REVUE DE STATISTIQUE APPLIQUÉE

J. LARRIEU

Méthodes d'analyse de la structure fine des débits

Revue de statistique appliquée, tome 4, nº 2 (1956), p. 39-62
<http://www.numdam.org/item?id=RSA_1956__4_2_39_0>

© Société française de statistique, 1956, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (http://www.sfds.asso.fr/publicat/rsa.htm) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

MÉTHODES D'ANALYSE DE LA STRUCTURE FINE DES DÉBITS

par

J. LARRIEU

Électricité de France
Service des Études et Recherches Hydrauliques

Quand Etienne HALPHEN entreprit en 1942, l'étude systématique des grandeurs hydrologiques, il fixa implicitement deux stades à ce travail.

On étudierait tout d'abord les grandes structures de l'hydrologie, débits annuels, mensuels, corrélation entre ensembles géographiques voisins. Cette première étape permettrait d'une part de se faire une idée nette du phénomène global et d'autre part d'utiliser les méthodes statistiques classiques.

HALPHEN avait bien vu en effet que l'étape suivante, celle de l'étude de la structure fine exigerait des outils statistiques nouveaux, adaptés à la nature très complexe des phénomènes.

En fait, la première partie du travail s'avéra plus difficile qu'il ne semblait a priori, et elle dura plus de huit ans. Huit années d'ailleurs très fructueuses puisqu'elles permirent la mise au point d'un grand nombre de techniques particulières, et la connaissance de faits jusque là inconnus. C'est dans cette période que naquirent la loi harmonique puis les lois A et B.

En 1953, HALPHEN aborda le second point de son programme, l'étude de la structure fine des débits. Il reprit à cet effet des techniques utilisées auparavant, mais sa disparition prématurée en 1954, l'empêcha d'en tirer tous les résultats qu'il pouvait légitimement en attendre.

Dans cette communication, nous voudrions rappeler les méthodes classiques qui permirent l'étude des structures grossières et montrer en quoi ces méthodes s'avèrent insuffisantes quand on veut étudier la structure fine des débits et leur répartition dans le temps.

Nous indiquerons, ensuite, les travaux qui ont utilisé des méthodes nouvelles pour cette analyse et nous montrerons pourquoi il était nécessaire de chercher d'autres procédés que ceux-là.

Enfin, nous donnerons les principes du nouveau mode d'approche conçu par HALPHEN; nous montrerons la façon dont nous exploitons cette méthode actuellement et nous dirons ce qu'il faut vraisemblablement en attendre.

NATURE STATISTIQUE DU PROBLÈME

Disons tout d'abord qu'il nous semble vain, voire dangereux, de vouloir trouver une signification profonde "métaphysique" aux schémas proposés. Ce serait

admettre qu'on peut connaître la nature intime des phénomènes. C'est vrai en Mathématiques où l'on se donne toujours, a priori, un certain nombre d'axiomes rigoureux; on peut alors valablement contrôler tout au long d'un raisonnement que les axiomes se trouvent vérifiés. Dans les sciences expérimentales où les axiomes ne nous sont pas connus, c'est chose très difficile. Tout au plus pourrons-nous déclarer dans certains cas simples : "Tel schéma représente bien le phénomène étudié, tel schéma le représente mal".

En outre, dire valablement que tel schéma représente bien tel phénomène implique l'une des deux conditions suivantes :

ou bien on possède des éléments de comparaison, soit dans le passé, soit dans le présent, avec des phénomènes de même nature où le schéma s'est avéré convenable,

ou bien on se trouve dans les conditions d'application de théorèmes généraux de la statistique permettant de justifier l'application à des phénomènes complexes de résultats simples. (Loi des grands nombres par exemple).

C'est là une évidence, peut-être, mais parfois oubliée car on est souvent conduit à utiliser des schémas sans aucun moyen de comparaison (Exemple, estimation de crues millénaires à partir de relevés portant sur quelques années) ou sans justification théorique suffisante (loi des débits maxima de GUMBEL).

ÉLÉMENTS STATISTIQUES UTILISÉS GÉNÉRALEMENT EN HYDROLOGIE

Les statisticiens hydrologues sont accoutumés à manier des débits journaliers, mensuels, annuels. Usuellement ces éléments sont considérés comme des variables aléatoires obéissant à des schémas classiques. D'une façon plus précise, on admet souvent que le débit d'un mois donné (si on s'intéresse aux débits mensuels) est tiré "au hasard" dans une urne dont la composition dépend essentiellement du mois considéré.

Il se pose alors un certain nombre de problèmes. Comment déterminer, "estimer" diront les statisticiens, l'urne dont nous tirons les débits ? Comment déterminer la période sur laquelle sera calculé le débit ?

ESTIMATION DES PARAMÈTRES

En admettant l'adéquation du schéma simpliste, "variable aléatoire tirée d'une urne" on peut se demander comment définir cette urne, c'est-à-dire comment définir la forme de la fonction de répartition à retenir et ensuite comment en déterminer les paramètres.

FORME DE LA FONCTION DE RÉPARTITION

Il n'existe pas en général de méthode pour adopter une forme de loi et l'intuition du statisticien joue ici un grand rôle.

La plupart du temps c'est l'allure du polygône de fréquence de l'échantillon étudié qui suggère la forme analytique convenable pour F (x) ou l'habitude du statisticien qui a déjà rencontré des variables du même genre. Si le polygône est

symétrique on adopte la plupart du temps une loi de GAUSS en raison des facilités analytiques qu'elle offre aux calculs.

Si le polygône est dissymétrique on adopte souvent une loi de GALTON ou une loi de PEARSON.

Admettons maintenant que la forme analytique de la fonction F soit déterminée. Il reste à trouver les valeurs des paramètres qui y figurent. Par exemple, dans la loi de GAUSS il faut estimer les valeurs de deux paramètres, la moyenne m et l'écart-type σ.

Nous n'allons pas faire ici une théorie de l'estimation, nous bornant simplement à signaler quelques difficultés généralement connues en théorie mais parfois négligées dans les applications.

On sait qu'on estime généralement m et σ à partir des formules suivantes :

$$m = \frac{\sum x_i}{n} \qquad \sigma = \sqrt{\frac{\sum (x_i - m)^2}{n}}$$

Or, on a parfois tendance à oublier les conditions de validité de ces formules.

Nous signalerons seulement les deux plus importantes :

1) Les observations effectuées sont indépendantes.

2) Elles appartiennent toutes à la même loi.

INDÉPENDANCE DES OBSERVATIONS

Les formules d'estimation usuellement employées supposent que l'échantillon étudié se compose d'éléments indépendants. S'il n'en est pas ainsi l'estimation de la moyenne reste valable, mais il n'en est plus de même pour l'estimation des moments d'ordre supérieur. En particulier l'estimation de l'écart-type est modifiée. Dans un cas de ce genre le fait d'utiliser un grand nombre d'observations donnera une fausse impression de sécurité.

HOMOGÉNÉITÉ DES OBSERVATIONS

Admettons maintenant que les débits considérés soient indépendants, ce qui n'est vrai d'ailleurs pour les débits moyens mensuels qu'en première approximation. Il ne serait pas correct pour autant de faire une estimation de paramètres à partir de leur totalité.

Il est, en effet, raisonnable de supposer que les débits mensuels suivent tous des lois de même forme analytique, par exemple des lois de GALTON. Mais il serait tout à fait inexact d'admettre que **tous** les débits moyens mensuels suivent la **même** loi de GALTON. La variation saisonnière des débits est bien connue; dans un régime nival par exemple, les débits d'été ont une moyenne plus grande que les débits d'hiver (Voir figure 1).

Par conséquent, mélanger l'ensemble des mois de l'année donnerait un résultat dénué de signification statistique.

Il convient donc dans le cas particulier que nous donnons comme exemple d'étudier la longueur des périodes dont on peut mélanger les débits mensuels. La plupart du temps, on considère chaque mois séparément. Quelquefois cependant,

on coupe l'année en deux et on mélange les mois d'hiver d'une part, les mois d'été d'autre part.

C'est d'ailleurs là une schématisation arbitraire car le passage d'une catégorie à l'autre ne s'effectue pas brutalement mais progressivement.

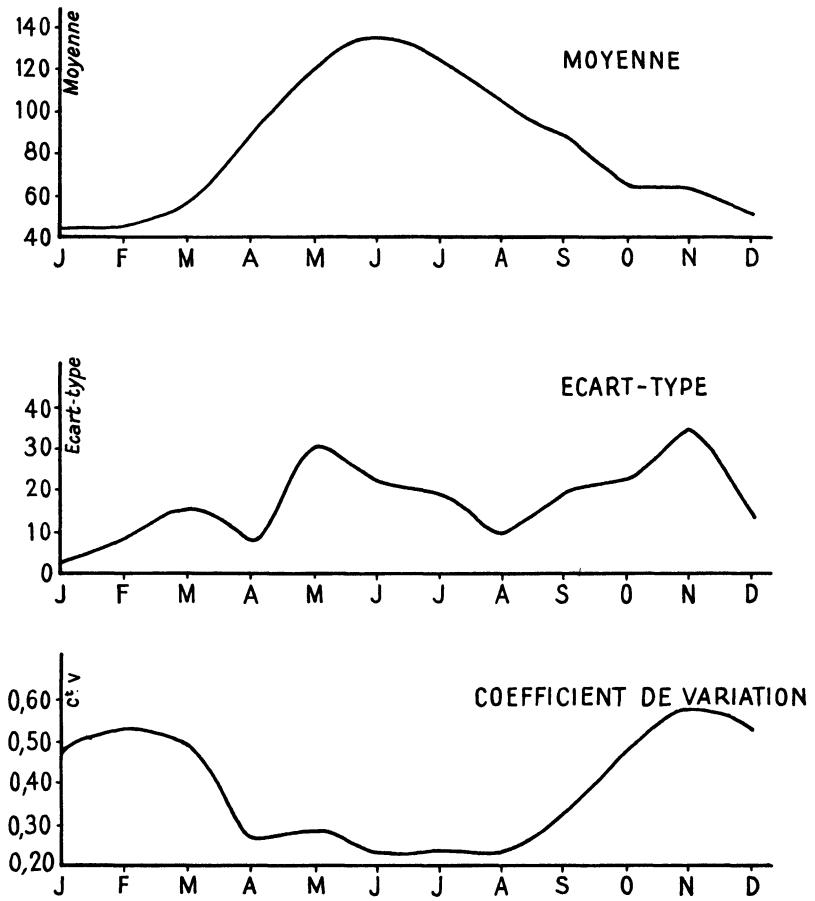


Fig. 1

Arve au Pont de Carouge

Variation des paramètres caractéristiques de la loi de probabilité des débits moyens mensuels

INADÉQUATION DES SCHÉMAS CLASSIQUES

Le schéma dont nous venons de parler s'avère généralement suffisant pour les débits mensuels, annuels et pour les crues annuelles. Mais le problème change du tout au tout quand on aborde ce qu'il est convenu d'appeler la structure fine des débits, c'est-à-dire l'étude des débits à l'échelle journalière ou horaire. Dans ce cas là, si on éprouve tant de difficultés pour ajuster des schémas d'urnes aux débits journaliers, c'est parce que ces schémas ne conviennent pas.

Ce qu'on pourrait songer à tirer d'une urne, ce n'est pas un point mais tout une courbe représentant l'évolution du débit au cours du temps. Il faudrait évidemment préciser la durée sur laquelle porterait la courbe aléatoire et les moyens de l'obtenir.

Pourquoi cette schématisation si naturelle n'a pas été exploitée plus tôt ? Parce qu'elle se heurte à des difficultés mathématiques considérables.

Nous allons essayer le plus simplement possible de montrer la nature de ces difficultés.

On sait qu'on appelle variable aléatoire discrète une quantité X qui peut prendre un certain nombre de valeurs x_1 x_2 ... x_p, chaque valeur x_i ayant la probabilité p de se réaliser au cours d'une épreuve, tirage dans une urne, par exemple. Il semble alors très simple et très naturel de dire : dans le cas où la quantité X peut prendre non pas des valeurs isolées x mais des ensembles de valeurs constituant des fonctions f_i (t) on affectera à chaque fonction f_i (t) une probabilité p_i comme on l'a fait précédemment.

Si on se place d'un point de vue général ceci est malheureusement impossible car il existe un nombre de fonctions trop grand pour qu'on puisse affecter un nombre p_1 à chaque fonction possible. Nous n'entrerons pas dans des détails théoriques au sujet de cette question mais donnerons un exemple familier pour montrer de quoi il s'agit. Vouloir affecter une probabilité p à toutes les fonctions possibles serait en quelque sorte équivalent à vouloir numéroter 100 boules avec les 50 premiers nombres. Il y en aurait nécessairement 50 qui ne pourraient pas être identifiées.

Les probabilistes essayent depuis de nombreuses années de tourner ces difficultés sans avoir bien réussi jusqu'à présent à obtenir des solutions utilisables dans la pratique courante.

De leur côté, les hydrologues ont bien senti la nécessité d'utiliser des méthodes nouvelles pour résoudre les difficultés inhérentes à l'analyse de la structure fine des débits. Nous indiquerons plus loin les procédés employés dans un certain nombre de cas concrets, mais auparavant nous définirons les éléments qui nous serviront constamment par la suite, à savoir les séries chronologiques.

SÉRIES CHRONOLOGIQUES

On appelle série chronologique une série d'observations u (t_1), u (t_2)... u (t_n)... faites à des instants t_1, t_2 ... t_n. S'il existe une relation fonctionnelle connue entre u et t la série chronologique n'est pas autre chose que la suite des valeurs prises par la fonction u (t) aux instants t_1, t_2, ... t_n.

Si au contraire la quantité u (t_i) n'est pas déterminée par un processus certain connu, on est en présence d'une véritable série chronologique. Il en est ainsi par exemple du cours d'une valeur boursière, du rendement annuel des récoltes de blé, des débits observés dans les rivières.

Remarquons pour ce dernier exemple qui nous intéresse spécialement ici, que la suite des débits est une série chronologique particulière un peu différente de celles que nous avons citées comme autres exemples. En effet, dans ces dernières, la suite était uniquement constituée par les valeurs u (t). Ainsi pour les rendements d'une récolte, il existe un chiffre par an. Au contraire, pour la suite des débits, il s'agit d'un phénomène continu que nous observons à des instants t_1, t_2 ... t_n. Cette remarque a deux conséquences notables.

1) Alors que dans les autres cas envisagés, la suite était constituée par des points isolés, dans le cas des débits, il s'agit d'une courbe continue sur laquelle on connait des points.

2) Nous ne pouvons jamais connaître complètement le processus étudié même dans le passé puisque nous n'avons pas d'éléments sur les valeurs qu'il prend entre les points d'observation. La plupart du temps, nous remplaçons le pro-

cessus réel par un processus variant par sauts brusques aux points de mesure, constant entre deux points observés et égal sur cet intervalle à la valeur mesurée par exemple au point précédent.

D'autre part, les séries que nous étudions peuvent être de durée finie ou infinie : **nous n'en étudions jamais que des tronçons finis** de durée plus ou moins longue.

Nous verrons plus loin l'importance de cette observation.

CARACTÉRISTIQUES D'UNE SÉRIE CHRONOLOGIQUE

Traditionnellement, on distingue quatre éléments principaux dans une série chronologique :

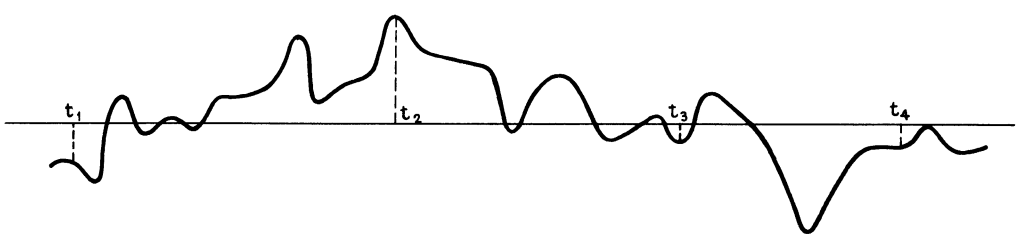
- la **tendance à long terme ou trend.**
- les mouvements réguliers qui se reproduisent d'une façon plus ou moins périodique appelés **composantes cycliques.**
- les mouvements périodiques appelés **composantes saisonnières.**

et enfin un reste appelé **résidu aléatoire**

Notre but n'est pas de présenter ici une étude détaillée des séries chronologiques en général, mais de rappeler des notions qui nous serviront par la suite.

Il convient de remarquer tout d'abord que la décomposition indiquée n'a pas une valeur absolue, mais une valeur relative.

Considérons par exemple la série ci-dessous :



Suivant l'intervalle de temps qu'on prend en considération, on trouve :

une tendance à long terme $\begin{cases} \text{fortement croissante (intervalle } t_1\ t_2) \\ \text{fortement décroissante (intervalle } t_2\ t_3) \\ \text{sensiblement constante (intervalle } t_1\ t_4) \end{cases}$

Si notre intérêt se porte sur les mouvements cycliques, nous distinguons sur l'intervalle t_1 t_4 un cycle assez ample auquel se superpose un feston de cycles courts d'amplitudes voisines.

Nous voyons bien ainsi la difficulté qui existe dans la définition objective des caractéristiques que nous étudions.

Ceci va nous mener au cœur du sujet en nous faisant pressentir les écueils auxquels on se heurte dans l'étude des processus aléatoires réels.

Précisons d'abord dans quel but sont effectuées ces recherches. Comme dans la plupart des études statistiques, il s'agit de **prévoir les états futurs à partir des états passés.**

Il est donc bien tentant de rechercher les périodes que pourrait présenter tel phénomène puisque la prévision de l'avenir s'en trouverait faite ipso facto à un résidu aléatoire près.

C'est pourquoi beaucoup d'analyses de séries chronologiques sont axées sur la recherche de périodicités. Nous allons d'abord examiner la plus caractéristique de ces méthodes.

MÉTHODES DES COMBINAISONS LINÉAIRES D'ORDONNÉES

Cette méthode créée par M. et Mme LABROUSTE se présente de la façon suivante :

Etant donné une courbe résultant de la superposition de sinusoïdes de périodes et d'amplitudes quelconques, on peut trouver les périodes, les phases et les amplitudes des sinusoïdes composantes, au moyen de combinaisons linéaires effectuées sur les ordonnées.

Sans entrer dans les détails, nous donnerons les grandes lignes de cette méthode.

Considérons une fonction qui se présente sous la forme suivante :

$$y = \sum a_i \cos(\omega_i x + \varphi_i)$$

Il est facile de vérifier que si on effectue des combinaisons linéaires du type suivant sur les y :

$$Y(x_0, m) = \sum_{p\,0}^{m} K_p\, y(x_0 + p) + y(x_0 - p)$$

les K_p étant des constantes qu'on sait déterminer, on trouve :

$$Y(x_0, m) = \sum \alpha_i \; a_i \cos(\omega_i x_0 + \varphi_i)$$

Autrement dit, les fonctions y et Y sont formées des mêmes sinusoïdes affectées d'amplitudes différentes a_i et α_i ai. M. et Mme LABROUSTE qualifient ces transformations de "non déphasantes", puisqu'elles conservent les périodes et les phases.

Leur intérêt consiste en ceci : pour des valeurs bien choisies des coefficients K les amplitudes α_i a_i sont toutes nulles sauf une qui est très grande. Par suite, la fonction transformée Y donne immédiatement la période et la phase correspondantes ω_i et φ_i. On obtient également l'amplitude a_i en divisant Y par α_i.

Il existe un mode de calcul permettant de rendre négligeables successivement les amplitudes a_1, a_2 ... a_i ... en prenant à chaque fois une combinaison convenable de coefficients K_p, de telle sorte qu'on détermine de proche en proche toutes les périodes et toutes les phases du phénomène.

Avec la méthode LABROUSTE, nous avons donc un procédé d'analyse des séries sinusoïdales très puissant quoique assez laborieux.

Cette méthode a été utilisée sur les séries chronologiques, en particulier les débits de la Seine à PARIS pour découvrir la loi de récurrence des crues catastrophiques.

Or l'hypothèse même que nous avons indiquée plus haut sur la méthode LABROUSTE en donne les limites d'emploi.

Si la courbe étudiée est formée par la superposition de sinusoïdes, les combinaisons linéaires d'ordonnées permettront d'expliciter ces sinusoïdes.

Si cette condition ne se trouve pas réalisée, nous ne pouvons tirer aucune conclusion des résultats obtenus.

La méthode LABROUSTE ne permet pas de tester si une série quelconque possède des périodes; elle permet de trouver les périodes d'une série qui en comporte. Si on croit démontrer ainsi l'existence de périodes, on prend en fait pour conclusion une hypothèse a priori, **en aucun cas confirmée par le résultat du calcul.**

On serait tenté de répondre à cette objection que si l'analyse de LABROUSTE permet de trouver des périodes c'est parce que ces périodes existent réellement.

Il n'en est rien, **car l'analyse effectuée porte sur un tronçon de la série seulement.**

Or, on montre que toute série chronologique non périodique peut, sur un intervalle borné quelconque, être approchée d'aussi près qu'on veut par une combinaison linéaire de sinusoïdes dont **les périodes et les phases sont pratiquement arbitraires.**

Par suite, l'analyse de LABROUSTE pratiquée sur un tel intervalle révèlerait l'existence de sinusoïdes. Bien mieux, suivant la manière dont s'effectuerait l'analyse, on trouverait des périodes et des phases différentes.

Si d'autre part, on effectuait l'analyse de LABROUSTE sur une autre partie de la série, on trouverait des sinusoides dont les caractéristiques, périodes, phases et amplitudes seraient encore différentes de celles qui auraient été calculées antérieurement.

Par suite, l'application brutale de la méthode LABROUSTE à une série chronologique dont **il est impossible de savoir a priori si elle est périodique ou non,** ne peut donner aucun résultat concluant en ce qui concerne la validité d'une extrapolation dans l'avenir.

Il existe un certain nombre d'autres méthodes basées sur des propriétés différentes des processus, mais qui se proposent également la recherche des périodes.

Il en est ainsi particulièrement des procédés d'analyse harmonique.

L'ANALYSE HARMONIQUE DES SÉRIES CHRONOLOGIQUES

L'analyse harmonique d'une série chronologique conduit à l'établissement d'un périodogramme qui fournit les valeurs des périodes éventuelles présentées par la série chronologique.

Nous indiquerons très rapidement le principe de cette méthode.

Considérons la série chronologique suivante :

$$u_1 \quad u_2 \quad \dots \quad u_i \quad \dots u_n$$

On calcule alors la quantité suivante :

$$S(\omega) = \sqrt{A^2 + B^2}$$

avec

$$A = \sum_i u_i \cos \omega_i \qquad B = \sum u_i \sin \omega_i$$

ceci pour différentes valeurs de ω. En portant sur un graphique ω en abscisse et

$S_i(\omega)$ en ordonnée on obtient un graphique appelé périodogramme de la série u_i. Or, on peut montrer que si la série u est de la forme :

$$u_i = \sum_j \alpha_j \sin \omega_j \, i$$

la quantité S (ω) est toujours très voisine de zéro sauf quand ω est égal à l'une des périodes ω_j, auquel cas S (ω) devient égal à α_j. Autrement dit :

$$S(\omega_j) = \alpha_j \qquad S(\omega) = \varepsilon \quad \text{si} \quad \omega \neq \omega_j \, , j = 1, 2, \ldots$$

Par suite le simple examen du périodogramme fournit immédiatement les valeurs des périodes ω_j et des amplitudes correspondantes α_j.

Mais cette méthode toute séduisante qu'elle paraisse donne lieu aux mêmes objections que la méthode LABROUSTE. Elle est en effet valable seulement si on sait a priori que la série étudiée est périodique.

Il existe à la vérité un test (test de SCHUSTER) donnant la probabilité pour que S excède une valeur donnée quand u_j est une variable aléatoire normale d'écart type σ. Mais ce test est valable **pour une valeur de S choisie au hasard.** Or, en fait, on choisit toujours les valeurs les plus grandes de S puisque ce sont elles précisément qui indiquent les périodes "valables" et d'autre part les variables u suivent souvent des lois non normales.

Par suite, les résultats qu'on obtient à partir de cette méthode sont illusoires et l'application erronée du test de SCHUSTER vient donner une confiance trompeuse dans les périodes mises en évidence.

Un auteur italien étudiant des séries de pluie sur une centaine d'années (1820-1921) trouve jusqu'à onze périodes s'échelonnant de trente cinq à deux ans. L'application incorrecte du test de SCHUSTER lui permet d'en trouver huit hautement significatives (niveau de signification 1°/°°) !

Là encore les résultats trouvés ne sont pas réellement significatifs parce qu'a priori, on ne sait pas si la série étudiée est périodique ou non.

Les méthodes que nous venons de passer en revue ont un caractère commun:

Elles se proposent de prévoir l'avenir par l'extrapolation de courbes effectivement observées dans le passé. Il y a là une sorte de définition fonctionnelle où la probabilité intervient très peu, et d'une façon d'ailleurs erronée comme nous le signalions plus haut. De telles méthodes utilisent en quelque sorte la probabilité pour que l'extrapolation soit valable! Cette probabilité est d'ailleurs souvent une question d' "impression" ou de "sentiment" chez les auteurs qui l'utilisent. Tel auteur ne dit-il pas "On **sent** que la probabilité est grande pour que les périodes observées dans le passé, se conservent dans l'avenir" !

Le mode d'analyse que nous proposons repose sur des principes tout à fait différents.

Nous cherchons à décomposer la série chronologique en éléments plus simples au moyen d'une "stylisation" analogue à certains procédés de lissage.

L'analyse nous fournit ainsi des éléments plus simples et plus maniables que la série initiale et nous facilite l'étude des lois de probabilité.

Nous chercherons ensuite la prévision de l'avenir non par une extrapolation des courbes observées dans le passé mais par l'utilisation des lois de probabilité, après avoir déterminé ces dernières sur la série dont nous disposons.

La première tâche consiste donc à remplacer la série observée par une autre dont les éléments seront calculés d'une façon mathématique. On obtient ainsi la représentation continue de la série considérée.

MÉTHODES DE REPRÉSENTATION CONTINUE DES SÉRIES CHRONOLOGIQUES

Une série chronologique peut être représentée d'une façon continue de deux manières.

1° On peut ajuster une courbe continue unique sur l'ensemble de la série observée fut-elle très longue.

2° On peut découper la série en tranches de longueurs égales, avec ou sans parties communes et effectuer un ajustement sur chaque tranche. C'est la méthode employée dans les calculs de moyennes mobiles.

D'une façon générale, la première méthode est assez peu commode car elle manque de souplesse. Le seul cas où elle peut donner satisfaction est celui où la série chronologique est terminée. Dans les autres cas, l'observation d'un seul point supplémentaire viendrait modifier la totalité de l'ajustement.

La méthode que nous utilisons appartient à la seconde catégorie. Elle n'est d'ailleurs pas nouvelle puisque M. FRECHET l'a déjà utilisée dans une étude qu'il avait effectuée sur la représentation des contours de crânes humains. Elle consiste à remplacer la suite des débits journaliers par des développements de FOURIER calculés mois par mois.

ANALYSE DE FOURIER SUR LES DÉBITS

Rappelons encore une fois ici très nettement qu'il ne s'agit pas dans cette méthode de chercher ou de trouver des périodicités mais simplement d'effectuer une représentation du processus. Pour clarifier les idées à ce sujet, nous rappellerons brièvement quelques propriétés mathématiques des séries de FOURIER.

Etant donnée une fonction f(x) continue et à variation bornée sur le segment $(o, 2\pi)$ la série suivante :

$$f^*(x) = a_0 + \sum_n a_n \cos n x + b_n \sin n x$$

où

$$a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(t)\, dt$$

$$a_n = \frac{1}{\pi}\int_0^{2\pi} f(t) \cos nt\, dt$$

$$b_n = \frac{1}{\pi}\int_0^{2\pi} f(t) \sin nt\, dt$$

est uniformément convergente en x et elle définit une fonction f (x) périodique de période (2π) égale sur le segment $(0, 2\pi)$ à la fonction non périodique f (x).

On peut d'ailleurs effectuer un développement analogue sur n'importe quel autre intervalle (a, b) en effectuant sur la variable x la transformation suivante:

$$x \longrightarrow 2\pi \frac{x - a}{b - a}$$

Nous allons donner deux exemples très simples: considérons sur l'intervalle $(-\pi, +\pi)$ la fonction y = x; la fonction suivante :

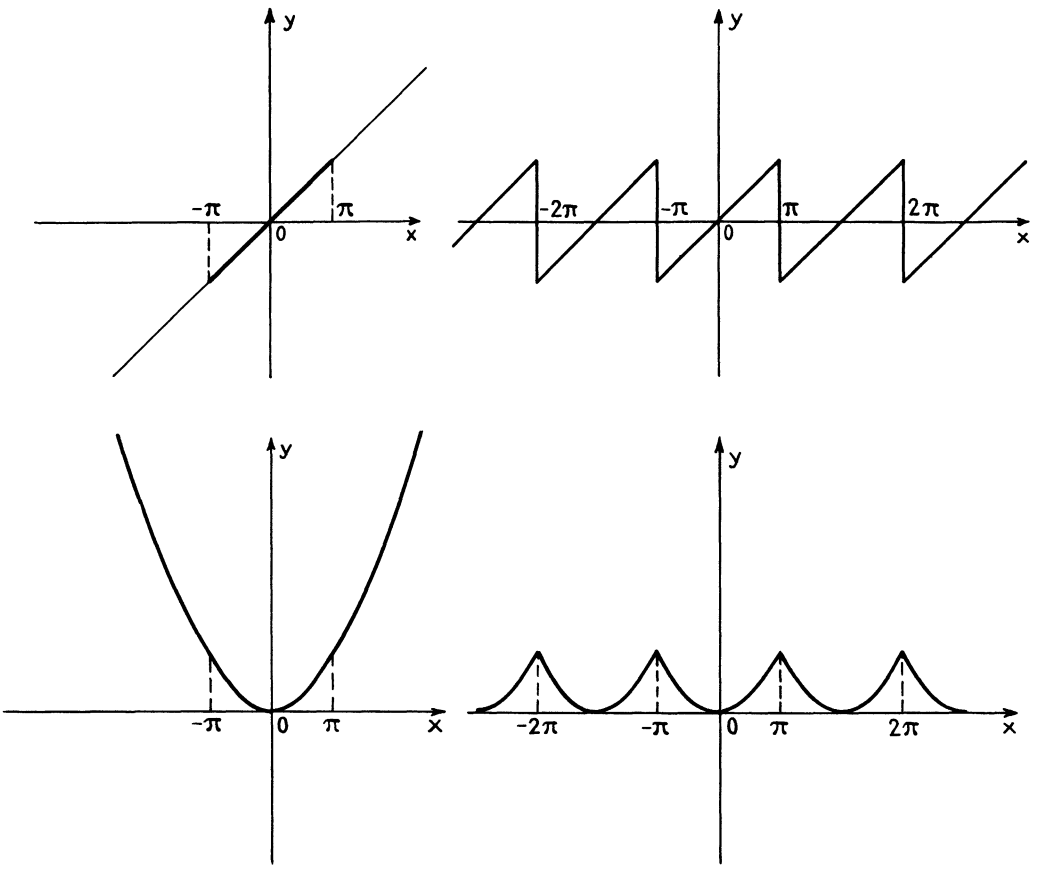
$$f^*(x) = 2 \sum_n (-I)^{n+I} \frac{\sin n x}{n}$$

est périodique et égale à x sur l'intervalle considéré.

Si nous prenons sur ce même intervalle $(-\pi , +\pi)$ la fonction $y = x^2$ la fonction suivante :

$$f^*(x) = \frac{\pi^2}{3} - 4 \sum^{\infty} (-I) \frac{\cos n x}{n^2}$$

est périodique et égale à x^2 sur l'intervalle $(-\pi , +\pi)$



Nous voyons donc qu'il est parfaitement possible de développer des fonctions non périodiques en séries périodiques. La seule précaution à prendre consiste à n'utiliser le développement trouvé que dans l'intervalle ayant servi à le définir.

Dans la suite nous appelons X_i le débit moyen logarithmique (1) journalier du jour i. Comme on l'a dit plus haut, on peut considérer le débit instantané logarithmique X (t) comme une fonction continue dont on connait les valeurs moyennes journalières. On aurait alors :

$$X_i = \int_{t_{i-I}}^{t_i} X(t)\, dt = X(\tau_i)$$

τ_i étant un instant compris entre t_i -I et t_i

Nous ne soulèverons pas de difficultés sur ce point qui relève uniquement de l'analyse mathématique.

(1) Cette appellation due à E. HALPHEN désigne le logarithme décimal du débit moyen journalier.

Disons simplement que nous construisons le développement de FOURIER de la fonction X (t) mois par mois à partir des valeurs X . Nous remplaçons les coefficients théoriques :

$$a_n = \frac{I}{\pi} \int X(t) \cos n t \, dt$$

par les valeurs suivantes :

$$a'_n = \frac{I}{\pi} \sum X \cos n t_i$$

On pourrait démontrer mathématiquement que cette substitution donne encore une série convergent vers X (t) moyennant certaines conditions.

Un problème important va se poser maintenant. Il ne peut être question évidemment de calculer tous les termes de la série. Par conséquent, on va être obligé de choisir le nombre de termes à retenir.

NOMBRE D'HARMONIQUES A RETENIR

On ne peut parler d'un choix objectif du nombre d'harmoniques nécessaires. Il faut bien voir le but que nous visons : représentation et description optimales du processus suivi par les débits.

Il s'agit là d'un optimum statistique qui ne se confond pas du tout avec un optimum analytique. En effet, si on voulait représenter "au mieux" analytiquement la fonction X (t), il faudrait calculer le nombre maximum possible d'harmoniques. On obtiendrait ainsi un ajustement très souple qui représenterait très bien chaque courbe mensuelle, mais d'un mois à l'autre, on trouverait des courbes de FOURIER très différentes entre elles.

Précisons donc ici le but poursuivi : Il s'agit d'isoler dans le processus une tendance systématique - quoique aléatoire - et une composante purement due au hasard dont nous étudierons séparément les lois de probabilité.

Or, si le nombre d'harmoniques est très grand, la composante purement aléatoire sera petite, tout l'aléa résiduel étant absorbé par le développement trigonométrique qui deviendra lui-même très variable d'un mois à l'autre.

En revanche, si on prend un nombre d'harmoniques très petit, la tendance sera stable et d'analyse plus simple mais le terme résiduel prendra une importance trop considérable.

Notre optimum se situera entre ces deux pôles extrêmes.

On pourrait le symboliser de la façon suivante. Considérons la quantité :

$$Z_i = X_i - a_0 - \sum (a_j \cos j t_i + b_j \sin j t_i)$$

La valeur n "la meilleure" de notre point de vue sera celle qui correspondra au minimum du coefficient de corrélation Z_i et Z_{i-1} .

D'une façon plus exigeante, on pourrait définir n par la condition que le corrélogramme de la série des Z_i se rapproche le plus possible du corrélogramme d'une série purement aléatoire. L'imprécision de ces définitions nous a conduits à les rejeter.

On voit que la notion d'optimalité est assez vague et qu'il n'est pas très facile d'en tirer parti concrètement en imposant la condition supplémentaire que n soit le même quels que soient les mois considérés.

En fait, on ne peut se décider à partir de critères aussi imprécis et on fait largement appel à des considérations empiriques.

Dans le cas concret qui nous intéressait nous avons choisi pour n la valeur 3 de sorte que notre formulation se présente de la manière suivante :

$$X_i^* = a_0 + a_1 \cos \omega t_i + b_1 \sin \omega t_i + a_2 \cos 2\omega t_i + b_2 \sin 2\omega t_i + a_3 \cos 3\omega t_i + b_3 \sin 3\omega t_i$$

ce qui peut s'écrire :

$$X_i^* = a_0 + A_i \cos (\omega t_i - \varphi 1) + A_2 \cos (2\omega t_i - \varphi 2) + A_3 \cos (3\omega t_i - \varphi 3)$$

Nous écrirons finalement le débit sous la forme suivante :

$$X_i = X_i^* + Z_{3i}$$

Z_{3i} étant un résidu aléatoire.

Ceci équivaut à considérer le débit X (t) comme la superposition d'un processus simple X^* où seuls les paramètres sont aléatoires et d'une variable aléatoire Z du type classique.

Avant de passer directement aux applications, nous signalerons une petite difficulté de réalisation pratique.

Comme nous le disions avec force plus haut, le processus étudié n'est pas périodique. Or le développement de FOURIER que nous utilisons est, lui, périodique, de par sa construction même. La conséquence pratique de cette différence de structure est la suivante. Aux deux bouts de chaque période étudiée, le développement de FOURIER s'écarte systématiquement de la courbe naturelle qu'il doit représenter. Nous avons pallié à cet inconvénient en effectuant la représentation, non pas sur la durée réelle de chaque mois, mais sur une période de 36 jours encadrant le mois calendaire et nous n'étudions la correspondance entre courbe ajustée et courbe naturelle que sur le mois réel en négligeant les deux extrémités "aberrantes" du système. L'avantage de cette méthode est double.

Nous obtenons une représentation qui est correcte et satisfaisante sur l'intervalle où nous la conservons et d'autre part les calculs pratiques se trouvent très simplifiés par le fait que 36 est le dixième de 360.... Il nous suffit de considérer dans le calcul des coefficients a_i et b_i des angles qui varient de multiples de 10°.

APPLICATIONS DE LA MÉTHODE

Nous avons appliqué cette méthode à la station de Pont-de-Carouge sur l'Arve.

On trouvera en annexe, un certain nombre de graphiques qui donnent, pour différents mois, les courbes de débits logarithmiques et les courbes des écarts entre ces courbes naturelles et les séries de FOURIER correspondantes arrêtées successivement au premier, au second... au sixième harmonique. (Figures 2 à 5).

On constatera ainsi que notre choix sur le nombre d'harmoniques à considérer, correspond à un compromis entre le désir d'avoir une autocorrélation minima et d'autre part, celui de conserver une série de FOURIER maniable.

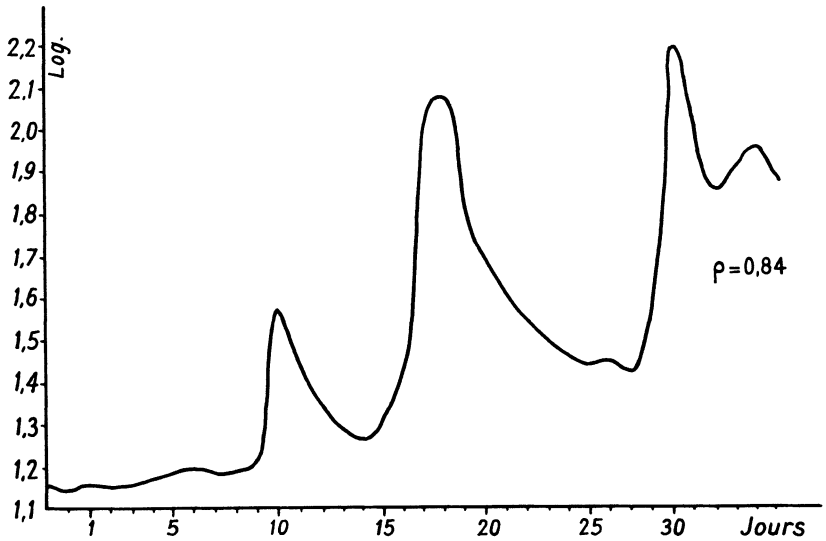
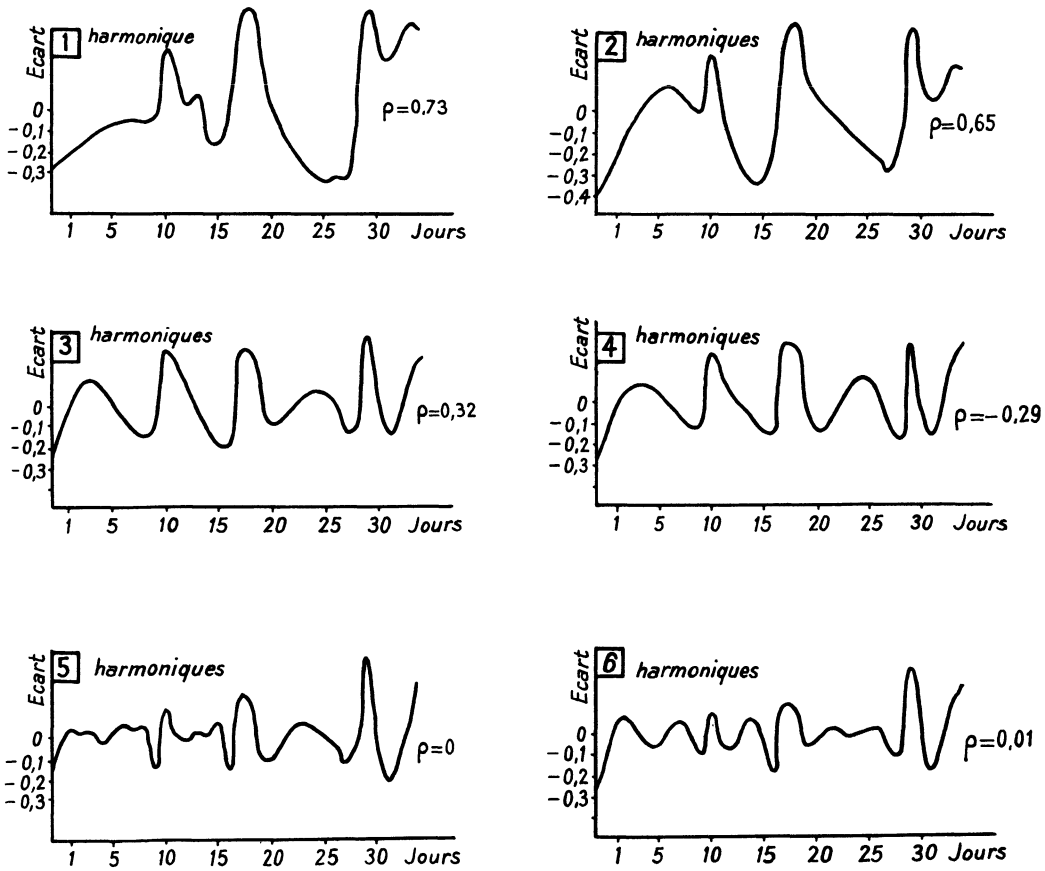


Fig. 2

Arve au Pont de Carouge Mars 1932

Courbe des débits logarithmiques

ρ : coefficient d'autocorrélation



Arve au Pont de Carouge Mars 1932

Ecart entre la courbe des débits logarithmiques et les séries de Fourier correspondantes

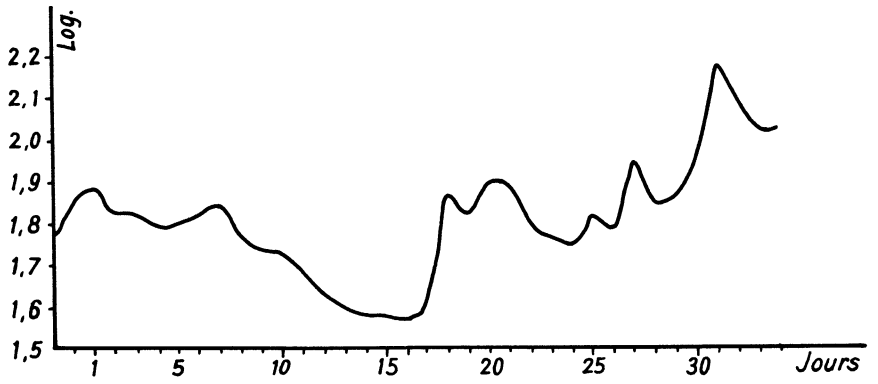
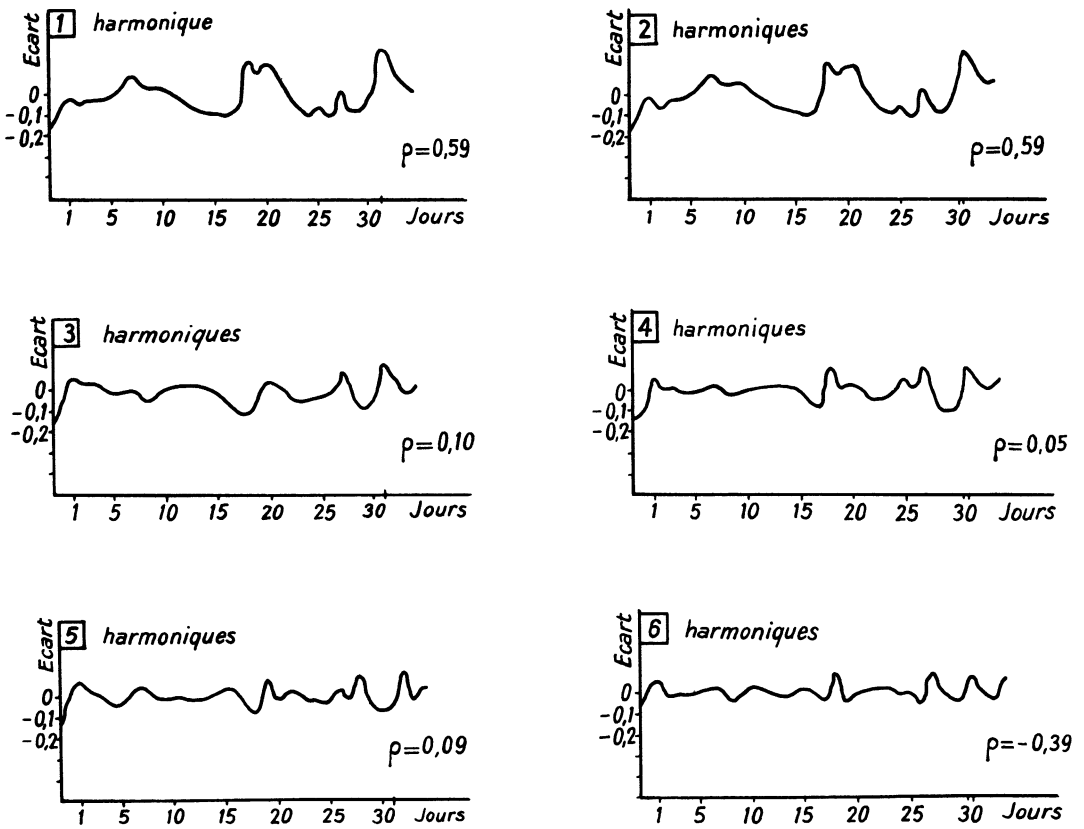


Fig. 3

Arve au Pont de Carouge Mai 1949

Courbe des débits logarithmiques



Arve au Pont de Carouge Mai 1949

Ecart entre la courbe des débits logarithmiques et les séries de Fourier correspondantes

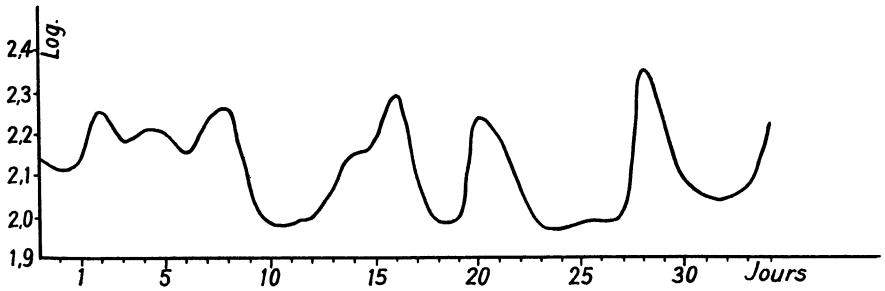
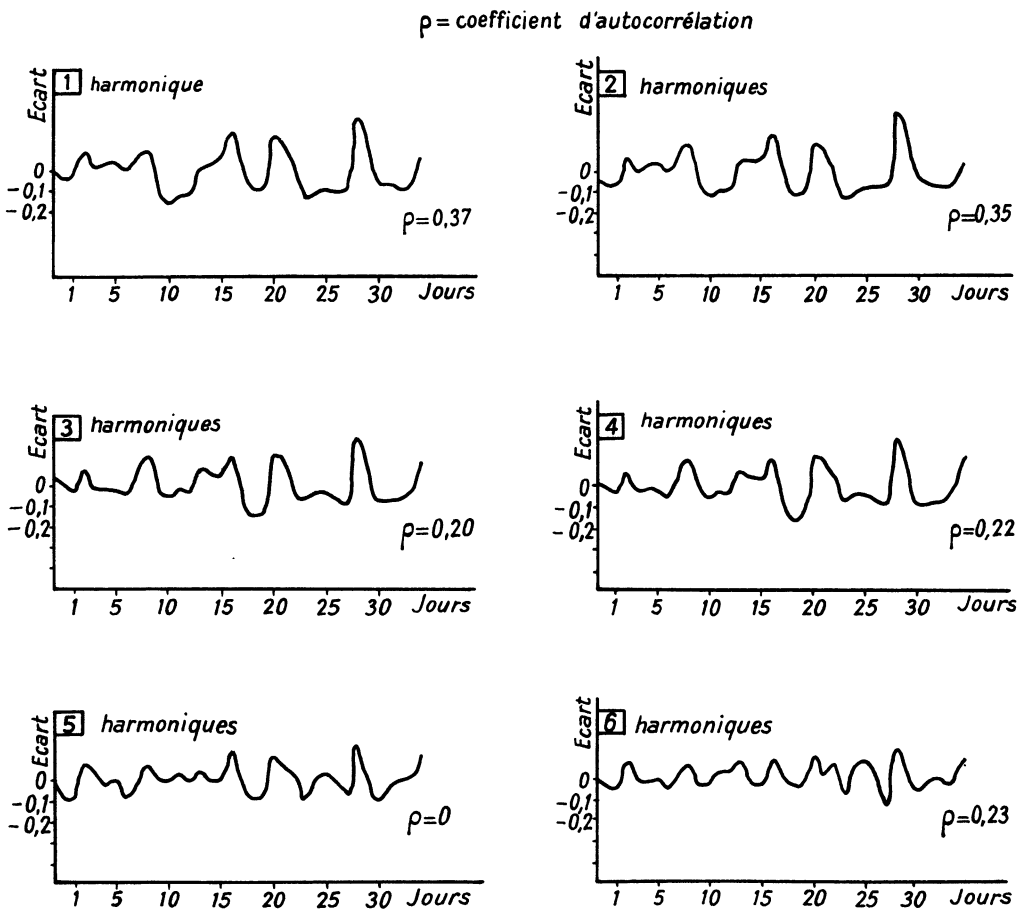


Fig. 4

Arve au Pont de Carouge Juillet 1931

Courbe des débits logarithmiques



Arve au Pont de Carouge Juillet 1931

Ecart entre la courbe des débits logarithmiques et les séries de Fourier correspondantes

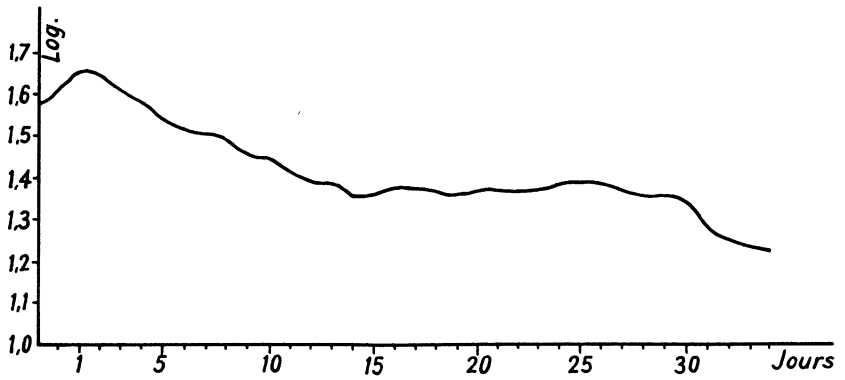
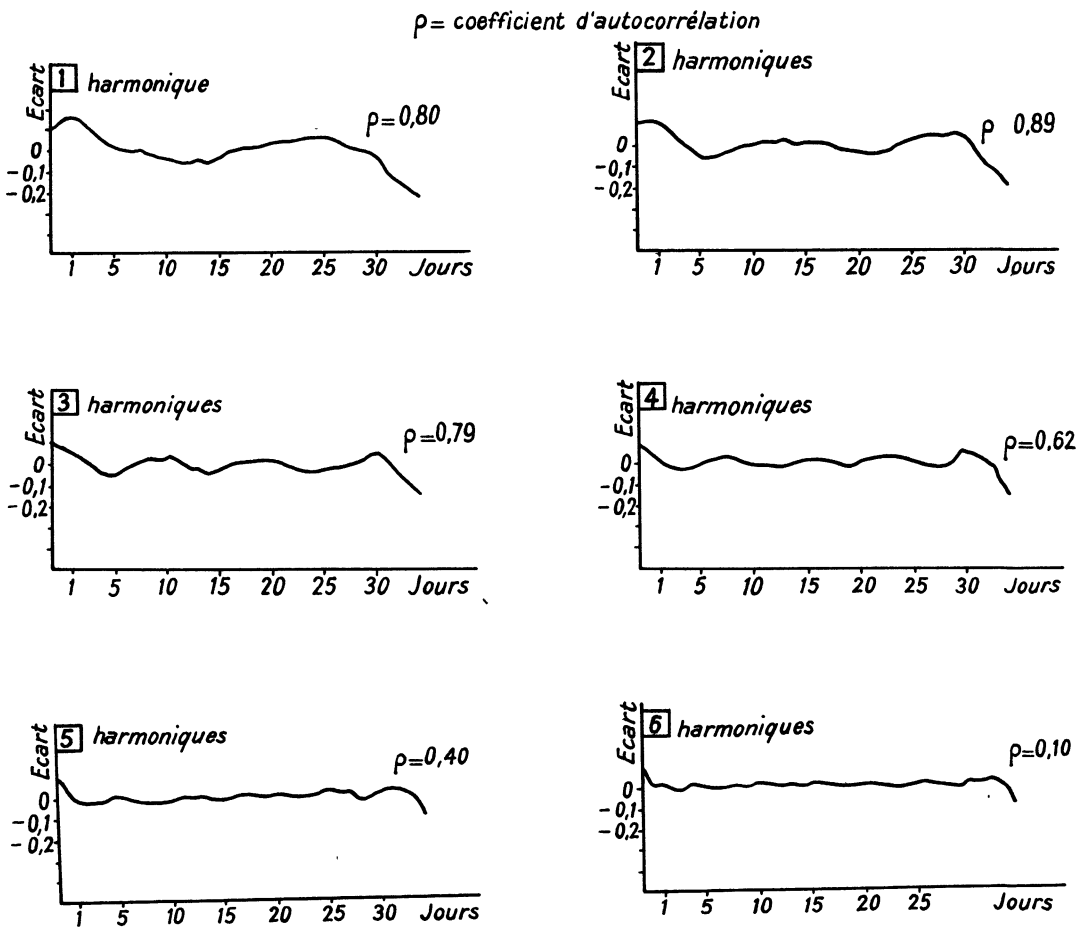


Fig. 5

Arve au Pont de Carouge Décembre 1933

Courbe des débits logarithmiques



Arve au Pont de Carouge Décembre 1933

Ecart entre la courbe des débits logarithmiques et les séries de Fourier correspondantes

Le tableau ci-dessous résume les éléments de notre choix à ce sujet. On y trouvera pour différents mois, les autocorrélations de la série des débits logarithmiques et des différents écarts Z_i

Mois	X	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
Novembre 1930	.81	.81	.16	.14	.04	.03	-.08
Juillet 1931	.47	.37	.35	.20	.22	0	-.23
Mars 1932	.84	.73	.65	.32	-.29	.0	.01
Décembre 1933	.98	.80	.89	.79	.62	.40	.10
Septembre 1939	.92	.75	.51	.19	.13	-.11	.28
Mars 1949	.91	.84	.56	.46	.36	.12	.06
Juin 1929	.80	.28	0	.03	-.01	0	-.02

Nous voyons que dans tous les cas, l'écart Z_3 a une autocorrélation beaucoup moins forte que le débit logarithmique lui-même.

Nous ferons cependant à cet égard deux remarques :

Tout d'abord, les autocorrélations que nous trouvons pour Z_3 sont faibles mais elles ne sont pas nulles, ni même négligeables.

D'autre part, il n'est pas possible de tirer une conclusion définitive de l'examen de sept mois seulement.

Ces restrictions mises à part, nous étudierons séparément la partie systématique et la partie résiduelle de notre décomposition.

PARTIE FONCTIONNELLE DE LA DÉCOMPOSITION

Nous appelons partie fonctionnelle de la décomposition, la série :

$$X^* (t) = a_0 + \sum_{i=1}^{i=3} A_i \cos (i \omega t - \varphi_i)$$

Nous avons pu vérifier sur un certain nombre de mois (janvier, mars, juin, novembre) que les éléments : a_0, A_1, A_2 et A_3 étaient peu corrélés entre eux et qu'ils suivent convenablement une loi harmonique. Les paramètres tg φ_1, tg φ_2 et tg φ_3 sont indépendants et suivent des lois normales. (Voir les figures 7 à 10).

Par suite, la partie systématique X^* est complètement définie, puisque sa forme analytique est donnée et que la loi de probabilité des paramètres qui y figurent est déterminée.

Il nous reste à considérer la partie résiduelle.

PARTIE RÉSIDUELLE DU DÉVELOPPEMENT

Le tableau que nous avons donné plus haut montre que l'autocorrélation de la série Z_3 n'est pas négligeable généralement. Nous en déduisons que le processus Z_3 n'est pas à valeurs indépendantes.

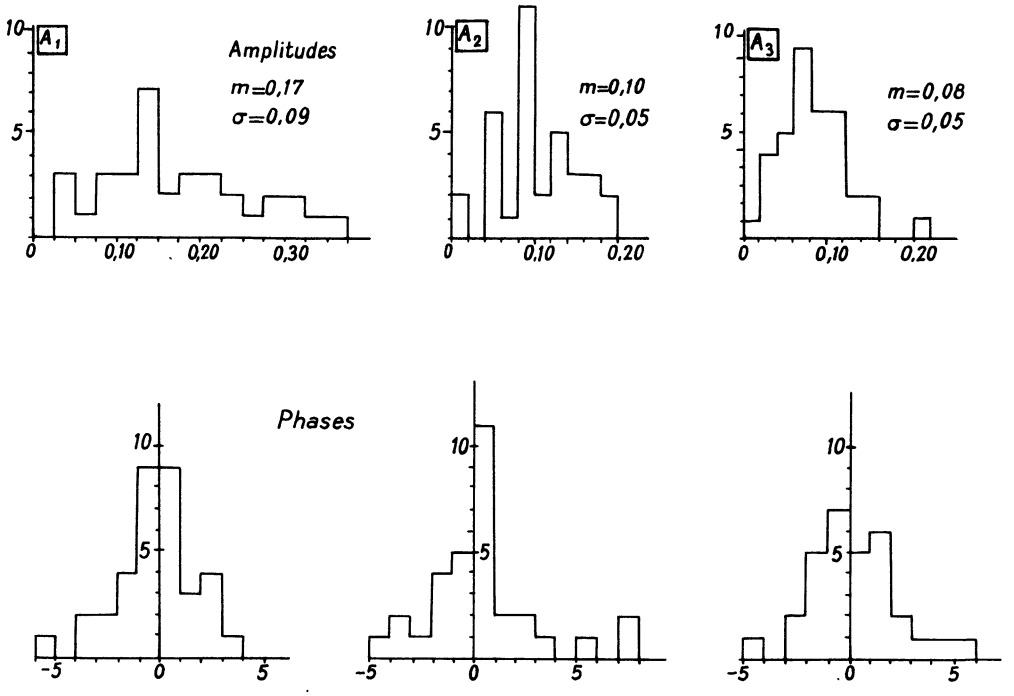


Fig. 6

Arve au Pont de Carouge Janvier

Polygone de fréquence des amplitudes et des phases des développements de Fourier

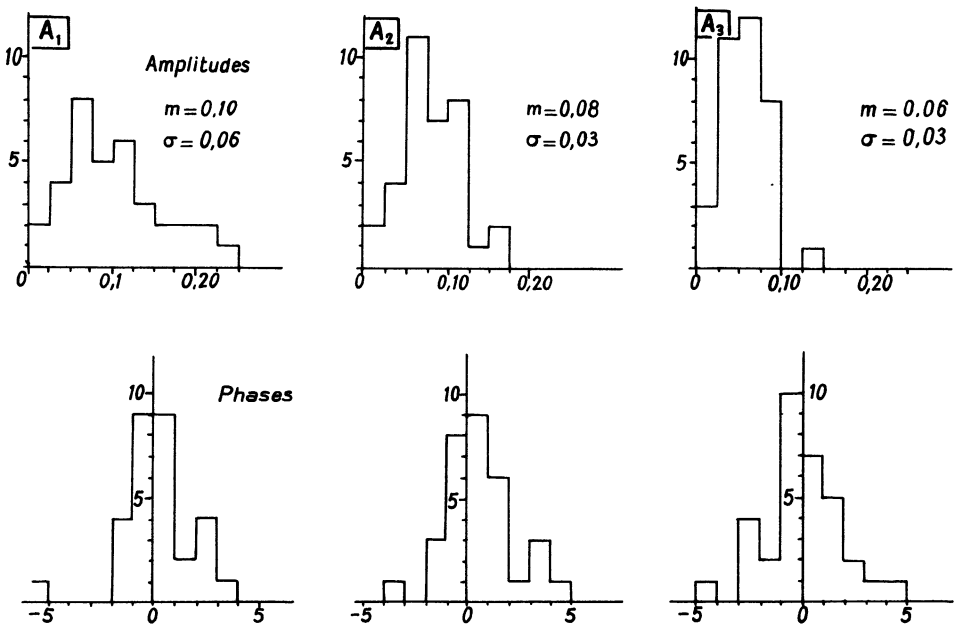


Fig. 7

Arve au Pont de Carouge Avril

Polygone de fréquence des amplitudes et des phases des développements de Fourier

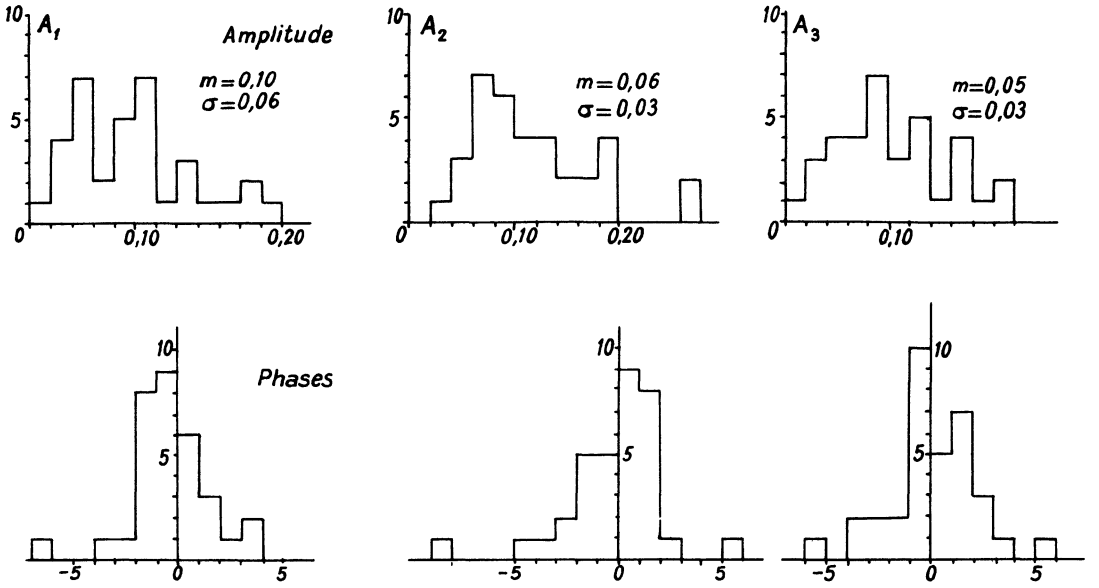


Fig. 8

Arve au Pont de Carouge Juin

Polygone de fréquence des amplitudes et des phases des développements de Fourier

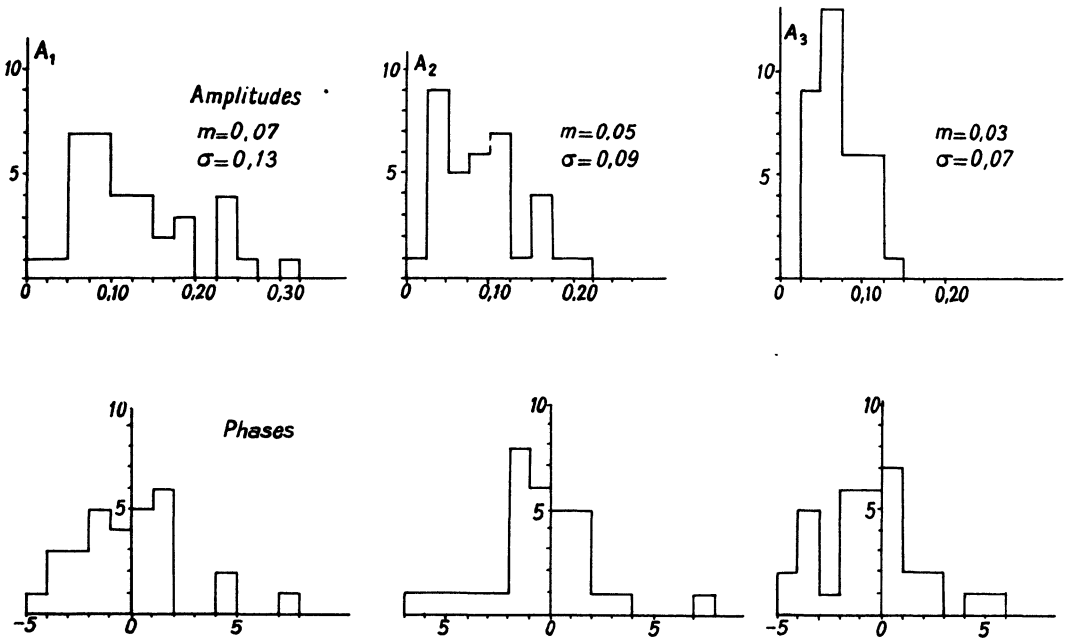


Fig. 9

Arve au Pont de Carouge Septembre

Polygone de fréquence des amplitudes et des phases des développements de Fourier

Cependant, on peut admettre qu'il obéit à un schéma simple de chaîne de MARKOV d'ordre 1 ou peut-être 2 au plus.

Une chaîne double serait définie par la dispersion σ et par les coefficients de corrélation ρ_1 et ρ_2 des couples $\left[Z_3(t),\ Z_3(t-1) \right]$ et $\left[Z_3(t),\ Z_3(t-2) \right]$ tandis qu'une chaîne simple serait définie simplement par la dispersion et par le coefficient de corrélation du couple $\left[Z_3(t),\ Z_3(t-1) \right]$.

On remarquera à ce sujet, qu'on ne devra pas estimer l'écart-type de Z_3 à partir de la formule classique :

$$\sigma^2 = \frac{\sum_{i=1}^{n} (Z_i - Z)^2}{n}$$

En raison de la liaison qui existe entre les Z_i, on devra utiliser la formule suivante si on ajuste un schéma de chaîne simple :

$$\sigma^2 = \frac{1}{n} \left[(Z_1 - \overline{Z})^2 + \frac{1}{1 - \rho^2} \sum_{i=2}^{n} \left[(Z_i - \overline{Z}) - \rho (Z_{i-1} - \overline{Z}) \right]^2 \right]$$

La formule serait plus complexe en cas de chaîne double. Nous ne l'écrirons pas.

INTERPRÉTATION ET UTILISATION DU SCHÉMA

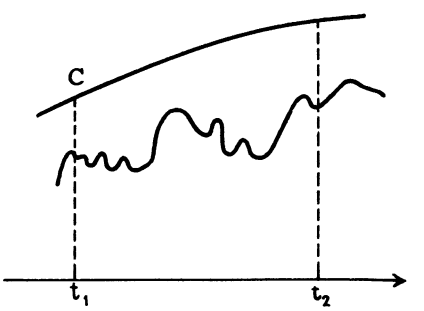
Arrivés à ce stade, nous pouvons commencer à donner une analyse très imparfaite encore de la structure fine des débits, mais nous n'avons encore abordé qu'une moitié du problème. Il nous faut maintenant tirer des informations pratiques de cette analyse.

Cette seconde partie n'est pas encore faite, mais nous l'étudions actuellement; c'est pourquoi les explications qui suivent sont nécessairement brèves et qualitatives.

Le problème que nous avions posé au début de cet exposé, est la prévision de l'avenir au moyen du passé. Plus concrètement, nous cherchons la loi de probabilité des débits journaliers au cours d'une période donnée.

La notion de processus stochastique ne nous permet pas de trouver directement cette loi de probabilité, mais elle nous permet de trouver la probabilité d'absorption par rapport à une courbe C imposée de la courbe des débits d'une période donnée.

Définissons la probabilité d'absorption; c'est la probabilité pour que la courbe des débits reste en dessous de la courbe C dans la période de temps (t_1, t_2). On voit que si on prend comme courbe C, une parallèle à l'axe des temps, la probabilité d'absorption correspondante n'est pas autre chose que la probabilité pour que dans période (t_1, t_2) les débits soient tous inférieurs à une constante c.



On conçoit qu'on puisse avoir à choisir différents types de courbes C suivant les études poursuivies. En particulier, un cas intéressant sera celui du remplissage des réservoirs où on pourra prendre la courbe d'alerte comme courbe C.

Il est vraisemblable que pour ce genre d'études on soit obligé d'utiliser des méthodes très approchées en raison de la complexité mathématique à laquelle on se heurterait à vouloir les traiter directement.

Dans cet exposé, nous avons essayé de mettre en évidence les insuffisances des méthodes classiques pour l'étude de la structure fine des débits. Quand on descend à l'échelle des débits journaliers ou horaires, en effet, des phénomènes invisibles ou peu perceptibles sur des débits mensuels deviennent prépondérants et rendent impossible l'usage de procédés simples.

Malheureusement, les méthodes plus perfectionnées n'ont pas toujours été cherchées dans la bonne voie et nous avons pu donner quelques exemples de conclusions tout à fait injustifiées.

Après Etienne HALPHEN, nous avons repris l'analyse des séries de débits journaliers par la méthode de représentation de FOURIER et nous avons cherché à en tirer quelques résultats concrets. Il importe ici de bien voir qu'il s'agit seulement de représentation et non pas d'explications. Nous n'avons pas éclairci le mystère de la nature des débits si tant est qu'on puisse y parvenir un jour.

Simplement, nous avons un procédé plus commode que certains autres et qui ne fait pas appel aux notions de périodicités.

L'application du procédé à des rivières différentes tant par le régime que par le bassin permettra peut-être de voir si certaines propriétés observées sur l'Arve sont encore valables ailleurs. Alors peut-être, nous pourrions commencer à tirer quelques conclusions quant à la valeur générale du procédé.

DISCUSSION (Président M. BATICLE)

M. le Président remercie M. LARRIEU de son exposé très clair et même pittoresque, ce qui est une grande qualité quand on a à travailler sur des questions aussi arides et croit qu'il a ouvert fenêtre sur un champ de recherches extrêmement intéressant, qui consiste, au fond, à essayer de réduire les caprices de la nature, ou au moins de les réduire à une petite variable aléatoire que l'on pourra mieux étudier que le phénomène en entier lui-même.

M. KIVELIOVITCH rappelle que s'étant aussi occupé des séries chronologiques, il a employé la méthode inverse de celle de M. LARRIEU partant de l'idée d'une courbe au hasard, il en a étudié les caractéristiques : les maxima, les minima, etc... et l'a comparée à une courbe expérimentale, il a cherché quelles étaient les conditions nécessaires pour que la courbe soit organisée ou non.

M. KIVELIOVITCH croit que quand on part, comme fait M. LARRIEU, d'une courbe organisée conduisant à une courbe désorganisée, on peut employer plusieurs développements, deux par exemple et ramener de nouveau à une courbe au hasard ces deux courbes qui peuvent être deux fonctions de distribution différentes. Il faut démontrer à l'avance qu'elles tendent toujours vers la même courbe au hasard pour caractériser le phénomène ; sinon, l'étude et la prévision, si celle-ci est possible, sont douteuses.

M. MORLAT remarque que la méthode choisie par HALPHEN, et que vient d'exposer M. LARRIEU a peut-être pour justification certaines structures physiques. C'est un peu le problème dont M. GIRAULT parlait tout à l'heure. On n'a pas pu en faire état ici, car ce sont des recherches encore informes.

M. LARRIEU observe qu'il ne s'agit pas par ce procédé-là de trouver une courbe au hasard, mais une courbe fonctionnelle et un résidu qui par définition, est la différence entre la courbe réelle et la courbe fonctionnelle.

M. KIVELIOVITCH estime que le choix entre trois périodes ou six périodes découle de la considération d'une courbe au hasard.

M. LARRIEU répond qu'il n'en est pas ainsi du fait de l'existence d'auto-corrélations.

M. KIVELIOVITCH rappelle que M. LARRIEU trouve que le nombre Z_3 est excessivement petit et qu'il ne peut le réduire à 0. Il y a donc là une petite organisation à voir, certes difficile à déceler.

M. LARRIEU est d'accord sur ce point mais remarque qu'il ne présente pas, dans le cas présent, un très gros intéret, et ceci pour deux raisons : d'abord parce que ce nombre est extrêmement petit ; par conséquent, le fait qu'il soit tout à fait organisé ou tout à fait désorganisé, n'introduit qu'une petite différence dans la plupart des cas ; ensuite parce que cette question est encore à l'étude, et qu'elle sera plus facile à discuter quand elle aura fait l'objet de nombreuses vérifications expérimentales.

Au point de vue de la forme topologique de la courbe au hasard, des maxima, des minima, etc... M. KIVELIOVITCH attire l'attention sur ce que l'on trouve que tout ce qui carcatérise qualitativement la courbe ne dépend pas de la fonction de distribution, mais de la courbe au hasard, et seulement des parties quantitatives quand on fait des coupures, des pentes, etc...

M. KIVELIOVITCH pense que, peut-être, en combinant les deux méthodes la sienne et celle de M. LARRIEU, on pourrait trouver une loi qui pourrait servir.

M. LARRIEU insiste sur le fait qu'il s'agit là d'une méthode encore dans l'enfance ; comme il le disait au début, Etienne HALPHEN, qui proposait d'étudier le processus des débits en deux périodes, avait prévu un temps relativement court pour la première partie: or, elle a duré 8 ans. La deuxième partie a été commencée en 1952 par HALPHEN, elle a été interrompue par sa mort. M. LARRIEU s'en occupe à l'heure actuelle : sans être pessimiste, il parait assez raisonnable de prévoir quelques années encore pour cette deuxième partie.

M. RISSER signale qu'une étude générale des séries chronologiques suggère qu'elles peuvent être regardées comme composées de trois éléments ou mouvements caractéristiques :

a) un mouvement de tendance générale ou de longue durée, auquel les statisticiens anglais donnent le nom de trend,

b) un mouvement saisonnier,

c) un mouvement cyclique.

Aux variations des chiffres bruts, résultantes de ces mouvements élémentaires, s'ajoute un résidu, issu des mouvements sporadiques et accidentels.

On est conduit à adopter pour la représentation du trend un polynome dont on évalue les coefficients par la méthode des moindres carrés, et l'on a recours - pour éviter l'inconvénient au passage d'un polynome de degré 1/0 à un polynome de degré (p + 1) à la méthode des moyennes mobiles.

Or, un tel procédé de lissage introduit une courbe sinusoïdale, comme l'ont observé tout d'abord YULE puis KITCHINE et SLUTZKY, et fausse la représentation du phénomène.

Dans ce cas, on est amené à préconiser l'emploi de l'équation aux différences x (t) - n (t - h) = o pour la figuration du phénomène, dont la solution est :

$$x_t = m + \sum_K \left(A_K \cos \frac{2\pi Kt}{h} + B_K \sin \frac{2\pi Kt}{h} \right)$$

où K varie de 1 à $\left(\frac{h-1}{2}\right)$ ou $\frac{h}{2}$, et en pratique à ne conserver qu'un développement limité,

$$x_t = m + \sum_{K=4}^{s} C_K \cos \left(\lambda_K t + \varphi_K \right)$$

M. LARRIEU répond que les phases se trouvent introduites dès le début de l'exposé mais sous la forme implicite classique :

$$a_n \cos_n \omega t + b_n \sin_n \omega t$$

M. RISSER rappelle que l'on s'est rendu compte que la méthode classique de recherche des périodicités dans le temps, dite analyse d'un périodigramme d'une série ne fournit guère d'indications précises dans beaucoup de cas.

Il signale que les études faites en 1947 par KANDALL au sujet de la distribution de $S^2 = A^2 + B^2$, et du nombre le plus rationnel d'harmoniques à utiliser.

M. LARRIEU répond que la méthode rappelée par M. RISSER n'est pas celle qu'il utilise, et qu'il a précisément essayé d'en montrer l'insuffisance.

En ce qui concerne le terme " Z ", M. LARRIEU explique qu'ayant des difficultés probabilistes assez sérieuses sur les débits, il a déplacé ces difficultés sur ce terme beaucoup plus maniable analytiquement.

M. RISSER estime que M. LARRIEU a préconisé pour l'étude de l'équation

$$u_{t+1} - a^u t = \varepsilon j_1$$

celle utilisée par YULE pour l'équation aux différences du second ordre

$$u_t + 2 - a^u_{t+1} - b_{u_t} = \varepsilon j_1 \qquad (\varepsilon j \text{ étant un élément aléatoire})$$

M. LARRIEU indique qu'en réalité il a effectué une représentation de FOURIER. D'ailleurs les séries chronologiques étudiées sont de bonnes séries, ne présentant pas trop de difficultés au point de vue mathématique.

M. LARRIEU précise qu'il ne cherche pas de périodes identiques ou différentes, car d'après, leur nature physique, les phénomènes de débits ne sont pas périodiques ; ce qu'il cherche, c'est une représentation de ces phénomènes : on aurait pu représenter les débits par des polynomes ou n'importe quel système de fonction orthogonales suffisamment complet.

M. RISSER signale également l'intérêt qu'il y aurait à examiner les intensités définies par :

$$S^2 \; A^2 + B^2 \left[\frac{2}{n} \sum u_j e \; \frac{2\pi i j}{u^2} \right]^2, \text{ avec } u_j = a \sin \frac{2\pi j}{\lambda} + b_j ,$$

l'élément bj ne renfermant aucun élément cyclique.

M. LARRIEU rappelle que d'après les résultats qu'il a donnés tout à l'heure, à savoir que toutes les séries chronologiques étudiées dans un intervalle quelconque, peuvent être approchées d'aussi près qu'on veut par une suite de sinusoïdes, la méthode des moyennes mobiles donnera toujours des périodes.

Il estime donc qu'on ne devrait jamais prendre en compte une période calculée sur un intervalle petit, parce que cette période peut être tout à fait accidentelle.

M. le Président conclut en rappelant que l'ajustement d'une chaîne de Markov au graphique des données d'observation n'a donné aucun résultat. "C'est pourquoi, dit M. BATICLE, je me permets de répéter qu'on a un peu essayé de discipliner les caprices de la nature en tenant compte de l'influence héréditaire du débit antérieur sur le débit postérieur."