

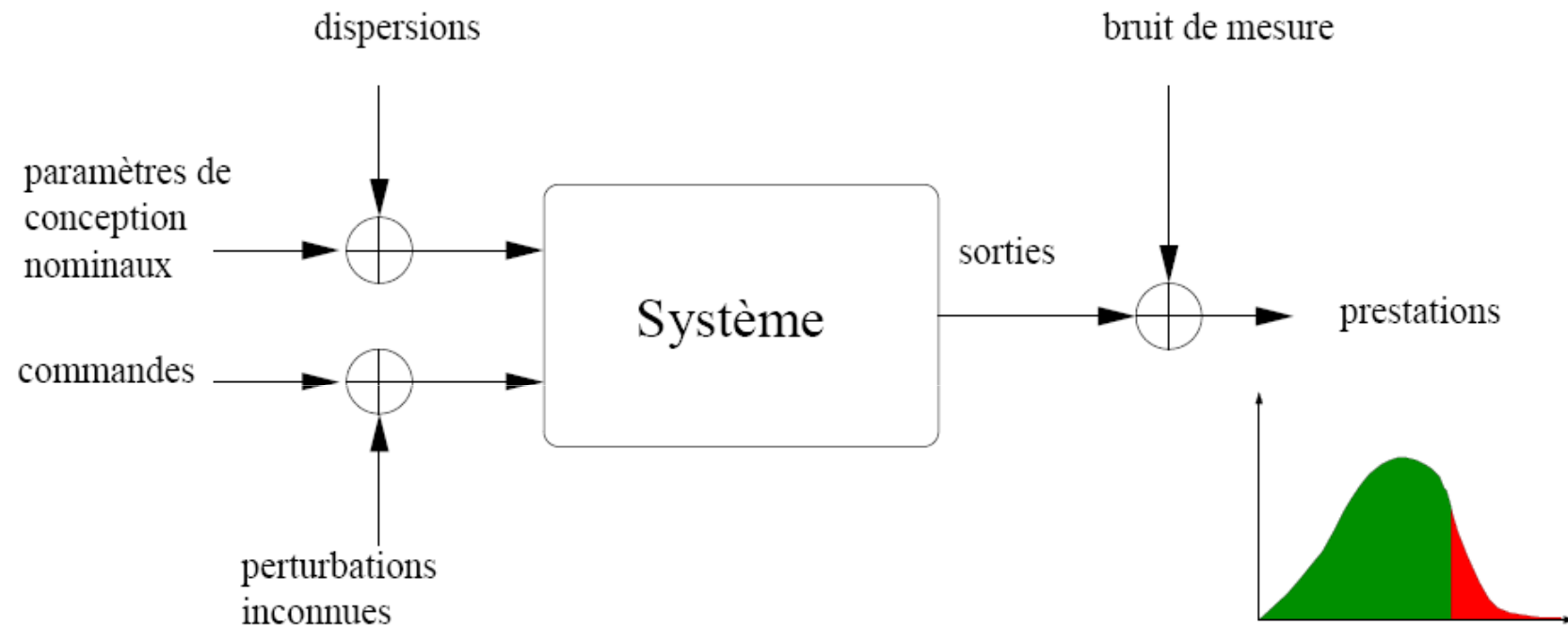
# **modélisation optimale – choix de modèles –**

Gilles Fleury

-----  
*Département SSE*

*EA2523 « Signaux et Systèmes Electroniques »*  
*SUPÉLEC*

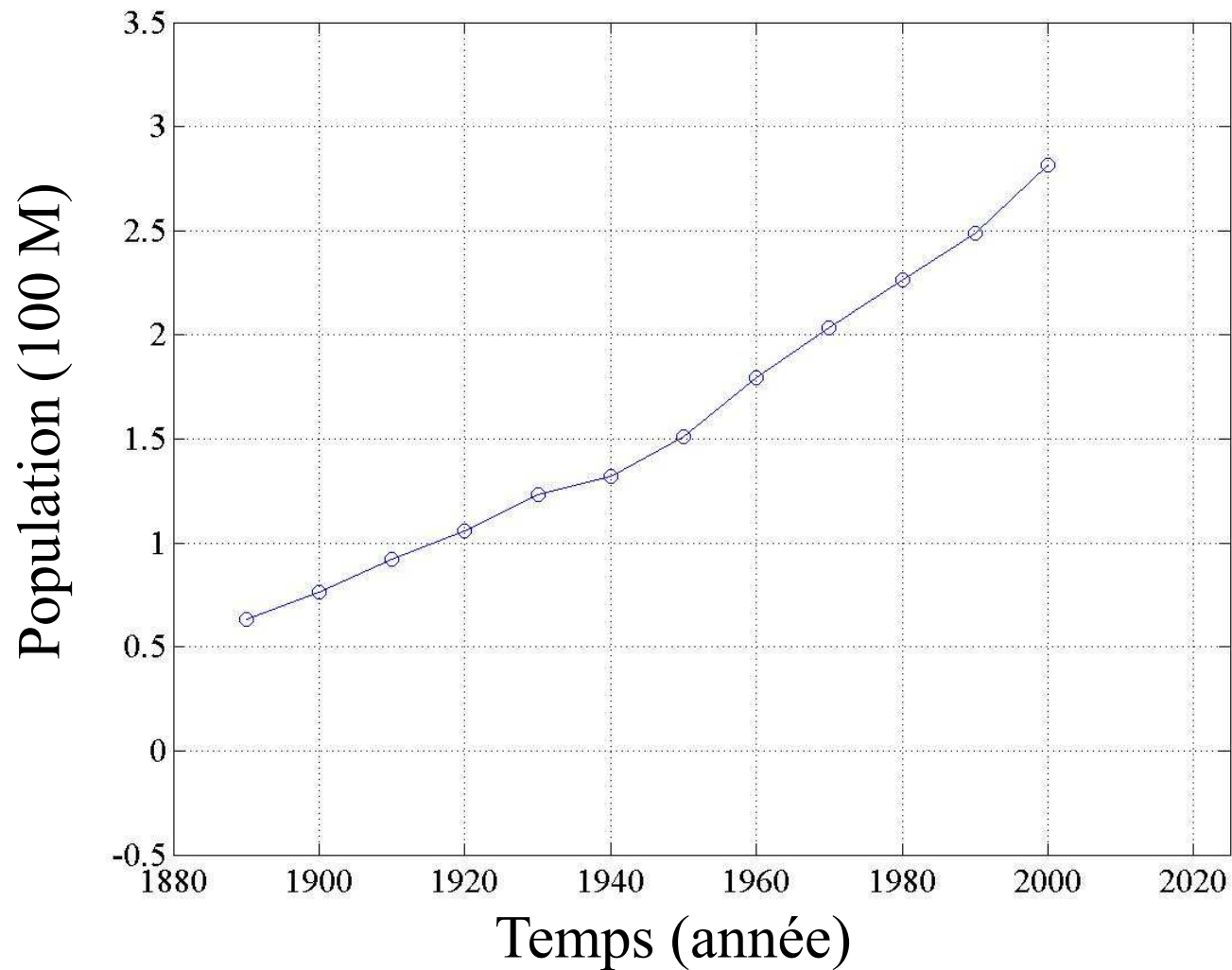
-----  
*01.07.2009*



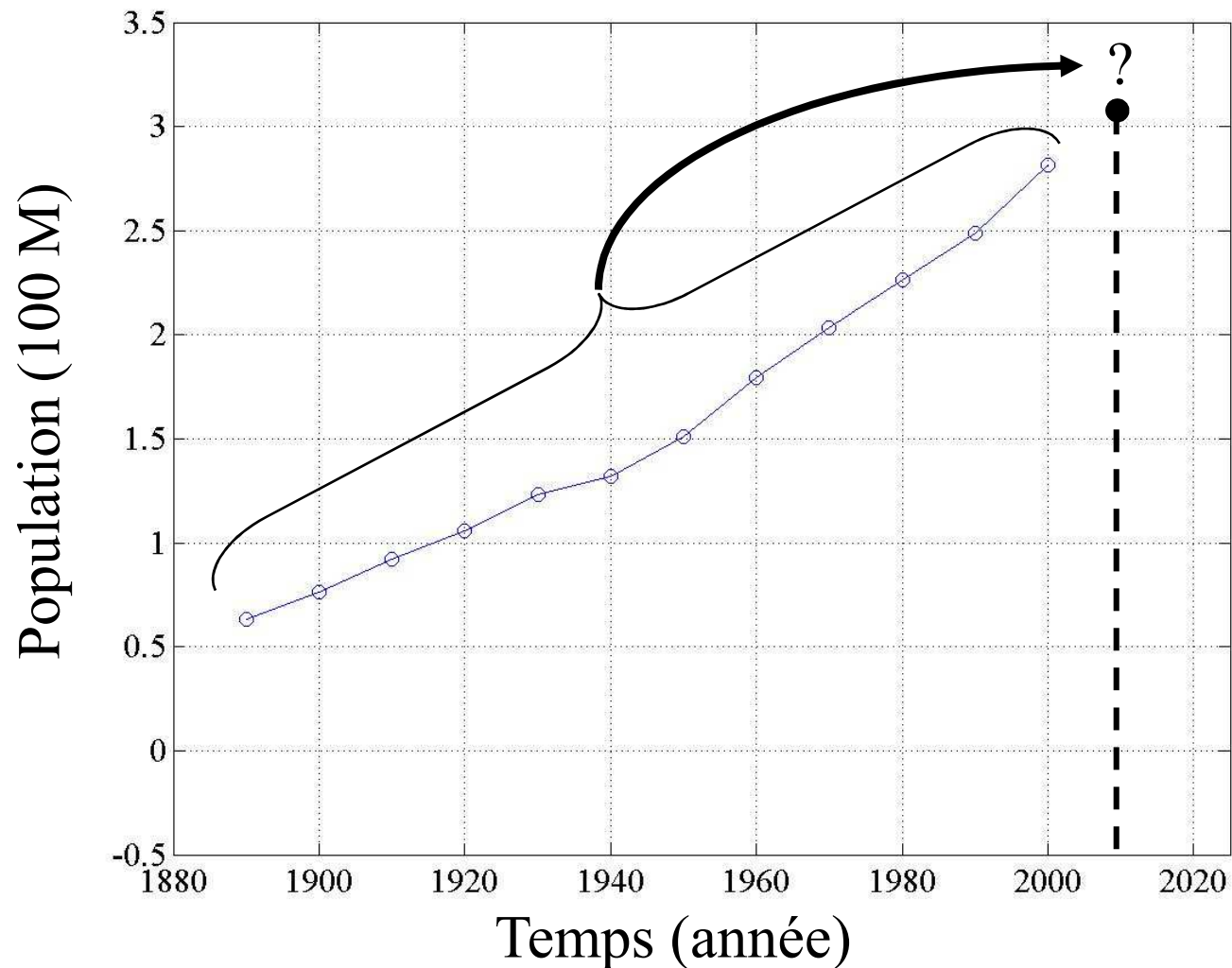
➡ **MODÉLISATION DES SYSTÈMES**

➡ **MODÉLISATION DES INCERTITUDES**

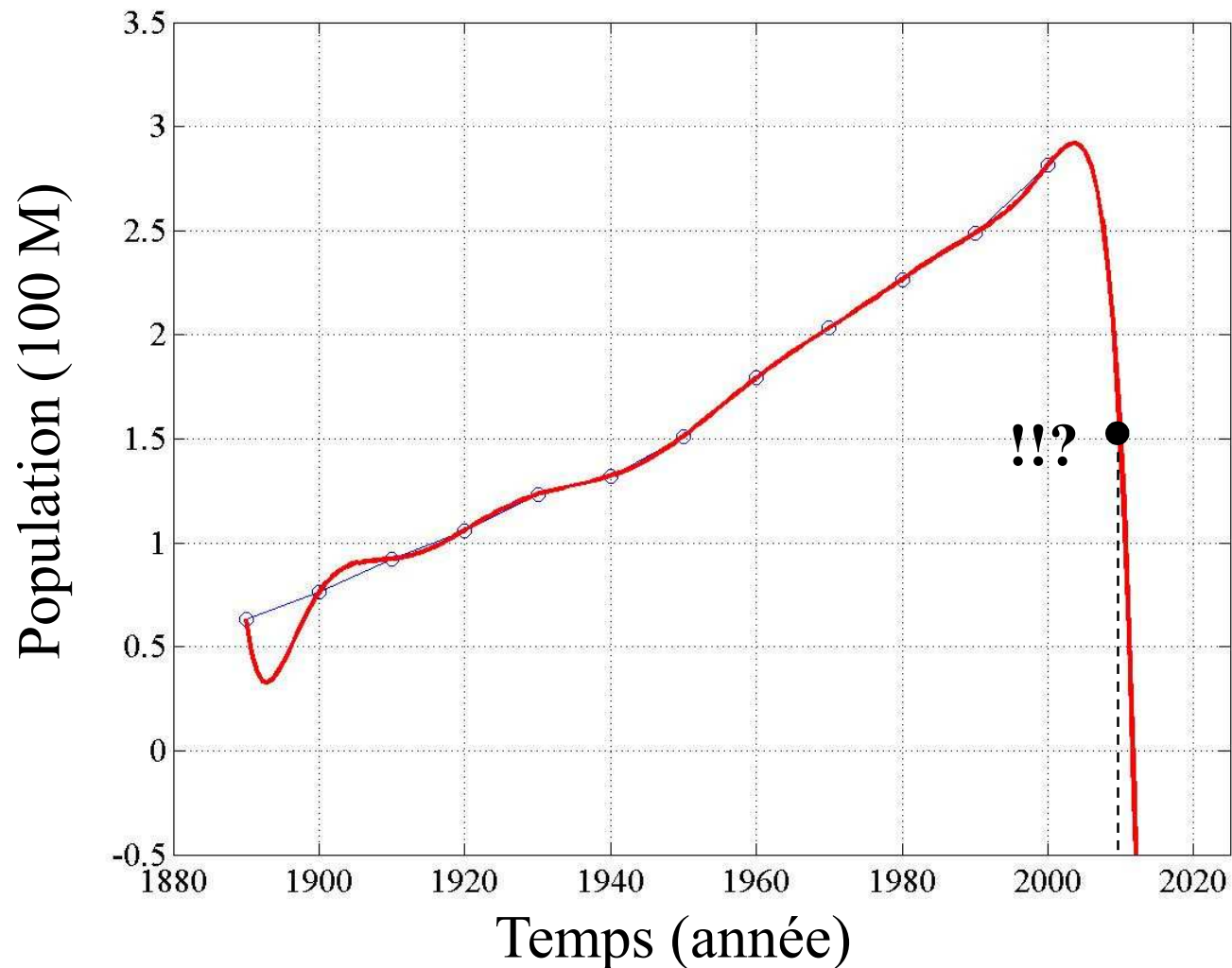
■ *Population des États-Unis*



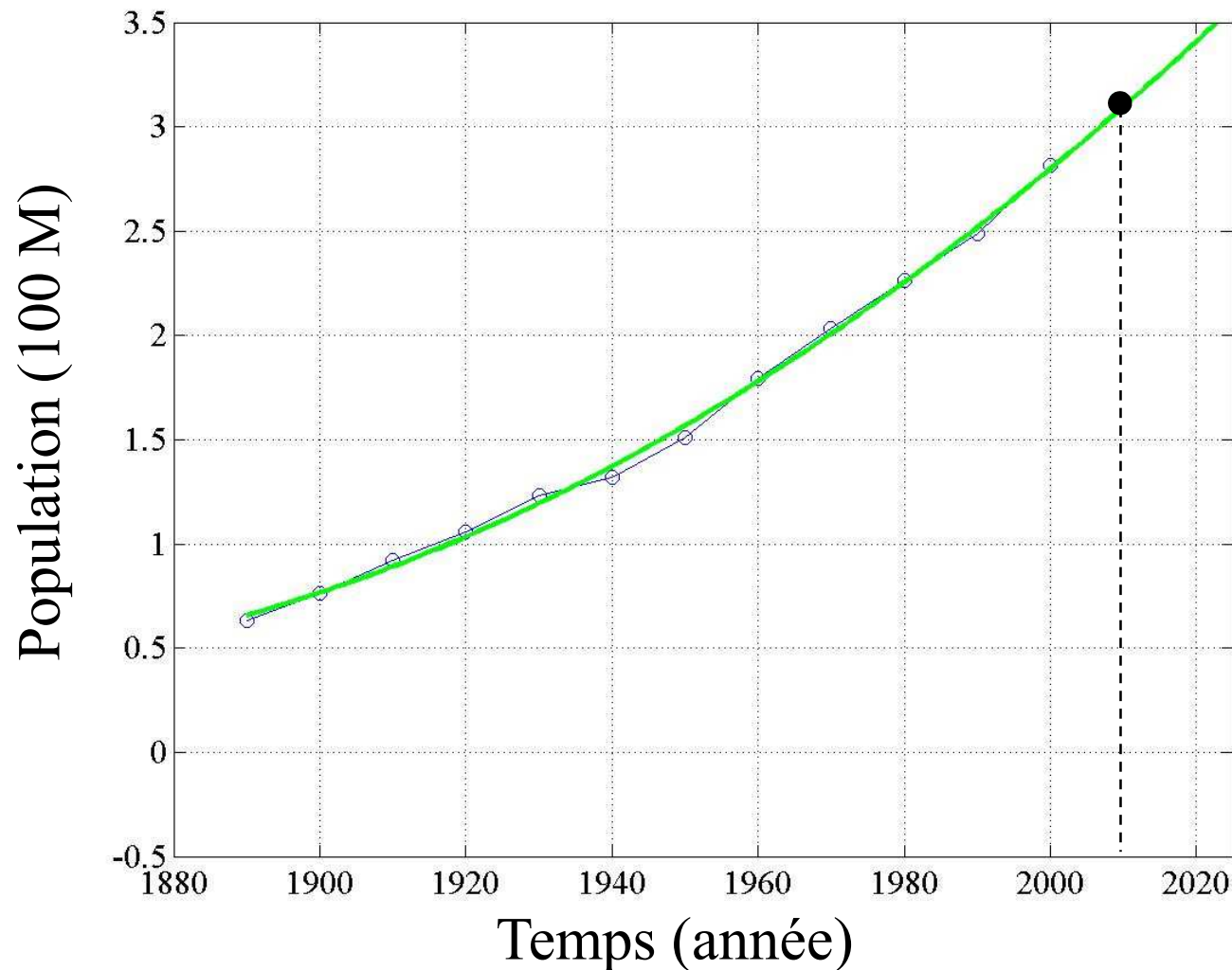
■ *Population des Etats-Unis (prédiction)*



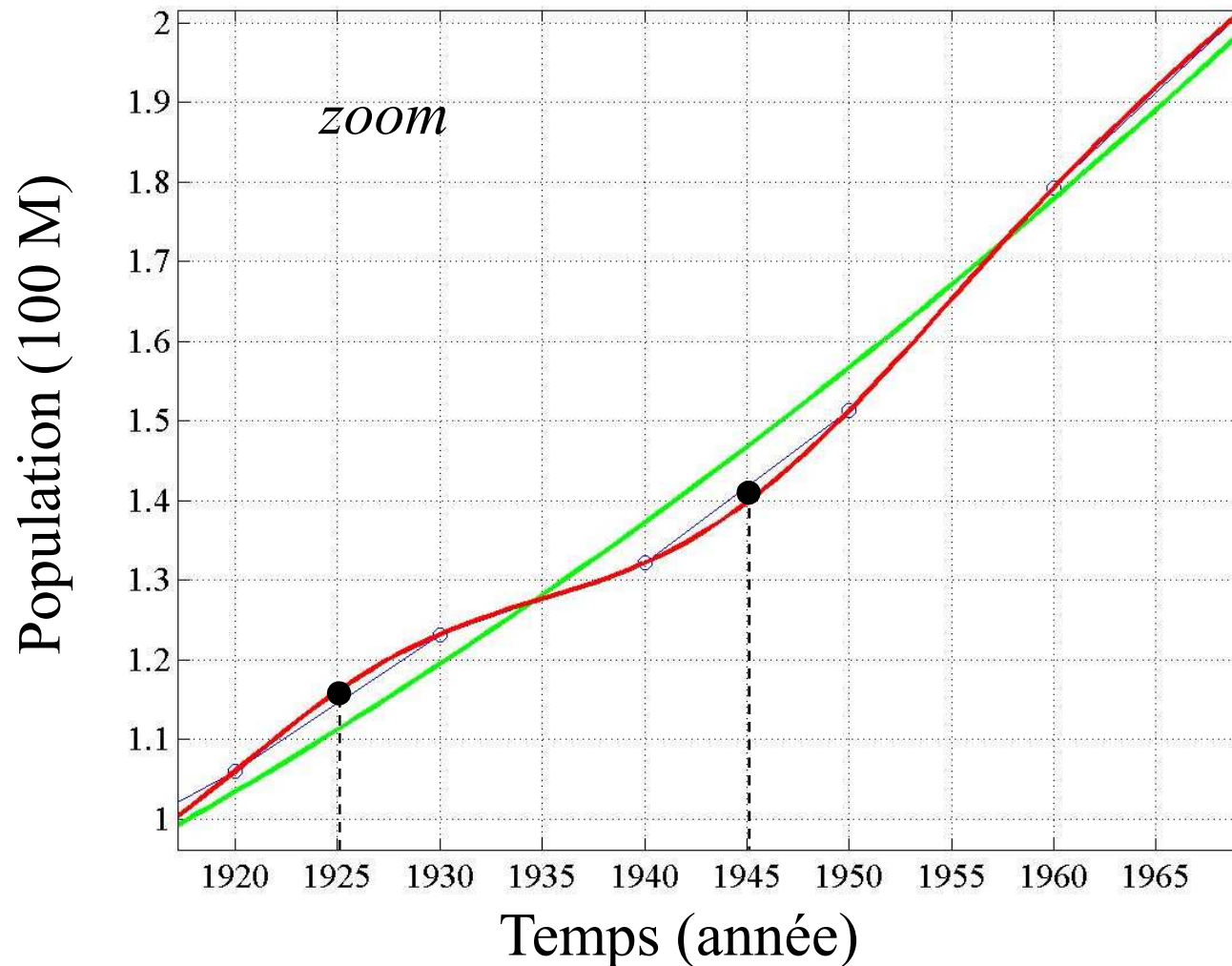
■ *Population des Etats-Unis (polynôme de Lagrange)*



- *Population des Etats-Unis (régression polynomiale, ordre 2)*



■ *Population des Etats-Unis (interpolation ?)*

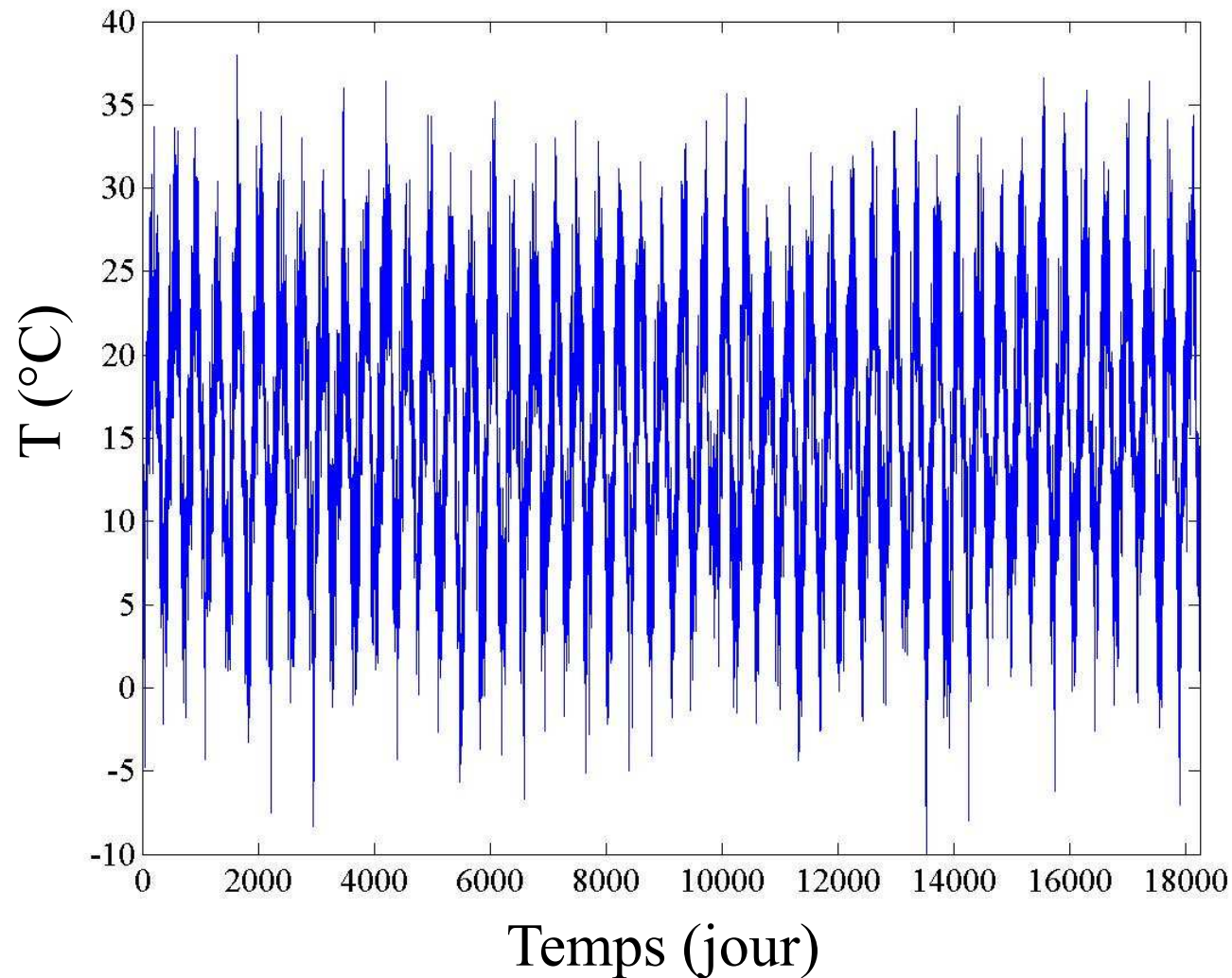


- Modélisation des incertitudes
  - approches paramétriques
  - approches non paramétriques
- Choix d'un modèle de système
  - choix de la complexité d'un modèle
  - choix d'un modèle de comportement
  - choix de la paramétrisation d'un modèle
  - vers des méta-modèles
- Aides à la prise de décision
  - Estimation d'évènements rares
  - Optimisation d'un système complexe
- Exemples d'application



- Modélisation des incertitudes
  - approches paramétriques
  - approches non paramétriques
- Choix d'un modèle de système
  - choix de la complexité d'un modèle
  - choix d'un modèle de comportement
  - choix de la paramétrisation d'un modèle
  - vers des méta-modèles
- Aides à la prise de décision
  - Estimation d'évènements rares
  - Optimisation d'un système complexe
- Exemples d'application

- *Température maximale au parc Montsouris (Paris)*



*signal  $y(t)$*

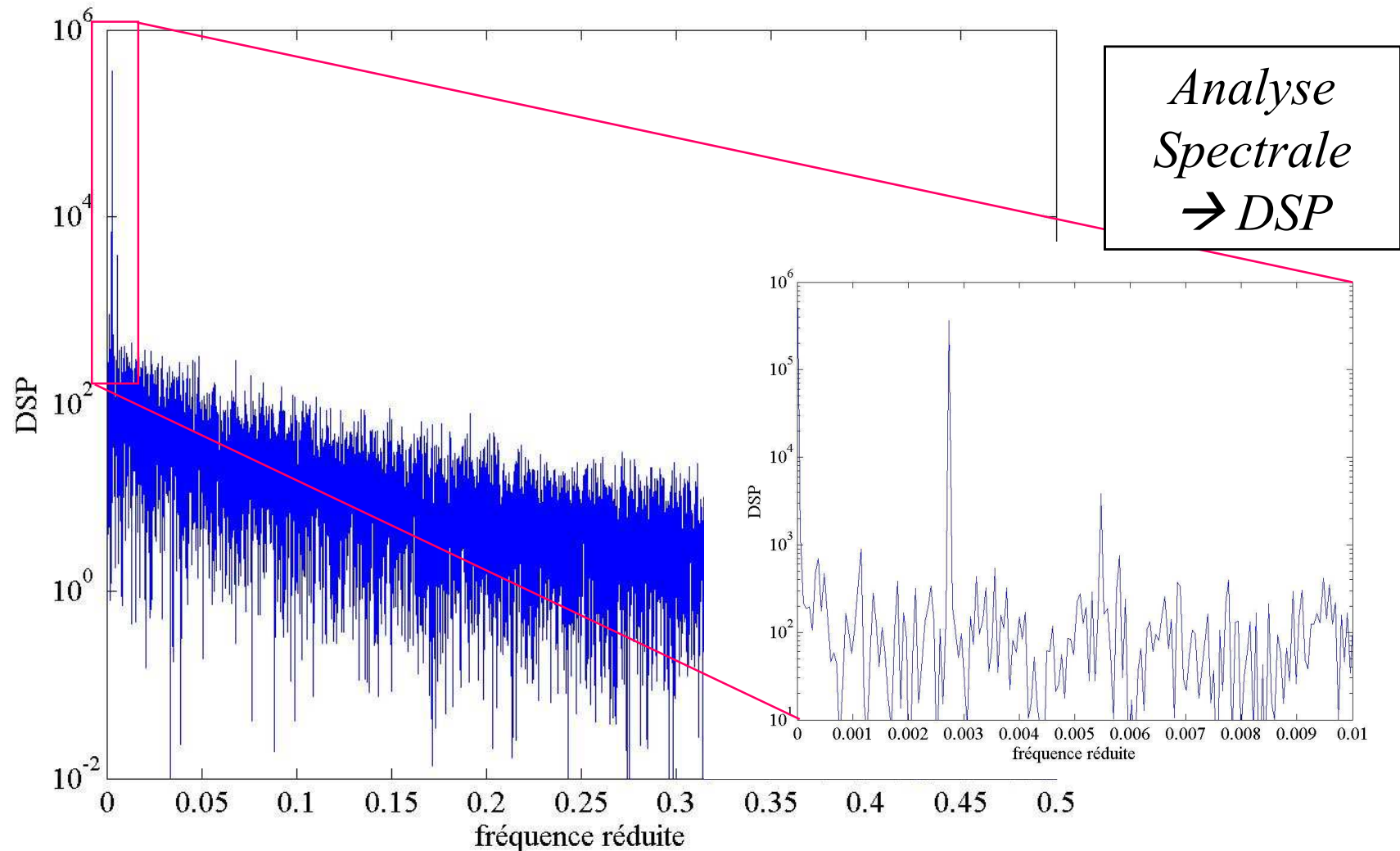
$T_e = 1 \text{ jour}$

$T = 50 \text{ ans}$

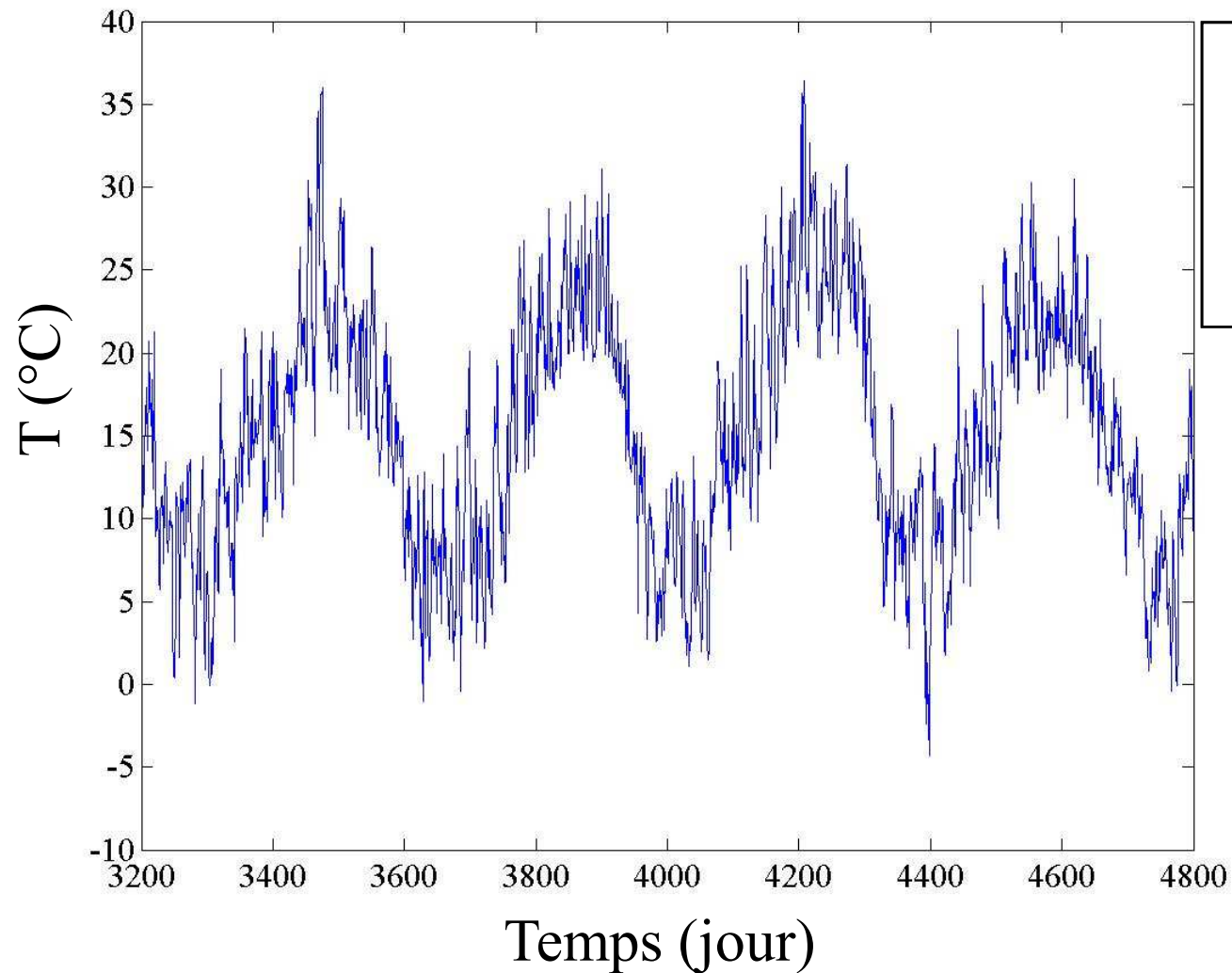


$N = 18263$

■ *Température maximale au parc Montsouris (Paris)*

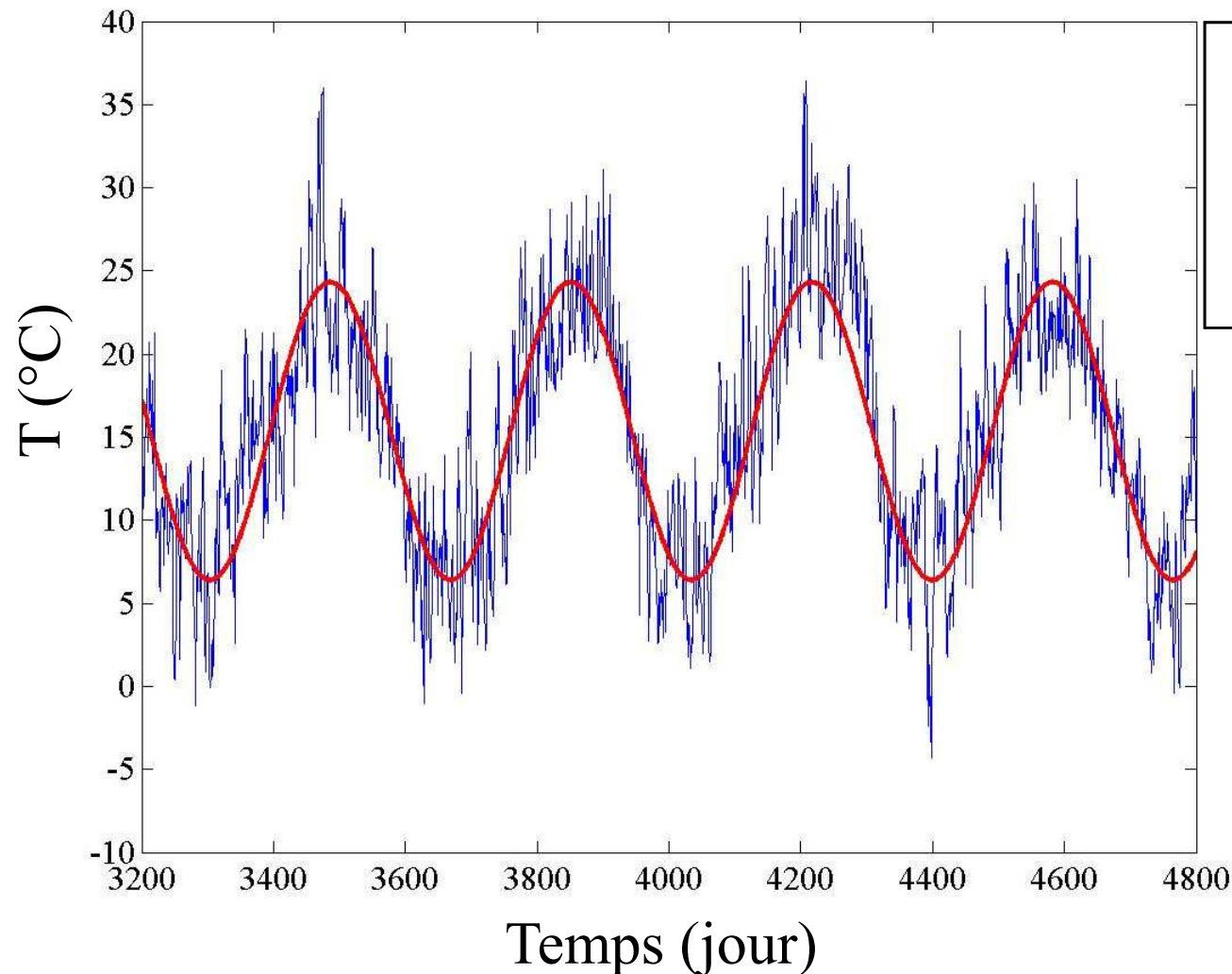


- *Température maximale au parc Montsouris (Paris)*



*Identification  
de  $s(t)$  :  
 $a \cos(2\pi ft + \phi) + m$*

■ *Température maximale au parc Montsouris (Paris)*



*Identification  
de  $s(t)$  :*

$$a \cos(2\pi f t + \phi) + m$$



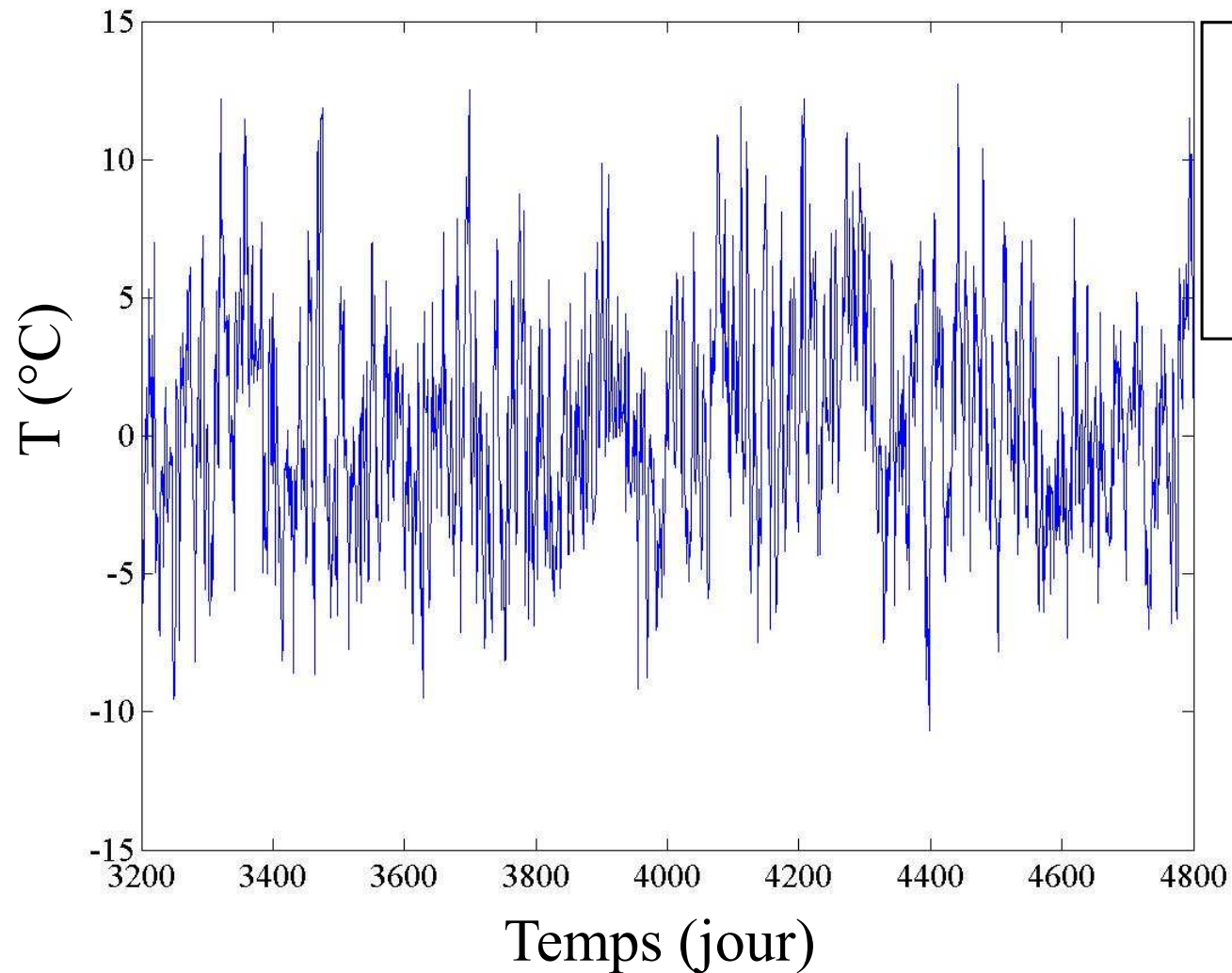
$$a = 8.9593 \text{ (}^{\circ}\text{C)}$$

$$f = 1.0009 \text{ (an}^{-1}\text{)}$$

$$\phi = -1.8361 \text{ (rad)}$$

$$m = 15.3545 \text{ (}^{\circ}\text{C)}$$

- *Température maximale au parc Montsouris (Paris)*



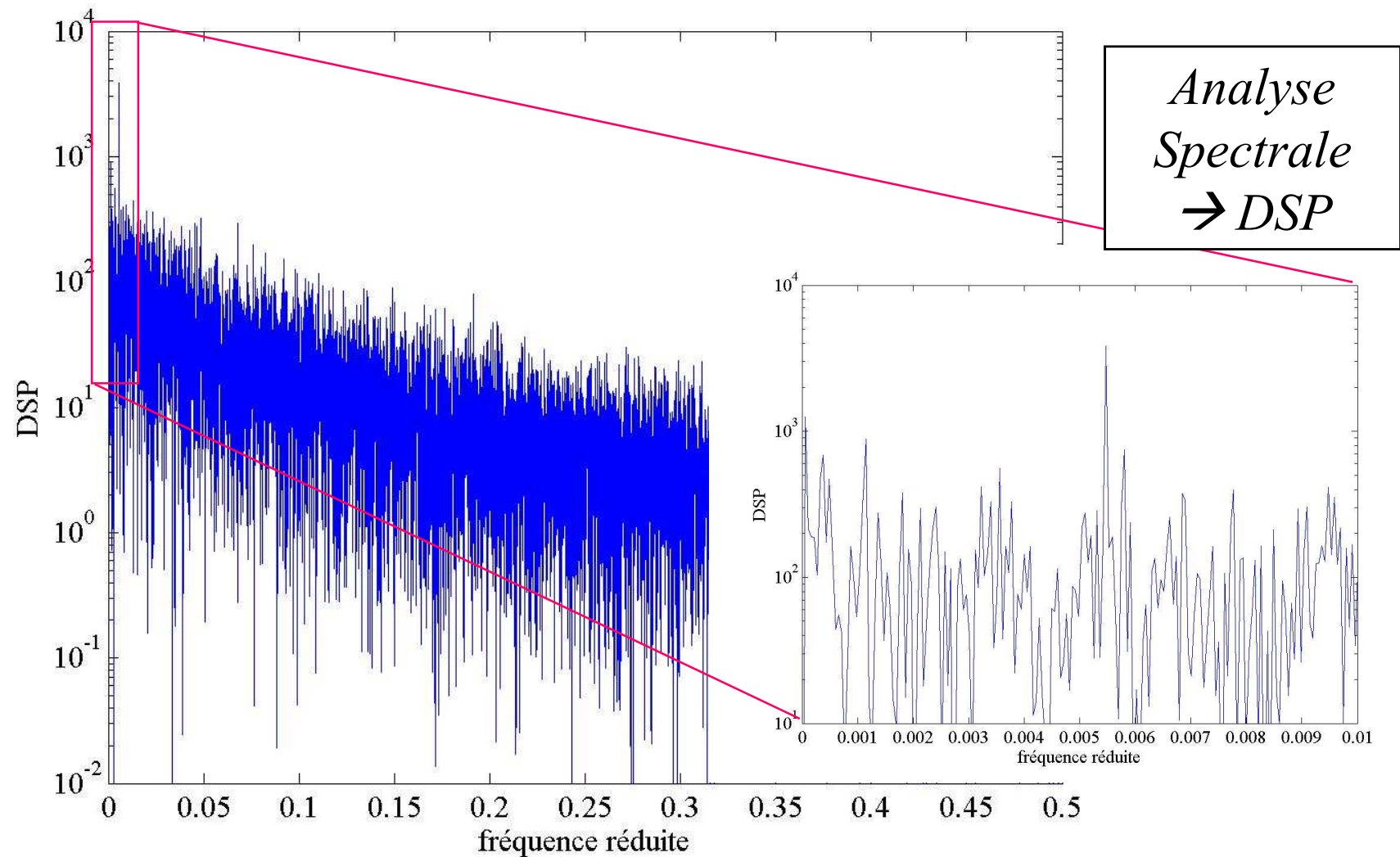
*Résidu  
d'identification  
 $y(t) = s(t) + r(t)$*



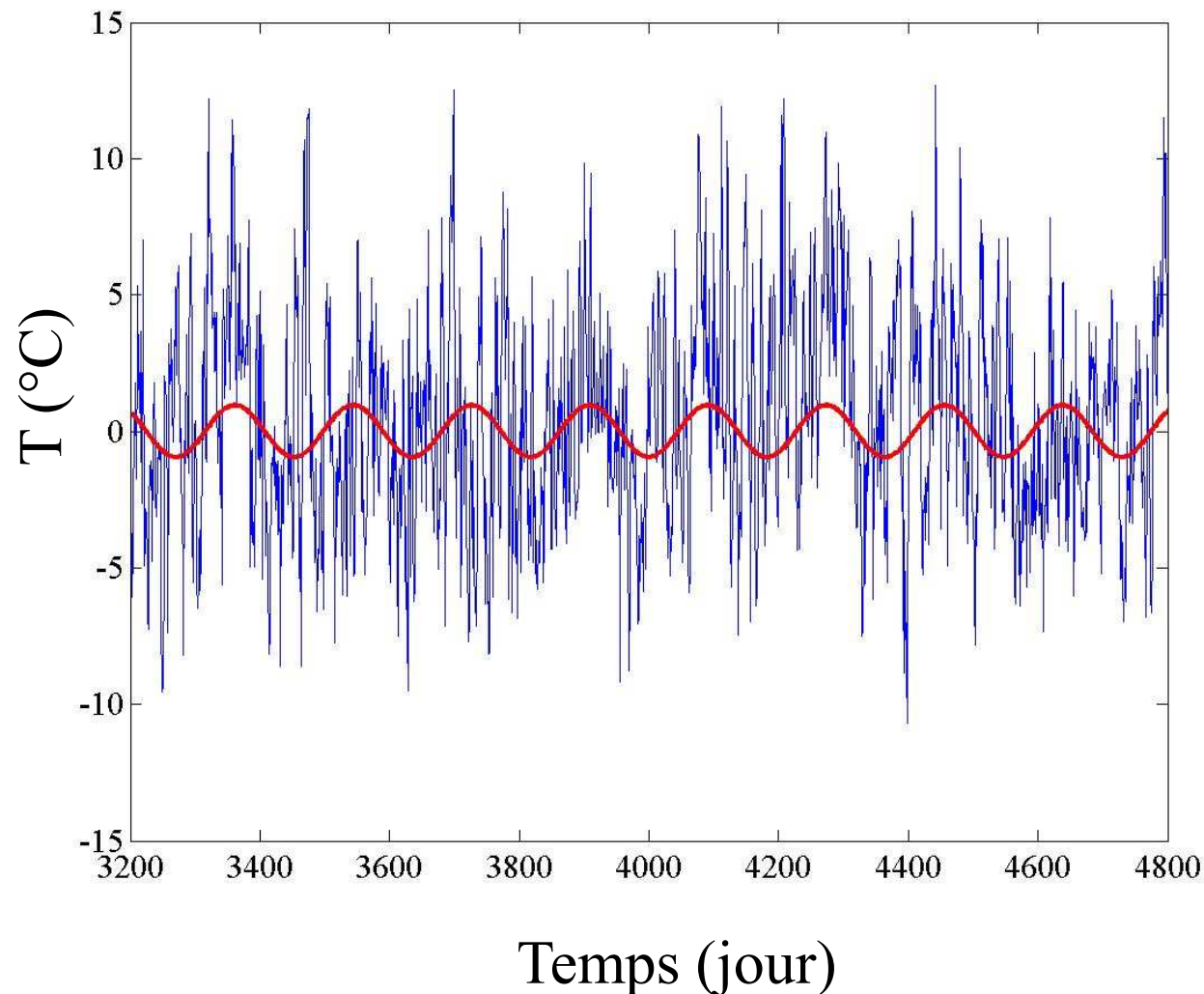
Caractérisation  
du résidu  $r(t)$



■ *Température maximale au parc Montsouris (Paris)*



■ *Température maximale au parc Montsouris (Paris)*



*Identification  
de  $s(t)$  :*

$$a \cos(2\pi ft + \phi) + m$$



$$a = 0.9514 \text{ (}^{\circ}\text{C)}$$

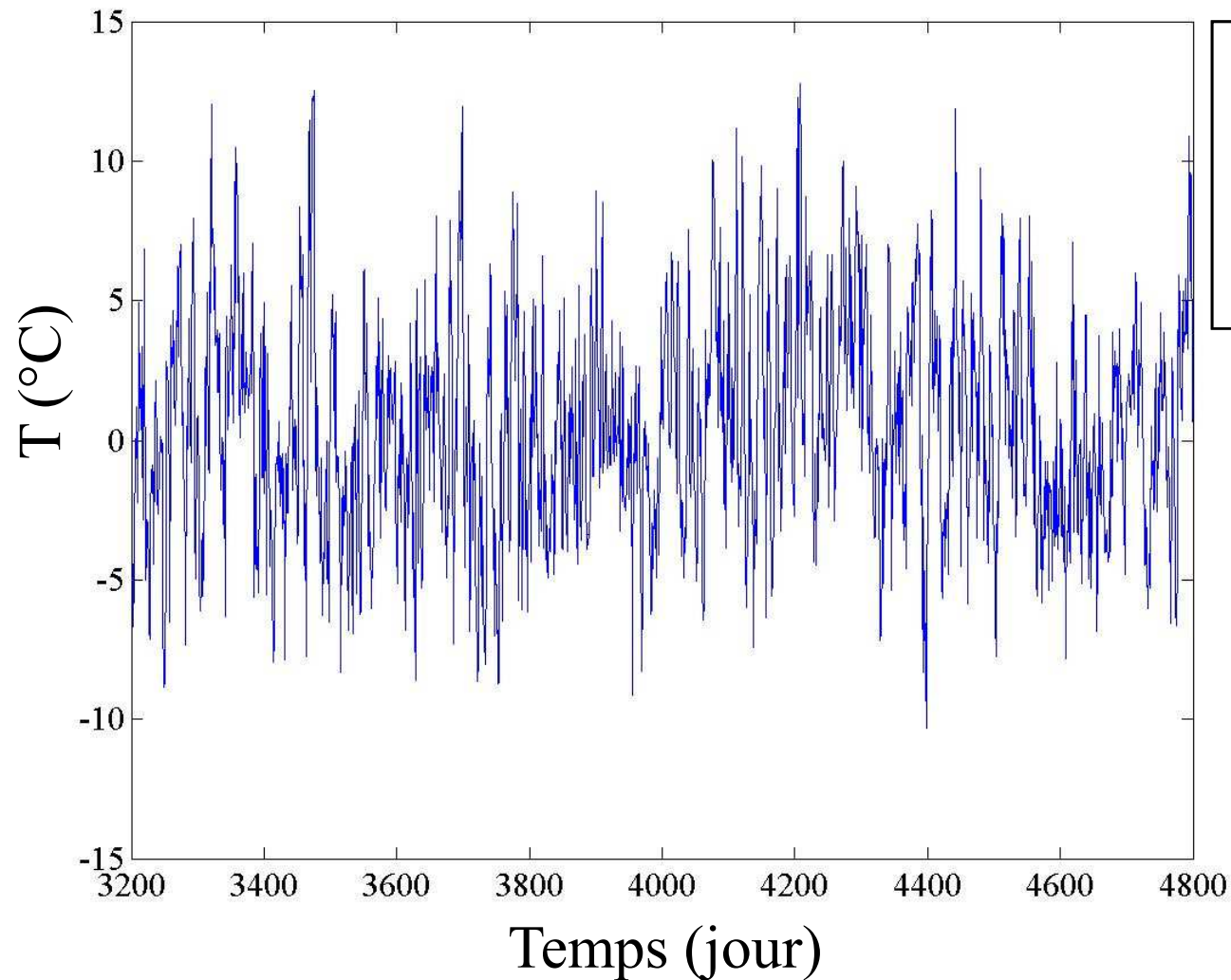
$$f = 0.4996 \text{ (an}^{-1}\text{)}$$

$$\phi = -1.1021 \text{ (rad)}$$

$$m = -0.0002 \text{ (}^{\circ}\text{C)}$$



- *Température maximale au parc Montsouris (Paris)*



*Choix de l'ordre  
d'un modèle :  
**compromis***

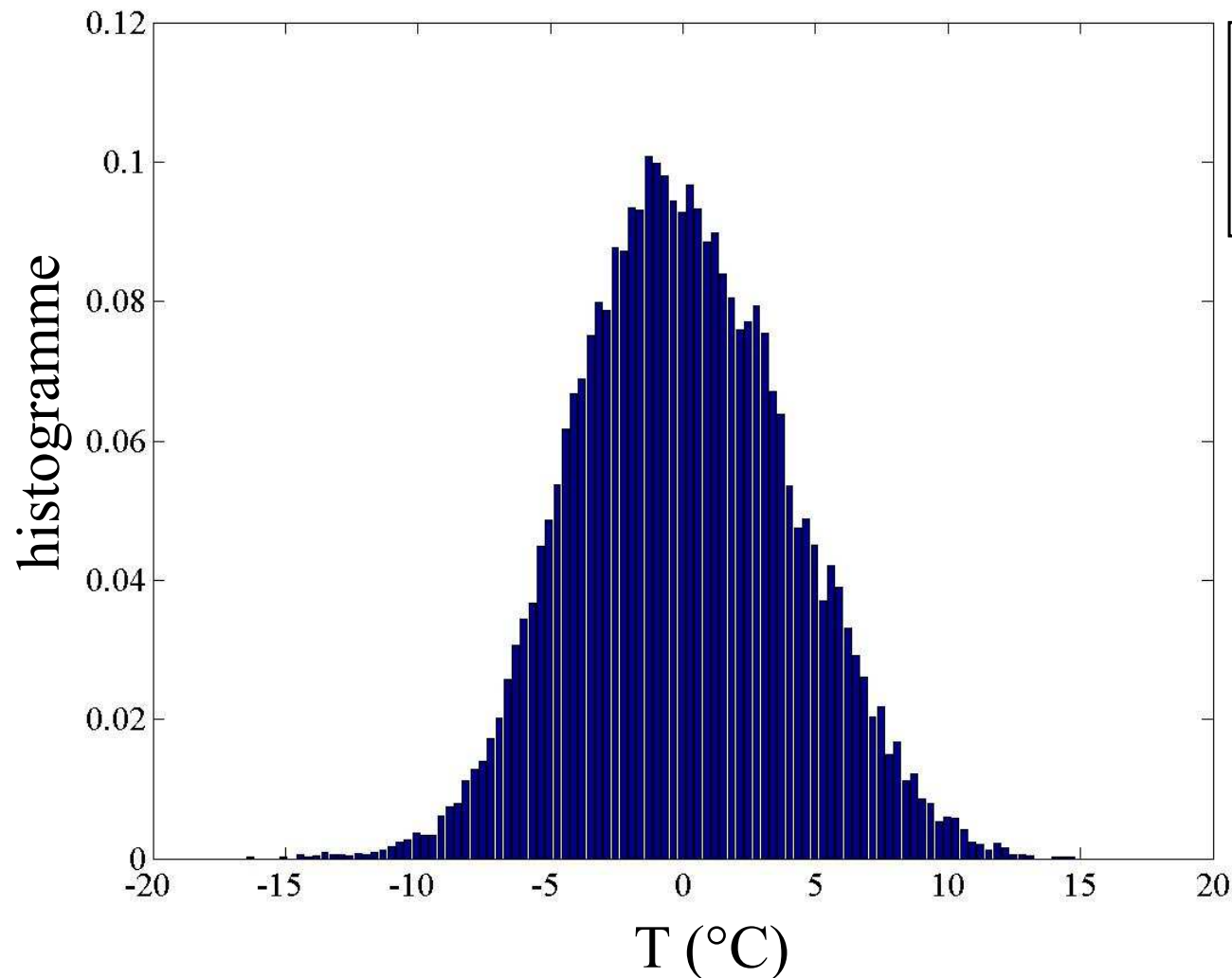


$$\sigma_1^2 = 2320$$

$$\sigma_2^2 = 552$$

$$\sigma_3^2 = 543$$

- *Température maximale au parc Montsouris (Paris)*



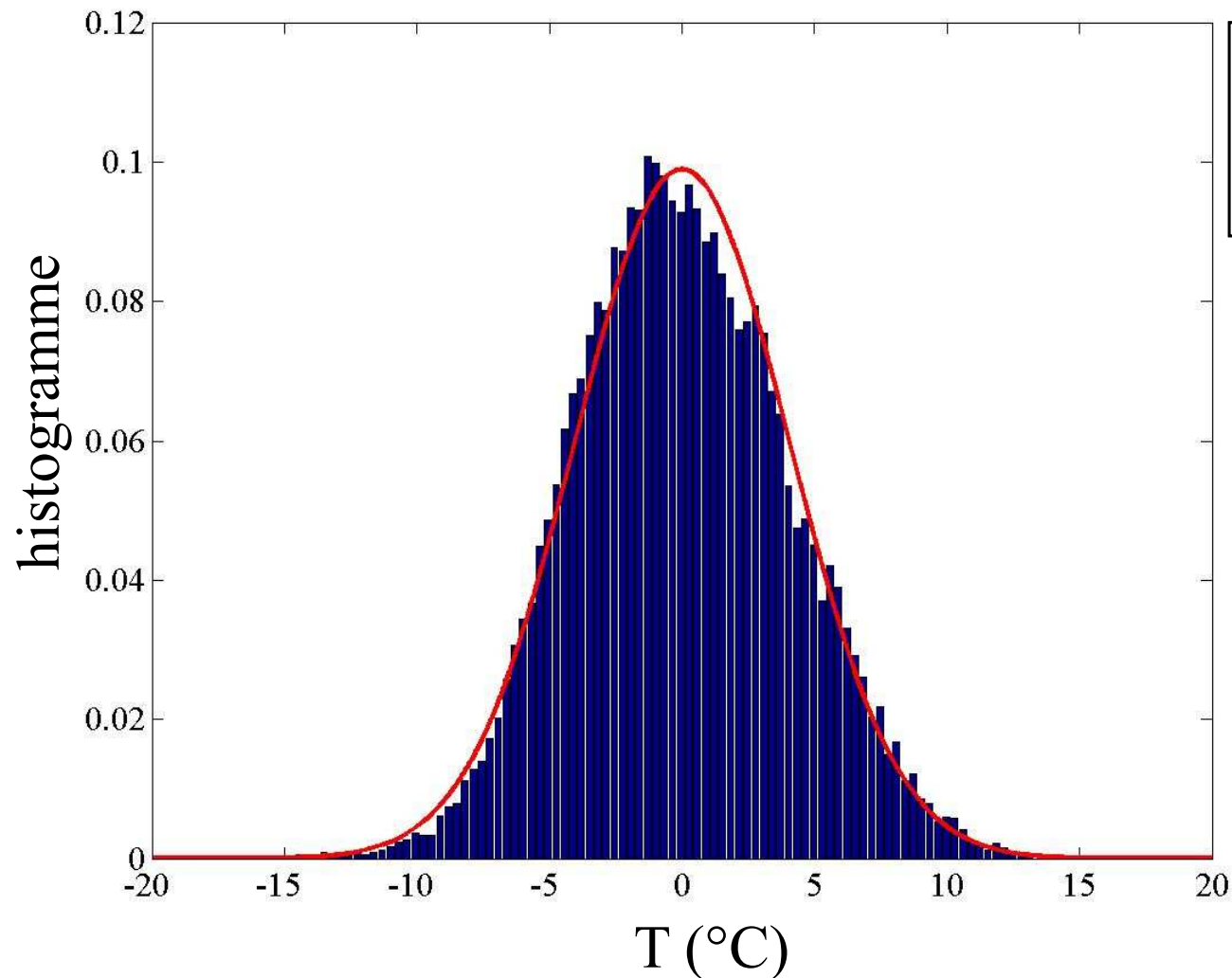
*Analyse  
du résidu*



estimateur de la ddp  
(densité de probabilité) :

- non paramétrique

■ *Température maximale au parc Montsouris (Paris)*



*Analyse  
du résidu*

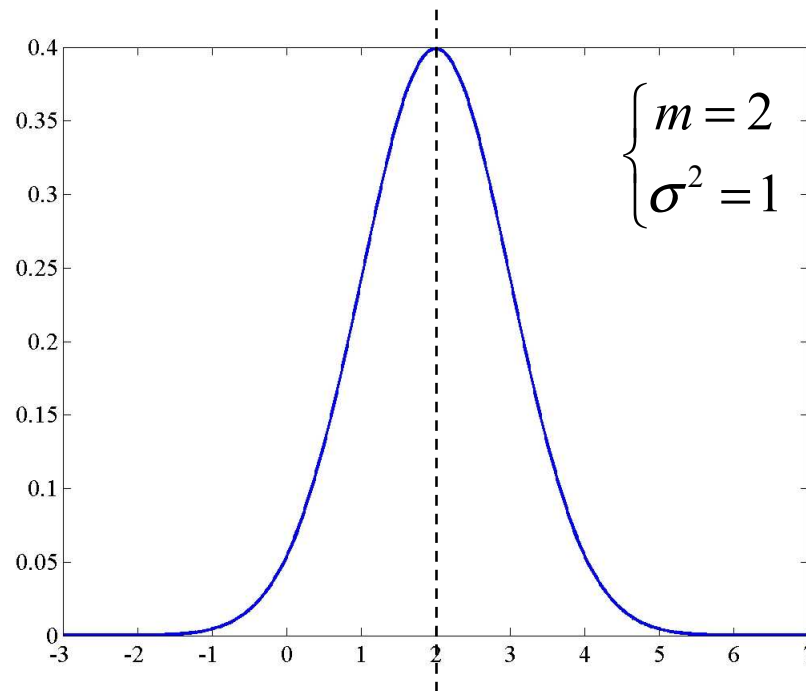


estimateur de la ddp  
(densité de probabilité) :

- non paramétrique
- paramétrique

### ■ Le modèle gaussien

$$\wp(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$



- calcul analytique possible
- théorème limite centrale

$$x = \frac{1}{N} \sum_{i=1}^N u_i \quad u_i \text{ IID}$$

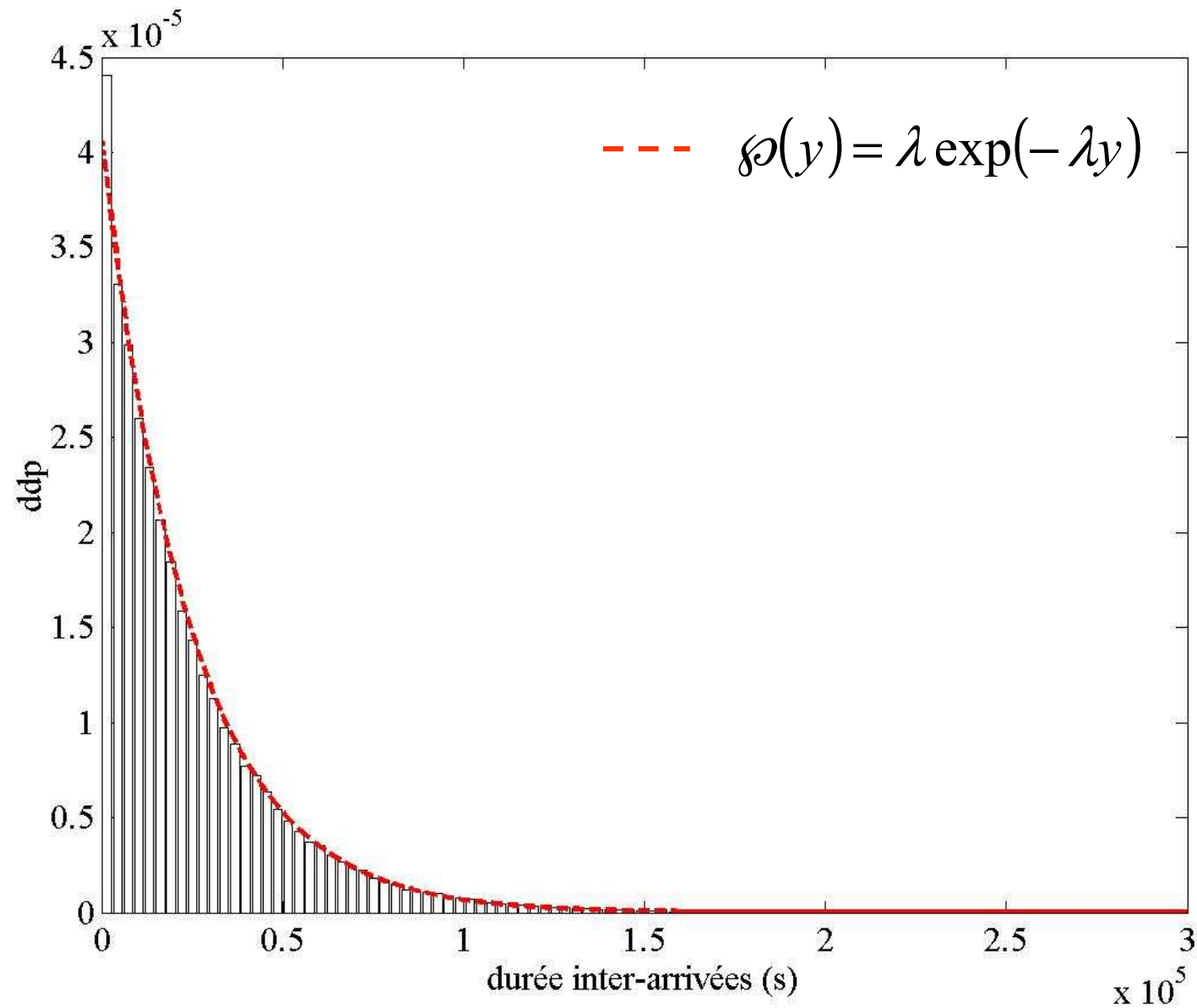
$$x \xrightarrow{N \rightarrow \infty} \wp(x) \text{ gaussien}$$

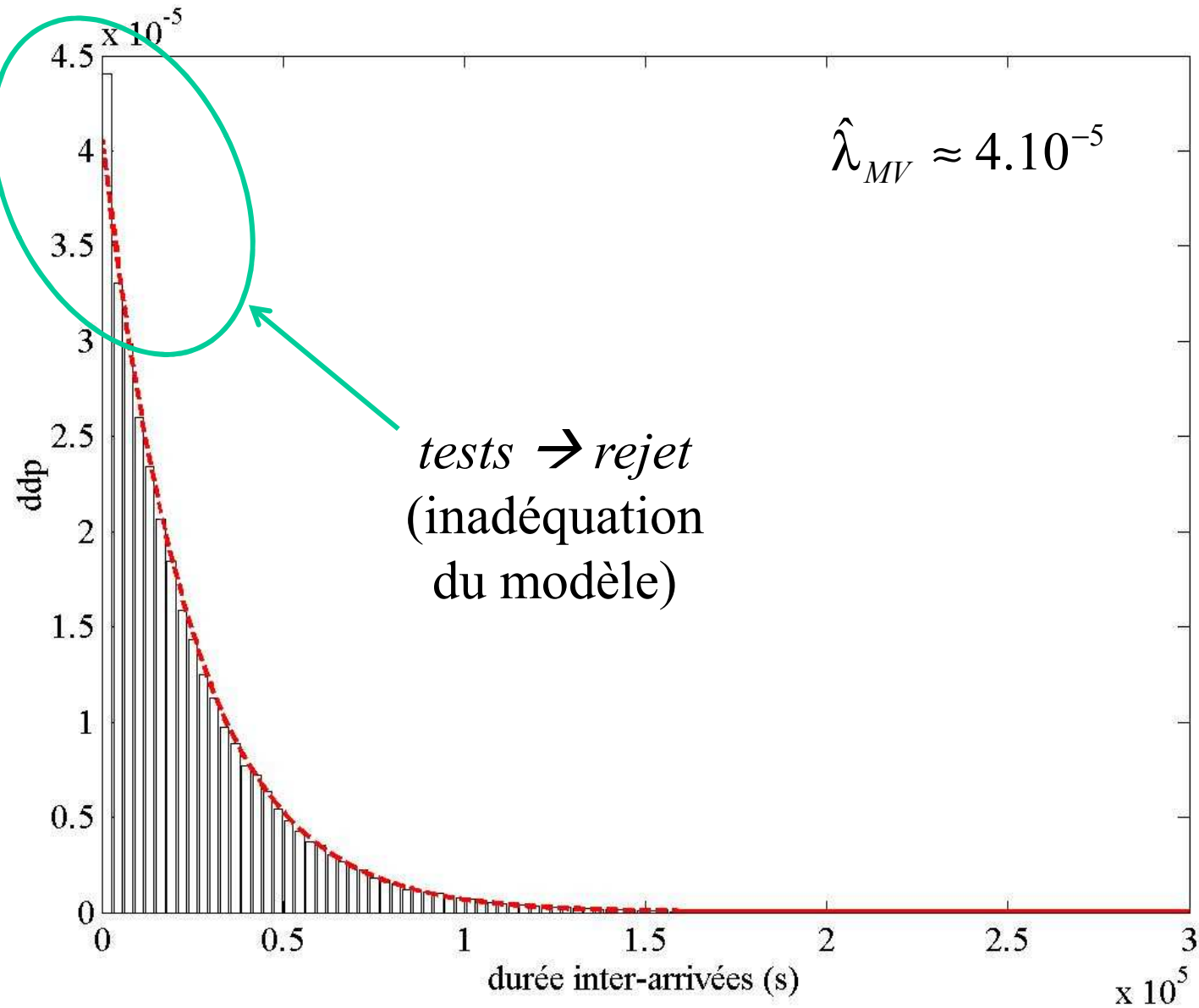
- MAXENT ( $H = -I$ )

$$H = -\int \wp(x) \log \wp(x) dx$$

$$\begin{cases} \max H \\ E\{\phi_i(x)\} = \mu_i \quad i \in \{1, 2, \dots\} \end{cases}$$

- *Autres modèles paramétriques*
  - lognormale, bêta (uniforme), gamma (exponentielle),...
- *Identification des paramètres de la loi*
  - méthode des moments
  - méthode du maximum de vraisemblance
  - approche bayésienne (a priori)
- *Tests d'ajustement*
  - test du Chi2
  - test de Kolmogorov-Smirnov
  - test de Cramer-von Mises
  - test de Wilcoxon-Mann-Whitney





- *Méthodes par noyaux* [Rosenblatt, Parzen, Masry]

$$\hat{\rho}(y) = \frac{1}{N} \sum_{k=1}^N \frac{1}{h} K\left(\frac{y - y_k}{h}\right)$$

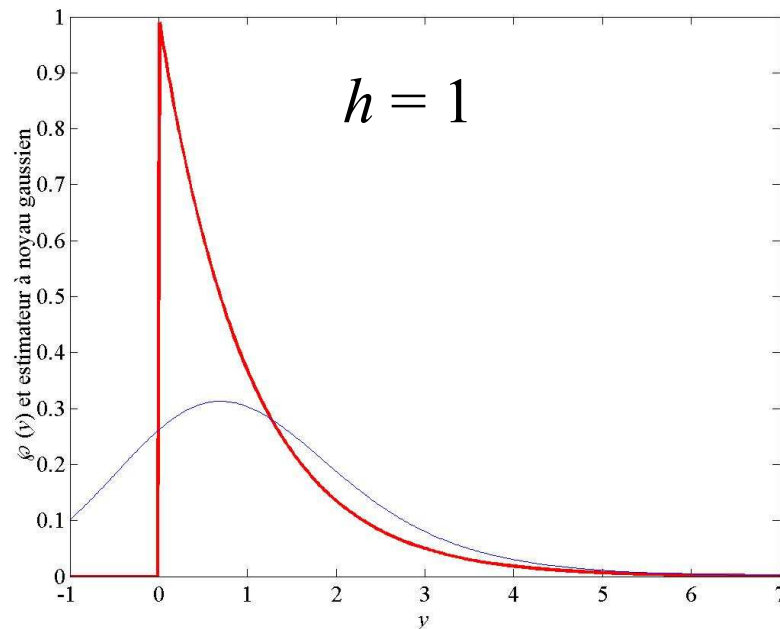
- *Choix du noyau ?*

$$\left\{ \begin{array}{ll} \int_{-\infty}^{+\infty} |K(y)| dy < \infty & \sup_y |K(y)| < \infty \\ \int_{-\infty}^{+\infty} K(y) dy = 1 & \lim_{|y| \rightarrow \infty} K(y) = 0 \end{array} \right.$$

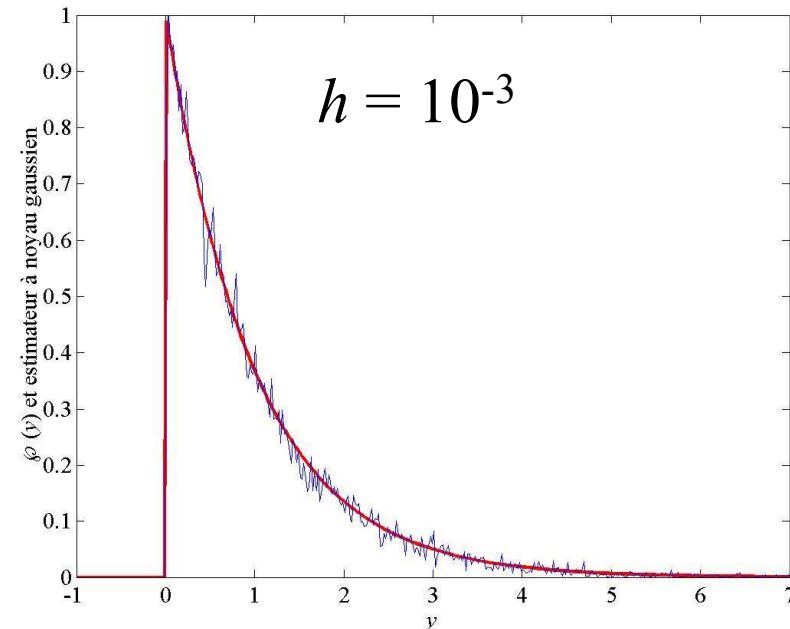
- *Choix de la largeur  $h$  ?*

$$\left\{ \begin{array}{l} \lim_N h = 0 \\ \lim_N N.h = \infty \end{array} \right.$$





biais important  
variance faible

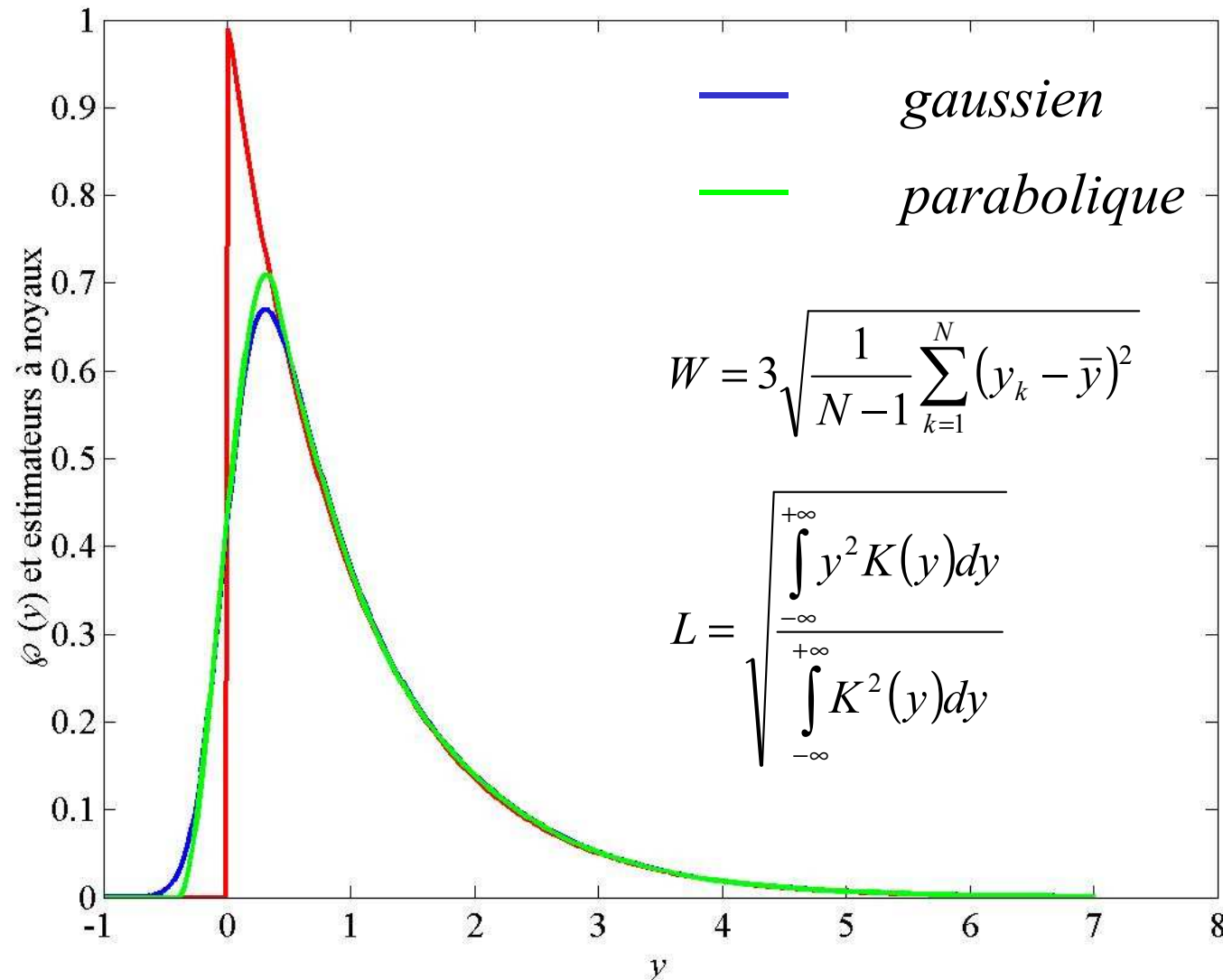


biais faible  
variance importante

Compromis  
biais-variance

[Jones]

$$MISE = E \left\{ \int_{-\infty}^{+\infty} (\rho(y) - \hat{\rho}(y))^2 dy \right\} \longrightarrow h^* = \left( \frac{\int_{-\infty}^{+\infty} K^2(y) dy}{N \left( \int_{-\infty}^{+\infty} (\rho^{(2)}(y))^2 dy \right) \left( \int_{-\infty}^{+\infty} y^2 K(y) dy \right)^2} \right)^{1/5}$$



$$\hat{h} = \frac{1}{2} \frac{W}{L} \frac{1}{N^{1/5}}$$

$$W = 3 \sqrt{\frac{1}{N-1} \sum_{k=1}^N (y_k - \bar{y})^2}$$

$$L = \sqrt{\frac{\int_{-\infty}^{+\infty} y^2 K(y) dy}{\int_{-\infty}^{+\infty} K^2(y) dy}}$$

■ *Méthode par noyau optimisée (IRINCORREL) [Rivoira]*

1. Estimer la ddp par une première approche par noyau classique

$$\hat{\wp}(y) = \frac{1}{N} \sum_{k=1}^N \frac{1}{h_1} K_1 \left( \frac{y - y_k}{h_1} \right) \quad h_1 = \frac{1}{2} \frac{W}{L_1} \frac{1}{N^{1/5}}$$

2. Estimer la dérivée seconde de la ddp au point  $y$

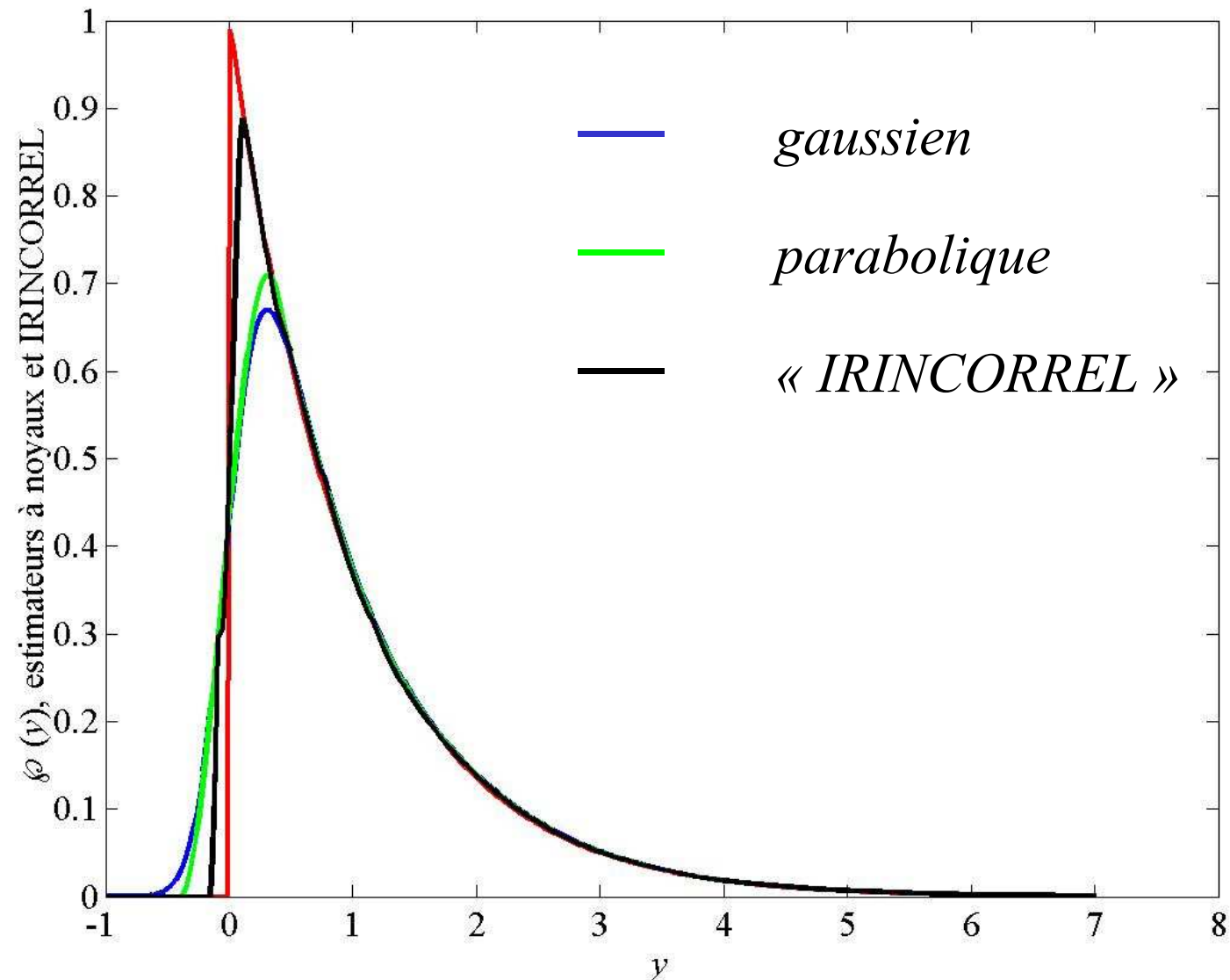
$$\hat{\wp}^{(2)}(y) = \frac{1}{N} \sum_{k=1}^N \frac{1}{h_2} K_2 \left( \frac{y - y_k}{h_2} \right) \quad h_2 = \frac{1}{2} \frac{W}{L_2} \frac{1}{N^{1/9}}$$

3. Estimer les largeurs optimales à partir de ces deux estimations

$$h_{opt}(y) = \left( L \frac{\hat{\wp}(y)}{\hat{\wp}^{(2)}(y)} \frac{1}{N} \right)^{1/5}$$

4. Estimer la ddp au point  $y$

$$\hat{\wp}_{opt}(y) = \frac{1}{N} \sum_{k=1}^N \frac{1}{h_{opt}(y)} K \left( \frac{y - y_k}{h_{opt}(y)} \right)$$

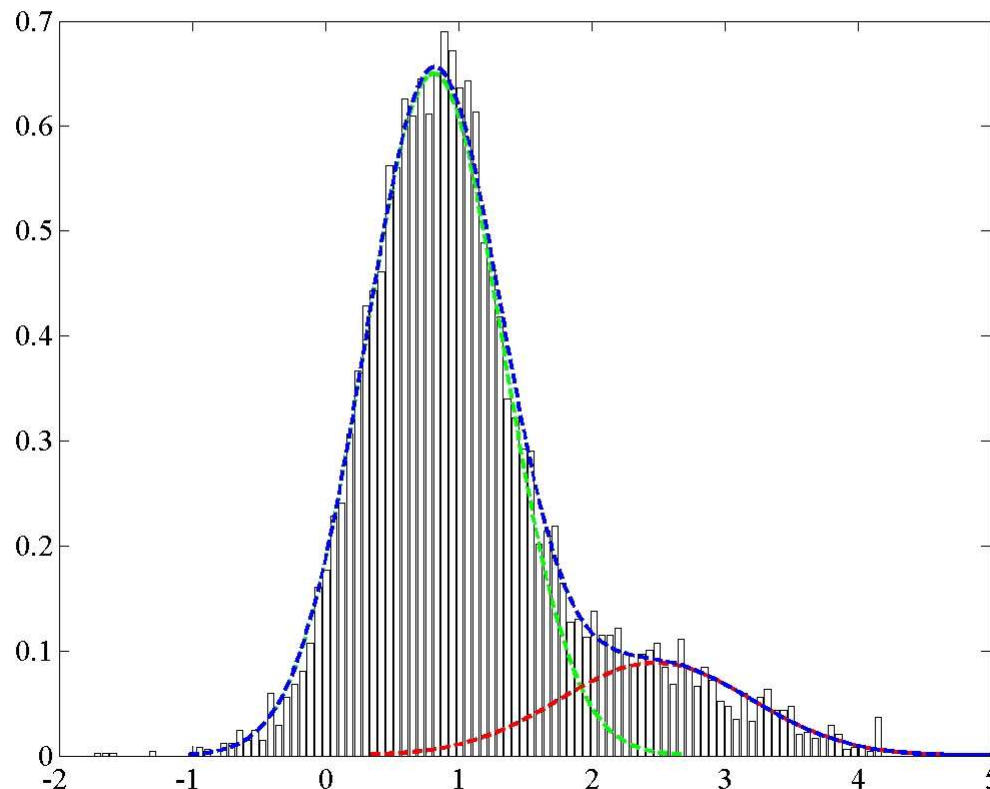


### ■ Entre deux : mélange de lois

$$\wp(y) = \sum_{i=1}^P \pi_i K\left(\frac{y - m_i}{\sigma_i}\right)$$

$K$  noyau (gaussien)

$$\sum_{i=1}^P \pi_i = 1$$



Identification :  
algorithme EM  
(MV, itératif)

$$\pi = 0.8404$$

$$m_1 = 0.8150$$

$$\sigma_1^2 = 0.2660$$

$$m_2 = 2.4792$$

$$\sigma_2^2 = 0.5188$$

■ *Polynômes de chaos*

$$y = \sum_{i=1}^P a_i P_i(\xi)$$

$P_i(.)$  polynôme

$\xi$  variable centrée réduite

- **orthogonalité** des polynômes

$$\langle P_i, P_j \rangle = \int_{-\infty}^{+\infty} P_i(\xi) P_j(\xi) \varphi(\xi) d\xi = \delta_{ij}$$

- dépend de la **mesure**  $\varphi(\xi)$  choisie  
 typ.  $\xi$  gaussienne centrée réduite  
 →  $P$  polynôme de Hermite

$$P_0(\xi) = 1$$

$$P_1(\xi) = \xi$$

$$P_2(\xi) = (\xi^2 - 1)/\sqrt{2}$$

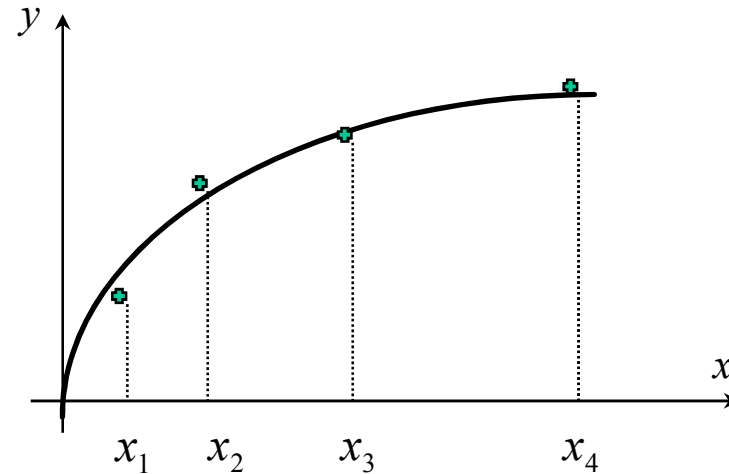
$$P_3(\xi) = (\xi^3 - 3\xi)/\sqrt{6}$$

...

### 1. Équation d'observation

$$y_i = f(x_i, \theta) + e_i$$

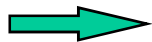
ex.:  $f(x, \theta) = \frac{\theta_1 x}{\theta_2 x + 1}$



### 2. Équation de mesure

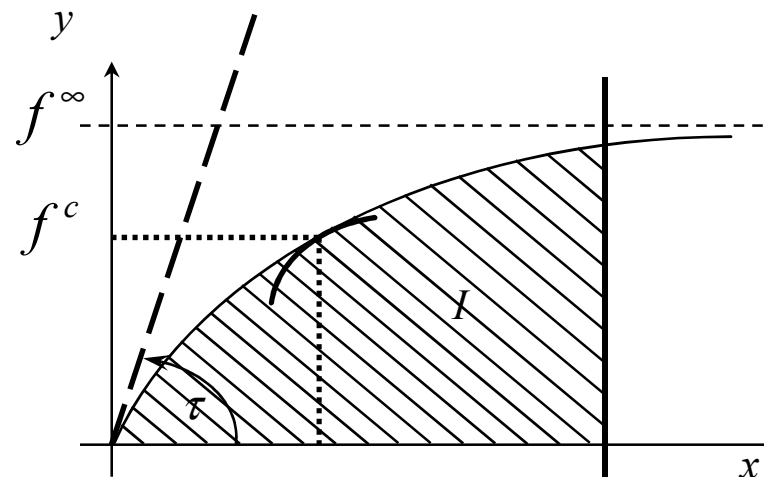
$$m = \mathcal{Q}(f)$$

dérivation, intégration  
extrapolation, interpolation ...



$$m = g(\theta)$$

ex.:  $m = f(\infty) = \frac{\theta_1}{\theta_2}$



$$\wp(\mathbf{e}) \xrightarrow{f, \xi} \wp(\boldsymbol{\theta}) \xrightarrow{g} \wp(m)$$

■ *Différentes approches*

➡ Expression littérale

$$\text{rq : } m = \theta_1 \theta_2 \quad \wp(m) = \frac{e^{cm/(1-c^2)\sigma_1\sigma_2}}{\pi\sigma_1\sigma_2\sqrt{1-c^2}} \int_0^{+\infty} \exp\left[-\frac{1}{2(1-c^2)}\left(\frac{m^2}{\sigma_1^2} \frac{1}{\theta_2^2} + \frac{\theta_2^2}{\sigma_2^2}\right)\right] \frac{d\theta_2}{\theta_2}$$

➡ Estimation des deux premiers moments [G.P.Y. Clarke]  
(biais, variance)

(développement en série de Taylor)

➡ Estimation de la DDP  $\wp(\boldsymbol{\theta})$  [A. Pazman]  
(géométrie différentielle)

➡ Approche Monte-Carlo (PMC, LHS, ... )



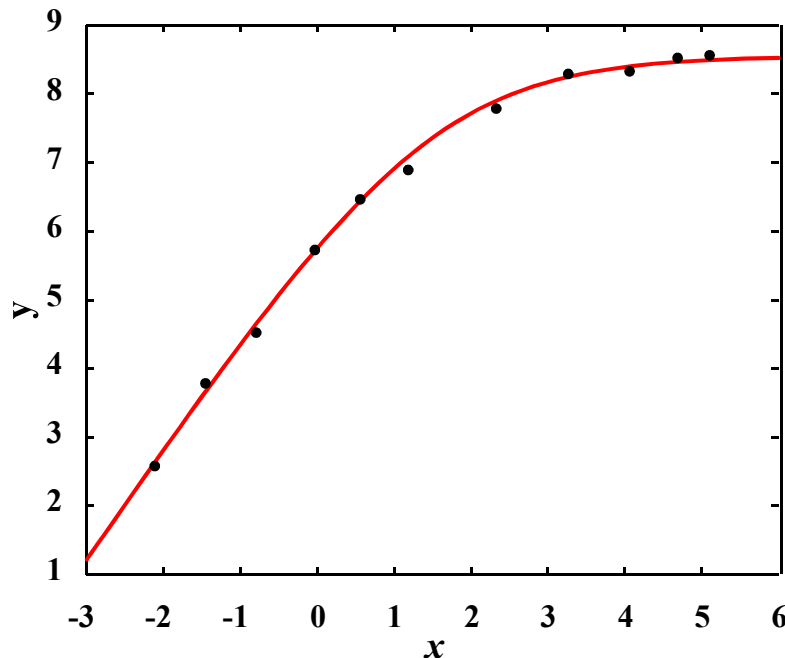
Approximation des deux premiers moments [Clarke]

*biais* 
$$\mathbf{b} = -\frac{\sigma^2}{2} K \begin{bmatrix} \text{tr } B^1 & \dots & \text{tr } B^p \end{bmatrix}^\top \quad (DL3)$$

*matrice de covariance* 
$$V = \sigma^2 K \left( Id + \sigma^2 (V_a + V_b + V_c) \right) K^\top \quad (DL5)$$

$$f(x, \theta) = \theta_4 - \theta_3 \log \left( 1 + \exp \left( -\frac{\theta_1 x + \theta_2}{\theta_3} \right) \right)$$

référence : Monte-Carlo



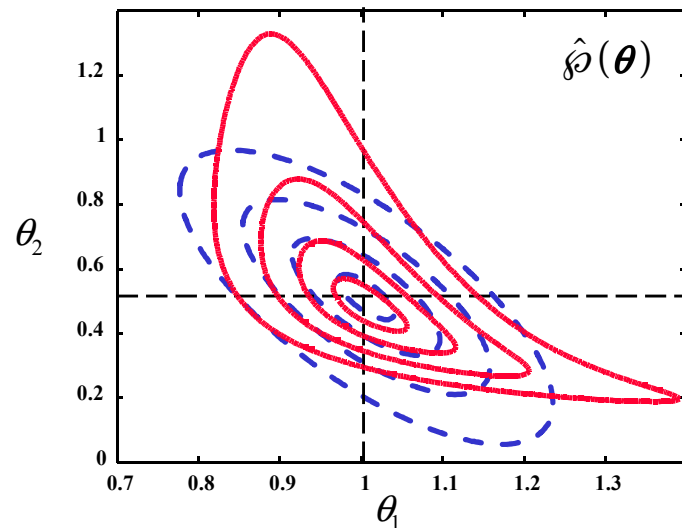
L 
$$\frac{V_{MC} - V_{Lin}}{V_{Lin}} \% = \begin{bmatrix} 16.2 & 27.4 & 17.1 & 14.5 \\ & 39.2 & 28.5 & 30.8 \\ & & 16.4 & 12.1 \\ & & & 5.1 \end{bmatrix}$$

NL 
$$\frac{V_{MC} - V_{NonLin4}}{V_{NonLin4}} \% = \begin{bmatrix} 3.6 & 5.5 & 3.8 & 4.9 \\ & 8.2 & 5.9 & 7.2 \\ & & 4.0 & 3.7 \\ & & & 2.1 \end{bmatrix}$$

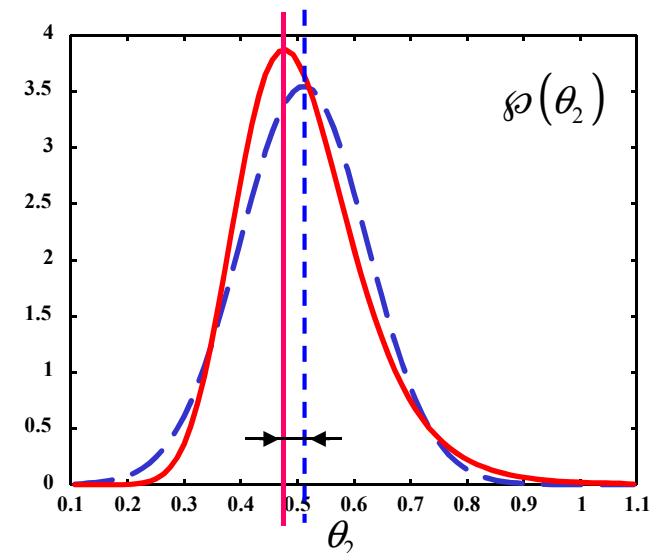
**Géométrie différentielle [Pazman]**

$$\hat{\wp}(\boldsymbol{\theta}) = \frac{\det \mathcal{Q}(\boldsymbol{\theta})}{(2\pi)^{N/2} \det^{1/2} M(\boldsymbol{\theta})} \exp\left(-\frac{1}{2\sigma^2} (f(\mathbf{x}, \boldsymbol{\theta}) - \mathbf{y})^T P(\boldsymbol{\theta})^T P(\boldsymbol{\theta}) (f(\mathbf{x}, \boldsymbol{\theta}) - \mathbf{y})\right)$$

$$f(x, \boldsymbol{\theta}) = \theta_1 (1 - \exp(-\theta_2 x))$$



*Marginalisation*



- Modélisation des incertitudes
  - approches paramétriques
  - approches non paramétriques
- Choix d'un modèle de système
  - choix de la complexité d'un modèle
  - choix d'un modèle de comportement
  - choix de la paramétrisation d'un modèle
  - vers des méta-modèles
- Aides à la prise de décision
  - Estimation d'évènements rares
  - Optimisation d'un système complexe
- Exemples d'application

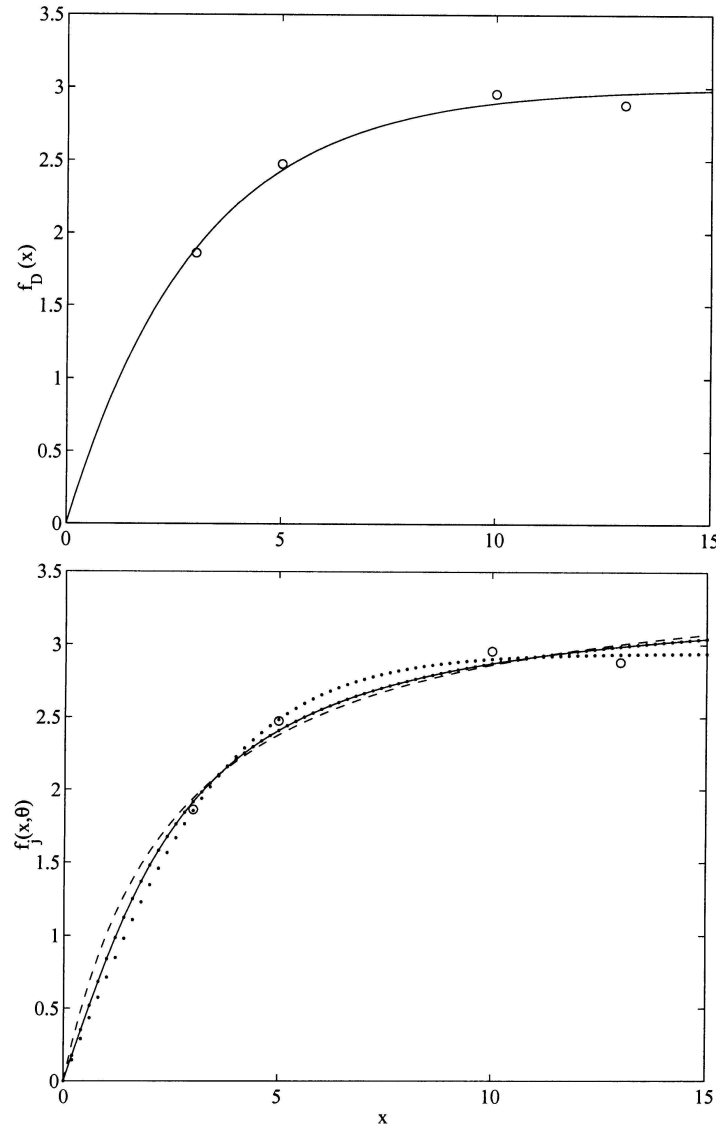
■ *Choix de modèle emboîtés*

- Nombreux critères dans la littérature (*choix de l'ordre*)
  - . VC
  - . AIC, MDL
  - . BIC, MAP
  - . KICc (optimalité non asymptotique) [Bekara]

■ *Choix de modèles quelconques*

- Critère de choix lié à un objectif donné
  - . Critères de type moindres carrés
  - . Distance entre densités de probabilité

### ■ Choix de structure de modèle



- « vrai » modèle (*inconnu*)

$$f_D(x) = \theta_1 (1 - \exp(-\theta_2 x))$$

- modèles candidats

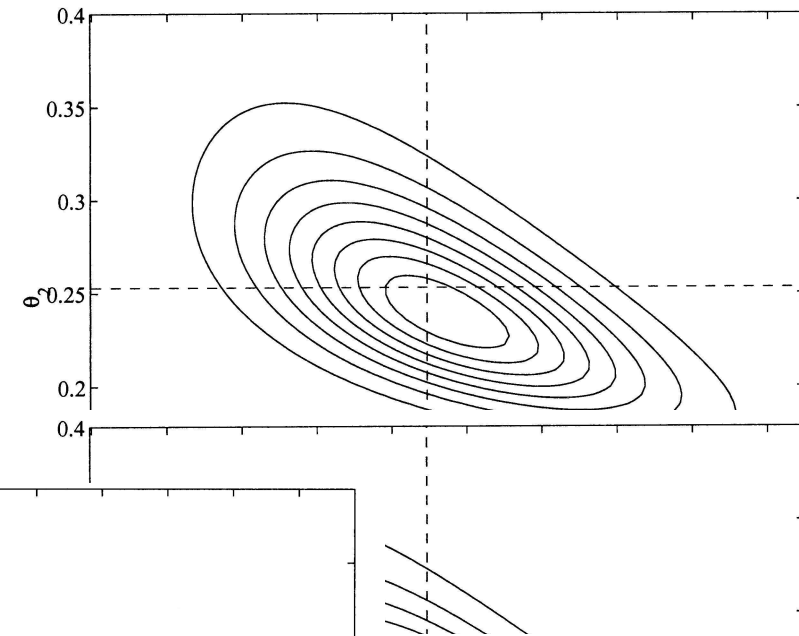
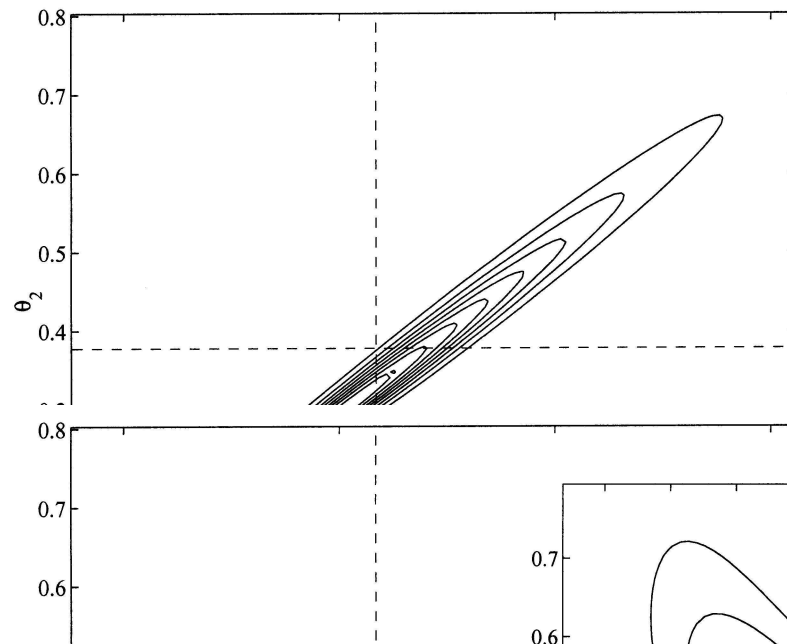
$$f_1(x) = \frac{\theta_1 x}{\theta_2 x + 1}$$

$$f_2(x) = \theta_1 \tanh(\theta_2 x)$$

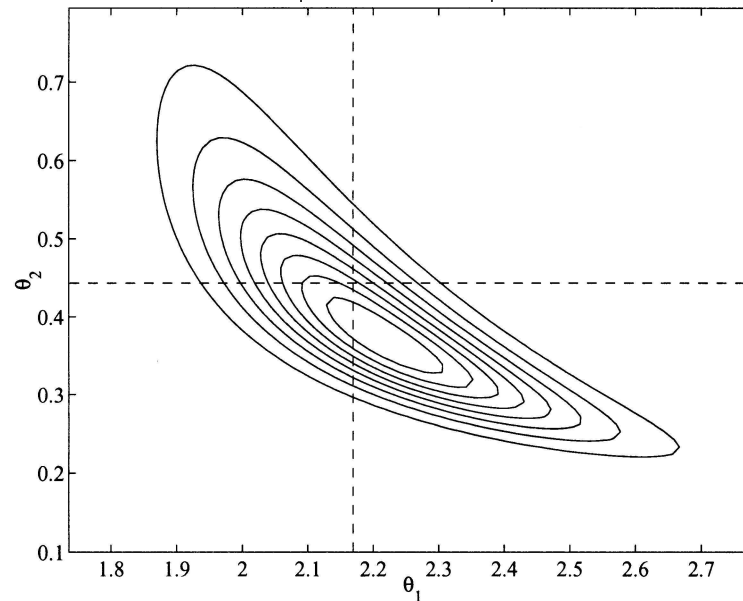
$$f_3(x) = \theta_1 \arctan(\theta_2 x)$$

	$f_1$	$f_2$	$f_3$
$MSE$	1.60	3.60	0.68

### ■ Choix de structure de modèle



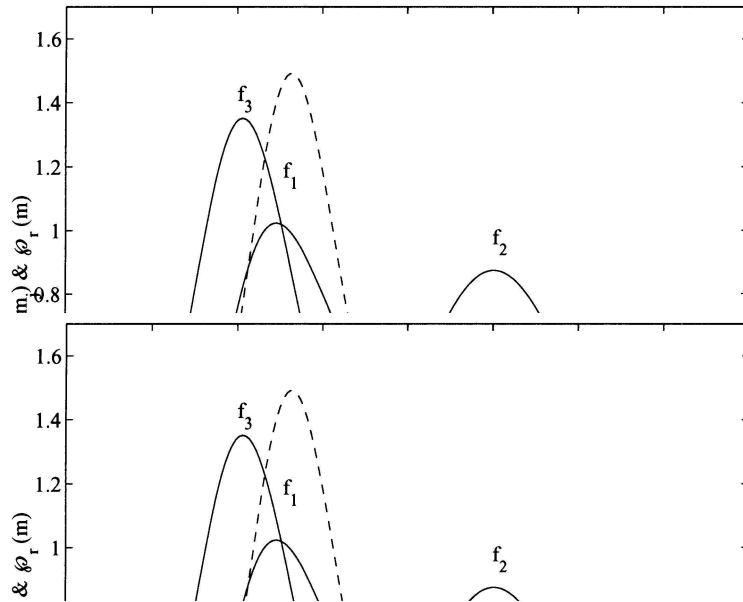
ddp des paramètres  
(Pazman) :  $\hat{\phi}_j(\hat{\theta})$



→ HPC

### ■ Choix de structure de modèle

1) ddp de l'objectif  $\hat{\phi}_j(m)$

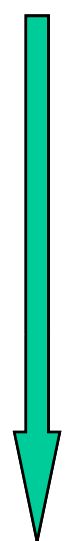


2) distance entre ddps  $\Delta_k$

	$f_1$	$f_2$	$f_3$
$\Delta_1$	0.317	1.233	0.688
$\Delta_2$	0.514	1.473	0.888
$\Delta_3$	0.095	1.377	0.258

	$m_S$	$m_A$	$m_I$
$f_1$	$\theta_1$	$\theta_1/\theta_2$	$\theta_1/\theta_2^2 (\theta_2 x_I - \log(1 + \theta_2 x_I))$
$f_2$	$\theta_1 \theta_2$	$\theta_1$	$\theta_1/\theta_2 \log(\cosh(\theta_2 x_I))$
$f_3$	$\theta_1 \theta_2$	$\theta_1 \pi/2$	$\theta_1 x_I \arctan(\theta_2 x_I) - \theta_1/2 \theta_2 \log(1 + \theta_2^2 x_I^2)$
	1	3	2

■ Choix de la paramétrisation d'une structure

	structure	$\frac{\bullet x}{\bullet x + \bullet}$	Code de génération formelle de modèles
	structure paramétrée	$\frac{\theta_1 x}{\theta_2 x + 1}$	Choix de la paramétrisation
	fonction	$\frac{3x}{4x + 1}$	Procédure d'estimation / d'identification

Pourquoi choisir  $f_\theta(x) = \frac{\theta_1 x}{\theta_2 x + 1}$  plutôt que  $f_\varphi(x) = \frac{x}{\varphi_1 x + \varphi_2}$  ?



→ Pour optimiser la phase d'identification des paramètres !

$$\text{reparamétrisation : } \theta = t(\varphi)$$

- *Influence sur la mesure*

- $t$  homéomorphisme
- $m$  dépend d'un  $\theta_i$



$$\mathcal{P}(m_\theta) = \mathcal{P}(m_\varphi)$$

“Optimal Nonlinear Modeling and Reparameterization”, *IEEE WISP*, 1999.

- *Comment optimiser la reparamétrisation*

- critère d'optimalité ?
- $\theta \longleftrightarrow ? \varphi$

Critère

critère de non-linéarité  $N_\theta$

- sources de non-linéarités :
- le lieu des solutions
  - le maillage de ce lieu

$$\text{var}(\hat{\theta}) = \sigma^2 K \left( Id + \underbrace{\sigma^2 V_a}_{\text{non-linéarité intrinsèque}} + \underbrace{\sigma^2 (V_b + V_c)}_{\text{paramétrisation}} \right) K^T$$

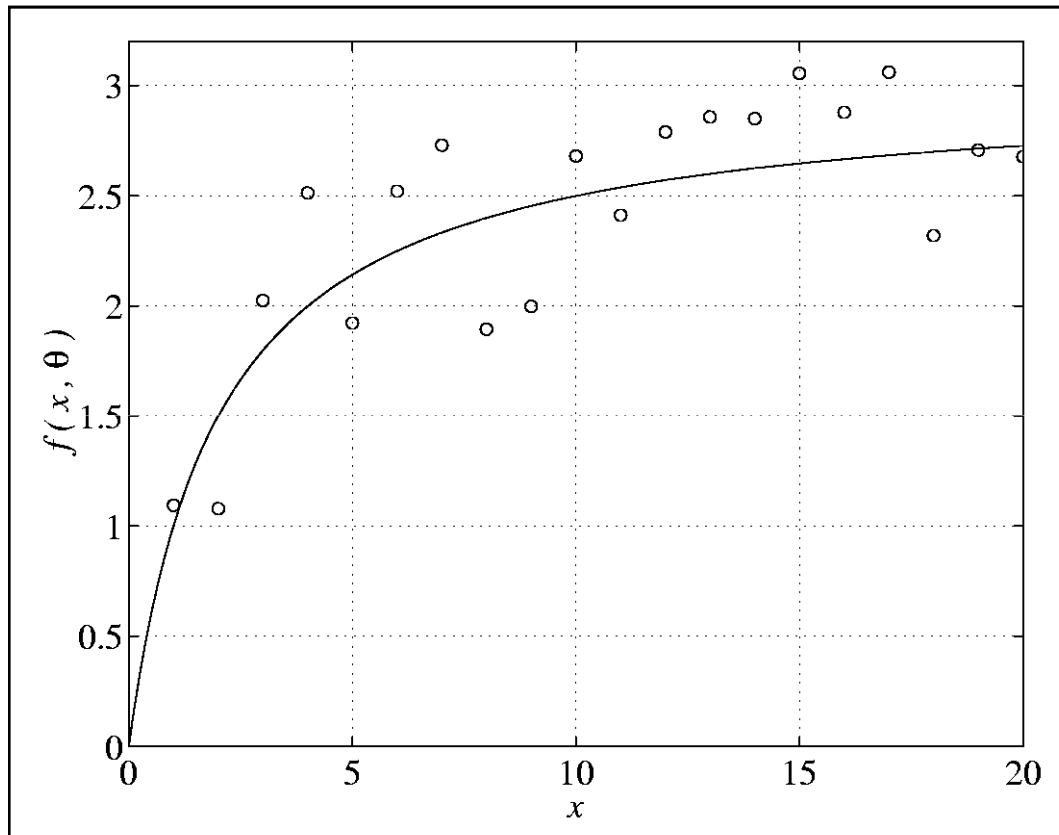
*non-linéarité intrinsèque*

*paramétrisation*

$\theta \longleftrightarrow ? \longrightarrow \varphi$

- linéaire  $\varphi = T\theta$   $\rightarrow$  inutile
- optimale  $\{\varphi_i = f(\pi_i, \theta)\}_{i \in [1, p]}$   $\rightarrow$  complexe
- quadratique  $\theta = AS \text{diag}[\Phi^2]$   $\rightarrow$  compromis

### ■ Exemple

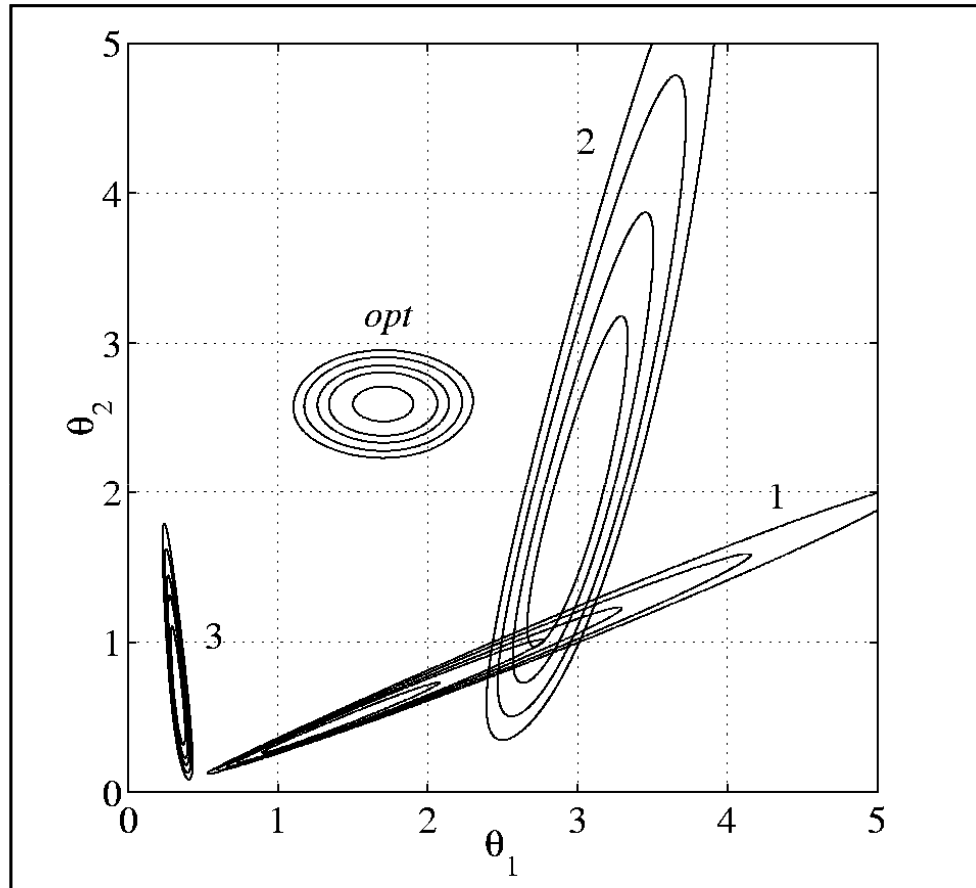


$$f(x, \theta) = \frac{\theta_1 x}{\theta_2 x + 1}$$

$$\begin{cases} \theta_0 = [3/2 & 1/2]^\top \\ \mathbf{x} = [1 & \dots & 20]^\top \\ \sigma^2 = 0.1 \end{cases}$$



$$f_\varphi(x) = \frac{(\pi_1 - \pi_2)\varphi_1\varphi_2 \cdot x}{(\pi_1\varphi_2 - \pi_2\varphi_1) \cdot x + \pi_1\pi_2(\varphi_1 - \varphi_2)}$$



densités de probabilité

structures	$N_{\theta}$	$ \rho $	$\bar{n}_{it}$	$\sigma_{\max}$
$\theta_1 x / (\theta_2 x + 1)$	4.304	0.997	16.2	X
$\theta_1 x / (x + \theta_2)$	1.408	0.969	13.9	6.426
$x / (\theta_1 x + \theta_2)$	0.797	0.808	12.0	0.634
quadratique	0.263	0.203	11.6	0.512
optimale	0.201	0.078	10.6	0.415

optimisation :

- *décorrélation*
- *temps de calcul*
- *robustesse numérique*

■ *Littérature*

⇒ Méthodes paramétriques ou non-paramétriques

⇒ Deux points de vue **équivalents** :

□ **Régression à noyau reproduisant**

- 1960 : splines, (**Schoenberg 1964, Duchon 1976–1979**)
- 1980 : RBF, (**Micchelli 1986, Powel 1987**)
- 1995 : SVM, (**Vapnik 1995**)
- 1997 : SVR, (**Smola 1997**)
- 1999 : SVR semi-paramétrique (**Smola 1999**)

□ **prédiction linéaire – Krigeage**

- 1950 : prédiction pour la recherche minière (**Krige 1951**)
- 1960 : krigeage, géostatistique (**Matheron 1963**) – École des Mines
- 1970 : krigeage intrinsèque (**Matheron 1971**)
- 1997 : prédiction par « processus gaussiens », (**Williams 1997, Neal 1997**)

■ *Krigeage* [Vazquez]

$$y_i = f(x_i, \theta) + e_i$$

$$y_i = f(x_i, \theta) + e_i + \omega(x_i)$$

erreur de modélisation

→ processus gaussien

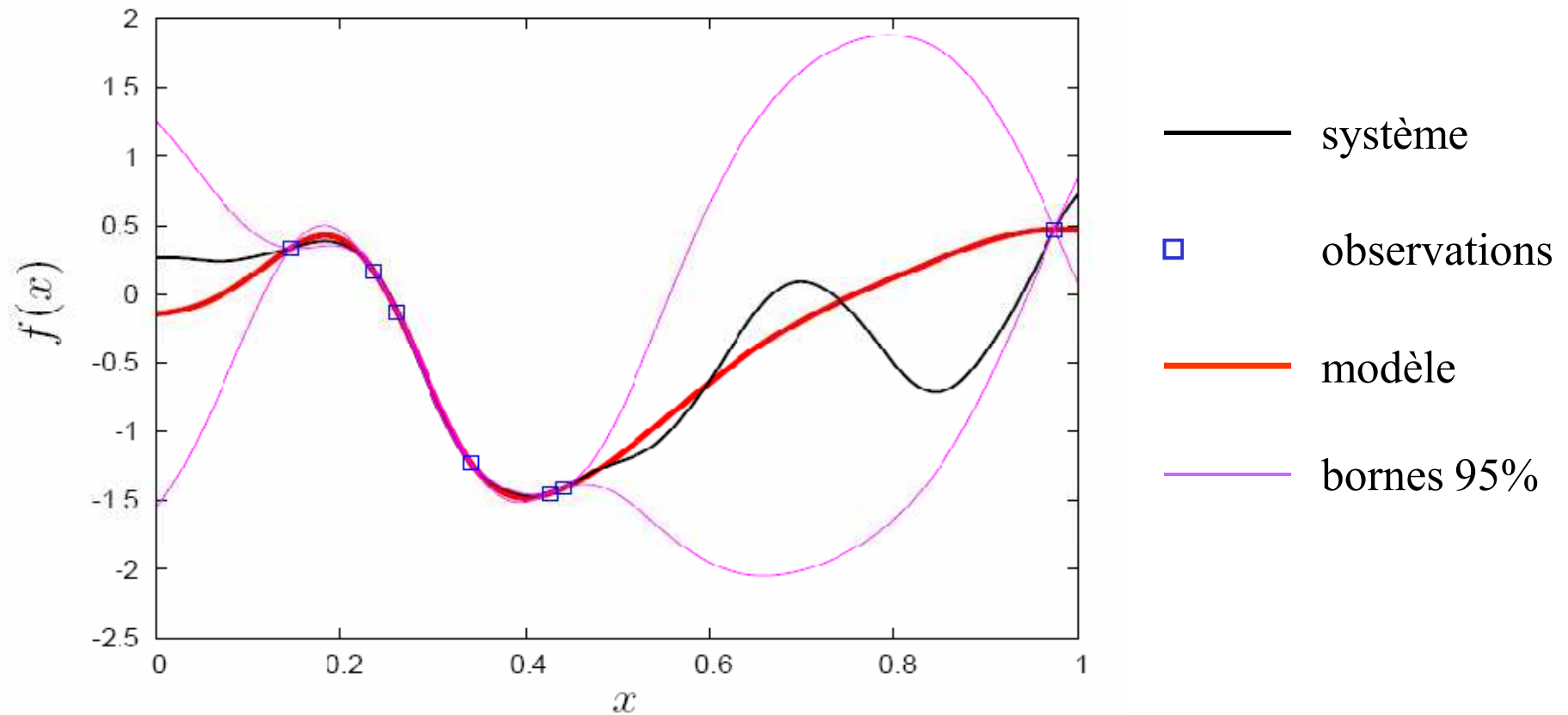
- moyenne nulle
- covariance  $k(x_i, x_j) = E\{\omega(x_i)\omega(x_j)\}$

→ modèle : structure de la covariance

$$k(x_i, x_j) = k(x_i - x_j) = k(h)$$

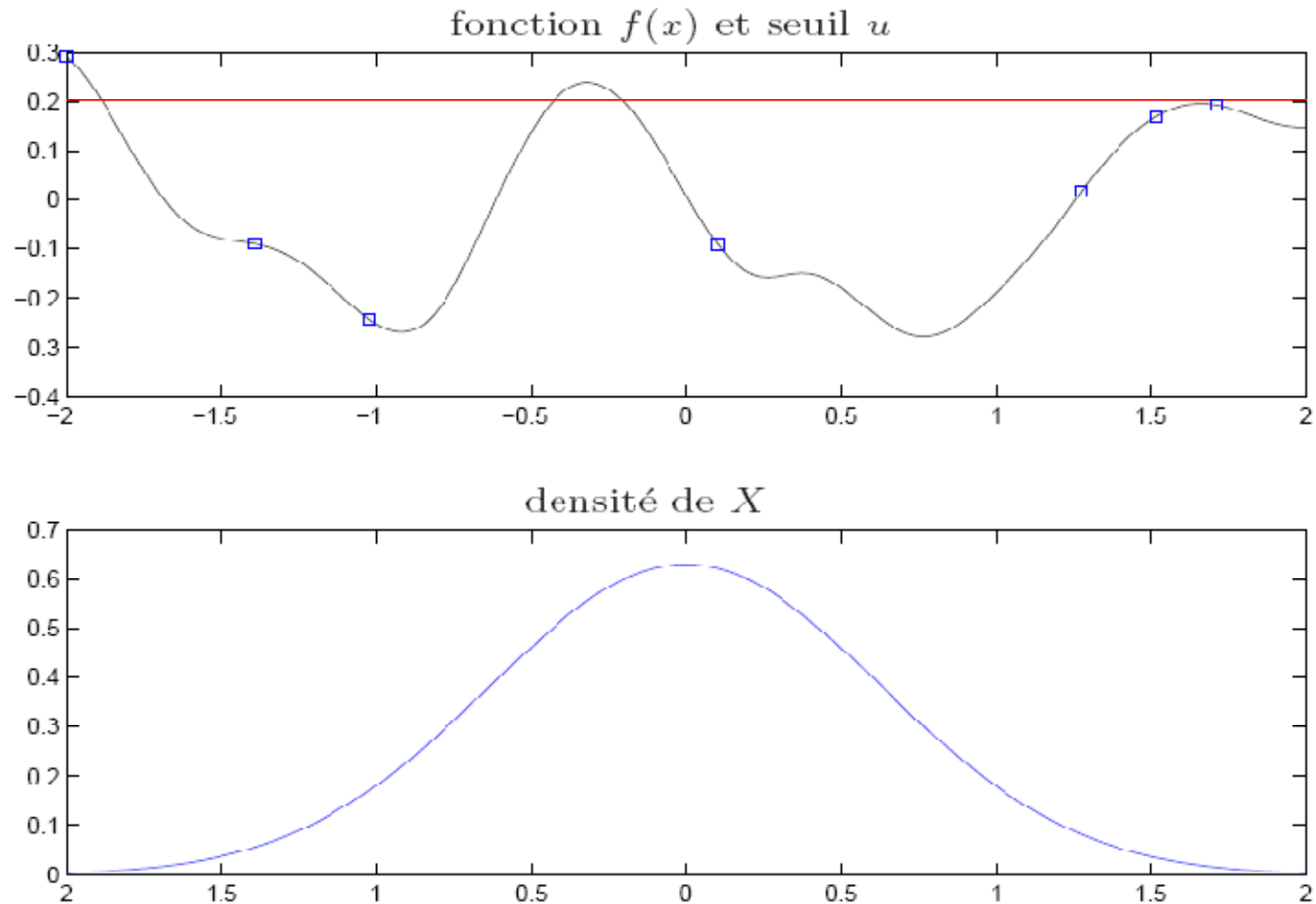
$$\text{ex. } k(h) = \exp(-h^2/\rho^2) \dots$$

■ *Exemple*

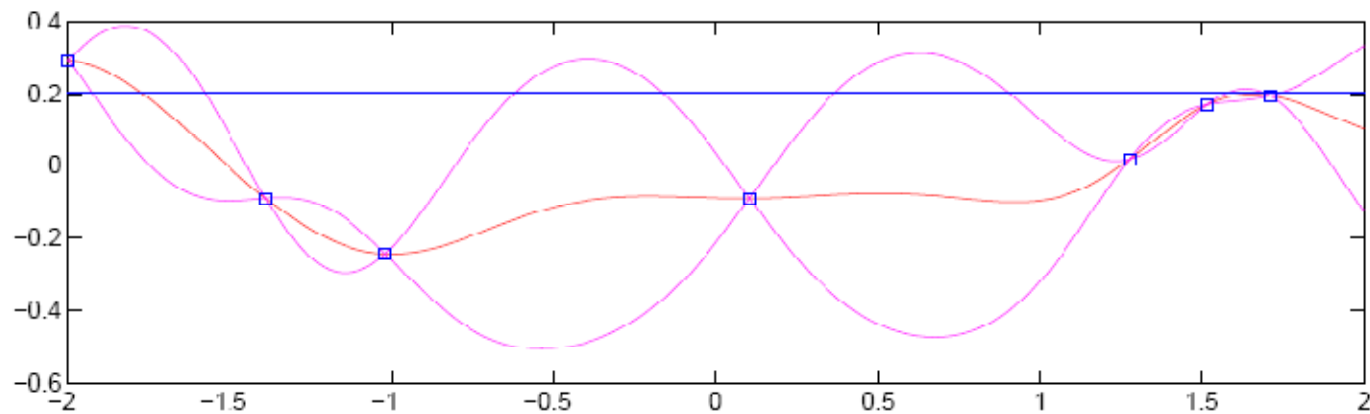


- Modélisation des incertitudes
  - approches paramétriques
  - approches non paramétriques
- Choix d'un modèle de système
  - choix de la complexité d'un modèle
  - choix d'un modèle de comportement
  - choix de la paramétrisation d'un modèle
  - vers des méta-modèles
- Aides à la prise de décision
  - Estimation d'évènements rares
  - Optimisation d'un système complexe
- Exemples d'application

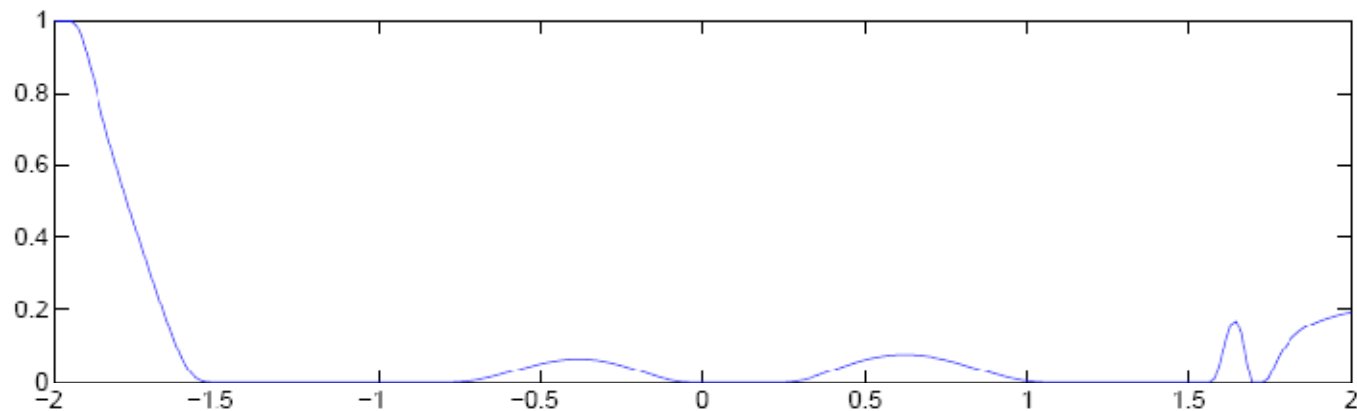




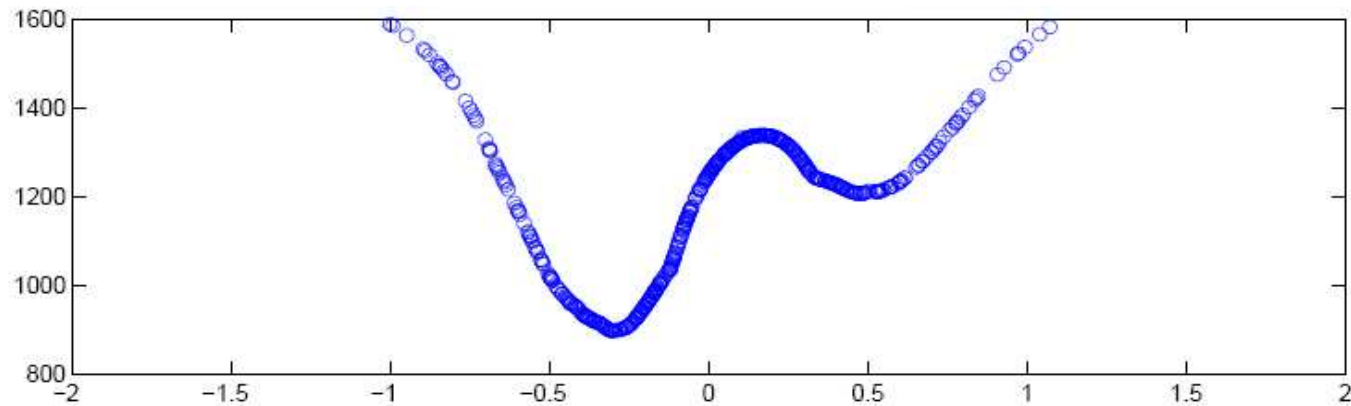
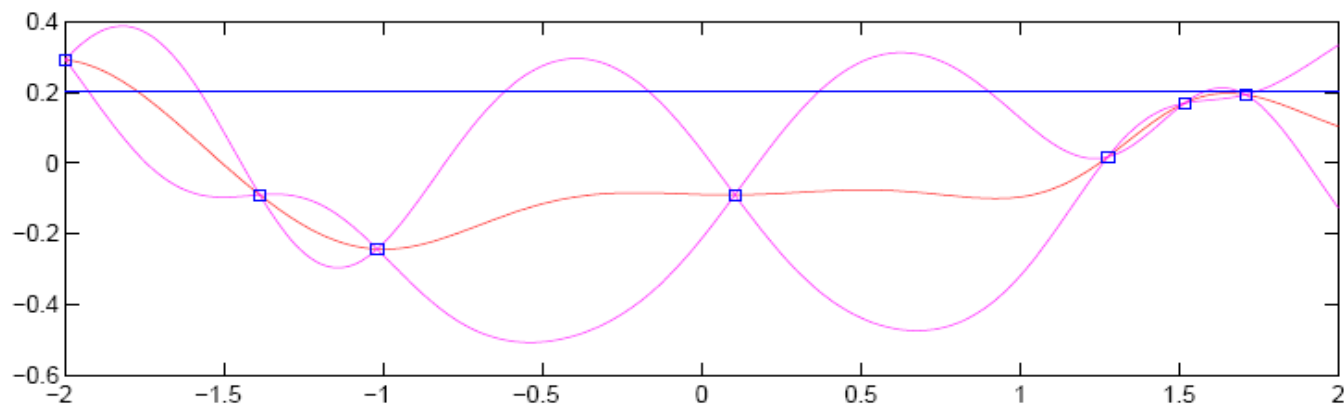
Prédiction et intervalles de confiance à 95%



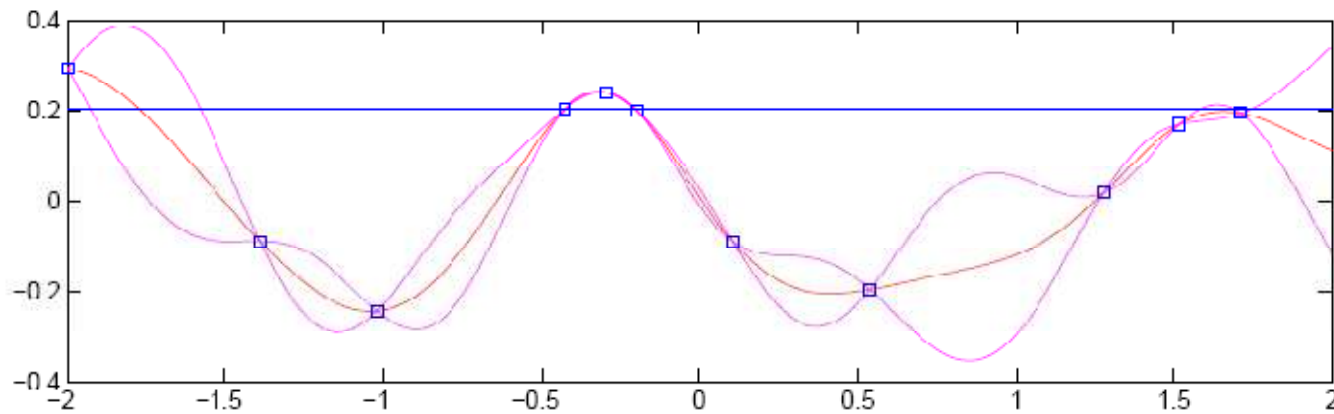
Probabilité de dépassement de seuil



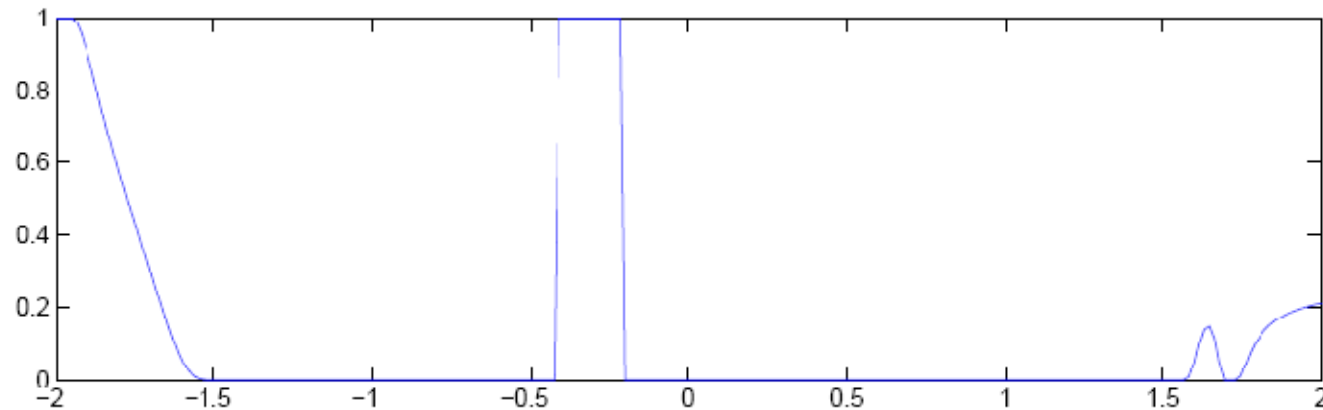
Prédiction et intervalles de confiance à 95%



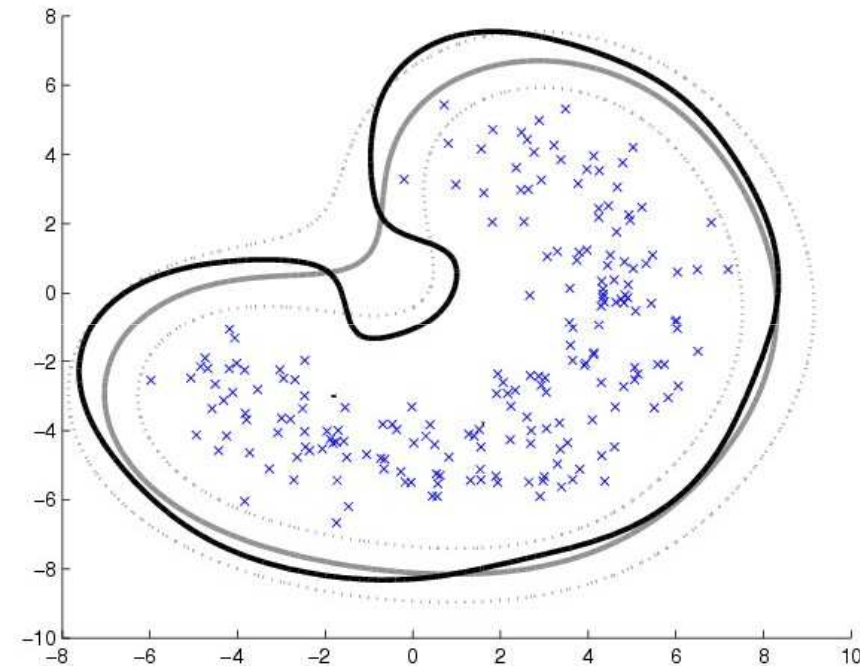
Prédiction et intervalles de confiance à 95%



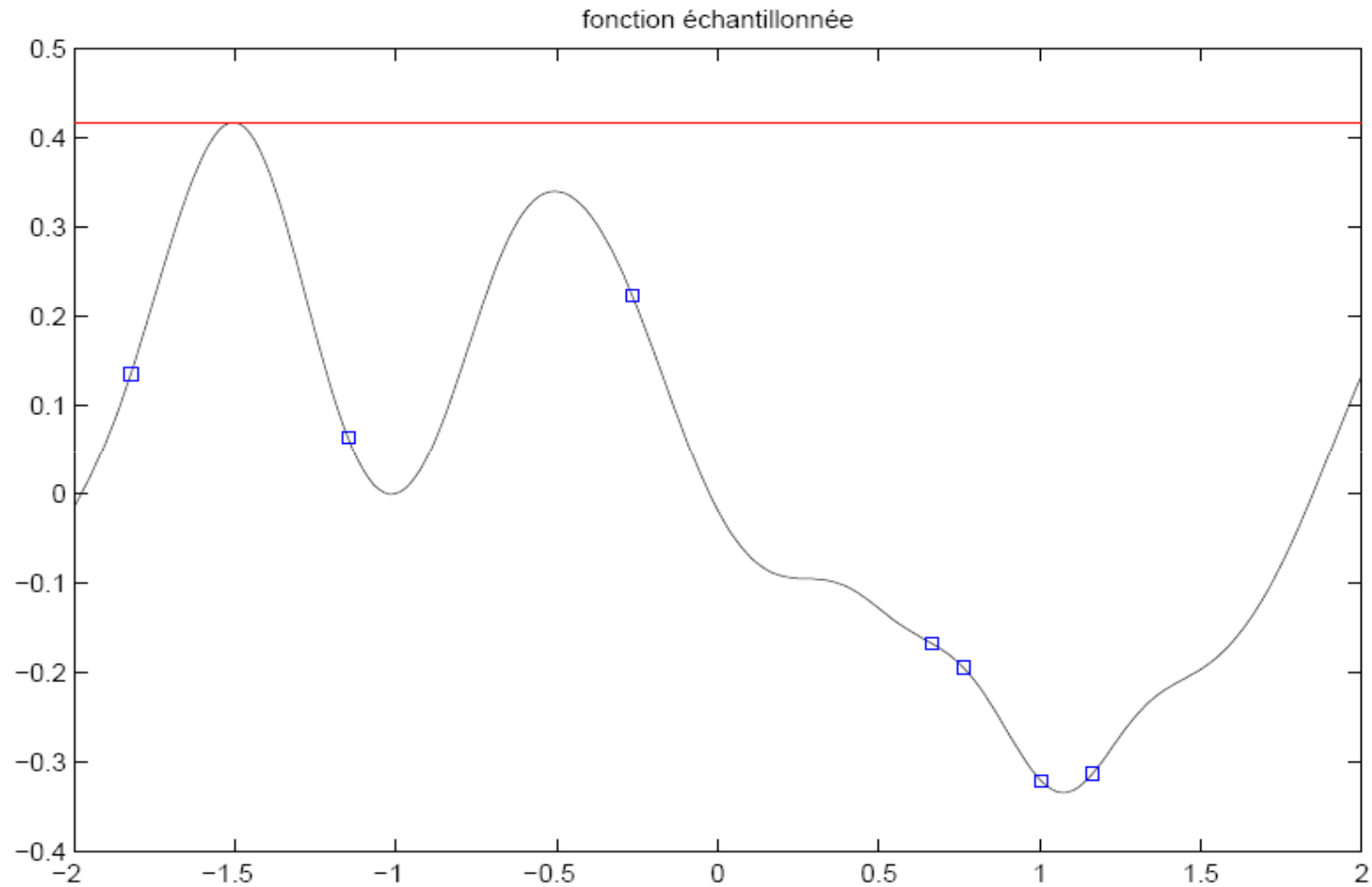
Probabilité de dépassement de seuil

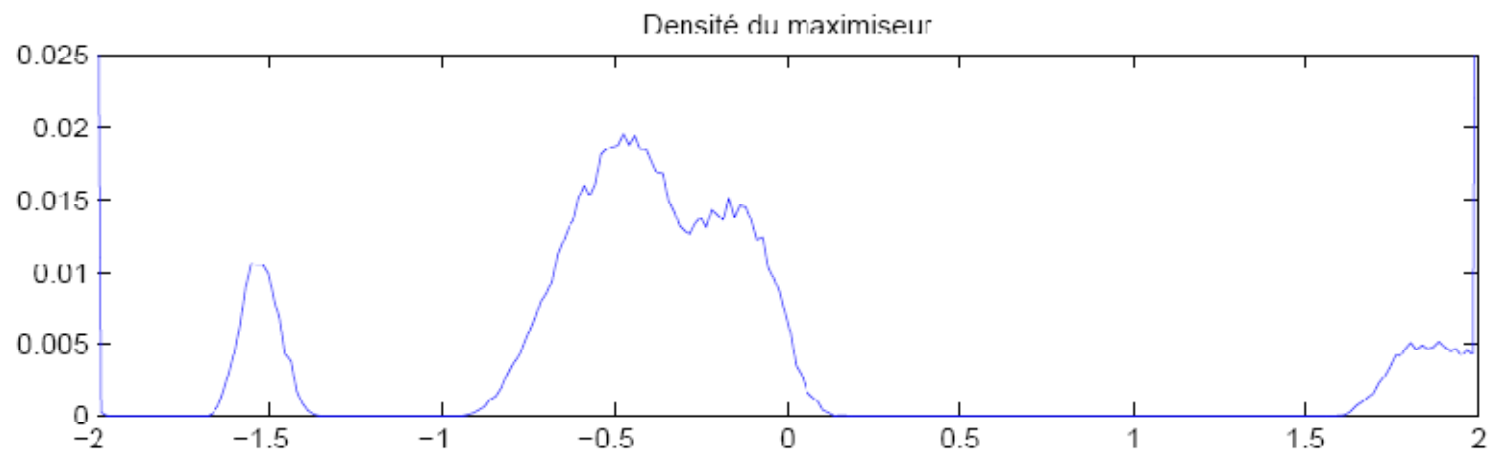
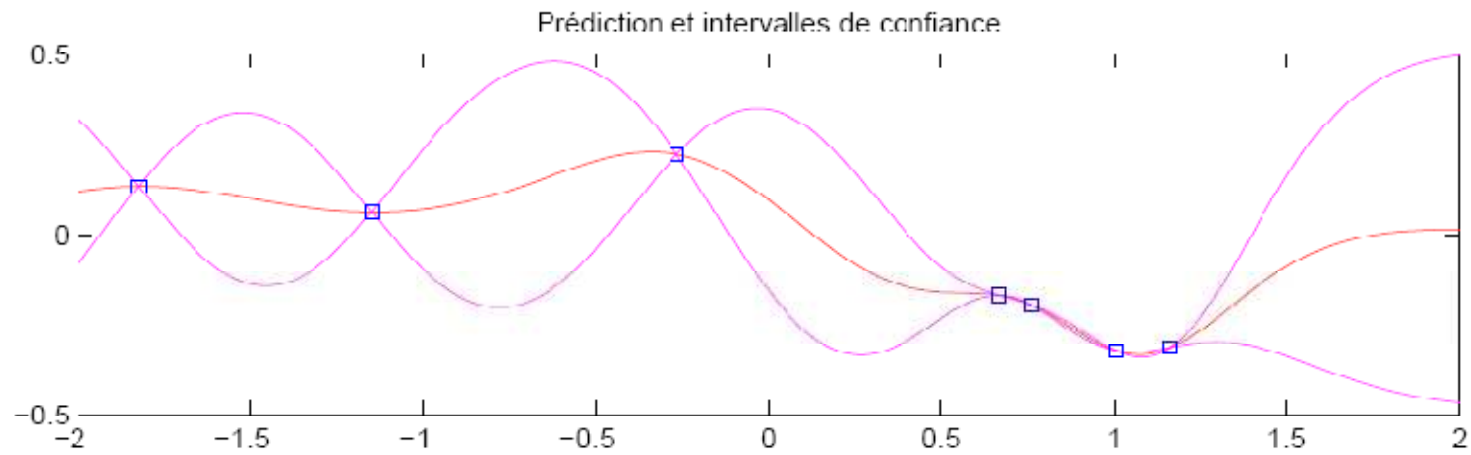


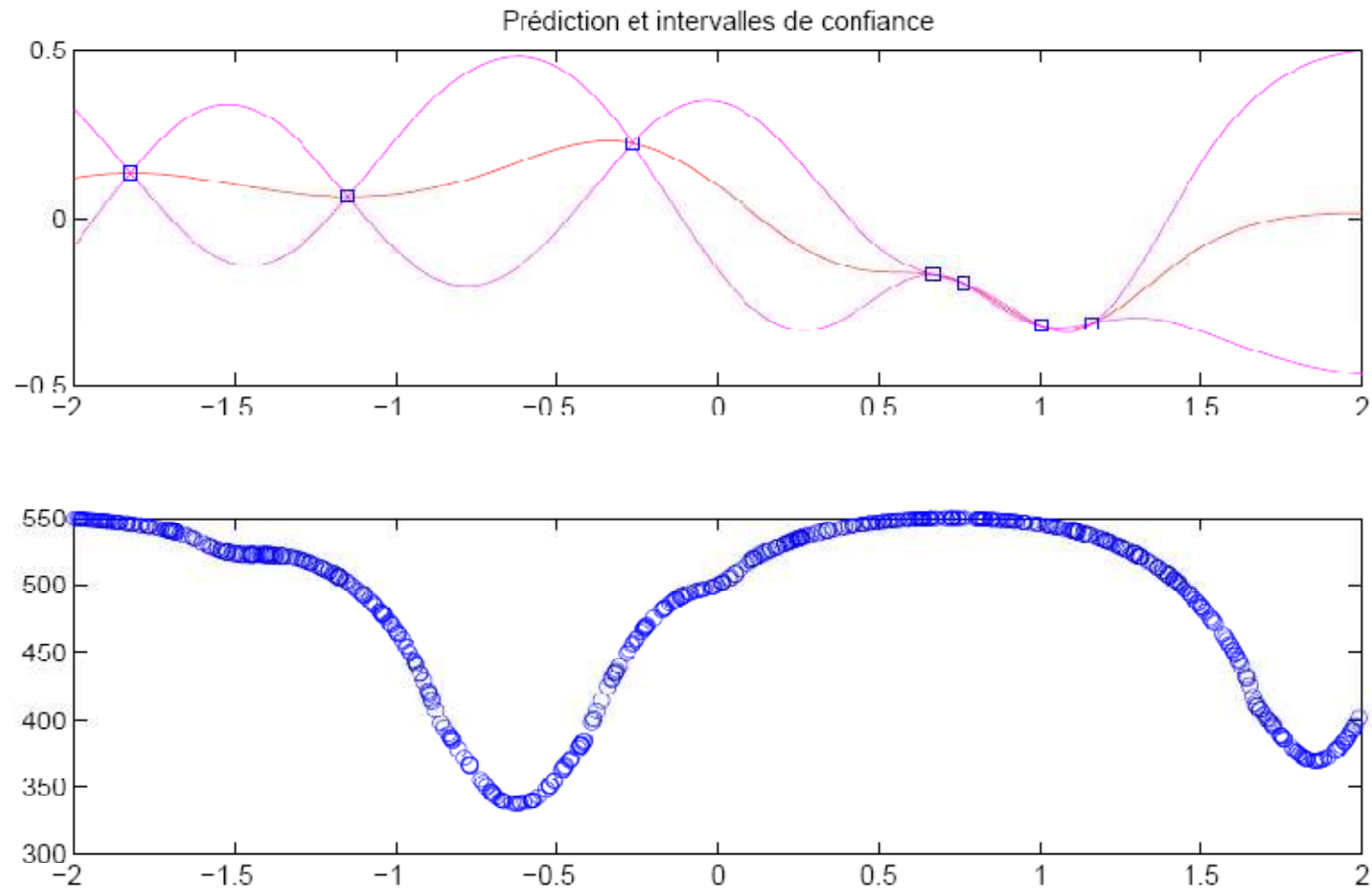
*Estimation d'un quantile à 99.9%  
à partir de 200 données par classification  
SVM et utilisation de la théorie statistique  
des valeurs extrêmes*



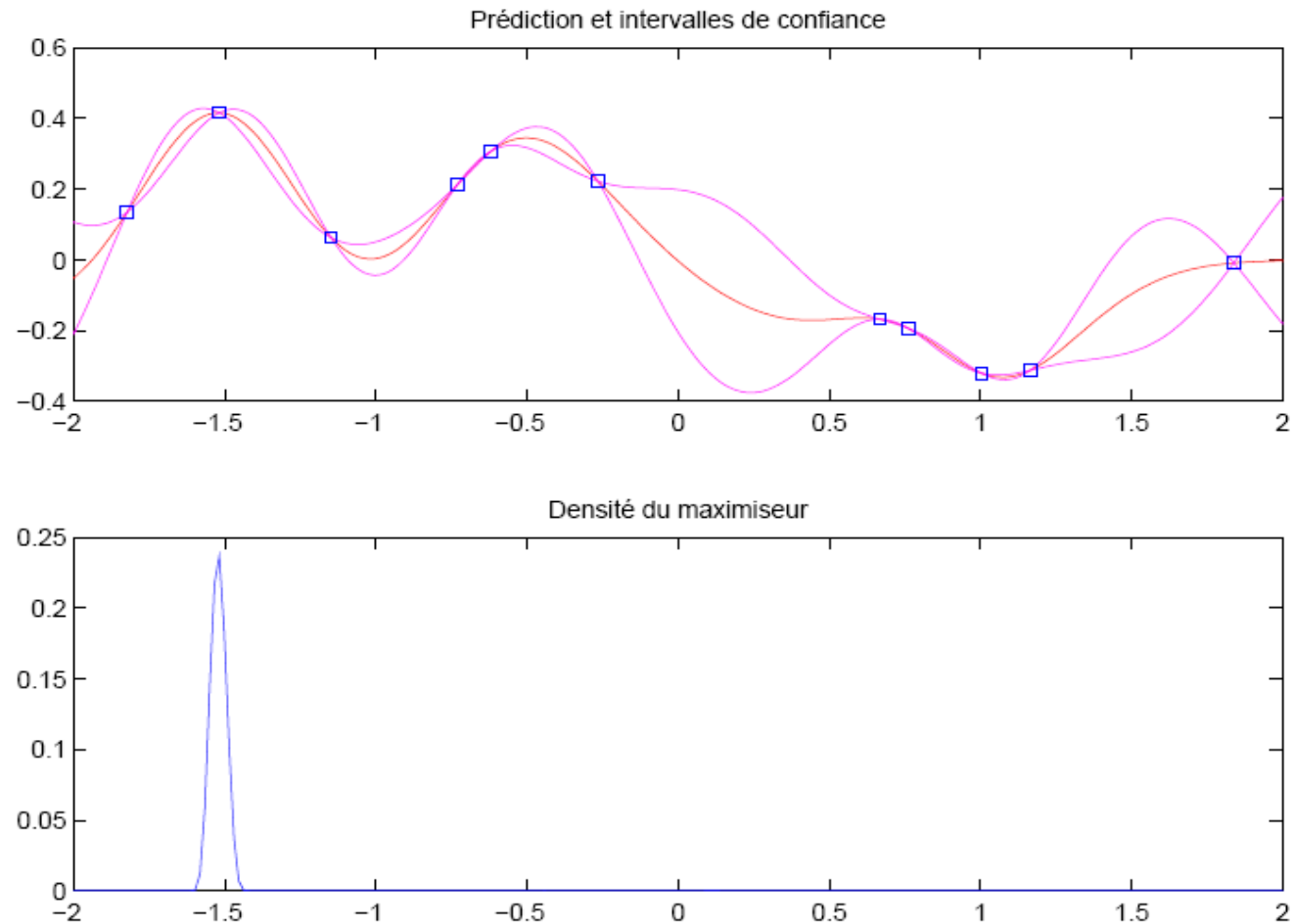
Détermination de quantiles multidimensionnels extrêmes et détection d'*outliers*  
[M. Piera-Martinez et E. Vazquez (2006)]







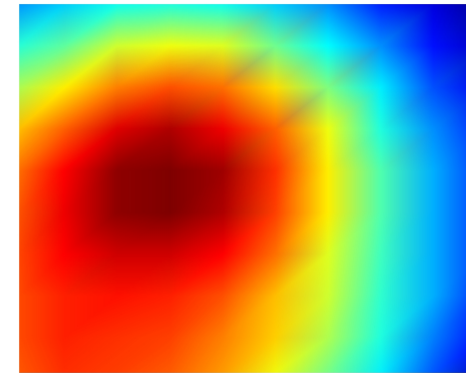
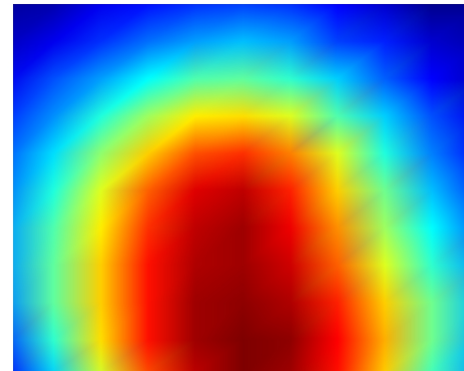




Informational Approach to Global Optimization **IAGO**  
[J. Villemonteix et E. Vazquez (2007)]

- Modélisation des incertitudes
  - approches paramétriques
  - approches non paramétriques
- Propagation des incertitudes
  - par approximations
  - par simulations
- Choix d'un modèle de système
  - choix d'un modèle de comportement
  - vers des méta-modèles
- Aides à la prise de décision
  - Estimation d'évènements rares
  - Optimisation d'un système complexe
- Exemples d'application

■ *Estimation de SAR (GSM)*



*Données :*

- champ électromagnétique mesuré  
(norme : 3D, **600 points**)
- 70 portables testés

*Méthodes :*

- réseaux de neurones
- **modélisation paramétrique** (11 ddl) → **20 points**

*Conclusion :*

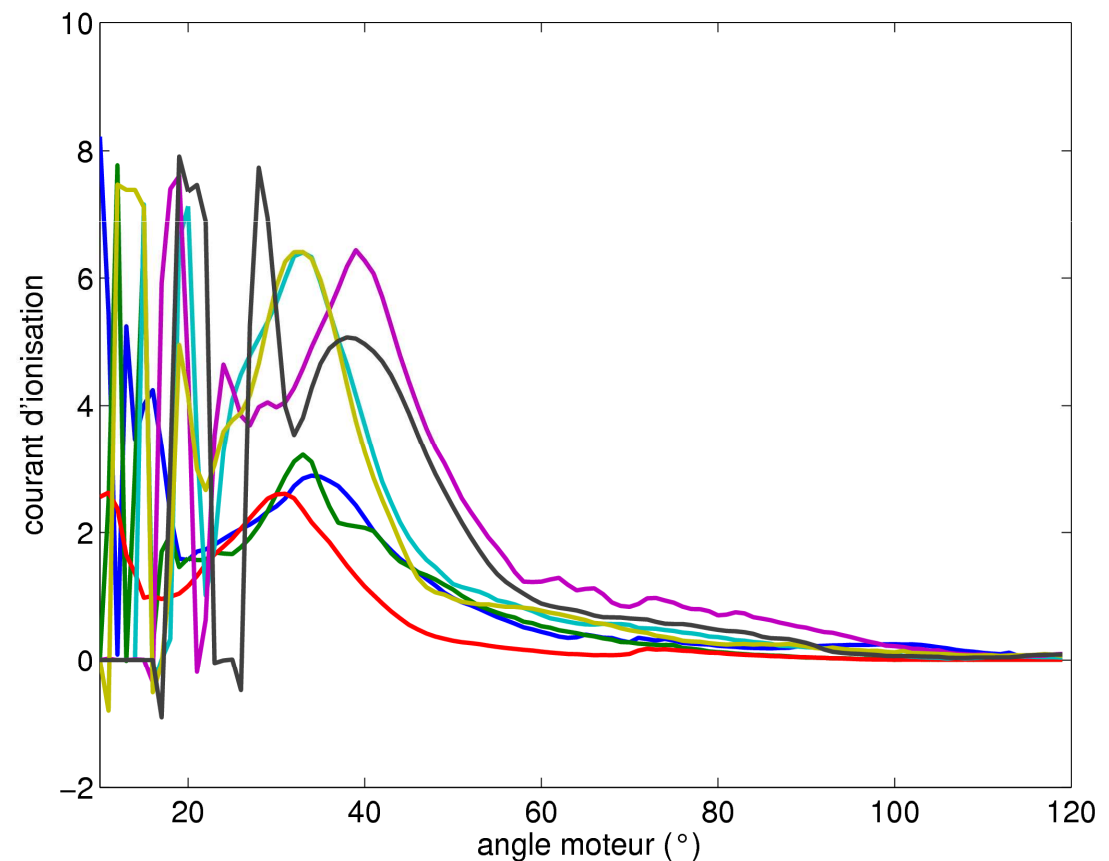
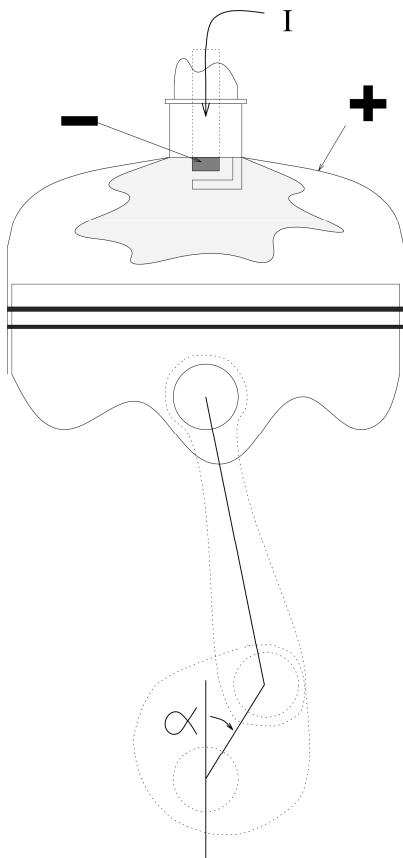
estimation du SAR avec une très bonne précision

### ■ Estimation de l'angle vilebrequin ( $P_{max}$ )

Problématiques de détection / estimation / classification / inversion

Estimation d'une grandeur d'intérêt  $y$  inaccessible directement

On dispose cependant de grandeurs  $x$  contenant de l'information sur  $y$



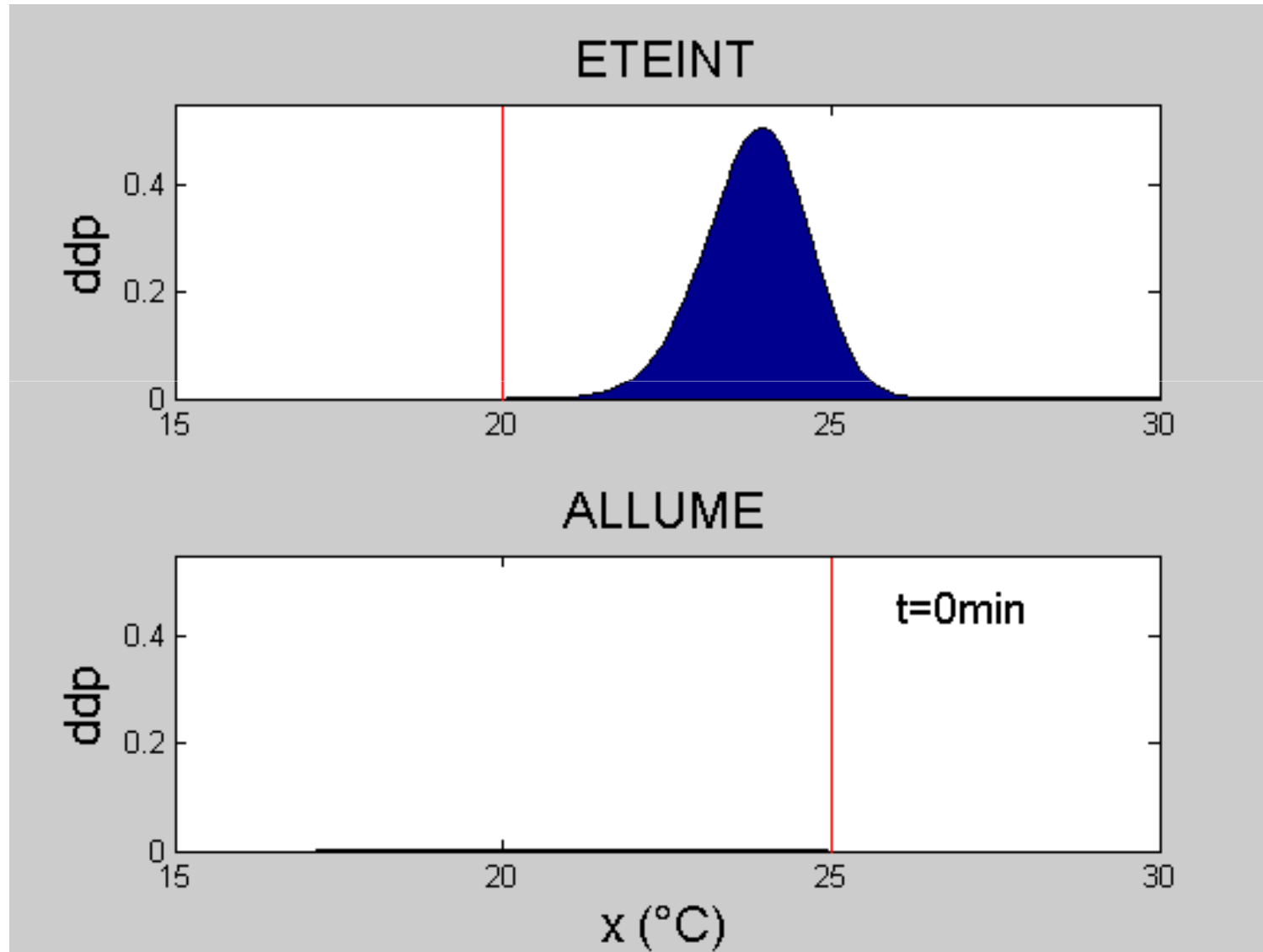
■ *Optimisation d'un conduit d'admission*



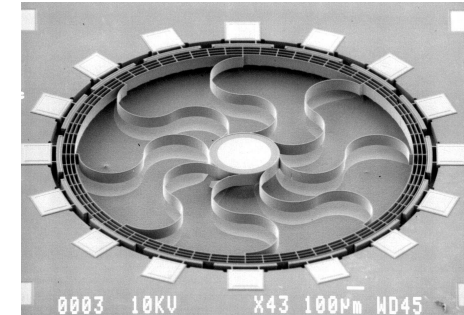
conduit d'admission

- Forme du conduit paramétrée (5 à 20 facteurs)
- Importance de la forme du conduit :
  - impact sur la performance moteur (débit d'air)
  - impact sur les émissions polluantes (turbulences)
- 1 simulation = 5 à 24 heures de calcul

■ *Systèmes hybrides stochastiques (thermostat)*



- L'aléatoire est assez incontournable du fait
  - de la *méconnaissance*
  - de *perturbations* de l'environnement
  - de la *miniaturisation* des composants
  - de la *complexité* croissante des systèmes
- L'aléatoire est un modèle qui permet
  - d'éviter la *surparamétrisation*
  - d'éviter de surestimer des *marges de sécurité*
  - de quantifier la *robustesse* de systèmes
- De nombreux travaux en cours
  - recherche
  - applications
  - regroupement de partenaires... (MASCOTNUM, ISRI, CSDL, ...)



**MERCI**  
**... questions**