

**Idris Abdou**

**Yohan Allouche**

**Renaud Champredonde**

**HYDROLOGIE STATISTIQUE  
ENSEEIHT -- Hydraulique et  
Mécanique des Fluides  
3Hy SE , Mastère Hy (2007-08)**

**Contrôle d' Hydrologie Statistique**

## **Table des matières :**

**Exo 1. Exercice de reconstitution de données par régression linéaire (pluies mensuelles à Mens et Roissard, bassin du Drac, Alpes).....3**

**TD 1 : Hydrologie statistique – analyse univariee – gumbel & poisson : Crues annuelle & crues extrêmes de la Garonne à Toulouse.....10**  
**Extension du TD .....13**

**TD2 : Exercice 2 : A.C.P. (Analyse en Composantes Principales).....16**

**TD3 - Bureau d'Etudes « Hyd.Stat.», Jan.2007. Analyse statistique de données de pluies & de débits.....22**

## Exo 1. Exercice de reconstitution de données par régression linéaire (pluies mensuelles à Mens et Roissard, bassin du Drac, Alpes)

Reconstitution des pluies mensuelles de Mens à partir de celles de Roissard:

- **Mars 1946:** vérifier et compléter la réponse du polycopié, si possible en refaisant vous-même le calcul de reconstitution, et donner aussi l'intervalle de confiance à 90% (à partir d'une table de la F de répartition de Gauss);
- **Mars 1940:** procéder de même... obtenir (i) la pluie reconstituée, et (ii) l'intervalle de confiance à 90

Pluies mensuelles observées sur deux stations de jaugeages dans les Alpes.

Le tableau de pluies mensuelles fourni sur ces deux stations alpines comportent des données manquantes en mars 1946 et 1940 à la station de Mens.

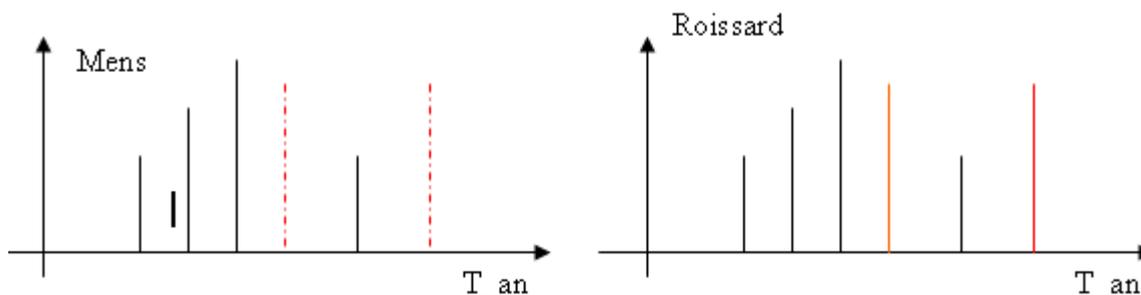


Illustration 1: Histogramme de Mens et Roissard en fonction du temps

a) Si on avait pas la deuxième station on aurait corrélé  $Y(t)$  de mens avec le temps

b) Ici comme on a une deuxième station (Roissard) on profite pour faire une estimation aléatoire.

Mens étant conditionné par Roissard on calque en terme de pluviométrie l'évolution de la station de Mens sur celle de Roissard. Ainsi on estime que quand la pluviométrie est importante sur Roissard, elle l'est aussi sur Mens et vice versa. Ce qui nous donne de tendance.

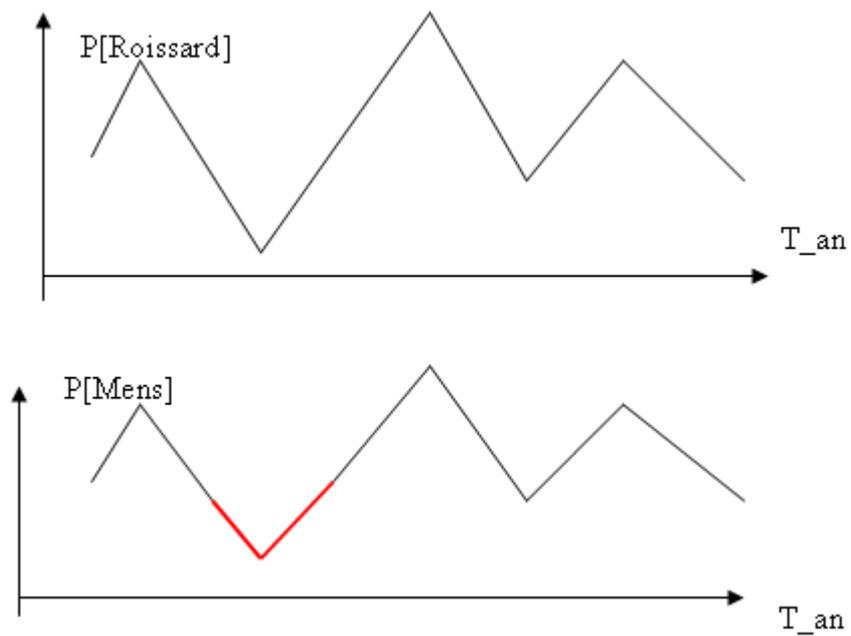
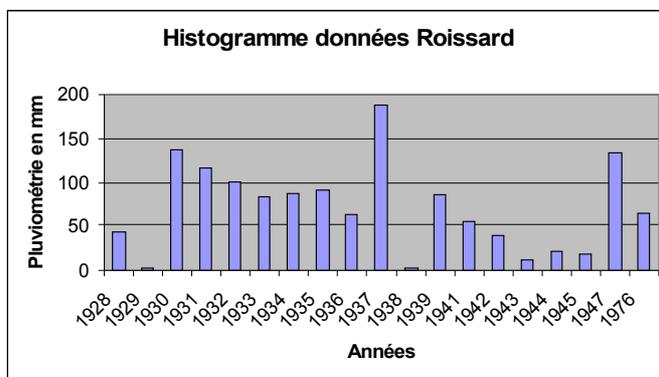
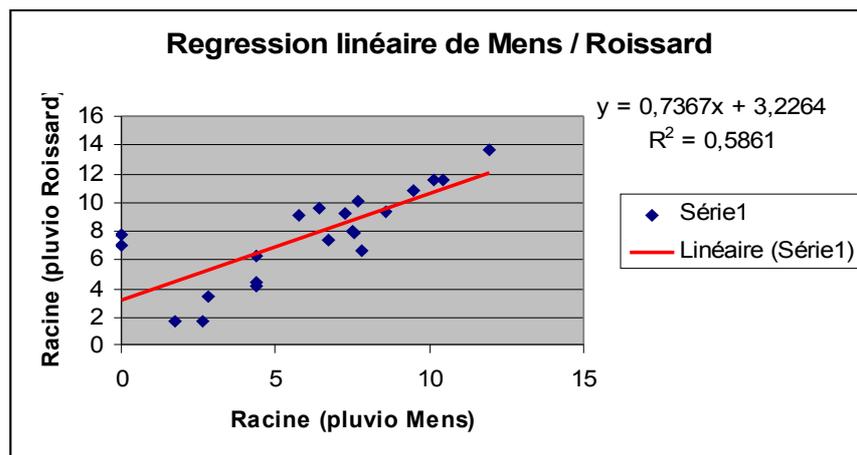
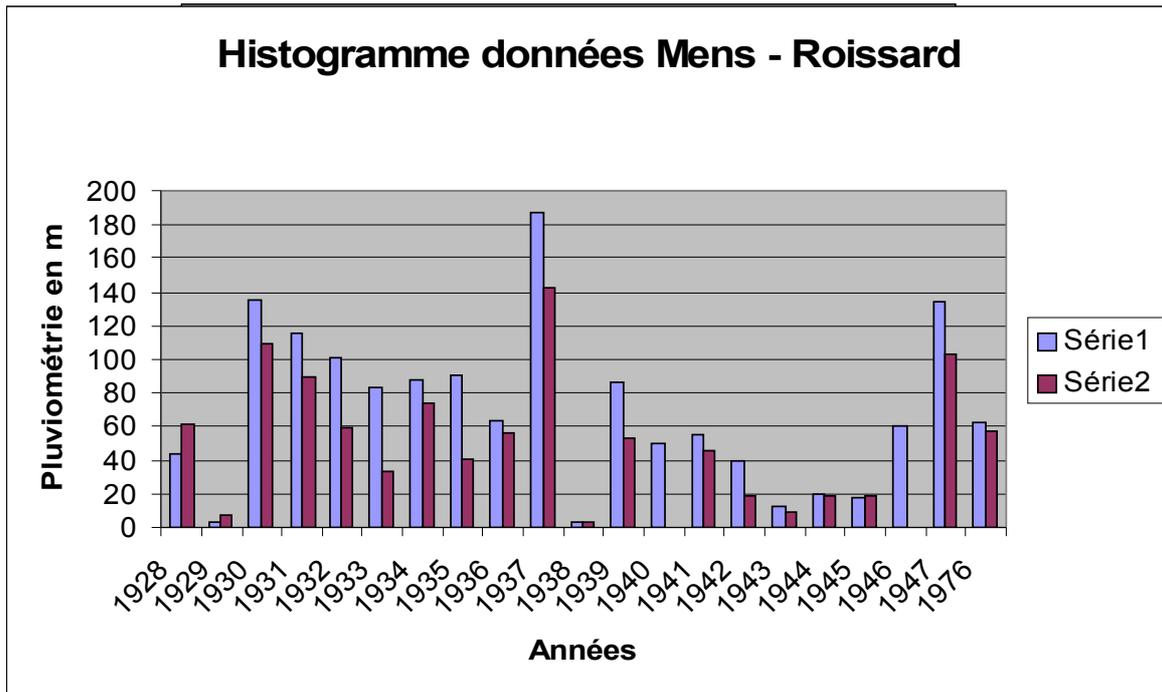


Illustration 2: Tendence des pluviométries de Mens et Roissard

### a) Analyse des données par extrapolation graphique

Histogramme de Mens et Roissard en fonction du temps :





- a) Si on avait pas la deuxième station on aurait corrélé  $Y(t)$  de mens avec le temps
- b) Ici comme on a une deuxième station (Roissard) on profite pour faire une estimation aléatoire

Mens étant conditionnée par Roissard on calcule en terme de pluviométrie l'évolution de la station de Mens sur celle de Roissard. Ainsi on estime que quand la pluviométrie est importante sur Roissard, elle l'est aussi sur Mens et vice versa. Ce qui nous donne de tendance.

## b) Résolution analytique :

Le travail consiste à la reconstitution des données de pluies à Mens à partir de celle de Roissard à la même période de l'année.

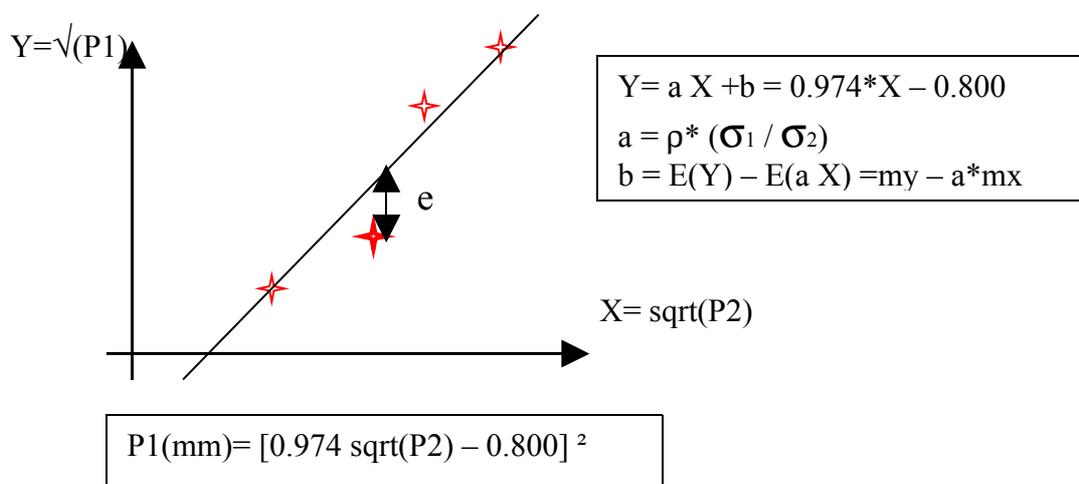
D'abord on annule les deux points manquants et on applique l'analyse sur les points connus.

Comme on connaît les données sur la station N°2 alors on fait la régression linéaire de X1/X2.

Soit  $Y = X1 = \sqrt{P1}$  : variable à expliquer (Mens)

$X = X2 = \sqrt{P2}$  : variable explicative (Roissard)

On trace le tableau des pluviométries donné



A.N Pour  $P_{\text{Roissard}} = 60 \text{ mm}$  on a  $P_{\text{Mens}} = 45,49 \text{ mm}$

$P_{\text{Roissard}} = 50 \text{ mm}$  on a  $P_{\text{Mens}} = 37,05 \text{ mm}$

### Méthode

On prends l'approximation Gaussienne la meilleure

Moments de Mens  $X = \sqrt{P1}$  :  $m1 = 6.7$  en  $\sqrt{\text{mm}}$  et  $\sigma_1 = 2.9$  en  $\sqrt{\text{mm}}$  Moments de Roissard  $Y = \sqrt{P2}$  :  $m2 = 7.7$  en  $\sqrt{\text{mm}}$  et  $\sigma_2 = 2.8$  en  $\sqrt{\text{mm}}$   $a = \rho * (\sigma_1 / \sigma_2) = 0.974$  et  $b = E(Y) - E(a X) = m_y - a * m_x = -0.8 \sqrt{\text{mm}}$

La droite de régression introduit une erreur ( dite d'estimation)

Par définition  $\hat{Y} = a X + b$  est la moyenne conditionnée sachant que  $\hat{Y} = E(Y/X)$  : théorème de Bayes et c'est le meilleur estimateur (  $\text{Min } \sigma_e^2$  ) si X, Y Gaussien.

Erreur  $e = Y - \hat{Y}$  est une V.A  
 $\sigma_e = \sqrt{(1 - \rho)} * \sigma_y$   
 $\rho_{x;y} = 0.94$  ( Mens – Roissard )

$\mathcal{E}_{rms} = \sigma_e / \sigma_y = \sqrt{(1 - \rho^2)}$

A.N  $\mathcal{E}_{rms} = 0.34$  Root means square error  
 Erreur relative à la normalisation

**Intervalle de confiance hachurée à I90% directement lié à  $\sigma_e$  :**

L'objectif est de caractériser une intervalle de confiance autour de la moyenne telle que la variable aléatoire ait une probabilité  $P = 90\%$  d'appartenir à cette intervalle à partir d'une table de la la fonction de répartition  $F(u)$  d'une variable gaussienne centrée réduite  $[N(0,1)]$

Table de la FdR de Gauss

F(0)	F(0.25)	F(0.52)	F(0.84)	F(1.28)	F(1.64)	F(2.32)	F(2.57)
0.50	0.60	0.70	0.80	0.90	0.95	0.99	0.995

L'intervalle à 90% de Probabilité (I90 %) est :

Proba (  $U \leq 1.64$  ) = 0.95

Proba (  $U \geq -1.64$  ) = 0.05 par symétrie de la loi

Donc

Proba (  $-1.68 \leq U \leq 1.64$  )  $\Rightarrow$  I90 % =  $[-1.68 ; +1.68 ]$  pour la valeur aléatoire centrée réduite  $N(0,1)$

Or  $X = mx + \sigma_x * U$

$U = 1.64$  pour I90 % =  $\hat{Y} + U * \sigma_e$

On obtient donc la variable aléatoire X gaussienne  $N[mx, \sigma_x^2]$

L'objectif est de générer un des points ( étoiles bleues) qui appartienne à l'intervalle I90% qu'on va corréler par  $\sigma_e$  puis reconstituer  $Y_n$  :

$Y_n \text{ reconstitué} = a X_n + b + (\sigma_e * G)$

G est une réplique d' une V.A  $N[0,1]$

$N[0,1]$  : Générateur d'une V.A gaussienne réduite

Cette méthode est intéressante car  $\sigma_x$  et  $\rho$  interviennent alors que la première méthode (dite de glissement) introduit beaucoup d'erreur (notamment dans le calcul de variance).

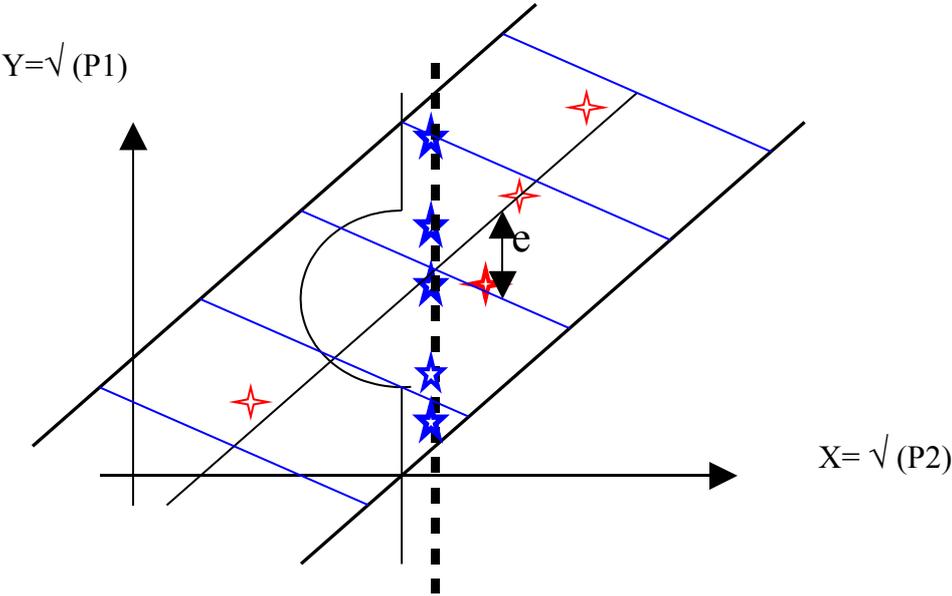


Figure 1 . Tableau pluies mensuelles de mars avril à Mens & Roissard (BV du Drac, Alpes)

Pluies Mensuelles en 2 stations d'un Bassin Versant du Drac (de 1928 à 1947, et en 1976)				
Années	S1 Mens		S2 Roissard	
	Mars	Avril	Mars	Avril
1928	61	84	44	132
1929	7	65	3	79
1930	109	53	135	115
1931	90	40	116	57
1932	59	67	101	89
1933	33	21	83	44
1934	74	135	88	130
1935	41	18	91	131
1936	56	132	64	132
1937	143	56	188	78
1938	3	19	3	7
1939	53	91	86	92
<b>1940</b>	<b>37.05</b>	<b>x</b>	<b>50</b>	<b>112</b>
1941	45	83	55	117
1942	19	23	40	42
1943	8	25	12	35
1944	19	30	20	30
1945	19	17	18	18
<b>1946</b>	<b>45.49</b>	<b>x</b>	<b>60</b>	<b>44</b>
1947	103	35	134	31
1976	57	60	62	65

## HYDROLOGIE STATISTIQUE – TD1: ANALYSE UNIVARIEE – GUMBEL & POISSON : CRUES ANNUELLE & CRUES EXTREMES DE LA GARONNE A TOULOUSE

1. *Quelle est la variable hydrologique étudiée (expliquez le terme crue)?*

La variable hydrologique étudiée est la hauteur d'eau. En utilisant la courbe de tarage on peut obtenir facilement les débits.

2. *Utilisez la FdR proposée pour obtenir la crue annuelle centennale (expliquez). Question subsidiaire: est-ce une loi de Gumbel ? (paramètres=?)*

Pour cela il convient au préalable de déterminer s'il s'agit d'une loi de Gumbel et s'il en est le cas, d'identifier les paramètres  $[\alpha, \beta]$  qui caractérisent la loi de Gumbel.

- La représentation graphique** laisse (à priori) à penser qu'il pourrait s'agir d'une loi de Gumbel. En effet nous avons là :  
une représentation à l'envers du graphe de la FdR:  $H=f[FdR(H)]$   
On lit sur l'axe des ordonnées la hauteur et sur l'axe des abscisses la FdR(H).  
On constate également que ces deux axes ont subi une transformation qui reste à déterminer.
- Pour confirmer qu'il s'agit bien d'une **loi de Gumbel** nous proposons la méthode qui consiste à tester 3 points appartenant à la séquence des hauteurs tracés. Si ces trois points sont alignés par suite de la transformation alors nous pourrions affirmer qu'il s'agit d'une loi de Gumbel.

Au préalable il convient de caractériser la transformation effectuée sur les axes.  
En faisant l'hypothèse que la sus transformation est une relation affine  
(type  $y= ax +b$ ) lorsque F(H) est une loi de Gumbel

### 1<sup>ère</sup> Etape : **Détermination des paramètres $[\alpha, \beta]$**

Remarque  $[\alpha, \beta]$  sont en mètre car la hauteur H est en mètre

$$F(H) = \exp[-\exp[-(H-\alpha)/\beta]] \quad \text{on pose } Y = (H-\alpha)/\beta$$

On détermine la pente (en prenant deux points éloignés sur la droite) et en déduit  $\beta = 0.9119$

Sachant  $\beta$  et un couple de point  $(x = F(H), y = H)$ ,

$$\text{on en déduit } \alpha = H - \beta * [-\ln[-\ln F]] = 1.9822$$

### 2<sup>ème</sup> étape: **Détermination des paramètres $(m_H, \sigma_H)$ : moments et écart type**

$$\sigma_H = 1.28 * \beta = 1.1672$$
$$m_H = \alpha + 0.45 * \sigma_H = 2.5075$$

**3<sup>ème</sup> étape : Test des points choisis**

On prends un trio des points (  $H_1=2m$ ,  $H_2= 4$ ,  $H_3=6m$ ) et détermine  $F(H_i)$  .  
On constate que ces points se retrouvent bien sur le diagramme de Gumbel  
donné dont l'équation est de type  $H_i=\beta * F(H_i) + \alpha$

**c) La crue annuelle centennale :**

On utilise la loi de Gumbel  $H= - \beta * [- \ln [- \ln F]] + \alpha = 6.2 \text{ m}$

**3°) La probabilité d'observer n crues supérieures à la crue centennale sur une période d'observation de 225 années**

Il s'agit là d'un évènement rare donc on utilisera la loi de Poisson sur un temps de retour  $T_r=100$  ans. Cette loi représente la probabilité d'observer n dépassement de la crue de temps de retour  $T_r=100$  précédemment calculé ( $H_{100}$ ). Voir 2°)

La loi de poisson s'écrit :

$$\text{Proba}(k=n) = ((\mu * T_D)^n / n!) * \exp (-\mu * T_D)$$
$$\mu = 1 / T_R$$
$$T_R = 100 \text{ ans et } T_D = 225 \text{ ans}$$

**a) Au moins une :**

$$P_0 = 0.105$$
$$\text{Proba}(k \geq 1) = 1 - P_0 = 89,4\%$$

**b) Au moins deux:**

$$P_1 = 0.237$$
$$\text{Proba}(k \geq 2) = 1 - P_0 - P_1 = 65,7\%$$

**c) Au moins trois :**

$$P_2 = 0.266$$
$$\text{Proba}(k \geq 3) = 1 - P_0 - P_1 - P_2 = 39.1\%$$

## Fichier.m Matlab du TD1 avec Résultat

```
clear all;  
close all;  
  
%  
% Question 0 : Est-ce une loi de Gumbel ?  
%  
alpha = 4;  
beta = 1;  
h1 = 2;  
h2 = 4;  
h3 = 6;  
f1 = exp(-exp(-(h1-alpha)/beta));  
f2 = exp(-exp(-(h2-alpha)/beta));  
f3 = exp(-exp(-(h3-alpha)/beta));  
f_transfo1 = -log(-log(f1));  
f_transfo2 = -log(-log(f2));  
f_transfo3 = -log(-log(f3));  
f_transfo = [f_transfo1,f_transfo2,f_transfo3];  
  
%  
% Question 1 : Identifier alpha et beta  
%  
% Pente réelle en la calculant sur le graphique  
beta_reel = ((8.25-2)/(-log(-log(0.998))+log(-log(0.15))))  
  
% Expression de alpha  
alpha_reel = 3.35-beta_reel*(-log(-log(0.80)))  
  
% D'où la hauteur de la crue centennale  
hauteur_crue = ((-log(-log(0.99))) * beta_reel) + alpha_reel
```

***Extension TD 1 : Résolution du problème par la méthode Numérique :***

Etape 1 : Numérisation des données de la Garonne

Etape 2 : Mise en place des deux vecteurs T\_date et H\_crue

Etape 3 : Calcul des moments d'ordre 1 des hauteurs :

Etape 4 : Classement des hauteurs par ordre croissant

Etape 5 : Fonction de répartition de la hauteur

Expression de Hazen

Etape 6 : Calcul de la hauteur centennale

Etape 7 : Ajustement par une loi de Gumbel - relation paramètres/moments de la loi de Gumbel

Etape 8 : Probabilité de non dépassement de la hauteur centennale (crue extrême)

## Fichier.m Matlab de Extension du TD1 avec Résultat

```
clear all;
close all;

%
% Question extension
%

% Chargement des données
load -ascii 'numerisation_donnees_garonne';
donnees = numerisation_donnees_garonne;
date = donnees(:,1);
hauteurs = donnees(:,2);
debits = donnees(:,3);
N=length(hauteurs);

% Calcul des moments d'ordre 1 des hauteurs :
moy_h = mean(hauteurs);
i = 1;
moment = zeros(N,1);
for i=1:N;
    moment(i) = mean(hauteurs(i)-moy_h);
end

% Classement des hauteurs par ordre croissant
[hauteurs_c i_c] = sort(hauteurs);
date_c = date(i_c);

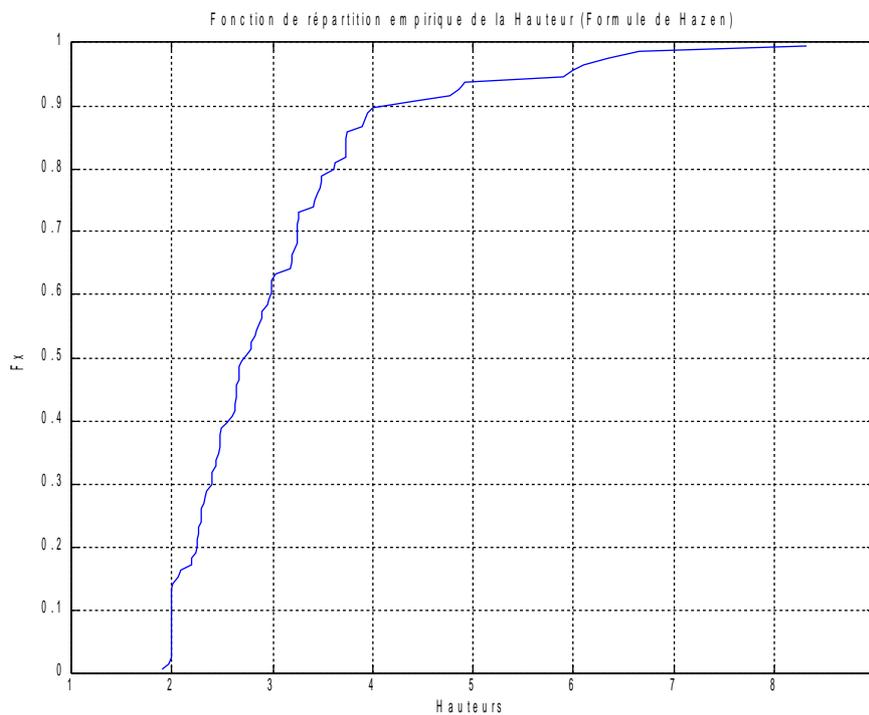
% Fonction de répartition de la hauteur
% Expression de Hazen
j = 1;
fx = zeros(N,1);
for j = 1:N;
    fx(j)=(j-0.5)/N;
end
plot(hauteurs_c,fx)
grid on
xlabel('Hauteurs')
ylabel('Fx')
title('Fonction de répartition empirique de la Hauteur (Formule de Hazen)')

% Calcul de la hauteur centennale
ecart_type_h = std(hauteurs); % Ecart-type sur les hauteurs
alpha = moy_h-0.45*ecart_type_h;
beta = ecart_type_h/1.28;
hauteur_crue = ((-log(-log(0.99))) * beta) + alpha
```

```
% Ajustement par une loi de Gumbel - relation paramètres/moments de la loi
% de Gumbel
a = pi / (sqrt(6)*ecart_type_h);
h0 = moy_h - 0.045*ecart_type_h;

% Probabilité de non dépassement de la hauteur centennale (crue extreme)
f_de_hauteur_crue = exp(-exp(-a*(hauteur_crue-h0)))
```

Ce script aboutit à la représentation suivante pour la fonction de répartition empirique de Hazen :



Nous obtenons ainsi les résultats suivants :

moy\_h =  
3.0503

hauteur\_crue =  
6.6054 mètres

a =  
1.1342

h0 =  
2.9994

f\_de\_hauteur\_crue = 0.9834

## TD2-Exo.2 : A.C.P. (Analyse en Composantes Principales)

**1.1 Question préalable** : quelle est la signification de la variable hydrologique analysée (débit « Q » -- ou débit spécifique « q ») ? A quel type de normalisation des débits cela correspond-il ?

Il s'agit des écoulements mensuels en 6 stations hygrométriques (probablement des exutoires de bassins versants. Dans ce cas ces mesures représentent des « quantités de pluies » relevées en mm sur les exutoires du bassin versant, soit des mm par unité d'aire.

**1.2 Moments simples.** Calculer la moyenne, la variance et l'écart-type de chaque variable (*en utilisant directement les données, ou bien encore, les sommes  $\Sigma$  données en annexe*).

Scripte matlab :

```
load -ascii 'donnees.csv';

donnees = donnees;
annee = donnees(:,1);
observation = donnees(:,2);
naguilhes = donnees(:,3);
lanoux = donnees(:,4);
izourt = donnees(:,5);
gnioure = donnees(:,6);
caillaouas = donnees(:,7);
bleu = donnees(:,8);

% Somme de l'ensemble des valeurs pour chacune des stations
somme1=sum(naguilhes);
somme2=sum(lanoux);
somme3=sum(izourt);
somme4=sum(gnioure);
somme5=sum(caillaouas);
somme6=sum(bleu);

% Moyenne des valeurs pour chacune des stations
moyenne1=mean(naguilhes);
moyenne2=mean(lanoux);
moyenne3=mean(izourt);
moyenne4=mean(gnioure);
moyenne5=mean(caillaouas);
moyenne6=mean(bleu);

% Ecart-type des valeurs pour chacune des stations
ecarttype1=std(naguilhes);
ecarttype2=std(lanoux);
ecarttype3=std(izourt);
ecarttype4=std(gnioure);
```

```

ecarttype5=std(caillaouas);
ecarttype6=std(bleu);

% On centre les valeurs
for i=1:23;
    centre1(i,1)=(naguilhes(i,1) - moyenne1)/ecarttype1;
    centre2(i,1)=(lanoux(i,1) - moyenne2)/ecarttype2;
    centre3(i,1)=(izourt(i,1) - moyenne3)/ecarttype3;
    centre4(i,1)=(gnoiure(i,1) - moyenne4)/ecarttype4;
    centre5(i,1)=(caillaouas(i,1) - moyenne5)/ecarttype5;
    centre6(i,1)=(bleu(i,1) - moyenne6)/ecarttype6;
end

% Variance des valeurs pour chacune des stations
variance1=var(naguilhes);
variance2=var(lanoux);
variance3=var(izourt);
variance4=var(gnoiure);
variance5=var(caillaouas);
variance6=var(bleu);
    
```

Les valeurs obtenues par ces différentes commandes sont regroupées dans le tableau ci dessous :

Station	Naguilhes	Lanoux	Izourt	Gnoiure	Caillaouas	Bleu
Moyenne	377,6	298,5	394,2	401.2	327.2	176.7
Ecart Type	93.2	71.7	78.9	85.3	115	97.8
Variance	8696.4	5143.5	6228.9	7281.1	13229	9569.1

**1.3 Matrice de corrélation.** Calculer la matrice de corrélation (i.e., la matrice de covariance des variables réduites). Remarques ?

Script Matlab :

```

X=[232  180   450   450   391   163
228   155   355   337   271   110
416   344   391   376   306   125
479   370   503   490   387   234
323   250   358   334   293   162
379   260   288   269   432   351
423   325   476   505   380   144
154   141   215   197   137   37
523   400   567   590   516   337
440   340   337   364   318   137
478   370   412   441   518   314
431   329   365   386   313   241
359   294   313   358   274   160
295   271   318   305   208   104
464   360   381   415   597   406
    
```

*Hydrologie statistique 2007-2008 M. Ababou*

```

366      285      451      428      228      139
472      353      478      489      377      223
383      310      396      404      215      66
370      320      423      449      242      95
417      359      403      447      372      181
334      238      393      400      197      87
447      370      471      459      348      170
273      242      322      335      205      78]
X =

```

```

232 180 450 450 391 163
228 155 355 337 271 110
416 344 391 376 306 125
479 370 503 490 387 234
323 250 358 334 293 162
379 260 288 269 432 351
423 325 476 505 380 144
154 141 215 197 137 37
523 400 567 590 516 337
440 340 337 364 318 137
478 370 412 441 518 314
431 329 365 386 313 241
359 294 313 358 274 160
295 271 318 305 208 104
464 360 381 415 597 406
366 285 451 428 228 139
472 353 478 489 377 223
383 310 396 404 215 66
370 320 423 449 242 95
417 359 403 447 372 181
334 238 393 400 197 87
447 370 471 459 348 170
273 242 322 335 205 78

```

```
>> Rqq=corr(X)
```

```
Rqq =
```

```

1.0000  0.9627  0.6363  0.7030  0.6708  0.6455
0.9627  1.0000  0.5998  0.6817  0.5775  0.5277
0.6363  0.5998  1.0000  0.9651  0.4732  0.3119
0.7030  0.6817  0.9651  1.0000  0.5347  0.3566
0.6708  0.5775  0.4732  0.5347  1.0000  0.9194
0.6455  0.5277  0.3119  0.3566  0.9194  1.0000

```

```
>> Cqq=cov(X)
```

```
Cqq =
```

1.0e+004 \*

```
0.8696 0.6439 0.4683 0.5594 0.7195 0.5888
0.6439 0.5144 0.3395 0.4172 0.4764 0.3702
0.4683 0.3395 0.6229 0.6499 0.4296 0.2408
0.5594 0.4172 0.6499 0.7281 0.5247 0.2976
0.7195 0.4764 0.4296 0.5247 1.3229 1.0344
0.5888 0.3702 0.2408 0.2976 1.0344 0.9569
```

**1.4 Diagonalisation de la matrice de corrélation.** Afin d'alléger les calculs, on donne en annexe la matrice diagonale D et la matrice de passage P. En déduire les valeurs propres, ainsi que les vecteurs propres ou « composantes principales ».

```
>> [P,D]=eig(Rqq)
```

```
%matrice de passage
```

```
P =
```

```
0.6457 -0.3174 0.2324 -0.4728 0.0129 0.4524
-0.5854 0.1670 -0.2377 -0.6211 -0.0655 0.4277
-0.3404 -0.4918 0.3251 0.3832 -0.4889 0.3883
0.3124 0.6476 -0.1879 0.2833 -0.4431 0.4137
0.0450 -0.3129 -0.6462 0.3498 0.4465 0.4010
-0.1578 0.3351 0.5753 0.2033 0.6007 0.3601
```

```
D =
```

```
%Matrice diagonale
```

```
0.0209 0 0 0 0 0
0 0.0246 0 0 0 0
0 0 0.0700 0 0 0
0 0 0 0.5537 0 0
0 0 0 0 1.1228 0
0 0 0 0 0 4.2079
```

On obtient donc le tableau des Cps suivants :

CP	1	2	3	4	5	6
Valeur	4.2079	1.1228	0.5537	0.0700	0.0246	0.0209

## 2. Analyses et applications

2. Montrer que, dans le cas présent, la CP1 représente les six variables avec un poids à peu près égal pour toutes.

La valeur CP1 est de 4,2079. Pour la représenter en terme de pourcentage d'information on effectue :

$$\%CP1=4,2079/6*100=70,13\%$$

CP1 contient donc environ 70% de l'information portée par les variances des 6 stations hydrologiques. Cependant cet axe ne sera pas un axe discriminant, en effet si on observe la dernière colonne de la matrice de passage, représentant les  $P_{ij}$  de CP1, on remarque que les valeurs sont très similaires. Il est donc fortement probable que les stations ne soient pas discriminées par cet axe.

La discrimination des stations se fera en apportant de l'information complémentaire (CP2, CP3 etc..).

2. Calculer le % de variance expliquée par les K premières CP, en faisant varier K de 1 à 6. En déduire que l'on ne perd que quelques % d'information en éliminant les CP4, CP5 et CP6.

On vient de calculer le % de variance pour CP1 (K=1).  
Les % pour les CP suivantes sont :

$$CP2=1.1228/6*100=18,71\%$$

$$CP3=0.5537 /6*100=9,23\%$$

$$CP4=0.0700/6*100=1,17\%$$

$$CP5=0.0246/6*100=0,41\%$$

$$CP6=0.0209/6*100=0,35\%$$

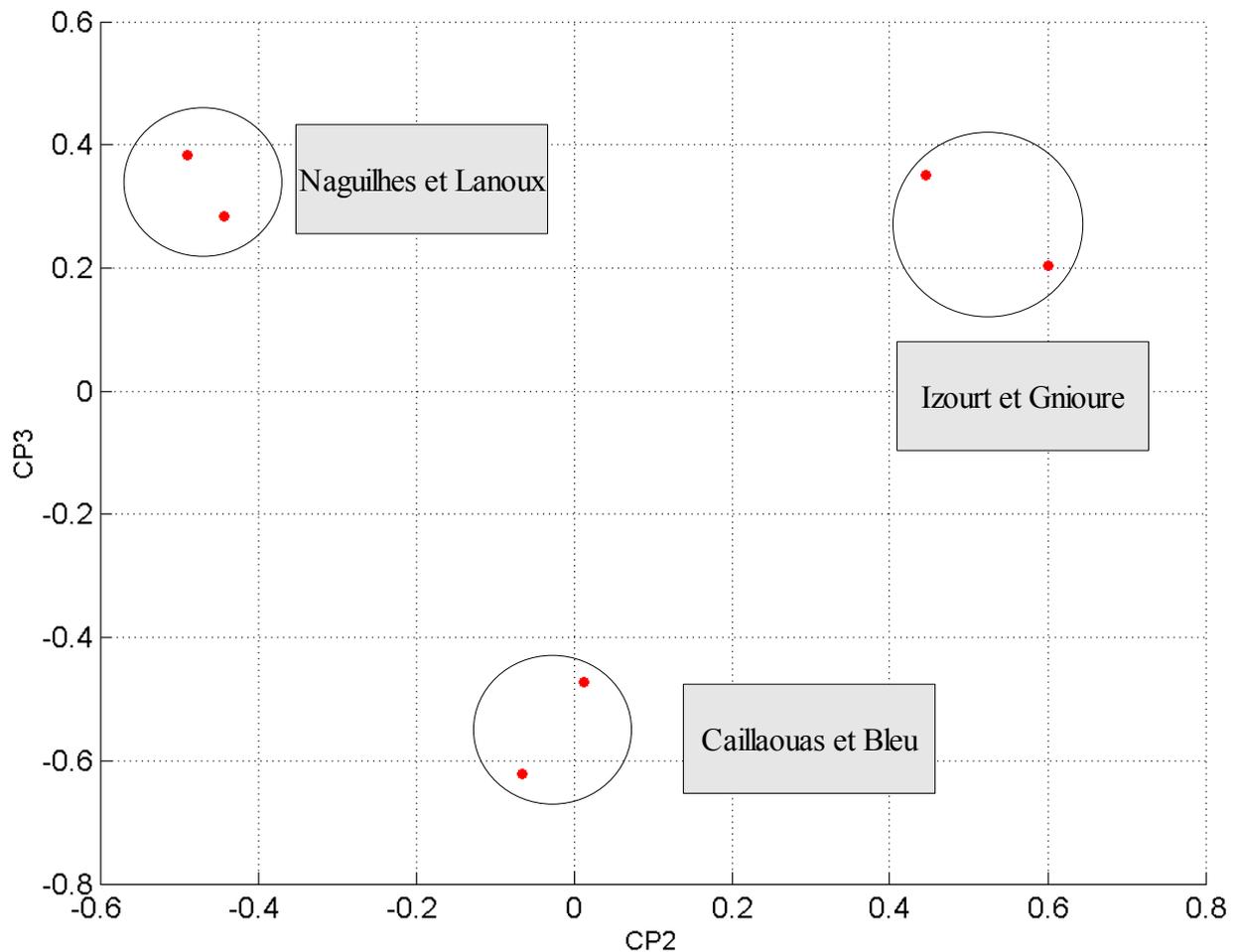
On obtient donc :

$$\text{Somme } CP1,CP2,CP3= 70,13+18,71+9,23=98,07\%$$

Les trois premiers axes comprennent donc plus de 98% de l'information disponible, les axes 4,5 et 6 semblent donc inutiles pour discriminer les stations. Il cependant signaler que ces axes peuvent être utiles dans le cas d'étude de population très proches.

De plus il faut signaler que les axes à utiliser seront dépendants des objectifs de l'étude. Si on cherche à discriminer spatialement des stations hydrologiques la sélection des axes choisis semble adaptée. Par contre si on essaye de déterminer des types de stations, l'axe CP1 sera peu utile, car représentatif de toutes les stations.

2. La **figure 1** représente les 6 stations de jaugeage de débits (variables 1,...,6) dans le plan des (CP2,CP3). Y a-t-il des regroupements possibles ? Que pouvez en déduire ?



**Figure 1 : Représentation graphique des 6 stations dans le plan (CP2,CP3)**

Sur le graphique on observe très nettement trois couples de stations. Pour identifier ces stations (comme reporté sur la figure), il suffit de repérer les "coordonnées" de chaque station dans la matrice de passage.

Ce fort appariement deux à deux des stations hydrologiques amènent à penser que ces couples sont situés sur le même exutoire de bassin versant, ces stations sont donc probablement relativement proche géographiquement l'une de l'autre à l'intérieur d'un couple.

## **TD3 - Bureau d'Etudes « Hyd.Stat. », Jan.2007** **ANALYSE STATISTIQUE DE DONNEES DE PLUIES & DE DÉBITS**

### *1 Calculer et interpréter des fonctions de covariances temporelles (auto et croisées) sur les chroniques hydrologiques données*

L'objectif de cette partie consiste à étudier les fonctions de covariances temporelles sur les chroniques hydrologiques de la source pyrénéenne Aliou. L'analyse de ces interactions mettra en évidence le caractère karstique de cette source. La première étape s'intéresse à la visualisation des données pluviométriques en parallèle avec l'hydrogramme associé correspondant à une sortie du bassin versant. Le script Matlab permettant la visualisation des variables hydrologiques est fournie ci-dessous :

```
clear all;

close all;

load -ascii 'PtQt_ALIOU_de75a95_3cols.txt';

tpq= PtQt_ALIOU_de75a95_3cols;

t=tpq(:,1);

p=tpq(:,2);

q=tpq(:,3);

subplot(4,1,1)

plot(p)

grid on

axis ij

xlabel('Temps en jours')

ylabel('Pluie en mm')

title('Hyétogramme')

hold on;

subplot(4,1,2)

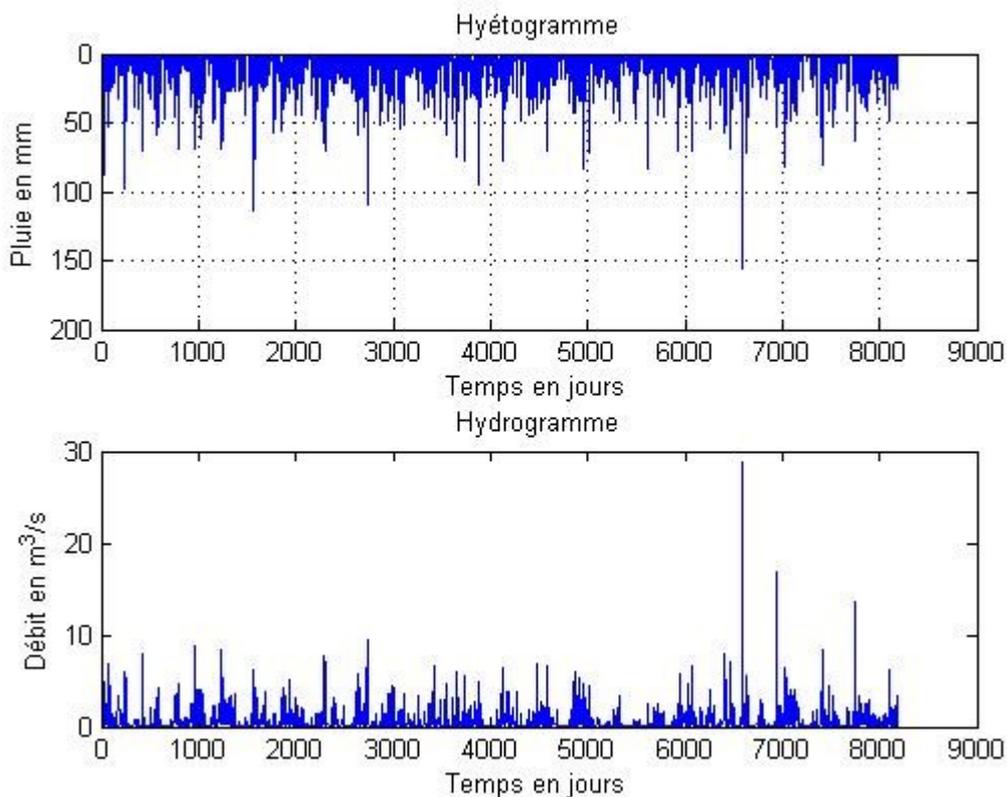
plot(t,q)

xlabel('Temps en jours')

ylabel('Débit en m^3/s')
```

```
title('Hydrogramme')
```

```
hold off;
```



**Figure 2 : hydrogramme et hyétogramme d'Aliou**

Ce premier graphique met en évidence une certaine corrélation visuelle entre les deux variables tracées. En effet la quantité d'eau tombée, tracée sur le graphique du haut, semble avoir une conséquence sur le débit à la sortie du bassin versant. C'est ainsi que nous observons pour certains maxima de précipitation, un maxima sur l'hydrogramme. La deuxième remarque s'attache à décrire un décalage entre ces deux maxima. En effet le débit réagit plus tard sur le graphique car nous observons un maximum de débit atteint après un maximum pluviométrique.

Cependant cette relation n'est pas présente sur la totalité des maxima pluviométriques et certains pics de quantité d'eau tombée semblent n'avoir aucune conséquence sur le débit à l'exutoire.

L'analyse ci-après permet de quantifier cette relation.

Pour exprimer la présence ou non d'une relation entre les deux variables hydrologiques tracées, nous allons exprimer l'autocorrélation tout d'abord puis ensuite la corrélation entre les deux variables.

Le script ci-dessous nous a permis de représenter le graphique de la fonction d'autocorrélation pour la variable hydrologique «précipitation», puis pour la fonction d'autocorrélation de la variable «débit». Le dernier graphique représente la fonction de covariance entre les deux variables hydrologiques.

```
clear all;

close all;

load -ascii 'QALIOU93_SEQ.txt';

load -ascii 'PALIOU93_SEQ.txt';

q= QALIOU93_SEQ; % Q pour 1 annee 93

p= PALIOU93_SEQ; % P pour 1 annee 93

maxlag = 350;

cpp=xcov(p,maxlag,'biased'); %'unbiased'

mcpp = length(cpp)

tau = [-maxlag:1:maxlag];

subplot(3,1,1)

plot(tau,cpp)

grid on

ymin = min(cpp);

ymax = max(cpp);

axis([-10 10 ymin ymax])

xlabel('Tau delai')

ylabel('Cpp')

title('Fonction d autocovariance')

cqq=xcov(q,maxlag,'biased'); %'unbiased'

subplot(3,1,2)

plot(tau,cqq)
```

```
grid on

ymin = min(cqq);

ymax = max(cqq);

axis([-100 100 ymin ymax])

xlabel('Tau delai')

ylabel('Cqq')

title('Fonction d autocovariance')

hold on

cpq=xcov(q,p,maxlag,'biased'); %'unbiased'

subplot(3,1,3)

plot(tau,cpq)

grid on

ymin = min(cpq);

ymax = max(cpq);

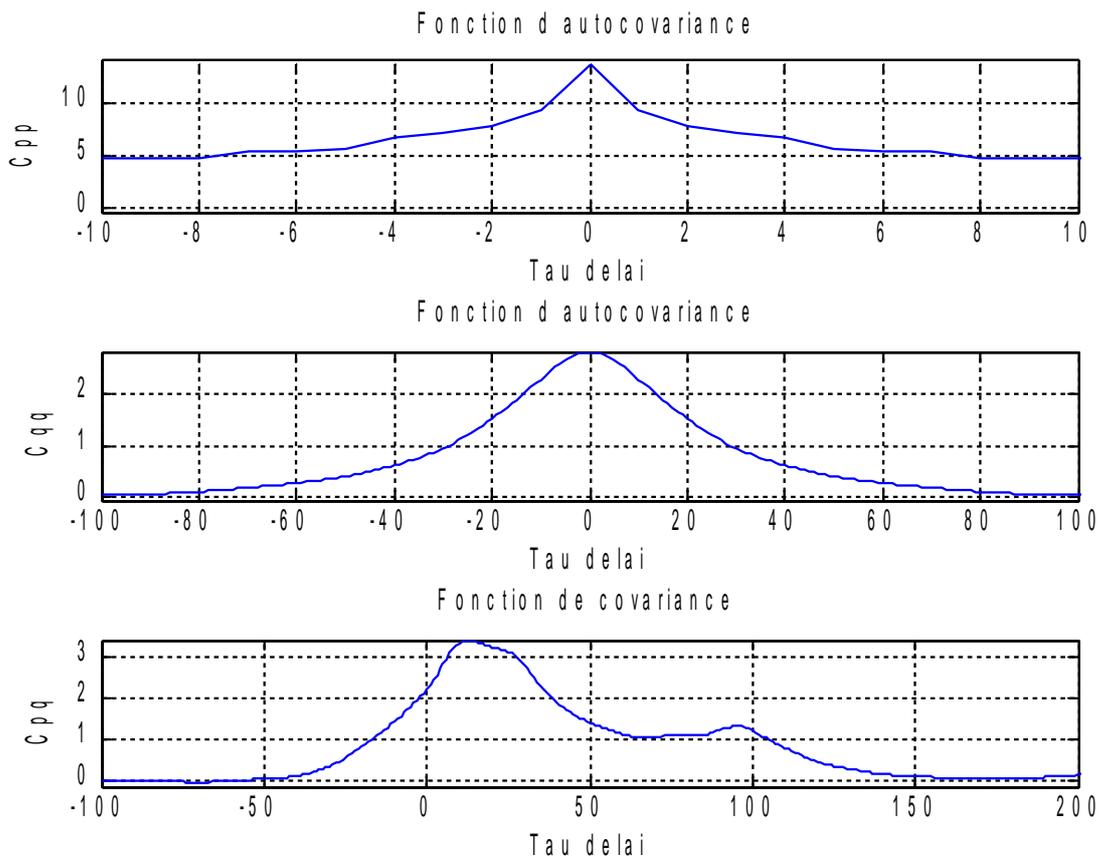
axis([-100 200 ymin ymax])

xlabel('Tau delai')

ylabel('Cpq')

title('Fonction de covariance')
```

**Ce script conduit aux trois graphiques ci-dessous :**



**Figure 3 : Fonction d'autocorrélation pour la variable hydrologique «précipitation», puis pour la fonction d'autocorrélation de la variable «débit». Fonction de covariance entre les deux variables hydrologiques.**

Sur le premier graphique, nous observons que la fonction d'autocovariance présente un maximum centré sur 0. La valeur maximale atteinte par la fonction d'autocovariance est de 5 aux extrémités. Les valeurs sont toutes positives ce permet d'affirmer qu'il y a uniquement corrélation positive. Aucun retard n'est observé.

Le second graphique présente la fonction d'autocorrélation de la variable hydrologique «débit». De la même façon que pour les précipitations, nous observons que la fonction est strictement positive. C'est qui permet d'annoncer une corrélation positive. Le maximum atteint en 0 n'indique aucun retard.

Les résultats de ces deux graphiques sont cohérents puisqu'ils donnent des résultats attendus à savoir que chaque variable hydrologique n'est corrélée avec elle-même uniquement de façon positive. Ceci signifie que le mouvement de la série hydrologique est identique pour chaque variable analysée, c'est-à-dire dans notre cas, deux fois le débit «qq» et deux fois les précipitations «pp».

Après s'être penché sur ces deux premiers graphiques, nous pouvons maintenant

analyser le graphique de la covariance entre les précipitations et le débit. Sur celui-ci, nous observons que le premier pic de la fonction de covariance est atteint pour un décalage compris entre 10 et 12 jours. A la suite de ce-dernier suit une première forte décorrélation jusqu'à une soixantaine de jours, c'est-à-dire l'équivalent de 3 mois. Puis la fonction de covariance croit de nouveau pour un second maxima moins fort que le premier. A la suite de cette deuxième corrélation la fonction décroît fortement.

Si l'on s'intéresse aux délais négatifs, nous remarquons que la fonction devient négative pour des délais de 60 jours. Ceci signifie qu'il y a corrélation négative ou encore anti-corrélation. Ces valeurs négatives correspondent bien aux délais correspondant au minimum de la fonction de covariance entre les deux pics.

De ces observations nous concluons qu'il existe donc une forte corrélation entre les deux variables hydrologiques précipitations et débit dont celle-ci est maximale pour un délai de 10 à 12 jours. Ce délai étant perceptible pour des valeurs positives, nous en déduisons la relation de causalité suivante : les précipitations ont une avance de 10 à 12 jours sur les débits. C'est donc la variable précipitation qui est une cause du débit.

Ensuite, nous avons montré qu'il existe une durée pour laquelle les variables sont les moins corrélées sur une période hydrologique cohérente (i.e approximativement 100 jours). Cette période atteint une valeur de 60 jours soit 2 mois approximativement. Le deuxième de corrélation aux alentours de 90 jours (du troisième mois) indique un regain de corrélation entre ces deux variables. L'ensemble des explications peuvent être mise en relation avec la nature karstique des variables hydrologiques analysées.

Le premier pic que l'on observe correspond certainement à la réaction rapide du bassin versant karstique. L'eau passe essentiellement dans les fractures du bassin versant. Le second pic correspond à la réponse lente du bassin. Ce pic permet d'estimer la réponse poreuse du bassin versant. Les deux échelles de temps sont nettement perceptibles sur ces deux pics. La réponse rapide avec le premier pic et la réaction du sous sol non fracturé plus lente dans le milieu poreux.

*The end...*