Chapitre 4

Les cascades multifractales et la longue dépendance

En hydrologie, tout comme en géophysique, en climatologie ou en économie, la planification expérimentale est impraticable en situation réelle, et l'on doit considérer les observations au fur et à mesure qu'elles se présentent. Le statisticien a tendance à vouloir modéliser les séries à l'aide de variables aléatoires indépendantes, car elles permettent une analyse statistique simple. Or les données hydrologiques sont influencées par plusieurs facteurs, dont certains échappent au contrôle et même à la connaissance de l'hydrologue. Travailler avec l'hypothèse que les données sont indépendantes est donc irréaliste ; au contraire, l'étude de la structure de dépendance de ces dernières pourrait restituer ou refléter la structure interne du phénomène pluvieux.

Mais la principale motivation de la modélisation en cascades multifractales est liée à l'invariance d'échelle du paramètre de décroissance algébrique, constatée dans la partie précédente. Les cascades multifractales sont les modèles les plus simples retranscrivant cette propriété. Mais rendent-ils compte d'autres propriétés statistiques constatées en pratique ? La longue dépendance est une propriété statistique non classique susceptible d'être présentée par tout système constitué d'un grand nombre de facteurs en interaction. Elle se caractérise par une décroissance de la fonction de corrélation avec le temps t si faible qu'elle n'en est plus sommable.

De telles propriétés ont été mises en évidence dans des séries financières ou en télécommunication sur des séries de trafic (Lang, 94 [54] ou Resnick, 97 [71]). Plusieurs hydrologues et statisticiens s'intéressant à des domaines de la géophysique ont constaté un comportement de longue dépendance au sein des séries étudiées.

Ces propriétés de longue dépendance ne sont pas décrites par les modèles usuels (ARMA, Markov), et l'élaboration de modèles possédant une telle propriété n'a commencé que dans les années 50, avec la constatation du phénomène de Hurst (Hurst, 51 [46]). Parmi ces modèles, on peut citer le mouvement Brownien fractionnaire ou encore les agrégations de processus AR. Les modèles en cascades multifractales ont fait l'objet d'une telle étude : il a été décelé une décroissance algébrique de la fonction de corrélation (Marsan, 96 [62]). Mais peut-on pour autant en conclure à la présence de longue dépendance ?

Le présent chapitre rappelle la définition mathématique de la longue dépendance, puis passe en revue les principales études des manifestations ce phénomène en hydrologie ces dernières décennies. Nous présenterons par la suite quelques modèles stochastiques rendant compte de la propriété de longue dépendance. Nous analyserons dans quelle mesure les cascades multifractales possèdent la propriété de longue dépendance. Enfin, la longue dépendance de quelques longues séries de débits sera estimée à l'aide d'une méthode statistique performante (le log-périodogramme global), et les résultats seront confrontés aux estimations issues d'outils statistiques plus classiques.

4.1 Notations

Avant de définir la longue dépendance, nous introduisons dans cette section les notations ainsi que les hypothèses choisies. Tout au long de cette partie, on adoptera les notations suivantes : la série d'observations hydrologiques aux instants i (cumul de pluie ou débit de rivière) sera notée $(X_i)_{i=1\cdots n}$. On considérera qu'elle constitue une suite de variables aléatoires identiquement distribuées de fonction de répartition F, et l'on ne s'intéresse qu'au cas où les moments d'ordres 1 et 2 existent :

$$\mu = E\left[X_i\right] < \infty \text{ et } \sigma^2 = Var\left[X_i\right] < \infty$$

Les covariances (resp. corrélations) aux instants i et j entre les observations X_i et X_j seront notées :

$$\gamma(i,j) = E\left[(X_i - \mu)(X_j - \mu)\right] \text{ (resp. } \rho(i,j) = \frac{1}{\sigma^2}\gamma(i,j)\text{)}$$

On se place dans le cadre de séries stationnaires. Sous cette hypothèse, les covariances (resp. corrélations) s'écrivent :

$$\gamma(i,j) = \gamma(|i-j|) \text{ (resp. } \rho(i,j) = \rho(|i-j|)\text{)}$$

4.2 Définitions de la longue dépendance

On trouve dans la littérature diverses définitions de la longue dépendance. On présente ici deux définitions asymptotiques (Taqqu, 00 [79]) qui caractérisent le comportement limite de la fonction de corrélation γ quand la distance tend vers l'infini. Comme nous l'avons énoncé précédemment, la longue dépendance se traduit par une décroissance algébrique de la fonction de corrélation γ vers zéro, et par une vitesse de convergence vers zéro si faible que $\sum_{-\infty}^{+\infty} \gamma(k)$ diverge :

Définition 27 *Soit $(X_i)_i$ une série stationnaire. On dit que $(X_i)_i$ est à longue dépendance si et seulement si :*

$$\sum_{-n}^{+n} \gamma(k) \sim n^{\alpha} L_1(n), \quad n \to \infty, \quad 0 < \alpha < 1$$

avec L_1 une fonction à variations lentes c'est-à-dire bornée sur tout intervalle et telle que :

$$\forall a > 0 \quad \frac{L(ax)}{L(x)} \overset{x\to\infty}{\longrightarrow} 1$$

Mais on trouve dans la littérature une autre définition qui ne lui est pas exactement équivalente :

Définition 28 *On dit que $(X_i)_i$ stationnaire est à longue dépendance si et seulement si :*
$\gamma(k) \sim k^{-\beta} L_2(k), \quad k \to \infty, \quad 0 < \beta < 1$
avec L_2 une fonction à variations lentes.

$(X_i)_i$ est encore dite à longue mémoire, à dépendance à long terme, ou à forte dépendance.

Pour citer des contre-exemples, les modèles couramment utilisés en hydrologie tels que les ARMA (Box et Jenkins, 70 [15]) ou les chaînes de Markov ne sont pas à longue dépendance. Ils sont à courte dépendance car leur fonction de corrélation décroît exponentiellement (en $e^{-\alpha t}$, $\alpha > 0$).

Ces deux définitions sont équivalentes dans le cas d'une fonction de corrélation γ monotone par la proposition suivante (Taqqu, 00 [79]) :

Proposition 29 *Si $\gamma(k)$ est asymptotiquement monotone quand $k \to \infty$, alors les deux définitions sont équivalentes et on a les relations suivantes :*
$\alpha = 1 - \beta, \quad L_1(x) = 2(1-\beta)^{-1} L_2(x)$

Remarquons que lorsqu'on s'intéresse à un comportement de longue dépendance, on ne cherche pas à quantifier les corrélations dans le but de constater si elles sont plus ou moins proches de zéro (par exemple en regardant les $\pm 2/\sqrt{n}-$intervalles de confiance du graphe de $\widehat{\rho}(k)$ vs k), mais plutôt à quantifier la vitesse de convergence de $\widehat{\rho}$ vers zéro quand la distance k tend vers l'infini. En effet, chacune des corrélations peut être arbitrairement petite et appartenir à l'intervalle de confiance. On ne peut donc pas visualiser la longue dépendance sur le corrélogramme. Par contre, comme nous le verrons en section 5.5.1, le graphe de la variance constitue un moyen d'accéder à l'estimation du paramètre de longue dépendance H. Cet outil est fondé sur le théorème suivant :

Proposition 30 *Soit $(X_i)_i$ une série à longue dépendance. On a alors :*

$$\lim_{n\to\infty} \left[\frac{Var\left(\overline{X}\right)}{c_\gamma . n^{-2H-2}} \right] = \frac{1}{H(2H-1)}$$

La variance de la moyenne empirique décroît vers zéro à une vitesse inférieure à n^{-1}. La vitesse de décroissance est proportionnelle à $n^{-\gamma}$, $\gamma = 2H + 2$.

4.3 La longue dépendance des séries hydrologiques

L'étude de la longue dépendance des séries hydrologiques a commencé avec les travaux de H.E. Hurst [46], qui s'intéressa à la validité de l'hypothèse d'indépendance adoptée en pratique pour la modélisation des débits maximum annuels.

S'intéressant à la capacité idéale d'un réservoir, c'est-à-dire à la taille optimale d'un réservoir conçu pour absorber les débits $X_1, X_2, \cdots, X_k$ arrivant entre les instants 1 et k, tout en ayant une vidange constante, un contenu identique aux instants 1 et k, et ne débordant jamais, il définit l'étendue R :

$$\begin{aligned} R(t,k) &= \max_{0\leq i\leq k}\left[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)\right] - \min_{0\leq i\leq k}\left[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)\right] \\ \text{avec } Y_t &= \sum_{i=1}^{t} X_i \end{aligned}$$

et S l'écart type des débits observés :

$$S(t,k) = \left[\frac{1}{k}\sum_{i=t+1}^{t+k}\left(X_i - \overline{X}_{t,k}\right)^2\right] \text{ où } \overline{X}_{t,k} = \frac{1}{k}\sum_{i=t+1}^{t+k} X_i$$

Sur la longue série des débits du Nil (constituée conjointement d'observations systématiques et paléohistoriques et couvrant les années 622 jusqu'à 1281), il constata un comportement linéaire de la courbe $\log - \log$ de l'étendue normalisée R/S par rapport au nombre d'observations k, et définit le coefficient de Hurst H comme la pente de cette dernière :

$$Q_t(k) = \frac{R(t,k)}{S(t,k)} \propto k^H$$

L'estimation du coefficient de Hurst H sur la série du Nil ainsi que d'autres séries de débits conduisit à des valeurs de l'ordre de 0.7. Sur d'autres séries d'observations géophysiques, des estimations de H du même ordre furent trouvées.

Les modèles statistiques classiques ne rendent pas compte de cette caractéristique. En effet, en 1951 Feller [34] montra que dans le cas de séries totalement indépendantes, le coefficient H était de $\frac{1}{2}$. Divers auteurs cherchèrent à trouver des modèles retranscrivant cet effet, dans le cadre des modèles stationnaires (étant donné que cet effet peut aussi être expliqué par une non-stationnarité ou un comportement pré-asymptotique de modèle autorégressif d'ordre 1 (Matalas, 67 [63])). Feller [34] s'orienta vers une structure de dépendance markovienne, mais Barnard en1956 [5] montra qu'une telle dépendance conduisait aussi à une estimation de $\frac{1}{2}$. Mandelbrot ([60] et [61]) fût le premier à proposer un modèle de longue dépendance stationnaire, le mouvement brownien fractionnaire, que nous présentons dans la section qui suit.

4.4 Le mouvement brownien fractionnaire

Le mouvement brownien fractionnaire, proposé par Mandelbrot [61], est un modèle à accroissements stationnaires et à longue dépendance. Nous rappelons succinctement la définition de ce modèle avant d'énoncer deux théorèmes permettant de distinguer les processus à longue dépendance des autres.

Définition 31 $(Y_t)_t$ *est un mouvement brownien* $B(t)$ *si et seulement si :*
1/ Y_t *est une variable aléatoire gaussienne*
2/ $Y_0 = 0$ *p.s*
3/ $(Y_t)_t$ *est à accroissements stationnaires*
4/ $E(Y_t - Y_s) = 0$
5/ $Var(Y_t - Y_s) = \sigma^2 |t - s|$

Définition 32 $(Y_t)_t$ *est un mouvement brownien fractionnaire* $B_H(t)$ *si et seulement si*
1/ Y_t *est une variable aléatoire gaussienne*
2/ $Y_0 = 0$ *p.s*
3/ $(Y_t)_t$ *est à accroissements stationnaires*
4/ $E(Y_t - Y_s) = 0$
5/ $Var(Y_t - Y_s) = \sigma^2 |t - s|^{2H}$

Proposition 33 *Un mouvement brownien fractionnaire* $B_H(t)$ *est à longue dépendance et son paramètre de longue dépendance* β *s'écrit (Taqqu, 00 [79]) :*

$$\beta = 2 - 2H \text{ (ou } \alpha = 2H - 1\text{)}$$

Les deux théorèmes limites suivants (Beran, 94 [10]) permettent de distinguer les processus à longue dépendance (exposant de Hurst supérieur à $\frac{1}{2}$) des processus à courte dépendance ($H = \frac{1}{2}$).

Théorème 34 *Soit* $(X_t)_t$ *tel que* ${X_t}^2$ *soit ergodique et* $t^{-\frac{1}{2}} \sum_{s=1}^{t} X_s \overset{t\to\infty}{\longrightarrow} B(t)$*. Alors,*
$k^{-\frac{1}{2}} Q_t(k) \overset{k\to\infty}{\longrightarrow} \xi$ *avec* ξ *v.a. non dégénérée.*

Les conditions de ce théorème sont vérifiées par la plupart des processus stationnaires, en particulier dans tous les cas où le Théorème de la Limite Centrale est vérifié, puisqu'il existe une variable aléatoire Z telle que :

$$k^{-\frac{1}{2}} Q_t(k) \overset{k\to\infty}{\longrightarrow} Z$$

Théorème 35 *Soit* $(X_t)_t$ *tel que* ${X_t}^2$ *soit ergodique et* $t^{-H} \sum_{s=1}^{t} X_s \overset{t\to\infty}{\longrightarrow} B_H(t)$*. Alors,*
$t^{-H} Q_t(k) \overset{k\to\infty}{\longrightarrow} \xi$ *avec* ξ *v.a. non dégénérée.*

Donc le graphe $(\log k, \log Q_t(k))$ a un comportement asymptotique rectiligne de pente H, exposant de Hurst, qui est égal à $\frac{1}{2}$ pour tout processus stationnaires à courte dépendance et supérieur à $\frac{1}{2}$ pour tout processus stationnaires à longue dépendance.

D'autres caractéristiques statistiques propres aux séries de débits ou d'extrêmes pluviométriques sont aussi retranscrites par le mouvement brownien fractionnaire. En effet, la plupart des séries de débits présentent de longues périodes de basses eaux suivies de longues périodes de hautes eaux. De plus, bien que ces séries semblent stationnaires et ne présentent ni cycle ni tendance, sur de petits intervalles de temps, on distingue de petits cycles et des tendances locales. Le mouvement brownien fractionnaire reproduit ces caractéristiques appelées "Effet Joseph" par Mandelbrot [61].

Plus généralement, vérifient la propriété de longue dépendance (voir *Annexe* **??**).

Parmi les modèles stationnaires de longue dépendance, citons les accroissements stationnaires de processus auto-similaires (d'exposant $H > \frac{1}{2}$), ou encore le modèle d'agrégation de processus auto-régressifs d'ordre 1 (Granger, 80 [40]). Ce dernier modèle possède l'avantage d'être simple à interpréter physiquement et pourrait répondre à l'interrogation de Klemes [52] sur la forme de mécanisme physique à la source de la retransmission de l'influence d'une température moyenne d'une année sur des dizaines ou des centaines d'années.

4.5 Les cascades sont-elles à longue dépendance ?

La pluie, comme tout autre phénomène géophysique, possède une grande variabilité, sur des gammes d'échelles (spatiales et temporelles) très étendues. Comme on l'a vu en *section 2.3.3*, les modèles les plus simples rendant compte de cette caractéristique sont les modèles en cascades multifractales.

L'objectif de cette section est de déterminer la structure de dépendance entre les cumuls désagrégés à une échelle donnée, et de comparer cette dernière à une situation où l'on a indépendance (de type processus ponctuel de Poisson). Les cascades multifractales possèdent-elles toutes la propriété de longue dépendance ? Quelles sont les conditions nécessaires sur les générateurs pour la retranscription de cette propriété ?

4.5.1 Définition de la cascade

On rappelle la définition d'une cascade multifractale : Une cascade multifractale de générateur η est une suite de variables aléatoires $(\mu_{j,k})$ définies par la relation de récurrence (figure 2.12 de la *section 2.3.3*) :

$$\begin{aligned} \mu_0 &= \mu_{0,0} > 0 \\ \text{et } \forall j &\geq 1 \quad \forall k = 0 \cdots 2^j - 1, \quad \mu_{j,k} = \mu_{j-1,\lfloor k/2 \rfloor} . \eta_{j,k} \end{aligned}$$

En désagrégeant jusqu'au niveau d'homogénéité N, on obtient une suite de variables aléatoires $(\mu_{N,k})_{k=0\cdots 2^N-1}$, les cumuls sont de $\mu_{N,0}$ sur $\left[0, \frac{T}{2^N}\right[$, ..., de $\mu_{N,2^N-1}$ sur $\left[(2^n-1)\frac{T}{2^N}, T\right[$.

Il est important de remarquer qu'en hydrologie, on ne s'intéresse pas à une désagrégation "infinie" (j tend vers l'infini) mais plutôt à un niveau de désagrégation limite N nommé niveau d'homogénéité de la cascade.

Hypothèse : Les générateurs $(\eta_{j,k})_{j=1\cdots N, k=0\cdots 2^j-1}$ sont indépendants deux à deux et identiquement distribués :

$$(\eta_{j,k})_{j=1\cdots N, k=0\cdots 2^j-1} \quad iid$$

On note e_1 et e_2 les moments d'ordre 1 et 2 du générateur η :

$$e_1 = E\eta \text{ et } e_2 = E\left(\eta^2\right)$$

4.5.2 Type de dépendance des cumuls

Notions de conservativité des cumuls

Lorsqu'on travaille sur des cumuls de pluie, une contrainte vient s'imposer dans la construction de la cascade : la condition de conservativité. En effet, une cascade de cumuls doit être conservative dans le sens où à chaque niveau j de son développement $j = 1 \cdots N$, la somme des cumuls de ce niveau vaut μ_0 :

$$\forall j = 1 \cdots N \quad \sum_{k=0}^{2^j-1} \mu_{j,k} = \mu_0$$

Sous cette condition, on ne peut supposer l'indépendance des générateurs par ligne. On a recours à une autre notion de conservativité, la conservativité en moyenne, qui s'écrit :

$$\forall j = 1 \cdots N \quad \sum_{k=0}^{2^j-1} E\mu_{j,k} = \mu_0$$

Ce choix est accrédité par le fait que la notion physique de conservativité ne prend son sens que pour des échelles macroscopiques. Ici, c'est la conservativité en moyenne qui nous intéresse car les intensités sont désagrégées à des échelles de plus en plus ténues.

Comme les générateurs d'une même ligne sont identiquement distribués, cette condition revient à :

$$\forall j = 1 \cdots N \quad 2^j.E\mu_{j,*} = \mu_0$$

Or, par itérations, le premier cumul désagrégé j fois $\mu_{j,0}$ s'écrit : $\mu_{j,0} = \mu_0 . \prod_{l=1}^{j} \eta_{l,0}$. La condition de conservativité devient, par indépendance entre les niveaux :

$$\forall j = 1 \cdots N \quad \prod_{l=1}^{j} e_1 = 2^{-j}$$

et revient à une condition sur le moment d'ordre 1 des générateurs :

$$e_1 = \frac{1}{2}$$

Remarque 36 *Si l'on impose aussi une conservativité sur les moments d'ordre 2 des cumuls :*

$$\forall j = 1 \cdots N \quad \sum_{k=0}^{2^j-1} E\mu_{j,k}^2 = \mu_0^2$$

on doit alors avoir : $\forall j = 1 \cdots N \quad \prod_{l=1}^{j} e_2 = 2^{-j}$
ce qui s'écrit : $\forall j = 1 \cdots N \quad e_2 = \frac{1}{2}$

On se contentera d'imposer ici des générateurs de moment d'ordre 2 inférieur à 1.

Variance des cumuls

La variance du premier cumul $\mu_{N,0}$ de la cascade développée à un niveau N s'écrit :

$$Var\left(\mu_{N,0}\right) = \mu_0^2\left[\prod_{j=1}^{N} e_2 - e_1^{2N}\right] \text{ car } \mu_{N,0} = \mu_0 . \prod_{j=1}^{N} \eta_{j,0}$$

et comme les $\left(\mu_{N,k}\right)_k$ sont identiquement distribués :

$$Var\left(\mu_{N,*}\right) = Var\left(\mu_{N,0}\right) = \mu_0^2\left[e_2^N - 2^{-2N}\right]$$

La variance des cumuls tend vers zéro lorsque N tend vers l'infini.

Covariance des cumuls

Soient deux cumuls $\mu_{N,k}$ et $\mu_{N,0}$ d'un même niveau de développement de la cascade N. Le nombre d'ancêtres différents $q\left(k\right)$ entre $\mu_{N,k}$ et $\mu_{N,0}$ s'écrit :

$$q\left(k\right) = \left\lfloor \frac{\log k}{\log 2} \right\rfloor$$

et le nombre d'ancêtres en commun entre $\mu_{N,k}$ et $\mu_{N,0}$ est $N - q\left(k\right) - 1$. On a :

$$\begin{aligned}
\mu_{N,0} &= \mu_{N-1,0}.\eta_{N,0} = \cdots = \mu_0 . \prod_{j=1}^{N} \eta_{j,0} \\
\mu_{N,k} &= \mu_{N-1,k}.\eta_{N,k} = \cdots = \mu_0 . \left(\prod_{j=1}^{N-q(k)-1} \eta_{j,0}\right)\left(\prod_{j=N-q}^{N} \eta_{j,*}\right) \text{ avec } * \neq 0
\end{aligned}$$

La covariance entre ces deux observations espacées de k s'écrit donc :

$$\gamma_{N0}\left(k\right) = Cov\left(\mu_{N,0}, \mu_{N,k}\right) = \mu_0^2 . Cov\left(XY_1, XY_2\right)$$

avec :

$$\begin{aligned}
X &= \prod_{j=1}^{N-q(k)-1} \eta_{j,0} \\
Y_1 &= \prod_{j=N-q(k)}^{N} \eta_{j,0} \\
Y_2 &= \prod_{j=N-q(k)}^{N} \eta_{j,*} \text{ avec } * \neq 0
\end{aligned}$$

Comme on est dans le cas d'indépendance des générateurs entre les niveaux j,

$$Cov\left(XY_1, XY_2\right) = VarX.EY_1.EY_2$$

et comme :

$$\begin{aligned} EY_1 &= EY_2 = 2^{-q(k)-1} \text{ (car } E\eta_{j,*} = e_1 = \frac{1}{2}) \\ VarX &= \prod_{j=1}^{N-q-1} e_{j,2} - \left(\prod_{j=1}^{N-q-1} e_1\right)^2 = e_2^{N-q(k)-1} - 2^{-2N+2q(k)+2} \end{aligned}$$

on obtient la covariance entre les deux observations $\mu_{N,0}$ et $\mu_{N,k}$:

$$\gamma_{N0}\left(k\right) = \mu_0^2 \left[2^{-2q(k)-2}.e_2^{N-q(k)-1} - 2^{-2N}\right]$$

Par stationnarité des cumuls du niveau N $\left(\mu_{N,j}\right)_{j=1\cdots 2^N-1}$, cette relation est vérifiée pour tout couple d'observations espacées de k observations ($\mu_{N,l}$ et $\mu_{N,l+k}$ avec $l = 0 \cdots 2^N - 1 - k$). A N fixé, la décroissance de γ avec k est donc en $\left(2^2 e_2\right)^{-q(k)}$ soit en $k^{-2-\log_2(e_2)}$. La décroissance de la covariance entre les intensités est donc algébrique de paramètre :

$$\log_2\left(e_2\right) + 2$$

Cette fonction de covariance étant monotone, les deux définitions de la longue dépendance citées en *section 4.2* sont équivalentes. On obtient donc le résultat suivant : une cascade multifractale conservative en moyenne est à longue dépendance si le moment d'ordre 2 du générateur vérifie :

$$0 < \log_2\left(e_2\right) + 2 < 1$$

c'est-à-dire si :

$$\frac{1}{4} < e_2 < \frac{1}{2}$$

son paramètre de longue dépendance s'écrit alors :

$$\beta = \log_2\left(e_2\right) + 2 \text{ (ou bien } \alpha = -1 - \log_2\left(e_2\right) \quad H = -\frac{\log_2\left(e_2\right)}{2})$$

4.5.3 Cascades particulières

Cascades log-Gamma

Nous présentons dans cette section une cascade particulière : celle dont les générateurs η suivent une loi log-Gamma $\log\Gamma\left(m, \alpha\right)$ ($\alpha > 1$ et $m \geq 1$) dont la densité est définie par :

$$f_{m,\alpha}(x) = \frac{\alpha^m}{\Gamma\left(m\right)}\left(\log x\right)^{m-1} x^{-\alpha-1} \qquad x \geq 1$$

Remarquons que :

- C'est une loi de type algébrique de paramètre de décroissance α.
- Elle est reliée à la loi Gamma par une simple transformation log :
  Si $X \sim \Gamma(m, \alpha)$ alors $\eta = e^X \sim \log\Gamma(m, \alpha)$

Le calcul de ses moments est simple :

$$\eta \sim \Gamma(m, \alpha) \qquad E(\eta^q) = \left(\frac{\alpha}{\alpha - q}\right)^m \qquad \forall 0 < q < \alpha - 1$$

On choisit les générateurs de la cascade comme suit : Les $(\eta_{j,k})_{j,k}$ sont des variables aléatoires *iid* $\Gamma(m, \alpha)$:

$$(\eta_{j,k})_{j,k} \overset{iid}{\sim} \log\Gamma(m, \alpha)$$

On veut une cascade de cumuls conservative en moyenne, donc le moment d'odre 1 de la loi des générateurs doit être égal à $\frac{1}{2}$:

$$\left(\frac{\alpha}{\alpha - 1}\right)^m = \frac{1}{2}$$

condition qui revient à une relation entre les paramètres α et m de la loi des générateurs :

$$\frac{1}{m} = -\log_2\left(\frac{\alpha}{\alpha - 1}\right)$$

Le moment d'ordre 2 se calcule aisément. Pour que la cascade soit conservative en moyenne, il doit être inférieur à $\frac{1}{2}$. Or, sous la condition précédente :

$$e_2 = \left(\frac{\alpha}{\alpha - 2}\right)^m = \frac{1}{2} \cdot \left(\frac{\alpha - 1}{\alpha - 2}\right)^m > \frac{1}{2}$$

On n'a donc pas de longue dépendance dans une cascade conservative log-Gamma.

Cascade log-Normale

La cascade log-Normale a été introduite par Kolmogorov, 62 [51] et Obhoukov, 62 [67] en turbulence. La loi des générateurs est définie par :

$$\eta = e^X \text{ avec } X \sim N(\gamma, \sigma^2)$$

La fonction génératrice des moments se calcule facilement :

$$E(\eta^q) = e^{\gamma q + \frac{\sigma^2}{2} \log 2 q^2}$$

En particulier,

$$\begin{aligned} e_1 &= e^{\gamma + \frac{\sigma^2}{2} \log 2} \\ \text{et } e_2 &= e^{2\gamma + \frac{3\sigma^2}{2} \log 2} \end{aligned}$$

Comme la cascade est conservative en moyenne ($e_1 = \frac{1}{2}$) :

$$e_2 = \left(\frac{1}{2}\right)^2 .e^{\frac{\sigma^2}{2}\log 2} = 2^{\frac{\sigma^2}{2}-2}$$

et $e_2 < \frac{1}{2}$ si et seulement si $\sigma^2 < 2$. Une cascade conservative en moyenne log-Normale est donc à longue dépendance si et seulement si son paramètre σ^2 est inférieur à 2. Dans ce cas, son paramètre de longue dépendance s'écrit :

$$\beta = \frac{\sigma^2}{2}$$

Cascade log-Poisson

Cette cascade a été introduite en turbulence par Dubrulle, 94 [28]. La loi des générateurs est définie par :

$$\eta = e^{\gamma + aX_\lambda} \text{ avec } X_\lambda \sim P(\lambda)$$

La fonction génératrice des moments se calcule :

$$E(\eta^q) = e^{\gamma q + \frac{\lambda}{\log 2}(2^{aq}-1)}$$

En particulier,

$$\begin{aligned} e_1 &= e^{\gamma + \frac{\lambda}{\log 2}(2^a - 1)} = \frac{1}{2} \\ \text{et } e_2 &= e^{2\gamma + \frac{\lambda}{\log 2}\left(2^{2a}-1\right)} = \left(\frac{1}{2}\right)^2 .e^{\frac{\lambda}{\log 2}\left(2^{2a}-2^{a+1}+1\right)} \end{aligned}$$

En choisissant a et λ tels que $0 < \lambda < \dfrac{(\log 2)^2}{2^{2a} - 2^{a+1} + 1}$, la cascade log-Poisson est à longue dépendance. Dans ce cas, son paramètre de longue dépendance s'écrit :

$$\beta = \frac{\lambda}{(\log 2)^2}\left(2^{2a} - 2^{a+1} + 1\right)$$

4.6 Application aux données

Dans cette section, on présente deux outils d'exploration et d'estimation de la longue dépendance couramment utilisés en statistique, avant de les appliquer à l'estimation de la longue dépendance de cascades multifractales simulées et de séries de débits.

4.6.1 Outils statistiques

Il existe dans la littérature de nombreux estimateurs de la longue dépendance. En tenant compte des résultats de l'étude comparée de leurs performances sur des séries simulées de Bardet et al. [4], on choisit dans cette section d'employer l'estimateur du log-périodogramme global. Les résultats seront ensuite confrontés aux estimations par l'écart ajusté réduit.

L'écart ajusté réduit : la statistique R/S

On a vu que cet estimateur, introduit par Hurst [46], est à l'origine de la découverte de la notion de longue dépendance (*section 4.3*). Il peut être utilisé pour estimer le paramètre de longue dépendance H ou bien, à cause de ses faibles performances, comme outil exploratoire pour détecter une longue dépendance. On rappelle son expression :

$$Q_t(k) = \frac{R(t,k)}{S(t,k)}$$

avec :

$$R(t,k) = \max_{0\le i\le k}\left[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)\right] - \min_{0\le i\le k}\left[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)\right]$$

et

$$S(t,k) = \left[\frac{1}{k}\sum_{i=t+1}^{t+k}\left(X_i - \overline{X}_{t,k}\right)^2\right] \text{ où } \overline{X}_{t,k} = \frac{1}{k}\sum_{i=t+1}^{t+k} X_i$$

La pente asymptotique du graphe $\log-\log$ de $(k, Q_t(k))$ constitue une estimation de l'exposant H de Hurst, directement relié, comme on l'a vu en *section 4.3*, à la détection de longue dépendance.

Mais lors de la mise en oeuvre de la méthode, deux problèmes essentiels apparaissent : quand le comportement asymptotique apparaît-il, et quelle est la qualité de l'estimateur $\widehat{H}$? Concernant cette dernière question, l'étude sur simulations de Lang [54] conduit à considérer cet estimateur comme peu fiable, car la variance d'estimation décroît peu avec la taille de l'échantillon.

Le log-périodogramme global

Cette méthode d'estimation a été proposée par Moulines et Soulier 98 [66]. Elle est basée sur le périodogramme de la série d'observations $(X_1, \cdots, X_N)$:

$$I_N(\lambda) = \frac{1}{2\pi N}\sum_{i=1}^{N}\left|X_j e^{-ij\lambda}\right|^2$$

où λ est la fréquence. Une série à longue dépendance possède une densité spectrale proportionnelle à $|\lambda|^{1-2H}$ au voisinage de l'origine. Comme $I(\lambda)$ est un estimateur de la densité spectrale g, que l'on écrit dans le cas stationnaire sous la forme semi-paramétrique :

$$g(\lambda) = g_*(\lambda)\,.\left|1 - e^{-i\lambda}\right|^{1-2H}$$

un estimateur de H découle d'une régression sur le graphe $\log-\log$ du périodogramme. Dans la méthode du $\log-$périodogramme, la régression est effectuée sur les fréquences de Fourier λ_j, où le périodogramme est calculable à partir des données observées :

$$\log I_N(\lambda_j) = \log g_*(0) + (1-2H)\log\left|1 - e^{-i\lambda_j}\right| + \log\left(\frac{g_*(\lambda_j)}{g_*(0)}\right) + \log\left(\frac{I_N(\lambda_j)}{g(\lambda_j)}\right)$$

La méthode du log $-$périodogramme global consiste à prendre en compte toutes les fréquences (et pas uniquement celles proches de zéro, comme dans la méthode du log $-$périodogramme local, où l'on veut négliger la part mémoire courte). Il est alors nécessaire d'estimer la part "mémoire courte" $\log(g_*)$. Cette dernière fonction est projetée sur la base des $h_j(\lambda_j) = \frac{\cos(j\lambda_j)}{\sqrt{\pi}}$ dont on ne retient que m premières composantes :

$$\log(g_*) = \sum_{j=0}^{m} \theta_j h_j$$

Dans la méthode FEXP, on choisit la base des $h_j(\lambda_j) = \cos\left(\frac{j\lambda_j}{\sqrt{\pi}}\right)$. Pour estimer H, on réalise ensuite une régression linéaire ordinaire :

$$Y_{N,k} = \left(1 - e^{-ix_k}\right)\left(1 - 2\widehat{H}\right) + \sum_{j=0}^{m} \widehat{\theta}_j h_j(x_k) + \varepsilon_{N,k}$$

avec $x_k = (2k+1)\pi/N$.

Contrairement à l'estimateur du rang ajusté, l'estimateur du log-périodogramme global conduit à des estimations fiables, le paramètre m étant calculé par un algorithme adaptatif. Pour plus de précisions sur la qualité de cet estimateur (et d'autres estimateurs de la longue dépendance), on pourra se reporter à l'ouvrage de Bardet et al. [4].

4.6.2 Séries simulées

Dans ce paragraphe, on estime la longue dépendance de séries de cascades multifractales simulées par le log-périodogramme global[1]. On simule 10 8192-échantillons de cascades log-Normale, log-Poisson et log-Gamma. Les résultats de l'estimation de la longue dépendance sont d'autant meilleurs que la longue dépendance est faible (cascade log-Gamma). Mais, sur les 10 cascades log-Normales simulées, l'écart à la valeur théorique $H - 0.5$ atteint en moyenne 0.06 (tableau 4.1). La taille de l'échantillon simulé (8192) est peut être trop faible pour interpréter les résultats du log périodogramme global. Dans le paragraphe suivant, les tailles de séries de débits varieront entre 5 840 et 28 470.

Série	$H-0.5$	$\overline{GLP}$	nb (*)	var	longueur
log-Normale$(\mu = 1, \sigma = 1)$	0.25	0.19	10	0.025	8192
log-Poisson$\left(\gamma = 0, a = 2, \lambda = \frac{(\log 2)^3}{9}\right)$	0.15	0.17	10	0.025	8192
log-Gamma$(n = 5, \alpha = 1)$	0	0.01	10	1.32*10^{-4}	8192

Tableau 4.1: Estimation par le log-périodogramme global sur des cascades simulées (* : nb=nombre d'échantillons simulés).

[1]Je remercie E. Moulines et J.M. Bardet qui ont fournit un programme Matlab de calcul de l'estimateur du log-périodogramme global (programme qui sera prochainement mis à disposition sur internet).

4.6.3 Séries de débits

L'estimation de la longue dépendance sur des séries pluviométriques est difficile (Lang, 94 [54]), car elles s'écartent du cadre de définition des estimateurs (non-normalité mais surtout intermittence des données). Cependant, si un comportement de longue dépendance de séries de débits peut provenir d'un comportement de longue dépendance des séries de précipitations, il peut aussi résulter d'un échange avec un gros réservoir régulateur. Il est donc intéressant d'étudier si les séries de débit possèdent la propriété de longue dépendance et de tenter de relier cette dernière aux caractéristiques physiques des bassins versants.

Dans cette section, on examine la forme de dépendance des longues séries de débits moyens journaliers de cours d'eau[2] déjà étudiées en section 3.5.5 (voir tableau 3.6). Parmi ces séries, quelques unes présentent des caractéristiques hydrogéologiques capables d'expliquer une éventuelle mémoire à long terme, telles que les rivières alimentées par une nappe dans la craie. Ces séries ont été étudiées par Lang [54] qui a estimé la longue dépendance par diverses méthodes statistiques, notamment celle de l'estimateur de l'écart ajusté réduit ; ces résultats seront donc comparés aux estimations de la présente étude.

Avant d'estimer la longue dépendance, il est nécessaire de désaisonnaliser ces séries (car la saisonnalité a un effet sur l'estimation). Cette opération est effectuée de la façon la plus simple : en soustrayant à chaque observation à la date t la moyenne de toutes les observations, à cette même date, sur toutes les années.

Les résultats du tableau 4.2 font clairement apparaître une longue dépendance des débits de rivières Authie, Herbissonne, Superbe, Suippe et Soude (puisque les estimations par le log-périodogramme global sont situées autour de 0.7). Lang [54] a constaté que ces rivières étaient des rivières sur craie, ce qui constitue une explication physique de la longue dépendance. De même, un groupe de grandes rivières apparaît : la Dordogne, la Drôme, la Garonne, la Seine, la Vienne et la Zorn présentent aussi une longue dépendance (estimations autour de 0.3). Ceci peut s'expliquer physiquement par une sorte d'effet réservoir. Les rivières de haute montagne, quant à elles, ne sont pas à longue dépendance (Averole, Isère et Ubaye). Contrairement aux estimations de l'écart ajusté réduit (Lang [54]), la proximité géographique des rivières est décrite par l'estimation du log-périodogramme global. En effet, les estimations des séries de l'Isère et de l'Averole sont proches, et il en est de même pour les deux séries de Corrèze et de Dordogne.

4.7 Conclusions

L'étude des longues séries de débits journaliers fait apparaître un résultat intéressant : Les rivières à gros volume et les rivières sur craie possèdent la propriété de longue dépendance. Les rivières de haute montagne, quant à elles, ne présentent pas la propriété de longue dépendance. On aurait pu suspecter cependant que les glaciers en amont soient à l'origine d'un effet réservoir en été, mais la forte saisonnalité de ces séries est peut être à l'origine de la non constatation de cet effet. L'estimateur du log-périodogramme fournit donc des résultats satisfaisant pour l'estimation de la longue dépendance, et ce malgré sa nature semi-paramétrique.

[2]Séries extraites de la base de données hydrologique HYDRO du Ministère de l'Agriculture.

Rivière	GLP ($\widehat{H}$)	Altitude	Longueur (ans)	VM/an(*)
Authie	0.75	12	27	234
Averole	-0.11	1950	16	64
Corrèze	0.15	465	47	177
Corrèze	0.15	101	72	671
Dordogne	0.26	780	61	117
Dordogne	0.25	173	76	3370
Drôme	0.25	237	76	90
Garonne	0.25	17	74	19872
Herbissonne	0.70	91	21	11
Isère	0.15	1831	40	59
Seine	0.25	26	60	8986
Seine	0.29	18	49	12874
Seine	0.32	148	31	2583
Soude	0.54	110	19	19
Suippe	0.70	60	23	23
Superbe	0.50	79	17	17
Ubaye	0.22	1133	78	78
Vienne	0.27	230	70	70
Zorn	0.32	147	67	67

Tableau 4.2: Estimations de la longue dépendance par le log-périodogramme global (* : volume moyen écoulé par an en hm^3).

Les cascades multifractales conservatives en moyenne ne sont pas toutes à longue dépendance bien que leur fonction de covariance décroisse lentement (décroissance algébrique). C'est la valeur du moment d'ordre deux du générateur sous condition de conservativité en moyenne qui permet de déterminer si la cascade est à longue dépendance ou non. La modélisation par cascade multifractale est donc, dans certains cas, adaptée à la modélisation des séries à longue dépendance.