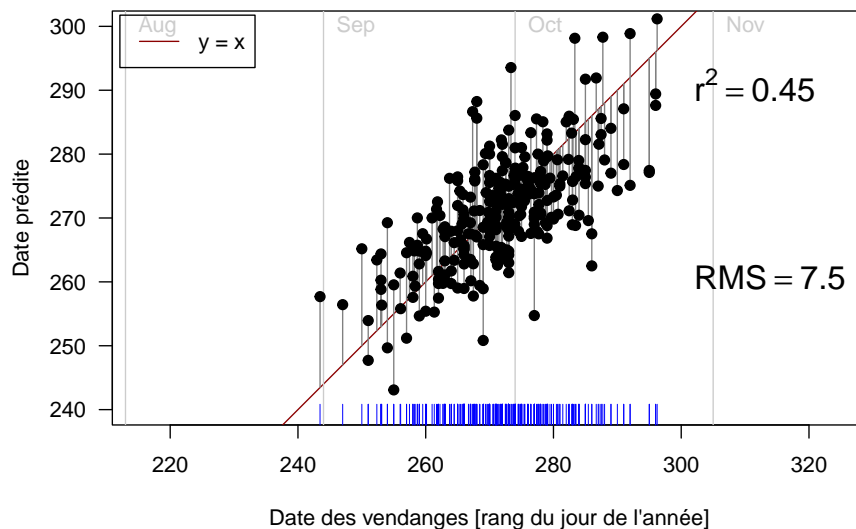
ON SORT ici un petit peu du cadre de cette fiche de TD puisque l’on n’utilise pas de façon *exclusive* les données de DE RÉAUMUR pour estimer les paramètres de son modèle. On cherche à exploiter le maximum de données disponibles pour avoir une idée de la « vraie » valeur des paramètres. En croisant (section 4.4 page 32) la longue série des températures journalières de Daniel ROUSSEAU [21] avec les données disponibles pour la date des vendanges [4] on peut exploiter 208 observations, soit un ordre de grandeur de plus. Notons que l’on pourrait craindre avoir une part de circularité ici puisque pour les séries les plus anciennes de 2009 [18] et 2013 [19] les dates des vendanges ont été utilisées pour calibrer les températures, mais ce n’est pas le cas pour la série récente de 2014 [21]. Cette dernière série n’est cependant pas corrigée de l’effet de l’urbanisation qui y est détectable à partir de 1900 environ, par conséquent les données postérieures à cette date ont été écartées. L’estimation ponctuelle des paramètres conduit au résultat suivant :

```
load(url(paste0(chmin, "Reaumet3.Rda")))
load(url(paste0(chmin, "Reauphe.Rda")))
Reauphe <- Reauphe[Reauphe$Year %in% unique(Reaumet$Year), ]
```

```
resnlm3 <- nlm(SSRR1, c(100, 2100))
save(resnlm3, file = "../web/donnees/tdr4R/resnlm3.Rda")

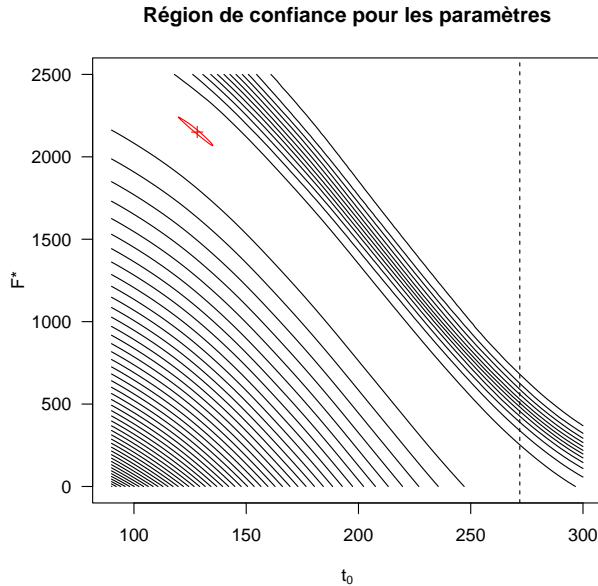
load(url(paste0(chmin, "resnlm3.Rda")))
invisible(SSRR1(resnlm3$estimate)) # pour calculer les valeurs prédites
y <- Reauphe$théo ; x <- Reauphe$Day_of_Year
plotpred(x, y, cex = 0.5)
```

Prédictions du modèle phénologique de de Réaumur (1735)



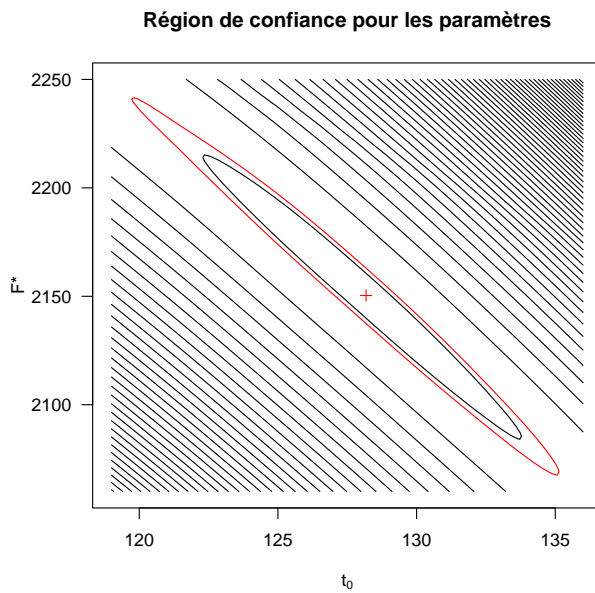
IL EST quand même surprenant qu'un modèle aussi simple à deux paramètres rende compte de près de la moitié de la variabilité résiduelle autour de la moyenne et avec une précision des prédictions de l'ordre de ± 4 jours alors que l'amplitude des dates de vendange observée ici est de 53 jours. Pour des prédictions plus précises on utilise des modèles plus sophistiqués qui distinguent trois phases (floraison, élongation cellulaire et accumulation de photoproduits) pour la prédiction de la date de véraison [3, annexe]. Grâce à la puissance de feu de ce jeu de données le « canyon vers la moyenne » est maintenant complètement comblé et on a une région de confiance bien ellipsoïdale :

```
load(url(paste0(chmin, "matSSR3.Rda")))
contour(p1seq, p2seq, matSSR3, nlevels = 50, drawlabels = FALSE,
        xlab = expression(t[0]), ylab = "F*", las = 1,
        main = "Région de confiance pour les paramètres")
load(url(paste0(chmin, "matSSR4.Rda")))
contour(p1seq, p2seq, matSSR4, levels = seuil, drawlabels = FALSE,
        add = TRUE, col = "red")
points(resnlm3$estimate[1], resnlm3$estimate[2], pch = 3, col = "red")
abline(v = mean(Reauphe$Day_of_Year), lty = 2)
```



EN zoomant sur la région de confiance on voit que l'on a une estimation ponctuelle pour t_0 de 130.6 jours, soit le 11 mai. L'intervalle de confiance marginal pour t_0 va de 121.9 (2 mai) à 139.4 (19 mai), soit une amplitude de 17.5 jours. On a donc une estimation de t_0 à plus ou moins 9 jours. L'estimation ponctuelle de F^* est de 2120° (2430°C) entre 2010° (2310°C) et 2215° (2540°C), on a donc une estimation à plus ou moins 120°C. Entre le 11 mai et la date moyenne des vendanges ($t = 272.4$ le 29 septembre) il y a 141 jours, soit 17.2°C en moyenne par jour pour que la véraison soit prédite par le modèle de DE RÉAUMUR à cette date.

```
load(url(paste0(chmin, "matSSR4.Rda")))
contour(p1seq, p2seq, matSSR4, nlevels = 50, drawlabels = FALSE,
        xlab = expression(t[0]), ylab = "F*", las = 1,
        main = "Région de confiance pour les paramètres")
contour(p1seq, p2seq, matSSR4, levels = seuil, drawlabels = FALSE,
        add = TRUE, col = "red")
points(resnlm3$estimate[1], resnlm3$estimate[2], pch = 3, col = "red")
abline(v = mean(Reauphe$Day_of_Year), lty = 2)
```

4 Annexes

4.1 Données phénologiques

4.1.1 Origine des données

LES données phénologiques auxquelles DE RÉAUMUR s'intéresse sont les dates des moissons et des vendanges, mais elles ne sont pas tabulées de façon très précise, par exemple [9, pp. 557-559] : « Cette année 1735 doit être mise au nombre des plus tardives ; la récolte des blés et celle des vins ont été faites au moins un mois plus tard que dans les années ordinaires. » Autre exemple en 1740 [14, p. 554] : « Les vignes du royaume se trouvèrent chargées d'une très grande quantité de raisins dont le mois d'août n'avança pas assez la maturité, le mois de septembre ne put l'achever. Jamais un beau mois d'octobre n'avait été plus nécessaire, les premiers jours firent espérer de bonnes vendanges, quoique tardives ; mais ce qu'il avait promis, se trouva démenti dès le 7, le matin la liqueur d'un thermomètres exposé à l'air libre, et attaché contre un arbre, n'était élevé que d'un degré au-dessus de la congélation ; aussi y eut-il de la gelée blanche, et même de la glace dans des endroits auxquels l'air avait communiqué plus aisément son degré de froid. »

C n'est donc pas directement dans les articles de DE RÉAUMUR que l'on trouvera des données phénologiques facilement exploitables. Heureusement, les séries de dates de vendanges, dont l'historien Alain GUERREAU donne une synthèse critique [16] très accessible et éclairante, sont un objet d'étude depuis au moins 1870. J'ai utilisé ici la base de données GHD [4] pour l'Île-de-France³ de 1735 à 1740 :

	A	B	C	M	N
1	Daux et al. 2012 Grape Harvest Date Data				
2					
3					
		Alsace	Auvergne	Île de France	Jura
383	1732	36,9		27,0	40,0
384	1733	43,4		19,0	42,0
385	1734	37,7		16,5	31,0
386	1735	47,9		40,5	49,0
387	1736	41,7		26,5	33,0
388	1737	34,7		20,0	33,0
389	1738	35,0		31,5	41,0
390	1739	33,7		30,0	44,0
391	1740	49,0		43,7	53,0
392	1741	45,9		22,3	37,8
393	1742	52,9		30,3	49,5

```
idf <- read.table(paste0(chmin, "GHD-IdF.csv"), sep = "\t", header = TRUE, dec = ",")
idf <- idf[!is.na(idf$IdF),]
```

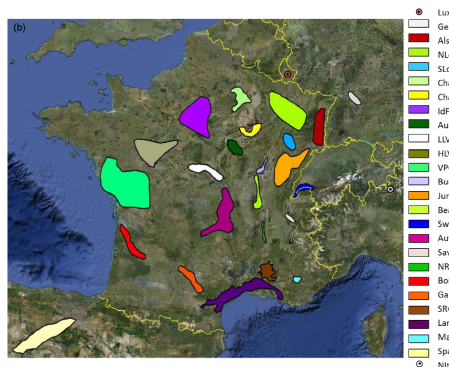
DEPUIS les travaux d'Alfred ANGOT [1] en 1883 il est d'usage [4] d'exprimer la date des vendanges en jour par un entier relatif au 1^{er} septembre⁴, valant 1 pour icelui. Les valeurs ne sont pas ici des entiers car il s'agit de séries composites⁵ de différentes localités qui pour l'Île-de-France correspondent à la région en violet clair dans la carte⁶ ci-après :

³Du temps de DE RÉAUMUR c'était le nom de l'Île Maurice, mais il s'agit ici tout simplement de la région parisienne où DE RÉAUMUR fit ses relevés.

⁴Pour homogénéiser on utilise toujours le calendrier grégorien, même si c'est un anachronisme avant 1583 en France.

⁵GHD-RCS : *Grape Harvest Date - Regional Composite Series*.

⁶Copie d'écran d'une partie de la figure 1(b) de [4].



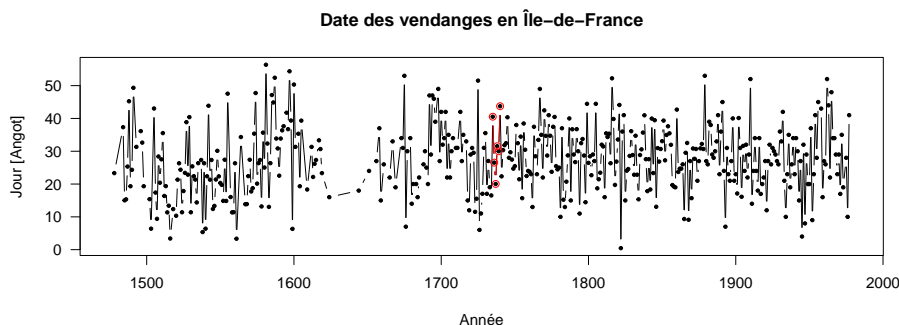
POUR la période qui nous intéresse, de 1735 à 1740 inclus, les localités concernées sont [4, matériel supplémentaire] :

- 1° Argenteuil (1735-1740);
- 2° Clamart (1735-1740);
- 2° Montmorency (1740 seulement).

4.1.2 Positionnement des données d'intérêt

LES dates observées ($n = 419$) pour les vendanges fluctuent sur une plage de l'ordre de 2 mois, les données d'intérêt (en rouge) ne présentent pas de caractère exceptionnel sur la chronique :

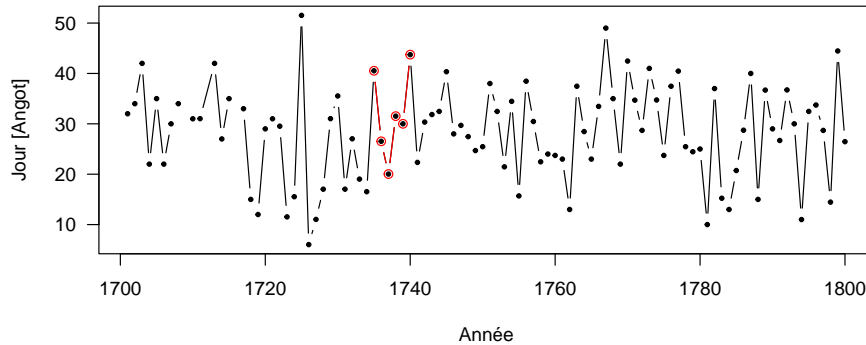
```
plot(idf, type = "b", las = 1, pch = 19, cex = 0.5,
     xlim = c(1475, 1980), ylab = "Jour [Angot]", xlab = "Année",
     main = "Date des vendanges en Île-de-France")
points(idf[idf$Year %in% 1735:1740, ], col = "red", type = "b")
```



SI l'on zoome sur le XVIII^e siècle, on peut constater, comme le rapporte DE RÉAUMUR (*cf. supra*), que les vendanges de 1735 et 1741 étaient plutôt tardives :

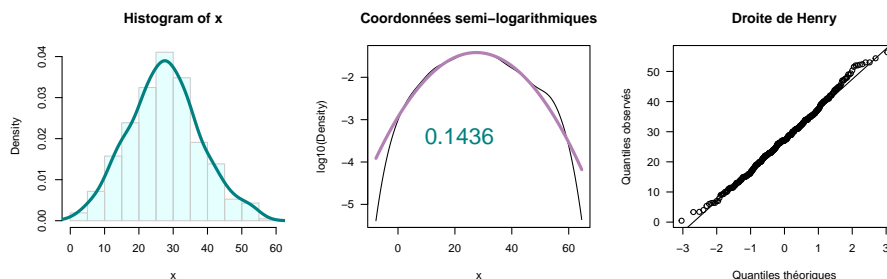
```
plot(idf[idf$Year %in% 1701:1800, ], type = "b", las = 1, pch = 19, cex = 0.5,
     xlim = c(1701, 1800), ylab = "Jour [Angot]", xlab = "Année",
     main = "Date des vendanges en Île-de-France")
points(idf[idf$Year %in% 1735:1740, ], col = "red", type = "b")
```

Date des vendanges en Île-de-France

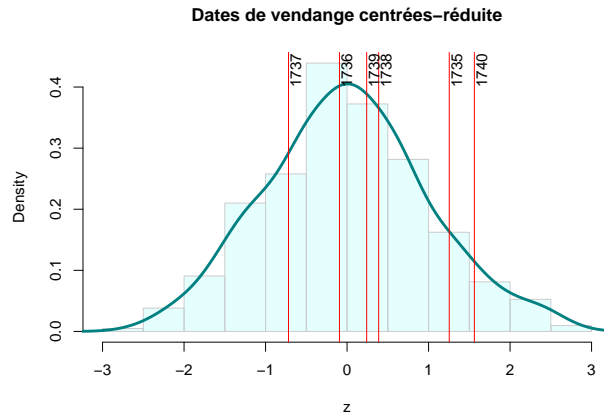


La distribution des dates de vendange est remarquablement normale, on est d'ailleurs dans l'incapacité de rejeter l'hypothèse nulle du test de SHAPIRO et WILK, alors qu'avec 419 points on dispose d'une puissance confortable. La date médiane de vendange correspond au 27 septembre. On note une très légère asymétrie à droite en raison d'un épaississement de la queue de distribution du coté des vendanges tardives à 50 jours :

```
data: x
Shapiro-Wilk normality test
W = 0.9957, p-value = 0.3077
```



Si l'on représente les dates de vendange après centrage et réduction, on constate que les années 1735 et 1740 correspondent bien à des vendanges tardives, sans être particulièrement exceptionnelles puisque l'on reste en deçà des 2σ :



4.1.3 Conversion en jour de l'année

DANS les modèles phénologiques mécanistes on utilise traditionnellement comme unité de temps le rang du jour de l'année. Il nous faut donc calculer le rang du 31 août de chaque année, qui vaut 243 ou 244 selon que l'année est bissextile ou non, et l'ajouter au jour « angotien » :

```
idf$Day_of_Year <- idf$IdF +
  as.numeric(format(as.POSIXct(paste(idf$Year, "-08-31", sep = "")), "%j"))
idf[idf$Year %in% 1735:1740, ]
```

Year	IdF	z	Day_of_Year	
382	1735	40.52857	1.25381059	283.5286
383	1736	26.52857	-0.09325033	270.5286
384	1737	20.02857	-0.71867147	263.0286
385	1738	31.52857	0.38784286	274.5286
386	1739	30.00000	0.24076580	273.0000
387	1740	43.72973	1.56182168	287.7297

4.1.4 Sauvegarde post-prétraitement

```
Reauphe <- idf[, c("Year", "Day_of_Year")]
comment(Reauphe) <- c("Time Stamp:", as.character(Sys.time()))
save(Reauphe, file = "tdr4R/Reauphe.Rda")
```

4.2 Données météorologiques publiées

4.2.1 Origine des données

LES données quotidiennes de température ont été publiées par DE RÉAUMUR pour les années 1735 à 1740, incluses, dans les mémoires de l'académie royale des sciences de PARIS [9, 10, 11, 12, 13, 14]. Une version digitalisée de ces articles au format PDF est disponible sur le site de l'académie (plus très royale) des sciences de PARIS⁷.

MISE EN GARDE. Il ne faudrait pas utiliser ce jeu de données pour une publication scientifique sérieuse, et ce pour deux raisons. Premièrement, il n'a fait l'objet que d'une saisie unique, qui plus est avec un tableur qui a la

⁷https://www.academie-sciences.fr/pdf/dossiers/Reaumur/Reaumur_publi.htm

fâcheuse habitude de proposer une complétion automatique, source d'erreur en cas d'inattention. Il faudrait *a minima* procéder à une seconde saisie indépendante pour consolider les résultats. Deuxièmement, les documents PDF ne sont toujours entièrement exploitables faute de lisibilité. Il faudrait revenir aux documents physiques pour essayer de lever les ambiguïtés qui, même si elles ne concernent très majoritairement que les fractions de degrés, mériteraient d'être corrigées.

J'AI découvert depuis la dernière mise en ligne⁸ du présent PDF l'existence de l'article de Daniel ROUSSEAU [20] qui a eu accès aux *notes manuscrites* des observations météorologiques de DE RÉAUMUR (section 4.3 page 30). Seules les 6 années de 1735 à 1740 ont été publiées et [20, p. 23] : « La comparaison avec le manuscrit a permis de vérifier que la publication était bien conforme, à de rares exceptions près, au contenu du manuscrit. » En confrontant mes valeurs avec celles issues des données manuscrites j'ai trouvé de nombreuses erreurs⁹ et il ne convient définitivement pas de les utiliser pour des études sérieuses.

Le fichier texte `ReaumurData.csv` est encodé en UTF-8 et contient des caractères diacrités dans les données de géolocalisation et dans le titre d'une colonne. Les colonnes sont séparées par des tabulations. La première ligne contient le titre des colonnes dont la signification est la suivante :

Year : l'année (**integer** de 1735 à 1740)

Month : le mois (**integer** de 1 à 12)

Day.of.Month : le jour du mois (**integer** de 1 à 31)

h.AM : l'heure d'observation du matin (**numeric** de 4.5 à 8.0)

loc.AM : la géolocalisation du matin (**character**)

d.AM : les degrés du matin (**numeric** de -10.25 à 18.5)

h.PM : l'heure d'observation de l'après-midi (**numeric** de 0.0 à 11.0)

loc.PM : la géolocalisation de l'après-midi (**character**)

d.PM : les degrés de l'après-midi (**numeric** de -9.5 à 29.0)

Notes.Réaumur : précisions de DE RÉAUMUR (**character**)

Notes.JLO : mes précisions (**character**)

Importation du fichier texte.

```
météo <- read.table(paste0(chmin, "ReaumurData.csv"), header = TRUE, sep = "\t", dec = ".")
# Calcul de la date au format POSIX
météo$Date <- with(météo, as.POSIXct(sprintf("%04d-%02d-%02d", Year, Month, Day.of.Month), tz = "UTC"))
# Sanity check:
all(diff(météo$Date) == 1) # TRUE si toutes les dates sont consécutives
[1] TRUE
```

Il y a 2192 lignes dans ce fichier. Il y a 6 années en tout, de 1735 à 1740, dont deux sont bissextiles. On peut le vérifier facilement avec la fonction `leap_year()` du paquet `lubridate` [15].

```
library(lubridate)
années <- 1735:1740
names(années) <- leap_year(années)
années
FALSE TRUE FALSE FALSE FALSE TRUE
1735 1736 1737 1738 1739 1740
```

⁸2024-01-20

⁹Détails dans <https://esb.univ-lyon1.fr/donnees/tdr4R/ReaumurData2.ods>

C'EST cohérent avec les tables de DE RÉAUMUR qui ne comportent un 29 février que pour les années 1736 et 1740 :

26 $1\frac{1}{2}$ 2	26 $8\frac{1}{2}$ $3\frac{3}{4}$
27 3 4	27 $8\frac{1}{2}$ $1\frac{1}{2}$
28 $2\frac{1}{2}$ 2	28 $2\frac{3}{4}$ $1\frac{1}{2}$
	& un peu plus bas.		29 $\frac{1}{2}$ $2\frac{1}{2}$
29 0 4			

Février 1736

Février 1740

Le nombre de lignes attendu dans le fichier est donc bien de $6 \times 365 + 2 = 2192$.

```
nrow(météo)
```

```
[1] 2192
```

4.2.2 Notation des degrés négatifs

DE RÉAUMUR rappelle à chaque fois la notation qu'il utilise pour noter les degrés négatifs, par exemple en 1740 [14, p. 539] : « Un avertissement qui a été mis à la tête de toutes les tables des volumes antérieurs, doit encore être placé avant celles de cette année ; à savoir, que lorsqu'une petite ligne se trouve au-dessus d'un chiffre, ce chiffre exprime des degrés au-dessous de la congélation, $\bar{6}$ exprime six degrés au-dessous de ce terme. »

POUR faciliter la lecture, les tables de DE RÉAUMUR comportent une ligne horizontale tous les cinq jours. Pour qu'il n'y ait pas d'ambiguïté cette ligne est supprimée si elle est suivie d'un degré négatif, comme par exemple les -0.75 observés le 16 février 1735 matin :

11 3	3 7
12 4	1 5
13 3	1 6
14 $3\frac{3}{4}$	8 3
15 1	1 1
16 $\frac{3}{4}$	1 1
17 $1\frac{3}{4}$	$2\frac{1}{2}$ 1
18 0	$2\frac{1}{2}$ $4\frac{1}{2}$
19 $1\frac{1}{4}$	1 $3\frac{3}{4}$
20 $4\frac{1}{2}$	à midi 8
21 $2\frac{1}{4}$ 0

J'en'ai trouvé qu'un seul cas d'erreur cléricale manifeste pour le 2 août 1739 au matin où les -13.5 rapportés correspondent en fait à +13.5 puisque la plus faible valeur observée par DE RÉAUMUR est de -10.25 le 10 janvier 1740 au matin :

[1739.] A O U S T.					
Idi.	J.	Degrés du Matin.		Degrés d'Après-midi.	
rés.		Heure.	Degrés.	Heure.	Degrés.
19	1	à 5 $\frac{1}{2}$	à 10 $\frac{1}{2}$	à 3	à 17
17	2	5 $\frac{1}{2}$	13 $\frac{1}{2}$	3	18 $\frac{1}{2}$
18 $\frac{1}{2}$	3	11	20
18	4	13	22 $\frac{1}{2}$
19 $\frac{1}{2}$	5	15 $\frac{1}{2}$	à midi $\frac{1}{2}$	22 $\frac{1}{2}$

ON comprend bien l'idée de la notation $\bar{6}$, ce sont 6 degrés sous la ligne du zéro du thermomètre. Attention au faux-ami ici : quand DE RÉAUMUR utilise l'expression « sous la ligne », comme dans le titre « Observations du thermomètre faites à Paris pendant l'année 1735 comparées avec celles qui ont été faites sous la ligne, à l'Île de France, à Alger et en quelques-unes de nos Îles de l'Amérique » de [9], cela signifie en fait la région équatoriale (*i.e.* sous la ligne de l'équateur).

IL semblerait qu'un avatar de cette notation ait encore été en usage récemment. Dans la figure ci-dessous extraite d'un article de 1975 [5], la notation $\bar{2} \cdot 75$ sur l'axe vertical de gauche doit se comprendre comme $-2 + 0.75 = -1.25$:

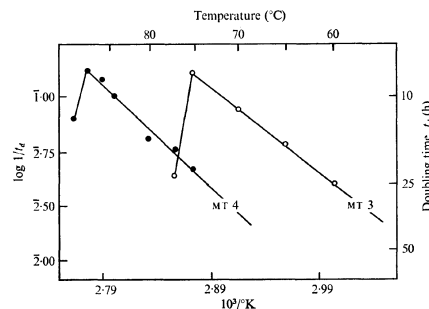


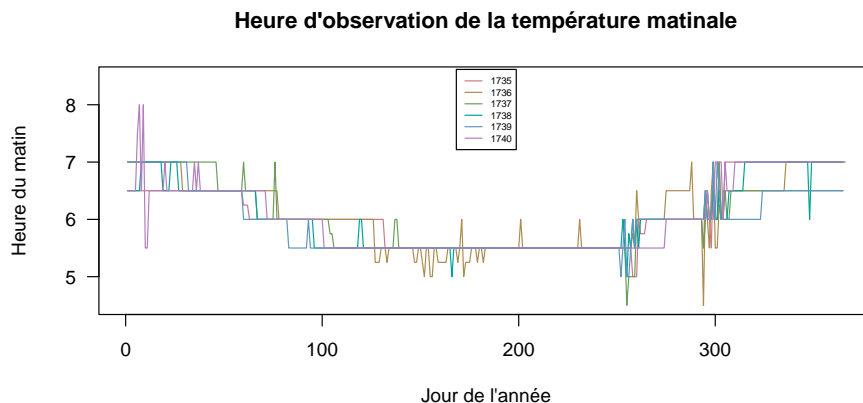
Fig. 1. Arrhenius plots of growth of MT3 and MT4 at pH 3.0.

4.2.3 Heures d'observation

LES heures d'observation sont données avec une précision de $\frac{1}{2}$ heure, parfois avec une précision de $\frac{1}{4}$ d'heure. Il y a en général une observation le matin et une le soir. Le 4 juin 1736 il y a deux observations le matin, et il y a 18 jours pour lesquels il y a deux observations l'après-midi¹⁰. En cas d'observation double avec deux températures différentes je n'ai conservé dans le fichier texte `ReaumurData.csv` qu'une valeur correspondant à l'heure d'observation la plus courante pour la période de l'année.
Heures du matin.

¹⁰ *viz.* 1735-05-21, 1736-01-22, 1736-05-06, 1736-05-30, 1736-07-12, 1736-08-31, 1736-09-03, 1736-09-28, 1736-09-30, 1736-10-01, 1736-11-08, 1736-12-02, 1736-12-09, 1737-02-14, 1737-02-25, 1737-04-17, 1738-04-03 et 1739-06-18.


```
# remplissage des trous
for(i in seq_len(nrow(météo))){
  if(is.na(météo[i, "h.AM"])) météo[i, "h.AM"] <- météo[i - 1, "h.AM"]
}
météo$doy <- as.integer(format(météo$Date, "%j"))
Years <- unique(météo$Year)
ny <- length(Years)
mycol <- hcl.colors(ny, "Dark 2")
plot(0, xlim = c(1, 366), ylim = c(4.5, 8.5), cex = 0, las = 1,
     xlab = "Jour de l'année", ylab = "Heure du matin",
     main = "Heure d'observation de la température matinale")
for(i in seq_len(ny)){
  with(météo[météo$Year == Years[i], ], {
    points(doy, h.AM, type = "l", col = mycol[i])
  })
}
legend("top", inset = 0.01, legend = Years, col = mycol, lwd = 1, cex = 0.5)
```



LES heures d'observation matinales ne posent pas de problème particulier d'interprétation. Elles partent en général de $5\frac{1}{2}$ h l'été pour augmenter jusqu'à 7 h l'hiver. Les périodes plus instables autour du 275^e et 300^e jour de l'année correspondent aux voyages de DE RÉAUMUR entre Paris et Réaumur.

LES heures d'observation de l'après-midi sont données de façon relative à midi, j'ai donc codé dans le fichier texte 0.0 pour midi. Il y a un problème d'interprétation pour les valeurs 10 et 11 dans les tables de DE RÉAUMUR. Il peut s'agir :

- 1° d'une valeur proche de midi, par exemple 11 h du matin ;
- 2° d'une valeur éloignée de midi, par exemple 23 h.

JE penche pour la première interprétation pour deux raisons. Premièrement, le 8 novembre 1736 DE RÉAUMUR donne deux valeurs pour l'après-midi à 11 h et 2 h, avec 11 h *avant* 2 h :

6 6	à Malhoue... 12
7 10 $14\frac{1}{2}$
8 $9\frac{3}{4}$	{ 11 12 2 10
9 7	
10 $10\frac{1}{2}$	2 $11\frac{1}{2}$ 13

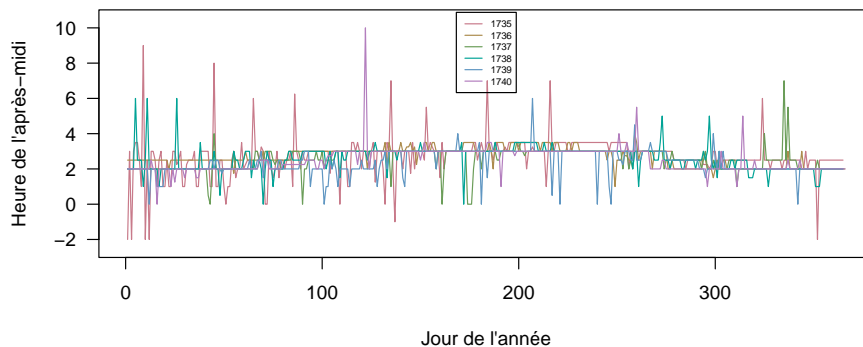
Deuxièmement, le 1^{er} mai 1740, pour la valeur 10 de l'après-midi, DE RÉAUMUR précise qu'il s'agit de 10 h *du soir* :

M A I. [1740.]					
J.	Degrés du matin.		Degr. d'après-midi.		J. D
	Heures.	Degrés.	Heures.	Degrés.	Heures.
1	à 5 $\frac{1}{2}$. . . à 5		10 du soir . . . à 5 $\frac{3}{4}$		1 à 5
2 5		3 5		2
3	à Charenton 2		2 6 $\frac{1}{2}$		3
4 2 $\frac{1}{2}$		3 6 $\frac{1}{2}$		4
5 2	 8		5

J'AI laissé dans le fichier texte les valeurs données par DE RÉAUMUR, je pense qu'il est préférable de recoder 11 h en -1 h et 10 h en -2 h, sauf pour le 1^{er} mai 1740.

```
# recodage 10 h et 11 h
météo[!is.na(météo$h.PM) & météo$h.PM == 10 & météo$Notes.Réaumur != "10 du soir", "h.PM"] <- -2.0
météo[!is.na(météo$h.PM) & météo$h.PM == 11, "h.PM"] <- -1.0
# remplissage des trous
for(i in seq_len(nrow(météo))){
  if(is.na(météo[i, "h.PM"])) météo[i, "h.PM"] <- météo[i - 1, "h.PM"]
}
plot(0, xlim = c(1, 366), ylim = c(-2.5, 10.5), cex = 0, las = 1,
     xlab = "Jour de l'année", ylab = "Heure de l'après-midi",
     main = "Heure d'observation de la température de l'après-midi")
for(i in seq_len(ny)){
  with(météo[météo$Year == Years[i], ], {
    points(doy, h.PM, type = "l", col = mycol[i])
  })
}
legend("top", inset = 0.01, legend = Years, col = mycol, lwd = 1, cex = 0.5)
```

Heure d'observation de la température de l'après-midi



COMME on peut le constater, les heures d'observation de l'après-midi sont beaucoup plus fluctuantes que celles de la matinée, elles sont en général autour de 3 h l'été et de 2 h l'hiver.

4.2.4 Géolocalisation

LES observations de DE RÉAUMUR se font principalement à Paris [7, p. 430] :
« Nous allons à présent rapporter les observations que nous avons faites à

Paris, rue St Thomas du Louvre, à l'hôtel d'Uzès. » Voir [20, p. 23, emplacement des mesures]. Il précise en 1735 [7, pp. 546-547] à propos de sa table qu'« [e]lle n'est cependant pas telle que je voudrais qu'elle fût ; il serait à souhaiter que toutes les observation fussent faites précisément dans le même lieu, et il m'est ordinaire de passer les mois de septembre et d'octobre en Poitou : je ne suis même pas assez sédentaire à Paris pendant le reste de l'année, je me tiens souvent à Charenton ; mais les différences qui ont été trouvées entre les observations que j'ai fait faire à Paris pendant que j'en étais absent, et celles que je faisais alors à Charenton, ces différences, dis-je, ne m'ont pas paru assez considérables pour que l'on ne puisse pas substituer les observations faites à Charenton, à celles qu'on a manqué de faire à Paris. »

QUAND DE RÉAUMUR se rend sur ses terres à Réaumur, il le fait début septembre (entre le 6 et le 9 du mois) en suivant toujours le même itinéraire donné dans le tableau ci-après. La distance totale est de l'ordre de 400 km, soit en moyenne 40 km par jour. À titre de de comparaison, ce même parcours se fait de nos jours en voiture en moins de 5 heures. On comprend mieux ici pourquoi DE RÉAUMUR a publié [6] en 1721 un mémoire intitulé : « Moyen de mettre les carrosses et les brelines en état de passer par des chemins plus étroits que des chemins ordinaires, et de se tirer plus aisément des ornières profondes. »

De	→	Via	→	À
Paris	→	Châtres (Arpajon 91290)	→	Étampes (91500)
Étampes (91500)	→	Toury (45210)	→	Artenay (45410)
Artenay (45410)	→	Orléans (45000)	→	Cléry (45370)
Cléry (45370)	→	Saint-Dié (Saint-Dyé 41500)	→	Blois (41000)
Blois (41000)	→	→	→	Amboise (37400)
Amboise (37400)	→	Tours (37000)	→	Langès (Langeais 37130)
Langès (Langeais 37130)	→	La Chapelle-Blanche (La Chapelle-sur-Loire 37140)	→	Saumur (49400)
Saumur (49400)	→	Montreuil (49260)	→	Thouars (79100)
Thouars (79100)	→	→	→	Bressuire (79300)
Bressuire (79300)	→	→	→	Réaumur (85700)

QUAND il redescend¹¹ à Paris, le mois d'octobre est bien avancé, avec un départ le 20 ou le 21 du mois selon les années. Il suit exactement le même parcours qu'à l'aller sauf qu'il fait étape à Toury au lieu d'Artenay.

4.2.5 Les degrés de température

LES valeurs données par DE RÉAUMUR sont d'une précision toujours au moins égale au degré, sauf le 17 juillet 1736 après-midi où il écrit « 19 ou 20 »¹² :

16 10 13
17 11 $\frac{2}{3}$ 19 ou 20
18 11	3 19
19	6 11	3 $\frac{1}{2}$ 20 $\frac{1}{2}$
20	5 $\frac{1}{2}$ 14 20

DE RÉAUMUR précise ses valeurs avec des fractions de degrés avec le plus souvent une précision au $\frac{1}{4}$ de degré, on trouve ainsi 2034 degrés entiers,

¹¹Monter ou descendre est une question d'altitude et non de latitude, or Réaumur est à 144 m [98-187] et Paris à 78 m [28-131].

¹²J'ai renseigné le fichier texte avec une valeur de 19.5 dans ce cas.

506 de $\frac{1}{4}$ de degré, 1356 de $\frac{1}{2}$ de degré et 333 de $\frac{3}{4}$ de degré, ce qui ensemble couvre 96.9 % des valeurs publiées. On trouve occasionnellement des fractions plus exotiques : $\frac{1}{8}$ 10 fois, $\frac{1}{6}$ 4 fois, $\frac{1}{5}$ 5 fois, $\frac{1}{3}$ 58 fois, $\frac{2}{5}$ 7 fois, $\frac{3}{5}$ 3 fois, $\frac{2}{3}$ 43 fois, $\frac{4}{5}$ 2 fois et $\frac{7}{8}$ 2 fois. Enfin, pour être complet, fin février 1936 il précise « un peu plus » ou « un peu moins » par rapport à la valeur donnée :

23 $1\frac{2}{3}$ $\frac{3}{4}$
24 $1\frac{1}{2}$ & un peu plus.	à $1\frac{1}{4}$ 2
25 $1\frac{1}{2}$ & un peu moins.	$2\frac{1}{2}$ $\frac{3}{2}$
26 $1\frac{1}{2}$ 2
27 3 4
28 $2\frac{1}{2}$ & un peu plus bas. 2

DE RÉAUMUR donne un résumé synthétique de sa table en donnant pour chaque mois les jours où les températures minimales et maximales ont été observées soit le matin, soit l'après-midi :

*RESULTATS DES TABLES PRÉCÉDENTES,
qui donnent les plus grands froids & les plus grands chauds
de chaque mois de 1735, soit du matin, soit de l'après-midi.*

Plus grand froid du matin.	Plus grand froid de l'après-midi.	Plus grand chaud du matin.	Plus grand chaud de l'après-midi.
JANVIER 1735.			
14° } à 7^h à $0^{\circ}\frac{1}{2}$ 15° }	13° } à h à $2^{\circ}\frac{1}{2}$ 14° }	30° à 7^h à $7^{\circ}\frac{1}{2}$	30° à 2^h à $8^{\circ}\frac{1}{2}$
FÉVRIER.			
5° à 7^h à $3^{\circ}\frac{2}{3}$	4° à 3^h à $0^{\circ}\frac{1}{2}$	21° à 7^h à $6^{\circ}\frac{1}{2}$	23° à 2^h à $8^{\circ}\frac{1}{2}$
MARS.			
3° } 4° } à 6^h à 0° 6° } 22° } 24° }	4° } à 2^h à $4^{\circ}\frac{1}{2}$ 5° }	7° à 6^h à $7^{\circ}\frac{1}{2}$	17° à 2^h à $18^{\circ}\frac{1}{2}$

C'EST d'ailleurs sous cette forme condensée que DE RÉAUMUR avait commencé à publier ses observations pour les années 1732 à 1734 [7, 8] avant qu'il ne décide en 1735 [9] : « Lorsque je lus la première de ces tables à l'académie, quelques uns de nos messieurs, et M. DE MAIRAN entre autres, pensèrent qu'il n'en fallait pas en rester là ; qu'il convenait de donner la suite complète des observations du thermomètre pour tous les jours de chaque année. La crainte de grossir nos volumes par des tables, qui semblent n'offrir rien d'agréable au lecteur, m'empêcha d'être de leur avis ; mais j'y suis revenu quand j'ai fait plus attention aux utilités qu'on pourra retirer de ces sortes de tables, surtout depuis que j'ai vu le nombre des observateurs du thermomètre se multipliait, et que nous avons lieu d'espérer d'avoir des observations faites dans toutes les parties du monde et dans leurs différents climats. »

IL N'Y A que 21 valeurs manquantes sur les 4384 attendues, soit un peu moins de 0.5 %. Pour avoir des séries complètes on va interpoler ces valeurs manquantes en prenant la moyenne du jour précédent et suivant :

```
for(i in seq_len(nrow(météo))){
  if(is.na(météo[i, "d.AM"])){
    météo[i, "d.AM"] <- 0.5*(météo[i - 1, "d.AM"] + météo[i + 1, "d.AM"])
  }
  if(is.na(météo[i, "d.PM"])){
    météo[i, "d.PM"] <- 0.5*(météo[i - 1, "d.PM"] + météo[i + 1, "d.PM"])
  }
}
```

C'E n'est pas tout à fait suffisant pour avoir des séries complètes parce que les 2 et 3 janvier 1737 après-midi il y a deux valeurs manquantes successives :

J A N V I E R. [1737.]					
J.	Degrés du Matin.		Degrés d'Après-midi.		J. L
	<i>Heure.</i>	<i>Degrés.</i>	<i>Heure.</i>	<i>Degrés.</i>	<i>Heure</i>
1	à 7.....à	4 $\frac{1}{2}$	à 2.....à	5 $\frac{1}{2}$	1 à 7
2	7.....	3			2 7
3	2 $\frac{1}{2}$			3 ..
4	1 $\frac{1}{2}$	2.....	3 $\frac{1}{2}$	4 ..
5	$\frac{1}{2}$	3 $\frac{1}{2}$	5 ..

J'ai forcé ces deux valeurs à +4.0 :

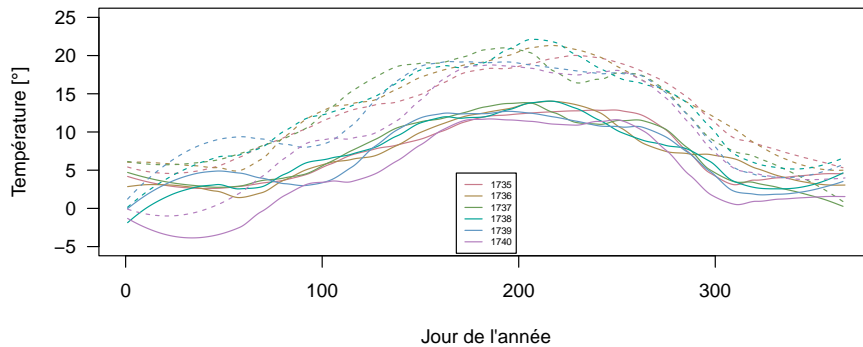
```
météo[météo$Date == as.Date("1737-01-02"), "d.PM"] <- +4.0
météo[météo$Date == as.Date("1737-01-03"), "d.PM"] <- +4.0
```

DE RÉAMUR explique de façon détaillée comment il calcule les températures moyennes journalières [9, p. 558] : « Comme les astronomes prennent un temps moyen pour mesurer la durée du temps vrai, j'ai pris pour chaque jour un degré de chaleur moyenne et cela en ajoutant les degrés du thermomètre qui ont exprimé la plus petite chaleur du matin aux degrés qui ont exprimé la plus grande chaleur de l'après-midi ; la moitié de cette somme m'a paru pouvoir être prise pour le degré de chaleur moyenne. Par exemple, le 3 avril 1735 à 6 heures du matin, la liqueur du thermomètre était à 4 degrés, et le même jour, à 3 heures de l'après-midi, elle était à 8 degrés ; ces degrés ajoutés ensemble donnent 12 degrés ; j'en prends la moitié, 6 degrés, pour l'expression du degré du chaud moyen du 3 avril. »

```
météo$Tmean <- with(météo, (d.AM + d.PM)/2.0)
météo[météo$Date == as.Date("1735-04-03"), c("d.AM", "d.PM", "Tmean")]
93  d.AM d.PM Tmean
    4     8      6

plot(0, xlim = c(1, 366), ylim = c(-5, 25), cex = 0, las = 1,
     xlab = "Jour de l'année", ylab = "Température [°]",
     main = "Variation annuelle des température du matin et de l'après-midi")
span <- 0.3
for(i in seq_len(ny)){
  with(météo[météo$Year == Years[i], ], {
    AM.lo <- loess(d.AM~doy, span = span)
    points(doy, predict(AM.lo, data.frame(doy = doy)), type = "l", col = mycol[i])
    PM.lo <- loess(d.PM~doy, span = span)
    points(doy, predict(PM.lo, data.frame(doy = doy)), type = "l", col = mycol[i],
           lty = 2)
  })
}
legend("bottom", inset = 0.01, legend = Years, col = mycol, lwd = 1, cex = 0.5)
```

Variation annuelle des température du matin et de l'après-midi




ON voit que le début de l'année 1740 a été particulièrement rigoureux, c'est d'ailleurs là, le 10 janvier, que DE RÉAUMUR observe la température la plus basse à -10.25 (soit -11.8 °C). Il écrit [14, 552] : « [...] l'année 1740 peut être mise au nombre de celles où la mortalité a été la plus grande au printemps dans le royaume. Dans la plupart de ses provinces, les campagnes ont perdu un nombre prodigieux d'habitants ; je connais des villages de Poitou à qui la moitié des leurs a été enlevée. »

4.2.6 Sauvegarde post-prétraitement

```
météo$jour <- format(météo$Date, "%A")
Reaumet <- météo
comment(Reaumet) <- c("Time Stamp:", as.character(Sys.time()))
save(Reaumet, file = "tdr4R/Reaumet.Rda")
```

4.3 Données météorologiques manuscrites

4.3.1 Origine des données

IL S'AGIT de la série établie par Daniel ROUSSEAU [20] avant conversion en degrés CELCIUS pour faciliter les comparaisons avec les données publiées. Ces données sont issues du déchiffrement¹³ des notes manuscrites de DE RÉAUMUR de 1732 à 1757. Les données sont disponibles dans la première annexe¹⁴ du site de l'article sous la forme d'un document PDF. La version Excel correspondante m'a été aimablement communiquée par l'auteur¹⁵. Je n'ai corrigé qu'une seule erreur cléricale pour la colonne H4 le 1734-01-19 (10;5 au lieu de 10,5). J'ai enregistré ce fichier sous RousseauD2019.ods pour l'importer dans  avec le code suivant.

```
library(readODS) ; library(lubridate)
fname <- "RousseauD2019.ods"
download.file(paste0(chmin, fname), fname)
dr <- read_ods(fname, as_tibble = FALSE, na = c("", NA),
```

¹³ Consulter les figures 3 et 5 de [20] pour avoir une idée de l'ampleur de la tâche.

¹⁴ https://lameteorologie.fr/issues/2019/105/meteo_2019_105_21

¹⁵ Communication personnelle du 2025-10-06 19:00 CEST, il s'agit du fichier nommé Réaumur Annexe 1 .xls créé le 2013-01-13 13:25:52 et modifié la dernière fois le 2023-12-22 10:29:26

```
col_type = "iiinnnnnnnncccnlnl")
dr[[19]] <- NULL # suppression dernière colonne vide
# Calcul de la date au format POSIX
dr$Date <- with(dr, as.POSIXct(sprintf("%04d-%02d-%02d", An, M, J), tz = "UTC"))
# Ajout de Year doy et Tmean
dr$Year <- year(dr$Date)
dr$doy <- yday(dr$Date)
dr$Tmean <- dr$Tmoy
```

LES informations codées par des attributs typographiques (*e.g.* caractères rouges pour les valeurs interpolées) sont perdues dans le fichier *.Rda mais toujours consultables dans le fichier *.ods.

4.3.2 Données manquantes

IL S'AGIT des valeurs manquantes après interpolation dans [20]. Les valeurs interpolées (en rouge dans le fichier *.ods) ne sont donc pas prises en compte ici.

```
sum(is.na(dr$Tmean))
[1] 63
rle(ifelse(is.na(dr$Tmean), T, F))
Run Length Encoding
lengths: int [1:8] 2 248 56 308 1 7 4 8779
values : logi [1:8] TRUE FALSE TRUE FALSE TRUE FALSE ...
```

IL Y A donc en tout 63 valeurs manquantes, voyons si cela est gênant pour l'estimation des paramètres du modèle de DE RÉAUMUR. Elles se répartissent en quatre plages :

- 1° 2 valeurs manquantes en tout début de série les 1 et 2 janvier 1732. Comme le t_0 du modèle de DE RÉAUMUR pour la date des vendanges est au second trimestre de l'année, ce n'est pas gênant.
- 2° 56 valeurs manquantes du 7 septembre au 1^{er} novembre 1732 [21, p. 41] : « [...] à cause d'un voyage de RÉAUMUR dans son domaine en VENDÉE pendant les vacances de l'Académie royale des sciences. » En 1732 les vendanges à PARIS ont eu lieu le 271^e jour de l'année, soit le 27 septembre. On ne pourra donc pas utiliser les données de 1732.
- 3° 1 valeur manquante le 6 septembre 1733.
- 4° 4 valeurs manquantes du 14 au 17 septembre 1733. En 1733 les vendanges à PARIS ont eu lieu le 262^e jour de l'année, soit le 19 septembre. C'est un petit peu long pour interpoler, on n'utilisera pas les données de 1733.

DE PLUS, l'année 1757 est incomplète et se termine le 30 septembre. En 1757 les vendanges à PARIS ont eu lieu le 273^e jour de l'année, soit le 30 septembre également. Cela risque de poser des problèmes lors de l'estimation des paramètres, on préférera censurer cette année également. On exploitera donc les données de 1734 à 1756, incluses, soit 23 années de suivi sans valeurs manquantes après interpolation.

```
dr <- dr[dr$Year %in% 1734:1756, ]
```

4.3.3 Le nombre d'observations

LES données publiées ne comportent généralement, à 19 exceptions près, que deux valeurs de température par jour alors que les données manuscrites en comptent jusqu'à quatre. On rajoute une colonne `nobs` pour le nombre d'observations journalière et on en regarde la distribution :

```
dr$nobs <- apply(dr[, paste0("T", 1:4)], 1, \x) sum(!is.na(x))
table(dr$nobs)
      2      3      4
2975 5332   94
```


CE SONT les trois observations par jour qui dominent largement. Les notes manuscrites sont donc bien plus riches que les données publiées.

4.3.4 Sauvegarde post-prétraitement

```
Reaumet <- dr
comment(Reaumet) <- c("Time Stamp:", as.character(Sys.time()))
save(Reaumet, file = "tdr4R/Reaumet2.Rda")
```

4.4 Données météorologiques de Rousseau 2024

4.4.1 Origine des données

CETTE série *extraordinaire* de 366 années consécutives de températures journalières à PARIS est décrite dans l'article de Daniel ROUSSEAU [21]. Elles sont directement disponibles en annexe de cet article sous la forme d'un fichier de type tableur sur le site de la revue¹⁶. J'ai enregistré ce fichier sous `RousseauD2024.ods` pour importer directement sous  [17] les données des 6 colonnes du deuxième feuillet dans la table `dr24`. J'ai transtypé en entier les colonnes des années, mois et jours pour être certain de leur classe.

```
library(readODS)
dr24 <- read_ods("../web/donnees/tdr4R/RousseauD2024.ods", sheet = 2)
dr24 <- dr24[, 1:6] # colonnes utiles
dr24$Année <- as.integer(dr24$Année)
dr24$Mois <- as.integer(dr24$Mois)
dr24$Jour <- as.integer(dr24$Jour)
```

4.4.2 Données manquantes

LES données manquantes sont décrites dans le dernier feuillet du fichier du tableur `meteo_2024_127_33_supp.xlsx` et seraient au nombre de 2904 en tout alors que je n'en compte que 2887 dans la table `dr24`. On ne va pas chipoter ici et prendre un « pied de pilote » conservatif en censurant les 21 années pour lesquelles il existe au moins une valeur manquante. On supprime également l'année 1658 qui est incomplète.

¹⁶https://lameteorologie.fr/issues/2024/127/meteo_2024_127_33. Le fichier `meteo_2024_127_33_supp.xlsx` a été créé le 2024-02-24 12:56:49 et modifié la dernière fois le 2025-01-28 07:20:53. Le premier feuillet du fichier précise qu'il s'agit de la « Version 1 du 28/01/2025 ».


```
# Suppression de 21 années avec au moins un NA
nay <- with(dr24, tapply(`T°C`, Année, \(x) any(is.na(x))))
censored <- as.integer(names(nay[nay]))
dr24 <- subset(dr24, !(Année %in% censored))
# Suppression de 1 année incomplète
ly <- with(dr24, tapply(`T°C`, Année, \(x) !(length(x) %in% 365:366)))
censored <- as.integer(names(ly[ly]))
dr24 <- subset(dr24, !(Année %in% censored))
```

IL nous reste à ce stade 344 années complètes sans valeurs manquantes. On ne va conserver que les années pour lesquelles la date des vendanges est documentée. On écarte également les données postérieures à 1900 qui sont biaisées par l'effet de l'îlot de chaleur urbain sur le site de MONTSOURIS [21, p. 29].

```
load(url(paste0(chmin, "Reauphe.Rda")))
yutil <- unique(Reauphe$Year)
dr24 <- subset(dr24, Année %in% yutil)
dr24 <- subset(dr24, Année < 1900)
```

IL nous reste à ce stade 208 années complètes sans valeurs manquantes pour lesquelles la date des vendanges est disponible.

4.4.3 Sauvegarde post-prétraitement

LES températures moyennes journalières sont exprimées en degrés CELCIUS, on les convertis dans les degrés utilisés par DE RÉAUMUR. On renomme la table en `Reaumet`, on change le nom des colonnes pour être homogène avec celles utilisées précédemment et on la sauve au format XDR [22] dans le fichier `Reaumet3.Rda`.

```
# On garde les colonnes utiles
dr24 <- dr24[, c("Année", "Jour", "T°C")]
# Changement du nom des colonnes
Reaumet <- dr24
colnames(Reaumet) <- c("Year", "doy", "Tmean")
# Conversion en degrés "Réaumur"
facC <- (5/9)*2.066
Reaumet$Tmean <- Reaumet$Tmean/facC
# Sauvegarde XDR
comment(Reaumet) <- c("Time Stamp:", as.character(Sys.time()))
save(Reaumet, file = "../web/donnees/tdr4R/Reaumet3.Rda")
```

References

- [1] A. Angot. Étude sur les vendanges en France. *Annales du Bureau central météorologique de France*, 1:B29–B120, 1883.
- [2] E.M.L. Beale. Confidence regions in non-linear estimation. *Journal of the Royal Statistical Society*, 22B:41–88, 1960.
- [3] I. Chuine, P. Yiou, N. Viovy, B. Seguin, V. Daux, and E. Le Roy Ladurie. Grape ripening as a past climate indicator. *Nature*, 432(7015):289–290, 2004.
- [4] V. Daux, I. Garcia de Cortazar-Atauri, P. Yiou, I. Chuine, E. Garnier, E. Le Roy Ladurie, O. Mestre, and J. Tardaguila. An open-database of grape harvest dates for climate research: data description and quality assessment. *Climate of the Past*, 8:1403–1418, 2012.

- [5] M. De Rosa, A. Gambacorta, and J.D. Bu'lock. Exteremly thermophilic acidophilic bacteria convergent with *Sulfolobus acidocaldarius*. *Journal of General Microbiology*, 86:156–164, 1975.
- [6] R.-A.F. de Réaumur. Moyen de mettre les carrosses et les brelines en état de passer par des chemins plus étroits que des chemins ordinaires, et de se tirer plus aisément des ornières profondes. *Mémoires de l'académie royale des sciences de Paris*, 1723:417–438, 1721.
- [7] R.-A.F. de Réaumur. Observations du thermomètre faites par M. COSSIGNY correspondant de l'Académie à l'Île de Bourbon à l'Île de France, à Madagascar et dans la route depuis l'Orient jusqu'à ces îles pendant l'année 1732 et partie de l'année 1733. Comparées avec les observation du thermomètre faites à Paris pendant le même temps. *Mémoires de l'académie royale des sciences de Paris*, 1735:417–438, 1733.
- [8] R.-A.F. de Réaumur. Suite des observations du thermomètre faite à l'Île de Bourbon par M. COSSIGNY, correspondant de l'académie et le résultat de celles de chaque mois faite à Paris pendant l'année 1734 avec un thermomètre pareil à celui de M. COSSIGNY. *Mémoires de l'académie royale des sciences de Paris*, 1736:553–563, 1734.
- [9] R.-A.F. de Réaumur. Observations du thermomètre faites à Paris pendant l'année 1735 comparées avec celles qui ont été faites sous la ligne, à l'Île de France, à Alger et en quelques-unes de nos Îles de l'Amérique. *Mémoires de l'académie royale des sciences de Paris*, 1738:545–576, 1735.
- [10] R.-A.F. de Réaumur. Observations du thermomètre faites à Paris pendant l'année 1736 comparées avec celles qui ont été faites pendant la même année dans différentes parties du monde. *Mémoires de l'académie royale des sciences de Paris*, 1739:469–502, 1736.
- [11] R.-A.F. de Réaumur. Observations du thermomètre faites à Paris pendant l'année 1737 comparées avec celles qui ont été faites dans les climats très différents de celui de Paris. *Mémoires de l'académie royale des sciences de Paris*, 1740:470–490, 1737.
- [12] R.-A.F. de Réaumur. Observations du thermomètre pendant l'année 1738 faites à Paris, à l'Île de France, à Pondichéry et au Sénégal, et la comparaison de ces observations. *Mémoires de l'académie royale des sciences de Paris*, 1740:387–403, 1738.
- [13] R.-A.F. de Réaumur. Observations du thermomètre pendant l'année 1739 faites à Paris et en différents pays. *Mémoires de l'académie royale des sciences de Paris*, 1741:447–466, 1739.
- [14] R.-A.F. de Réaumur. Observations du thermomètre faites en 1740 à Paris, et dans d'autres endroits, soit du Royaume, soit des Pays étrangers. *Mémoires de l'académie royale des sciences de Paris*, 1742:539–566, 1740.
- [15] G. Grolemond and H. Wickham. Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3):1–25, 2011.

- [16] A. Guerreau. Climat et vendanges (XIV^e-XIX^e siècles) : révisions et compléments. *Histoire & Mesure*, 10:89–147, 1995.
- [17] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024.
- [18] D. Rousseau. Les températures mensuelles en région parisienne de 1676 à 2008. *La Météorologie*, 2009(67):43–55, 2009.
- [19] D. Rousseau. Les moyennes mensuelles de températures à Paris de 1658 à 1675. D’Ismaël Boulliau à Louis Morin. *La Météorologie*, 2013(81):11–22, 2013.
- [20] D. Rousseau. Le cahier d’observations météorologiques de Réaumur. ses mesures de températures de 1732 à 1757. *La Météorologie*, (105):21–28, 2019.
- [21] D. Rousseau. La série des températures journalières à Paris de 1658 à 2023. *La Météorologie*, 2024(127):33–42, 2024.
- [22] Sun Microsystems. XDR: external data representation standard. RFC 1014. Technical report, Network Working Group, 1987.
- [23] Yang Xiang, Sylvain Gubian, Brian Suomela, and Julia Hoeng. Generalized simulated annealing for efficient global optimization: the GenSA package for R. *The R Journal Volume 5/1, June 2013*, 2013.