



Les crues de la Vienne à Nouâtre  
Evaluation probabiliste robuste

Rapport remis à la

Société COSEA

par la

Société de Calcul Mathématique SA

*en application de votre commande, notifiée le 14 août 2013*

*rédaction : Bernard Beauzamy, avec la collaboration de :*

*Sophie Davin (chef de projet), Miriam Basso, Damien Raffanel, Olga Zeydina, Philippe Crouzet*

## Résumé Opérationnel

La question posée est l'évaluation de la probabilité de crues extrêmes pour la Vienne, juste après son confluent avec la Creuse.

On dispose pour cela de relevés, plus ou moins anciens, plus ou moins précis, plus ou moins corrects, faits en un certain nombre de stations, dont les périodes d'activité sont différentes.

\*\*\*\*\*

La méthode communément utilisée est l'ajustement aux données existantes (sans tenir compte de leur imprécision) d'une "loi de Gumbel" ; c'est cette méthode qui est utilisée dans le document [Atlas]. La loi de Gumbel n'est utilisée que parce qu'elle ne fait intervenir qu'un petit nombre de paramètres et par conséquent l'ajustement est facile. Mais cette loi est entièrement dépourvue de signification physique. En outre, l'ajustement est fait à partir des points les plus nombreux (c'est-à-dire les crues modérées) et son extrapolation aux crues extrêmes est hasardeuse ; elle est particulièrement mauvaise pour le document [Atlas]. Ajoutons pour terminer que cet ajustement, dans le document [Atlas], est fait après passage aux logarithmes itérés, méthode qui génère des incertitudes très importantes et cette façon de procéder n'est pas recommandée par les spécialistes du domaine.

\*\*\*\*\*

A cette méthode d'ajustement, nous substituons des méthodes probabilistes robustes, que nous avons développées dans le cadre de contrats avec la Caisse Centrale de Réassurance. La seule hypothèse que requièrent ces méthodes est celle-ci : plus la crue est importante, plus sa probabilité est faible. Cette hypothèse est réaliste pour les phénomènes extrêmes.

Voici les conclusions obtenues en ce qui concerne les crues de la Vienne à Nouâtre :

débit (m <sup>3</sup> /s)	proba rencontrer au moins une fois en 100 ans	durée de retour (ans)
2700-2800	0,46	160
2800-2900	0,38	210
2900-3000	0,31	270
3000-3100	0,25	340
3100-3200	0,20	440
3200-3300	0,16	580
3300-3400	0,12	800
3400-3500	0,06	1760

\*\*\*\*\*

Travaillons maintenant sur des données récentes (quarante ans), les seules dont on dispose pour toutes les stations en même temps. Nous avons voulu savoir d'où provenaient les fortes crues de la Vienne à Nouâtre : est-ce de la Vienne elle-même (station d'Ingrandes, juste en amont du confluent) ? Est-ce de la Creuse (station de Leugny, juste en amont du confluent), ou bien des deux à la fois ? Le document [Atlas] affirme qu'il y a très peu de différences entre les deux rivières ; elles auraient leurs crues en même temps. Mais ceci est contredit par les données disponibles.

Bien évidemment, le débit à Nouâtre est la somme des débits à Ingrandes et Leugny. Mais les données disponibles montrent qu'il n'y a de forte crue à Nouâtre que s'il y a simultanément forte crue à Ingrandes et Leugny, avec des débits à peu près égaux pour les deux (en d'autres termes, la Vienne amont seule, ou bien la Creuse seule, sont incapables de donner une forte crue à Nouâtre).

Dans un premier temps, nous travaillons sur les données de débit maximum, extraites de la Banque Hydro.

Elles montrent que, pour les fortes crues (supérieures à 700 m<sup>3</sup>/s), il y a beaucoup moins de concomitances que ne l'affirme le document [Atlas], qui s'est probablement fondé sur l'ensemble des crues.

Pour les fortes crues, nous avons 51 enregistrements au total, 11 concomitances (représentant 22 enregistrements) et donc 29 enregistrements isolés : l'une des rivières est en crue, l'autre non.

Sur les 11 concomitances, il y en a 5 où la Vienne est plus forte et 6 où c'est la Creuse. Sur les 29 enregistrements isolés, 17 concernent la Vienne (la Creuse n'a pas eu de crue) et 12 concernent la Creuse (la Vienne n'a pas eu de crue).

Pour les fortes crues, on constate donc qu'il n'y a pas de relation claire entre la Vienne et la Creuse ; leurs crues ne sont pas automatiquement simultanées, loin de là.

Ces deux phénomènes expliquent bien pourquoi la durée de retour de crues extrêmes pour la Vienne à Nouâtre est très grande (439 ans pour la tranche 3100-3200 m<sup>3</sup>/s) : pour qu'il y ait une crue extrême à cette place, il faut une crue extrême pour la Vienne à Ingrandes, pour la Creuse à Leugny, et il faut que ces crues extrêmes aient lieu en même temps, ce qui n'est pas si fréquent. On constate par exemple que, depuis 1974 (date des enregistrements communs), les maxima à Ingrandes et à Lagny n'ont pas eu lieu en même temps.

Voici l'histogramme conjoint de la Vienne à Ingrandes et de la Creuse à Leugny :

Vienne\Creuse	900-1000	1000-1100	1100-1200	1200-1300	1300-1400
900-1000	0	1	1	1	0
1000-1100	0	0	2	0	0
1100-1200	0	2	0	1	0
1200-1300	0	0	1	0	0
1300-1400	0	1	0	1	0

Ce tableau se lit de la manière suivante : les débits de la Vienne sont en ligne et ceux de la Creuse en colonne. La case  $(i, j)$  comporte le nombre d'occurrences où le débit  $i$  pour la Vienne et le débit  $j$  pour la Creuse ont été observés simultanément. Par exemple, le nombre 2 sur la ligne 1100-1200 et la colonne 1000-1100 signifie que, à deux reprises, on a observé Vienne = 1100-1200 et Creuse = 1000-1100.

La dispersion des nombres dans le tableau montre bien que les rivières sont peu liées pour les crues extrêmes. Si elles l'étaient, tous les nombres seraient proches d'une diagonale. Or ici on voit, par exemple, que pour Vienne = 1300-1400, on a des observations de la Creuse à 1000-1100 et à 1200-1300. On voit aussi que pour les grandes valeurs de la Creuse on peut avoir plusieurs valeurs pour la Vienne, et inversement.

Dans un deuxième temps, nous avons voulu valider les données de débit maximum de la Banque Hydro, en les comparant aux données de débit journalier, en principe plus fiables. Voir Annexe 3 pour les définitions. Cette validation se révèle satisfaisante pour les trois stations qui nous intéressent.

Sur les 73 crues validées sur la Vienne à Nouâtre, c'est-à-dire les situations où le débit dépasse les 791 m<sup>3</sup>/s (quantile 80 des débits maxima mensuels instantanés à Nouâtre), nous obtenons les résultats suivants :

	Creuse à Leugny	Vienne à Ingrandes	simultanément
Nombre de crues	55	62	44
Pourcentage	75	85	60

On constate ainsi, encore une fois, que les crues simultanées de la Creuse à Leugny et de la Vienne à Ingrandes ne sont pas si fréquentes.

Même si ces données sont récentes, et ne mentionnent pas la grande crue de 1792, elles montrent clairement deux choses : pour avoir une crue extrême de la Vienne en aval, il faut simultanément des crues extrêmes de la Vienne en amont et de la Creuse, et la concomitance de ces crues extrêmes est rare. Ceci se comprend assez facilement : une crue extrême représente de fortes précipitations pendant une durée suffisante, et les très fortes précipitations sont par essence très localisées.

Les méthodes probabilistes que nous développons permettent de rendre compte de l'imprécision sur les durées de retour, beaucoup plus importante lorsque les données n'ont que cinquante ans. Une méthode probabiliste (à la différence d'un ajustement) s'accommode de situations où les données sont rares, et sait en tirer les conclusions en terme de précision du résultat : moins les données sont nombreuses et plus le résultat est imprécis.

### **Organisation du présent rapport :**

- La première partie, de nature probabiliste, est consacrée à l'estimation des crues extrêmes à Nouâtre ;
- La seconde partie, de nature déterministe, compare les données des différentes stations ;
- Les troisième et quatrième parties sont analogues à la première et sont consacrées aux crues en amont du confluent.

Nous donnons en outre trois annexes :

- Annexe I : La loi de Gumbel et son utilisation dans le document [Atlas] ;
- Annexe II : Les méthodes mathématiques pour les phénomènes extrêmes ;
- Annexe III : Vérification des données pour chaque station.

## Table des matières

I.	Introduction .....	10
II.	L'objectif .....	10
III.	La situation de la station de Nouâtre .....	11
IV.	Saisonnalité.....	13
	Première partie .....	14
	La Vienne à Nouâtre.....	14
I.	Difficultés liées aux données .....	15
I.	L'histogramme des crues à Nouâtre.....	17
II.	Les débits douteux.....	19
III.	Impact des travaux.....	21
IV.	Situations extrêmes.....	22
A.	Définition de la situation extrême.....	22
B.	Critique de la méthode habituellement utilisée .....	22
C.	Approche théorique.....	22
D.	Description théorique de la méthode.....	23
E.	Formulation mathématique .....	23
F.	Calculs explicites.....	24
G.	Les résultats .....	25
H.	Application à l'estimation absolue des probabilités des phénomènes extrêmes .	26
1.	Estimation de la probabilité $P\{X \geq 1900\}$ .....	27
2.	Probabilités absolues par jour.....	28
3.	Modèle "bons et mauvais jours".....	28
4.	Résultats par classe.....	29
5.	Le concept de "durée de retour".....	29
I.	Variante : seuil à 2 000 m <sup>3</sup> /s.....	30
	Seconde Partie .....	32
	Concomitances .....	32
I.	Introduction .....	33
II.	La Vienne à Ingrandes et Châtellerault.....	33
A.	Méthodologie.....	33
B.	Concomitance.....	34
C.	Comparaison des niveaux.....	34
III.	La Vienne à Nouâtre et la Vienne à Ingrandes .....	35
A.	Concomitances .....	35

B.	Comparaison des niveaux.....	35
IV.	La Creuse à Leugny et la Vienne à Nouâtre .....	35
A.	Concomitances .....	35
B.	Comparaison des niveaux.....	35
V.	La Vienne à Ingrandes et la Creuse à Leugny .....	36
A.	Concomitances .....	36
B.	Comparaison des niveaux.....	36
C.	Cas des crues fortes .....	36
Troisième Partie.....		39
La Vienne à Ingrandes.....		39
I.	Introduction .....	40
II.	Données .....	40
III.	Crues extrêmes.....	41
Quatrième Partie .....		43
La Creuse à Leugny .....		43
I.	Introduction .....	44
II.	Les données .....	44
III.	Crues extrêmes.....	45
Références .....		47
Annexe I .....		48
La loi de Gumbel et son utilisation dans le document [Atlas] .....		48
A.	Présentation générale.....	48
B.	La loi de Gumbel.....	48
C.	Ajustement de la loi aux données.....	50
D.	L'utilisation dans le document [Atlas] .....	52
Annexe II.....		54
Les méthodes mathématiques pour les phénomènes extrêmes .....		54
I.	Présentation mathématique du problème.....	54
II.	L'hypothèse mathématique de base.....	55
III.	La stationnarité de la loi.....	56
IV.	Le choix de la période de référence .....	57
V.	Les données de travail.....	57
VI.	Calcul de l'intégrale d'un monôme.....	60
A.	La méthode de Peter Robinson.....	61
B.	Une méthode rapide et totalement déterministe .....	62

C.	Cas de deux variables .....	62
D.	Cas général .....	62
VII.	Un premier changement de variable .....	64
VIII.	Un second changement de variables .....	65
IX.	Le problème explicite .....	66
X.	Calcul par récurrence .....	66
Annexe III .....		70
Vérification des données pour chaque station.....		70
I.	Disponibilité des données .....	70
II.	Comparaison des dates de crues .....	71



## I. Introduction

Le présent document est rédigé par la Société de Calcul Mathématique SA en réponse à la commande notifiée par la Société COSEA le 14 août 2013 (référence : notre proposition technique et financière datée du 06/08/2013) : évaluation des débits extrêmes pour les rivières Vienne et Creuse.

## II. L'objectif

Il s'agit de parvenir à une évaluation de la probabilité des débits extrêmes pour la Vienne, au moyen de méthodes robustes, scientifiquement indiscutables, et tenant compte des incertitudes sur les données d'entrée. Nous veillerons donc à bien détailler les données que nous utilisons et les raisonnements que nous tenons.

Le phénomène de crue se définit comme la montée des eaux hors du lit mineur (délimité par des berges franches), par suite de l'augmentation temporaire du débit du cours d'eau. Cette définition est trop imprécise pour une évaluation ; en revanche elle qualifie l'effet de la crue qui peut poser problème : l'inondation du lit majeur dans lequel de nombreuses activités peuvent être exercées et perturbées du fait de la crue, avec des dommages d'autant plus grands que la crue sera plus marquée (hauteur d'eau atteinte) et plus durable.

Une crue se caractérise ainsi par de nombreux paramètres ; il ne peut y avoir une définition simple et univoque :

- Le débit de la rivière, qui est nécessairement supérieur au débit moyen interannuel (et qui peut être plusieurs fois ce débit moyen, selon le régime hydrologique) ; il est mesuré (de manière idéale !) sur un plan vertical, mais il n'est pas constant tout au long de la crue ;
- La durée de retour de ce débit (exprimée en année ou par une probabilité d'occurrence pendant une période de temps donnée) ;
- L'inondation, que l'on peut considérer comme l'effet de la crue contre lequel on veut se prémunir ou se préparer, se caractérise par :
  - La hauteur atteinte par l'eau en un endroit donné, par exemple sous un pont ; elle dépend des caractéristique locales d'écoulement ;
  - Le débordement par submersion d'un ouvrage de protection (digue par exemple) ;
  - La superficie recouverte par l'eau (aire de la zone inondée et durée de submersion), qui est difficile à définir et varie avec le temps ;

- La durée de la crue, qui est également difficile à définir : qu'est-ce que le "retour à la normale" ?

Pour évaluer un risque ou une durée de retour, la seule grandeur qui peut être manipulée est le débit du cours d'eau, car c'est une grandeur conservative qui autorise l'analyse de valeurs en divers points du système (ici la Vienne avant la confluence avec la Creuse, la Creuse et la Vienne à l'aval par exemple). Les autres informations, comme la hauteur d'eau atteinte, sont dépendantes de facteurs locaux et largement imprévisibles.

Par exemple, pour un même débit dans une rivière avec de digues de protection, la hauteur atteinte va dépendre de la continuité de la digue, de l'ouverture volontaire de pertuis pour inonder des casiers de protection ou de la rupture accidentelle de la digue. En revanche, le débit restera à peu près constant.

Passer d'un débit à une hauteur d'eau, ou à une surface couverte, est certainement très difficile, et ne peut être fait que grâce à une connaissance du terrain sur place. En revanche, l'estimation d'une durée de retour pour un débit ne requiert que des outils mathématiques, comme nous allons le voir.

Nous travaillons donc sur un unique paramètre, qui est le débit : celui qui est retenu est le maximum journalier (maximum enregistré sur une période de 24 h). Le débit maximum est effectivement celui qui présente la relation la plus étroite avec le risque d'inondation par submersion. Il n'est pas possible sans modélisation de relier un débit de pointe dans un bief principal avec les risques d'inondation par remous dans des annexes hydrauliques (petits affluents par exemple) : ils dépendent de la pente et de la charge hydraulique dans les annexes.

Le travail présenté ici est complètement analogue avec ce qui se fait pour les températures : température maximale sur 24 h. Notre objectif est d'associer une loi de probabilité à chaque valeur de débit, compte-tenu de l'historique disponible. C'est un problème de nature purement probabiliste.

### **III. La situation de la station de Nouâtre**

La station de Nouâtre (L7000610 de la banque Hydro) est située en aval du confluent de la Vienne et de la Creuse, au pont de Nouâtre (Figure 1).

Elle mesure le débit sur un bassin versant de 19 920 km<sup>2</sup> (la Vienne draine en totalité, à sa confluence avec la Loire, 21 161 km<sup>2</sup>, soit plus de 94% de la superficie totale mesurée à Nouâtre).

Cette station est elle-même en aval du confluent avec la Creuse, qui a été munie d'un barrage (Maisons-Rouges), édifié en 1920 et effacé en avril 1999. Ce barrage, d'une hauteur de 3,80 m, avait essentiellement un impact hydromorphologique et biologique et

certainement une faible influence sur le régime des crues bien que créant un plan d'eau sur plus de 15 km cumulé de rivière.

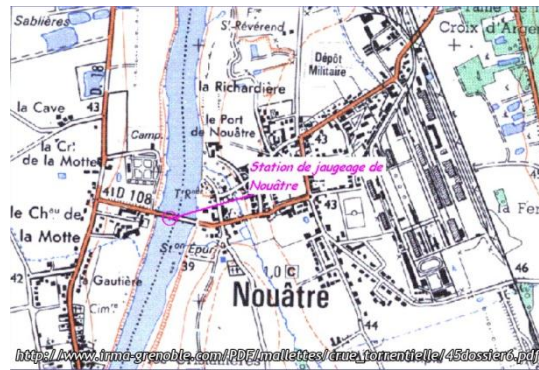


Figure 1 : Extrait de carte de position de la station hydrométrique de la Vienne à Nouâtre

Source : [http://www.donnees.centre.developpement-durable.gouv.fr/Sta\\_Jau/images/Z9999999.htm#L7000610](http://www.donnees.centre.developpement-durable.gouv.fr/Sta_Jau/images/Z9999999.htm#L7000610).

La situation globale la station et de la confluence sont présentées sur la vue qui suit :

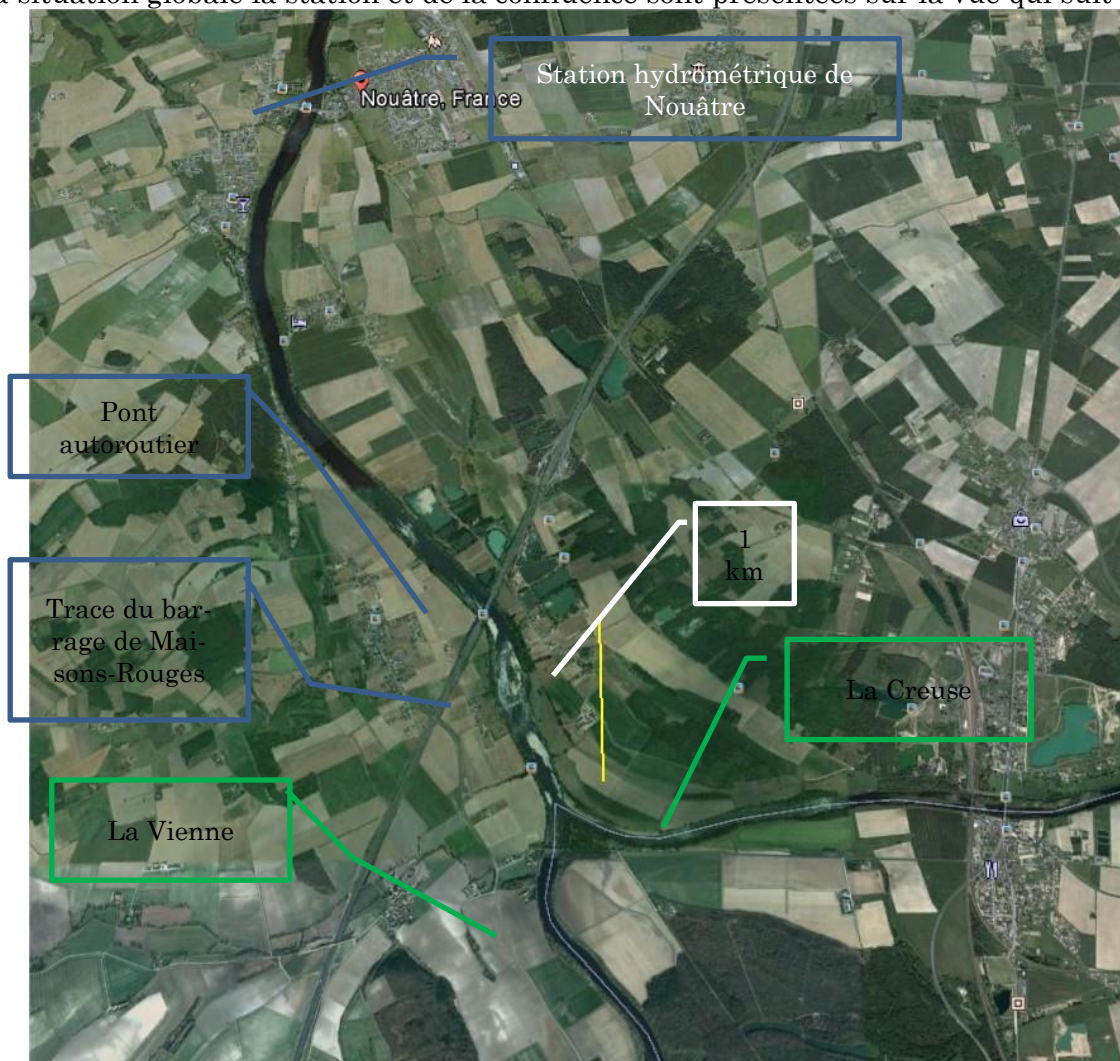


Figure 2 : Vue satellite de la région étudiée

Source : capture d'écran GoogleEarth

Les crues relevées à la station de Nouâtre peuvent donc provenir de la Vienne en amont, ou de la Creuse, ou des deux. Pour évaluer les débits attendus, nous le ferons d'abord :

- ensemble pour les deux rivières ;
- séparément pour chacune.

## IV. Saisonnalité

Nous voulons évaluer la probabilité de débits extrêmes pour une rivière. Une probabilité peut être considérée grossièrement comme le quotient de deux nombres : au numérateur, on met le nombre de cas "favorables" et au dénominateur le nombre de cas "possibles". "Favorables" signifie ici : situations où le débit est élevé et "possibles" signifie ici : toutes situations d'observation.

C'est ici que se pose le problème de la saisonnalité des crues. Pour bien comprendre ceci, faisons référence aux températures très élevées : si on veut déterminer leur probabilité, il ne sera pas honnête de considérer toute l'année, car la probabilité de telles températures en hiver est infime (jamais observées). Il faut donc se restreindre à une plage limitée de mois, par exemple de mai à octobre.

Le même raisonnement vaut pour les crues : s'il y a des mois où les crues ne se produisent jamais, il faut les éliminer du décompte des situations possibles (le dénominateur de la probabilité), sans quoi nous allons minorer celle-ci d'une manière factice et malhonnête.

A partir des fichiers remis par COSEA (fichiers provenant de la banque "Hydro"), nous avons déterminé les mois où les crues se sont produites, sur les diverses stations :

saisonnalité												
mois	1	2	3	4	5	6	7	8	9	10	11	12
Le Clain à Dissay	X	X	X	X	X					X	X	X
La Creuse à Ciron	X	X	X	X	X	X					X	X
La Creuse à Leugny	X	X	X	X	X	X					X	X
La Gartempe à Montmorillon	X	X	X	X	X	X	X				X	X
La Vienne à Châtelleraut	X	X	X	X	X	X	X			X	X	X
La Vienne à Ingrandes	X	X	X	X	X	X	X			X	X	X
La Vienne à Nouâtre	X	X	X	X	X	X					X	X

Tableau 1 : la saisonnalité des crues selon les stations

Nous en concluons que les crues peuvent se produire dix mois sur douze. Nous sommes sur ce point en accord avec le document [Atlas] qui concluait que les fortes crues peuvent se produire à peu près n'importe quand. Mais notre restriction "dix mois sur douze" est plus précise et majore les probabilités (nous nous pénalisons en calculant ainsi).

Les calculs de saisonnalité, pour chaque station, se trouvent dans le second onglet ("sheets(2)") du fichier de données propre à cette station.

## **Première partie**

### **La Vienne à Nouâtre**

## I. Difficultés liées aux données

Les données récentes disponibles sont celles de la banque Hydro. Celle-ci procure des débits maxima journaliers et des débits moyens journaliers. Dans tous les cas, pour les stations concernées, le débit n'est pas une mesure, mais le résultat d'un calcul à partir d'une observation de la hauteur d'eau et de l'application d'une "courbe de tarage" expérimentale reliant chaque débit mesuré à la hauteur observée simultanément.

Une courbe de tarage est faite à partir de mesures de débit (jaugeages) effectués à une certaine date, soit systématique (bas débits) soit à l'opportunité d'une crue. Compte-tenu des changements dans la morphologie d'un cours d'eau (par exemple remaniements du lit lors de crues, dépôts en basses eaux, travaux, etc.) il ne peut pas y avoir une relation totalement univoque entre débit et hauteur ; évidemment les stations sont situées en raison de la meilleure stabilité possible de la section.

La plus ou moins grande confiance des débits publiés dépend donc de la qualité et du nombre de jaugeages (les points sur la courbe de tarage) et leur proximité pour des débits proches.

Les débits récents ont tous été mesurés avec la méthode résumée ci-dessus, les valeurs publiées étant réajustées au fur et à mesure du complément de la courbe de tarage. La publication des débits comporte donc des mentions "validé", "provisoire" ou "douteux" selon la confiance que le gestionnaire accorde à la valeur publiée, notamment en raison des possible détarages de la station au cours du temps.

Seuls les débits publiés depuis la mise en service de la station peuvent être considérés comme faisant partie d'une série homogène du point de vue de la mesure ; en revanche des données historiques antérieures sont nécessairement reconstituées sans que l'on puisse les qualifier exactement. C'est le cas des valeurs des 18<sup>ème</sup> et 19<sup>ème</sup> siècle et du début du 20<sup>ème</sup> utilisées ici.

Dans le cas présent, seules les données extrêmes journalières ont été prises en compte. Il peut y avoir des différences considérables entre les débits moyens journaliers et le débit maximal, principalement pour les hautes eaux ; par exemple si une crue extrême de courte durée se produit pendant la nuit, elle est mécaniquement écrêtée par le calcul de moyenne ! A l'Annexe III, nous montrons comment valider les données de débit max, en les comparant aux données de débit moyen.

Pour la station de Nouâtre, par exemple, les valeurs extrêmes du débit journalier instantané, du débit moyen et de la hauteur n'ont pas été observées simultanément.

maximums connus (par la banque HYDRO)

débit instantané maximal (m <sup>3</sup> /s)	2480.	1 janvier 1962 00:00
hauteur maximale instantanée (mm)	8610	8 janvier 1982 12:41
débit journalier maximal (m <sup>3</sup> /s)	2300.	14 janvier 1962

Figure 3 : maxima à Nouâtre

Source : Banque Hydro <http://www.hydro.eaufrance.fr/presentation/procedure.php>

La différence significative de dates entre le débit instantané et moyen journalier pourrait poser des problèmes dans l'analyse de simultanéité des crues de la Vienne et de la Creuse, si le temps de trajet de l'onde de crue était significatif.

Heureusement, les distances entre les stations sont acceptables :

- De Leugny (Creuse) à la confluence, environ 22 km ;
- De Ingrandes (Vienne) à la confluence, environ 17 km ;
- De la Confluence à Nouâtre, environ 6 km.

En admettant en crue une vitesse de l'eau de 2 à 4 m/s, on a donc des temps de trajet de la crue de l'eau (l'onde de niveau pourrait se propager plus vite) variant entre une heure et demie et 3 ou 4 heures pour la station la plus distante et la confluence. Nous considérons donc les données comme valides, et, pour nous, une "concomitance" sera le fait que les événements ont eu lieu le même jour, ou bien avec une journée d'intervalle.

Le fichier de données sur lequel nous travaillons est Vienne\_NouatreComplete.xlsx. Il comporte dans le premier onglet les données brutes, provenant de la banque "Hydro". Nous disposons de 69 relevés, depuis la grande crue de 1792 jusqu'à aujourd'hui.

Les difficultés sont les suivantes :

- Pour la grande crue de 1792, c'est certainement une hauteur d'eau (approximative) qui a été enregistrée ; la conversion en débit est douteuse ;
- Même chose pour les quatre données du 19<sup>ème</sup> siècle ;
- Les données du 19<sup>ème</sup> siècle concernent des crues ayant un débit d'au moins 2200 m<sup>3</sup>/s ; on n'est pas sûr qu'il n'y a pas eu de crues dans l'intervalle 1900-2200 ;
- Il est certain que les crues faibles et moyennes du 19<sup>ème</sup> siècle n'ont pas été enregistrées, mais notre méthode de calcul n'utilise pas cette information ; elle utilise seulement les données extrêmes ;

- La grande crue de 1792 se situe pendant le "petit âge glaciaire" (entre 1300 et 1860) ; il est généralement admis que les précipitations neigeuses étaient plus importantes qu'aujourd'hui, et donc les crues plus importantes (mais cette crue a eu lieu le 1<sup>er</sup> juillet !)
- Il y a eu divers travaux au cours du 20<sup>ème</sup> siècle, qui ont pu modifier le cours de la rivière, et donc (en principe) altérer le débit de crue. Ceci est étudié plus loin.

Nous tenons compte de l'imprécision des données par notre méthode (voir plus bas). Pour les deux derniers points, on peut admettre que les résultats que nous obtenons sont conservateurs : les durées de retour des crues sont plus longues que nous ne l'annonçons (et donc les crues encore moins probables).

## I. L'histogramme des crues à Nouâtre

A partir de l'historique, nous répartissons les débits par tranche de 100 m<sup>3</sup>/seconde.

La valeur minimale est 385 et la valeur maximale est 3 375 (l'unité sera constamment le mètre cube par seconde). La plus forte observation jamais faite tombe donc dans la classe 3300-3400, mais nous introduisons par principe une classe de plus : 3400-3500, car rien ne dit qu'une crue encore plus forte ne puisse jamais être observée. La classe la plus basse est 300-400 ; nous disposons donc de 32 classes ; le résultat de la répartition se lit dans le tableau suivant :

débit	nb occurrences	histo	cumul
300-400	1	0,0198	0,0198
400-500	0	0,0099	0,0297
500-600	1	0,0198	0,0495
600-700	3	0,0396	0,0891
700-800	3	0,0396	0,1287
800-900	3	0,0396	0,1683
900-1000	3	0,0396	0,2079
1000-1100	2	0,0297	0,2376
1100-1200	2	0,0297	0,2673
1200-1300	5	0,0594	0,3267
1300-1400	3	0,0396	0,3663
1400-1500	4	0,0495	0,4158
1500-1600	3	0,0396	0,4554
1600-1700	3	0,0396	0,4950
1700-1800	3	0,0396	0,5347
1800-1900	0	0,0099	0,5446
1900-2000	6	0,0693	0,6139
2000-2100	2	0,0297	0,6436
2100-2200	5	0,0594	0,7030

2200-2300	6	0,0693	0,7723
2300-2400	3	0,0396	0,8119
2400-2500	3	0,0396	0,8515
2500-2600	1	0,0198	0,8713
2600-2700	3	0,0396	0,9109
2700-2800	0	0,0099	0,9208
2800-2900	0	0,0099	0,9307
2900-3000	0	0,0099	0,9406
3000-3100	0	0,0099	0,9505
3100-3200	0	0,0099	0,9604
3200-3300	0	0,0099	0,9703
3300-3400	1	0,0198	0,9901
3400-3500	0	0,0099	1,0000
total	69	1,0000	

*Tableau 1 : histogramme des débits de la Vienne à Nouâtre*

En première colonne, le débit, par tranche de 100 m<sup>3</sup>/s.

En seconde colonne, le nombre de fois où ce débit a été enregistré (le total fait 69).

En troisième colonne, la probabilité déduite du nombre d'observations. La formule de calcul est :

$$p = \frac{n + 1}{N + K}$$

où  $n$  est le nombre d'observations (colonne 2),  $N$  le nombre total d'observations (ici 69) et  $K$  le nombre de classes (ici 34).

Cette formule de calcul est démontrée dans le livre [NMP] ; elle est plus correcte que la formule habituellement employée :  $p = \frac{n}{N}$ . En effet, si une valeur n'a jamais été observée, la formule usuelle lui attribue la probabilité 0, ce qui n'est certainement pas correct. Par exemple, dans le cas présent, aucune observation n'a donné un débit maximum journalier entre 1800 et 1900, mais il serait absurde d'en déduire que ces débits sont impossibles ; simplement, notre échantillon est trop pauvre pour qu'ils apparaissent.

La somme des probabilités de la troisième colonne fait 1. La quatrième colonne donne le cumul : sommes partielles depuis la première ligne. La dernière ligne fait 1.

Voici la représentation des probabilités sous forme d'histogramme :

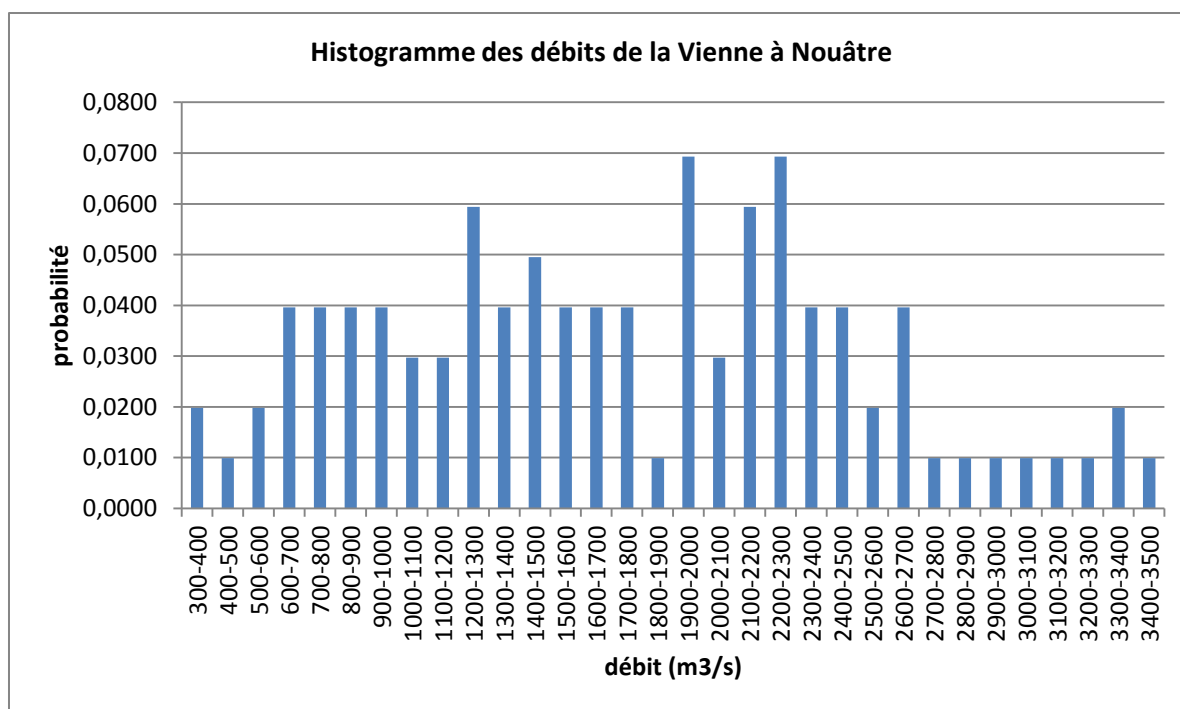


Figure 2 : l'histogramme des crues de la Vienne à Nouâtre

Cette représentation est purement factuelle : elle est obtenue à partir des données enregistrées. La simplification qui est faite : prendre des tranches de 100 m<sup>3</sup>/s est très appropriée au besoin traité ici ; car elle permet de tenir compte des incertitudes. Un débit n'est jamais calculé exactement ; on ne sait pas si l'écoulement était totalement laminaire, on n'est pas sûr de la validité de la relation entre hauteur et débit ; on ne sait pas si la courbe de tarage est correctement documentée pour la hauteur observée. Ce découpage en tranches en rend compte. Il sera en outre la base du raisonnement probabiliste qui permettra l'évaluation des probabilités pour les phénomènes extrêmes.

Les calculs sont très simples et ils sont disponibles sous forme de macros VBA dans le fichier Vienne\_NouatreComplete.xlsm (onglet "développeur").

## II. Les débits douteux

Selon les règles de la Banque Hydro, la qualificatif n'implique pas que la valeur soit erronée, mais que le gestionnaire ne dispose pas des éléments pour les déclarer "validés".

En particulier, des travaux de réfection du pont franchissant la Vienne à Nouâtre ont entraîné un dysfonctionnement de la station (juillet 2005, selon la Banque Hydro). Les valeurs douteuses en 1980 et 1981 n'ont pas pu être expliquées ; en revanche l'absence de hauteurs dans la Banque Hydro entre 1958 et 1966 est certainement liée à l'enregistrement sur papier des hauteurs et à leur traitement manuel

La qualification des valeurs est reportée dans la figure ci-dessous :

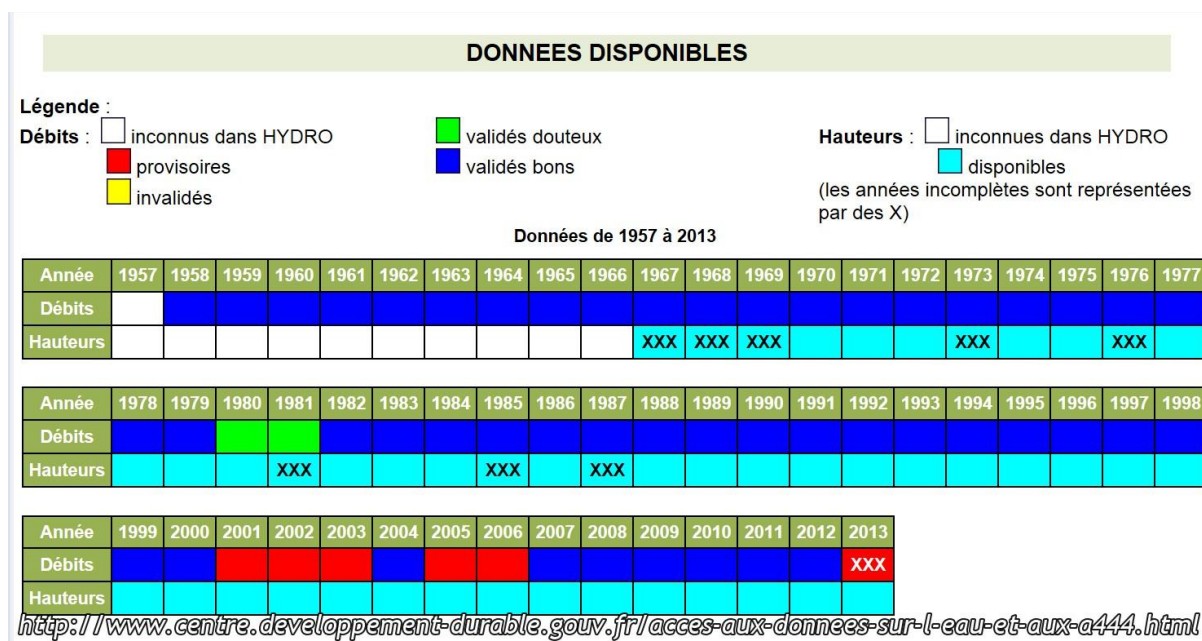


Figure 4: Qualification des données de la Banque Hydro à Nouâtre

Source : Banque Hydro

Dans le relevé des débits, ceux-ci sont présentés comme "douteux" pour cinq occurrences, à savoir :

- 29/12/1999 : 2120 m<sup>3</sup>/s
- 22/04/2008 : 1450 m<sup>3</sup>/s
- 25/01/2009 : 1200 m<sup>3</sup>/s
- 02/01/2010 : 791 m<sup>3</sup>/s
- 07/12/2010 : 643 m<sup>3</sup>/s

Comme nous ne nous intéressons qu'aux fortes crues (de débit au moins égal à 1 900 m<sup>3</sup>/s, voir ci-dessous), une seule valeur est concernée, à savoir la première. Mais la modification d'une seule valeur dans le tableau d'occurrences n'aura pas d'impact significatif. Nous pouvons donc considérer les données comme suffisamment fiables, indépendamment du fait que des travaux ont pu être réalisés pour réguler les cours des différents fleuves.

### III. Impact des travaux

Les bassins de la Vienne et de la Creuse ont été équipés en barrages hydro-électriques et à fonctions multiples au cours du 20<sup>ème</sup> siècle.

Sur la Creuse, les ouvrages d'Eguzon (1926, 58 hm<sup>3</sup>, La Roche au Moines, 1932, 4 hm<sup>3</sup> et la Roche Bat-l'Aigue (1977, petit volume) sont susceptibles de modifier les caractéristiques d'écoulement. Par exemple, la vidange d'Eguzon en 1988 a été suivie d'un remplissage au détriment des crues.

Sur la Vienne, il existe un grand nombre d'ouvrages dont trois majeurs : Lavaud-Gelade (21 hm<sup>3</sup>) et St Marc (20 hm<sup>3</sup>) sur le Taurion, Vassivières (106 hm<sup>3</sup>) sur la Maulde.

Source : <http://www.hydro.eaufrance.fr/presentation/procedure.php>

Le régime de la Vienne est très artificialisé, mais celle-ci est surtout sensible en étiage. Depuis fin 1997, un soutien d'étiage est réalisé pour assurer le refroidissement de la centrale nucléaire de Civaux, située en amont de Châtellerault et en aval de Nouâtre. Ainsi, à Ingrandes, le QMNA<sub>5</sub> (débit mensuel minimal annuel qui se produit en moyenne une fois tous les 5 ans) était pour la période 1918-1997 de 20,9 m<sup>3</sup>/s, il est de 23,4 m<sup>3</sup>/s depuis 1997.

Le soutien d'étiage consiste en pratique à stocker de l'eau de crue pour la restituer en étiage. En fait, les volumes mis en jeu ne sont pas du même ordre de grandeur. Une journée de crue à 2 000 m<sup>3</sup>/s implique un volume de 172 hm<sup>3</sup> qui ne peut être significativement altéré que par des remplissage préalables de grands barrages. Toutefois, pour des raisons évidentes de sécurité des ouvrages, la fraction prélevée sur une crue reste faible pour que le niveau ne monte pas trop vite. De plus, ceux-ci sont situés très en amont, où les débits de crue sont très largement moindres.

En première approximation, on peut donc considérer que l'influence des remplissages de barrage sur les crues les plus fortes est négligeable, au sens des calculs présentés ici.

Il est possible que les changements d'occupation des terres, notamment le couvert forestier, qui a largement augmenté au cours du siècle passé, aient une influence sur la constitution des crues. Aucune donnée n'est disponible ; certains experts considèrent que les conditions ayant présidé à la crue de 1792 et aux suivantes ne se présenteraient plus aujourd'hui de manière similaire et que les mêmes précipitations ne résulteraient plus en le même débit. Nous n'avons pas de compétence sur ces questions. On peut estimer que, en faisant les calculs avec le débit de l'époque, nous nous pénalisons : les probabilités que nous obtenons sont trop élevées et les durées de retour que nous indiquons sont trop courtes, ce qui constitue un facteur de sécurité.

## IV. Situations extrêmes

### A. Définition de la situation extrême

On s'intéressera, dans la suite, aux situations de très fort débit. L'analyse qui vient d'être faite rend compte de tous les débits, à partir de 300 m<sup>3</sup>/s, mais nous souhaitons nous intéresser aux fortes crues. Vu l'aspect de l'histogramme ci-dessus, nous considérerons dans l'avenir les débits  $\geq 1900$  m<sup>3</sup>/s. Cela revient à considérer les trente plus fortes crues connues. On peut certainement faire varier ce seuil : une variante avec un seuil fixé à 2 000 m<sup>3</sup>/s est donnée plus loin.

Deux remarques sont essentielles :

- Lorsqu'on cherche à déterminer la probabilité de crues extrêmes, il est nécessaire de fixer un seuil ; si on considère toutes les crues, même minimales, on fausse le résultat, car les crues minimales se produisent beaucoup plus souvent.
- Si on fixe le seuil trop haut, on ne dispose plus de données en nombre suffisant. Par exemple, si on fixe le seuil à 3 300 m<sup>3</sup>/s (la plus forte crue jamais observée), la seule estimation que l'on puisse faire est  $p = \frac{n+1}{N+K} = \frac{2}{69+32} = \frac{2}{101} \approx 0.02$ , mais cette estimation n'est pas correcte, parce qu'elle n'utilise pas les observations portant sur les classes précédentes, comme nous allons le voir.

### B. Critique de la méthode habituellement utilisée

Cette critique est donnée en Annexe I.

### C. Approche théorique

On se fixe donc le seuil de 1 900 m<sup>3</sup>/s, et on cherche à évaluer la probabilité de chacune des classes au-delà de ce seuil 1900-2000, ..., 3400-3500. Chacune de ces probabilités est inconnue et donc sera elle-même traitée comme une variable aléatoire. Nous notons  $X_{19}, \dots, X_{34}$  la variable aléatoire qui caractérise la probabilité de la classe correspondante.

Il s'agit de probabilités conditionnelles : sachant que la crue dépasse 1900 m<sup>3</sup> par seconde, quelle est la probabilité qu'elle tombe dans l'une des 16 classes 1900-2000, ..., 3400-3500 ? C'est la répartition entre ces 16 classes qui nous intéresse ici, et la somme de ces 16 probabilités doit faire 1.

La méthode que nous employons a été développée par la SCM à partir de 2009 dans le cadre de contrats avec la Caisse Centrale de Réassurance. Elle est détaillée en Annexe II. Les principales étapes ont été les suivantes :

- 2009 : Mise en œuvre d'une méthode pratique d'implémentation, à base de tirages de Monte-Carlo, par Peter Robinson (vainqueur du prix offert conjointement par la CCR et la SCM)

[http://www.scmsa.eu/RMM/ART\\_2010\\_Peter\\_Robinson\\_Efficient\\_Integration.pdf](http://www.scmsa.eu/RMM/ART_2010_Peter_Robinson_Efficient_Integration.pdf)

- 2013 : Définition d'une méthode explicite et déterministe, par la SCM.

Les travaux préliminaires sont décrits dans la fiche de compétences "risques" de la SCM : [http://scmsa.eu/fiches/SCM\\_risques.pdf](http://scmsa.eu/fiches/SCM_risques.pdf), avec des liens vers les articles scientifiques correspondants.

#### *D. Description théorique de la méthode*

Elle est décrite de manière détaillée en Annexe II.

L'unique hypothèse que fait notre méthode est que les probabilités de crues doivent être décroissantes avec leur amplitude : plus le débit est fort, plus la probabilité est faible, ce qui est certainement conforme à l'intuition. Ceci n'est certainement pas vrai pour les crues minimales : on voit ainsi pourquoi il faut fixer un seuil. Inversement, ceci explique pourquoi considérer uniquement la crue maximale n'est pas correct : on lui donne la probabilité  $\frac{2}{100}$ , alors qu'on donnera la probabilité  $\frac{1}{100}$  à une crue dans la gamme 3200-3300, qui ne s'est jamais rencontrée.

#### *E. Formulation mathématique*

Nous avons à déterminer la loi conjointe de variables aléatoires  $X_{19}, \dots, X_{34}$  sous les hypothèses :

$$X_{19} + \dots + X_{34} = 1$$

$$X_{19} \geq \dots \geq X_{34} \geq 0$$

et connaissant le nombre d'observations de chacune des classes de crue, à savoir :

débit	nb occurrences
1900-2000	6
2000-2100	2
2100-2200	5
2200-2300	6
2300-2400	3
2400-2500	3
2500-2600	1
2600-2700	3
2700-2800	0
2800-2900	0
2900-3000	0
3000-3100	0
3100-3200	0
3200-3300	0
3300-3400	1
3400-3500	0

Tableau 3 : le nombre d'occurrences de chaque classe

Nous avons donc 16 classes et 30 observations. On notera  $n_{19} = 6$  le nombre d'observations correspondant à la classe 1900-2000, et ainsi de suite pour les autres.

### F. Calculs explicites

Les calculs sont réalisés dans le fichier Excel :  
 Vienne\_Nouatre\_extreme\_seuil\_1900.xlsm  
 et les algorithmes sont programmés en VBA.

Ils requièrent le calcul d'une intégrale multiple sur un simplexe, à savoir :

$$I(n_{19}, \dots, n_{34}) = \int \dots \int_S x_{19}^{n_{19}} \dots x_{34}^{n_{34}} dx_{19} \dots dx_{34} \quad (1)$$

où  $S$  est le simplexe :

$$S = \{(x_{19}, \dots, x_{34}) ; x_{19} \geq \dots \geq x_{34} \geq 0, x_{19} + \dots + x_{34} = 1\} \quad (2)$$

Une fois ce calcul fait, la probabilité de chacune des variables aléatoires s'exprime simplement :

$$p_{19} = E(X_{19}) = \frac{I(n_{19} + 1, n_{20}, \dots, n_{34})}{I(n_{19}, n_{20}, \dots, n_{34})}$$

et plus généralement, pour  $k = 19, \dots, 34$

$$p_k = E(X_k) = \frac{I(n_{19}, \dots, n_{k-1}, n_k + 1, n_{k+1}, \dots, n_{34})}{I(n_{19}, n_{20}, \dots, n_{34})}$$

Ces formules, en effet, résultent du calcul ordinaire des lois marginales à partir de la loi conjointe.

Le calcul de l'intégrale (1) sur le simplexe (2) est extrêmement délicat en pratique, car les fonctions à intégrer sont très petites. La méthode suivie est décrite en annexe.

### G. Les résultats

Ils sont donnés par le tableau suivant :

débit	nb occurrences	proba calculée	proba observée
1900-2000	6	0,2009	0,1522
2000-2100	2	0,1480	0,0652
2100-2200	5	0,1278	0,1304
2200-2300	6	0,1104	0,1522
2300-2400	3	0,0906	0,0870
2400-2500	3	0,0745	0,0870
2500-2600	1	0,0597	0,0435
2600-2700	3	0,0490	0,0870
2700-2800	0	0,0366	0,0217
2800-2900	0	0,0284	0,0217
2900-3000	0	0,0223	0,0217
3000-3100	0	0,0175	0,0217
3100-3200	0	0,0136	0,0217
3200-3300	0	0,0102	0,0217
3300-3400	1	0,0073	0,0435
3400-3500	0	0,0034	0,0217
total	30	1,000	1,0000

Tableau 4 : les probabilités conditionnelles pour les crues extrêmes

La colonne 3, "probabilités calculées", comporte les résultats obtenus par la méthode ci-dessus. La colonne 4, "probabilités observées", est simplement le résultat du calcul grossier  $p = \frac{n+1}{N+K}$  avec ici  $N = 30$  (nombre total d'observations) et  $K = 16$  (nombre de classes).

Voici la présentation sous forme d'histogramme :

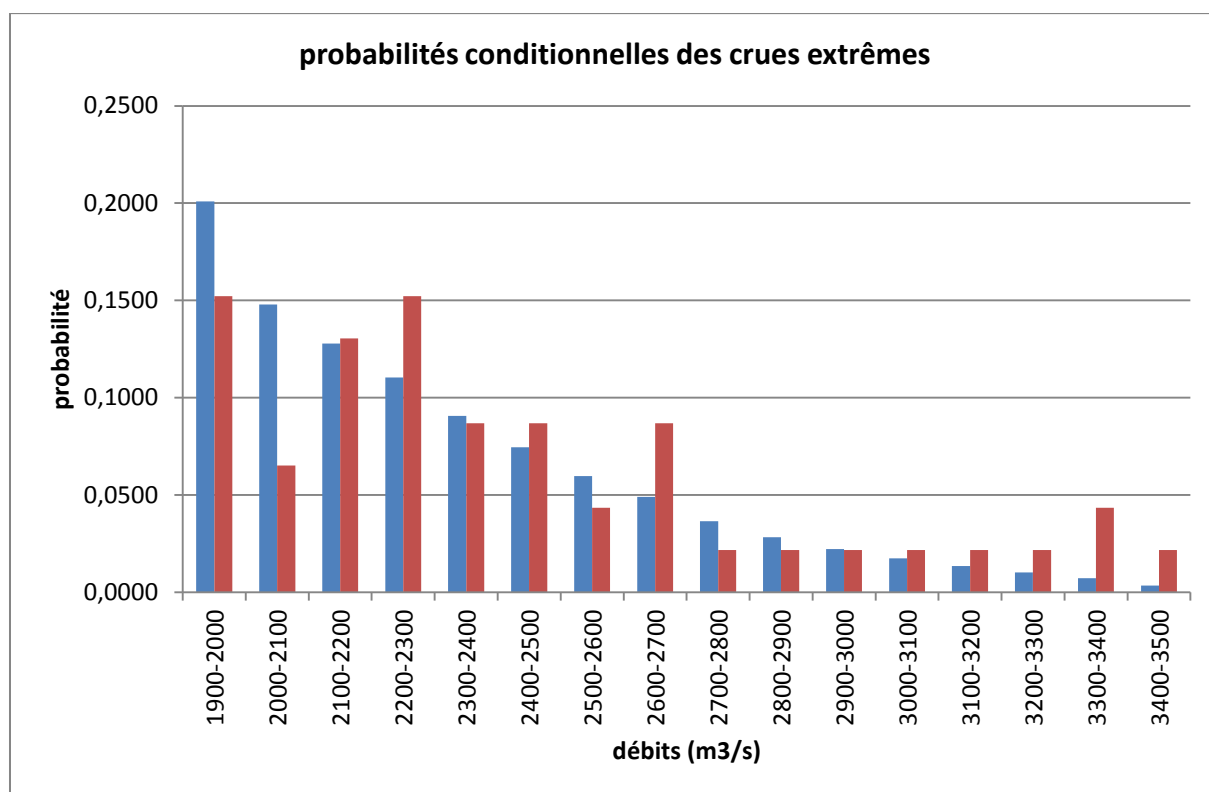


Figure 5 : les probabilités conditionnelles des crues extrêmes à Nouâtre

En rouge : probabilités observées ; elles sont erratiques, parce que l'échantillon est trop faible (pas assez d'observations) ;

En bleu : probabilités calculées ; elles sont décroissantes, comme la théorie le demande.

#### H. Application à l'estimation absolue des probabilités des phénomènes extrêmes

Répetons-le : les probabilités obtenues au paragraphe précédent sont des probabilités conditionnelles. Elles s'énoncent ainsi : sachant que le débit est supérieur ou égal à 1 900 m<sup>3</sup>/s, quelle est la probabilité qu'il soit entre 1 900 et 2 000, entre 2 000 et 2 100, etc. Nous avons calculé, pour  $k = 19, \dots, 35$  :

$$P\{100 k \leq X < 100(k+1) \mid X \geq 1900\}$$

Nous voudrions connaître la probabilité absolue de chaque classe, c'est-à-dire :

$$P\{100 k \leq X < 100(k+1)\}$$

Pour cela, selon la formule usuelle des probabilités conditionnelles, il suffit de multiplier par  $P\{X \geq 1900\}$ , puisque :

$$P\{100k \leq X < 100(k+1)\} = P\{100k \leq X < 100(k+1) | X \geq 1900\} \times P\{X \geq 1900\}$$

Il nous reste donc à déterminer la probabilité absolue de dépassement du seuil 1900 m<sup>3</sup>/s, et c'est là que la saisonnalité intervient.

### 1. Estimation de la probabilité $P\{X \geq 1900\}$

Elle se fait en utilisant la formule habituelle  $p = \frac{n+1}{N+K}$ , où ici :

$n$  est le nombre d'occurrences du dépassement du seuil, à savoir  $n = 30$  ;

$K$  est le nombre de classes, à savoir  $K = 2$  (on est au-dessous du seuil ou au-dessus) ;

$N$  est le nombre total d'observations pertinentes.

Le nombre total de jours d'observation est le nombre de journées écoulées depuis la première crue conservée en mémoire, à savoir le 01/07/1792, jusqu'à aujourd'hui.

Excel n'étant pas capable de traiter des dates antérieures à 1900, nous raisonnons comme suit :

Du 01/07/1792 au 01/07/1992, nous avons deux siècles, soit 200 ans, soit :

$$N_1 = 200 \times 365,25 = 73\,050 \text{ jours}$$

Du 01/07/1992, au 17/08/2013 (date du premier rapport), nous avons  $N_2 = 7\,717$  jours (calcul fait par Excel au moyen de la fonction DateDif).

Au total, depuis la première crue conservée en mémoire, nous avons :

$$N_1 + N_2 = 80\,767 \text{ jours.}$$

Mais tous ne sont pas pertinents pour "héberger" une crue ; nous avons vu plus haut au paragraphe "saisonnalité" que seulement dix mois sur douze l'étaient. Nous prendrons donc :

$$N = \frac{10}{12}(N_1 + N_2) \approx 67\,300 \text{ jours.}$$

Nous obtenons donc l'estimation :

$$P\{X \geq 1900\} = \frac{n+1}{N+K} = 0,00046$$

## 2. Probabilités absolues par jour

Le calcul est fait dans l'onglet 2 "probabilités absolues" du fichier :

Vienne\_Nouatre\_extreme\_seuil\_1900.xlsm

et, comme tous les autres calculs, il est programmé sous forme de macros VBA. Pour chaque classe de crue, nous obtenons donc, par jour, les probabilités suivantes :

débit	nb occurrences	proba cond	proba absolue
1900-2000	6	0,2009	9,25346E-05
2000-2100	2	0,1480	6,81654E-05
2100-2200	5	0,1278	5,8877E-05
2200-2300	6	0,1104	5,08437E-05
2300-2400	3	0,0906	4,17258E-05
2400-2500	3	0,0745	3,43125E-05
2500-2600	1	0,0597	2,74806E-05
2600-2700	3	0,0490	2,25743E-05
2700-2800	0	0,0366	1,68469E-05
2800-2900	0	0,0284	1,30589E-05
2900-3000	0	0,0223	1,02634E-05
3000-3100	0	0,0175	8,05909E-06
3100-3200	0	0,0136	6,24372E-06
3200-3300	0	0,0102	4,70238E-06
3300-3400	1	0,0073	3,36405E-06
3400-3500	0	0,0034	1,55821E-06

Tableau 6 : les probabilités absolues pour les crues de la Vienne à Nouâtre

Si  $p$  est la probabilité d'un événement par jour, la probabilité de ne pas le rencontrer, par jour, est  $1-p$  et la probabilité de ne pas le rencontrer en une année est  $q_1 = (1-p)^{365,25}$  ; la probabilité de ne pas le rencontrer en cent ans est  $q_{100} = (1-p)^{36525}$  et la probabilité de le rencontrer au moins une fois en cent ans est  $1-q_{100}$ .

## 3. Modèle "bons et mauvais jours"

Les calculs ci-dessus sont faits dans le cadre de ce que nous appelons (voir le livre [NMP]) un modèle "bons et mauvais jours". Un "bon" jour est un jour sans crue, un mauvais est un jour avec. On suppose que les crues peuvent frapper de manière aléatoire et sans mémoire : si un jour a été mauvais, le suivant n'a pas plus de raison d'être bon ou d'être mauvais. Ce type de modèle est également adopté pour les séismes (voir [NMP]) ;

les spécialistes de chaque discipline font souvent des objections, en disant qu'il s'agit de décharges d'énergie et que les jours ne peuvent être indépendants.

Si on se borne (comme le fait le mathématicien) à regarder les tableaux d'occurrences, on constate que ce modèle "bons jours, mauvais jours" décrit bien la réalité. Tout se passe comme si les crues frappaient de manière aléatoire, sans mémoire, avec une très faible probabilité. Nous n'avons donc aucune raison de le remettre en cause.

#### 4. Résultats par classe

Voici les résultats pour cent ans :

débit	proba abs	pas en un jour	pas en cent ans	au moins une fois en cent ans
1900-2000	9,2535E-05	0,999907	0,03	0,97
2000-2100	6,8165E-05	0,999932	0,08	0,92
2100-2200	5,8877E-05	0,999941	0,12	0,88
2200-2300	5,0844E-05	0,999949	0,16	0,84
2300-2400	4,1726E-05	0,999958	0,22	0,78
2400-2500	3,4312E-05	0,999966	0,29	0,71
2500-2600	2,7481E-05	0,999973	0,37	0,63
2600-2700	2,2574E-05	0,999977	0,44	0,56
2700-2800	1,6847E-05	0,999983	0,54	0,46
2800-2900	1,3059E-05	0,999987	0,62	0,38
2900-3000	1,0263E-05	0,999990	0,69	0,31
3000-3100	8,0591E-06	0,999992	0,75	0,25
3100-3200	6,2437E-06	0,999994	0,80	0,20
3200-3300	4,7024E-06	0,999995	0,84	0,16
3300-3400	3,364E-06	0,999997	0,88	0,12
3400-3500	1,5582E-06	0,999998	0,94	0,06

Tableau 7 : probabilités de crue sur cent ans

Ces résultats figurent dans l'onglet 2 du fichier  
Vienne\_Nouatre\_extreme\_seuil\_1900.xlsm

#### 5. Le concept de "durée de retour"

Elle désigne le laps de temps moyen qui s'écoule entre deux occurrences du phénomène ; beaucoup de gens ayant tendance à croire que "durée de retour" signifie périodicité, il vaudrait mieux insister sur "durée de retour moyenne" (en abrégé DRM).

A partir de la probabilité absolue d'occurrence d'un phénomène, notée  $p$ , la loi de la durée de retour est facile à calculer. La probabilité qu'il s'écoule  $k$  jours entre deux occurrences consécutives est :

$$P\{D = k\} = (1 - p)^{k-1} p$$

et la DRM, exprimée en jours, est l'espérance de cette loi :

$$DRM = \sum_{k=0}^{+\infty} k(1 - p)^{k-1} p = \frac{1}{p}$$

Si on veut l'exprimer en années, il faut diviser par 365 ; voici les résultats par classe :

débit	durée de retour (années)
1900-2000	30
2000-2100	40
2100-2200	47
2200-2300	54
2300-2400	66
2400-2500	80
2500-2600	100
2600-2700	121
2700-2800	163
2800-2900	210
2900-3000	267
3000-3100	340
3100-3200	439
3200-3300	583
3300-3400	814
3400-3500	1758

*Tableau 8 : les différentes durées de retour pour les crues à Nouâtre*

Nous avons dit plus haut que le seuil de 1900 m<sup>3</sup>/s, choisi pour qualifier une crue d'"extrême" était assez arbitraire. Nous analysons maintenant une variante.

### *I. Variante : seuil à 2 000 m<sup>3</sup>/s*

Nous reprenons les mêmes calculs, le seuil "crue importante" étant désormais fixé à 2 000 m<sup>3</sup>/s et non plus à 1 900. L'ensemble des calculs est fait dans le fichier :

Vienne\_Nouatre\_extreme\_seuil\_2000.xlsm

Voici les résultats :

débit	nb occurrences	proba condit	proba absolue	au moins 1 en 100 ans	durée de retour (ans)
2000-2100	2	0,1979	7,35E-05	0,9318	37
2100-2200	5	0,1642	6,10E-05	0,8923	45
2200-2300	6	0,1389	5,16E-05	0,8481	53
2300-2400	3	0,1117	4,15E-05	0,7802	66
2400-2500	3	0,0907	3,37E-05	0,7080	81
2500-2600	1	0,0720	2,68E-05	0,6236	102
2600-2700	3	0,0589	2,19E-05	0,5503	125
2700-2800	0	0,0437	1,62E-05	0,4474	169
2800-2900	0	0,0338	1,26E-05	0,3677	218
2900-3000	0	0,0265	9,85E-06	0,3021	278
3000-3100	0	0,0208	7,72E-06	0,2458	355
3100-3200	0	0,0161	5,98E-06	0,1962	458
3200-3300	0	0,0121	4,50E-06	0,1516	609
3300-3400	1	0,0087	3,22E-06	0,1109	851
3400-3500	0	0,0040	1,49E-06	0,0530	1838

*Tableau 9 : cas où le seuil est fixé à 2000 m<sup>3</sup>/s*

Les résultats ne sont pas significativement différents des précédents.

## **Seconde Partie**

### **Concomitances**

## I. Introduction

Dans la Première Partie, nous avons travaillé sur les données de la Vienne à Nouâtre, telles qu'elles sont fournies. Nous allons maintenant vouloir aborder la question de la répartition entre les deux rivières, Vienne et Creuse, et y a-t-il concomitance : les crues de la Vienne sont-elles simultanées aux crues de la Creuse ? Comme nous disposons de plusieurs stations sur chaque rivière, nous pouvons aussi effectuer une analyse de la qualité des données.

Une "concomitance" sera une situation où les crues ont été relevées simultanément aux deux endroits ; nous acceptons une différence de date d'une journée (pour tenir compte du fait que la crue peut s'étendre sur plusieurs jours, les relevés être pris à minuit dans un cas, à une heure du matin dans l'autre, etc.), sans prendre un délai plus grand compte tenu du temps bref de passage de l'eau d'une station à une autre exposé précédemment.

Nos conclusions ici sont assez significativement différentes de celles du document [Atlas], qui écrit page 6 :

"Les similarités importantes entre le bassin de la Creuse et celui de la Vienne amont conduisent à une assez forte concomitance entre les ondes de crue provenant de la Creuse amont et de la Vienne amont."

## II. La Vienne à Ingrandes et Châtelleraut

### A. Méthodologie

Nous comparons ici les relevés faits pour la Vienne à Ingrandes et Châtelleraut et nous nous interrogeons sur deux questions :

- Y a-t-il concomitance ? Les crues relevées à l'une des stations le sont-elles aussi à l'autre, comme ce devrait être le cas, puisque la distance entre les deux villes est de l'ordre de 7,5 km ?
- Y a-t-il comparaison des niveaux de crue ? La Vienne, à cet endroit, coule du sud au nord, et Ingrandes (altitude 44 m) est au nord de Châtelleraut (altitude 52 m), à courte distance. On s'attendrait donc à ce que les résultats soient similaires, avec un débit supérieur à Ingrandes, situé en aval de Châtelleraut.

Les réponses à ces questions sont les suivantes :

## *B. Concomitance*

Nous avons au total 101 enregistrements pour les deux stations.

De 1919 à 1938 (38 enregistrements), concomitance parfaite ;

De 1961 à 1999 (39 enregistrements), seule Ingrandes est mesurée

De 1999 à 2010 (24 enregistrements), concomitance parfaite, sauf :

05/01/2003	Vienne Ingrandes
04/02/2003	Vienne Châtelleraut

On peut penser qu'il s'agit d'une erreur de report quant au numéro du mois. Il ne s'agit pas de petites crues (759 et 825 m<sup>3</sup>/s respectivement) et elles auraient dû être enregistrées par les deux stations.

Pour 9 enregistrements, on lit "validité année = provisoire".

Voir fichier Vienne\_Châtelleraut\_Ingrandes.xlsm pour les résultats complets. Les concomitances sont en bleu.

## *C. Comparaison des niveaux*

De 1919 à 1938, les stations sont toutes deux relevées (19 enregistrements en commun), les débits sont présentés comme bons, mais Châtelleraut est en permanence supérieur à Ingrandes, ce qui est anormal.

De 1938 à 1961, aucune crue n'est enregistrée, ni pour l'un ni pour l'autre.

De 1961 à 1999, seule Ingrandes est enregistrée (39 enregistrements) et Châtelleraut ne l'est pas.

Depuis 1999, les deux stations sont enregistrées. Sur les 11 enregistrements communs, Ingrandes apparaît comme plus fort 6 fois et Châtelleraut 5 fois (mais le débit est présenté comme "douteux").

En résumé, il est possible que ces enregistrements n'aient jamais été comparés ; ils ne donnent pas une impression de qualité. L'analyse de validation que nous menons à l'Annexe III ne donne pas des résultats favorables pour Châtelleraut.

**Remarque importante :** il serait bon de présenter ces commentaires aux responsables de la banque "Hydro" et de recueillir leurs explications.

### **III. La Vienne à Nouâtre et la Vienne à Ingrandes**

Nouâtre est en aval d'Ingrandes.

#### *A. Concomitances*

Il y a 130 relevés pour les deux stations, mais seulement 36 concomitances (représentant 72 relevés en commun). Il y a donc 58 relevés isolés : l'une des stations apparaît, mais pas l'autre. De 1928 à 1928, seul Ingrandes apparaît ; de 1944 à 1959, seul Nouâtre apparaît.

#### *B. Comparaison des niveaux*

Pour toutes les situations de concomitance, le débit à Nouâtre est plus fort que le débit à Ingrandes, ce qui est satisfaisant.

Voir fichier *Vienne\_Ingrandes\_Nouatre.xlsm* pour les résultats complets. Les concomitances sont en bleu.

### **IV. La Creuse à Leugny et la Vienne à Nouâtre**

Nouâtre est en aval de Leugny, sur la même rivière, puisque le confluent est à Nouâtre.

#### *A. Concomitances*

Il y a, depuis 1965, 91 enregistrements portant sur les deux stations et 34 concomitances, correspondant à 68 enregistrements satisfaisants. Il reste 23 enregistrements isolés : l'une des stations a vu une crue, l'autre non.

#### *B. Comparaison des niveaux*

Pour toutes les concomitances, Nouâtre est supérieur à Leugny : ceci est satisfaisant.

Voir fichier *Creuse\_Leugny\_Vienne\_Nouatre.xlsm* ; les concomitances sont en bleu.

## V. La Vienne à Ingrandes et la Creuse à Leugny

Il s'agit ici de deux rivières distinctes, observées en amont de leur confluent. Il n'y a donc pas de raison a priori pour qu'elles aient les mêmes crues, ou pour que leurs niveaux soient comparables dans un sens ou dans l'autre.

Nous avons, depuis 1965, 94 enregistrements portant sur l'ensemble des deux stations.

### A. Concomitances

Il y a 24 concomitances, correspondant à 48 enregistrements communs. Il reste donc 46 enregistrements isolés, soit environ 50% des enregistrements.

### B. Comparaison des niveaux

Sur les 24 concomitances, il y a 17 situations où la Vienne est plus forte que la Creuse et 7 situations où la Creuse est plus forte que la Vienne.

En résumé, il n'y a pas de relation claire entre la Vienne et la Creuse en amont de leur confluent. Il n'est certainement pas exact que les crues de l'une se produisent en même temps que les crues de l'autre.

### C. Cas des crues fortes

Nous avons voulu voir si les choses étaient différentes pour les crues fortes : supérieures à 700 m<sup>3</sup>/s. Nous avons alors 51 enregistrements au total, 11 concomitances (représentant 22 enregistrements) et donc 29 enregistrements isolés.

Sur les 11 concomitances, il y en a 5 où la Vienne est plus forte et 6 où c'est la Creuse.

Sur les 29 enregistrements isolés, 17 concernent la Vienne (la Creuse n'a pas eu de crue) et 12 concernent la Creuse (la Vienne n'a pas eu de crue).

La conclusion est la même pour les fortes crues : il n'y a pas de relation claire entre la Vienne et la Creuse ; leurs crues ne sont pas automatiquement simultanées, loin de là. Nous allons voir ceci plus en détail.

Regroupons les fortes crues de la Vienne à Nouâtre (crues  $\geq 1900$  m<sup>3</sup>/s) et comparons-les, du point de vue des dates, avec les crues de la Vienne à Ingrandes et de la Creuse à Leugny :

Vienne à Nouâtre		Vienne à Ingrandes		Creuse à Leugny	
2010	22/03/1974	1050	21/03/1974	1170	01/03/1974
1920	25/01/1978	1140	25/01/1978	1200	01/05/1981
2170	08/01/1982	1300	08/01/1982	1200	01/01/1982
2220	18/12/1982	1230	21/12/1982	1160	19/12/1982
1990	20/03/1988	1120	19/03/1988	1040	20/03/1988
1990	16/02/1990	985	16/02/1990	1180	16/02/1990
2220	07/01/1994	1340	07/01/1994	1010	07/01/1994
2120	29/12/1999	1140	29/12/1999	1090	29/12/1999
1940	02/05/2001	964	02/05/2001	1030	02/05/2001
2030	06/03/2006	928	05/03/2006	1240	06/03/2006
2170	03/03/2007	1070	03/03/2007	1100	03/03/2007

Tableau 1 : les fortes crues pour la Vienne et la Creuse

Evidemment, le débit à Nouâtre est la somme des débits à Leugny et à Ingrandes (aux erreurs de mesure près !). Il y a probablement une erreur de report pour la première ligne en ce qui concerne la date à Leugny : on lit 01/03/1974 alors que la crue a été enregistrée ailleurs le 21 ou le 22/03 (voir à ce propos l'Annexe III).

Mais, ce qui est vraiment intéressant, c'est la constatation suivante : en cas de crue majeure à Nouâtre, les débits à Ingrandes et Leugny sont sensiblement égaux. Autrement dit, il n'y a jamais de crue majeure à Nouâtre sans que l'on ait simultanément une crue majeure à Ingrandes et une crue majeure à Leugny.

Or, les crues majeures pour la Vienne et la Creuse ne sont pas réellement simultanées. Voici le tableau des occurrences, avec les dates :

Vienne à Nouâtre		Vienne à Ingrandes		Creuse à Leugny	
2010	22/03/1974	1050	21/03/1974	1170	01/03/1974
1920	25/01/1978	1140	25/01/1978	1200	01/05/1981
2170	08/01/1982	1300	08/01/1982	1200	01/01/1982
<b>2220</b>	18/12/1982	1230	21/12/1982	1160	19/12/1982
1990	20/03/1988	1120	19/03/1988	1040	20/03/1988
1990	16/02/1990	985	16/02/1990	1180	16/02/1990
<b>2220</b>	07/01/1994	<b>1340</b>	07/01/1994	1010	07/01/1994
2120	29/12/1999	1140	29/12/1999	1090	29/12/1999
1940	02/05/2001	964	02/05/2001	1030	02/05/2001
2030	06/03/2006	928	05/03/2006	<b>1240</b>	06/03/2006
2170	03/03/2007	1070	03/03/2007	1100	03/03/2007

Tableau 2 : les occurrences des fortes crues

En rouge et caractère gras, les maxima enregistrés pour chaque rivière : on constate qu'ils ne sont pas simultanés.

Voici maintenant l'histogramme conjoint des deux rivières : Vienne à Ingrandes et Creuse à Leugny :

Vienne\Creuse	900-1000	1000-1100	1100-1200	1200-1300	1300-1400
900-1000	0	1	1	1	0
1000-1100	0	0	2	0	0
1100-1200	0	2	0	1	0
1200-1300	0	0	1	0	0
1300-1400	0	1	0	1	0

*Tableau 3 : l'histogramme conjoint de la Vienne à Ingrandes et de la Creuse à Leugny*

Ce tableau se lit de la manière suivante : les débits de la Vienne sont en ligne et ceux de la Creuse en colonne. La case  $(i, j)$  comporte le nombre d'occurrences où le débit  $i$  pour la Vienne et le débit  $j$  pour la Creuse ont été observés simultanément. Par exemple, le nombre 2 sur la ligne 1100-1200 et la colonne 1000-1100 signifie que, à deux reprises, on a observé Vienne = 1100-1200 et Creuse = 1000-1100.

La dispersion des nombres dans le tableau montre bien que les rivières sont peu liées pour les crues extrêmes. Si elles l'étaient, tous les nombres seraient situés sur une diagonale. Or ici on voit, par exemple, que pour Vienne = 1300-1400, on a des observations de la Creuse à 1000-1100 et à 1200-1300.

## **Troisième Partie**

### **La Vienne à Ingrandes**

## I. Introduction

L'intérêt de cette analyse tient au fait qu'elle porte sur la Vienne seule, avant le confluent avec la Creuse. L'objectif est ici de voir la part des crues au confluent qui est imputable à la Vienne seule.

## II. Données

Nous disposons de 70 enregistrements de débits maximaux journaliers ; le plus ancien date de 1919. Il y a donc une différence majeure avec les enregistrements de Nouâtre, qui remontent à 1792.

Le débit minimum dans ces enregistrements est de 196 m<sup>3</sup>/s et le débit maximum est de 1 510 m<sup>3</sup>/s.

Voici le tableau des crues, rangées par classes :

débit	nb occurrences	histo occurrences
100-200	1	0,0143
200-300	1	0,0143
300-400	1	0,0143
400-500	9	0,1286
500-600	5	0,0714
600-700	5	0,0714
700-800	6	0,0857
800-900	11	0,1571
900-1000	11	0,1571
1000-1100	5	0,0714
1100-1200	6	0,0857
1200-1300	1	0,0143
1300-1400	3	0,0429
1400-1500	4	0,0571
1500-1600	1	0,0143
1600-1700	0	0,0000

*Tableau 1 : les données de crues par classe*

Voici l'histogramme des occurrences (pour chaque occurrence, le nombre d'occurrences est divisé par le nombre total, ici 70) :

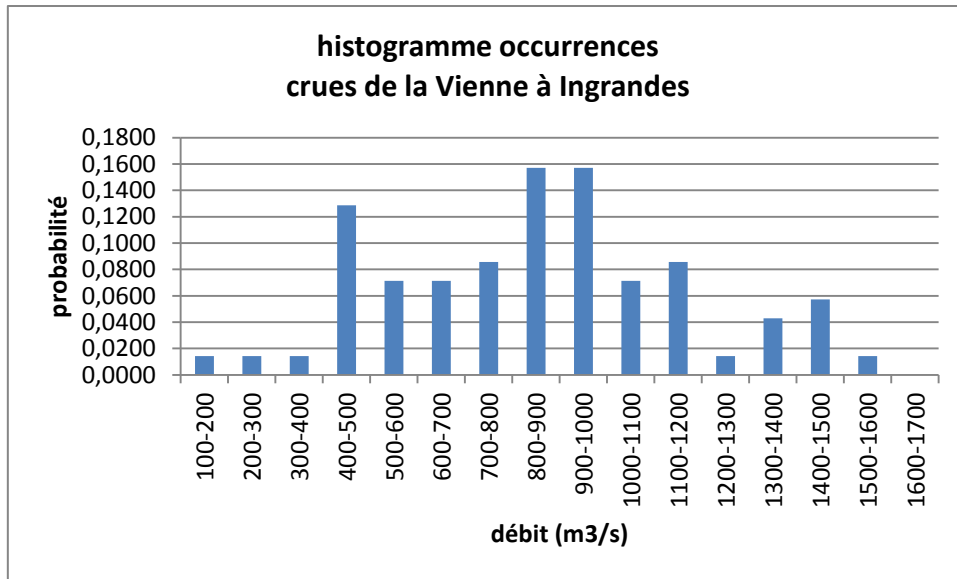


Figure 2 : l'histogramme des occurrences

Il y a une certaine bizarrerie dans cet histogramme : les faibles valeurs sont peu représentées ; cela tient probablement au fait que l'échantillon n'est pas suffisamment important.

Le seuil pour les crues fortes sera fixé à 1 000 m3/s.

### III. Crues extrêmes

On suit le même raisonnement que dans la Première Partie. On commence par calculer des probabilités conditionnelles de crues, sachant que le seuil est dépassé. Voici l'histogramme de ces probabilités conditionnelles :

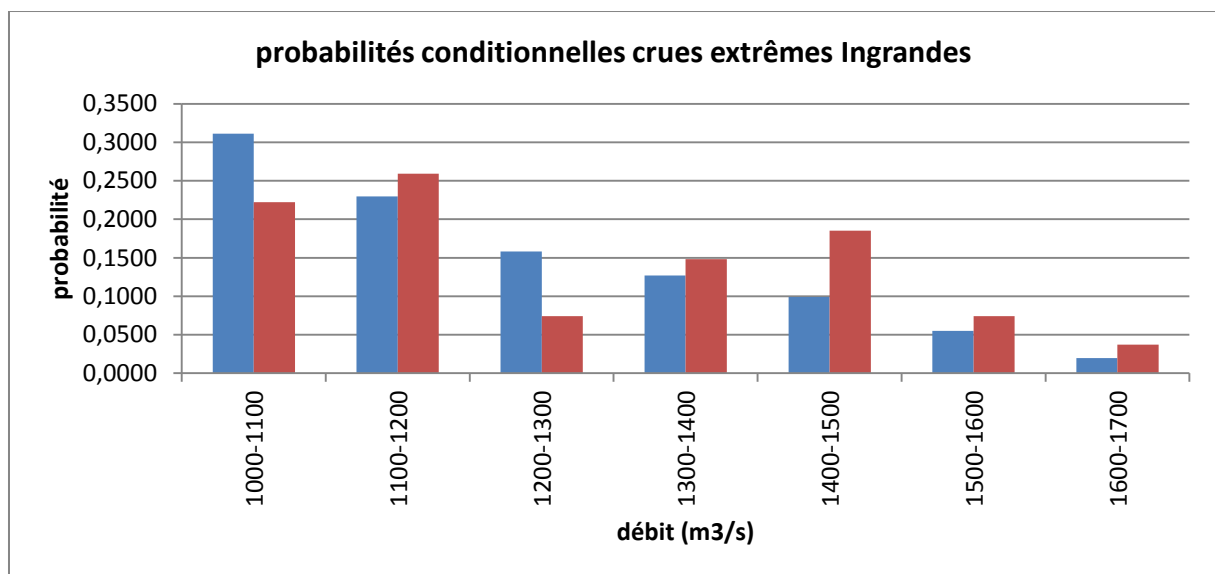


Figure 3 : histogramme des probabilités conditionnelles des crues à Ingrandes

En rouge, ce sont les valeurs observées (nécessairement erratiques) et en bleu les valeurs calculées (nécessairement décroissantes).

On calcule ensuite les probabilités absolues, en multipliant par la probabilité de dépasser le seuil. Les valeurs retenues sont :

nombre de jours retenus, depuis le 31/03/1919	34476
Proba X>1000	0,000608996

On en déduit la probabilité d'avoir au moins une crue en cent ans et la durée de retour pour chaque crue. Les calculs sont détaillés dans les différents onglets du fichier Excel Vienne\_Ingrandes.xlsm.

Voici le tableau des résultats :

classe	nb occurrences	proba calculée	proba observée	proba absolue par jour	proba au moins une fois en 100 ans	durée de retour (ans)
1000-1100	5	0,3111	0,2222	0,000189	0,9990	14
1100-1200	6	0,2297	0,2593	0,000140	0,9939	20
1200-1300	1	0,1580	0,0741	0,000096	0,9702	28
1300-1400	3	0,1271	0,1481	0,000077	0,9407	35
1400-1500	4	0,0991	0,1852	0,000060	0,8895	45
1500-1600	1	0,0552	0,0741	0,000034	0,7068	82
1600-1700	0	0,0198	0,0370	0,000012	0,3555	228

*Tableau 4 : la durée de retour des crues de la Vienne à Ingrandes*

## **Quatrième Partie**

### **La Creuse à Leugny**

## I. Introduction

La station de Leugny est située sur la Creuse, tout près du confluent avec la Vienne ; le but de l'analyse est de déterminer la part que prend la Creuse dans les crues de la Vienne au confluent.

## II. Les données

Elles sont enregistrées depuis 1965 ; on dispose de 47 relevés ; le débit minimum est 205 m<sup>3</sup>/s et le débit maximum est 1240 m<sup>3</sup>/s.

Saisonnalité : tous les mois de janvier à juin, puis novembre et décembre.

Voici le tableau des crues, rangées par tranches :

débit	nb occurrences
200-300	3
300-400	2
400-500	9
500-600	5
600-700	5
700-800	6
800-900	3
900-1000	3
1000-1100	4
1100-1200	4
1200-1300	3
1300-1400	0

*Tableau 1 : les crues de la Creuse à Leugny*

et voici l'histogramme des occurrences correspondant :

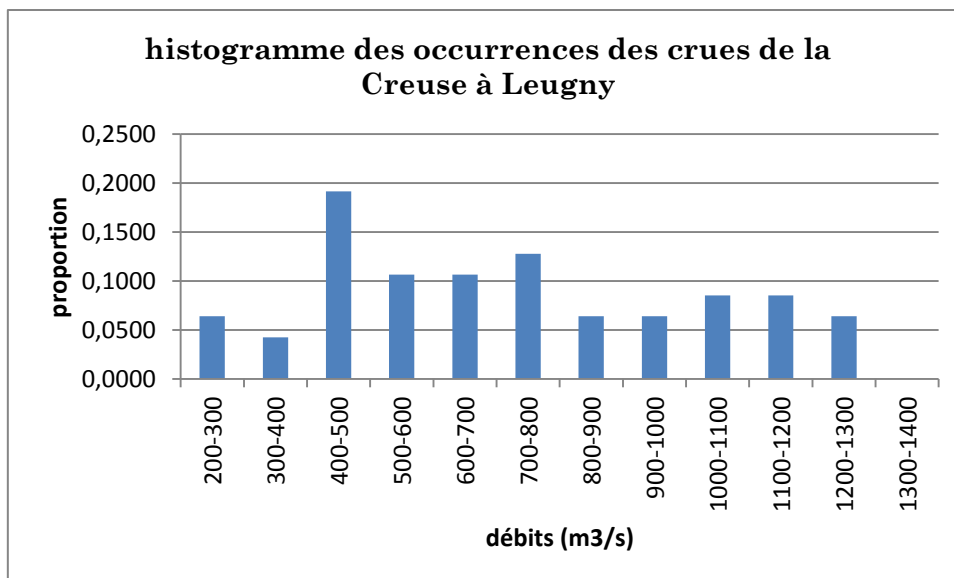


Figure 2 : histogramme des occurrences de crues de la Creuse à Leugny

Cet histogramme apparaît moins bizarre que celui de la Vienne à Ingrandes.

Le seuil pour les valeurs extrêmes sera fixé à 1 000 m<sup>3</sup>/s, ce qui est identique au précédent.

### III. Crues extrêmes

La démarche est la même que précédemment. On commence par calculer les probabilités conditionnelles des crues, sachant que le seuil est dépassé.

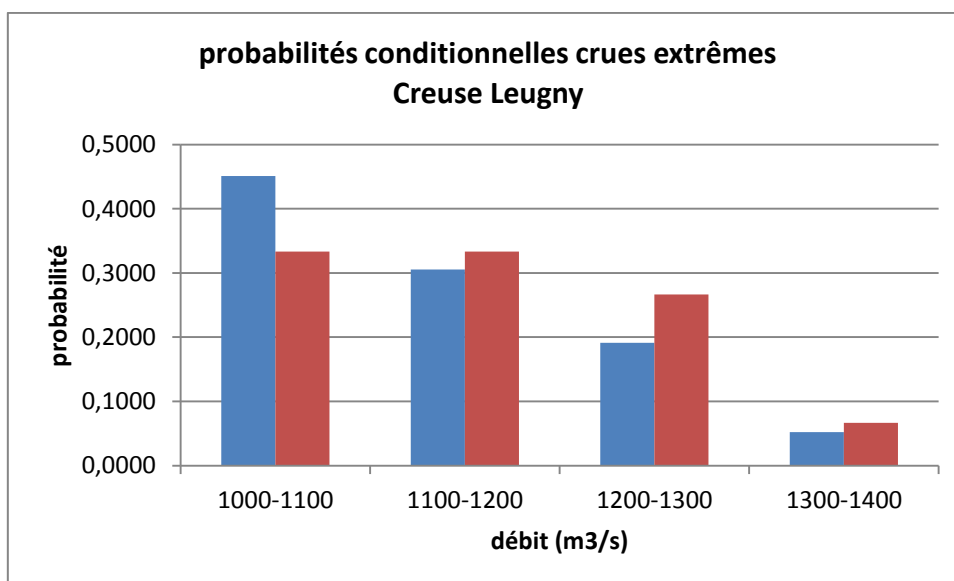


Figure 3 : histogramme des probabilités conditionnelles des crues extrêmes à Leugny



## Références

[Atlas] Atlas des zones inondables de la Vienne en Indre-et-Loire, Rapport technique de l'étude hydraulique, avril 2005

[MPPR] Bernard Beauzamy : Méthodes Probabilistes pour l'étude des phénomènes réels. Ouvrage édité et commercialisé par la Société de Calcul Mathématique SA, ISBN 2-9521458-0-6, ISSN 1767-1175. Mars 2004.

[NMP] Bernard Beauzamy : Nouvelles Méthodes Probabilistes pour l'évaluation des risques. Ouvrage édité et commercialisé par la Société de Calcul Mathématique SA. ISBN 978-2-9521458-4-8. ISSN 1767-1175, avril 2010.

## Annexe I

### La loi de Gumbel et son utilisation dans le document [Atlas]

#### A. Présentation générale

Nous commençons par une présentation générale de la loi de Gumbel et nous montrons pourquoi elle est inadaptée à l'évaluation des durées de retour des phénomènes extrêmes. Le paragraphe qui suit est reproduit de [MGSR].

Un autre aspect de la tyrannie de la précision porte sur les modèles, et non plus sur les données. Les ingénieurs veulent à toute force, c'est dans leur culture, disposer de modèles suivant des lois précises. Nous allons en passer en revue la plus communément utilisée.

#### B. La loi de Gumbel

Un exemple est celui de la loi de Gumbel, qui est une distribution de probabilité souvent utilisée dans la prédiction des valeurs extrêmes, parce qu'elle ne dépend que d'un petit nombre de paramètres, et est donc facile à caler.

On dit qu'une variable aléatoire  $G$  suit une loi de Gumbel de paramètres  $a, b$  si sa fonction de répartition est donnée par la formule :

$$F(x) = \exp\left\{-\exp\left\{-\frac{x-a}{b}\right\}\right\} \quad (1)$$

La densité de probabilité s'obtient par dérivation :

$$f(x) = \frac{1}{b} \exp\left\{-\frac{x-a}{b}\right\} \exp\left\{-\exp\left\{-\frac{x-a}{b}\right\}\right\}$$

L'espérance vaut :

$$E(G) = a + \gamma b$$

où  $\gamma$  est la constante d'Euler ( $\gamma \sim 0.5772\dots$ ).

La variance vaut :

$$\sigma^2(G) = \frac{\pi^2 b^2}{6}$$

Voici le graphe de la densité pour  $a=0$  et  $b=1$ ; la fonction décroît plus vite du côté gauche que du côté droit, du fait de la double exponentielle.

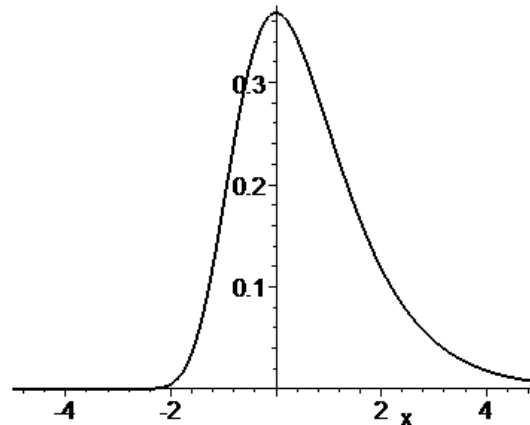


Figure 2 : la densité de Gumbel

On lit dans Wikipedia (Gumbel) : "*La distribution de Gumbel peut, par exemple, servir à prévoir le niveau des crues d'un fleuve, si on possède le relevé des débits sur dix ans [...].*"

*La distribution de Gumbel est une bonne approximation de la loi du maximum d'un échantillon de variables aléatoires indépendantes et de même loi, dès que cette loi appartient, précisément, au domaine d'attraction de la loi de Gumbel. Parmi les lois appartenant au domaine d'attraction de la loi de Gumbel, on compte la loi exponentielle.*"

L'idée est ici que la partie gauche de la courbe, décroissant très vite, peut être un bon modèle pour les phénomènes extrêmes, dont la probabilité décroît très vite à mesure que le caractère extrême s'accroît.

Concrètement, pour les phénomènes extrêmes, il s'agit donc de "renverser" la courbe (symétrie par rapport à  $Oy$ , pour que la partie décroissante soit la partie droite et non la partie gauche) et de chercher, en réglant les paramètres  $a, b$ , la courbe de Gumbel qui approche le mieux la courbe de la figure 1.

On y parvient évidemment (à condition de définir ce qu'on entend par "qui approche le mieux"), mais le choix de la forme (1) est totalement arbitraire ; il n'a aucune justification, ni physique ni probabiliste. Et lorsque Wikipedia écrit "c'est une bonne approximation...dès lors que la loi à approximer appartient au domaine d'approximation", c'est une tautologie, car personne ne sait a priori si la loi qui caractérise le phénomène extrême que l'on veut décrire appartient ou non au "domaine d'approximation" de la loi de Gumbel !

Voyons en pratique comment procéder pour représenter nos données au moyen d'une loi de Gumbel.

### C. Ajustement de la loi aux données

Tout d'abord, comme il nous faut une décroissance forte vers la droite et non vers la gauche, nous prendrons :

$$f(x) = \frac{1}{b} \exp\left\{\frac{x-a}{b}\right\} \exp\left\{-\exp\left\{\frac{x-a}{b}\right\}\right\} \quad (2)$$

et :

$$F(x) = 1 - \exp\left\{-\exp\left\{\frac{x-a}{b}\right\}\right\} \quad (3)$$

Nous disposons d'un tableau du type suivant (le sujet est ici des températures) :

temp	nb jours	temp	nb jours	temp	nb jours	temp	nb jours	temp	nb jours	temp	nb jours
35	9	36	3	37	3	38	1	39	0	40	0
35,2	7	36,2	7	37,2	4	38,2	1	39,2	0	40,2	0
35,4	8	36,4	6	37,4	1	38,4	1	39,4	3	40,4	1
35,6	11	36,6	6	37,6	2	38,6	2	39,6	0	40,6	0
35,8	6	36,8	4	37,8	2	38,8	0	39,8	0	40,8	0

Tableau 3 : fréquence des températures extrêmes

Nous avons ici  $K=30$  valeurs possibles de température, par pas de  $0.2^\circ\text{C}$ . Nous notons  $t_k$  ( $k=1,\dots,K$ ) les températures (colonnes de gauche) et  $n_k$  le nombre d'occurrences (colonnes de droite). Le nombre total d'observations est  $N = \sum_{k=1}^K n_k = 88$ . La probabilité de chaque température est :

$$p_k = \frac{n_k}{N}$$

L'ajustement de la loi de Gumbel aux données disponibles peut être réalisé de plusieurs manières possibles :

- A partir de la densité  $f$  (formule (2)) :

On cherche les valeurs de  $a, b$  qui minimisent la quantité :

$$Q_1 = \sum_{k=1}^K (f(t_k) - p_k)^2 \quad (4)$$

Pour cela, on écrit les deux dérivées partielles, par rapport à  $a$  et  $b$  et on résout numériquement le système obtenu. Le choix de la forme (4) est standard, mais n'est pas justifié pour autant : pourquoi une somme de carrés ? Voir [NMP] pour d'autres critiques.

– A partir de la fonction de répartition (formule (3)).

Pour cela, on commence par définir la fonction de répartition empirique (à partir de l'échantillon) ; elle est définie par :

$$\begin{aligned} F_e(t) &= 0 && \text{si } t < t_1 \\ F_e(t) &= p_1 && \text{si } t_1 \leq t < t_2 \\ F_e(t) &= p_1 + p_2 && \text{si } t_2 \leq t < t_3 \\ &\dots && \\ F_e(t) &= p_1 + \dots + p_k && \text{si } t_k \leq t < t_{k+1} \\ &\dots && \\ F_e(t) &= 1 && \text{si } t_K \leq t \end{aligned}$$

Il faut ensuite normaliser la fonction  $F$  de la formule (3) pour la limiter à l'intervalle  $[t_1, t_K]$  ; pour cela, nous posons :

$$F_1(t) = \frac{F(t)}{\int_{t_1}^{t_K} F(x) dx} \quad \text{si } t_1 \leq t \leq t_K, \quad 0 \text{ sinon.}$$

Ceci fait, le choix de  $a$  et  $b$  se fait en minimisant la quantité :

$$Q_2 = \int_{t_1}^{t_K} |F_e(t) - F_1(t)|^2 dt$$

ce qui requiert d'écrire deux dérivées partielles comme précédemment.

On constate qu'il y a deux approches principales possibles (sur la fonction de répartition ou sur la densité), que pour chacune le choix de la norme retenue est arbitraire (somme de carrés) et que de toute façon les calculs seront complexes. Mais le handicap principal est celui que nous avons dit au début : le choix même de la forme de la fonction de Gumbel est arbitraire.

## Référence

[MGSR] Bernard Beauzamy : Méthodes probabilistes pour la gestion des situations rares, en préparation.

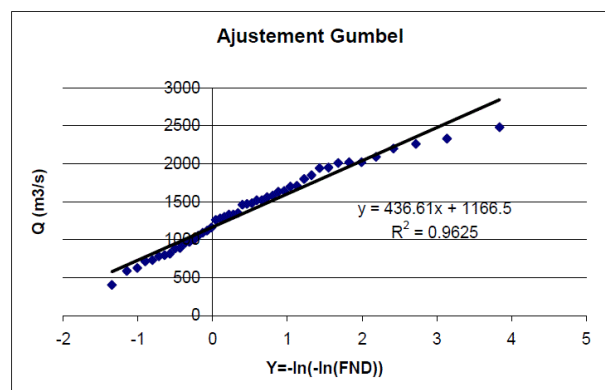
### D. L'utilisation dans le document [Atlas]

Elle est faite par un ajustement au sens des moindres carrés, comme expliqué ci-dessous (extrait du document [Atlas]).

#### 2.2.2 Ajustement des débits de la station de Nouâtre à une loi de Gumbel

L'analyse de la station de Nouâtre permet d'ajuster les valeurs des débits mesurés pendant près de 50 ans à une loi double exponentielle de Gumbel, moyennant l'affectation préalable, à chaque valeur maximale de débit enregistrée chaque année, d'une fréquence expérimentale (dite de Weibull) selon son rang de classement.

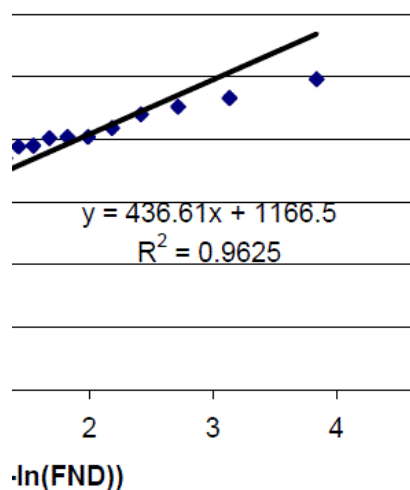
$$Q_T = -436,61 \cdot \ln\left(-\ln\left(1 - \frac{1}{T}\right)\right) + 1166,5$$



Il ne faut pas oublier qu'ici on travaille en "log(log)", et que toute imprécision sur l'ajustement a des conséquences extrêmement importantes sur le résultat.

L'ajustement est réalisé pour la partie centrale de la courbe, mais si on se limite aux valeurs extrêmes :

#### it Gumbel



on constate que l'ajustement est très mauvais : la droite tracée passe très au-dessus des points enregistrés, ce qui signifie que, pour un débit fixé (mettons l'horizontale 2500 m<sup>3</sup>/s), la durée de retour est beaucoup plus grande que ce que les auteurs n'écrivent.

Ici, nous avons deux problèmes distincts :

- L'utilisation de la loi de Gumbel, de toute manière, est inappropriée pour les phénomènes extrêmes, comme expliqué plus haut ;
- L'ajustement qui est fait par les auteurs de [Atlas] est inadéquat : il aurait mieux valu ajuster le graphe de la loi de Gumbel aux données, au moyen de l'une des méthodes présentées plus haut, que vouloir ajuster une droite après passage au log-log : cette dernière méthode amplifie considérablement les erreurs d'ajustement.

En effet, si :

$$|\log(\log(x)) - a| < \varepsilon$$

cela signifie :

$$a - \varepsilon < \log(\log(x)) < a + \varepsilon$$

et donc :

$$\exp(a - \varepsilon) < \log(x) < \exp(a + \varepsilon)$$

et finalement :

$$\exp(\exp(a - \varepsilon)) < x < \exp(\exp(a + \varepsilon))$$

Par exemple, si  $a = 1$  et  $\varepsilon = 0.1$ , on trouve  $11.7 < x < 20.1$  : la précision est très mauvaise.

Les auteurs de [Atlas] admettent eux-mêmes certaines limitations :

### 2.2.3 Ajustement à une loi normal à la station de Chinon

L'estimation du débit centennal à la station Nouâtre est en limite de validité compte-tenu de la longueur de la chronique disponible à cette station. Il est intéressant d'analyser les chroniques de hauteurs historiques disponibles à Nouâtre et surtout à Chinon pour bénéficier d'une chronique beaucoup plus longue (210 ans) et affermir l'estimation des valeurs des débits de référence.

## Annexe II

### Les méthodes mathématiques pour les phénomènes extrêmes

Nous détaillons maintenant les méthodes mathématiques que nous mettons en œuvre ; Le chapitre qui suit est reproduit de notre travail [MGSR].

#### I. Présentation mathématique du problème

Dans ce chapitre, nous présentons la théorie mathématique qui permet, à partir d'un petit nombre d'observations, d'assigner une probabilité à des événements rares, voire même qui n'ont jamais été observés.

Cette théorie permet de s'affranchir des défauts relevés au chapitre précédent : elle ne fait aucune hypothèse sur la forme des lois (Gaussiennes, exponentielles, log normales, etc.).

Le cadre de travail est le suivant : on dispose d'un certain nombre d'observations, d'un certain nombre de valeurs. Nous travaillerons sur l'exemple des températures à Paris : il s'agit du maximum journalier (température la plus élevée relevée au cours de la journée), observé depuis le 1er janvier 1873.

Comme nous nous intéressons aux phénomènes extrêmes, nous nous limiterons aux températures supérieures ou égales à 35°C : il s'agit d'un choix arbitraire. On aurait pu choisir 34°C ou 36°C : la théorie aurait été la même ; le résultat aurait été différent.

La donnée de base est donc un tableau où on lit le nombre d'observations d'une température donnée, pendant toute la période concernée. Voici ce tableau, sous forme préliminaire : il sera modifié par la suite, comme nous le verrons.

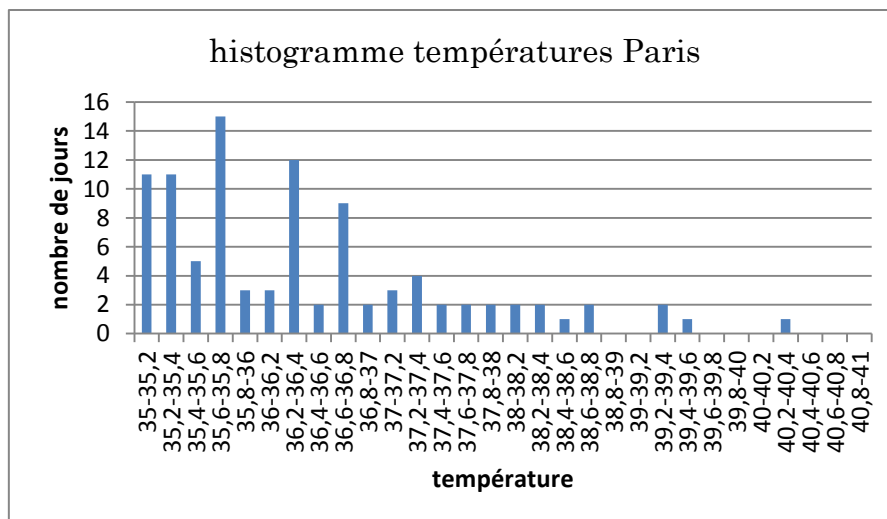


Tableau 1 : histogramme des températures élevées à Paris depuis 1873

Les températures, en abscisse, sont représentées par pas de 0.2°C, et sous la forme  $a \leq t < b$  (dans la classe 35-35.2, on a toutes les températures vérifiant  $35 \leq t < 35.2$ ). En ordonnée, on a le nombre de jours, depuis 1873, où cette température a été observée.

Notons bien ceci : nous avons besoin d'une liste d'observations pour travailler, même si la température dont nous voulons estimer la probabilité n'a jamais été observée. Par exemple, la température 42°C n'a jamais été observée, mais des températures plus basses l'ont été.

Aucune théorie mathématique ne permet de fixer une probabilité pour un événement qui n'a jamais été observé, ni d'autres événements de même nature. Si le tableau d'observations est vide, nous ne savons rien faire. Par exemple, on ne saurait pas attribuer une probabilité à l'événement "les martiens vont débarquer", parce que nous n'avons aucune observation, ni de cet événement, ni d'événements de même nature.

La première remarque que nous puissions faire sur le tableau 1 est qu'il est irrégulier. Par exemple, il y a beaucoup moins d'observations dans l'intervalle 35.4 – 35.6 que dans l'intervalle 35.6 – 35.8. Ceci est dû au fait que l'historique d'observation est insuffisant.

## II. L'hypothèse mathématique de base

L'hypothèse mathématique de base que nous ferons (et c'est la seule) est que la probabilité du phénomène doit décroître avec sa gravité : plus une situation est extrême et plus elle est rare. On acceptera volontiers l'idée que, à Paris, il est plus rare de rencontrer 40°C que 39.8°C.

Cette hypothèse paraît acceptable et légitime, mais elle requiert réflexion :

Ce n'est certainement pas vrai pour toutes les températures. Pour les températures usuelles, la probabilité n'est pas une fonction décroissante de la valeur. Bien au contraire, on rencontre la même situation de températures exceptionnelles avec les températures basses.

Il se peut que ce ne soit pas vrai pour certains phénomènes. Par exemple, il se pourrait qu'une température supérieure à 40°C requière une configuration (de pression atmosphérique, de pluviométrie, etc.) beaucoup plus fréquente qu'une configuration voisine de 36°C ; les données dont nous disposons ne confortent pas une telle conclusion.

L'hypothèse de décroissance est la seule que l'on puisse valablement faire sans connaissance particulière du phénomène. Elle nous permet d'affirmer que, dans le tableau 1, c'est à cause de l'insuffisance des observations que le graphe n'est pas décroissant.

En d'autres termes, nous postulons que, si on disposait d'un historique des observations sur mille années, au lieu de 140, le graphe du tableau 1 serait décroissant, ou, du moins, serait moins irrégulier.

### **III. La stationnarité de la loi**

Nous voulons répondre à une question du type : quelle est la probabilité de rencontrer une température supérieure à 38°C à Paris, au cours d'une année donnée, ou, par exemple, au cours du 22<sup>ème</sup> siècle ?

Mais, comme nous utilisons des données qui remontent à 1873, cela suppose que la loi soit stationnaire, ce qui signifie que cette probabilité ne varie pas d'une année sur l'autre. Est-ce bien vrai ?

La réponse est que nous n'en savons rien. Indépendamment d'un hypothétique réchauffement climatique global, qui concernerait la planète tout entière, nous avons une situation locale qui peut être affectée par de nombreux éléments, par exemple l'industrialisation et la désindustrialisation. N'oublions pas aussi qu'il s'agit d'une mesure ponctuelle (Paris Montsouris), qui peut être affectée par des circonstances locales (construction d'usines ou d'immeubles à proximité, etc.).

Nous admettrons cette stationnarité pour les besoins de la construction que nous allons faire (il s'agit avant tout de présenter une méthode), mais nous admettons volontiers qu'elle demande vérification. Les choses seraient plus assurées s'il s'agissait de moyennes sur tout un territoire.

## IV. Le choix de la période de référence

Ici, nous avons clairement plusieurs choix possibles, parce qu'il est évident que la température n'est pas stationnaire sur toute l'année ! Il est très peu probable de rencontrer 35°C en janvier à Paris (en fait, cela ne s'est jamais vu), mais, formellement, ce n'est pas impossible.

Parmi les choix possibles, nous avons les situations suivantes :

- Nous travaillons sur l'année entière. Nous allons utiliser toutes les données disponibles, sans nous soucier du mois pendant lequel elles ont été recueillies, et le résultat obtenu sera la probabilité d'avoir telle température à Paris au cours d'une année. Le nombre d'observations de température  $\geq 35^\circ\text{C}$  est alors de 97 : c'est le tableau précédent. Le nombre total d'observations est 51 224.
- Nous travaillons sur une "période chaude", par exemple 15 juillet – 15 août. Nous utiliserons uniquement les données provenant de cette période, et le résultat sera la probabilité d'avoir telle température à Paris au cours de cette période. Le nombre d'observations de température  $\geq 35^\circ\text{C}$  est alors de 60 et le nombre total d'observations est de 4 480 pour cette période.

Nous choisirons la seconde version pour présenter la méthode. Toutes les versions sont correctes, mais ne correspondent pas à la même question et ne donnent pas le même résultat. Des températures supérieures à 35°C ont été observées à Paris en juin, juillet, août, septembre. Il est important de bien préciser le cadre de travail.

## V. Les données de travail

Nous aurons donc les données suivantes :

$N = 4\,480$  : c'est le nombre total d'observations, c'est-à-dire le nombre de jours entre le 15 juillet et le 15 août (bornes incluses), du 1<sup>er</sup> janvier 1873 au 31 avril 2013.

$K = 30$ : c'est le nombre de valeurs possibles que nous considérons pour la température, entre 35°C et 41°C. Le  $k$ -ème intervalle est défini par :

$$v_k = 35 + 0.2 \times (k - 1)$$

où  $v_k$  est la borne de gauche de l'intervalle,  $k = 1, \dots, 30$ .

Le tableau des occurrences est le suivant :

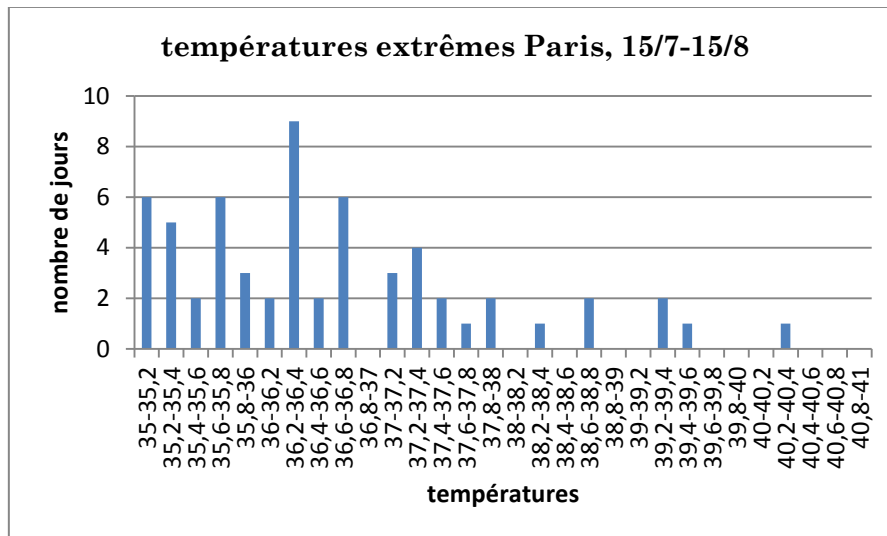
k	temp	nb de jours	k	temp	nb de jours
1	35-35,2	6	16	38-38,2	0
2	35,2-35,4	5	17	38,2-38,4	1
3	35,4-35,6	2	18	38,4-38,6	0
4	35,6-35,8	6	19	38,6-38,8	2
5	35,8-36	3	20	38,8-39	0
6	36-36,2	2	21	39-39,2	0
7	36,2-36,4	9	22	39,2-39,4	2
8	36,4-36,6	2	23	39,4-39,6	1
9	36,6-36,8	6	24	39,6-39,8	0
10	36,8-37	0	25	39,8-40	0
11	37-37,2	3	26	40-40,2	0
12	37,2-37,4	4	27	40,2-40,4	1
13	37,4-37,6	2	28	40,4-40,6	0
14	37,6-37,8	1	29	40,6-40,8	0
15	37,8-38	2	30	40,8-41	0

Tableau 2 : nombre d'occurrences de chaque température

Pour  $k = 1, \dots, K$ , nous notons  $n_k$  le nombre d'occurrences de la valeur  $v_k$  : c'est la troisième et la sixième colonne du tableau ci-dessus.

Nous avons un total de  $\sum_{k=1}^K n_k = 60$  jours pendant lesquels la température a été  $\geq 35^\circ\text{C}$ .

Voici la représentation sous forme d'histogramme :



Histogramme 3 : les températures extrêmes à Paris du 15 juillet au 15 août

De manière logique, nous procédons en deux temps :

Première étape : nous estimons la probabilité de dépasser la valeur 35°C. Pour cela, nous rappelons le résultat démontré dans [MPPR], chapitre 14 : si  $n$  accidents ont été observés sur  $N$  expériences, le taux de risque a pour densité :

$$f_{n,N}(\lambda) = c\lambda^n(1-\lambda)^{N-n}$$

où  $c$  est une constante de normalisation (pour que l'intégrale soit égale à 1), qui vaut :

$$c = \frac{(N+1)!}{n!(N-n)!}$$

Ici, le fait de dépasser 35°C est considéré comme un "accident", et nous avons  $N = 4\,480$ ,  $n = 60$ .

L'espérance de la loi vaut :

$$E = \frac{n+1}{N+2} = \frac{61}{4482} = 1.36 \times 10^{-2}$$

Seconde étape : il faut déterminer la probabilité de chaque classe de température au dessus de 35°C, en supposant que l'on dépasse ce seuil ; c'est donc une probabilité conditionnelle.

Pour  $k = 1, \dots, K$ , nous notons  $X_k$  la variable aléatoire représentant la probabilité du  $k$ -ème intervalle  $\nu_k$ . Cette probabilité est inconnue (c'est précisément ce que l'on cherche à estimer) et il est donc légitime de la considérer comme une variable aléatoire. Il y a donc ici une difficulté de vocabulaire : ce qu'on cherche à estimer est une probabilité, et on sera donc amené à parler de "loi de probabilité de la probabilité". Nous avons déjà rencontré cette difficulté dans [NMP] avec les taux de risque.

Nous noterons  $p_k = EX_k$  l'espérance de la variable  $X_k$  : c'est l'estimation que l'on sera amené à faire pour la variable  $X_k$ . Lorsque nous dirons "la probabilité de l'intervalle  $\nu_1$  est de tant", cela se référera à la valeur que nous aurons calculée pour  $\nu_1$ . Si nous disons la probabilité de l'intervalle  $\nu_1$  est, à 95%, comprise entre telle valeur et telle valeur, cela se réfère à la loi de  $X_1$ .

Ici, nous avons la même situation qu'à la première étape, avec  $K$  classes d'accidents (toutes les températures supérieures à 35°C). La variable  $X_k$  répond à la question : quelle est la probabilité de tomber dans le  $k$ -ème intervalle de température, sachant que la température est  $\geq 35^\circ\text{C}$  ? On a donc :

$$\sum_{k=1}^K X_k = 1$$

La loi conjointe de notre ensemble  $(X_1, \dots, X_K)$  sera donnée par la densité :

$$f(\lambda_1, \dots, \lambda_K) = c 1_S(\lambda_1, \dots, \lambda_K) \lambda_1^{n_1} \dots \lambda_K^{n_K} \quad (1)$$

où:

- $1_S$  est la fonction indicatrice du simplexe :

$$S = \left\{ (\lambda_1, \dots, \lambda_K); \lambda_1 \geq \dots \geq \lambda_K \geq 0, \sum_{k=1}^K \lambda_k = 1 \right\},$$

c'est-à-dire la fonction qui vaut 1 dans cet ensemble et 0 en dehors ;

- $c$  est une constante de normalisation : l'intégrale multiple de  $f(\lambda_1, \dots, \lambda_K)$  vaut 1.

Prise séparément, chaque  $X_k$  aura pour densité la loi marginale :

$$f_k(\lambda) = \int_S \dots \int f(\lambda_1, \dots, \lambda_{k-1}, \lambda, \lambda_{k+1}, \dots, \lambda_K) d\lambda_1 \dots d\lambda_{k-1} d\lambda_{k+1} \dots d\lambda_K$$

et la valeur que nous assignerons à chaque  $p_k$  sera l'espérance :

$$\overline{p_k} = \int_0^1 \lambda f_k(\lambda) d\lambda$$

## VI. Calcul de l'intégrale d'un monôme

Que ce soit pour évaluer l'intégrale multiple de la fonction en (1) l'espérance de l'une des lois marginales, le problème est le même ; nous devons calculer l'intégrale d'un monôme :

$$M(\lambda_1, \dots, \lambda_K) = \lambda_1^{n_1} \dots \lambda_K^{n_K}$$

sur le simplexe  $S$ .

A l'origine, ceci a été fait par l'auteur en 2009 en utilisant le calcul symbolique

[http://www.scmsa.eu/RMM/BB\\_rare\\_events\\_2009\\_08.pdf](http://www.scmsa.eu/RMM/BB_rare_events_2009_08.pdf)

mais cette méthode était très lente et ne pouvait être mise en oeuvre que pour des valeurs faibles des  $n_k$ .

Ensuite, en 2010; une nouvelle approche a été introduite par Peter Robinson [http://www.scmsa.eu/RMM/ART\\_2010\\_Peter\\_Robinson\\_Efficient\\_Integration.pdf](http://www.scmsa.eu/RMM/ART_2010_Peter_Robinson_Efficient_Integration.pdf) elle reposait sur une méthode de Monte-Carlo.

### A. La méthode de Peter Robinson

Elle repose sur l'idée suivante : il faut disposer de points "bien répartis" dans le simplexe  $S$  ; si  $(\lambda_1^{(j)}, \dots, \lambda_K^{(j)})$ ,  $j = 1, \dots, J$ , sont ces points, alors l'intégrale  $\int_S M(\lambda_1, \dots, \lambda_K)$  pourra être approchée par la somme  $\frac{1}{J} \sum_{j=1}^J M(\lambda_1^{(j)}, \dots, \lambda_K^{(j)})$  ; l'énoncé précis est que la somme converge vers l'intégrale lorsque  $J \rightarrow +\infty$ .

La difficulté consiste à obtenir ces points bien répartis. Pour cela, on procède de la manière suivante :

Pour  $j = 1, \dots, J$  :

On génère  $K - 1$  nombres  $(x_1^{(j)}, \dots, x_{K-1}^{(j)})$  selon une loi uniforme sur  $0-1$ .

On les range par ordre croissant et on les note  $u_1^{(j)} < \dots < u_{K-1}^{(j)}$

On calcule les différences consécutives :

$$d_1^{(j)} = u_1^{(j)}, d_2^{(j)} = u_2^{(j)} - u_1^{(j)}, \dots, d_k^{(j)} = u_k^{(j)} - u_{k-1}^{(j)}, \dots, d_K^{(j)} = 1 - u_{K-1}^{(j)}$$

On range ces différences par ordre décroissant :  $\lambda_1^{(j)} > \dots > \lambda_K^{(j)}$  : ce seront les points voulus.

La référence pour la justification théorique de ces résultats est :

Devroye, L (1981). Laws of the iterated logarithm for order statistics of uniform spacings. Annals of Probability Vol. 9, No. 5, 860-867.

Cette méthode marche bien si le nombre d'observations et le nombre d'intervalles n'est pas trop élevé (voir le livre [NMP] pour les limitations des méthodes de Monte-Carlo).

### B. Une méthode rapide et totalement déterministe

Nous présentons maintenant une nouvelle approche qui permet un calcul rapide des intégrales multiples, qui restent à la base de notre méthode. Cette approche n'utilise aucune procédure de type Monte-Carlo ; elle est complètement explicite et déterministe.

Nous allons commencer par bien définir ce que sont ces intégrales multiples sur un simplexe.

### C. Cas de deux variables

Commençons par le cas de deux variables. Nous avons à calculer :

$$I = \int_S \lambda_1^{n_1} \lambda_2^{n_2} d\lambda_1 d\lambda_2 \quad (1)$$

où

$$S = \{(\lambda_1, \lambda_2) ; \lambda_1 \geq \lambda_2 \geq 0, \lambda_1 + \lambda_2 = 1\} \quad (2)$$

Il en résulte que  $\lambda_2 = 1 - \lambda_1$  et les conditions  $0 \leq 1 - \lambda_1 \leq \lambda_1 \leq 1$  impliquent :

$$\frac{1}{2} \leq \lambda_1 \leq 1$$

L'intégrale (1) s'écrit donc :

$$I = \int_{1/2}^1 \lambda_1^{n_1} (1 - \lambda_1)^{n_2} d\lambda_1 \quad (3)$$

### D. Cas général

Nous allons être amenés à faire différents changements de variables. Considérons donc le monôme :

$$M = x_1^{n_1} \cdots x_K^{n_K} \quad (1)$$

dont nous voulons calculer l'intégrale sur le simplexe :

$$S = \left\{ (x_1, \dots, x_K) ; x_1 \geq \dots \geq x_K \geq 0, \sum_{k=1}^K x_k = 1 \right\} \quad (2)$$

que nous noterons :

$$I = \int_S M$$

Par définition,

$$x_K = 1 - x_1 - \dots - x_{K-1}$$

et donc  $M$  est une fonction de  $K-1$  variables :

$$M = x_1^{n_1} \dots x_{K-1}^{n_{K-1}} (1 - x_1 - \dots - x_{K-1})^{n_K}$$

L'inégalité  $x_K \leq x_{K-1}$  se traduit par :

$$1 - x_1 - \dots - x_{K-1} \leq x_{K-1}$$

et donc :

$$1 - x_1 - \dots - x_{K-2} \leq 2x_{K-1}$$

D'où l'intervalle de variation pour  $x_{K-1}$  :

$$\frac{1}{2}(1 - x_1 - \dots - x_{K-2}) \leq x_{K-1} \leq x_{K-2}$$

La condition  $\frac{1}{2}(1 - x_1 - \dots - x_{K-2}) \leq x_{K-1}$  ci-dessus implique :

$$1 - x_1 - \dots - x_{K-3} \leq 3x_{K-2}$$

d'où l'intervalle de variation pour  $x_{K-2}$  :

$$\frac{1}{3}(1 - x_1 - \dots - x_{K-3}) \leq x_{K-2} \leq x_{K-3}$$

et ainsi de suite pour tout  $j$  :

$$\frac{1}{K-j}(1-x_1-\dots-x_j) \leq x_{j+1} \leq x_j$$

jusqu'à :

$$\frac{1}{K-1}(1-x_1) \leq x_2 \leq x_1$$

ce qui requiert :

$$\frac{1}{K} \leq x_1 \leq 1$$

L'intégrale  $I$  est donc une intégrale multiple ordinaire, portant sur  $K-1$  variables, assujetties aux inégalités que nous venons d'écrire.

Nous allons maintenant voir comment calculer explicitement cette intégrale.

## VII. Un premier changement de variable

Nous introduisons de nouvelles variables :

$$y_K = x_K$$

$$y_k = x_k - x_{k+1}, \quad k = 1, \dots, K-1.$$

On a donc  $y_k \geq 0$ ,  $k = 1, \dots, K$ .

A l'inverse, les  $x_k$  peuvent être calculés à partir des  $y_k$  :

$$x_K = y_K$$

$$x_k = y_k + y_{k+1} + \dots + y_K, \quad k = 1, \dots, K.$$

Si nous utilisons toutes les variables, le Jacobien de la transformation des  $x_k$  en  $y_k$  est égal à 1 ; ce ne sera pas le cas si la  $K$ -ème est substituée à partir des précédentes. Quoi qu'il en soit, la valeur de ce Jacobien ne nous intéresse pas, parce que nous devons calculer des quotients :

$$\frac{\int \lambda f}{\int f}$$

et que le changement de variable se fait de la même façon au numérateur et au dénominateur.

La condition  $\sum_{k=1}^K x_k = 1$  devient :

$$y_K + y_K + y_{K-1} + \dots + y_K + \dots + y_k + \dots + y_K + \dots + y_1 = 1$$

c'est à dire :

$$Ky_K + (K-1)y_{K-1} + \dots + ky_k + \dots + y_1 = 1.$$

Avec ces nouvelles variables, le monôme  $M$  devient :

$$P(y_1, \dots, y_K) = \left( \sum_{j=1}^K y_j \right)^{n_1} \dots \left( \sum_{j=k}^K y_j \right)^{n_k} \dots \left( \sum_{j=K}^K y_j \right)^{n_K}$$

Nous devons intégrer le polynôme  $P(y_1, \dots, y_K)$  sur le simplexe :

$$S_1 = \left\{ (y_1, \dots, y_K) ; y_k \geq 0, \sum_{k=1}^K ky_k = 1 \right\}$$

## VIII. Un second changement de variables

Faisons encore un changement de variables ; posons :

$$z_k = ky_k, \quad k = 1, \dots, K.$$

Le Jacobien de cette transformation vaut  $K!$  si l'on utilise toutes les variables.

Le polynôme  $P(y_1, \dots, y_K)$  devient :

$$Q(z_1, \dots, z_K) = \left( \sum_{j=1}^K \frac{z_j}{j} \right)^{n_1} \dots \left( \sum_{j=k}^K \frac{z_j}{j} \right)^{n_k} \dots \left( \sum_{j=K}^K \frac{z_j}{j} \right)^{n_K}$$

et l'ensemble d'intégration est maintenant :

$$S_2 = \left\{ (z_1, \dots, z_K) ; z_k \geq 0, \sum_{k=1}^K z_k = 1 \right\}$$

## IX. Le problème explicite

Nous voulons calculer l'intégrale :

$$I = \frac{1}{K!} \int_{S_2} \left( \sum_{j=1}^K \frac{z_j}{j} \right)^{n_1} \cdots \left( \sum_{j=k}^K \frac{z_j}{j} \right)^{n_k} \cdots \left( \sum_{j=K}^K \frac{z_j}{j} \right)^{n_K} dz_1 \cdots dz_K \quad (1)$$

Nous rappelons que, d'après [MPPR], chapitre 14, §9, on a :

$$\int_{S_2} z_1^{n_1} \cdots z_K^{n_K} dz_1 \cdots dz_K = \frac{n_1! \cdots n_K!}{(N + K - 1)!} \quad (2)$$

avec  $N = n_1 + \cdots + n_K$ .

## X. Calcul par récurrence

Nous allons établir le résultat suivant :

**Théorème.** – *L'intégrale :*

$$I = \frac{1}{K!} \int_{S_2} \left( \sum_{j=1}^K \frac{z_j}{j} \right)^{n_1} \cdots \left( \sum_{j=k}^K \frac{z_j}{j} \right)^{n_k} \cdots \left( \sum_{j=K}^K \frac{z_j}{j} \right)^{n_K} dz_1 \cdots dz_K$$

sur le simplexe

$$S_2 = \left\{ (z_1, \dots, z_K); z_k \geq 0, \sum_{k=1}^K z_k = 1 \right\}$$

vaut :

$$I = \frac{1}{K!(N + K - 1)!} \sum_{m_1=0}^{n_1} \binom{n_1}{m_1} \frac{m_1!}{1^{m_1}} \sum_{m_2=0}^{v_2 - m_1} \binom{v_2 - m_1}{m_2} \frac{m_2!}{2^{m_2}} \cdots \sum_{m_k=0}^{v_k - \mu_{k-1}} \binom{v_k - \mu_{k-1}}{m_k} \frac{m_k!}{k^{m_k}} \cdots \\ \times \sum_{m_{K-1}=0}^{v_{K-1} - \mu_{K-2}} \binom{v_{K-1} - \mu_{K-2}}{m_{K-1}} \frac{m_{K-1}!}{(K-1)^{m_{K-1}}} \frac{(v_K - \mu_{K-1})!}{K^{v_K - \mu_{K-1}}}$$

où l'on a noté :

$$\nu_k = n_1 + \dots + n_k, \quad \mu_k = m_1 + \dots + m_k.$$

### Démonstration du théorème

Elle se fait par récurrence. Nous écrivons :

$$\left( \sum_{j=1}^K \frac{z_j}{j} \right)^{n_1} = \left( z_1 + \sum_{j=2}^K \frac{z_j}{j} \right)^{n_1} = \sum_{m_1=0}^{n_1} \binom{n_1}{m_1} z_1^{m_1} \left( \sum_{j=2}^K \frac{z_j}{j} \right)^{n_1-m_1}$$

et donc :

$$Q(z_1, \dots, z_K) = \sum_{m_1=0}^{n_1} \binom{n_1}{m_1} z_1^{m_1} \left( \sum_{j=2}^K \frac{z_j}{j} \right)^{n_1+n_2-m_1} \left( \sum_{j=3}^K \frac{z_j}{j} \right)^{n_3} \dots \left( \sum_{j=k}^K \frac{z_j}{j} \right)^{n_k} \dots \left( \sum_{j=K}^K \frac{z_j}{j} \right)^{n_K} \quad (3)$$

c'est à dire :

$$Q(z_1, \dots, z_K) = \sum_{m_1=0}^{n_1} \binom{n_1}{m_1} z_1^{m_1} Q_2$$

avec :

$$Q_2(z_2, \dots, z_K; m_1) = \left( \sum_{j=2}^K \frac{z_j}{j} \right)^{n_1+n_2-m_1} \left( \sum_{j=3}^K \frac{z_j}{j} \right)^{n_3} \dots \left( \sum_{j=k}^K \frac{z_j}{j} \right)^{n_k} \dots \left( \sum_{j=K}^K \frac{z_j}{j} \right)^{n_K}$$

De même :

$$\left( \sum_{j=2}^K \frac{z_j}{j} \right)^{n_1-m_1+n_2} = \left( \frac{z_2}{2} + \sum_{j=3}^K \frac{z_j}{j} \right)^{n_1+n_2-m_1} = \sum_{m_2=0}^{n_1+n_2-m_1} \binom{n_1+n_2-m_1}{m_2} \left( \frac{z_2}{2} \right)^{m_2} \left( \sum_{j=3}^K \frac{z_j}{j} \right)^{n_1+n_2-m_1-m_2}$$

ce qui donne :

$$Q_2 = \sum_{m_2=0}^{n_1+n_2-m_1} \binom{n_1+n_2-m_1}{m_2} \left( \frac{z_2}{2} \right)^{m_2} \left( \sum_{j=3}^K \frac{z_j}{j} \right)^{n_1+n_2+n_3-m_1-m_2} \left( \sum_{j=4}^K \frac{z_j}{j} \right)^{n_4} \dots \left( \sum_{j=k}^K \frac{z_j}{j} \right)^{n_k} \dots \left( \sum_{j=K}^K \frac{z_j}{j} \right)^{n_K}$$

En répétant le calcul, nous obtenons :

$$Q_2 = \sum_{m_2=0}^{n_1+n_2-m_1} \binom{n_1+n_2-m_1}{m_2} \left( \frac{z_2}{2} \right)^{m_2} Q_3$$

avec :

$$Q_3(z_3, \dots, z_K; m_1, m_2) = \left( \sum_{j=3}^K \frac{z_j}{j} \right)^{v_3 - \mu_2} \left( \sum_{j=4}^K \frac{z_j}{j} \right)^{n_4} \dots \left( \sum_{j=k}^K \frac{z_j}{j} \right)^{n_k} \dots \left( \sum_{j=K}^K \frac{z_j}{j} \right)^{n_K}$$

Plus g en eralement :

$$Q = \sum_{m_1=0}^{n_1} \binom{v_1}{m_1} z_1^{m_1} \sum_{m_2=0}^{v_2 - \mu_1} \binom{v_2 - \mu_1}{m_2} \left( \frac{z_2}{2} \right)^{m_2} \dots \sum_{m_{k-1}}^{v_{k-1} - \mu_{k-2}} \binom{v_{k-1} - \mu_{k-2}}{m_{k-1}} \left( \frac{z_{k-1}}{k-1} \right)^{m_{k-1}} Q_k$$

avec :

$$Q_k(z_k, \dots, z_K; m_1, \dots, m_{k-1}) = \left( \sum_{j=k}^K \frac{z_j}{j} \right)^{v_k - m_{k-1}} \left( \sum_{j=k+1}^K \frac{z_j}{j} \right)^{n_{k+1}} \dots \left( \sum_{j=K}^K \frac{z_j}{j} \right)^{n_K}$$

$$Q_{K-1} = \left( \frac{z_{K-1}}{K-1} + \frac{z_K}{K} \right)^{v_{K-1} - \mu_{K-2}} \left( \frac{z_K}{K} \right)^{n_K}$$

ce qui donne :

$$Q_{K-1} = \sum_{m_{K-1}}^{v_{K-1} - \mu_{K-2}} \binom{v_{K-1} - \mu_{K-2}}{m_{K-1}} \left( \frac{z_{K-1}}{K-1} \right)^{m_{K-1}} Q_K$$

avec :

$$Q_K(z_K; m_1, \dots, m_{K-1}) = \left( \frac{z_K}{K} \right)^{v_K - \mu_{K-1}}$$

Tous les coefficients sont positifs et il ne peut y avoir d'annulation. Chaque polyn ome apporte sa propre contribution. Nous notons  $EQ_j$  la contribution du  $j$ -  eme polyn ome.

Nous obtenons :

$$EQ_K(m_1, \dots, m_{K-1}) = \frac{(v_K - \mu_{K-1})!}{K^{v_K - \mu_{K-1}}}$$

Nous remarquons que cette quantit e ne d epend que de  $\mu_{K-1} = m_1 + \dots + m_{K-1}$  ; ceci sera utile en pratique.

Aussi :

$$EQ_{K-1}(\mu_{K-2}) = \sum_{m_{K-1}=0}^{v_{K-1}-\mu_{K-2}} \binom{v_{K-1}-\mu_{K-2}}{m_{K-1}} \frac{m_{K-1}!}{(K-1)^{m_{K-1}}} EQ_K(\mu_{K-1})$$

et plus généralement :

$$EQ_{k-1}(\mu_{k-2}) = \sum_{m_{k-1}=0}^{v_{k-1}-\mu_{k-2}} \binom{v_{k-1}-\mu_{k-2}}{m_{k-1}} \frac{m_{k-1}!}{(k-1)^{m_{k-1}}} EQ_k(\mu_{k-1})$$

et :

$$EQ = EQ_1 = \sum_{m_1=0}^{n_1} \binom{v_1}{m_1} \frac{m_1!}{1^{m_1}} EQ_2(\mu_1)$$

Finalement, en utilisant les formules (1) et (2), nous obtenons :

$$I = \frac{EQ}{(N+K-1)!K!}$$

ce qui montre le théorème.

## Annexe III

### Vérification des données pour chaque station

Nous allons, pour chaque station qui nous intéresse, travailler sur les débits moyens et non plus, comme précédemment, sur des débits maximaux. Ceci va nous fournir des possibilités de vérification et vient compléter les analyses déjà faites, qui procédaient par comparaison entre stations.

Nous commençons par évaluer la disponibilité des données de débit moyen.

#### I. Disponibilité des données

Les données étudiées ici sont les valeurs de débits moyens journaliers sur tout l'historique disponible pour chacune des 7 stations (ce ne sont plus les débits max) :

- La Creuse à Ciron ;
- La Creuse à Leugny ;
- Le Clain à Dissay ;
- La Vienne à Châtelleraut ;
- La Vienne à Ingrandes ;
- La Vienne à Nouâtre ;
- La Gartempe à Montmorillon.

Nous cherchons dans un premier temps à savoir si les historiques disponibles pour chacune de ces stations sont complets.

Le tableau suivant récapitule ces résultats :

Cours d'eau	Station	Date début historique	Date fin historique	nb données théorique	nb données manquantes	% données manquantes
Clain	Dissay	01/01/1965	26/08/2013	17770	1964	11,05%
Creuse	Ciron	01/01/1971	29/07/2013	15551	8	0,05%
Creuse	Leugny	01/01/1964	26/08/2013	18136	850	4,69%
Gartempe	Montmorillon	01/01/1955	29/08/2013	21426	2162	10,09%
Vienne	Chatelleraut	01/01/1918	24/08/2013	34935	21951	62,83%
Vienne	Ingrandes	01/01/1918	29/08/2013	34940	7061	20,21%
Vienne	Nouâtre	01/01/1958	29/08/2013	20330	0	0,00%

Toutes les stations à part la Creuse à Ciron et la Vienne à Nouâtre ont des données manquantes.

Le résultat le plus mauvais est obtenu pour la Vienne à Châtelleraut, avec plus de 62% de données manquantes. Pour cette station, aucune mesure n'a été relevée entre le 1<sup>er</sup> janvier 1939 et le 1<sup>er</sup> janvier 1999. Nous avons déjà relevé des anomalies plus haut à propos de cette station. Mais en ce qui concerne la Vienne à Nouâtre, qui est l'objet principal de ce rapport, les choses sont satisfaisantes.

## II. Comparaison des dates de crues

Nous allons maintenant chercher à savoir si les débits maximaux, utilisés plus haut, correspondent bien à des débits moyens élevés, et aux mêmes dates.

Les valeurs de ces deux débits ne sont évidemment pas les mêmes, mais on peut penser qu'une crue est aussi détectable sur la valeur du débit moyen journalier, de la manière suivante.

Pour chaque date de crue identifiée au moyen du débit maximum, nous relevons la valeur du débit moyen journalier correspondant. Nous comparons ensuite cette valeur à la valeur de débit maximum utilisée pour l'étude. Si le débit moyen journalier correspond à plus de 75% du débit maximum, on considère que la crue est validée à cette date. Il se peut bien sûr qu'une crue soit particulièrement brève, auquel cas elle échapperait à cette méthode.

Par exemple, si nous étudions le cas de la Vienne à Nouâtre le 1<sup>er</sup> janvier 1962, nous avons un débit maximal instantané de 2480 m<sup>3</sup>/s à cette date pour cette station. En revanche, les données journalières donnent un débit moyen de 230 m<sup>3</sup>/s pour cette même station à cette même date, ce qui représente moins de 10% de la valeur utilisée dans les calculs. Il y a donc une incohérence de date sur cette valeur.

Si l'on observe les valeurs de débits moyens journaliers pour l'ensemble du mois de janvier 1962 pour cette station, la valeur la plus élevée a été relevée le 14 janvier, avec un débit moyen de 2300 m<sup>3</sup>/s. Cette valeur correspond à 93% de la valeur maximale instantanée enregistrée pour ce mois. On peut donc supposer que la crue relevée a en fait eu lieu à cette date.

Considérons un autre exemple pour cette même station. Le 21 février 1967, le débit maximal instantané enregistré est de 1140 m<sup>3</sup>/s. Le débit moyen de cette date est de 1010 m<sup>3</sup>/s, soit 89 % du débit maximal instantané enregistré. La date de cette crue est donc validée dans ce cas.

Nous faisons cette vérification pour l'ensemble des 7 stations. Les résultats obtenus sont les suivants :

	pourcentage de données validées
La Creuse à Ciron	51%
La Creuse à Leugny	93%
Le Clain à Dissay	56%
La Vienne à Châtelleraut	80%
La Vienne à Ingrandes	75%
La Vienne à Nouâtre	91%
La Gartempe à Montmorillon	48%

Les résultats les moins bons sont obtenus pour la Gartempe à Montmorillon avec 48% de données validées, et les meilleurs correspondent à la Creuse à Leugny avec 93% de données validées. Aucune des 7 stations dont nous disposons n'a été totalement validée. Mais les résultats sont plutôt bons pour les trois stations qui nous intéressent particulièrement : Nouâtre surtout, Ingrandes et Leugny.

Dans de nombreux cas, les données de débit maximal instantané qui sont erronées correspondent à une valeur attribuée au 1<sup>er</sup> jour du mois. On peut supposer que, dans ce cas de figure, le capteur a enregistré la valeur maximale sans lui assigner de date, et que les techniciens traitant les données, ne disposant pas de la date exacte de la crue, l'ont fixée au 1<sup>er</sup> jour du mois.

Il est parfois possible de retrouver la crue en question en s'appuyant sur les débits journaliers moyens, mais cela n'est pas toujours le cas. En outre, on n'est pas assuré que les données de débit journalier moyen soient de meilleure qualité : on ne peut que constater une discordance.