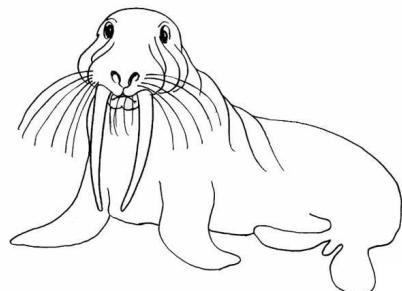


Engineering Applications of POT, GeV and Compound Poisson Models : A Unified approach via Bayesian Thinking

Eric Parent, Jacques Bernier & Jean-Jacques Boreux



UMR 518 Math. Info. App.
ENGREF/INRA/INAPG

équipe MOdélisation, Risque, Statistique, Environnement

*Statistical Modeling of Extreme in Data Assimilation and Filtering Approaches
Strasbourg , 23-26 June, 2008*

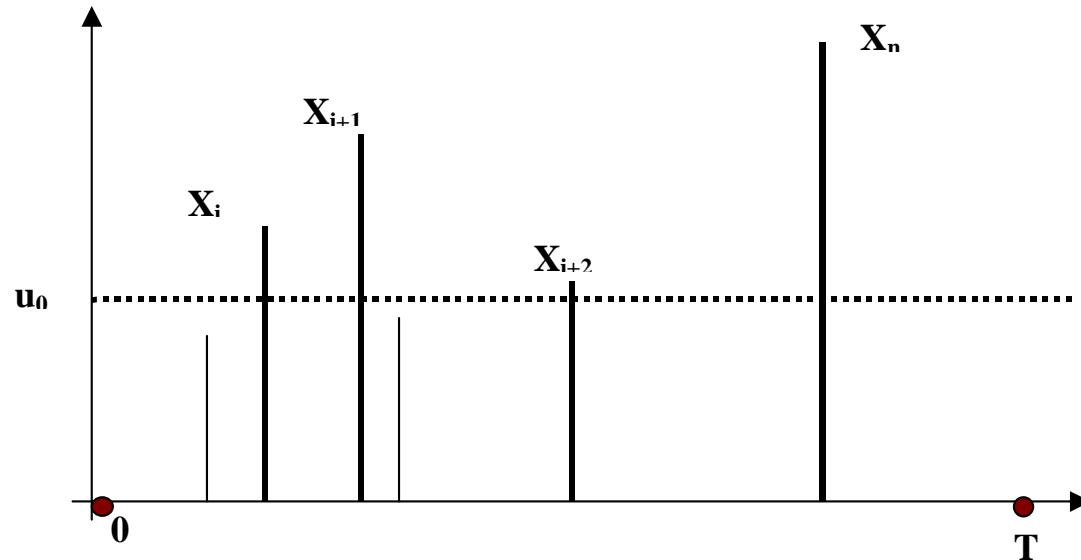
Eric.Parent@agroparistech.fr

Jacques.bernier2@wanadoo.fr

Some nice things to do with Exponentially marked Poisson process

- Application 1 : Fully Observed , Pickands' POT!
- Application 2 : Max Observed, GeV (Gumbel)
- Application 3 : Sum Observed instead of max
- Application 4 : App3+Go Hierarchical

Two parameters only for the trajectory of a marked Poisson Process



$$[N_T = n \mid \mu, T] = \exp(-\mu T) \frac{(\mu T)^n}{n!}$$

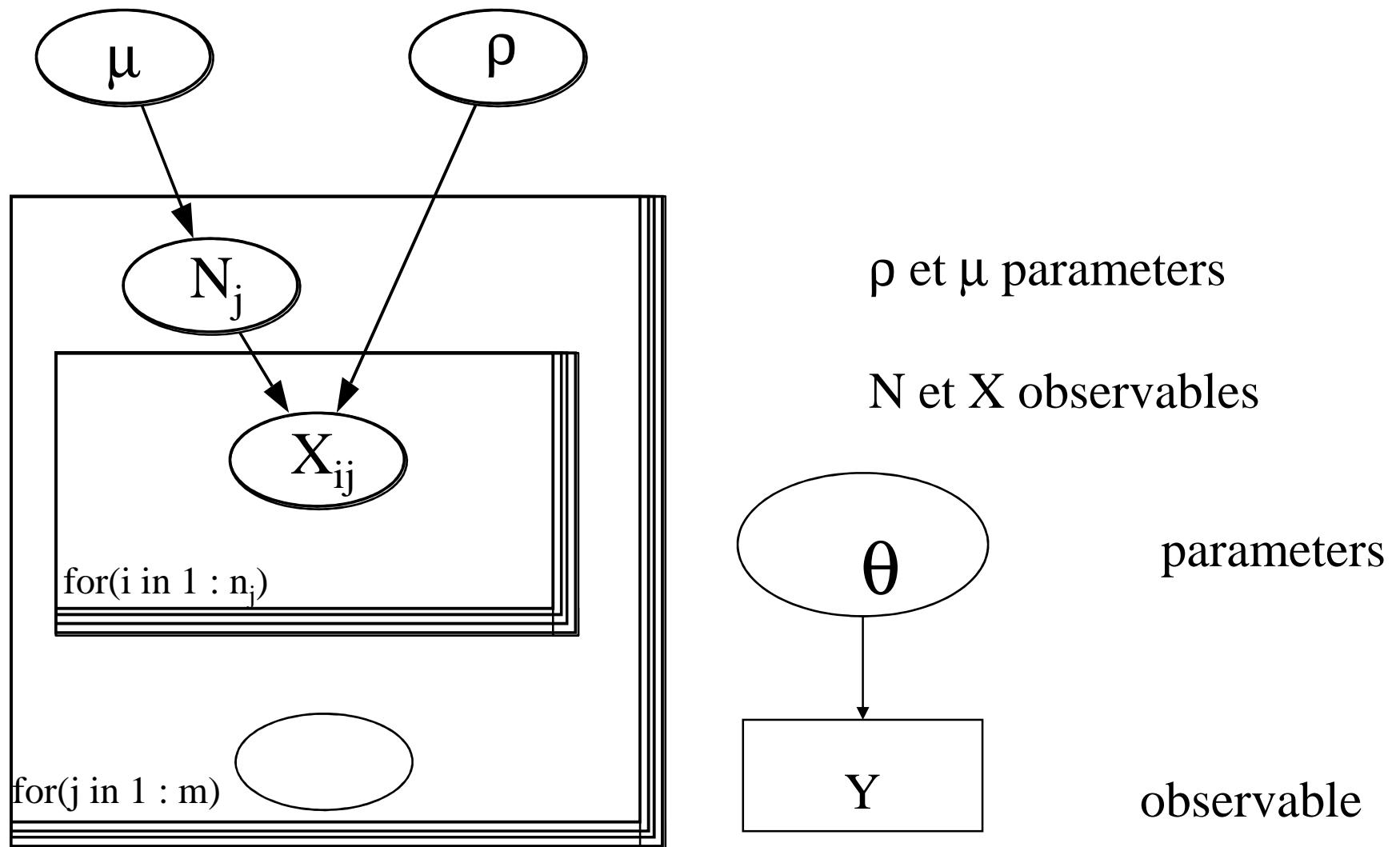
$$[X_i \leq x \mid \rho, X_i > u_0] = 1 - \exp(-\rho(x - u_0))$$

$$[X_i \leq x \mid \rho, X_i > u_0] = 1 - \exp((\rho / \beta) \log(1 - \beta(x - u_0))) \quad \text{si } \beta \neq 0$$

Application 1 : Rains at Bar/Seine

*The point Poisson process
with exponential marks is
completely observed*

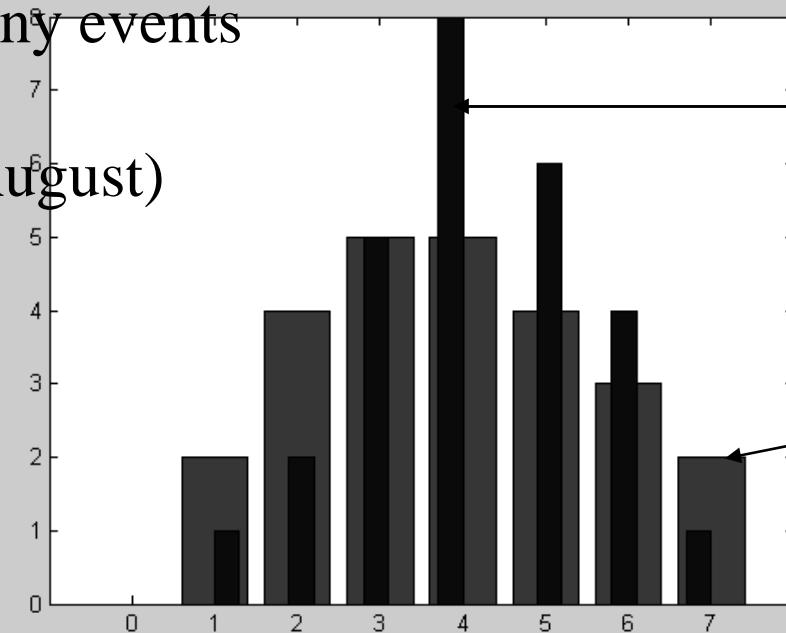
The exponentially marked point Poisson process is completely observed



date	pluie tombée (en mm) par averse					
1975	20.2	1.0	12.0	34.5	1.0	25.1
1976	77.1					
1977	2.4	24.4	17.3	89.4		
1978	1.2	5.0	2.7	1.4	16.6	0.9
1979	14.7	14.6	39.8			
1980	4.5	0.9	4.6	38.5		
1981	2.2	19.3	6.4	0.9	6.6	53.9
1982	1.0	4.8	12.9	0.8		
1983	4.2	5.3	13.5			
1984	13.1	5.4	6.2	4.8		
1985	2.2	30.4	2.0	6.6	19.0	
1986	2.5	12.0	14.2			
1987	0.6	8.7	1.0	8.1	3.5	
1988	31.4	0.8	0.5			
1989	16.0	1.1	11.8	9.3		
1990	7.4	4.2				
1991	0.9	24.7	9.2	14.0		
1992	1.4	1.0	47.0	5.1	16.1	
1993	20.0	0.7	20.0	7.8		
1994	11.3	15.0	4.7	1.5	14.3	34.0
1995	13.3	17.0	23.0			
1996	1.2	55.0				
1997	12.3	6.1	1.4	6.9	79.0	
1998	9.4	9.1	0.5	4.8	2.1	
1999	1.0	16.5	21.1	5.3	8.0	6.0
2000	23.3	67.3	13.0	7.0		
2001	25.2	3.4	6.4	39.5	21.0	

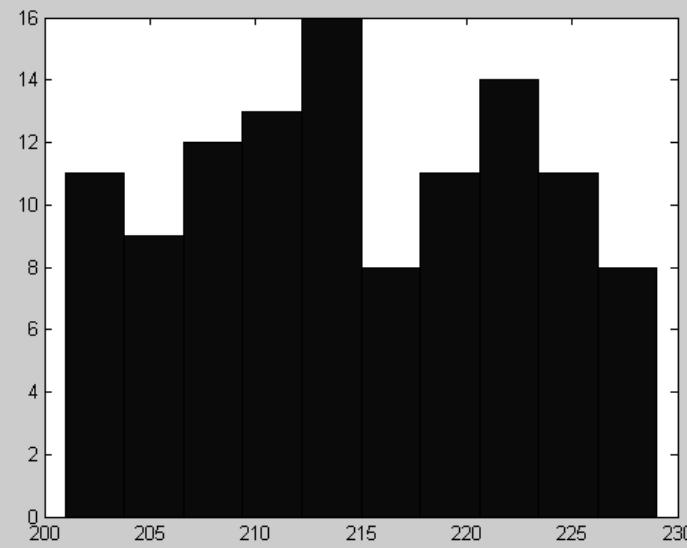
TAB. 1 – Quantité de pluie tombée au cours d'averses à la station de Bar sur Seine entre le 19 juillet et le 18 août pour les années 1975 à 2001

Number of rainy events
at Bar/Seine
(15 July-15 August)



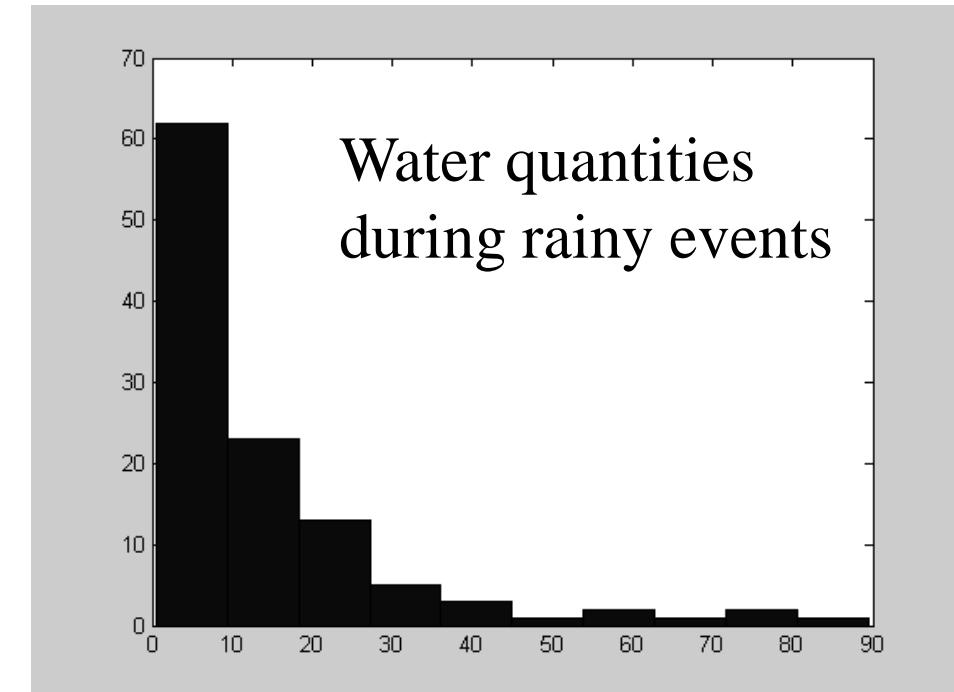
Histogram of
The total number
of rain events

Expected values for
A Poisson dist. with
Average number = 4



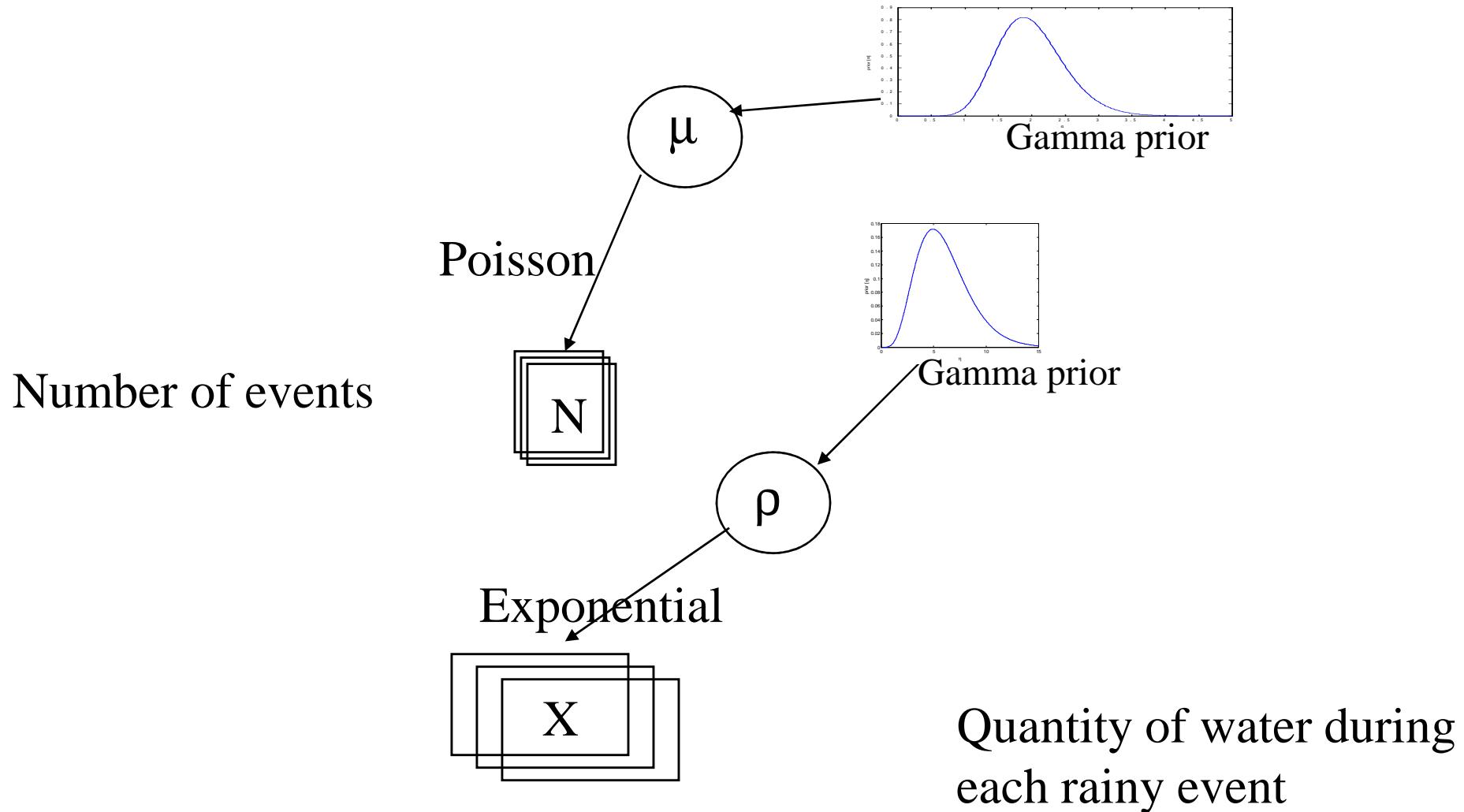
Starting date of rainy events

Water quantities
during rainy events



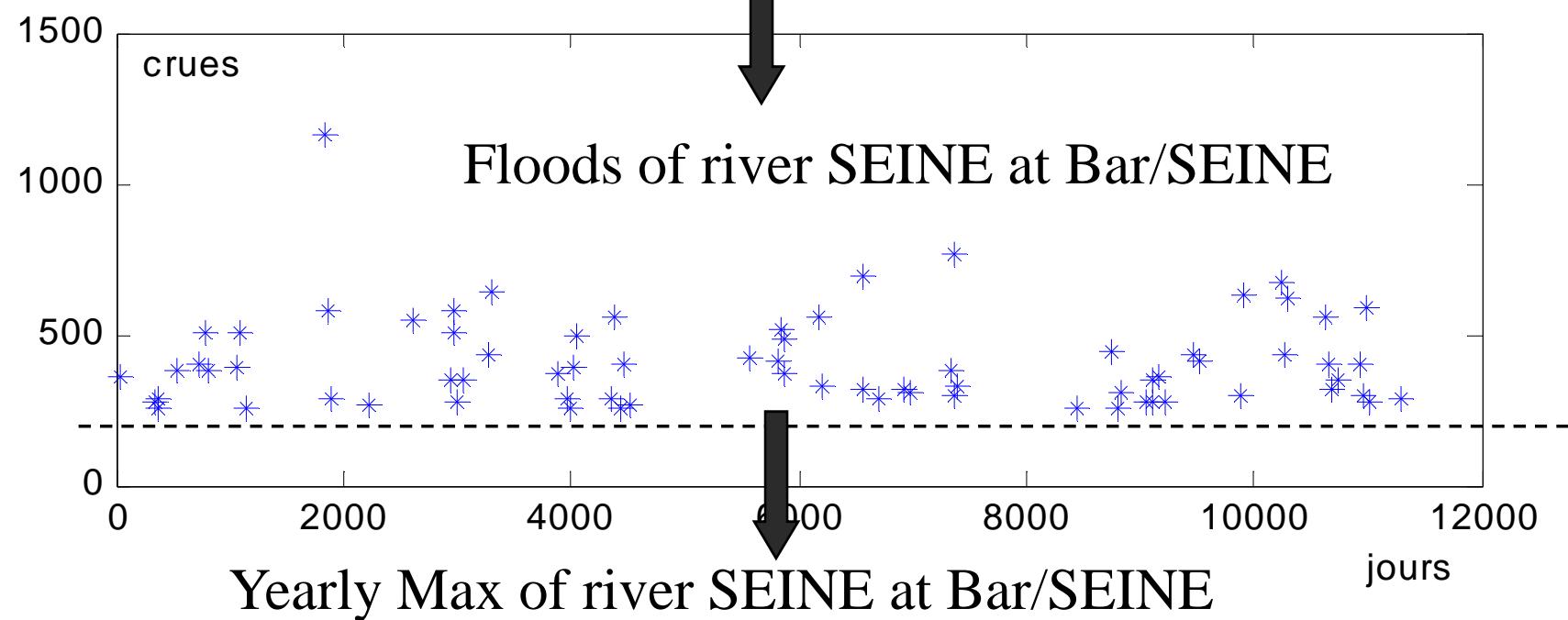
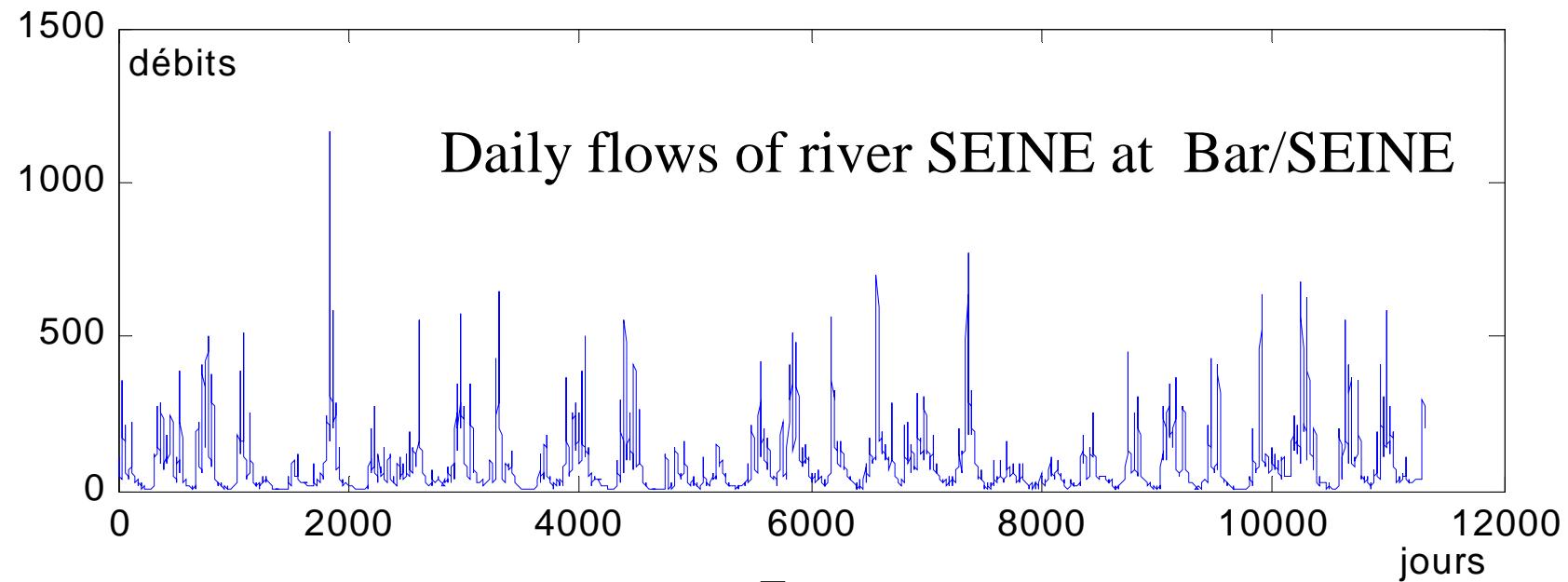
Bayesian inference easy!

The conjugacy « miracle »



Application 2 : Max annual flow at Bar/Seine

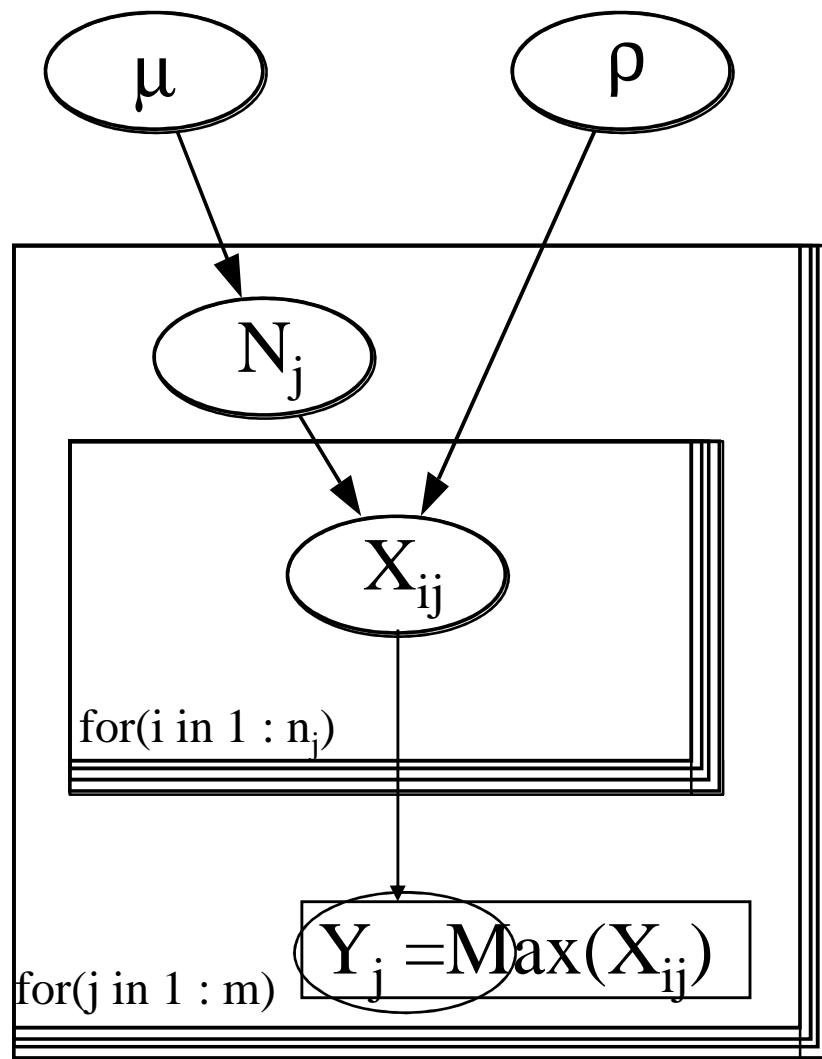
*The point Poisson process
with exponential marks is
incompletely observed
(only the max per year)*



1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
360	384	509	251	240	1163	269	550	579	642	369
1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971
498	553	159	147	513	561	330	697	305	768	162
1972	1973	1974	1975	1976	1977	1978	1979	1980		
184	446	347	432	413	635	675	553	587		

TAB. 3 – Débit Maximum annuel à la station de Bar sur Seine pour les années 1950 à 1980 (en dixièmes de mm sur le bassin-versant)

The Exponentially marked Poisson process is only partially observed



ρ and μ parameters

N and X latent variables

Y observed

Hierarchical scheme

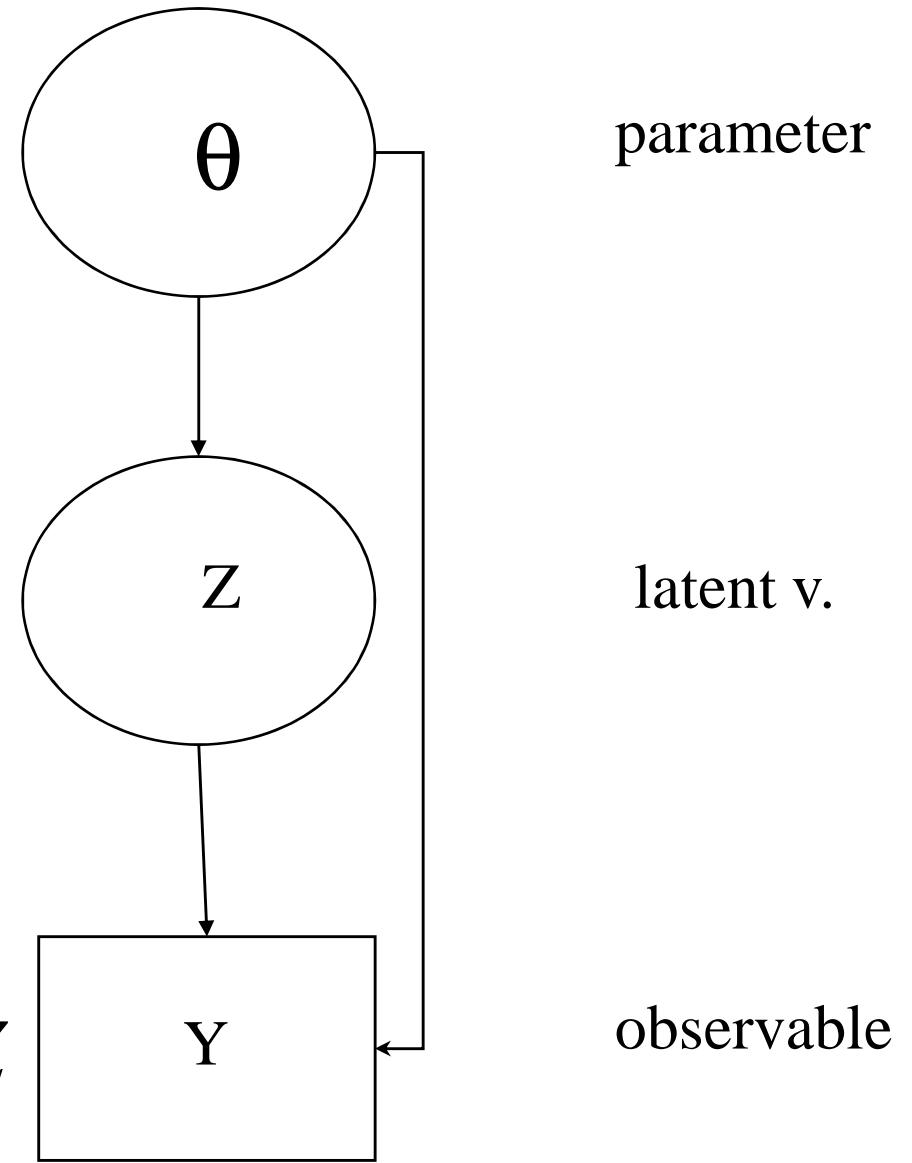
Formulation

$$[Z|\theta]$$

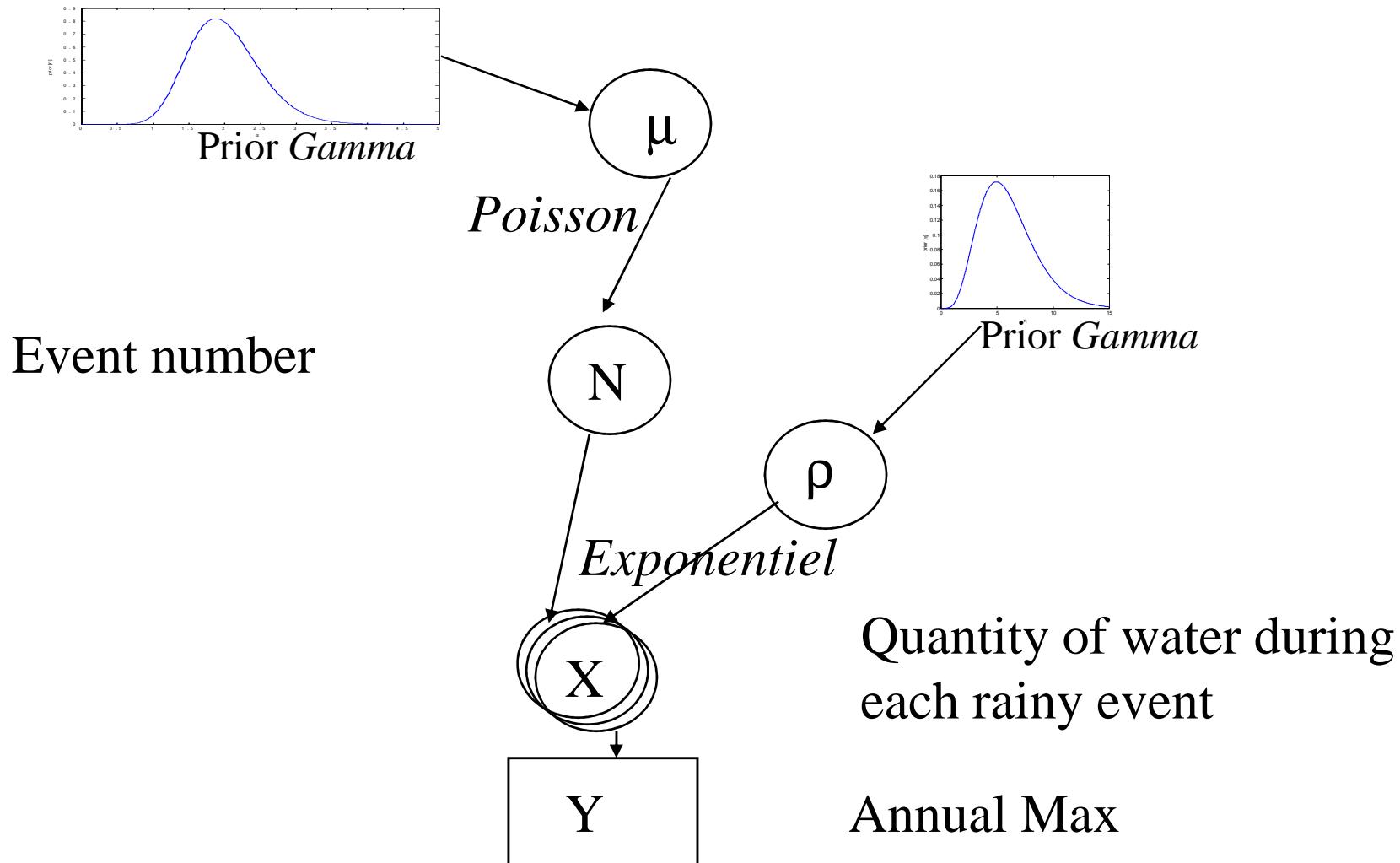
$$[Y|Z,\theta]$$

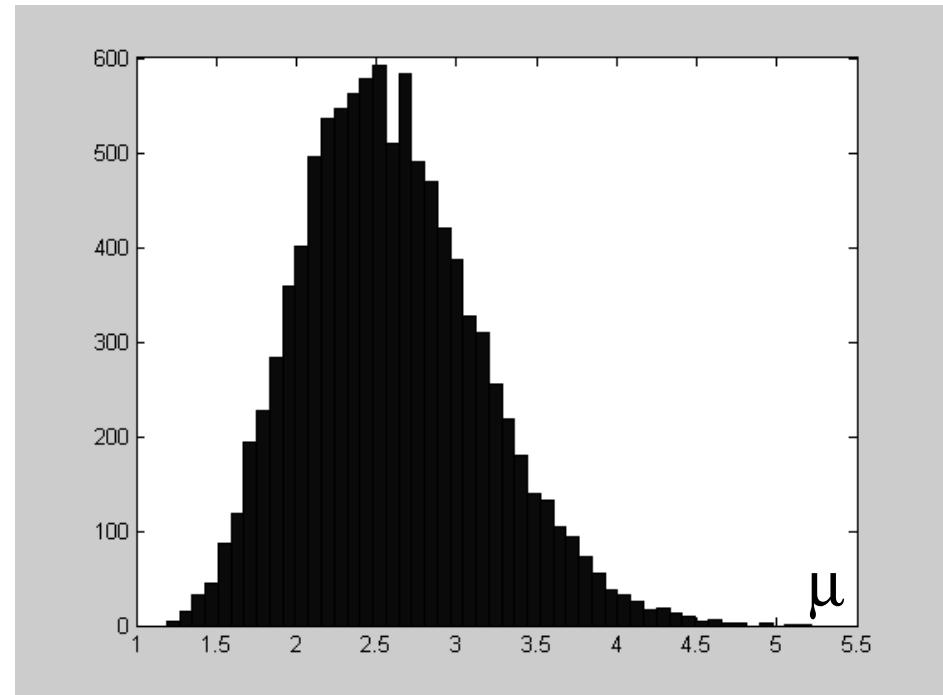
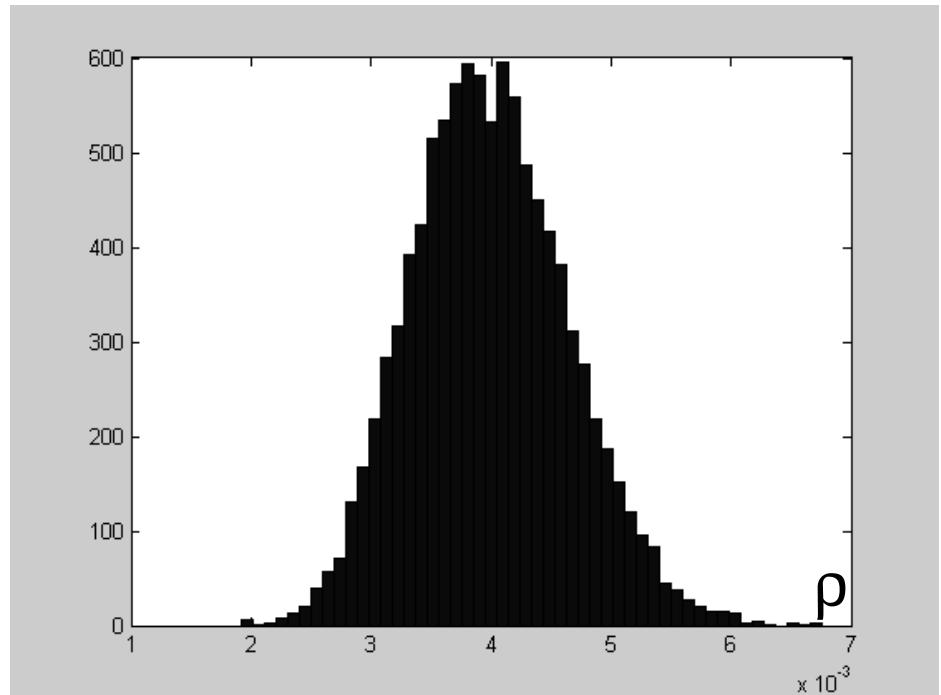
likelihood

$$[Y|\theta] = \int_z [Y|Z,\theta] \times [Z|\theta] dZ$$

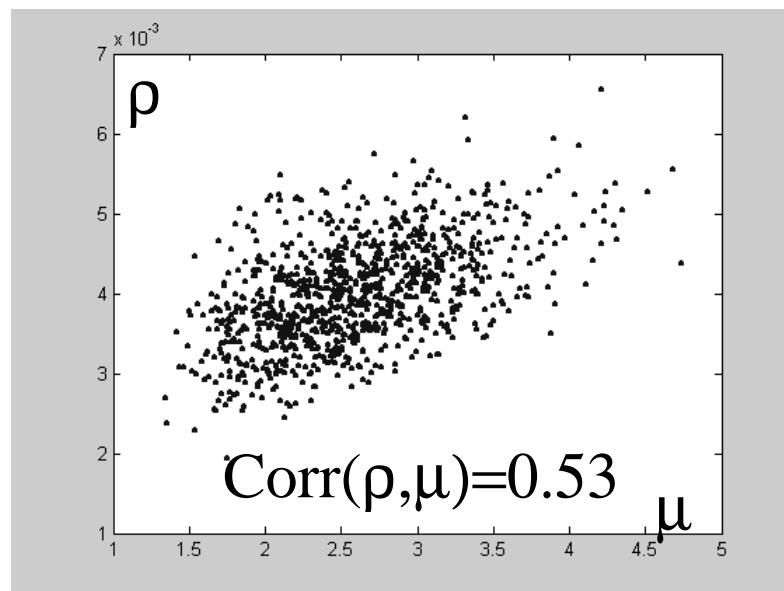


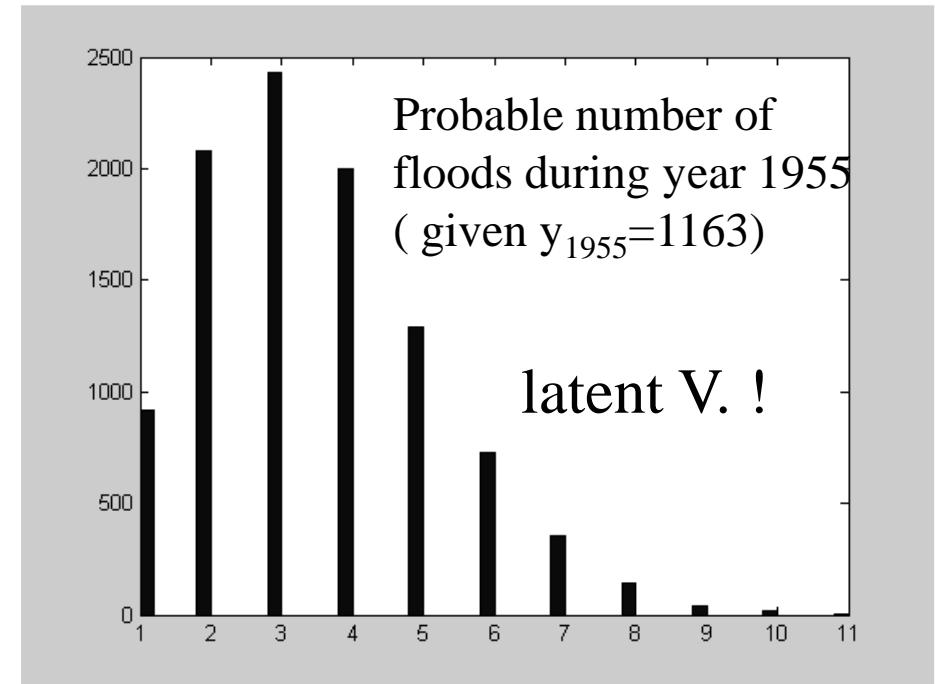
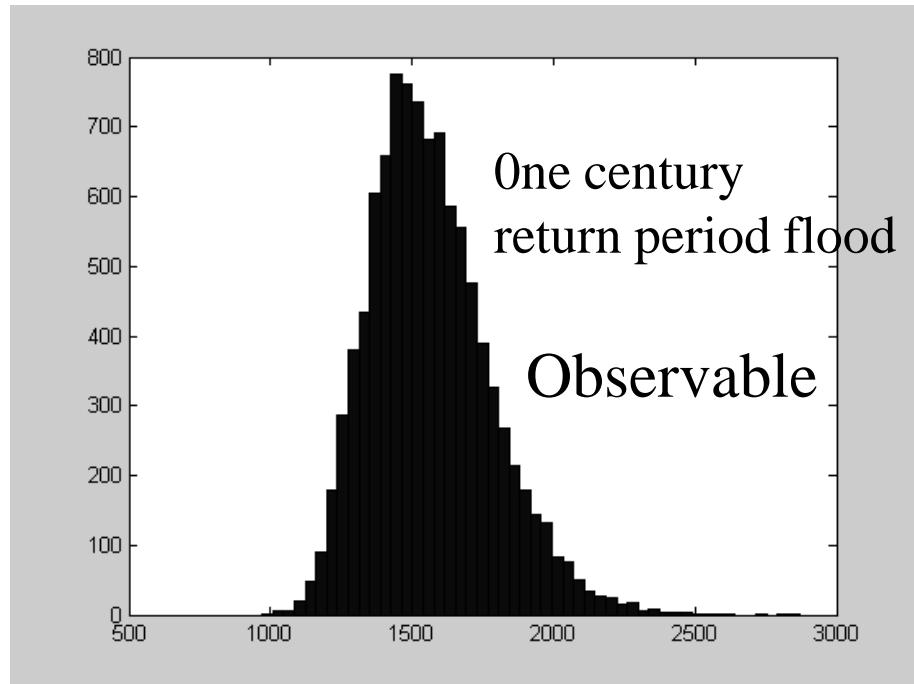
Bayesian inference by Data Augmentation =« reconstruction » $[X|Y,N,\rho,\mu]$ is a truncated exp!



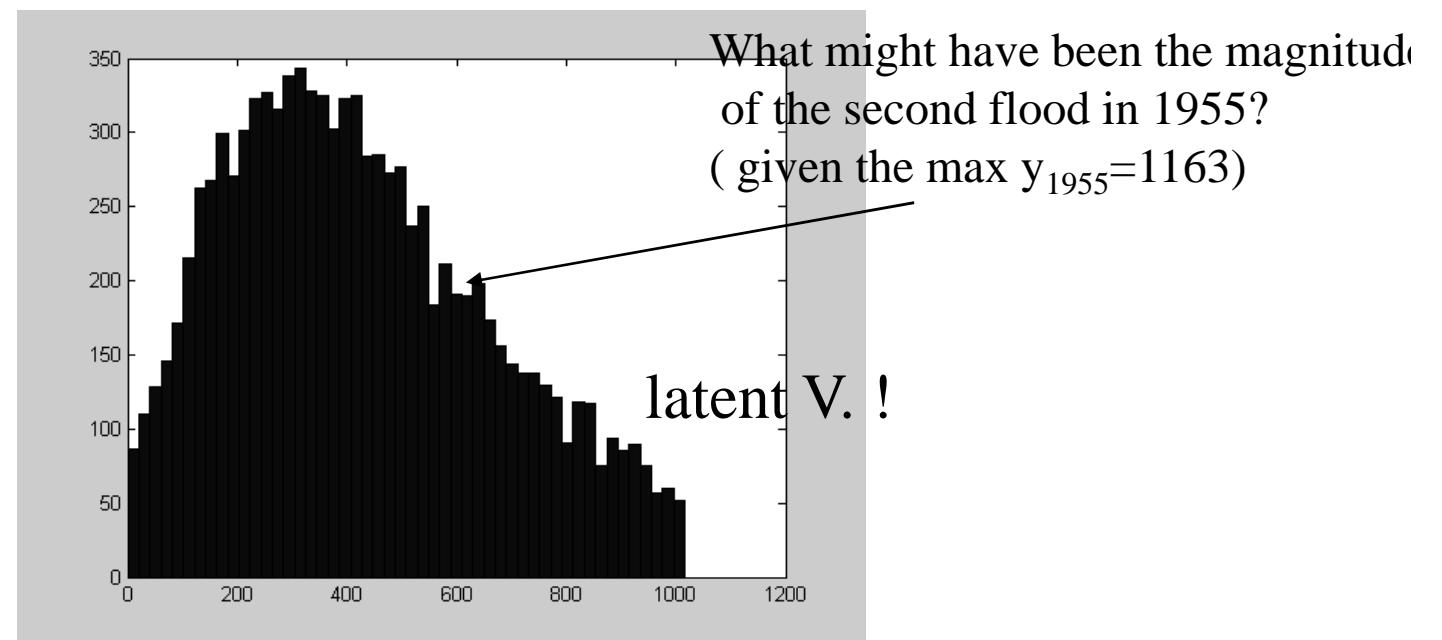


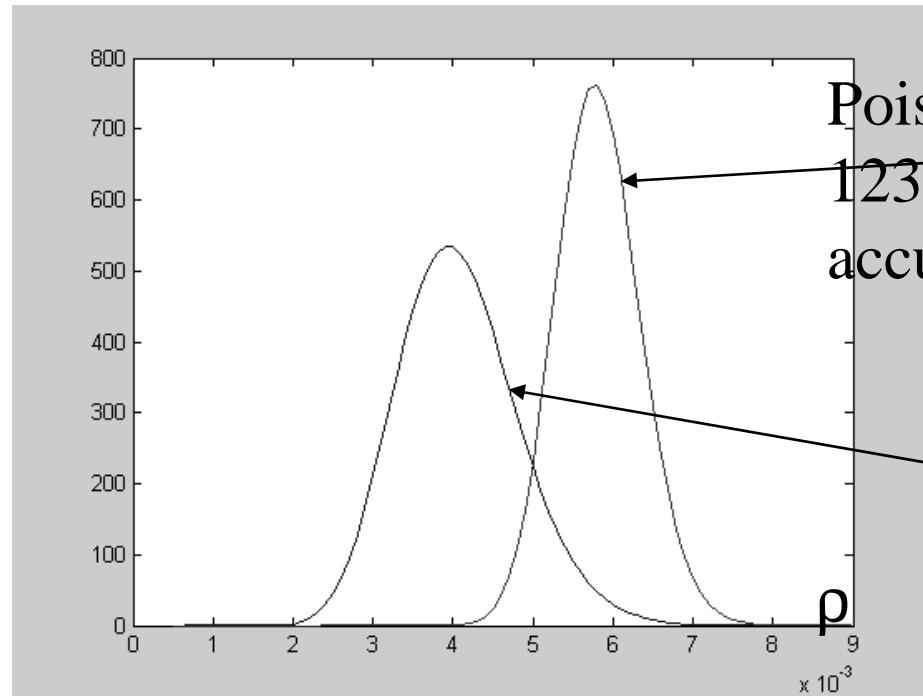
Estimations on max of P.P. with Exp. marks (GUMBEL model)





PREDICTIONS thanks to the latent P.P with Exponential marks





Gumbel model
for annual max

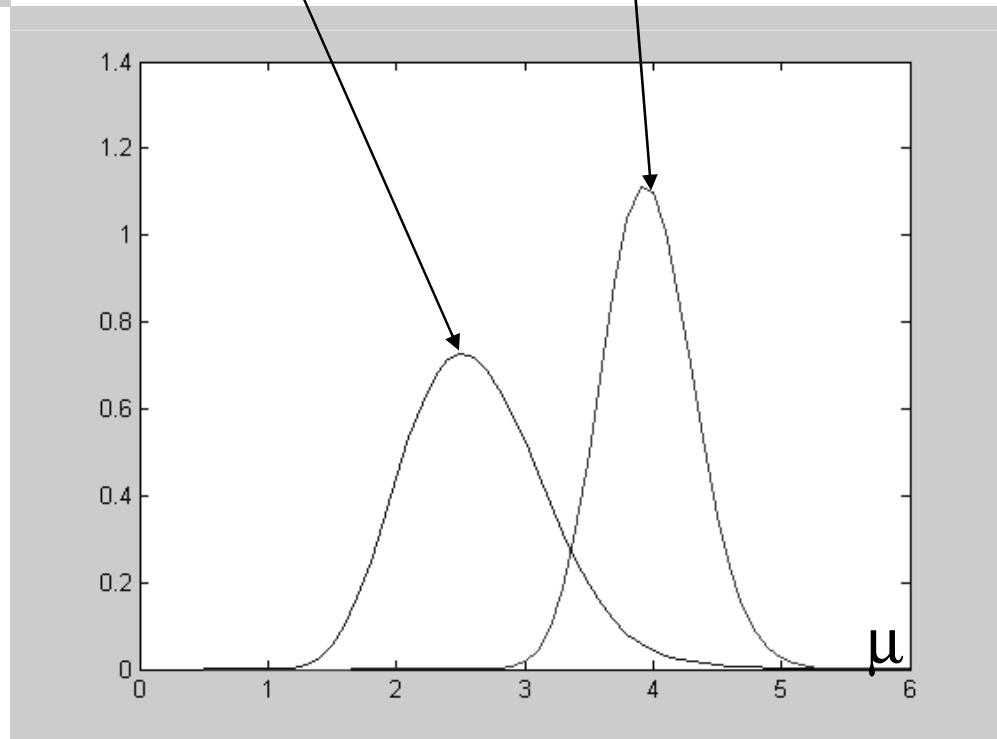
Large events of 1955

3 floods

17 january: 1163 mm/10

11 february: 583 mm/10

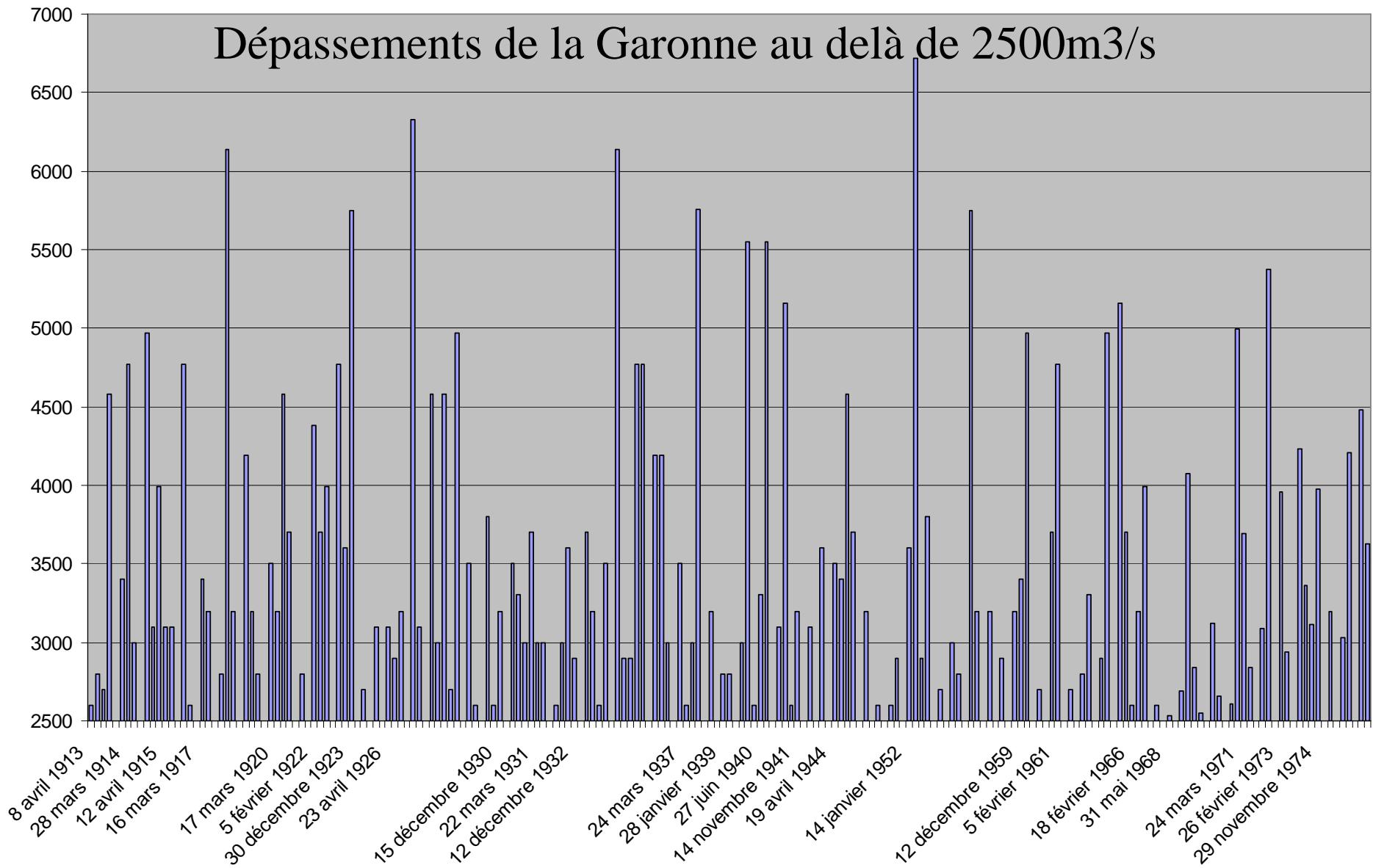
28 february: 282 mm/10



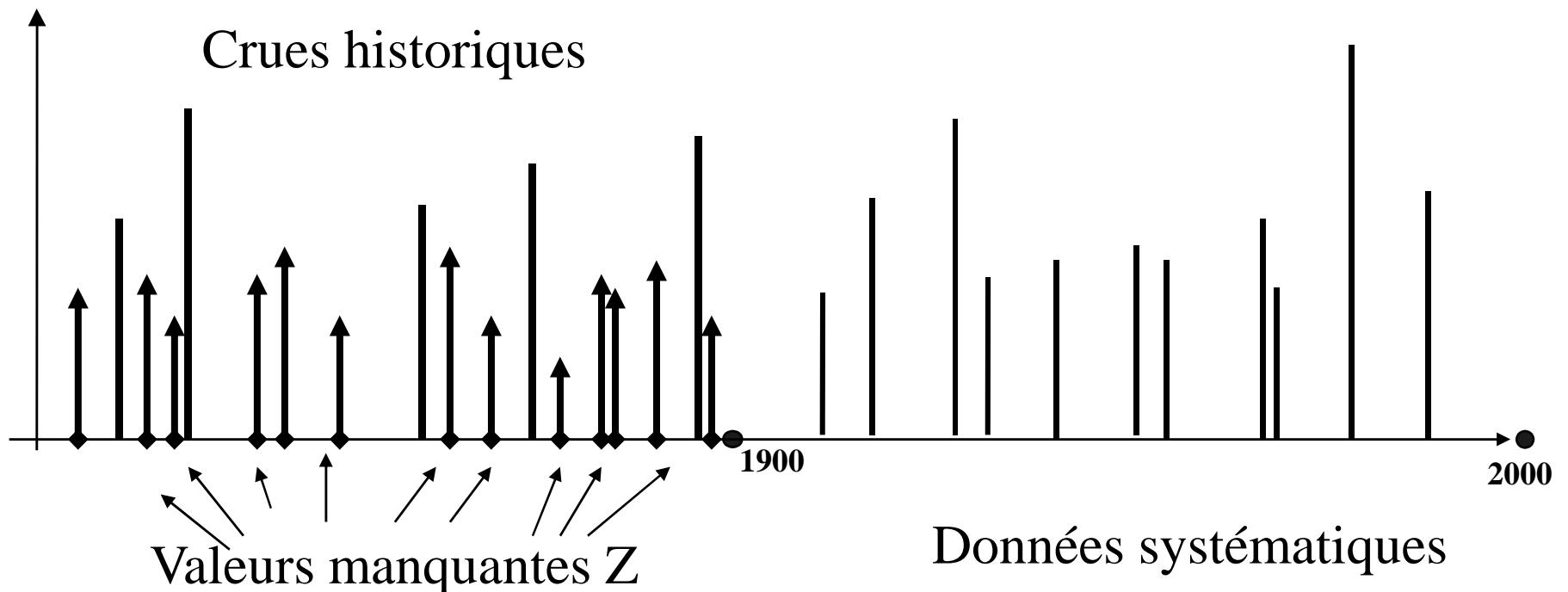
Records historiques de la Garonne à Mas d'Agenais

DATE	HAUTEUR (m)	DEBIT ESTIME (m3/s)
Avril 1770	10.34	7000 à 7400
Sept. 1772		6300
Mars 1783		7000 à 7200
Mai 1827		6500
Mai 1835		6400
Janv. 1843		6500
Juin 1855	9.96	7000
Mai 1856	9.62	6200
Juin 1856	9.88	6600
Juin 1875	10.56	7000 à 7500 (peut-être 8000)
Janv. 1879	9.62	6300
Fev. 1879	10.02	7000
Mai 1918	9.51	6000
Mars 1927	9.97	6700
Mars 1930	10.72	7000 à 7500 (peut-être 8000)
Mars 1935	9.95	6700
Fev. 1952	10.26	6000 à 7000
Janv. 1955	9.32	5200 à 5700

DEBIT q(i) (m³/s)



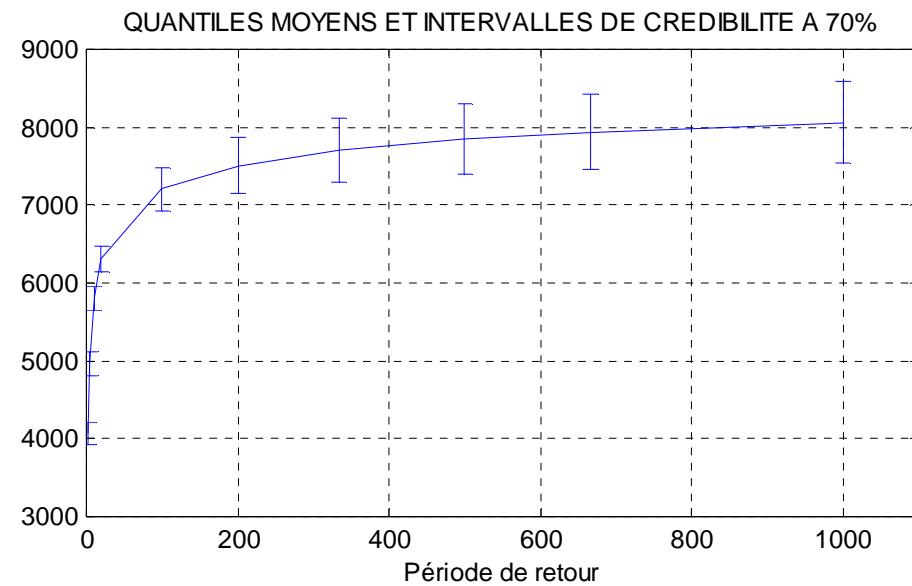
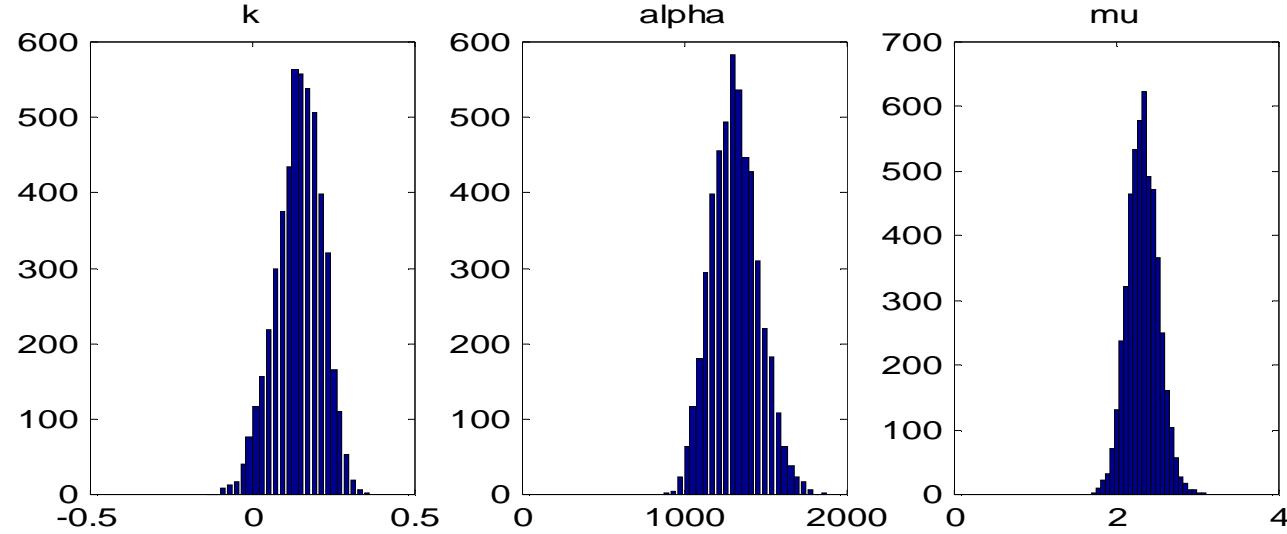
Flood and latent variables



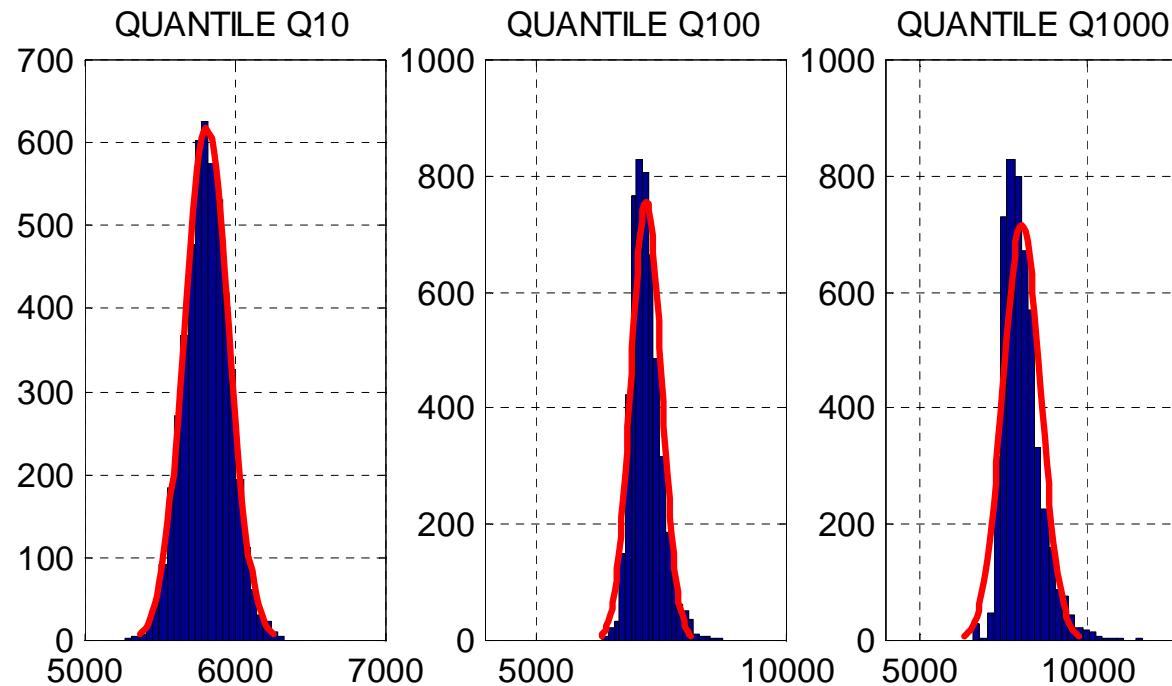
Data Augmentation algorithm

- Generate K , the number of missing data
 - Draw Z for missing flood data
 - Compute posterior pdf of parameters θ on augmented sample
-
- Poisson pdf with parameter μ_H
- Truncated GEV with parameter θ
- Same pdf structure as the already collected data

Bayesian Estimation with historical data



Estimation sur données historiques

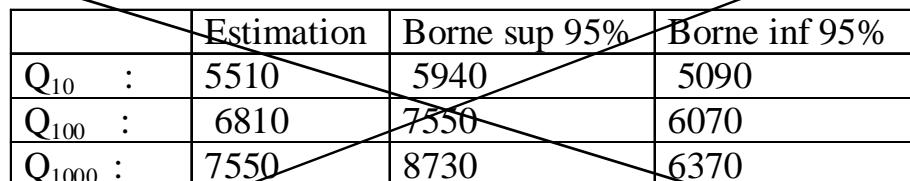


	Moyenne	Médiane	Borne Sup 95%	Borne Inf 95%
mu	2.33	2.32	2.64	2.01
k	0.22	0.22	0.32	0.11
alpha	1477.63	1468.41	1739.05	1234.42
Q ₁₀	5808.35	5807.08	6049.74	5564.52
Q ₁₀₀	7205.26	7166.80	7759.84	6805.37
Q ₁₀₀₀	8060.79	7967.45	9101.08	7375.53

Conclusions Mas d Agenais

	Médiane	Borne Sup 95%	Borne Inf 95%
Q_{10}	5520 5810	5990 6050	5190 5565
Q_{100}	7004 7170	8450 7760	6350 6805
Q_{1000}	8020 7970	10880 9100	6930 7375

Les données historiques ne changent guère l'estimation de la valeur centrale, mais réduisent par 2 sur ce cas l'incertitude d'estimation des valeurs de projets



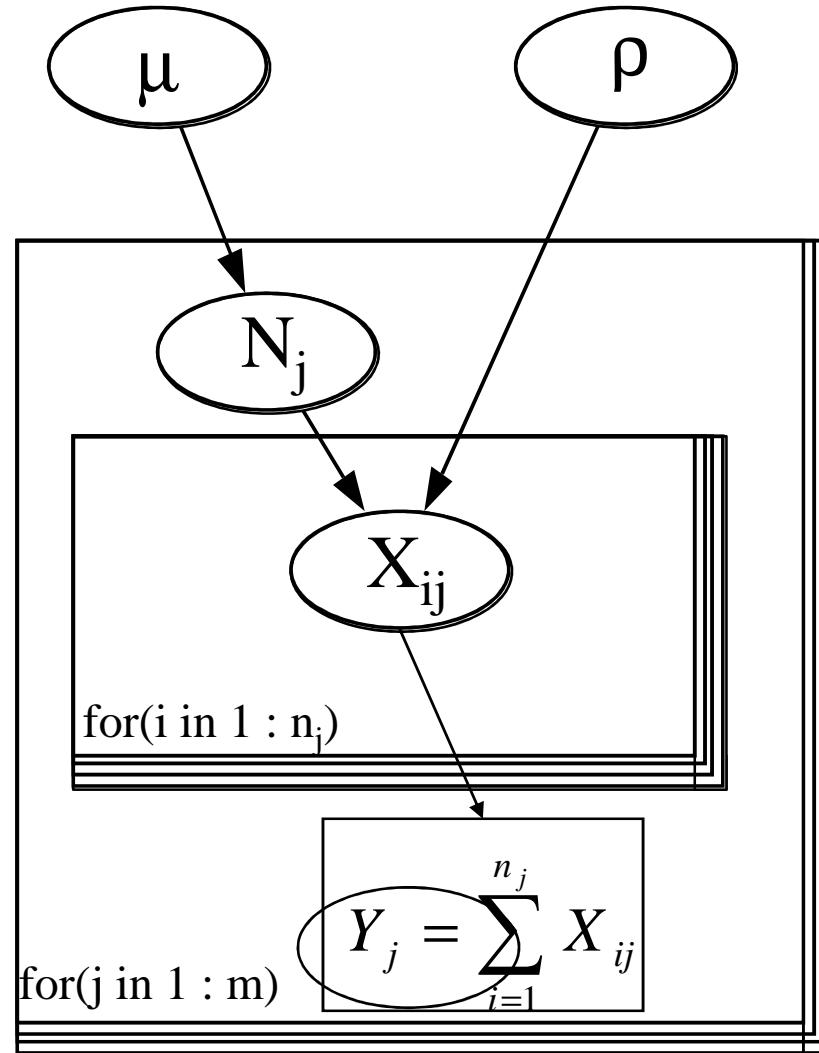
	Estimation	Borne sup 95%	Borne inf 95%
$Q_{10} :$	5510	5940	5090
$Q_{100} :$	6810	7550	6070
$Q_{1000} :$	7550	8730	6370

Max Vrais

Application 3 : Montly rains at Ghezala dam, Tunisie

*The point Poisson process
with exponential marks is
incompletely observed
(only the sum per period)*

The Exponentially marked Poisson process is only partially observed



	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978
<i>Février</i>	55.7	99.9	105.9	186.9	75.6	150.0	145.1	156.6	80.3	23.8	102.2
<i>Août</i>	0	0.6	0	0	1.6	10.2	0	23.6	10.2	29.7	13.2
	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
<i>Février</i>	123.8	65.0	65.5	73.8	26.0	102.6	91	106.1	164.9	82.2	65.8
<i>Août</i>	0	20.5	0	14.1	0	2.4	0	0	0.6	0	0
	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
<i>Février</i>	34.2	136.4	126.0	71.2	79.6	0.7	291.3	87.2	61.3	76.8	50.4
<i>Août</i>	33.3	7.5	3.4	0	0.1	31.5	8.6	4.3	52.2	1.4	0
											0.2

TAB. 5 – Précipitations mensuelles de Février et d'Août au barrage de Ghezala

Hierarchical scheme (again...)

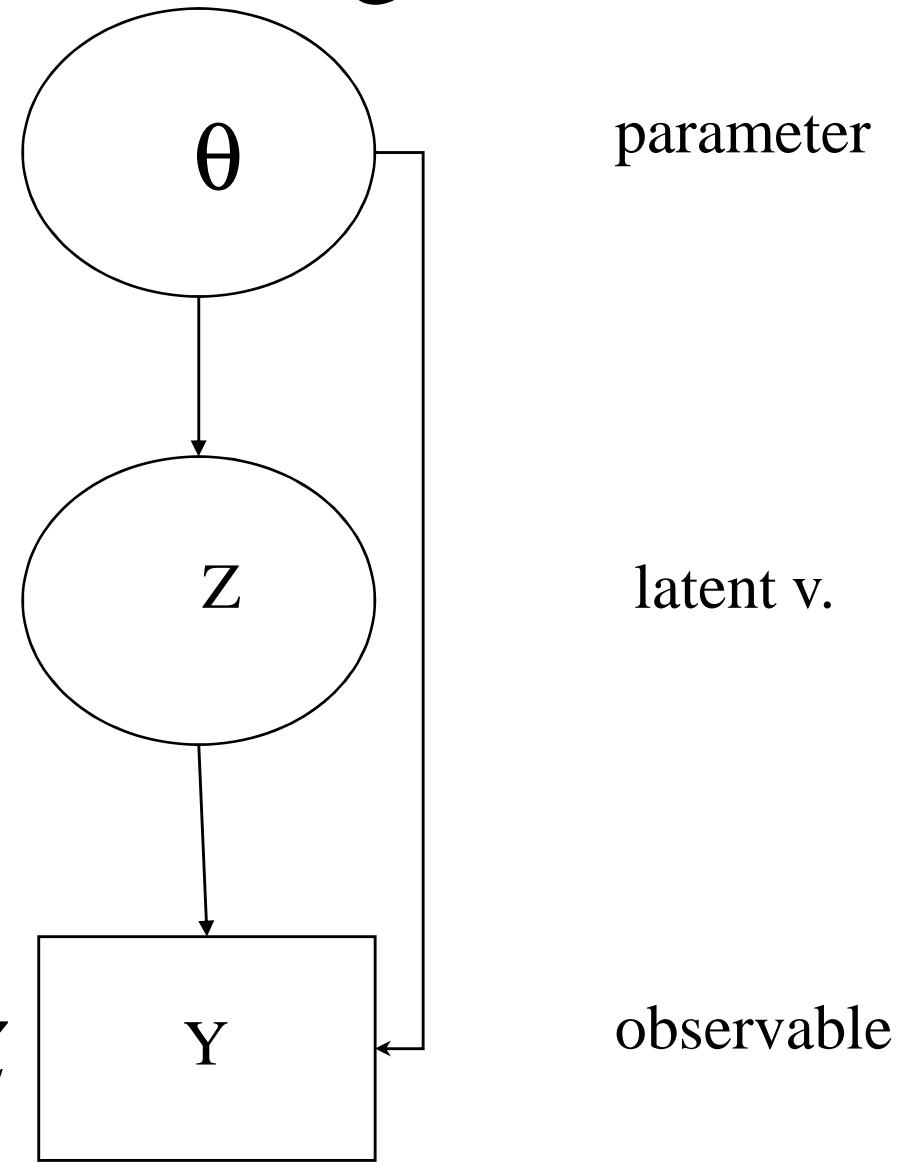
Formulation

$$[Z|\theta]$$

$$[Y|Z,\theta]$$

likelihood

$$[Y|\theta] = \int_z [Y|Z,\theta] \times [Z|\theta] dZ$$



La « loi des fuites »: Y~fuitepdf(ρ , μ)

- Def : Compound Poisson of exp. marks

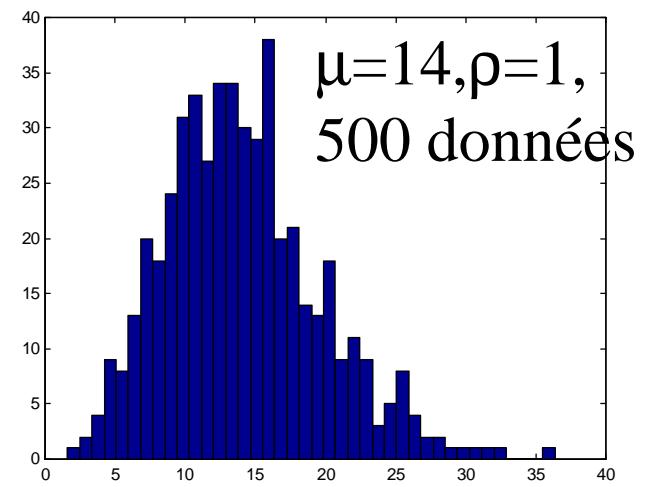
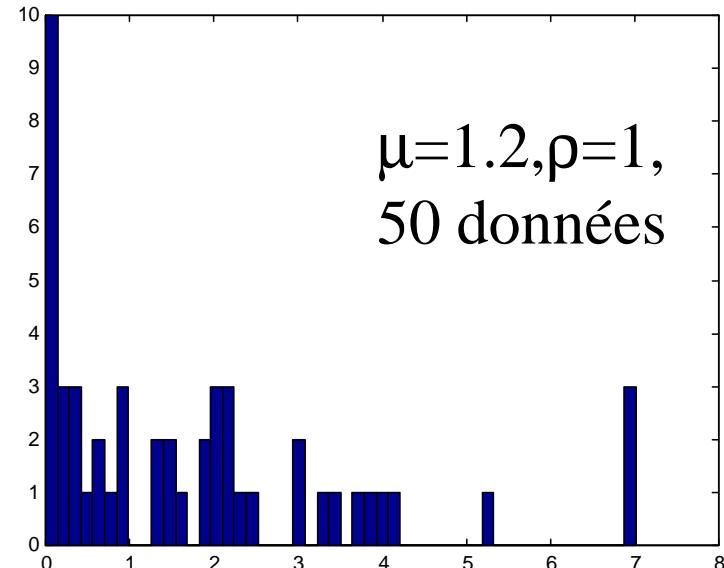
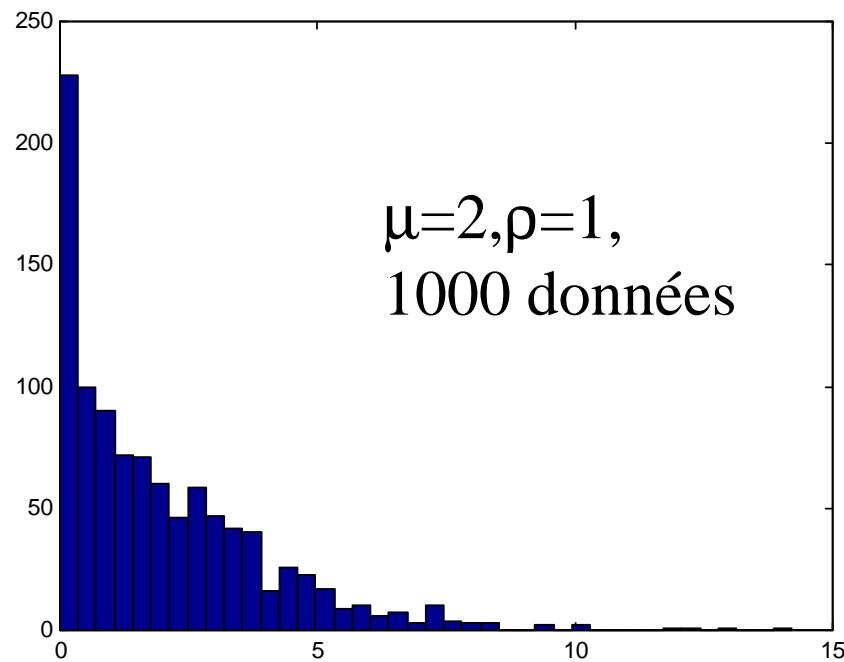
$$Y = \sum_{i=1}^N X_i; X_i \sim \exp(\rho), N \sim \text{Pois}(\mu)$$

- Other définition : Poisson convolution of gamma variates

$$[Y = y | \rho = 1, \mu] = \begin{cases} \sum_{k=1}^{\infty} \left(e^{-\mu} \frac{\mu^k}{k!} \right) \frac{y^{k-1}}{(k-1)!} e^{-y} 1_{y>0} \\ e^{-\mu} 1_{y=0} \end{cases}$$

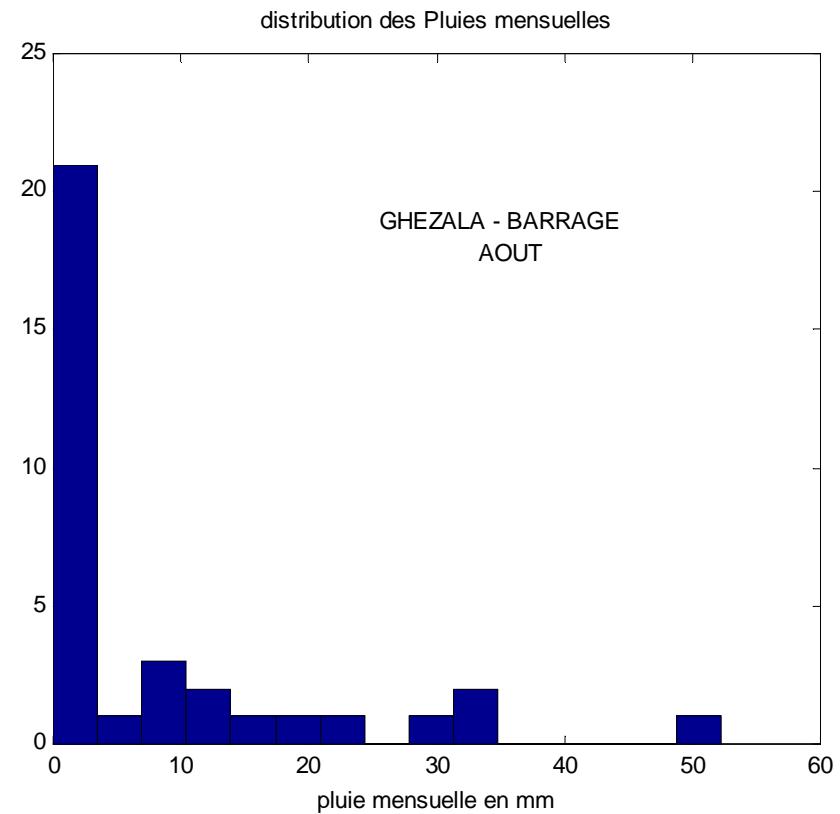
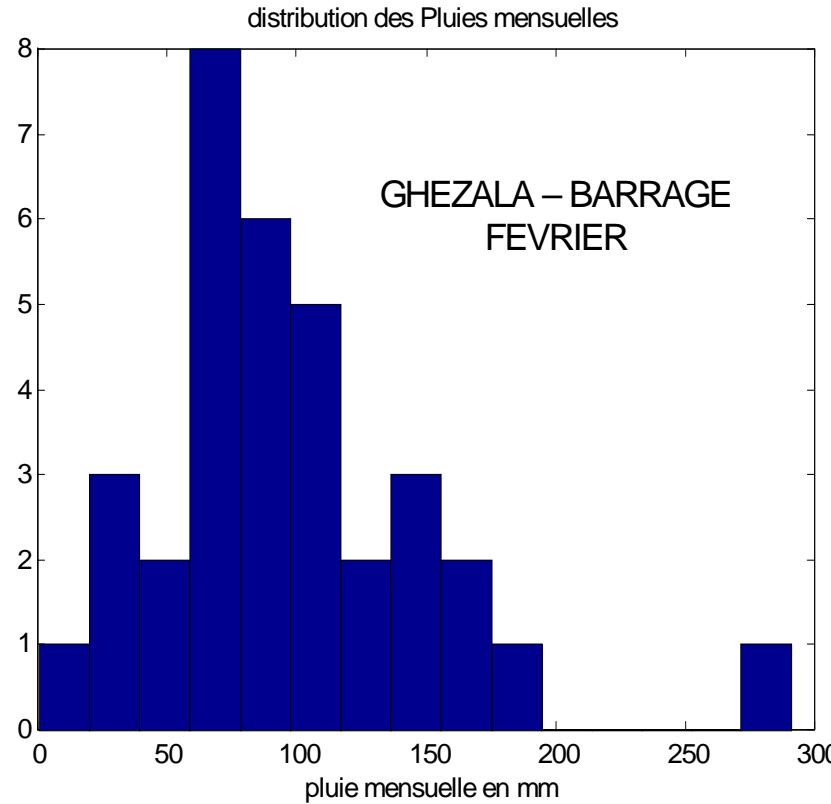
Simulations

```
N=poissrnd(mu,1,1000)
for i=1:length(N)
    if N(i)==0
        y(i)=0;
    else
        y(i)=gamrnd(N(i),1);
    end,
end
```



Real data: monthly rain

- a bucket at Ghezala-dam : february, august



Loi des fuites

- Moments

$$E(Y|\rho, \mu) = \mu \times \rho$$

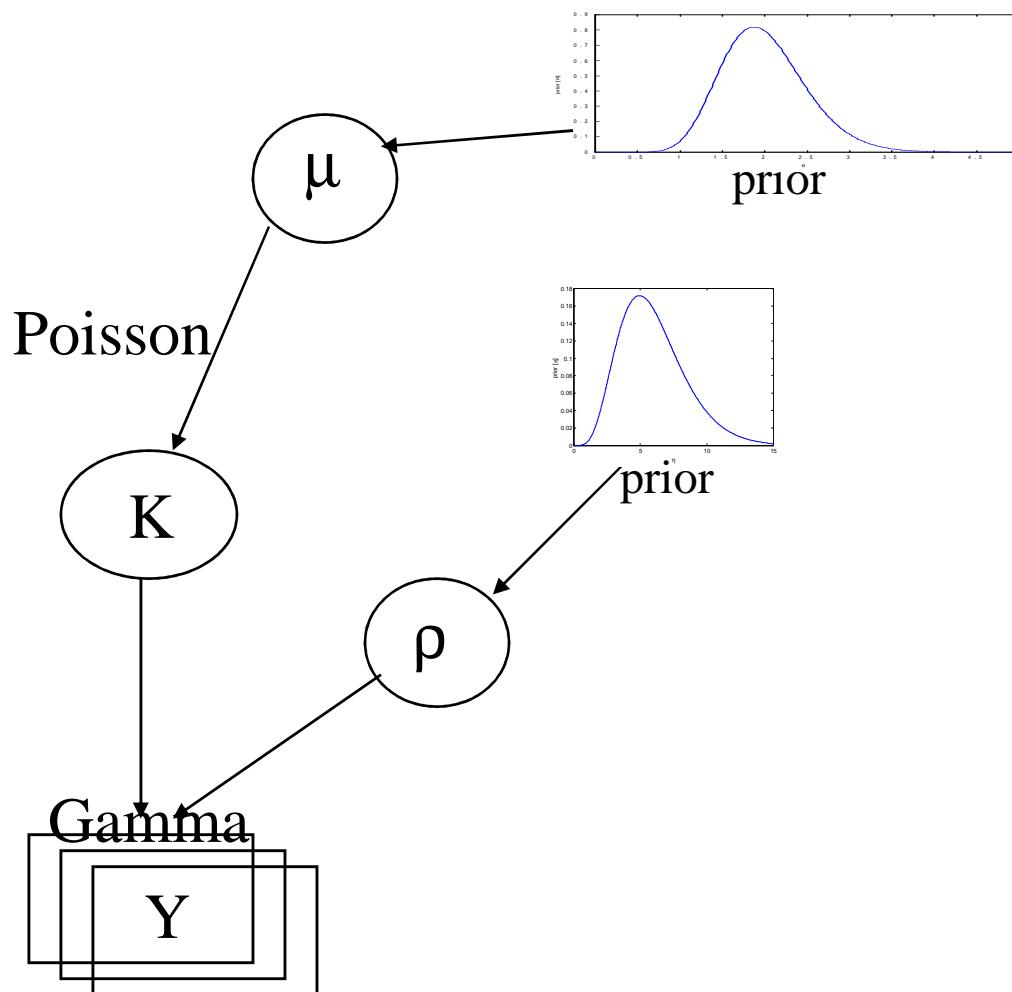
$$\text{Var}(Y|\rho, \mu) = 2\mu \times \rho^2$$

- f.c. $\psi_Y(s) = E(e^{isY}|\rho, \mu)$

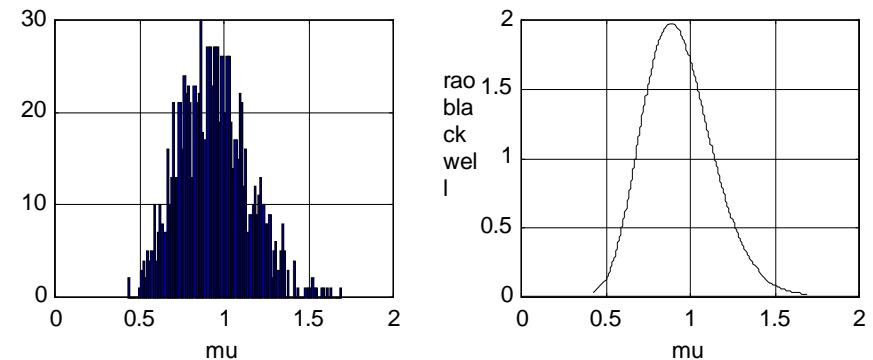
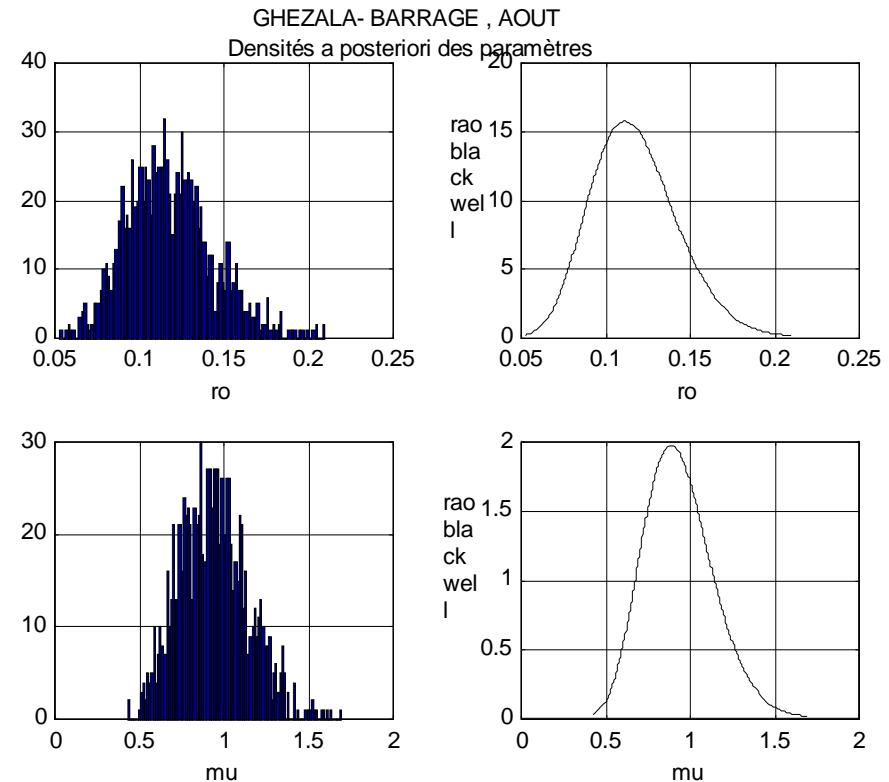
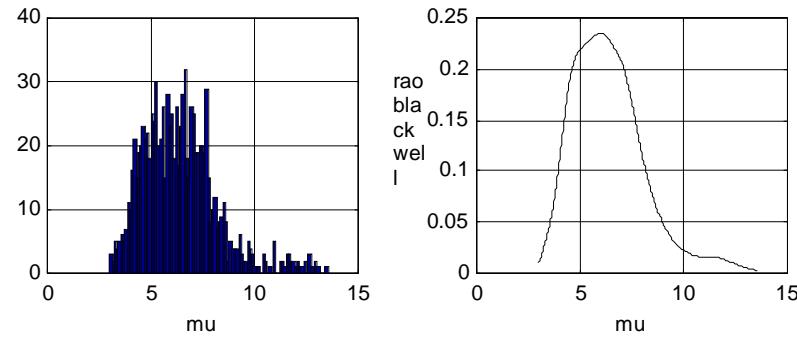
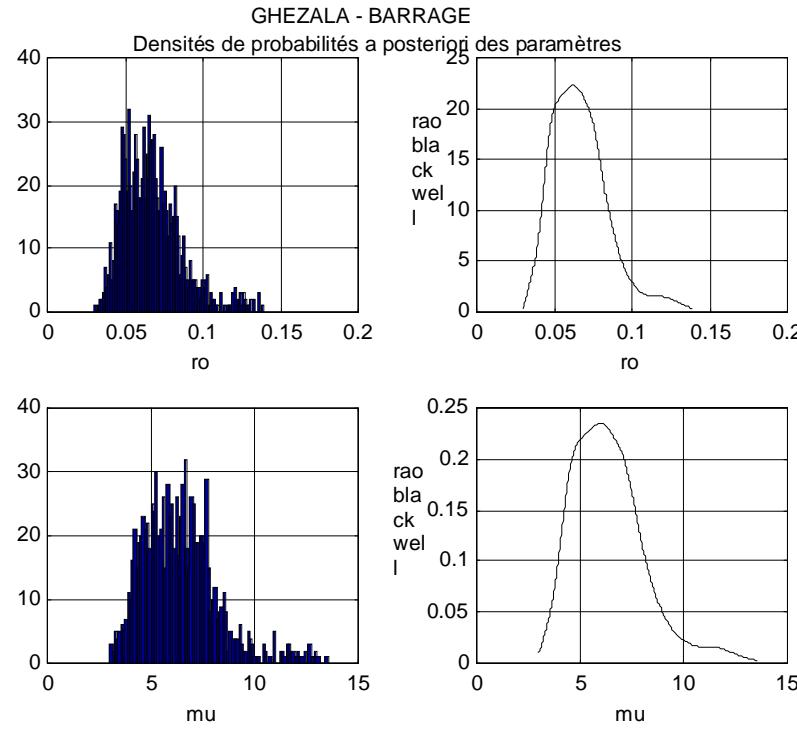
$$\psi_Y(s) = e^{-\mu} + e^{isy} \sum_{k=1}^{\infty} \left(e^{-\mu} \frac{\mu^k}{k!} \right) \frac{y^{k-1}}{(k-1)!} e^{-y}$$

$$\psi_Y(s) = \exp\left(\mu \frac{is}{1-is}\right)$$

Bayesian inference not difficult!



GHEZALA :Posterior parameter Distributions (MCMC Gibbs)



$$\hat{\mu} = 7; \hat{\rho} = 0.065; \hat{\rho}^{-1} = 15.4 \text{mm} \quad \text{En February}$$

$$\hat{\mu} = 0.9; \hat{\rho} = 0.12; \hat{\rho}^{-1} = 8.3 \text{mm} \quad \text{En August}$$

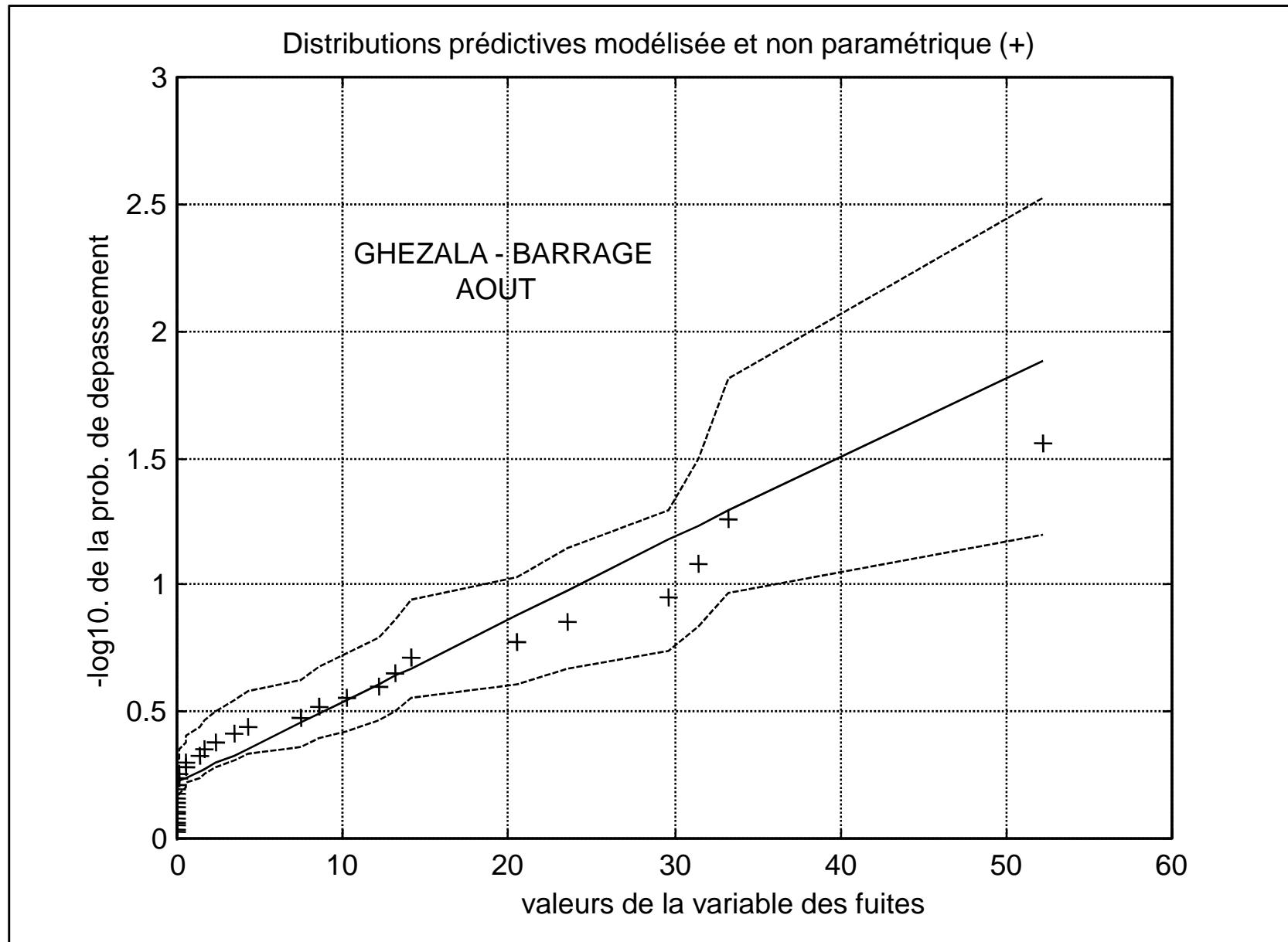
Modèle des fuites

- Various hydrological situations (zero inflated Ghezala summer dist.)
- Help of Tom Bayes (relying on conditional structure)
- Convenient Interpretation of parameters + credibility bands

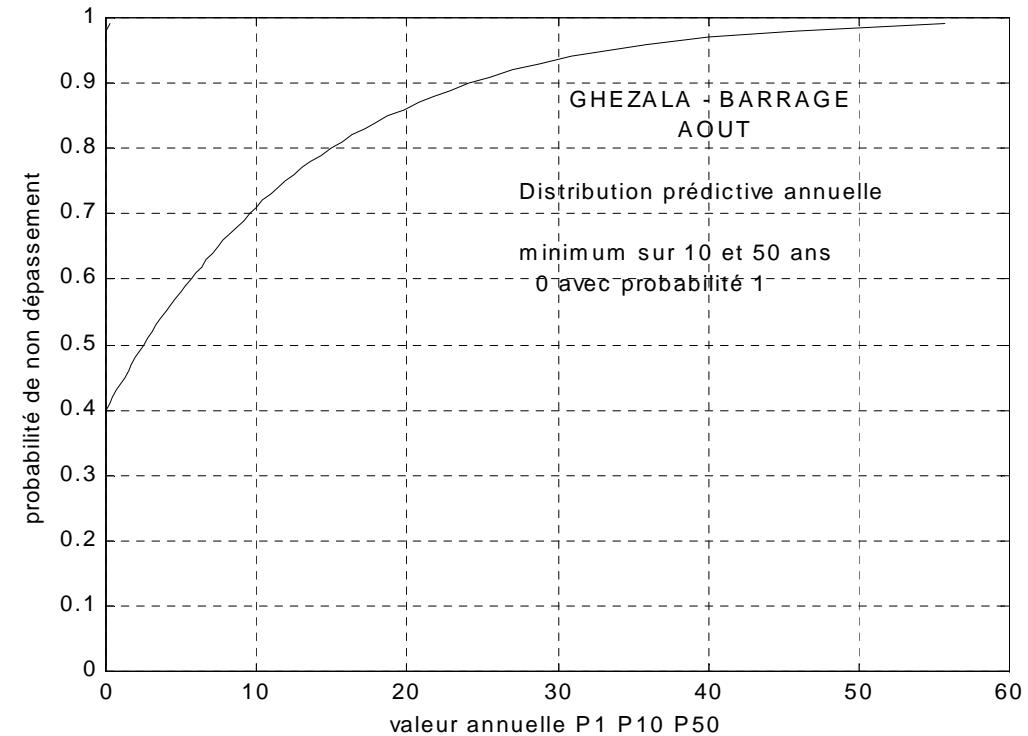
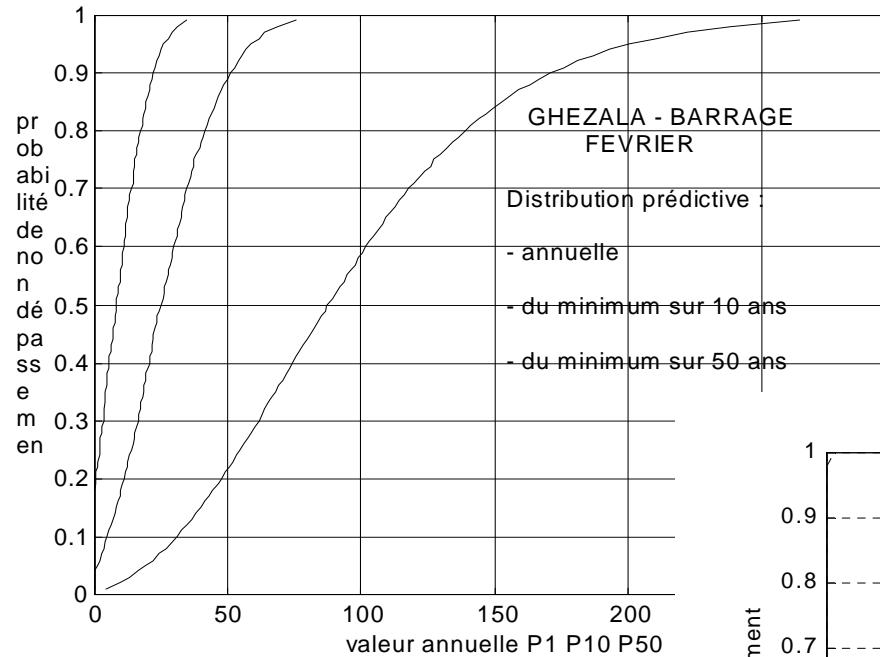
$\hat{\mu} = 7; \hat{\rho} = 0.065; \hat{\rho}^{-1} = 15.4 \text{mm}$ In February

$\hat{\mu} = 0.9; \hat{\rho} = 0.12; \hat{\rho}^{-1} = 8.3 \text{mm}$ In August

Predictive Validation of la “ loi des fuites ”



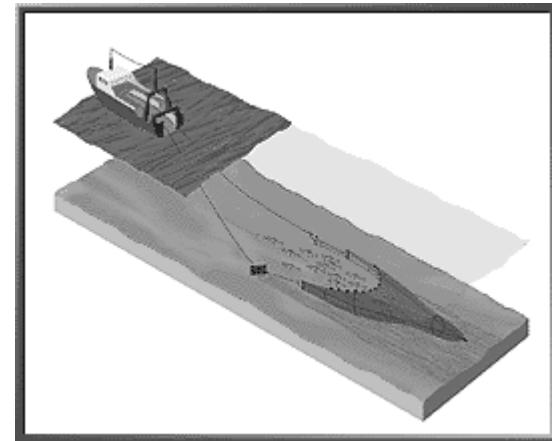
Distributions prédictives cumulées annuelles, sur 10 et 50 ans



Application 4 : Trawl data in St-Laurence bay, Canada

*A compound Poisson distribution
for zero-inflated data*

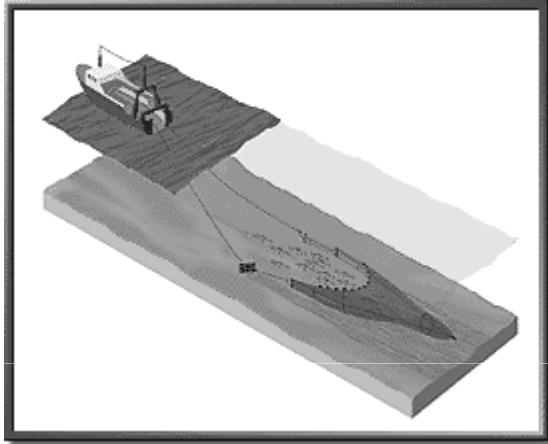
Context



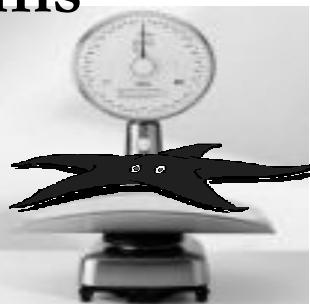
Yearly scientific catches
with a trawl in the bay

Data collection

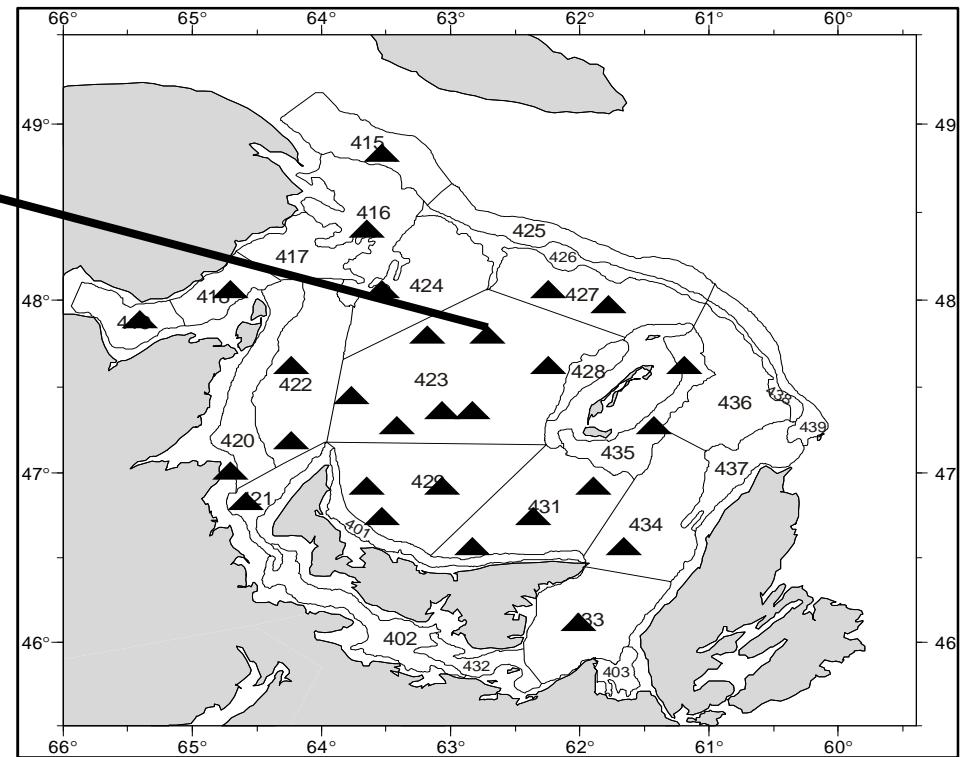
Every September since 1971



Standard trawl sweep of 3,24kms



Biomass weighted

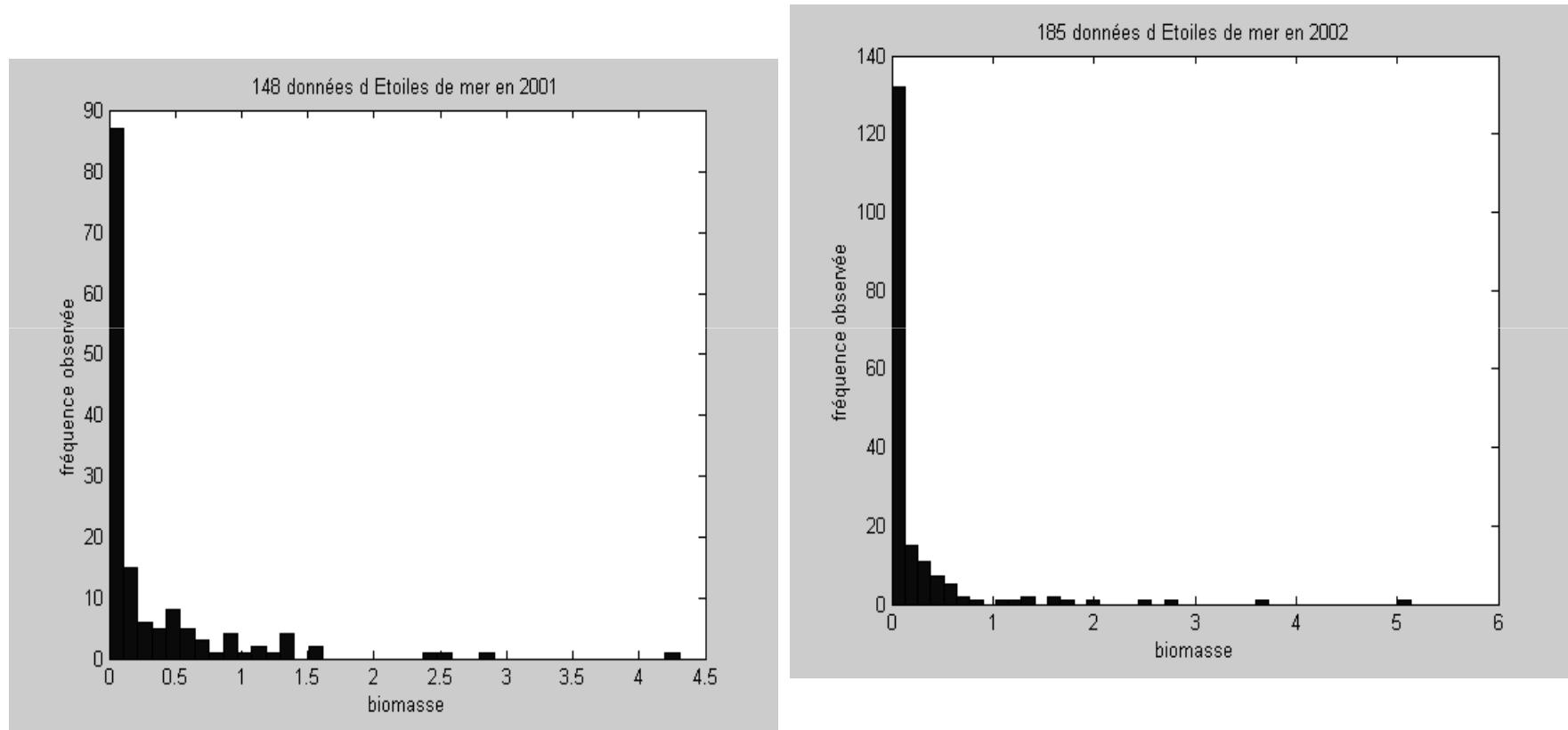


Sampling design

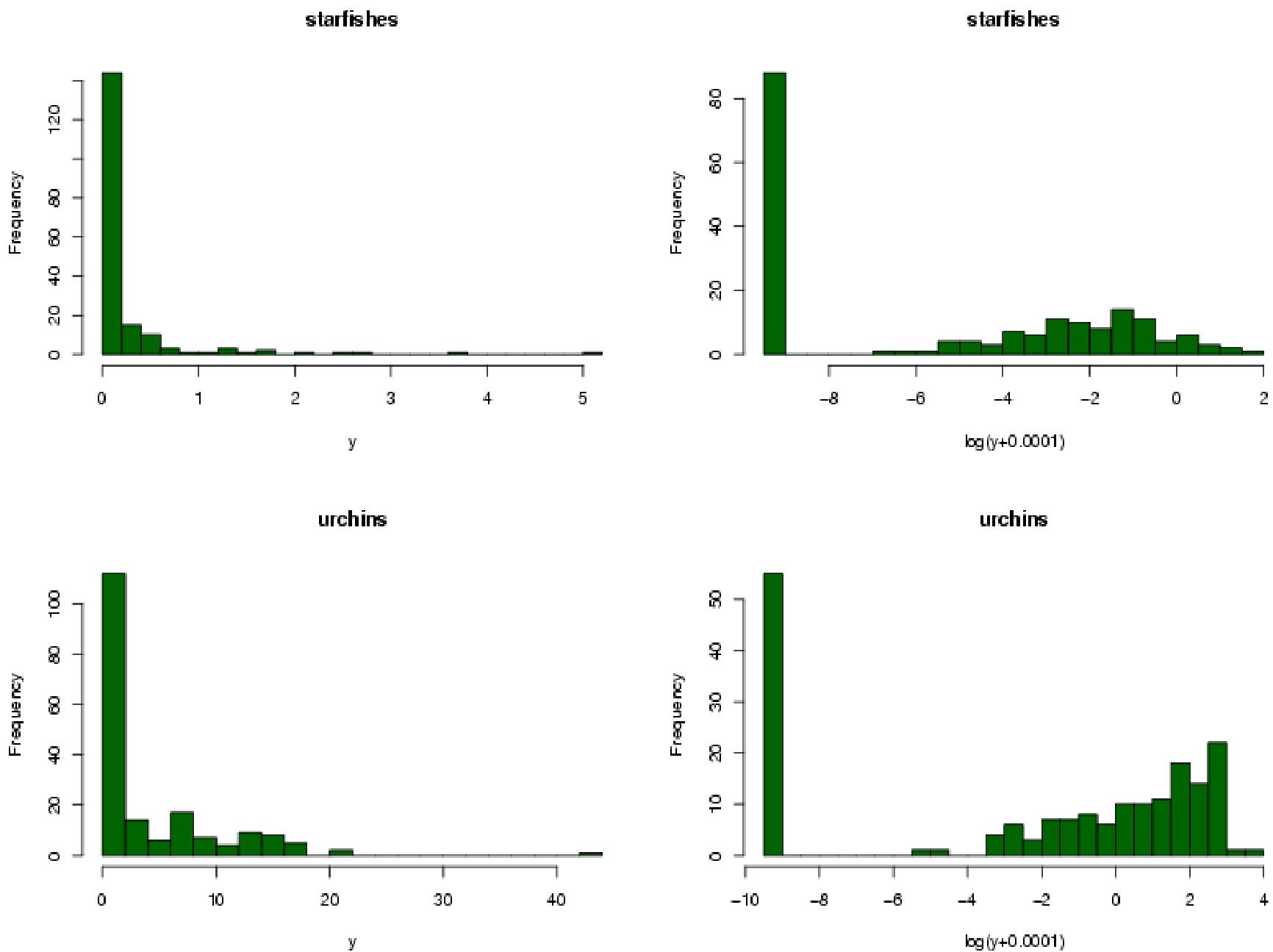
strate	latitude	longitude	depth	temperature	totconsum	primary	starfish
401	46,5003	-61,999	26,5	14,9	0,19407908	gravel with occasional sand patches	0,456
401	46,8368	-63,8482	25	11,4	0,13635805	coarse sand	0
401	46,6475	-63,7232	19	11,7	0,15432443	coarse sand	0
401	46,5448	-63,253	30,5	3,7	0,5569532	coarse sand	0
401	46,4845	-62,73	19,5	13,9	0,09576385	gravel with occasional sand patches	0,002
402	46,0482	-63,4303	19,5	18,2	0,01246536	pelite	0
402	45,9542	-63,4572	17	16,8	0,05808044	pelite	0,001
402	45,8683	-63,2725	15	16,3		pelite	0
403	45,8712	-61,8452	28	13,4	0,69245353	pelite	0,024
403	45,7037	-61,7698	18,5	19,3	1,31965306	gravel with occasional sand patches	0
403	45,7772	-61,6212	26,5	17,6	2,11126564	pelite	0
403	45,8392	-61,6037	24	14,8	0,64985972	gravel with occasional sand patches	0
415	48,5592	-63,1627	207	4,7	0,17525881	fine sand	0
415	48,8453	-63,6463	251	5,1	0,37630839	pelite	0
415	48,9662	-63,8677	313,5	5,4	0,16415286	pelite	0
415	48,9933	-64,253	190	4,1	0,28090004	pelite	0
415	48,8862	-63,9892	228,5	4,6	0,29882976	pelite	0
415	48,851	-63,8143	200	4,4	0,15000818	pelite	0
416	48,5197	-63,3615	125,5	3,4	0,03399162	glacial drift	0
416	48,5487	-64,0977	100,5	1,9	0,02270026	pelite	0
416	48,6447	-63,9293	128	2,8	0,02925672	coarse sand	0,021
416	48,6887	-63,6857	135,5	3,2	0,05543526	glacial drift	0,026
416	48,653	-63,631	140,5	2,6	0,04041808	glacial drift	0,074

The Challenges...

- Define a random structure to represent bottom-trawl surveys data



- Find a parsimonious alternative to a mixture model with delta distribution

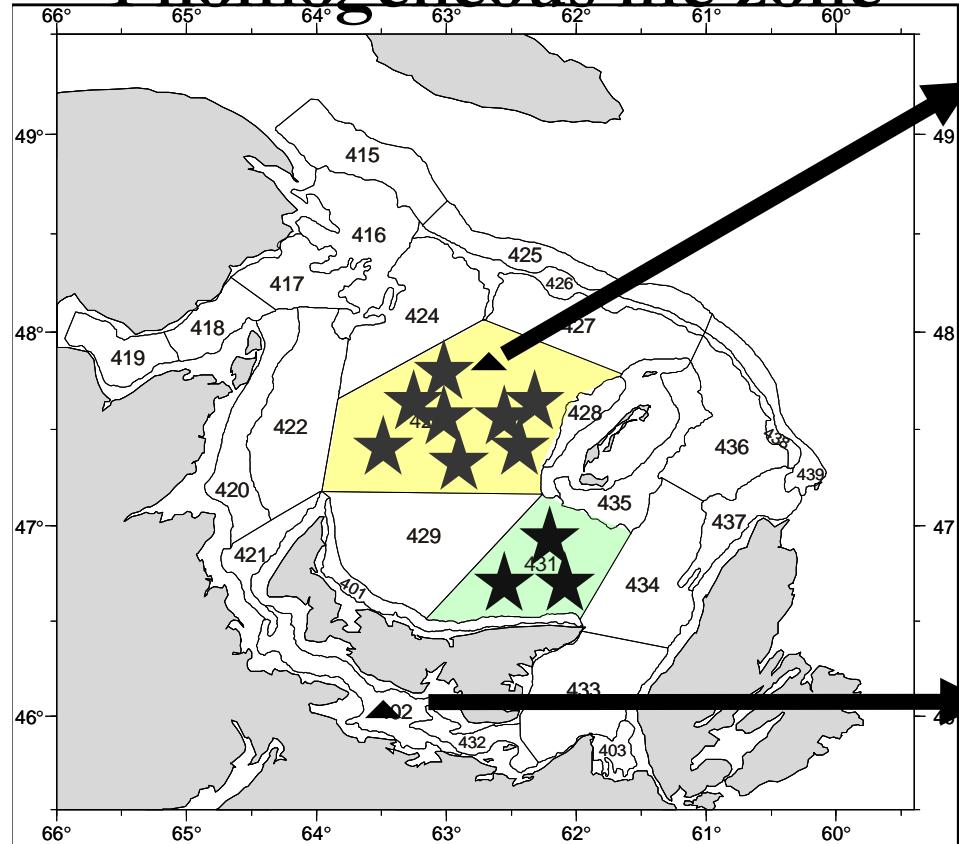


Hypotheses behind la « loi des fuites »

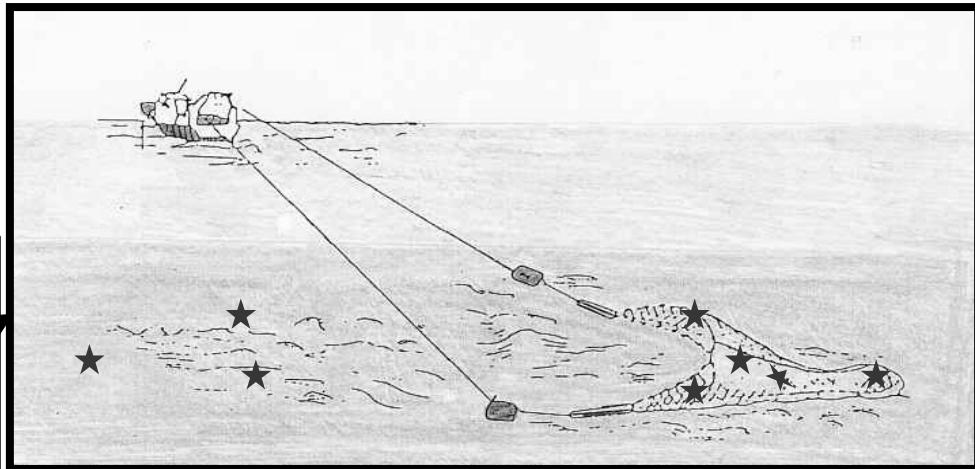
1 strata

1

1 homogeneous life zone



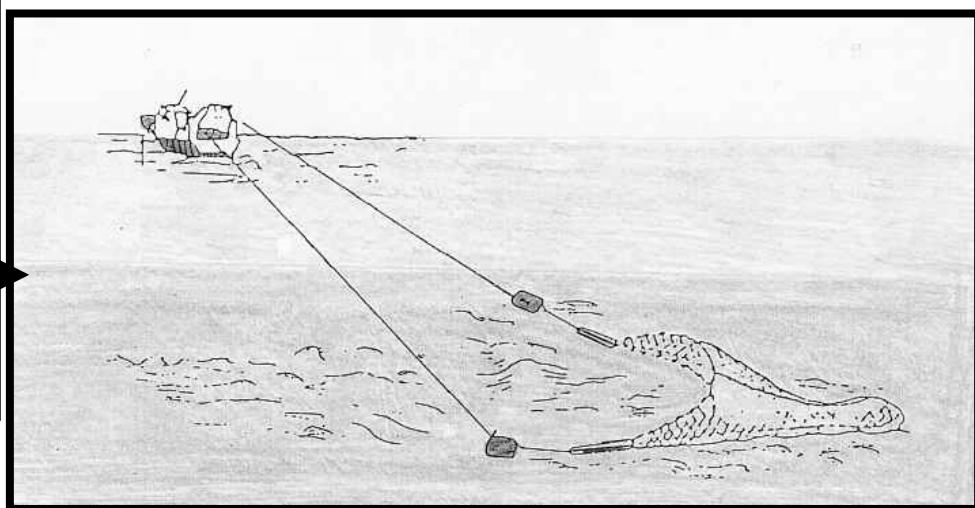
\star = a clump with a random biomass quantity



Case 1: Some biomass is collected

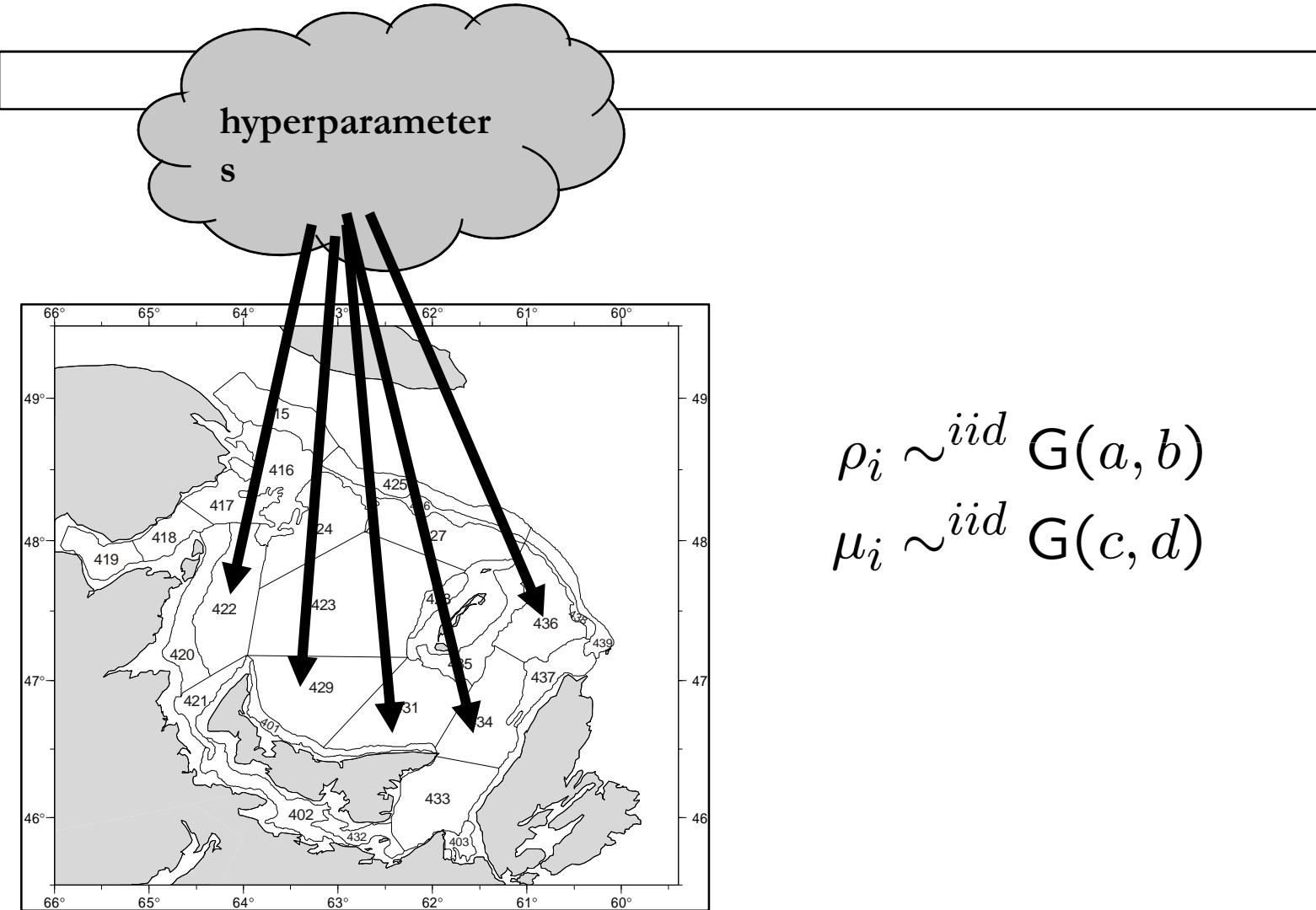
二

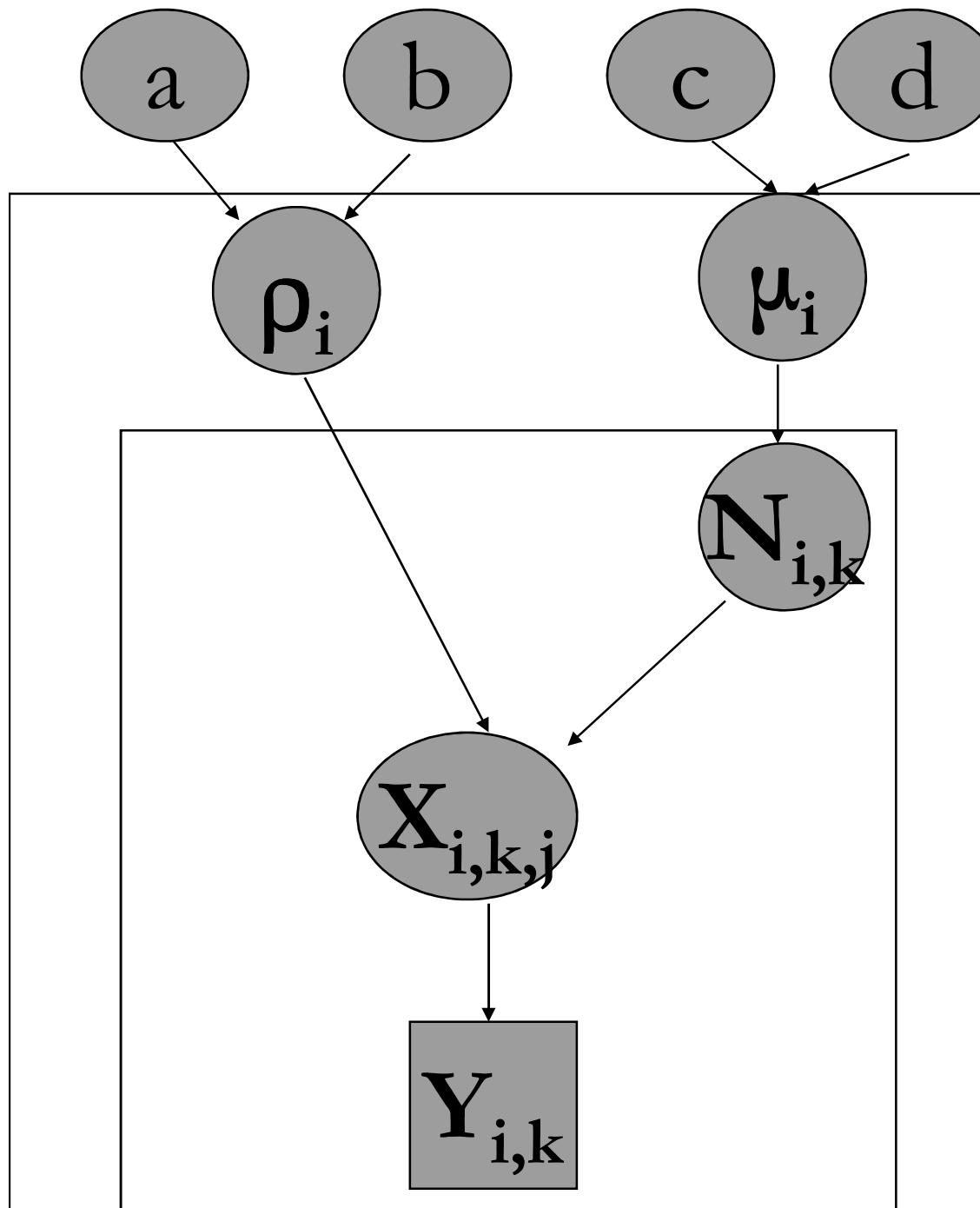
The sum in each clump



Case 2: No biomass is caught

Spatial Model 1 : Régionalisation





Between strata
variability

$$\rho_i \sim^{iid} G(a, b)$$

$$\mu_i \sim^{iid} G(c, d)$$

Within strata
variability : la
loi des fuites

$$N_{i,k} \sim^{iid} \text{Pois}(\mu_i)$$

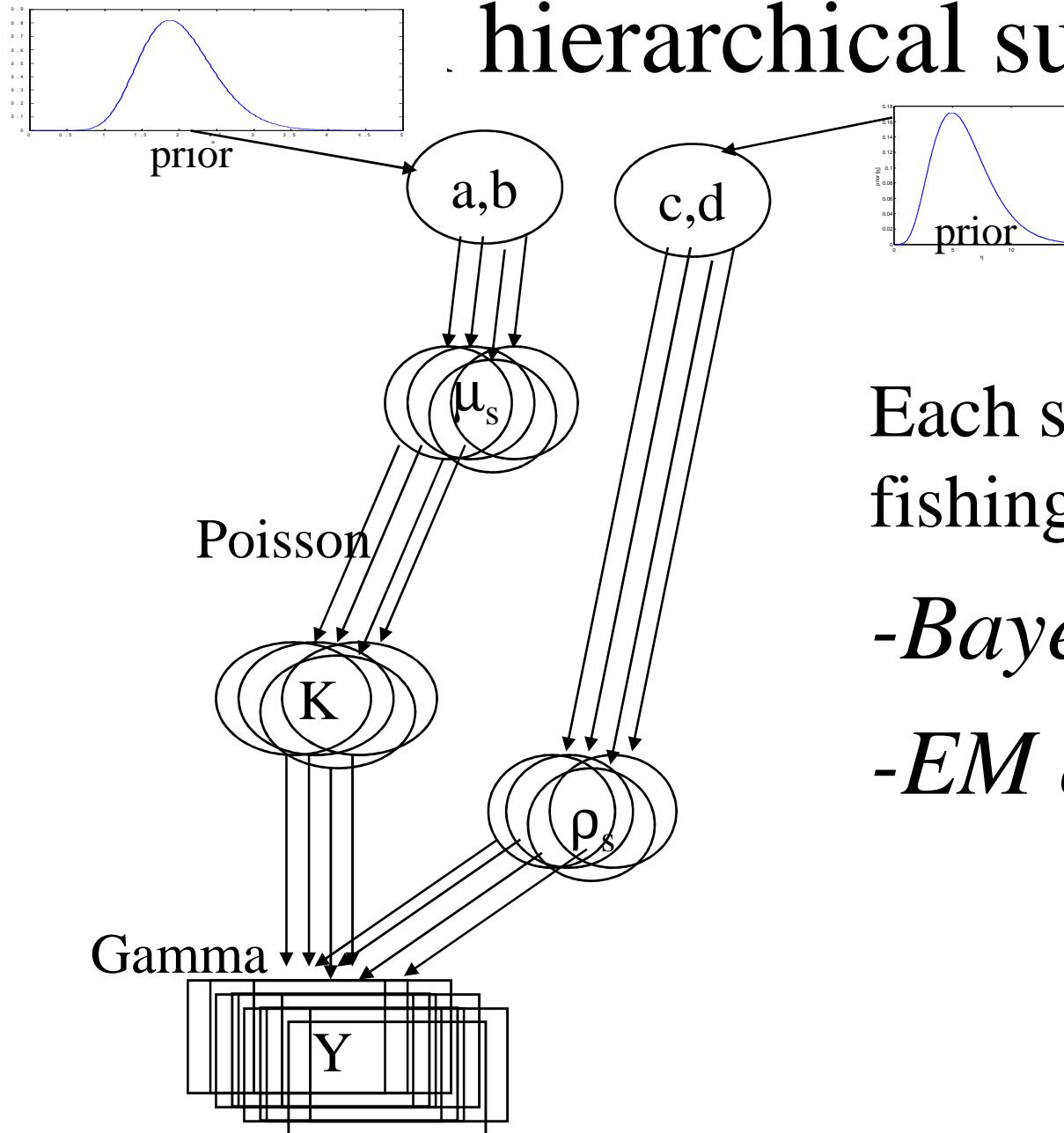
$$X_{i,k,j} \sim^{iid} \text{Exp}(\rho_i)$$

$$Y_{i,k} = \sum_{j=1}^{N_{i,k}} X_{i,k,j}$$

$$[Y_{i,k}|N_{i,k} = 0] \sim \delta_0$$

$$[Y_{i,k}|N_{i,k} = n] \sim G(n, \rho_i)$$

Inference of a mixed effect model hierarchical submodels



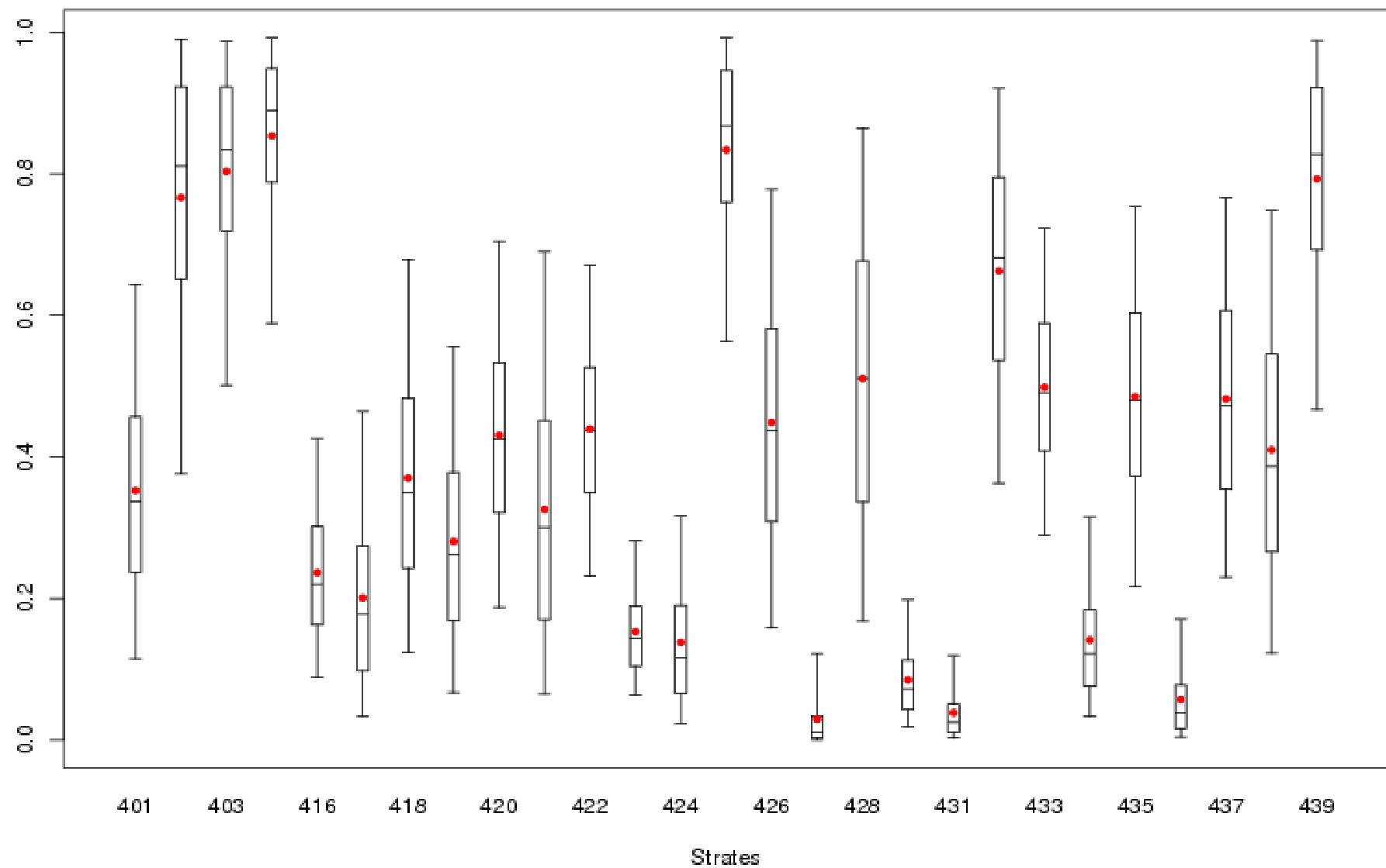
Each strata s is a fishing zone (μ_s, ρ_s)

-Bayesian techniques

-EM algorithms

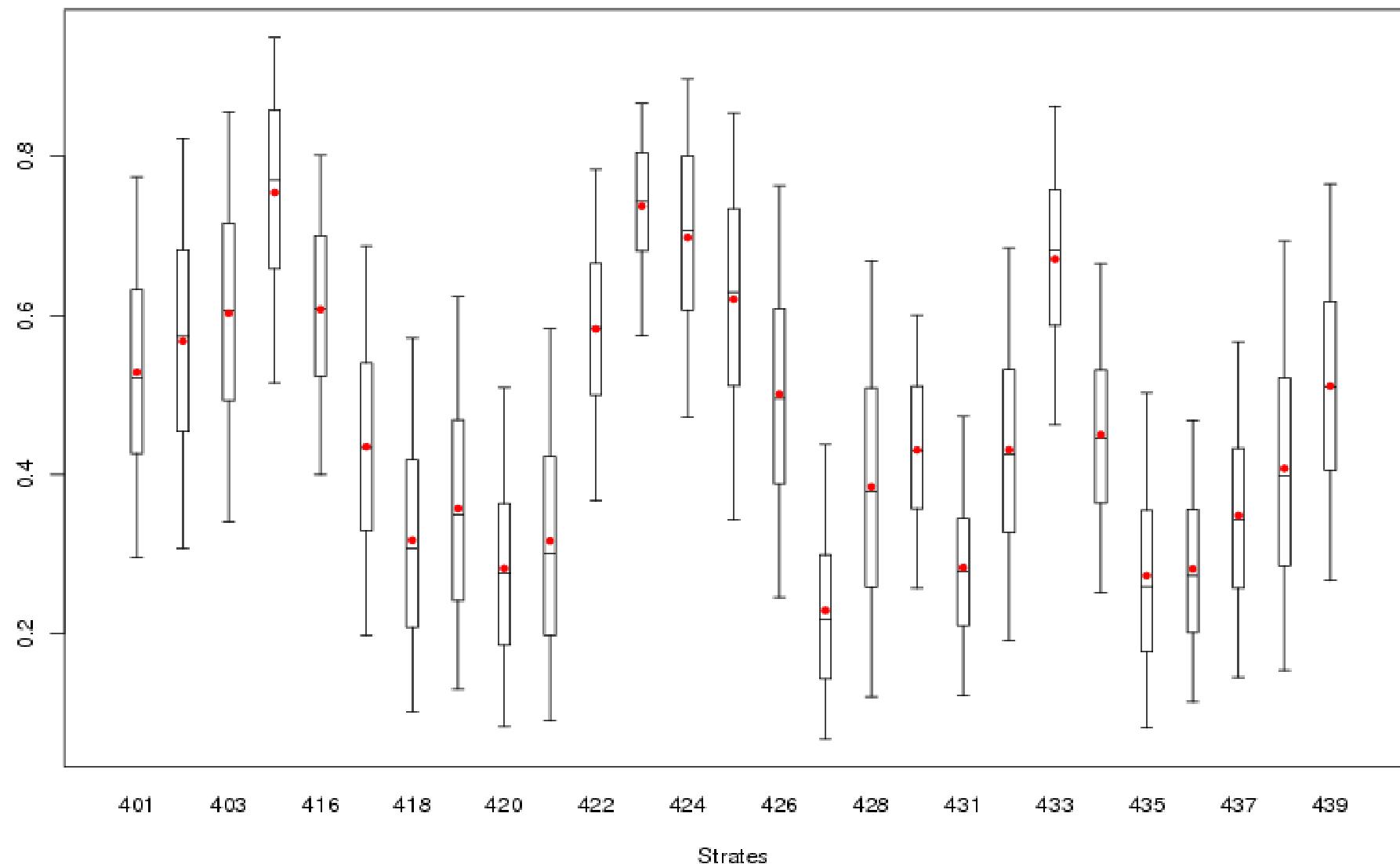
Urchins in 2002

Probabilité conditionnelle d'absence de l'espèce

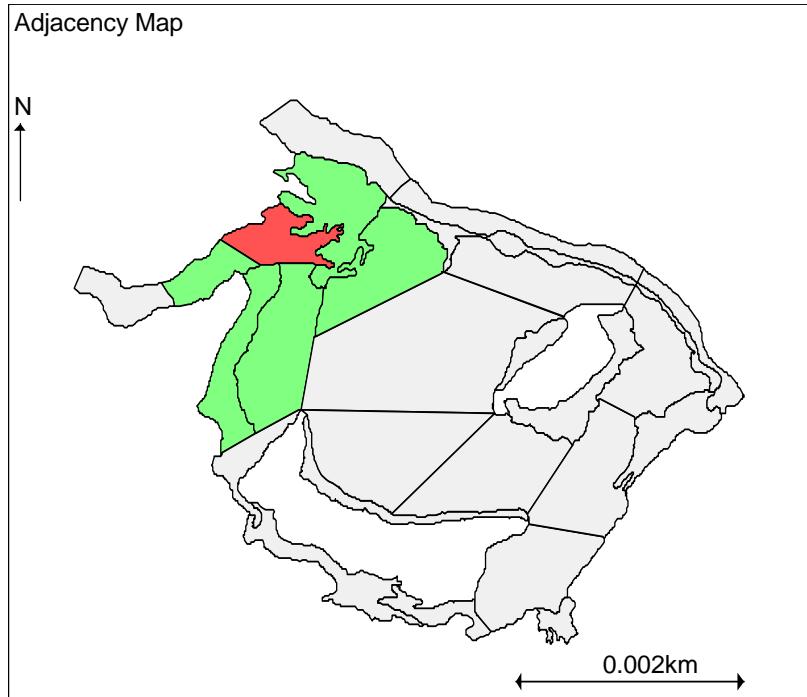


Starfish in 2002

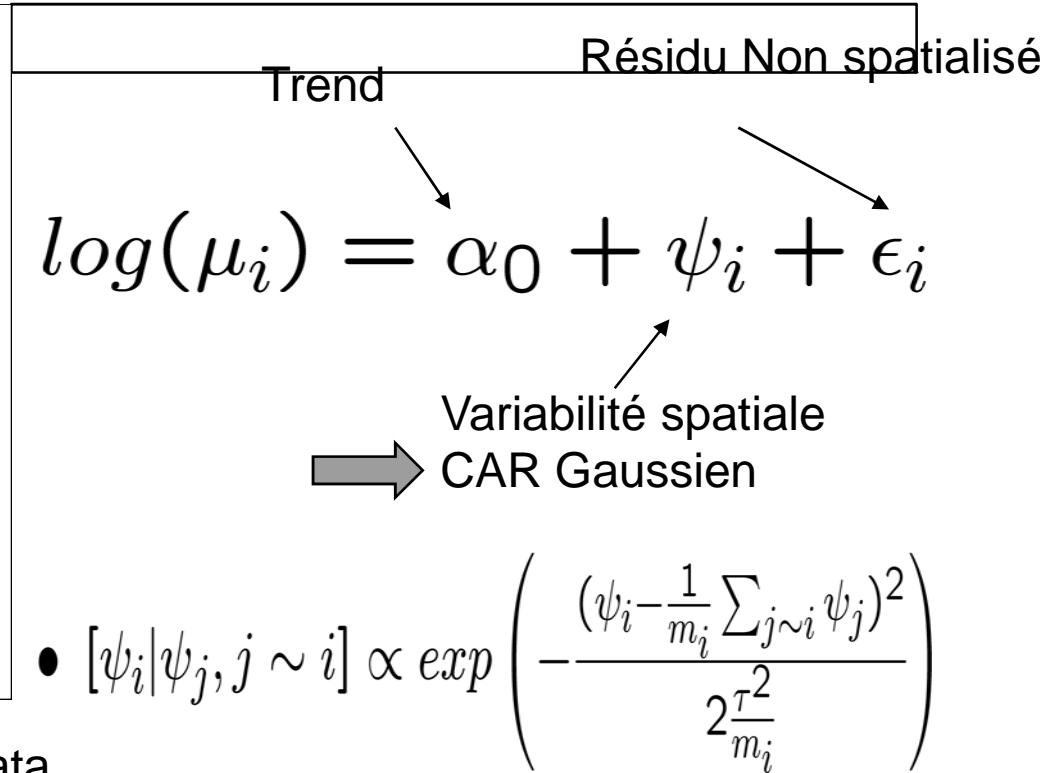
Probabilité conditionnelle d'absence de l'espèce



Spatial model 2: IAR on a lattice

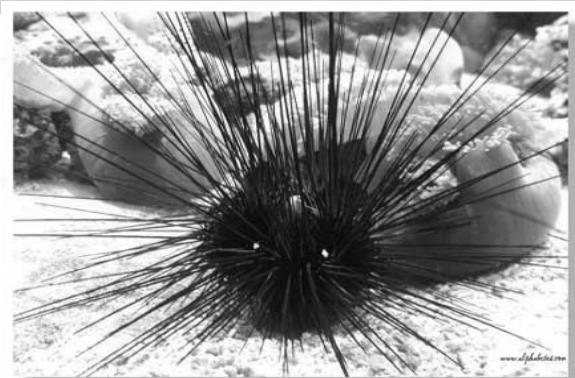
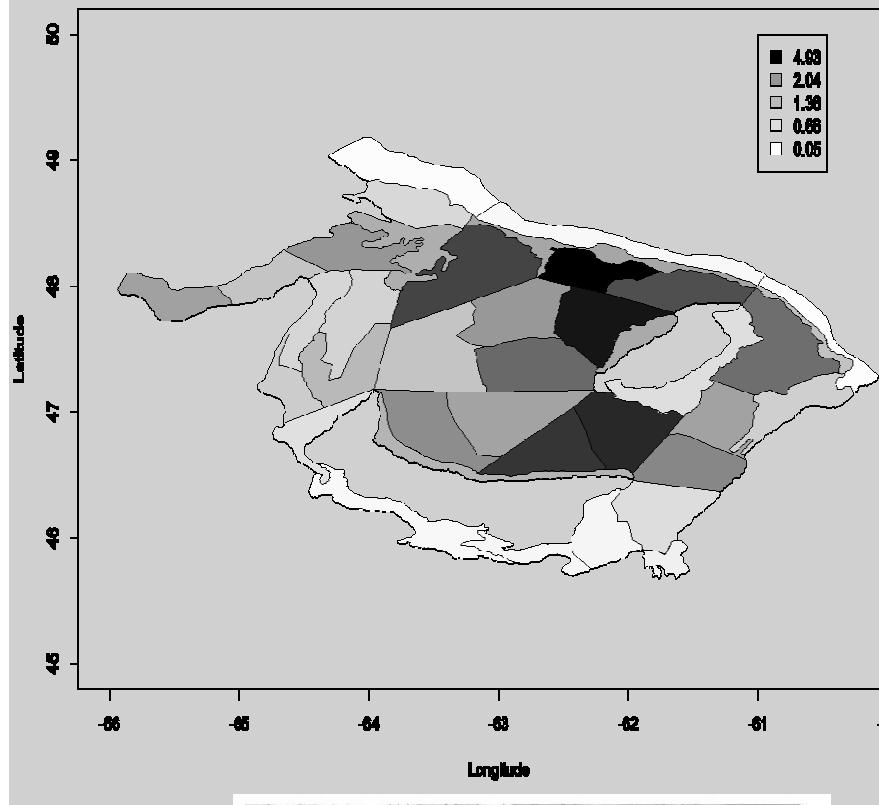


Neighboring cohérency between strata



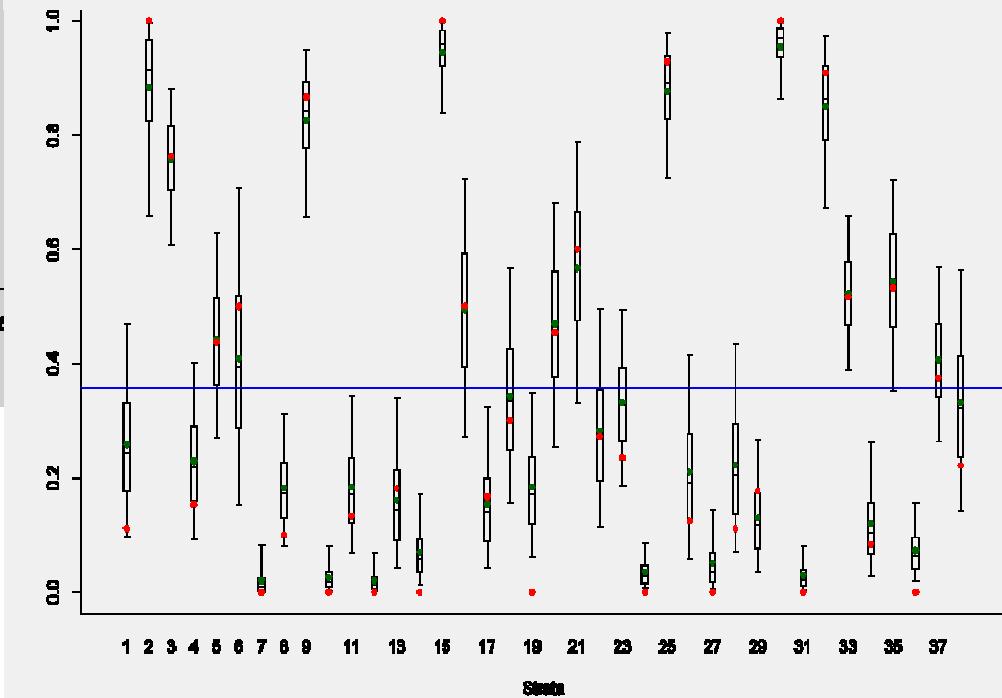
- $\epsilon_i \sim^{iid} \mathcal{N}(0, \sigma^2)$

Carte des moyennes a posteriori pour μ_i

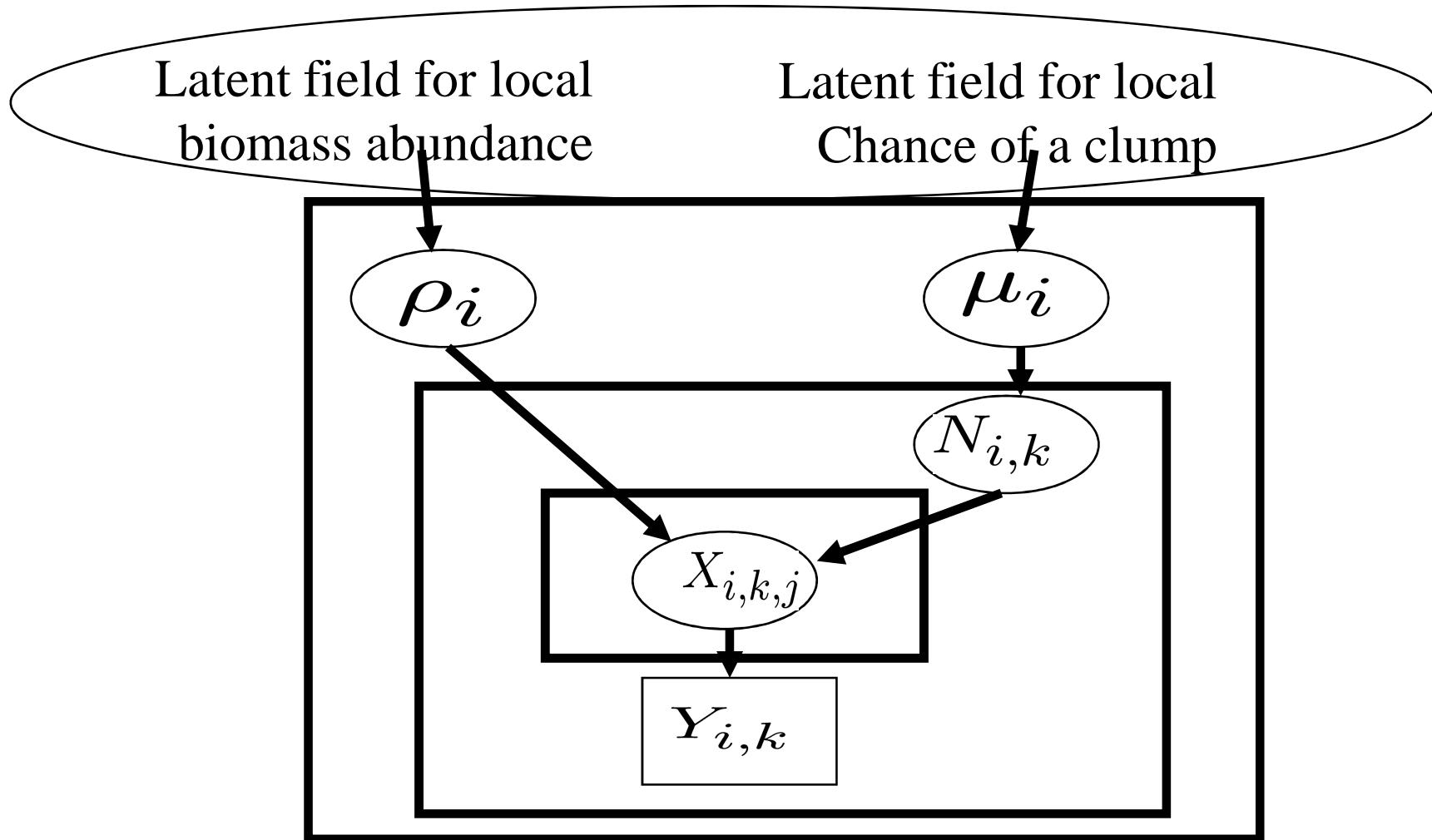


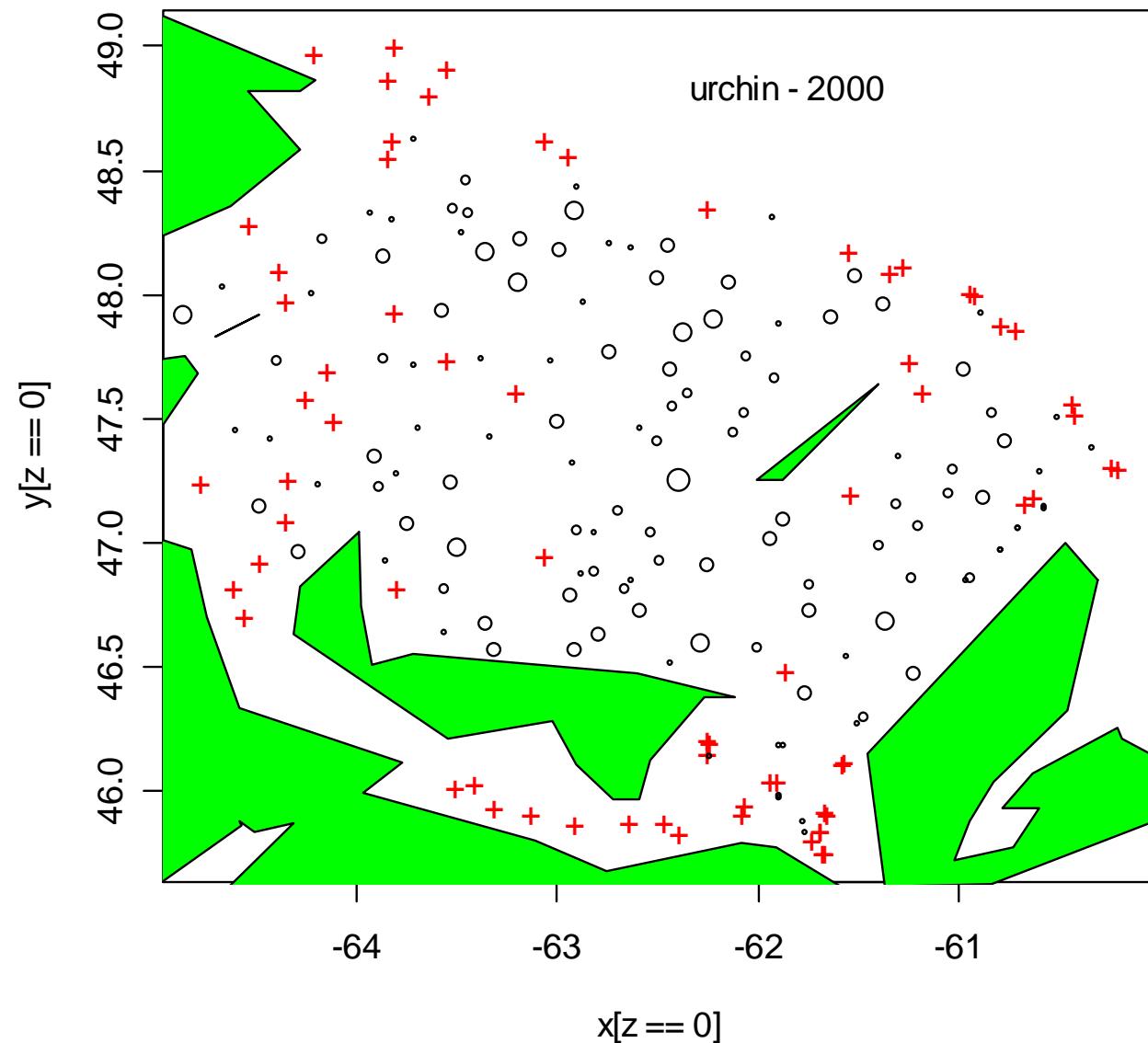
Inférence sur μ_i

Distributions a posteriori pour $\exp(-\mu_i)$

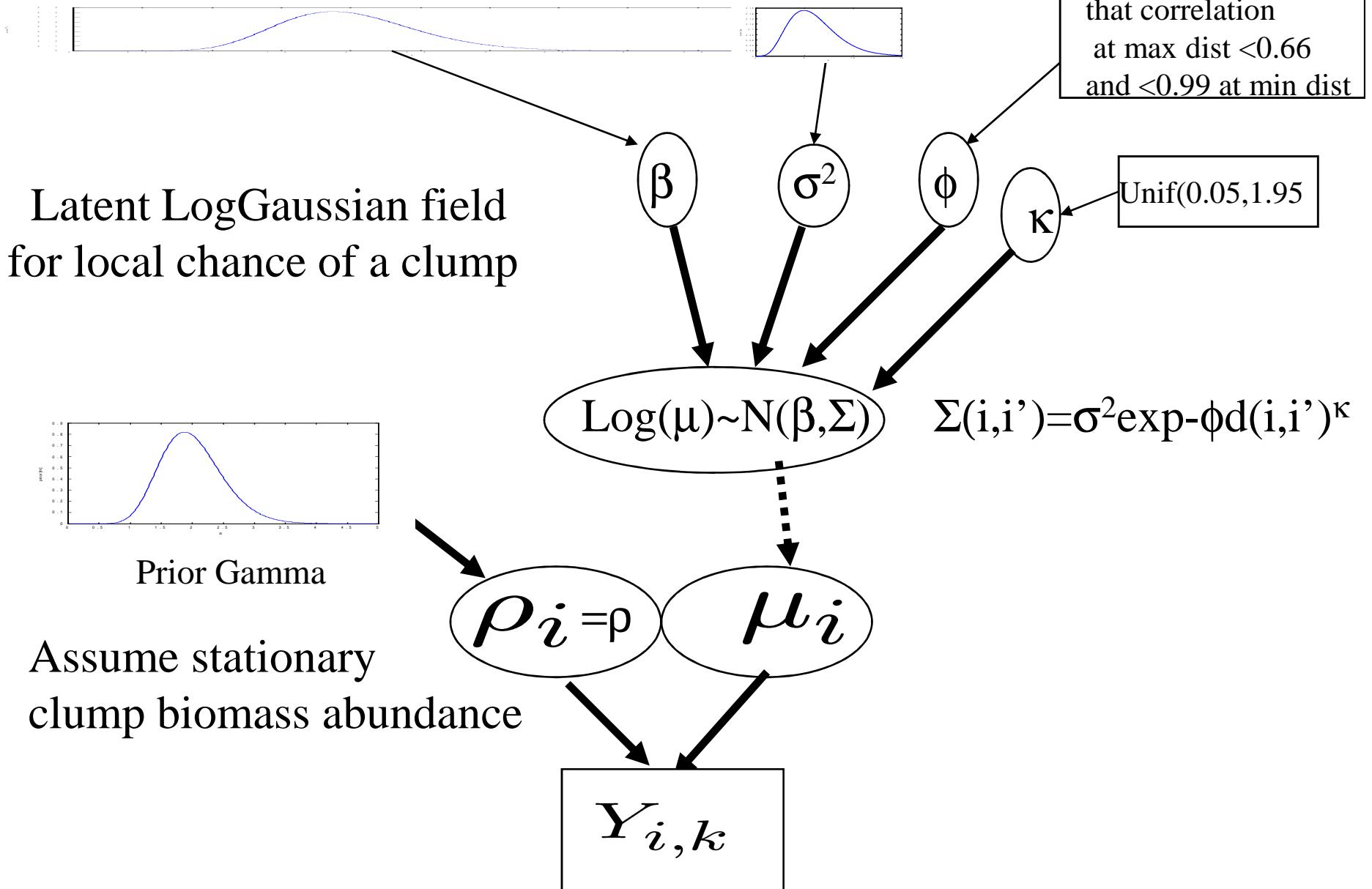


Spatial model 3: A three stage hierarchical model

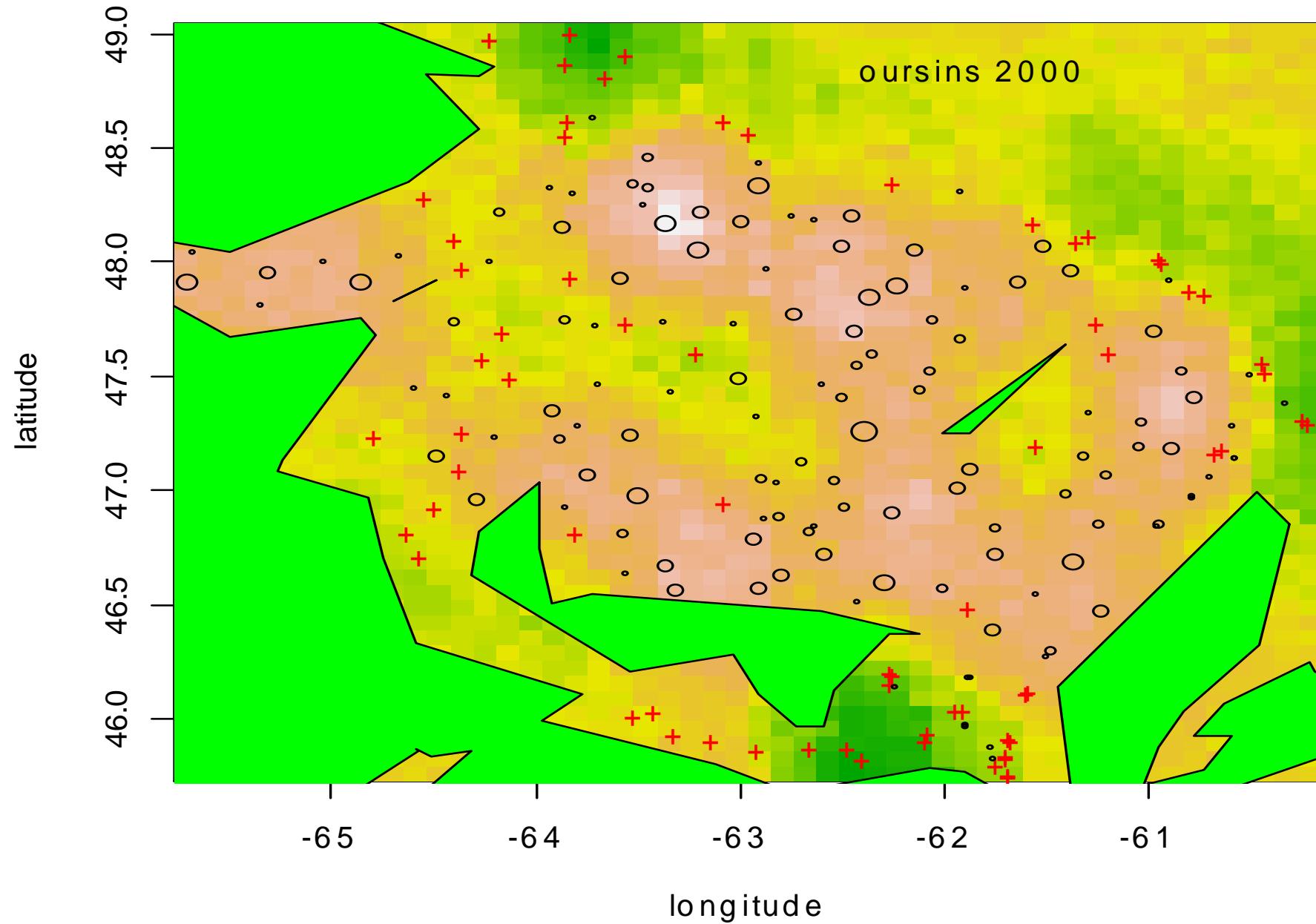




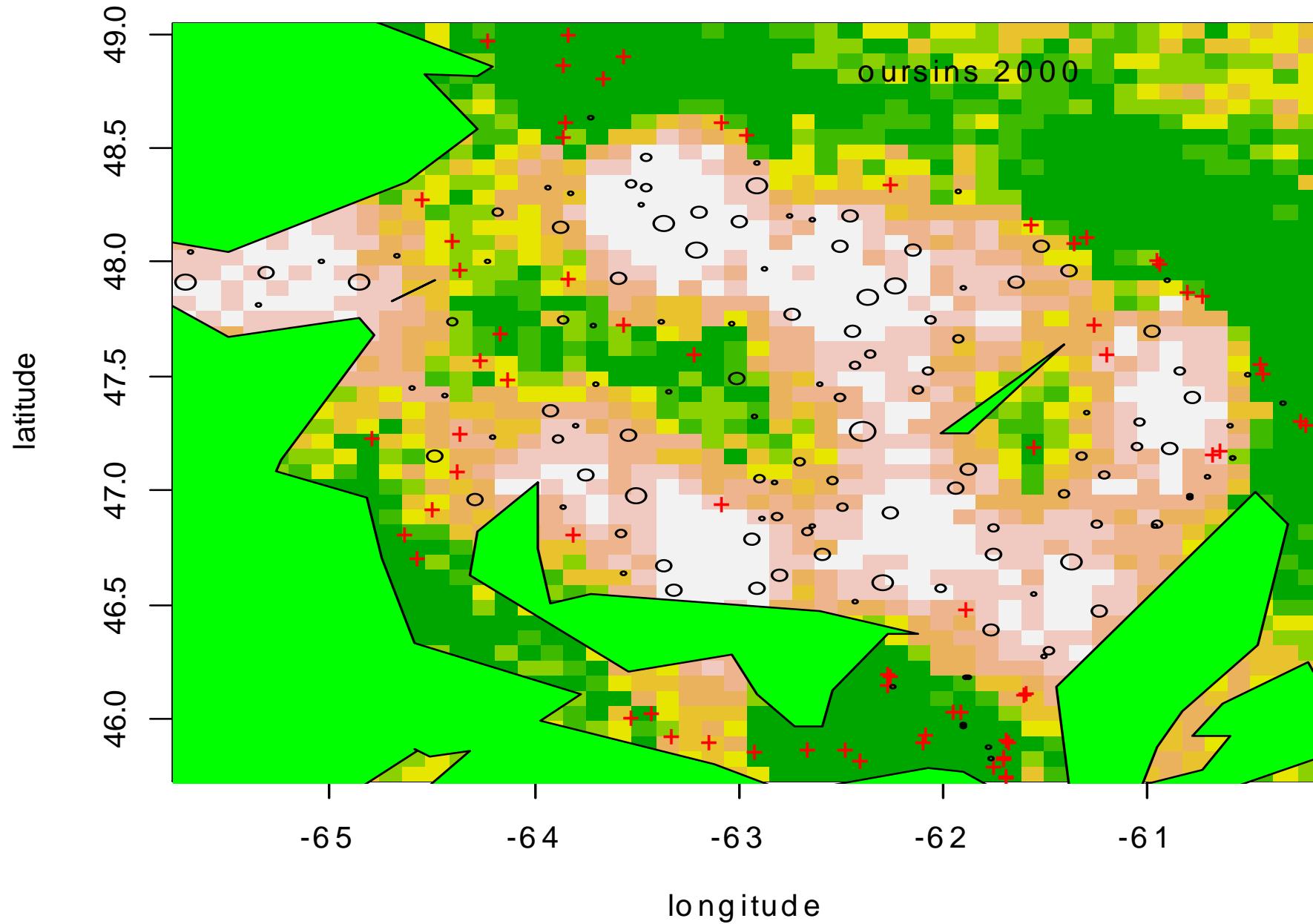
Model Structure



Champ Gaussien latent Moyen



Mediane de la repartition spatiale predictive de la biomasse



La « loi des fuites »

Conclusions

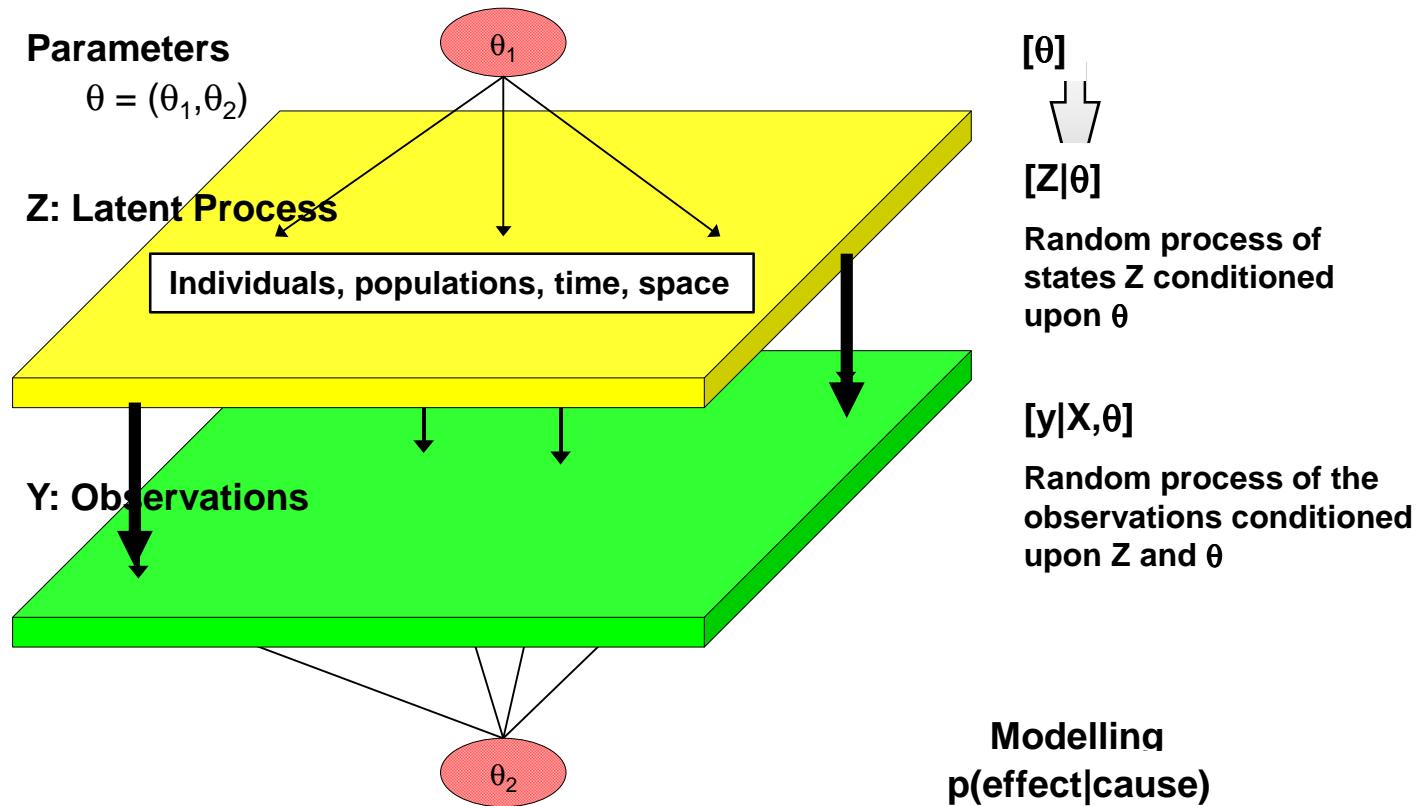
- *Parsimonious*
- *Conceptual interpretation*
- *Inference benefits from conditional structure*
- *Infinitely divisible*
- *Lego brick for various constructions*

Further tracks

- Multivariate constructions
- Spatio-dynamics modelling
- Normal Convolutions
 - following *Calder, Wickle, Craigmiles...*
- Gamma Convolutions
 - following *Wolpert, R.L. and Ickstadt, K. (1998). Poisson/Gamma random field models for spatial statistics. Biometrika, 85, 251-267.*
- Markov Point Processes
 - following *Berliner*

Conditionnal modelling strategy in Hierarchical Bayesian Models

- **Modelling :** Capacity to accommodate complexity
Parameters, states process and observations process can be modelled independently



General Comments

- *Easy for modelling*

- Thinking conditionally breaks an uneasy problem into tractable pieces
- convenient to incorporate knowledge within the various layers of the structure
- easy to adapt/improve because locally defined
- mind over-modelling...

- *Handy for inference*

- inputs similar to outputs and probability distributions
- probabilistic formalization by reverse conditionning
- software available (AppliBUGS ENGREF since 2007)

Quelques bonnes lectures

- Redécouvrir la théorie du Risque en Environnement . Revue de l 'AIGREF, n° spécial sur les risques en 2003 , pages 18 à 24.
- Bernier J., Parent E., Boreux. JJ. (2000), *Statistique pour l 'Environnement.* LAVOISIER, Eds TEC et DOC.
- Bernier J., Parent E. (2007), Le *Raisonnement Bayésien* . Springer Verlag France