

Ajustement du modèle de DE RÉAUMUR (1735) aux données de DE RÉAUMUR (1735-1740)

Pr Jean R. Lobry

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

1 Ajustement du modèle de de Réaumur

1.1 Le modèle

LE modèle de DE RÉAUMUR [8] 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 exprimées en degrés Celcius, 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, donc entre 1 et 365 ou 366 selon que l'année est bissextile ou non.

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

1.2 Les données

IL s'agit des données de température publiées par DE RÉAUMUR (*cf.* annexe page 17) et des dates de vendange (*cf.* annexe page 13) pour les années 1735 à 1740 incluses pour la région Île-de-France. Les importer dans  avec :

```
chmin <- "https://pbil.univ-lyon1.fr/R/donnees/tdr4R/"
load(url(paste0(chmin, "Reaumet.Rda")))
load(url(paste0(chmin, "Reauphe.Rda")))
```

1.3 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 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)
}
```

1.4 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 `Reauphe` :

```
Reauphe$théo <- NA # colonne des valeurs prédites
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)))
}
```

1.5 Estimation des paramètres

ON utilise ici la fonction de base `R::nlm()` pour trouver la valeur des paramètres qui minimise la somme des carrés des écarts. Voir la fiche de TD¹ « ré-

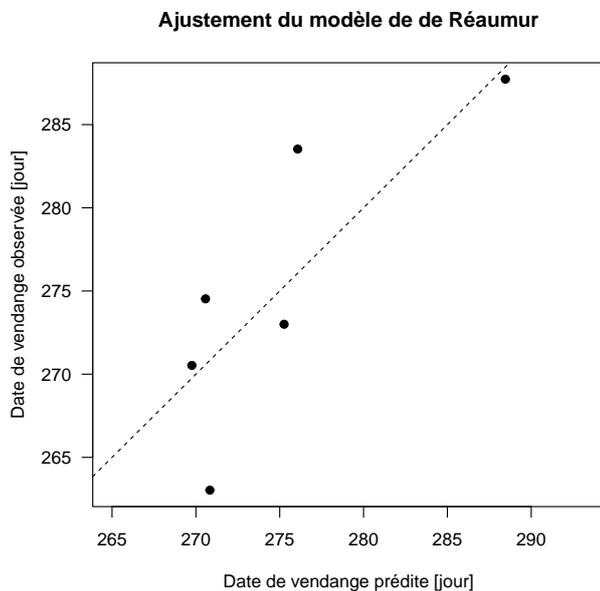
¹tdr46 : <https://pbil.univ-lyon1.fr/R/pdf/tdr46.pdf>

gression non linéaire » pour plus de détails sur les sorties de cette fonction.

```
(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 $F^* = 2006.4$ (soit environ 15 °C en moyenne pour les 134 jours entre le 21 mai et la date moyenne des vendanges le 2 octobre). 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 :

```
invisible(SSRR1(resnlm$estimate)) # pour calculer les valeurs prédites
x <- Reauphe$théo ; y <- Reauphe$Day_of_Year
plot(x, y, asp = 1, las = 1, main = "Ajustement du modèle de de Réaumur",
     pch = 19, xlab = "Date de vendange prédite [jour]",
     ylab = "Date de vendange observée [jour]")
abline(c(0, 1), lty = 2)
```



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.

1.6 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,

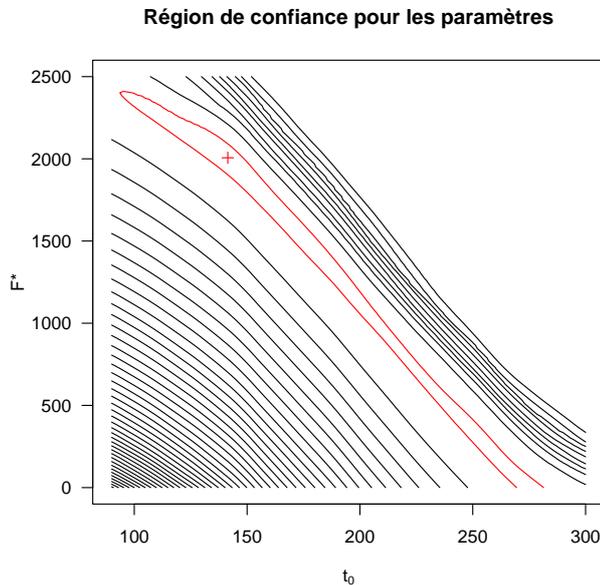
$$\theta \mid S(\theta) \leq S(\hat{\theta}) \left(1 + \frac{p}{n-p} F_{p;n-p}^{\alpha}\right)$$

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. 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, $\hat{\theta}$, de la valeur des paramètres. La courbe en rouge donne la région de confiance pour les paramètres :

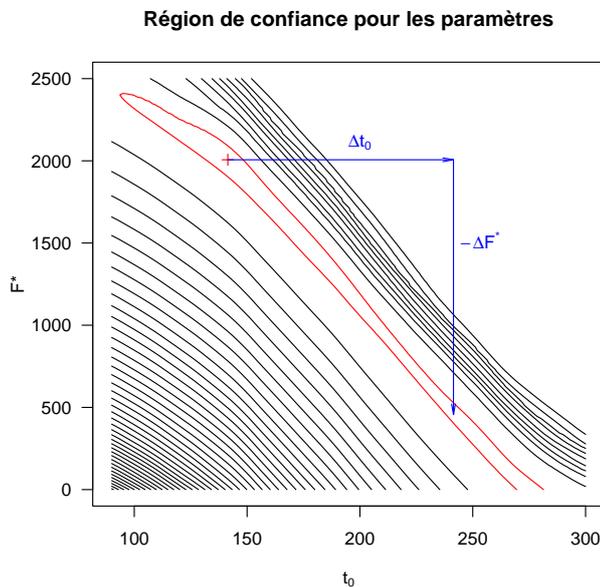
```
load(url(paste0(chmin, "matSSR.Rda")))
contour(p1seq, p2seq, matSSR, nlevels = 50, drawlabels = FALSE,
        xlab = expression(t[0]), ylab = "F*", las = 1,
        main = "Région de confiance pour les paramètres")
contour(p1seq, p2seq, matSSR, levels = seuil, drawlabels = FALSE,
        add = TRUE,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



1.6.1 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 :

```
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")
arrows(resnlm$estimate[1], resnlm$estimate[2], resnlm$estimate[1] + 100, resnlm$estimate[2],
        code = 2, angle = 10, length = 0.1, col = "blue")
text(200, 2000, expression(Delta*t[0]), pos = 3, col = "blue")
arrows(resnlm$estimate[1] + 100, resnlm$estimate[2], resnlm$estimate[1] + 100, resnlm$estimate[2] - 1550,
        code = 2, angle = 10, length = 0.1, col = "blue")
text(240, 1500, expression(-Delta*F~plain("*")), pos = 4, col = "blue")
```



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.

1.6.2 Présence de mesa

DANS le quadrant 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 :

```
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")
(resnlmMesa <- nlm(SSRR1, c(250, 2000)))

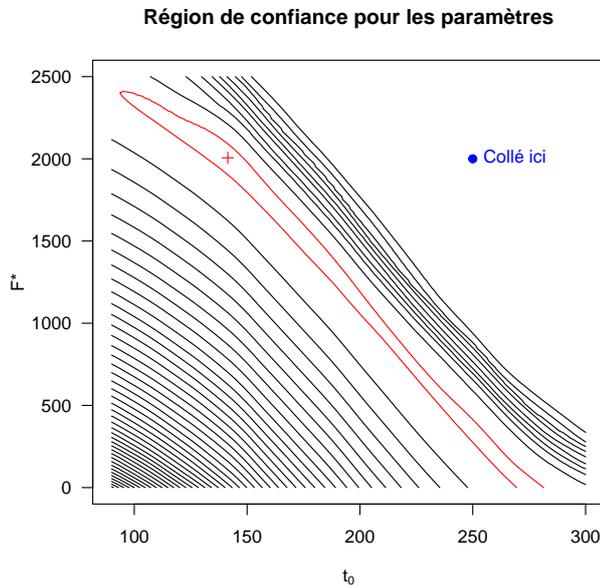
$minimum
[1] 48580.39
$estimate
[1] 250 2000

$gradient
[1] 0 0

$code
[1] 1

$iterations
[1] 0

points(resnlmMesa$estimate[1], resnlmMesa$estimate[2], pch = 19, col = "blue")
text(resnlmMesa$estimate[1], resnlmMesa$estimate[2], "Collé ici", col = "blue", pos = 4)
```

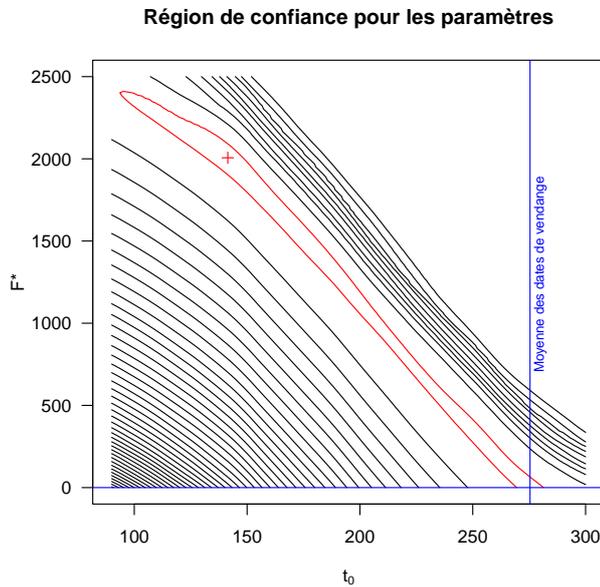


Ceci illustre tout l'intérêt d'avoir une bonne estimation initiale des paramètres quand on fait de la régression non-linéaire. L'origine de cette mesa est que lorsque l'on prend des seuils F^* trop élevés, il devient alors impossible de les atteindre en sommant tous les degrés disponibles. Dans ce cas de figure la fonction `dddR1()` renvoie pour t_d la plus grande date disponible dans le jeu de données (ici 365), la somme des carrés des écarts est alors insensible aux variations des paramètres.

1.6.3 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 :

```
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")
moy <- mean(Reauphe$Day_of_Year)
abline(v = moy, col = "blue")
abline(h = 0, col = "blue")
text(moy, 700, "Moyenne des dates de vendange", srt = 90, col = "blue", pos = 4, cex = 0.75)
```



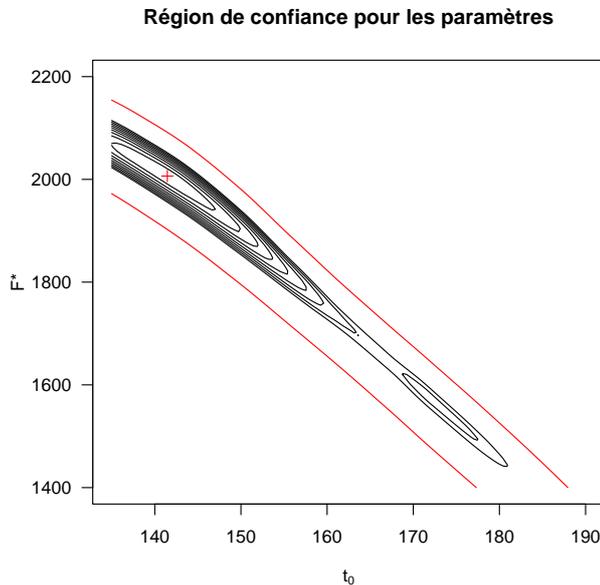
LE modèle de DE RÉAUMUR à 2 paramètres ne fait donc pas significativement mieux que le modèle à 1 paramètre qui résume les données par leur moyenne. On dit que le modèle est sur-paramétré : avec 5 points observés il n'y a pas suffisamment de données pour estimer de façon fiable les paramètres du modèle.

1.6.4 Minimum local

SI l'on zoome 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")

load(url(paste0(chmin, "minloc.Rda")))
contour(p1seq, p2seq, matSSR,
        levels = seq(resnlm$minimum, 2*resnlm$minimum, le = 10),
        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")
```



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 :

```
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")
badguess <- c(230, 800)
(resnlmMinLoc <- nlm(SSRR1, badguess))

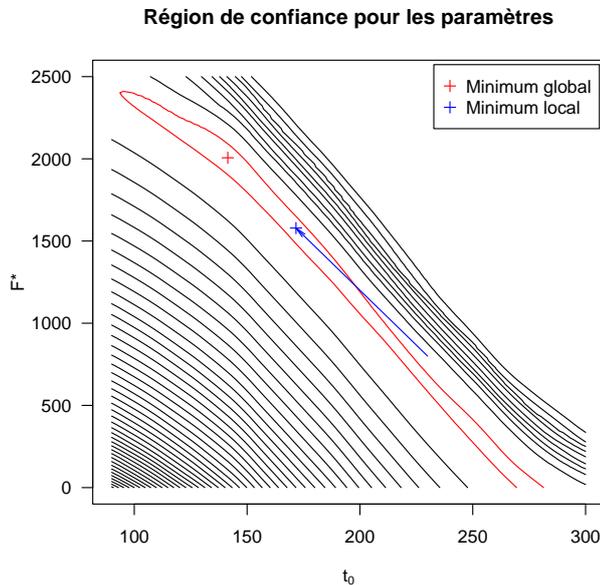
$minimum
[1] 254.638
$estimate
[1] 171.6747 1578.7346

$gradient
[1] 8.277779e-10 -1.854293e-09

$code
[1] 1

$iterations
[1] 25

points(resnlmMinLoc$estimate[1], resnlmMinLoc$estimate[2], pch = 3, col = "blue")
arrows(badguess[1], badguess[2], resnlmMinLoc$estimate[1], resnlmMinLoc$estimate[2],
       angle = 10, length = 0.1, col = "blue", lwd = 1)
legend("topright", inset = 0.01, legend = c("Minimum global", "Minimum local"),
      col = c("red", "blue"), pch = 3)
```



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 palier ce problème on peut tenter d'utiliser des algorithmes de type « recuit simulé » (*simulated annealing*) avec par exemple la fonction `GenSA()` du paquet  éponyme [15] mais les temps de calcul sont alors 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
```

2 Structure des résidus

SI on compare les valeurs observées et les valeurs prédites, on peut constater que le modèle de DE RÉAUMUR a tendance à sous-estimer la variabilité inter-annuelle. C'est une tendance bien connue des modèles mécanistes utilisés en phénologie [1]. Par exemple dans notre cas l'amplitude des valeurs prédites est de 18.7 jours contre 24.7 jours pour celle des valeurs observées :

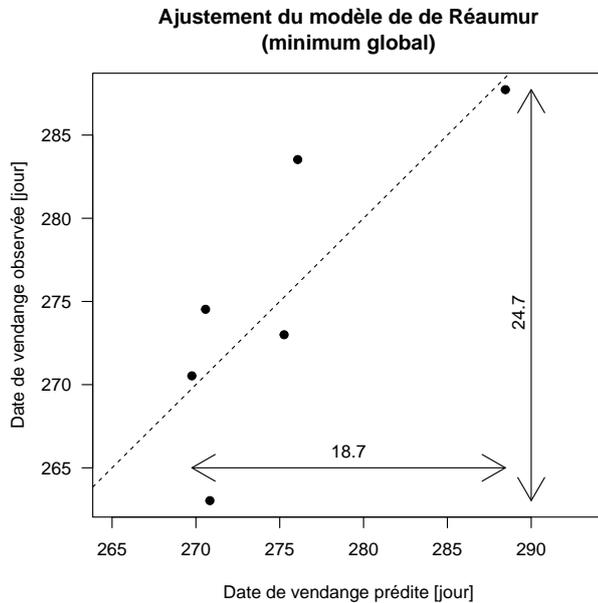
```
invisible(SSRR1(resnlm$estimate)) # pour calculer les valeurs prédites
x <- Reauphe$théo ; y <- Reauphe$Day_of_Year
plot(x, y, asp = 1, las = 1, main = "Ajustement du modèle de de Réaumur\n(minimum global)",
     pch = 19, xlab = "Date de vendange prédite [jour]",
     ylab = "Date de vendange observée [jour]")
abline(c(0, 1), lty = 2)
```

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

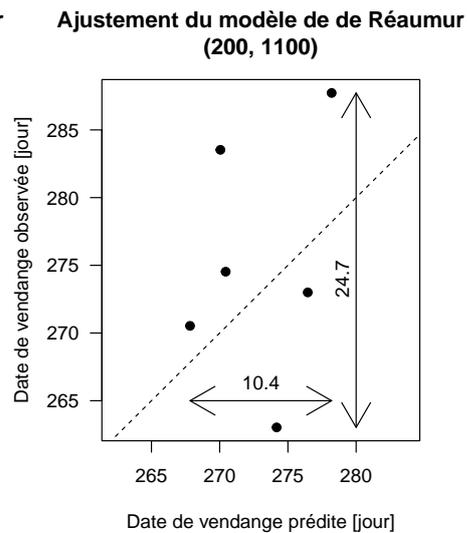
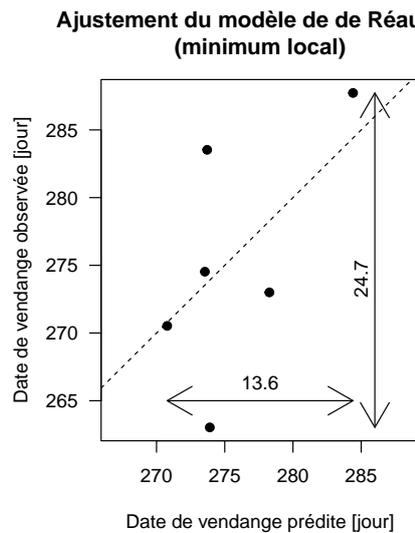
```

arrows(min(x), 265, max(x), 265, code = 3)
text(0.5*(max(x) + min(x)), 265, signif(diff(range(x)), 3), pos = 3)
arrows(290, min(y), 290, max(y), code = 3)
text(290, 0.5*(max(y) + min(y)), signif(diff(range(y)), 3), pos = 2, srt = 90)

```



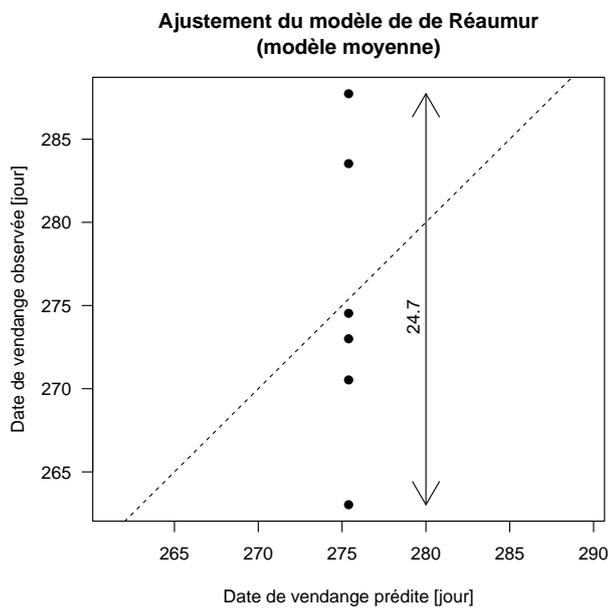
A U niveau des résidus ce qui est généralement observé c'est que les dates précoces ne sont pas prédites suffisamment précocement et les dates tardives pas suffisamment tardivement. Ce n'est pas bien visible ici parce qu'il n'y a pas beaucoup de points, mais on va forcer le trait en prenant d'autres valeurs des paramètres tout en restant dans la région de confiance : celle du minimum local ($t_0 = 171.7$ et $F^* = 1579$) et une autre un peu plus loin ($t_0 = 200$ et $F^* = 1100$).



QUAND on augmente t_0 et que l'on diminue F^* pour compenser et rester dans la région de confiance, on réduit l'amplitude des dates de jours prédits : on passe de 18.7 à 13.6 jours puis 10.4 jours. On peut pousser le bouchon jusqu'au cas extrême où le modèle de DE RÉAUMUR se contente de résumer les valeurs par leur moyenne ($t_0 = 275.4$, $F^* = 0$) :

```

y <- Reauphe$Day_of_Year
x <- rep(mean(y), length(y))
plot(x, y, asp = 1, las = 1, main = "Ajustement du modèle de de Réaumur\n(modèle moyenne)",
     pch = 19, xlab = "Date de vendange prédite [jour]",
     ylab = "Date de vendange observée [jour]")
abline(c(0, 1), lty = 2)
arrows(280, min(y), 280, max(y), code = 3)
text(280, 0.5*(max(y) + min(y)), signif(diff(range(y)), 3), pos = 2, srt = 90)
    
```



IL n'y a plus aucune variabilité en abscisse. Toutes les dates précoces sont prédites comme insuffisamment précoces, toutes les dates tardives sont prédites comme insuffisamment tardives. La structure des résidus peut donc s'interpréter comme une conséquence mécanique de la forte corrélation structurelle entre t_0 et F^* et qui sera d'autant plus marquée que l'on est proche de la sur-paramétrisation.

3 Annexe données phéno

3.1 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 [8, pp. 557-559] : « [c]ette 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 [13, p. 554] : « [l]es 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. »

Ce 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 [14] très accessible et éclairante, sont un objet d'étude depuis au moins 1870. J'ai utilisé ici la base de données GHD [3] 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	Ile 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$Year %in% 1735:1740, ]
```

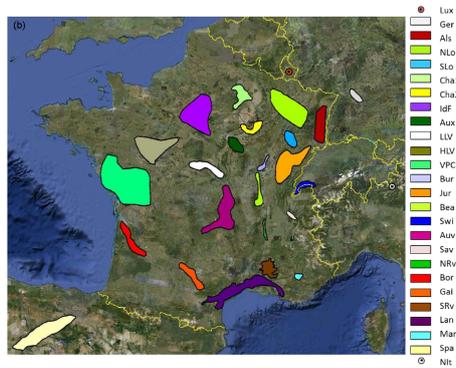
```
Year IdF
382 1735 40.52857
383 1736 26.52857
384 1737 20.02857
385 1738 31.52857
386 1739 30.00000
387 1740 43.72973
```

DEPUIS les travaux d'Alfred ANGOT en 1883 il est d'usage [3] d'exprimer la date des vendanges en jour par un entier relatif au 1^{er} septembre⁴, valant

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.

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 :



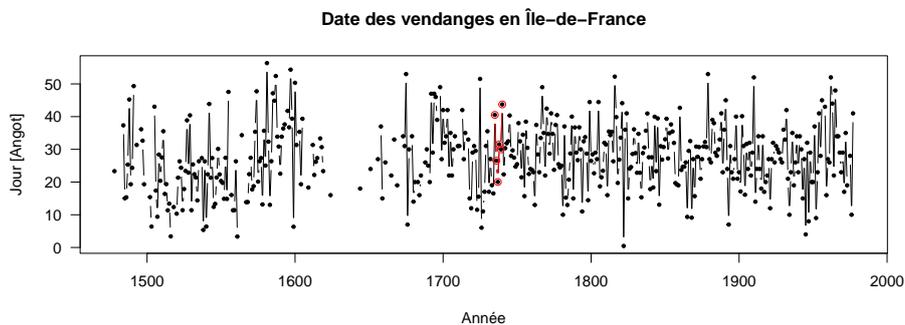
POUR la période qui nous intéresse, de 1735 à 1740 inclus, les localités concernées sont [3, matériel supplémentaire] :

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

3.2 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")
```

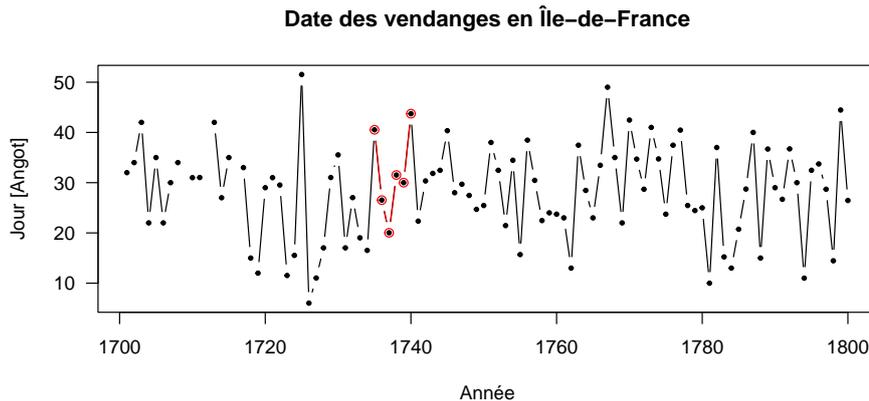


⁵GHD-RCS : *Grape Harvest Date - Regional Composite Series*.

⁶Copie d'écran d'une partie de la figure 1(b) de [3].

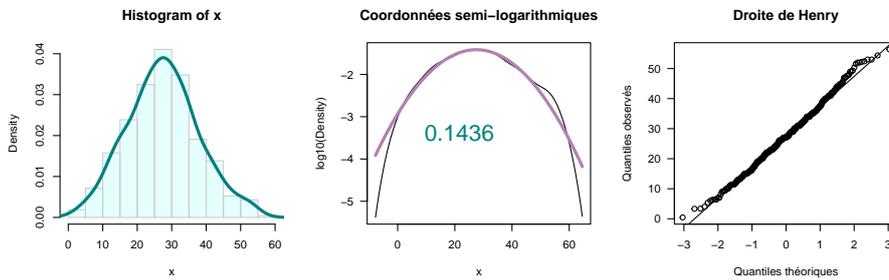
SI on zoome sur le XVIII^e siècle, on peut constater, comme le rapporte DE RÉAUMUR (*cf. supra*), que les vendanges de 1735 et 1741 étaient plutôt tardives :

```
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")
```



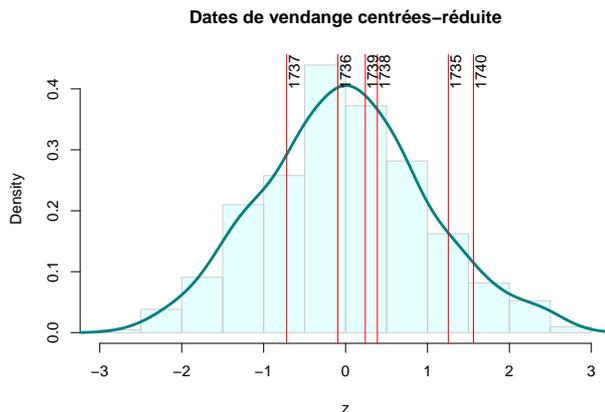
LA distribution des dates de vendange est remarquablement normale, on est d'ailleurs dans l'incapacité de rejeter l'hypothèse nulle du test de SHAPIRO et WILK, alors qu'avec 419 points on dispose d'une puissance confortable. La date médiane de vendange correspond au 27 septembre. On note une très légère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

```
library(e1071) # pour skewness()
myplot <- function(x, mylwd = 3, mycol = rgb(0, 0.5, 0.5), myborder = grey(0.8),
                  mycolhist = rgb(0, 1, 1, 0.1), mycolnorm = rgb(0.7, 0.5, 0.7), ...){
  old.par <- par(no.readonly = TRUE)
  on.exit(par(old.par))
  par(mfrow = c(1, 3))
  hx <- hist(x, plot = FALSE)
  dx <- density(x)
  hist(x, freq = FALSE, border = myborder, col = mycolhist, ylim = c(0, max(hx$density, dx$y)), ...)
  lines(dx, lwd = mylwd, col = mycol)
  plot(dx$x, log10(dx$y), type = "l", main = "Coordonnées semi-logarithmiques", xlab = "x",
       ylab = "log10(Density)", las = 1)
  legend("center", legend = round(skewness(x), 4), cex = 2, text.col = mycol, bty = "n",
        adj = 0.5)
  lines(dx$x, log10(dnorm(dx$x, mean(x), sd(x))), col = mycolnorm, lwd = mylwd)
  qqnorm(x, main = "Droite de Henry", xlab = "Quantiles théoriques",
        ylab = "Quantiles observés", las = 1)
  qqline(x)
}
x <- idf[complete.cases(idf), "IdF"]
myplot(x)
shapiro.test(x)
      Shapiro-Wilk normality test
data:  x
W = 0.9957, p-value = 0.3077
```



SI 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 deça de 2σ :

```
mycol <- rgb(0, 0.5, 0.5); myborder <- grey(0.8) ; mycolhist <- rgb(0, 1, 1, 0.1)
idf$z <- scale(idf$IdF)
hx <- hist(idf$z, plot = FALSE)
dx <- density(idf$z, na.rm = TRUE)
hist(idf$z, freq = FALSE, border = myborder, col = mycolhist,
     ylim = c(0, max(hx$density, dx$y)), xlab = "z",
     main = "Dates de vendange centrées-réduite")
lines(dx, lwd = 3, col = mycol)
abline(v = idf[idf$Year %in% 1735:1740, "z"], col = "red")
text(x = idf[idf$Year %in% 1735:1740, "z"], y = 0.4,
     idf[idf$Year %in% 1735:1740, "Year"], srt = 90, pos = 4, xpd = NA)
```



3.3 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

3.4 Sauvegarde post-prétraitement

```
Reauphe <- idf[idf$Year %in% 1735:1740, c("Year", "Day_of_Year")]
comment(Reauphe) <- c("Time Stamp:", as.character(Sys.time()))
save(Reauphe, file = "Reauphe.Rda")
```

4 Annexe données météo

4.1 Origine des données

4.1.1 Sources

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 [8, 9, 10, 11, 12, 13]. Une version digitalisée de ces articles au format PDF est disponible sur le site de l'académie (plus très royale) des sciences de Paris⁷.

4.1.2 Mise en garde - *Note of caution*

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 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.

4.1.3 Le fichier texte

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 °C 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*)

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

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

Notes.Réaumur : précisions de DE RÉAUMUR (character)

Notes.JLO : mes précisions (character)

4.1.4 Importation du fichier texte

```
météo <- read.table(paste0(chmin, "ReaumurData.csv"), header = TRUE, sep = "\t", dec = ".")
# Calcul de la date au format POSIX
météo$Date <- with(météo, as.POSIXct(sprintf("%04d-%02d-%02d", Year, Month, Day.of.Month)))
# Sanity check:
all(diff(météo$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 en essayant de convertir un 29 février de chaque année au format POSIX :

```
for(i in 1735:1740){
  tryres <- try(as.POSIXct(paste(i, "-02-29", sep = "")), silent = TRUE)
  if(inherits(tryres, "try-error")){
    cat(paste(i, "non bissextile\n"))
  } else {
    cat(paste(i, "bissextile\n"))
  }
}
1735 non bissextile
1736 bissextile
1737 non bissextile
1738 non bissextile
1739 non bissextile
1740 bissextile
```

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

Février 1736

26 $8\frac{1}{2}$ $3\frac{3}{4}$
27 $8\frac{1}{2}$ $1\frac{1}{2}$
28 $2\frac{3}{4}$ $1\frac{1}{2}$
29 $\frac{1}{4}$ $2\frac{1}{2}$

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 Notation des degrés négatifs

DE RÉAUMUR rappelle à chaque fois la notation qu'il utilise pour noter les degrés négatifs, par exemple en 1740 [13, p. 539] : « [u]n 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 °C 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

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

[1739.]		A O U S T.			
nidi.	J.	Degrés du Matin.		Degrés d'Après-midi.	
rés.		Heures.	Degrés.	Heures.	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	$18\frac{1}{4}$
$18\frac{1}{2}$	3	11	20
18	4	13	$22\frac{1}{4}$
$19\frac{1}{2}$	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 [8], 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 [4], la notation $2 \cdot 75$ sur l'axe vertical de gauche doit se comprendre comme $-2 + 0.75 = -1.25$:

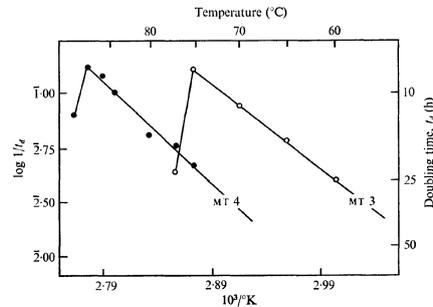


Fig. 1. Arrhenius plots of growth of MT3 and MT4 at pH 3.0.

4.3 Heures d'observation

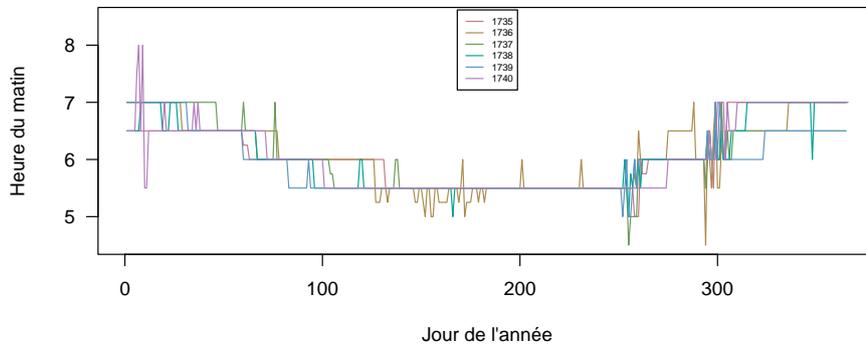
LES heures d'observation sont données avec une précision de $\frac{1}{2}$ heure, parfois avec une précision de $\frac{1}{4}$ d'heure. Il y a en général une observation le matin et une le soir. Le 4 juin 1736 il y a deux observations le matin, et il y a 18 jours pour lesquels il y a deux observations l'après-midi⁸. 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.

4.3.1 Heures du matin

```
# 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)
```

⁸ 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.

Heure d'observation de la température matinale



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.

4.3.2 Heures de l'après-midi

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

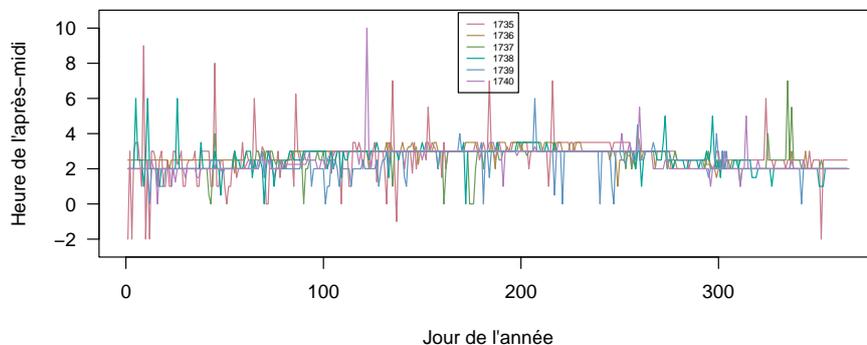
Deuxièmement, le 1^{er} mai 1740, pour la valeur 10 de l'après-midi, DE RÉAUMUR précise qu'il s'agit de 10 h *du soir* :

M A I. [1740.]					
J.	Degrés du matin.		Degr. d'après-midi.		J. D
	Heures.	Degrés.	Heures.	Degrés.	Heu
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.

4.4 Géolocalisation

LES observations de DE RÉAUMUR se font principalement à Paris [6, p. 430] : « [n]ous allons à présent rapporter les observations que nous avons faites à Paris, rue S^t Thomas du Louvre, à l'hôtel d'Uzès⁹ [...] ». Il précise en 1735 [6,

⁹Cela ne peut pas être l'hôtel d'Uzès bâti en 1767 par LEDOUX puisque DE RÉAUMUR écrit ceci en 1733.

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 observations 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 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 [5] en 1721 un mémoire intitulé : « [m]oyen 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-Dyé (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 redescend¹⁰ 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.

4.5 Les degrés de température

4.5.1 Unités

AVEC son thermomètre DE RÉAUMUR utilisait une échelle allant de 0 pour la température de fusion de la glace (qui est de 0 °C) à 80 pour la température d'ébullition de l'éthanol (qui est de 78.4 °C), on est donc très proche des degrés Celsius. Il y a une source de confusion possible ici parce qu'une échelle dite de Réaumur a été introduite ensuite (vers 1750) avec 80 unités entre le point de fusion de la glace et celui de l'ébullition de l'eau.

4.5.2 Précision

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 20¹¹ » :

¹⁰ Monter 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].

¹¹ J'ai renseigné le fichier texte avec une valeur de 19.5 dans ce cas.

16 10 13
17 11 $\frac{2}{3}$ 19 ou 20
18 11	3..... 19
19	6..... 11	3 $\frac{1}{2}$ 20 $\frac{1}{4}$
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, 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 4 $\frac{1}{2}$ & un peu plus.	à 1 $\frac{3}{4}$ 2
25 4 $\frac{1}{2}$ & un peu moins.	2 $\frac{1}{2}$ $\frac{2}{3}$
26 1 $\frac{1}{2}$ 2
27 3 4
28 2 $\frac{1}{2}$ & un peu plus bas. 2

4.5.3 Résumé des tableaux

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 :

*RÉSULTATS 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 ^d $\frac{1}{2}$ 15. }	13. } à 1 ^h à 2 ^d $\frac{1}{2}$ 14. }	30. à 7 ^h à 7 ^d $\frac{1}{2}$	30. à 2 ^h à 8 ^d $\frac{1}{2}$
FÉVRIER.			
5. à 7 ^h à 3 ^d $\frac{2}{3}$	4. à 3 ^h à 0 ^d $\frac{1}{2}$	21. à 7 ^h à 6 ^d $\frac{1}{2}$	23. à 2 ^h $\frac{1}{2}$ à 8 ^d $\frac{1}{2}$
MARS.			
3. } à 6 ^h à 0 ^d 4. } 6. } 22. } 24. }	4. } à 2 ^h $\frac{1}{2}$ à 4 ^d $\frac{1}{2}$ 5. }	7. à 6 ^h à 7 ^d $\frac{1}{2}$	17. à 2 ^h à 18 ^d $\frac{1}{2}$

C'EST d'ailleurs sous cette forme condensée que DE RÉAUMUR avait commencé à publier ses observations pour les années 1732 à 1734 [6, 7] avant qu'il ne

décide en 1735 [8] : « [l]orsque 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. »

4.5.4 Valeurs manquantes

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"])
  }
}
```

CE 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.	E
	Heures.	Degrés.	Heures.	Degrés.		Heur
1	à 7.....	à 4½	à 2.....	à 5½	1	à 7
2	7.....	3			2	7
3	2½			3	..
4	1½	2.....	3½	4	..
5	½	3½	5	..

J'ai forcé ces deux valeurs à +4.0 °C :

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

4.5.5 Température moyenne

DE RÉAMUR explique de façon détaillée comment il calcule les températures moyennes journalières [8, p. 558] : « [c]omme 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,

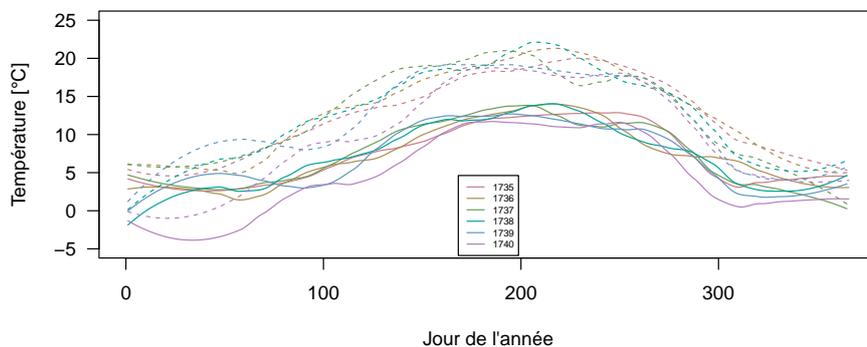
à 3 heures de l'après-midi, elle était à 8 degrés ; ces degrés ajoutés ensemble donnent 12 degrés ; j'en prends la moitié, 6 degrés, pour l'expression du degré du chaud moyen du 3 avril. »

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

4.5.6 Chroniques

```
plot(0, xlim = c(1, 366), ylim = c(-5, 25), cex = 0, las = 1,
     xlab = "Jour de l'année", ylab = "Température [°C]",
     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 = doym)), type = "l", col = mycol[i])
    PM.lo <- loess(d.PM-doy, span = span)
    points(doy, predict(PM.lo, data.frame(doy = doym)), 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 °C. Il écrit [13, 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. »

4.6 Sauvegarde post-prétraitement

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

References

- [1] D. Basler. Evaluating phenological models for the prediction of leaf-out dates in six temperate tree species across central Europe. *Agricultural and Forest Meteorology*, 217:10–21, 2016.
- [2] E.M.L. Beale. Confidence regions in non-linear estimation. *Journal of the Royal Statistical Society*, 22B:41–88, 1960.
- [3] 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.
- [4] M. De Rosa, A. Gambacorta, and J.D. Bu'lock. Extremely thermophilic acidophilic bacteria convergent with *Sulfolobus acidocaldarius*. *Journal of General Microbiology*, 86:156–164, 1975.
- [5] 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.
- [6] 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 observations du thermomètre faites à Paris pendant le même temps. *Mémoires de l'académie royale des sciences de Paris*, 1735:417–438, 1733.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.

- [12] 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.
- [13] 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.
- [14] A. Guerreau. Climat et vendanges (XIV^e-XIX^e siècles) : révisions et compléments. *Histoire & Mesure*, 10:89–147, 1995.
- [15] 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.