

Cah. ORSTOM, sér. Biol., n° 11 - avril 1970.

LES NÉMATODES PHYTOPARASITES DES RIZIÈRES INONDÉES DE CÔTE D'IVOIRE

II. - ESSAI D'ESTIMATION DE L'IMPORTANCE DES POPULATIONS

PAR

G. MERNY * et J. DÉJARDIN **

RÉSUMÉ

L'étude des populations de quelques espèces de nématodes contenues dans 100 unités échantillons par hectare, prélevées dans deux rizières du nord de la Côte d'Ivoire a montré que les populations n'étaient que très rarement distribuées au hasard. Elles sont presque toutes agrégées et vérifient la même loi de Taylor dont l'indice d'agrégation est 1,65. Presque toutes les distributions observées s'ajustent à une loi binomiale négative ou en sont assez voisines.

Si l'on excepte les populations faibles, la normalisation des distributions est possible par transformation logarithmique $y = \log(x + x_0)$, dans laquelle x_0 est une grandeur variant de 0,1 à 20, qu'on détermine graphiquement pour chaque distribution.

L'estimation de la moyenne et de ses limites de confiance est possible sur les données transformées. La retransformation introduit un biais important dans cette estimation, qu'on pallie en appliquant aux grandeurs ainsi calculées un coefficient C obtenu en calculant la somme d'une série convergente d'après une formule élaborée par NEYMAN & SCOTT (1960), faisant entrer en ligne de compte le nombre d'unités d'échantillonnage, la variance des données transformées et le nombre de degrés de liberté de cette variance. Cette méthode n'est pas valable pour les populations très faibles pour lesquelles les données ne sont pas normalisables.

L'étude d'un certain nombre de séries de 10 unités d'échantillonnage a montré que cette méthode ne permettait une estimation acceptable de la moyenne que pour des populations importantes. Quand une certaine précision est nécessaire, l'étude d'une centaine d'unités à l'hectare est indispensable.

* Laboratoire de Nématologie. Centre O.R.S.T.O.M. d'Adiopodoumé. B.P. 20, Abidjan (Côte d'Ivoire).

** Service de Biométrie. Services Scientifiques Centraux de l'O.R.S.T.O.M., Bondy (France).

SUMMARY

The study of the populations of some nematode species contained in 100 sampling units per hectare collected from two flooded rice fields in the north of the Ivory Coast showed that the populations are almost never randomly distributed. They are generally aggregated and follow the same Taylor law with an aggregation indice of 1.65. Almost all the observed distributions fit the negative binomial or a similar law.

If weak populations are excepted, the distributions can be normalised by a logarithmic transformation of the data : $y = \log(x + x_0)$, x_0 being a number varying between 0.1 and 20, determined graphically for each distribution.

The estimation of the mean and its confidence limits is possible with the transformed data. Retransformation introduces an important bias in the estimation, which can be overcome by applying a coefficient C obtained by developing a convergent series according to the formula given by NEYMAN & SCOTT (1960), taking into account the number of sampling units, the variance of the transformed data and the number of degrees of freedom of this variance. This method is not valuable for very weak populations, the data of which cannot be normalised.

The study of series of 10 sampling units only led to a suitable estimation, by the method described above, for large populations. When an accurate estimation is wanted, especially for moderate or moderately weak populations, about 100 sampling units must be used.

INTRODUCTION

Il est fréquent que, exposant les résultats d'une enquête sur les nématodes présents dans une culture ou un biotope, des auteurs donnent des estimations de l'importance des populations de telle ou telle espèce. Ces estimations résultent de l'extrapolation, à un poids ou un volume donné de sol, du comptage effectué sur une ou plusieurs parties aliquotes d'un ou plusieurs échantillons. Malheureusement, les auteurs n'indiquent pas toujours le nombre d'unités d'échantillonnage étudiées ni leur mode de prélèvement. Bien souvent, un seul nombre est fourni sans qu'on puisse savoir quel crédit on peut lui accorder ni dans quelles limites il est valable.

Il est souvent très important de connaître, avec une certaine exactitude, la taille d'une population de nématodes parasites présents dans une aire donnée, soit que la dynamique de cette population fasse l'objet même de l'étude, soit qu'ayant appliqué un traitement quelconque on désire en mesurer les effets, soit encore qu'avant d'entreprendre une certaine culture on veuille savoir s'il existe une population de parasites représentant, pour elle, une menace grave.

Dans les cas précis d'*Heterodera rostochiensis* et *Heterodera schachtii*, qui font l'objet d'une surveillance systématique dans plusieurs pays du nord de l'Europe, les méthodes d'estimation des populations de kystes dans le sol ont été décrites par divers auteurs (OOSTENBRINK, 1950 ; GOFFART, 1962). Elles consistent toutes à faire, au champ, un grand nombre de petits prélèvements, à les mélanger dans un échantillon global représentant une superficie donnée et à compter des parties aliquotes de cet échantillon. JONES (1955) a montré que les valeurs obtenues avec les parties aliquotes pouvaient être considérées comme se distribuant selon une loi voisine de celle de POISSON. Ainsi que le fait remarquer JONES, cette méthode n'est valable que pour une évaluation grossière, suffisante quand on effectue une enquête à but agronomique. Elle est insuffi-

sante quand on veut connaître avec précision l'infestation d'un terrain parce qu'elle ne tient pas compte des variations de la population d'un point à un autre du terrain. WIDDOWSON (1962) a étudié la population de kystes d'*Heterodera rostochiensis* dans un terrain de 25 ares, divisé en 364 carrés, deux échantillons globaux étant étudiés dans chaque carré. La variance entre les carrés correspondait dans ce cas, à un coefficient de variation de 155%.

Lorsqu'on étudie non plus des kystes mais des nématodes mobiles dans le sol, on se heurte à une nouvelle difficulté : si l'on veut mélanger le sol de façon à obtenir un échantillon homogène, on a toutes chances de tuer une forte proportion des nématodes présents. Il peut arriver aussi, comme c'est le cas dans les rizières, qu'on soit en présence d'une terre argileuse difficile, sinon impossible, à bien mélanger.

L'un des auteurs (MERNY, 1970) ayant, au cours d'une enquête portant sur 72 points de prélèvements, répartis dans deux régions de la Côte d'Ivoire, dressé une liste des espèces phytoparasites trouvées dans les rizières, la question s'est posée de déterminer celles qui étaient vraiment des parasites actifs du riz. En dehors des espèces ayant un mode de vie totalement ou partiellement endophyte et dont le parasitisme est prouvé par le seul fait qu'on les extrait des racines, des hypothèses peuvent être faites en se basant sur deux critères : la fréquence et l'abondance.

En effet, si une espèce rarement rencontrée autour des racines d'une autre plante est fréquente autour de celles du riz, il y a forte présomption pour qu'elle soit un parasite actif de celui-ci. Par ailleurs, on imaginerait mal un parasite abondant autour des racines du riz et qui ne tirerait pas sa subsistance de celles-ci.

La fréquence des différentes espèces ayant été déterminée, il restait à connaître leur abondance. Au cours de l'enquête, une première approximation avait été faite ne portant que sur une unité-échantillon par point de prélèvement. Une rizière est un écosystème qui semble homogène, relativement peu étendu et aux limites nettement définies. On pouvait donc s'attendre à une certaine homogénéité des populations de nématodes, mais il y avait fort peu de chances pour qu'une seule unité-échantillon puisse être considérée comme représentative de l'ensemble.

Il nous a donc paru essentiel, avant toute étude quantitative, de définir la méthode à employer et d'évaluer la confiance à accorder aux résultats ainsi obtenus. Nos études ont porté sur les lois de distribution constatées, la normalisation des données et l'estimation des populations moyennes, le but final étant de connaître le nombre d'unités-échantillons nécessaires à l'obtention d'une précision donnée.

LOIS DE DISTRIBUTION

Les spécialistes de l'étude des populations animales distinguent trois sortes de distribution :

1. — *Distribution régulière* où chaque individu occupe toujours la même surface ou, plus exactement, le même volume de sol. Des échantillons issus d'une population distribuée de cette façon montreront une distribution uniforme du nombre de nématodes par unité prélevée. Elle n'est pratiquement jamais observée.

2. — *Distribution au hasard*. Les individus se répartissent de façon aléatoire dans le sol. La présence d'un individu en un point n'influence nullement la répartition des autres et la loi de distribution observée pour le nombre d'individus par unité-échantillon sera une loi de Poisson.

Pour cette loi, la fréquence simple du nombre x d'exemplaires observés, f_x , est :
 $f_x = e^{-\mu} \frac{\mu^x}{x!}$, avec $\mu > 0$. Sa moyenne et sa variance sont toutes deux égales à μ .

3. — *Distribution agrégée.* Les individus ne se répartissent pas complètement au hasard dans le sol, mais forment des amas de plus forte densité. Des échantillons extraits de telles populations montreront des lois de distribution du type contagieux dont la plus courante est la loi binomiale négative. Pour cette loi la fréquence simple f_x est le terme général du développement du binôme $(q - p)^{-k}$, soit :

$$f_x = \frac{\Gamma(k + x)}{x! \Gamma(k)} \left(\frac{1}{q}\right)^k \left(\frac{p}{q}\right)^k$$

avec : $k > 0, p > 0, q = 1 + p$.

Sa moyenne est : $\mu = kp$ et sa variance $\sigma^2 = kpq = \mu q = \mu + \frac{\mu^2}{k}$.

Nous avons, à deux reprises, une fois pendant l'intercampagne et une autre fois à l'époque de la maturation du riz, récolté des échantillons de $n = 100$ prélèvements dans une rizière d'environ 1 ha. La première fois, les prélèvements ont eu lieu à la ferme semencière de Nambingué en juillet 1965, c'est-à-dire pendant l'intercampagne, et la seconde à la station de recherches de Ferkéssédougou en décembre 1965, au moment de la maturation du riz. Les nématodes ont été extraits de 125 cm³ de chaque prise et les individus appartenant aux espèces phytoparasites ont été comptés sur 1/5 des nématodes extraits.

Les résultats de ces observations sont représentés par les figures 1 et 2.

Le simple examen des diagrammes de la figure 1, résumant les observations de Nambingué, montre qu'on n'est pas généralement en présence de distributions de POISSON, ainsi que le confirme le tableau I dans lequel, sauf pour *Criconemoides palustris*, la variance est toujours très largement supérieure à la moyenne alors que ces deux quantités doivent être sensiblement égales dans le cas d'échantillons extraits d'une distribution « au hasard ».

TABLEAU I

Espèce	Moyenne	Variance
<i>Hirschmanniella spinicaudata</i>	14,32	592,12
<i>Trichotylenchus rhopalocercus</i>	23,12	395,18
<i>Heterodera sacchari</i>	0,20	0,53
<i>Xiphinema</i> sp. « ae »	0,67	2,53
<i>Criconemoides palustris</i>	0,03	0,029

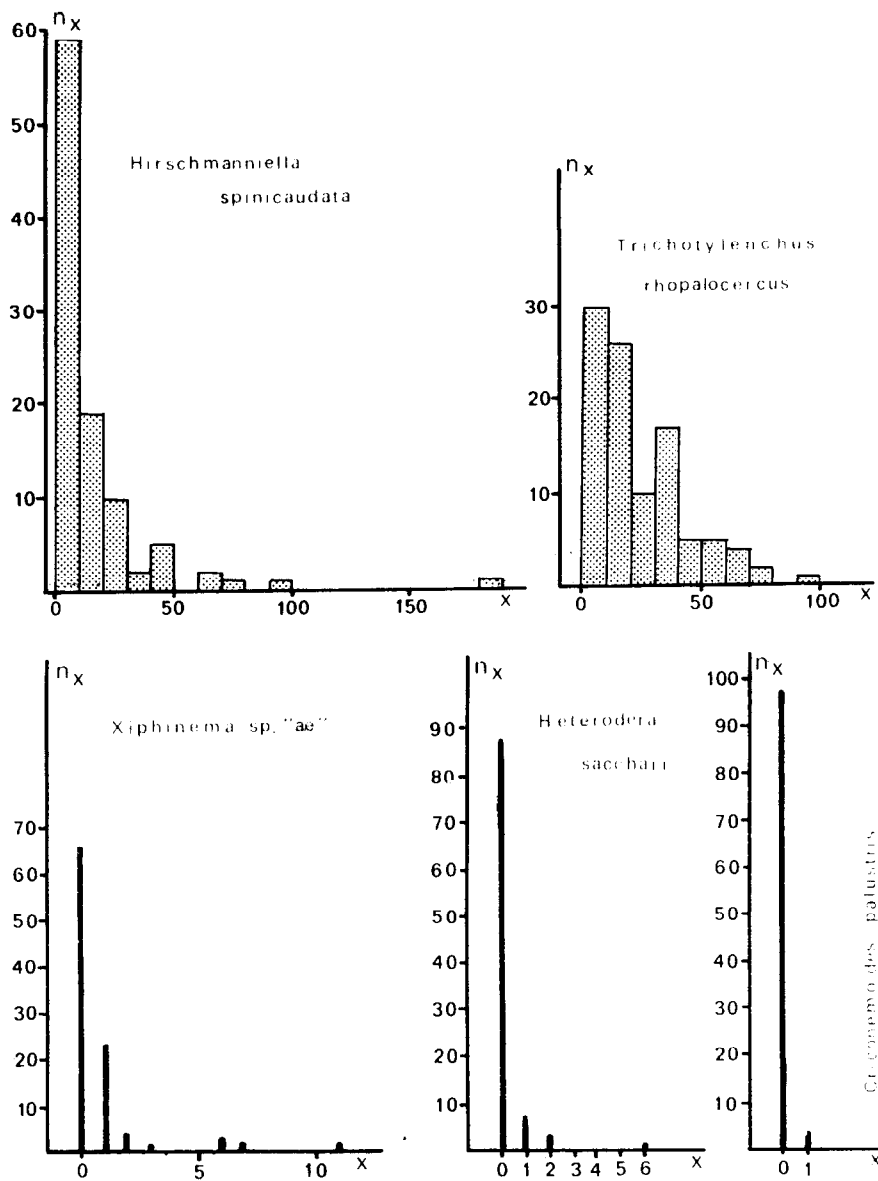


FIG. 1. — Distribution de cinq espèces dans 100 unités-échantillons (Nambingué). x : nombre d'individus dans une unité-échantillon ; nx : nombre d'unités-échantillons contenant x individus.

Pour *Criconemoides palustris* la distribution observée s'ajuste exactement à une distribution de POISSON :

- χ^2 de comparaison de la variance à la moyenne 95,7 à 99 d.d.l., non significatif ;
- Effectifs théoriques : 97,046 et 2,954, comparés aux effectifs observés : 97 et 3.

Dans le cas d'*Hirschmanniella spinicaudata*, *Xiphinema* sp «ae» et *Heterodera sacchari*, l'ajustement à une loi binomiale négative a été tenté, il est satisfaisant (tests de KOLMOGOROFF-SMIRNOV non significatifs au seuil 5% : $|d_{\max}| = 0,082 ; 0,060 ; 0,009$, tous $< 0,136$).

Les observations faites à Ferkéssédougou, en décembre 1965, ont donné des résultats du même ordre (fig. 2 et tableau II).

L'ajustement à une loi binomiale négative a été tenté pour *Xiphinema* sp et pour *Hirschmanniella spinicaudata*. Pour la première espèce, il est bon, pour la seconde, le test de KOLMOGOROFF-SMIRNOV est significatif au seuil 5% ($|d_{\max}| = 0,146 > 0,136$). Cependant, pour le moment, nous admettons le modèle binomial négatif compte tenu du fait qu'il s'agit d'un test pris parmi d'autres et que la différence est loin d'être significative à 1% ($0,146 < 0,163$). Ceci est d'ailleurs de peu d'importance car, ainsi qu'il sera montré au paragraphe suivant, la normalisation des résultats s'effectue de la même manière.

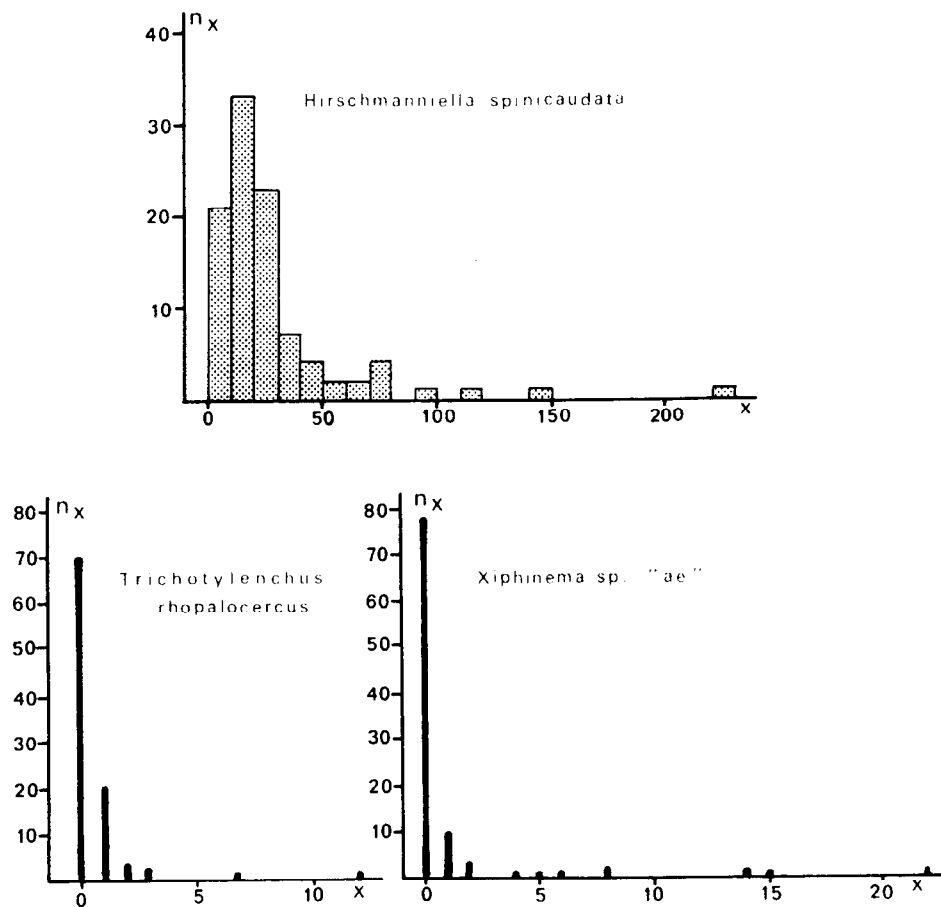


Fig. 2. — Distribution de trois espèces dans 100 unités-échantillons (Ferkéssédougou). x : nombre d'individus dans une unité-échantillon ; n_x : nombre d'unités-échantillons contenant x individus.

Pour *Trichotylenchus rhopalocercus*, aussi bien à Nambingué qu'à Ferkéssédougou, l'ajustement à une loi binomiale négative est satisfaisant (tests de KOLMOGOROFF-SMIRNOV non significatifs au seuil 5% : $|d_{\max}| = 0,068$ et $0,043 < 0,136$).

TABLEAU II

Espèce	Moyenne	Variance
<i>Hirschmanniella spinicaudata</i>	27,13	974,24
<i>Trichotylenchus rhopalocercus</i>	0,57	2,268
<i>Xiphinema</i> sp. « ac »	1,08	12,116

NORMALISATION DES DONNÉES

Etude de la liaison moyenne-variance. Loi de Taylor.

Il est bien certain que, pour définir une population aussi variable, la seule moyenne du nombre d'individus par unité-échantillon est insuffisante et même, dans certains cas, de très peu de signification. Il est indispensable d'indiquer le crédit qu'on peut accorder au nombre ainsi énoncé en calculant son intervalle de confiance, calcul qui ne peut s'effectuer classiquement que sur une population normalement distribuée ou, pour le moins, s'ajustant d'une manière suffisamment satisfaisante à une loi normale.

La première opération consiste donc à normaliser les données brutes en leur faisant subir une transformation.

Le simple vu des figures 1 et 2 montre que l'hypothèse de normalité doit être écartée. De plus les quelques moyennes et variances qui viennent d'être données montrent que ces deux caractéristiques ne sont pas indépendantes. Nous avons donc, pour trouver la transformation, étudié la liaison qui existe entre la moyenne et la variance en utilisant d'une part les caractéristiques qui viennent d'être fournies et, d'autre part, celles de 77 échantillons de $n = 10$ prélèvements non détaillés ici (1).

La figure 3 traduit graphiquement les résultats observés. Le nuage de points est parfaitement rectilinéaire, il n'est pas besoin de test pour l'admettre. La pente de la droite moyenne est de 1,650 et son ordonnée à l'origine de 0,6290 (= log 4,256). La variance est donc liée à la moyenne par la relation : $s^2 = 4,26 (\bar{x})^{1,65}$.

Nous trouvons ici une remarquable vérification de la loi de TAYLOR (1961) $s^2 = a\bar{x}^b$ pour des intervalles extrêmement importants (moyenne : de 0,03 à 75 ; variance : de 0,03 à 12 700) rarement observés, à notre connaissance, jusqu'à ce jour. L'indice d'agrégation, selon TAYLOR, est de 1,65.

(1) Les données numériques seront fournies sur simple demande, faite au Service de Biométrie, S.S.C., 70, route d'Aulnay, 93-Bondy (France).

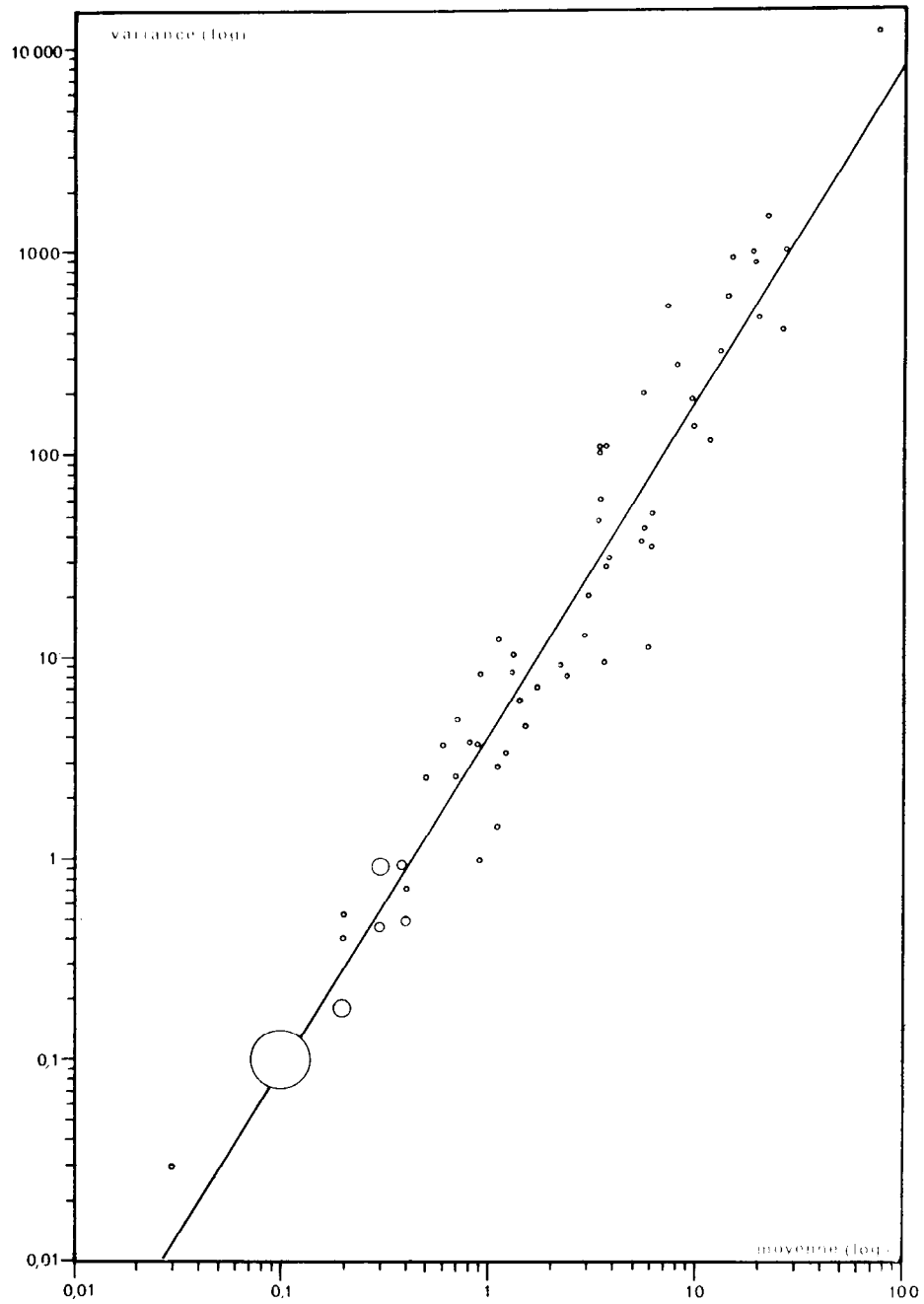


FIG. 3. — Liaison entre la moyenne et la variance dans les échantillons observés (vérification de la loi de Taylor).

Normalisation proprement dite.

Nous pouvons toujours, en utilisant la transformation $(x + x_0)^p$ avec $p = 1 - \frac{b}{2}$ et tabulée pour certaines valeurs de p par HEALY & TAYLOR (1962), stabiliser les variances ; nous trouverions ici $y = (x + x_0)^{0,2}$. Toutefois il ne nous sera jamais possible de normaliser, même simplement de rendre symétriques, des distributions présentant un effectif de la classe 0 très important. Un exemple limite nous est fourni par la distribution de *Criconemoides palustris* donnée ci-dessus. Ces distributions sont caractérisées par des moyennes faibles. D'autre part sur le plan pratique le calcul de $y = (x + x_0)^{0,2}$ est mal commode, la détermination de x_0 n'étant pas facile. Nous nous sommes donc attachés à rechercher une transformation normalisante plus simple, valable pour des distributions présentant des moyennes suffisamment élevées de telle façon que la normalité, ou presque, de la distribution de la variable transformée puisse être atteinte.

La figure 4 est la reprise d'une partie du graphique de la figure 3 : ne sont portés sur ce graphique que les points correspondants aux moyennes supérieures à 3 et variances supérieures à 20. La droite tracée en pointillé est la droite générale d'ajustement de tout le nuage de la figure 3 et celle tracée en trait plein la droite de pente 2 passant par le point moyen du nuage partiel : aucun test n'est indispensable pour admettre la validité de l'ajustement de cette droite de pente 2 qui correspond à la transformation logarithmique : nous avons donc là une solution simple à la transformation des données du haut du graphique : la transformation logarithmique $y = \log(x + x_0)$, x_0 étant une constante différente de zéro, variable, comme nous allons le voir, d'une population à l'autre et qui a pour effet de faire disparaître l'effet sur le log des valeurs $x = 0$ que l'on observe pratiquement toujours. Il convient toutefois, dans chaque cas particulier, de s'assurer de la validité de cette transformation qui peut, si elle est appliquée sans discernement, conduire à des mécomptes comme nous le verrons ci-dessous.

F. J. ANSCOMBE (1949) a signalé la transformation logarithmique $y = \log(x + x_0)$ pour la loi binomiale négative. La constante x_0 est alors égale à $\frac{k}{2}$, elle peut être estimée par itération mais les calculs sont longs.

Nous avons choisi d'opérer par la méthode graphique : elle consiste à tracer des droites de Henri en faisant varier x_0 . Ceci permet de choisir la valeur de ce paramètre pour laquelle les points s'alignent au mieux. La figure 5 donne deux exemples d'évaluation de x_0 par cette méthode. Elle est assez empirique mais largement suffisante dans la pratique.

Il a pu être rapidement constaté que les points ne pouvaient pas s'aligner pour les populations à moyenne faible. Ceci revient à dire que les données issues de ces populations ne peuvent être normalisées, ce que nous avons déjà annoncé, et que le calcul de l'intervalle de confiance de la moyenne \bar{y} est impossible par normalisation. Ce cas est malheureusement fréquent. Le calcul reste toutefois possible, nous le verrons, lorsque les moyennes sont très faibles et que le modèle de Poisson peut être adopté.

Pour les populations où x_0 a pu être déterminé, nous avons trouvé des valeurs variant de 0,1 à 20.

Le calcul d'ajustement à une loi normale a été effectué pour les populations d'*Hirschmanniella spinicaudata* représentées aux figures 1 et 2, dont l'une s'ajustait à une loi binomiale négative et l'autre de façon douteuse. Dans les deux cas, les données transformées s'ajustaient, de façon satisfaisante, à une loi normale (tests de Kolmogoroff-Smirnov non significatifs à 5% : $|d_{\max}| = 0,047$ et $0,075 < 0,136$).

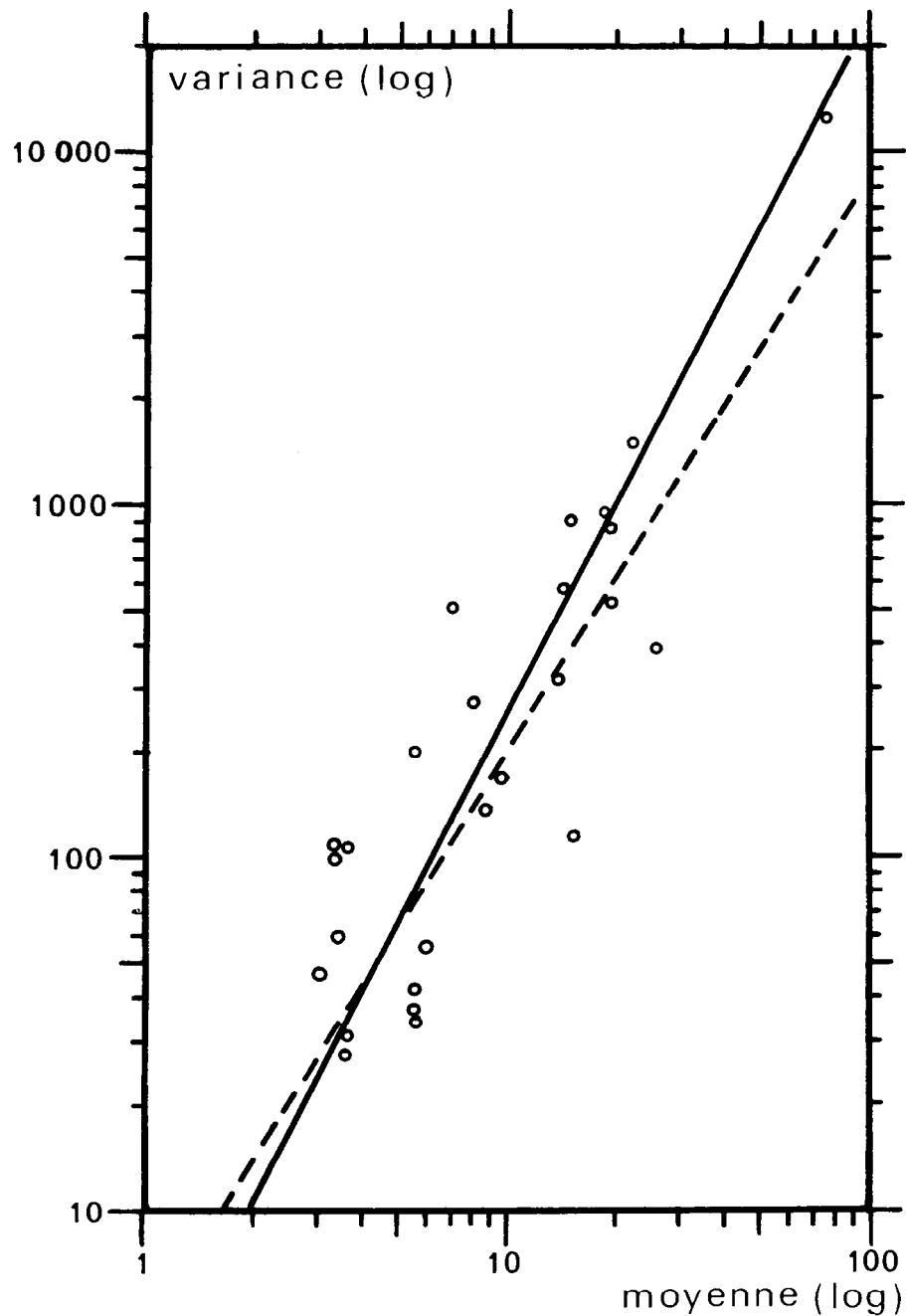


FIG. 4. — Reprise de la partie supérieure du graphique de la figure 3.

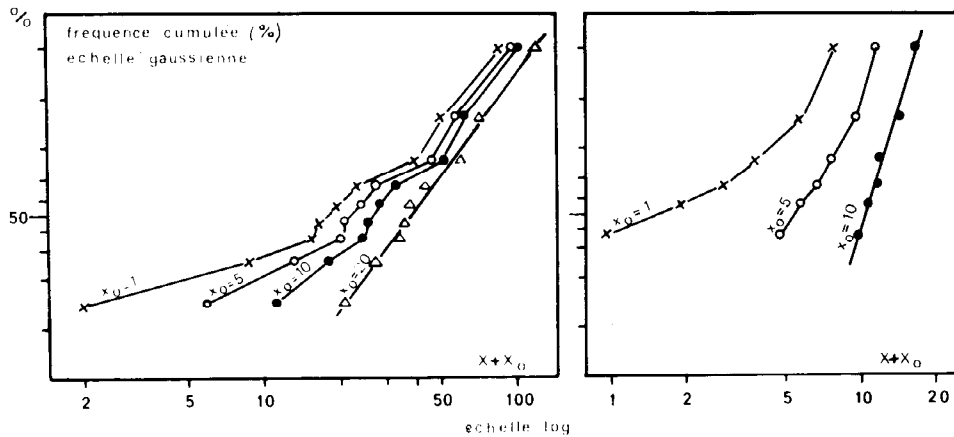


Fig. 5. — Deux cas de détermination graphique de x_0 .

ESTIMATION DE LA MOYENNE
 ET CALCUL DE SON INTERVALLE DE CONFIANCE
 A PARTIR DES DONNÉES TRANSFORMÉES EN LOG

La transformation $y = \log(x + x_0)$ permet de normaliser, ou à peu près, la distribution. Les calculs seront donc conduits sur la variable y et nous obtenons une moyenne \bar{y} , un carré moyen s_y^2 et deux limites de confiance :

$$l_i = \bar{y} - t \frac{s_y}{\sqrt{n}} \quad \text{et} \quad l_s = \bar{y} + t \frac{s_y}{\sqrt{n}}$$

pour la moyenne de y , t étant la variable de Student à $n - 1$ degrés de liberté au seuil choisi. La retransformation simple de la moyenne \bar{y} conduit à :

$$\bar{x}' = 10^{\bar{y}} - x_0$$

qui est une estimation biaisée de la moyenne de x . Il en est de même des retransformées des limites $b_i = 10^{l_i} - x_0$ et $b_s = 10^{l_s} - x_0$.

La présence de ce biais a notamment été signalée par QUENOUILLE (1953) qui a proposé une formule de correction. Le problème général de la correction du biais dû à la transformation a été étudié par NEYMAN & SCOTT (1960). Il est important lorsque l'obtention d'estimation sans biais de μ , et c'est le cas ici, est essentielle. SOUTHWOOD (1966) conseille d'ailleurs d'utiliser dans de tels cas la moyenne arithmétique des données non transformées mais il n'est pas alors possible d'obtenir des limites de confiance.

L'utilisation des formules fournies par NEYMAN & SCOTT permet l'obtention de limites de confiance et d'estimations sans biais : elles sont calculées en multipliant les retransformées simples par un coefficient de correction C dont l'expression est pour le cas qui nous intéresse :

$$C = \sum_{i=0}^{\infty} \frac{\left[\frac{\text{Ln } 10}{2} \cdot \frac{n-1}{n} \cdot s_y^2 \right]^i}{i! \prod_{k=1}^i \left(1 + 2 \frac{k-1}{n-1} \right)}$$

Le calcul de C, bien qu'un peu long, peut se faire assez aisément de la façon suivante :

1. Calculer :

$$A = \frac{2,65095 (n-1) S_{2y}}{n}$$

n est la taille de l'échantillon,

S_{2y} est la somme des carrés des écarts de y (numérateur de s_y^2).

$$(2,65095 = \frac{1}{2} (\text{Ln } 10)^2).$$

2. Calculer les termes successifs :

$$t_{i+1} = \frac{A}{(i+1)(2i+n-1)} t_i$$

pour $i = 0, 1, 2, \dots$ avec $t_0 = 1$ jusqu'à ce que le dernier terme calculé soit suffisamment petit.

3. Calculer la somme des t : c'est le coefficient C.

Citons, à titre d'exemple, un calcul qui a exigé neuf cycles. Il s'agit de *Hirschmaniella spinicaudata* à Nambingué dont la moyenne et la variance brute ont déjà été fournies ci-dessus (tableau I). La transformation utilisée a été $y = \log(x + 0,5)$. Les caractéristiques de la distribution de cette dernière variable sont :

$$n = 100 \quad \bar{y} = 0,74428 \quad s_y^2 = 43,24710/99 = 0,43684.$$

$$A = \frac{2,65095 \times 99 \times 43,24710}{100} = 113,49944$$

$$i = 0 \quad t_0 = 1,00000$$

$$t_1 = \frac{113,49944}{1 \times 99} \times 1,00000 = 1,14646$$

$$i = 1$$

$$t_2 = \frac{113,49944}{2 \times 101} \times 1,14646 = 0,64417$$

$$i = 2$$

$$t_3 = \frac{113,49944}{3 \times 103} \times 0,64417 = 0,23661$$

$$i = 3$$

$$t_4 = \frac{113,49944}{4 \times 105} \times 0,23661 = 0,06394$$

$$\begin{aligned}
 i = 4 & & t_5 &= \frac{113,49944}{5 \times 107} \times 0,06394 = 0,01356 \\
 i = 5 & & t_6 &= \frac{113,49944}{6 \times 109} = 0,01356 = 0,00235 \\
 i = 6 & & t_7 &= \frac{113,49944}{7 \times 111} \times 0,00235 = 0,00034 \\
 i = 7 & & t_8 &= \frac{113,49944}{8 \times 113} \times 0,00034 = 0,00004 \\
 i = 8 & & t_9 &= \frac{113,49944}{9 \times 115} \times 0,00004 = 0,00000
 \end{aligned}$$

$$C = 3,107$$

Il est possible de simplifier ce calcul en considérant comme égal à 1 le produit

$$\prod_{k=1}^i \left(1 + 2 \frac{k-1}{n-1} \right)$$

qui figure au dénominateur des termes de la somme fournissant le coefficient C. Il reste :

$$c = \sum_{i=0}^{\infty} \frac{\left[\frac{(\text{Ln } 10)^2}{2} \cdot \frac{n-1}{n} \cdot s_y^2 \right]^i}{i!}$$

qui n'est autre que :

$$10^{1,15129 \frac{n-1}{n} s_y^2} \quad (1,15129 = \frac{1}{2} \text{ Ln } 10)$$

C'est la formule de correction proposée par QUENOUILLE.

Le produit considéré comme égal à 1 est en fait un peu supérieur, il en résulte que c est supérieur au coefficient exact de correction C. Toutefois l'approximation sera d'autant meilleure que n sera élevé et (ou) s_y^2 faible. Appliquant cette dernière formule à l'exemple que nous venons de considérer nous trouvons :

$$c = \text{antilog} \left(1,15129 \times \frac{99}{100} \times 0,43684 \right) = \text{antilog} (0,4979) = 3,147$$

soit une différence de 0,040 en valeur absolue ou 1,3% en valeur relative. Il semble bien que dans un bon nombre de cas pratiques, le coefficient c soit suffisant, bien qu'il introduise un léger biais.

Il convient toutefois d'être prudent avec les échantillons de petite taille, surtout si la variance est grande, ou non normalisés par la transformation logarithmique. Un exemple va illustrer cette remarque. Les valeurs observées sont les suivantes, classées

par ordre de grandeur croissant :

0 — 0 — 2 — 5 — 14 — 22 — 85 — 97 — 182 — 347.

Les caractéristiques de cet échantillon sont : $n = 10$, $\bar{x} = 75,4$, $s_x^2 = 12.669$.

La transformation log paraît *a priori* appropriée. La figure 6 donne le polygone de fréquence cumulée avec $x_0 = 5$ qui est la valeur paraissant le mieux convenir. La courbe obtenue n'est pas une droite, mais présente une déformation systématique (sigmoïde) qui indique une distribution aplatie, voire même bimodale (mélange de deux populations): elle n'est pas log-normale. Si nous appliquons sans discernement les formules établies ci-dessus à la distribution de $y (= \log(x + 5))$ nous trouvons :

$$\begin{array}{lll} \bar{y} = 1,47344 & & s_y = 0,78845 \\ b_i = 1,891 & \bar{x}' = 24,74 & b_s = 123,4 \\ c = 6,560 & & C = 5,177 \\ Cb_i = 9,79 & C\bar{x}' = 128,09 & Cb_s = 638,8 \end{array}$$

La moyenne retransformée corrigée est bien supérieure à la moyenne arithmétique et la borne supérieure dépasse largement la plus grande observation.

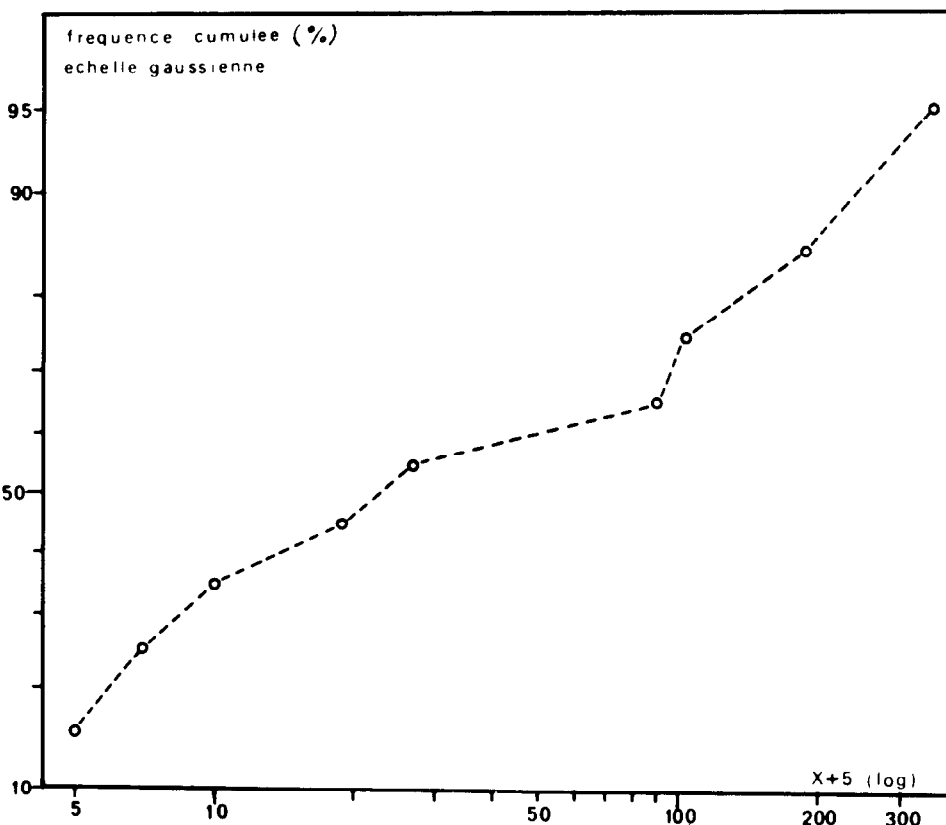


FIG. 6. — Polygone de fréquence cumulée d'un échantillon avec $x_0 = 5$. La courbe présente une déformation systématique (sigmoïde).

Il convient donc d'être prudent dans l'application de la méthode et de s'assurer que dans chaque cas particulier les conditions de validité sont remplies. De plus, lorsque

l'échantillon est de petite taille et (ou) que la variance est grande, le calcul du coefficient C s'impose. Une autre conclusion doit être tirée de cet exemple : les échantillons petits sont d'une précision très médiocre. Nous reviendrons ultérieurement sur ce point.

Appliquée à *Hirschmanniella spinicaudata* à Nambingué (échantillon A) et à Ferkéssédougou (échantillon B), la méthode fournit les résultats suivants :

$$(y = \log (x + 0,5))$$

Echan- tillon	\bar{x}	\bar{y}	\bar{x}'	b_1	b_s	C	Cb_1	$C\bar{x}'$	Cb_s
A	14,32	0,74428	5,05	3,59	7,02	3,107	11,15	15,69	21,81
B	27,13	1,27983	18,55	15,49	22,19	1,458	22,58	27,05	32,35

Si nous extrapolons les trois dernières colonnes au nombre d'individus par litre de sol, nous obtenons :

Echan- tillon	Limite inférieure	Moyenne	Limite supérieure
A	446	628	872
B	903	1 082	1 294

La méthode qui consiste à étudier un échantillon de 100 prélèvements dans une rizière d'environ un hectare permet donc d'avoir une évaluation satisfaisante des populations de nématodes qui s'y trouvent lorsque les données sont normalisables par transformation logarithmique, autrement dit que la moyenne est suffisamment élevée, c'est-à-dire lorsque les populations ont une certaine importance.

ESTIMATION DE LA MOYENNE ET CALCUL DE SON INTERVALLE DE CONFIANCE POUR DES DISTRIBUTIONS A FAIBLE MOYENNE

Pour les distributions à moyenne faible, deux cas doivent être envisagés : celui où la distribution est assimilable à une loi de Poisson et celui où elle ne l'est pas. Dans le premier cas, c'est-à-dire celui des distributions pour lesquelles le modèle de Poisson peut être adopté, le problème de la détermination d'un intervalle de confiance pour la moyenne est résolu. La solution est exposée dans différents ouvrages de statistique, en particulier HALD (1952) : elle utilise la relation qui existe entre la loi de Poisson et la loi de χ^2 . Les formules à utiliser sont les suivantes, pour un intervalle de confiance à $1 - \alpha$ symétrique ($\alpha/2$ de part et d'autre) :

$$\text{limite supérieure : } \frac{1}{2} \chi^2_{1 - \frac{\alpha}{2}} \quad \nu = 2(x + 1)$$

$$\text{limite inférieure : } \frac{1}{2} \chi^2_{\frac{\alpha}{2}} \quad \nu = 2x$$

x étant le nombre observé.

Prenons pour illustrer cette méthode un petit échantillon dont la distribution de fréquence est :

x	Effectif
0	7
1	2
2	1

$$\begin{aligned} \bar{x} &= 0,4 \\ s_x^2 &= 0,489 \\ n &= 10 \end{aligned}$$

Le modèle de POISSON convient.

Nous avons ici 10 observations extraites de la même loi de POISSON de paramètre μ . Utilisant le théorème selon lequel la somme de n variables de POISSON indépendantes de paramètres $\mu_1, \mu_2, \dots, \mu_n$ est une variable de POISSON de paramètre $\sum_{i=1}^n \mu_i$, nous allons chercher des limites pour le paramètre $\mu' = n\mu$ en partant du total observé $s_1 = \sum n_i x_i$, puis en déduire des limites pour μ .

Dans notre exemple, $n = 10$, $s_1 = 4$, nous prendrons $\alpha = 5\%$.

— Limite pour $\mu' = 10\mu$:

$$\text{supérieure : } \frac{1}{2} \chi^2_{0,975} \quad \nu = 2(4 + 1) = 10$$

$$\text{soit } \frac{1}{2} 20,5 = 10,25$$

$$\text{inférieure : } \frac{1}{2} \chi^2_{0,025} \quad \nu = 2 \times 4 = 8$$

$$\text{soit } \frac{1}{2} 2,18 = 1,09$$

— Limite pour μ :

$$\text{inférieure : } \frac{1,09}{10} = 0,11$$

$$\text{supérieure : } \frac{10,25}{10} = 1,02.$$

Pour les autres cas, moyennes faibles ou peu élevées et modèle de Poisson inacceptable, nous n'avons pas trouvé de méthode fournissant un intervalle de confiance à la moyenne. De toute façon, il s'agit toujours, même dans le cas du modèle de Poisson, de faibles populations et la fréquence des prélèvements sans nématodes paraît avoir autant d'intérêt pratique que le nombre moyen de nématodes par prélèvement, cette fréquence est en effet à peu près proportionnelle à la moyenne dans de tels cas, tout au moins pour des lois de Poisson et binomiale négative.

TAILLE D'ÉCHANTILLON NÉCESSAIRE
A L'OBTENTION D'UNE PRÉCISION DONNÉE

Nous avons vu ci-dessus que le prélèvement d'échantillons de taille $n = 100$ était satisfaisant dans le cas de normalisation possible par transformation log. Toutefois, le prélèvement et l'étude de 100 unités-échantillons prennent un temps considérable et ne peuvent être adoptés comme pratique courante. Il faut les réserver aux cas où l'on a des raisons particulières d'évaluer la population aussi exactement que possible.

TABLEAU III
Taille n de l'échantillon

Précision p	Moyenne \bar{x}'	Constante x_0			
		0,1	1	10	20
0,20	1	649	2 128	64 160	233 821
	10	548	649	2 128	4 779
	20	542	592	1 200	2 128
0,30	1	292	950	28 519	103 924
	10	247	292	950	2 128
	20	245	267	537	950
0,40	1	167	537	16 045	58 460
	10	142	167	537	1 200
	20	141	153	305	537

REM. Le tableau est donné pour la valeur $s_y^2 = 1$ de la variance de la variable transformée.

Les valeurs de n étant proportionnelles à s_y^2 , il suffit de multiplier la valeur tabulaire par s_y^2 pour obtenir la taille d'échantillon recherché.

Les valeurs les plus fréquemment rencontrées pour s_y^2 se situent entre 0,25 et 0,50.

Nous sommes ainsi tout naturellement amenés à nous demander quelle est la taille n nécessaire à l'obtention d'une précision donnée ou, inversement, quelle précision attendre avec un échantillon de taille fixée.

La précision (p) peut se définir comme une fraction de la moyenne qui, appliquée à celle-ci en plus et en moins, établira les limites de son intervalle de confiance. Autrement dit, la moyenne sera estimée à 100 $p\%$ près.

TABLEAU IV
Valeurs de la précision p attendue

Taille d'échantillon n	Moyenne \bar{x}'	Variance (1) s_y^2	Constante x_0			
			0,1	1	10	20
10	1	0,25	1,01	1,84	10,12	19,32
		0,50	1,59	2,89	15,91	30,37
	10	0,25	0,93	1,01	1,84	2,76
		0,50	1,46	1,59	2,89	4,34
	20	0,25	0,92	0,97	1,38	1,84
		0,50	1,45	1,52	2,17	2,89
30	1	0,25	0,49	0,89	4,87	9,31
		0,50	0,71	1,62	8,93	17,06
	10	0,25	0,45	0,49	0,89	1,33
		0,50	0,65	0,89	1,62	2,44
	20	0,25	0,44	0,46	0,66	0,89
		0,50	0,65	0,85	1,22	1,62
100	1	0,25	0,25	0,46	2,53	4,84
		0,50	0,36	0,66	3,61	6,90
	10	0,25	0,23	0,25	0,46	0,69
		0,50	0,33	0,36	0,66	0,99
	20	0,25	0,23	0,24	0,35	0,46
		0,50	0,33	0,34	0,49	0,66

(1) Variance de la variable transformée.

En écrivant que la différence entre Cb_s et Cb_l vaut $2p \mu$ et en estimant μ par $C\bar{x}'$ nous arrivons à :

$$n = \frac{t^2 s_y^2}{\left\{ \log \left[\frac{p \bar{x}'}{\bar{x}' + x_0} + \sqrt{1 + \left(\frac{p \bar{x}'}{\bar{x}' + x_0} \right)^2} \right] \right\}^2}$$

(t étant la variable de Student à $n - 1$ d.d.l.), qui fournit n en fonction des caractéristiques et de la précision que l'on veut,

ou à :

$$p = \frac{1}{2} \left(1 + \frac{x_0}{\bar{x}'} \right) \left[\text{antilog} \left(\frac{ts_y}{\sqrt{n}} \right) - \frac{1}{\text{antilog} \left(\frac{ts_y}{\sqrt{n}} \right)} \right]$$

qui fournit la précision que l'on peut espérer en fonction de n et des caractéristiques.

Il est bien évident que ces formules ne sont pas totalement rigoureuses, puisque pour les écrire nous avons utilisé des estimations mais, dans la pratique, elles sont suffisantes pour fournir une indication valable sur la taille d'échantillon à prélever ou sur la précision à attendre.

Les tableaux III et IV fournissent quelques résultats : taille d'échantillon (tableau III) ou précision attendue (tableau IV) qui peuvent servir d'indication à la conduite ultérieure du travail. Le niveau de confiance est dans tous les cas égal à 95%.

POPULATIONS OBSERVÉES

Expérimentalement, la mauvaise précision obtenue avec des échantillons de petite taille a pu être constatée : pendant la campagne rizicole de 1965, avant que les calculs ci-dessus aient été effectués, nous avons, arbitrairement, fixé à dix le nombre de prélèvements par rizière. Après détermination de x_0 par la méthode graphique, chaque fois qu'elle s'est avérée possible, les intervalles de confiance ont été calculés pour un certain nombre de populations. Le tableau V donne le détail de quelques résultats ainsi obtenus. Il est évident que la méthode consistant à opérer dix prélèvements par rizière, si elle peut être considérée comme modérément satisfaisante pour évaluer des populations relativement importantes (1 000 individus au litre de sol, et plus) devient rapidement beaucoup trop imprécise quand les populations diminuent, l'erreur relative croissant alors très vite.

Si, dans la plupart des cas, il est impossible d'avancer avec une précision satisfaisante une valeur de la population moyenne, l'étude statistique de dix unités d'échantillonnage permet cependant de classer les espèces observées dans une échelle de valeurs. Nous définissons ses divers degrés en nous basant sur le fait que la détermination de x_0 et, partant, la normalisation ont été ou non possibles et sur la valeur de la limite inférieure de l'intervalle de confiance exprimée en nombre d'individus au litre de sol.

TABLEAU V

Estimation de la moyenne et de son intervalle de confiance dans des échantillons de 10 unités

Espèce	n	x _o	Cb ₁	C \bar{x} '	Cb _s	Au litre de sol		
						Limite inférieure	Moyenne	Limite supérieure
<i>Hirschmanniella spinicaudata</i>	10	20	19,62	30,09	43,28	785	1 204	1 731
»	10	20	18,03	29,27	43,75	721	1 171	1 750
»	10	0,1	6,50	16,90	43,40	260	676	1 736
»	10	10	0,40	2,94	6,07	16	118	243
»	10	1	0,53	2,29	5,61	21	92	224
<i>Trichotylenchus rhopalocercus</i>	10	10	0,67	3,10	5,97	27	124	239
»	10	2	1,08	4,44	10,73	43	178	429
<i>Xiphinema</i> sp. « ae »	10	1	1,27	3,01	5,95	51	120	238
<i>Criconemoides palustris</i>	10	10	2,42	5,68	9,76	92	227	390
»	10	10	4,76	5,49	6,25	190	220	250
<i>Criconemoides curvatus</i>	10	10	3,63	17,31	41,96	145	692	1 678
<i>Helicotylenchus dihystra</i>	10	10	5,97	16,77	34,02	239	791	1 361
» sp.	10	10	1,19	7,15	15,93	48	286	637

La définition des classes d'abondance qui composeront une telle échelle comporte inévitablement une certaine part d'arbitraire. Cependant, nous basant sur les études effectuées sur un sujet semblable par FRONTIER (1969), nous nous sommes fixé de définir ces classes en suivant une progression géométrique de raison 4.

Les classes d'abondance ont donc été définies de la façon suivante :

- Très faible : Normalisation impossible, trop grande variabilité ou valeurs $x = 0$ en nombre ≥ 4 .
- Faible : Limite inférieure de l'intervalle de confiance < 25 .
- Moyenne : Limite inférieure comprise entre 25 et 99.
- Forte : Limite inférieure comprise entre 100 et 399.
- Très forte : Limite inférieure ≥ 400 .

Ainsi définie, cette échelle rend compte de valeurs minimales, une population n'étant classée à un niveau donné que si l'on a la certitude qu'elle atteigne au moins une certaine valeur. Si elle ne peut fournir une idée précise de l'importance de la population d'une espèce en un point donné, elle doit permettre, cependant, de comparer plusieurs espèces dont les populations ont été étudiées en un certain nombre de points (tableau VI).

TABLEAU VI

Nombre de points où une espèce a été trouvée à un certain niveau de population

Espèces	Populations				
	Très faibles	Faibles	Moyennes	Fortes	Très fortes
<i>Hirschmanniella spinicaudata</i>	4	3		2	3
<i>Trichotylenchus rhopalocercus</i>	3		3	2 *	
<i>Tylenchorhynchus elegans</i> (en mélange avec <i>T. palustris</i>)					1
<i>Aphelenchus avenae</i>	6		1		
<i>Xiphinema</i> sp. « ae »	10	1	2		
<i>Xiphinema rotundatum</i>	3				
<i>Criconemoides palustris</i>	6		1	1	
<i>Criconemoides curvatus</i>	1			1	
<i>Helicotylenchus flatus</i>	2	1		1	
<i>Helicotylenchus dihystra</i>	1				
<i>Hemicriconemoides cocophilus</i>	1				
<i>Rotylenchulus</i> sp.	1				
<i>Paratylenchus aquaticus</i>	3				

* Dont une en mélange avec *T. elegans*.

Les populations exophytes d'*Hirschmanniella spinicaudata* sont très variables, souvent très faibles ou faibles, elles peuvent atteindre de très fortes densités.

La même variabilité se rencontre chez *Trichotylenchus rhopalocercus*, qu'on trouve parfois en populations très faibles mais, plus souvent, moyennes ou fortes. Il s'agit d'ailleurs de l'une des espèces les plus fréquentes dans les rizières et son parasitisme à l'égard du riz a été démontré expérimentalement.

Tylenchorhynchus elegans et *T. palustris* ont été rencontrés en mélange, en population très forte, ce qui rend probable leur liaison parasitaire avec le riz.

Aphelenchus avenae et *Xiphinema* sp « ae », bien que fréquents se présentent, en général, sous forme de populations très faibles, rarement faibles ou moyennes et leur liaison parasitaire avec le riz reste douteuse.

C'est le cas, également, pour *Criconemoides palustris* et *C. curvalus* dont la fréquence est, de plus, sensiblement inférieure à celle des deux espèces précédentes.

Ces quatre dernières espèces, si elles sont assez souvent présentes dans les rizières, n'y rencontrent pas les conditions optimales pour le développement de populations importantes.

Helicotylenchus flatus a été rencontré, dans un cas, sous la forme d'une forte population et il est, de ce fait, probable que cette espèce soit un parasite actif du riz.

La seule population d'*Helicotylenchus dihystra* dont l'évaluation ait pu être tentée était très faible. Son parasitisme reste donc douteux.

Enfin, *Xiphinema rotundatum*, *Hemicriconemoides cocophilus*, *Rotylenchulus* sp. et *Paratylenchus aquaticus* n'ont été rencontrés qu'en populations très faibles et, s'il reste possible que, dans certains cas, l'une ou l'autre de ces espèces soit parasite du riz, la rizière ne constitue pas leur biotope de prédilection et leur présence peut y être considérée, la plupart du temps, comme accidentelle.

CONCLUSION

Contrairement à ce à quoi on aurait pu s'attendre, dans un milieu apparemment aussi homogène qu'une rizière de dimensions relativement faibles, les estimations des populations de nématodes phytoparasites varient beaucoup d'une unité-échantillon à l'autre, ce qui nous amène à avancer l'hypothèse d'une distribution agrégée des animaux.

Ce n'est qu'avec la plus grande prudence qu'on peut avancer un nombre exprimant la taille d'une population, la précision de ce nombre variant grandement avec l'importance de celle-ci. La plupart des distributions observées appartenant à la loi binomiale négative ou en étant voisine, le calcul de l'intervalle de confiance de la moyenne observée ne peut être effectué qu'après transformation logarithmique des données, qui normalise suffisamment les distributions. Le biais introduit par la retransformation de la moyenne et des limites de l'intervalle de confiance peut être pallié en leur appliquant un coefficient dont le calcul est relativement aisé.

Pour des populations importantes, plus de 1 000 individus au litre de sol, l'examen de dix unités d'échantillonnage suffit à estimer valablement une moyenne. Pour des populations de moyenne importance, l'étude d'un plus grand nombre d'unités d'échantillonnage devient nécessaire, ce nombre pouvant aller jusqu'à 100.

Enfin, pour des populations faibles, la normalisation des données est impossible et l'estimation de la moyenne illusoire. La proportion d'échantillons ne contenant aucun individu constitue alors le meilleur critère.

En résumé, si au cours d'une enquête extensive, on étudie dix unités d'échantillonnage par hectare, ce qui représente déjà un gros travail, il ne sera possible d'estimer une moyenne que pour les fortes populations. Pour les autres, on pourra tout juste affirmer qu'elles sont moyennes, faibles ou très faibles, encore cette affirmation sera-t-elle entachée de subjectivité. Il sera nécessaire d'étudier une centaine d'unités quand on voudra estimer une population moyenne avec une certaine exactitude, par exemple pour mettre en train un essai biologique ou pour effectuer une étude écologique suivie.

BIBLIOGRAPHIE

- ANDREWARTHA (H. G.) — 1961 — *Introduction to the study of animal population*. Methuen & Co Ltd, London. xvii + 281 p.
- ANSCOMBE (F. J.) — 1949 — The statistical analysis of insect counts based on the negative binomial distribution. *Biometrics* 5, 165-173.
- FRONTIER (S.) — 1969 — Sur une méthode d'analyse faunistique rapide du zooplancton. *J. exp. mar. Biol. Ecol.* 3, 18-26.
- GOFFART (H.) — 1962 — Sampling and extraction methods for cyst forming nematodes of the genus *Heterodera*. In : « *Progress in soil Zoology* ». Ed. Murphy P. W. Butterworths London, 268-278.
- HALD (A.) — 1952 — *Statistical theory with engineering applications*. New York. London. J. Wiley & Sons. xii + 783 p.
- HEALY (M. J. R.), TAYLOR (R. L.) — 1962 — Tables for power-law transformations. *Biometrika* 49, 557-559.
- JONES (F. G. W.) — 1955 — Quantitative methods for the estimation of cyst-forming nematodes (*Heterodera* spp.) in soil. In « *Soil Zoology* ». Ed. D.K.McE. Kevan. London Butterworths, 394-402.
- MERNY (G.) — 1970 — Les nématodes phytoparasites des rizières inondées de Côte d'Ivoire. I. Les espèces présentes. *Cahiers O.R.S.T.O.M.*, sér. Biol., 11, 3-43.
- NEYMAN (J.), SCOTT (E.) — 1960 — Correction for bias introduced by a transformation of variables. *Ann. Math. Stat.* 31, 643-655.
- OOSTENBRINK (M.) — 1950 — Het aardappelaaltje (*H. rostochiensis* Wollenweber) een gevaarlijke parasiet voor de eenzijdige aardappelcultuur *Versl. PIZiekt. Dienst Wageningen* 115, 230 p.
- QUENOUILLE (M. H.) — 1953 — *The design and analysis of experiment*. London. Ch. Griffin. xiii + 356 p.
- SOUTHWOOD (T. R. E.) — 1966 — *Ecological Methods*. London. Methuen & Co. xviii + 391 p.
- TAYLOR (R. L.) — 1961 — Aggregation, variance and the mean. *Nature*, London, 189, 732-735.
- WIDDOWSON (E.) — 1962 — The estimation of soil populations of *Heterodera rostochiensis* Woll. In : « *Progress in Soil Zoology* », ed. by P. W. Murphy. London, Butterworths, 59-64.