

Conférence de recyclage en Hydrologie

Juin 1967

ANALYSE du REGIME des PLUVES et des DÉBITS :
COURBES CHRONOLOGIQUES et MONOTONES,
AJUSTEMENTS STATISTIQUES

par

Pierre DUBREUIL

Directeur de Recherches ORSTOM

Chef du Département de la Recherche Appliquée
au Service Hydrologique de l'ORSTOM

Conférence du mercredi
7 Juin après-midi

1. BASES, METHODES et OBJECTIFS de l'ANALYSE des REGIMES de PLUIES et DEBITS

1.1 - Les bases de l'analyse

L'analyse des régimes des pluies et des débits s'effectue à partir des données d'observations élaborées et étendues (si nécessaire) dont il a été question dans une précédente conférence (1).

L'analyse des régimes revêt deux aspects, deux phases successives, comme l'indique le titre de cette conférence. Tout d'abord une phase classique de représentation par courbes chronologiques et monotones des données élaborées. Cette phase s'inscrit théoriquement sans solution de continuité derrière l'élaboration puisqu'elle ne procède vraiment qu'à une représentation et à un constat des formes élaborées. Au cours de cette phase, ne sont traitées que les données d'observations élaborées et critiquées, les lacunes de courte durée (quelques jours, au plus un mois, dans une série annuelle) y sont comblées, bien entendu, par corrélation, par analogie ou par interpolation.

Ensuite, la seconde phase procède à l'aide de l'outil "probabilités et statistiques". Cette phase n'est plus un constat mais une tentative d'épuiser toute l'information pour déboucher sur des conclusions et des prévisions. Le traitement intéresse alors toutes les données, y compris les données de séries courtes étendues selon les procédés décrits précédemment (1).

Le décalage entre les deux phases de l'analyse se reflète dans la pratique. La représentation chronologique ou monotone peut être considérée comme une activité continue, selon une périodicité annuelle (révision des moyennes interannuelles, des séries classées ... etc ...), susceptible de s'inscrire après l'élaboration et la critique des données brutes ; elle est réalisée tantôt par le service gestionnaire du réseau, tantôt par l'hydrologue d'analyse (s'il est distinct de ce service). La phase d'analyse statistique, elle, ne peut être opérée qu'après un laps de temps suffisant pour que les séries observées atteignent des durées point trop courtes ; la variance des phénomènes pluies et débits est telle que, même en FRANCE pays tempéré, la longueur minimale d'un échantillon est bien de 10 à

(1) "Elaboration, critique et extension des relevés hydro-météorologiques bruts".

15 ans si l'on ne veut pas aboutir à des résultats trop dispersés (1). La phase d'analyse s'intègre généralement avec les travaux d'extension qui procèdent des mêmes méthodes ; elle est l'oeuvre du service d'analyse hydrologique. A l'encontre de la phase classique, on ne procède à une nouvelle analyse que tous les 5 ans environ, au minimum, lorsque les longueurs d'échantillon ont été allongées de manière significative pour que la reprise de l'ajustement statistique puisse offrir un intérêt et avoir un sens (amélioration de la connaissance du régime) (1).

Il est même en pratique assez rare que l'analyse statistique des données hydrométéorologiques soit effectuée systématiquement. Les services d'analyse hydrologique n'en ont ni le temps ni les moyens, car il s'agit là de travaux longs que seuls peuvent mettre en oeuvre des spécialistes. Une telle exploitation systématique peut constituer un objectif idéal : réalisation d'une monographie hydrologique complète d'un bassin et révision périodique. Aujourd'hui encore cet objectif n'est pas atteint et l'on procède au coup par coup : l'hydrologue répond aux questions des utilisateurs à l'occasion des projets d'aménagements impliquant usage ou consommation d'eau.

1.2 - Les méthodes de l'analyse

Les courbes chronologiques et monotones sont établies selon des procédés simples et classiques, dont il est fait état lors de l'exposé qui suit ; mais il ne peut être question de parler de méthodologie spécifique.

Par contre, l'introduction des probabilités et de la statistique dans le traitement de l'information hydrométéorologique constitue un apport méthodologique particulier. Deux conférences viennent d'être consacrées à cet apport pour mettre en évidence :

- a) le caractère particulier de l'information hydrologique et l'adaptation conséquente des processus statistiques à son traitement (échelle de temps des variables, liaison ou indépendance de ces mêmes variables),
- b) le phénomène essentiel de l'échantillonnage réduit de l'information traitée et les conséquences qui en découlent (erreur d'échantillonnage, intervalle de confiance).

(1) Se reporter aux conférences sur le traitement statistique de l'information et les erreurs d'échantillonnage.

L'outil statistique est un outil dynamique et explorateur qui va plus loin que la représentation classique. Il ne faut cependant pas le considérer comme une panacée susceptible de répondre à toutes les questions des utilisateurs. L'ajustement statistique d'une variable hydrologique a ses limites : telle loi n'est compatible qu'avec telle variable, l'extrapolation au-delà des limites explorées est assez vite dénuée de précision.

1.3 - Les objectifs de l'analyse des régimes

L'analyse des données hydrométéorologiques vise à définir les régimes à travers certaines valeurs caractéristiques et représentatives. L'analyse dégage ces valeurs. L'hydrologue procède alors à deux démarches, selon le cas :

- a) l'une, purement analytique bien souvent, pour répondre à une question précise d'un utilisateur ; par exemple quel sera le débit maximal de crue de telle fréquence au site de barrage projeté ?,
- b) l'autre, à vocation plus synthétique, pour définir les régimes des pluies et des débits dans l'espace, selon une optique géographique de classification des régimes homogènes.

Les valeurs caractéristiques de régime sont de deux sortes : valeurs centrales et valeurs extrêmes.

Les valeurs centrales caractérisent l'abondance des régimes et leur irrégularité ; les couples les plus utilisés sont tantôt la médiane M_e et l'intervalle interquartile IQ , tantôt la moyenne $\bar{M}$ et l'écart-type S . Les premiers offrent un réel intérêt pour caractériser les échantillons très dispersés ou sur lesquels aucun ajustement n'est possible (choix des valeurs expérimentales).

Les valeurs extrêmes sont en hydrologie les débits de crues et d'étiages. En pluviométrie, les hauteurs journalières ou les intensités de précipitations, peuvent être considérées comme des valeurs centrales et extrêmes à la fois.

On limite le développement de cette conférence aux valeurs hydrologiques centrales et aux valeurs pluviométriques ; les étiages font, en effet, l'objet de développements particuliers en fin de session et il a été convenu que les crues ne seraient traitées qu'au cours de la prochaine session de recyclage.

2. REPRESENTATIONS CHRONOLOGIQUES et MONOTONES CLASSIQUES

Est dite monotone toute représentation ne tenant pas compte de la chronologie de l'évènement mais qui, dans un intervalle de temps fixé, considère l'ensemble des valeurs prises par l'évènement selon un classement croissant ou décroissant.

On peut dire, en quelque sorte, que les données élaborées, telles que nous les avons présentées dans une précédente conférence, constituent la première représentation classique. De là, certaines modifications simples permettent d'autres représentations : l'objectif étant ici d'offrir une - ou plusieurs - illustration la plus évocatrice des données élaborées sous divers angles. On conçoit ainsi qu'à côté des tableaux, les meilleures illustrations soient données par des graphiques (évocation chronologique ou monotone d'un phénomène ponctuel) et par des cartes (représentation spatiale de l'évolution d'un phénomène à partir de plusieurs relevés ponctuels).

On trouve dans les principaux manuels d'Hydrologie (Cf. Bibliographie) une description détaillée des principales courbes chronologiques et monstones. On se contente ici, en rappelant les plus intéressantes de ces diverses représentations, de formuler quelques remarques sur leur établissement et leur utilisation.

2.1 - Représentations chronologiques

A quelques exceptions près les formes sont les mêmes que l'on représente des pluies ou des débits.

2.1.1 - Le graphique chronologique

Le graphique classique le plus répandu est celui qui représente le déroulement chronologique, soit des pluies, soit des débits en une station, représentation des valeurs journalières, mensuelles ou annuelles. Cette représentation est typique de l'Annuaire Hydrologique ; on la trouve dans tous les ouvrages de climatologie. Elle offre une vision du phénomène permettant une définition du régime à la station dans le cadre d'une classification géographique : climat océanique, continental, ou régime pluvio-nival, tropical de transition ... etc ...

Cette représentation concerne tantôt une période déterminée, tantôt une moyenne de périodes ; on y fait figurer dans le premier cas les valeurs moyennes, dans le second les extrêmes observés ce qui permet de situer, d'une part l'abondance de la période, d'autre part la variabilité du régime.

Enfin, ce mode de représentation est bâti sur les valeurs absolues ou sur les valeurs relatives (en % ou ‰ du total annuel par exemple).

2.1.2 - Le réseau d'isohyètes

Plus intéressante peut être la carte des hauteurs de précipitations sur un bassin ou une région donnée ; le tracé des courbes isohyètes permet la mise en évidence de l'influence de facteurs comme l'altitude et l'exposition aux vents dominants. Un réseau d'isohyètes n'a de sens que s'il est tracé pour une durée d'observations égale en tous les postes pris en compte. Pour l'évaluation de la hauteur moyenne interannuelle de précipitations, cela nécessite généralement une homogénéisation des données (extension des séries courtes ...). La signification d'une courbe isohyète dépend également de la précision des valeurs ponctuelles, bases du tracé. Par exemple, pour un réseau interannuel, si l'on veut une garantie du tracé de 90 %, on doit considérer que chaque valeur ponctuelle $\bar{P}$ est susceptible de varier dans un intervalle de confiance à 90 %, défini par $\pm 2 S$, S étant l'écart-type de la série des $\bar{P}$. Ainsi pour une station ayant, sur 49 ans, une moyenne $\bar{P}$ de 750 mm et un écart-type de 175 mm, la vraie valeur de $\bar{P}$ se situe dans l'intervalle :

$$750 \pm \frac{2 \times 175}{\sqrt{49}} \quad (1) \quad (IC \ 90 \ \%)$$

soit entre 700 et 800 mm.

Dans ces conditions, le tracé des isohyètes inter-annuelles n'est possible que de 100 en 100 mm ; la position des intermédiaires, de 50 en 50 mm, n'est pas significative.

(1) La série $\bar{P}$ est supposée normale

2.1.3 - La courbe cumulée

Le phénomène observé n'est plus représenté par la succession de ses valeurs à un instant déterminé mais par le total cumulé de celles-ci sur une certaine période. La courbe cumulée est d'un emploi classique avec les débits; elle permet de présenter les volumes écoulés et son emploi est courant pour l'estimation du débit de régularisation ou de la capacité d'un réservoir.

2.2 - Représentations monotones

La plus classique est celle des débits moyens journaliers classés sur une année; c'est une illustration courante d'Annuaire en complément et en surcharge, du graphique chronologique. La courbe de débits classés offre par la prise en compte de certains points, d'abscisses choisies, la possibilité de caractériser le régime hydrologique: débits caractéristiques de n jours, n mois. Cette courbe est également utilisée dans l'établissement des projets de réservoir. Son seul inconvénient est son caractère monotone, car l'on néglige ainsi l'influence saisonnière très importante dans tous les problèmes d'équipement de cours d'eau.

Une autre représentation de nature monotone est celle des courbes intensités-durées des précipitations, puisqu'elle repose sur un découpage du hyétogramme en tranches de temps fixées a priori, découpage qui conduit à un classement par tranches. Nous avons déjà émis dans une précédente conférence quelques réserves sur ces courbes. Il faut ici faire état d'une extension spatiale des courbes intensités-durées qui permet l'établissement d'abaques "intensités-durées-surfaces".

Ce procédé de traitement des précipitations a deux inconvénients :

- a) il n'exploite que les relevés effectués en S stations durant k années, négligeant l'existence possible de quelques stations observées durant $n > k$ années,
- b) il mélange, selon la règle dite des stations-années, des observations non indépendantes de hauteurs de précipitations recueillies en des postes voisins, sans en tenir compte sur le plan des fréquences.

La réponse au problème : sachant qu'il tombe p millimètres d'eau en une station S pendant le temps T avec la fréquence F , combien de millimètres $p' < p$ recueille-t-on sur une surface quelconque ayant le même régime de précipitations que la station S ? , ne peut être valablement fournie par le faisceau de courbes I-D-S. Ce problème de l'abattement requiert à la fois :

- a) un classement des événements survenus sur la surface étudiée avec mise en évidence du rapport des hauteurs ponctuelles observées à la hauteur moyenne calculée,
- b) une analyse statistique de la hauteur ponctuelle au poste pluviométrique représentatif de la surface et observé sur une longue durée,
- c) une extension des résultats de l'étude des pluies sur la surface à l'aide de l'ajustement statistique fourni par la série ponctuelle.

La méthodologie de l'étude de l'abattement a été bien améliorée récemment, mais elle ne peut pas encore être considérée comme définitive tant qu'elle n'aura pas été employée sur de nombreux échantillons d'origines diverses ; des modifications de détail sont prévisibles.

L'étude de l'abattement sort du domaine du simple constat ; elle est dynamique et cherche à répondre à une question difficile. Son dynamisme provient du fait qu'elle est déjà, et en partie, une méthode statistique.

3. Les AJUSTEMENTS de LOIS de PROBABILITE sur des VARIABLES INDEPENDANTES

L'hydrologue ne peut se contenter du constat offert par les représentations chronologiques et monotones classiques.

Essayer d'ajuster une loi de probabilité sur un échantillon de longueur réduite, c'est avoir pour objet d'approcher la population infinie contenant cet échantillon, en faisant le postulat d'homogénéité entre l'échantillon et la population (invariance dans l'échelle temps). Les limites de signification de cet ajustement vous ont été exposées.

A partir de cet approche de la population d'une quelconque variable hydrologique, on peut répondre à une foule de questions sur le comportement de cette variable dans des conditions déterminées à l'avance.

La plupart des problèmes posés par l'emploi de la statistique en hydrométéorologie peuvent être évoqués en examinant la réponse à la question : définir la valeur prise par un caractère hydrologique ou pluviométrique pour une probabilité fixée à l'avance.

3.1 - L'indépendance des variables

Avant toute chose, il faut rappeler qu'à ce premier stade, simple et classique, d'emploi de la statistique, l'hydrologue ne travaille que sur des variables indépendantes.

Les caractères hydrométéorologiques sont rarement indépendants surtout dans leur réalisation instantanée. L'indépendance est admissible pour les valeurs moyennes (ou totales) relatives à une certaine période : l'année ou le mois (en considérant les mois séparément et non tous ensemble). On peut l'admettre mais l'on verra plus loin (paragraphe 4) qu'en stricte logique cette indépendance est peut-être relative. A l'échelle journalière, il est évident que l'on ne peut parler d'indépendance ni entre les débits ni entre les pluies de 2 jours consécutifs.

Cependant, tout est une question de définition de variable. Considérer comme variable, par exemple, la hauteur de précipitations tombée en un jour quelconque de l'année, c'est admettre implicitement que les précipitations pouvant survenir tous les jours de l'année appartiennent à la même population. Ce qui est faux étant donné l'existence d'un effet saisonnier. Au contraire, si l'on désire seulement connaître la hauteur de précipitations de fréquence donnée, la définition de la variable est restreinte, mais on peut la considérer comme indépendante. Seul inconvénient : la loi de probabilité ainsi ajustée ne peut délivrer d'échantillon fictif par tirage au sort dans une table de nombres au hasard. On ne peut donc pas reconstituer la chronique de l'événement. La perte d'information est comparable à celle qui, par destruction de la chronologie, résulte de l'emploi des représentations monotones.

Une même restriction dans le choix de la variable "la valeur prise pour une fréquence donnée" permet d'admettre l'hypothèse de l'indépendance pour toutes variables hydrologiques ou pluviométriques.

Dans le processus de définition de la valeur d'une variable pour une fréquence donnée, il y a lieu de s'arrêter sur 3 points particuliers méritant un développement, sachant que erreur d'échantillonnage et intervalle de confiance ont été traités précédemment. Ces points sont les suivants : choix de la loi de probabilité, test d'adéquation de la loi sur l'échantillon, signification d'une faible probabilité.

3.2 - Choix de la loi de probabilité

Deux remarques préalables et fondamentales :

- a) selon le théorème central limite, une combinaison linéaire de N variables aléatoires est distribuée normalement quand N tend vers l'infini, quelle que soit la loi propre de ces variables, si elles sont indépendantes et si leurs valeurs ne sont point trop différentes,
- b) en dehors de la loi normale, on ne peut guère traiter les problèmes d'intervalle de confiance, de gain dans les extensions de série, de corrélations entre variables.

Ces deux remarques appellent une conclusion : on essaye toujours d'ajuster une loi normale et parmi les lois dissymétriques, si nécessaire, on prend de préférence celles qui permettent un retour à la normalité par changement de variable : logarithme (loi de Galton) ou racine nème.

Les variables, somme de variables, ont d'autant plus de chance d'être distribuées normalement que leurs composantes sont nombreuses et point trop différentes, c'est-à-dire que les régimes sont abondants.

En d'autres termes, que la variable ait pour base le débit moyen journalier ou la hauteur journalière ponctuelle de précipitations, la loi de probabilité ajustable à l'échantillon est d'autant plus près de la normalité que :

- a) le régime des pluies ou des débits est abondant,
- b) la variable concerne une longue échelle de temps.

Le tableau joint évoque pour les principales variables hydrologiques et pluviométriques les lois de probabilité qui sont généralement utilisées (simple évocation non-exhaustive).

Quand on procède à la recherche d'une loi pour un échantillon dans une région mal connue, ou n'ayant pas encore fait l'objet d'un tel travail, on peut faire un essai sur papier à échelle gaussique par report des points expérimentaux (fréquence expérimentale égale à :

$$\frac{R}{N+1} \quad \text{ou} \quad \frac{R-1}{N}$$

avec le rang R et le nombre de valeurs n).

La dissymétrie plus ou moins accusée, permet de guider le choix vers une loi satisfaisante.

Entre plusieurs lois possibles, il faut choisir celle dont le nombre de paramètres est le plus réduit et dont le calcul desdits paramètres est le plus aisé.

L'étude d'une variable pour un mois déterminé peut se heurter à l'effet saisonnier. Pour un mois de transition, dans les régimes pluvieux à saisons bien tranchées, l'échantillon peut ne pas être homogène et comprendre p valeurs de saison des pluies et q valeurs de saison sèche selon que les pluies ont commencé en avance ou en retard. Cette hétérogénéité introduit une dissymétrie visible sur le polygone des fréquences par l'apparition de 2 modes ; il faut traiter chaque série indépendamment en prenant une fréquence expérimentale pour le passage de l'une à l'autre (rapport q/p ou p/q précédent).

Le choix d'une loi de probabilité étant arrêté, il faut estimer les paramètres de cette loi à partir de l'échantillon. Cette estimation, à laquelle l'on procède soit par la méthode du maximum de vraisemblance, soit par l'emploi d'estimateurs, donne des résultats qui sont évidemment entachés de l'erreur d'échantillonnage. De plus dans certaines séries très dispersées, l'emploi d'estimateurs de puissance élevée (moments d'ordre 3 ou 4, par exemple)

QUELQUES EXEMPLES d'AJUSTEMENT sur des VARIABLES INDEPENDANTES

Définition de la variable	Régime	Loi de probabilité
Hauteur annuelle de précipitation P	Abondant (océanique, équatorial)	Normale
ou		
Module de cours d'eau Q	Déficitaire (méditerranéen, tropical, aride)	Dissymétrique : Pearson III Retour à la normalité possible avec : Galton, $\sqrt[n]{P}$
Hauteur mensuelle de précipitations	Très abondant	Normale (cas rare)
ou		
Débit moyen mensuel	Abondant ou déficitaire	Dissymétrique Loi tronquée (si nombreuses valeurs nulles) Loi bimodale (mois de transition)
Hauteur journalière de précipitations	Quel qu'il soit	Loi dissymétrique : Pearson III, Goodrich, Galton, ...
Débit maximal de crue	- id -	Loi dissymétrique : Pearson III, Goodrich, Galton, ... Loi des valeurs extrêmes : Gumbel (ou Fréchet)

peut aboutir à des valeurs inadéquates des paramètres. Ainsi pour une loi de Galton, il est souvent préférable de chercher graphiquement le premier paramètre Z_0 à ajouter à la variable Z pour que son logarithme suive une loi normale $[Y = a \log (Z-Z_0) + b]$.

Du meilleur procédé d'estimation des paramètres découle bien souvent une meilleure adéquation de la loi et un ajustement plus facile à mettre en oeuvre. On ne peut pas dire que ces problèmes d'estimation soient tous résolus ; des modifications et des améliorations sont encore possibles surtout pour les lois dissymétriques (Goodrich, Pearson, Gumbel et leurs variantes).

Les méthodes d'estimation sont du domaine de la statistique pure, mais le caractère très particulier des échantillons de l'hydrologue impose à ce dernier un travail personnel de recherche des meilleurs estimateurs.

3.3 - L'adéquation de la loi sur l'échantillon

En exagérant légèrement, on pourrait dire que sur un échantillon de taille réduite (c'est toujours le cas en hydrologie, où les séries dépassent rarement quelques dizaines d'années) doté d'une variance importante, une loi de probabilité est toujours ajustable et que seule la multiplication de la taille, par 2, 3 ou plus, permettrait de vérifier l'adéquation de cet ajustement.

Or, d'échantillons de 10, 20 ou 30 ans, on essaye de tirer des conclusions centenaires et millénaires. L'extrapolation est d'autant plus hasardeuse que les diverses lois divergent généralement entre elles pour les faibles fréquences, alors qu'il est difficile de choisir la plus adéquate au vu d'ajustements sur les échantillons réduits possédés.

Le test d'adéquation le plus communément utilisé est déduit de l'étude d'une distribution du χ^2 de PEARSON. Il a fait l'objet d'abondants développements dans la littérature. La trop grande souplesse d'emploi rend peu significative la conclusion que l'on en tire. Pour les échantillons réduits de l'hydrologue, quelques règles d'usage s'imposent :

- a) le découpage en classes, s'il n'est pas imposé par un échantillon classé avant analyse, laissé à l'arbitraire peut apporter des probabilités $P(\chi^2)$ très variables donc non significatives. On recommande un découpage en classes égales en probabilités théoriques, avec un minimum d'effectif de 5 valeurs théoriques par classe,
- b) la perte d'information due au groupage en classes de l'échantillon peut être compensée en effectuant, en complément, un test des suites (plus de 100 valeurs au moins pour qu'il soit indépendant de celui du χ^2), ou un test des signes,
- c) la probabilité $P(\chi^2)$ sur un échantillon de population inconnue représente l'adéquation ; elle a donc 50 % de chances de n'être pas comprise entre 0,25 et 0,75, 10 % de chances de n'être pas comprise entre 0,05 et 0,95. Une trop forte probabilité est aussi troublante qu'une trop faible.

Le choix du seuil d'acceptation ou de refus d'un ajustement est libre ; on peut le faire varier avec la qualité des observations ; il est de coutume de prendre 0,05.

Le test du χ^2 n'est qu'un test. Le négliger est une erreur ; tenir trop rigide ment compte des conclusions que l'on en tire, également. Mais son intérêt est indéniable non seulement pour l'adéquation mais aussi, par exemple pour la comparaison d'échantillons supposés appartenir à une même population (test d'homogénéité).

3.4 - Signification d'une faible probabilité

Après choix d'une loi, ajustement et test d'adéquation, on peut enfin répondre : pour telle fréquence, la variable indépendante étudiée vaut telle valeur, à l'intérieur de tel intervalle de confiance.

Il nous semble utile de préciser la signification d'une faible probabilité p (1/100 et moins). Pour la rendre claire, on parle souvent de récurrence, ou période de retour, R - inverse de la fréquence F .

Mais attention, si la crue de fréquence 1/1000° (millénaire) a une récurrence de mille ans en moyenne, cela ne veut pas dire que cet événement apparaisse régulièrement avec cette périodicité. Bien au contraire, cette apparition reste aléatoire et irrégulière sans que cela détruise les hypothèses d'ajustement.

Le nombre N d'années, séparant 2 apparitions successives d'une valeur supérieure à la valeur X de la variable pour la probabilité p, est une variable aléatoire qui, pour p petit, obéit à :

$$\text{Prob } (N \leq n) = 1 - e^{-np}$$

L'évènement X de probabilité $p = 0,001$ à 63 % de chance de survenir en 1 000 ans, 39 % en 500 ans, 18 % en 200 ans ... etc ... Le risque d'apparition d'un événement rare n'est pas négligeable pour une certaine période fixée. C'est un point à prendre en considération par exemple lors de l'exécution de grands travaux en rivière (risque d'une crue durant la période de dérivation ...).

4. Le PROBLEME de la DEPENDANCE des VARIABLES

Admettre l'indépendance des variables, c'est se limiter à rechercher la valeur prise pour une certaine fréquence. Il peut être intéressant de pousser plus loin la connaissance du régime de ces variables hydrologiques et pluviométriques. On peut, en effet, désirer conserver la chronique de l'évènement et pouvoir la reconstituer par tirage au sort. La loi de probabilité à prendre en considération n'est plus simple car la variable, dans sa pleine acceptation - pluie ou débit - n'est plus alors indépendante.

La liaison entre les débits d'une rivière au cours de journées successives est évidemment très serrée. C'est pourquoi tout ajustement d'une loi de probabilité sur la série annuelle des débits classés conduit à des fréquences fausses sous l'angle probabiliste.

En matière d'analyse de débits, on travaille très rarement sur les débits journaliers ; plus couramment on s'intéresse aux débits mensuels ou aux modules. Les débits

mensuels successifs ne sont pas indépendants (effet saisonnier) ; les modules paraissent l'être bien que certains doutes soient possibles.

Les mêmes remarques affectent les variables "pluies". La liaison entre hauteurs journalières de précipitation successives bien que lâche existe. L'effet de persistance (un jour de pluie en appelle un autre, un jour sec également appelle un autre jour sec) est communément admis. Sur les hauteurs mensuelles et annuelles, l'effet saisonnier et l'indépendance relative peuvent être également indiqués.

Admettre la dépendance des variables, c'est reconnaître qu'elles sont liées au temps. Une suite de réalisations d'une variable aléatoire $X(t)$ fonction du temps, pour des intervalles constants de celui-ci, est une série chronologique ordonnée en processus stochastique.

Une série chronologique est la superposition d'une tendance à long terme, d'une composante cyclique et d'une composante aléatoire. Impossible sur les faibles échantillons hydrologiques de déceler une tendance (d'où le postulat conservé de l'invariance climatique ...) ; inutile de rechercher la composante cyclique car elle apparaîtrait toute seule (méthode de la moyenne mobile, effet SLUTSKY ...) et n'a donc pas de sens si une cause physique ne la justifie a priori (rare pour les variables hydropluviométriques) ; reste la composante aléatoire qui est primordiale en hydrologie.

Alors que les représentations chronologiques et monotones classiques demandent peu de moyens à l'hydrologue (une feuille, un crayon, une règle à calcul), les ajustements statistiques de variables indépendantes ne sont possibles qu'en procédant aux calculs sur machines, électromécaniques au moins. La mise en oeuvre des processus stochastiques, représentatifs des variables liées au temps, exige un pas de plus dans l'escalade des moyens : la calculatrice électronique, utile à la statistique simple, est indispensable. Cette limitation matérielle et la difficulté de maniement de ces processus expliquent que leur mise en oeuvre soit très récente en hydrologie.

Les modèles de processus stochastiques sont nombreux, du simple au plus complexe selon que l'effet d'auto-corrélation remonte plus ou moins loin dans le temps.

On utilise maintenant divers processus pour l'étude des pluies journalières ou des débits (succession d'étiage, phases sèches, phases pluvieuses ...). Le schéma de dépendance est auto-regressif, de la forme :

$$X_{t+2} = a X_{t+1} + b X_t \dots + \varepsilon_{t+2}$$

avec ε variable normale centrée réduite indépendante de ses états antérieurs.

La réalisation future de la variable dépend de l'état présent seul (chaîne de MARKOFF) ou également des états passés.

Une dépendance de 5 jours a été admise pour le régime des précipitations dans le Sud-Est de la FRANCE.

Les difficultés d'emploi des processus stochastiques tiennent d'une part à la complexité des calculs d'estimation des paramètres, d'autre part au manque de test d'adéquation significatif. L'erreur d'échantillonnage intervient également sur la confiance à accorder aux résultats des calculs basés sur des séries de faible durée.

Cependant, l'intérêt de ces processus est certain ils permettent une analyse plus fine des phénomènes pluies et débits ; ils offrent la possibilité de reconstituer des échantillons fictifs pour des opérations de simulation (gestion de réservoir). Leur mise en oeuvre, appuyée par l'expérience des hydrologues praticiens, doit s'améliorer régulièrement et devenir monnaie aussi courante que les lois dissymétriques, l'usage de l'ordinateur étant sous-entendu.

5. CORRELATIONS PLUIES-DEBITS

Il ne faut pas que le développement de cet exposé laisse croire que l'analyse des régimes des pluies et des débits est menée, selon des procédés identiques, sur l'une ou l'autre des séries de variables sans se préoccuper des liaisons qui existent entre elles.

La liaison pluie-débit est bien connue. Dès la mise en oeuvre des données hydrométéorologiques brutes, elle est utilisée pour procéder aux extensions de séries hydrologiques courtes à partir de séries pluviométriques longues.

Ce paragraphe n'a donc pour objet que de rappeler l'existence de cette liaison et son intérêt en hydrologie, intérêt double pour le moins :

- permettre l'extension des données, donc valoriser l'information,
- offrir une explication du phénomène de l'écoulement (apparition, intensité, déficit et coefficient d'écoulement etc ...).

Au niveau de l'ajustement statistique, l'objectif est de travailler sur une variable normale ou susceptible d'être réduite à la normale par changement de variable. Au niveau de la corrélation, il en est de même.

La richesse d'une corrélation n'est totalement exploitable (gain d'années, intervalle de confiance ...) que si cette liaison, cette régression est linéaire. La chose va de soi lorsque les variables considérées sont normales ; rien n'est évident quand elles sont dissymétriques. Certaines conditions peuvent être édictées pour justifier la linéarité d'une régression entre deux variables suivant des lois de Pearson III, par exemple. Cependant, dès que les variables sont par trop dissymétriques, la régression n'est plus linéaire et les méthodes de corrélation à employer sont purement graphiques (déviations résiduelles ...).

Quelles que soient la nature de la corrélation et la méthode employée, il faut bien avoir présent à l'esprit le caractère probabiliste de la corrélation déduite de la confrontation de deux échantillons réduits. Ce caractère se manifeste, en quelque sorte, comme suit :

- a) l'équation de régression, comme les paramètres estimés dans un ajustement de loi, n'est qu'une estimation de l'équation liant les deux populations. On peut calculer et définir l'intervalle de confiance autour de la droite de régression à l'intérieur duquel on a un

certain pourcentage de chance de trouver la vraie droite,

- b) prendre des valeurs de la variable Y, de série courte, à l'aide de l'équation de régression la liant à X, de série longue, pour les valeurs de X, sans concomitance de Y, c'est négliger la dispersion résiduelle. Il importe de redistribuer au hasard, autour de la droite de régression, les points ainsi calculés. On peut employer la méthode de MONTE CARLO pour effectuer un tirage au sort dans la loi de probabilité conditionnelle, ce qui fournit un certain résidu fictif à ajouter à la valeur déduite de la régression. Une autre méthode permet le calcul direct des nouveaux paramètres de la série Y (estimation par la méthode du maximum de vraisemblance) et du gain d'information, sans négliger la dispersion résiduelle (emploi de la notion de quantité d'information de discrimination).

Toutes les méthodes de corrélation calculées nécessitent des moyens matériels importants (calculatrice électronique). Les méthodes graphiques sont plus souples mais elles ne sont pas généralisables et ne permettent pas l'estimation du gain lors d'une extension.

A côté de la corrélation simple entre deux variables, il faut mentionner les corrélations multiples d'un grand intérêt en hydrologie analytique pour la recherche des facteurs conditionnels de l'écoulement et l'estimation du poids relatif de chaque facteur dans la corrélation. Le module ou le débit de crue d'une rivière dépend de nombreuses variables : précipitation sur le bassin, températures, état de saturation préalable des terrains ... etc ... Si la méthode des déviations résiduelles permet l'approche graphique, il est préférable de recourir au calcul quand on peut linéariser l'ensemble des corrélations partielles.

Dans certains cas, le choix des facteurs conditionnels n'est pas évident. Lorsque l'on procède à des expérimentations en hydrologie (simulation de pluie, modification du couvert végétal ...), la recherche de l'influence significative de tel ou tel facteur, soupçonné être conditionnel, se fait en appliquant les méthodes d'analyse de variances. Attention, l'analyse de variance n'est possible et utile que si un plan d'expériences a été dressé a priori, posant clairement le problème et les facteurs à tester. Il s'agit là d'un domaine assez particulier de l'hydrologie analytique, qui n'a pas encore été fortement exploré. Mais, dans l'avenir, la recherche hydrologique s'orientera certainement de plus en plus dans cette direction pour expliquer le phénomène "débit" qu'elle se contente, aujourd'hui, d'observer, de représenter ou d'interpréter sans recours systématique aux causes.

CONCLUSION

Ce texte, sans illustration, a pu paraître aride et terne. Il nous a semblé que la reproduction d'exemples d'ajustement, de corrélation ou de représentation chronologique et monotone offrait un intérêt limité puisque de tels procédés sont abondamment décrits dans la littérature hydrologique. Certes, cette littérature est d'une telle abondance aujourd'hui qu'il est difficile de s'y retrouver d'autant que les ouvrages les plus généraux ne peuvent prétendre être complets et se trouvent souvent périmés dès leur sortie tellement progresse vite la science hydrologique. Les développements les plus récents sur les dernières méthodes doivent être recherchés dans la littérature spécialisée : communications aux réunions techniques de divers organismes, notes de calcul pour tel projet ... etc ... L'exploration en est difficile. La bibliographie que nous proposons pour pallier notre "mutisme" ne peut donc prétendre être exhaustive. Suffit-il pour être excusé que notre conclusion affirme le caractère dynamique et en perpétuel renouvellement de la science hydrologique d'aujourd'hui ?

BIBLIOGRAPHIE INDICATIVE

Nota : Se reporter aux indications correspondantes des deux conférences relatives au traitement statistique de l'information et aux erreurs d'échantillonnage pour les références bibliographiques concernant les probabilités et les statistiques en général.

A - OUVRAGES GENERAUX

1. W.M.O. "Guide to Hydrometeorological practices" n° 168 TP 82 - 1965

L'analyse hydrologique y fait l'objet d'une annexe A de 67 pages d'un contenu classique assez complet, dans lequel prédomine les conceptions américaines un peu systématiques.

2. H. CRISOLLET, B. GUILMET, R. ARLERY "Climatologie, méthodes et pratiques" Edition Gauthier-Villars - Paris 1962

La deuxième partie de cet ouvrage est entièrement consacrée au "Traitement statistique des observations" ; elle est assez complète et très claire quant au choix des lois de probabilités, au calcul des paramètres et à leur signification. Elle est peut-être insuffisamment développée en ce qui concerne les lois dissymétriques, les séries chronologiques et les corrélations multiples.

3. G. NEMENIERAS "L'Hydrologie de l'Ingénieur" Edition Eyrolles

La présentation des courbes chronologiques et monotones classiques y est complète et bien illustrée ; on s'y reportera avec fruit tant pour les pluies (chap. II.C) que pour les débits (chap. VII.A).

4. M. ROCHE "Hydrologie de surface" Edition Gauthier-Villars et ORSTOM

Une introduction assez détaillée y aborde l'emploi de la statistique et du calcul des probabilités en hydrologie. Des exemples d'application sont largement décrits dans les chapitres I "Précipitations" et X "Etude des modules".

5. M. ROCHE "Interprétation et utilisation des données hydrologiques" Edition Riber - Cours professé à l'ENGREF - 1966

Des exemples d'ajustement statistique et des réflexions sur l'indépendance des variables peuvent être utilement consultés dans les chapitres II "Distribution statistique des pluies ponctuelles", III "La pluie dans l'espace" (problème de l'abattement) et VI "Extension des données hydrologiques pour la connaissance des apports".

6. J. BERNIER "Cours sur les applications des méthodes statistiques aux travaux de recherche expérimentale" C.R.E.C. - E.D.F.

Toutes les méthodes statistiques y sont exposées dans leur logique ; le passage aux problèmes purement hydrologiques est facile ; ce cours est une synthèse habile et utile des ouvrages généraux et classiques de la statistique dont celui de MORICE et CHARTIER (INSEE) "Méthode Statistique" est le meilleur exemple.

B - QUELQUES REFERENCES SPECIFIQUES

7. M.G. KENDALL et A. STUART "The Advanced theory of statistics" Edition ch. Griffin et Cie, Londres

Dans cette source, on a utilisé ici tout particulièrement l'exposé du volume II sur la théorie du test du χ^2 .

8. Y. BRUNET-MORET "Le test du χ^2 de PEARSON" ORSTOM, Service Hydrologique Note technique n° 3. Diffusion interne - Novembre 1966

Cette note développe le point de vue de Kendall dans l'optique hydrologique.

9. J. BERNIER, J.F. MALHERBE "L'analyse fréquentielle régionale des variables hydrologiques" C.R.E.C. - E.D.F. (Comm. des Débits de la S.H.F. Avril 1966)

Nouvelle méthode d'extension des séries courtes avec estimation des paramètres par le maximum de vraisemblance et utilisation de la quantité d'information de discrimination.

10. Y. BRUNET-MORET, M. ROCHE "Etude théorique et méthodologique de l'abattement des pluies" ORSTOM - Cahiers série Hydrologie n° 4 - Mai 1966

Le dernier état des recherches sur l'abattement des précipitations.

11. Y. BRUNET-MORET "Influence du corps de l'averse sur le ruissellement d'un petit bassin" ORSTOM - Cahiers série Hydrologie n° 3 - Octobre 1965

Un exemple démonstratif de la souplesse des méthodes graphiques pour résoudre les problèmes de corrélation multiple (facteurs conditionnels de l'écoulement).

12. GRARD "La prévision des crues de l'Ain" S.H.M. Alpes E.D.F. (Comm. des Débits S.H.F. Mai 1967)

Un exemple d'application de la méthode de calcul des corrélations partielles et multiples pour l'estimation des causes de crues.

13. D. DUBAND "Composition des lois de probabilités des précipitations journalières" E.D.F. - D.T.G. Grenoble (Comm. des Débits S.H.F. Mai 1967)

Une tentative d'introduction de la persistance (processus d'auto-regression) dans l'étude des pluies selon un schéma de Pearson III.