INRA, Paris, 2015

Estimation of model parameters

David Makowski
INRA

Parameters

$$f\left(x,\theta\right)$$

Equations | Input variables | **Parameters**

« **A parameter** is a numerical value which is not calculated by the model and is not measured »

Parameter estimation

« aims at approximating the parameter values by using *experimental data* and/or *expert knowledge* »

It is important because

« *Model performances* depend on the accuracy of the parameter estimates »

The Bayesians and the Frequentists

For Frequentists,

- parameters are fixed
- parameters are estimated by points values for a given dataset
- estimation is performed by using data only

For Bayesians,

- parameters are defined as random variables
- parameters are estimated by distributions for a given dataset
- estimation is performed from both data and prior information
- computations more complex, but results more intuitive

Four steps for estimating parameters

1. How many and which parameters should be estimated?

→ In simple models, all parameters

→ In complex models, a subset of parameters is estimated

2. What kind of information is available?

→ Data

→ Prior information (expert knowledge, litterature)

3. Which estimation method?

→ Ordinary least squares

→ Weighted/Generalized least squares, maximum likelihood

→ Bayesian method

4. What is the accuracy of the parameter estimator?

→ Theoretical consideration, variances, residuals

Four estimation problems

Pb.A: One parameter

Pb.B: Linear model with 2 parameters

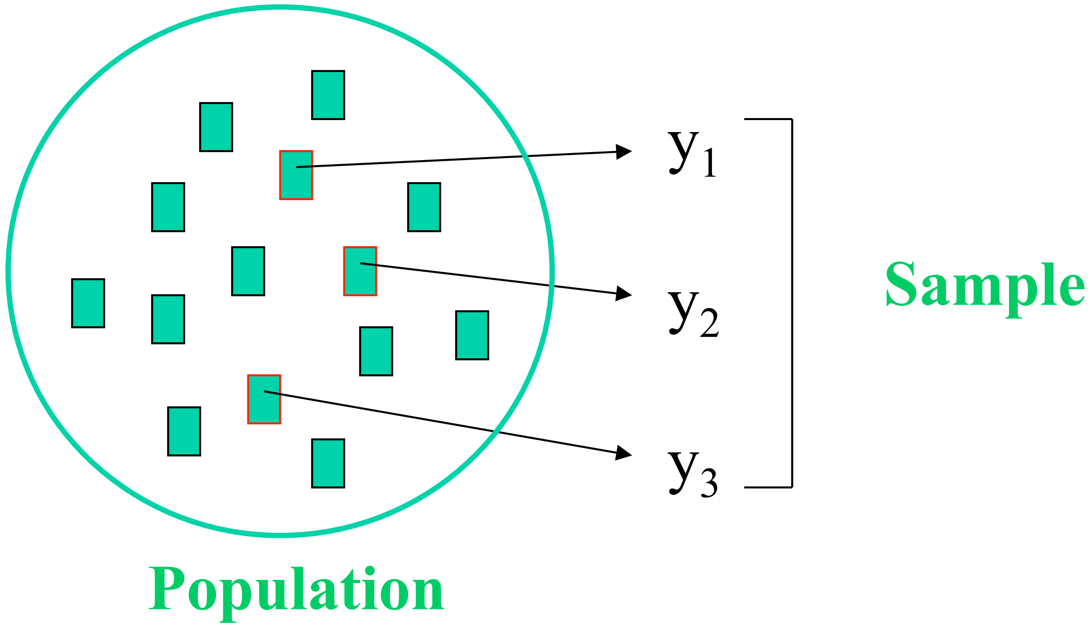
Pb.C: Non linear model with 4-6 parameters

Pb.D: Non linear model with 18 parameters

Problem A

« Estimation of the average oilseed rape yield in 2004 in a small area

from 3 yield measurements collected in three different plots »



Step 1. Which parameters?

A single parameter, the average yield in the considered area, noted θ.

Step 2. What kind of information?

Available information: a ***sample*** of three measures collected in three plots from the ***population of plots*** of interest

Step 3. Which method?

An estimator of the average yield is:

$$\hat{\theta} = \frac{y_1 + y_2 + y_3}{3}$$

Example :

- If y_1=30, y_2=39 et y_3=35, the estimated average yield is **34.7** q/ha.
- If y_1=32, y_2=38 et y_3=39, the estimated average yield is **36.3** q/ha.

« An estimator is a function relating the parameter to the observations »

Data set 1 $\longrightarrow$ $\hat{\theta}_1$

Data set 2 $\longrightarrow$ $\hat{\theta}_2$

Data set N $\longrightarrow$ $\hat{\theta}_N$

Step 4. Is the estimator accurate?

$$E\left[\left(\hat{\theta}-\theta\right)^2\right] = \left[E\left(\hat{\theta}\right)-\theta\right]^2 + \operatorname{var}\left(\hat{\theta}\right)$$

Mean squared error | **Bias²** | **Variance**

Step 4. Is the estimator accurate?

a. Theoretical consideration

« Under some assumptions, our estimator is ***unbiased*** and of ***minimum variance*** among the unbiaised estimators »

Step 4. Is the estimator accurate?

b. Estimator variance

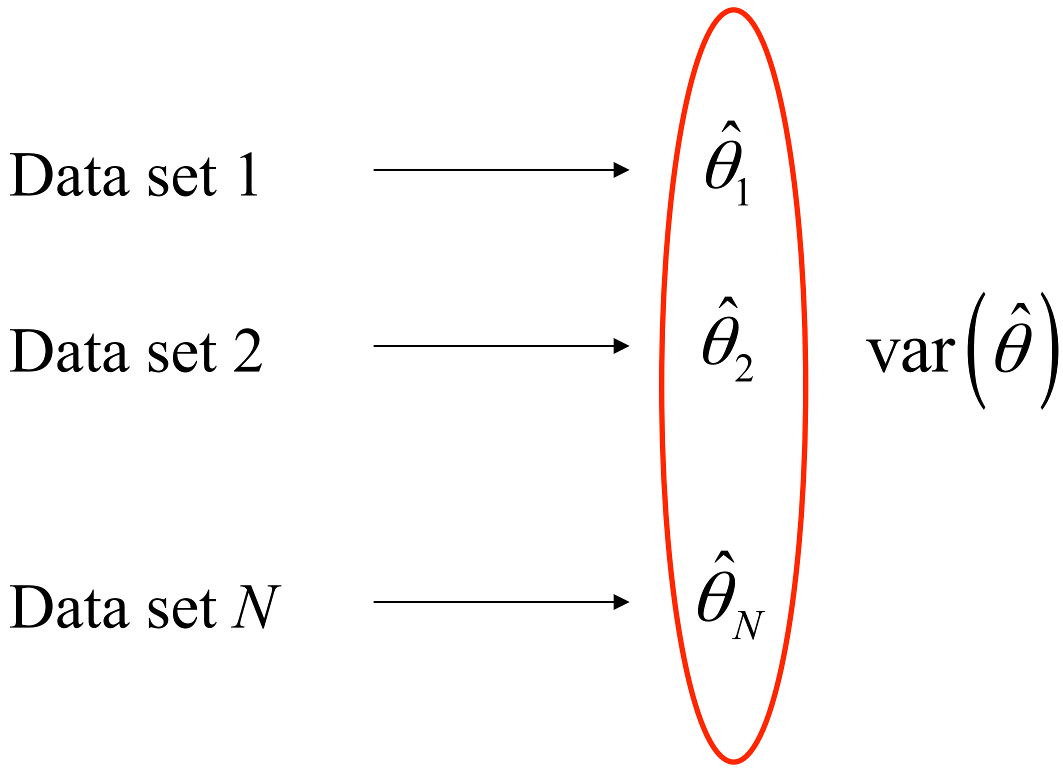
$\mathrm{var}\left(\hat{\theta}\right)$ can be estimated from data

Example :

- If y_1=30, y_2=39 and y_3=35, the estimated variance is 6.78 q²/ha², standard deviation=2.6 q/ha.

- If y_1=32, y_2=38 and y_3=39, the estimated variance is 4.78 q²/ha², standard deviation=2.19 q/ha.

The variance of an estimator measures its variability across datasets



Four estimation problems

Pb.A: One parameter

Pb.B: Linear model with 2 parameters

Pb.C: Non linear model with 4-6 parameters

Pb.D: Non linear model with 18 parameters

Problem B

« Estimation of the parameters of the model $f(x;\theta_1,\theta_2)$ »

$$f(x;\theta_1,\theta_2) = \theta_1 + \theta_2\, x$$

Nitrogen uptake in oilseed rape crop

Nitrogen fertilizer dose

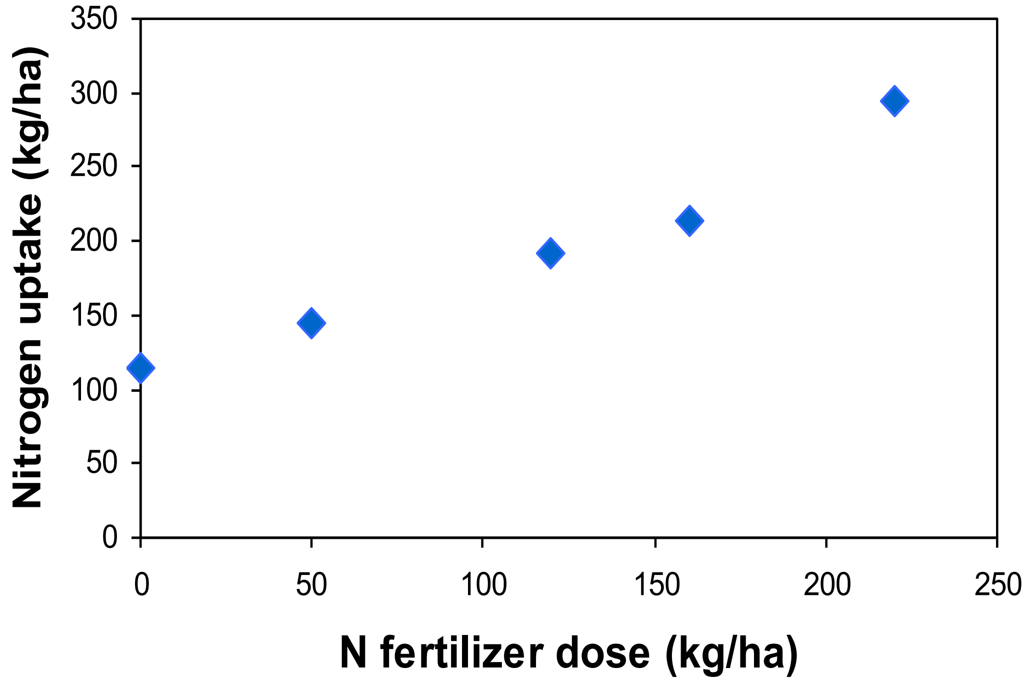
This model computes nitrogen uptake in function of fertilizer dose

Step 1. Which parameters?

Two model parameters: θ_1 and θ_2

Step 2. What kind of information?

A ***sample*** of 5 nitrogen uptake measurements obtained in 5 plots in the ***population of interest*** (an area in France)



Step 3. Which method?

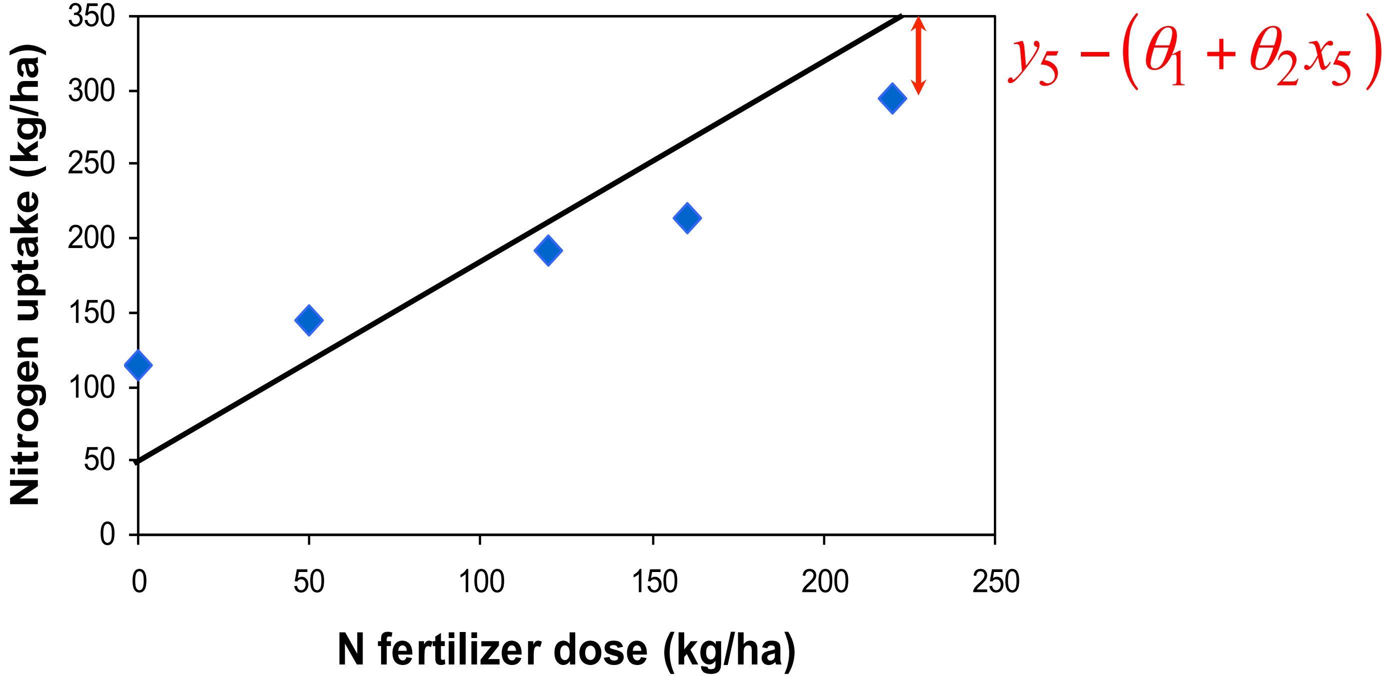
Ordinary least squares

The parameter estimators are the values of θ_1 and θ_2 minimizing

$$\sum_{i=1}^{N}\left(y_i - \theta_1 - \theta_2 x_i\right)^2$$

that is

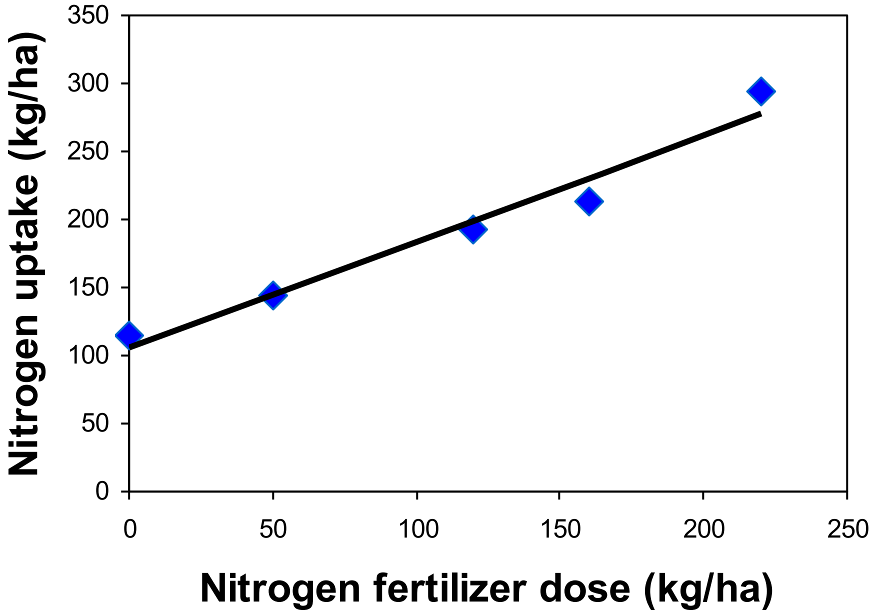
$$\hat{\theta}_2 = \frac{\sum_{i=1}^{N}\left(y_i - \bar{Y}.\right)\left(x_i - \bar{X}.\right)}{\sum_{i=1}^{N}\left(x_i - \bar{X}.\right)^2} \qquad \hat{\theta}_1 = \bar{Y}. - \hat{\theta}_2\,\bar{X}.$$



Nitrogen uptake (kg/ha)
N fertilizer dose (kg/ha)
0
50
100
150
200
250
300
350
0
50
100
150
200
250
$y_5 - (\theta_1 + \theta_2 x_5)$

Here, with our 5 measurements, we got $\hat{\theta}_1 = 106.01 \text{ kg.ha}^{-1}$

$$\hat{\theta}_2 = 0.78 \text{ kg.kg}^{-1}$$



Step 4. Are these estimators accurate?

$$E\left[\left(\hat{\theta}-\theta\right)^2\right]=\left[E\left(\hat{\theta}\right)-\theta\right]^2+\operatorname{var}\left(\hat{\theta}\right)$$

Mean squared error **Bias²** **Variance**

Step 4. Are these estimators accurate?

a. Theoretical aspect

« Under some assumptions,these estimators are ***unbiased*** and with ***minimum variances*** among the unbiased estimators ».

Assumptions are:

- ***independance*** of the model errors,
- ***homogeneity*** of the model error variances.

Step 4. Are these estimators accurate?

b. Variances of the estimators

Estimation of $\mathrm{var}\left(\hat{\theta}\right)$ from the data

$$\sqrt{\mathrm{var}\left(\hat{\theta}_1\right)} = 11.99 \ \ \mathrm{kg.ha^{-1}}$$

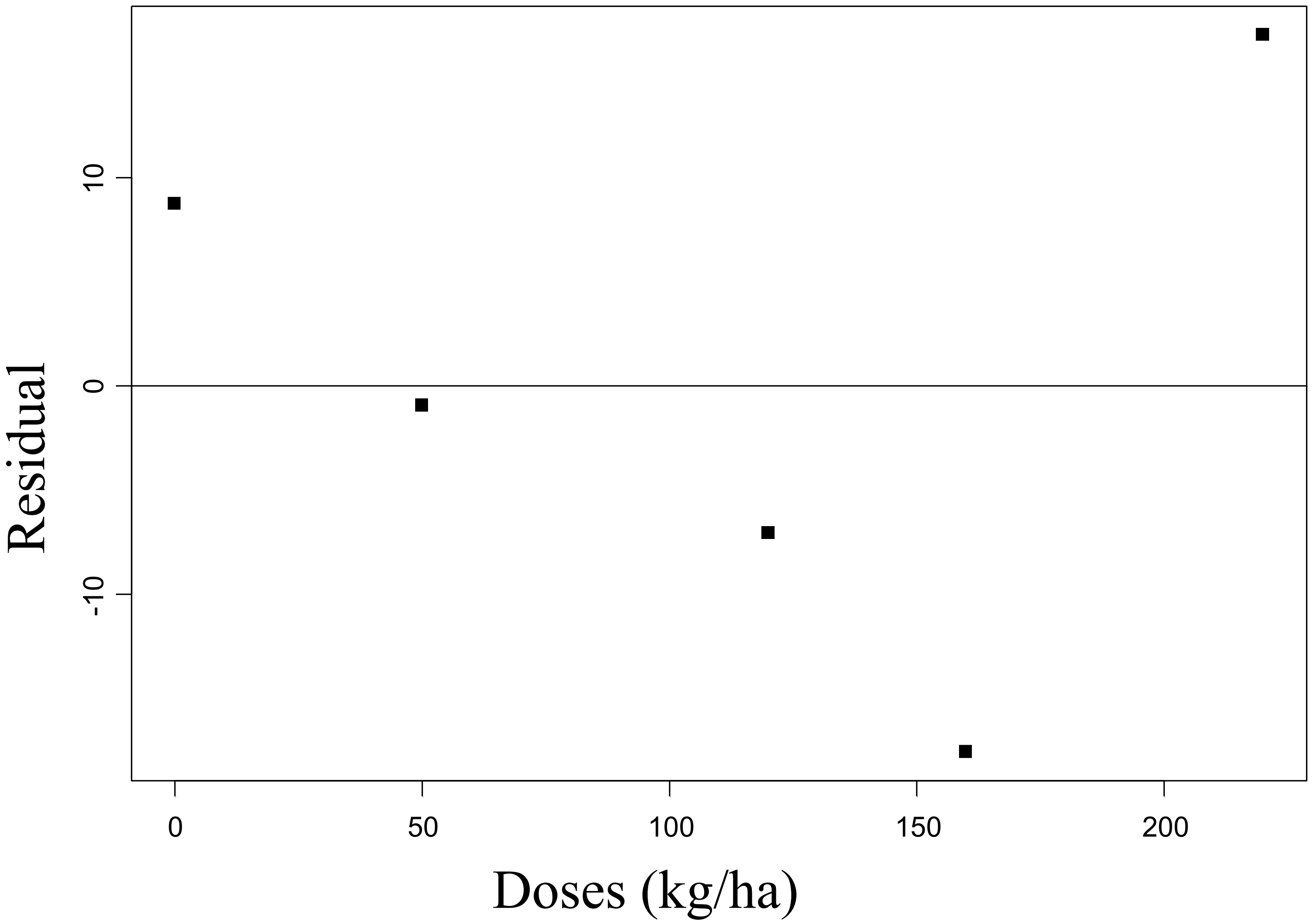
$$\sqrt{\mathrm{var}\left(\hat{\theta}_2\right)} = 0.09 \ \ \mathrm{kg.kg^{-1}}$$

Step 4. Are these estimators accurate?

c. Analysis of the residuals

$$r_i = y_i - \left(\hat{\theta}_1 + \hat{\theta}_2 x_i\right), \quad i = 1,\ldots,5$$

Useful to check the independance of the model errors and variance homogeneity



Residual
10
0
-10
0
50
100
150
200
Doses (kg/ha)

R code

```
DOSE<-c(0,50,120,160,220)
NABS<-c(114.75,144.0,192.38,213,294.16)
DATA<-data.frame(DOSE,NABS)

Fit<-lm(NABS~DOSE,data=DATA)

print(summary(Fit))
plot(DOSE,Fit$residuals,ylab="Residual",ylab="Dose",pch=15)
abline(0,0)
```

Comments on the first two problems

Four steps

1. Which parameters?
2. What kind of information?
3. Which estimation method?
4. Accuracy of the parameter estimators?

Comments on the first two problems

It was easy because

- Linear model: $\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_p x_p$

  → Analytic relationships between estimators and data

- Number of data > Number of parameters

- Only one type of measurements

- No prior information about parameter values

- Sotwares are available for the computations (SAS, R, ModelMaker...).

It can be much more difficult

- Non linear models: $\neq \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_p x_p$

  → No analytic relationship between the estimators and the data

- The number of data can be low compared to the number of parameters

- Complex dataset

  → several types of observations, correlated observations

- Prior information about parameter values

Four estimation problems

Pb.A: One parameter

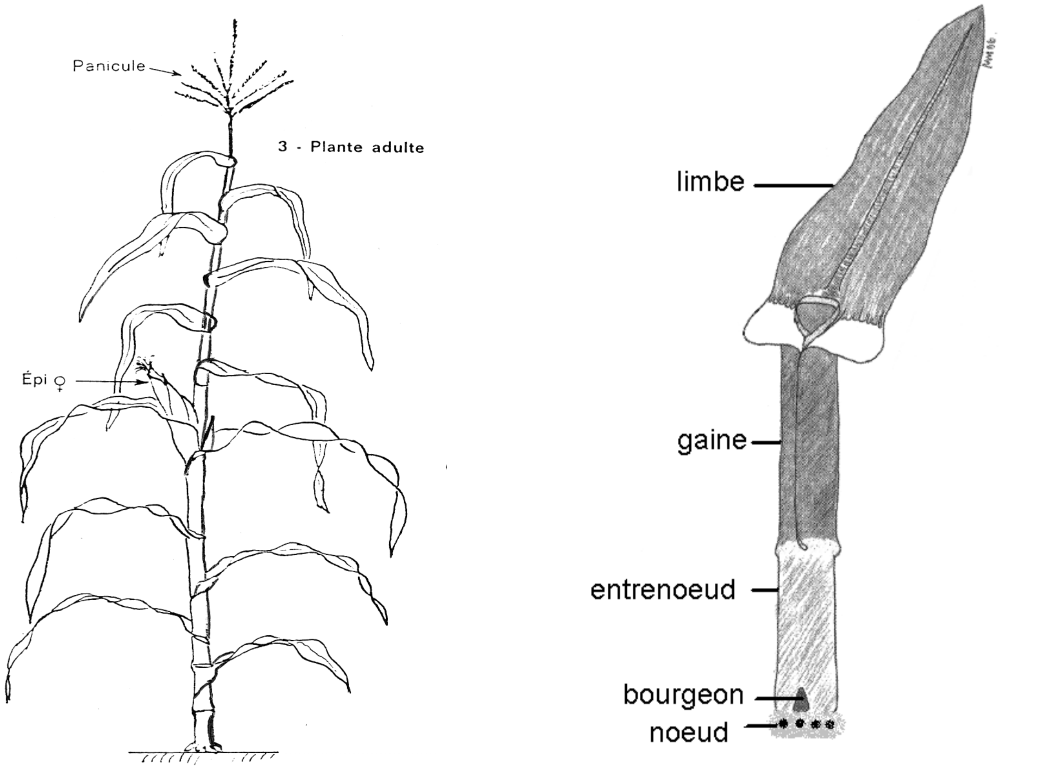
Pb.B: Linear model with 2 parameters

Pb.C: Non linear model with 4-6 parameters

Pb.D: Non linear model with 18 parameters

Problem C

Estimation of the parameters of a model simulating maize leaf length in function of temperature sum



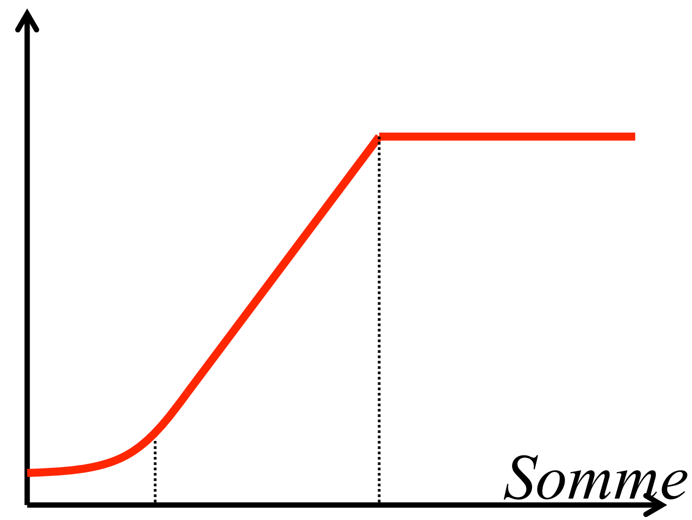
Représentation d'un plan de maïs et d'un phytomère
(adapté d'après Scanlon et Freeling 1997).

Le modèle (pour un phytomère)

$$f(x;\theta) = L_{MIN} e^{R_1(x - T_0)} \qquad T_0 < x \leq T_1$$

$$f(x;\theta) = \alpha + \beta(x - T_1) \qquad T_1 < x \leq T_2$$

$$f(x;\theta) = L_{MAX} \qquad T_2 < x$$



Quels paramètres doit-on estimer ?

Quels paramètres doit-on estimer ?

• *Contraintes pour assurer la continuité de la fonction:*

$$\alpha = L_{MIN} e^{R_1 (T_1 - T_0)} \qquad L_{MAX} = \alpha + \beta (T_2 - T_1)$$

• *Contrainte pour assurer la continuité de la dérivé entre la première et la deuxième phase:*

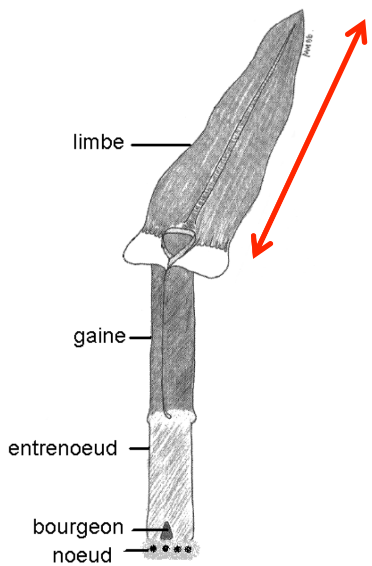
$$\beta = L_{MIN} R_1 e^{R_1 (T_1 - T_0)}$$

• *On peut fixer L_{MIN} à une valeur initiale.*

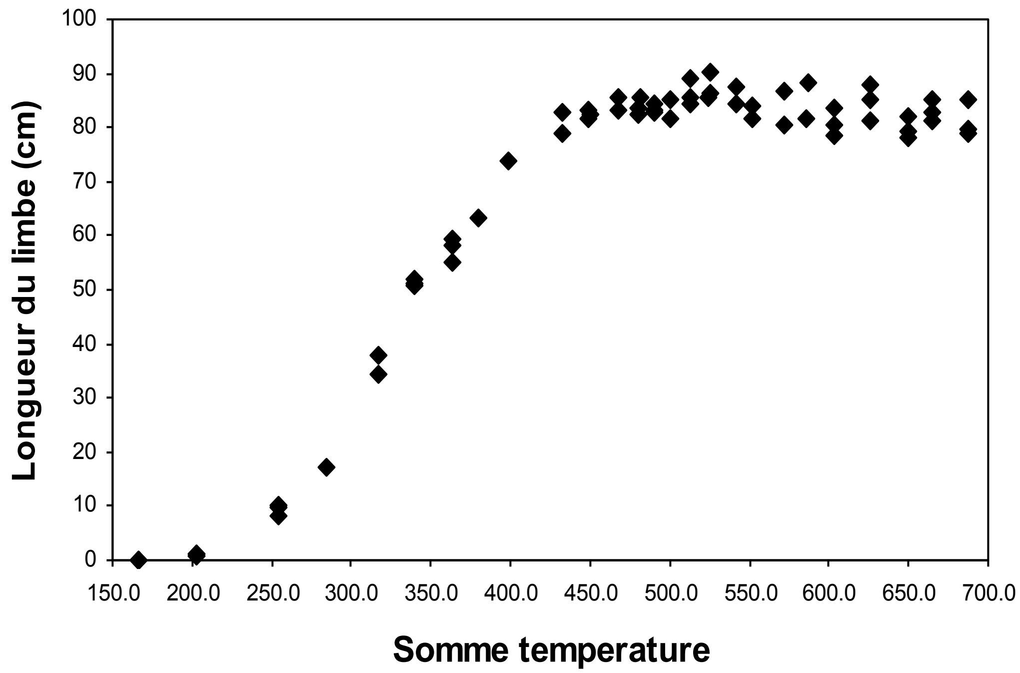
→ **Quatre paramètres à estimer : R_1, T_0, T_1, T_2 .**

Quelle information utiliser ?

- Mesures de longueurs de limbe obtenues sur une parcelle pour le 9ème phytomère.
- Mesures réalisées 2 à 3 fois par semaine.
- Une à trois mesures obtenues pour chaque date sur différentes plantes.



Phytomère 9



Quelle méthode d'estimation utiliser ?

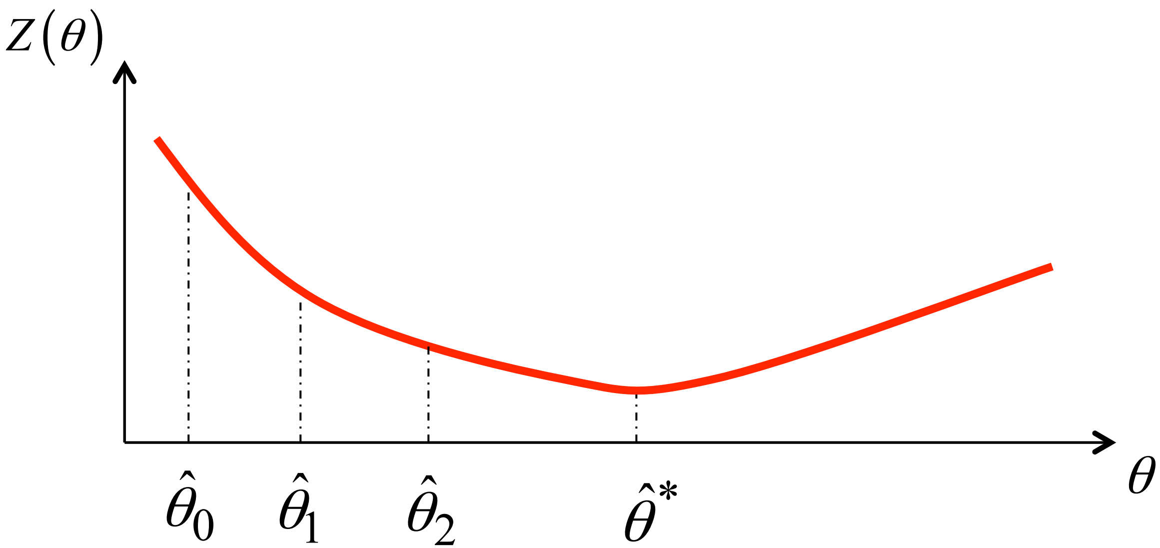
1er possibilité : La méthode des moindres carrés ordinaires

Trouver la valeur de θ qui minimise : $$Z(\theta) = \sum_{i=1}^{N} \left[y_i - f\left(x_i;\theta\right) \right]^2$$

Problème :

- le modèle est non linéaire,
- on ne peut pas trouver l'expression analytique des estimateurs.

Appliquer la méthode des MCO avec un algorithme itératif

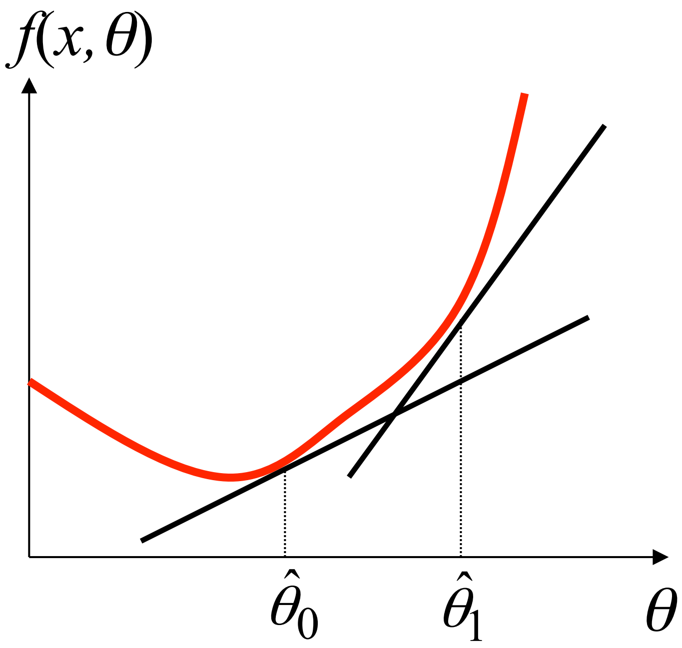


L'algorithme de Gauss-Newton

1. On définit une valeur initiale $\hat{\theta}_0$

2. On linéarise le modèle par un dév. de Taylor

$$f(x;\theta) \approx f\left(x;\hat{\theta}_0\right) + \sum_{j=1}^{p} \left.\frac{\partial f(x;\theta)}{\partial \theta_j}\right|_{\hat{\theta}_0} \left(\theta_j - \hat{\theta}_{0j}\right)$$



3. On calcule l'estimateur des moindres carrés avec le modèle linéarisé → $\hat{\theta}_1$

4. Retour à l'étape 1 en remplaçant $\hat{\theta}_0$ par $\hat{\theta}_1$.

Arrêt si $\sum_{i=1}^{N}\left[y_i - f\left(x_i;\hat{\theta}_{k+1}\right)\right]^2 - \sum_{i=1}^{N}\left[y_i - f\left(x_i;\hat{\theta}_k\right)\right]^2$ est négligeable.

L'algorithme de Gauss-Newton

Questions

$$f\left(x;\theta\right) = e^{\theta x}$$

- Linéariser le modèle à l'aide d'un développement de Taylor

$$f\left(x;\theta\right) \approx f\left(x;\hat{\theta}_0\right) + \left.\frac{df\left(x;\theta\right)}{d\theta}\right|_{\hat{\theta}_0} \left(\theta - \hat{\theta}_0\right)$$

- Exprimer le modèle sous la forme : $A + B\ \theta$

L'algorithme de Gauss-Newton

$$f\left(x;\theta\right)=e^{\theta x}$$

$$e^{\theta x}\approx e^{\hat{\theta}_0 x}+\left(\theta-\hat{\theta}_0\right)xe^{\hat{\theta}_0 x}$$

$$e^{\theta x}\approx e^{\hat{\theta}_0 x}-\hat{\theta}_0 xe^{\hat{\theta}_0 x}+xe^{\hat{\theta}_0 x}\times\theta$$

Estimateur: $\hat{\theta}_1=\hat{\theta}_0+\dfrac{\sum_{i=1}^{N}x_i e^{\hat{\theta}_0 x_i}\left(Y_i-e^{\hat{\theta}_0 x_i}\right)}{\sum_{i=1}^{N}x_i^2 e^{2\hat{\theta}_0 x_i}}$

L’algorithme de Gauss-Newton

Aspects pratiques

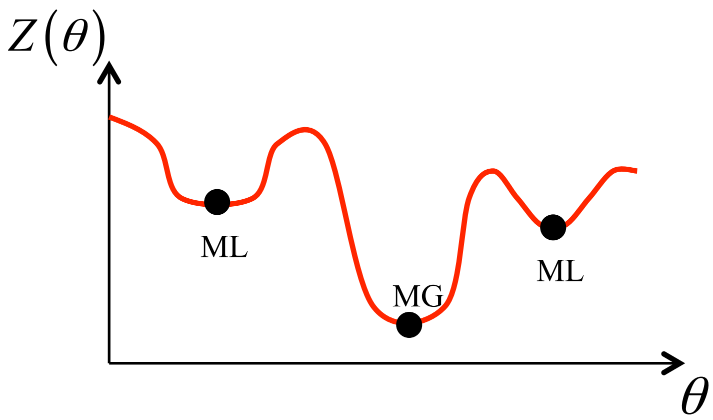
- On utilise un logiciel statistique (SAS, S+, R, MatLab…)

- On donne en entrée:

  - des données,
  - un modèle,
  - des valeurs initiales des paramètres.

- Le logiciel fournit en sortie les valeurs estimées des paramètres.

Minimum locaux et minimum globaux



→ Essayez plusieurs valeurs initiales !

Quelle méthode d'estimation utiliser ?

1er possibilité : La méthode des moindres carrés ordinaires

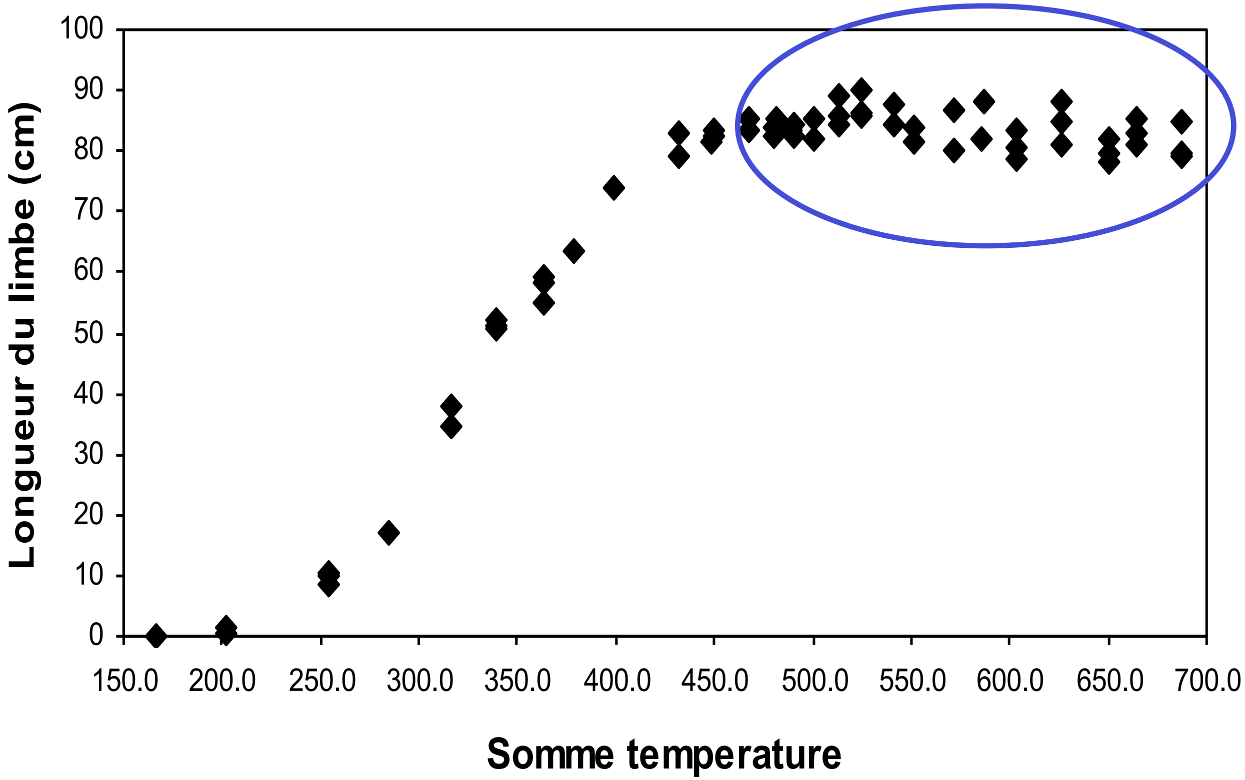
Trouver la valeur de θ qui minimise : $Z(\theta) = \sum_{i=1}^{N} \left[y_i - f(x_i;\theta) \right]^2$

Inconvénient :

- Les estimateurs ne sont pas de variances minimales si les résidus ont des variances hétérogènes.

- Or ici, les erreurs de mesures sont plus grandes pour les limbes de grandes tailles.

Variance plus grande ici.



Quelle méthode d'estimation utiliser ?

La méthode des moindres carrés pondérés

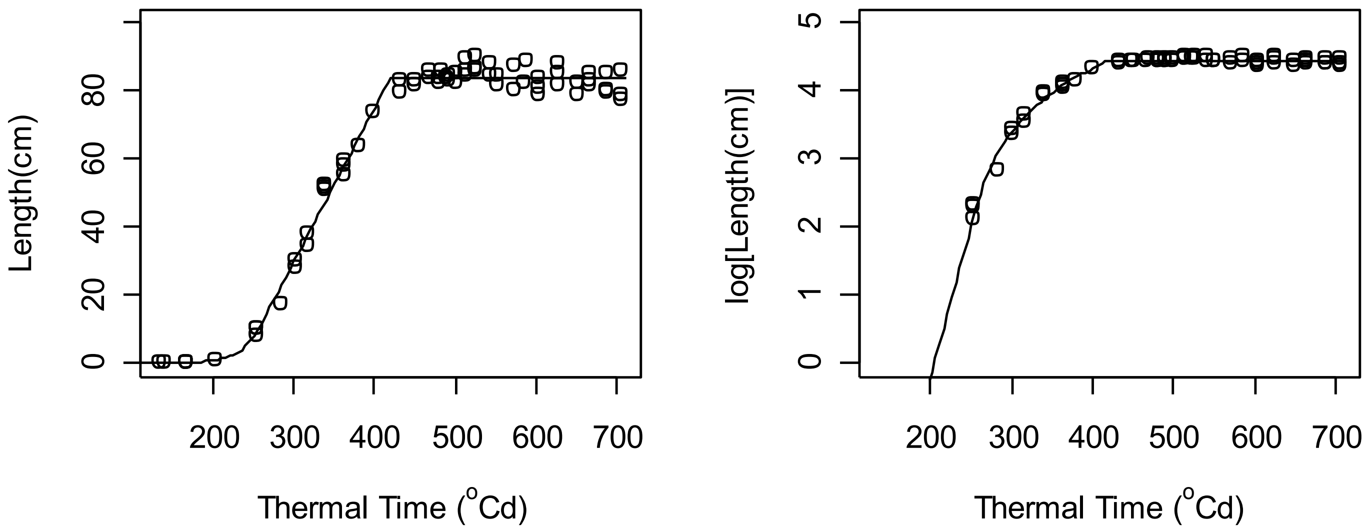
Trouver la valeur de θ qui minimise : $Z(\theta) = \sum_{i=1}^{N} \frac{\left[y_i - f(x_i;\theta)\right]^2}{\sigma_i^2}$

avec $\sigma_i^2 = \frac{1}{K-1} \sum_{k=1}^{K} \left(y_{ik} - \bar{y}_{i\cdot}\right)^2$

Définir la variance des résidus comme une fonction croissante de la longueur du limbe

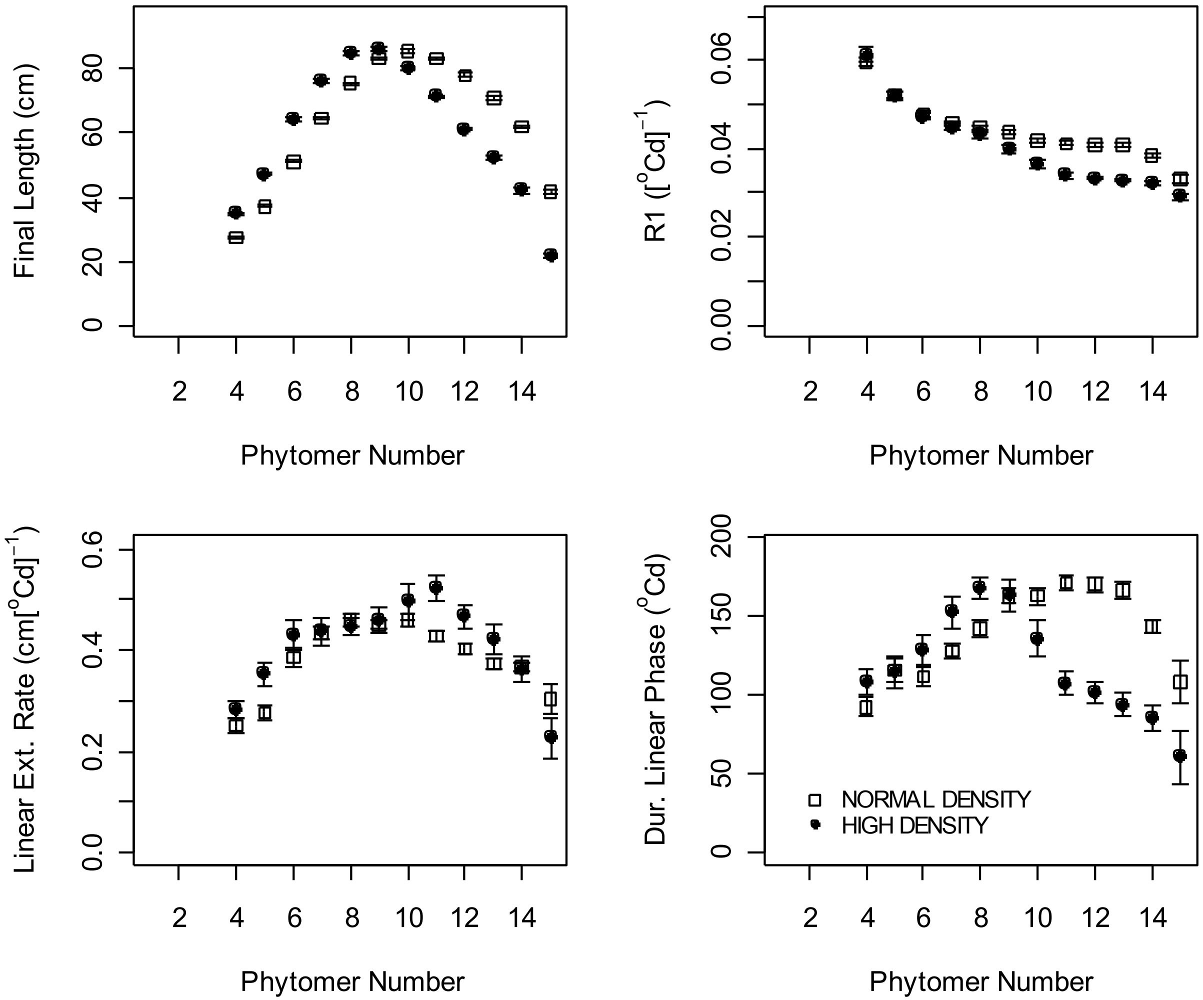
$$\mathrm{var}\left[y_i - f(x_i;\theta)\right] = \sigma^2 f(x_i;\theta)^\tau$$

On estime θ, σ et τ à partir des données.



Length(cm)
0
20
40
60
80
200
300
400
500
600
700
Thermal Time (°Cd)
log[Length(cm)]
0
1
2
3
4
5
200
300
400
500
600
700
Thermal Time (°Cd)

Ces estimateurs sont-ils précis ?

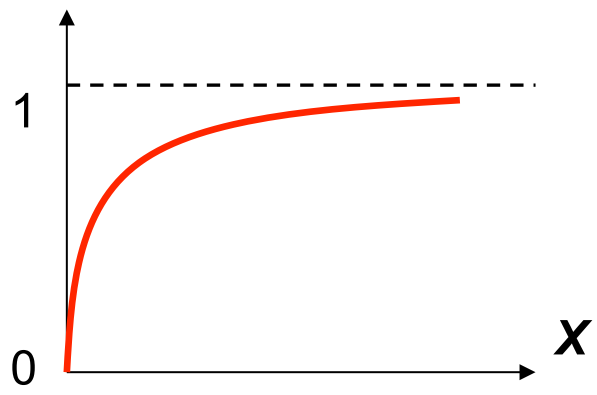


Exercise

Estimation of the parameter of a non linear model

- Non linear model predicting relative yield as a function of a factor *x* (amount of soil mineral N)

$$f(x,\theta) = \left[1 - \exp(-\theta \times x)\right]$$



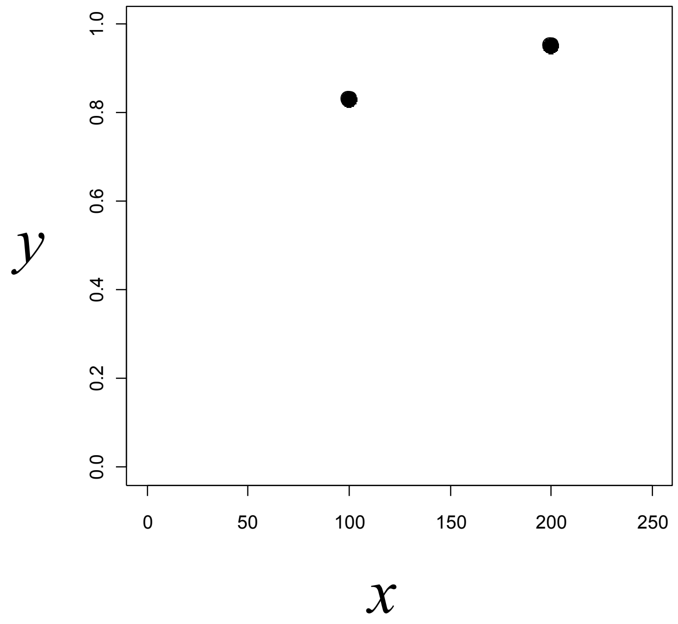
- One parameter θ: the growth rate

Data

Two measurements of relative yield y_1 and y_2 are available:

y_1 = 0.83 for x_1 = 100 kg/ha

y_2 = 0.95 for x_2 = 200 kg/ha



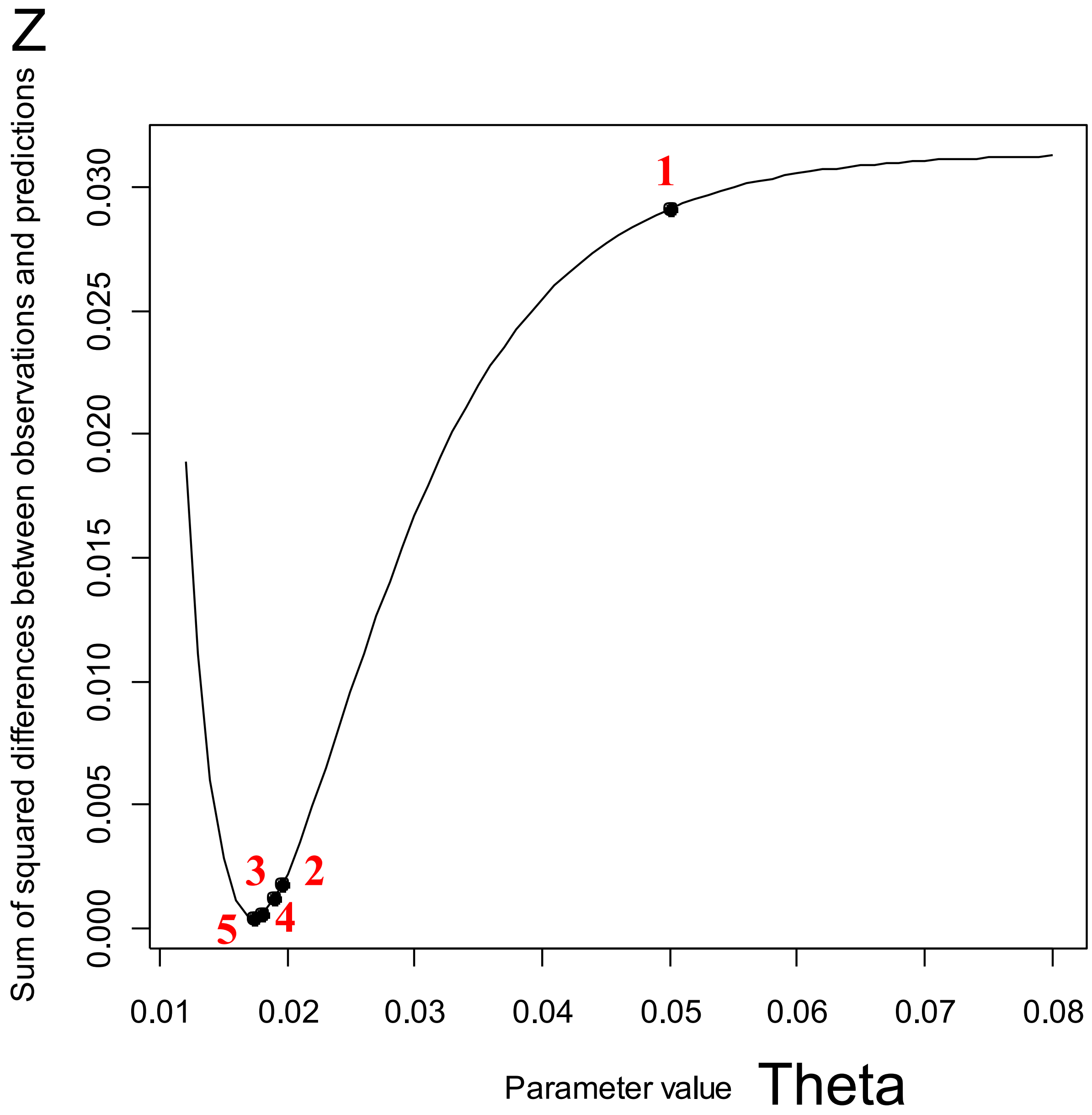
Question

- Estimate the parameter by ordinary least squares

```
x<-c(100, 200)

y<-c(0.83, 0.95)

TAB<-data.frame(x,y)
```

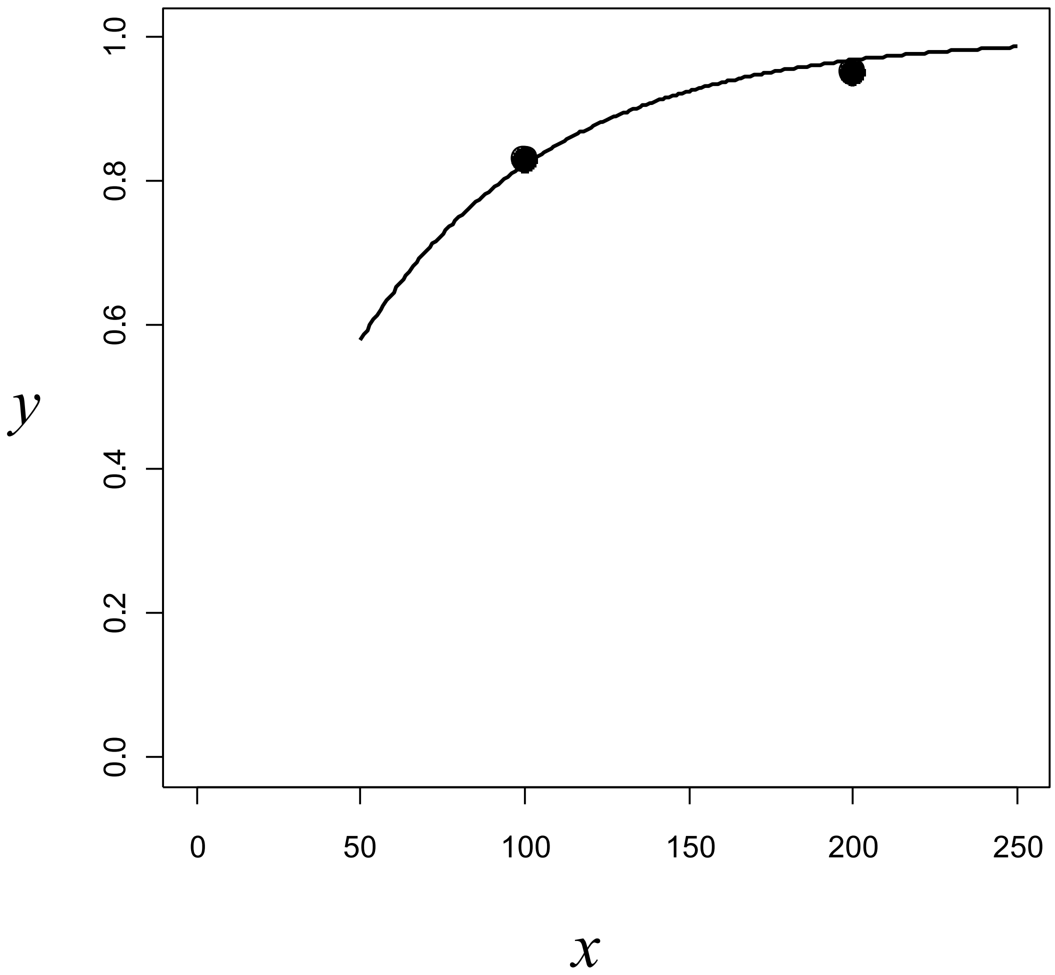
```
x<-c(100, 200)

y<-c(0.83, 0.95)

TAB<-data.frame(x,y)

Fit<-nls(y~1-exp(-Theta*x), data=TAB, start=list(Theta=0.05), trace=T)

print(summary(Fit))
```

```
> Fit<-nls(y~1-exp(-Theta*x), data=TAB, start=list(Theta=0.05), trace=T)
0.02914996 :  0.05
0.001751208 :  0.01959284
0.001160448 :  0.01897615
0.0005836208 :  0.01810939
0.0003974804 :  0.01732987
0.0003974661 :  0.01733614
0.0003974661 :  0.01733635

> print(summary(Fit))

Formula: y ~ 1 - exp(-Theta * x)

Parameters:
      Estimate Std. Error t value Pr(>|t|)
Theta 0.017336   0.001064   16.29    0.039 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01994 on 1 degrees of freedom

Number of iterations to convergence: 6
```


Z
Sum of squared differences between observations and predictions
0.000
0.005
0.010
0.015
0.020
0.025
0.030
1
2
3
4
5
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
Parameter value
Theta

```
X.vec<-50:250

Y.vec<-1-exp(-coef(Fit)[1]*X.vec)

plot(x,y, xlim=c(0, 250), pch=19, cex=2, ylim=c(0,1))

lines(X.vec, Y.vec, lwd=2)
```

$$\hat{\theta} = 0.0173$$



Four estimation problems

<u>**Pb.A:**</u> **One parameter**

<u>**Pb.B:**</u> **Linear model with 2 parameters**

<u>**Pb.C:**</u> **Non linear model with 4-6 parameters**

<u>**Pb.D:**</u> **Non linear model with 18 parameters**

Problem D

Estimation of the parameters of a model simulating winter wheat growth between January and May

(Jeuffroy et Recous, 1999)

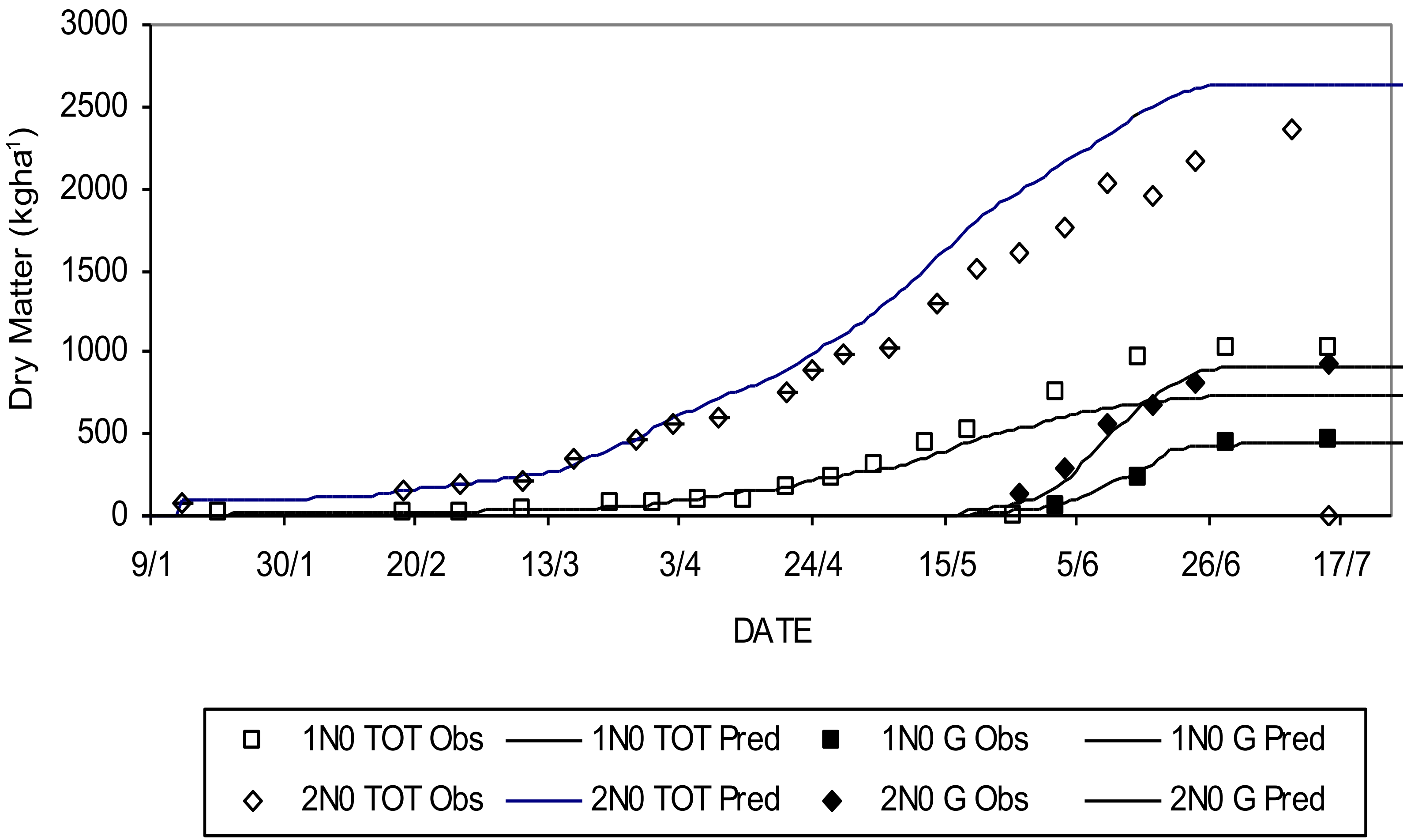
State variables simulated at a daily time step

- Crop above-ground biomass (dry matter) (kg/ha) → MS_t
- Nitrogen uptake (kg/ha), → QN_t
- Leaf Area Index → LAI_t

Input variables:

- Global daily radiation → RG_t
- Average daily temperature, → T_t
- Initial values of biomass and nitrogen uptake → MS_0, QN_0

Simulations of wheat biomass using the AZODYN dynamic crop model



Few model equations

$$MS_j = MS_{j-1} + \left(E_{b\max} \times ft_{j-1} \times Ei_{j-1} \times C \times RG_{j-1} \right)$$

$$Ei_{j-1} = E_{i\max} \left[1 - \exp\left(-K \times LAI_{j-1} \right) \right]$$

$$LAI_{j-1} = D \times QNc_{j-1}$$

$$MS_j = MS_{j-1} + \left\{ E_{b\max} \times C \times E_{i\max} \left[1 - \exp\left(-K \times D \times QNc_{j-1} \right) \right] \times ft_{j-1} \times RG_{j-1} \right\}$$

18 parameters

Parameter	Meaning	Initial value	Range
Ebmax	Efficiency of radiation conversion	3.3 g/MJ	1.8-4
K	Coefficient of radiation extinction	0.72	0.6-0.8
D	Ratio LAI / Critical nitrogen uptake	0.028	0.02-0.045
Vmax	Maximum rate of nitrogen uptake	0.5 kg/ha/dj	0.2-0.7
C	PAR/RG	0.48	
Tmin	Minimum temperature for photosynthesis	0 °C	
Topt	Optimum temperature for photosynthesis	15 °C	
Tmax	Maximum température for photosynthesis	40 °C	
Eimax	Efficiency of radiation interception	0.96	
Tep-flo	Time between two stages	150 dj	
E		1.55 t/ha	
F		4.4 %	
G		5.35 %	
H		-0.442	
L		2 t/ha	
M		6 %	
N		8.3 %	
P		-0.44	

The two expressions of a dynamic model

1: Dynamic system model

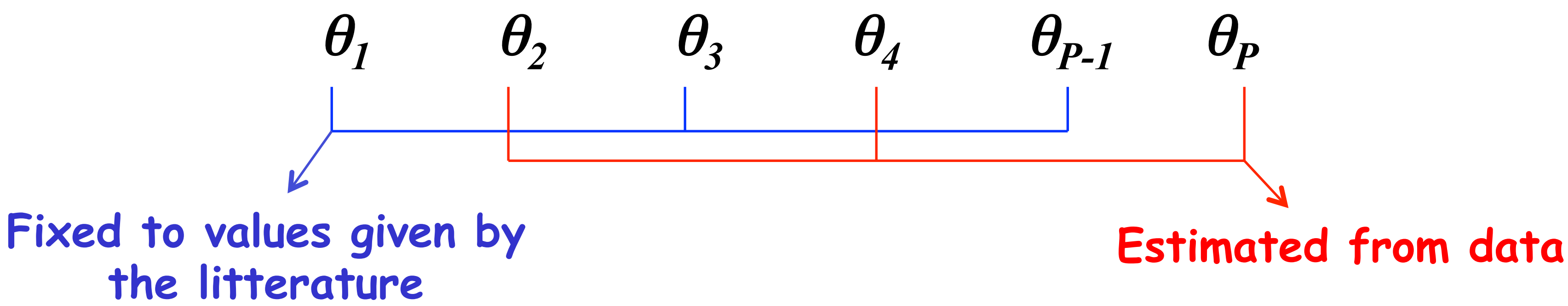
$$MS_t = MS_{t-1} + g\left(X_{t-1};\theta\right)$$

2: Response model

$$MS_t = f\left(t, X;\theta\right)$$

Step 1. Which parameters?

A subset of parameters must be selected



- **Numerical problems if all the parameters are estimated**

- **Not a good idea anyway**

  → High estimator variances

  → High prediction errors

New issue: How to select the subset of parameters

i. **from the literature**

ii. **by analyzing the model equations**

iii. **by sensitivity analysis**

iv. **from data**

i. from the literature

« Determine the parameters whose values are not well known ».

Drawbacks :

- can be quite subjective
- the available papers are not always relevant

ii. by analyzing the model equations

« Identify the parameters which cannot be simultaneously estimated »

$$MS_j = MS_{j-1} + \left(E_{b\max} \times ft_{j-1} \times Ei_{j-1} \times C \times RG_{j-1}\right)$$

$$Ei_{j-1} = E_{i\max}\left[1 - \exp\left(-K \times LAI_{j-1}\right)\right]$$

$$LAI_{j-1} = D \times QNc_{j-1}$$

$$MS_j = MS_{j-1} + \left\{E_{b\max} \times C \times E_{i\max}\left[1 - \exp\left(-K \times D \times QNc_{j-1}\right)\right] \times ft_{j-1} \times RG_{j-1}\right\}$$

Case 1: only observed values of MS are available

Case 2: observed values of MS and LAI are available

$$MS_j = MS_{j-1} + \left(E_{b\max} \times ft_{j-1} \times Ei_{j-1} \times C \times RG_{j-1}\right)$$

$$Ei_{j-1} = E_{i\max}\left[1 - \exp\left(-K \times LAI_{j-1}\right)\right]$$

$$LAI_{j-1} = D \times QNc_{j-1}$$

$$MS_j = MS_{j-1} + \left\{E_{b\max} \times C \times E_{i\max}\left[1 - \exp\left(-K \times D \times QNc_{j-1}\right)\right] \times ft_{j-1} \times RG_{j-1}\right\}$$

Case 1: only observed values of MS are available

It is not possible to estimate simultaneously
- the 3 parameters $E_{b\max}, C, E_{i\max}$
- the 2 parameters K, D

Case 2: observed values of MS and LAI are available

It is not possible to estimate simultaneously the 3 parameters $E_{b\max}, C, E_{i\max}$

iii. by sensitivity analysis

« Select the parameters that strongly influence the model outputs »

Drawbacks:

A sensitivity threshold must be defined.

Does not prevent from lack of identifiability .

iv. from data

« Select the parameters leading to the best model predictions »

Number of estimated parameters	MSEPcv
1	$MSEP_1$
2	$MSEP_2$
3	$MSEP_3$
…	…
P	$MSEP_P$

Step 1. Which parameters?

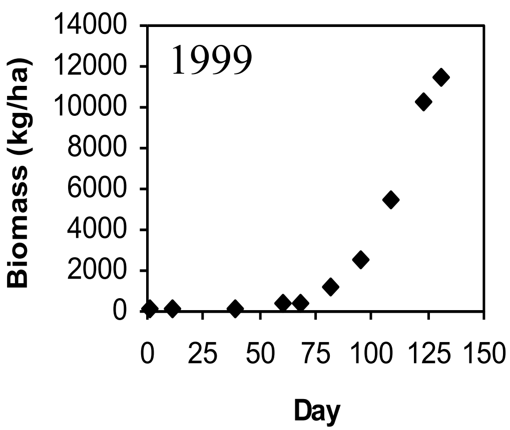
- **13 parameters** were fixed to values provided by the literature.
- **One parameter** was fixed after an analysis of the model equation.
- **Four parameters** were estimated from data: E_{BMAX}, D, K and V_{MAX}

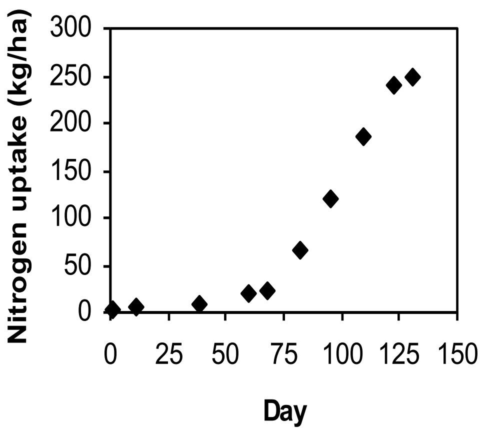
Step 2. What kind of information?

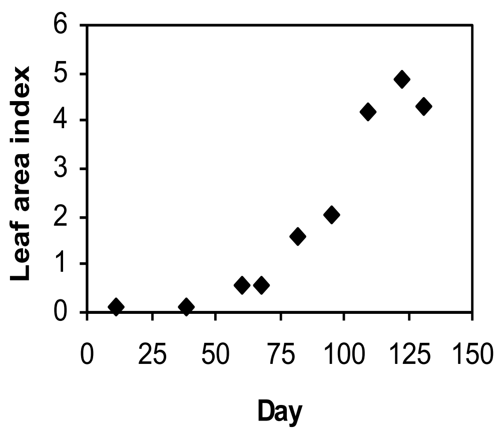
• Measurements of wheat **biomass**, of **leaf area index** and of **nitrogen uptake** for one site (Grignon) and 6 years.

• Ten dates of measurement each year.

• Three replicates at each date. Replicates were averaged.







Step 3. Which estimation method?

1st option: Ordinary least squares

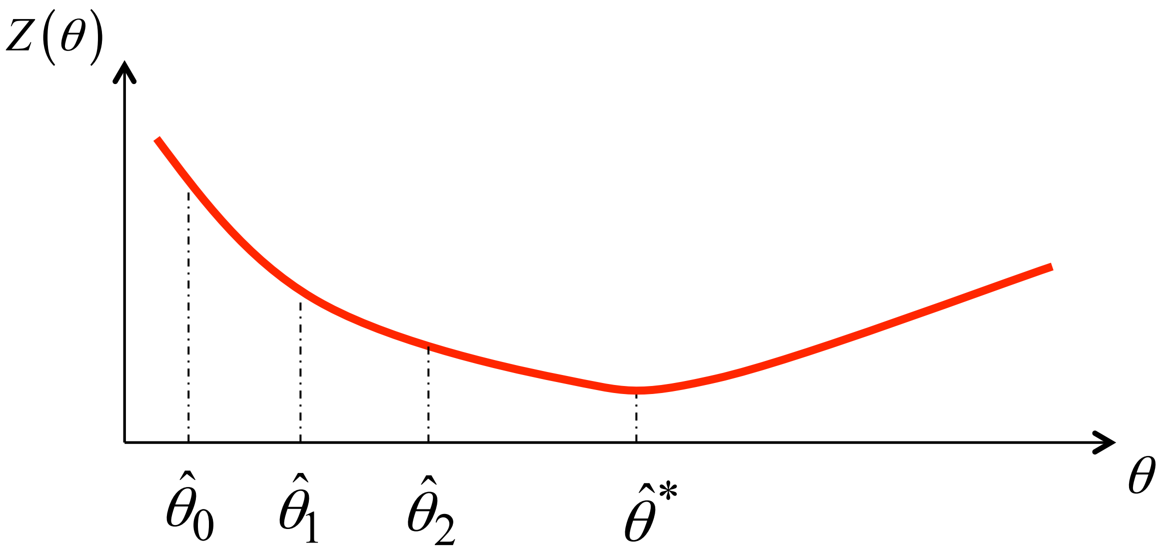
Find θ minimizing: $$Z(\theta) = \sum_{i=1}^{N} \left[y_i - f(t_i, x_i; \theta) \right]^2$$

Difficulties:

- non linear model

- no analytical expression for the estimators

Minimization using an iterative algorithm



Step 3. Which estimation method?

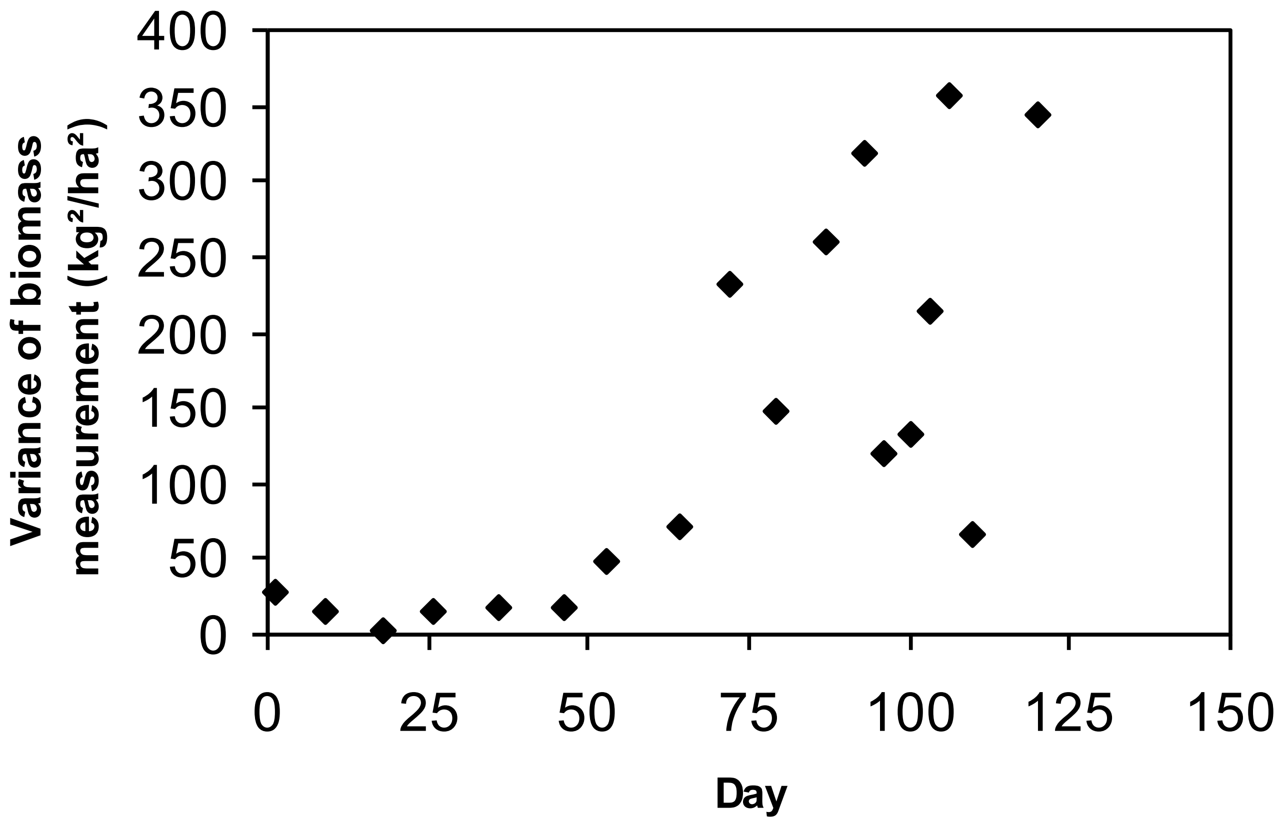
1st option: Ordinary least squares

Find θ minimizing: $$Z(\theta) = \sum_{i=1}^{N} \left[y_i - f\left(t_i, x_i; \theta\right) \right]^2$$

This method performs well when the model error variances are constant

This is not very realistic here.

Here, the variances of the observations are variable



Step 3. Which method?

2nd option: Weighted least squares

Find the value of θ minimizing:

$$Z(\theta) = \sum_{i=1}^{N} \frac{\left[y_i - f(t_i, x_i; \theta)\right]^2}{\sigma_i^2}$$

$$\text{with } \hat{\sigma}_i^2 = \frac{1}{K(K-1)} \sum_{k=1}^{K} (y_{ik} - y_i)^2$$

Weighted least squares

minimizing

$$Z_{MCP}(\theta) = \sum_{i=1}^{6}\sum_{j=1}^{10}\frac{\left[y_{ij}^{MS} - f^{MS}\left(t_j, x_i; \theta\right)\right]^2}{\hat{\sigma}_{MS.ij}^2} + \sum_{i=1}^{6}\sum_{j=1}^{10}\frac{\left[y_{ij}^{N} - f^{N}\left(t_j, x_i; \theta\right)\right]^2}{\hat{\sigma}_{N.ij}^2} + \sum_{i=1}^{6}\sum_{j=1}^{10}\frac{\left[y_{ij}^{L} - f^{L}\left(t_j, x_i; \theta\right)\right]^2}{\hat{\sigma}_{L.ij}^2}$$

$$\hat{\sigma}_{MS.ij}^2 = \frac{1}{K\left(K-1\right)}\sum_{k=1}^{K}\left[y_{ijk}^{MS} - y_{ij}^{MS}\right]^2$$

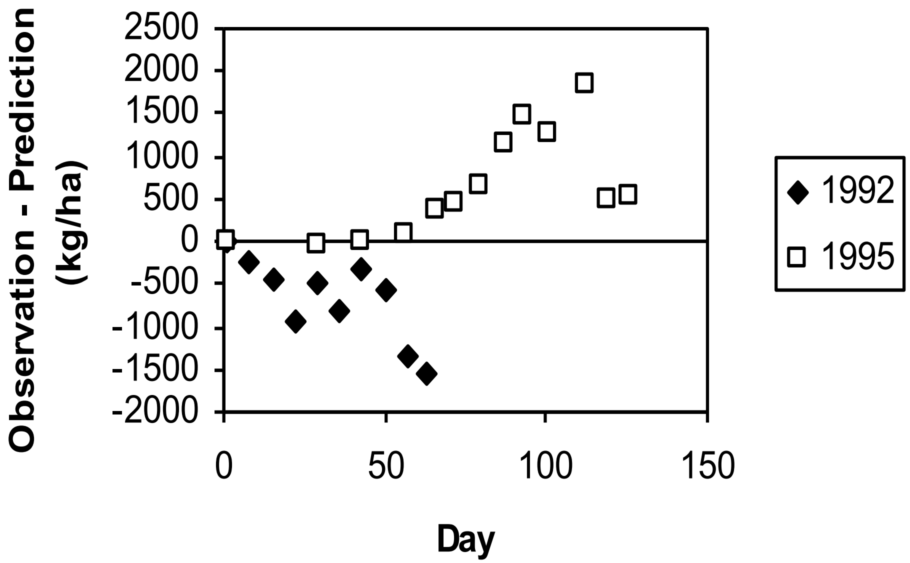
Results of step 3: weighted least square estimates of the 4 model parameters

Parameter	Initial value	Estimated value
E_{BMAX} (g/MJ)	3.3	3.29 (0.11)
D	0.028	0.037 (0.06)
K	0.72	0.74 (0.001)
V_{MAX} (kg/ha/dj)	0.5	0.38 (0.02)

Step 4. Are these estimators accurate?

Model residuals

Residuals are not independant →



Methods for taking correlations into account

- Generalized least squares
- Mixed models

Conclusion

On procède en plusieurs étapes

1. Quels paramètres estimer ?

→ Dans les cas simples, on peut tout estimer.

→ Dans les cas complexes, il faut faire une sélection.

2. Quelle information disponible ?

→ Les données

→ Information a priori

3. Quelle méthode d'estimation ?

→ Moindres carrés ordinaires,

→ Moindres carrés pondérés/généralisés,

→ Méthodes Bayésiennes…

4. Quelle est la précision des estimateurs ?

→ Aspects théoriques, variances, résidus.

Quelques références

Agresti A. 1990. *Categorical data analysis.* Wiley

Makowski D, Monod H. 2011. *Analyse statistique des risques agro-environnementaux*. Springer

McCulloch C.E., Searle S.R. 2001. *Generalized, linear, and mixed models*. Wiley

Seber GAF, Wild CJ. 2003. *Nonlinear regression*. Wiley.

Wallach D, Makowski D, Jones JW. 2006. *Working with dynamic crop models*. Elsevier.

Collection
Statistique
et probabilités
appliquées

David Makowski, Hervé Monod

Analyse statistique des risques agro-environnementaux

Études de cas

Cette collection met à la disposition du public intéressé par la statistique (étudiants, enseignants, chercheurs) des ouvrages qui concilient effort pédagogique et travail permanent de mise à jour. Cette démarche implique de prendre en compte de façon sélective et critique les renouvellements des concepts, des champs d'application et des outils de traitement. Seules une compréhension profonde et une appropriation des connaissances permettront de s'adapter aux évolutions qui n'ont pas fini de bouleverser cette discipline.

Conçu comme un véritable manuel pratique, ce livre est une introduction aux méthodes statistiques les plus couramment utilisées pour l'analyse des risques agro-environnementaux. Celles-ci peuvent être regroupées au sein de trois grandes sections :

- La modélisation des risques en fonction de facteurs environnementaux et anthropiques (modèle linéaire, modèle linéaire généralisé, modèle non linéaire, modèle hiérarchique, régression quantile) ;
- L'optimisation de décisions ou de règles de décision pour mieux gérer les risques, en intégrant des variables décisionnelles dans les modèles (optimisation de seuils de décision, optimisation par simulation, analyses ROC) ;
- L'analyse et la communication des incertitudes associées aux modèles (estimation et description de distributions de probabilité, assimilation de données, analyse de sensibilité).

L'utilisation de chaque méthode est illustrée par une ou plusieurs applications à des problèmes concrets (pollution de l'eau par les nitrates, invasion par des espèces nuisibles, flux de gènes d'une culture OGM vers une culture non OGM, etc.). Les programmes informatiques R ou WinBUGS utilisés dans les exemples sont présentés et commentés en détail. A la fin de chaque chapitre, des exercices permettront aux lecteurs de tester leur compréhension des méthodes étudiées.

00 € TTC

ISBN : 978-2-8178-0250-3

9 782817 802503

springer.com

Makowski • Monod

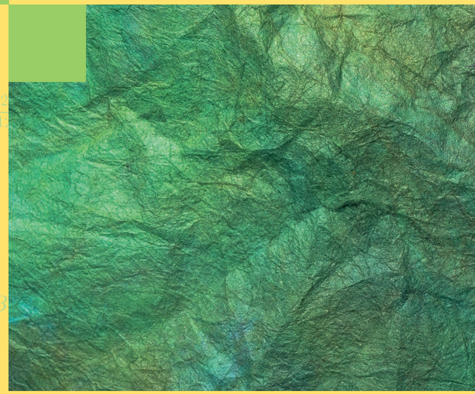
Analyse statistique des risques agro-environnementaux

Collection
Statistique
et probabilités
appliquées

David Makowski, Hervé Monod

Analyse statistique des risques agro-environnementaux

Études de cas



$\phi(n) = \left(1 - \frac{1}{12}\right)\left(1 - \frac{2}{12}\right)$

$P(A) = \frac{25}{216}$ et $P(B$

$(\forall B \in \beta_{\mathbb{R}})\ P_X(B) = P(X^{-1}(B))$
$= P(\{\omega \in \Omega \mid X(\omega) \in B\})$

Springer

$P(A_2) = 1 - P(\overline{A_2}) = 1 - \left(\frac{35}{36}\right)^{24} \approx 0,491.$