

Exercice n° HA 0801 - Corrigé

Estimation du débit de crue de temps de retour 30 ans par la méthode statistique à partir de 3 séries (annuelle, tronquée et gonflée) – Application au bassin versant bassin versant du Parimbot (FR, Suisse)

Données de l'exercice :

L'exercice porte sur le bassin versant du Parimbot (6.75 km²). Les données nécessaires à la réalisation de cet exercice se trouvent dans le tableau 1 (il s'agit des débits instantanés enregistrés sur plusieurs années et plus grand que 0.35 m³/s). Ces données sont aussi regroupées dans le fichier Excel « HA0801_enonce.xls » ou dans le fichier « HA0801_feuillecalcul.xls » à compléter. Les résultats sont aussi disponibles sur le fichier Excel « HA0801_corrige.xls ».

Question 1 : Construction des séries de maxima annuels, gonflée et tronquée

A partir de la série de débits de l'énoncé il est facile de construire les trois séries de débits demandées. On obtient :

- Une série de 20 débits de pointe annuels. Une valeur de débit de pointe a été supprimée du fait de sa trop grande faiblesse.
- Une série gonflée comprenant 39 valeurs de débit choisis parmi les deux plus grands débits de pointe pour chaque année (« série gonflée »).
- Une série de 48 débits de pointe ayant une valeur supérieure à 6.00 m³/s (« série tronquée »).

Question 2 : Estimation du débit de pointe de temps de retour 30 ans par la méthode statistique – Ajustement de Gumbel

☉ Méthode à appliquer : ajustement statistique d'une série de données – distribution de Gumbel

L'objectif de cet exercice est d'estimer les débits de pointes (débits maximaux) correspondants à un certain temps de retour, c'est-à-dire à une certaine probabilité d'apparition donnée.

L'analyse fréquentielle d'une longue série de débits maximaux permet d'estimer le temps de retour d'une valeur particulière. Cette prédiction repose sur la définition et la mise en œuvre d'un **modèle fréquentiel** qui est une équation décrivant (modélisant) le comportement statistique d'un processus. Ces modèles décrivent la **probabilité d'apparition** d'un événement de valeur donnée. C'est du choix du modèle fréquentiel (et plus particulièrement de son type) que dépendra la validité des résultats de l'analyse fréquentielle.

Un modèle fréquentiel très souvent utilisé pour décrire le comportement statistique des valeurs extrêmes est la distribution statistique de Gumbel (loi double exponentielle ou loi de Gumbel). La fonction de répartition de la loi de Gumbel $F(x)$ s'exprime de la manière suivante :

$$F(x) = \exp\left(-\exp\left(-\frac{x-a}{b}\right)\right) \quad (1) \quad \text{avec la variable réduite suivante : } u = \frac{x-a}{b} \quad (2)$$

où a et b sont les paramètres du modèle de Gumbel.

La distribution s'écrit alors de la manière suivante :

$$F(x) = \exp(-\exp(-u)) \quad (3) \quad \text{et } u = -\ln(-\ln(F(x))). \quad (4)$$

L'avantage d'utiliser la variable réduite est que l'expression d'un quantile est alors linéaire ($x_q = a + bu_q$).

En conséquence, dès lors que les points de la série à ajuster peuvent être reportés dans un système d'axes $x-u$, il est possible d'ajuster une droite qui passe le mieux par ces points et d'en déduire les deux paramètres a et b de la loi. L'estimation des paramètres a et b de l'ajustement peut se faire graphiquement (ajustement à l'œil ou à l'aide d'une régression statistique), ou selon une méthode mathématique comme celle des moments (cf. ci-dessous).

En pratique il s'agit essentiellement d'estimer la probabilité de non dépassement $F(x_i)$ qu'il convient d'attribuer à chaque valeur x_i . Il existe de nombreuses formules d'estimation de la fonction de répartition $\hat{F}(x)$ à l'aide de la fréquence empirique. Elles reposent toutes sur un tri de la série par valeurs croissantes permettant d'associer à chaque valeur son rang r . Des simulations ont montré que pour la loi de Gumbel, il faut utiliser la fréquence empirique de Hazen :

$$F(x_{[r]}) = \frac{r-0.5}{n} \quad (5)$$

où r est le rang dans la série de données classée par valeurs croissantes, n est la taille de l'échantillon, $x_{[r]}$ la valeur de rang r .

Rappelons encore que le temps de retour T d'un événement est défini comme étant l'inverse de la fréquence d'apparition de l'événement. Soit :

$$T = \frac{1}{1 - F_Q(x_Q)} \quad (6)$$

A l'aide de l'ajustement, il est alors possible d'estimer le débit de pointe pour un temps de retour donné.

☉ Méthode à appliquer : Méthode des moments

La méthode des moments consiste à évaluer les moments des échantillons avec les moments théoriques de la loi¹. Par la méthode des moments les paramètres a et b sont calculés d'après les formules :

$$\begin{cases} \hat{b} = \frac{\sqrt{6}}{\pi} \hat{\sigma} \\ \hat{a} = \hat{\mu} - \hat{b}\gamma \end{cases} \quad \text{avec } \gamma = 0.5772 \text{ (constante d'Euler)}. \quad (7)$$

avec

σ : écart-type des valeurs composant l'échantillon.

μ : moyenne de l'échantillon.

Dès lors il est possible d'estimer les débits dont la représentation graphique est une droite d'équation :

$$\hat{Q} = \hat{a} + \hat{b} \cdot u \quad (8)$$

avec :

u : variable réduite (cf. équation (4)).

¹ Autrement dit : la méthode des moments consiste à évaluer les caractéristiques de la distribution (empirique) des échantillons avec les caractéristiques théoriques de la loi. Les caractéristiques utilisées pour décrire une distribution sont les moments dont les plus connus sont la moyenne et la variance (la moyenne d'une distribution est appelée premier moment, la variance est le deuxième moment). Les moments d'ordre plus élevés sont utiles pour caractériser d'autres aspects de la distribution. Le troisième moment est par exemple lié à l'asymétrie ou la dyssymétrie.

⊙ Démarche et résultats

Etape 1 : Préparation de la série de données des débits de pointe.

- Trier les valeurs dans l'ordre croissant.
- Attribuer un rang à chaque valeur.

Etape 2 : Calcul de la fréquence empirique pour chaque rang (Hazen, équation (5)).

Etape 3 : Calcul de la variable réduite « u » du Gumbel (équation (4)).

Etape 4 : Représentation graphique des couples (u_i, x_i) de la série à ajuster (figure 1).

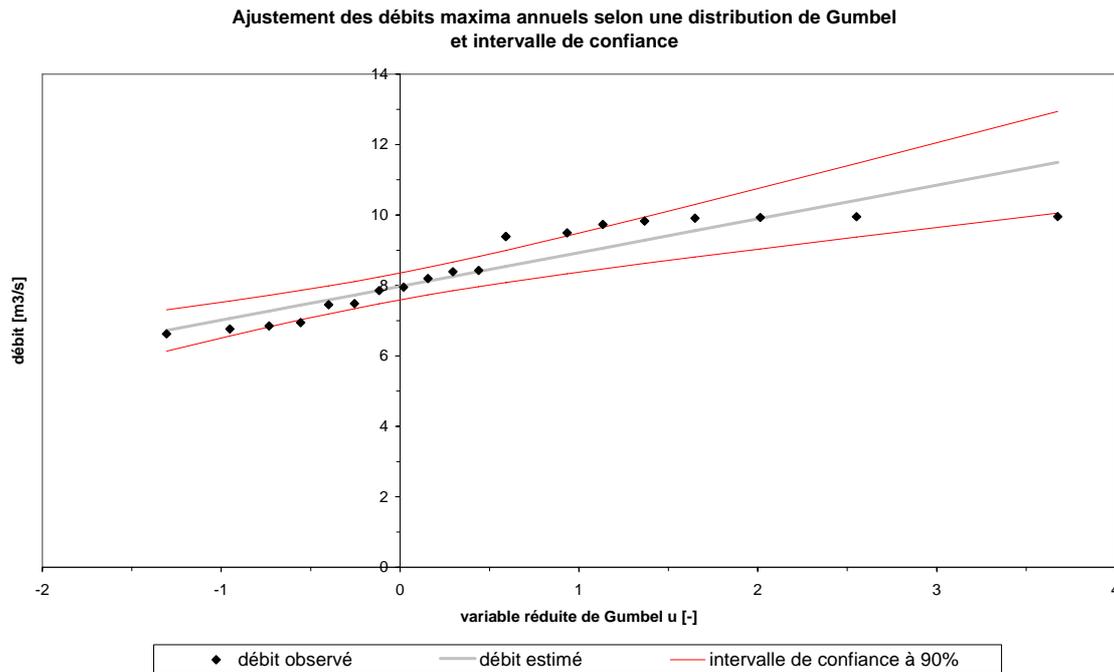


Figure 1. Ajustement graphique du modèle (calcul des paramètres « a » et « b » de la droite d'ajustement de Gumbel par la méthode des moments)

Etape 5 : Ajustement d'une relation linéaire de type $x_q = a + bu_q$ aux couples (u_i, x_i) (figure 1).

Avec un ajustement par la méthode des moments, on a alors pour ce premier ajustement une estimation des paramètres a_1 et b_1 :

$$a_1 = 167 \text{ et } b_1 = 38.2$$

Vérifier de manière visuelle l'ajustement. On peut remarquer que l'ajustement est médiocre pour les grandes valeurs de u . A ce stade, il serait nécessaire de vérifier statistiquement que les valeurs observées sont estimées « de manière satisfaisante » à l'aide de tests d'ajustement appropriés.

Etape 6 : Calcul de l'intervalle de confiance à l'aide de la formulation de l'énoncé (voir figure).

Etape 7 : Utilisation du modèle statistique pour estimer des débits de pointe de différents temps de retour T . Pour $T=30$ ans, on suit les étapes suivantes :

- Calcul de la fréquence de non-dépassement d'après la relation (6) :

$$F(Q_p(T)) = 1 - \frac{1}{T} = 1 - \frac{1}{30} = 0,97$$

- Calcul de la variable réduite de Gumbel correspondante d'après la relation (4) :

$$u = -\ln\left(-\ln\left(F(Q_p(T))\right)\right) = -\ln\left(-\ln(0,97)\right) = 3,38$$

- Calcul du quantile correspondant d'après la relation linéaire (avec a et b fournis par l'étape 5 précédente) :

$$Q_p(30) = a_1 + b_1 \cdot u_{30} = 8 + 1 \cdot 0.97 = 11.2 \quad \text{m}^3/\text{s}$$

Question 3 : Application aux séries tronquée et gonflée

⊙ Méthode à appliquer : : ajustement statistique d'une série de données et corrections fréquentielles (voir rappels-énoncé).

Lorsque la série à ajuster est gonflée ou tronquée, la démarche décrites ci-dessus peut être appliquée mais en apportant des corrections aux formulations des probabilités cumulées de non-dépassement et aux paramètres de la droite d'ajustement (voir rappels-énoncé).

⊙ Résultats

Les débits de pointe de temps de retour 30 ans ainsi que les paramètres a et b pour les ajustements précédents se trouvent dans le tableau 1. Les résultats sont représentés dans la figure 2. On peut constater que les valeurs de débit sont peu différentes.

Tableau 1 : Débits de pointe de temps de retour 30 ans et paramètres a et b des ajustements des trois séries de débit.

	Série maxima annuels	Série gonflée	Série tronquée
a	8	8,1	8
b	1	0,9	1,2
Q_p [m ³ /s]	11,2	11,3	11,9

Comparaison des ajustement des débits

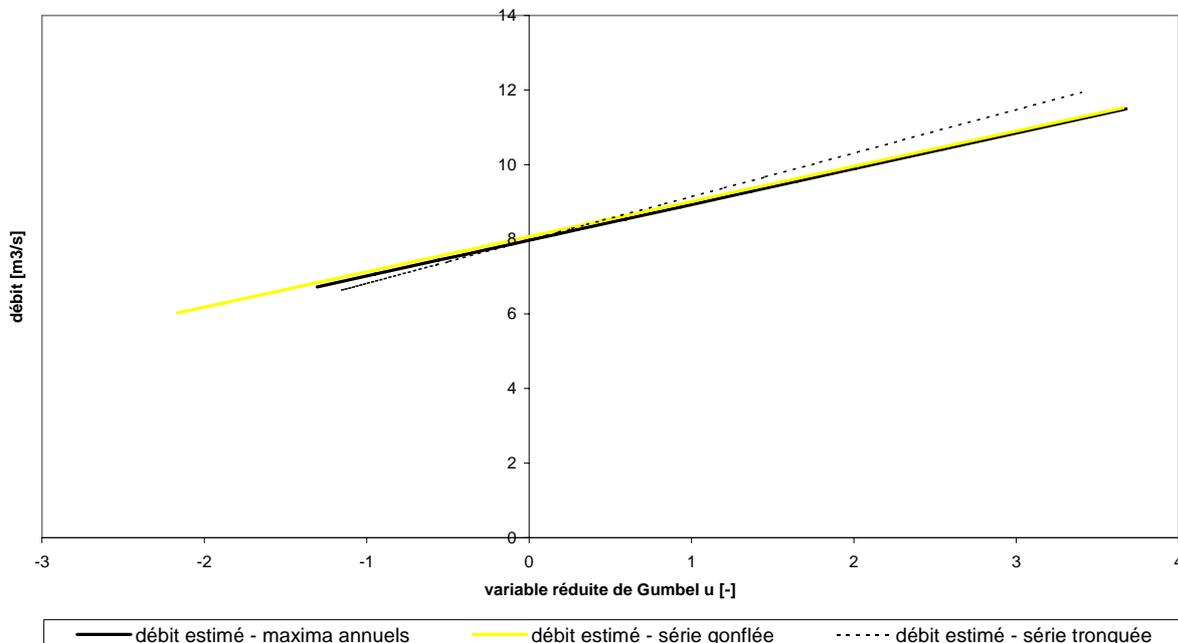


Figure 2: Comparaison des ajustements à l'aide de la série des maxima annuels, de la série gonflée et tronquée