

Exercice no HA 0810 - Corrigé

Régionalisation des débits de pointe pour une région hydrologique relativement homogène

Données de l'exercice

Le fichier Excel « HA0810_enonce.xls » donne les principales caractéristiques hydrologiques et physiographiques des quatre groupes de bassins versants (G0, G1, G2 et G3).

L'exercice porte sur trois bassins versants :

- La Mentue à Yvonnand (surface : 105 km², altitude moyenne : 679 m).
- Le Lyssbach (surface : 22.5 km², altitude moyenne : 616 m).
- Un bassin versant hypothétique non jaugé (surface : 100 km², altitude moyenne : 700 m).

Le fichier « HA0810_enonce.xls » donne les séries de crues annuelles des deux premiers bassins. Le fichier « L-moments-théoriques » donne quelques informations sur les lois théoriques (Gumbel, GEV, lognormale, etc.). Il existe aussi dans le fichier Excel « HA0810_feuillecalcul.xls » des feuilles de calcul compléter. Les résultats sont disponibles sur le fichier Excel « HA0810_corrige.xls ».

Question 1 : Méthode de « Index Flood »

Rappelons que trois étapes sont nécessaires pour aboutir à une estimation des débits de pointe selon la méthode index flood:

- Délimitation de régions (sur la base de critères géographiques, physiographiques, climatiques, etc.) et vérification de leur homogénéité en vérifiant que certaines statistiques des débits de pointe annuels observés (coefficients de variation et d'asymétrie) peuvent être considérées comme les mêmes pour tous les bassins de la région.
- Pour chacune de ces régions, ajustement d'une loi statistique à la distribution normalisée des débits de pointe annuels (i.e. débit de pointe divisé par le débit de pointe annuel moyen). Cette distribution normalisée, représentative de la région est aussi appelée "courbe de croissance".
- Enfin, détermination de l'expression permettant d'obtenir pour chacun des sites le débit de pointe annuel moyen (facteur d'échelle) à partir de certaines caractéristiques physiographiques ou climatiques des bassins versants. Cette expression peut être obtenue notamment par régression simple ou multiple.

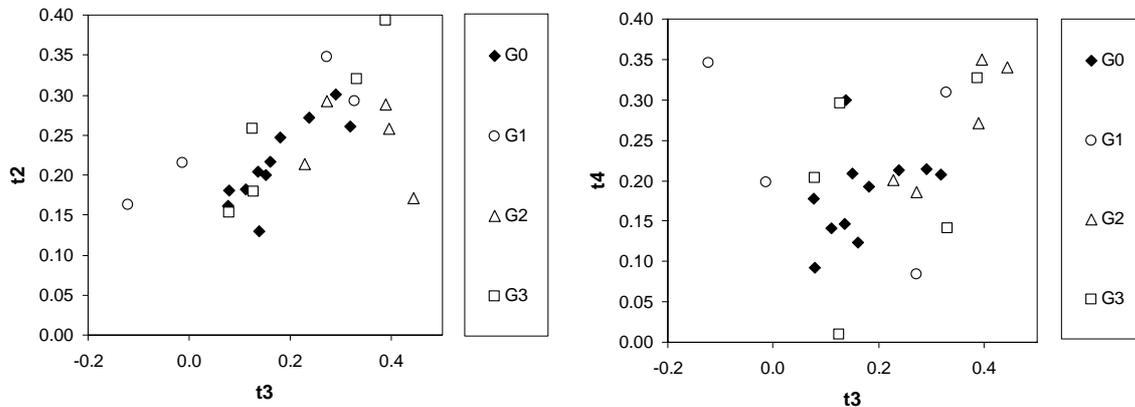
La distribution des débits de pointe recherchée au site i est alors le produit du débit de pointe annuel moyen du site i et de la courbe de croissance de la région correspondante.

Question 2 : Evaluation de l'homogénéité et choix d'une loi théorique sur la base du diagramme des L-moments.

Les diagrammes des L-moments (ou plus exactement les diagrammes des L-coefficients de variation t_2 , d'asymétrie t_3 et d'aplatissement t_4) sont similaires aux diagrammes des moments classiques, mais ont l'avantage d'être moins sensibles aux fluctuations d'échantillonnage. Ils permettent d'évaluer le degré d'homogénéité d'une région et de choisir une distribution théorique pour construire une courbe de croissance.

On admettra que la région formée par les bassins du groupe G0 est homogène. Autrement dit, si ces bassins disposaient chacun d'une série de mesures infiniment longue, ils seraient tous représentés par le même point dans ces diagrammes. La dispersion des points observée est donc attribuée aux fluctuations d'échantillonnage.

Les diagrammes des L-moments construits pour les groupes G0, G1, G2 et G3 sont présentés ci-dessous.



Sur la base de ces diagrammes, on peut faire quelques remarques :

a) Groupe G1 :

La dispersion des points du groupe G1 (bassins disposant de moins de 10 ans de mesure) dans les deux diagrammes ci-dessus est plus grande que celle du groupe G0.

Les causes principales de cette dispersion sont les fluctuations d'échantillonnage importantes dues à des séries de mesures très courtes sur les bassins du groupe G1.

L'acquisition de nouvelles données aux stations du groupe G1 aura pour effet de réduire cette dispersion. On ne peut donc pas conclure, dans ce cas, que la dispersion est due à des réponses hydrologiques particulières, et donc à un phénomène d'hétérogénéité du groupe G1. De même, il n'est pas possible d'affirmer que l'incorporation des bassins du groupe G1 dans le groupe G0 augmente l'hétérogénéité de ce dernier.

b) Groupe G2 :

Pour les bassins du groupe G2 (bassins préalpins) on observe un décalage du nuage de points vers des valeurs de t_3 et t_4 plus grandes (les valeurs de t_2 restant sensiblement identiques). Contrairement à celles du groupe G1, les séries de mesures du groupe G2 ne sont pas significativement plus courtes que celle du groupe G0. L'incorporation des bassins du groupe G2 au groupe G0 a donc pour effet d'augmenter l'hétérogénéité (augmentation de la taille du nuage de points non imputable aux fluctuations d'échantillonnage).

Cependant les valeurs du groupe G2 sont peu dispersées (de même ordre de grandeur que celles du groupe G0). Compte tenu des données à disposition, on peut s'attendre à ce que les bassins du groupe G2 appartiennent à une autre région homogène (une vérification formelle de l'homogénéité peut se faire au moyen de tests statistiques comme ceux de Chowdhury ou de Hosking et Wallis).

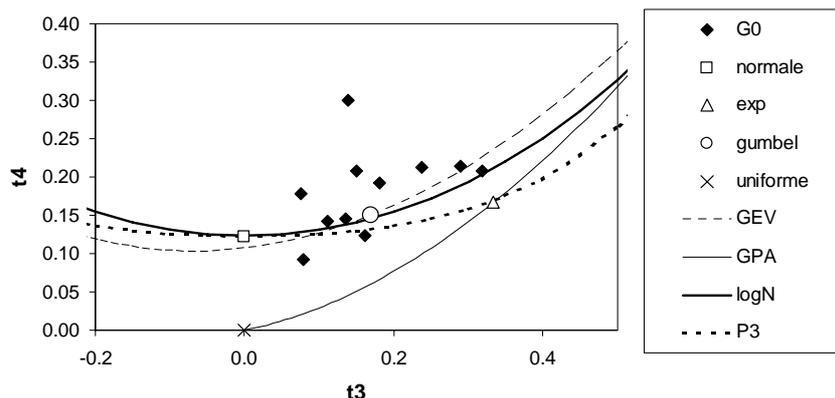
Les différences au niveau des L-coefficients entre le groupe G0 et G2 sont le reflet des différences au niveau des réactions hydrologiques entre les bassins versants du Plateau et ceux des Préalpes. En effet, les bassins préalpins sont caractérisés par : (1) la quasi-imperméabilité des surfaces, (2) des pentes élevées (diminution du temps de concentration) et (3) des températures plus basses que sur le plateau (précipitations plus souvent sous forme de neige en hiver). Les crues des bassins préalpins se produisent ainsi généralement en été à la suite de précipitations convectives. En revanche, sur le plateau, les crues annuelles sont essentiellement provoquées par des précipitations hivernales de type frontal.

c) Groupe G3 :

Enfin, pour les bassins du groupe G3 (petits bassins versants du plateau) on observe que les points sont plus dispersés que ceux du groupe G0, sans pour autant que l'on puisse y observer un décalage dans l'un ou l'autre des axes t_2 , t_3 ou t_4 (contrairement aux bassins du groupe G2). Par ailleurs la dispersion plus élevée n'est pas imputable à des fluctuations d'échantillonnage importantes (les séries de mesures du groupe G3 ne sont pas significativement plus courtes que celle du groupe G0). Les bassins du groupe G3 ne constituent donc apparemment pas une région homogène et, l'inclusion de ces bassins dans le groupe G0 va augmenter l'hétérogénéité de ce dernier. Ceci s'explique par le fait que la variabilité des caractéristiques (et par conséquent des réponses hydrologiques) des petits bassins (moins de 10 km²) est en général beaucoup plus grande que pour les grands bassins versants.

d) Lois théoriques pour le groupe G0 :

La figure ci-dessous montre le diagramme des L-coefficients d'asymétrie (t_3) et d'aplatissement (t_4) pour le groupe G0. On y a également représenté les valeurs théoriques de certaines lois de distributions utilisées pour l'analyse fréquentielle (Gumbel, GEV, Exponentielle, etc.). On y observe que les distributions de Gumbel (à deux paramètres) et GEV (à trois paramètres) constituent de bons candidats pour modéliser la courbe de croissance des bassins du groupe G0. Le choix entre la loi de Gumbel et la loi GEV revient à décider si l'on fixe le paramètre k à 0 ou non. Une vérification formelle de l'adéquation peut se faire au moyen de tests statistiques comme ceux de Chowdhury ou de Hosking et Wallis ou par le test de Kolmogorov-Smirnov (présenté par exemple dans [Meylan et Musy, 1998]).



Question 3. Construction d'une courbe de croissance

Pour construire la courbe de croissance des bassins du groupe G0, une des approches possibles est de calculer des valeurs centrales des L-coefficients de variation et de déterminer les paramètres de la loi théorique retenue (cf. cours).

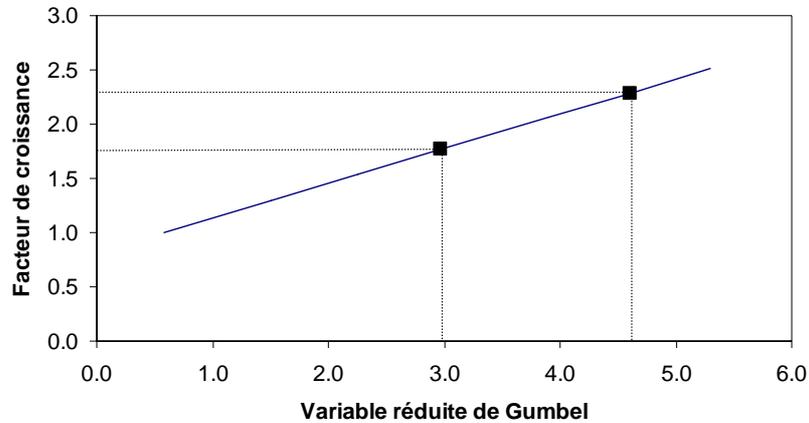
Pour calculer ces valeurs centrales et afin de tenir compte de l'effet de l'incertitude associé au nombre d'années de mesures à disposition, la moyenne pondérée (où les poids sont inversement proportionnels à la longueur des séries observées) est préférable à la moyenne arithmétique simple. On a ainsi pour le groupe G0:

	t2	t3	t4
Valeurs pondérées =	0.221	0.171	0.174

On obtient les paramètres suivants :

- Pour une loi de Gumbel : $k = 0$, $b = t_2 / \ln(2) = 0,319$, et $a = 1 - 0,5772 \cdot b = 0.815$
- Pour une loi de distribution GEV les paramètres estimés sont très peu différents d'une distribution de Gumbel qui est donc tout à fait appropriée dans ce cas.

La figure ci-dessous donne la courbe de croissance. Cette courbe permet de lire le facteur de croissance, qui est le rapport entre le débit de pointe de temps de retour T et le débit de pointe annuel moyen. Pour un temps de retour 100 ans, il est de 2.28. Autrement dit, il faut multiplier le débit de pointe annuel moyen par 2.28 pour obtenir le débit centennal. Lorsque T=20 ans, le facteur de croissance est de 1.76.



Question 4. Modélisation du débit de pointe annuel des bassins du groupe G0

Pour modéliser le débit de pointe annuel (Q_p), on doit disposer d'un jeu de débits mesurés pour un certain nombre de bassins versants jaugés ainsi qu'un certain nombre de variables physiographiques et/ou climatiques pour ces mêmes bassins. Dans notre cas, on dispose de la surface (A) et de l'altitude moyenne des bassins versants (ALT). On testera ici seulement **l'effet de la surface**.

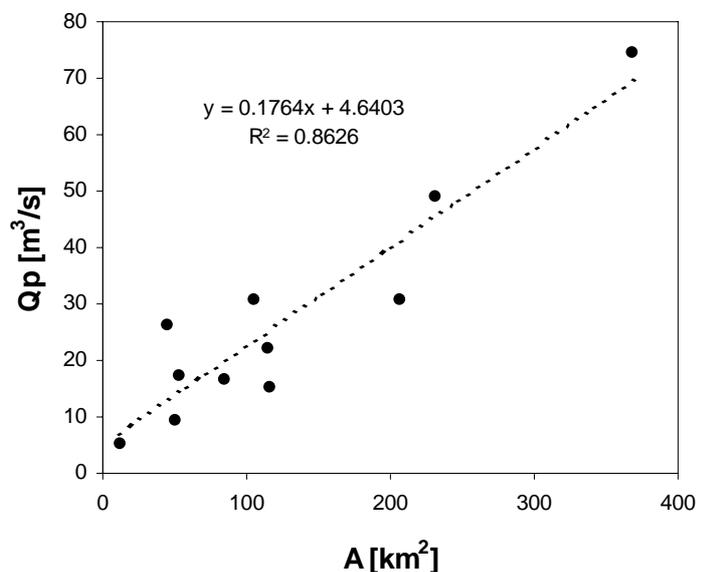
Le modèle testé est donc :

$$Q_p = f(A).$$

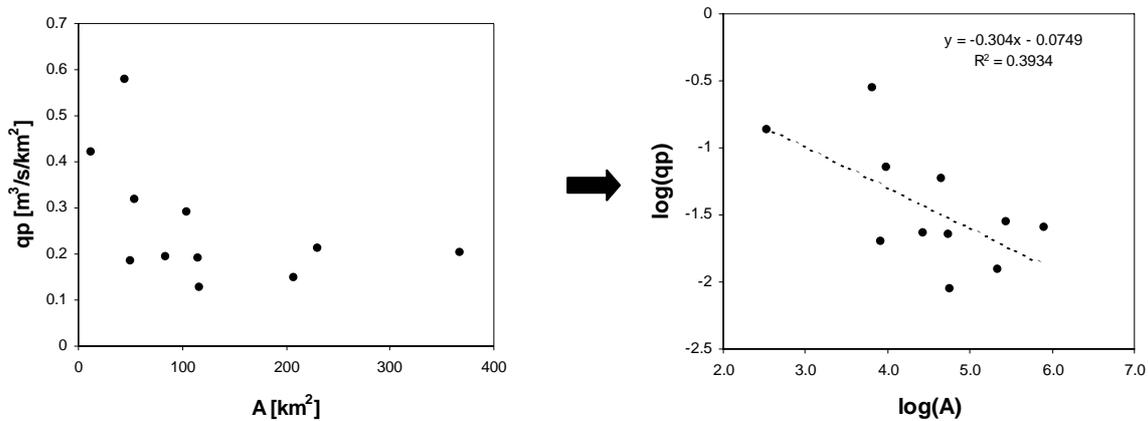
Une approche courante en régionalisation est d'estimer Q_p par régression linéaire simple (une seule variable explicative) ou multiple (plusieurs variables explicatives). C'est une technique statistique extrêmement populaire dans de nombreuses applications mais elle implique un certain nombre d'hypothèses qu'il convient de vérifier, par exemple, la linéarité de la relation entre la variable expliquée et la (ou les) variable(s) explicative(s). De plus, particulièrement en régression multiple, il convient de se méfier de l'apparente simplicité de cette méthode, et ne pas aboutir à des interprétations erronées des résultats. Le meilleur modèle n'est pas nécessairement celui qui maximise la valeur absolue du coefficient de corrélation.

La **régression linéaire** entre le débit Q_p et la surface A (voir figure ci-contre) semble être correcte.

Il faut cependant préciser que ce modèle ne contraint pas le débit à être positif. En effet, si l'ordonnée à l'origine de la droite de régression est négative, les petits bassins auront également un débit négatif. Dans le cas présent ce n'est pas le cas, mais la situation opposée (ordonnée à l'origine négative) peut très bien se produire avec un autre jeu de donnée.



Une alternative est de représenter la relation entre le débit spécifique $q_p = Q_p/A$ et la surface (figure ci-dessous). Il apparaît ici clairement que l'hypothèse de linéarité n'est plus valable et qu'une transformation logarithmique de Q_p et de A permet de mieux respecter cette hypothèse (voir figure ci-dessous).



Le passage aux logarithmes a aussi l'avantage de garantir une certaine cohérence des estimations. Ainsi quelle que soit la valeur de la variable A , la valeur du débit Q_p (ou du débit spécifique q_p) n'est jamais inférieure à 0.

On a alors la relation suivante:

$$\ln(q_p) = K_0 + K_1 \cdot \ln(A)$$

où K_0 et K_1 sont les paramètres de la régression linéaire à estimer.

Cette formule est équivalente à :

$$q_p = \exp(K_0) \cdot A^{K_1}$$

ou encore à :

$$Q_p = \exp(K_0) \cdot A^{K_1+1}$$

Dans notre cas nous obtenons :

$$\hat{K}_0 = -0.075 \text{ et } \hat{K}_1 = -0.30$$

Remarque : Il s'agit en fait d'estimateurs, basés sur les données, des paramètres K_0 et K_1 et non pas des vraies valeurs, qui sont inconnues.

La valeur de \hat{K}_1 obtenue est compatible avec l'hypothèse de décroissance du débit spécifique avec la surface du bassin versant. Le coefficient de corrélation est de -0.62. En régression linéaire, il convient de vérifier que le coefficient K_1 associé à la variable $\log(A)$ est significativement différent de 0. Si ce n'est pas le cas, on rejette l'hypothèse que $\log(q_p)$ dépend linéairement de $\log(A)$. Ceci se fait au moyen de la statistique t (test de Student).

$$t = \left| \frac{\hat{K}_1}{s^2(\hat{K}_1)} \right|$$

Où s^2 est la variance d'échantillonnage de \hat{K}_1 .

Cette valeur doit être supérieure à une valeur critique, donnée par la loi de Student pour un seuil de signification α et un nombre de degré de liberté $n-2$ (où n est le nombre de bassins versants). Lorsque $\alpha = 5\%$ et $n = 11$, cette valeur critique est de 2.26, alors que $t = 2.42$. On accepte donc l'hypothèse que $\log(q_p)$ dépend linéairement de $\log(A)$.

Le modèle est donc :

$$Q_p = 0,93 \cdot A^{0.70}$$

Une investigation plus poussée est possible dans cet exemple. On peut par exemple vérifier, sur la base de cet échantillon, qu'un modèle comprenant simultanément à la fois la surface et l'altitude moyenne n'est pas significativement meilleur que le modèle basé sur la surface uniquement. Le coefficient de corrélation est certes passé de -0.62 à -0.67 (augmentation en valeur absolue) mais le coefficient associé à la variable $\log(ALT)$ n'est pas significativement différent de 0 dans ce modèle (test basé sur la distribution de Student mentionné ci-dessus). Plus de détails concernant la régression linéaire et la sélection des variables peuvent être obtenus dans des ouvrages *ad hoc*. Une description des principaux éléments de la régression linéaire est aussi donnée dans [Meylan et Musy, 1998]. Des logiciels statistiques comme SPLUS disposent de méthodes de sélection de variables explicatives en régression multiple.

Question 5. Calcul du débit de temps de retour 20 ans

Tout d'abord notons que la variante A est une variante exclusivement locale (elle n'utilise que l'information disponible au bassin étudié).

A l'opposé, la variante C qui est une variante quasiment exclusivement régionale. Elle se base en effet sur la courbe de croissance régionale et le modèle de régression régional. La seule information locale est la surface du bassin versant. Dans cette variante, on multiplie le débit de pointe annuel moyen calculé avec le modèle régressif du point 2 (23.7 et 8.1 pour respectivement la Mentue à Yvonnand et le Lyssbach) par le facteur de croissance correspondant à un temps de retour 20 ans (1.76).

La variante B est une variante mixte. Ici on multiplie le débit de pointe annuel moyen **observé** (30.69 et 4.02 pour respectivement la Mentue à Yvonnand et le Lyssbach) par le facteur de croissance. Il est clair que pour le bassin hypothétique non jaugé, les variantes A et B ne sont pas possibles.

Le calcul selon les 3 variantes donne les résultats suivants

Bassin versant	Variante A	Variante B	Variante C
Mentue Yvonnand	50.9	54.1	41.8
Lyssbach	7.96	7.09	14.3
Bassin hypothétique	Estimation impossible	Estimation impossible	40.4

En admettant que les trois bassins sont situés dans une même région relativement homogène, le choix du meilleur modèle dépend essentiellement de la longueur de la série observée à disposition :

- Pour le cas du bassin hypothétique, le choix est trivial (variante C) puisqu'il n'y a pas d'observation.
- Pour la Mentue à Yvonnand, il y a 31 années d'observation, ce qui est généralement suffisant pour estimer le débit de temps 20 ans en se basant uniquement sur les informations locales. La variante A est donc la meilleure *a priori*.
- Enfin, pour le Lyssbach, il n'y a que 5 ans de mesures. Une estimation par ajustement fréquentiel en se basant uniquement sur des données locales est ici hasardeuse, en particulier si l'on se base sur la méthode des moments (l'estimation de l'écart type du débit de pointe annuel requise par cette méthode comporte beaucoup d'incertitudes pour de si courte série de mesures). Il est donc préférable d'utiliser la courbe de croissance régionale, qui intègre l'information des bassins voisins. Par contre, la moyenne observée, même calculée sur 5 ans de mesures, est sans doute plus précise que la moyenne calculée par le modèle du point 2. On choisira donc ici la variante B.