

## 2

### ÉCHANTILLONNAGE DE VARIABLES BACTÉRIOLOGIQUES DANS LES MILIEUX AQUATIQUES

M. TROUSSELLIER\*, B. BALEUX\* ET P. ANDRÉ\*\*

\* Laboratoire d'Hydrobiologie, U.A.CNRS 694, Université des Sciences et Techniques du Languedoc, 34060 MONTPELLIER (FRANCE)

\*\* Ecole Polytechnique, CP 6079, Succursale A, MONTREAL, H3C 3A7 (CANADA)

**RÉSUMÉ** - A partir d'échantillons d'eau provenant de différents niveaux d'observation d'une lagune saumâtre, des solutions statistiques sont proposées pour l'optimisation de l'échantillonnage des abondances en bactéries revivifiables. La variabilité entre les dénombrements est étudiée au moyen de la loi des puissances de Taylor. L'optimisation des plans d'échantillonnage comprenant plusieurs sources de variabilité est réalisée par l'analyse de variance hiérarchisée et par l'estimation des coûts d'échantillonnage de chacune des sources. Cette procédure d'optimisation amène à un accroissement important de l'efficacité relative du plan d'échantillonnage et permet de distinguer des différences significatives entre les concentrations bactériennes. L'effet de la réduction de l'effort d'échantillonnage total sur la précision de ces concentrations moyennes met en évidence l'intérêt scientifique et économique d'un choix *a priori* du niveau de précision approprié aux études.

*Mots clés* : bactéries, écosystèmes aquatiques, stratégies d'échantillonnage, analyse de variance, précision/coût.

**ABSTRACT** - Using water samples from several observation levels of a brackish lagoon, statistical solutions are offered for the sampling optimisation of bacteria viable counts. The variability between plates count is studied by the mean of the Taylor's power law. The sampling strategy optimisation, including several sources of variability, is realized by the way of hierarchical analysis of variance and the estimation of the sampling cost of each source. This optimisation procedure result in a great increase of the relative efficiency of the sampling design and allow to obtain significant differences between the bacterial concentrations. The effect of the reduction of the total sampling effort on the precision of these average concentrations pointed out the scientific and economic interest to choose *a priori* the suitable level of precision to the studies.

*Key words* : bacteria, aquatic ecosystems, sampling strategy, analysis of variance, precision/cost.

### INTRODUCTION

Souvent ignorées, quelquefois évoquées, rarement prises en compte, les questions d'échantillonnage sont cependant d'une grande importance lors de l'étude des variables bactériologiques dans les milieux aquatiques (Maul *et al.*, Hugues *et al.*, 1983).

La plupart des hypothèses de travail en écologie bactérienne sont basées sur le recueil, la description et l'analyse de paramètres bactériologiques pour une échelle d'espace et/ou de temps caractéristique de la problématique initiale. Ces objectifs ne peuvent être valablement atteints que dans la mesure où la variabilité étrangère au niveau d'observation choisi est prise en compte et/ou maîtrisée.

Le premier problème, commun à toutes les disciplines biologiques, mais très amplifié en bactériologie, est l'étendue et la diversité des échelles d'observations existant dans les milieux aquatiques. Pour relativiser notre perception du "monde bactérien" un parallé-

lisme a été tenté avec le "monde humain". (fig. 1). Par exemple, à l'échelle des longueurs, vouloir échantillonner des bactéries sur une distance de 1 m correspond au fait d'échantillonner des hommes sur la distance routière existant entre Montpellier et Brest. A l'échelle des surfaces, l'échantillonnage de bactéries sur une surface de 1 m<sup>2</sup> est équivalent à celui d'êtres humains sur le territoire de l'Egypte. Enfin à l'échelle des volumes, l'éloignement du "monde bactérien" s'accroît encore, puisque l'observation de bactéries dans 1 m<sup>3</sup> d'eau serait comparable à celle d'individus humains, s'ils s'en trouvaient, dans le volume des océans et mers réunis.

Dans ce travail, n'est abordée que l'échelle des surfaces, car la plupart de nos travaux se situent en zone côtière peu profonde et en milieu lagunaire, où il est possible de faire abstraction de la troisième dimension, c'est-à-dire la profondeur. La position approximative de l'exemple étudié pour illustrer certains aspects de l'optimisation d'un plan d'échantillonnage de variables bactériologiques est reportée sur l'échelle des surfaces (fig. 1). L'optimisation dont les résultats sont présentés dans cette publication sera recherchée en référence au critère de précision statistique de la réponse au problème posé mais aussi et simultanément aux contraintes liées au coût ou à l'effort d'échantillonnage disponible.

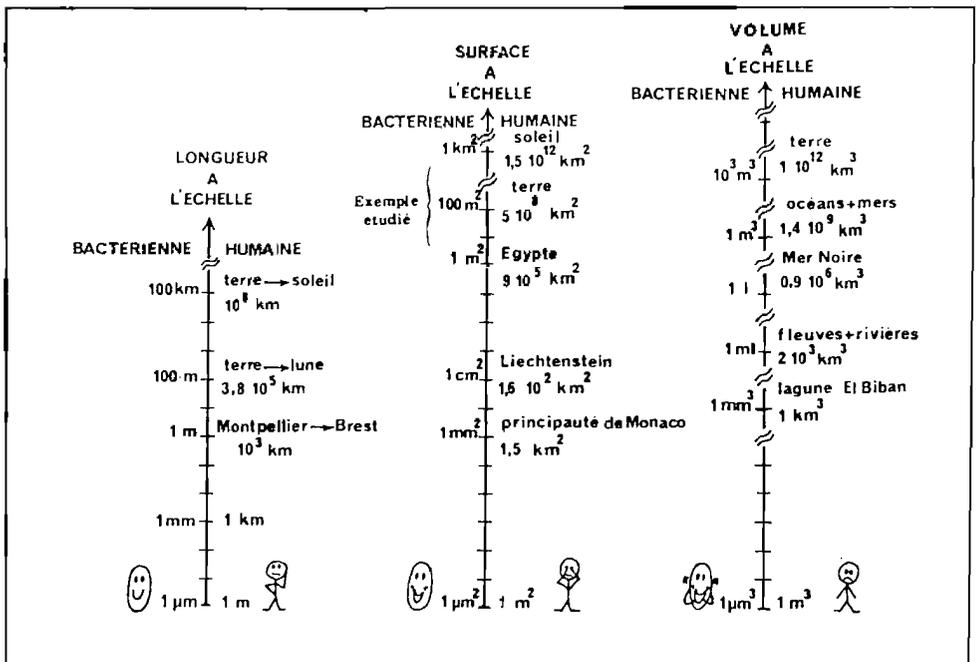


Figure 1 : Echelles comparées des longueurs, surfaces et volumes du "monde bactérien" et du "monde humain".

## MATÉRIEL ET METHODES

La baie qui constitue le niveau d'observation supérieur de cette étude est située sur la frange nord du bassin de Thau qui est un étang de la façade méditerranéenne française. Elle est soumise entre autre, au rejet d'eaux usées urbaines traitées par bassin de lagunage dont on trouvera une description dans un article précédent (Baleux et Trousselier, 1983).

Quotidiennement y est ainsi déversé un volume d'eau d'environ 1000 m<sup>3</sup>, constituant un apport important de matières inertes et vivantes (par exemple plusieurs centaines de kilos de phyto et zooplancton et une dizaine de kilos d'ammoniaque par jour).

L'impact de ces apports sur le milieu aquatique que figure cette baie, est un sujet de nos recherches, tant du point de vue des modifications des communautés bactériennes autochtones que du point de vue du devenir des bactéries d'origine continentale qui y parviennent. Dans cette optique, qui inclut à court terme le suivi de populations bactériennes et à moyen terme la description des communautés, la stratégie d'échantillonnage concerne avant tout les variables du type: abondance en bactéries revivifiables sur différents milieux de culture dont Marine agar 2216 Difco (Ma), Nutrient agar bioMérieux (Na), Milieu Eimhjellen (Em) (Eimhjellen, 1965), gélose lactosée au T.T.C. et Tergitol (CF) et D-Coccosel agar (DF). Cependant la procédure d'optimisation qui sera utilisée pourrait s'appliquer à tout autre paramètre bactériologique tel que les comptages en épifluorescence ou en immunofluorescence.

Comme pour évaluer les effets de ces apports continentaux il est indispensable d'observer le milieu récepteur pendant plusieurs années, il est *a priori* rentable de réaliser une première série de campagnes de pré-échantillonnage afin de posséder les éléments de choix d'une stratégie de suivi spatio-temporel.

Le schéma initial retenu pour chacune des campagnes de pré-échantillonnage est composé de 5 stations localisées chacune au hasard à l'intérieur de 5 zones de la baie de dimensions approximativement équivalentes définies par rapport à leur éloignement du rejet des bassins de lagunage (niveau supérieur d'observation = 1<sup>er</sup> niveau de variabilité = variabilité à l'échelle de la baie ou inter stations), de 5 prélèvements par station (2<sup>e</sup> niveau de variabilité = variabilité inter-prélèvements) et de 2 à 5 mesures par prélèvement (3<sup>e</sup> niveau de variabilité = variabilité inter-mesures), après dilution ou filtration suivant le type de bactéries recherchées et leur concentration (fig. 2).

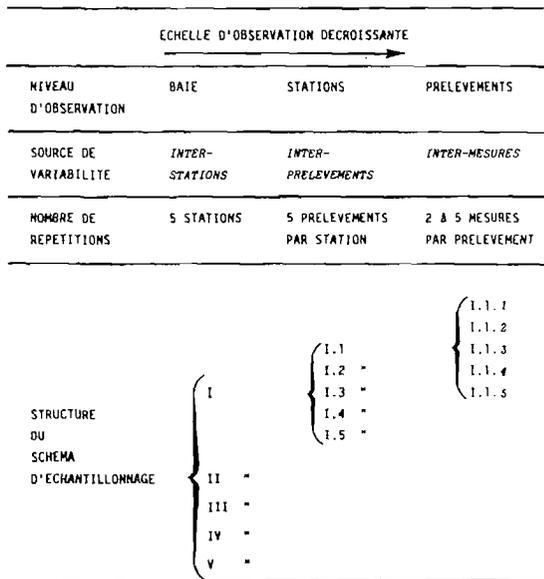


Figure 2 : Schéma d'une campagne de pré-échantillonnage.

Mais quels que soient les exemples choisis, il existe une source de variabilité commune à tous les plans d'échantillonnage : la variabilité inter-mesures due surtout à l'hétérogénéité du prélèvement à l'échelle des bactéries et, en partie, aux manipulations, dont l'appréciation constituera la première partie de cet article avant de considérer simultanément les 3 niveaux de variabilité observés.

## RESULTATS ET DISCUSSION

### *Variabilité inter-mesures ou variabilité résiduelle*

Sans vouloir détailler les différentes sources de cette variabilité inhérentes à la méthode de dénombrement, dont on trouvera une bonne appréciation dans les travaux de Kaper *et al.* (1978), nous avons étudié les valeurs absolues et relatives de la variabilité liée à la répétition des mesures d'abondance des bactéries des eaux de l'étang de Thau sur différents milieux de culture. Ces valeurs sont illustrées sur la figure 3 a et b. La distribution des valeurs absolues de variance inter-mesures après transformation logarithmique des données brutes est représentée sur la figure 3a. Ces valeurs, comme la plupart de celles rapportées dans la littérature sont inférieures à  $0.1 \log_{10}$ . Elles correspondent à des pourcentages de la variance totale en général faibles (fig. 3b) mais dont l'intensité doit être évaluée avec prudence. C'est ce qu'illustre la dernière valeur reportée sur cette figure et qui concerne la variance des coliformes fécaux recensés non plus dans un étang mais à titre de comparaison dans une rivière (Ardèche). La variabilité inter-mesures relative est très forte. Mais comme sa valeur absolue est très faible cette variance relative n'est pas le signe que la méthode de recensement deviendrait soudainement imprécise. Elle est due à une très faible variabilité de la concentration bactérienne aux niveaux d'observation supérieurs (rivière à fort débit, très homogène). Bien que la variance inter-mesures n'intervienne que faiblement dans la variance totale, elle peut cependant être étudiée dans le but de chercher les moyens éventuels de la réduire et donc d'augmenter la précision des mesures.

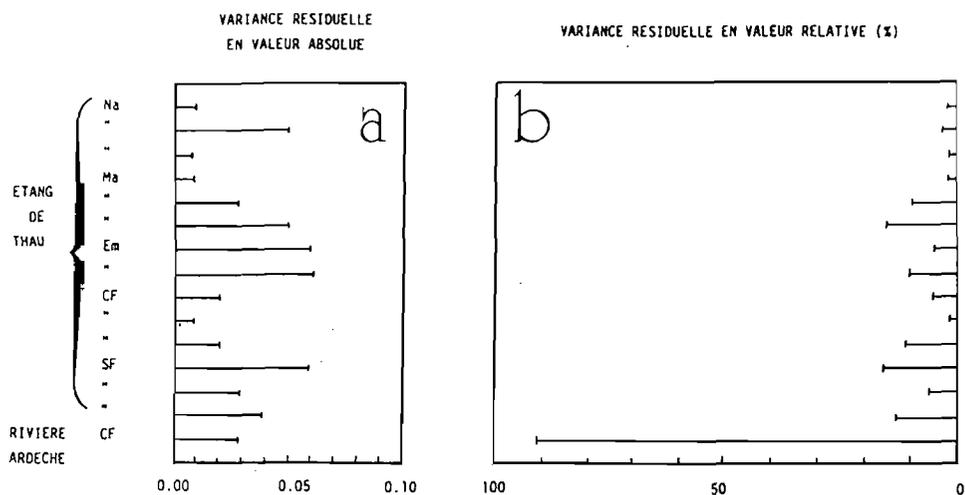


Figure 3 : Variations résiduelles en valeur absolue (a) et relative (b) (% de la variance totale) obtenues sur les dénombrements de différents types de bactéries au cours de trois campagnes de pré-échantillonnage dans l'étang de Thau et pour un transect horizontal dans la rivière Ardèche.

L'examen des variances  $S^2_x$  et des moyennes  $\bar{x}$  des séries de dénombrements des bactéries hétérotrophes revivifiables montre que très fréquemment les mesures suivent une distribution contagieuse ( $S^2_x \gg \bar{x}$ ). Dès lors l'utilisation des formules de précision issues d'une loi de Poisson ou d'une loi normale, auxquelles se réfèrent la plupart des auteurs, n'est pas justifiée. Par ailleurs il est difficile et parfois impossible d'ajuster des lois théoriques connues aux distributions observées et ces ajustements n'aboutissent pas toujours à une solution unique. (Baleux et Troussellier, 1982; Esterby, 1982). Par contre, la relation entre la variance et la moyenne de séries de mesures peut être facilement analysée grâce à la loi des puissances de Taylor (Taylor, 1961).

Traduite en échelle logarithmique, cette relation permet d'apprécier rapidement et facilement :

- 1) le type de distribution des abondances : hasard, régulier, contagieux ;
- 2) la mesure du degré de contagion, ainsi que le type de transformation à utiliser pour stabiliser la relation variance-moyenne par le calcul de la pente de la relation. Cette transformation aboutit très souvent à une normalisation des données brutes ;
- 3) le degré de précision en fonction de la concentration moyenne en bactéries  $\bar{x}$  et du nombre de mesures (n).

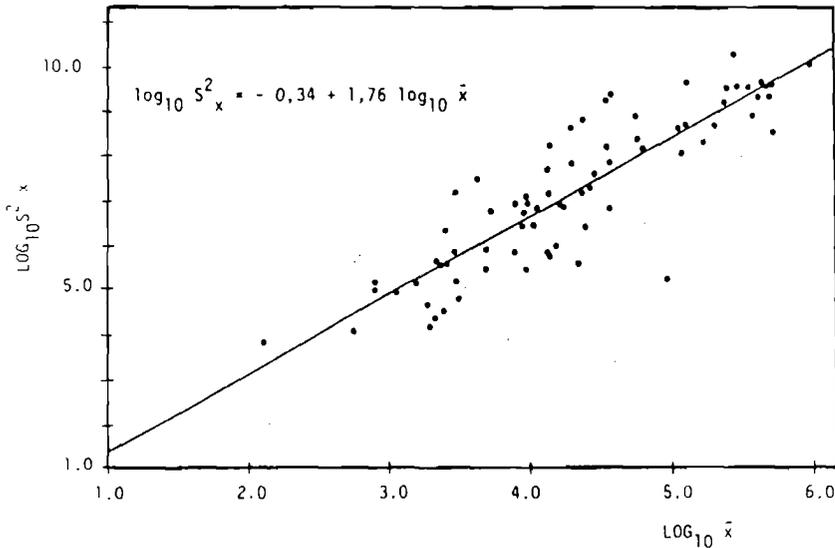


Figure 4 : Relation entre la variance ( $S^2_x$ ) et la moyenne ( $\bar{x}$ ) de séries de dénombrements de bactéries cultivant sur Marine agar, obtenues à partir de 75 prélèvements d'eau de l'étang de Thau.

A titre d'exemple la relation qui a été obtenue pour les dénombrements de bactéries hétérotrophes revivifiables sur Marine agar est illustrée sur la figure 4. Cette relation a été calculée à partir de 75 prélèvements sur chacun desquels ont été réalisés 5 mesures. Sans entrer dans le détail des formules statistiques nécessaires à leur calcul, rapportées dans un article précédent (Troussellier *et al.*, 1984), les abaques de précision déduits de cette relation sont présentés sur la figure 5. C'est ainsi que pour obtenir une précision de 10 % autour de la concentration moyenne en bactéries, il faut réaliser environ 15, 9, 5 ou 3 dénombrements si respectivement le nombre moyen de bactéries est d'environ  $10^2$ ,  $10^3$ ,  $10^4$ , ou  $10^5$  par ml. Ainsi, en pratiquant 2 à 3 mesures par prélèvement, il faut avoir à

l'esprit que la précision théorique sera élevée pour une forte concentration bactérienne, mais plus faible si la concentration bactérienne est peu élevée.

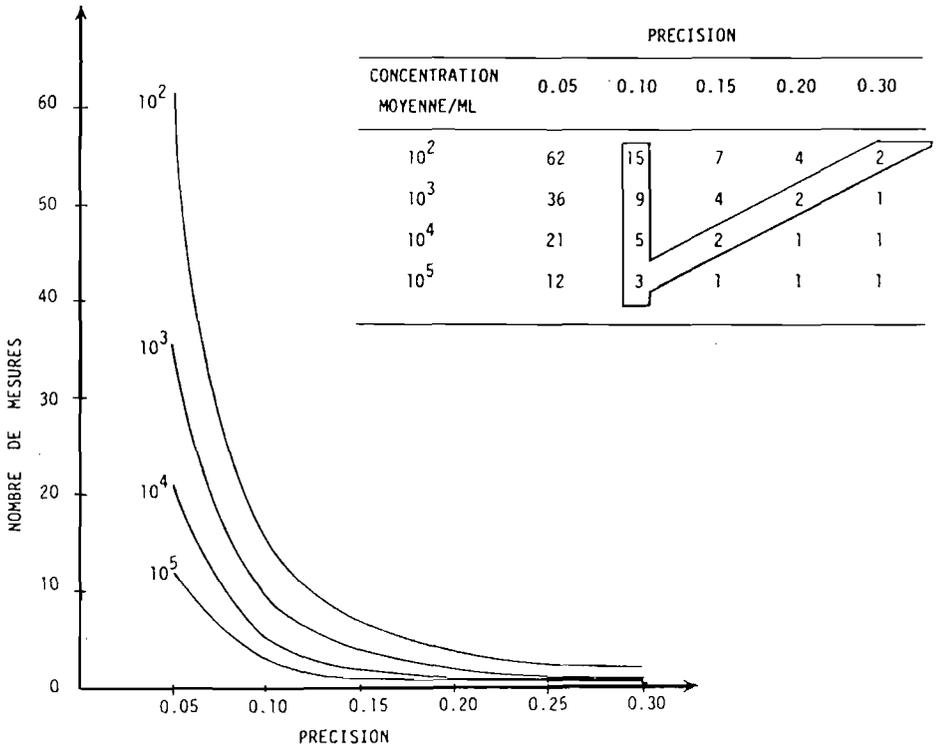


Figure 5 : Nombre de mesures à effectuer par prélèvement en fonction de la précision souhaitée et compte tenu de la concentration moyenne par ml des bactéries cultivant sur Marine agar.

La loi de Taylor et ses développements peuvent être appliqués à n'importe quel type de variables bactériologiques. Pour éventuellement généraliser les abaques de précision obtenus, il serait indispensable d'analyser d'autres prélèvements d'origine géographique différents. La loi des puissances de Taylor constitue donc une approche performante aux problèmes d'échantillonnage mais elle ne peut être employée qu'en considérant un seul niveau de variabilité à la fois. Elle ne permet donc pas de répartir de façon optimale un effort d'échantillonnage global entre différents niveaux de variabilité considérés simultanément.

**Optimisation d'une stratégie d'échantillonnage concernant plusieurs niveaux de variabilité**

En plus de la variabilité résiduelle dont l'origine, rappelons-le, se trouve au niveau des mesures, deux autres sources de variabilité, dues à deux niveaux d'observation, sont prises en compte dans cette étude : celle entre stations (observation de la baie) et celle entre prélèvements (observation des stations).

Chacune de ces sources de variabilité étant hiérarchiquement imbriquée l'une dans l'autre, leur analyse simultanée passe par une analyse de variance hiérarchisée (Snedecor and Cochran, 1967). Afin de pouvoir utiliser ce modèle, les données brutes ont été transformées en log 10 de façon à ce que les variances soient homogènes et que les données tendent vers la normalité. Les résultats de cette analyse de variance, à partir des dénombrements de bactéries cultivant sur Marine agar, sont reportés dans le tableau 1. La variance inter-mesures est toujours la plus faible et elle varie relativement peu d'une campagne à l'autre. La source de variance la plus élevée demeure la variabilité inter-stations. De plus, la différence entre un niveau supérieur de variation et son niveau inférieur a toujours été significative.

	SOURCE DE VARIATION	DDL	SCE	CM	F	PARAMETRES ESTIMES	COMPOSANTES DE LA VARIANCE	POURCENTAGES	
1ère CAMPAGNE	STATION (S)	4	95.09	23.77	40.01**	$25 S_s^2 + 5 S_p^2 + S_m^2$	$S_s^2 = 0.93$	87.8	
	PRELEVEMENT (P) 20		11.88	0.59	47.25**		$5 S_p^2 + S_m^2$	$S_p^2 = 0.11$	11.0
	MESURE (M) 100		1.25	0.01			$S_m^2$	$S_m^2 = 0.01$	1.2
2ème CAMPAGNE	STATION (S)	4	53.87	13.47	26.46**	-	$S_s^2 = 0.52$	80.0	
	PRELEVEMENT (P) 20		10.58	0.53	17.05**	-	$S_p^2 = 0.10$	15.4	
	MESURE (M) 100		3.10	0.03		-	$S_m^2 = 0.03$	4.6	
3ème CAMPAGNE	STATION (S)	4	21.75	5.43	25.17**	-	$S_s^2 = 0.21$	72.0	
	PRELEVEMENT (P) 20		4.32	0.21	4.51**	-	$S_p^2 = 0.03$	11.5	
	MESURE (M) 100		4.79	0.05		-	$S_m^2 = 0.05$	16.5	

Tableau 1 : Résultats de l'analyse de variance hiérarchisée réalisée sur les données de 3 campagnes de pré-échantillonnage sur l'étang de Thau. La variable bactériologique est la concentration par ml de bactéries cultivant sur Marine agar (DDL = degrés de liberté ; S.C.E. = sommes des carrés des écarts ; C.M. = carrés moyens ; \*\* = différence significative au seuil de 1%).

A partir de cette analyse de variance, un nouveau plan d'échantillonnage peut être défini qui tiendra compte d'une part de l'importance relative de chacune des sources de variabilité et d'autre part, et c'est là l'intérêt de la méthode, de l'effort ou du coût qu'il est possible de consacrer à cette étude.

L'expression des deux critères de base : effort d'échantillonnage global (1) et variance de la moyenne globale des concentrations bactériennes (2) servant à optimiser le plan d'échantillonnage est :

$$(1) ET = n_s E_s + n_s n_p E_p + n_s n_p n_m E_m$$

où  $n_s$  = nombre de stations ;  $n_p$  = nombre de prélèvements ;  $n_m$  = nombre de mesures ; ET = effort total ;  $E_s$  = effort relatif aux stations ;  $E_p$  = effort relatif aux prélèvements ;  $E_m$  = effort relatif aux mesures.

$$(2) S_x^2 = \frac{S_s^2}{n_s} + \frac{S_p^2}{n_s n_p} + \frac{S_m^2}{n_s n_p n_m}$$

où

$S_s^2$  = estimation de la variation inter-stations ;

$S_p^2$  = estimation de la variation inter-prélèvements ;

$S_m^2$  = estimation de la variance inter-mesures.

La stratégie la plus efficace est celle qui minimise le produit de ces deux critères (Snedecor and Cochran, 1967). Les solutions théoriques fournissant le nombre de prélèvements par station et le nombre de mesures par prélèvements sont : celles qui minimisent le produit

$$ET \times S_{\bar{x}}^2 : n_p = (E_s/E_p \times S_p^2/S_s^2)^{1/2};$$

$$n_m = (E_p/E_m \times S_m^2/S_p^2)^{1/2}$$

Le nombre de stations à échantillonner est alors obtenu en résolvant l'équation de l'effort global ou celle de la variance globale suivant le critère considéré comme dominant.

Lorsque les descripteurs bactériologiques sont des abondances en bactéries revivifiables, le critère dominant est la quantité de mesures réalisables quasi-extemporanément. En effet, la conservation des prélèvements même réfrigérée ou congelée peut modifier la concentration bactérienne initiale. (McLeod and Calcott, 1976 ; Nelson and Parkinson, 1978 ; Kaper *et al.*, 1978).

Dans le cas présent, nous avons considéré que l'effort exigé pour l'obtention d'une mesure est environ 15 fois plus important que celui nécessaire à la réalisation d'un prélèvement. Par rapport à ce dernier, l'effort de visite d'une station est estimé comme étant 5 fois plus élevé. Dans notre étude, l'effort est estimé en unité de temps. En utilisant ces données et celles fournies par l'analyse de variance, les nombres théoriques de prélèvements et de mesures sont trouvés égaux à 1. Pour tenir compte des risques liés aux manipulations, ces valeurs ont été fixées à 2.

Le dernier paramètre à déterminer est l'effort global pouvant être consacré à une campagne.

Dans un premier temps, considérons l'effort d'échantillonnage maximal que nous pouvons investir, soit environ 2 000 unités d'effort d'échantillonnage. Dans ce cas, le nombre de stations à échantillonner, de façon à obtenir la variance minimum et donc la meilleure précision de la concentration moyenne en bactéries à l'échelle de la baie, est de 18 (tab. 2). Cette valeur est identique pour les trois campagnes réalisées non seulement pour les bactéries cultivant sur Marine agar mais aussi pour les bactéries cultivant sur les autres milieux. Sur ce même tableau est rapportée la précision ainsi obtenue. Celle-ci est très bonne, moins de 6 %, en se souvenant cependant que cette précision est calculée sur des données transformées en  $\log_{10}$ . Enfin l'accroissement de l'efficacité relative (Sokal and Rohlf, 1981) par le plan optimisé se situe entre 200 et 250 % !

	1 <sup>re</sup> Campagne	2 <sup>e</sup> campagne	3 <sup>e</sup> campagne
$n_s$	18	18	18
$S_{\bar{x}}^2$ optimisé	0,055	0,032	0,013
$\bar{x}$ (en log)	3,91	4,84	3,94
P (%)	6,0	3,7	2,9
$S_{\bar{x}}^2$ observé	0,19	0,11	0,04
Accroissement de l'efficacité relative (%)	245	244	208

Tableau 2 : Résultats de l'optimisation du plan d'échantillonnage des bactéries cultivant sur marine agar  $E \approx 2000$ ,  $n_p = n_m = 2$ ), compte tenu des variances estimées grâce au plan de pré-échantillonnage du tableau 1. P% = précision ( $P = 1 / x \times \sqrt{S_{\bar{x}}^2}$ ); accroissement de l'efficacité relative ( $\Delta RE = RE - 100$  % où  $RE = (S_{\bar{x}}^2 \text{ observé} / S_{\bar{x}}^2 \text{ optimisé}) \times 100$ ).

L'effet d'un tel accroissement de l'efficacité du plan d'échantillonnage pour un même effort global que le plan initial de pré-échantillonnage est bien visualisé par l'examen des intervalles de confiance à 95 % des concentrations moyennes en bactéries pour chacune des campagnes (fig. 6). La diminution des intervalles de confiance est évidente pour la stratégie optimisée et ainsi le dénombrement moyen qui, dans cet exemple, est celui des bactéries cultivant sur Marine agar, apparaît alors significativement différent lors des trois campagnes.

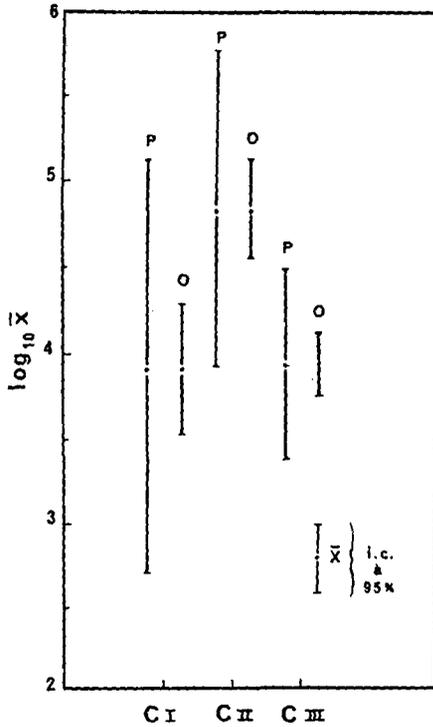


Figure 6 : Confrontation de la précision des estimations d'abondance des bactéries cultivant sur Marine agar entre le plan d'échantillonnage optimisé (O) et le plan pré-échantillonnage (P). Dans les deux cas  $ET \approx 2000$ .

Cependant échantillonner 18 stations dans une baie d'environ 1 km<sup>2</sup> peut paraître exagéré.

Qu'advient-il alors de la précision si l'effort d'échantillonnage global est diminué ? En d'autres termes, dans quelle mesure la précision des concentrations moyennes en bactéries sera-t-elle affectée si le nombre de stations implantées dans la baie est réduit ?

L'évolution de la précision du dénombrement des bactéries cultivant sur Marine agar, pour une échelle d'effort d'échantillonnage allant de 0 à 2 000, effort maximum réalisable par notre unité de recherche, est représentée sur la figure 7. Le gain de précision tend rapidement vers une asymptote lorsque l'effort d'échantillonnage augmente ou ce qui est similaire lorsque le nombre de stations augmente. Si par exemple l'obtention de 95 % de la précision maximale, pour les bactéries cultivant sur Marine agar, est considérée comme une précision satisfaisante —ce qui correspond à environ 10 % de précision sur la moyenne— le nombre de stations à échantillonner passe de 18 à environ 5 soit une réduction de 3,5 fois l'effort d'échantillonnage.