

Caractérisation des dynamiques de pluie dans la région de Boeny

L'étude menée ci-après aborde le traitement de données de précipitations relevées au quotidien en un lieu, pour lesquelles on veut ajuster une loi de distribution statistique. Afin de réaliser cet objectif, des analyses et des traitements visant à établir les caractéristiques de ces données seront nécessaires.

Comme on pourra le voir, les premières observations des données montrent les caractères aléatoires et très erratiques des variations de la quantité de pluie. De telles caractéristiques sont généralement observées et admises depuis longtemps. Elles ont donné lieu dans la littérature à de nombreux travaux d'analyse et de modélisation adoptant l'approche stochastique [57][58][59][60][61]. Au-delà des constatations que l'on retrouve largement dans de nombreuses publications, l'étude plus approfondie de la structure probabiliste associée à ces données et les caractérisations du type de dynamique de variations observées doivent permettre d'améliorer la réflexion sur l'ajustement de la loi ou la modélisation de ces précipitations.

Dans ce cas, l'intensité des précipitations et le temps d'attente entre deux pluies seront considérées comme deux variables aléatoires dont on déterminera les lois de probabilité.

3.1 Courbes des individus et valeurs propres par ACP

La figure 10 présente la courbe d'évolution des valeurs propres. D'après Kaiser, on prend la valeur propre supérieure à 1 ; on retient donc les axes F1 et F2.

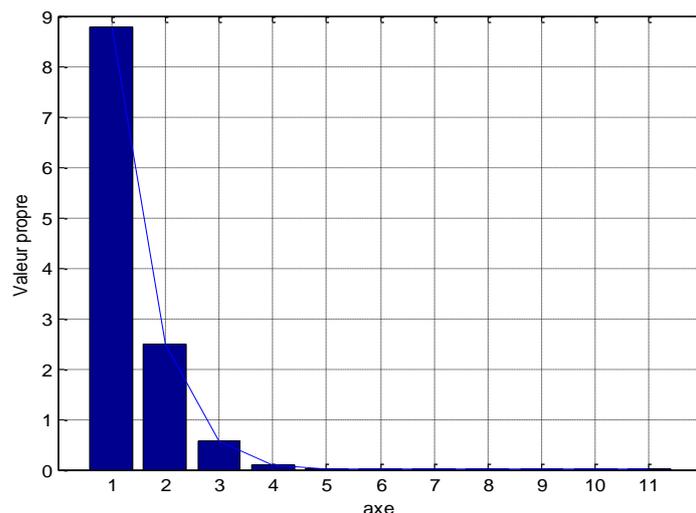


Figure 10 : Courbe d'évolution des valeurs propres

Nous avons réalisé l'Analyse en Composante Principale (ACP) pour définir la zone représentée sur la figure 11. Pour ce faire, nous prenons la moyenne climatologique mensuelle de la pluviométrie de 1979 à 2015.

On maille de $0,5^\circ$ à $0,5^\circ$. Pour le calcul en ACP, les individus sont les nœuds dénommés A1 à A7 ; B1 à B7 ; C1 à C7; D1 à D7; E1 à E7; F1 à F7; G1 à G7; H1 à H7 et I1 à I7. Ces individus ont des longitudes et des latitudes bien spécifiques. Par contre, les mois sont les variables.

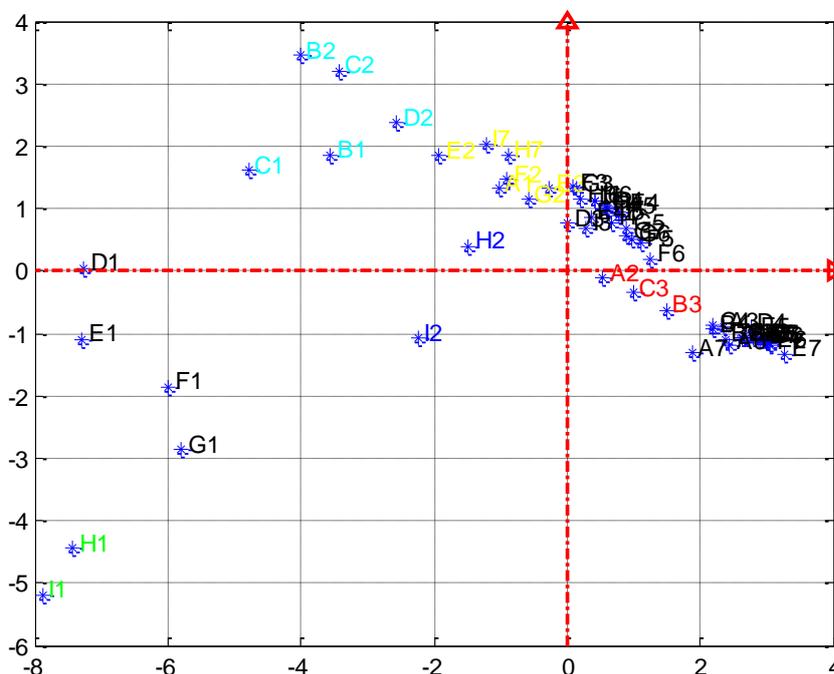


Figure 11 : Représentation des individus dans le plan factoriel F1 et F2

On peut diviser notre zone d'étude en neuf zones, comme indiquée ci dessous :

- Zone 1: H1 et I1
- Zone 2: F1 et G1
- Zone 3: D1 et E1
- Zone 4: H2 et I2
- Zone 5: A1, E2, E3, F2, G2, H7 et I7
- Zone 6: B1, B2, C1, C2 et D2
- Zone 7: A2, B3 et C3
- Zone 8: D3, E4, E5, F3, F4, F5, F6, F7, G3, G4, G5, G6, G7, H3, H4, H5, H6, I3, I4, I5 et I6
- Zone 9: A3, A4, A5, A6, A7, B4, B5, B6, B7, C4, C5, C6, C7, D4, D5, D6, D7, E6 et E7

3.2 Détermination des années sèches et humides [62] [63] [64] [65]

[66]

La variabilité interannuelle du régime pluviométrique peut être analysée à partir de la distribution des années sèches et des années humides. Mais les définitions d'une année

sèche (ou déficitaire) et d'une année humide (ou excédentaire) sont très diverses. En effet, on peut définir une année sèche ou humide, soit par rapport aux quantiles, soit par rapport aux écarts types ou, en pourcentage, par rapport à la médiane ou encore à la moyenne. Dans la présente étude, une année sèche ou une année humide est définie par rapport à l'indice de Lamb (l'écart à la moyenne normalisé par l'écart type) qui s'exprime par:

$$ISP(i) = \frac{[H(i)-m]}{\sigma} \quad (3.1)$$

Avec $H(i)$: cumul moyen annuel pour l'année i .

m : la moyenne de la série.

σ : l'écart type de la série

ISP : Indice Standardisé des Précipitations.

Le calcul de cet indice permet de déterminer la sévérité de la sécheresse selon différentes classes :

- $ISP > 2$ Humidité extrême
- $1 < ISP < 2$ Humidité forte
- $0 < ISP < 1$ Humidité modérée (3.2)
- $-1 < ISP < 0$ Sécheresse modérée
- $-2 < ISP < -1$ Sécheresse forte
- $ISP < -2$ Sécheresse extrême

Ainsi, une année sera considérée comme normale (ou modérée) si son indice $ISP(i)$ est compris entre -1 et +1. Elle sera dite humide si son indice est supérieur à +1 et sèche en deçà de -1. Cet intervalle reste critiquable puisqu'il est relativement faible de sorte que les années normales sont très peu nombreuses. Toute fois, il permet de bien distinguer les années sèches et les années humides.

3.3 Moyenne arithmétique [67] [68]

D'après le langage courant, la moyenne réfère en général à la moyenne arithmétique. Elle est exprimée par la formule :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.3)$$

où n est le nombre total d'observations de l'échantillon à étudier.

Son écart-type empirique est de :

$$\sigma = \sqrt{\frac{(x_i - \bar{x})^2}{n}} \quad (3.4)$$

3.4 Moyenne glissante ou moyenne mobile [69] [70]

La moyenne glissante est un type de moyenne statistique utilisé pour analyser des séries ordonnées de données. Elle supprime les fluctuations transitoires de façon à souligner les tendances plus long terme. Cette moyenne est dite mobile car elle est recalculée de façon continue, en utilisant à chaque calcul un sous ensemble.

La moyenne mobile simple est calculée par la formule :

$$\bar{x} = \frac{1}{N} \sum_{k=0}^{N-1} x_{n-k} \text{ ou}$$

$$\bar{x}_n = \bar{x}_{n-1} + \frac{\bar{x}_n - \bar{x}_{n-N}}{N} \quad (3.5)$$

Où $N \leq n$ et N est le nombre de valeurs dans le sous groupe consécutif.

L'avantage d'une moyenne mobile est de lisser les éventuels écarts accidentels.

La courbe en histogramme de la figure 12 permet de savoir la nature pluviométrique de chaque année en moyennant les données de toute la zone d'étude.

Les deux droites $x = 1$ et $x = -1$ délimitent les années normales avec les années excédentaires ou déficitaires (trop humide ou trop sèche).

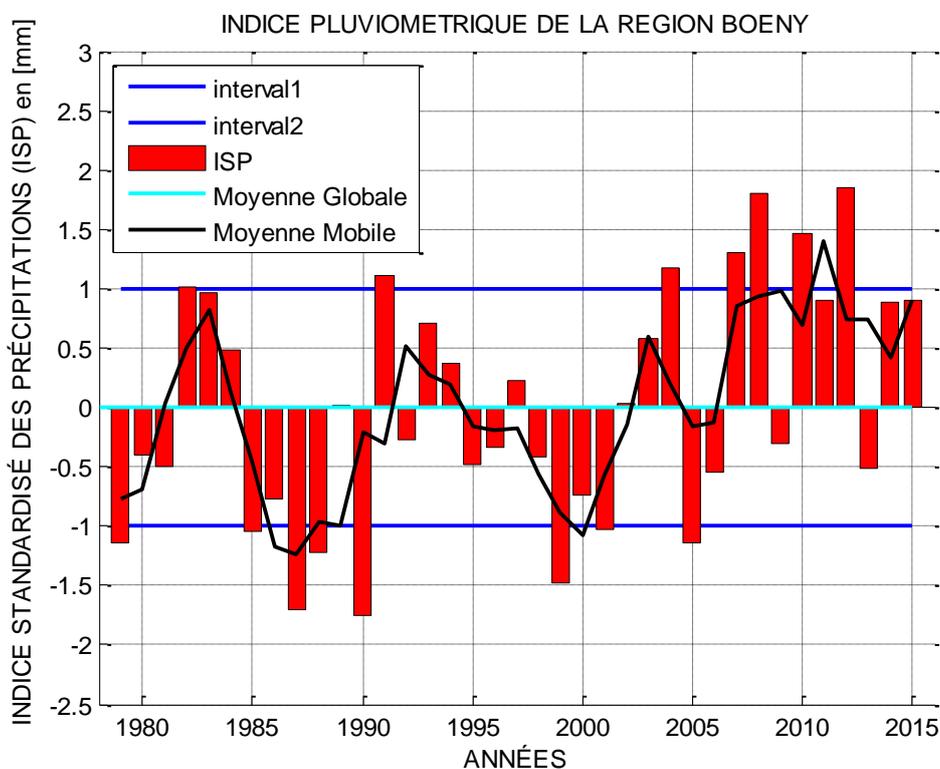


Figure 12 : Représentation graphique des ISP.

Les données pluviométriques enregistrées dans la région de Boeny, à travers l'indice standardisé de précipitations ISP (Tableau 1), caractérisent une situation majoritairement dominée par des situations modérées qui représentent le 56,76 % des cas, dont : 27,03% pour l'humidité modérée et 29,73% pour la sécheresse modérée. L'humidité forte représente le 18,92% des cas, tandis que la sécheresse forte marque le 21,62 %. Le cas de l'année normale représente 2,70%.

La moyenne mobile nous dit qu'entre 1985 et 1990 il y a un déficit de pluie. Il en est de même de 1998 à 2001. Par contre entre 2004 et 2012 il y a un excès de pluie.

D'après le tableau 1, les années 1979, 1980, 1981, 1985, 1986, 1987, 1988, 1998, 1999, 2000 et 2001 sont particulièrement concernées par le déficit pluviométrique annuel. Cette persistance de la sécheresse, sur plusieurs années, a provoqué de grands déséquilibres au niveau de la pluviométrie du Boeny. Une suite successive de deux ou plusieurs années de sécheresse modérée est plus dramatique et catastrophiquement sévère pour le réchauffement global et même pour l'agriculture.

Parmi d'autres années, les années 2008, 2010 et 2012 sont très marquées par sa supériorité en abondance pluviométrique, ces années représentent une humidité forte. Tandis que les années 1987, 1990 et 1999 sont très particulières à cause des déficits pluviométriques (sécheresse fortes), ce sont les années les plus sèches depuis 1979 jusqu'à 2015 dans notre zone d'étude.

Tableau 1 : Indice standardisé de précipitation (ISP) de l'année 1979 à 2015

ANNEE	INDICE ISP	NATURE DE L'ANNEE
1979	-1,15	Sècheresse forte
1980	-0,409	Sècheresse modérée
1981	-0,503	Sècheresse modérée
1982	+1,02	Humidité forte
1983	+0,967	Humidité modérée
1984	+0,479	Humidité modérée
1985	-1,04	Sècheresse forte
1986	-0,769	Sécheresse modérée
1987	-1,7	Sécheresse forte
1988	-1,22	Sècheresse forte
1989	0	Normale
1990	-1,76	Sècheresse forte
1991	+1,11	Humidité forte
1992	-0,27	Sècheresse modérée
1993	+0,716	Humidité modérée
1994	+0,365	Humidité modérée
1995	-0,487	Sécheresse modérée
1996	-0,344	Sécheresse modérée
1997	+0,234	Humidité modérée
1998	-0,416	Sécheresse modérée
1999	-1,48	Sècheresse forte
2000	-0,735	Sècheresse modérée
2001	-1,03	Sècheresse forte
2002	+0,0332	Humidité modérée
2003	+0,5969	Humidité modérée
2004	+1,18	Humidité forte
2005	-1,14	Sècheresse forte
2006	-0,541	Sécheresse modérée
2007	+1,3	Humidité forte
2008	+1,8	Humidité forte
2009	-0,3	Sècheresse modérée
2010	+1,46	Humidité forte
2011	+0,9	Humidité modérée
2012	+1,85	Humidité forte
2013	-0,507	Sècheresse modérée
2014	+0,881	Humidité modérée
2015	+0,8935	Humidité modérée

Et l'année 1989 représente l'année normale durant ces 37 ans.

La persistance de ce déficit pluviométrique nous incite à étudier la tendance des données pluviométrique dans la région de Boeny.

3.5 Test statistique

3.5.1 Techniques non-paramétriques pour l'analyse des séries chronologiques

Les techniques non-paramétriques présentées ci-après sont applicables aux cas où les données X_i d'une série chronologique peuvent être écrites sous la forme suivante :

$$X_i = f(t_i) + \varepsilon_i \quad (3.6)$$

Où $f(t_i)$ est une fonction continue croissante ou décroissante, et ε_i les résidus, de même distribution, ont une moyenne nulle.

3.5.2 Test sur la tendance

En premier lieu, il est nécessaire de définir les hypothèses H_0 telles qu'elles sont testées par les méthodes statistiques dédiées à cette tâche :

- l'hypothèse nulle H_0 pour les tests de tendance (ou de stationnarité) est qu'il n'y a pas de tendance significative dans les données.
- l'hypothèse nulle H_0 pour les tests de saut de moyenne/médiane ou de différence dans les moyennes/médianes (tests de rupture de pente) est qu'il n'y a pas de saut ou de différence significative entre les moyennes/médianes de deux portions d'une même série chronologique de données.

3.5.2.1 Test de Mann Kendall

Le test de Mann-Kendall est un test non paramétrique de tendance monotonique.

Cette méthode permet de tester s'il existe une tendance monotonique dans une série de données chronologiques. Le test est non paramétrique et ne demande donc pas de vérification de la distribution des données. Le principe du test est le remplacement des n valeurs de la série temporelle $(x_1, x_2, x_3, \dots, x_n)$ par leur rang relatif $(r_1, r_2, r_3, \dots, r_n)$.

Le test statistique S est alors le suivant :

$$S = \sum_{i=1}^{n-1} [\sum_{j=i+1}^n \text{sgn}(r_j - r_i)] \quad (3.7)$$

Où $\text{sgn}(x) = 1$ pour $x > 0$

$$\text{sgn}(x) = 0 \text{ pour } x = 0 \quad (3.8)$$

$$\text{sgn}(x) = -1 \text{ pour } x < 0$$

Si l'hypothèse nulle H_0 est vraie, S approche d'une distribution normale avec la valeur de moyenne m et de la variance Var , respectivement :

$$m = 0$$

$$Var = \frac{(n)(n-1)(2n+5)}{18} \quad (3.9)$$

La statistique Z est alors de la forme (les valeurs critiques du test statistique pour différents niveaux de confiance sont reportées dans les tables de probabilités de distribution normale)

$$Z = \frac{|S|}{\sqrt{Var}} \quad (3.10)$$

La statistique S est significative au risque de 5 % si $Z > 1,96$ et hautement significative au risque de 1% si $Z > 2,57$ (distribution normale) [71] [72] [73] [74] [75].

Une valeur positive de S indique une tendance à l'augmentation et vice versa.

Ainsi, pour notre série de données la figure 13 représente les différentes valeurs de « S » sous la forme progressive (de 1979 à 2015) et le résultat du test dans le tableau 2.

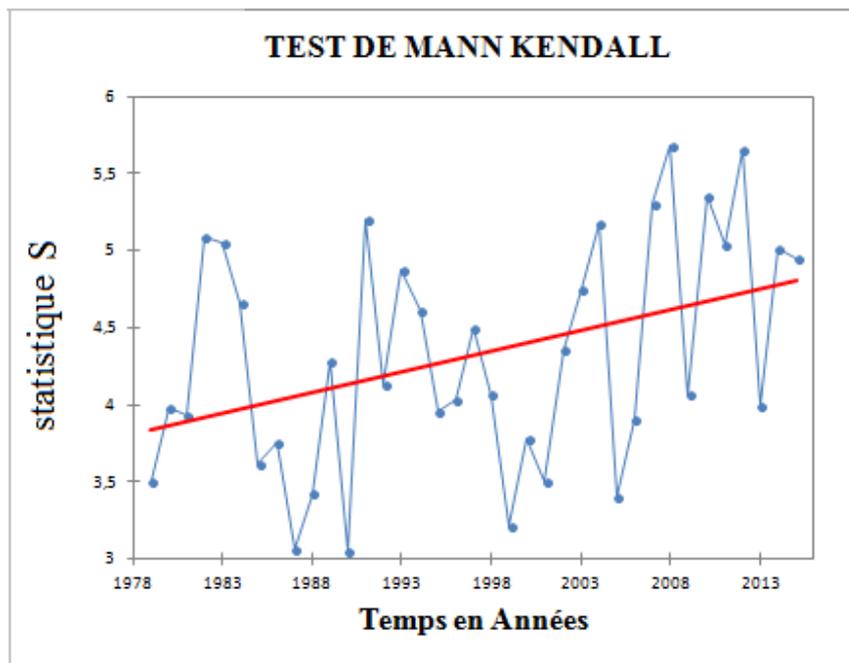


Figure 13 : Mann Kendall.

Test de tendance de Mann-Kendall / Test bilatéral :

Tableau 2 : Résultats du test

Tau de Kendall	0,2402
S	160,0000
p-value (bilatérale)	0,0370
alpha	0,05

La p-value est calculée suivant une méthode exacte.

Interprétation du test :

H_0 : Il n'y a pas de tendance dans la série

Ha : Il existe une tendance dans la série

Etant donné que la p-value calculée est inférieure au niveau de signification $\alpha=0,05$, l'hypothèse nulle H_0 est rejetée donc l'hypothèse alternative H_a est retenue bien que le risque de rejeter l'hypothèse nulle H_0 est inférieur à 3,70%.

Bref le test de Mann Kendall révèle que les données pluviométriques présentent une tendance significative linéairement croissante.

3.5.3 Test de rupture

3.5.3.1 Test de Pettitt

Le test de Pettitt dérive du test de Mann-Kendall mais il est dédié à la détection d'une rupture que plutôt d'une tendance. Le principe de ce test non paramétrique est d'évaluer si les couples de valeurs ont tendance à être rangés préférentiellement dans un sens, les couples étant formés par sélection d'une valeur de chaque côté d'un point de rupture hypothétique k (k est un entier compris entre 1 et $n-1$). Le test statistique de Pettitt $U(k)$ est de la forme :

$$U(k) = \sum_{i=1}^k [\sum_{j=k+1}^n \text{sgn}(x_j - x_i)] \quad (3.11)$$

Où $\text{sgn}(x) = 1$ pour $x > 0$

$$\text{sgn}(x) = 0 \text{ pour } x = 0 \quad (3.12)$$

$$\text{sgn}(x) = -1 \text{ pour } x < 0$$

La statistique Z de ce test vaut : $Z = \max |U(k)|$ avec $k = 1, 2, 3, \dots, n$ (3.13)

Soit α , représentant un risque de première espèce ;

H_0 est rejetée si une certaine probabilité calculée à partir d'une série chronologique étudiée est inférieure à α [76] [77] [78].

La fonction de répartition de Z peut-être approchée asymptotiquement par :

$$F(z) = P(Z > z) = 2e^{\left(\frac{-6z^2}{n^3+n^2}\right)} \quad (3.14)$$

Le test de Pettitt permet en outre de faire une estimation du point de rupture \hat{k} tel que :

$$\hat{k} = \text{Arg max } |U(k)| \quad (3.15)$$

Avec $k = 1, 2, 3, \dots, n$

Le seuil de signification est donné par la quantité suivante :

$$Q = \pm \sqrt{\frac{-(n^3+n^2) \log\left(\frac{\alpha}{2}\right)}{6}} \quad (3.16)$$

(Source: Partenariat 2010 Domaine : Savoirs Action n°8 Janv. 2011 - p 45/166)

Nous avons deux maxima (figure 14) qui sont très proches en valeur, l'année 2001 et l'année 2006. Le maximum des valeurs absolues de U est égale à 172 (ou bien $\max|U| = 172$)

$$H_0 \text{ est rejetée si } F(z) = P(172 > z) = 2e^{\frac{-6z^2}{n^3+n^2}} < \alpha \quad (3.17)$$

$$\text{d'où } z \text{ doit vérifier les deux conditions } z > \sqrt{\frac{-(n^3+n^2)\log(\frac{\alpha}{2})}{6}} \text{ et } z < 172 \quad (3.18)$$

De plus, α doit être inférieur à 2, car $\log(\frac{\alpha}{2})$ devrait être inférieur à zéro. Ces conditions ne seront pas remplies qu'à partir de $\alpha > 1,4$.

Conclusion : Si l'on va rejeter l'hypothèse H_0 , il y aura un risque d'erreur de décision de l'ordre de 15%. (C'est trop élevé !).

Les tests numériques par la méthode de Pettitt ne nous permettent pas d'affirmer l'existence d'une date de rupture au sein des données pluviométriques. Donc, la série pourrait être stationnaire. Cette situation nous amène à appliquer un autre test (test de CUSUM) pour essayer de trouver l'existence d'un point de rupture.

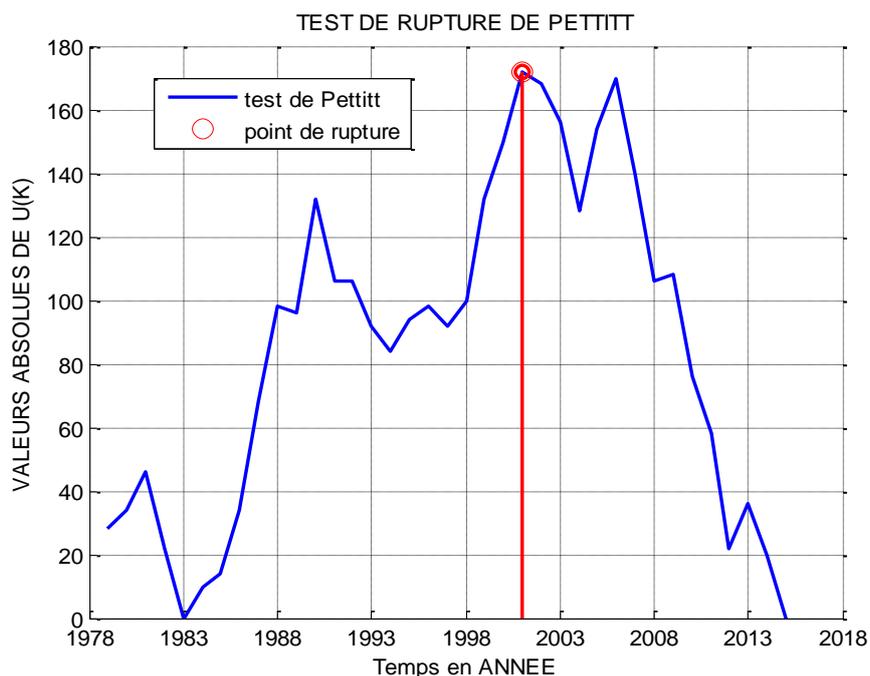


Figure 14 : Test de Pettitt pour la rupture.

3.5.2.2 Test de CUSUM (somme cumulée) [79][80][81]

Le test CUSUM (ou somme cumulée) est « un test non paramétrique de rupture ». Il permet de tester si les moyennes de deux parties d'une série chronologique sont significativement différentes (pour une date de changement inconnue).

Etant donné une série chronologique $(x_1, x_2, x_3, \dots, x_n)$, le test statistique est défini par :

$$V_K = \sum_{i=1}^k \text{sgn}(x_i - x_{\text{médiane}}) \quad k = 1, 2, 3, \dots, n \quad (3.19)$$

Où $\text{sgn}(x) = 1$ pour $x > 0$

$\text{sgn}(x) = 0$ pour $x = 0$ et $x_{\text{médiane}}$ est la valeur médiane de la série x_i .

$\text{sgn}(x) = -1$ pour $x < 0$

La distribution de V_k suit la loi statistique de Kolmogorov-Smirnov pour deux échantillons

($KS = (2/n) \cdot \max |V_K|$) avec les valeurs critiques de $\max(|V_K|)$ correspondant à :

$$\alpha = 0,10 \rightarrow 1,22 \sqrt{n}$$

$$\alpha = 0,05 \rightarrow 1,36 \sqrt{n} \quad (3.20)$$

$$\alpha = 0,01 \rightarrow 1,63 \sqrt{n}$$

Une valeur négative de V_k indique que la dernière partie de la série chronologique (partie la plus récente) a une moyenne plus grande que la partie plus ancienne et inversement.

La figure 15 nous donne le test de rupture de CUSUM (valeurs absolues de $V(k)$ en fonction du temps exprimé en années).

C'est le rang $k = 23$ qui a donné le $V(k)$ max, correspondant à l'année 2001.

La valeur de $V(2001) = +6 > 0$, signifie que la dernière partie de la série (2002-2015) a une moyenne plus petite que la partie (1979-2000).

Le maximum de la valeur absolue de $V(k)$ est égale à 6, et la valeur de $KS = 0,3243$.

$\text{Alpha} = 0,10$ valeur critique = 1,710

$\text{Alpha} = 0,05$ valeur critique = 1,917 (3.21)

$\text{Alpha} = 0,01$ valeur critique = 2,298

La valeur critique max de $V(k)$ est égale à 6, supérieure à la valeur $1,63\sqrt{n} = 2,298$.

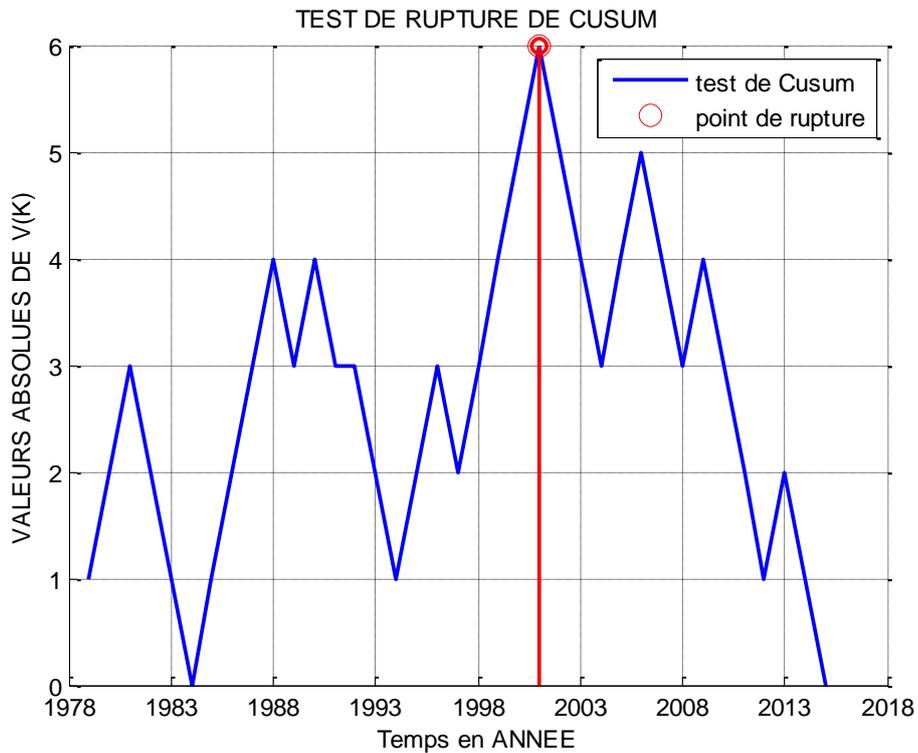


Figure 15 : Test de CUSUM pour la rupture.

Ainsi pour le test CUSUM, l'année 2001 est une année de rupture avec un risque de 1%.

Malgré le résultat du Test de CUSUM sur l'existence de rupture au sein des données pluviométriques, on peut supposer « à priori » la validité du résultat du test de Pettitt sur la stationnarité des données. Vue la durée des données de 1979 à 2015, soit 37 années, on peut dire que les données pluviométriques montre une tendance et une rupture. Le test de CUSUM présente une hausse des valeurs de pluie pour les cinq dernières années : 2011, 2012, 2013, 2014 et 2015 (figure 15).