Estimation des indicateurs de dépendance

ACI

10 mai 2006

Mesure de la dépendance

Structuration des extrêmes : dépendance spatiale des extrêmes $\max(Z(s))$ et $\max(Z(s+h))$ sont dépendants si h est petit.
Calcul de

$$\theta(h) = \theta(Z(s), Z(s+h))$$

Si

- $1 \leq \theta(h) < 2$, $Z(s)$ et $Z(s+h)$ sont asymptotiquement dépendants
- $\theta(h) = 2$, $Z(s)$ et $Z(s+h)$ sont asymptotiquement indépendants ou indépendants

Estimation de $\theta(h)$

4 approches :

- *Maximum de vraisemblance,*
- *fonction de dépendance*
- *madogramme transformé*
- *madogramme puissance*

I) Maximum de vraisemblance (Schlather-Tawn (2003))

- k répétitions indépendantes $(Z_i^{(1)}, Z_i^{(2)})$
- marginales Fréchet
- loi de $\max(Z^{(1)}, Z^{(2)}) \sim \exp(-\theta/x)$, au-dessus d'un seuil z ;
- $t = \exp(-1/z)$

Estimateur du maximum de vraisemblance censuré :

$$\widehat{\theta}_{MV} = \frac{\mathrm{card}\left\{ j : \max_{i\in\{1,2\}} (Z_i^{(j)} \bar{X}_i) > z \right\}}{\sum_{j=1}^{k} \left[\max\left\{ z, \max_{i\in\{1,2\}} (Z_i^{(j)} \bar{X}_i) \right\} \right]^{-1}}$$

avec $\bar{X}_i = k^{-1} \sum_{j=1}^{k} \frac{1}{Z_i^{(j)}}$ $\qquad$ $(\widehat{\theta}_{MV} = 1$ quand $t = 0)$.

II) Estimation non paramétrique de la fonction de dépendance

Loi bivariée des valeurs extrêmes :

$$G(x,y) = exp\left\{-\left(\frac{1}{x}+\frac{1}{y}\right)A\left(\frac{x}{x+y}\right)\right\}$$

où $A(\cdot)$ fonction convexe de $[0,1]$ sur $[1/2,1]$ (Pickands (1981)).

On a : $\theta = 2A(\frac{1}{2})$.

$\leadsto$ Estimation non paramétrique de $A(\cdot)$; (Capéraà, Fougères and Genest (1997) ; R *evd* package, Stephenson (2003))

III) Estimation par le madogramme (Poncet, Cooley and Naveau (2005))

Madogramme :

$$\nu(h) = \frac{1}{2}\mathbb{E} \mid Z(s+h) - Z(s) \mid$$

Si $Z(\cdot)$ processus max-stable marginales Fréchet .
Madogramme transformé

$$\nu_F(h) \equiv \frac{1}{2}\mathbb{E} \mid F(Z(s+h)) - F(Z(s)) \mid$$

De $\mid X - Y \mid = 2\max(X, Y) - (X + Y)$, on tire

$$\nu_F(h) = \frac{\theta(h) - 1}{2(\theta(h) + 1)}$$

et

$$\boxed{\widehat{\theta}_{PCN}(h) = \frac{1 + 2\hat{\nu_F}(h)}{1 - 2\hat{\nu_F}(h)}}$$

IV) Estimateur par moindres carrés

$\lambda \in \mathbb{R}$, $\nu_F^{\lambda}(h) = \frac{1}{2}\mathbb{E} \mid F^{\lambda}(Z(s+h)) - F^{\lambda}(Z(s)) \mid$

$$\boxed{\nu_F^{\lambda}(h) = \frac{\lambda(\theta - 1)}{(\theta + \lambda)(\lambda + 1)}}$$

$\lambda_1, \ldots, \lambda_m$ réels , $g_{\theta(h)}(\lambda_i) = \dfrac{\theta(h) - 1}{\theta(h) + \lambda_i}$ et $y_i(h) = \dfrac{1 + \lambda_i}{\lambda_i} \nu_F^{\lambda_i}(h)$

$\rightsquigarrow$ estimateur (non-linéaire) des Moindres Carrés

$$\boxed{\widehat{\theta}_{MC}(h) = \mathrm{argmin}_{\theta} \sum_{i=1}^{m} (y_i(h) - g_{\theta}(\lambda_i))^2}$$

Etude sur simulations

2 modèles :

- (a) modèle Gaussien (*indépendance asymptotique*)
- (b) modèle tempête (*dépendance asymptotique*)

2 approches :

- *classique* : répétitions sur les couples $(Z(s), Z(s+h))$
- *spatial* : utilise l'ergodicité et la stationnarité de $Z(\cdot)$: les estimateurs sont calculés sur une seule réalisation

2 transformations :

- *Fonction de répartition empirique*
- *Fonction de répartition estimée par GEV*

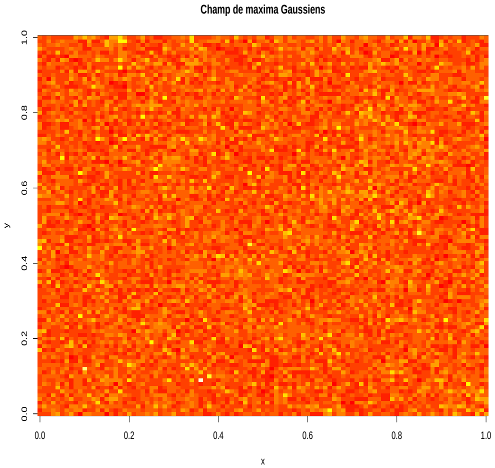
Exemple

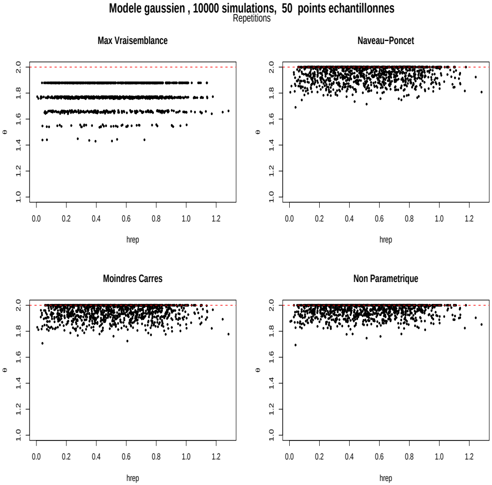
Champ Gaussien stationnaire isotrope de fonction de corrélation $\rho(h)$

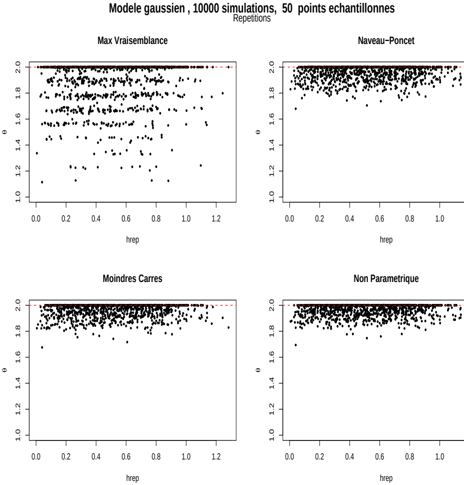
- $\theta(h) = 2$

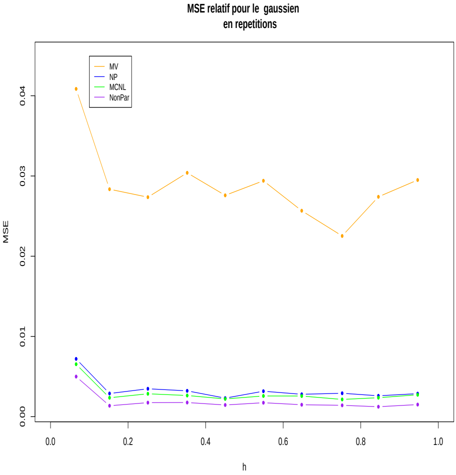
Asymptotiquement indépendant

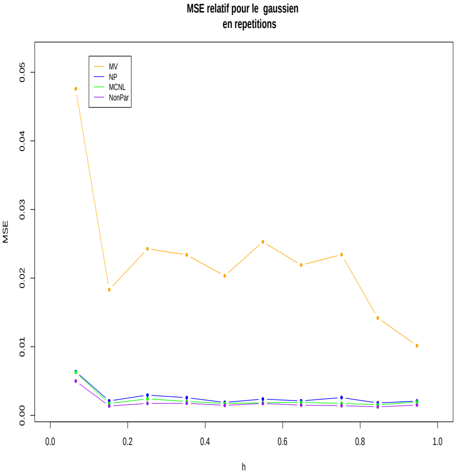
100 Champs Gaussien de fonction de corrélation exponentielle, de portée $a = 0.2$, carré unité, maximum en chaque point.

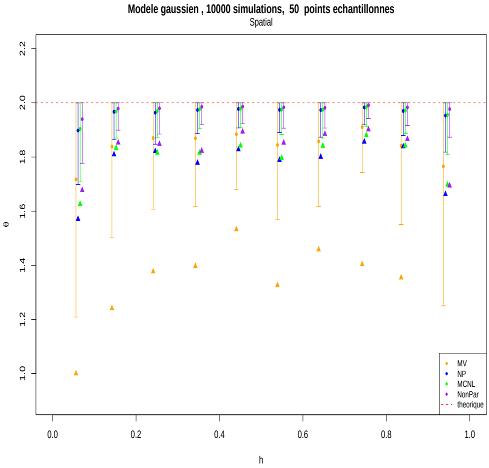

Champ de maxima Gaussiens
y
x
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0

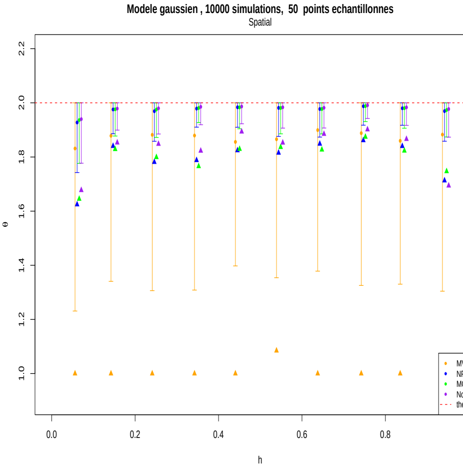

Modele gaussien , 10000 simulations, 50 points echantillonnes
Repetitions
Max Vraisemblance
Naveau−Poncet
Moindres Carres
Non Parametrique
θ
hrep

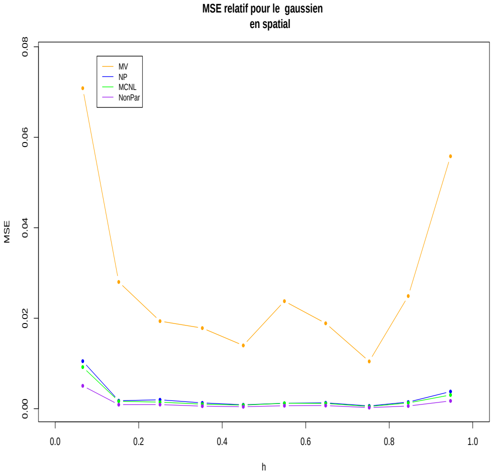

Modele gaussien , 10000 simulations, 50 points echantillonnes
Repetitions
Max Vraisemblance
Naveau−Poncet
Moindres Carres
Non Parametrique
θ
hrep

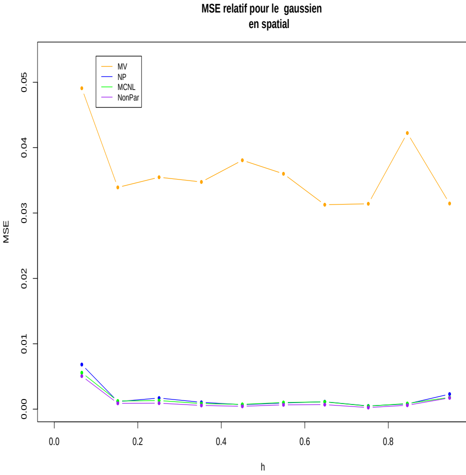

MSE relatif pour le gaussien
en repetitions
MV
NP
MCNL
NonPar
MSE
0.00
0.01
0.02
0.03
0.04
0.0
0.2
0.4
0.6
0.8
1.0
h


MSE relatif pour le gaussien
en repetitions
MV
NP
MCNL
NonPar
MSE
0.00
0.01
0.02
0.03
0.04
0.05
0.0
0.2
0.4
0.6
0.8
1.0
h


Modele gaussien , 10000 simulations, 50 points echantillonnes
Spatial
θ
h
MV
NP
MCNL
NonPar
theorique


Modele gaussien , 10000 simulations, 50 points echantillonnes
Spatial
θ
h


MSE relatif pour le gaussien
en spatial
MV
NP
MCNL
NonPar
MSE
0.00
0.02
0.04
0.06
0.08
0.0
0.2
0.4
0.6
0.8
1.0
h


MSE relatif pour le gaussien
en spatial
MV
NP
MCNL
NonPar
MSE
0.00
0.01
0.02
0.03
0.04
0.05
0.0
0.2
0.4
0.6
0.8
h

Modèles tempête

φ une fonction aléatoire t.q. $\mu = E\left(\int_{\mathbf{R}^2} \max(0, \varphi(x))dx\right) \in]0; \infty[$
π processus de Poisson sur $\mathbb{R}^2 \times]0; +\infty[$, d'intensité $\mu^{-1} dy d\zeta^{-2} d\zeta$

$$Z(s) = \sup_{(y,\zeta)\in\pi} \zeta \varphi_{y,\zeta}(s - y)$$

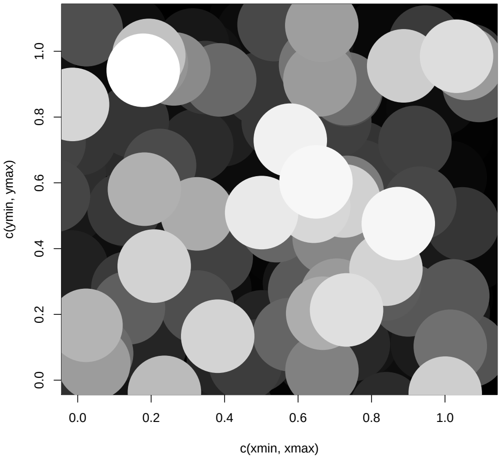
où $\varphi_{y,\zeta}$ répliques i.i.d de φ.
Z processus max-stable stationnaire à marginales Fréchet unité.
φ est la forme de la tempête,
aléatoire (disques de rayon aléatoire ...)
déterministe (bi-Gaussienne de corrélation fixée).
ζ : l'intensité de la tempête
y est le centre de la tempête.

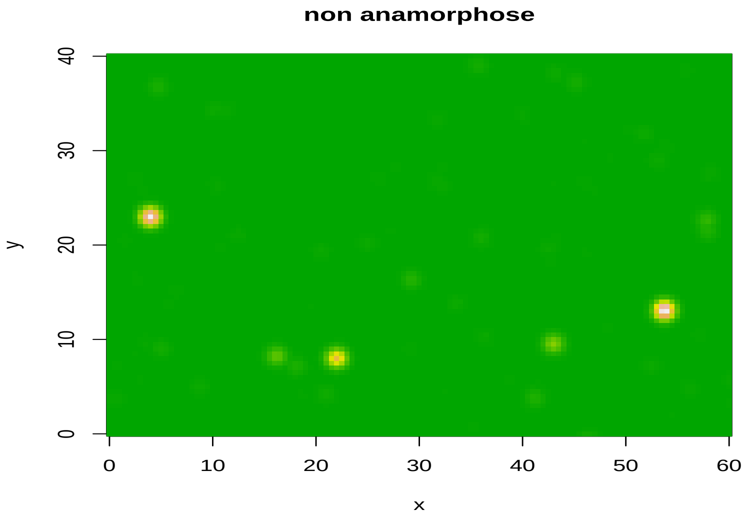
- Disques de rayon aléatoires

$$\theta(h) = 1 + \frac{1}{2}\min(1, h)\left(3 - \min(1, h)^2\right)$$

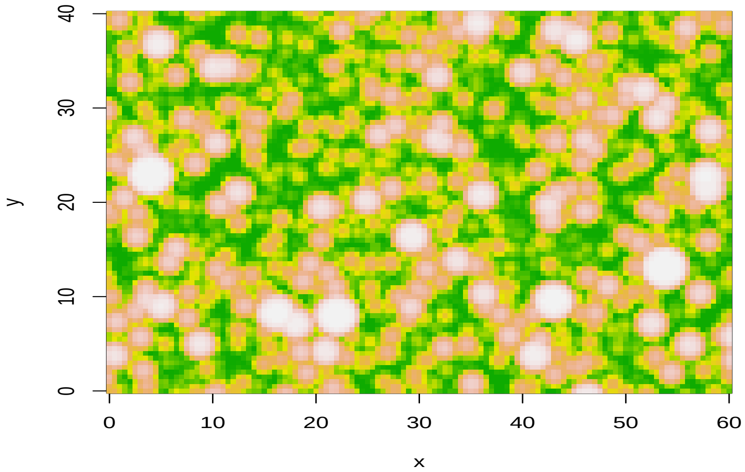
- tempête Gaussienne de matrice Σ

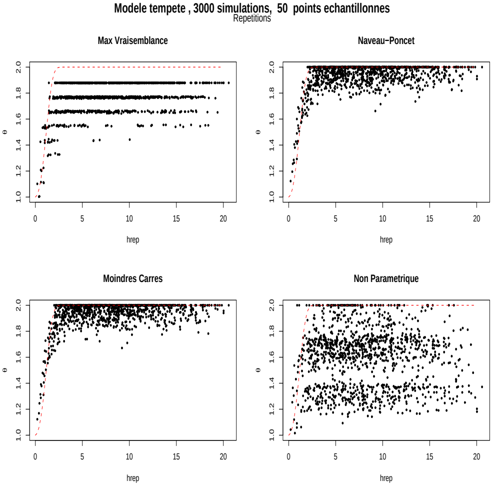
$$\theta(h) = 2\Phi(h^t \Sigma^{-1} h/2)$$

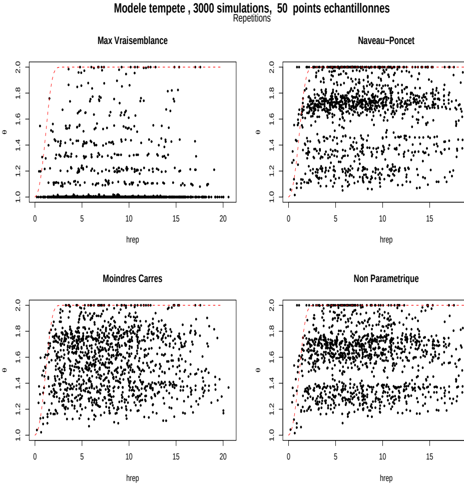

c(ymin, ymax)
0.0
0.2
0.4
0.6
0.8
1.0
c(xmin, xmax)

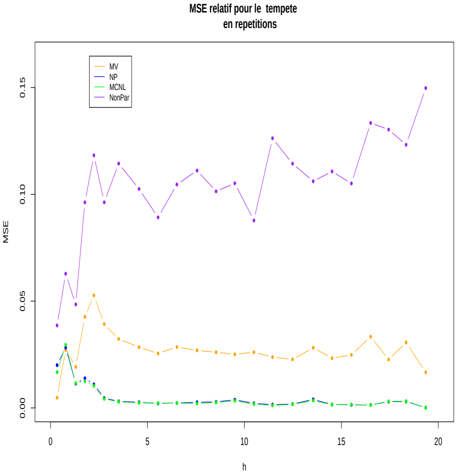

non anamorphose
y
0
10
20
30
40
x
0
10
20
30
40
50
60

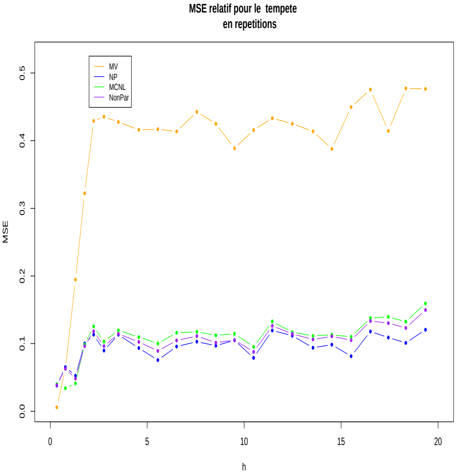
anamorphose

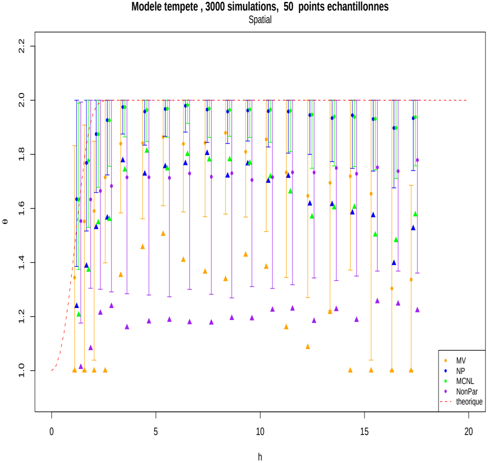


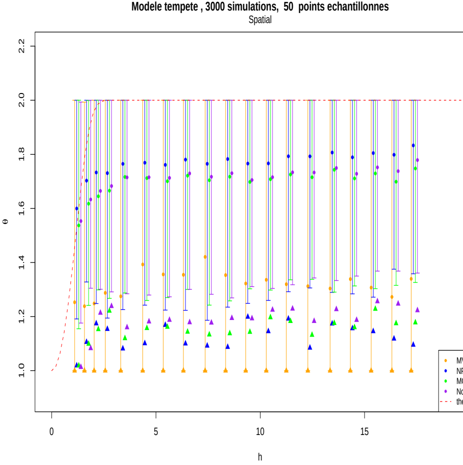

Modele tempete , 3000 simulations, 50 points echantillonnes
Repetitions
Max Vraisemblance
Naveau−Poncet
Moindres Carres
Non Parametrique
θ
hrep

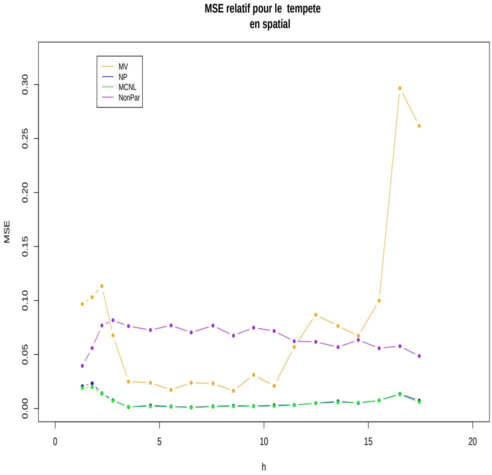

Modele tempete , 3000 simulations, 50 points echantillonnes
Repetitions
Max Vraisemblance
Naveau−Poncet
Moindres Carres
Non Parametrique
θ
hrep

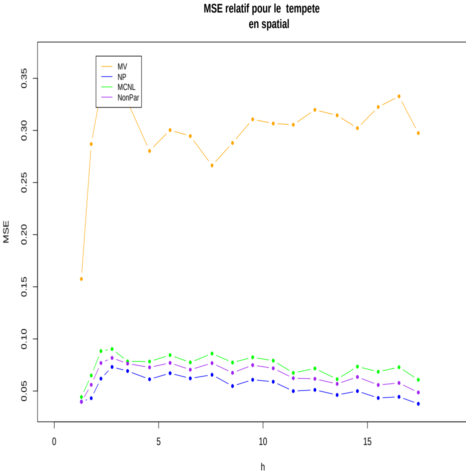

MSE relatif pour le tempete
en repetitions
MV
NP
MCNL
NonPar
MSE
0.00
0.05
0.10
0.15
0
5
10
15
20
h


MSE relatif pour le tempete
en repetitions
MV
NP
MCNL
NonPar
MSE
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20
h


Modele tempete , 3000 simulations, 50 points echantillonnes
Spatial
θ
h
MV
NP
MCNL
NonPar
theorique


Modele tempete , 3000 simulations, 50 points echantillonnes
Spatial
θ
h


MSE relatif pour le tempete
en spatial
MV
NP
MCNL
NonPar
MSE
h
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
5
10
15
20


MSE relatif pour le tempete
en spatial
MV
NP
MCNL
NonPar
MSE
h
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
5
10
15

Conclusions

- **Modèle Gaussien** :

- peu de différence uniforme-GEV, sauf MV

- Approche classique : non paramétrique, MC, puis NCP

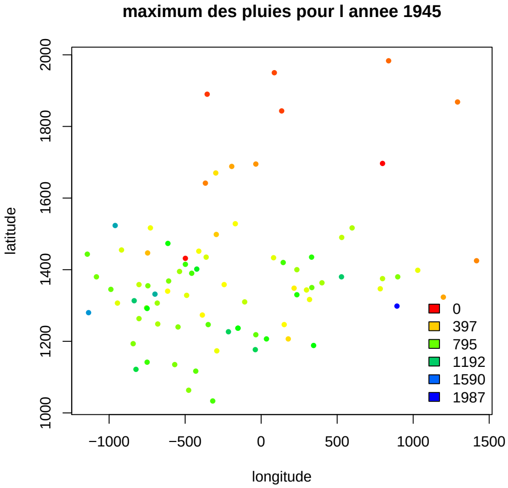
- Appproche Spatiale : dégradation pour les petites et les moyennes distances (moins de couples)

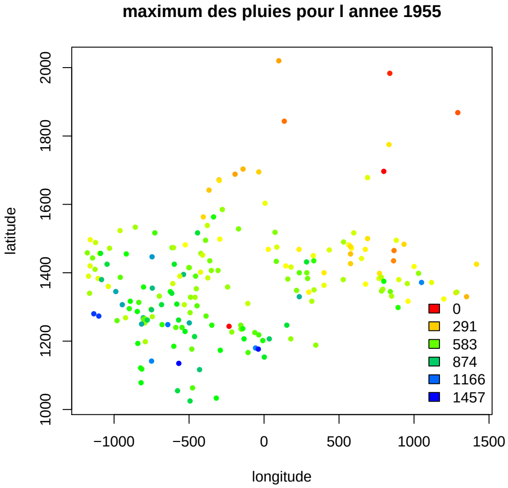
- **Modèle tempête** :

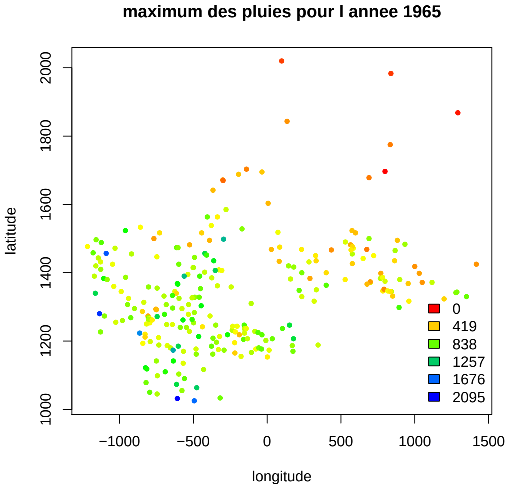
- très nette dégradation avec la GEV

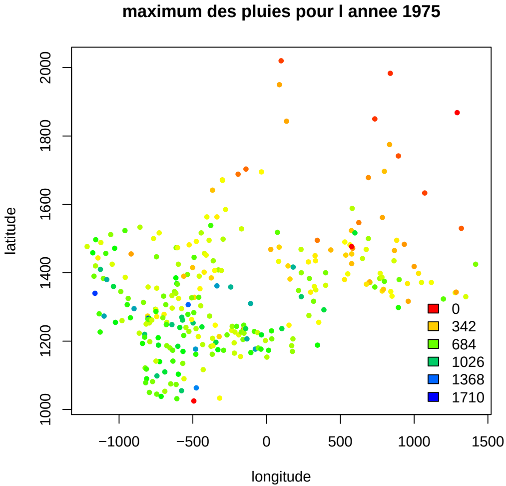
- Approche classique : MC,NCP,MV

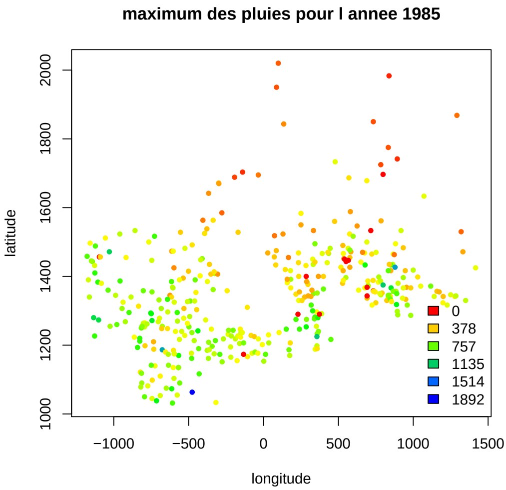
- Approche Spatiale : idem, à peine dégradé.

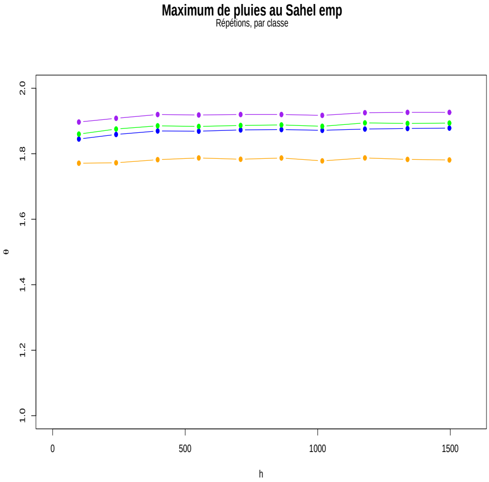

maximum des pluies pour l annee 1945
latitude
longitude
2000
1800
1600
1400
1200
1000
−1000
−500
0
500
1000
1500
0
397
795
1192
1590
1987

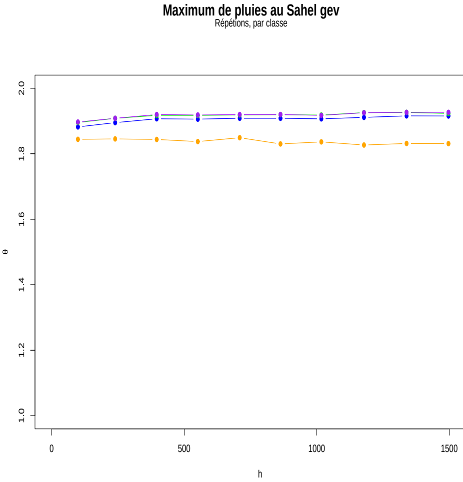

maximum des pluies pour l annee 1955
latitude
1000
1200
1400
1600
1800
2000
−1000
−500
0
500
1000
1500
longitude
0
291
583
874
1166
1457

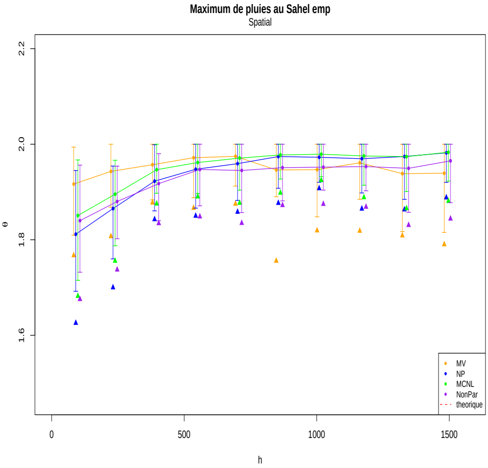

maximum des pluies pour l annee 1965
latitude
longitude
2000
1800
1600
1400
1200
1000
−1000
−500
0
500
1000
1500
0
419
838
1257
1676
2095

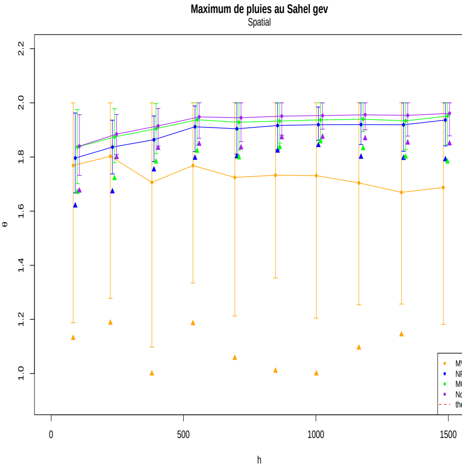

maximum des pluies pour l annee 1975
latitude
longitude
2000
1800
1600
1400
1200
1000
−1000
−500
0
500
1000
1500
0
342
684
1026
1368
1710


maximum des pluies pour l annee 1985
latitude
longitude
2000
1800
1600
1400
1200
1000
−1000
−500
0
500
1000
1500
0
378
757
1135
1514
1892


Maximum de pluies au Sahel emp
Répétions, par classe
θ
1.0
1.2
1.4
1.6
1.8
2.0
0
500
1000
1500
h


Maximum de pluies au Sahel gev
Répétions, par classe
θ
1.0
1.2
1.4
1.6
1.8
2.0
0
500
1000
1500
h


Maximum de pluies au Sahel emp
Spatial
θ
h
MV
NP
MCNL
NonPar
theorique


Maximum de pluies au Sahel gev
Spatial
θ
h

Conclusions

- Plus variable avec la GEV
- estimateurs concordants (sauf MV)
- faible dépendance à petites distances ?

Coefficient η

Champ Gaussien stationnaire isotrope de fonction de corrélation $\rho(h)$:

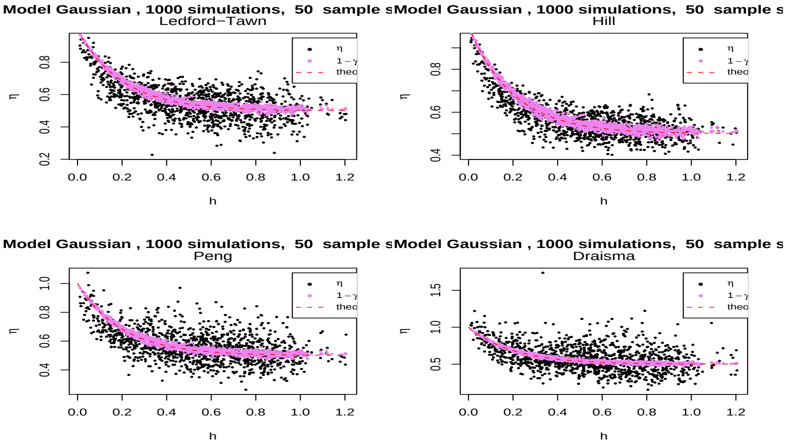
$$\eta(h) = \frac{1}{2}(1 + \rho(h))$$

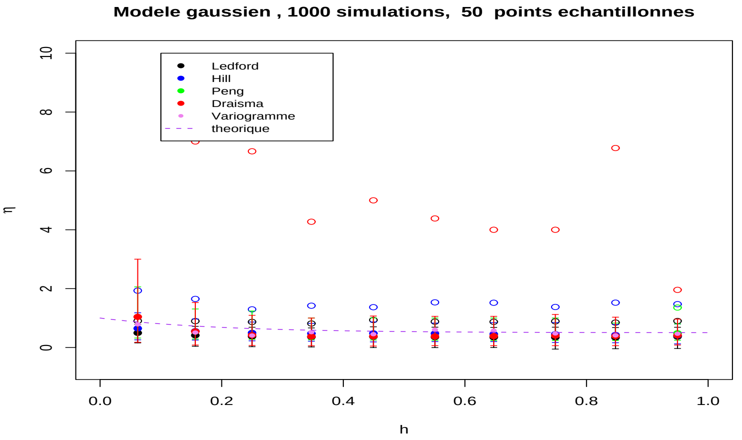
Modèle tempête (asymptotiquement dépendant)

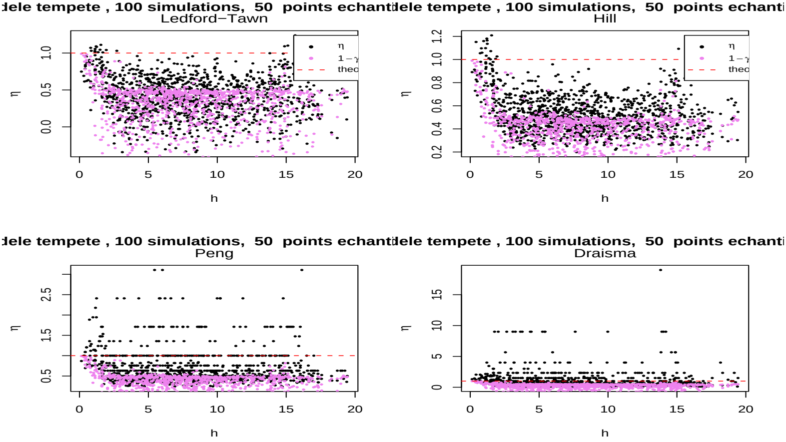
$$\eta(h) = \frac{1}{2}$$

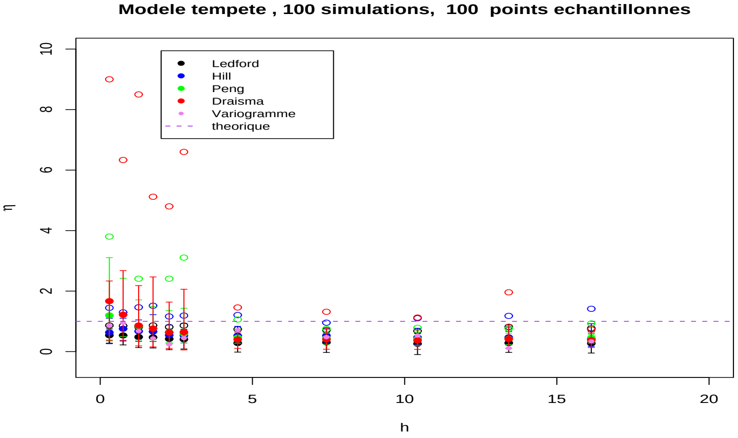
4 estimateurs :

Ledford and Tawn (1996), Hill (1975), Peng (1999), Draisma, Drees, Ferreira and de Haan (2001).


Model Gaussian , 1000 simulations, 50 sample s
Ledford−Tawn
Model Gaussian , 1000 simulations, 50 sample s
Hill
Model Gaussian , 1000 simulations, 50 sample s
Peng
Model Gaussian , 1000 simulations, 50 sample s
Draisma
η
1 − γ
theo
h
0.0 0.2 0.4 0.6 0.8 1.0 1.2


Modele gaussien , 1000 simulations, 50 points echantillonnes
Ledford
Hill
Peng
Draisma
Variogramme
theorique
η
h
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10


dele tempete , 100 simulations, 50 points echanti
Ledford−Tawn
Hill
Peng
Draisma
η
1 − γ
theo
h
η


Modele tempete , 100 simulations, 100 points echantillonnes
Ledford
Hill
Peng
Draisma
Variogramme
theorique
η
h
0
2
4
6
8
10
0
5
10
15
20

Conclusions ?

- **Modèle tempête** : tous trop faibles dès que $h > 0$;

- Approche classique et spatiale : pareil

- **Modèle Gaussien** : donnent la bonne valeur

variogramme meilleur (moins de variabilité)

Conclusions ?

- $\theta(h)$, $\eta(h)$: caractériser le type de dépendance des extrêmes
- dépendance asymptotique / indépendance asymptotique : bonnes notions ?
- influence de la dépendance spatiale dans l'estimation ?
- En déduire un modèle adéquat pour la simulation