

Ajustement du modèle de croissance de Monod aux données de Monod (ou pas)

J.R. Lobry

Ajustement par intégration numérique et régression non-linéaire du
modèle de croissance de Monod aux données de Monod.



Courtesy of Cold Spring Harbor
Laboratory Archives
Noncommercial, educational use only.
Jacques Monod
(1910-1976)

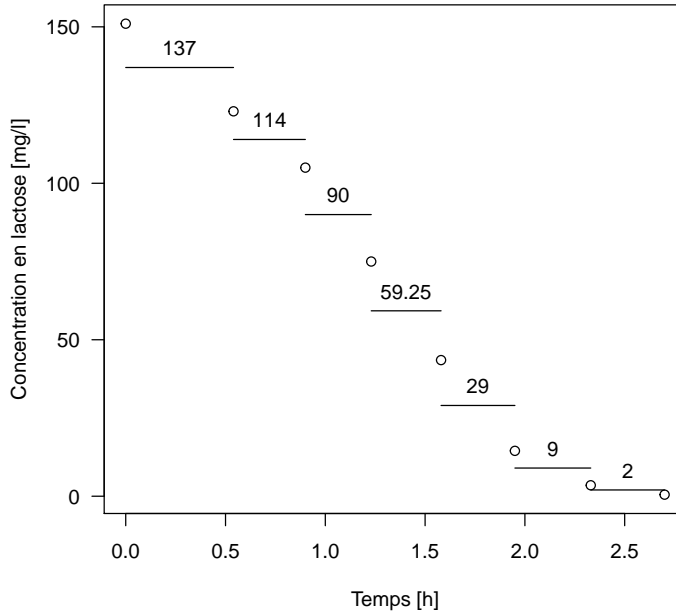
1 Les données

Elles sont extraites de l'expérience I de la table XIX p 74 (*cf* figure 1 page 9)
et de la figure 18 p 71 (*cf* figure 2 page 9) de la thèse de Jacques Monod [3, 4]
et reproduites ici dans la table 1 page 2.

Attention, comme me l'a signalé par courriel Eduardo A. Sánchez Torres le
2020-12-13 20:22, il y a un très sérieux problème dans la reconstitution de
ce jeu de données. Dans mon souvenir j'ai utilisé la figure 18 pour estimer la
concentration terminale de lactose à 0.5 mg.l^{-1} et déduit de proche en proche
les concentrations aux bornes des intervalles de temps :

```
m41 <- read.table("http://pbil.univ-lyon1.fr/R/donnees/Monod1941.txt", sep = "\t", header = TRUE)
with(m41, plot(t, s, xlab = "Temps [h]", ylab = "Concentration en lactose [mg/l]",
              las = 1, main = "Comparaison des valeurs"))
for(i in 1:(nrow(m41) - 1)){
  smean <- with(m41, (s[i] + s[i + 1])/2)
  tmean <- with(m41, (t[i] + t[i + 1])/2)
  with(m41, segments(x0 = t[i], y0 = smean, x1 = t[i + 1], y1 = smean))
  text(tmean, smean, smean, pos = 3)
}
```

Comparaison des valeurs



On retrouve bien les valeurs moyennes de la table XIX (e.g 137, 114, 90) mais il y a un affreux 59.25 au lieu du 43 attendu. Il y a donc eu une erreur lors de la reconstitution de ce jeu de données. Je l'ai laissé tel quel dans la suite du document, mais on ne peut pas considérer qu'il s'agit des données originelles de Jacques Monod.

	t	x	s
1	0.00	15.5	151.0
2	0.54	23.0	123.0
3	0.90	30.0	105.0
4	1.23	38.8	75.0
5	1.58	48.5	43.5
6	1.95	58.3	14.5
7	2.33	61.3	3.5
8	2.70	62.5	0.5

TABLE 1 – Croissance de *Escherichia coli* avec le lactose comme substrat limitant. t est le temps exprimé en heures. x est la biomasse dans les unités de DO de Monod, on a 1 unité de DO \approx 0.75 mg de biomasse sèche par litre. s est la concentration en lactose exprimée en mg.l^{-1}

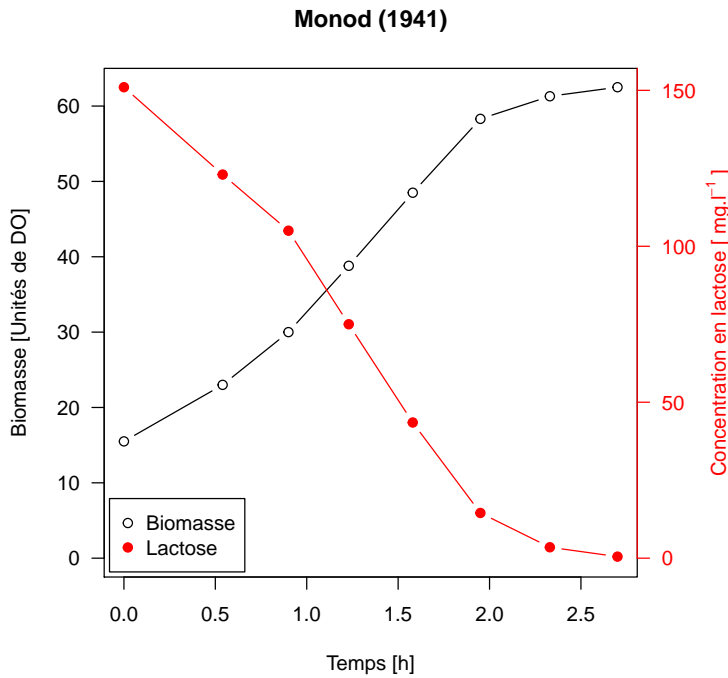
Représenter sur un même graphique la croissance de la biomasse et la consommation du substrat au cours du temps. C'est un exercice amusant parce que l'on a besoin deux échelles différentes en ordonnée, mais sous R on en vient vite à bout :

```
par(mar = c(5,4,4,4) + 0.1)
plot(x = m41$t, y = m41$x, las = 1, type = "b", xlab = "Temps [h]",
```

```

ylab = "Biomasse [Unités de DO]", ylim = c(0, max(m41$x)),
main = "Monod (1941)"
yscale <- max(m41$x)/max(m41$s)
lines(x = m41$t, y = yscale*m41$s, type = "b", pch = 19, col = "red")
axis(side = 4, at = yscale*pretty(m41$s), labels = pretty(m41$s), las = 1, col = "red",
col.axis = "red")
mtext(text = expression(paste("Concentration en lactose [ mg.", l^-1, " ]", sep = "")),
line = 3, side = 4, col = "red")
legend("bottomleft", inset = 0.01, legend = c("Biomasse","Lactose"),
pch = c(1,19), col = c("black","red"))

```



2 Le modèle

Le modèle de Monod s'écrit sous la forme d'un système de deux équations différentielles,

$$\begin{cases} \frac{dx}{dt} = x \frac{\mu_m s}{K_s + s} \\ \frac{ds}{dt} = -\frac{1}{Y} \frac{dx}{dt} \end{cases} \quad (1)$$

où x est la biomasse au temps t , s la concentration en substrat limitant au temps t , Y le rendement de la croissance, μ_m le taux de croissance maximum, et K_s la concentration en substrat telle que le taux de croissance soit demi-maximal. Une expérience en cuvette (*batch experiment*) consiste typiquement, à partir d'un inoculum ($x = x_0$) et d'une concentration initiale en substrat ($s = s_0$) au temps initial ($t = 0$), à mesurer l'évolution de la biomasse, du substrat, ou des deux, au cours du temps, jusqu'à ce que la biomasse finale ($x = x_m$) soit atteinte quand tout le substrat est consommé ($s = 0$).

On suppose en général que le rendement est constant au cours de la croissance, ce qui est raisonnable quand le substrat limitant est un aliment énergétique [5]. Dans ce cas on estime simplement le rendement par :

$$Y = \frac{x_m - x_0}{s_0} \quad (2)$$

```
xm <- max(m41$x)
x0 <- min(m41$x)
s0 <- max(m41$s)
(Y <- (xm-x0)/s0)
[1] 0.3112583
```

Nous avons ici avec les données de la table 1 :

- $x_m = 62.5$
- $x_0 = 15.5$
- $s_0 = 151$
- Et donc $Y \approx 0.31$

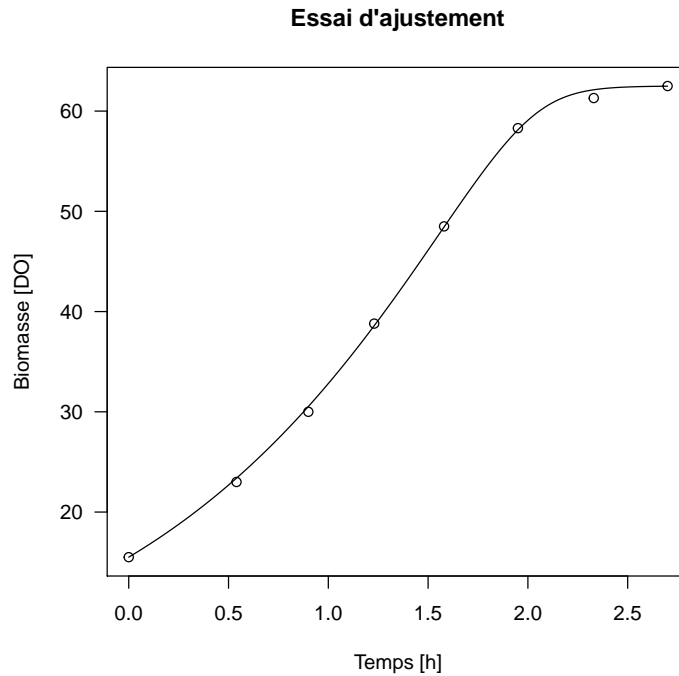
Le système 1 se simplifie alors en une simple équation différentielle :

$$\frac{dx}{dt} = x \frac{\mu_m \frac{x_m - x}{Y}}{K_s + \frac{x_m - x}{Y}} \quad (3)$$

3 Intégration du modèle

On essaye d'intégrer le modèle avec les valeurs des paramètres données dans [2], c'est à dire $\mu_m = 0.878 \text{ h}^{-1}$ et $K_s = 21.2 \text{ mg.l}^{-1}$:

```
library(deSolve)
monod <- function(t, y, parms){
  s <- (xm - y[1])/Y
  dy = y[1]*parms["mum"]*s/(parms["ks"] + s)
  return( list( dy ) )
}
theo <- lsoda(y = c(x = x0),
times = seq(from = min(m41$t), to = max(m41$t), length = 255),
func = monod,
parms = c(mum = 0.878, ks = 21.2))
plot(x = theo[,"time"], y = theo[,"x"], type = "l", xlab = "Temps [h]", ylab = "Biomasse [D0]", las = 1,
main = "Essai d'ajustement")
points(m41$t, m41$x)
```



L'ajustement est plutôt bon, mais comment retrouver la valeur des paramètres ?

4 Estimation des paramètres

On estime les paramètres au moindres carrés ordinaires, on commence par définir une fonction qui calcule la somme des carrés des écarts :

```
sce <- function(p){
  mum <- p[1]
  ks <- p[2]
  obs <- m41$x
  theo <- lsoda(y = c(x = x0),
  times = m41$t,
  func = monod,
  parms = c(mum = mum, ks = ks))
  return( sum((obs - theo[, "x"])^2) )
}
(scemin <- sce(c(0.878, 21.2)))
[1] 1.193019
```

On trouve 1.193 à comparer à 1.18 donné dans [2]. On minimise cette fonction :

```
(nlmfit <- nlm(sce, p = c(1,10)))
$minimum
[1] 1.1902
$estimate
[1] 0.8787674 21.3808949
$gradient
[1] -5.686562e-07 8.276995e-09
$code
```

```
[1] 1
$iterations
[1] 22
```

On trouve donc un minimum de 1.1902 pour $\mu_m = 0.8788 \text{ h}^{-1}$ et $K_s = 21.3809 \text{ mg.l}^{-1}$, ce qui est cohérent avec [2].

5 Région de confiance pour les paramètres

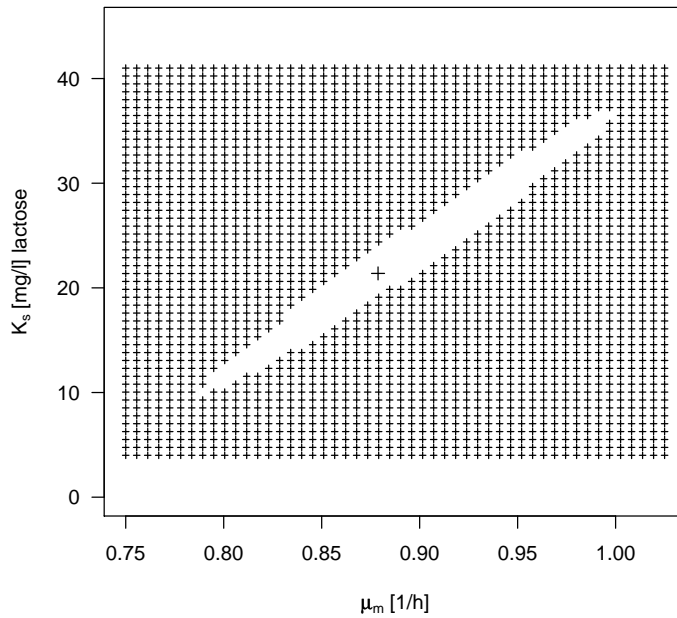
Une région de confiance pour la valeur des paramètres avec un risque de première espèce α est donnée [1] par l'ensemble des valeurs des paramètres telles que la somme des carrés des résidus n'excède pas un seuil donné,

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

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 le critère soit minimal.

On peut faire une exploration systématique :

```
scemin <- nlmfit$minimum
n <- nrow(m41)
p <- 2
alpha = 0.05
seuil <- scemin*( 1 + (p*qt(p = 1 - alpha, df1 = p, df2 = n - p))/(n - p) )
museq <- seq(from = 0.75, to = 1.025, length = 50)
ksseq <- seq(from = 4, to = 41, length = 50 )
scegrid <- matrix(nrow = length(museq), ncol = length(ksseq))
plot(nlmfit$estimate[1], nlmfit$estimate[2], pch = 3,
xlim = c(0.75, 1.025), ylim = c(0, 45), las = 1,
xlab = expression(paste(mu[m], " [1/h]")),
ylab = expression(paste(K[s], " [mg/l] lactose")))
for( i in 1:length(museq) ){
  for( j in 1:length(ksseq)){
    scegrid[i,j] <- sce(c(museq[i],ksseq[j]))
    if(scegrid[i,j] > seuil){
      points(museq[i],ksseq[j], pch = 3, cex = 0.5)
    }
  }
}
```



Puis utiliser `contour()` pour couper au niveau du seuil :

```
contour(x = museq,y=ksseq,z=scegrid, levels= seuil, drawlabels = FALSE,
xlab = expression(paste(mu[m], " [1/h]")),
ylab = expression(paste(K[s], " [mg/l] lactose")),
main ="Région de confiance des paramètres")
points(nlmfit$estimate[1], nlmfit$estimate[2], pch =3)
```

Région de confiance des paramètres

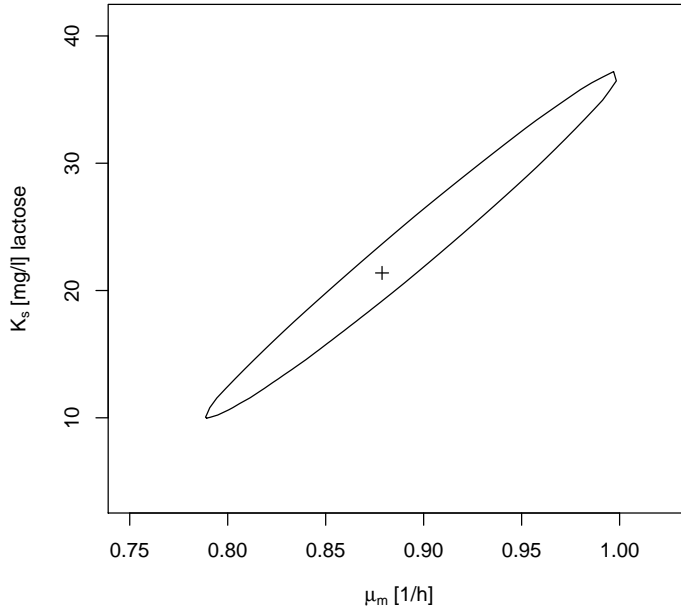


TABLEAU XIX.

Taux de croissance et temps de division des cultures de *B. coli* en fonction de la concentration moyenne du milieu en lactose.

Intervalle de densité	Intervalle de temps	Concentration moyenne	Taux de croissance moyen	Temps de division moyen
<i>Culture I :</i>				
61,3-62,5	0,37	2	0,07	860'
38,3-61,3	0,38	9	0,19	550'
48,5-58,3	0,37	29	0,72	85'
38,8-48,5	0,35	43	0,92	65'
30,0-38,8	0,33	90	1,13	53'
23,0-30,0	0,36	114	1,06	56'
15,5-23,0	0,54	137	1,05	57'

FIGURE 1 – Copie d'écran d'une partie de la table XIX page 74 de la thèse de Jacques Monod [3, 4].

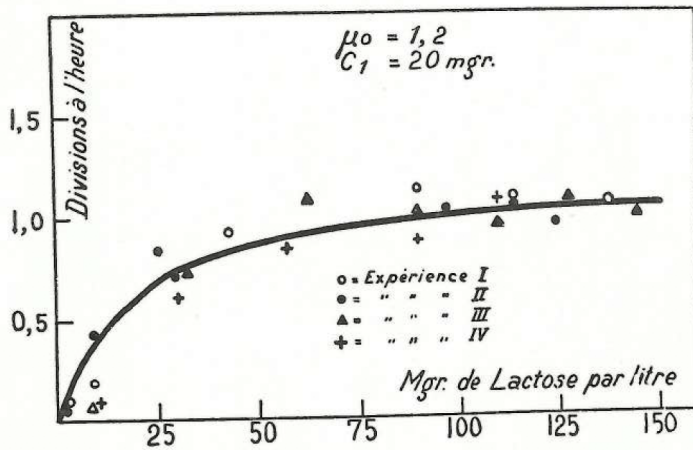


FIG. 18. — Taux de croissance des cultures de *B. coli* en milieu synthétique, en fonction de la concentration du milieu en lactose.

FIGURE 2 – Copie d'écran figure 18 page 71 de la thèse de Jacques Monod [3, 4].

Références

- [1] E.M.L. Beale. Confidence regions in non-linear estimation. *Journal of the Royal Statistical Society*, 22B :41–88, 1960.
- [2] J.R. Lobry and J.-P. Flandrois. Comparison of estimates of monod’s growth model parameters from the same data set. *Binary Computing in Microbiology*, 3 :20–23, 1991.
- [3] J. Monod. *Recherches sur la croissance des cultures bactériennes*. PhD thesis, Paris, 1941.
- [4] J. Monod. *Recherches sur la croissance des cultures bactériennes*. Herman, Paris, France, 1942.
- [5] U. Wanner and T. Egli. Dynamics of microbial growth and cell composition in batch culture. *FEMS Microbiology Reviews*, 75 :19–44, 1990.