

Statistiques extrêmes

Watiotienne Henry

July 22, 2024

Contents

1	Objectif	2
2	Analyse exploratoire	2
2.1	Description des données	2
2.2	Test de Stationnarité	4
2.3	Nettoyage des données	4
3	Analyse des valeurs extrêmes	5
3.1	Modèle GEV	5
3.1.1	Estimation des paramètres	5
3.1.2	Niveau de retour	6
3.2	Approche GPD	6
3.2.1	Estimation des paramètres	6
3.2.2	Niveau de retour	7
3.2.3	Validation du modèle	8
3.3	Modèle à seuil avec clustering	9
3.3.1	Ajustement du modèle	9
3.3.2	Niveau de retour	10
3.3.3	Validation du modèle	11
3.4	Modèle non stationnaire	11
3.4.1	Ajustement du modèle	11
3.4.2	Comparaison du modèle	12
4	Conclusion	12
5	Annexe Code R	13

1 Objectif

L'objectif principal de ce rapport est d'analyser les données climatiques, avec une attention particulière portée sur l'étude des valeurs extrêmes. Ce sujet est très important dans le contexte du réchauffement climatique actuel, qui s'accélère et entraîne une hausse des phénomènes météorologiques extrêmes. Ainsi, les principaux objectifs de ce rapport sont :

1. Réaliser une analyse exploratoire des données pour comprendre la distribution et les tendances des températures au fil du temps.
2. Sélectionner et ajuster des modèles de valeurs extrêmes adaptés à nos données.
3. Estimer des niveaux de retour, ce qui implique de calculer la probabilité d'occurrence d'événements météorologiques extrêmes sur différentes périodes.

2 Analyse exploratoire

Cette phase Analyse exploratoire a pour objectif de comprendre la structure et les caractéristiques principales de notre jeu de données, avant d'entreprendre des analyses statistiques.

2.1 Description des données

Nous nous intéressons aux données de température de la ville de Brindisi, située en Italie.

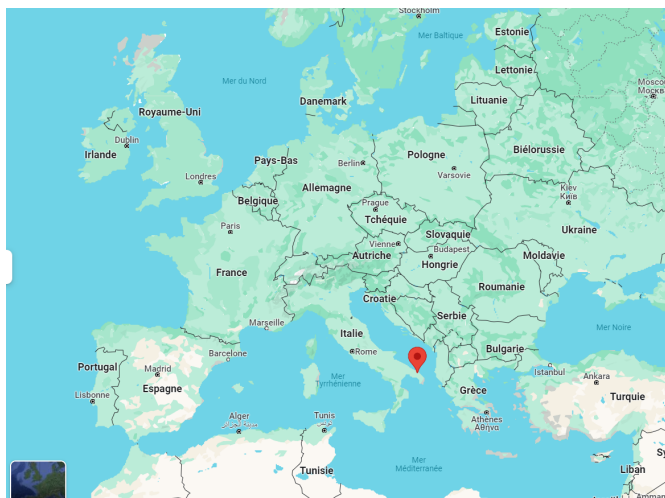


Figure 2.1.1: Localisation de la ville de Brindisi

Les données sont journalières et répertoriées depuis 1951 jusqu'à 2023. Il y a 287 ce qui ne représente que environ 1% de notre jeu de données données manquantes, nous verrons dans la suite comment l'on gère ses données manquantes. Les valeurs des températures s'étendent de 0.1°C à 44.5°C . On s'intéresse a la distribution de nos données.

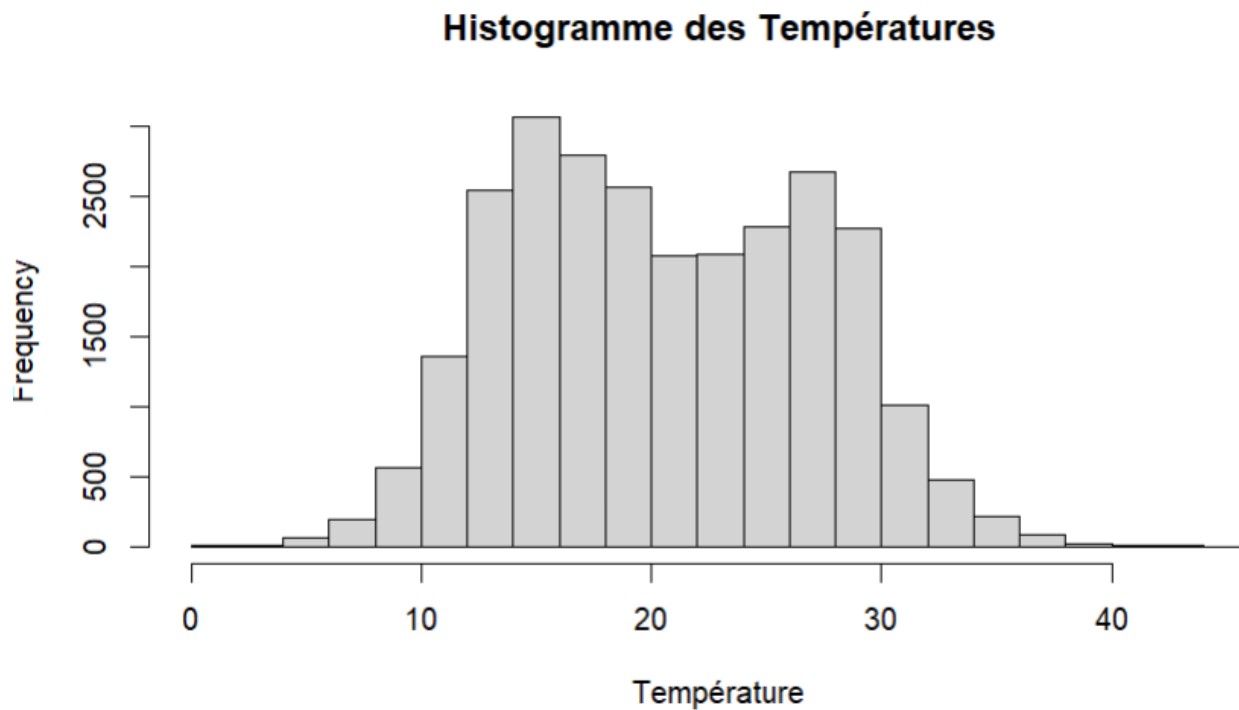


Figure 2.1.2: Histogramme de nos données

La plupart de nos températures semble être situées entre 10 et 30°C.
Nous allons maintenant représenter nos séries de températures.

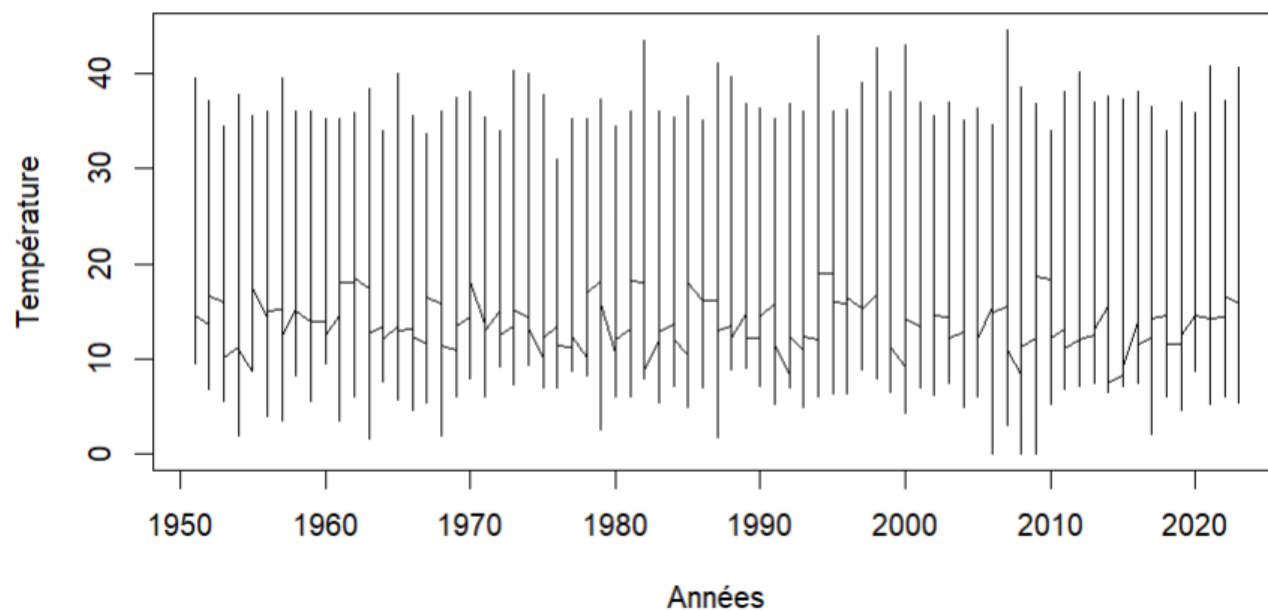


Figure 2.1.3: Représentation graphique de nos données

Le graphique montre les températures à Brindisi de 1951 à 2023, avec une ligne représentant la température moyenne annuelle et des barres verticales indiquant les températures maximales et minimales pour chaque année. Il n'y a pas de tendance claire à l'augmentation ou à la diminution sur le long terme, ce qui suggère une stationnarité de la série temporelle des températures.

Dans la suite nous nous intéresserons aux maximums des températures par an. Nous pouvons représenté nos nouvelles.

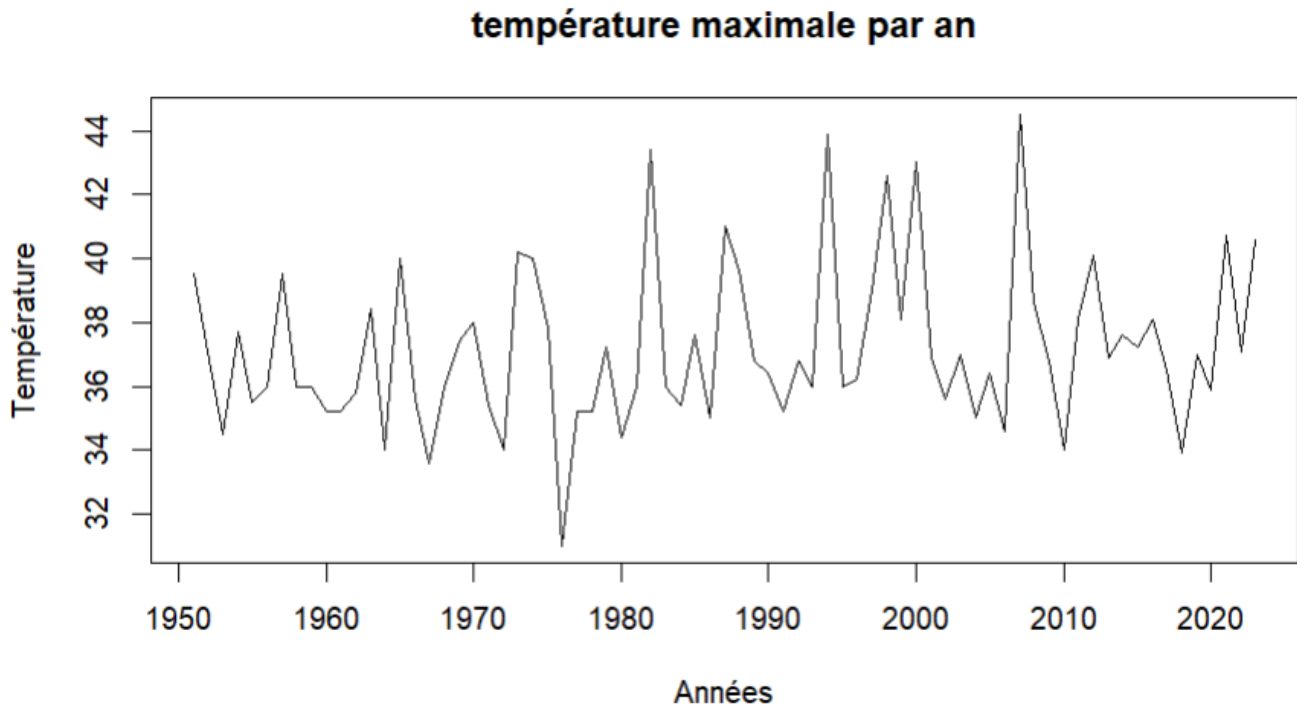


Figure 2.1.4: Température maximale par an

Une nouvelle fois la série ne semble pas avoir de tendance claire ce qui suggère à nouveau une stationnarité.

2.2 Test de Stationnarité

L'objectif est de confirmer la stationnarité de notre série temporelle, une condition préalable à l'utilisation de certains modèles statistiques. Le test de Dickey-Fuller augmenté, appliqué aux séries des températures maximales annuelles ainsi qu'à la série initiale, a produit des valeurs-p très faibles. Ces résultats nous permettent de rejeter l'hypothèse nulle, qui est la non-stationnarité des séries, nous conduisant à conclure à leur stationnarité.

2.3 Nettoyage des données

Nous avons identifié des valeurs manquantes dans notre série de données. Plutôt que de les exclure, ce qui entraînerait un décalage temporel puisque notre analyse repose sur une base de 365 jours par an, nous optons pour leur remplacement par la valeur moyenne de 20 °C. Cette méthode permet

de maintenir la cohérence de l'intervalle de temps sans introduire de distorsion due à l'absence de données.

3 Analyse des valeurs extrêmes

Cette section se consacre à l'analyse des valeurs extrêmes des données climatiques. Nous explorerons divers modèles et approches pour caractériser et prédire ces événements. Le Modèle de Valeurs Extrêmes Généralisées (GEV) et l'Approche de la Distribution Pareto Généralisée (GPD) sont des méthodes statistiques classiques pour modéliser les extrêmes. Nous étudierons également l'application d'un modèle à seuil avec clustering pour identifier des régimes extrêmes distincts, ainsi qu'un modèle non stationnaire pour tenir compte des changements climatiques dans l'analyse des extrêmes. Nous étudierons des niveaux de retour pour chacun de ces modèles.

3.1 Modèle GEV

3.1.1 Estimation des paramètres

Dans un modèle GEV on cherche à estimer 3 paramètres :

- **Localisation** (loc) Définit le centre ou le seuil de la distribution des événements extrêmes.
- **Échelle** (scale) Contrôle la dispersion des valeurs extrêmes.
- **Forme** (shape) Indique la forme de la queue de la distribution, caractérisant comment les probabilités décroissent pour les valeurs extrêmes.

Pour estimer les paramètres, nous utilisons la fonction `fgev` dans R, qui utilise la méthode du maximum de vraisemblance. Cette technique sélectionne les valeurs des paramètres qui maximisent la probabilité des données observées sous le modèle GEV. Les intervalles de confiance pour ces estimations sont ensuite obtenus via la fonction `confint`. On obtient les résultats suivant :

Paramètre	2,5%	97,5%
loc	35,5698586	36,66954115
scale	1,8210354	2,56944639
shape	-0,2240389	0,03256558

Table 1: Intervalles de confiance à 95% pour les paramètres du modèle GEV

Les intervalles de confiance indiquent une estimation pour le paramètre de localisation (loc) entre environ 35,57 et 36,67, ce qui suggère une valeur centrale autour de laquelle les températures maximales annuelles sont distribuées. L'échelle (scale), entre 1,82 et 2,57, montre la variabilité de ces extrêmes. Le paramètre de forme (shape) varie de -0,224 à 0,033, indiquant une queue de distribution légèrement penchée vers les valeurs négatives mais proche de zéro.

3.1.2 Niveau de retour

On s'intéresse au niveau de retour sur 20 ans et 100 ans en utilisant la méthode Delta pour l'estimation. Cependant, les résultats obtenus avec cette méthode présentent des intervalles de confiance à 95% très larges ce qui les rend inutiles d'un point de vue prévision et utilité climatique.

Pour un niveau de retour de 20 ans, l'intervalle de confiance à 95% est [6.298722, 50.85702]. Cette intervalle de confiance est bien trop large pour fournir des informations utiles sur les événements climatiques extrêmes sur une période de 20 ans.

La situation est encore moins favorable pour un niveau de retour de 100 ans, où l'intervalle de confiance à 95% est [-15.40843, 62.27198].

Face à ces larges intervalles de confiance, il est évident que notre modèle actuel ne permet pas d'obtenir des estimations précises des niveaux de retour extrêmes. Nous devons donc rechercher un modèle plus adapté pour améliorer la précision de nos estimations.

3.2 Approche GPD

3.2.1 Estimation des paramètres

Dans un modèle GPD (Generalized Pareto Distribution), notre objectif est d'estimer la forme (shape) et l'échelle (scale) de la distribution des valeurs extrêmes. Le modèle GPD utilise également une méthode de maximum de vraisemblance pour estimer ces paramètres.

Cependant, dans notre approche, nous allons définir un seuil (threshold) manuellement pour considérer les températures comme étant extrêmes. Pour déterminer ce seuil, nous nous appuyons sur le graphique suivant :

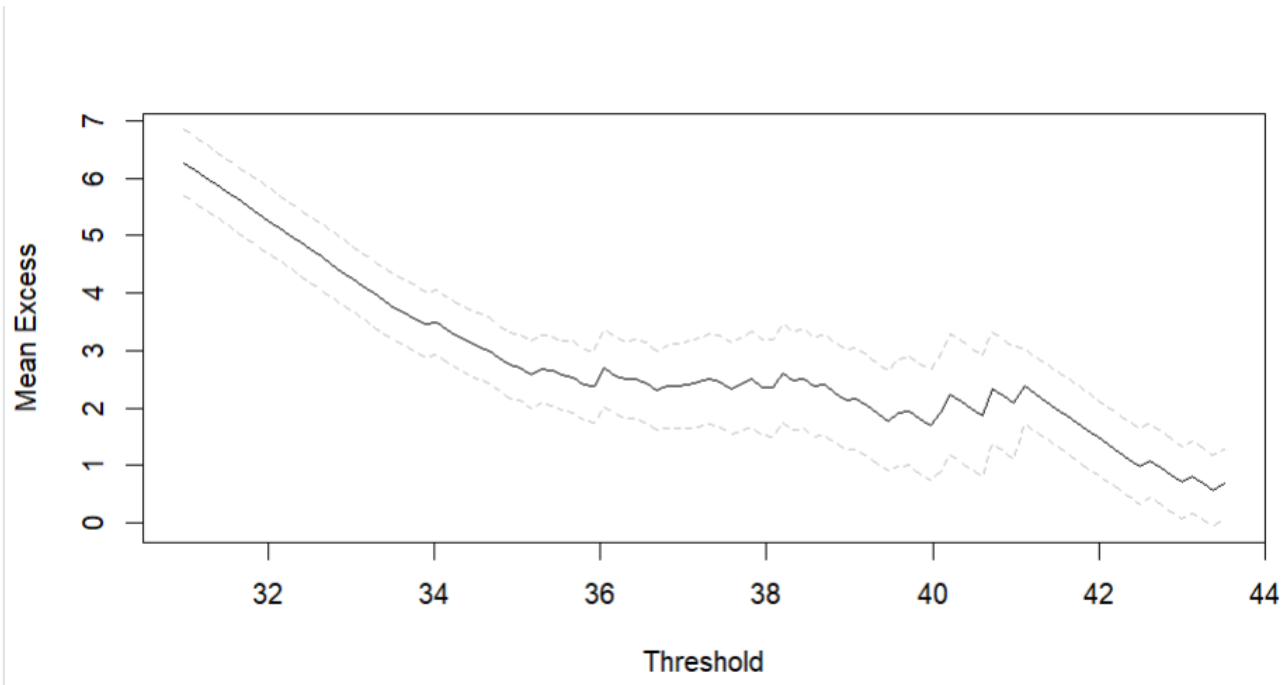


Figure 3.2.1: Détermination du seuil

Le seuil est atteint lorsque la courbe n'est plus linéaire. On le fixera donc à 37°C. On peut visualiser nos données avec ce seuil pour se convaincre qu'il est raisonnable.

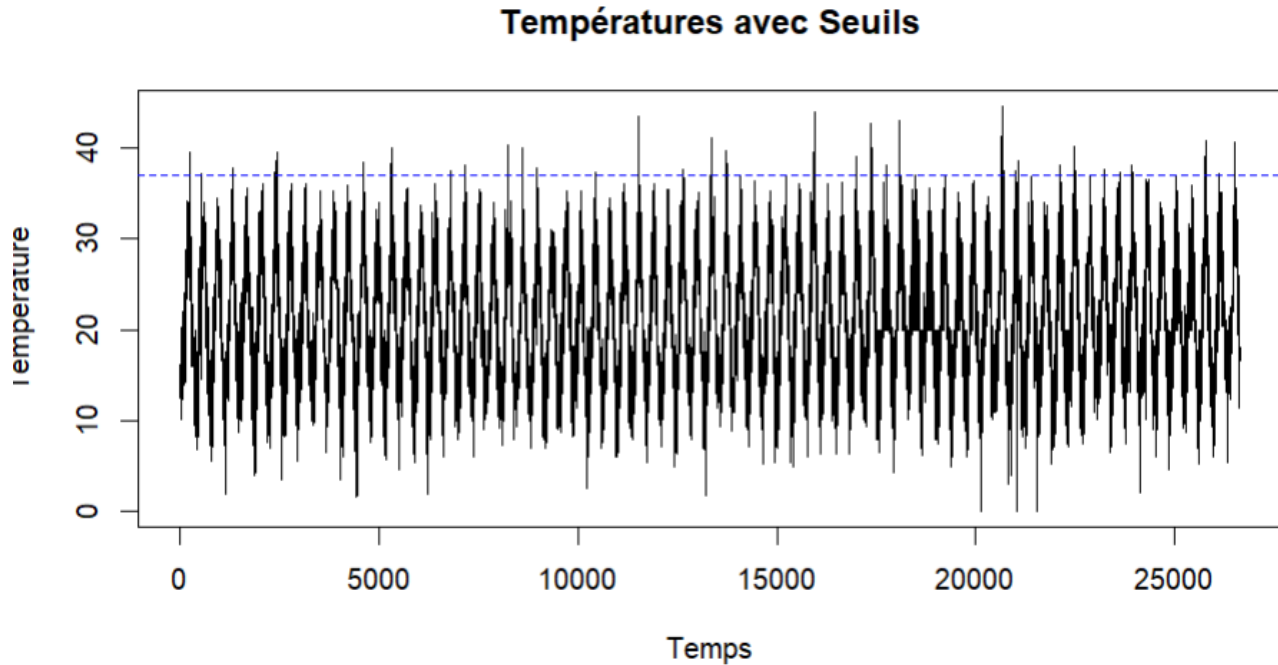


Figure 3.2.2: Données avec seuil

Les estimations des paramètres du modèle GPD avec un seuil à 37°C avec leurs intervalles de confiance à 95

Paramètre	Estimation	Intervalle de Confiance à 95%
Échelle (Scale)	2.05864	[1.27835, 3.83874]
Forme (Shape)	-0.08921	[-0.37695, 0.19854]

On remarque que les intervalles de confiance du modèle GPD sont plus larges que ceux du modèle GEV. Cependant, les résultats obtenus à partir des deux méthodes restent cohérents, car les intervalles de confiance des deux modèles ne s'excluent pas mutuellement.

3.2.2 Niveau de retour

Nous nous intéressons maintenant au niveau de retour en utilisant la méthode manuelle de calcul, que nous vérifierons ensuite en utilisant la fonction `fpot`. Les résultats obtenus sont très similaires entre les deux approches les estimations sont les suivantes:

Ces résultats du modèle GPD sont plus intéressants que ceux du modèle GEV car ils sont plus interprétables et revêtent un intérêt climatique plus significatif. Les intervalles de confiance sont moins larges, ce qui les rend plus fiables pour l'estimation des niveaux de retour extrêmes.

Période de Retour	Intervalle de Confiance à 95%
20 ans	[40.95344, 43.44656]
100 ans	[41.79, 47.4]

Pour l'interprétation des niveaux de retour pour 20 ans par exemple cela signifie que si nous observons les futures données de température sur une période de 20 ans, nous nous attendons à ce que au moins une valeur se situe entre environ 40.95°C et 43.45°C avec une probabilité de 95%. C'est la même interprétation pour le retour sur 100 ans mais on peut noter que l'intervalle est plus décalé vers les hautes températures ce qui paraît logique puisque en 100 ans on a plus de chance d'être confronté à des températures plus extrêmes que sur 20 ans.

3.2.3 Validation du modèle

On s'intéresse à la validation du modèle en étudiant les graphiques suivants :

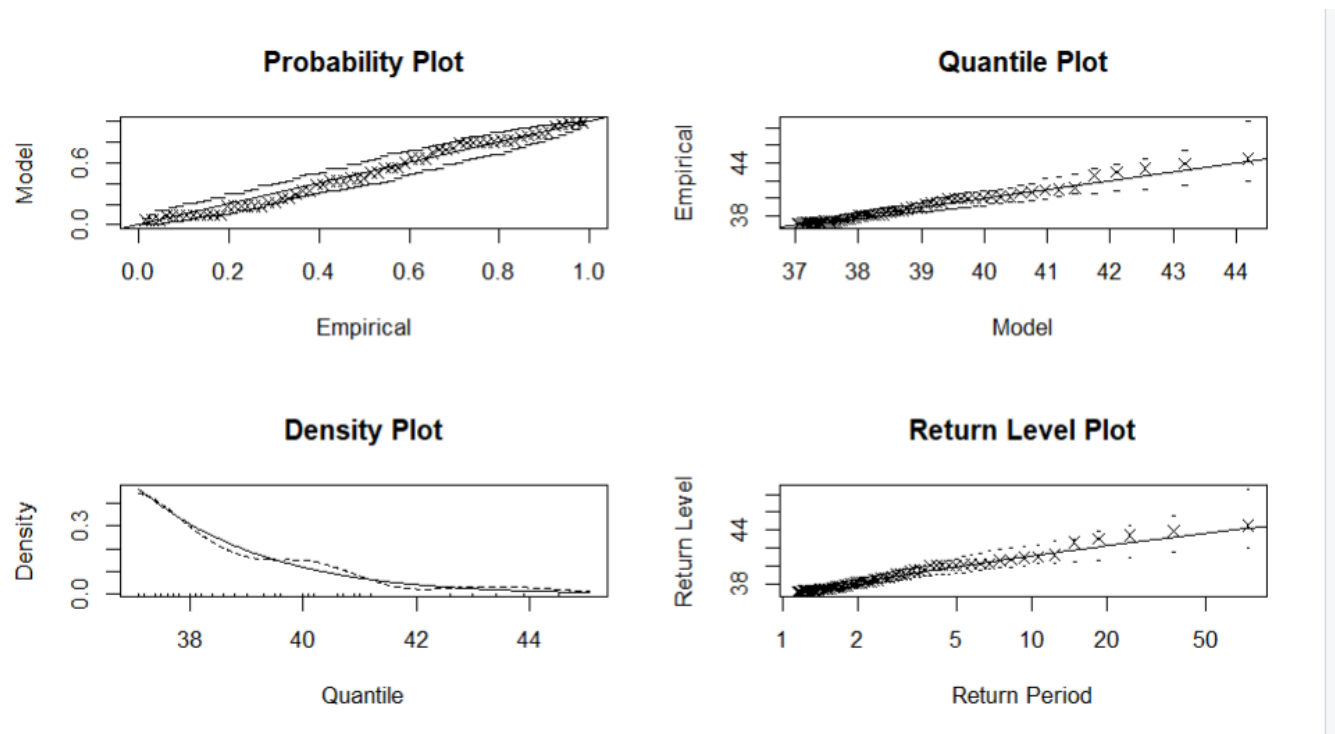


Figure 3.2.3: Validation du modèle

Explication des graphiques :

Les Probability Plot et Quantile Plot montrent un bon ajustement du modèle GPD aux données empiriques, les points se placent bien le long de la diagonale, indiquant que le modèle suit la tendance des données. Le Density Plot révèle que la densité estimée par le modèle et celle observée dans les données sont proches, ce qui renforce l'idée que le modèle s'ajuste bien aux données. Enfin, le Return Level Plot, avec des intervalles de confiance étroits, confirme que le modèle prédit avec

précision le niveau de retour pour différentes périodes. En résumé, le modèle GPD semble très bien adapté aux données et donne des résultats très satisfaisants.

3.3 Modèle à seuil avec clustering

Cette méthode se concentre sur les événements qui dépassent un certain seuil prédéterminé, considérés comme extrêmes. Elle intègre en plus la notion de clustering pour regrouper ces événements en fonction de leur proximité temporelle, permettant ainsi de distinguer les occurrences isolées des séquences d'événements extrêmes. On garde le même seuil que précédemment c'est à dire 37°C . On définit également de 'longueur de run' r , par exemple $r=2$, on établit qu'au moins deux valeurs non extrêmes consécutives sont nécessaires pour clôturer un cluster d'événements extrêmes en cours.

On applique cette méthode avec $r=2$ et $r=4$ on visualise les clusters suivant:

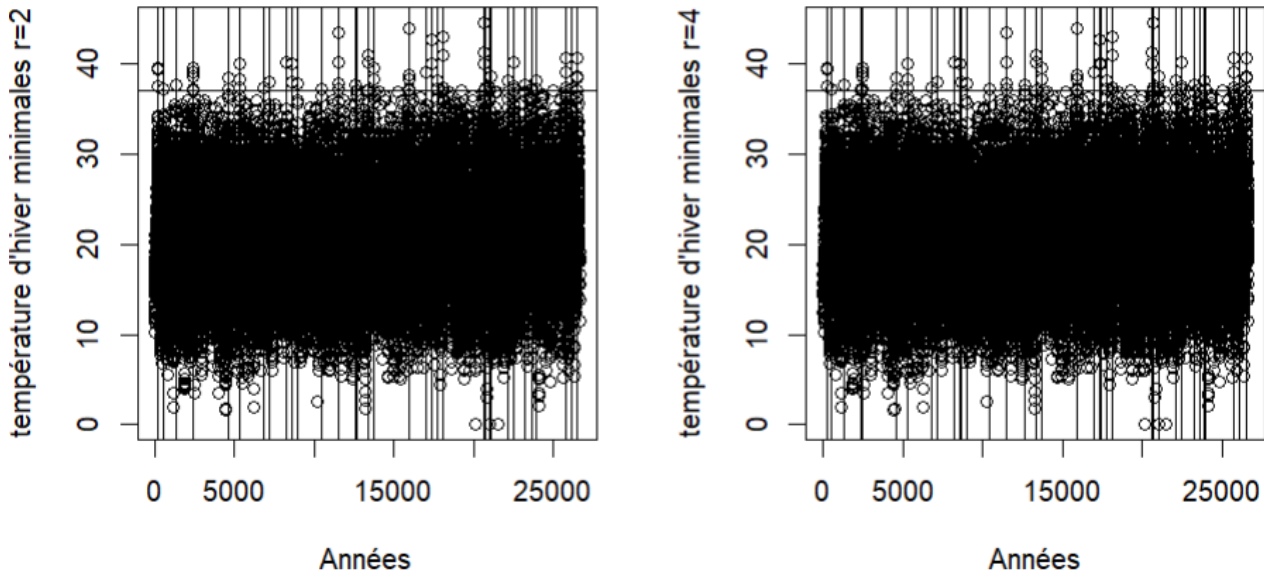


Figure 3.3.1: Représentation des clusters

3.3.1 Ajustement du modèle

On applique donc notre modèle et l'on obtient les résultats suivants :

r	Number of Clusters	Extremal Index	Scale Estimate (CI)	Shape Estimate (CI)
4	44	0.6769	2.5475 [1.3631, 3.7319]	-0.1737 [-0.5363, 0.1889]
2	47	0.7231	2.6140 [1.4792, 3.7488]	-0.2009 [-0.5327, 0.1309]

Table 2: Résumé des estimations et des intervalles de confiance pour les paramètres du modèle avec différents r .

L'observation d'un nombre légèrement plus élevé de clusters avec $r=2$ comparé à $r=4$ est cohérente avec les attentes : une longueur de run plus courte signifie que moins de valeurs non extrêmes consécutives sont nécessaires pour clore un cluster, conduisant ainsi à identifier plus de clusters. Les estimations de scale et shape sont cohérentes avec les estimations des modèles précédent.

Concernant l'indice extremal, sa valeur varie entre les deux analyses, avec une valeur plus élevée pour $r=2$. Cet indice mesure la tendance des données extrêmes à se regrouper : plus l'indice est proche de 1, moins il y a de dépendance entre les événements extrêmes, signifiant que les extrêmes sont plus dispersés. Une valeur plus élevée de l'indice extremal pour $r=2$ suggère donc une légère diminution de la dépendance ou une plus grande indépendance entre les événements extrêmes.

3.3.2 Niveau de retour

Période de Retour	Intervalle de Confiance à 95% pour rlevel
100 ans ($r=2$)	[42.03, 46.72]
100 ans ($r=4$)	[41.83, 47.11]
20 ans ($r=2$)	[41.05, 43.41]
20 ans ($r=4$)	[40.90, 43.40]

Table 3: Intervalle de confiance à 95% pour les niveaux de retour de 100 ans et de 20 ans, pour différentes longueurs de run.

On observe que le modèle avec $r=4$ génère des intervalles de confiance plus larges que pour $r=2$, ce qui peut le rendre moins précis, même s'il demeure interprétable. En revanche, pour $r=2$, on obtient des résultats plus serrés que ceux obtenus dans l'approche GPD, ce qui indique une meilleure précision par rapport à l'approche GPD. De plus, ces intervalles de confiance ne se chevauchent pas mutuellement avec ceux obtenus par d'autres méthodes, ce qui renforce la fiabilité de cette approche. Ainsi, le modèle avec $r=2$ apparaît légèrement meilleur pour l'estimation des niveaux de retour par rapport à l'approche GPD.

3.3.3 Validation du modèle

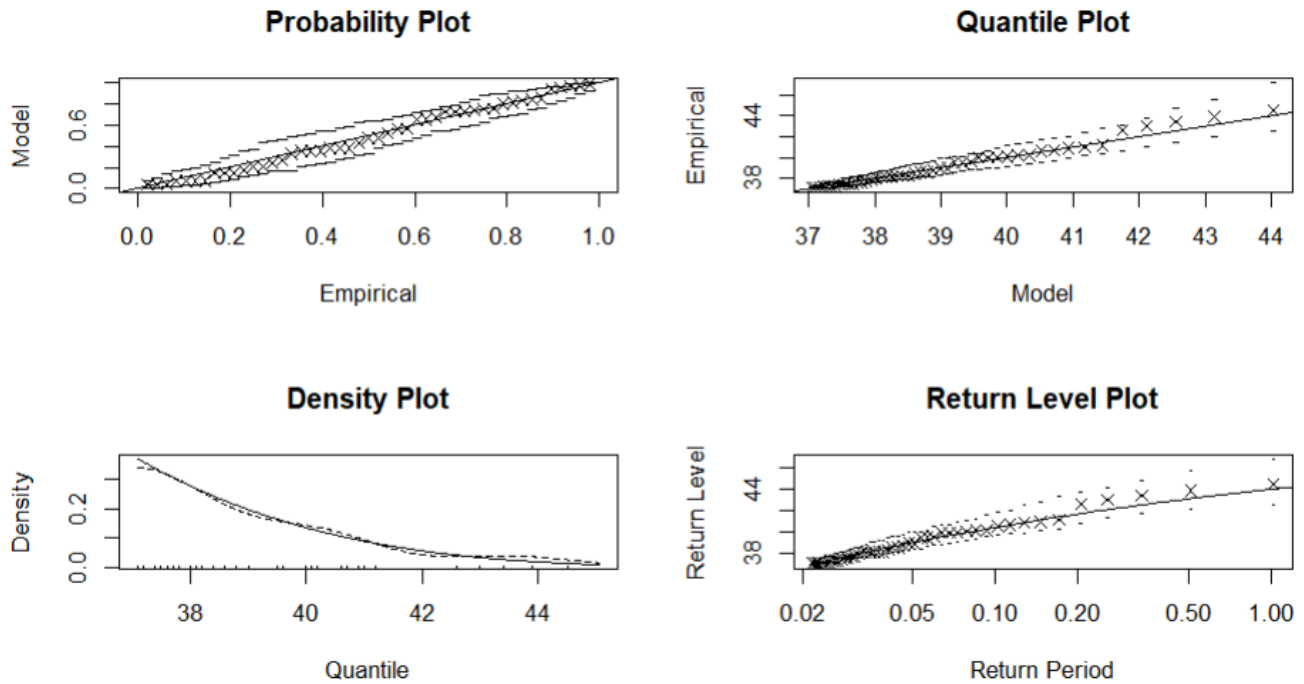


Figure 3.3.2: Validation du modèle avec $r=2$

Nous arrivons à la même conclusion qu'avec l'approche GPD : le modèle montre un ajustement satisfaisant. Cependant, nous observons une amélioration notable dans l'ajustement du density plot. Ce modèle se révèle donc parfaitement adapté à nos données.

3.4 Modèle non stationnaire

Même si nous avons vu que les données ne sont stationnaires d'après notre test précédent. Cependant on connaît l'impact du réchauffement climatique c'est pourquoi nous allons mettre en oeuvre un modèle qui se base sur le modèle GEV auquel on ajoute une tendance linéaire.

3.4.1 Ajustement du modèle

On visualise nos données avec la tendance linéaire. On travaille sur les maximum par années.

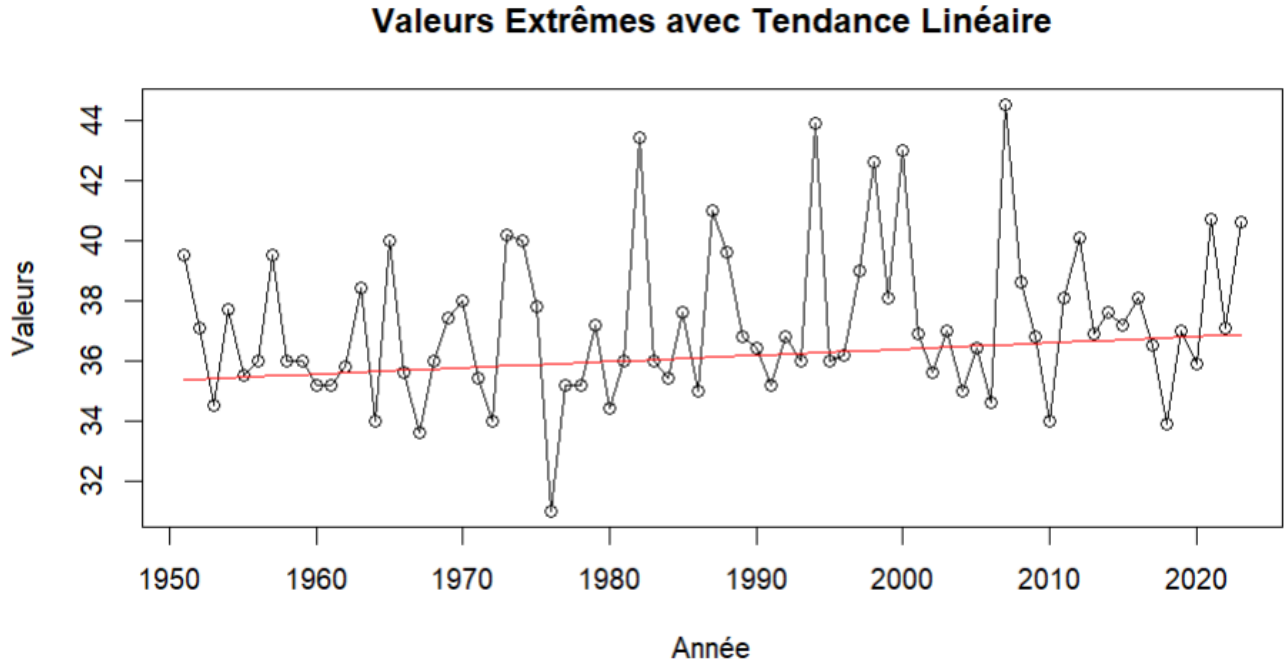


Figure 3.4.1: Visualisation des données extrêmes avec la tendance

La tendance linéaire apparaît relativement plate, ce qui est cohérent avec la conclusion précédente indiquant la stationnarité des données. Il n’y a pas de tendance claire.

3.4.2 Comparaison du modèle

Nous avons comparé le modèle avec tendance linéaire à un modèle sans tendance en utilisant un test de déviance, qui a donné une p-valeur de 0,1. Cela signifie que l’ajout d’une tendance linéaire n’améliore pas significativement l’ajustement du modèle. Ce résultat était prévu, étant donné l’apparence plate de la tendance dans les données observées et la stationnarité de nos données.

4 Conclusion

Nous avons exploré plusieurs modèles, pour capturer les caractéristiques des événements extrêmes. Les modèles ont été évalués pour leur ajustement et leur capacité à estimer les niveaux de retour. Malgré la stationnarité des données signalée par les tests de Dickey-Fuller augmenté, un modèle GEV avec tendance linéaire a été considéré pour tenir compte de la non-stationnarité potentielle due au réchauffement climatique. Cependant, le test de déviance a révélé que la tendance linéaire n’apportait pas d’amélioration significative, soulignant l’adéquation des modèles stationnaires dans ce contexte. Les résultats ont indiqué que l’approche GPD, en particulier avec un seuil et clustering, offrait une grande amélioration par rapport aux modèles GEV classiques, fournissant des estimations plus précises pour les niveaux de retour.

5 Annexe Code R

```
1 #=====
2 # Chargement des bibliotheques necessaires
3 #=====
4
5 library(summarytools)
6 library(evd)
7 library(tseries)
8 library(ismev)
9
10 # =====
11 # Preparation des donnees
12 # =====
13
14 data <- read.csv("C:/Users/hrywa/OneDrive/Bureau/M2/Statistiques_spatiales/
15   BRINDISI.txt", sep="")
16 data[is.na(data)] <- 20
17
18 # =====
19 # Analyse exploratoire des donnees
20 # =====
21
22 # Histogramme des Temperatures
23 hist(data$TX, xlab='Temperature', main='Histogramme des Temperatures')
24
25 # Resume des donnees
26 view(dfSummary(data))
27
28 # Graphique de serie temporelle
29 plot(data$Years, data$TX, type="l", xlab='Annees', ylab='Temperature')
30
31 # Calcul des maxima annuels
32 annees <- unique(data$Years)
33 max_annee <- numeric(length(annees))
34 for (i in seq_along(annees)) {
35   max_annee[i] <- max(data[data$Years == annees[i], ]$TX)
36 }
37 plot(annees, max_annee, type="l", xlab='Annees', ylab='Temperature', main="
38   temperature maximale par an")
39
40 # =====
41 # Test de stationnarite
42 # =====
43
44 # Application du test ADF sur les donnees
45 result <- adf.test(data$TX)
46 print(result)
47
48 result2 <- adf.test(max_annee)
49 print(result2)
```

```

48
49 # =====
50 # Analyse GEV des maximums par annee
51 # =====
52
53 # Ajustement et analyse GEV
54 gev.fit <- fgev(max_annee)
55 plot(profile(gev.fit, which = names(gev.fit$estimate[1])))
56 plot(profile(gev.fit, which = names(gev.fit$estimate[2])))
57 plot(profile(gev.fit, which = names(gev.fit$estimate[3])))
58 confint(gev.fit)
59
60 #=====
61 # Calcul des niveaux de retour
62 #=====
63
64 variance=gev.fit$var.cov
65 scale=gev.fit$estimate[2]
66 shape=gev.fit$estimate[3]
67 loc=gev.fit$estimate[1]
68
69 # Niveaux de retour pour T=20 et T=100 ans avec methode Delta
70 y20 <- -log(1 - 1/20)
71 y100 <- -log(1 - 1/100)
72
73 # Niveau de retour pour T=20 et T=100 ans
74 Z20 <- loc - (scale / shape) * (1 - (y20 ** shape))
75 Z100 <- loc - (scale / shape) * (1 - (y100 ** shape))
76
77 # Methode Delta pour T=20
78 delta_20 <- c(1, shape ** -1 * (1 - (-log(1 - 1/20)) ** shape),
79               scale * shape ** -2 * (1 - (-log(1 - 1/20)) ** shape) - scale *
80               shape ** -1 * (-log(1 - 1/20)) ** shape * log(-log(1 - 1/20)))
81 var_20 <- t(delta_20) %*% variance %*% delta_20
82
83 # Methode Delta pour T=100
84 delta_100 <- c(1, shape ** -1 * (1 - (-log(1 - 1/100)) ** shape),
85               scale * shape ** -2 * (1 - (-log(1 - 1/100)) ** shape) - scale *
86               shape ** -1 * (-log(1 - 1/100)) ** shape * log(-log(1 - 1/100)))
87 var_100 <- t(delta_100) %*% variance %*% delta_100
88
89 # Calcul des intervalles de confiance a 95%
90 alpha <- 0.05
91 z_alpha <- qnorm(1 - alpha/2)
92
93 # Intervalle de confiance pour Z20
94 lower_Z20 <- Z20 - z_alpha * sqrt(var_20)
95 upper_Z20 <- Z20 + z_alpha * sqrt(var_20)
96
97 # Intervalle de confiance pour Z100

```

```

97 lower_Z100 <- Z100 - z_alpha * sqrt(var_100)
98 upper_Z100 <- Z100 + z_alpha * sqrt(var_100)
99
100 # Affichage des intervalles de confiance
101 cat("Intervalle de confiance a 95% pour Z20 : [", lower_Z20, ", ", upper_Z20, "
    ]\n")
102 cat("Intervalle de confiance a 95% pour Z100 : [", lower_Z100, ", ", upper_Z100
    , "]\n")
103
104
105 # =====
106 # Approche GPD
107 # =====
108
109 # Preparation de la serie temporelle pour l'approche GPD
110 temperature_ts <- ts(max_annee, start=1951, end=2023)
111 mrlplot(temperature_ts) # Pour identifier visuellement le seuil
112 plot(data$TX, type='l', main="Températures avec Seuils", xlab="Temps", ylab="
    Temperature")
113 abline(h=37, col="blue", lty=2)
114
115 # Modele GPD
116 mod <- fpot(data$TX, 37)
117 mod
118
119 #=====
120 #calcul a la main des niveau de retour
121 #=====
122 # Niveau de retour pour T=20 et T=100 ans
123 C= sum(data$TX > 37)/length(data$TX)
124
125 R100=37+(1.96179 /(-0.05828 ))*((100*365*C)^-0.05828 -1) # *365 car periode
    de retour en annees et nos donnees sont journaliere
126 R100
127
128 R20=37+(1.96179 /(-0.05828 ))*((20*365*C)^-0.05828 -1) # *365 car periode
    de retour en annees et nos donnees sont journaliere
129 R20
130
131 matrice=mod$var.cov
132 zeta=-0.05828 #la shape de mod
133 V=cbind(c(zeta*(1-zeta)/length(data$TX),0,0),rbind(rep(0,2),mod$var.cov))
134 delta3 <- c(
135     1.96179 ^(-0.05828 ) * C^(-0.05828 - 1),
136     1 / -0.05828 * ((100 * 365 * C)^-0.05828 - 1),
137     -1.96179 * (-0.05828 )^(-2) * ((100 * 365 * C)^-0.05828 - 1) + 1.96179
        * 1 / (-0.05828 ) * (100 * 365 * C)^(-0.05828 ) * log(100 * 365 * C)
138 )
139 V_x100=t(delta3)%*% V %*%delta3
140 V_x100
141

```

```

142 delta4 <- c(
143   1.96179 ^(-0.05828 ) * C^(-0.05828 - 1),
144   1 / -0.05828 * ((20 * 365 * C)^-0.05828 - 1),
145   -1.96179 * (-0.05828 )^(-2) * ((20 * 365 * C)^-0.05828 - 1) + 1.96179
146     * 1 / (-0.05828 ) * (20 * 365 * C)^(-0.05828 ) * log(20 * 365 * C)
147 )
148 V_x20=t(delta4)%*% V %*%delta4
149 V_x20
150 #verification avec fpot
151 Mod100=fpot(data$TX,37,npp=365,mper=100)
152 Mod20=fpot(data$TX,37,npp=365,mper=20)
153
154
155 # =====
156 # Affichage des niveaux de retour et variances pour T=20 et T=100 ans
157 # =====
158 Mod100
159 R100
160 V_x100
161
162 #intervalle de confiance [ rlevel+/- 1.96* standart errors]
163
164 Mod20
165 R20
166 V_x20
167
168 #=====
169 #VAlidation du modele
170 #=====
171 par(mfrow=c(2,2))
172 plot(Mod100)
173
174 plot(Mod20)
175
176 #=====
177 #Autre modele a seuil avec cluster
178 #=====
179
180 #reresentation des donnees
181 plot(data$TX, type='l', main="Temperatures avec Seuils", xlab="Temps", ylab="
  Temperature")
182 abline(h=37, col="blue", lty=2)
183
184
185 #Parametres
186 r1=2
187 r2=4
188 u1=37
189
190 par(mfrow=c(1,2))

```



```

191 #representation des clusters
192 clusters(data$TX,u=u1,r=r1,plot=TRUE,col="white",xlab="Annees",ylab="
    temperature d'hiver minimales r=2")
193 clusters(data$TX,u=u1,r=r2,plot=TRUE,col="white",xlab="Annees",ylab="
    temperature d'hiver minimales r=4")
194
195
196 #Estimation du modele GPD pour les donnees au-dessus du seuil avec differents r
197 Mod1=fpot(data$TX,threshold = u1,cmax=TRUE,r=r1)
198 Mod2=fpot(data$TX,threshold = u1,cmax=TRUE,r=r2)
199
200
201 # Estimation de l'indice extremal theta pour chaque modele
202 exi(data$TX,u=u1,r=r1)
203 exi(data$TX,u=u1,r=r2)
204
205 #les niveaux de retour
206 fpot(data$TX,u1,npp=365,mper=100,cmax=TRUE,r=r1)
207
208 fpot(data$TX,u1,npp=365,mper=100,cmax=TRUE,r=r2)
209
210 fpot(data$TX,u1,npp=365,mper=20,cmax=TRUE,r=r1)
211
212 fpot(data$TX,u1,npp=365,mper=20,cmax=TRUE,r=r2)
213
214 #=====
215 #Validation du modele
216 #=====
217 par(mfrow=c(2,2))
218 plot(Mod1)
219 #=====
220 #Si les donnees sont non stationnaire
221 #=====
222
223
224 # Ajustement d'un modele GEV avec une tendance lineaire sur les donnees de
    temperature
225 Annees <- 1951:2023
226 mod3 <- fgev(max_annee, nsloc =1: length(max_annee))
227 mod3
228
229 # Trace des donnees et de la tendance estimee par le modele
230 plot(Annees,max_annee, type = 'o', main = "Valeurs Extremes avec Tendance
    Lineaire", xlab = "Annee", ylab = "Valeurs")
231 Mu <- mod3$estimate[1] + mod3$estimate[2] * (Annees - min(Annees)) #
    Ajustement pour la tendance lineaire
232 lines(Annees, Mu, type = "l", col = "red")
233
234
235
236 #Calcul de vraisemblance

```

```

237 l1=-sum(log(mod3$estimate[3])+(1+1/mod3$estimate[4])*log(1+mod3$estimate[4]*(max
    _annee-Mu)/mod3$estimate[3])+(1+mod3$estimate[4]*((max_annee-Mu)/mod3$
    estimate[3]))^(-1/mod3$estimate[4]))
238
239 #verification avec la fonction
240 Mod3bis=gev.fit(max_annee,ydat=matrix(1:length(max_annee),ncol=1),mul=1)
241 #nllh single numeric giving the negative log-likelihood value. 117.9764 la
    vraisemblance
242
243
244
245 #autre modele classique pour comparaison
246 m=length(max_annee)
247 Mod_iid=fgev(max_annee)
248 l2=-m*log(Mod_iid$estimate[2])-(1+1/Mod_iid$estimate[3])*sum(log(1+Mod_iid$
    estimate[3]*
249
    (max_annee-
    Mod_iid$estimate[1])/Mod_iid$estimate[2]))-sum((1+Mod_iid$estimate[3]*(max_
    annee-Mod_iid$estimate[1])/Mod_iid$estimate[2])^(-1/Mod_iid$estimate[3]))
250
251 # Calcul de la deviance
252 deviance <- 2 * (l1 - l2) #du grand moins le petit #a comparer a qchisq
    (0.95,1)
253
254 # Nombre de degres de liberte (difference dans le nombre de parametres)
255 df <- length(mod3$estimate) - length(Mod_iid$estimate)
256
257 # Test du Chi-carre
258 p_value <- pchisq(deviance, df, lower.tail = FALSE)
259
260 # Afficher les resultats
261 print(c("Deviance" = deviance, "P-Value" = p_value))

```