Rapport - Valeurs Extrêmes

Alexis Guyonvarch - Philippe Real

16 janvier, 2020

Abstract

This is my abstract.

Contents

1 Introduction

1.1 Rappel des notions fondamentales de la théorie des valeurs extrêmes

On étudie un n-échantillon iid de v.a $X_1, X_2, ..., X_n$ de même loi, de fonction de répartition F, et on considère la v.a $X_{n,n} = max(X_1, X_2, ..., X_n)$ dont on cherche à décrire statistiquement le comportement asymptotique.

Par la propriété d'indépendance de l'échantillon, la fonction de répartition la loi du max $X_{n,n}$ peut s'écrire aussi F^n.

La théorie des valeurs extrêmes s'appuie sur la notion de "Domaine Maximum d'attraction" MDA(G) où G est la fonction de répartition d'une v.a non dégénérée (i.e associée à une va non constante) limite en loi (?) de la suite "rescalée?" $X^*_{n,n} = \frac{X_{n,n}-b_n}{a_n}$.

Plus précisemment, on dit que F appartient au domaine d'attraction DMA(G) de G ssi il existe deux suites a_n>0 et b_n telles que : $\lim_{n\to+\infty} F^n(a_n x + b_n) = G(x)\ \forall x \in \mathbb{R}$ De manière équivalent : $X^*_{n,n} \xrightarrow{d} Y$ où Y est une loi ayant pour fonction de répartition G.

Le domaine d'attraction est unique pour les paramètre de position et de dimensionnement (scale) fixés.

Le théorème fondamentale de la théorie, théorême de Fisher-Tippett-Gnedenko ou théorème des valeurs extrêmes stipule qu'il n'y a que 3 grandes familles (ou ensemble) de domaines d'attractions ou dit autrement 3 grandes familles de fonctions de répartitions limites pour $X^*_{n,n}$. Qui sont les familles de Gumbel, Frêchet et Weibull. Ces familles de lois des extrêmes ont des profiles et des comportement asymptotique différents.

Si la loi limite $F \in MDA(G)$ alors la lois de répartition G est de l'une des formes suivantes :

- Fréchet: définit pour $\gamma > 0$ par $\phi^F(x) = exp(-\frac{(x-b)}{a})^{-1/\gamma}$ si x > b et 0 sinon.
- Weibull: définit pour $\gamma > 0$ par $\phi^W(x) = exp(\frac{(x-b)}{a})^{1/\gamma}$ si x < b et 1 sinon.
- Gumbel: définit par $\phi^G(x) = exp(-exp(-\frac{(x-b)}{a}))$ si x $\in \mathbb{R}$.

En se basant sur le signe de γ on distingue les troisn domaines d'attraction:

- Si $\gamma < 0$ F est dit appartenir au domaine d'attraction de type Weibull $F \in MDA(Weibull)$.
- Si $\gamma = 0$ F est dit appartenir au domaine d'attraction de type Gumbel $F \in MDA(Gumbel)$.
- Si $\gamma > 0$ F est dit appartenir au domaine d'attraction de type Fréchet $F \in MDA(Fréchet)$.

The Generalized Extreme Value distribution : GEV

1.2 Plan et description

Dans chacune des 3 partie du rappert, on va étudier une série et chercher à déterminer la distribution limite des valeurs extrêmes. On commencera par une decription statistique de chaque série étudié puis on va chercher à estimer les paramètres scale: $a = \lim_{n\to+\infty} a_n$ et shape: $b = \lim_{n\to+\infty} b_n$ qui sont associés aux différentes catégories de lois limites et à la série étudiée.

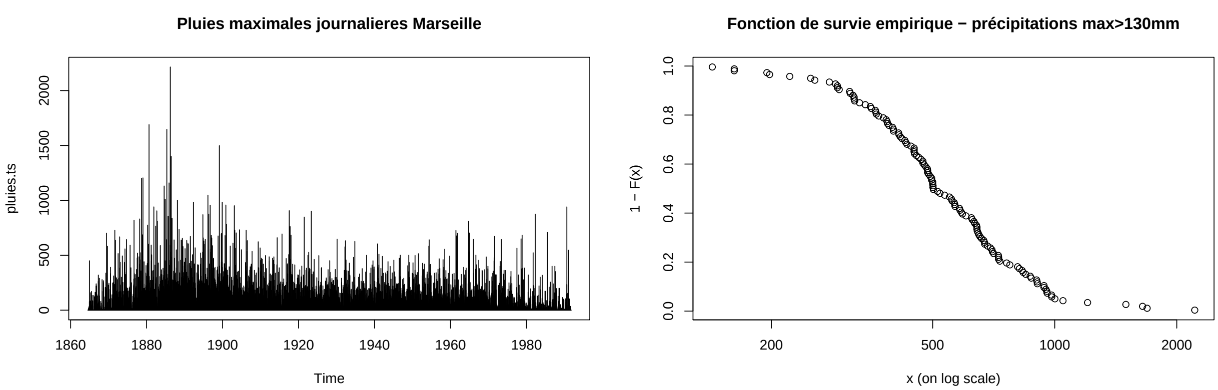
On déterminera ensuite la famille la plus appropriée à la série et le domaine d'attraction associé. On terminera notre étude avec une estimation de la probabilité et du niveau de retour. [associé à chacune de ces familles ou distribution de la loi des extrêmes]

2 Partie I - Comparaison des approches GEV et GPD sur les données *Marseille*

On s'intéresse dans cette première partie, à la pluviométrie dans la vile de Marseille sur la période allant de 1864 à 1991.

Cette étude est réalisée à patir des données d'accumulations pluviométriques quotidiennes de 10^{-1} mm dans la vile de Marseille pendant 127 ans d'observations (1864–1991). La première date est le 1er août 1864. Tous les 29 février ont été supprimés. Pour cet ensemble de données, on va utiliser les deux approches (GEV et GPD) et les comparer.

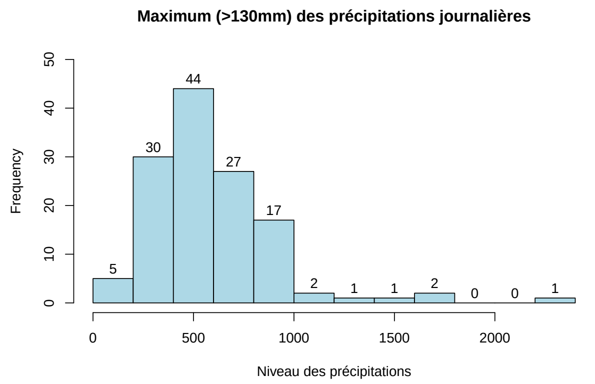
2.1 Lecture des données - description statistique

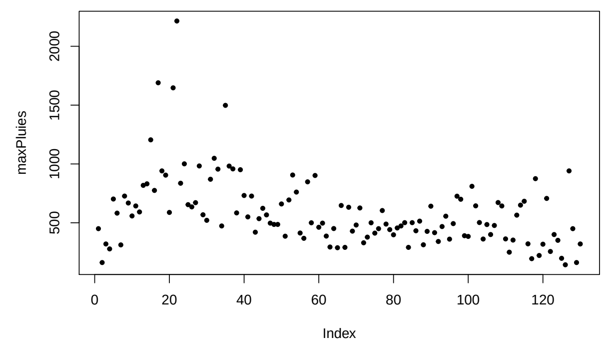


- summary, quantiles, histogramme des précipitations max>130 mm (cf. slides Intro)

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   143.0   398.5   501.5   588.1   698.5  2215.0

##     0%    25%    50%    75%   100%
##  143.0  398.5  501.5  698.5 2215.0
```

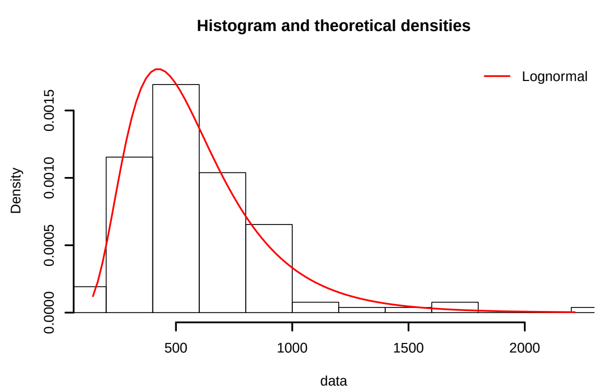


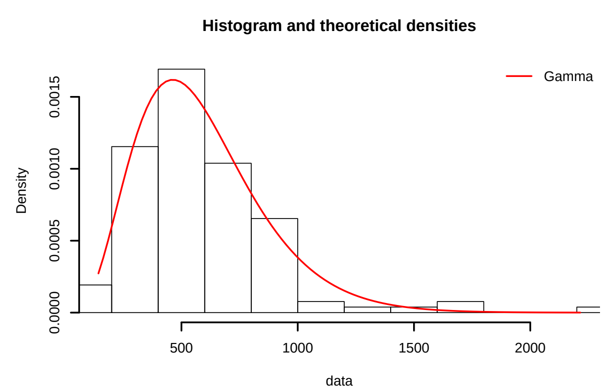


- Comparaison à une loi Log-normale et Gamma

```
## Fitting of the distribution ' lnorm ' by maximum likelihood
## Parameters :
##          estimate Std. Error
## meanlog 6.2664063 0.04097632
## sdlog   0.4672019 0.02897403
## Loglikelihood:  -900.1656   AIC:  1804.331   BIC:  1810.066
## Correlation matrix:
##         meanlog sdlog
## meanlog       1     0
## sdlog         0     1
```

```
## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters :
##          estimate   Std. Error
## shape 4.687551118 0.5154367576
## rate  0.007970481 0.0009102688
## Loglikelihood:  -903.2824   AIC:  1810.565   BIC:  1816.3
## Correlation matrix:
##           shape      rate
## shape 1.0000000 0.9371161
## rate  0.9371161 1.0000000
```






```
## [1] 1954.438
```

```
## [1] 1954.438
```

```
## [1] 1636.377
```

```
## [1] 1636.377
```

```
## [1] 103.5641
```

```
## [1] 2014.004
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   143.0   398.5   501.5   588.1   698.5  2215.0
```

2.2 Choix du niveau des blocs (GEV) et du seuillage (GPD)

- Ajustement GEV sur les max annuels

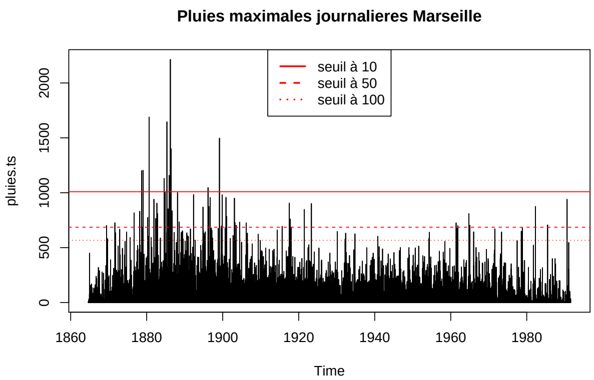
```
##
## Call: fgev(x = PluieMatrix)
## Deviance: 1756.301
##
## Estimates
##       loc      scale      shape
## 461.51437  196.76399    0.09805
##
## Standard Errors
##      loc     scale     shape
## 19.49785  14.46937   0.06024
##
## Optimization Information
##   Convergence: successful
##   Function Evaluations: 47
##   Gradient Evaluations: 29
```

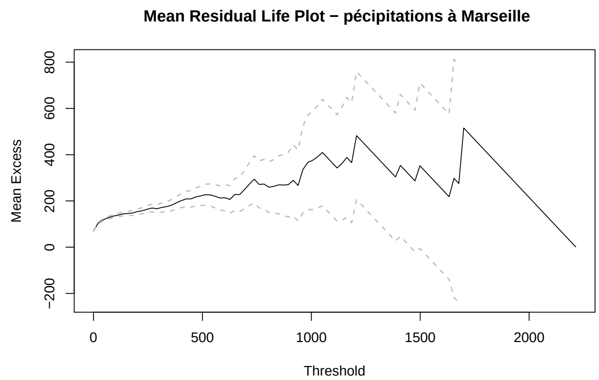
2.2.1 Choix du seuil pour l'approche GPD

- function findthresh of the package evir (p38 cours III)

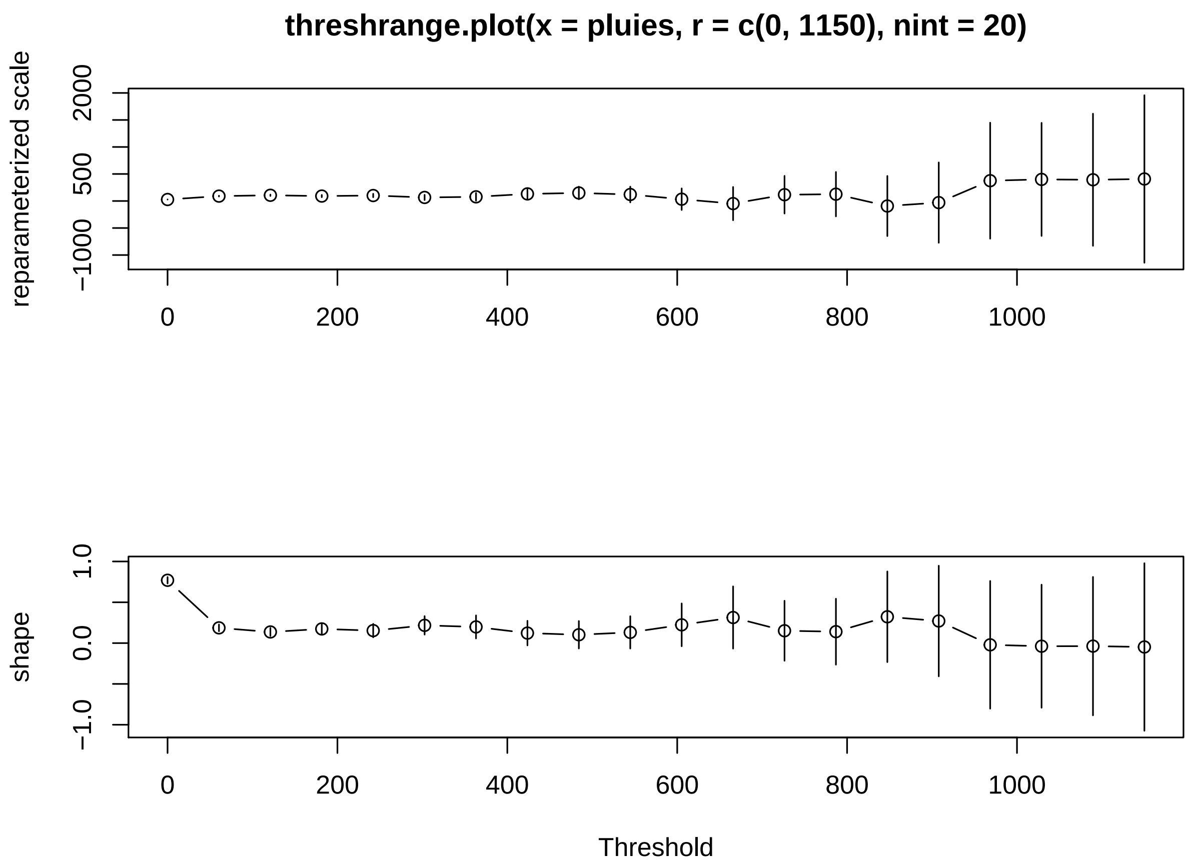
```
## [1] 1690 1009  567
```

```
##  n=10  n=50 n=100
##  1009   686   567
```



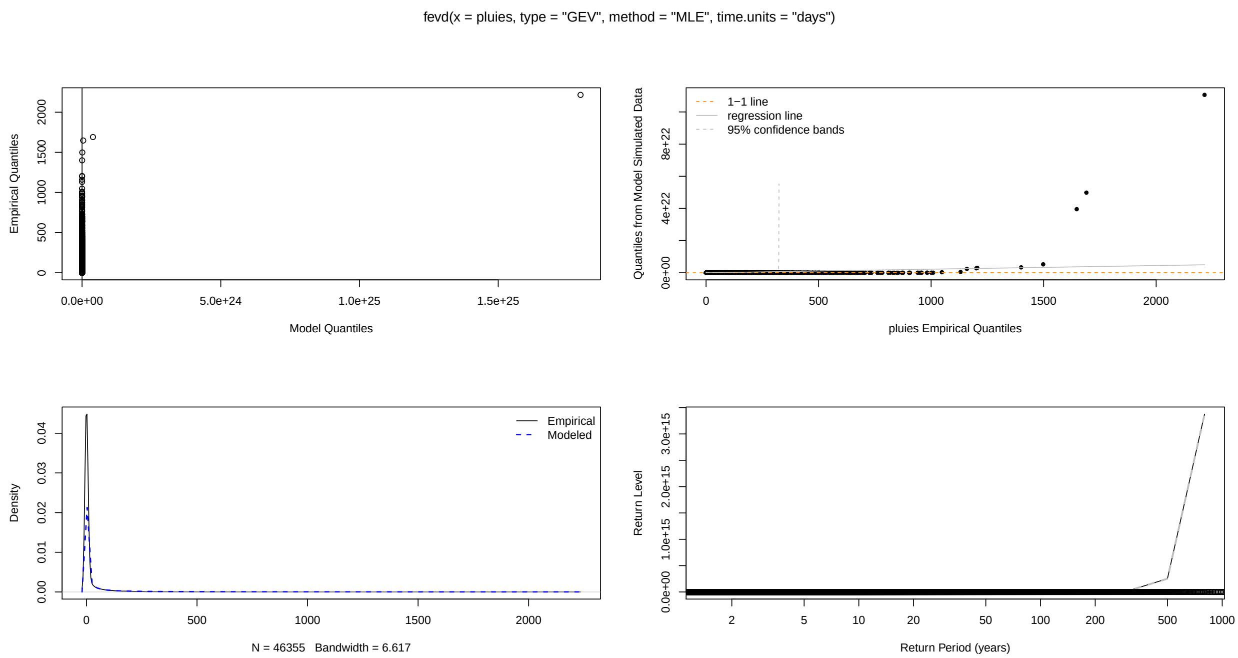


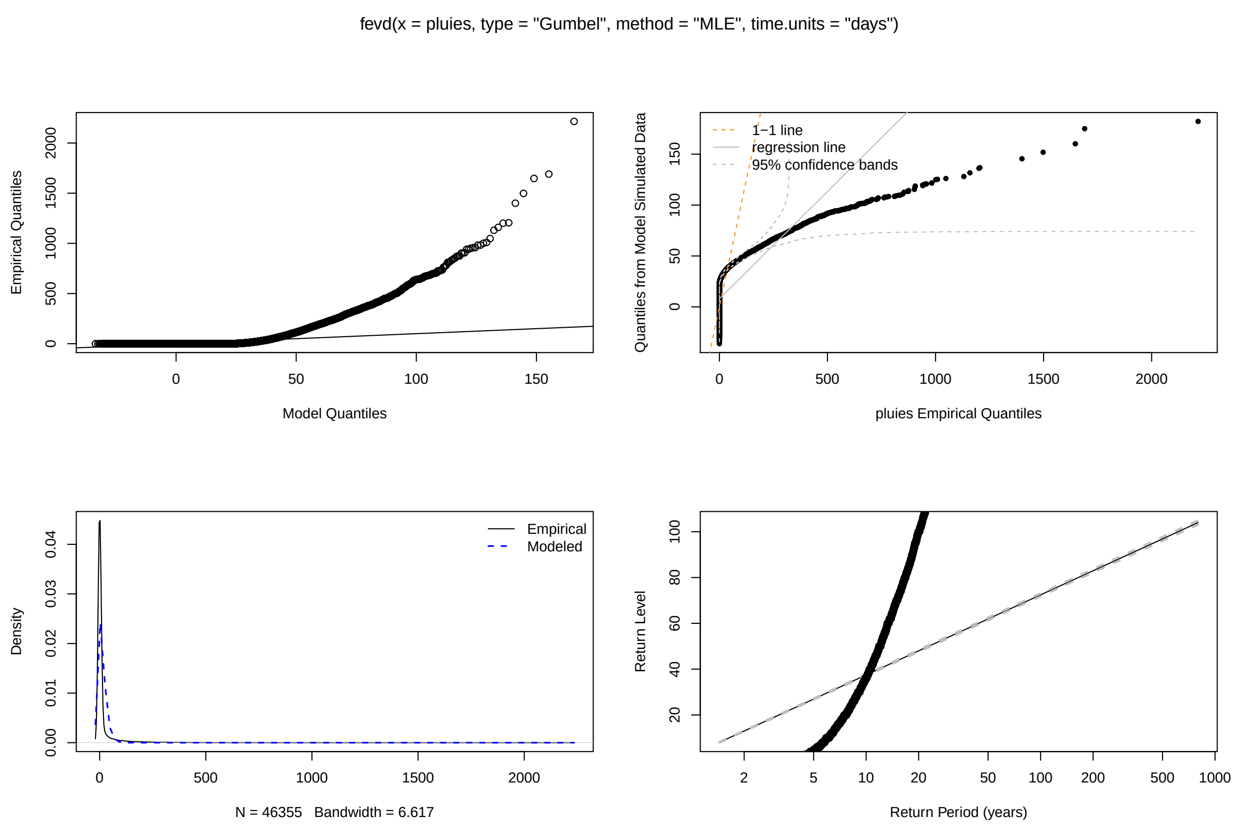
seuil à 600 580 cf p57 cours GPD



2.3 Estimation des paramètres shape et- scale par MLE et PWM.

2.3.1 Approche GEV





2.3.2 Approche GPD

- Estimation des paramètres par maximum de vraissamblance dans l'approche GPD (p47 cours GPD)

On utilise la fonction R fved avec 2 la méthode MLE et des moments

Méthode	scale	shape
GPD - MLE GV	165.9205569	0.228288
GPD - MLE Exp	214.9886364	NA
GPD - Moment GV	165.9205339	0.2282358

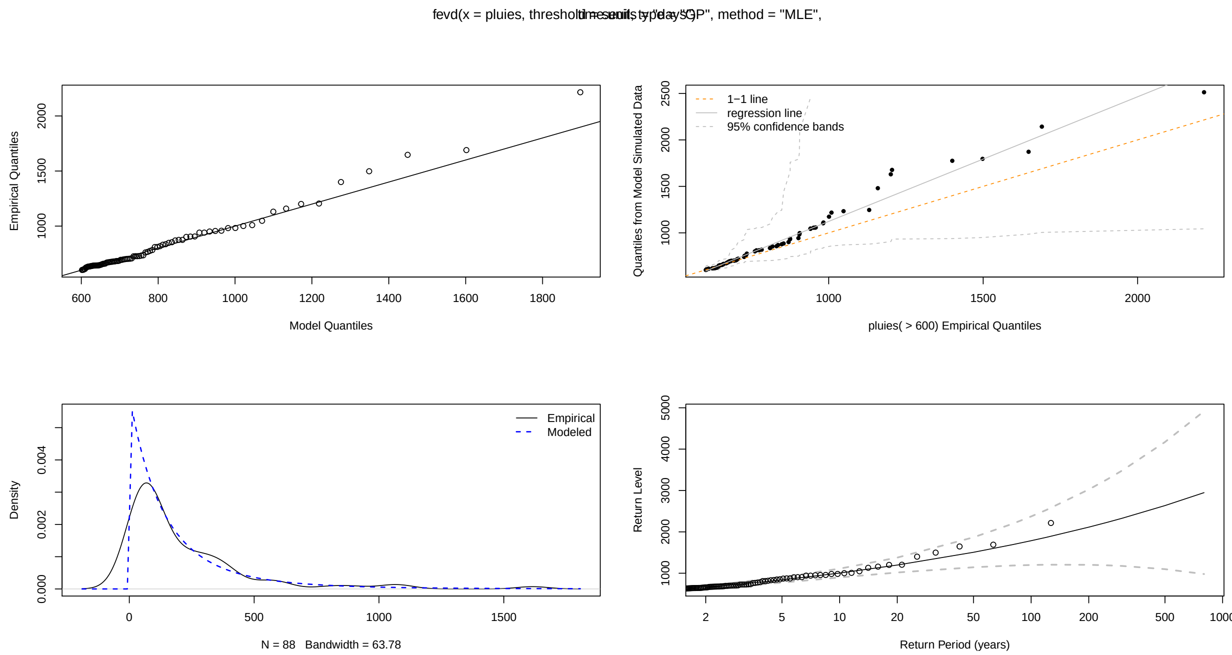
- Approximations des paramètres *scale* et *shape* et Intervalles de confiance

```
## fevd(x = pluies, threshold = seuil, type = "GP", method = "MLE",
##     time.units = "days")
##
## [1] "Normal Approx."
##
##       95% lower CI   Estimate 95% upper CI
## scale 111.51172894 165.920557  220.3293848
## shape  -0.03140201   0.228288    0.4879779
```

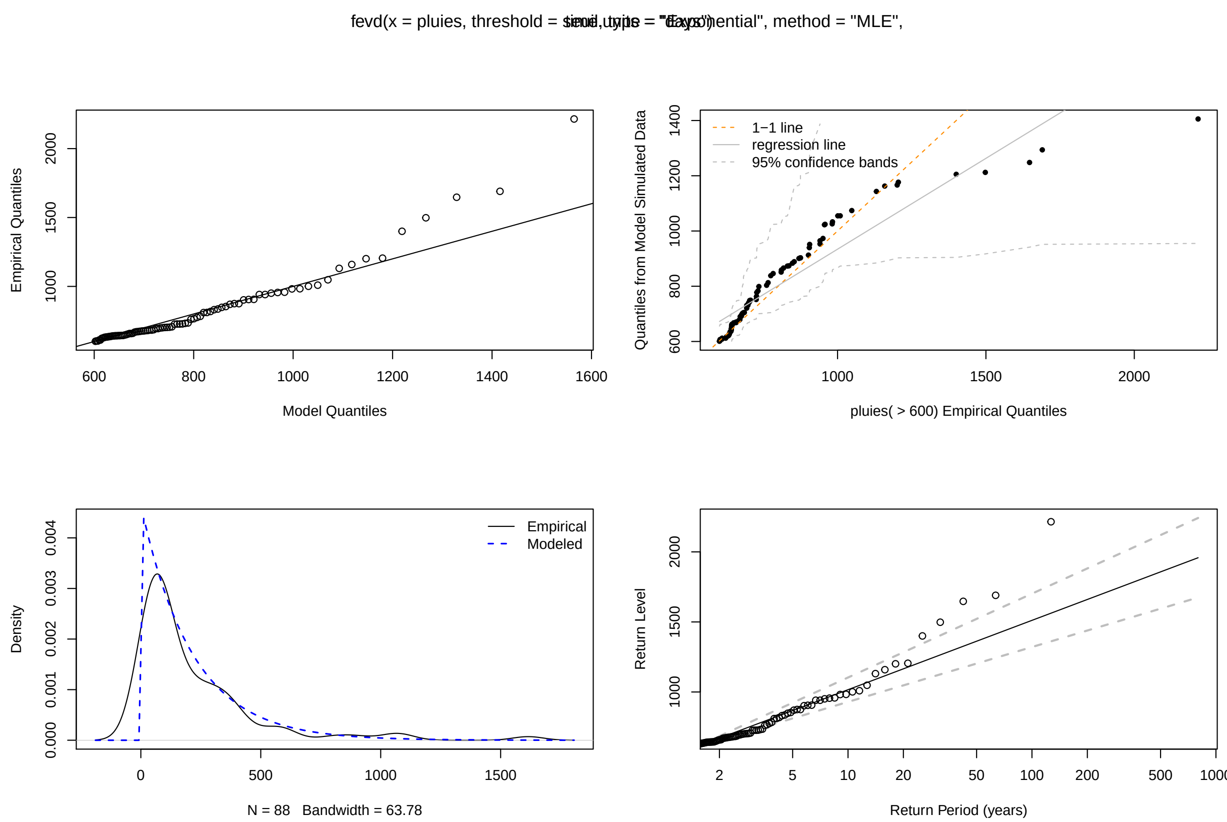
```
## fevd(x = pluies, threshold = seuil, type = "Exponential", method = "MLE",
##     time.units = "days")
##
## [1] "Normal Approx."
##
## [1] "scale: 214.989"
##
## [1] "95% Confidence Interval: (170.0705, 259.9067)"
```

```
## fevd(x = pluies, threshold = seuil, type = "GP", method = "Lmoments",
##     time.units = "days")
##
## [1] "Parametric Bootstrap"
## 502  iterations
##
##               2.5%    Estimate       97.5%
## scale 123.11771289 165.9205339 239.3638740
## shape  -0.06220906   0.2282358   0.4101039
```

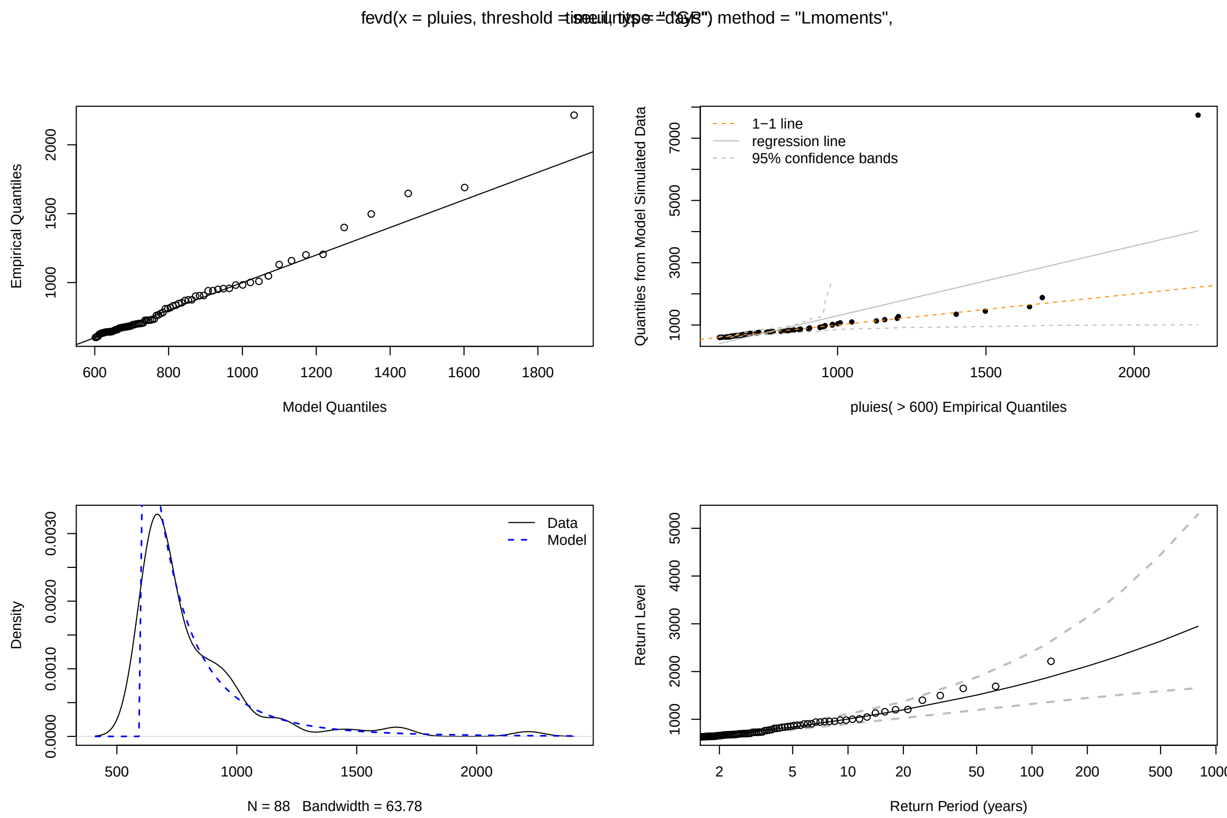
- Graphique - GPD estimé par MLE



- Graphique - Exponentiel estimé par MLE



- Graphique - GPD estimé par la PWD



2.4 Domaine d'attraction

2.5 Niveau de retour et interprétation

- MEV
- GPD

2.6 Estimation du niveau de retour correspondant à une période 100 ans et 1000 ans

- MEV

Méthode	100-year return	IC 95% borne Inf	IC 95% borne Sup
MLE GV	3.4406279×10^{10}	3.4406273×10^{10}	3.4406285×10^{10}
MLE Gumbel	72.3568861	71.7050733	73.0086989

Méthode	1000-year return	IC 95% borne Inf	IC 95% borne Sup
MLE GV	1.1569643×10^{16}	1.156964×10^{16}	1.1569646×10^{16}

Méthode	1000-year return	IC 95% borne Inf	IC 95% borne Sup
MLE Gumbel	107.397057	106.438766	108.355348

Méthode	2-year level	20-year level	100-year level	1000-year level
MLE GV	2.4827974	4.2862327×10^{6}	3.4406279×10^{10}	1.1569643×10^{16}
MLE Exponentiel	8.056698	47.6012529	72.3568861	107.397057

- GPD

Méthode	100-year return	IC 95% borne Inf	IC 95% borne Sup
MLE GV	1786.085834	1199.2759901	2372.8956779
MLE Exponentiel	1511.3378209	1320.9298455	1701.7457962
Moment GV	1785.9337272	1282.7348793	2400.5168149

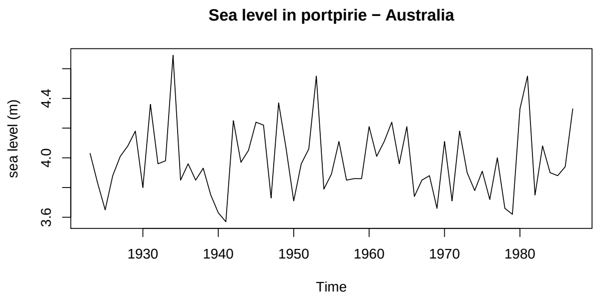
Méthode	1000-year return	IC 95% borne Inf	IC 95% borne Sup
MLE GV	3108.967893	904.9970456	5312.9387403
MLE Exponentiel	2006.3674501	1712.5317587	2300.2031416
Moment GV	3108.4367836	1741.833706	6188.1521965

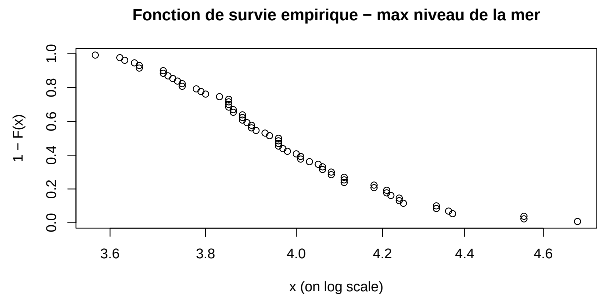
Méthode	2-year level	20-year level	100-year level	1000-year level
MLE GV	656.3291902	1197.9143253	1786.085834	3108.967893
MLE Exponentiel	670.2973295	1165.3269588	1511.3378209	2006.3674501
Moment GV	656.3286961	1785.9337272	1785.9337272	3108.4367836

2.7 Conclusion - Comaparaison GPD et GEV

3 Partie II - Approche GEV - Données *portpirie*

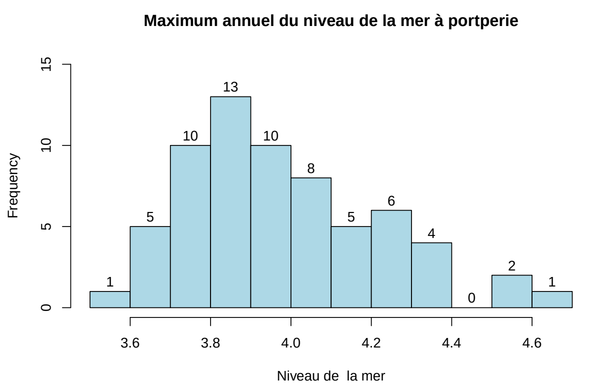
3.1 Lecture des données - description statistique

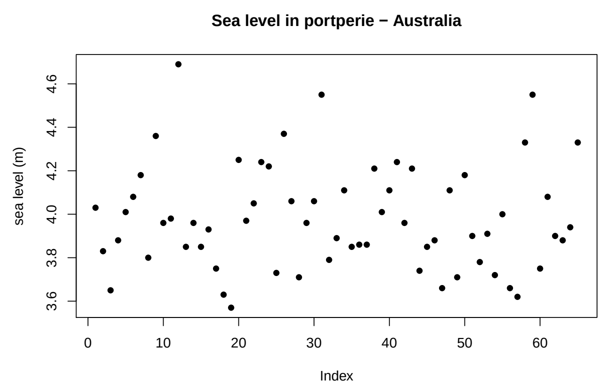





```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.570   3.830   3.960   3.981   4.110   4.690
```

```
##   0%  25%  50%  75% 100%
## 3.57 3.83 3.96 4.11 4.69
```



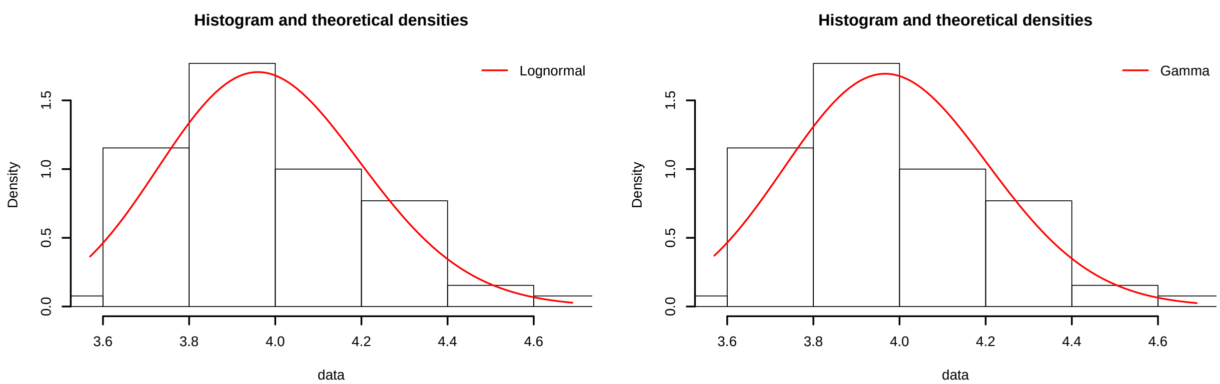


- Comparaison à une loi Log-normale et Gamma

```
## Fitting of the distribution ' lnorm ' by maximum likelihood
## Parameters :
##           estimate  Std. Error
## meanlog 1.37968039 0.007310659
## sdlog   0.05894042 0.005162722
## Loglikelihood:  2.119604   AIC:  -0.2392087   BIC:  4.109566
## Correlation matrix:
##               meanlog         sdlog
## meanlog  1.000000e+00 -1.466606e-14
## sdlog   -1.466606e-14  1.000000e+00
```

```
## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters :
##        estimate Std. Error
```

```
## shape 284.90664   49.94611
## rate   71.57345   12.55834
## Loglikelihood:  1.746143   AIC:  0.5077143   BIC:  4.856489
## Correlation matrix:
##           shape      rate
## shape 1.0000000 0.9991226
## rate  0.9991226 1.0000000
```




```
## [1] 4.688572

## [1] 4.688572

## [1] 4.674685

## [1] 4.674685

## [1] 3.236599

## [1] 4.853634

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.570   3.830   3.960   3.981   4.110   4.690
```

3.2 Choix des blocs

3.3 Estimation des paramètres par MLE et PWM.

Méthode	scale	shape
GEV - MLE GV	3.8747499	0.198044
GEV - MLE Gumbel	3.8694436	0.1948895

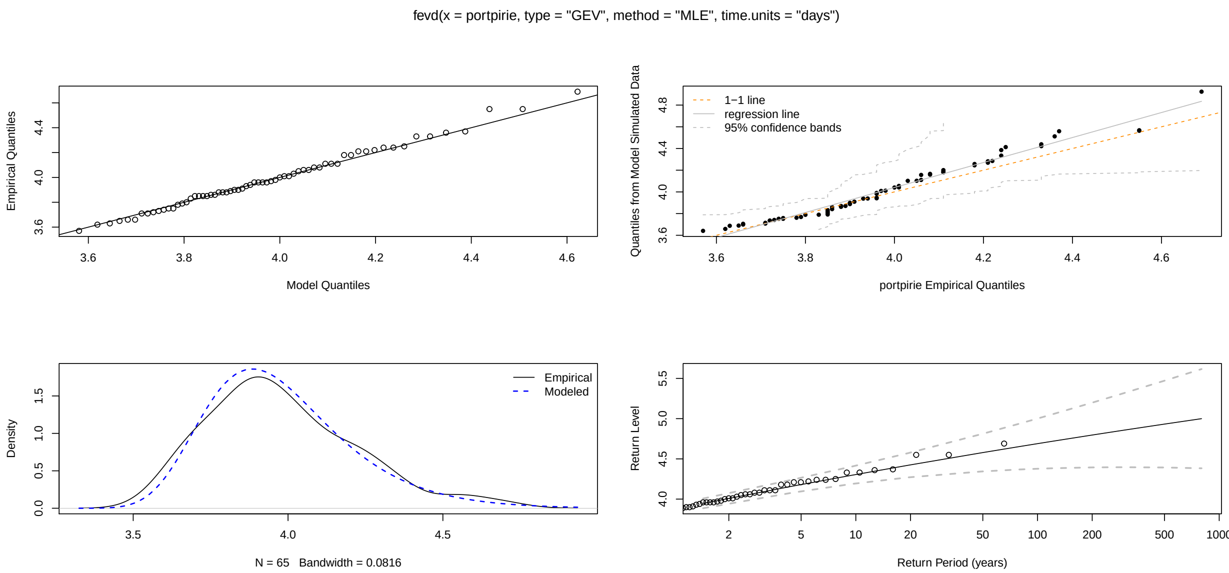
- Approximations des paramètres *scale* et *shape* et Intervalles de confiance

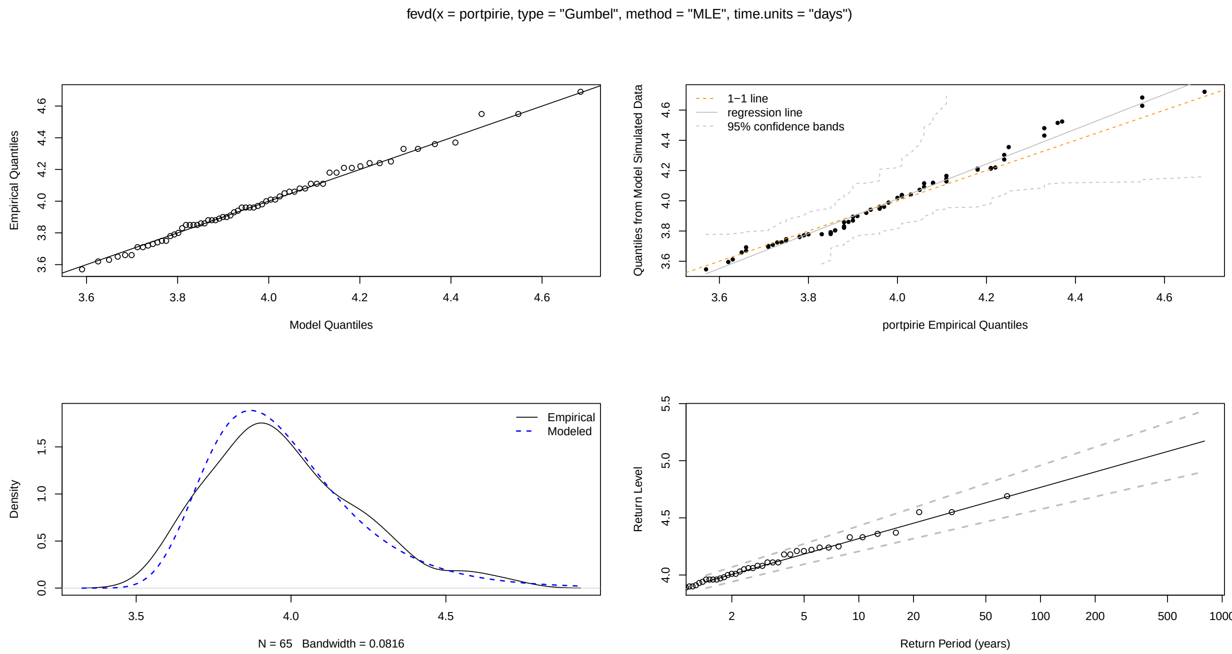
```
## fevd(x = portpirie, type = "GEV", method = "MLE", time.units = "days")
##
```

```
## [1] "Normal Approx."
##
##           95% lower CI   Estimate 95% upper CI
## location    3.8200037  3.8747499    3.9294960
## scale       0.1583586  0.1980440    0.2377293
## shape      -0.2426841 -0.0501095    0.1424651
```

```
## fevd(x = portpirie, type = "Gumbel", method = "MLE", time.units = "days")
##
## [1] "Normal Approx."
##
##           95% lower CI  Estimate 95% upper CI
## location    3.8194765 3.8694436     3.919411
## scale       0.1579369 0.1948895     0.231842
```

- Graphiques





3.4 Domaine d'attraction

3.5 Niveau de retour et interprétation

3.6 Estimation du niveau de retour correspondant à une période 100 ans et 1000 ans

Méthode	100-year return	IC 95% borne Inf	IC 95% borne Sup
MLE GV	4.6884038	4.3771254	4.9996822
MLE Gumbel	4.7659643	4.5741585	4.95777

Méthode	1000-year return	IC 95% borne Inf	IC 95% borne Sup
MLE GV	5.0310589	4.3764574	5.6856604
MLE Gumbel	5.2155949	4.9403707	5.490819

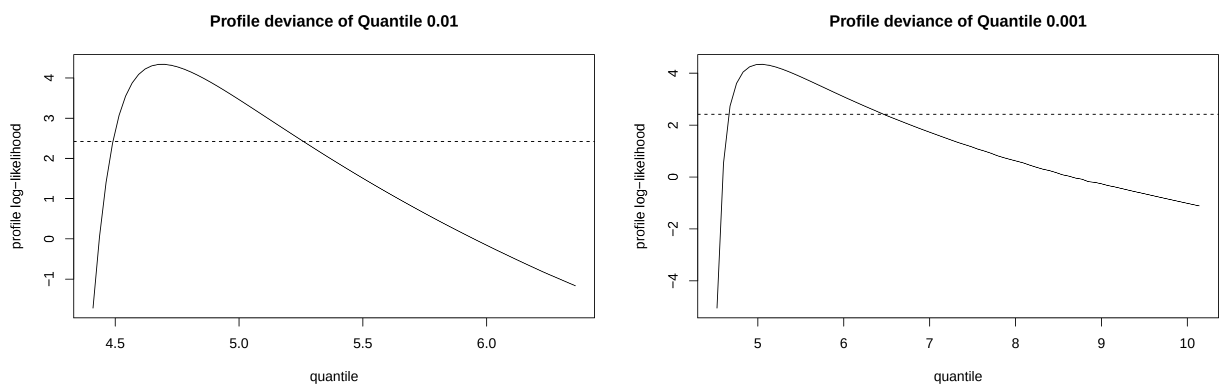
Méthode	2-year level	20-year level	100-year level	1000-year level
MLE GV	3.946673	4.4212978	4.6884038	5.0310589
MLE Gumbel	3.9408732	4.4483034	4.7659643	5.2155949

```
##
## Call: fgev(x = portpirie)
## Deviance: -8.678117
##
## Estimates
##      loc     scale     shape
```

```
##  3.87475   0.19805  -0.05012
##
## Standard Errors
##     loc    scale    shape
## 0.02793  0.02025  0.09826
##
## Optimization Information
##   Convergence: successful
##   Function Evaluations: 33
##   Gradient Evaluations: 8
```

```
## [1] "profiling quantile"
```

```
## [1] "profiling quantile"
```



3.7 Conclusion

4 Partie III - Approche GPD - Données *temps100m*

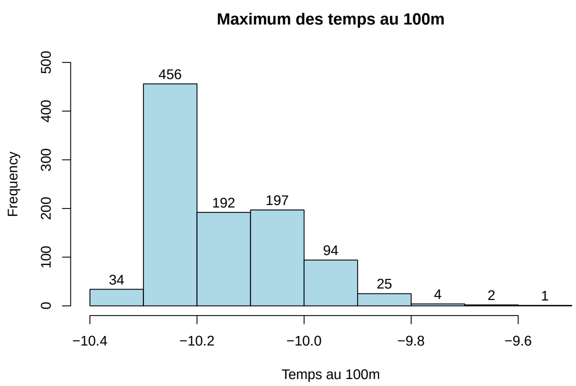
4.1 Lecture des données - description statistique

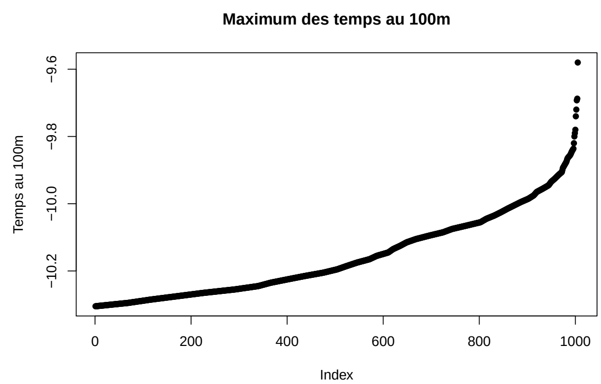
Ici on s'intéresse à la loi des miniùuns. Aussi pour étudier la loi des maximum ici on va inverser et regarder les temps en prenant l'opposé.

```
## [1] "numeric"
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -10.30  -10.26  -10.20  -10.16  -10.07   -9.58
```

```
##       0%      25%      50%      75%     100%
## -10.3049 -10.2610 -10.1955 -10.0715  -9.5800
```



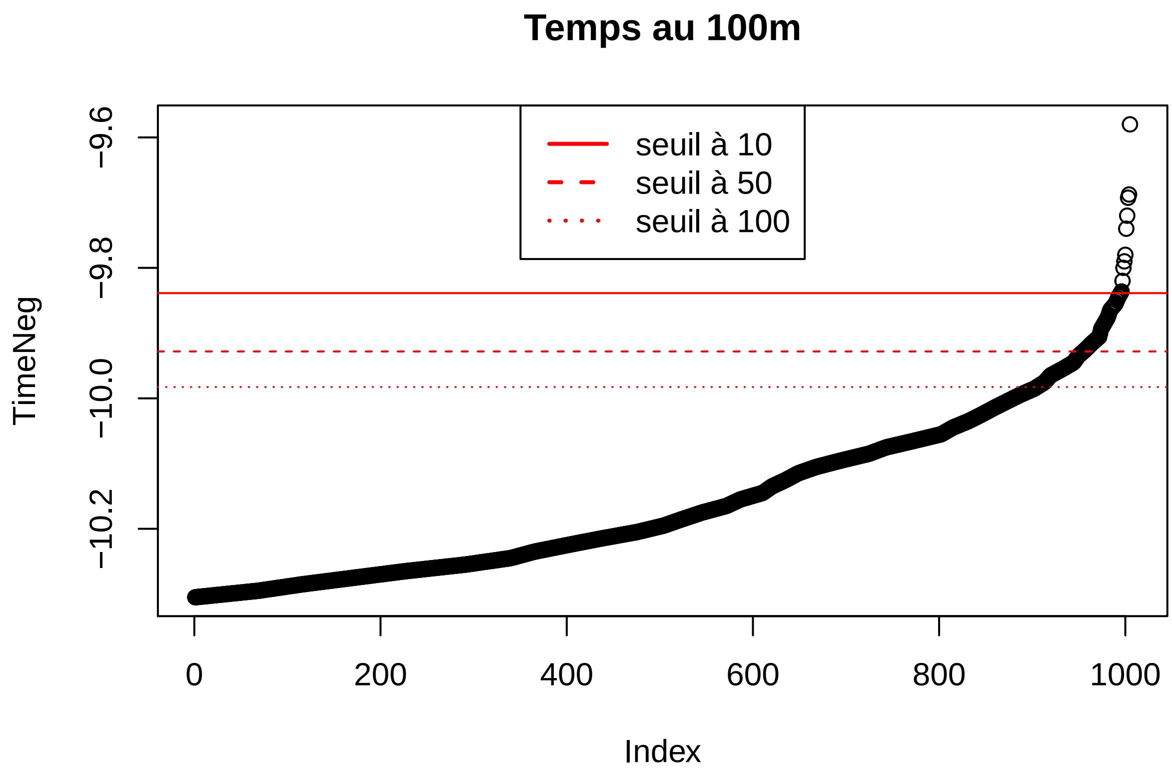


4.2 Choix du niveau de seuillage

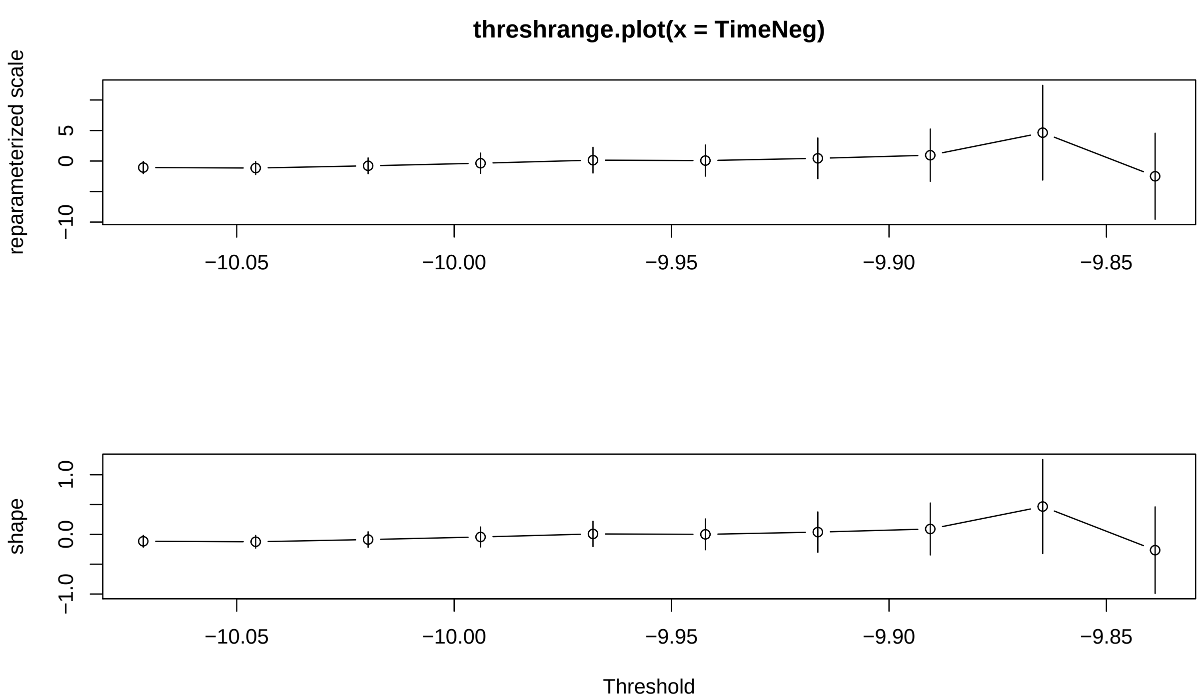
- function findthresh of the package evir (p38 cours III)

```
## [1] -9.6875 -9.8387 -9.9827
```

```
##    n=10    n=50   n=100
## -9.8387 -9.9281 -9.9827
```



- A partir des graphes tcplot: partie constante du graphe.



On obtient un seuil paproximatif de -9.92

4.3 Estimation des paramètre par MLE et PWM.

On utilise la fonction R fved avec 2 la méthode MLE et des moments

Méthode	scale	shape
MLE GV	0.0673474	0.0281371
MLE Exponentiel	0.0692932	NA
Moment GV	0.0650986	0.0605335

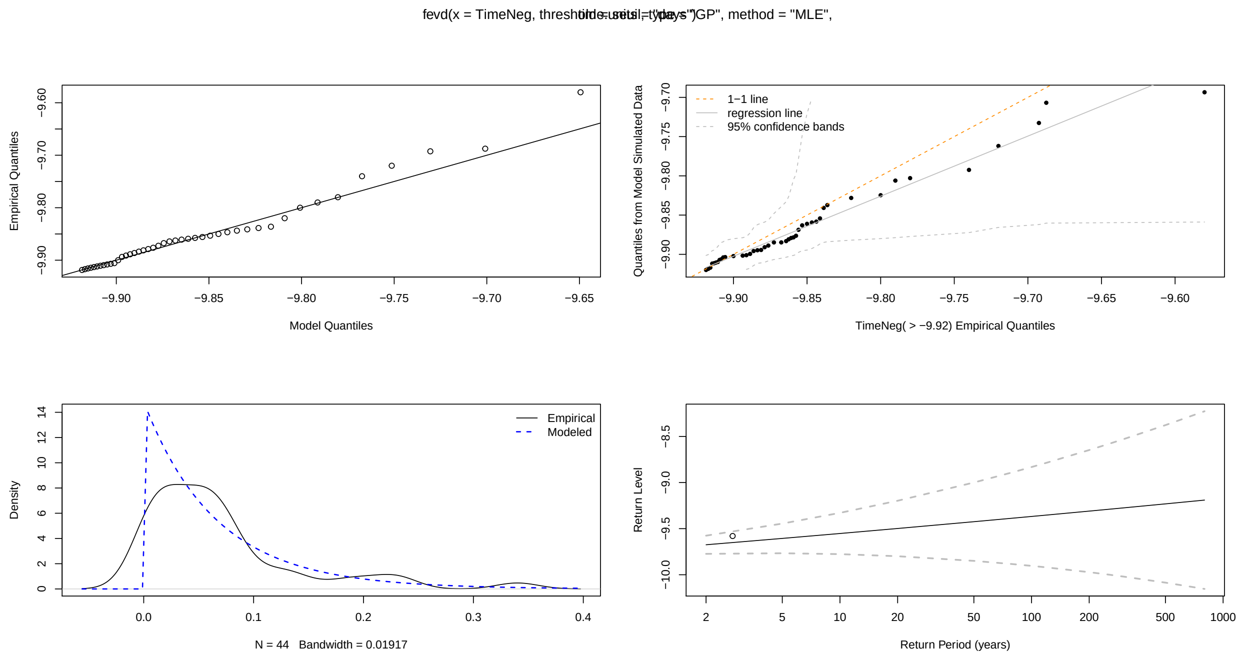
- Approximations des paramètres *scale* et *shape* et Intervalles de confiance

```
## fevd(x = TimeNeg, threshold = seuil, type = "GP", method = "MLE",
##     time.units = "days")
##
## [1] "Normal Approx."
##
##       95% lower CI   Estimate 95% upper CI
## scale   0.03782059 0.06734740    0.0968742
## shape  -0.29624524 0.02813714    0.3525195
```

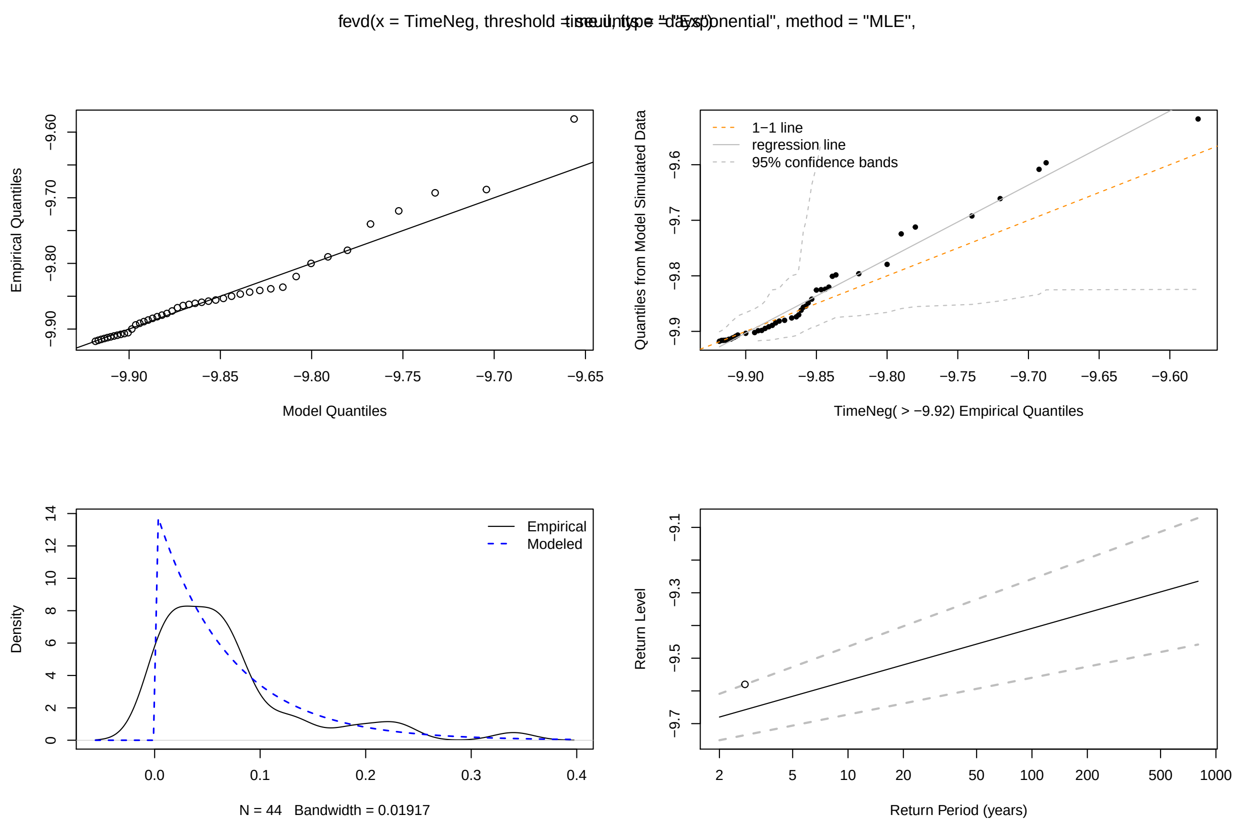
```
## fevd(x = TimeNeg, threshold = seuil, type = "Exponential", method = "MLE",
##     time.units = "days")
##
## [1] "Normal Approx."
##
## [1] "scale: 0.069"
##
## [1] "95% Confidence Interval: (0.0488, 0.0898)"
```

```
## fevd(x = TimeNeg, threshold = seuil, type = "GP", method = "Lmoments",
##     time.units = "days")
##
## [1] "Parametric Bootstrap"
## 502  iterations
##
##              2.5%   Estimate     97.5%
## scale  0.04253295 0.06509862 0.1033830
## shape -0.38019988 0.06053348 0.3374264
```

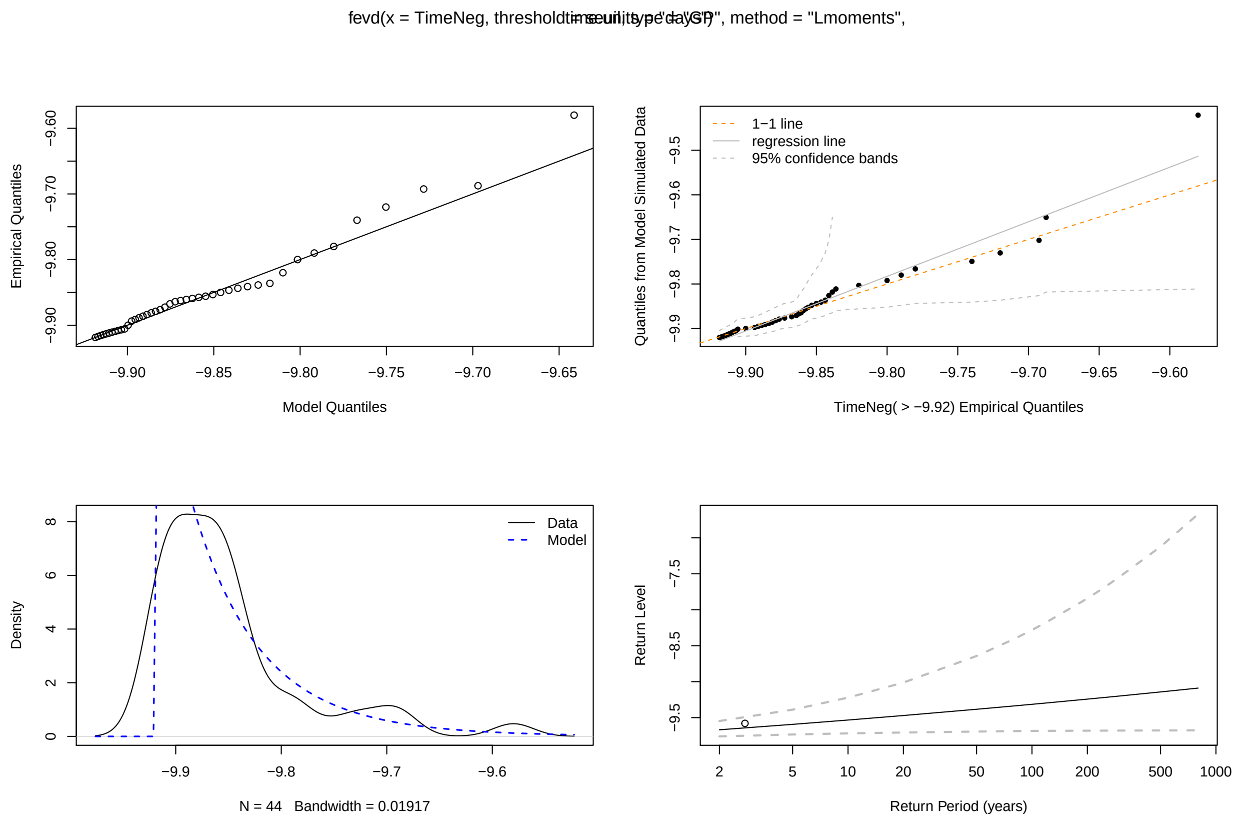
- Graphique - GPD estimé par MLE



- Graphique - Exponentiel estimé par MLE



- Graphique - GPD estimé par les moments



4.4 Domaine d'attraction

4.5 Niveau de retour et interprétation

4.6 Estimation du niveau de retour correspondant à une période 100 ans et 1000 ans

Méthode	100-year return	IC 95% borne Inf	IC 95% borne Sup
MLE GV	-9.3678391	-9.9037681	-8.8319101
MLE Exponentiel	-9.4088104	-9.5597601	-9.2578607
Moment GV	-9.3146077	-9.7133054	-8.1410653

Méthode	1000-year return	IC 95% borne Inf	IC 95% borne Sup
MLE GV	-9.1706745	-10.1891766	-8.1521724
MLE Exponentiel	-9.249257	-9.4473214	-9.0511925
Moment GV	-9.0632179	-9.6658323	-5.830045

Méthode	2-year level	20-year level	100-year level	1000-year level
MLE GV	-9.6690185	-9.4706362	-9.3146077	-9.0632179

Méthode	2-year level	20-year level	100-year level	1000-year level
MLE Exponentiel	-9.6798869	-9.5203335	-9.4088104	-9.249257
Moment GV	-9.6690185	-9.3146077	-9.3146077	-9.0632179

4.7 Conclusion