

Analyse des composants du bilan hydrologique par régression multiple

Autor(en): **Mathey, Bernard / Morel, Gérard**

Objektyp: **Article**

Zeitschrift: **Bulletin de la Société Neuchâteloise des Sciences Naturelles**

Band (Jahr): **97 (1974)**

PDF erstellt am: **25.09.2024**

Persistenter Link: <https://doi.org/10.5169/seals-89063>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

ANALYSE DES COMPOSANTS DU BILAN HYDROLOGIQUE PAR RÉGRESSION MULTIPLE

par

BERNARD MATHEY¹ et GÉRARD MOREL

AVEC 5 FIGURES

1. INTRODUCTION

L'analyse des résultats expérimentaux par régression multiple est de plus en plus fréquente en hydrologie. Deux exemples d'étude des paramètres du bilan par régression multiple sont envisagés ici, qui donnent un aperçu des possibilités d'emploi de cette méthode en hydrogéologie.

On voudra bien excuser les auteurs du rappel de quelques lois fondamentales de statistique, ainsi que de comparer les avantages réciproques des méthodes déterministes et statistiques, en insistant sur les limites de ces dernières.

2. MODÈLES STATISTIQUES ET DÉTERMINISTES

L'idée de recourir à des lois mathématiques pour décrire le comportement des phénomènes physiques est à la base de toute recherche scientifique.

Pour certaines sciences, en hydrologie en particulier, les causes des phénomènes sont souvent multiples et si complexes que rares sont ceux que l'on décrit actuellement par des lois parfaitement déterministes, sans se placer dans des conditions expérimentales extrêmement particulières et simplifiées.

Si les méthodes statistiques ne partent pas de l'analyse rigoureuse d'un phénomène, analyse à laquelle elles peuvent d'ailleurs conduire, elles permettent cependant de maîtriser un phénomène à l'intérieur de limites bien définies.

¹ Centre d'hydrogéologie, 11, rue Emile-Argand, CH - 2000 Neuchâtel 7.

Par exemple, on sait par l'observation que le débit d'une source est lié au gradient hydraulique de la nappe dont elle est l'exutoire. Le niveau de la nappe est lié aux précipitations antécédentes, mais d'autres facteurs, tels que la température de l'air, dont dépendent l'activité végétale et l'évapotranspiration, la perméabilité du terrain et l'humidité du sol avant les précipitations, sont extrêmement importants et se répercuteront sur le débit de la source.

Si les lois physiques définissant les écoulements en zone saturée sont bien développées malgré quelques hypothèses simplificatrices, celles qui décrivent les phénomènes de l'infiltration et de l'écoulement dans la zone non saturée font actuellement l'objet d'études nombreuses et doivent donc être utilisées avec certaines réserves.

Une étude de l'ensemble des paramètres agissant sur le débit de la source par régression multiple, permettra dans ce cas de définir statistiquement l'importance relative de chacun d'eux sur le niveau de la nappe ou le débit de la source.

Simultanément les résultats de la régression multiple fournissent un modèle statistique qui sera utilisable dans la prévision des niveaux et des débits, à l'exception des cas où il y a modification des prélèvements et des apports.

3. LA RÉGRESSION MULTIPLE

3.1. Définitions

Considérons qu'il existe une relation fonctionnelle entre les variables X_1, X_2, \dots, X_N et que l'on exprime cette relation par l'équation suivante :

$$X_1 = a_1 + a_2X_2 + a_3X_3 + \dots + a_NX_N + C \quad (1)$$

où $a_1, a_2, a_3, \dots, a_N$ et C sont des constantes.

C'est une équation linéaire à plusieurs variables, où X_1 représente la variable dépendante et X_2, X_3, \dots, X_N , les variables indépendantes.

Si $N = 2$, cette équation représente une droite, et si $N = 3$, un plan. Si le nombre de variables excède 4, on se trouve dans des espaces de dimension quatre, cinq (SPIEGEL).

L'estimation de l'équation (1), exprimant la liaison de plusieurs variables entre elles à partir de données expérimentales, par la méthode des moindres carrés, est appelée régression linéaire multiple.

Si la relation entre les variables n'est pas linéaire, on parle alors de régression non linéaire multiple. La relation entre les variables pourra prendre par exemple la forme suivante :

$$X_1 = a_1 + a_2X_2 + b_2X_2^2 + a_3X_3 + b_3X_3^2 + \dots + a_NX_N + b_NX_N^2 \quad (2)$$

De même que l'on exprime la corrélation linéaire simple par un coefficient, on définit le rapport de la variation expliquée à la variation totale par un coefficient de corrélation linéaire multiple, R .

On trouvera dans la littérature (SPIEGEL 1961, p. 270) les équations normales dont la résolution permet d'obtenir les termes constants des équations de régression.

3.2. *Calcul des équations de régression*

Les variables dépendantes et indépendantes étant fixées, on obtient les termes constants des équations de régression par résolution des équations normales.

Le recours à une calculatrice électronique permet ici une résolution des équations très rapide, d'autant plus qu'on trouve dans la littérature de nombreux programmes adaptés à ce genre de problèmes.

Le nombre de variables possibles est alors fonction des capacités de la calculatrice. Les exemples envisagés ont été traités à l'aide du programme REGRE sur une calculatrice IBM 1130, donnant la possibilité d'effectuer une régression multiple avec 22 variables.

3.3. *Limites de la méthode*

L'utilisation des régressions multiples en hydrologie est très souvent destinée à des travaux de prévision. Il résulte de considérations théoriques que les équations de régression ne pourront avoir un caractère prévisionnel uniquement dans les cas où les variables sont distribuées normalement (STAMMERS 1966). En remplaçant la variable par son logarithme, on contournera la difficulté dans le cas où sa distribution est log-normale.

Lorsque la régression est non linéaire, on peut être tenté d'abuser des degrés. Au-delà du deuxième degré, la précision devient souvent illusoire : si la variance des écarts entre les valeurs calculées et mesurées diminue avec l'élévation des degrés du polynôme, elle ne suit pas nécessairement cette tendance pour les points où les valeurs sont inconnues (MATHERON 1972).

D'autres méthodes se développent (WALLIS 1965, MATHERON *op. cit.*) qui, plus rigoureuses, permettront de se libérer de ces erreurs. La régression multiple n'est qu'une étape dans l'analyse des phénomènes naturels.

4. BILAN DES EAUX DE SURFACE : LE BASSIN DU SEYON / NE

4.1. *Introduction*

C'est la nécessité de prévoir le débit des cours d'eau qui voit actuellement de nombreux chercheurs recourir à la régression multiple pour l'étude du bilan des eaux de surface.

La régression multiple remplace alors avantageusement les méthodes de corrélation graphiques telle que la méthode coaxiale (LINSLEY et *al.* 1949). Le développement de fonctions analytiques appropriées pour définir les relations pluie-débit se poursuit, mais ces dernières ne sont pas encore parfaitement opérationnelles (SITTNER et *al.* 1969).

4.2. *Le bassin du Seyon*

Le bassin du Seyon ($S \cong 30 \text{ km}^2$) occupe la partie centrale d'un synclinal tertiaire séparant la première et la deuxième chaîne du Jura neuchâtelois. Les formations secondaires (calcaires) et tertiaires (marnes) sont recouvertes de dépôts glaciaires, en général imperméables à l'exception de quelques moraines et dépôts alluvionnaires graveleux occupés par des nappes d'eau souterraines. La réaction de la rivière aux précipitations est très rapide.

4.3. *Bilan des eaux de surface*

Les relations classiques pour l'étude du bilan des eaux de surface font intervenir les variables suivantes :

- Un indice d'antécédence des précipitations (Antécédent Precipitation Index, API), logarithmique (LINSLEY *op. cit.*) et défini par la relation suivante :

$$\text{API}(j) = \text{API}(j-1) \cdot F + P(j-1)$$

où F est un facteur de récession valant en général 0.9 et $P(j)$ les précipitations du jour j .

- Un indice mensuel ou hebdomadaire, défini par le numéro d'ordre du mois ou de la semaine où se déroule l'averse.
- La durée des précipitations, exprimée en heure ou en jour.
- La lame d'eau tombée durant l'averse, en mm.
- Une variable dépendante qui est la lame d'eau exprimée en mm, écoulée à l'exutoire pour une averse donnée.

La souplesse de la méthode permettrait de faire intervenir d'autres éléments tels que l'humidité du sol, avant et au cours de l'averse, en plusieurs points du bassin, ou la température moyenne de l'air par exemple.

Pour le bassin du Seyon, 72 averses, réparties au cours de l'année hydrologique, ont été sélectionnées pour 25 années d'observations des débits. Pour des raisons de simplification et vu la répartition régulière des averses dans le bassin, API, durée de l'averse et lame d'eau tombée ont été étudiées en une seule station. Là encore, la régression multiple permettrait d'introduire les données pluviométriques de plusieurs stations simultanément.

Les données de pluie et de débit étant journalières, il était impossible d'étudier le bilan pour des périodes inférieures à 24 heures, d'où une certaine imprécision dans l'estimation de la durée des averses. La lame d'eau écoulée pour chaque averse a été obtenue par planimétrie des graphiques des débits, après décomposition de l'hydrogramme, compte tenu de la recharge des eaux souterraines.

4.4. Recherche et optimalisation des équations de régression

Soit : X_1 la lame d'eau écoulée
 X_2 l'API
 X_3 le numéro d'ordre de la semaine
 X_4 la durée de l'averse
 X_5 la lame d'eau tombée

On recherche d'abord les coefficients de l'équation de régression linéaire reliant ces 5 variables. On trouve pour le bassin du Seyon :

$$X_1 = 0,26X_2 + 0,05X_3 + 2,10X_4 + 0,37X_5 - 15,91 \quad (4)$$

où $R = 0,71$.

L'augmentation progressive du nombre de variables sous la forme de produit et de puissance des 4 variables indépendantes, en procédant par « tâtonnement raisonné », permet d'optimiser la valeur du coefficient de corrélation multiple. Certains termes de l'équation sont également écartés ou modifiés en testant la valeur du coefficient de régression partielle. Le polynôme sélectionné définitivement pour représenter l'équation du bilan des averses dans le bassin du Seyon vaut :

$$\begin{aligned} X_1 = & -1,54X_2 - 7,24X_3 - 5,65X_4 + 1,10X_5 \\ & + 0,17X_2X_3 + 0,52X_2X_4 - 0,0031X_2X_5 \\ & - 0,10X_3X_4 - 0,021X_3X_5 + 0,00037X_4X_5 \\ & - 0,0063X_2^2 + 0,12X_3^2 + 0,00073X_5^2 \\ & - 0,0021X_2^2X_3^2 - 0,0020X_2^2X_4 - 0,010X_2X_3X_4 \\ & + 0,0057X_3^2X_4 + \frac{66 \cdot 24}{1 + 0,001X_2^2} + 17,42 \end{aligned} \quad (5)$$

où $R = 0,78$.

On constate que l'amélioration du coefficient de corrélation multiple est peu importante, malgré la complexité de l'équation 5. Utilisée à des fins prévisionnelles, cette dernière ne présentera pour ainsi dire aucune amélioration par rapport à l'équation 4.

5. INFLUENCE DES FACTEURS MÉTÉOROLOGIQUES SUR L'ÉVAPOTRANSPIRATION : LE BASSIN SUPÉRIEUR DE L'AREUSE

5.1. Situation

Le bassin occupe une région calcaire du Jura helvétique plissé. Il s'étend sur la vallée de La Brévine (polje de 85 km²) et sur la partie orientale de la vallée des Verrières. L'une et l'autre sont formées par deux synclinaux qui se relaient (BURGER 1959).

Les calcaires du Crétacé et du Malm, de quelque 500 m d'épaisseur, reposent sur les marnes de l'Argovien qui constituent le soubassement imperméable généralisé.

La décharge des eaux souterraines s'opère par une seule grande source : la source de l'Areuse, jaillissant au village de Saint-Sulpice, à 790 m d'altitude.

Superficie du bassin	129 km ²
Altitude moyenne	1115 m
Pluviosité moyenne	1510 mm
Débit moyen de la source	4,68 m ³ s ⁻¹ (1959-1971)

5.2. Recherche de l'évapotranspiration réelle

L'établissement de formules mathématiques déterministes exprimant l'évapotranspiration réelle (ETR) à partir des paramètres hydro-météorologiques (précipitation, insolation, température, etc.) a fait l'objet de nombreux travaux (THORNTHWAITE, SZESZTAY, etc.).

Dans les bassins karstiques élevés du Jura suisse et français, l'ETR calculée selon ces formules est en général inférieure à l'ETR déduite du bilan hydrogéologique à partir de l'équation ci-dessous (DELAROZIÈRE 1968, TRIPET 1971) :

$$P = ETR + Q + \Delta ws + \Delta wn \quad (6)$$

- où
- P = les précipitations
 - Q = le volume d'eau écoulé à l'exutoire
 - Δws = la variation des réserves souterraines
 - Δwn = la variation des réserves sous forme de neige

En effet, les méthodes classiques de calcul de l'ETR sont souvent trop générales ou font intervenir des paramètres peu connus, comme les réserves en eau du sol, dont le rôle est très réduit dans les régions karstiques, peu couvertes.

L'emploi des régressions multiples va permettre la recherche d'une formule de calcul de l'ETR, à partir des paramètres hydrométéorologiques, qui s'adaptera mieux au bassin.

Dans le bassin de l'Areuse, à la suite des travaux de TRIPET (*op. cit.*), les paramètres du bilan (P, Q, Δws , Δwn) sont bien connus. On constate que l'ETR déduite du bilan est rarement égale à l'évapotranspiration potentielle (sauf pour les mois très pluvieux).

5.3. Recherche de l'ETR par régression multiple

Les paramètres hydrométéorologiques mensuels suivants ont été retenus :

- P = les précipitations du mois
- T_m = la température moyenne du mois
- T_{m-1} = la température du mois précédent
- I_g = l'irradiation globale
- H_j = l'humidité relative moyenne
- T_s = le nombre de jours où l'humidité relative est inférieure à 50%
- I_p = la densité moyenne des précipitations

API = l'indice d'antécédence des précipitations mensuel moyen
 T_p = la durée moyenne des précipitations
 P_a = la pression atmosphérique moyenne
 V = la vitesse moyenne du vent

soit au total 11 variables indépendantes et une variable dépendante (ETR), dont les valeurs mensuelles des mois d'été (mai à octobre) ont été établies pour le bassin supérieur de l'Areuse entre 1960 et 1969.

La normalité des variables avait préalablement été testée graphiquement (fig. 1 à 3). Une sélection progressive des variables importantes selon la valeur des coefficients de corrélation partiels et multiples, et optimisation par élévation à la puissance ou recherche des racines carrées des variables sélectionnées, a permis d'établir l'équation définitive suivante :

$$ETR_m = 0,52 P + 3,25 T_m + 1,19 T_{m-1} + 1,43 \sqrt{I_g} - 0,015 API_m^2 - 0,07 I_p^2 - 54,2 \quad (7)$$

avec un coefficient de corrélation multiple $R = 0,89$.

5.4. Importance relative des variables sur l'ETR

Certaines variables comme la vitesse du vent [V], la pression atmosphérique [P_a] présentait des coefficients de régression partiels très bas ; elles ont été rapidement écartées.

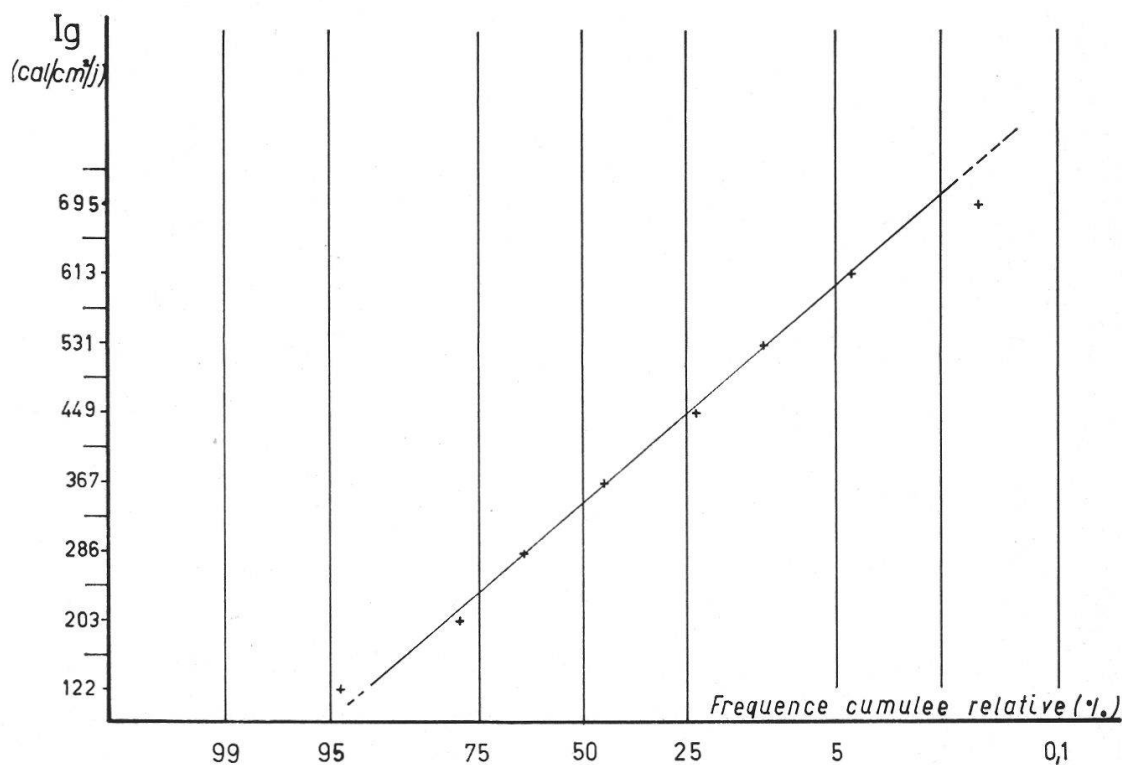


Fig. 1. Distribution de l'irradiation globale mensuelle moyenne dans le bassin de l'Areuse (1960-1969).

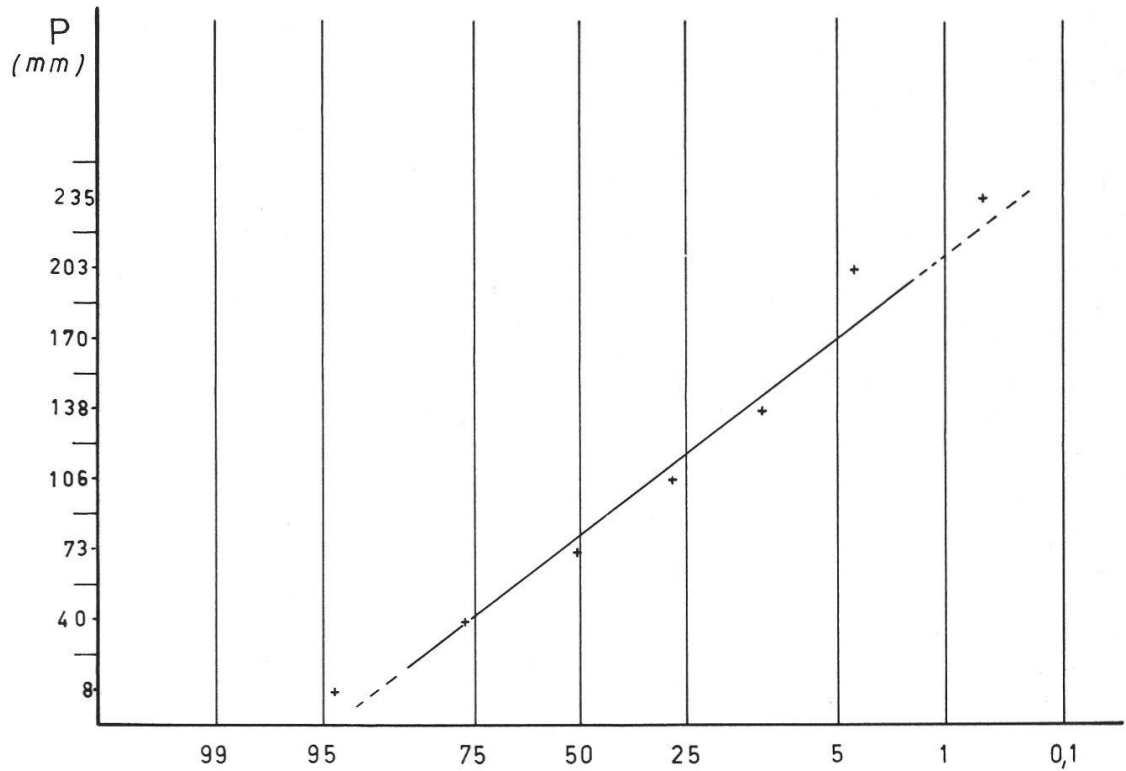


Fig. 2. Distribution des précipitations mensuelles moyenne dans le bassin de l'Areuse.
Test de normalité.

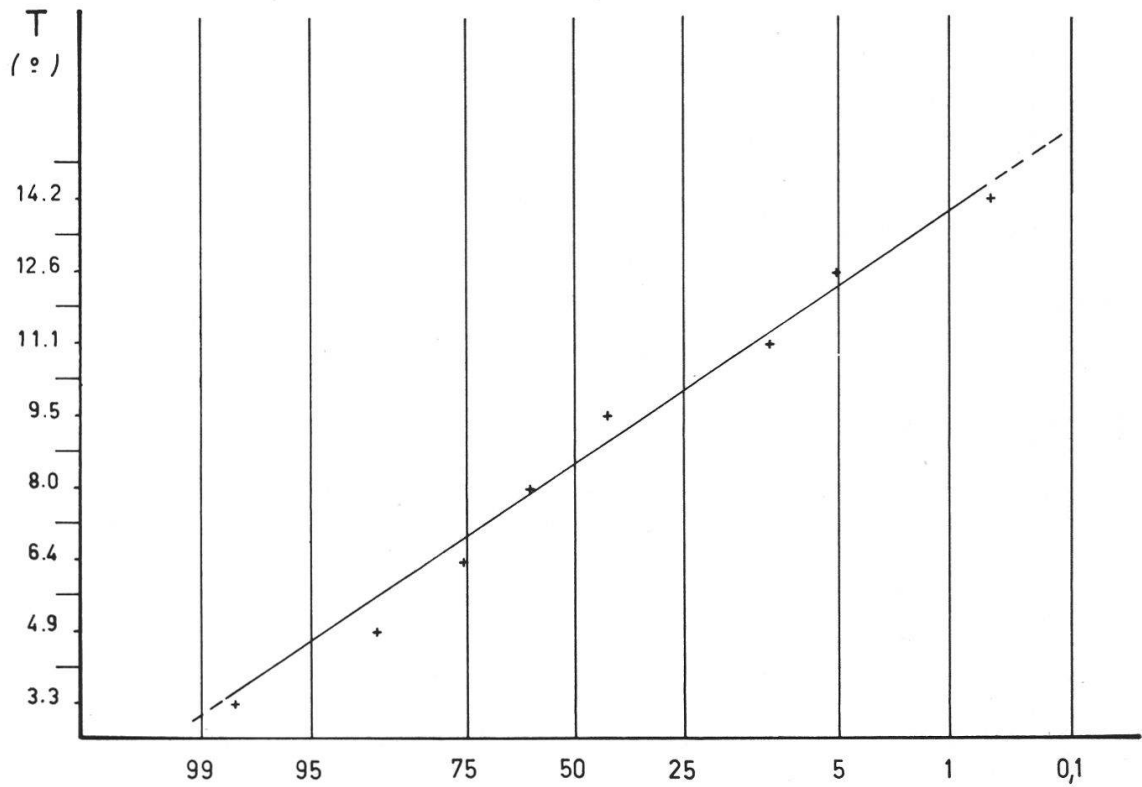


Fig. 3. Distribution des températures mensuelles moyenne dans le bassin de l'Areuse.
Test de normalité.

Le rôle de l'humidité relative $[H_j]$ s'est avéré également très peu important.

En revanche, les paramètres liés aux précipitations, à la température moyenne et à l'irradiation globale semblent jouer un rôle prépondérant dans les phénomènes d'évapotranspiration. L'effet non négligeable de la température du mois précédent peut surprendre; l'introduction de variables telles que API_{m-1} et API_{m-2} améliore également le coefficient de corrélation multiple (0,90). Ceci traduit l'influence retardée des températures et précipitations sur le développement des végétaux et par conséquent sur l'évapotranspiration.

Remarque :

Le calcul de l'ETR n'a pas été envisagé pour les mois d'hiver. En effet, Δw_n est connu approximativement et la notion d'API perd partiellement sa signification.

5.5. Discussion

Cette formule de calcul de l'ETR (7) ne peut être appliquée qu'à des bassins présentant des conditions géologiques et climatiques voisines.

En revanche, elle est beaucoup mieux adaptée aux conditions locales (fig. 4) que les formules classiques, très simplifiées quand elles sont générales ou trop complexes quand elles veulent tenir compte de toutes les situations particulières.

6. CONCLUSIONS

L'utilisation des régressions multiples dans le cas où les variables sont distribuées normalement reste un excellent moyen pour estimer l'influence relative des paramètres hydrologiques dans des situations où la complexité des phénomènes ne permet pas l'établissement de lois déterministes.

Dans le bassin supérieur de l'Areuse, par une sélection des paramètres les plus importants, cette méthode a permis d'établir une formule de calcul relativement simple et très satisfaisante de l'évapotranspiration mensuelle. Cette démarche présente également un avantage certain pour le choix des paramètres à prendre en considération lors de l'établissement d'un réseau d'observation hydrologique.

Remerciements

L'étude des eaux souterraines du bassin supérieur de l'Areuse entre dans le cadre de recherches financées par un comité d'étude comprenant l'Office fédéral de l'économie hydraulique, l'Etat de Neuchâtel, ainsi que les villes de La Chaux-de-Fonds et du Locle.

Nos remerciements vont à ces institutions et à leurs représentants.

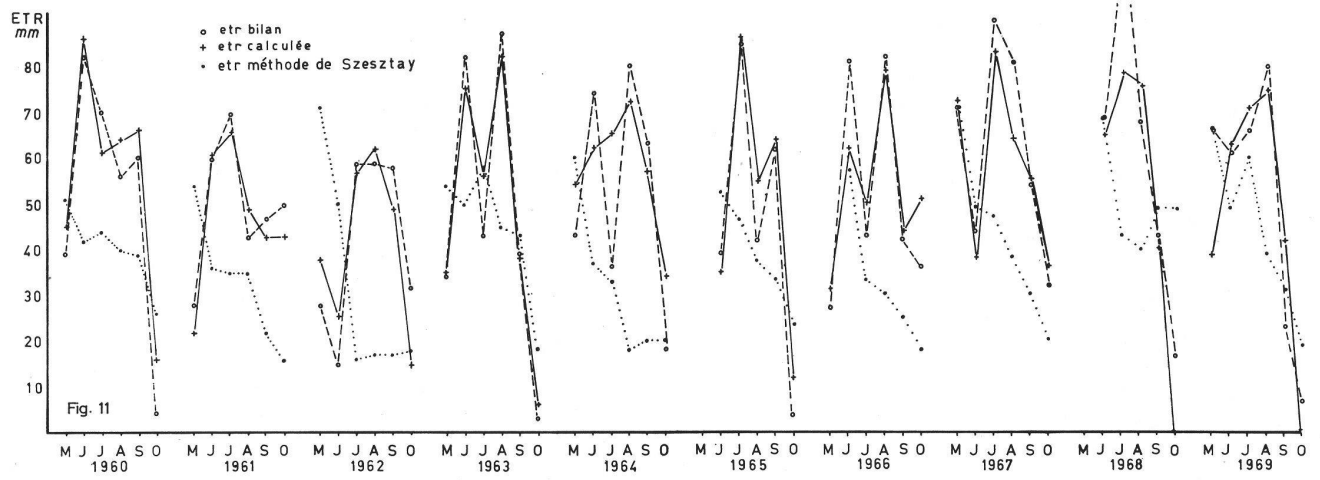


Fig. 4. Evapotranspiration réelle dans le bassin supérieur de l'Areuse. En trait plein : valeurs observées ; - - - : valeurs calculées par régression multiple ; : valeurs calculées par la méthode de SZESZTAY.

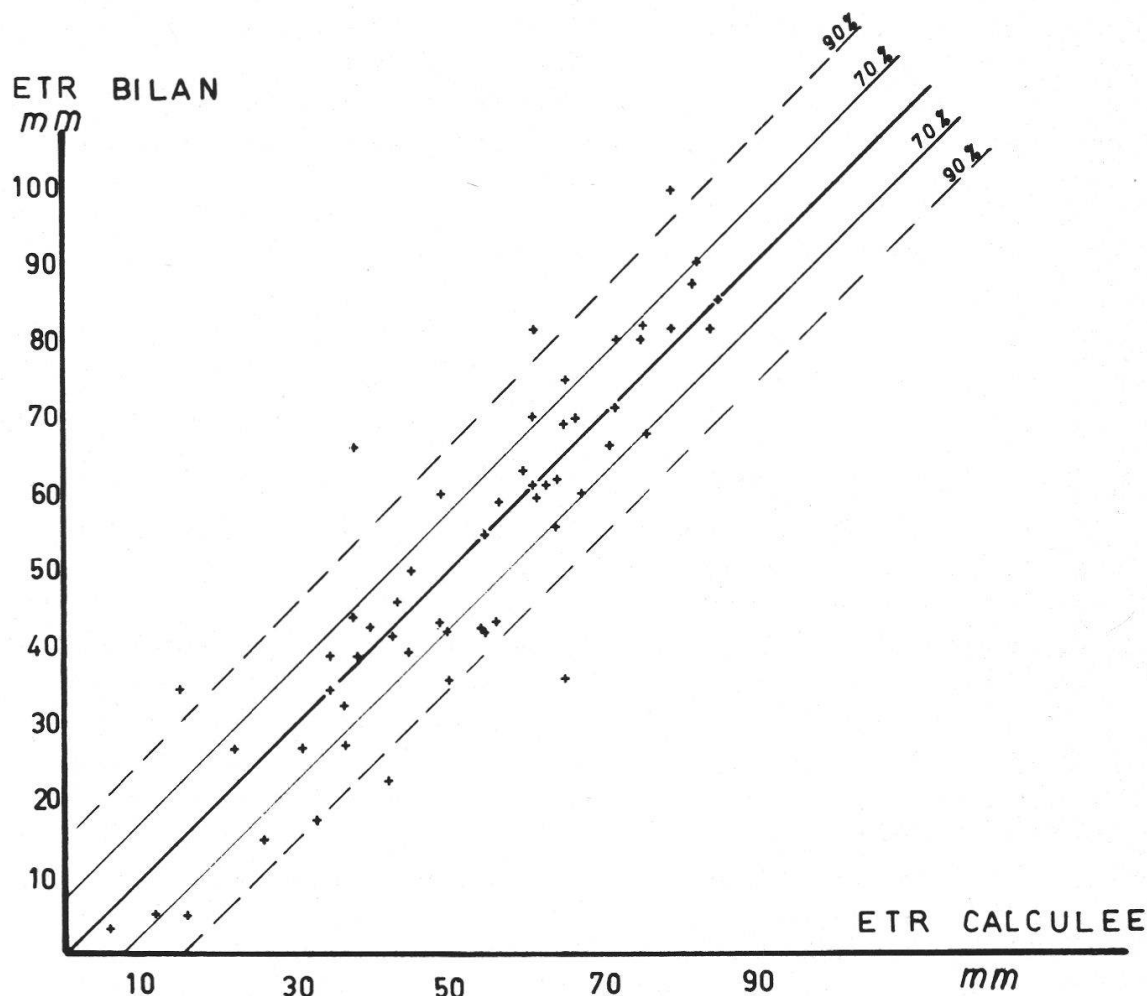


Fig. 5. Evapotranspiration réelle calculée et observée dans le bassin supérieur de l'Areuse ; largeur de la bande de confiance. Le 90% des valeurs tombe à l'intérieur des tirets.

Résumé

Deux applications des régressions multiples sont envisagées. L'une comprend le calcul du bilan des averse dans le bassin du Seyon pour l'établissement des relations pluie-débit.

Dans la seconde, on recherche une formule de calcul de l'évapotranspiration réelle mensuelle pour un bassin karstique élevé (bassin supérieur de l'Areuse). La méthode permet de rechercher l'influence relative des paramètres hydrométéorologiques sur l'évapotranspiration réelle. En particulier, on constate l'effet non négligeable de la température moyenne du mois précédent.

Summary

Two applications of the multiple regressions are considered. The first deals with the calculation of the rainfall-runoff relations in the River Seyon basin (30 km²).

In the second a calculation formula of the monthly real evapotranspiration in looked for, for a elevated karstic basin.

The method allows the research on the relative influence of hydro-meteorological parameters on the real evapotranspiration to be undertaken.

In particular a relative importance of the effect of the average temperature of the previous month is noticed.

BIBLIOGRAPHIE

- BOX, G. E. P. and JENKINS, G. M. — (1970). *Time Series Analysis, Forecasting and Control*. 553 pp. (Edit. Holden Day).
- BURGER, A. — (1959). *Hydrogéologie du bassin de l'Areuse*. Thèse, Univ. Neuchâtel, 304 pp.
- DELAROZIÈRE, O. — (1968). *Contribution à l'étude du bassin du Doubs ; climatologie, hydrologie et déficit d'écoulement rapportés aux unités géologiques*. B.R.G.M., Lyon, 68 SGL 162 JAL, 82 pp.
- GUILLOT, P. — (1971). *Quelques remarques sur l'utilisation des modèles mathématiques pour la prévision en hydrologie et météorologie*. *La Houille Blanche* 3 : 209-217.
- HAAN, C. T. and ALLEN, D. M. — (1972). *Comparison of Multiple Regression and Principal component Regression for predicting Water in Kentucky*. *Water Resources Research* 8 (6) : 1593-1596.
- IBM — (1970) : 1130 Scientific Subroutine Package Programmers Manual. E. GH 20 - 0252-4, pp. 23-29, 151-156.
- LINSLEY, R. K., KOHLER, M. A. and PAULHUS, J. L. H. — (1949). *Applied Hydrology*, pp. 405-443 (McGraw Hill).
- MATHERON, G. — (1972). *Théorie des variables régionalisées, II, le krigeage universel*. *Traité d'informatique en géologie*, pp. 362-378 (Masson).
- SPIEGEL, M. R. — (1961). *Theory and Problems of Statistics*. 359 pp. (Schaum publishing Co.).
- STAMMERS, W. M. — (1966). *The application of Multivariate Techniques in Hydrology*. *Statistical Methods in Hydrology, Proceeding of Hydrology. Symposium No. 5, Mc Gill University, Canada*, pp. 255-267.
- TRIPET, J.-P. — (1972). *Etude hydrogéologique du bassin supérieur de l'Areuse (Jura neuchâtelois)*. Thèse, Univ. Neuchâtel, 183 pp.
- WALLIS, J. R. — (1965). *Multivariate Statistical Methods in Hydrology — A Comparison Using Data of Know Functional Relationship*. *Water Resources Research* 1 (4) : 447-461.
-