

THESE DE DOCTORAT  
DE L'UNIVERSITE PARIS VII  
Spécialité : BIOMATHEMATIQUES

**Contributions méthodologiques à un réseau de surveillance  
bactériologique de l'environnement marin littoral**

par

**Benoît BELIAEFF**

*Soutenu le 18 septembre 1992 devant le jury composé de MM. :*

**Alain-Jacques VALLERON  
Jean-Yves MARY  
Philippe GROS  
Marc TROUSSELLIER  
Alain LAUREC**

## Avant-Propos

*J'ai avant tout l'immense plaisir de dédier ma thèse à Claude, Anna et Louise sans qui ce document n'aurait sans doute pas vu le jour. A mes parents, sans lesquels je ne l'aurais pas vu.*

*Alain-Jacques VALLERON est le Directeur de l'Unité de Recherches Biomathématiques et Biostatistiques à l'université Paris VII et responsable de la formation doctorale dont j'ai pu bénéficier. Je le remercie vivement de m'avoir assisté lors de la phase difficile de la recherche d'un sujet de thèse et de présider aujourd'hui ce jury .*

*Jean-Yves MARY est le directeur de ma thèse. Malgré ses multiples occupations professionnelles, il a toujours su se rendre disponible pour m'écouter ou lire et corriger mes écrits. Sa grande compétence dans le domaine des statistiques me fut maintes fois utile ainsi que son soutien logistique particulièrement efficace. Je le remercie très chaleureusement d'être présent dans le jury aujourd'hui.*

*Alain LAUREC m'a offert l'opportunité d'effectuer ma thèse au sein de l'IFREMER, alors qu'il en était le Directeur des Ressources Vivantes et est à l'origine du sujet. Il me fait le grand honneur de participer au jury. Pour tout cela, je le remercie très vivement.*

*Philippe GROS est chercheur au laboratoire d'écologie benthique du centre IFREMER de Brest. J'ai pu apprécier sa rigueur, et son esprit, lors de la correction toujours rapide et néanmoins méticuleuse de certains de mes écrits dont le moindre n'est pas ce travail de thèse, qu'il rapporte aujourd'hui. Je lui adresse mes plus vifs remerciements.*

*Marc TROUSSELLIER est chercheur au laboratoire d'hydrobiologie marine et continentale de la faculté des Sciences de Montpellier. Ses différentes publications dans le domaine délicat de l'échantillonnage bactériologique en milieu marin m'ont été une aide précieuse au cours de mon travail, ainsi que les discussions que nous avons pu échanger à ce sujet. Il est également rapporteur de la thèse. Qu'il trouve ici le témoignage de ma gratitude.*

*C'est à Gérard LASSERRE, Directeur du laboratoire cité plus haut, que je dois mon orientation biomathématique. Je mesure aujourd'hui la pertinence de ses conseils. Il a toujours suivi mon parcours avec une amicale bienveillance. Ce jour, je suis heureux de lui adresser ma profonde reconnaissance.*

*Jean-Paul BERTHOME et Laurence MIOSSEC, responsable du REMI, m'ont accueilli au sein du laboratoire de surveillance du milieu (centre IFREMER de Nantes) et m'ont fourni le confort logistique nécessaire au bon déroulement de mon travail. Je les en remercie vivement.*

*Je tiens à remercier Alain MERCKELBAGH, Directeur de l'Environnement Littoral à l'IFREMER, pour m'avoir fait bénéficier d'un prolongement de bourse, me permettant ainsi de boucler ce travail.*

*Martial CATHERINE m'a fait profiter de son expérience du terrain et m'a amicalement permis de m'intéresser aux données de l'enquête de salubrité bactériologique de la plage Benoît. Qu'il en soit très chaleureusement remercié.*

*Un grand merci à Jacques DUPOND pour avoir mis à ma disposition les données conductance-métriques et pour avoir toujours répondu à mes interrogations bactériologiques.*

*J'exprime ma gratitude à Louis GIBOIRE qui a assuré l'essentiel de l'iconographie de ce travail, ainsi qu'à Christine Le PAUL pour sa contribution dans ce domaine et à Jean-Paul BEUCHER pour son efficace et amical soutien informatique.*

*Je ne peux citer toutes les personnes de l'IFREMER m'ayant aidé au cours de ces trois ans. Que Bernadette MINIER, Frédérique CHEVALIER et Françoise DUMONT soient remerciées au nom de tous.*

*Enfin à cette occasion, je souhaite que tous mes amis reçoivent le témoignage de ma profonde estime.*

## SOMMAIRE

INTRODUCTIONI. PRÉSENTATION GÉNÉRALE

<b>1. La surveillance bactériologique en milieu marin côtier</b>	<b>13</b>
<b>1.1. La surveillance de l'environnement marin</b>	<b>13</b>
<b>1.2. Le réseau de surveillance microbiologique</b>	<b>14</b>
1.2.1. <i>Historique et objectifs</i>	14
1.2.2. <i>Stratégies</i>	15
<b>1.3. Risque sanitaire et épidémiologie</b>	<b>16</b>
1.3.1. <i>Les germes pathogènes et les pathologies afférentes</i>	16
1.3.2. <i>Quelques données épidémiologiques</i>	17
<b>1.4. Les coliformes fécaux, indicateurs de contamination fécale en milieu marin</b>	<b>18</b>
1.4.1. <i>Biologie bactérienne : quelques généralités</i>	19
1.4.2. <i>La disparition des indicateurs fécaux dans l'eau de mer</i>	23
1.4.3. <i>Le devenir dans les sédiments</i>	26
1.4.4. <i>Propriétés physiologiques des bivalves et impact sur les données bactériologiques</i>	26
<b>1.5. Réglementations</b>	<b>31</b>
<b>1.6. Conclusion</b>	<b>32</b>
<b>2. Présentation des méthodes</b>	<b>34</b>
<b>2.1. Introduction</b>	<b>34</b>
<b>2.2. Techniques analytiques</b>	<b>34</b>
2.2.1. <i>Préparation de l'échantillon</i>	34
2.2.2. <i>Méthode de dilution des tubes</i>	34
2.2.3. <i>Conductance-métrie</i>	37

<b>2.3. L'estimateur du "Nombre le Plus Probable"</b>	39
2.3.1. <i>La fonction de vraisemblance</i>	39
2.3.2. <i>Estimation ponctuelle</i>	41
2.3.3. <i>Biais</i>	44
2.3.4. <i>Précision</i>	45
2.3.5. <i>Tests de comparaison entre deux estimations NPP</i>	48
<b>2.4. Stratégies d'échantillonnage</b>	55
2.4.1. <i>Introduction</i>	55
2.4.2. <i>Echantillonnage aléatoire simple (E.A.S.)</i>	55
2.4.3. <i>Echantillonnage à plusieurs niveaux</i>	57
2.4.4. <i>Echantillonnage composite</i>	61
2.4.5. <i>Echantillonnage systématique et krigeage</i>	62

## III. PRECISION DE L'ESTIMATION DE LA CONCENTRATION EN COLIFORMES FECAUX

<b>1. Estimation de la concentration bactérienne en routine</b>	71
<b>1.1. Introduction</b>	71
<b>1.2. Estimation ponctuelle et biais</b>	71
1.2.1. <i>Comparaison des estimations (modèle de Poisson vs. distribution binomiale)</i>	71
1.2.2. <i>Biais</i>	
<b>1.3. Estimation par intervalles</b>	75
<b>1.4. Discussion</b>	79
1.4.1. <i>Estimation ponctuelle et biais</i>	79
1.4.2. <i>Précision</i>	80
<b>2. Choix d'une combinaison de tubes et de dilutions</b>	83
<b>2.1. Choix des dilutions</b>	83
<b>2.2. Comparaison entre deux résultats NPP</b>	88
2.2.1. <i>Test du rapport des vraisemblances</i>	88
2.2.2. <i>Test "exact" de comparaison entre deux estimations NPP</i>	88
<b>2.3. Abaques précision-coût</b>	98
2.3.1. <i>Problématique</i>	98
2.3.2. <i>Détermination des axes</i>	98
2.3.3. <i>Exemples d'utilisation des abaques</i>	99

2.4. Discussion	101
3. Utilisation d'une méthode conductance-métrique	103
3.1. Introduction	103
3.2. Méthodes	103
3.2.1. Protocole analytique	103
3.2.2. Régression	103
3.3. Résultats	105
3.3.1. Régressions	105
3.3.2. Test de comparaison des droites	109
3.3.3. Exemple numérique	109
3.4. Discussion	110
4. Conclusion	112

III. REPRESENTATIVITE DE L'ECHANTILLONNAGE DANS LE  
CONTEXTE DU RESEAU DE SURVEILLANCE  
MICROBIOLOGIQUE

1. Enquête pilote à l'échelle locale	114
1.1. Problématique	114
1.2. Description de la zone d'étude et protocoles	114
1.2.1. Choix de la zone	114
1.2.2. Variables recueillies	117
1.2.3. Préenquête	117
1.2.4. Protocole de l'enquête pilote	121
1.2.5. Traitement des données	122
1.3. Résultats	
1.4. Discussion	
1.4.1. Intérêt du plan d'échantillonnage à plusieurs niveaux	126
1.4.2. Niveau de contamination	129
1.4.3. Les différentes sources de variabilité	130
1.5. Conclusion	133

<b>2. Estimation de la contamination fécale à l'échelle d'un secteur de surveillance</b>	134
<b>2.1. Examen des données du réseau de surveillance</b>	134
2.1.1. <i>Problématique</i>	134
2.1.2. <i>Méthodes d'évaluation de la contamination à l'échelle d'un secteur</i>	134
2.1.3. <i>Des exemples : résultats et discussion</i>	137
2.1.4. <i>Détermination d'un niveau moyen et optimisation de l'effort dans une zone homogène</i>	148
<b>2.2. Structure spatiale de la contamination fécale et cartographie</b>	157
2.2.1. <i>Introduction</i>	157
2.2.2. <i>Matériel et méthodes</i>	157
2.2.3. <i>Résultats</i>	163
2.2.4. <i>Discussion</i>	170
<b>2.3. Conclusion</b>	177
<b>3. Salubrité des zones conchylicoles.</b>	
<b>Un exemple : la plage Benoît (La Baule, Loire-Atlantique)</b>	178
<b>3.1. Problématique</b>	178
<b>3.2. Classement de salubrité de la plage Benoît</b>	178
3.2.1. <i>Description de la zone et protocole</i>	178
3.2.2. <i>Méthodes de traitement des données</i>	181
3.2.3. <i>Résultats</i>	182
3.2.4. <i>Formulation de l'avis</i>	188
<b>3.3. Critique de l'arrêté</b>	191
3.3.1. <i>Couverture spatiale</i>	191
3.3.2. <i>Fréquence d'échantillonnage</i>	191
3.3.3. <i>Fondement probabiliste du texte</i>	192
<b>3.4. Evaluation du risque consommateur</b>	195
3.4.1. <i>Définitions</i>	195
3.4.2. <i>Procédures d'échantillonnage relatives à la conformité à une norme</i>	196
3.4.3. <i>Limites des procédures</i>	199
<b>3.5. Conclusion</b>	203

## RECOMMANDATIONS ET CONCLUSION GÉNÉRALE

<u>BIBLIOGRAPHIE</u>	212
----------------------	-----

<u>ANNEXES</u>	232
I.1 Tableau d'allocation d'effort d'échantillonnage (REMI)	233
I.2 Descriptif de la base de données	234
I.3 Arrêté du 12 octobre 1976	235
I.4 Descriptif des programmes informatiques	236
II.1 Tables NPP (2/5/2) et (9/9/9)	237
II.2 Résultats de la régression linéaire	239

## INTRODUCTION

Au lendemain de la conférence des Nations unies sur l'environnement et le développement (CNUED), dite "Sommet de la planète Terre", la surveillance et les réseaux qui en constituent l'infrastructure apparaissent les outils indispensables à l'évaluation de la dégradation de l'environnement.

L'estimation de la concentration en contaminants dans les milieux qui nous entourent est entachée d'erreurs. La connaissance de l'erreur permet aux scientifiques d'assortir la diffusion des résultats obtenus par les programmes de surveillance de leurs limites d'interprétation.

Comme le souligne Kempf dans le Monde Diplomatique (juin 1992) cette attitude rigoureuse laisse place à l'argumentation diplomatique. Ainsi, l'auteur note la réserve de l'actuel Président des Etats-Unis au cours de ce sommet en raison de l'incertitude des "constructions scientifiques" sur la réalité des changements climatiques, et plus généralement il précise que "l'articulation entre science et politique, si elle ne résume pas la problématique écologique, en est une prémisses incontournable".

Toutes proportions gardées, on retrouve cette dualité dans le contexte de la surveillance bactériologique du littoral français, sujet de nos investigations doctorales, où la clarté d'une démarche, par exemple la formulation d'un avis sur la salubrité d'une zone de production conchylicole, peut induire des réactions polémiques, parfois légitimes.

La recherche de l'adéquation de l'information obtenue par le réseau microbiologique (REMI) de l'IFREMER aux objectifs préalablement définis a été notre motivation tout au long de nos travaux de thèse, menés depuis trois ans. Encouragé par l'existence d'études méthodologiques antérieures sur l'échantillonnage bactériologique en milieux aquatiques (*e.g.* El-Shaarawi *et al.*, 1981; Troussellier *et al.*, 1989; Maul *et al.*, 1989), il s'est agi de "traquer" l'incertitude (connaissance des limites) et de projeter les moyens de la réduire.

Dessiner le cadre conceptuel de la surveillance bactériologique en exprimant clairement les objectifs poursuivis est une base nécessaire justifiant les développements méthodologiques ultérieurs. Avec la présentation des méthodes d'estimation, au laboratoire et sur le terrain, de la concentration en coliformes fécaux (descripteur bactériologique retenu), ceci constitue le premier chapitre de notre travail de thèse.

Le deuxième chapitre traite des propriétés caractéristiques de l'estimateur attaché au dénombrement des germes et en particulier de sa précision, dans les conditions d'application en routine de la technique analytique préconisée : la méthode de dilution des tubes.

L'étude de la représentativité de l'échantillonnage à différentes échelles spatio-temporelles est assortie d'une identification de macrostructures dans la recherche de l'hétérogénéité spatiale de la contamination à grande échelle. Cette éventuelle hétérogénéité intervient également dans la problématique de la conformité de résultats bactériologiques à une réglementation, analysée ici et dont l'analyse clôt le troisième et dernier chapitre.

Des recommandations destinées à améliorer la stratégie de surveillance pour les différents objectifs et une réflexion sur la finalité de la surveillance bactériologique accompagnent la conclusion de ce travail.

## I. PRESENTATION GENERALE

## **1. La surveillance bactériologique en milieu marin côtier**

### **1.1. La surveillance de l'environnement marin**

Dans le contexte de l'étude et de la protection de l'environnement marin, la surveillance est l'évaluation répétée d'une activité ou d'un contaminant ou de ses effets directs ou indirects (ICES, 1989).

La surveillance s'exerce dans le cadre de réseaux. La définition des objectifs de la surveillance est sans doute la phase la plus importante de la chronologie de la mise en place d'un réseau. L'annexe 1 du huitième rapport annuel d'activité de la commission d'Oslo fournit les définitions d'objectifs suivants (Joanny & Chaussepied in RNO, 1988) :

- (a) l'évaluation des risques possibles pour la santé humaine,
- (b) l'évaluation de la nocivité pour les ressources vivantes et la vie marine (écosystème),
- (c) l'évaluation du niveau de pollution marine existant,
- (d) l'évaluation de l'efficacité des mesures prises dans le cadre des conventions en vue de la diminution de la pollution marine.

Trois réseaux sont gérés par l'IFREMER. Ce sont, dans l'ordre chronologique de mise en place:

- Le Réseau National d'Observation de la qualité du milieu marin (RNO), créé en 1974. La surveillance porte sur les paramètres généraux de qualité (température, salinité, sels nutritifs, etc...) et sur des polluants divers (métaux, PCB, hydrocarbures,...). La mission assignée au RNO correspond aux objectifs (c) et (d), élargis à l'évaluation des tendances des paramètres et polluants, analysés dans la colonne d'eau (Joanny et Chaussepied in RNO, 1988).

- Le réseau de surveillance phytoplanctonique (REPHY, 1984). Trois missions incombent à ce réseau, correspondant respectivement aux objectifs (a), (b) et (c): la protection de la Santé Publique, la protection des cheptels, et la recherche des niveaux et tendances de la concentration phytoplanctonique.

- Les objectifs "Protection de la santé humaine" (a) et "Détermination des niveaux et des tendances" (c) ont prélué à la restructuration du réseau de surveillance microbiologique (REMI) en 1989.

## **1.2. Le réseau de surveillance microbiologique (REMI)**

### *1.2.1. Historique et objectifs*

Afin de garantir la salubrité bactériologique des coquillages pour le consommateur, un contrôle bactériologique a été mis en place au début du siècle. Développée au départ par des laboratoires privés à l'initiative de négociants parisiens, cette surveillance a été reprise et institutionnalisée par l'Office Scientifique et Technique des Pêches Maritimes en 1919, puis par les organismes qui lui ont succédé, l'ISTPM<sup>1</sup> et l'IFREMER (Miossec, 1990).

La dégradation du milieu marin est un phénomène contemporain, dont l'ampleur est ressentie avec beaucoup d'acuité dans le domaine côtier. Actuellement, la préoccupation majeure des gestionnaires de l'environnement, qu'ils soient techniciens ou hommes politiques, est d'évaluer, voire de restaurer la qualité du milieu. L'état bactériologique du littoral est un des paramètres sensibilisateurs, si l'on en croit le succès du Pavillon Bleu Européen, label décerné aux communes présentant des plages bactériologiquement salubres et motivant ainsi les efforts d'assainissement (Garans & Courtois, 1991).

---

<sup>1</sup> Institut Scientifique et Technique des Pêches Maritimes

D'ambition nationale, l'objectif "environnemental" du REMI est la connaissance du niveau et des tendances de la contamination bactériologique du littoral. La concentration en indicateurs fécaux est alors considérée comme un indice de la qualité du milieu.

Des stratégies distinctes ont été élaborées pour répondre aux deux objectifs précédemment définis.

### *1.2.2. Stratégies*

La surveillance microbiologique couvre tout le littoral français. Le littoral français a été découpé initialement en 92 secteurs : 56 secteurs conchylicoles et 36 secteurs non-conchylicoles. Ces chiffres sont en évolution dans la mesure où des secteurs sont devenus conchylicoles et où d'autres vont le devenir.

Le tableau I.1 présente les différentes structures du réseau de surveillance microbiologique ainsi que les stratégies correspondantes. Le détail du calcul de l'allocation spatiale de l'effort d'échantillonnage par secteur, pour la surveillance et l'intervention, est donné par ailleurs (annexe I.1). Notons toutefois qu'il est fondé pour la surveillance sur un indice prenant en compte la production conchylicole d'un secteur donné relativement à la production nationale ainsi que la contamination moyenne en coliformes fécaux du secteur, avec un effort minimal de deux stations pour les secteurs non-conchylicoles. Pour l'intervention, le critère théorique d'allocation est le pourcentage de mesures en coliformes fécaux supérieures à une valeur critique, également affectée de la production relative du secteur (Miossec, 1990). Les 92 secteurs regroupent finalement 345 points de surveillance et 278 points d'intervention, dont la cartographie détaillée par secteur est fournie par Miossec & Raffin (1990). Ces chiffres sont également soumis à fluctuations du fait des contraintes opérationnelles de différentes natures afférentes à l'échantillonnage, dont une des plus évidentes est la disponibilité en coquillages au point.

Tab.I.1 : Structure du réseau de surveillance microbiologique (d'après Miossec, 1990)

	STRUCTURE PÉRENNE		STRUCTURE NON PÉRENNE	
	Surveillance	Intervention (alerte)	Études occasionnelles	Études de zones
Objectifs	Niveaux et tendances de la <b>qualité du milieu</b>	<b>Santé publique</b> Salubrité des coquillages	<b>Santé publique</b> Salubrité des zones à potentialités aquacoles ( <b>Milieu</b> )	Respect de la <b>réglementation</b>
Indicateurs	Coliformes fécaux	Coliformes fécaux*	Coliformes fécaux*	Coliformes fécaux*
Zones concernées	56 secteurs conchylicoles 36 secteurs non conchylicoles	56 secteurs conchylicoles	(1) Zones où nouveaux rejets (2) Zones à potentialités aquacoles	Zones à classer ou à reclasser
Fréquence	Mensuelle	Après déclenchement, au minimum hebdomadaire	(1) Fonction des caractéristiques du rejet (2) Au minimum mensuelle	26 prélèvements "espacés régulièrement" sur un an
Positionnement des stations	Systématique (linéaire)	Suivant un gradient de décontamination	(1) Suivant un gradient de décontamination (2) Systématique (quadrillage)	Systématique (quadrillage)

\* avec recherche de salmonelles

Les données récoltées sur le terrain selon les stratégies précédemment décrites sont stockées dans une base de données dont un descriptif plus que sommaire est fourni par ailleurs (annexe I.2).

### 1.3. Risques sanitaires et épidémiologie

#### 1.3.1. Les germes pathogènes et les pathologies afférentes

Les bivalves sont cultivés généralement dans des zones abritées, estuaires ou baies. Ces zones sont soumises à des rejets de natures diverses et en particulier à des apports nutritifs (nitrates, phosphates) nécessaires à la rentabilité de la production conchylicole. S'agissant parfois également de sites fortement urbanisés, ils sont lieux de rejets d'eaux usées domestiques, où les agents pathogènes rencontrés sont des bactéries ou des virus.

La plupart des bactéries rencontrées dans les eaux usées, dont la charge bactérienne peut atteindre  $10^9$  à  $10^{10}$  cellules.l<sup>-1</sup>, ne sont pas dangereuses pour l'homme; les espèces considérées

comme pathogènes à transmission hydrique sont réparties en quatre genres : *Salmonella* (bacilles de la typhoïde, des paratyphoïdes A et B et de diverses gastro-entérites), *Shigella* (bacilles dysentériques), *Escherichia* (essentiellement *E. coli* ou colibacille) parmi les Entérobactéries, et *Vibrio* (vibrion du choléra) parmi les Vibrionacées (Gauthier & Piétri, 1989).

Pour les pathologies virales, il s'agit principalement du virus de l'hépatite A et du virus de *Norwalk* ou apparenté *Norwalk*, responsable de gastro-entérites, mais de nombreux autres virus peuvent être rencontrés (e.g. voir Larkin & Hunt, 1982).

Le problème sanitaire majeur provient du fait que les coquillages filtreurs, ainsi potentiellement vecteurs de germes pathogènes, sont consommés crus ou après une légère cuisson.

### 1.3.2. Quelques données épidémiologiques

Les données portant sur les toxi-infections alimentaires collectives (TIAC) concernent en particulier les États-Unis (Prieur *et al.*, 1990) : 11 % des épidémies d'intoxications alimentaires y sont imputables aux produits de la mer, dont 2% pour les mollusques (Bryan in Prieur *et al.*, 1990). En France, les données du ministère de la Santé font état de 611 foyers de TIAC déclarées en 1989 pour 11146 malades recensés. Les produits de la mer seraient responsables de 6,7% de ces TIAC, dont moins de 3% auraient pour agent causal les salmonelles (Poggi, 1990). Les pathologies sont aussi bien bactériennes que virales. Ainsi, la consommation de coquillages contaminés par des virus est à l'origine de nombreux cas d'hépatites et de gastro-entérites (Schwartzbrod, 1990).

Le risque sanitaire lié à la consommation des coquillages ne peut réellement être apprécié qu'au travers d'études épidémiologiques rigoureuses. Malheureusement, peu d'études existent à ce jour; Poggi constatait en 1986 que l'absence de statistiques françaises fiables, régionales ou nationales, interdisait toute approche épidémiologique des risques présentés par la

consommation des fruits de mer. Les moyens considérables à mettre en oeuvre pour obtenir des résultats significatifs (Block & Prieur, 1987; PNUE/OMS, 1991) sont sans doute l'explication d'une telle carence.

Pour les eaux de baignade, l'existence d'une relation significative entre la fréquence des pathologies bactériennes et la contamination bactérienne des eaux de mer, a pu être montrée par divers auteurs (Cabelli *et al.*, 1982; Fattal *et al.*, 1986).

Concernant les coquillages, citons une étude de la NOAA<sup>1</sup> et de l'EPA<sup>2</sup> impliquant 1393 volontaires (White, 1989) : la morbidité chez les consommateurs d'huîtres prélevées en limite de zone insalubre est significativement plus importante que chez les témoins non-consommateurs d'huîtres. Notons que les consommateurs d'huîtres d'une zone notoirement saine sont également significativement plus malades que les témoins. Ceci ne doit pas être interprété comme "manger des coquillages rend malade", mais plutôt comme un exemple de risque "consommateur" non négligeable : il est probable de décréter une zone saine alors qu'elle ne l'est pas.

#### **1.4. Les coliformes fécaux, indicateurs de contamination fécale en milieu marin**

L'utilisation des indicateurs de contamination fécale est rendue nécessaire de par la difficulté à détecter, et *a fortiori* à dénombrer les germes pathogènes présents simultanément dans le rejet en mer (Trollope, 1984). Les faibles teneurs de ces derniers et la lourdeur et le caractère onéreux des techniques analytiques à mettre en oeuvre pour les mesurer en sont responsables. *A contrario*, les coliformes fécaux se trouvent en très grand nombre dans les eaux usées ( $10^8$  à  $10^9$  cell.l<sup>-1</sup>) et la numération de ces bactéries est plus simple et plus rapide (Gauthier & Piétri, 1989).

---

<sup>1</sup> National Oceanic and Atmospheric Administration

<sup>2</sup> Environmental Protection Agency

La quantité de coliformes fécaux présents dans les coquillages au moment de leur prélèvement est l'aboutissement d'une chaîne complexe de processus (Fig. I.1), se déroulant dans plusieurs compartiments, sur laquelle interagissent des facteurs abiotiques (hydrodynamisme, météorologie,...) et des facteurs biotiques (survie des bactéries, physiologie des bactéries et des bivalves).

Bissonnette *et al.* (1975) ont déjà souligné l'importance fondamentale de la connaissance des paramètres de la survie des bactéries en eau douce pour une interprétation rigoureuse des données. De façon analogue, une meilleure compréhension de la mesure de la concentration en coliformes fécaux dans les coquillages passe par l'examen du devenir de ces indicateurs dans les différents compartiments figurés. Au préalable, les principales caractéristiques biologiques de ces bactéries sont rappelées.

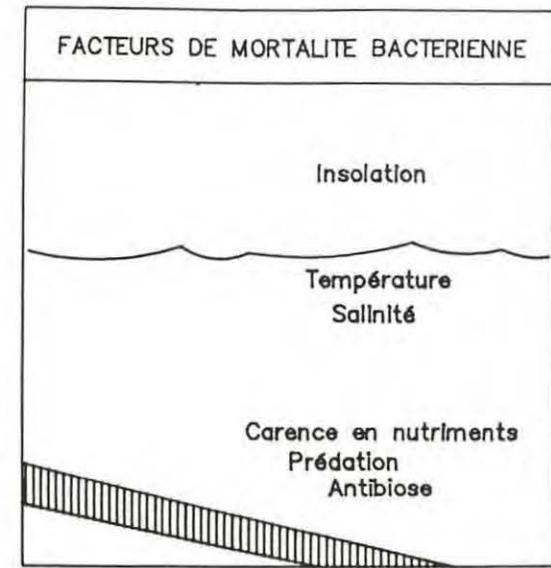
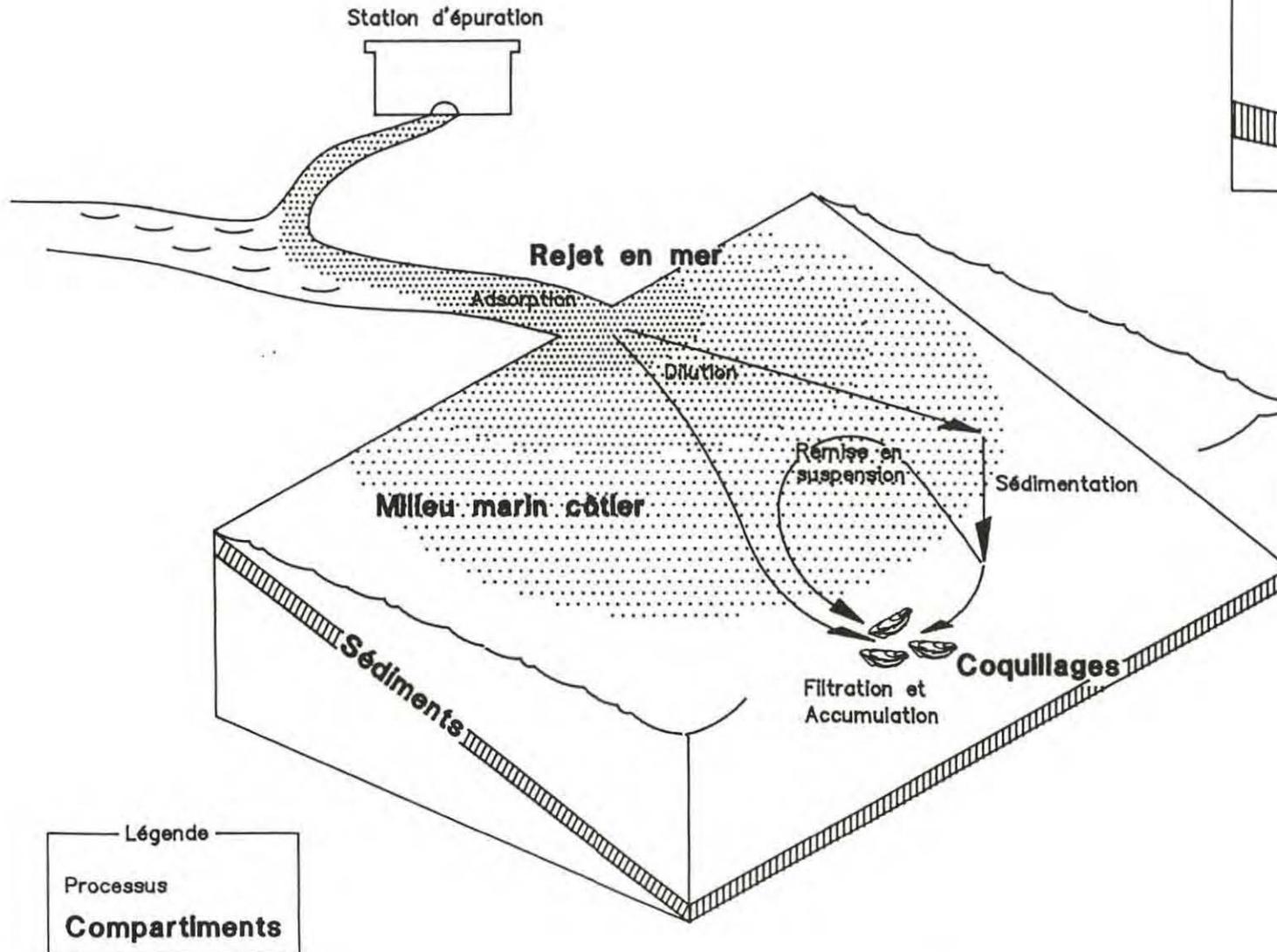
#### *1.4.1. Biologie bactérienne : quelques généralités*

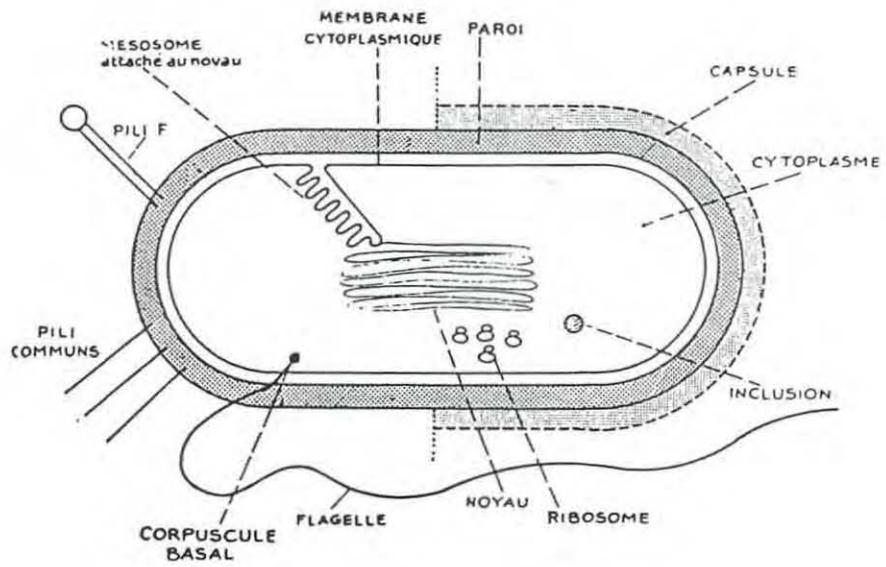
Les éléments de bactériologie donnés ici sont tirés de Anonyme (1972).

Les bactéries sont des Protistes inférieurs (classe des Schizomycètes), dits Procaryotes : il s'agit de petites cellules dans lesquelles le noyau est un unique chromosome nu, dépourvu de membrane (Fig. I.2).

Le groupe bactérien présentant le plus d'intérêt pour notre propos est le groupe des Entérobactéries; elles sont hétérotrophes : elles exigent pour se développer des composés organiques. Certaines de ces bactéries jouent chez l'homme un rôle biologique important : elles synthétisent des vitamines en quantité excédant largement leurs propres exigences, disponibles pour l'hôte. D'autre part, elles suscitent la formation d'anticorps gamma globulines spécialisées présentes dans le sérum humain normal. Les genres principaux de ce groupe sont *Escherichia*, *Citrobacter*, *Proteus*, *Klebsiella*, *Enterobacter*, *Shigella* et *Salmonella*; toutes les entérobactéries ne sont pas bénéfiques pour l'homme et nous avons déjà noté le risque sanitaire lié à la présence des deux derniers genres.

**Fig. I.1** : Schéma décrivant les différents processus aboutissant à la contamination fécale des coquillages en milieu marin à partir d'un rejet polluant





**Fig. 1.2** : Anatomie bactérienne (d'après Anonyme, 1972)

Les Entérobactéries sont Gram - : la structure chimique de leur paroi fait qu'elles ne prennent pas la coloration Gram; elles sont aérobies, anaérobies facultatives et utilisent le glucose par métabolisme fermentaire avec ou sans production de gaz. Ce sont des organismes de 2 à 4  $\mu\text{m}$  de long sur 0,4 à 0,6  $\mu\text{m}$  de large.

Parmi elles, *Escherichia coli*, hôte naturel des voies intestinales des mammifères supérieurs, secrètent les colicines, substances capables d'inhiber la multiplication excessive et dangereuse d'autres espèces bactériennes.

*E. coli* est le représentant le plus important et le plus étudié du groupe des Coliformes fécaux. Colibacille découvert en 1885 par Escherich, il fermente le lactose avec production de gaz, propriété utilisée pour son dénombrement (voir plus loin).

La reproduction des cellules se fait par scissiparité : une cellule-mère se scinde transversalement pour donner naissance à deux cellules-filles. La croissance bactérienne en général et de *E. coli* en particulier en milieu liquide se décompose en plusieurs phases : après une phase de latence, puis une phase d'accélération, la colonie bactérienne suit une croissance exponentielle. En coordonnée semi-logarithmique, la courbe de croissance devient alors une droite dont la pente est le taux de croissance, noté  $\mu$ ;  $\mu$  est défini comme le nombre de divisions (ou de "doublements" ou de "générations") par unité de temps. Dans des conditions normales de développement, le temps de génération (inverse du taux de croissance) d'*E. coli* est de 20' ( $\mu = 3 \text{ générations.h}^{-1}$ ).

Suite à cette phase exponentielle,  $\mu$  diminue puis s'annule marquant la fin du développement de la colonie bactérienne et annonçant sa disparition.

Les coliformes fécaux ne rencontrent évidemment pas les conditions optimales pour leur développement en milieu marin; ils y subissent des agressions de différentes natures : température, ensoleillement, stress osmotique,... (Fig. I.1). Le compartiment dans lequel ils se

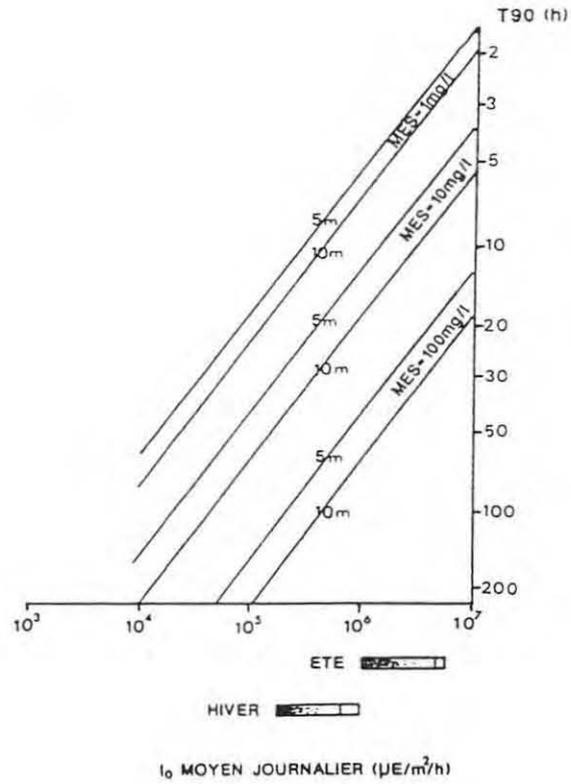
trouveront est la clef de leur survie dans le milieu marin. Notons par ailleurs que cette survie est également bien sûr conditionnée par l'état physiologique de la bactérie au moment du rejet mais également par les conditions physico-chimiques des milieux dans lesquels elle se trouvait (Munro, 1988), par exemple dans les eaux de rejet d'une station d'épuration d'eau.

#### 1.4.2. La disparition des indicateurs fécaux dans l'eau de mer

A partir d'une hypothèse de cinétique du premier ordre de disparition de la population bactérienne, on détermine le T90 : paramètre classiquement utilisé (e.g. Plusquellec, 1984), il s'agit du temps nécessaire pour voir disparaître 90% de la population bactérienne initiale. On distingue ici la disparition des bactéries en mer de la mortalité naturelle, en y adjoignant principalement la dilution du flux bactérien lors de son rejet en mer, phénomène souvent considéré comme prépondérant (e.g. Baleux *et al.*, 1988); ceci peut être visualisé en confrontant la courbe de décroissance de la contamination *in situ* à la courbe de dilution théorique de l'effluent (Bonfont *et al.*, 1986; Pommepuy *et al.*, 1987; Pommepuy & Guillaud, 1992).

De Rouville & Quetin (1983) soulignent que si la décroissance des bactéries en mer est un phénomène distinct de la mortalité stricte, elle ne constitue pas pour autant un processus physique rigoureux et identifié : la disparition et par suite le T90 sont dépendants de facteurs externes de différentes natures (Mitchell, 1968; Bellair *et al.*, 1977). Ainsi, Pommepuy *et al.* (1990) établissent des relations entre la profondeur, la turbidité, l'intensité lumineuse et le T90. Par exemple, pour une insolation estivale ( $10^6 \mu\text{E}\cdot\text{m}^{-2}\cdot\text{h}^{-1}$ ) et une eau peu turbide ( $1 \text{ mg}\cdot\text{l}^{-1}$ ), le T90 des bactéries fécales à 5 m de profondeur est voisin de 5 h (Fig. I.3). Gauthier & Piétri (1989) indiquent un T90 pour les coliformes fécaux, estimé *in situ*, voisin de 3h pendant le jour. La prédation et l'antibiose sont également des facteurs de mortalité des indicateurs fécaux (Munro, 1988).

Un bon indicateur devrait subir idéalement une disparition moins rapide, de durée mesurée par exemple par le T90, que celle des pathogènes dont il témoigne du risque de présence (Geldreich, 1966; Metcalf *et al.*, 1973; Elliott & Colwell, 1985).



**Fig. I.3** : Relations entre la profondeur, la turbidité, l'intensité lumineuse et les T90 (d'après Pommepuy *et al.*, 1990)

L'étude *in situ* de l'évolution comparée de la concentration en coliformes fécaux et de la concentration en germes pathogènes se heurte au choix d'une technique analytique de dénombrement. Dawe & Penrose (1978) ont les premiers montré l'existence d'un état de dormance pour les coliformes rejetés en milieu marin; ces bactéries replacées dans un milieu de culture adéquate pouvaient ensuite reformer des colonies. Colwell *et al.* (1985) montrent que remis en culture après la phase de dormance, *Vibrio cholerae* conserve son pouvoir virulent. Le lecteur aura compris l'impact sanitaire de tels résultats qui ont rendu obsolète l'hypothèse de disparition à court terme des bactéries fécales en milieu marin et ont expliqué les différences parfois très importantes de résultats entre les méthodes de comptages directs et indirects.

Ainsi, Xu *et al.* (1982), étudiant la survie de *Escherichia coli* et de *Vibrio cholerae*, montrent que ces deux bactéries, en milieu estuarien, deviennent non cultivables par les méthodes classiques de mises en culture (boîtes de Pétri en milieu solide et estimations du nombre le plus probable en milieu liquide) mais restent viables et dénombrables par immunofluorescence combinée à la coloration de l'ADN bactérien par l'acridine orange. Soulignons au passage que *Vibrio cholerae* conserve sa virulence même en état de dormance (Colwell *et al.*, 1985).

Un autre obstacle à la recherche de liaisons entre indicateurs et pathogènes est imputable à l'interaction des conditions environnementales sur les mesures. L'intensité de la liaison sera fonction des valeurs prises par tel ou tel facteur. Rhodes & Kator (1988), s'intéressant à la survie d' *E. coli* et de *Salmonella* spp en estuaire, mettent l'accent sur l'interaction des facteurs physiques, et en particulier la température de l'eau sur les résultats. On trouvera par ailleurs dans Roszak & Colwell (1987) une revue bibliographique des facteurs de survie bactérienne dans les milieux aquatiques.

Les exemples précédents attestent de la difficulté à mettre en évidence des relations quantitatives voire qualitatives entre indicateurs de contamination fécale, en particulier les coliformes, et les germes reconnus pathogènes. Un corollaire est que les coliformes fécaux ne

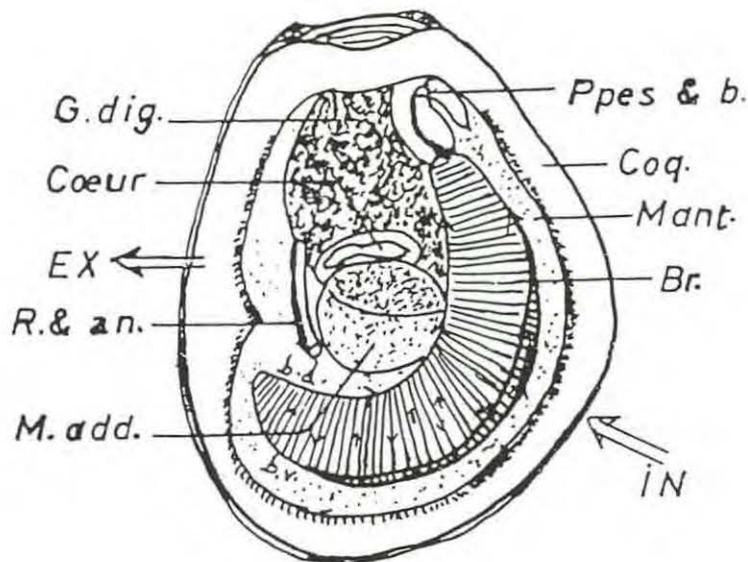
sont en général pas considérés comme de bons indicateurs (Gallagher & Spino, 1968); les streptocoques fécaux leur sont souvent préférés (Cohen & Shuval, 1972; Dutka, 1973; Cornax *et al.*, 1991). On trouvera néanmoins dans Bonde (1977) un exemple de relation significative entre la concentration en coliformes fécaux et le pourcentage d'occurrence en *Salmonella* spp pour des mesures effectuées en estuaires.

#### 1.4.3. Le devenir dans les sédiments

Les bactéries rejetées sont pour une majeure partie adsorbées sur des particules en suspension, soumises à la sédimentation (Mitchell & Chamberlin, 1974; Paerl, 1975; Dan & Stone, 1991). Les indicateurs fécaux ainsi que les pathogènes présents dans les sédiments peuvent y trouver des conditions favorables à leur survie du fait de l'abondance en matière organique (Gerba & McLeod, 1976; Erckenbrecher, 1981; Pommepuy *et al.*, 1989). De ce fait, des corrélations significatives ont pu être établies entre coliformes fécaux et virus dans les sédiments (LaBelle *et al.*, 1980). Si la durée de survie des indicateurs dans l'eau de mer est de l'ordre du jour, la semaine paraît un ordre de grandeur raisonnable dans les sédiments pour les bactéries d'origine entérique (Le Guyader *et al.*, 1991). Ainsi la remise en suspension de sédiments peut conduire à une recontamination du milieu.

#### 1.4.4. Propriétés physiologiques des bivalves et impact sur les données bactériologiques

La contamination des mollusques bivalves s'effectue au travers de l'ingestion soit des bactéries fécales directement, soit de particules, sur lesquelles sont adsorbées les germes fécaux, après remise en suspension ou non des sédiments : elles sont filtrées au niveau des branchies, également organe respiratoire, où elles sont englobées dans un mucus. Les plus grosses sont évacuées sous forme de pseudo-fèces, les autres sont conduites jusqu'aux palpes labiaux puis à la bouche pour être finalement orientées vers le tractus digestif de ces mollusques (Fig. I.4). Ceci entraîne *in fine* la concentration des bactéries dans le tractus et plus particulièrement dans l'intestin postérieur, où elles peuvent trouver des conditions propices à



**Fig. I.4 :** Anatomie générale schématisée d'une huître plate (*Ostrea edulis*) ouverte. *Ppes & b.* : palpes labiaux et bouche; *Coq.* : valve creuse; *mant.* : manteau; *Br.* : branchie. *IN* : zone d'inhalation de l'eau (côté ventral). *b.d.* et *b.v.* : bord dorsal et bord ventral des branchies; *M. add.* : muscle adducteur; *R. & an.* : rectum et anus; *EX* : zone de rejet de l'eau (côté dorsal); *G. dig.* : glande digestive (pseudo hépato-pancréas). Les flèches sur les branchies indiquent les directions suivies par les particules tamisées (schéma inspiré de Walne *in* Marteil, 1976).

leur survie voire à leur développement (Prieur, 1981; Prieur *et al.*, 1990), tout en gardant leur pouvoir infectieux (Plusquellec, 1984).

La capacité des bivalves à concentrer de nombreux polluants, parmi lesquels les contaminants bactériens, est bien connue et utilisée dans les pays ayant la volonté et les moyens d'effectuer une surveillance systématique de leur littoral. Le *Mussel Watch* américain (Goldberg *et al.*, 1978) constitue la première utilisation des bivalves comme "organismes sentinelles" à une échelle nationale, pour en particulier la détermination des concentrations en métaux lourds dans les eaux marines littorales.

La concentration des bactéries dans le mollusque bivalve est une caractéristique nécessaire de l'indicateur correspondant mais non suffisante. Encore faut-il que le facteur de concentration soit fonction de la teneur bactérienne dans le milieu. Suite à de nombreuses études, Cabelli et Heffernan (in Plusquellec, 1984) montrent que le taux de bactéries dans les tissus du clam (*Mercenaria mercenaria*) atteint un équilibre fonction du taux de particules en suspension. Considérant que la majeure partie des bactéries sont adsorbées sur le matériel particulaire, comme souligné plus haut (§ 1.4.3), les bivalves peuvent être considérés comme indicateurs du niveau de contamination des eaux.

Les cinétiques de contamination et de décontamination des coquillages filtreurs sont fonction principalement de l'espèce considérée, de la température de l'eau et de la concentration en particules de l'eau. Paoletti (in Plusquellec, 1984) indique une durée de deux heures pour atteindre la concentration maximale dans le cas de moules, *in vitro*. La durée de décontamination maximale serait voisine de 24 heures (voir la revue bibliographique de Richards, 1988). Prieur *et al.* (1990) indiquent que la plupart des auteurs situent la durée de décontamination entre 24 et 48 h. Ces chiffres nous donnent la mesure de l'intégration temporelle maximale réalisée par les mollusques. Pour sa part, Plusquellec (1984) a pu montrer des variations similaires entre les mesures horaires de concentration en coliformes fécaux dans les moules et dans l'eau de mer. S'ils garantissent l'indépendance stochastique des mesures

bactériologiques pour des pas de temps rapprochés, ces résultats indiquent certainement la nécessité de l'augmentation de la fréquence d'échantillonnage pour la détermination d'un niveau moyen de la qualité bactériologique du milieu. Delattre & Delesmont (1981) indiquent également la nécessité d'analyses répétées dans le temps pour l'appréciation de la qualité sanitaire d'une zone littorale.

Prieur *et al.* (1990) mentionnent des ordres de grandeur de taux d'enrichissement par rapport à l'eau variant de 10 pour les coliformes fécaux à plus de 100 pour les streptocoques fécaux. La tendance des streptocoques à se présenter en amas pourrait ainsi favoriser leur rétention (Delattre & Delesmont, 1981).

Le taux d'enrichissement pour un coquillage donné est fonction du taux de filtration du mollusque utilisé. Héral (1989) définit le taux de filtration comme la quantité d'eau épurée à 100% par le mollusque, par heure et par gramme de poids sec (gps) de la chair de l'animal. L'animal pouvant continuer à pomper sans filtrer, la filtration ne doit pas être confondue avec le pompage (circulation de l'eau au travers du corps de l'animal). Toutes conditions étant égales par ailleurs, le taux de filtration peut fortement varier d'une espèce de bivalve à l'autre (Bonadona *et al.*, 1990). L'huître japonaise (*Crassostrea gigas*) filtre cinq fois plus que l'huître plate (*Ostrea edulis*), pour la même température (Héral, 1989).

L'exigence de couverture totale du littoral par le réseau microbiologique entraîne l'utilisation d'espèces de bivalves différentes pour une raison évidente de disponibilité en coquillages. Ainsi si les filtreurs (filtration dans la masse d'eau) huître creuse (*Crassostrea gigas*) et moule (*Mytilus edulis*) sont majoritairement utilisés, des déposivores de surface (filtration dans le sédiment) tels que palourdes (*Venerupis*) et coques (*Cerastoderma edule*) le sont également, avec des taux de filtration différents. Ceci est un frein important aux comparaisons inter-stations ou inter-secteurs de surveillance : la différence entre deux mesures de contamination mesurée dans deux coquillages différents peut n'être que le reflet des variations inter-espèces pour une même dégradation bactériologique de l'environnement marin.

Mais l'espèce n'est pas le seul facteur en cause dans la variation du taux de filtration du bivalve. Ces facteurs sont nombreux et variés (Walne, 1972) : température, salinité, oxygène dissous, concentration et nature des éléments en suspension, taille, âge et condition physiologique du mollusque, etc... Ajoutons que deux animaux de même taille, dans les mêmes conditions, peuvent présenter des taux de filtration différents et que chez un même animal cette vitesse n'est pas constante (Marteil, 1976).

En illustration, Fiala-Médioni et Copello (*in* Héral, 1989) montrent que le taux de filtration pour *Crassostrea gigas* (huître creuse) à 15°C, varie de 2,1 à 12,3 l.h<sup>-1</sup>.gps<sup>-1</sup>, quand la concentration en cellules phytoplanctoniques (source de nourriture) passe de 27,5 10<sup>6</sup> à 0,1 10<sup>6</sup> cell.l<sup>-1</sup>.

Plusquellec *et al.* (1990) ont pu observer que le taux d'enrichissement présente des fluctuations saisonnières. En particulier, la maturation sexuelle du mollusque a des répercussions très importantes sur son taux de filtration. Pour Deslous-Paoli *et al.* (1987), les évolutions saisonnières de la filtration correspondent aux cycles de maturation des gonades des différents bivalves : les plus forts taux de filtration sont enregistrés au moment de la maturité des gamètes et des pontes consécutives, quel que soit le mollusque considéré. Par exemple, la moule (*Mytilus edulis*) présente un taux de filtration moyen de  $3,4 \pm 1,5$  l/gps/h entre juin et février (hors période de fin de gamétogénèse et ponte), contre  $13,3 \pm 5,3$  l/gps/h entre février et mai. Des pics post-hivernaux de concentration en coliformes fécaux, indices de la contamination bactériologique du milieu, seraient alors la manifestation indirecte d'une activité physiologique chez le mollusque considéré. Hussong *et al.* (1981) ont pu effectivement observer une saisonnalité de la concentration des coliformes par les huîtres.

La fluctuation de la mesure de la concentration des indicateurs dans les coquillages pourrait être amplifiée non seulement par la multiplicité des facteurs agissant sur la physiologie des coquillages mais également par leurs interactions. Par exemple, la concentration des particules

en suspension dans l'eau va agir différemment sur le taux de filtration selon l'état physiologique du mollusque (Berthomé, comm. pers.).

Tous ces facteurs de variabilité, auxquels on peut adjoindre l'interaction des variables environnementales sur la technique de comptage bactérien préconisé, font que la mesure de la concentration en coliformes fécaux dans les coquillages est un signal particulièrement bruité, rendant difficile l'atteinte des objectifs prédéfinis.

Malgré ces limites, dont il est difficile de quantifier la portée sur les données du réseau microbiologique, l'utilisation du bivalve permet néanmoins de s'affranchir dans une certaine mesure de la forte variabilité spatiale de la mesure des coliformes dans l'eau de mer (Ayles *et al.*, 1978; Turick *et al.*, 1988). Cela répond également aux deux objectifs du réseau microbiologique :

- le mollusque est un organisme sentinelle au sens du *Mussel watch* (voir début § 1.4.4), mais employé ici à l'évaluation des niveaux et tendances de la contamination bactériologique du milieu,

- le mollusque est également une denrée dont la dégradation de la qualité bactériologique fait encourir un risque sanitaire aux consommateurs de coquillages, même si des liaisons satisfaisantes entre les concentrations en coliformes fécaux et en organismes pathogènes, mesurées dans les coquillages filtreurs, n'ont pu être mises en évidence (Hood *et al.*, 1983; voir la revue bibliographique dans Prieur, 1990).

### **1.5. Réglementations**

En dépit de la carence constatée en études nationales permettant d'établir un lien entre la concentration en indicateurs fécaux et la morbidité due à l'ingestion des bivalves, l'objectif de protection de la santé humaine nécessite de recourir à des normes.

Les textes réglementaires existant permettent ainsi l'examen de la conformité des résultats bactériologiques à des valeurs-guides ou -seuils. Même si ces valeurs peuvent être entachées d'un certain arbitraire, elles existent néanmoins et fournissent matière à des formulations d'avis et à des prises de décision, que ce soit pour l'évaluation de la qualité du milieu ou bien pour garantir la salubrité de la denrée.

En application du décret du 20 août 1939 relatif à la salubrité des coquillages, l'arrêté du 12 octobre 1976 (annexe I.3) précise les conditions régissant le classement de salubrité des zones conchylicoles, en fixant une méthode analytique de référence; les coliformes fécaux sont dénombrés indirectement au moyen de la méthode de dilution des tubes, dont l'interprétation numérique est une estimation du "Nombre le Plus Probable" (NPP) de germes dans un volume unitaire de l'échantillon à analyser. En fixant une méthode d'analyse, le législateur fixe le degré de précision de la mesure. Nous étudierons par la suite les conséquences que ceci peut avoir en termes de prise de décision (§ III.3). L'arrêté du 12 octobre 1976 est le seul document en droit français qui permette actuellement le classement de salubrité des zones conchylicoles (Catherine, 1992).

La directive du Conseil (CEE) du 30 octobre 1979 (1979) précise les conditions relatives à la qualité des eaux conchylicoles afin d'éviter des disparités entre les états membres, pouvant créer des conditions de concurrence ayant une incidence directe sur le fonctionnement du marché commun. Catherine (1992) note que cette directive est peu appliquée.

Citons enfin la récente Directive du conseil (CEE) du 15 juillet 1991 indiquant les règles sanitaires régissant la production et la mise sur le marché des mollusques bivalves vivants (Catherine, 1992).

## **1.6. Conclusion**

L'exposé des différents objectifs du réseau de surveillance microbiologique aboutit à la définition de stratégies correspondantes. Dans ce contexte, la description du devenir des

coliformes fécaux en milieu marin est un préalable nécessaire à l'interprétation des données collectées, dans la mesure où des sources de variabilité de la mesure se répartissent dans chaque compartiment et pour chaque processus, conduisant à un signal sans doute très bruité.

Le choix de la technique de comptage bactérien n'est pas anodin et en particulier quand il s'agit d'évaluer la qualité bactériologique du milieu marin dans un objectif de protection de la santé humaine. Le réseau de surveillance microbiologique fait appel à des dénombrements indirects; ceci est imposé par le respect des standards en matière d'énumération des coliformes fécaux dans les coquillages, repris par les textes réglementaires.

Le prochain chapitre constitue la description des étapes suivantes vers l'obtention de la mesure : stratégies d'échantillonnage sur le terrain et analyse au laboratoire. Une étude approfondie de l'interprétation numérique de la méthode de dilution des tubes, à laquelle font référence les textes réglementaires est également conduite.

## 2. Présentation des méthodes

### 2.1. Introduction

L'optimisation du réseau de surveillance microbiologique passe au préalable par :

- l'examen approfondi de la technique analytique utilisée,
- le choix de stratégies d'échantillonnage permettant les nécessaires études exploratoires à différentes échelles spatio-temporelles.

### 2.2. Techniques analytiques

#### 2.2.1. Préparation de l'échantillon

La concentration en coliformes fécaux, exprimée en CF/100 ml, est mesurée dans les coquillages. Pour chaque échantillon, les coquilles sont nettoyées puis la chair et le liquide intervalvaire sont recueillis jusqu'à obtenir un poids minimum de 25 g. Un broyat est réalisé à l'aide d'un homogénéisateur de type *Waring Blendor*, après dilution au 1/3 avec de la tryptone-sel.

A partir de ce broyat, le dénombrement indirect des coliformes fécaux a été effectué à l'aide de deux techniques : l'ensemencement en milieu liquide de séries de dilutions successives de l'échantillon et une technique conductance-métrique.

#### 2.2.2. Méthode de dilution des tubes

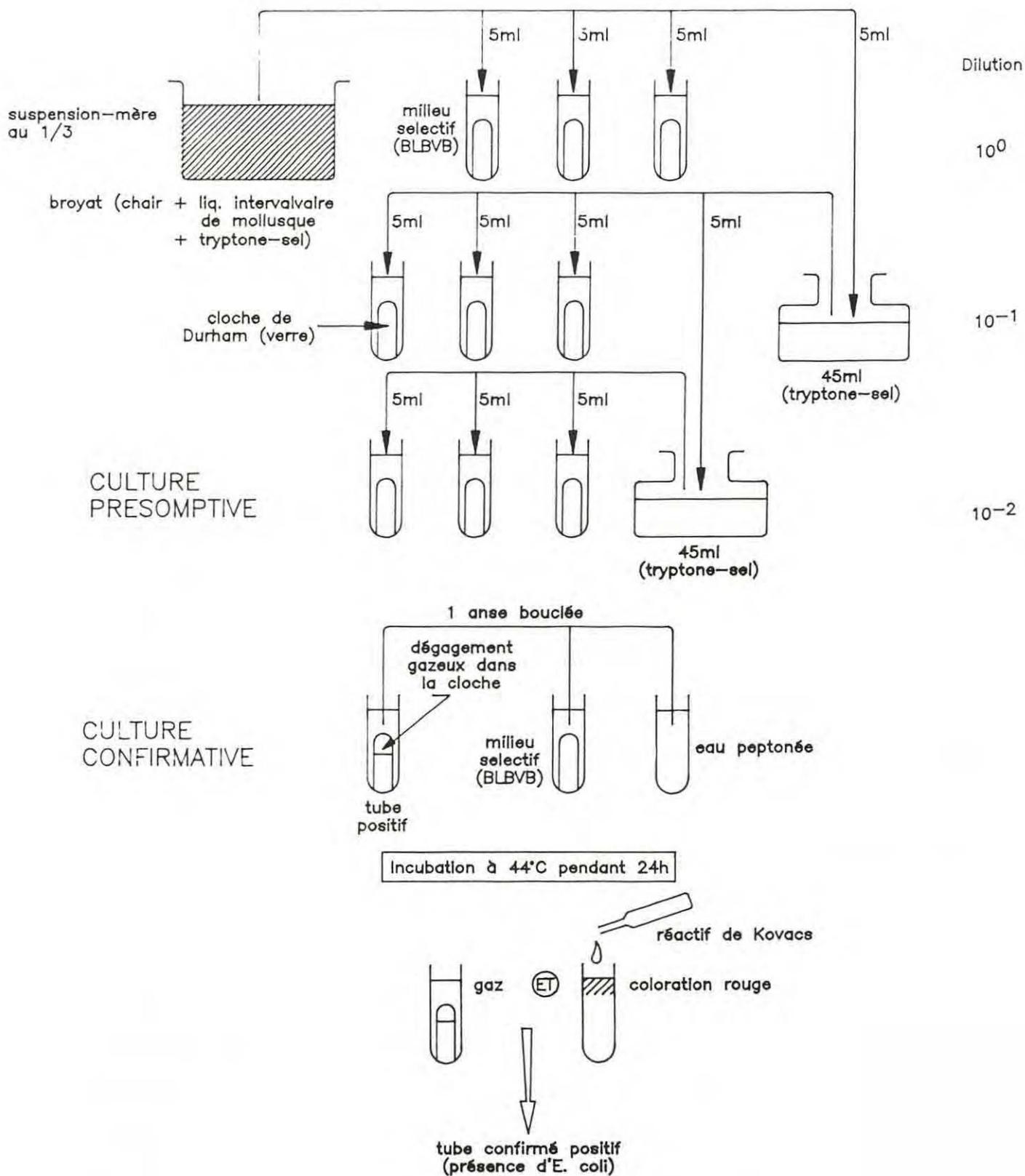
Le principe de la méthode exposée ci-après est tiré de la norme française homologuée V 45-110, portant sur le dénombrement des coliformes fécaux dans les eaux conchyliques et dans les coquillages marins vivants. La norme définit les coliformes fécaux, ou *Escherichia coli* présumés, comme des coliformes thermo-tolérants, qui, à la température de 44°C fermentent le lactose avec production de gaz et donnent de l'indole à partir du tryptophane.

Les modifications vis-à-vis de la norme reposent sur les concentrations en milieux utilisés et les volumes d'ensemencement; la norme préconise d'ensemencer 10 ml de broyat de coquillages dans 5 ml de milieu sélectif triple concentration. L'application de ceci au laboratoire entraîne la formation systématique d'un bouchon particulière dans le tube inoculé. Ceci peut induire des conditions d'anaérobiose, propices au développement d'espèces telles que *Clostridium perfringens*, permettant *in fine* l'occurrence de faux-positifs. En outre, compte tenu de la proportion importante d'inoculum dans le tube après ensemencement, on peut s'interroger sur le maintien du caractère sélectif du milieu (Dupond, comm. pers.). Les différentes étapes de l'analyse sont les suivantes (Fig. I.5) :

- Ensemencement de trois tubes contenant un milieu sélectif liquide (bouillon lactosé bilié au vert brillant) à double concentration avec une quantité déterminée de l'échantillon pour essai. Dans le cas des coquillages, la suspension est diluée au 1/3 avec une solution tryptone-sel. Dans les mêmes conditions sont ensemencées des dilutions décimales de concentrations appropriées, à partir de l'échantillon pour essai. La première dilution est obtenue en introduisant 5 ml de la solution mère dans un flacon contenant 45 ml de diluant. Des prélèvements de 5 ml dans la solution au 1/10 ainsi obtenue servent ensuite à ensemencer les tubes de la dilution correspondante. La dilution au 1/100 est obtenue en introduisant 5 ml du flacon précédent dans un flacon de 45 ml de diluant, etc... L'utilisation d'un volume d'inoculum de 5 ml pour toutes les dilutions est justifié eu égard au caractère fortement particulière de l'échantillon. Pipeter 0,5 ml, et *a fortiori* 0,05 ml, conduirait à des erreurs importantes sur le volume d'inoculum pour les dilutions élevées. Pour la même raison un volume de 5 ml est préféré à un volume de 1 ml pour l'inoculum.

- Trois séries de tubes de dilutions appropriées sont incubées à 37°C pendant 48h.

- A partir de chaque tube de milieu sélectif dans lequel est observé un dégagement gazeux (tube lactose +), ou douteux, on effectue des subcultures (repiquages) dans un tube contenant



**Fig. I.5 :** Description de la méthode de dilution des tubes appliquée à la mesure des coliformes fécaux dans les coquillages (3 dilutions et trois tubes par dilution; volume d'inoculum constant égal à 5 ml)

un milieu sélectif liquide simple concentration et dans un tube contenant de l'eau peptonée; ces tubes sont incubés à 44°C pendant 24h.

- Les tubes dont les repiquages ont donné à la fois une production de gaz dans le milieu sélectif et une production d'indole dans l'eau peptonée sont dénombrés; ce sont les tubes positifs, à partir desquels est estimée la concentration bactérienne dans la suspension mère, au moyen de l'estimateur du "Nombre le Plus Probable" (NPP), dont nous exposons la théorie plus loin.

### 2.2.3. Conductance-métrie

#### Principe et dispositif

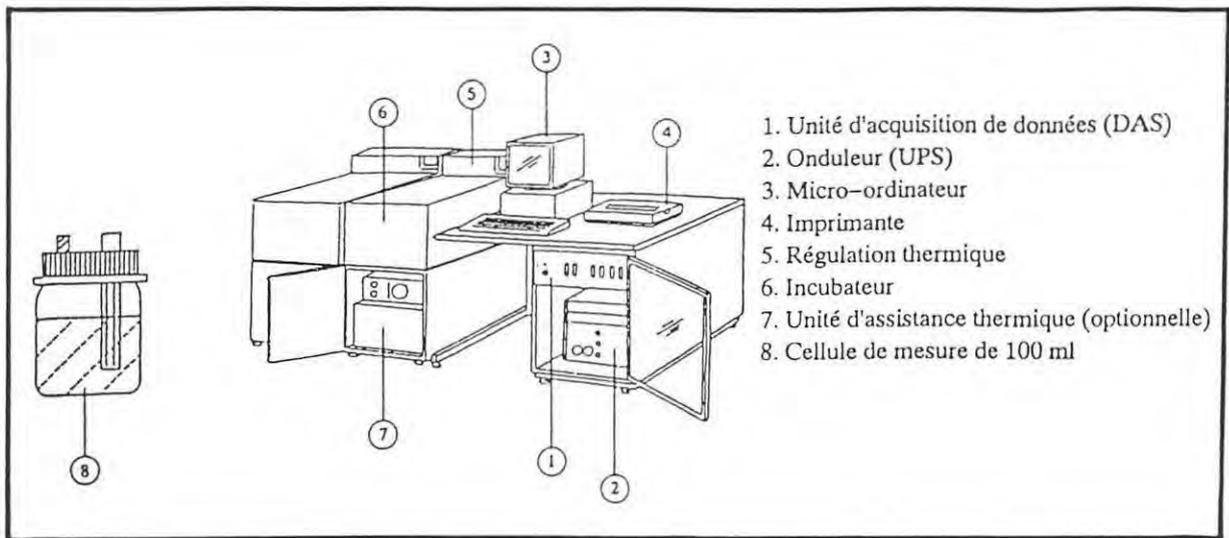
Introduites dans un milieu de culture approprié, les bactéries utilisent pour leur croissance les substances présentes et libèrent des métabolites dont certains jouent le rôle de dipôles électriques. Il s'ensuit une augmentation de la conductance du milieu. Celle-ci est mesurée en faisant passer un courant alternatif de 10 KHz entre deux électrodes (Dupont *et al.*, 1989). La figure I.6 présente le dispositif de l'analyseur microbiologique "Malthus", utilisé au laboratoire de méthodologie analytique du centre IFREMER de Nantes. Cet appareil a été conçu par la société *Malthus Instruments Ltd* (William Clowes Street, Burslem, Stoke-on-Trent, Angleterre); il est fabriqué par *Radiometer* (Copenhague, Danemark) et commercialisé en France par *Radiometer Tacussel* (Paris).

#### Analyse

10 ml de la suspension à analyser sont inoculés dans une cellule contenant du milieu MCM<sup>1</sup>, fabriqué et commercialisé par la société *Malthus*. Dans la cellule plongent deux électrodes (Fig. I.6), permettant la mesure conductance-métrique en continu au cours de l'incubation au bain-marie à 44°C. Différents auteurs, pour diverses espèces bactériennes dans des milieux variés, ont déjà relié une mesure conductance-métrique (ou impédance-métrique)

---

<sup>1</sup>*Malthus Coliform Medium*



**Fig. I.6 :** Description de l'analyseur microbiologique Malthus (d'après Dupont *et al.*, 1989)

dans un milieu de culture à un dénombrement bactérien direct ou indirect au moyen d'une régression linéaire (*e.g.* Firstenberg-eden, 1983; Firstenberg-eden & Tricarico, 1983; Gibson *et al.*, 1984 et Ogden, 1986, spécifiquement pour l'appareil "Malthus"). Pour leur part, Dupont *et al.* (1989) montrent que le temps de détection (TD) de la phase de croissance exponentielle de la courbe conductancemétrique - marquant l'arrêt de la phase de latence - s'exprime linéairement en fonction du logarithme décimal de la concentration en coliformes fécaux. La concentration en indicateur est prédite à l'aide de la droite de régression, pour le temps de détection obtenu.

### 2.3. L'estimateur du "Nombre le Plus Probable"

Le résultat de la méthode de dilution des tubes s'exprime en nombre de tubes positifs par dilution, fournissant ainsi un "nombre caractéristique". Pour chaque tube, on dispose d'une information binaire : le tube est positif ou non, il y a croissance bactérienne ou non. Les nombres totaux de tubes positifs et négatifs par dilution permettent l'estimation de la concentration en microorganismes. Elle peut être ponctuelle et dans ce cas on parlera d'estimation NPP, ou bien par intervalles et on parlera de l'intervalle de confiance de la densité vraie inconnue.

#### 2.3.1. La fonction de vraisemblance

Phelps, en 1908, a le premier tenté d'interpréter les résultats de la méthode de dilution des tubes (Pipes, 1977). Il suggérait d'utiliser comme estimation de la concentration bactérienne dans un échantillon l'inverse de la dilution la plus élevée dans laquelle un résultat positif avait été obtenu. Il est clair que ceci n'était pas satisfaisant et la première méthode rigoureuse d'interprétation a été donnée par McCrady (1915). De nombreuses contributions théoriques ont ensuite été apportées (Greenwood & Yule, 1917; Stein, 1919; Reed, 1925; Halvorson & Ziegler, 1933a; Hoskins, 1933).

Soit  $V$  le nombre de volumes unitaires dans l'échantillon à analyser et soit  $\delta$  le nombre de bactéries inconnu dans l'échantillon, dont on cherche une estimation  $\hat{\delta}$ . Si l'échantillon est parfaitement mélangé et si l'erreur de mesure des volumes est négligeable, la probabilité pour

une bactérie donnée parmi les  $\delta$  de ne pas être contenue dans un volume unitaire, prélevé dans l'échantillon s'écrit :

$$P_0 = \frac{V-1}{V}$$

et la probabilité qu'un volume unitaire tiré parmi  $V$  contenant  $\delta$  bactéries soit lui-même dépourvu de bactéries s'écrit donc :

$$P = \left(\frac{V-1}{V}\right)^\delta$$

Le nombre  $n_b$  de bactéries contenu dans un volume unitaire tiré dans  $V$  est une variable aléatoire. Ainsi la probabilité d'obtenir  $n_b$  microorganismes dans un volume unitaire est donnée par  $C_\delta^{n_b} \left(1 - \frac{1}{V}\right)^{\delta-n_b} \left(\frac{1}{V}\right)^{n_b}$ ,  $(n_b+1)$ ième terme de l'expansion du binôme  $\left[\frac{(V-1)+1}{V}\right]^\delta$ .

Soit  $n$  le nombre de volumes prélevés parmi  $V$ , la probabilité d'obtenir  $n_v$  volumes parmi  $n$  contenant au moins une bactérie s'écrit  $C_n^{n_v} (1-P)^{n_v} P^{(n-n_v)}$ , avec  $P$  défini plus haut. Si  $V$  est grand ou, en d'autres termes, si la probabilité pour une bactérie donnée d'être incluse dans le volume unitaire prélevé est faible, l'approximation par une distribution de Poisson est justifiée. Ainsi, si  $V$  est grand, la probabilité d'obtenir un prélèvement stérile, ou un tube ne présentant pas de croissance, s'écrit :

$$P = e^{-\delta/V}$$

La généralisation au cas de plusieurs dilutions s'écrit simplement à partir des notations suivantes :

Soit,

$\delta$  : la densité bactérienne, en nombre de bactéries par unité de volume

$k$  : le nombre de dilutions

$n_i$  : le nombre de tubes pour la  $i^{\text{ème}}$  dilution

$p_i$  : le nombre de tubes positifs pour la  $i^{\text{ème}}$  dilution

$q_i$  : l'inoculum pour la  $i^{\text{ème}}$  dilution

La probabilité d'observer un tube négatif à la dilution  $i$  est  $P = e^{-q_i d}$ . Par conséquent, la probabilité d'obtenir  $p_i$  tubes positifs à chaque dilution  $i$  est :

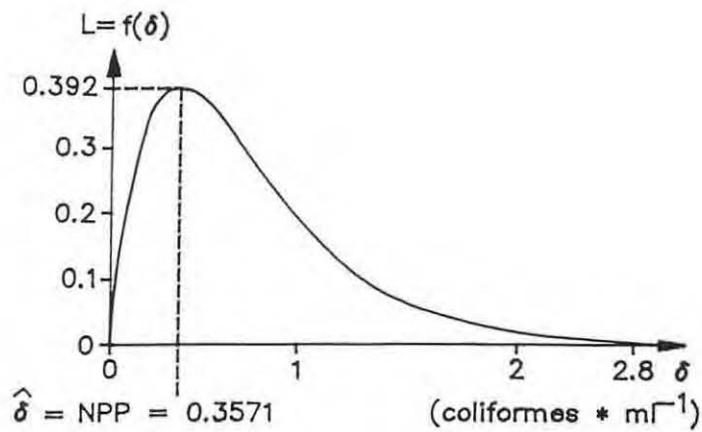
$$L_{\delta} = f(\delta) = \prod_{i=1}^k C_{n_i}^{p_i} (1 - e^{-q_i \delta})^{p_i} (e^{-q_i \delta})^{(n_i - p_i)} \quad (1)$$

Pour une densité réelle donnée, le résultat de l'ensemencement des tubes pour les différentes dilutions de l'échantillon, et donc l'estimation NPP afférente, est une variable aléatoire, dont la probabilité est donnée par la formule (1). On parlera de distribution de probabilité de l'estimation NPP, connaissant  $\delta$ .

Pour un résultat caractéristique donné,  $L_{\delta}$  ou  $f(\delta)$  est la vraisemblance d'avoir obtenu  $p_1, p_2, \dots, p_k$  tubes positifs et donc un résultat caractéristique donné, pour la densité  $\delta$ ; notons que l'ensemble des valeurs de  $f(\delta)$  sur  $\delta$  décrivant  $[0, +\infty]$  ne représente pas une distribution de probabilité. L'estimation ponctuelle de la densité est donnée par la valeur rendant  $L_{\delta}$  maximum (Fig. I.7), ou en d'autres termes rendant le résultat caractéristique **le plus vraisemblable**. McCrady, (1915) a nommé cette densité *Most probable number*, choix plutôt fâcheux puisqu'il ne s'agit pas en effet d'une distribution de probabilité : l'estimateur du Nombre le Plus Probable (NPP) est l'estimateur du maximum de vraisemblance, dont la théorie fut développée plus tard par Fisher (Swaroop, 1938). Il n'a alors pas été rebaptisé, malgré l'essai de Halvorson et Moeglein qui avaient proposé *most likely bacterial density* (El Shaarawi & Pipes, 1982). Cochran (1950) note également que l'adjectif "probable" ne convient pas dans la mesure où il a ici un sens différent de celui donné dans la théorie des probabilités.

### 2.3.2. Estimation ponctuelle

Les hypothèses justifiant l'utilisation de la fonction de vraisemblance sont les suivantes (Cochran, 1950; Taswell, 1981) :



**Fig. I.7** : Vraisemblance  $f(\delta)$  fonction de la concentration en coliformes fécaux ( $\delta$ ) pour le résultat caractéristique 1/0/0 ( $p_1 = 1, p_2 = p_3 = 0; n_1 = n_2 = n_3 = 3; q_3 = q_2/10 = q_1/100 = 0,01$  ml); dans ce cas, l'estimation NPP est égale à  $0,3571$  CF.ml<sup>-1</sup>

- les organismes sont distribués aléatoirement dans l'échantillon,
- tout tube présentant une croissance contient au moins une bactérie et un tube se révélant négatif ne contient pas de germes.

Annuler  $\frac{\partial[\text{Log}(L_\delta)]}{\partial\delta}$  est une condition nécessaire pour rendre  $L_\delta$  maximum (Fig. I.7). Ceci conduit à l'estimation NPP, notée  $\hat{\delta}$ , en résolvant l'équation de vraisemblance :

$$\sum_{i=1}^k \left( \frac{p_i}{1-e^{-q_i\hat{\delta}}} - n_i \right) q_i = 0$$

$\hat{\delta}$  est calculée par la méthode d'itérations successives de Newton dont nous rappelons ici l'algorithme (Press *et al.*, 1986) :

$$\hat{\delta}_{j+1} = \hat{\delta}_j - \left[ \frac{\partial[\text{Log}(L_\delta)]}{\partial\delta} \right]_{\hat{\delta}_j} / \left[ \frac{\partial^2[\text{Log}(L_\delta)]}{\partial\delta^2} \right]_{\hat{\delta}_{j+1}}$$

La mesure de la dispersion des densités autour de l'estimation ponctuelle NPP est donnée par l'expression suivante de la variance asymptotique (Fisher *in* Swaroop, 1938) :

$$\sigma^2 = - \frac{1}{\sqrt{\frac{\partial^2[\text{Log}(L_\delta)]}{\partial\delta^2}}}$$

et est alors estimée par :

$$\hat{\sigma}^2 = \frac{1}{\sqrt{\sum_{i=1}^k \frac{p_i q_i e^{-q_i \hat{\delta}}}{(1-e^{-q_i \hat{\delta}})^2}}}$$

*Cas où l'approximation par la loi de Poisson est non justifiée*

Si les volumes prélevés ne sont pas petits devant l'échantillon, l'approximation par la loi de Poisson n'est plus valable.

Soit,

$V_i$  : volume de l'échantillon pour la  $i^{\text{ème}}$  dilution

$q_i$  : inoculum pour la  $i^{\text{ème}}$  dilution

$r_i$  : rapport de dilution pour la  $i^{\text{ème}}$  dilution

La fonction de vraisemblance devient :

$$L_{\delta} = f(\delta) = \prod_{i=1}^k C_{n_i}^{p_i} \left[ 1 - \left( \frac{V_i - q_i}{V_i} \right)^{\delta V_i r_i} \right]^{p_i} \left[ \left( \frac{V_i - q_i}{V_i} \right)^{\delta V_i r_i} \right]^{(n_i - p_i)}$$

et l'estimation NPP,  $\widehat{\delta}$ , est donnée par la résolution de l'équation de vraisemblance suivante, moyennant les hypothèses précédemment émises :

$$\sum_{i=1}^k \left\{ V_i r_i \text{Log} \left( \frac{V_i - q_i}{V_i} \right) \left[ n_i - \frac{p_i}{1 - \exp \left( \widehat{\delta} V_i r_i \text{Log} \left( \frac{V_i - q_i}{V_i} \right) \right)} \right] \right\} = 0$$

### 2.3.3. Biais

Soit  $f_{R_q}$  la spécification de la fonction de vraisemblance pour une combinaison particulière des nombres de dilutions et de tubes par dilution conduisant au résultat caractéristique  $R_q$ . Etant donné une densité vraie  $\delta$ , chaque résultat caractéristique  $R_q$  (*i.e.* le nombre de tubes positifs par dilution), et par conséquent chaque estimation NPP,  $\widehat{\delta}_{R_q}$ , a une probabilité  $f_{R_q}(\delta)$  d'être obtenue. La somme des produits  $f_{R_q}(\delta) \widehat{\delta}_{R_q}$  est limitée par les résultats extrêmes ("tous les tubes négatifs" et "tous les tubes positifs"), pour lesquels aucune estimation réaliste n'est possible. La distribution de probabilité est ainsi tronquée bilatéralement; néanmoins  $\delta$  sera choisi tel que :

$$\sum_{R_q} f_{R_q}(\delta) \text{ proche de } 1$$

D'où la formule du biais :

$$b = \frac{\sum_{R_q} f_{R_q}(\delta) \widehat{\delta}_{R_q}}{\sum_{R_q} f_{R_q}(\delta)}$$

et celle du biais relatif :

$$b_r = \frac{b}{\delta}$$

Des formules analogues sont obtenues si l'on considère les valeurs après transformation logarithmique (Prost & Hugues, 1982).

#### 2.3.4. précision

La précision de l'estimation NPP a très tôt fait l'objet de nombreux travaux. Citons parmi les précurseurs : Wolman & Weaver (1917) et Halvorson & Ziegler (1933b et 1933c).

La méthode de détermination de limites de confiance préconisée ici s'appuie sur le théorème de Bayes, dont Von Mises (1942) rappelle les fondements théoriques.

Soient  $d_1$  et  $d_2$ , respectivement, les limites de confiance bayésiennes inférieure et supérieure, et  $\alpha_1$  et  $\alpha_2$  les probabilités correspondantes. Alors, l'hypothèse de distribution uniforme de la densité *a priori* étant admise (Hugues, 1981; Fontvieille & Maul, 1985), on peut écrire :

$$\alpha_1 = \frac{\int_0^{d_1} L_\delta d\delta}{\int_0^\infty L_\delta d\delta} \quad \text{et} \quad 1 - \alpha_2 = \frac{\int_0^{d_2} L_\delta d\delta}{\int_0^\infty L_\delta d\delta}$$

La fonction de vraisemblance est intégrable analytiquement sur la densité. Ainsi la probabilité  $\alpha$  pour que la densité réelle  $\delta$  ne dépasse pas une valeur donnée peut se calculer directement, comme le montrent les développements suivants :

La fonction de vraisemblance (1) peut également s'écrire:

$$L_{\delta} = \left( \prod_{i=1}^k C_{n_i}^{p_i} \right) e^{-\left[ \sum_{i=1}^k (n_i - p_i) q_i \right] \delta} \left[ \prod_{i=1}^k (1 - e^{-q_i \delta})^{p_i} \right]$$

Le dernier facteur peut être calculé comme suit, en utilisant la formule du binôme de Newton :

$$\prod_{i=1}^k (1 - e^{-q_i \delta})^{p_i} = \prod_{i=1}^k \left[ \sum_{a_i=0}^{p_i} C_{p_i}^{a_i} (-e^{-q_i \delta})^{a_i} \right]$$

où  $a_i$  est l'indice de sommation pour la  $i$ ème dilution. Pour simplifier l'expression, notons :

$$\sum_{(a_i)=0}^{(p_i)} \equiv \sum_{a_1=0}^{p_1} \sum_{a_2=0}^{p_2} \dots \sum_{a_k=0}^{p_k}$$

Alors,

$$\prod_{i=1}^k (1 - e^{-q_i \delta})^{p_i} = \sum_{(a_i)=0}^{(p_i)} \left[ \left( \prod_{i=1}^k C_{p_i}^{a_i} \right) (-1)^{\left( \sum_{i=1}^k a_i \right)} e^{-\left( \sum_{i=1}^k a_i q_i \right) \delta} \right]$$

Notons  $Q_i = \sum_{i=1}^k (a_i + n_i - p_i) q_i$ , il vient :

$$L_{\delta} = \left( \prod_{i=1}^k C_{n_i}^{p_i} \right) \sum_{(a_i)=0}^{(p_i)} \left[ \left( \prod_{i=1}^k C_{p_i}^{a_i} \right) (-1)^{\left( \sum_{i=1}^k a_i \right)} e^{-Q_i \delta} \right]$$

On peut ensuite déterminer une primitive de  $L_{\delta}$ , soit :

$$\Lambda_{\delta} = \left( \prod_{i=1}^k C_{n_i}^{p_i} \right) \sum_{(a_i)=0}^{(p_i)} \left[ \left( \prod_{i=1}^k C_{p_i}^{a_i} \right) (-1)^{\left( \sum_{i=1}^k a_i \right)} \left( -\frac{e^{-Q_i \delta}}{Q_i} \right) \right]$$

Notons  $\Gamma_{\delta}$ , une autre primitive telle que  $\Gamma_{\delta} = \Lambda_{\delta} - \Lambda_0$ . Soit alors  $\alpha$  la probabilité pour  $\delta$  de ne pas dépasser une valeur donnée  $d_0$  :

$$\alpha = \frac{\int_0^{d_0} L_\delta d\delta}{\int_0^\infty L_\delta d\delta} = \frac{\Gamma_{d_0} - \Gamma_0}{\Gamma_\infty - \Gamma_0} = 1 - \frac{\Lambda_{d_0}}{\Lambda_0}$$

Et finalement,

$$\alpha = 1 - \frac{\sum_{(a_i)=0}^{(p_i)} \left[ \left( \prod_{i=1}^k c_{p_i}^{a_i} \right) (-1)^{\left( \sum_{i=1}^k a_i \right)} \left( -\frac{e^{-Q_i d_0}}{Q_i} \right) \right]}{\sum_{(a_i)=0}^{(p_i)} \left[ \left( \prod_{i=1}^k c_{p_i}^{a_i} \right) (-1)^{\left( \sum_{i=1}^k a_i \right)} \right]} \quad (2)$$

Pour les probabilités  $\alpha_1$  et  $\alpha_2$ , les limites de confiance, respectivement  $d_1$  et  $d_2$ , sont obtenues par la méthode itérative de Newton, déjà présentée plus haut.

### Choix des probabilités $\alpha_1$ et $\alpha_2$

Pour calculer les limites de confiance, les probabilités  $\alpha_1$  et  $\alpha_2$  ont d'abord été prédéterminées avec, classiquement  $\alpha_1 = \alpha_2 = 0,025$ . Il a ensuite été pris en compte le fait que les vraisemblances respectives des limites diffèrent sensiblement, du fait de la dissymétrie de la fonction de vraisemblance. Par conséquent, on cherche les densités  $d_1$  et  $d_2$  de même valeur de vraisemblance  $f_0$ , telles que  $\alpha_1 + \alpha_2 = 0,05$ ;  $d_1$  et  $d_2$  sont les racines de l'équation  $f(d) = f_0$ , conduisant à des bornes inégales de part et d'autre du mode, c'est-à-dire de l'estimation NPP proprement dite. On définit ainsi un critère objectif, dit d'équivraisemblance, pour déterminer les probabilités.

### Cas des résultats extrêmes

Une limite inférieure de densité s'obtient à partir de la formule (2) quand tous les tubes sont négatifs. Naturellement, la limite inférieure est la concentration nulle.

Dans le cas où une croissance bactérienne est révélée dans tous les tubes, la formule ne peut être utilisée dans la mesure où la borne d'intégration est théoriquement infinie. En pratique, l'infini est impossible du fait de l'encombrement stérique de la population bactérienne. On considère alors arbitrairement que la borne d'intégration  $w$  recherchée est telle que  $f(w) = 0,05$ , avec  $p_1 = n_1, p_2 = n_2, \dots, p_k = n_k$ . Ainsi, la probabilité du résultat "tous les tubes sont négatifs" lorsque la vraie densité est  $w$  est de cinq chances sur 100. La limite inférieure de l'intervalle est ensuite déterminée itérativement, par utilisation de la méthode d'intégration numérique des trapèzes (*e.g.* voir l'algorithme dans Press *et al.*, 1986).

#### *2.3.5. Tests de comparaison de deux estimations NPP*

Les méthodes mentionnées dans la littérature s'appuient sur le test paramétrique de différence entre deux moyennes observées, après transformation logarithmique des estimations (Cochran, 1950; Prost & Hugues, 1982). Les conditions requises pour l'utilisation du test sont les suivantes (Prost & Hugues, 1982) :

- le nombre de dilutions et le nombre de tubes par dilution sont identiques; en outre le nombre de tubes doit être supérieur à 20,
- les estimations NPP, indépendantes, ont des logarithmes compris dans la gamme de densités - définie comme l'intervalle  $l_2$  par les auteurs - telle qu'on puisse considérer qu'ils ne proviennent pas d'une distribution s'écartant par trop d'une distribution normale.

Pour  $n$  tubes par dilution, Cochran (1950) donne une expression approchée de l'erreur approchée du logarithme de l'estimation NPP :

$$\widehat{E.S.} = 0,55 \sqrt{\frac{\log(a)}{n}}$$

où  $a$  est le facteur de dilution. Pour  $a = 10$ , l'auteur préconise  $\widehat{E.S.} = \frac{0,58}{\sqrt{n}}$ .

Soit le corps d'hypothèses suivant :

$$\begin{cases} H_0 : \delta_1 = \delta_2 = \delta_0 \\ H_1 : \delta_1 \neq \delta_2 \end{cases}$$

et soient  $\widehat{\delta}_1$  et  $\widehat{\delta}_2$  les estimations NPP respectives des densités réelles  $\delta_1$  et  $\delta_2$ . La statistique considérée est la différence centrée et réduite, supposée distribuée normalement, entre les logarithmes des estimations NPP observées :

$$\varepsilon = \frac{|\log(\widehat{\delta}_1) - \log(\widehat{\delta}_2)|}{\widehat{E.S.}}$$

Cette méthode ne peut être utilisée pour des analyses effectuées à partir d'un faible nombre de tubes, comme 3 ou 5 tubes pour les analyses standards (*e.g.* ISO, 1984). Deux tests sont alors proposés : le rapport des vraisemblances et un test fondé sur la construction de la courbe de probabilité cumulée empirique de la différence entre deux estimations NPP sous l'hypothèse nulle.

#### Test du rapport des vraisemblances

Soient  $\widehat{\delta}_1$  et  $\widehat{\delta}_2$  les estimations NPP respectives des densités réelles  $\delta_1$  et  $\delta_2$  et soient les mêmes hypothèses :

$$\begin{cases} H_0 : \delta_1 = \delta_2 = \delta_0 \\ H_1 : \delta_1 \neq \delta_2 \end{cases}$$

L'estimation  $\widehat{\delta}_0$  commune est obtenue en formant le "nombre caractéristique somme" (Hugues, 1981). Il s'agit d'additionner le nombre de tubes positifs pour les deux résultats aux dilutions correspondantes. Par exemple, pour trois dilutions et trois tubes par dilution, ont été

obtenus les nombres caractéristiques (2/1/0) et (3/1/0). Alors la meilleure estimation commune est obtenue en formant le nombre ou résultat somme (5/2/0) et en calculant l'estimation NPP correspondante, pour trois dilutions et six tubes par dilution. Déjà Halvorson & Ziegler (1934) avaient utilisé ce procédé pour moyenner leurs observations. Les conditions d'utilisation sont la répartition aléatoire des bactéries dans l'échantillon (homogénéité de l'échantillon à analyser) - et l'indépendance des mesures. L'estimation NPP commune nécessite d'avoir les mêmes combinaisons de nombres de dilutions et de tubes par dilution et les mêmes volumes d'inoculum. Dans ce cas l'estimation commune est obtenue à l'aide de la moyenne géométrique des deux densités, donnant des résultats très comparables. Quoiqu'il en soit la même procédure analytique est en général à la base de l'obtention des deux résultats à comparer.

On forme le rapport  $\lambda$  de la vraisemblance de la différence sous  $H_0$  sur la vraisemblance sous  $H_1$ , en remarquant que la vraisemblance de la différence de deux estimations indépendantes est égale au produit des deux vraisemblances :

$$\lambda = \frac{f_1(\widehat{\delta}_0)f_2(\widehat{\delta}_0)}{f_1(\widehat{\delta}_1)f_2(\widehat{\delta}_2)}$$

où  $f_1(\widehat{\delta}_1)$  et  $f_2(\widehat{\delta}_2)$  sont les fonctions de vraisemblance des deux nombres caractéristiques conduisant respectivement aux estimations NPP  $\widehat{\delta}_1$  et  $\widehat{\delta}_2$ .  $\lambda$  est le rapport des maxima de vraisemblance (RMV).

Soit alors le test  $H_0$  contre  $H_1$ , par théorème, sous  $H_0$  la fonction  $-2\text{Log } \lambda$  suit asymptotiquement une loi du  $\chi^2$  à 1 degré de liberté (Tassi, 1989).

Afin de pouvoir utiliser le test, le théorème doit pouvoir être vérifié pour un petit nombre de tubes par dilution, limitant le nombre de résultats caractéristiques possibles. Deux approches différentes, devant conduire à la même conclusion, ont été conduites :

- la construction de la distribution de  $-2\text{Log } \lambda$  sous  $H_0$ , pour une estimation commune  $\widehat{\delta}_0$ , en calculant tous les  $-2\text{Log } \lambda_{ij}$  possibles avec :

$$\lambda_{ij} = \frac{f_i(\widehat{\delta}_0)f_j(\widehat{\delta}_0)}{f_i(\widehat{\delta}_i)f_j(\widehat{\delta}_j)}$$

où  $i$  et  $j$  décrivent l'ensemble des résultats non extrêmes possibles, et où  $f_i$  et  $f_j$  et  $\widehat{\delta}_i$  et  $\widehat{\delta}_j$  sont respectivement les fonctions de vraisemblance et les estimations correspondantes. La probabilité de chaque  $(-2\text{Log } \lambda_{ij})$  sous  $H_0$  est donnée par le numérateur de l'expression ci-dessus. On construit ensuite la courbe de probabilité cumulée.  $\widehat{\delta}_0$  est choisie telle que les probabilités des résultats caractéristiques extrêmes pour cette concentration soient négligeables.

- par simulation : soit la densité  $d_0$  - on se place ainsi sous  $H_0$  - donnée égale à une valeur  $\widehat{\delta}_0$  utilisée précédemment résultant du choix de deux observations. On recrée l'opération de sous-échantillonnage dans l'échantillon à analyser : pour chaque dilution, on génère aléatoirement des nombres de bactéries dans les tubes correspondant, sous l'hypothèse d'une distribution de Poisson de paramètre  $d_0/\text{facteur de dilution}$  (bibliothèque Fortran NAG<sup>1</sup>). Pour deux simulations successives, on obtient deux résultats caractéristiques exprimés en nombres de tubes positifs par dilution et les estimations NPP correspondantes  $\widehat{\delta}_1$  et  $\widehat{\delta}_2$ . On forme :

$$\lambda = \frac{f_1(d_0)f_2(d_0)}{f_1(\widehat{\delta}_1)f_2(\widehat{\delta}_2)}$$

On effectue ainsi 2000 simulations (1000 couples de deux simulations). La courbe de fréquence cumulée de la distribution des  $\lambda$  est alors construite.

Dans les deux cas, le 95ème percentile de la distribution cumulée obtenue est comparé à celui du  $\chi^2$  à 1 degré de liberté (3,84).

---

<sup>1</sup> Numerical Algorithm Group

Test "exact" de comparaison entre deux estimations NPP

Soient  $\widehat{\delta}_1$  et  $\widehat{\delta}_2$  deux estimations dont on se demande si elles proviennent de la même population. On considère le même corps d'hypothèses que précédemment dans le cas bilatéral. L'estimation commune  $\widehat{\delta}_0$  est obtenue de façon similaire.

Le principe du test est le suivant : sous l'hypothèse nulle on construit la distribution de probabilité cumulée de la différence entre deux estimations NPP. Pour une combinaison de nombres de dilutions et de tubes par dilution donnée, toutes les différences possibles sont effectuées, soient N différences avec :

$$N = \left[ \left( \prod_{i=1}^k (n_i + 1) \right) - 2 \right]^2$$

Retrancher 2 au nombre d'estimations possibles est lié à l'existence des résultats extrêmes, pour lesquels, comme déjà souligné, nous ne disposons pas d'estimations réalistes. De la même façon que ci-dessus, on doit donc choisir la gamme de concentration de telle sorte que la distribution de la statistique ne soit pas tronquée.

Le calcul de la probabilité attachée à chaque différence s'effectue comme suit : soient deux résultats caractéristiques i et j, les estimations  $\widehat{\delta}_i$  et  $\widehat{\delta}_j$  étant supposées indépendantes, la probabilité d'observer la différence  $\widehat{\delta}_i - \widehat{\delta}_j$  sous  $H_0$  est égal au produit de la probabilité d'observer le résultat i sous  $H_0$  avec la probabilité d'observer le résultat j sous  $H_0$ . Soit donc,

$$p_{ij} = f_i(\widehat{\delta}_0) f_j(\widehat{\delta}_0)$$

Les N différences sont ordonnées et les probabilités correspondantes sommées. Pour un risque de première espèce donné, on définit ainsi un "intervalle de confiance" de la différence sous l'hypothèse nulle, symétrique et centré sur la différence nulle.

La règle de décision est la suivante : si la différence observée se situe hors de cet intervalle, on rejette l'hypothèse nulle et on conclue à l'existence de deux populations distinctes au risque  $\alpha$  de se tromper; en d'autres termes, sous  $H_0$ , il y avait moins de  $100\alpha\%$  de chances pour qu'on observe une telle différence sous l'hypothèse nulle, alors rejetée.

Dans le cas unilatéral, on a le corps d'hypothèse suivant :

$$\begin{cases} H_0 : \delta_1 = \delta_2 = \delta_0 \\ H_1 : \delta_1 < \delta_2 \end{cases}$$

La valeur critique de différence est alors donnée par le fractile à  $100\alpha\%$  de la distribution de probabilité cumulée.

L'appellation test "exact" vient uniquement de l'analogie existant entre le calcul d'une fréquence théorique pour un tableau de contingence 2x2 avec le calcul de la probabilité d'une différence sous  $H_0$ . Rappelons que le *Fisher exact probability test* (Siegel, 1956) permet de connaître la signification d'un ensemble de fréquences observées dans une table 2x2. Hormis le procédé de calcul des probabilités, les deux tests diffèrent sensiblement. Le test de Fisher est non paramétrique en ce sens qu'il ne fait aucunement appel à une quelconque loi de probabilité. Pour notre test, la distribution de probabilité des résultats caractéristiques sous  $H_0$  est conditionnée par le mode de répartition des bactéries dans l'échantillon, caractérisée, par exemple, par le paramètre  $\lambda$  de la loi de Poisson.

### Programmes informatiques

Le *Nomogram* (outil permettant la détermination graphique de l'estimation NPP) de Dalla Valle (1941) a vécu et depuis Parnow (1972) le NPP a fait l'objet de nombreux programmes informatiques (e.g. Wyshak & Detre, 1972; Russek & Colwell, 1983). La liste des différents programmes informatiques utilisés pour notre propos et leur tâche respective est fournie en annexe. Ils ont été écrits en Fortran V et implémentés d'une part sur une station Sun (O.S. 4.0)

et d'autre part sur un CYBER 992 (Centre de calcul de Brest). Le programme de détermination des intervalles de confiance bayésiens non symétriques en probabilité a été écrit en Turbo-Pascal 5.0.

## 2.4. Stratégies d'échantillonnage

### 2.4.1. Introduction

Les variations spatiales ou temporelles du descripteur, à différentes échelles, ne pourront être perceptibles et interprétées comme résultant de l'action de tel ou tel facteur que si leur amplitude est compatible avec le degré de précision connu, ou d'ordre de grandeur connu. Ainsi, tout protocole d'échantillonnage afférent à une problématique donnée doit intégrer cette notion, dans un souci d'efficacité, qu'il s'agisse d'étudier la cinétique de contamination ou de décontamination des coquillages *in vitro* ou bien d'évaluer l'impact bactériologique des rejets d'une station d'épuration d'eau en milieu marin.

En d'autres termes on doit être attentif à ce que l'imprécision du dénombrement ne vienne pas diminuer de façon trop sensible la résolution de la perception d'un phénomène dans l'espace ou dans le temps.

Le fil conducteur de toute étude bactériologique sur le terrain est la recherche des paramètres de la distribution de la population étudiée, qu'ils s'agissent d'estimations de moyennes, de fractiles, ou de variance. Seul un plan d'échantillonnage rigoureux permet l'inférence sur ces paramètres. Une liste des principaux plans utilisables dans le domaine étudié est proposée ci-dessous. Les avantages et inconvénients, généraux ou spécifiques, des différentes techniques de sondage sont décrits et les estimateurs respectifs explicités.

### 2.4.2. Echantillonnage aléatoire simple (E.A.S.)

#### Définitions et estimateurs

L'échantillonnage aléatoire simple est une méthode qui consiste à prélever au hasard et de façon indépendante  $n$  unités d'échantillonnage d'une population de  $N$  éléments,  $N$  pouvant tendre vers l'infini. Chaque élément de la population possède la même probabilité de faire partie

Soit  $(y_1, y_2, \dots, y_i, \dots, y_n)$  un échantillon sans répétition et  $f = \frac{n}{N}$  le taux d'échantillonnage,

alors  $\hat{\mu} = \frac{\sum_{i=1}^n y_i}{n}$  est une estimation non biaisée de la moyenne de la population  $\mu = \frac{\sum_{i=1}^N y_i}{N}$ , et

$\hat{V}(\hat{\mu}) = \frac{s^2}{n} (1 - f)$  est une estimation non biaisée de la variance de l'estimateur, avec

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{\mu})^2}{n - 1}.$$

Dès lors que  $N$  est grand et que  $n$  est petit devant  $N$ , l'estimation de la variance de l'estimateur pour un plan sans remise est considérée égale à celle pour un plan avec remise; ceci est en pratique très souvent le cas.

#### Avantages, inconvénients et contraintes

Aucune information *a priori* n'est requise pour l'utilisation d'un tel plan, dont la mise en oeuvre se révèle simple dès lors que la liste des éléments de la population est disponible ou facile à dresser. Pratiquement, cette dernière condition n'est que peu fréquemment réalisée dans les études bactériologiques sur le terrain : au nombre de prélèvements unitaires réalisables sur des aires représentant parfois des dizaines de km<sup>2</sup>, s'ajoute le nombre de combinaisons possibles de constitutions d'un échantillon unité. Si la zone d'étude est suffisamment réduite pour permettre un dénombrement exhaustif des unités d'échantillonnage, ce qui représente néanmoins une contrainte opérationnelle très importante, on s'expose à l'existence de phénomènes d'autocorrélation et par suite d'une redondance trop importante de l'information à l'issue du tirage au sort. A grande échelle, l'échantillonnage par degrés lui sera préféré.

### 2.4.3. Echantillonnage à plusieurs niveaux

#### Définition

Dans ce type de plan, dérivé de l'échantillonnage par grappes, chacune des  $N$  unités de la population, aussi appelées unités primaires ou grappes, se compose de  $M_i$  sous-unités plus petites, aussi appelées unités secondaires, qui peuvent elles-mêmes comporter  $K_{ij}$  unités tertiaires, et ainsi de suite... A chaque niveau un échantillonnage aléatoire peut être effectué. Si un niveau seulement est échantillonné, on parle de sous-échantillonnage, ou d'échantillonnage du premier degré, et pour trois niveaux d'un échantillonnage du troisième degré.

L'échantillonnage du premier degré consiste donc à prélever de façon aléatoire  $n$  unités primaires ou grappes parmi les  $N$  de la population et à mesurer les  $M_i$  sous-unités, c'est-à-dire toutes les unités secondaires des grappes sélectionnées.

L'échantillonnage à plusieurs niveaux constitue une extension du principe du sous-échantillonnage. Ainsi, par exemple, au troisième degré,  $n$  unités primaires sont sélectionnées de façon aléatoire parmi les  $N$  de la population,  $m_i$  unités secondaires sont prélevées des  $M_i$  composant chaque unité primaire sélectionnée, enfin  $k_{ij}$  unités tertiaires sont retenues aléatoirement parmi les  $K_{ij}$  composant chaque unité secondaire prélevée.

#### Estimateurs d'un plan d'échantillonnage du troisième degré pour un descripteur quantitatif

Soient,

$N$  : effectif de la population (nombre total d'unités primaires ou prélèvements réalisables)

$n$  : nombre d'unités primaires composant l'échantillon

$M_i$  : nombre d'unités secondaires composant la  $i^{\text{ème}}$  unité primaire

$m_i$  : nombre d'unités secondaires échantillonnées dans la  $i^{\text{ème}}$  unité primaire

$f = f_1 = \frac{n}{N}$  : taux d'échantillonnage des unités primaires

$f_{2i} = \frac{m_i}{M_i}$  : taux d'échantillonnage des unités secondaires dans la  $i^{\text{ème}}$  unité primaire

$K_{ij}$  : nombre total d'unités tertiaires composant la  $j^{\text{ème}}$  unité secondaire de la  $i^{\text{ème}}$  unité primaire

$k_{ij}$  : nombre d'unités tertiaires échantillonnées dans la  $j^{\text{ème}}$  unité secondaire de la  $i^{\text{ème}}$  unité primaire

$f_{3ij} = \frac{k_{ij}}{K_{ij}}$  : taux d'échantillonnage des unités tertiaires dans la  $j^{\text{ème}}$  unité secondaire de la  $i^{\text{ème}}$  unité primaire :

Soient les notations suivantes :

*Moyennes aux niveaux*

<i>primaire</i>	<i>secondaire</i>	<i>Observations</i>
$\bar{y}_1$	$\bar{y}_{11}$ $\bar{y}_{1j}$ $\bar{y}_{1m_1}$	$Y_{111}, \dots, Y_{11u}, \dots, Y_{11k_{11}}$ $Y_{1j1}, \dots, Y_{1ju}, \dots, Y_{1jk_{11}}$ $Y_{1m_11}, \dots, Y_{1m_1u}, \dots, Y_{1m_1k_{11}}$
$\bar{y}_i$	$\bar{y}_{i1}$ $\bar{y}_{ij}$ $\bar{y}_{im_i}$	$Y_{i11}, \dots, Y_{i1u}, \dots, Y_{ik_{i1}}$ $Y_{ij1}, \dots, Y_{iju}, \dots, Y_{ijk_{ij}}$ $Y_{im_i1}, \dots, Y_{im_iu}, \dots, Y_{im_ik_{im_i}}$
$\bar{y}_n$	$\bar{y}_{n1}$ $\bar{y}_{nj}$ $\bar{y}_{nm_n}$	$Y_{n11}, \dots, Y_{n1u}, \dots, Y_{nk_{n1}}$ $Y_{nj1}, \dots, Y_{nju}, \dots, Y_{njk_{nj}}$ $Y_{nm_n1}, \dots, Y_{nm_nu}, \dots, Y_{njk_{nm_n}}$

On notera  $\bar{y}$  comme l'estimateur et l'estimation de  $\mu$ , moyenne par élément dans la population :  $\bar{y} = \hat{\mu}$ .

Pour des unités primaires de même taille ( $M_i = C^{te}$ ,  $f_{2i} = C^{te}$ ) et des unités secondaires de même taille ( $K_{ij} = C^{te}$ ,  $f_{3ij} = C^{te}$ ), les résultats suivants sont obtenus (Cochran, 1977) :

$$\bar{\bar{y}} = \frac{\sum_{i=1}^n \sum_{j=1}^m \sum_{u=1}^k y_{iju}}{nmk} : \text{estimation non biaisée de la moyenne par élément } [E(\bar{\bar{y}}) = \mu].$$

$$\hat{V}(\bar{\bar{y}}) = \frac{1-f_1}{n} s_1^2 + \frac{f_1(1-f_2)}{nm} s_2^2 + \frac{f_1 f_2 (1-f_3)}{nmk} s_3^2 : \text{estimation non biaisée de la variance de}$$

$\bar{\bar{y}}$  où,

$$s_1^2 = \frac{\sum_{i=1}^n (\bar{y}_i - \bar{\bar{y}})^2}{n-1}$$

$$s_2^2 = \frac{\sum_{i=1}^n \sum_{j=1}^m (\bar{y}_{ij} - \bar{y}_i)^2}{n(m-1)}$$

$$s_3^2 = \frac{\sum_{i=1}^n \sum_{j=1}^m \sum_{u=1}^k (y_{iju} - \bar{y}_{ij})^2}{nm(k-1)}$$

Cochran (1977) indique que pour un taux faible  $f_1$  d'échantillonnage des unités primaires, le terme de variance entre unités primaires prend une part essentielle dans le calcul de  $\hat{V}(\bar{\bar{y}})$ , qui s'écrit alors :

$$\hat{V}(\bar{\bar{y}}) \approx \frac{s_1^2}{n} = \frac{\sum_{i=1}^n (\bar{y}_i - \bar{\bar{y}})^2}{n(n-1)}$$

### Optimisation du plan d'échantillonnage (troisième degré)

Soient  $C$  le coût total de l'opération d'échantillonnage ( $C_0$ , coût forfaitaire indépendant de l'acte d'échantillonnage, non compris) et  $c_1$ ,  $c_2$  et  $c_3$  les coûts respectivement relatifs aux éléments, grappes, et supergrappes. Pour une fonction de coût de la forme :

$$C = c_1 n + c_2 nm + c_3 nmk,$$

les valeurs optimales de  $k$  et de  $m$  s'écrivent :

$$k_{\text{opt}} = \frac{s_3}{\sqrt{s_2^2 - s_3^2 / k'}} \sqrt{c_2 / c_3} \quad \text{et} \quad m_{\text{opt}} = \frac{\sqrt{s_2^2 - s_3^2 / k'}}{\sqrt{s_1^2 - s_2^2 / m'}} \sqrt{c_1 / c_2} \quad \text{où } m' \text{ et } k' \text{ sont}$$

respectivement le nombre de grappes et le nombre d'éléments par grappe d'un prééchantillon ou d'un échantillon prélevé lors d'une étude antérieure (Scherrer, 1983).

#### Avantages, inconvénients et contraintes

Scherrer (1983) note que l'échantillonnage par degré s'impose lorsqu'il est impossible d'inventorier tous les éléments de la population et qu'il est possible d'énumérer les éléments de grappes sélectionnées de façon aléatoire. Plus généralement, peuvent être envisagées des procédures d'échantillonnage à plusieurs niveaux dès l'instant qu'à chaque niveau (sauf au dernier) les unités d'échantillonnage sont trop vastes pour être étudiées en totalité. Ceci rend l'utilisation de ce plan aléatoire tout à fait justifié pour le prélèvement unitaire réalisé pour obtenir une valeur du descripteur (concentration en coliformes fécaux pour 100 ml de broyat de chair et de liquide intervalvaire). Le nombre de niveaux et la taille des unités aux différents niveaux sont fonctions des dimensions de l'aire d'échantillonnage; le coût total est fonction, en outre, du taux d'échantillonnage à chaque degré.

Le protocole final est un compromis entre ces trois paramètres. Des contraintes viennent guider le choix dans le cas de prélèvements de coquillages dans l'espace : l'échelle spatiale représentant le dernier niveau doit être suffisamment réduite pour que l'on puisse considérer que le prélèvement effectué est ponctuel. Un exemple d'application est donné ultérieurement (voir enquête pilote), où l'aire au troisième niveau est voisine du m<sup>2</sup>. Pour des raisons de simplicité dans les formules et calculs, il conviendrait de s'arrêter au troisième niveau; le nombre d'unités pour un niveau est alors dépendant de l'échelle spatiale correspondante.

#### 2.4.4. Echantillonnage composite

##### Définition

L'échantillonnage composite consiste à prélever aléatoirement (*e.g.* E.A.S. ou par degrés)  $J$  éléments parmi  $N$  possibles puis à sélectionner  $K$  échantillons à partir de chaque élément et enfin à mélanger les  $JK$  le plus parfaitement possible. Cet échantillon composite peut alors être divisé en  $R$  sous-échantillons dont  $r$  sont tirés au sort. Une mesure est effectuée à partir des  $r$  unités d'échantillonnage obtenues, par opposition à la démarche classique où chaque élément est mesuré.

##### Estimateurs

Soit le modèle :

$$X_{jk} = \mu + b_j + e_{jk}$$

où  $X_{jk}$  est l'observation du  $k^{\text{ième}}$  échantillon, issu du  $j^{\text{ième}}$  élément, parmi  $J$ ,  $\{e_{ij}\}$  est un ensemble de variables aléatoires indépendantes de moyenne zéro et de variance  $\sigma_e^2$ , appelée variance intra-élément. On suppose également que :

$$\sum_{j=1}^N b_j = 0 \text{ et } \frac{\left( \sum_{j=1}^N b_j^2 \right)}{N} = \sigma_b^2, \text{ variance inter-éléments}$$

De plus,

$$\hat{\mu} = \frac{\left( \sum_{i=1}^r \sum_{j=1}^J \sum_{k=1}^K \alpha_{ijk} X_{jk} \right)}{r}$$

où  $\{\alpha_{ijk}\}$  est un ensemble de variables aléatoires donnant la proportion du  $(j,k)^{\text{ième}}$  échantillon, incorporée dans le  $i^{\text{ième}}$  sous-échantillon, supposées indépendantes de  $\{X_{jk}\}$ , avec :  $E(\alpha_{ijk}) = \mu_\alpha$  et  $V(\alpha_{ijk}) = \sigma_\alpha^2$

Une estimation de  $\mu$  est donnée par la moyenne des  $r$  valeurs obtenues pour les sous-échantillons composites  $y_1, y_2, \dots, y_r$  :

$$\hat{\mu} = \frac{\sum_{i=1}^r y_i}{r}$$

Brown et Fisher (1972) et Rohde (1976) donnent l'expression d'une estimation de la variance de la moyenne de l'échantillon composite. On détermine ainsi la précision de  $\hat{\mu}$ , estimateur de la moyenne dans la population. En considérant que l'homogénéité du mélange conduisant à l'échantillon est parfaite ( $\sigma_a^2 = 0$ ) et que  $N$  tend vers l'infini, l'estimation s'écrit :

$$\hat{V}(\hat{\mu}) = \frac{1}{J} \left( \hat{\sigma}_b^2 + \frac{\hat{\sigma}_e^2}{K} \right)$$

La variance peut alors être estimée si l'on connaît des estimations de  $\sigma_b^2$  et de  $\sigma_e^2$ .

### Avantages et inconvénients

On conçoit donc qu'un tel plan puisse être particulièrement rentable si le coût analytique est important, ce qui est le cas, par exemple, pour la mesure des concentrations métalliques dans les moules. Une telle stratégie est néanmoins à déconseiller si l'objectif n'est pas tant le niveau moyen de concentration de tel ou tel polluant que la recherche des maxima sur la zone étudiée. En outre, comme le fait remarquer Gilbert (1987), même si le mélange est parfait l'hétérogénéité des échantillons initialement tirés au sort se retrouve dans les échantillons à mesurer finalement.

#### *2.4.5. Echantillonnage systématique et krigeage*

### Définition

L'échantillonnage systématique est une technique qui consiste à tirer au hasard un  $i^{\text{ème}}$  élément, situé entre le premier et le  $p^{\text{ème}}$  de la population puis à prélever systématiquement le  $(i+p)^{\text{ème}}$ ,  $(i+2p)^{\text{ème}}$ , ...,  $(i+(n-1)p)^{\text{ème}}$  élément de la population. Les rangs des  $n$  unités sont

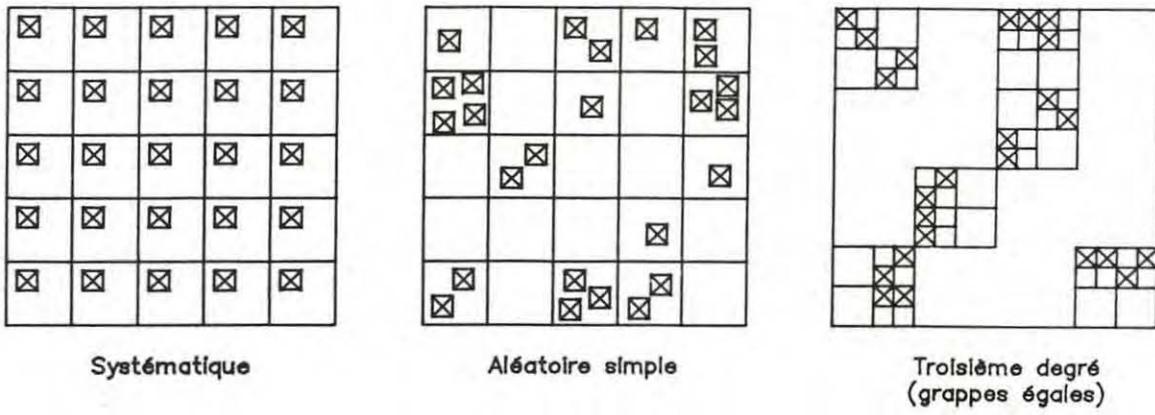
ainsi en progression arithmétique dont la base est un nombre aléatoire  $i$  et la raison un nombre  $p$  calculé de telle sorte que l'échantillon se répartisse uniformément sur toute la population (voir figure I.8 pour une représentation schématique des différents plans utilisés). Contrairement à l'E.A.S., les unités ne sont pas prélevées de façon indépendante puisque le choix du premier élément détermine la composition de tout l'échantillon (Scherrer, 1983). Pour la dimension temporelle, l'échantillonnage systématique est assuré par la régularité des prélèvements, donc à fréquence constante.

Dans le cas de prélèvements de coquillages, on ne peut inventorier de façon exhaustive l'ensemble des unités d'échantillonnage. Mais on peut toujours appliquer un maillage sur la zone étudiée, qu'elle corresponde à un transect linéaire ou à un plan, et effectuer un prélèvement aux noeuds.

L'échantillonnage systématique est un plan souvent utilisé quand on ne dispose d'aucune information *a priori*. Il se révélera néfaste dès lors que le phénomène étudié, que ce soit dans le temps ou dans l'espace, est en phase avec un autre phénomène (Scherrer, 1983). Par exemple, des mesures en coliformes fécaux sur des prélèvements de coquillages effectués toujours en vives-eaux conduiront à une sur-estimation importante de la contamination si la remise en suspension de sédiments réservoirs de germes est exacerbée à cette période. Inversement, aux périodes de mortes-eaux la faible contamination subséquente pourrait être dangereuse. Le seul cas ne conduisant pas à un biais est celui où  $p$  est un entier impair multiple de la demi-période du phénomène (Cochran, 1977). Par exemple, on pourrait envisager d'échantillonner systématiquement aux mortes-eaux et aux vives-eaux.

### Estimateurs

Dans le cas d'éléments disposés de façon aléatoire au sein d'une population, les estimateurs du plan systématique sont identiques à ceux de l'aléatoire simple. Gros (1988) souligne qu'un tel protocole ne permet pas d'obtenir une estimation réaliste de la précision des résultats :



**Fig. I.8 :** Présentation des plans d'échantillonnage utilisés

considérant qu'un maillage forme une grappe spatialement discontinue et constitue donc la réalisation unique de l'acte d'échantillonnage, on ne peut alors estimer la variance inter-grappes. Ceci rend théoriquement le calcul des estimateurs impossible. Un recours est fourni par l'estimateur du krigeage, outil de la géostatistique, dont sont donnés quelques résultats théoriques ci après.

La méthode d'interpolation du krigeage utilise la théorie probabiliste des fonctions aléatoires (F.A.). L'ensemble des valeurs du descripteur est interprété comme une réalisation d'une F.A.  $Z(x)$ , où  $Z(x)$  peut être considérée comme un ensemble de variables aléatoires  $Z(x_i)$  définies en chaque point  $x_i$  du domaine  $D$  étudié :  $Z(x) = \{Z(x_i), \forall x_i \in D\}$  (Journel & Huijbregts, 1978).

Cette réalisation est unique pour une campagne donnée et constitue un obstacle à l'inférence statistique, c'est-à-dire à l'estimation des premiers moments de la F.A., ce qui conduirait à réfuter tout intérêt opérationnel à cette technique (Journel & Huijbregts, 1978).

Ainsi des hypothèses doivent être émises sur la F.A., hypothèses devant être suffisamment faibles pour couvrir le plus grand nombre de problèmes pratiques possible (Delhomme, 1976).

Au sens strict, une F.A. est dite stationnaire lorsque sa distribution spatiale est invariante par translation. L'inférence statistique est alors possible en remplaçant les moyennes, que l'on est normalement amené à faire sur l'ensemble des réalisations disponibles, par des moyennes spatiales sur l'unique réalisation.

Moins contraignante est la stationnarité d'ordre 2 ou stationnarité faible, qui requiert l'invariance des deux premiers moments :

$$E[Z(x)] = m, \forall x;$$

$$E[(Z(x+h)-m)(Z(x)-m)] = C(h), \text{ ne dépendant que de la distance } h$$

Néanmoins, le développement des calculs d'estimation optimale n'exige pas que la F.A. elle-même mais seulement ses accroissements soient stationnaires d'ordre 2. Ce cas est encore moins restrictif puisqu'alors l'existence de la covariance n'est plus exigée. Cette hypothèse, qualifiée d'"intrinsèque", requiert les conditions suivantes :

- l'espérance mathématique de la F.A. existe et ne dépend pas du point d'appui  $x$ ,

$$E[Z(x)] = m \text{ et donc } E[Z(x+h)-Z(x)] = 0, \forall x;$$

- pour tout vecteur  $h$ , l'accroissement  $[Z(x+h)-Z(x)]$  a une variance finie, qui ne dépend pas du point d'appui  $x$ ,

$$\text{Var}[Z(x+h)-Z(x)] = E\{[Z(x+h)-Z(x)]^2\} = 2\gamma(h), \forall x.$$

La fonction  $\gamma(h)$  est le demi-variogramme, encore appelée "variogramme" par abus de langage; elle constitue un résumé quantitatif de l'information structurale disponible (Journel & Huijbregts, 1978). Le demi-variogramme est estimé par le calcul du demi-variogramme expérimental :

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} (z(x_i+h)-z(x_i))^2$$

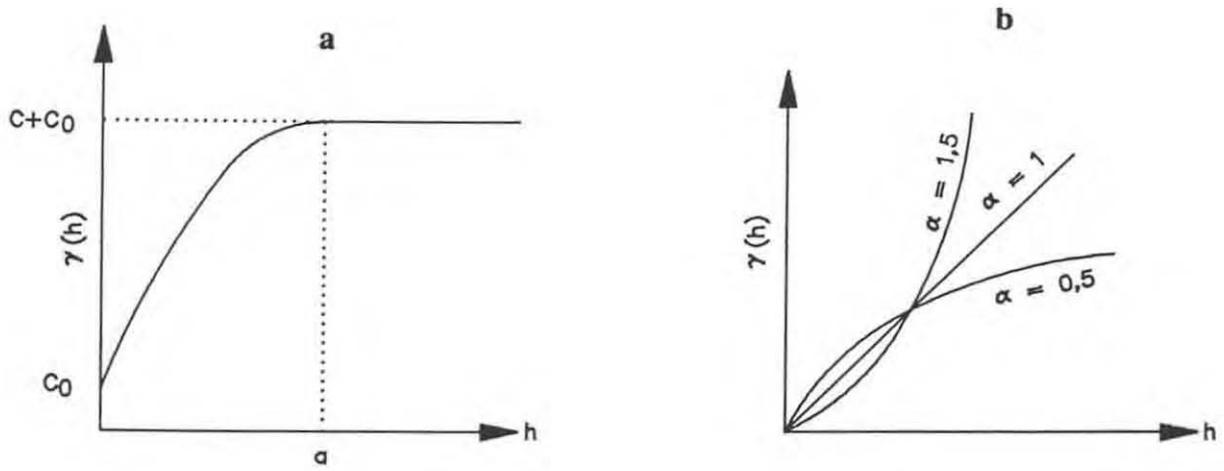
où  $z(x_i)$  représente la valeur du descripteur observée au point  $i$ , et  $n(h)$  le nombre de couples de points distants de  $h$ . Un variogramme théorique peut être ajusté à partir des valeurs du variogramme expérimental, pour chaque pas ou classe de distance.

Les modèles théoriques de variogramme utilisés pour notre propos sont :

Le modèle sphérique (Fig. 1.9a) :

$$\gamma(h) = c_0 + c \left[ \frac{3h}{2a} - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right] \text{ pour } 0 < h \leq a$$

$$\gamma(h) = c_0 + c \quad \text{pour } h > a$$



**Fig. 1.9** : Présentation des deux types de variogrammes utilisés (a : modèle sphérique; b : modèle en  $h^\alpha$ )

Trois paramètres interviennent dans ce variogramme :

- $c_0$  constitue la "pépite" du phénomène. Cette discontinuité à l'origine signifie qu'il existe une part de variabilité non expliquée par la fonction de structure, et attribuable soit à une microrégionalisation d'échelle très inférieure à l'espacement des données, soit à des erreurs de mesure, soit encore à des erreurs dans le positionnement des points de prélèvement (Delhomme, 1976; Cressie, 1988). Des prélèvements supplémentaires effectués localement permettent la comparaison de la pépite extrapolée à partir du variogramme à la mesure de la variabilité locale.

- Au delà d'une certaine distance, le variogramme atteint une valeur limite  $\gamma^{(\infty)}$  (Journel & Huijbregts, 1978), représentée ici par la quantité  $(c_0 + c)$ , et appelée "palier". Cette valeur est la variance entre points ne s'influençant pas mutuellement (Gros & Hamon, 1989).

- $a$  est la "portée" du phénomène, c'est-à-dire la distance pour laquelle le variogramme atteint le palier. La portée renseigne sur l'extension de la zone d'influence d'un point expérimental (Journel & Huijbregts, 1978).

#### Le modèle en $h^\alpha$ (Fig. 1.9b)

Il s'agit dans ce cas d'un variogramme non borné, d'équation  $\gamma(h) = wh^\alpha$  ( $0 < \alpha < 2$ ) auquel il est également possible d'adjoindre une pépite, de signification strictement équivalente à celle décrite pour le modèle sphérique.

D'autres types de variogramme existent et le lecteur intéressé pourra se reporter utilement à Journel & Huijbregts (1978).

Après ajustement d'un modèle idoine, les estimations sont obtenues par krigeage ponctuel aux noeuds d'un maillage régulier. Ce maillage est à définir *a posteriori* à partir de sa propre connaissance du phénomène, et de l'allure du variogramme.

L'estimateur du krigeage ponctuel s'écrit comme une combinaison linéaire de  $\alpha$  observations. Au point  $x_0$ ,  $z(x_0)$  est estimé par

$$\hat{z}(x_0) = \sum_{\alpha} \lambda_{\alpha} z(x_{\alpha})$$

où l'ensemble des points  $\alpha$  constitue l'ensemble informant,  $x_0$  représentant le point à informer.

La variance d'estimation locale  $E[z - \hat{z}]^2$  est minimisée sous la contrainte de non-biais,  $E[z - \hat{z}] = 0$  et l'expression générale de la variance de krigeage s'écrit à l'aide des notations précédentes :

$$\sigma_k^2 = 2 \sum_{\alpha} \lambda_{\alpha} \gamma(x_{\alpha}, x_0) - \gamma(0) - \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} \gamma(x_{\alpha}, x_{\beta})$$

La qualité de l'estimation dépend de trois géométries (Petitgas & Poulard, 1989) :

- la structure de la variable  $\gamma(h)$ ,
- la localisation relative du voisinage considéré parmi l'ensemble des voisinages  $\{\gamma(x_{\alpha}, x_{\beta})\}$ .
- la localisation relative du point  $x_0$  parmi l'ensemble des points constituant son voisinage  $\{\gamma(x_{\alpha}, x_0)\}$ . Une extrémité du vecteur  $h$  est fixée au point  $x_0$ , l'autre décrivant le voisinage.

Cette formulation exprime en particulier que les zones où l'effort d'échantillonnage est moins bien réparti sont moins bien informées, et présentent des variances de krigeage ponctuel plus importantes. Par exemple, le calcul de la variance de krigeage en un point situé à la périphérie de la zone s'appuiera sur des distances importantes entre les  $x_{\alpha}$  et  $x_0$  pour un même nombre de points  $x_{\alpha}$ , conduisant à des termes  $\gamma(x_{\alpha}, x_0)$  élevés et à une estimation imprécise *in fine*.

II. PRECISION DE L'ESTIMATION DE LA  
CONCENTRATION EN COLIFORMES FECAUX

## 1. Estimation de la densité bactérienne en routine

### 1.1. Introduction

En routine, dans le contexte du réseau microbiologique, les estimations NPP sont obtenues pour trois tubes par dilution décimale et le plus souvent trois dilutions.

L'utilisation d'un faible nombre de tubes est imputable à diverses contraintes :

- Le budget de fonctionnement du réseau : environ 10000 analyses sont réalisées annuellement pour ce réseau.
- La capacité opérationnelle des analystes ( en temps et en matériel).
- L'existence de textes réglementaires fixant cette technique d'analyse, en matière d'appréciation de la qualité bactériologique du milieu mesurée dans les coquillages.

L'optimisation de la stratégie de surveillance passe au préalable par la connaissance des propriétés statistiques de l'estimateur utilisé, en particulier le biais et la précision, puis par la recherche de leur amélioration.

Ceci sera le plus souvent illustré au travers de la confrontation des résultats obtenus pour trois tubes par dilution et cinq tubes par dilution.

### 1.2. Estimation ponctuelle et biais

#### 1.2.1. Comparaison des estimations (modèle de Poisson vs. distribution binomiale)

Le tableau II.1 (resp. II.2) donne les estimations NPP en utilisant le modèle de Poisson (§ I.2.3.2, form. (1)), pour trois dilutions et trois tubes par dilution (resp. pour 3 dilutions et 5 tubes par dilution). Les résultats NPP en utilisant la distribution binomiale (§ I.2.3.2, form. (2)) sont exprimés comme une proportion de l'estimation NPP, obtenue à partir du

Tab. II.1 : Estimation NPP (Distribution de Poisson), vraisemblance, limites de confiance (Intervalle bayésien symétrique en probabilité),  $R$  rapport des estimations NPP (Distribution binomiale<sup>a</sup> sur distribution de Poisson) dans le cas ( $n_1 = n_2 = n_3 = 3$ ;  $q_3 = q_2/10 = q_1/100 = 0,01$  ml)

Résultat Caractéristique	NPP	Vraisemblance	Limites de confiance		$R$
			95%	99%	
000	/	/	< 0,9	< 1,4	/
001	0,3	0,003	0,07 - 1,7	0,03 - 2,2	0,986
010	0,3	0,034	0,07 - 1,7	0,03 - 2,3	0,986
020	0,6	0,002	0,34 - 2,7	0,21 - 3,4	0,987
100	0,4	0,392	0,09 - 2,1	0,04 - 2,8	0,985
101	0,7	0,006	0,22 - 2,7	0,12 - 3,5	0,985
110	0,7	0,065	0,22 - 2,8	0,12 - 3,6	0,985
111	1,1	0,002	0,41 - 3,4	0,25 - 4,3	0,985
120	1,1	0,006	0,41 - 3,5	0,25 - 4,4	0,986
121	1,5	$0,3 \cdot 10^{-3}$	0,62 - 4,1	0,41 - 5,1	0,986
130	1,6	$0,3 \cdot 10^{-3}$	0,63 - 4,2	0,42 - 5,2	0,987
200	0,9	0,319	0,29 - 3,0	0,33 - 6,4	0,984
211	2,0	0,006	0,82 - 6,0	0,54 - 7,6	0,983
220	2,1	0,023	0,85 - 6,3	0,56 - 7,9	0,984
221	2,8	0,002	1,2 - 7,4	0,82 - 9,2	0,984
230	2,9	0,002	1,2 - 7,7	0,84 - 9,6	0,985
300	2,3	0,341	0,87 - 13	0,52 - 18	0,974
301	3,9	0,031	1,5 - 18	0,95 - 23	0,964
302	6,4	0,002	2,4 - 23	1,6 - 29	0,953
310	4,3	0,374	1,7 - 21	1,0 - 29	0,951
311	7,5	0,066	2,8 - 28	1,8 - 36	0,949
312	12	0,007	4,4 - 35	2,9 - 44	0,950
320	9,3	0,328	3,3 - 39	2,1 - 51	0,949
321	15	0,125	5,6 - 51	3,6 - 65	0,949
322	21	0,025	8,6 - 64	5,7 - 81	0,949
323	29	0,002	13 - 79	8,6 - 99	0,949
330	24	0,366	9,1 - 140	5,5 - 190	0,949
331	46	0,428	18 - 240	11 - 320	0,949
332	110	0,444	38 - 480	24 - 640	0,949
333	/	/	> 111	> 67	

\*: Modèle binomial:  $V_1 = 250$  ml,  $V_2 = V_3 = 50$  ml.

Tab. II.2 : Estimation NPP (Distribution de Poisson), vraisemblance, limites de confiance (Intervalle bayésien symétrique en probabilité),  $R$  rapport des estimations NPP (Distribution binomiale<sup>a</sup> sur distribution de Poisson) dans le cas ( $n_1 = n_2 = n_3 = 5$ ;  $q_3 = q_2/10 = q_1/100 = 0,01$  ml)

Résultat Caractéristique	NPP	Vraisemblance	Limites de confiance		$R$
			95%	99%	
000	/	/	< 0,5	< 0,8	/
001	0,2	0,003	0,04 - 1,0	0,02 - 1,3	0,986
010	0,2	0,033	0,04 - 1,0	0,02 - 1,4	0,986
011	0,4	0,4 10 <sup>-3</sup>	0,11 - 1,3	0,06 - 1,7	0,986
020	0,4	0,002	0,11 - 1,3	0,06 - 1,7	0,986
100	0,2	0,365	0,05 - 1,1	0,02 - 1,5	0,985
101	0,4	0,005	0,12 - 1,5	0,07 - 1,9	0,985
110	0,4	0,055	0,13 - 1,5	0,07 - 1,9	0,986
111	0,6	0,001	0,22 - 1,8	0,14 - 2,3	0,986
120	0,6	0,006	0,22 - 1,8	0,14 - 2,3	0,986
121	0,8	0,3 10 <sup>-3</sup>	0,33 - 2,1	0,22 - 2,6	0,986
130	0,8	0,4 10 <sup>-3</sup>	0,34 - 2,2	0,22 - 2,7	0,986
200	0,4	0,266	0,14 - 1,7	0,08 - 2,2	0,985
201	0,7	0,007	0,25 - 2,0	0,15 - 2,6	0,985
210	0,7	0,77	0,25 - 2,1	0,15 - 2,6	0,985
211	0,9	0,003	0,37 - 2,4	0,25 - 3,0	0,985
220	0,9	0,013	0,38 - 2,5	0,25 - 3,1	0,985
221	1,2	0,7 10 <sup>-3</sup>	0,52 - 2,8	0,36 - 3,5	0,985
230	1,2	0,001	0,52 - 2,9	0,36 - 3,5	0,986
300	0,8	0,217	0,28 - 2,4	0,17 - 3,1	0,984
301	1,1	0,010	0,43 - 2,9	0,28 - 3,6	0,983
302	1,3	0,2 10 <sup>-3</sup>	0,59 - 3,4	0,41 - 4,2	0,983
310	1,1	0,105	0,43 - 3,0	0,29 - 3,7	0,984
311	1,4	0,006	0,60 - 3,4	0,41 - 4,3	0,984
312	1,7	0,2 10 <sup>-3</sup>	0,78 - 3,9	0,56 - 4,8	0,984
320	1,4	0,027	0,61 - 3,5	0,42 - 4,3	0,984
321	1,7	0,002	0,79 - 4,0	0,57 - 4,9	0,984
330	1,7	0,005	0,80 - 4,1	0,57 - 5,0	0,985
331	2,1	0,4 10 <sup>-3</sup>	1,0 - 4,6	0,74 - 5,6	0,985
340	2,1	0,5 10 <sup>-3</sup>	1,0 - 4,7	0,75 - 5,7	0,985
400	1,3	0,087	0,52 - 3,8	0,34 - 4,9	0,982
401	1,7	0,014	0,72 - 4,5	0,50 - 5,7	0,981
402	2,1	0,5 10 <sup>-3</sup>	0,95 - 5,3	0,68 - 6,6	0,980
410	1,7	0,148	0,73 - 4,7	0,51 - 5,9	0,981
411	2,1	0,014	0,97 - 5,5	0,69 - 6,8	0,981
412	2,6	0,7 10 <sup>-3</sup>	1,2 - 6,3	0,90 - 7,8	0,980
420	2,2	0,063	0,99 - 5,6	0,71 - 7,0	0,981
421	2,6	0,008	1,3 - 6,5	0,92 - 8,0	0,981
422	3,2	0,4 10 <sup>-3</sup>	1,6 - 7,4	1,2 - 9,1	0,980
430	2,7	0,017	1,3 - 6,7	0,94 - 9,3	0,982
431	3,3	0,003	1,6 - 7,7	1,2 - 9,4	0,981
432	3,9	0,2 10 <sup>-3</sup>	1,9 - 8,7	1,5 - 11	0,980
440	3,4	0,003	1,6 - 7,9	1,2 - 9,7	0,982
441	4,0	0,6 10 <sup>-3</sup>	2,0 - 9,0	1,5 - 11	0,981

450	4,1	$0,3 \cdot 10^{-3}$	2,1 - 9,3	1,6 - 11	0,982
500	2,3	0,166	1,0 - 8,7	0,68 - 12	0,974
501	3,1	0,023	1,4 - 11	0,98 - 14	0,969
502	4,3	0,002	1,9 - 14	1,4 - 18	0,962
510	3,3	0,264	1,5 - 12	1,0 - 16	0,969
511	4,6	0,052	2,0 - 15	1,4 - 20	0,961
512	6,3	0,006	2,8 - 18	2,6 - 27	0,953
520	4,9	0,260	2,2 - 17	1,5 - 22	0,959
521	7,0	0,079	3,0 - 21	2,1 - 27	0,952
522	9,4	0,013	4,1 - 25	2,9 - 31	0,950
523	12	0,002	5,4 - 29	3,9 - 36	0,949
530	7,9	0,226	3,3 - 25	2,3 - 32	0,951
531	11	0,110	4,6 - 30	3,2 - 38	0,949
532	14	0,029	6,2 - 36	4,4 - 44	0,949
533	17	0,005	8,1 - 41	5,9 - 51	0,949
534	21	$0,5 \cdot 10^{-3}$	10 - 48	7,6 - 58	0,949
540	13	0,199	5,4 - 39	3,7 - 50	0,949
541	17	0,161	7,5 - 48	5,3 - 61	0,949
542	22	0,070	10 - 58	7,2 - 72	0,949
543	28	0,020	13 - 69	9,7 - 86	0,949
544	35	0,004	17 - 82	13 - 101	0,949
550	24	0,187	10 - 94	7,1 - 126	0,949
551	35	0,312	15 - 130	11 - 180	0,949
552	54	0,337	23 - 190	16 - 250	0,949
553	92	0,345	37 - 300	26 - 380	0,949
554	160	0,410	66 - 530	44 - 690	0,949
555	/	/	> 150	> 100	/

\*: Modèle binomial:  $V_1 = 250$  ml,  $V_2 = V_3 = 50$  ml

modèle de Poisson. Ce rapport varie de 0,949 à 0,986. pour 3 (Tab. II.1) et 5 tubes (Tab. II.2).

### 1.2.2. Biais

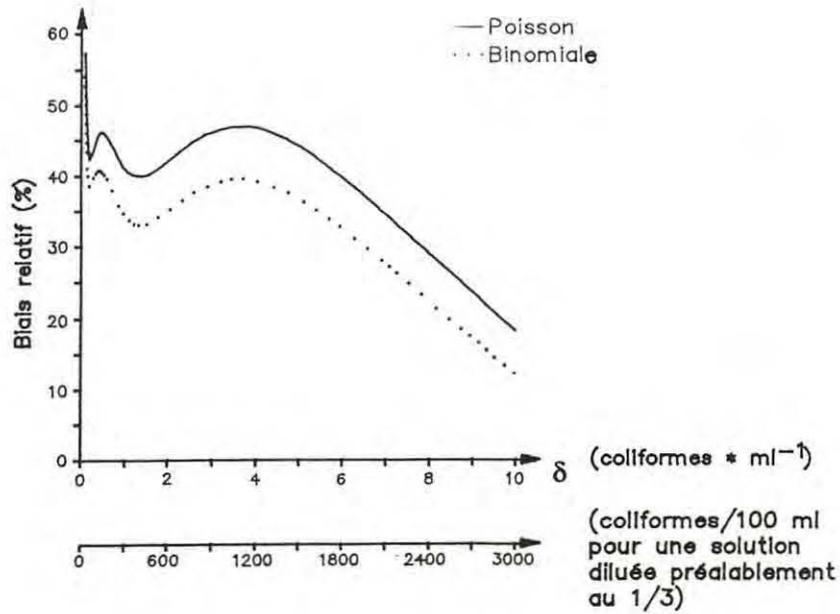
Le biais relatif est exprimé en fonction de la densité bactérienne pour trois dilutions, avec  $n = 3$  et  $n = 5$  en utilisant le modèle de Poisson (Fig. II.1) et pour trois dilutions et trois tubes par dilution, en utilisant le modèle binomial et le modèle de Poisson (Fig. II.2). Pour ces figures, l'axe des abscisses est limité par la densité  $\delta$ , telle que  $f(\delta) = 0,05$  dans le cas où tous les tubes sont positifs, ( $\forall i \in \{1,2,3\}, p_i = n_i$ ). Pour trois tubes par dilution, le biais relatif maximum (46,9%), si l'on ne tient pas compte des valeurs de densité au voisinage de 0, est obtenu pour  $\delta = 3,65$  coliformes.ml<sup>-1</sup>. Dans la gamme de concentration considérée ([0 - 2790] en CF/100 ml pour une solution préalablement diluée au 1/3, soit [0 - 9,3] en coliformes.ml<sup>-1</sup>), le biais relatif est supérieur ou égal à 20%.

### 1.3. Estimation par intervalles

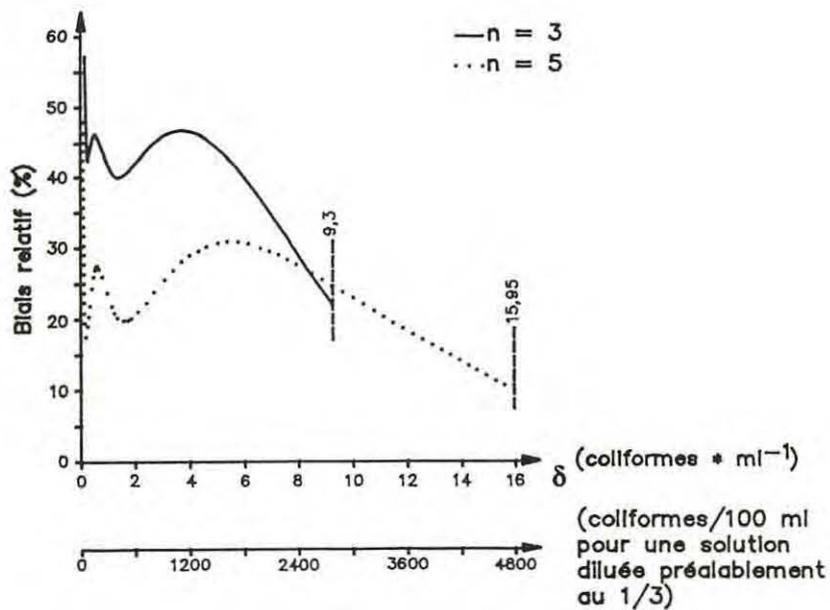
Les intervalles de confiance bayésiens symétriques en probabilité, à 95% et 99%, ont été calculés pour trois dilutions et trois tubes et trois dilutions et cinq tubes (Tab. II.1 et II.2). La liste des nombres caractéristiques proposée est identique à celle fournie par De Man (1977), selon un critère fondé sur les probabilités cumulées pour cinq mesures indépendantes.

Le tableau II.3 présente les résultats relatifs aux intervalles de confiance bayésiens non symétriques en probabilités. Le critère d'équivraisemblance conduit à des intervalles plus étroits que les intervalles bayésiens symétriques en probabilité. La détermination des limites de confiance est illustrée pour le résultat caractéristique 1/0/0, pour trois tubes par dilution (Fig. II.3).

Les densités limites ( $w$ ) nécessaires à l'intégration numérique dans le cas où tous les tubes sont positifs (résultats '333' et '555'), sont respectivement 407,7 et 457,5 coliformes.ml<sup>-1</sup>.



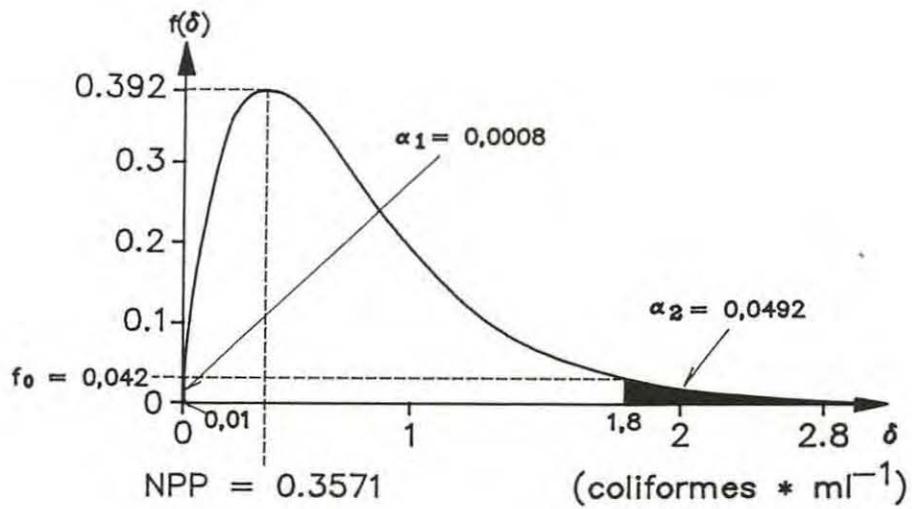
**Fig. II.1 :** Biais relatif (%) de l'estimation NPP (distribution de Poisson) en fonction de la distribution utilisée, Poisson ou binomiale) dans le cas de trois tubes par dilution ( $q_3 = q_2/10 = q_1/100 = 0,05$  ml)



**Fig. II.2 :** Biais relatif (%) de l'estimation NPP (distribution de Poisson) en fonction de la concentration dans le cas de trois et cinq tubes par dilution ( $q_3 = q_2/10 = q_1/100 = 0,05$  ml)

Tab. II.3 : Limites de confiance à 95% pour les intervalles bayésiens symétriques en probabilité et non-symétriques en probabilité avec le critère d'équivraisemblance, probabilités correspondantes des limites inférieure et supérieure ( $\text{lim}_{\text{inf}}$  et  $\text{lim}_{\text{sup}}$ ), et vraisemblance  $f_0$  (valeur de vraisemblance commune pour les deux limites de confiance) pour  $n_1 = n_2 = n_3 = 3$  et  $q_3 = q_2/10 = q_1/100$  ml.

Résultat Caract.	Interv. Sym.		Interv. non Sym.		Prob. corresp.		$f_0$
	$\text{lim}_{\text{inf}}$	$\text{lim}_{\text{sup}}$	$\text{lim}_{\text{inf}}$	$\text{lim}_{\text{sup}}$	< $\text{lim}_{\text{inf}}$	> $\text{lim}_{\text{sup}}$	
000		<0,9		< 0,9	0,0000	0,0500	/
001	0,07	1,7	0,01	1,4	0,0009	0,0491	0,0004
010	0,07	1,7	0,01	1,4	0,0009	0,0491	0,004
020	0,34	2,7	0,09	2,0	0,0037	0,0463	0,0002
100	0,09	2,1	0,01	1,8	0,0008	0,0492	0,042
101	0,22	2,7	0,11	2,4	0,0034	0,0466	0,064
110	0,22	2,8	0,11	2,4	0,0034	0,0466	0,008
111	0,41	3,4	0,26	3,1	0,0056	0,0444	0,0002
120	0,41	3,5	0,26	3,1	0,0056	0,0444	0,0008
121	0,62	4,1	0,45	3,8	0,0072	0,0428	0,00003
130	0,63	4,2	0,46	3,8	0,0072	0,0428	0,00004
200	0,29	3,8	0,13	3,3	0,0027	0,0473	0,035
201	0,52	4,8	0,31	4,3	0,0045	0,0455	0,001
210	0,53	5,0	0,32	4,4	0,0045	0,0455	0,014
211	0,82	6,0	0,56	5,4	0,0059	0,0441	0,0008
220	0,85	6,3	0,58	5,6	0,0058	0,0442	0,003
221	1,2	7,4	0,88	6,8	0,0070	0,0430	0,0002
230	1,2	7,7	0,91	7,0	0,0069	0,0431	0,0003
300	0,87	13	0,38	10,8	0,0017	0,0482	0,028
301	1,5	18	0,81	15	0,0026	0,0474	0,003
302	2,4	23	2,5	20	0,0039	0,0461	0,0002
310	0,7	21	0,84	18	0,0022	0,0478	0,037
311	2,8	28	1,6	25	0,0034	0,0466	0,008
312	1,4	35	2,9	32	0,0051	0,0449	0,0008
320	3,3	39	1,8	34	0,0026	0,0474	0,036
321	5,6	51	3,4	45	0,0040	0,0460	0,015
322	8,6	64	5,9	58	0,0056	0,0444	0,003
323	13	79	9,2	72	0,0068	0,0432	0,0003
330	9,1	140	4,0	117	0,0015	0,0485	0,029
331	18	240	8,8	207	0,0020	0,0480	0,042
332	38	480	19	413	0,0023	0,0477	0,047
333	> 111		> 111			< 0,05	/



**Fig. II.3** : Détermination des limites de l'intervalle bayésien non symétrique en probabilité pour le résultat caractéristique 1/0/0 ( $q_3 = q_2/10 = q_1/100 = 0,01$  ml)

## 1.4. Discussion

### 1.4.1. Estimation ponctuelle et biais

Il convient de souligner au préalable avec d'autres auteurs (Pipes, 1977; Maul *et al.*, 1981; Block *et al.*, 1982) qu'à la méthode analytique utilisée ne correspond pas qu'une seule technique d'interprétation numérique. Ainsi, Taswell (1981), dénombrant des cellules immunocompétentes, préconise l'estimateur du  $\chi^2$  minimum fournissant des estimations moins biaisées et de variances comparables par rapport à l'estimateur du maximum de vraisemblance (résultats obtenus par simulation).

Les estimations NPP obtenues (Tab. II.1 et II.2) sont très proches de celles de De Man (1975, 1977), avec de légères différences dues probablement à des erreurs d'arrondi. L'utilisation de la distribution binomiale conduit à des estimations inférieures, comme le montre le rapport *R*. Une différence maximale de 5% a pu être notée entre les deux estimations, ce qui peut être considéré comme négligeable en termes de précision d'estimation NPP.

Sous l'hypothèse de répartition aléatoire des bactéries dans l'échantillon à analyser, le modèle de Poisson n'est alors pas à remettre en question avec les volumes utilisés au laboratoire pour les analyses de routine.

L'estimateur NPP conduit à une surestimation du titre bactérien. Comme déjà noté par Prost & Hugues (1982), le biais est d'autant plus grand que le nombre de tubes est réduit (Fig. II.2). L'estimateur du "Nombre le Plus Probable" est asymptotiquement non biaisé, ce qui n'est pas toujours le cas (Tassi, 1989). Prost et Hugues précisent également que la transformation logarithmique diminue le biais, comme toute opération visant à rendre plus symétrique la distribution de la variable NPP.

Le biais relatif n'apparaît pas constant avec l'augmentation de la concentration, pour un nombre donné de tubes par dilution et pour une gamme de concentration donnée. Ceci va à l'encontre de ce qu'ont suggéré Mc Carthy *et al.* (1958). Les auteurs indiquent un facteur

correctif du biais ne dépendant que du nombre (constant) de tubes par dilution :  $e^{-\frac{0,805}{n}}$ . Par exemple, pour  $n = 3$  tubes, on trouve que pour une estimation de 258 cellules pour 100 ml de broyat, l'estimation corrigée est :  $258 * 0,765 = 197$  cellules/ 100 ml, ce qui correspondrait à un biais relatif de 31%. Le calcul montre qu'il est proche de 46%, soit un facteur correctif de 0,685. On peut considérer qu'à l'échelle de l'imprécision de l'estimation le biais relatif est constant et de l'ordre de 40% pour 3 tubes et trois dilutions.

Par ailleurs, l'utilisation de la distribution binomiale, au lieu du modèle de Poisson, conduit à un biais relatif inférieur (Fig. II.1). La réduction du biais apparaît constante, de 5 à 8%, en fonction de la densité. Ici encore cette amélioration n'est pas significative.

Dans le cas où la répartition des bactéries dans l'échantillon à analyser n'est plus aléatoire mais contagieuse, *a contrario* l'estimation NPP à partir du modèle de Poisson sous-jacent est une sous-estimation de la concentration réelle, sous-estimation d'autant plus grave que l'agrégation des bactéries est un phénomène plus marqué. Il est plausible que ceci soit réalisé dans le cas d'un dénombrement dans les coquillages compte tenu de la nature fortement particulière de l'échantillon, propice à l'adsorption.

#### 1.4.2. Précision

La détermination des intervalles de confiance obtenus dans les tables est fondée sur le théorème de Bayes. Mac Crady (1915) a utilisé ce théorème dans sa forme discrète pour définir la probabilité qu'une vraie densité avait d'être inférieure à une valeur donnée. Greenwood et Yule (1917) ont généralisé ce principe à plusieurs dilutions et plusieurs tubes par dilution. L'application de cette méthode suggère que toutes les densités de 0 à une limite supérieure  $w$  sont également probables, ou en d'autres termes, que la distribution de probabilité *a priori* de la concentration est uniforme. D'autres auteurs (Von Mises, 1942; Fontvieille & Maul, 1985; Prost & Hugues, 1982) ont décrit et utilisé l'hypothèse d'équiprobabilité *a priori*. Les intervalles obtenus par De Man (1977) diffèrent légèrement de ceux calculés ici. Ces différences

procédure d'intégration numérique par rapport à l'utilisation de la formulation analytique de l'intégrale de la fonction de vraisemblance, pour le modèle de Poisson.

Pour le résultat "tous les tubes sont négatifs" ('000'), le calcul donne le même intervalle, pour  $\alpha = 0,05$ , que Woodward (1957) pour trois tubes par dilution. Quand tous les tubes sont positifs, une limite inférieure approximative est donnée après intégration numérique, pour un seuil d'intégration arbitrairement fixé à la valeur de concentration qui a  $p = 95\%$  de chances d'être dépassée dans ce cas. Cette limite est bien sûr sensible à  $p$ , pour  $f(w) = p$ . Pour illustrer ce point, pour  $p = 0,9$  la limite inférieure était  $100 \text{ coliformes.ml}^{-1}$  (au lieu de 111) avec  $w = 336,7 \text{ coliformes.ml}^{-1}$  (au lieu de 407,7). Il s'agit donc d'une limite tout à fait arbitraire, ce qui donne *in fine* un intervalle arbitraire. Ce type de problème ne devrait pas être posé si les dilutions appropriées ont été appliquées à l'échantillon.

En pratique, l'information *a priori* sur la contamination de l'échantillon n'est pas toujours disponible. Et même quand elle l'est, la procédure de laboratoire pour des analyses de routine est souvent un compromis entre un petit nombre de dilutions, et un pari raisonnable qui ne devrait pas conduire trop souvent au résultat "tous les tubes sont positifs".

De Man fut critiqué parce qu'il avait précédemment calculé des intervalles bayésiens (De Man, 1975 & 1977). On peut effectivement s'interroger sur le bien fondé de l'utilisation des statistiques bayésiennes quand la distribution *a priori* est choisie uniforme par manque de connaissance ! On trouvera dans Eisenhart & Wilson (1943) une discussion sur ce point. De Man (1983) proposa de nouveaux intervalles, cette fois obtenus par la méthode des intervalles de confiance *stricto sensu* (Kendall & Stuart, 1961), comme l'avait réalisé auparavant Von Mises (1942) et Woodward (1957), ne prenant en compte la fréquence relative des évènements. Von Mises (1942) montre que dans ce cas, si l'on doit statuer sur les chances d'avoir raison en effectuant une assertion du type " $\delta_1 < \delta < \delta_2$ , si tel résultat caractéristique est obtenu", on doit de toute façon recourir à une hypothèse de distribution *a priori* de la densité dans l'échantillon.

Il est alors difficile d'établir que la première méthode est meilleure que la seconde, ou *vice-versa*. Comme Kendall et Stuart l'ont souligné, "*the two different methods are consciously or unconsciously either answering different questions or resting on different postulates*". En choisissant une gamme de concentration, l'analyste doit se référer souvent à une grossière connaissance *a priori* du niveau de contamination dans l'échantillon. Cependant, une fois ce choix fait, il est supposé n'avoir aucune idée de la vraie densité dans la suspension à analyser et l'hypothèse d'uniformité de la distribution *a priori* paraît légitime.

Notons enfin que Loyer & Hamilton (1984) ont défini un troisième type d'intervalle, *Sterne-type intervals*, dont le calcul se rapproche de celui des *confidence intervals*. Les auteurs indiquent que le désavantage majeur de cette méthode est la non-possibilité de fournir un intervalle de confiance pour les résultats caractéristiques peu vraisemblables.

Le critère d'équivraisemblance, proposé dans ce travail, fournit des intervalles bayésiens plus étroits que les intervalles bayésiens symétriques en probabilité (Tab. II.3). En effet, la limite supérieure est bien plus décalée vers la gauche que la limite inférieure ne l'est vers la droite, du fait de la dissymétrie droite de la fonction de vraisemblance (Fig. II.3). Le résultat est une forte dissymétrie entre les deux probabilités extrêmes correspondant en moyenne à une probabilité à droite de 4,6% : la restriction de l'intervalle entraîne le rejet des fortes densités considérées comme peu vraisemblables.

Bien que ce critère conduise à des limites moins arbitraires et des densités plus vraisemblables, il devrait être noté qu'il donne un intervalle moins "sûr" dans le contexte de la protection de la santé humaine. En se restreignant, l'intervalle rejette les fortes densités, considérées alors comme "improbables".

## 2. Choix d'une combinaison de tubes et de dilutions

### 2.1. Choix des dilutions

Les dilutions décimales nous sont imposées par le respect des normes en matière de dénombrement des coliformes fécaux en milieu liquide (*e.g.* ISO<sup>1</sup>, 1984). Notons cependant que divers auteurs mentionnent l'augmentation de précision consécutive à la diminution du facteur de dilution (Cochran, 1950; Cowell & Morisetti, 1969), tous les autres paramètres étant fixés par ailleurs. En effet, dans le cas de la dilution décimale la majeure partie de l'information se trouve obtenue pour une dilution, alors que dans le cas d'un facteur de dilution égal à 2 l'information se répartit sur plusieurs dilutions. Naturellement cette augmentation de précision s'accompagne d'une diminution de la gamme de concentration "explorable" : on se résigne à obtenir plus souvent des résultats extrêmes, donc sans possibilité d'estimation NPP, et particulièrement en routine où pour l'estimation de la contamination sur certains zones montrant de fortes variabilités dans le temps, l'analyste effectue d'emblée une à deux dilutions décimales supplémentaires.

Or les tables sont données pour trois dilutions décimales de l'échantillon. Quelle est alors l'interprétation numérique du résultat caractéristique obtenu *in fine* ? La norme ISO-7251 (1984), concernant le dénombrement d'*Escherichia coli* présumée, à l'aide de la "technique du Nombre le Plus Probable" fixe une règle de choix des trois dilutions à retenir dans le cas où plus de trois dilutions décimales sont effectuées. Le but de ce paragraphe est d'attirer l'attention du lecteur sur le fait que la règle est erronée. Deux exemples de la norme sont repris à dessein en comparant les valeurs suggérées par la règle, d'une part à la valeur "vraie" obtenue par estimation en tenant compte de toutes les dilutions, et d'autre part à l'estimation pour trois dilutions décimales à partir du choix correct des dilutions (Tab. II.4).

---

<sup>1</sup> International Standardization Organization

Tab. II.4 : Critiques de cas empruntés à la norme ISO-7251 (1984)

Cas	Au plus deux tubes positifs par dilution	Trois tubes négatifs pour au moins deux dilutions
Règle du choix de trois dilutions (ISO-7251)	Choisir les trois dilutions les plus élevées	Retenir la dilution la moins forte sans tubes positifs et les deux moins fortes précédentes
Résultat caractéristique	2/2/1/1/0	2/2/1/0/0
Estimation NPP* - règle ISO (RC)	7,4 (?/?/1/1/0)	14,7 (?/2/1/0/?)
Estimation NPP* "vraie" (RC)	0,35 (2/2/1/1/0)	2,757... (2/2/1/0/0)
Estimation NPP* correcte pour trois dilutions (RC)	0,28 (2/2/1/?/?)	2,763... (2/2/1/?/?)

\* : en CF.ml<sup>-1</sup>

A l'évidence la règle n'est pas satisfaisante : le rapport entre l'estimation NPP à laquelle conduit la norme et la vraie estimation en considérant toutes les dilutions est proche de 21 pour le premier cas, et supérieur à 5 pour le deuxième. Il s'agit d'erreurs pour le moins non négligeables, qu'il convient d'analyser plus profondément.

Le principe de base à l'élaboration d'une règle concernant le choix le plus judicieux de trois dilutions pourrait être énoncé comme ceci : "en choisissant trois dilutions parmi un nombre de dilutions supérieur, on perd de l'information; le meilleur choix sera celui qui en fait perdre le moins".

Considérons ainsi le premier cas : l'information importante - qu'il ne faut pas perdre ! - est que chacune des deux dilutions les moins fortes ne présente une série de tubes tous positifs; ainsi l'interprétation numérique sera satisfaisante si l'on a pris en compte le fait qu'il est beaucoup plus important de savoir qu'on a obtenu deux tubes positifs parmi trois tubes pour les deux premières dilutions que de connaître le nombre de tubes positifs pour la dilution la plus forte, qui est sans doute nul, voire égal à 1, au regard des quatre premiers résultats (2/2/1/1/...). En termes de vraisemblance, obtenir deux tubes positifs pour les deux premières dilutions est

beaucoup moins probable que d'en obtenir trois : c'est la raison pour laquelle il faut le noter; par contre, obtenir autre chose que 0 pour la dilution la plus forte est très peu vraisemblable (Tab. II.5). En termes cette fois d'estimation, le calcul montre bien l'importance de la prise en compte des deux premières dilutions sur l'interprétation numérique du résultat caractéristique.

**Tab. II.5** : Sensibilité de l'estimation NPP aux résultats caractéristiques voisins. NPP : estimation du Nombre le Plus Probable en CF.ml<sup>-1</sup>; L : vraisemblance correspondante.

Dilutions					NPP	L
10 <sup>0</sup>	10 <sup>-1</sup>	10 <sup>-2</sup>	10 <sup>-3</sup>	10 <sup>-4</sup>		
2	2	1	1	0	3,5 *	4 10 <sup>-5</sup>
3	2	1	1	0	20,4	6 10 <sup>-3</sup>
3	3	1	1	0	74,1	6 10 <sup>-2</sup>
2	2	1	0	0	2,8	2 10 <sup>-3</sup>
2	2	1	0	1	3,5	< 10 <sup>-5</sup>
2	2	1	1	1	4,2	< 10 <sup>-5</sup>
2	2	1	2	0	4,2	< 10 <sup>-5</sup>
2	2	1	2	1	5,0	< 10 <sup>-5</sup>

\* : Estimation de référence

L'étude du deuxième exemple (Tab. II.4) repose rigoureusement sur le même raisonnement. Par conséquent, il n'y a pas lieu de distinguer le deuxième cas du premier.

La solution préconisée par la norme pour le résultat caractéristique 3/3/0/0 est également fautive, mais prête beaucoup moins à conséquence quantitativement. La norme préconise de retenir le résultat 3/3/0 (la dilution la moins forte ne présentant aucun tube positif et les deux moins fortes précédentes) pour une estimation NPP correspondante de 24 CF.ml<sup>-1</sup>. L'estimation "vraie" et l'estimation pour le résultat caractéristique 3/0/0 sont égales (23,1 CF.ml<sup>-1</sup>), à l'arrondi près. En effet, la deuxième série étant toute positive, il est beaucoup moins vraisemblable d'observer un autre résultat que celui observé pour la première série, que d'avoir un résultat autre que 0 pour la dernière série. Par conséquent, le fait d'avoir

obtenu 0 pour la dernière dilution est une information légèrement plus importante que celle d'avoir obtenu 3 pour la première; selon le principe de base, il faut retenir le résultat de la plus forte dilution.

La norme fait état d'une dernière règle : quand une dilution au moins comporte trois tubes positifs, le texte précise que la dilution plus élevée comportant trois tubes doit être retenue, ainsi que les deux dilutions les plus fortes consécutives. Notons ici que l'exemple précédent s'inscrit également dans ce cas et cette fois la règle conduit au bon résultat. Cette règle peut encore être prise en défaut. Soit le résultat 3/3/2/1/0, la règle préconise de conserver le résultat 3/2/1; or implicitement conserver le résultat 2/1/0 exprime que l'on a bel et bien obtenu tous les tubes positifs pour les deux premières dilutions et c'est donc le résultat caractéristique qu'il faut conserver.

Enfin pour être tout à fait complet, il subsiste le cas où une dilution moins forte que celle présentant trois tubes positifs, ne présente pas elle-même trois tubes positifs. Imaginons le résultat suivant : 3/2/3/1/0. En faisant abstraction des éventuels biais analytiques, conduisant à des aberrations dans le nombre de tubes par dilution, le résultat obtenu n'est pas totalement saugrenu : la vraisemblance de l'estimation NPP correspondante s'établit à  $2 \cdot 10^{-4}$ .

D'après la norme, on devrait conserver le résultat 3/1/0 pour une estimation NPP de 427,3 CF.ml<sup>-1</sup>. L'estimation "vraie" donne 36,0 et le résultat 2/3/1 également. Le résultat 3/2/3 donne 29,2 CF.ml<sup>-1</sup>. Ainsi, le choix du résultat 2/3/1 doit être retenu. En conservant 3/1/0, on présuppose arbitrairement que les deux dilutions les moins fortes fournissent trois tubes positifs, entraînant la surestimation importante observée ( $\approx 1200\%$ ), et donc que les deux tubes positifs de la deuxième série sont peu crédibles. Faut-il pour autant attribuer un crédit plus large aux trois tubes de la dilution suivante ? Le choix de 2/3/1 prend en compte ce flou, en fournissant de surcroît l'information de l'avant-dernière dilution.

Cet examen long mais exhaustif de la norme nous permet de voir se dessiner une règle particulièrement simple et courte que nous énonçons ci-après. Cette règle est généralisable à un nombre de tubes (ou de cupules dans le cas des micro-plaques) supérieur par dilution, et donc par exemple à 5 tubes par dilution.

**Une dilution au moins fournit trois tubes positifs**

*Sélectionner la dilution la moins forte ne fournissant pas trois tubes positifs et les deux suivantes. Dans le cas où cette dilution ne fournit aucun tube positif, sélectionner la dilution précédente et les deux suivantes.*

3/3/2/1/0 -> retenir 2/1/0

3/2/3/1/0 -> retenir 2/3/1

3/3/3/0/0 -> retenir 3/0/0

**Aucune dilution ne fournit trois tubes positifs**

*Sélectionner les trois dilutions les moins fortes*

2/2/1/1/0 et 2/2/1/0/0 -> retenir 2/2/1

La solution la plus satisfaisante est bien sûr de regarder dans les tables NPP correspondant aux nombres de dilutions correspondants. Celles-ci pourront être obtenues sur simple demande auprès de l'auteur sur disque magnétique ou sur papier. Les programmes permettant de construire ces tables sont également disponibles (annexe I.4).

## 2.2. Comparaison entre deux résultats NPP

### 2.2.1. Test du rapport des vraisemblances

Préalablement à l'utilisation du test paramétrique du rapport des vraisemblances, l'adéquation de la distribution du rapport des maxima de vraisemblance (RMV) à un  $\chi^2$  à 1 degré de liberté doit pouvoir être vérifiée. Deux approches différentes ont été conduites et explicitées par ailleurs (§ I.2.3.5). Le fractile correspondant à la probabilité 0,05 d'être dépassés est de 5,89 pour la courbe "vraie" (Fig. II.4) et la valeur obtenue par simulation est de 6,88, écart notable dû aux faibles nombres de simulation par classes de valeurs de rapport.

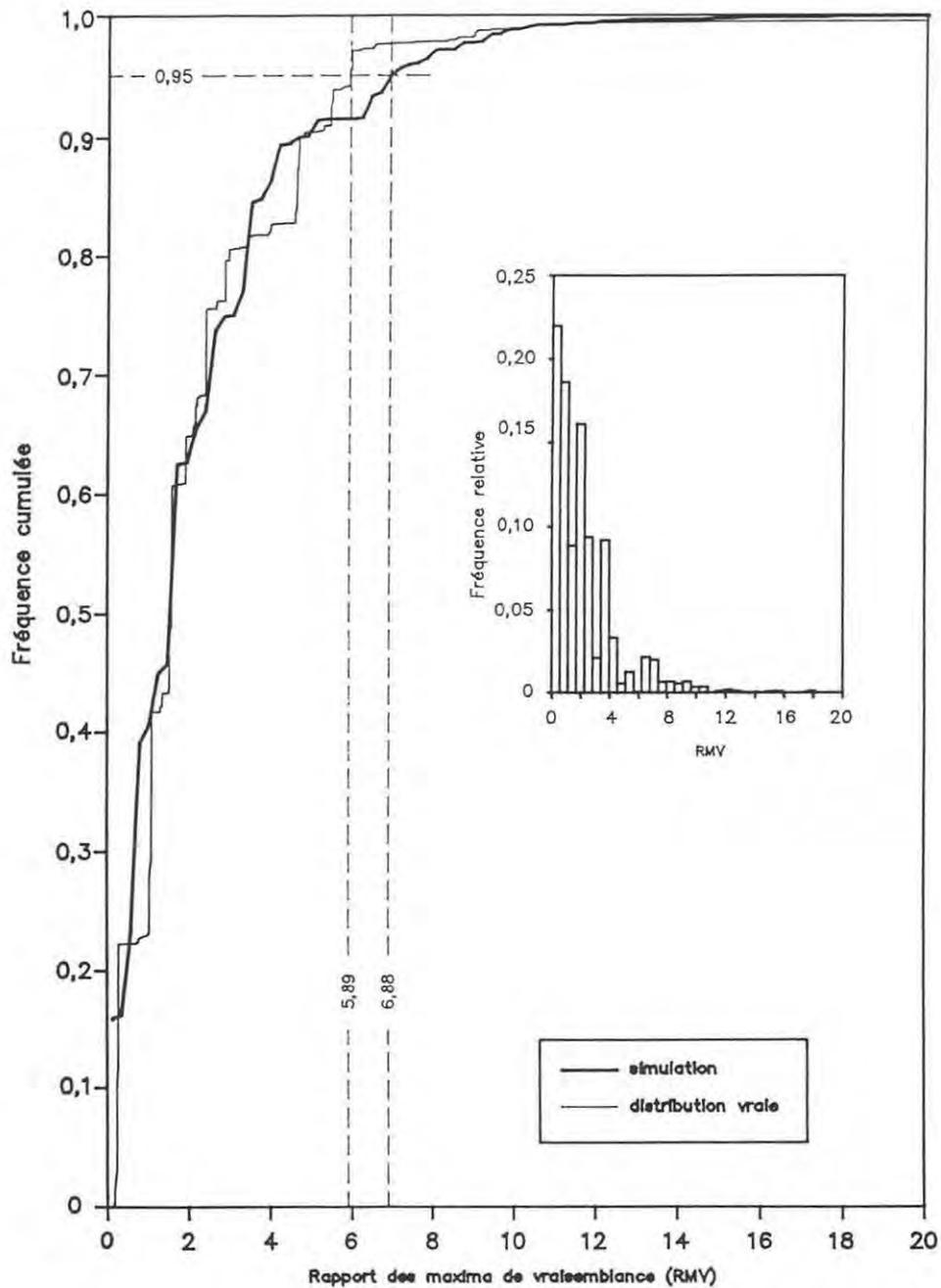
Ainsi si le test du rapport des vraisemblances est utilisé, on est amené à juger trop souvent que deux estimations sont différentes alors qu'elle ne le sont pas forcément : ceci entraîne une augmentation non justifiée de l'erreur de première espèce  $\alpha$ ; les résultats afférents ne sont pas montrés ici. Maul *et al.* (1985b) donnent une application bactériologique du test du rapport des vraisemblances.

### 2.2.2. Test "exact" de comparaison entre deux estimations NPP

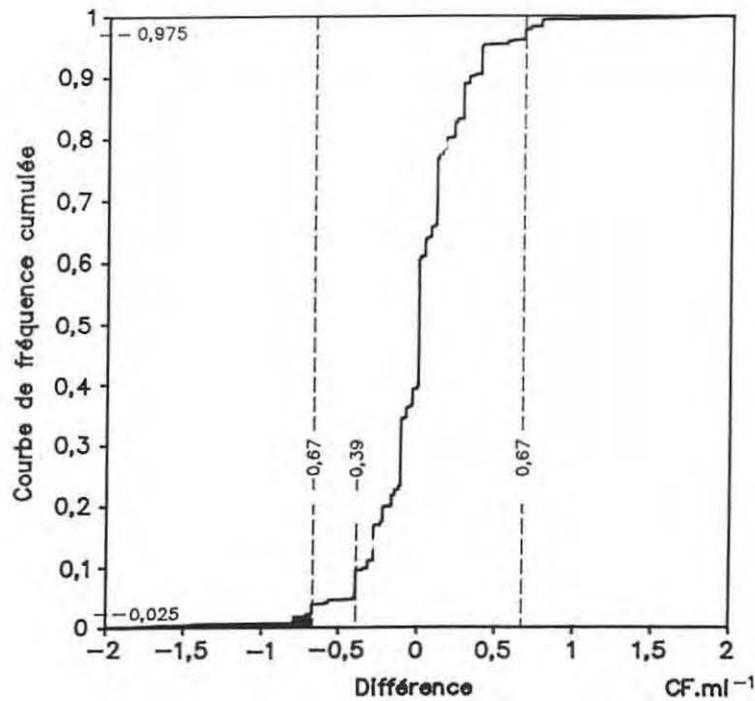
Deux courbes de fréquence cumulée de la différence entre deux estimations NPP, sous l'hypothèse nulle d'une densité réelle commune, sont présentées (Fig. II.5 et II.6).

La différence entre les estimations correspondant aux résultats caractéristiques 1/0/0 et 3/0/0 (respectivement 0,07 et 0,46 CF.ml<sup>-1</sup> pour  $q_3 = q_2/10 = q_1/100 = 0,05$  ml) est non significative (Fig. II.5) : elle se situe à l'extérieur de l'intervalle de confiance centré de la différence, sous l'hypothèse d'une densité commune; la densité commune est l'estimation NPP correspondant au nombre caractéristique somme 4/0/0, soit 0,18 CF.ml<sup>-1</sup>.

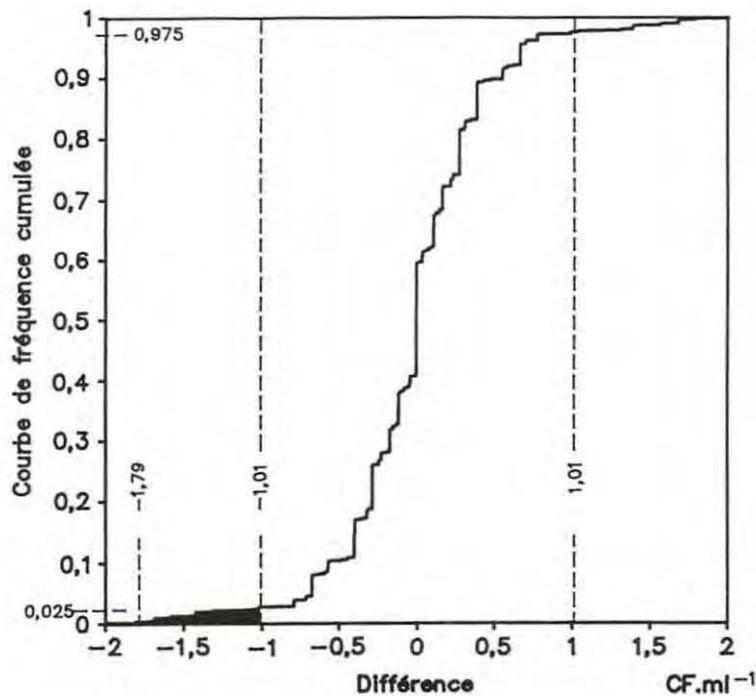
La différence entre les estimations correspondant aux résultats caractéristiques 1/0/0 et 3/2/0, soit 1,79 CF.ml<sup>-1</sup> en valeur absolue, est cette fois significative (Fig. II.6), pour une densité commune de 0,29 CF.ml<sup>-1</sup> (résultat caractéristique 4/2/0).



**Fig. II.4 :** Distributions de fréquence cumulée "vraie" et simulée du rapport des maxima de vraisemblance sous l'hypothèse nulle pour trois tubes par dilution décimale; l'histogramme dans l'encadré correspond aux fréquences de la distribution simulée



**Fig. II.5 :** Distribution de fréquence cumulée de la différence entre deux estimations NPP sous l'hypothèse nulle d'une densité commune pour les deux observations correspondant aux résultats caractéristiques 1/0/0 et 3/0/0 (Différence = -0,39 CF.ml<sup>-1</sup>)

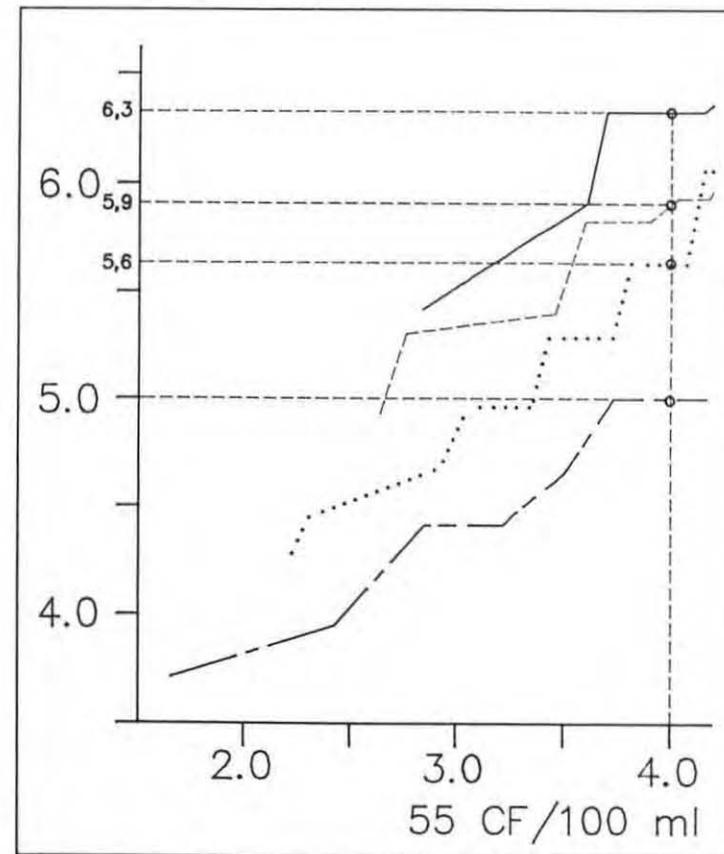
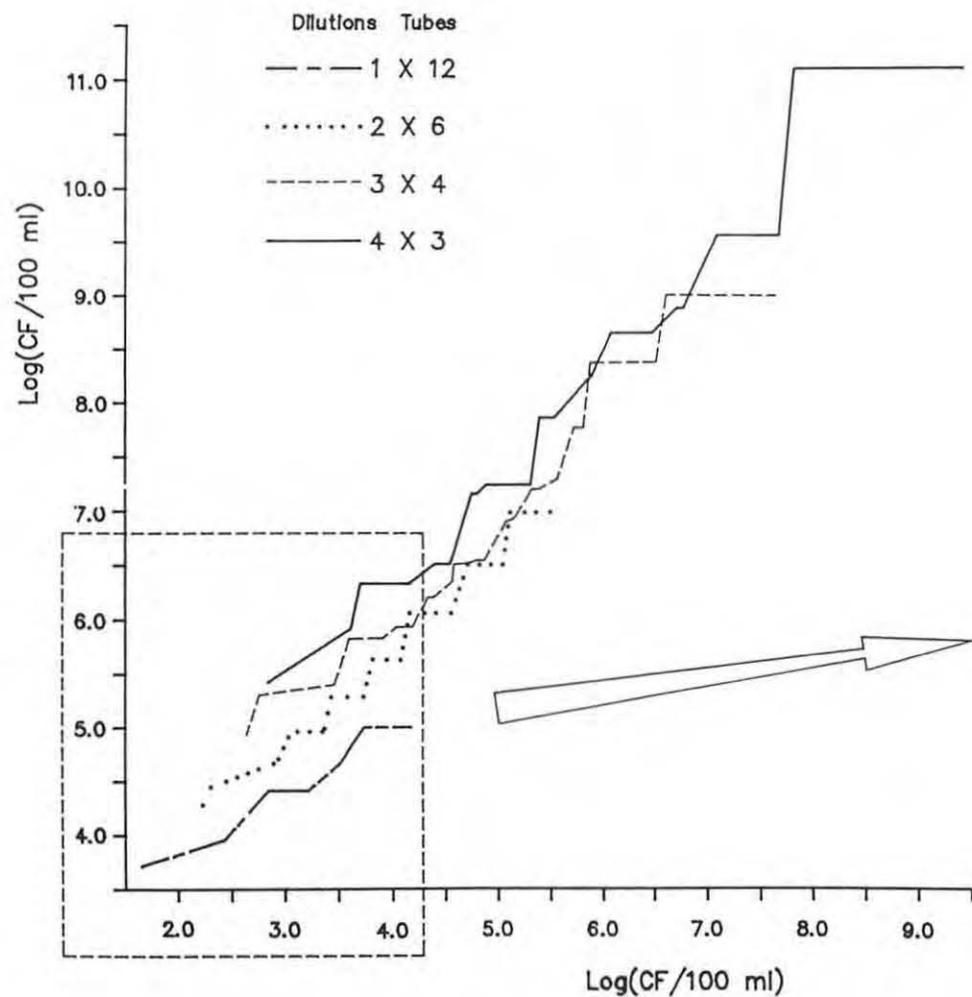


**Fig. II.6 :** Distribution de fréquence cumulée de la différence entre deux estimations NPP sous l'hypothèse nulle d'une densité commune pour les deux observations correspondant aux résultats caractéristiques 1/0/0 et 3/2/0 (Différence = -1,79 CF.ml<sup>-1</sup>)

Les résultats des tests sont présentés sous deux formes selon que l'on s'intéresse aux estimations NPP ou aux résultats caractéristiques, exprimés donc en nombre de tubes positifs par dilution.

Dans le premier cas, la ligne brisée obtenue pour une combinaison particulière de tubes et de dilutions est le lieu géométrique des limites de non-discrimination (Fig. II.7); pour une limite de non-discrimination donnée, l'estimation NPP de plus faible valeur, lue en ordonnée, est significativement différente de l'estimation NPP correspondante en abscisse. Les paliers observés s'expliquent naturellement par le caractère fortement discontinu de la distribution des estimations NPP. L'abscisse déterminant l'arrêt de la ligne brisée est telle que, pour des abscisses supérieures, il n'y ait pas d'estimations NPP supérieures qui soient significativement différentes. Par souci de meilleure clarté de la représentation, les axes ont pour unité le logarithme népérien de la concentration en coliformes fécaux.

Le choix des combinaisons de nombres de tubes et de dilutions n'est pas anodin : le nombre total de tubes par dilution est le même quelque soit la combinaison. Il apparaît que globalement il est d'autant plus facile de distinguer deux estimations que l'on dispose de moins de dilutions pour un même nombre total de tubes. Ceci est illustré par un exemple dans l'encadré de la figure. Pour une densité estimée de 55 CF/100 ml ( $4,0 \text{ Log}(\text{CF}/100 \text{ ml})$ ), les premières estimations jugées significativement différentes sont dans l'ordre du "pouvoir discriminant" croissant 148, 270, 365 et 545 CF/100 ml. Bien sûr, une telle comparaison n'est possible que pour une gamme de concentration très réduite, ici telle que l'utilisation d'une seule série de dilution ne conduise pas au résultat "douze tubes positifs", ce qui a 5% de chances de se produire quand la densité réelle est de 91 CF/100 ml. Le corollaire est qu'il est nécessaire d'avoir une connaissance *a priori* de la concentration réelle. Notons que si l'échantillon retenu pour ensemercer les douze tubes n'est pas la suspension-mère, mais une dilution de cette suspension, le problème est bilatéral : il faut éviter également le résultat "douze tubes négatifs", si l'on désire obtenir une estimation ponctuelle de la concentration.



**Fig. II.7** : Limites de non-discrimination (lignes brisées) pour diverses combinaisons de nombres de dilutions et de tubes par dilution, présentant au total le même nombre de tubes. L'ordonnée d'un point particulier d'une ligne brisée correspond à la plus faible estimation NPP qui soit supérieure à l'estimation NPP correspondante en abscisse

Le deuxième type de représentation est beaucoup plus visuel : il s'agit de la matrice des croisements des résultats caractéristiques entre eux, pour une combinaison donnée des nombres de dilution et de tubes par dilution. Les deux combinaisons utilisées sont étudiées : pour trois séries de dilutions, trois (Fig. II.8a) et cinq tubes (Fig. II.8b) par dilution. Les résultats des tests de comparaison de résultats caractéristiques peu vraisemblables (*e.g.* 2/0/2 ou 5/2/4) ne sont pas présentés ici, mais ont été calculés par ailleurs et peuvent être fournis. Ces résultats sont d'un intérêt pratique moindre. Les estimations NPP correspondant aux résultats présentés dans les tables ont des probabilités supérieures à 0,002 et 0,017 pour respectivement 3 et 5 tubes par dilution.

Ces matrices ou tableaux de test sont d'une utilisation simple : l'aire ombrée représente l'ensemble des croisements entre résultats caractéristiques deux à deux dont les estimations NPP afférentes et respectives sont jugées non significativement différentes. Par abus de langage, les "résultats caractéristiques" sont non significativement différents. La diagonale inférieure donne les résultats du test bilatéral et la diagonale supérieure ceux du test unilatéral. Les degrés de signification des tests sont également indiqués; les cases non ombrées et où ne figurent pas de chiffres correspondent à un degré de signification inférieur ou égal à 1 pour 1000 ( $p < 0,001$ ). Pour le cas bilatéral "\*" indique une différence non significative pour le test "exact", révélée significative par le test classique (Cochran, 1950; voir § I.2.3.5) au seuil de 5%. Ce dernier test conduit à rejeter sensiblement plus souvent l'hypothèse nulle, alors que ses conditions d'utilisation ne sont pas respectées pour les nombres de tubes utilisés.

Par exemple, pour trois tubes par dilution, 1/0/0 est non significativement différent de 3/0/0 mais est différent de 3/2/0, compte tenu des courbes de fréquence cumulée construites sous l'hypothèse d'une densité commune (Fig. II.5 et II.6), au seuil de 5%. 1/0/0 est différent de 3/2/0 pour un degré de signification  $p$  inférieur à 0,001. Pour cinq tubes, 2/0/0 n'est pas différent de 5/0/0 mais est différent du résultat 5/1/0 (pour les estimations voir Tab. II.2), au seuil de 5% pour un degré de signification compris entre 5 et 1%.

a 3X3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
1: 0/0/1														5	5	5	5	1	1	1	1	1	1	1				
2: 0/1/0														5	5	5	5	1	1	1	1	1	1	1				
3: 1/0/0																5	5	5	5	1	1	1	1					
4: 0/2/0																		5	5	5	1	1	1	1				
5: 1/0/1																		5	5	5	1	1	1	1				
6: 1/1/0																		5	5	5	1	1	1	1				
7: 2/0/0																			5	5	5	1	1					
8: 1/1/1																			5	5	5	5	1	1	1			
9: 1/2/0																			5	5	5	5	1	1	1			
10: 2/0/1 *	*																				5	5	5	5	1			
11: 2/1/0 *	*	*																			5	5	5	5	1			
12: 2/1/1 *	*	*	*																		5	5	5	5	1			
13: 2/2/0 *	*	*	*	*																	5	5	5	5	1			
14: 3/0/0 *	*	*	*	*	*																		5	5	5	1		
15: 2/2/1 *	*	*	*	*	*	*																	5	5	5	1		
16: 2/3/0 *	*	*	*	*	*	*	*																	5	5	1		
17: 3/0/1	5	*	*	*	*	*	*																		5	5	1	
18: 3/1/0	5	5	*	*	*	*	*																			5	1	
19: 3/0/2	5	5	5	5	*	*	*	*	*																	5	1	
20: 3/1/1	5	5	5	5	*	*	*	*	*	*																5	1	
21: 3/2/0	1	1	1	1	5	5	5	5	5	*	*																1	
22: 3/1/2	1	1	1	1	5	5	5	5	5	5	*	*	*	*													5	
23: 3/2/1	1	1	1	1	1	1	5	5	5	5	5	*	*	*	*	*											5	
24: 3/2/2	1	1	1	1	1	1	5	5	5	5	5	*	*	*	*	*	*										5	
25: 3/3/0							1	1	1	1	1	1	1	5	5	5	5	*	*									
26: 3/3/1														5	1	1	5	*	*	*	*							
27: 3/3/2																			1	1	1	5	5	5	*			

**Fig. II.8 :** Table des tests "exacts" (a: trois dilutions et trois tubes par dilution; b: trois dilutions et cinq tubes par dilution); les diagonales supérieure et inférieure représentent respectivement les résultats correspondant aux tests unilatéraux et bilatéraux avec les degrés de signification p correspondant (aire ombrée :  $p > 0,05$ ; "5" :  $0,1 < p < 0,5$ ; "1" :  $0,01 < p < 0,001$ ; cases vides :  $p < 0,001$ ). "\*" : test classique (Cochran, 1950) significatif au seuil de 5%

## b 3X5

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27		
1: 1/0/0								5	5	5	5	5	1	1	1	1													
2: 1/1/0									5	5	5	5	5	1	1														
3: 2/0/0										5	5	5	5	1	1														
4: 2/1/0 *													5	5	1	1	1	1											
5: 3/0/0 *														5	5	5	1	1	1										
6: 3/1/0 *															5	5	1	1	1										
7: 4/0/0 *																5	5	1	1	1									
8: 3/2/0 *																	5	1	1	1	1								
9: 4/1/0 * * *																		5	5	5	1	1	1						
10: 4/2/0	5	*	*															5	5	1	1	1	1	1					
11: 5/0/0	5	*	*	*															5	5	5	5	1	1	1				
12: 4/3/0	5	*	*	*	*															5	5	5	1	1	1	1			
13: 5/1/0	5	5	*	*	*																5	5	5	5	5	1	1		
14: 5/1/1	1	5	5	5	5	*	*	*														5	5	5	5	5	1		
15: 5/2/0	1	1	1	5	5	*	*	*															5	5	5	5	1		
16: 5/2/1	1	1	1	5	5	5	5	5	*	*																5	1	1	
17: 5/3/0		1	1	1	1	5	5	5	5	*	*															5	5	1	
18: 5/3/1			1	1	1	1	5	5	5	5	*	*	*														5	1	
19: 5/4/0					1	1	1	1	1	5	5	5	*														5	5	1
20: 5/4/1						1	1	1	1	5	5	5	5	*	*													5	1
21: 5/4/2							1	1	1	1	5	5	5	*	*													5	1
22: 5/5/0										1	1	5	1	5	5	*													1
23: 5/4/3											1	1	5	1	5	5	*	*	*										1
24: 5/5/1												1	1	5	5	5	*	*											5
25: 5/5/2															1	1	5	5	5	*	*								
26: 5/5/3																1	1	1	5	5	5	*	*	*					
27: 5/5/4																				1	1	1	1	5	5	*			

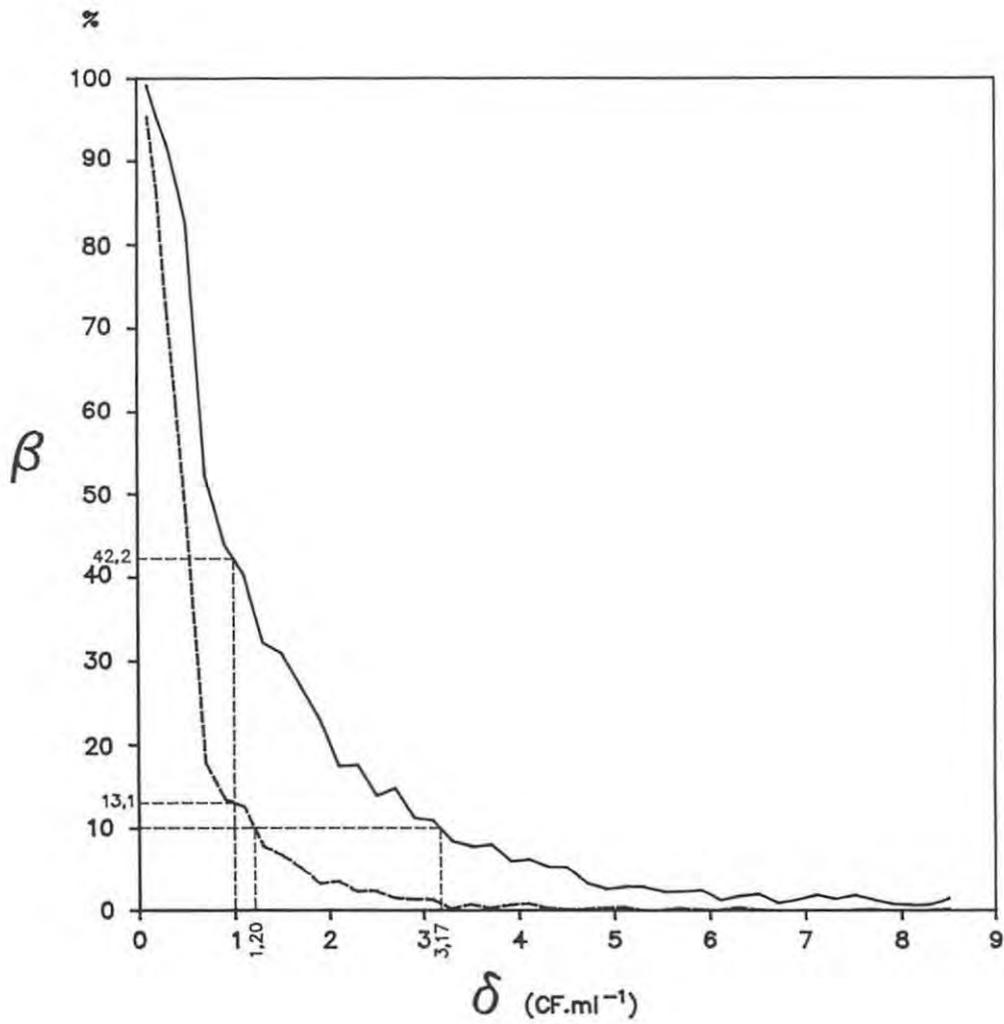
Fig. II.8 : Table des tests "exacts" (a: trois dilutions et trois tubes par dilution; b: trois dilutions et cinq tubes par dilution); les diagonales supérieure et inférieure représentent respectivement les résultats correspondant aux tests unilatéraux et bilatéraux avec les degrés de signification p correspondant (aire ombrée :  $p > 0,05$ ; "5" :  $0,1 < p < 0,5$ ; "1" :  $0,01 < p < 0,001$ ; cases vides :  $p < 0,001$ ). "\*" : test classique (Cochran, 1950) significatif au seuil de 5%

La proportion aire ombrée/aire totale des tableaux autorise la comparaison entre les deux combinaisons. Le gain de pouvoir discriminant obtenu pour l'utilisation de cinq tubes par dilution apparaît nettement. Cette vision ne reflète pas l'aspect quantitatif des estimations NPP correspondantes : elle est donc biaisée, mais dans le bon sens car l'écart entre les estimations NPP correspondant aux résultats extrêmes est plus important pour cinq tubes que pour trois tubes par dilution - la gamme de concentration possiblement explorable est plus large; ainsi l'aire ombrée est en réalité plus étroite et allongée.

L'augmentation du nombre de tubes par dilution augmente à l'évidence la puissance du test. Les courbes de puissance correspondant à trois tubes par dilution et cinq tubes par dilution ont pu être simulées pour des différences croissantes en concentration (Fig. II.9).

Chaque point d'une courbe est obtenu de la façon suivante : soient  $d_1$  et  $d_2$  fixées, avec  $d_2 > d_1$ , un résultat caractéristique est simulé (voir § I.2.3.5) à partir de chacune de ces deux densités. Alors pour une différence  $d_2 - d_1$  donnée (en abscisse), on obtient deux estimations NPP  $\hat{\delta}_1$  et  $\hat{\delta}_2$  et on effectue le test "exact" correspondant. Après avoir réalisé 1000 fois cette procédure, on dénombre le nombre de fois où la différence  $\hat{\delta}_2 - \hat{\delta}_1$  est apparue non significative, c'est-à-dire où l'hypothèse nulle n'est pas rejetée. Le rapport à 1000 de ce nombre donne l'ordonnée d'un point.

En utilisant trois tubes, la probabilité de ne pas distinguer une différence de 1 CF.ml<sup>-1</sup> avec  $\hat{\delta}_1 = 0,1$  CF.ml<sup>-1</sup> est de 0,42 alors qu'elle n'est que de 0,13 pour cinq tubes. Inversement, une erreur de seconde espèce de 10% est obtenue pour une différence de 1,2 CF.ml<sup>-1</sup> pour cinq tubes et de 3,2 CF.ml<sup>-1</sup>.



**Fig. II.9** : Courbes de probabilité d'accepter à tort l'hypothèse nulle (pas de différence significative entre deux estimations) pour trois tubes (trait plein) et cinq tubes par dilution (trait pointillé)

## 2.3. Abaques précision-coût

### 2.3.1. Problématique

La quête d'une meilleure précision passe par une augmentation du coût de l'analyse. Existe-t-il des optima coût-précision en matière de dénombrement indirect des microorganismes au moyen de l'estimateur NPP ? Pour un choix de combinaison donné, existe-t-il une plage de concentration sensiblement plus "chère" à analyser? Quelle est l'évolution de la courbe coût-précision avec l'augmentation du nombre de tubes par dilution? Autant de questions auxquelles la construction graphique imaginée plus loin tente d'apporter des réponses.

### 2.3.2. Détermination des axes

Un point (coût,précision) d'une courbe est construit à partir de 1000 simulations de résultats caractéristiques pour une concentration donnée en coliformes ( avec  $q_3 = q_2/10 = q_1/100 = 0,05/3$  ml) . Le principe de la simulation a été exposé par ailleurs (§ I.2.3.5). L'abscisse du point est le coût moyen des mille résultats et l'ordonnée la précision moyenne.

L'axe des abscisses représente le coût d'une analyse en unité monétaire (UM), en l'occurrence le franc. Nous avons vu précédemment (§ I.2.2.2) que la mesure des coliformes fécaux n'était réalisée qu'à l'issue de deux opérations successives : une première où un tube positif signifie la présence de coliformes totaux (incubation à 37°C) et une deuxième où un tube positif est lié plus spécifiquement à la présence de coliformes fécaux, thermo-tolérants pour une température d'incubation de 44°C, après repiquage. Tout tube positif à l'issue de la première phase est repiqué pour la seconde. On considère que la simulation d'un résultat caractéristique nous donne en fait la concentration en coliformes fécaux et que, par conséquent, on ne peut préjuger du résultat caractéristique donnant au préalable la concentration en coliformes totaux.

Ainsi, nous admettrons que le résultat caractéristique est le même dans les deux cas. Le coût est alors surestimé pour toutes les combinaisons envisagées.

Le coût unitaire, pour un résultat caractéristique donné et pour une combinaison donnée, est alors donné par l'équation :

$$\text{Coût} = 3 \cdot C_t \cdot n_t + 2 \cdot C_t \cdot t_m + C_d \cdot d_m$$

avec  $C_t$  le coût de l'ensemencement d'un tube,  $3 \cdot n_t$  le nombre de tubes total pour les trois dilutions,  $t_m$  la moyenne du nombre de tubes repiqués,  $2 \cdot C_t \cdot t_m$  le coût moyen du repiquage (2 tubes sont systématiquement repiqués par tube positif),  $C_d$  le coût d'une dilution, et  $d_m$  le nombre moyen de dilutions obtenu à l'issue de 1000 simulations pour une combinaison donnée.

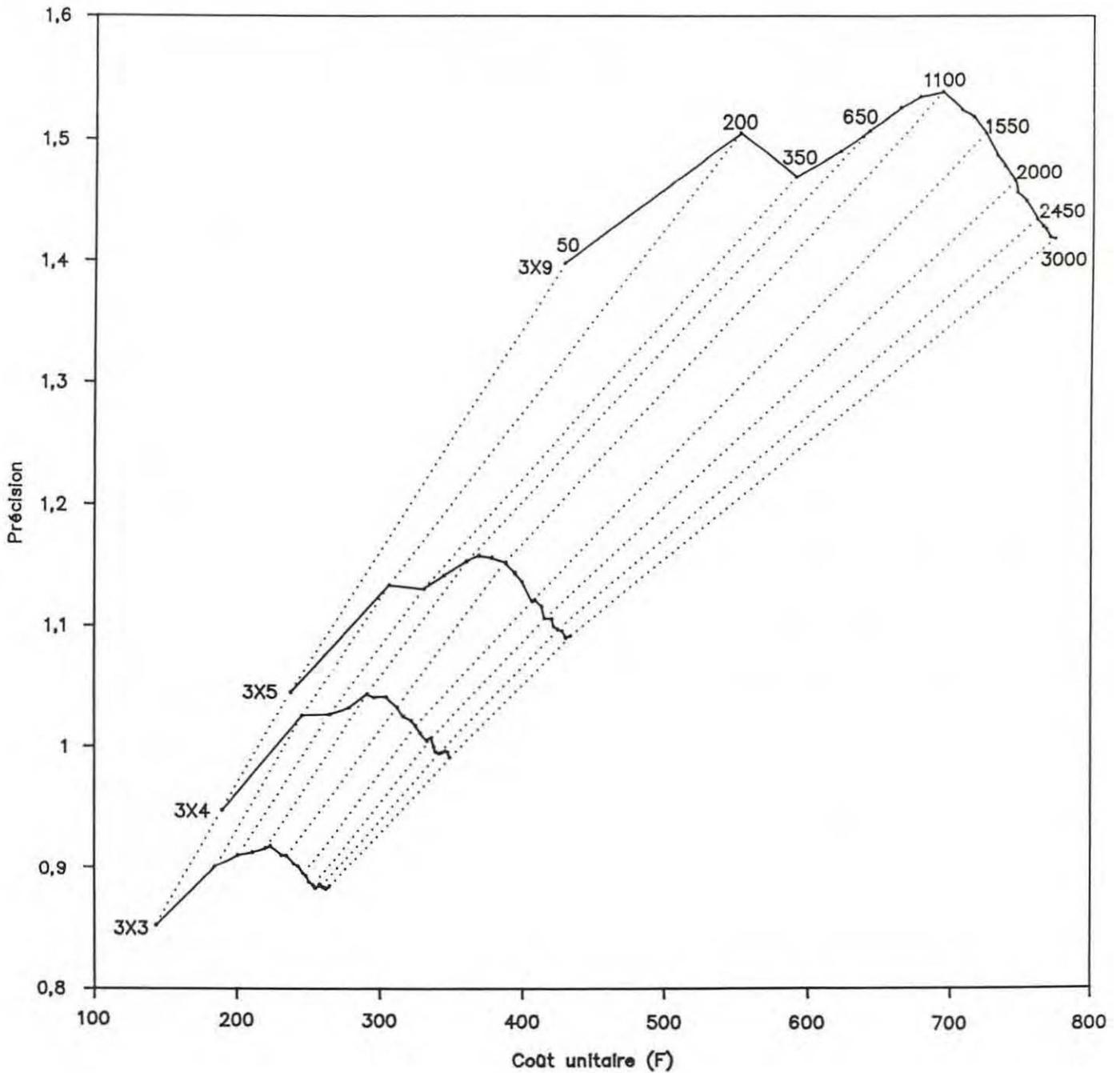
En ordonnée, la précision choisie est l'inverse de la moyenne arithmétique des logarithmes décimaux des rapports de la limite supérieure (Ls) à la limite inférieure (Li) pour les différentes estimations obtenues par simulation. On peut écrire :

$$P = \frac{1}{\log\left(\frac{Ls}{Li}\right)}$$

Cette mesure de la précision comparée à la moyenne des écarts au carré présente l'avantage d'être peu sensible aux résultats extrêmes obtenus par simulation : pour un nombre de dilutions donné et des nombres de tubes par dilution donnés,  $\log(Ls/Li)$  est stable; par exemple pour trois dilutions et trois tubes par dilution,  $0,9 < \log(Ls/Li) < 1,3$ .

### 2.3.3. Exemples d'utilisation des abaques

Pour une concentration réelle de 350 CF/100 ml le coût unitaire moyen est de 200 UM pour une précision de 0,9 alors qu'il est de 330 UM pour une précision de 1,1 (Fig. II.10). Pour 9 tubes par dilution, la précision obtenue est de 1,5 environ ce qui correspond grossièrement à un



**Fig. II.10** : Abaques coût-précision pour les estimations NPP après simulation de résultats caractéristiques à partir de densités réelles (en CF/ 100 ml)

rapport de 5 entre  $L_s$  et  $L_i$ . Ce rapport est voisin de 13 pour trois tubes, pour un coût trois fois moindre en moyenne.

#### 2.4. Discussion

La méthode de dilution des tubes ne doit pas être considérée comme une méthode rigide, mais plutôt comme une technique adaptable aux besoins en précision par le truchement du jeu sur le nombre de tubes par dilution, le nombre de dilutions et le facteur de dilution.

Ainsi, Cochran (1950) fournit un critère de planification des séries de dilution, fondé sur le nombre minimal de bactéries que doit contenir le plus grand volume d'échantillon (la plus faible dilution) et le nombre maximal de bactéries que doit contenir le plus faible volume (la plus forte dilution). On parie préalablement que la concentration réelle inconnue se trouve entre deux nombres, puis les volumes sont fixés pour satisfaire au critère. Quand le nombre de dilutions est fixé, comme dans notre cas, ceci revient à fixer le facteur de dilution.

Russek et Colwell (1983), pour leur part, indiquent que le nombre de tubes utilisés par dilution doit évoluer en fonction de l'intérêt de la dilution considérée, sur le plan de la précision de l'estimation. Ainsi l'effort de sous-échantillonnage dans le broyat doit porter naturellement sur la dilution intermédiaire dans le cas de trois dilutions, et *a fortiori* décimales. Par conséquent, dans le contexte des réseaux, il serait plus judicieux d'utiliser la combinaison ( $n_1 = 2$ ,  $n_2 = 5$ ,  $n_3 = 2$ ) pour une précision accrue dans une gamme de concentration éloignée des résultats extrêmes (0/0/0 et 2/5/2). La table correspondante est fournie en annexe II.1. A l'extrême, nous avons montré le gain de précision apportée par l'utilisation d'une unique dilution (Fig. II.7).

Ces différentes solutions pour une utilisation raisonnée de la méthode de dilution des tubes ont en commun le fait qu'elles reposent sur la même condition initiale : on doit pouvoir avoir une idée *a priori* pas trop mauvaise de la concentration bactérienne dans l'échantillon à analyser, ce qui n'est pas toujours évident suivant les secteurs de surveillance. En outre, une inflation de

nouvelles combinaisons selon les secteurs ne facilite pas la tâche des analystes microbiologistes. On augmente également les chances, en resserrant la gamme de concentration explorable, de laisser échapper un pic important de contamination.

L'augmentation du nombre de tubes par dilution de trois à cinq pour trois décimales entraîne celle de la précision. On ne distingue pas de véritable optimum coût-précision, même si on observe un pic entre 200 et 650 CF/100 ml visible surtout pour la combinaison 3x9 (trois dilutions et neuf tubes par dilution). Cet accident est imputable à un changement assez net dans la nature des résultats caractéristiques simulés, les résultats de type '9/x/x' moins précis devenant plus fréquents. Par exemple, les valeurs de  $\log(L_s/L_i)$  pour les résultats 8/3/0 (estimation NPP = 154 CF/100 ml) et 9/4/0 (estimation NPP = 327 CF/100 ml) sont respectivement 4,01 et 4,98... En outre, quelle que soit la combinaison envisagée une densité réelle voisine de 1000 CF/100 ml génère les résultats les plus précis pour des raisons analogues. Notons enfin que la plus mauvaise précision est obtenue pour les densités extrêmes de la gamme envisagée : une présomption de fortes valeurs, suite par exemple, à l'examen de données historiques doit inciter à effectuer une dilution supplémentaire.

L'augmentation de la précision avec le nombre de tubes par dilution se traduit concrètement par un gain de puissance pour le test "exact" de comparaison entre deux estimations NPP, seul test utilisable pour les nombres de tubes utilisés en routine. Il permet la comparaison de deux estimations NPP indépendantes et peut être appliqué à l'alerte bactériologique, dans le cas unilatéral pour répondre par exemple à la question : l'augmentation de contamination observée à telle station est-elle significative ?

Pour autant, le descripteur reste imprécis et des différences fines de contamination spatiale dans le temps ou dans l'espace ne seront mises en évidence qu'au prix d'un effort d'échantillonnage important.

### 3. Utilisation d'une méthode conductance-métrique

#### 3.1. Introduction

Disposant du temps de détection de la courbe conductance-métrique (§ 1.2.2.3), l'analyste désire connaître la concentration en coliformes fécaux correspondante; ceci nécessite l'utilisation d'un outil prévisionnel. A cette fin est construite la droite de régression du logarithme décimal du nombre de coliformes fécaux pour 100 ml de broyat de chair et de liquide intervalvaire (CF) en temps de détection.

#### 3.2. Méthodes

##### 3.2.1. Protocole analytique

A partir d'un même broyat de chair et de liquide intervalvaire de coquillages, sont réalisées deux mesures : l'estimation du "Nombre le Plus Probable" et la mesure d'un temps de détection d'une courbe conductimétrique (TD). Les volumes d'inoculum sont suffisamment petits devant le volume total du broyat pour considérer que les nombres de bactéries dans les différents volumes prélevés sont indépendants.

##### 3.2.2. Régression

Seuls les couples de mesures toutes deux positives sont conservés. Les premiers essais de régression linéaire sur l'ensemble des résultats ont montré que pour les faibles concentrations, proches du seuil de détection de la méthode NPP ( $1,26 \log(\text{CF})$ ), la gamme de variation du temps de détection croît sensiblement vers les forts temps de détection (10-12 h). Ceci contribue à diminuer la pente en valeur absolue, puisque pour des temps de détection élevés, on ne dispose pas de valeurs de concentrations inférieures au seuil de détection. Ceci amène à déterminer *a posteriori* un temps de détection limite, qui doit être choisi de telle façon que la concentration correspondante soit proche du seuil de détection de la méthode conductimétrique

( $1,48 \log(CF)$ ). Finalement, seuls les couples présentant des temps de détection inférieurs ou égaux à 9,2 h seront considérés dans le calcul de régression.

Un modèle de régression linéaire est ajusté pour chaque type de coquillage. La décomposition de la somme des carrés des ordonnées observées ( $\log(CF)$ ) permet la construction d'un tableau d'analyse de la variance, dans lequel figurent les parts de variance (carrés moyens) attribuées respectivement à la régression et aux écarts à la régression. Le test F permet de juger de la signification du modèle préconisé. L'utilisation du test de signification de la régression requiert la normalité des résidus, recherchée par ailleurs au moyen du test de Kolmogoroff-Smirnoff. Notons que ce test n'est pas le plus indiqué, celui de Shapiro-Wilk étant reconnu le plus puissant (Kendall & Stuart, 1961) dont Tassi (1989) donne une description détaillée.

Le bien-fondé de l'utilisation du modèle linéaire est évalué. Pour ce faire nous n'utilisons pas le test classique indiqué par Schwartz (1963), car ce test nécessite d'avoir plusieurs valeurs de concentration pour chaque valeur de temps de détection (Draper & Smith, 1966), ce qui n'est pas le cas dans notre étude. On pourrait pallier ce problème en effectuant des regroupements en classe, ou bien en simulant des valeurs de  $y$  pour un  $x$  donné. Mais dans un cas comme dans l'autre, on est amené à émettre une hypothèse de normalité des  $y$  pour un  $x$  donné. Voulant s'affranchir de cette condition, on utilise le test non paramétrique des séquences pour un échantillon (Siegel, 1956) : on dénombre les séquences des signes des résidus (écarts entre les observations et les valeurs prédites) pour les temps de détection ordonnés. Une statistique  $z$  est ensuite calculée à partir de ce nombre et des nombres de résidus positifs et négatifs; la distribution de cette statistique approche la variable normale réduite dans le cas où le nombre de + ou de - est supérieur à 20; ainsi, si  $|z|$  est supérieure à 1,96, il y avait moins de 5% de chances qu'on observe une telle valeur sous l'hypothèse d'une répartition aléatoire des résidus autour de 0, ou de façon équivalente sous l'hypothèse d'une répartition aléatoire des observations autour

de la droite de régression, lieu des prédictions. Il convient de souligner la faible puissance de ce test.

Trois droites de régression sont obtenues dont on se demande si elles peuvent être considérées comme semblables. Les différentes étapes de la méthode de comparaison, décrite dans l'annexe II.2, sont les suivantes :

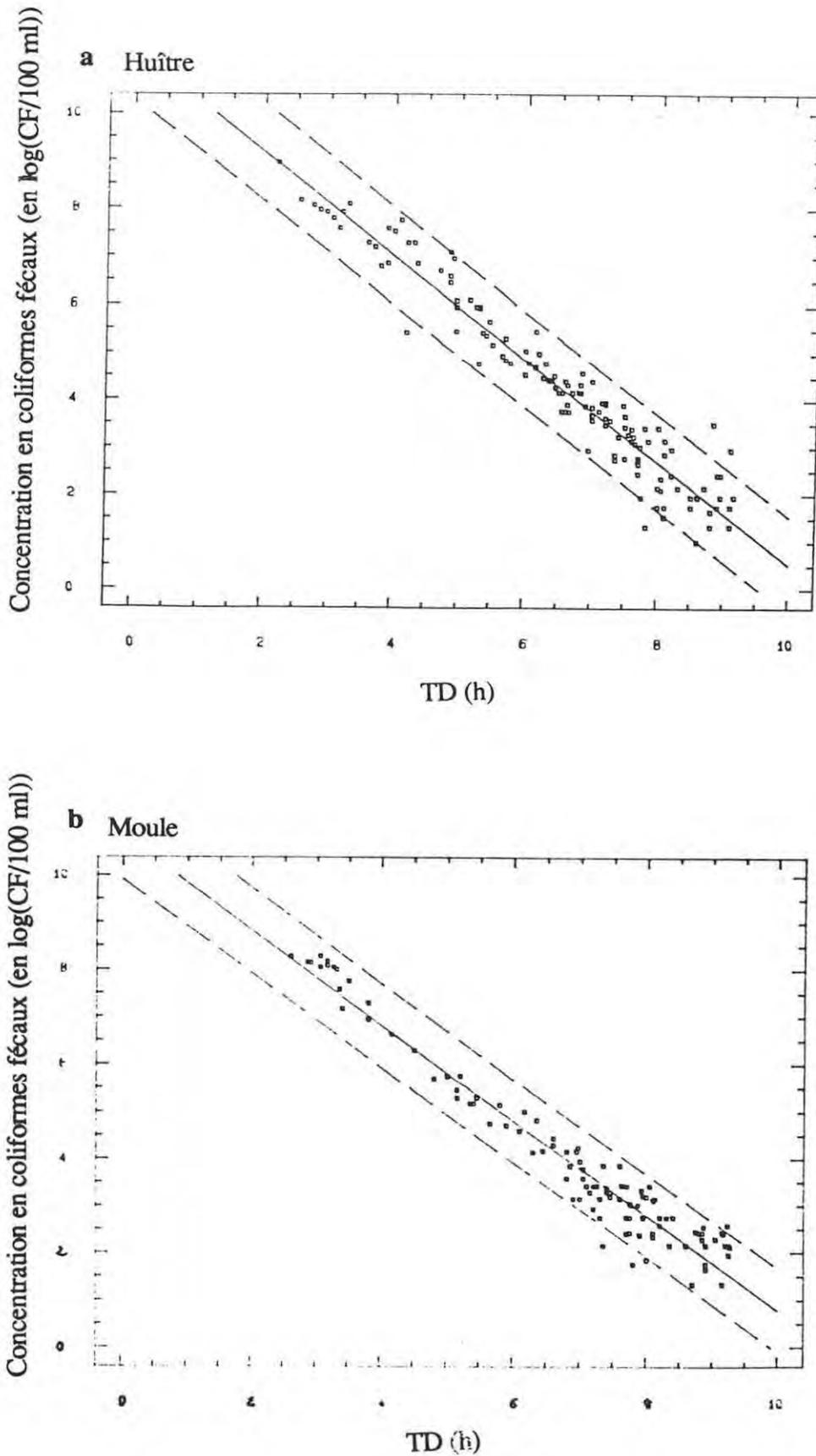
- comparaison des variances résiduelles au moyen du test de Bartlett,
- si le test est non significatif : comparaison des pentes ou coefficients de régression,
- si les pentes ne diffèrent pas significativement : comparaison des ordonnées à l'origine.

### **3.3. Résultats**

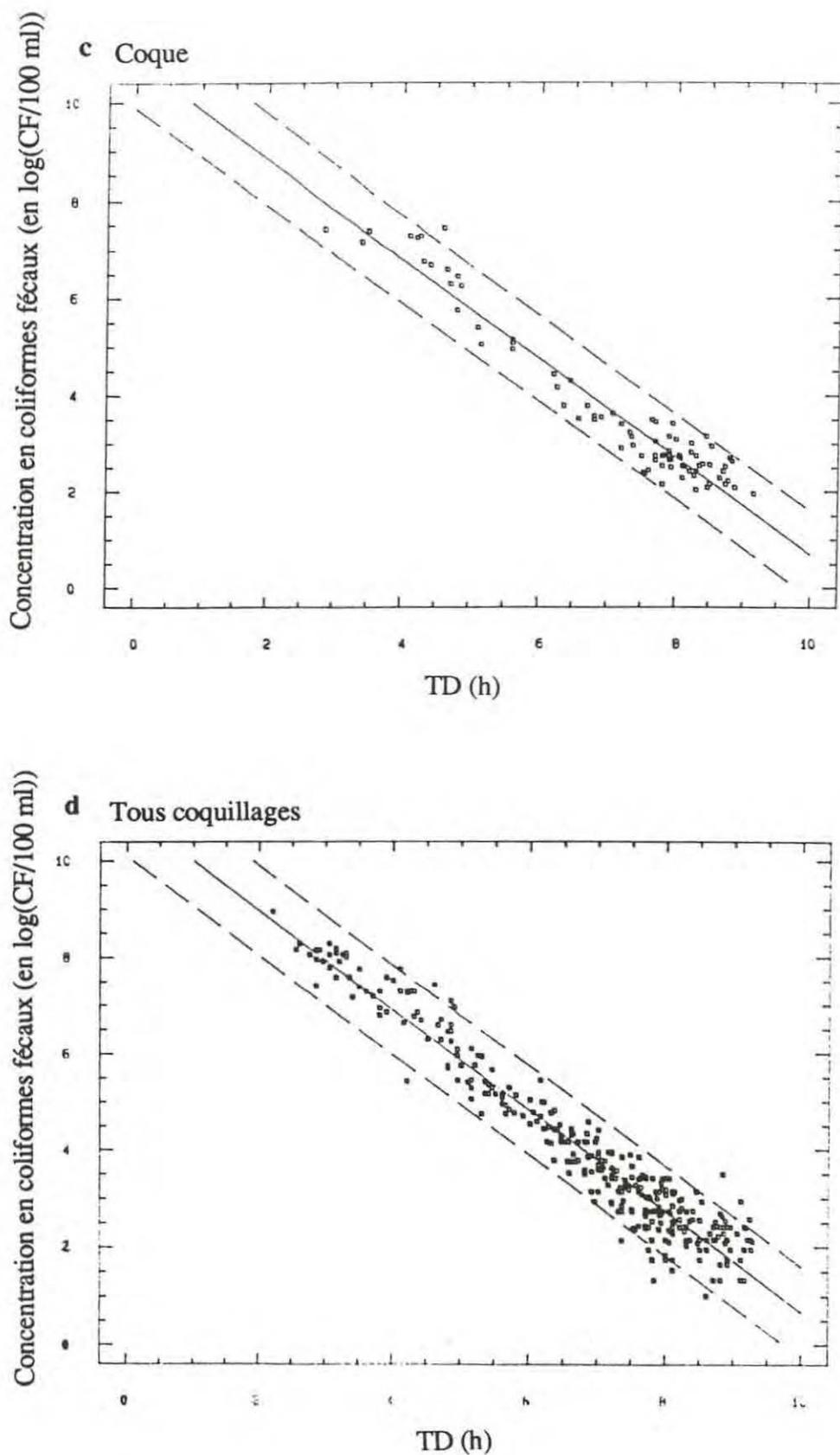
#### *3.3.1. Régressions*

Pour chaque régression, sont donnés dans le tableau ci-dessous les coefficients de régression et leur erreur-type, l'erreur-type de la valeur prédite, la part de variance expliquée par le modèle et le résultat du test des séquences sur le signe des résidus. Chaque droite de régression est représentée (Fig. II.11a, b, c et d), avec le nuage des observations et l'enveloppe de confiance de la prédiction (voir annexe II.2).

Les tableaux d'analyse de variance permettant de tester la signification des modèles sont donnés en annexe II.2, ainsi que les représentations des résidus, en fonction des temps de détection. Compte tenu des valeurs de F calculées pour chaque régression et des degrés de liberté correspondants (Tab. II.6), les test de signification de la régression se révèlent hautement significatifs ( $p < 10^{-6}$ ). Les tests de Kolmogoroff-Smirnoff sont tous non significatifs : on ne peut donc rejeter l'hypothèse de conformité de la distribution des résidus à une loi normale.



**Fig. II.11** : Régressions de la concentration en coliformes fécaux (transformation logarithmique de l'estimation NPP en  $\log(\text{CF}/100 \text{ ml})$ ) sur le temps de détection (TD en heures) pour différents coquillages (a : huître; b : moule; c : coque; d : tous coquillages)



**Fig. II.11** : Régressions de la concentration en coliformes fécaux (transformation logarithmique de l'estimation NPP en  $\log(\text{CF}/100 \text{ ml})$ ) sur le temps de détection (TD en heures) pour différents coquillages (a : huître; b : moule; c : coque; d : tous coquillages)

**Tab. II.6** : Résultats de la régression. n : nombre de couples; b ( $s_b$ ) : pente et (écart-type); a ( $s_a$ ) : ordonnée à l'origine et (écart-type);  $s_{\hat{y}}$  : écart-type de prédiction;  $R^2$  : pourcentage de variance expliquée par le modèle.

	Huître	Moule	Coque	Total
n	132	115	84	331
b ( $s_b$ )	-1,08 (0,03)	-1,01 (0,02)	-1,02 (0,03)	-1,04 (0,01)
a ( $s_a$ )	11,33 (0,17)	10,85 (0,16)	10,85 (0,22)	11,05 (0,10)
$s_{\hat{y}}$	0,50	0,44	0,44	0,47
$R^2$	93,1%	94,7%	92,9%	93,7%
Test des séquences	Non Significatif	Non Significatif	Très Significatif	Non Significatif

L'équation de la droite s'écrit :

$$\widehat{\log(\text{CF})} = a + b * (\text{TD})$$

où a et b sont les valeurs des coefficients lues dans le tableau.

En pratique, on désire obtenir la mesure de la concentration en coliformes fécaux en échelle arithmétique. Calculer simplement l'antilogarithme de l'estimation de  $\log(\text{CF})$  obtenue conduit à une sous-estimation de cette mesure. En effet, après transformation logarithmique les fortes valeurs de CF ont moins de poids dans le calcul de la régression que les petites valeurs, ce qui conduit à une diminution de la pente. Beauchamp & Olson (1973) donnent un facteur de correction de biais  $\psi(\hat{\sigma}_{y,x}^2/2)$  dans l'hypothèse d'une distribution normale de la variable expliquée (ici  $\log(\text{CF})$ ), où  $\hat{\sigma}_{y,x}^2$  est l'estimation non biaisée de la variance de  $\log(\text{CF})$  égale au carré moyen des écarts à la régression noté  $s_{y,x}^2$  en annexe II.2. L'estimateur non biaisé et de variance minimum de la concentration en coliformes fécaux s'écrit alors :

$$\widehat{\text{CF}} = 10^{(a+b*\text{TD})} \psi(\hat{\sigma}_{y,x}^2/2)$$

Beauchamp & Olson (1973) donnent une expression approchée du facteur de correction de la forme  $10^{\hat{\sigma}_{yx}^2/2} \{1-g(1/n, 1/n^2)\}$ . Pour la taille d'échantillon considérée ici ( $n=331$ ), le terme représenté par la fonction  $g$  est petit devant 1 (au maximum  $5 \cdot 10^{-3}$ ). L'estimation s'écrit :

$$\widehat{CF} = 10^{(a+b*TD+\hat{\sigma}_{yx}^2/2)}$$

Les limites de confiance de la valeur prédite après transformation sont données par l'antilogarithme des limites de confiance de la valeur prédite avant transformation. Soient finalement les limites inférieure et supérieure après transformation :

$$CF_{inf} = 10^{[\widehat{\log(CF)} - 1.96*s_y^*]} \text{ et } CF_{sup} = 10^{[\widehat{\log(CF)} + 1.96*s_y^*]}$$

### 3.3.2. Test de comparaison des droites

Le test de Bartlett d'homogénéité des variances est non significatif (statistique = 1,01;  $p = 0,26$ ). Les calculs afférents au test de comparaison des pentes donnent un  $F_{322}^2 < 1$ . On ne peut naturellement conclure à une différence significative entre les trois droites.

### 3.3.3. Exemple numérique

Soit la droite de régression obtenue pour l'ensemble des coquillages :

$$\log(CF) = -1,04*TD + 11,05$$

Soit donné  $TD = 8h$ ,

$$\log(CF) = 2,73 \log(CF/100 \text{ ml})$$

Intervalle de confiance de la valeur prédite :

$$2,73 - 1,96*0,47 < \log(CF) < 2,73 + 1,96*0,47$$

$$1,81 < \log(CF) < 3,65$$

L'équation de transformation en métrique arithmétique est :  $CF = 10^{(11,16 - 1,04*TD)}$ , comprenant un facteur correctif de biais proche de 1,29 ( $10^{(0,47^2/2)}$ ).

Soit donc après transformation,  $CF = 692$  coliformes/100 ml et  $65 < CF < 4467$ .

### 3.4. Discussion

L'utilisation d'un modèle de régression linéaire pour notre propos pose quelques problèmes qu'il convient de souligner même s'il paraît difficile de se passer de l'outil prédictif simple ainsi obtenu, en pratique.

Le problème de l'utilisation de l'estimation NPP comme variable expliquée dans la régression provient du caractère fortement discret de sa distribution pour trois tubes par dilution. Néanmoins, des répliquats ont été effectués, permettant ainsi d'obtenir une estimation plus précise de la concentration (moyennes arithmétiques des logarithmes) pour un temps de détection moyen (moyenne arithmétique), et une distribution de valeurs plus continue. Par ailleurs, l'erreur de mesure du signal conductance-métrique est négligeable par rapport à la variance résiduelle de la concentration en coliformes fécaux : le modèle de régression décrit en annexe n'est alors pas à remettre en cause; en particulier le coefficient de régression calculé estime sans biais le coefficient de régression réel.

Les résidus sont d'aspects satisfaisants, mis à part pour les coques où le test des séquences se révèle significatif (Tab. II.6). L'interprétation en termes de métabolisme bactérien est difficile à cerner. Plus prosaïquement, remarquons qu'il s'agit de la régression calculée avec le moins d'observations; plus précisément ce nuage apparaît déséquilibré par rapport aux autres nuages avec une part prépondérante des observations pour des temps de détection supérieurs à 7 heures (Fig. II.11d). On peut ainsi imaginer que ces observations aient plus de poids lors des calculs, au détriment de l'allure des résidus.

On ne peut conclure à une différence significative entre les pentes des trois droites. Ceci autorise l'utilisation d'une unique droite de régression - celle obtenue en groupant l'ensemble des observations, tous coquillages confondus; ceci a pour conséquence appréciable de faciliter le travail de conversion, en pratique. Notons que, *a priori*, il n'y avait pas de raison pour que ceci soit réalisable : en effet, on pourrait supposer qu'il existe une interaction entre la croissance

bactérienne et la composition de la solution constituée du milieu de culture et du broyat de coquillage, cette composition variant d'un type de coquillage à l'autre (si ce n'est d'un coquillage à l'autre...).

Il sera enfin noté la forte imprécision de la valeur prédite illustrée par les intervalles de confiance présentés dans l'exemple numérique. Ceux-ci ne peuvent être naturellement que supérieurs à ceux des estimations NPP, déjà importants. En pratique, une estimation plus précise pourra être obtenue en effectuant un duplicat par broyat; ceci permet en outre de pallier la défection d'une cellule (électrodes endommagés, mauvais contacts, etc...).

#### 4. Conclusion

Dans les conditions d'utilisation de la méthode de dilution des tubes pour le réseau de surveillance microbiologique, l'estimateur du "Nombre le Plus Probable" est particulièrement imprécis; ceci est caractérisé par l'étendue des intervalles bayésiens. L'estimateur du maximum de vraisemblance permet d'inférer une estimation de la concentration bactérienne, dont l'aspect ponctuel occulte par trop la forte incertitude de l'estimation. Ce point avait été déjà souligné par Cowell et Morisetti (1969).

L'imprécision est également illustrée par la faible puissance du test "exact" de comparaison entre deux estimations indépendantes. Un recours réaliste consisterait à augmenter le nombre de tubes par dilution, mais ceci induit naturellement une augmentation du coût.

L'utilisation de la méthode conductance-métrique, moins onéreuse à l'usage, facilite en outre grandement la répétition des mesures à partir d'un même échantillon, conduisant ainsi à une estimation plus précise de la concentration en coliformes fécaux dans le broyat de chair de coquillages. Avec deux bains d'incubation, le système actuel permet 56 analyses simultanées, ce qui était inenvisageable, en temps et en moyens, avec la technique classique de dilution des tubes. Cette propriété a été mise à profit dans la suite de notre travail.

III. REPRESENTATIVITE DE L'ECHANTILLONNAGE  
DANS LE CONTEXTE DU RESEAU DE SURVEILLANCE  
MICROBIOLOGIQUE

## **1. Enquête pilote à l'échelle locale**

### **1.1. Problématique**

Le but de l'enquête pilote est de permettre l'évaluation qualitative et/ou quantitative des différentes sources de variabilité spatio-temporelle de l'estimation de la densité en coliformes fécaux, mesurée dans les coquillages à l'échelle d'une station de prélèvement du réseau microbiologique. Comme Maul *et al.* (1985a) le soulignent, ou en d'autres termes Breittmayer (1978), "la relativisation des différentes sources de variation inhérentes au contexte spatio-temporel doit servir de référence pour la mise au point d'une stratégie de prélèvement".

Frontier (1983) définit le mot "représentatif" en ces termes : "un échantillonnage est représentatif d'une strate donnée s'il donne des estimations non biaisées des paramètres de distribution des descripteurs".

L'existence de micro-distributions très marquées à l'échelle locale peut induire un biais très important dans l'évaluation de la contamination dans le contexte du réseau microbiologique. Ainsi, les résultats obtenus doivent permettre d'évaluer la représentativité du prélèvement à l'échelle d'un point de prélèvement, préalable indispensable à l'étude des variabilités à des échelles supérieures.

Après une description du cadre de l'étude et des méthodes utilisées, une préenquête permet de hiérarchiser les différentes sources de variabilité devant servir à adopter la stratégie d'échantillonnage la plus efficace (optimum coût-précision) pour suivre l'évolution du niveau de contamination sur la zone.

### **1.2. Description de la zone d'étude et protocoles**

#### *1.2.1. Choix de la zone*

Les critères de choix de la zone d'étude répondent à divers impératifs :

- la proximité vis-à-vis du lieu des analyses microbiologiques,
- l'accès au gisement quel que soit le coefficient de marée,
- la topographie : l'aire doit être suffisamment plane pour qu'on puisse considérer que les coquillages sont exondés simultanément et donc que le facteur "durée d'exondaison" n'intervient pas dans les éventuelles différences spatiales observées,
- la forte densité en coquillages : leur présence est ainsi assurée sur toute la période d'étude,
- la zone doit pouvoir être identifiée au voisinage d'un point du réseau de surveillance microbiologique : d'une part pour disposer de données historiques et d'autre part pour pouvoir confronter les résultats obtenus par le réseau aux résultats de l'enquête pilote.

Compte tenu de ces différents critères, le site choisi est un gisement naturel de moules, se situant au large de la plage du Closillot, à Piriac/Mer (Loire-Atlantique), au Sud de la Baie de Vilaine (Fig. III.1). L'aire d'étude est un carré de 30 m de côté.

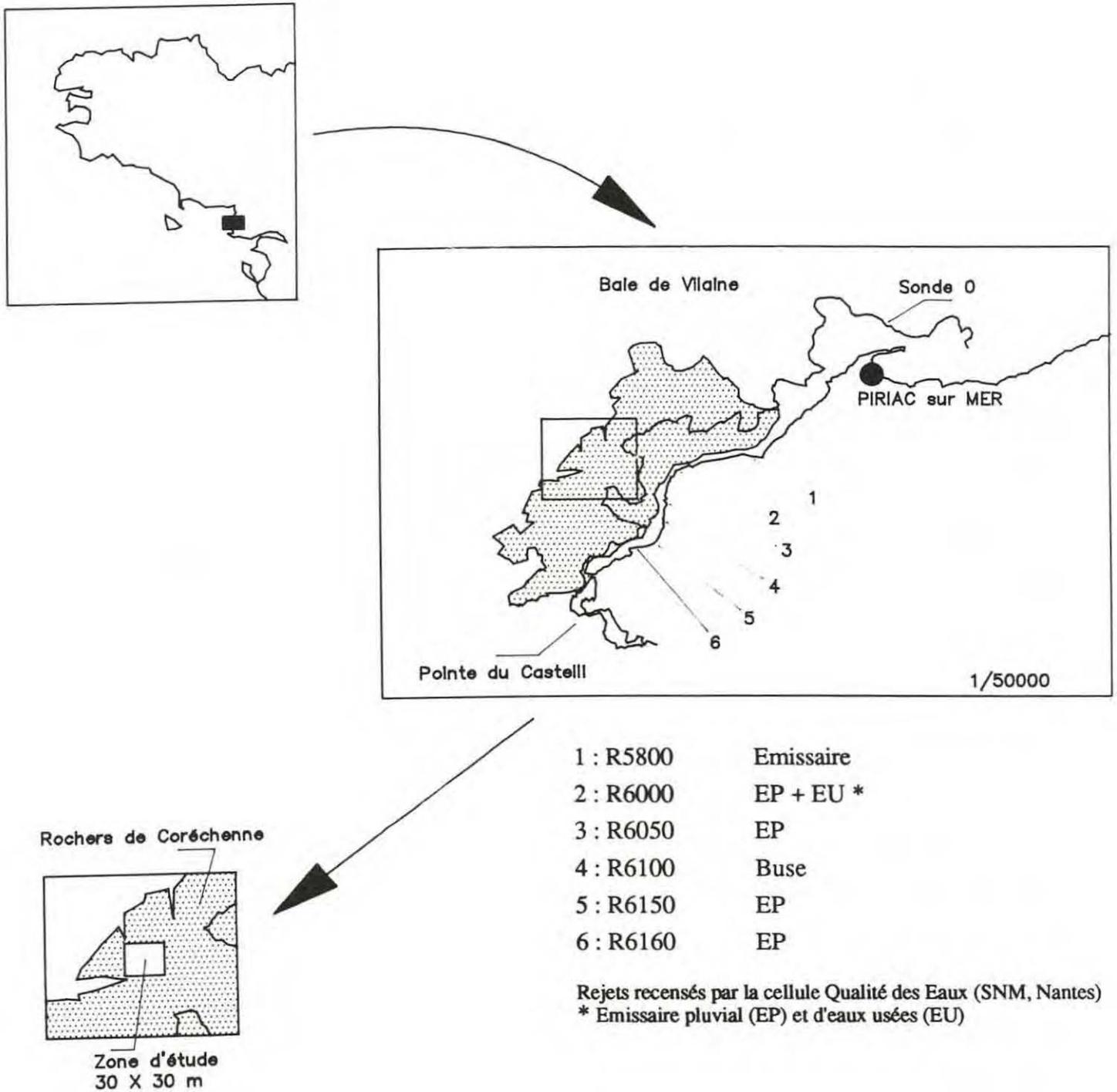
Concernant le réseau d'assainissement local, l'épuration des eaux usées de la commune de Piriac/Mer s'effectue à la station intercommunale (SIVOM<sup>1</sup> de Piriac/Mer et SIVOM de la Baule). Elle est constituée d'une unité physico-chimique suivi d'un lagunage de finition (125000 équivalents habitants). Les effluents sont rejetés par un émissaire en mer situé dans la baie du Scall, située plus au sud, hors figure. Les sources potentielles de pollution sur la zone ont été recensées (DDASS<sup>2</sup>, 1990; SNM<sup>3</sup> de Nantes, 1990). Les émissaires identifiés sont reportés sur la figure. Pour l'assainissement autonome, la DDASS (1990) note qu'il est difficile à mettre en place à cause de la nature rocheuse du sous-sol et offre peu de garanties quant à la protection des eaux marines, en particulier en période estivale, où la population de la commune croît de plus de 1000%: de 1300 à 15000-25000 habitants.

---

<sup>1</sup>Syndicat Intercommunal à Vocations Multiples

<sup>2</sup>Direction Départementales des Affaires Sanitaires et Sociales

<sup>3</sup>Service de Navigation et Maritime



**Fig. III.1** : Présentation générale de la zone d'étude (Piriac/Mer, Loire-Atlantique)

La pollution fécale sur la zone est significative comme en témoignent les données historiques obtenues pour l'année 1989 (Fig. III.2).

Dans la baie de Vilaine, les courants de marée et les courants résiduels sont faibles (respectivement 0,4 et 0,2 m.s<sup>-1</sup>). La circulation résiduelle est essentiellement dominée par le vent (De Nadaillac & Breton, 1986).

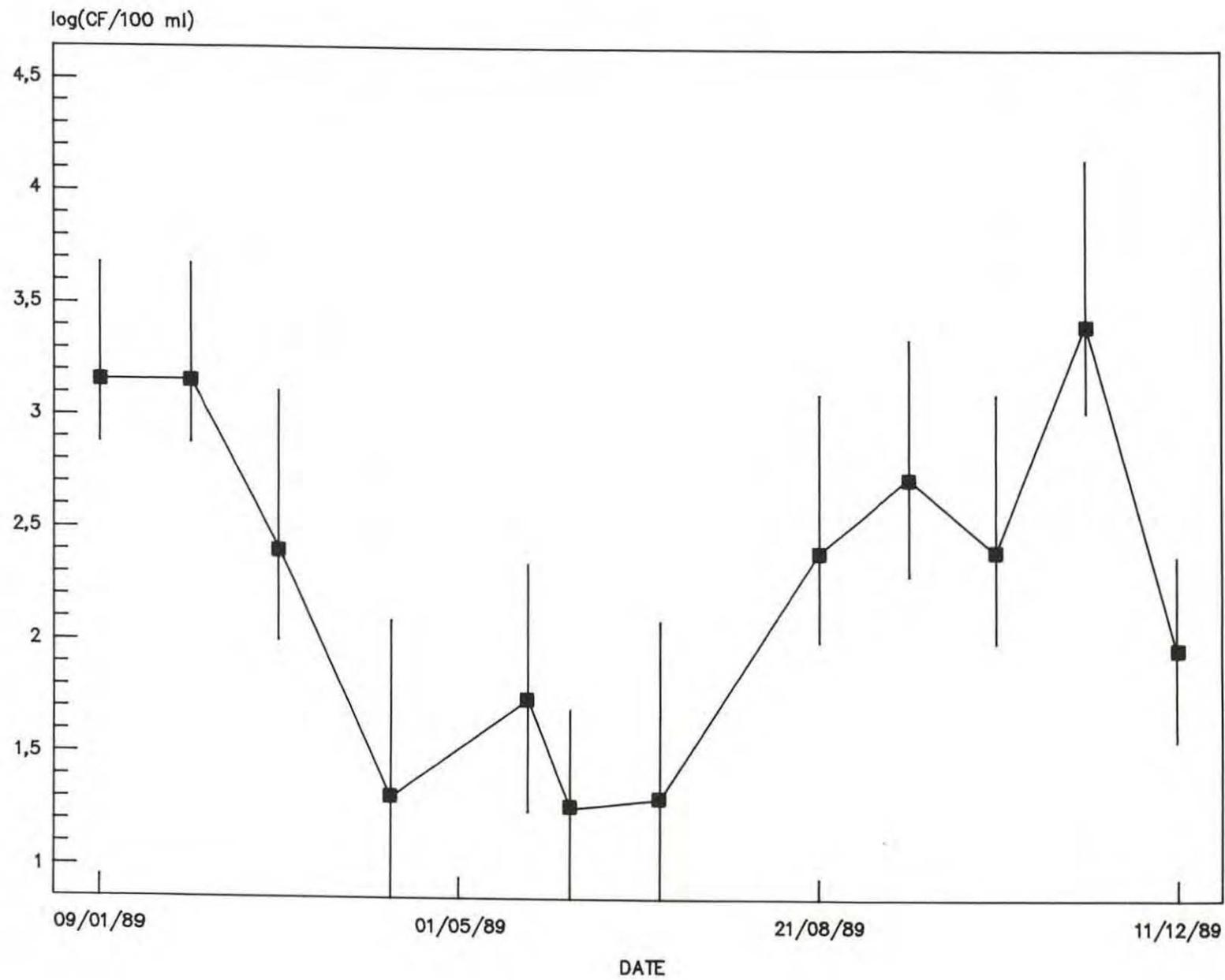
### 1.2.2. Variables recueillies

- variables physiques : à chaque sortie, ont été mesurées la température (thermomètre à mercure) et la densité (densimètre) de l'eau afin d'obtenir une mesure de la salinité, en se reportant à la table idoine (Leroy *et al.*, 1977). L'échantillon d'eau est prélevé en surface à proximité immédiate du gisement de moules. Sur le même échantillon, la turbidité est mesurée au moyen d'un turbidimètre *Hach*.

- Météorologie pour la période d'étude (Station météorologique de Saint-Nazaire) : l'insolation cumulée journalière en 1/10h, la hauteur des précipitations en mm, la direction (rose des vents de 36) et la vitesse (en m.s<sup>-1</sup>) du vent moyen maximal (valeur maximale du vent moyen calculé sur 6 mn).

### 1.2.3. Préenquête

Deux campagnes ont été réalisées les 4 et 19 décembre 1989. La stratégie d'échantillonnage sur le terrain correspond à un plan d'échantillonnage du troisième degré (Tab. III.1). Le tirage au sort des unités aux différents niveaux est effectué à l'aide d'un générateur de nombres pseudo-aléatoires. De plus, la variabilité analytique est estimée à partir de trois réplicats par prélèvement : pour chaque broyat réalisé à partir d'un échantillon unité (un prélèvement) trois mesures supposées indépendantes sont effectuées. La méthode conductancemétrique (§ 1.2.2.3) permet ainsi d'obtenir 54 mesures du logarithme décimal de la concentration en coliformes fécaux, descripteur retenu. Il s'agit également de la méthode de dénombrement indirect retenue pour l'enquête pilote proprement dite. Dans un premier temps, l'homogénéité spatiale des



**Fig. III.2 :** Données historiques : évolution de la concentration en coliformes fécaux pour l'année 1989 au Closillot, point de suivi du réseau de surveillance (REMI)

résultats est évaluée au travers du calcul des nombres optimaux d'unités aux différents degrés, où seuls les résultats de la campagne du 19 décembre ont été utilisés, celle du 4 décembre présentant une contamination très faible et donc trop de résultats nuls.

Tab. III.1 : Plan d'échantillonnage du 3<sup>ième</sup> degré (préenquête)

Niveau d'observation	Voisinage d'un point du réseau de surveillance	Prélèvements
Source de variabilité	Inter-stations	Inter-mesures
Nombre de répétitions	$n = 3; m = 2; k = 3$	3 (triplicats)
Structure du schéma d'échantillonnage	1	1.1.1 (1), (2) et (3) etc...
	...	
	2	
	3	
	3.2.1 3.2.2 3.2.3	

Suite aux résultats obtenus à partir de la stratégie prédéfinie, le calcul d'optimisation de l'effort aux différents niveaux dans l'espace nécessite l'estimation des différents coûts afférents.

Le coût total C se décompose en :

- $C_0$  : coût du trajet aller-retour sur le site d'échantillonnage; ce coût est indépendant de l'effort d'échantillonnage.

- $C_1$  et  $C_2$  : coûts relatifs respectivement aux supergroupes (premier niveau) et aux groupes (deuxième niveau), comprenant le repérage de l'unité considérée, l'accès à cette unité et l'inventaire des éléments au sein des groupes sélectionnés.

-  $C_3$  : coût relatif aux prélèvements, comprenant le repérage et l'accès aux unités du troisième niveau, mais également le coût des mesures effectuées à partir de ces éléments à la plus petite échelle spatiale.

$C_0$  est estimé à 75 F; compte tenu des dimensions de la zone, les coûts  $C_1$  et  $C_2$  sont négligeables; la préparation du broyat est estimée à 60 F et le coût d'une mesure à 5 F. Soit alors le coût total :  $C = 75 + nmk (60+r*5)$  F, si  $r$  est le nombre de répliqués par prélèvement. Le coût total d'une campagne de la préenquête est alors estimé à 1245 F.

A partir de ces estimations de coûts et de variances (Tab. III.2) et à l'aide des formules d'optimisation (§ I.2.4.3), on obtient  $m_{opt} = 1$  et  $k_{opt} = 1$  (les valeurs obtenues en réalité sont inférieures à 1). Ces valeurs reflètent l'homogénéité spatiale du descripteur à l'échelle considérée. La très faible variance inter-unités primaires peut également signifier la restitution satisfaisante de l'hétérogénéité globale par chaque unité primaire.

**Tab. III.2** : Tableau récapitulatif des différentes estimations pour le calcul d'optimisation (avec  $m' = 2$  et  $k' = 3$ )

	Coûts (F)	Variances
1 <sup>er</sup> degré	$c_1 = \epsilon$	0,0033
2 <sup>ième</sup> degré	$c_2 = \epsilon$	0,0393
3 <sup>ième</sup> degré	$c_3 = 75$	0,0325

Ceci nous autorise dans un deuxième temps à comparer la variabilité spatiale, tous niveaux confondus, à la variabilité analytique au moyen d'une analyse de variance hiérarchisée (Tab. III.3). Le lecteur trouvera les résultats théoriques afférents à cette technique classique dans Snedecor & Cochran (1957).

Tab. III.3 : Analyse de variance hiérarchisée (SCE : Somme des Carrés des Ecart; dl : degrés de libertés; CM : Carré Moyen; p : probabilité)

Source de variation	SCE	dl	CM	F	p
Entre prélèvements	1,303	17	0,077	0,621	0,85
Entre mesures	4,442	36	0,123		
Totale	5,745	53			

La variance inter-prélèvements étant inférieure à la variance inter-mesures, la totalité de la variabilité est ici exprimée par l'erreur de mesure. Ceci pourrait nous amener légitimement à n'effectuer qu'un seul prélèvement donnant lieu à plusieurs répliquats.

En matière d'échantillonnage de variables bactériologiques dans les milieux aquatiques, l'augmentation de la variance avec la moyenne d'un échantillon a souvent été observée (*e.g.* Troussellier *et al.*, 1986). Le niveau moyen de contamination était faible pour la préenquête (126 CF/100 ml); ainsi plusieurs prélèvements ont été effectués par sortie afin de pouvoir quantifier la variabilité spatiale à l'occasion d'éventuelles fortes contaminations.

#### 1.2.4. Protocole de l'enquête pilote

26 sorties ont été effectuées au cours de l'année 1990, soient deux par mois, une en vives-eaux et une en mortes-eaux (respectivement pour des coefficients de marée supérieurs et inférieurs à 70), mis à part au mois de mars où quatre sorties ont été réalisées dans le but d'obtenir une estimation plus précise de la variance intra-mensuelle. La préenquête a permis de définir les taux d'échantillonnage du plan du second degré utilisé : 3 unités primaires parmi 9 sont tirées au sort et pour le deuxième degré (correspondant au troisième degré de la préenquête) deux unités secondaires parmi 81 sont tirées au sort. Il vient (voir § I.2.4.3) :

$$f_1 = \frac{1}{3} \text{ et } f_2 = \frac{2}{81}$$

Un duplicat est effectué à partir de chaque prélèvement.

Le très faible taux d'échantillonnage pour le deuxième degré permet d'obtenir une échelle spatiale quasi-punctuelle pour le deuxième niveau, facilitant ainsi l'acte de prélèvement unitaire à cette échelle. Le tirage au sort des unités est effectué à l'aide d'un programme informatique pour chaque sortie; en effet pour des problèmes de disponibilité des coquillages, on ne peut échantillonner à points fixes, compte tenu du nombre de sorties effectuées (26).

D'autre part, nous avons procédé comme suit pour l'échantillonnage composite : à chacune des six stations du plan d'échantillonnage du second degré sont prélevées quatre moules, considérées comme représentant un élément. Les six éléments ainsi prélevés sont mélangés. Le mélange, soient 24 moules, est divisé en deux. Un broyat homogène est ensuite effectué à partir de chacun des deux sous-échantillons composites ainsi constitués. Deux mesures sont effectuées par broyat. Les paramètres intervenant dans l'estimation de la variance de l'estimateur (§ 1.2.4.4) sont donc :  $J = 6$  et  $K = 1$ ;  $\sigma_b^2$  et  $\sigma_c^2$  sont estimées égales à l'estimation de la variance inter-unités secondaires du plan d'échantillonnage du deuxième degré  $s_2^2$  pour la campagne d'échantillonnage correspondante.

#### *1.2.5. Traitement des données*

Le caractère aléatoire de la série est évalué au moyen du test des séquences (Siegel, 1956) : il s'agit ici de déterminer si le nombre de séquences de valeurs supérieures ou inférieures à la médiane peut être considéré comme aléatoire (hypothèse nulle) ou non (hypothèse alternative). Les variabilités mensuelles et intramensuelles sont estimées et comparées au moyen d'une analyse de variance hiérarchisée. L'influence du coefficient de marée est évaluée au moyen d'un test du  $\chi^2$ , après codage en réponse/non-réponse, du fait d'un nombre important de valeurs en deçà du seuil de détection.

Disposant d'une série courte, il nous est apparu non nécessaire de recourir aux méthodes classiques de comparaison des séries telles que celles faisant appel aux modèles auto-régressifs; de plus l'hypothèse de stationnarité requise avec ces outils est délicate à admettre avec nos

données bactériologiques. La liaison du descripteur avec les facteurs météorologiques (pluie et insolation) est alors testée au moyen du coefficient de corrélation sur les rangs de Spearman (Lebart *et al.*, 1982).

### 1.3. Résultats

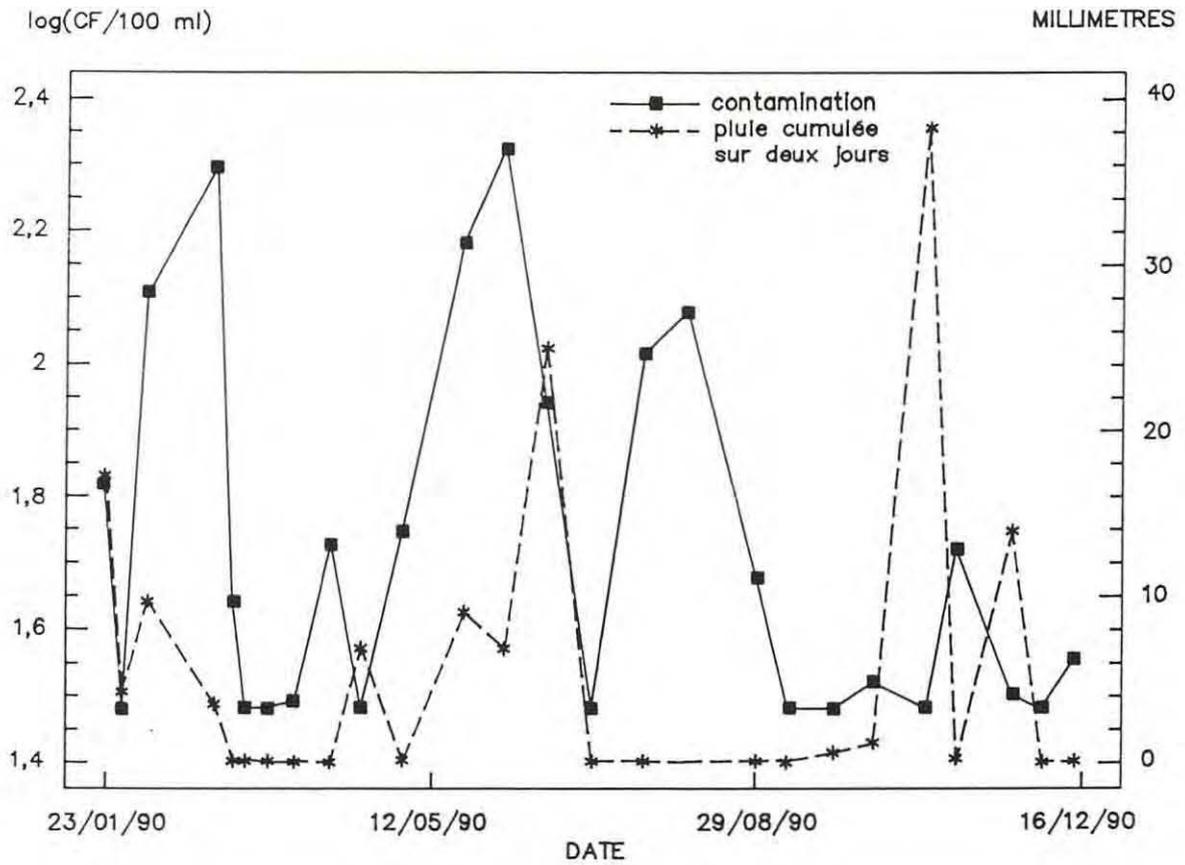
On ne peut rejeter l'hypothèse selon laquelle les observations se répartissent aléatoirement de part et d'autre de la médiane : on ne distingue pas de tendance significative dans la séquence des observations au seuil de 5% (médiane = 1,60 log(CF/100 ml); statistique  $z = -0,60$ ;  $p = 0,55$ ).

L'observation des composantes de variance fait ressortir la part plus importante prise par la composante de variance intramensuelle et par la composante de l'erreur (Tab. III.4), composée essentiellement de la variabilité analytique, comme il a pu être mis en évidence dans la préenquête.

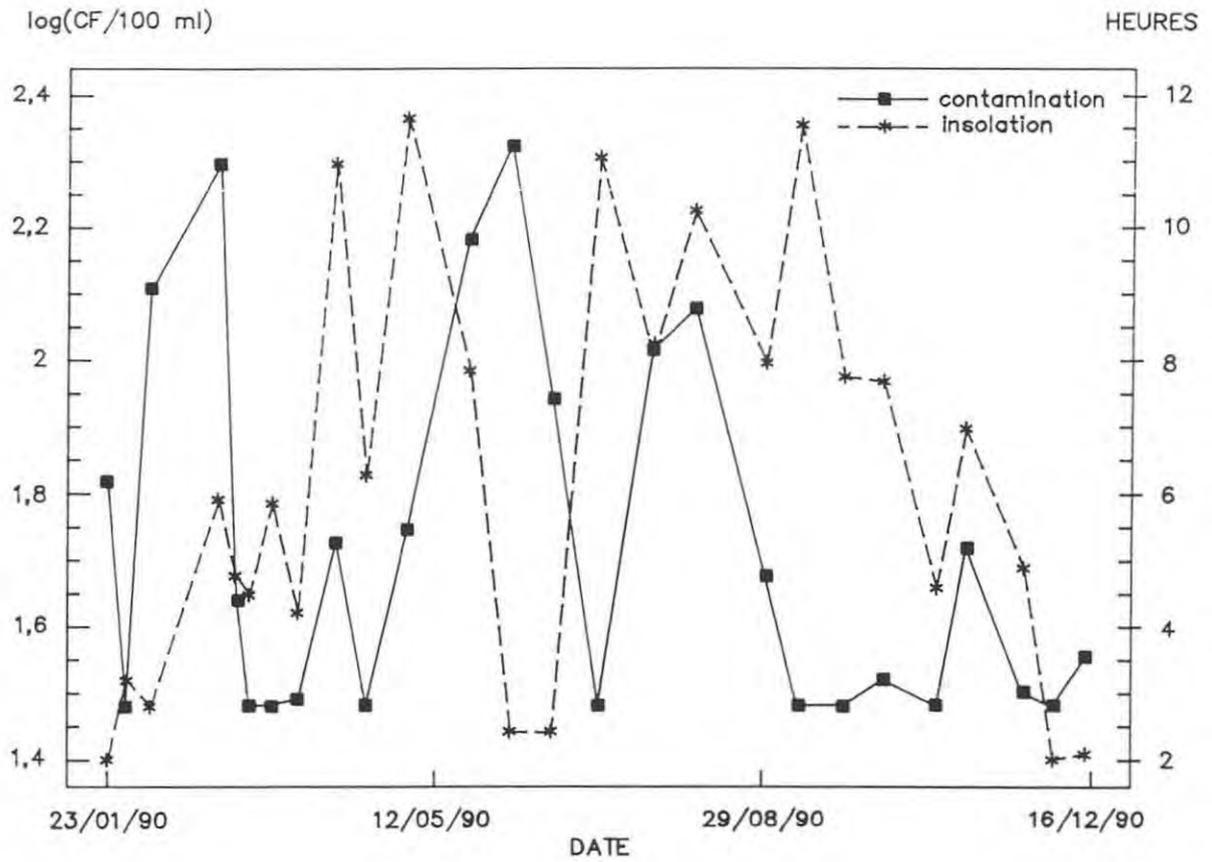
**Tab. III.4** : Analyse de variance hiérarchisée (SCE : somme du carré des écarts; D.L. : degrés de liberté; CM : carré moyen; % : pourcentage de la composante de variance à la variance totale)

Source de variation	SCE	D.L.	CM	Composantes de la variance	%
Mois	8,74	11	0,795	0,030	20
Intra-mois	5,20	12	0,434	0,063	42
Résiduelle	6,83	120	0,057	0,057	38

Le test du  $\chi^2$  est non significatif ( $\chi^2 = 1,27$ ) : on ne met pas en évidence de dépendance entre le coefficient de marée et la colimétrie (Tab. III.5; les chiffres entre parenthèses indiquent les fréquences théoriques).



**Fig. III.3** : Evolutions de la contamination (concentration en coliformes fécaux) et de la pluviométrie au cours de la période d'étude



**Fig. III.4** : Evolutions de la contamination (concentration en coliformes fécaux) et de l'insolation au cours de la période d'étude

**Tab. III.5 :** Tableau de contingence 2 X 2 (C = coefficient de marée)

	Détection	Non détection
C > 70	39 (35,5)	39 (42,5)
C < 70	32 (35,5)	46 (42,5)

Par ailleurs, il n'apparaît pas de dépendance entre les observations et les facteurs météorologiques (Tab. III.6 & Fig. III.3 et III.4), pas plus qu'on ne met en évidence de liaisons avec la salinité et la turbidité (résultats non présentés).

**Tab 3.6 :** Résultats des corrélations de rang de Spearman (PL1 : pluie du jour; PL2 : pluie cumulée sur deux jours;  $r_s$  : coefficient de corrélation de Spearman et p la probabilité associée)

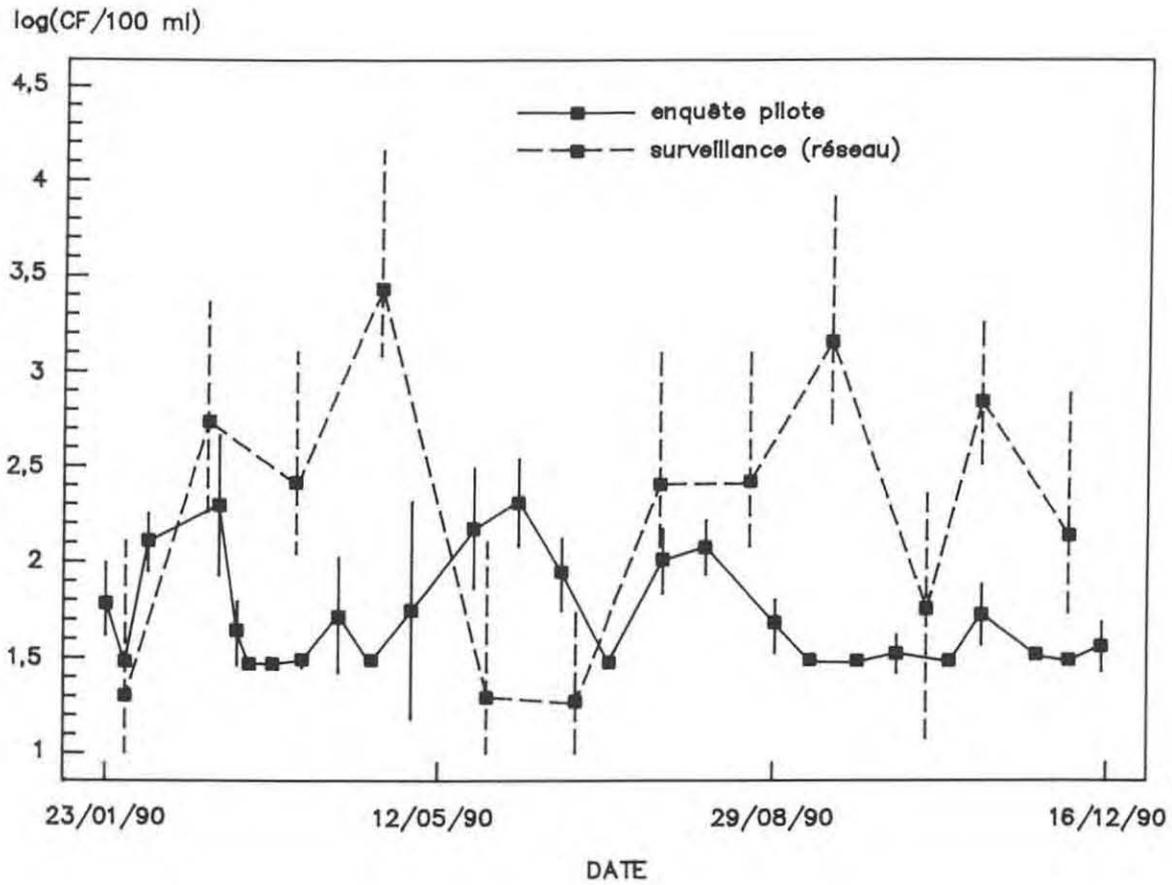
	insolation ( $10^{-1}$ h)	PL1 ( $10^{-1}$ mm)	PL2 ( $10^{-1}$ mm)
$r_s$	0,07	0,11	0,13
p	0,79	0,66	0,61

L'évolution de la concentration moyenne obtenue à partir des résultats de l'échantillonnage par degré présente un profil remarquablement différent de celui obtenu dans le cadre du réseau de surveillance (Fig. III.5). La comparaison à l'évolution de la moyenne obtenue à partir de l'échantillonnage composite est naturellement plus satisfaisante (Fig. III.6).

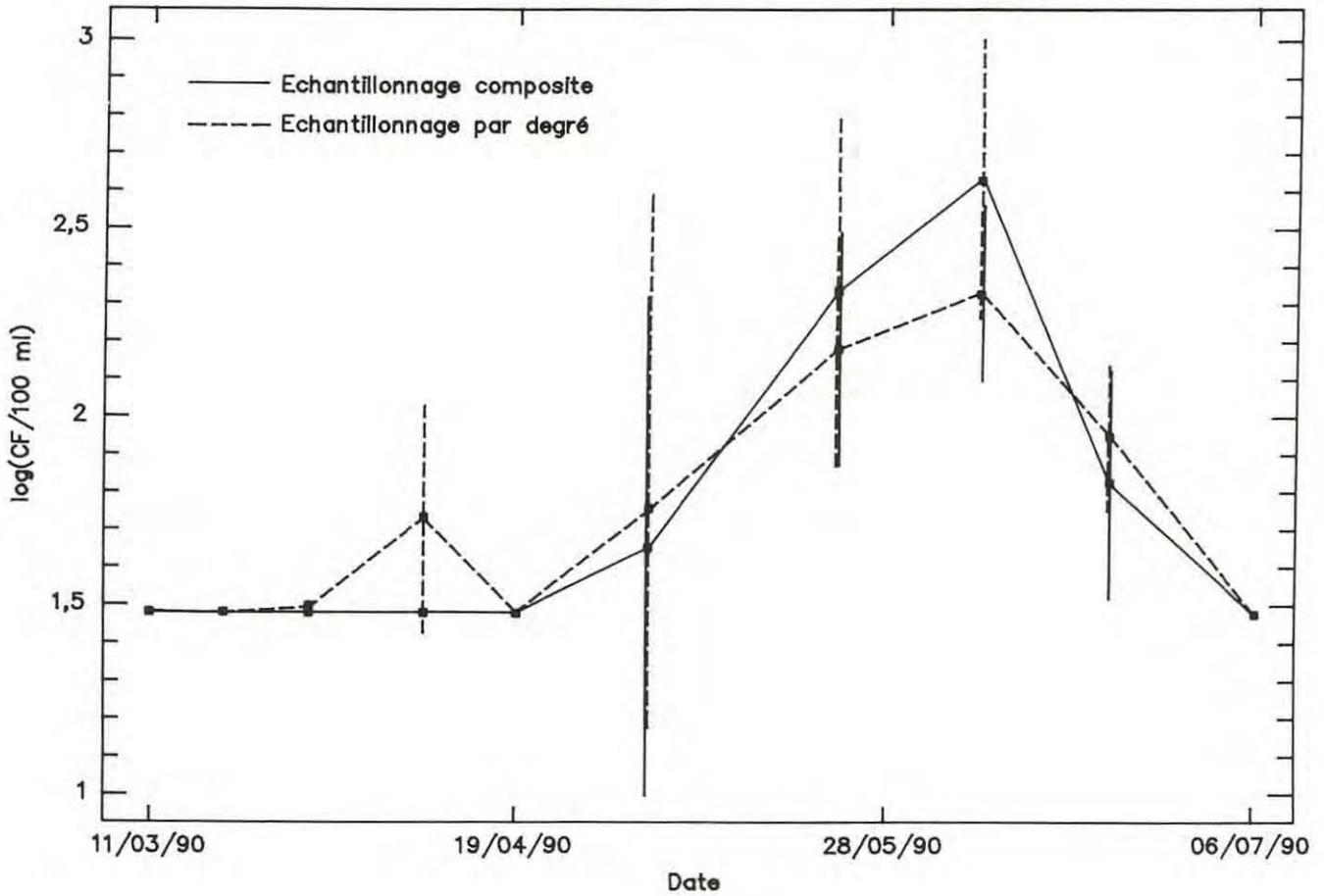
## 1.4. Discussion

### 1.4.1. Intérêt du plan d'échantillonnage à plusieurs niveaux

Le plan d'échantillonnage du deuxième degré pratiqué au cours de l'enquête pilote est de mise en oeuvre facile à l'échelle étudiée. L'échantillonnage de l'unité du deuxième niveau est facilité par le choix d'un taux d'échantillonnage idoine; en effet, l'aire de l'unité correspondant environ à 2 m<sup>2</sup>, ce niveau est alors considéré comme représentant l'échelle ponctuelle. On s'affranchit ainsi de l'inventaire exhaustif des éléments du dernier niveau, nécessaire en théorie.



**Fig. III.5** : Evolutions comparées de la contamination fécale (concentration en coliformes fécaux) sur le site d'étude estimées par le réseau de surveillance (une mesure NPP par mois) et lors de l'enquête pilote (six prélèvements et deux mesures conductancémétriques par prélèvement par quinzaine)



**Fig. III.6 :** Evolutions comparées de la contamination fécale (concentration en coliformes fécaux) sur le site d'étude estimée à partir d'un plan d'échantillonnage du deuxième degré (six prélèvements, deux mesures conductancémétriques par prélèvement) et dans le cas de la construction d'un échantillon composite (deux sous-échantillons composites et deux mesures conductancémétriques par sous-échantillon)

Ce type de plan peut être appliqué à des échelles plus importantes : Gros et Hamon (1989) ont, par exemple, utilisé ce plan comme procédure de tirage aléatoire à l'intérieur de strates sur une zone de bouchots.

#### *1.4.2. Niveau de contamination*

On ne distingue pas de tendance à l'augmentation de la colimétrie en période estivale. Ceci semble cohérent avec le débouché sur le site d'émissaires d'eaux pluviales principalement, logiquement responsables de pollution sur le site en cas de pluie. Néanmoins, ce point n'a pas été vérifié hors période sèche et plus précisément, il n'y a pas coïncidence entre les fortes pluviométries et d'éventuels pics de contamination, pas plus qu'une corrélation négative significative n'a été relevée avec l'insolation, considérée comme un facteur important de la mortalité naturelle des bactéries en milieu marin (Breittmayer & Gauthier, 1979; Gauthier *et al.*, 1991; El-Sharkawi *et al.*, 1991). En outre, la valeur du coefficient de marée ne semble pas avoir d'influence sur le niveau de contamination, explicable sans doute par la sédimentologie et le faible hydrodynamisme locaux. La zone est soumise à des apports faibles et le non discernement de tendances (test des séquences non significatif) en atteste le caractère sporadique. Les faibles courants de marée remettent en cause l'hypothèse selon laquelle la pollution proviendrait de la plage du Canot de Sauvetage, zone plus contaminée (DDASS Loire-Atlantique, 1990).

Il est clair que l'influence d'un facteur donné ne pourra s'exercer que pour un niveau moyen de contamination éloigné du seuil de détection de la méthode analytique. Or la moitié environ des mesures effectuées sont inférieures à ce seuil (Tab. III.5) et la contamination maximale observée sur l'année est de  $2,86 \log(\text{CF}/100 \text{ ml})$  soit 724 CF/100 ml. On pourrait considérer que ces résultats remettent en cause le choix de la zone pour l'étude de fluctuations temporelles. Nous avons déjà énuméré les raisons de ce choix (§ 1.2.1) : notamment, les données historiques du site montraient une contamination significative avec des pics situés en hiver (Fig. III.2). Des pics significatifs, cette fois printaniers et automnaux, sont également

observés dans la séquence des observations du réseau microbiologique. Comment expliquer une dissemblance aussi importante entre les deux séries ? On peut émettre l'hypothèse d'une forte interaction entre les résultats obtenus et les méthodes analytiques utilisées : sous certaines conditions environnementales, telle méthode, de par le milieu de culture utilisé, serait plus à même de faciliter la revivification de bactéries stressées que telle autre (Troussellier, comm. pers.). Ceci semble corroboré par des valeurs supérieures obtenues par la méthode conductancemétrique uniquement en été. L'interaction du facteur insolation sur la colimétrie serait alors une explication. Il conviendrait de pouvoir vérifier ces hypothèses. Quoiqu'il en soit, la forte sensibilité du profil de contamination à la méthode analytique retenue doit inciter à interpréter les données du réseau avec prudence. Ce problème surgit avec une acuité toute particulière si l'on considère la portée sanitaire de la concentration en coliformes fécaux, indicateurs de la présence probable de germes pathogènes.

#### *1.4.3. Les différentes sources de variabilité*

La variabilité spatiale à l'échelle de ce point de prélèvement ne semble être que le reflet de la variabilité analytique comme le montre l'analyse des composantes de la variance réalisée sur les résultats de la préenquête. Troussellier *et al.* (1986) mettent en garde contre une interprétation abusive des fortes valeurs relatives de variance inter-mesures, en constatant que ce n'est pas tant la forte imprécision de la mesure qui est en cause - la variance résiduelle absolue restant ici faible (0,123 log(CF/100 ml)) - mais surtout la très faible variabilité du descripteur à des échelles supérieures, ce qui est le cas pour notre étude.

A cette échelle la zone paraît donc homogène et ne montre pas de phénomènes de microdistributions, pour des résultats supérieurs au seuil de détection, souvent évoqués dans les estimations d'abondance en germes fécaux (*e.g.* Lebaron *et al.*, 1990). Une conséquence pratique de ce résultat est qu'il est aussi précis et moins onéreux d'effectuer des répliquats à partir d'un même broyat que de multiplier les échantillons sur le terrain, contrairement à ce que

Baleux & Troussellier (1982) ont préconisé pour l'échantillonnage dans les sédiments lagunaires de l'étang de Prévost, à petite échelle.

L'échantillonnage composite apparaît comme une alternative intéressante : le coût est trois fois moindre que pour l'échantillonnage du deuxième degré, pour une précision environ 1,6 fois inférieure. Des réserves doivent être émises sur ce résultat : on n'a pas accès à une véritable estimation de la variance inter-éléments et on ne dispose pas d'un mélange parfait ( $\sigma_{\alpha}^2 \neq 0$ ); cette dernière condition serait mieux remplie si on prélevait plus de moules par élément pour un même nombre de sous-échantillons composites. Il serait intéressant d'utiliser cette stratégie dans le cadre des études de zone à plus grande échelle, si l'objectif est plus d'obtenir un niveau moyen de la contamination à moindre coût que la quête des maxima sur le site, à cause du lissage réalisé par l'échantillonnage composite.

La variabilité analytique est imputable *a priori* à différentes causes, en pratique indissociables l'une de l'autre :

- La variabilité de la mesure entre coquillages. La préparation de l'échantillon à analyser requiert l'utilisation d'une dizaine de moules, choisies de calibre identique, limitant ainsi les fluctuations d'un individu à l'autre. En effet, l'activité de filtration de l'animal et l'accumulation résultante prennent en compte de nombreux facteurs, dont la taille de l'individu. A côté des variations physiologiques existent également des comportements individuels totalement aléatoires et donc par nature imprévisibles, comportement dans le milieu et hors du milieu avant l'analyse. A notre connaissance, aucune donnée n'existe sur la variabilité inter-individus s'agissant de mesure de la colimétrie dans les coquillages.

- La variabilité liée au comportement des cellules bactériennes dans l'animal et dans le milieu de culture. Par exemple, on peut penser que la mesure du temps de détection est affectée par ces phénomènes encore complètement aléatoires, déjà décrits par Bousquet et Devoucoux (1982) dans le cas de l'analyse des eaux destinées à la consommation humaine.

- La variabilité liée à la répartition des bactéries dans le broyat, problème déjà évoqué précédemment; elle est caractérisée par la précision de l'estimateur NPP et elle peut être sous-estimée dans le cas d'une répartition agrégative des bactéries.

Notons que l'utilisation d'huîtres comme bioaccumulateurs requiert un nombre moins important d'individus (4 à 5), que l'utilisation des coques (une vingtaine d'individus) ou des moules (une dizaine d'individus). Ceci pourrait être la cause d'une variabilité analytique plus importante, et même pour des huîtres de même calibre étant entendu que le rapport pondéral chair/coquille est reconnu plus variable chez *Crassostrea gigas*, huître creuse.

Divers auteurs (Laurec *et al.*, 1979; Maul *et al.*, 1982; Lebaron *et al.*, 1989; Ibanez, 1990) ont souligné les liaisons existant entre les différentes échelles spatio-temporelles. Ainsi, les fluctuations temporelles observées peuvent n'être que le reflet des variations dans l'espace ou dans notre cas le fait uniquement de la variabilité analytique : le "bruit" résultant peut masquer totalement l'évolution réelle de la contamination dans l'espace. Ce n'est pas le cas dans notre étude qui montre des variations significatives (Fig. III.5) avec une composante de variance temporelle non négligeable (Tab. III.4). La non prise en compte de la variabilité intra-mensuelle peut masquer l'existence de fluctuations type saisonnières et suggère que la fréquence d'échantillonnage retenue peut ne pas être adaptée à la détermination de tendances sur de futures séries à long terme. En outre qu'en est-il de la variabilité journalière, voire horaire ? Divers auteurs ont montré qu'elles pouvaient être non négligeables (Plusquellec, 1984; Lebaron *et al.*, 1989; Catherine, 1992). En particulier, Catherine (1992) observe une remarquable concordance entre le niveau de contamination et le coefficient de marée dans une zone de marais (Bassin du Mes, Loire-Atlantique).

### 1.5. Conclusion

La connaissance des différentes sources de variabilité à l'échelle locale, c'est-à-dire au voisinage d'une station du réseau de surveillance, permet de jauger la validité de l'échantillonnage pratiqué dans ce contexte.

A cette échelle, la variabilité analytique prime sur les variabilités spatiales, dont les valeurs absolues sont particulièrement faibles, allant ainsi dans le sens d'une intégration de la variabilité spatiale réalisée par les mollusques bivalves. Cette hypothèse était préalable à l'utilisation de ces organismes sentinelles. L'échantillonnage composite apparaît comme économiquement rentable s'agissant de connaître le niveau moyen de contamination sur une zone donnée.

Cette étude ne permet pas de trancher pour l'échelle temporelle la mieux adaptée au descripteur; on note cependant de plus fortes variations de contamination dans un mois qu'entre mois. La fréquence pourrait ainsi se révéler insuffisante pour déceler d'éventuelles fluctuations saisonnières. La comparaison des évolutions "enquête pilote" et "réseau de surveillance" met en exergue la forte sensibilité du profil de contamination à la méthode analytique préconisée.

## 2. Estimation de la contamination fécale à l'échelle d'un secteur

### 2.1. Examen des données du réseau de surveillance

Les données examinées ici sont collectées par les laboratoires côtiers de l'IFREMER qui assurent le prélèvement sur le terrain et l'analyse bactériologique au laboratoire.

#### *2.1.1. Problématique*

On désire appréhender la représentativité d'un échantillon à l'échelle d'un secteur de surveillance. Ceci ne peut s'effectuer qu'au travers d'un effort important réalisé au travers de plusieurs campagnes sur un même site. Les résultats obtenus ne pourront en outre être extrapolés à l'ensemble des secteurs surveillés, chacun ayant sa spécificité, la distinction la plus importante étant sans conteste l'opposition d'une mer sans marée, la mer Méditerranée, à l'Océan Atlantique, et à la Manche, mer macrotidale.

Au préalable, il semble souhaitable de pouvoir tirer quelques enseignements de l'examen de séries chronologiques obtenues dans le cadre du réseau de surveillance, à l'échelle d'un secteur. En particulier, on s'efforcera de répondre aux questions suivantes :

Pour chaque point d'un même secteur, l'évolution temporelle observée n'est-elle pas le reflet de la variabilité analytique? Peut-on alors identifier des "accidents" dans la série? Peut-on considérer que les profils d'évolution des points pris deux à deux sont similaires? Existe-t-il des tendances? Il ne s'agit pas ici de traiter exhaustivement les données du réseau de surveillance, mais de proposer quelques méthodes robustes pour un début d'interprétation de la contamination observée à l'échelle d'un secteur.

#### *2.1.2. Méthodes d'évaluation de la contamination à l'échelle d'un secteur*

Au préalable et pour chaque point du secteur, une représentation graphique de la contamination fécale, en échelle logarithmique, permet de visualiser la série des observations.

Les éventuels accidents dans les profils de contamination sont identifiés au moyen du test de comparaison entre deux estimations NPP, à l'aide des tables de discrimination des résultats NPP (§ II.2.2.2). Dans le cas d'une différence significative au seuil de 5%, une flèche pointe sur l'accident de contamination.

Naturellement, un tel procédé est sans "mémoire" : il ne prend pas en compte la séquence des événements; il est complété par un test des séquences de part et d'autre de la médiane; ce test est sensible aux tendances à court terme dans les séries. Il est effectué à l'aide du logiciel Statgraphics sur PC.

En outre, la variance de la série observée est comparée à une estimation de la variabilité analytique, après transformation logarithmique des données. Celle-ci est estimée à partir d'un échantillon de résultats caractéristiques simulés. La simulation s'opère de la façon suivante : on calcule la moyenne géométrique des  $n$  observations composant la série. A partir de cette estimation,  $n$  résultats caractéristiques sont simulés selon la méthode précédemment décrite (§ I.2.3.5), conduisant à  $n$  estimations NPP. On forme le rapport de la variance "observée" à la variance "simulée" et un test F permet finalement la comparaison des deux estimations de la variance au seuil de 5%, pour les nombres de degré de liberté correspondants ( $n-1$ ). Si le test se révèle significatif, on dira qu'il existe un effet "temps" ou que les fluctuations temporelles sont suffisamment importantes pour se dégager du bruit analytique. Une étude de comparaison inter-laboratoires (Beliaeff & Miossec, non publ.) montre que la variabilité des résultats intra-laboratoires est du même ordre que la variance de résultats ainsi simulés, ce qui laisserait à penser que le modèle de Poisson est satisfaisant pour décrire l'abondance des germes mesurés dans un broyat de coquillages.

Pour l'ensemble des points constituant le maillage d'un bassin, on désire savoir si les différentes séries observées peuvent être considérées comme homogènes, dans le but ultime de n'en considérer qu'une seule. Pour cela, on s'attachera non pas à une quelconque comparaison

des valeurs entre elles, mais plutôt à une comparaison des profils deux à deux. La méthode utilisée est l'analyse des lignes polygonales (De Carolis in Ibanez, 1991).

Notre choix a porté sur cette méthode, déjà utilisée pour l'étude intra-site de séries chronologiques de benthos (Ibanez, 1991), car elle permet de comparer des séries de tailles différentes et pour des dates d'observation différentes.

L'équation d'une ligne polygonale est :  $y = \sum_{r=1}^k a_r |x - x_r|$ , avec les coefficients  $a_r$  :

$$a_1 = \frac{1}{2} \left( \frac{y_2 - y_1}{x_2 - x_1} + \frac{y_1 + y_k}{x_k - x_1} \right)$$

$$a_j = \frac{1}{2} \left( \frac{y_{j+1} - y_j}{x_{j+1} - x_j} - \frac{y_j - y_{j-1}}{x_j - x_{j-1}} \right)$$

$$a_k = \frac{1}{2} \left( \frac{y_1 + y_k}{x_k - x_1} - \frac{y_k - y_{k-1}}{x_k - x_{k-1}} \right)$$

On vérifie facilement l'identité de l'équation de la ligne polygonale donnée ci-dessus avec l'équation de la droite passant par les deux points  $(x_j, y_j)$  et  $(x_{j+1}, y_{j+1})$  :

$$y = \frac{y_{j+1} - y_j}{x_{j+1} - x_j} x + \frac{x_{j+1}y_j - x_jy_{j+1}}{x_{j+1} - x_j}$$

En outre, on remarquera que, pour tout  $j$  différent de 1 et de  $k$ , le coefficient  $a_j$ , qui par ailleurs ne correspond pas à une pente de la ligne polygonale, se calcule comme la différence de deux pentes consécutives.

La forme de toute ligne polygonale est donnée par la distribution statistique des  $a_r$  successifs. Leur comparaison est effectuée à l'aide de deux tests non paramétriques, décrits par Siegel (1956) :

- Le test U de Mann & Whitney : l'hypothèse nulle est que les deux séries ont la même distribution. Par construction du test, l'hypothèse alternative est la différence statistique entre les deux médianes. Le test de la médiane, utilisé également par Ibanez (1991) teste également

l'identité des médianes et est relativement peu puissant pour les tailles d'échantillon considérées ici; dans ces conditions ce test est redondant.

- Le test des séquences de Wald-Wolfowitz : sous l'hypothèse nulle, les deux séries ont la même distribution; l'hypothèse alternative est que les deux séries diffèrent pour quelque paramètre que ce soit (tendance centrale, dispersion, asymétrie,...).

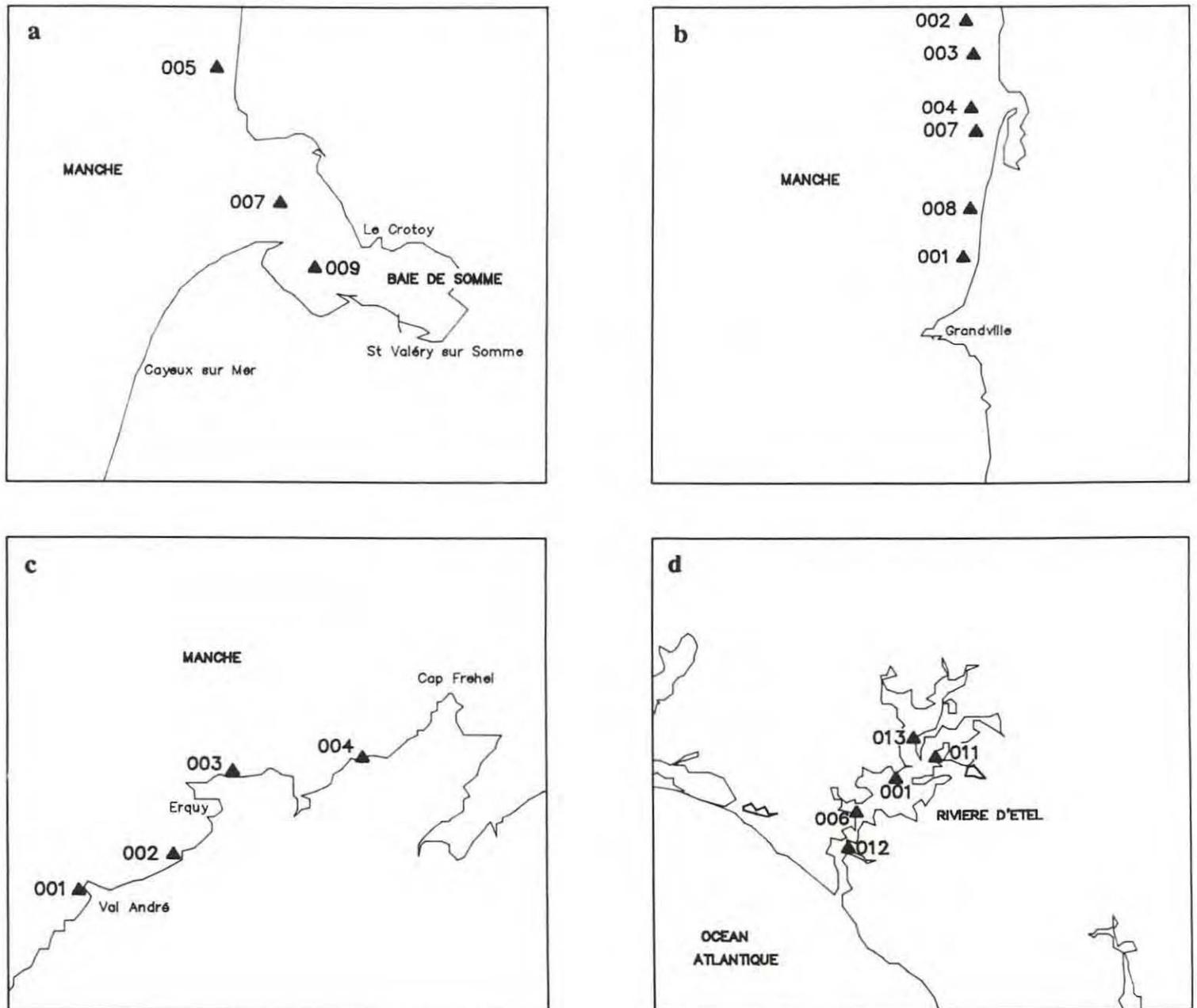
Ibanez (1991) considère que deux séries de benthos ne peuvent être considérées comme hétérogènes si aucun des tests ne fait rejeter l'hypothèse nulle pour un niveau de signification de 5 ou 10%. Nous adopterons cette règle pour un niveau de 5%.

### 2.1.3. Des exemples : résultats et discussion

Pour chaque secteur (Fig. III.7a, b, c et d) est représentée l'évolution de la contamination aux différentes stations constituant le secteur, avec l'identification des pics. De plus est donné le tableau des résultats de la comparaison entre variance temporelle ( $s^2$ ) et variance résiduelle ( $s_R^2$ ) et du test des séquences. Enfin, la diagonale supérieure de la matrice des résultats de la comparaison des profils d'évolution est fournie.

Les symboles utilisés dans les tableaux sont :

NS	Non significatif au seuil de $p=0,05$
*	F significatif au seuil de $p=0,05$
**	F significatif au seuil de $p=0,01$
***	F significatif au seuil de $p=0,001$
XX	Différence significative de profils entre deux stations



**Fig. III.7 :** Représentation des secteurs étudiés avec le positionnement des points de surveillance (a : Baie de Somme; b : Hauteville-Granville; c : Erquy; d : Rivière d'Étel) - D'après Miossec & Raffin (1990)

Tab. III.6 : Résultats du secteur 006 (Baie de Somme)

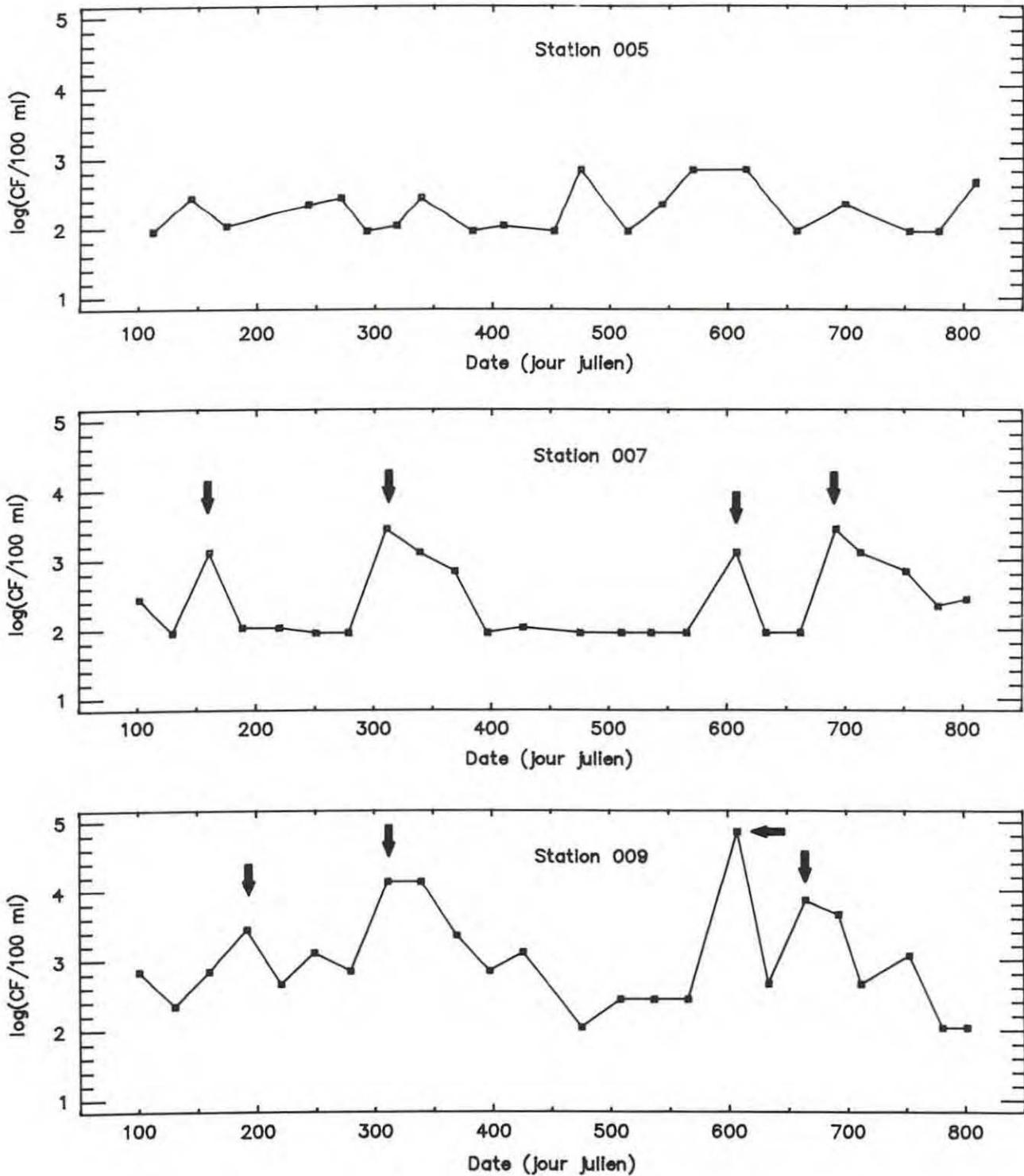
Station	$s^2$	$s_R^2$	ddl	F	Runs
5	0,109	0,128	23	< 1	NS
7	0,304	0,088	23	3,45 **	NS
9	0,531	0,102	23	5,21 ***	NS

	7	9
5	XX	XX
7		XX

Pour le secteur 006, l'observation des évolutions de la contamination fécale du point 9 au point 5 (Fig. III.8) montre un gradient important correspondant au sens amont-aval dans la baie de Somme (Fig. III.7a). On passe ainsi de pics très significatifs en amont à, en aval, une évolution reflétant plus un bruit "analytique" que de véritables variations temporelles, comme le montre le résultat du test de comparaison des variances pour la station 5.

La comparaison 2 à 2 des trois profils met en évidence les fortes dissemblances constatées visuellement : le test de comparaison est toujours significatif. Ainsi on rejette l'hypothèse selon laquelle les lignes brisées considérées, prises deux à deux, présenteraient simultanément des pentes successives de mêmes médianes, de mêmes tendances générales et dont les distributions seraient pareillement asymétriques.

S'agissant de suivre la décroissance amont-aval de la contamination, d'un impact maximal sur la zone jusqu'à l'extinction du phénomène, le choix des stations paraît judicieux, même si on peut se demander si la station 5 ne devrait pas être plus amont. Par contre, la détermination d'un niveau moyen de la qualité bactériologique du milieu peut être remise en cause dans la mesure où nous n'avons pas affaire aux mêmes distributions statistiques d'un point à un autre, et où il conviendrait, par exemple, de pondérer les résultats obtenus à chaque station par un coefficient prenant en compte l'étendue de la zone "sale" (station 9), de la zone



**Fig. III.8 :** Evolution de la contamination fécale - Secteur 006 (Baie de Somme)

"moyennement sale" (station 7) et de la zone "propre" (station 5), pour une zone totale d'aire restant à définir.

Remarquons néanmoins que si les niveaux sont différents pour les stations 7 et 9, les pics de contamination sont en phase; ainsi une solution moins drastique et plus pragmatique que de scinder la zone en trois stations indépendantes consisterait à écarter la station 5 et à considérer comme représentative la série moyenne obtenue à partir des stations 7 et 9.

Tab. III.7 : Résultats du secteur 021 (Hauteville-Granville)

Station	$s^2$	$s_R^2$	ddl	F	Runs
1	0,494	0,095	15	5,20 **	NS
2	0,346	0,099	19	3,49 **	NS
3	0,324	0,116	17	2,79 *	NS
4	0,249	0,103	21	2,42 *	NS
7	0,233	0,091	21	2,56 *	NS
8	0,279	0,079	14	3,53 *	NS

	2	3	4	7	8
1				XX	XX
2				XX	
3					
4					
7					

Les profils observés pour le secteur 021 (Fig. III.9) ne présentent pas de tendance générale, mais plutôt une succession de pics significatifs de contamination, caractérisés par des fluctuations temporelles discernables de la variabilité analytique. Compte tenu des résultats des tests et du positionnement géographique des stations (Fig. III.7b), la zone circonscrivant les stations 2,3 et 4 sera considérée comme non hétérogène.

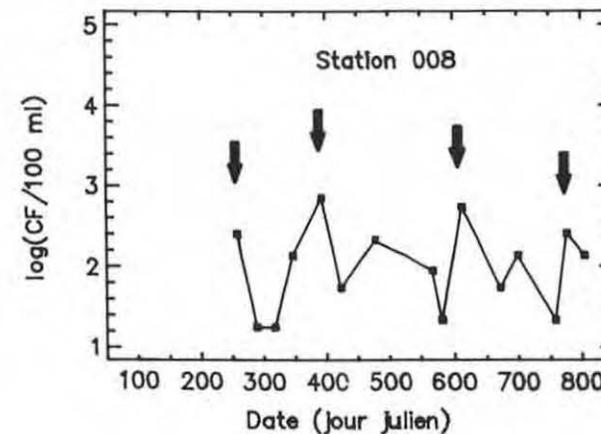
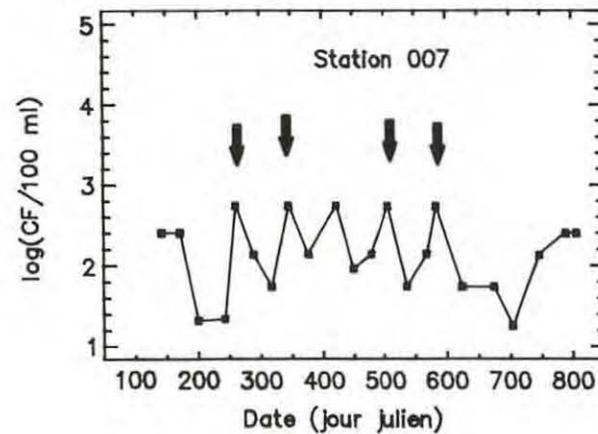
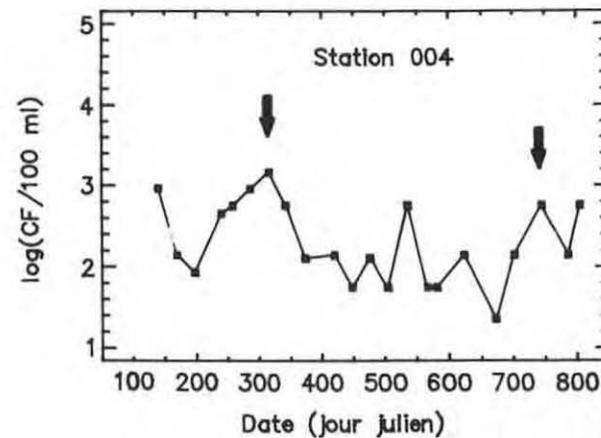
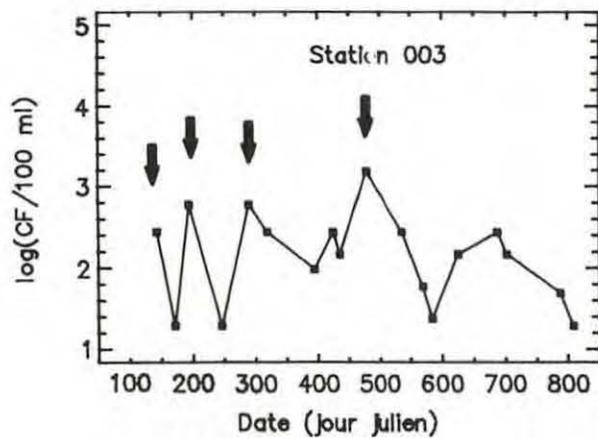
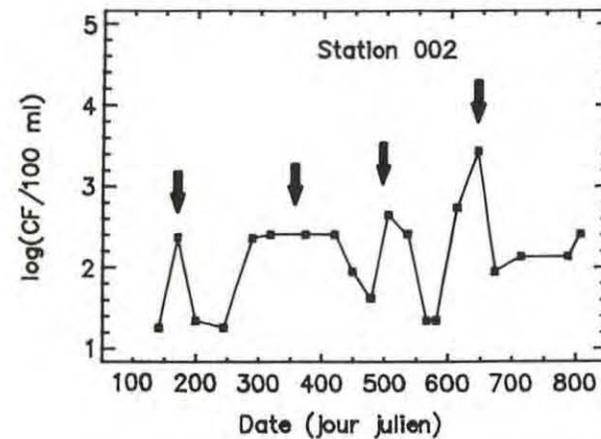
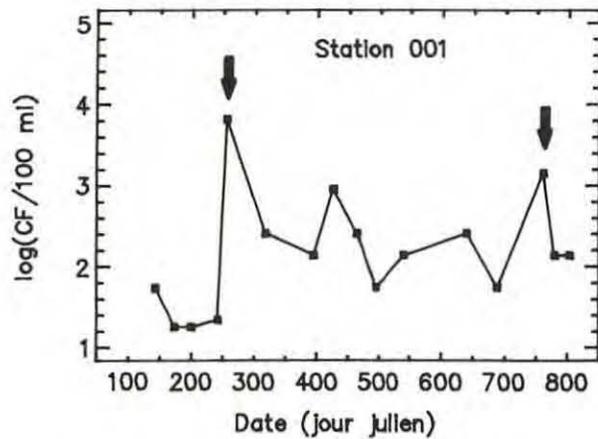


Fig. III.9 : Evolution de la contamination fécale - Secteur 021 (Hauteville-Granville)

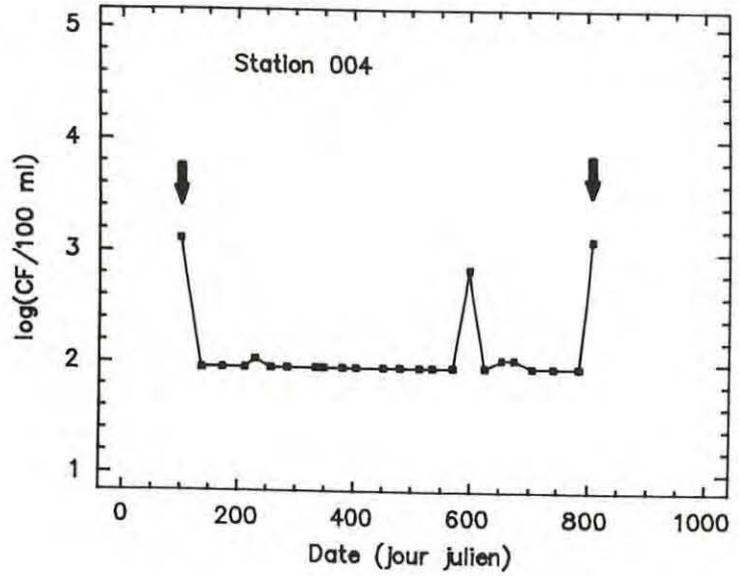
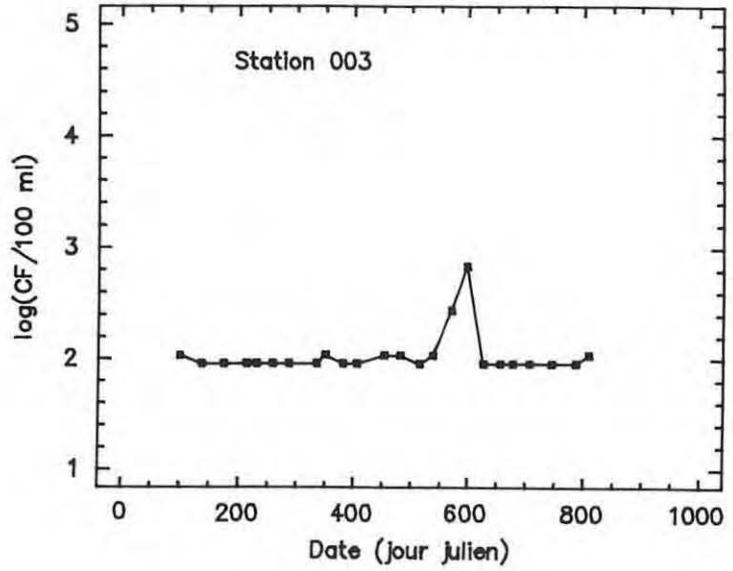
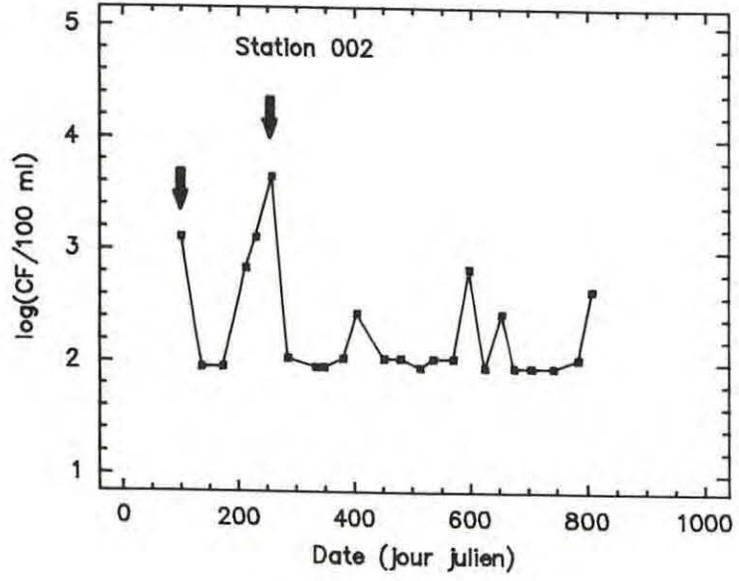
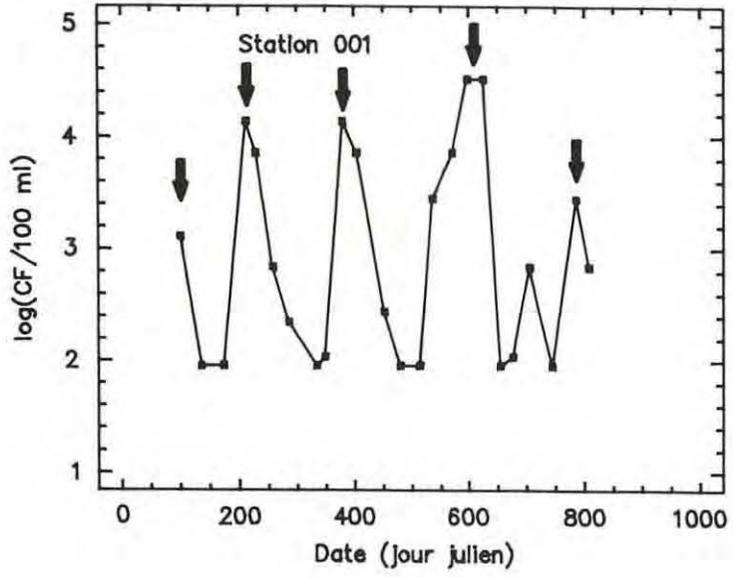
Tab III.8 : Résultats du secteur 026 (Erquy)

Station	$s^2$	$s_R^2$	ddl	F	Runs
1	0,869	0,105	23	8,28 ***	NS
2	0,254	0,081	23	3,14 **	NS
3	0,040	0,075	23	< 1	NS
4	0,130	0,148	23	< 1	NS

	2	3	4
1		XX	XX
2			
3			XX

Seules les stations 1 et 2 du secteur d'Erquy (026) témoignent d'une variabilité temporelle significative et plus particulièrement la station 1. Pour les stations 3 et 4 on se situe le plus souvent en deçà du seuil de détection analytique. On ne peut ainsi considérer que l'ensemble de ces quatre stations soit représentatif de la contamination du bassin d'Erquy, ce qui ressort de l'examen des profils d'évolution (Fig. III.10) et du tableau des comparaisons des profils deux à deux. L'observation de la carte correspondante (Fig. III.7c) montre que, grossièrement, la maille retenue ici est de l'ordre de 5 km. Ceci suggère que les dimensions de la zone où on pourrait tenter avec le même effort de déterminer un niveau de contamination moyen devraient être circonscrites à un domaine plus homogène.

Fig. III.10 : Evolution de la contamination fécale - Secteur 026 (Erquy)



Tab. III.9 : Résultats pour le secteur 047 (Rivière d'Etel)

Station	$s^2$	$s_R^2$	ddl	F	Runs
1	0,520	0,095	23	5,47 ***	NS
6	0,456	0,076	23	6,00 ***	NS
11	0,222	0,117	23	1,90 (NS)	NS
12	0,286	0,030	23	9,53 ***	NS
13	0,234	0,076	23	3,08 **	NS

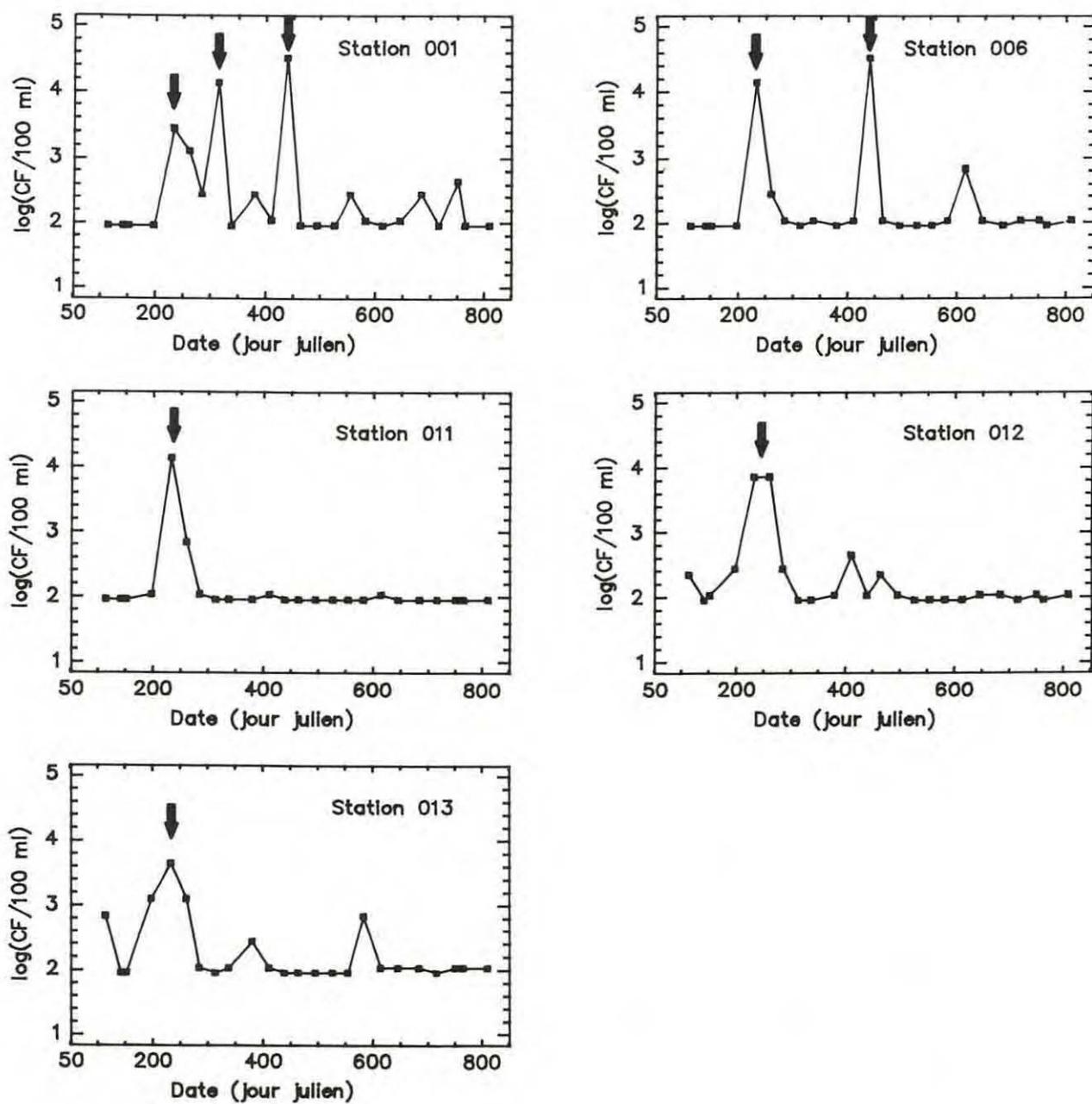
	6	11	12	13
1	*	XX	*	XX
6		*	*	XX
11			*	XX
12				*

Les fluctuations observées à la station 11 du secteur 047 (Fig. III.11) sont plutôt de l'ordre d'un bruit de fond témoin d'une faible contamination, malgré l'observation du pic estival que l'on retrouve à toutes les stations. La matrice de comparaison indique que le profil de la station 13 est très différent des autres; à un degré moindre la station 11 apparaît également comme détachée. La détermination d'un niveau moyen à partir des résultats aux stations 1, 6 et 12 peut alors être envisagé; elles sont localisées suivant un gradient amont-aval (Fig. III.7d) avec une maille approximative de 2 km.

### Discussion

La recherche de sous-zones homogènes est un préalable nécessaire à la recherche d'une concentration moyenne sur un secteur donné

El-Shaarawi *et al.* (1981) ont défini des sous-zones homogènes à partir d'une hypothèse de distribution de Poisson dans chaque zone ainsi déterminée. L'indice de dispersion de Fisher testant l'ajustement des données à cette distribution fournit alors un critère objectif de partition spatiale. L'analyse des lignes polygonales permet de s'affranchir de cette hypothèse et en outre



**Fig. III.11 :** Evolution de la contamination fécale - Secteur 047 (Rivière d'Etel)

ne prend pas en compte directement les valeurs absolues des estimations mais la succession des pentes.

Ce test nous fournit un moyen pour juger du bien fondé de la localisation de telle ou telle station, et donc *in fine* une ébauche de jugement sur la représentativité de l'échantillonnage à l'échelle d'un bassin de surveillance, dans l'objectif de la détermination d'un niveau moyen non biaisé à cette échelle. Les exemples exhibés ici soulignent la trop grande maille spatiale utilisée dans ce but : il est illusoire en terme d'information de vouloir obtenir une moyenne à partir de stations présentant des différences d'évolutions aussi marquées, comme le montre par exemple l'examen du secteur 006.

On peut réussir à isoler dans un bassin un sous-groupe de stations où les évolutions sont jugées non différentes, comme pour la rivière d'Etel (secteur 047); dans ce cas, la question importante est de savoir si l'image obtenue à chaque station résulte d'une même source de contamination, caractérisée par un flux de pollution unique, ou bien les évolutions ne sont-elles pas le reflet de flux bien différenciés mais dont les caractéristiques sont régies par les mêmes causes. Par exemple, trois stations situées indépendamment l'une de l'autre à l'embouchure de trois rivières, distantes de plusieurs km, vont présenter des profils non significativement différents du fait que les flux de pollution ont pour origine la même affluence estivale, ou la même pluviométrie. Dans ces conditions, la valeur moyenne ne représente non pas une estimation de l'impact moyen de la contamination fécale sur la totalité du secteur, mais plutôt la réponse moyenne sur un type de site particulier à un facteur donné, en faisant abstraction des éventuelles macro-structures sur l'ensemble de la zone. Il convient ainsi de s'interroger sur le sens des valeurs moyennes obtenues.

Notons que jamais le test des séquences n'est significatif, ce qui montrerait le caractère sporadique de la contamination - on ne distingue pas de tendances à court-terme. Si des fluctuations saisonnières existent, la fréquence mensuelle n'est probablement pas apte à la mettre en évidence.

#### 2.1.4. Détermination d'un niveau moyen et optimisation de l'effort dans une zone homogène

Afin d'illustrer notre propos, les calculs seront effectués sur les données de la rivière d'Etel pour le secteur, dont on suppose l'homogénéité, et constitué des stations 1, 6 et 12 (Fig. III.7d).

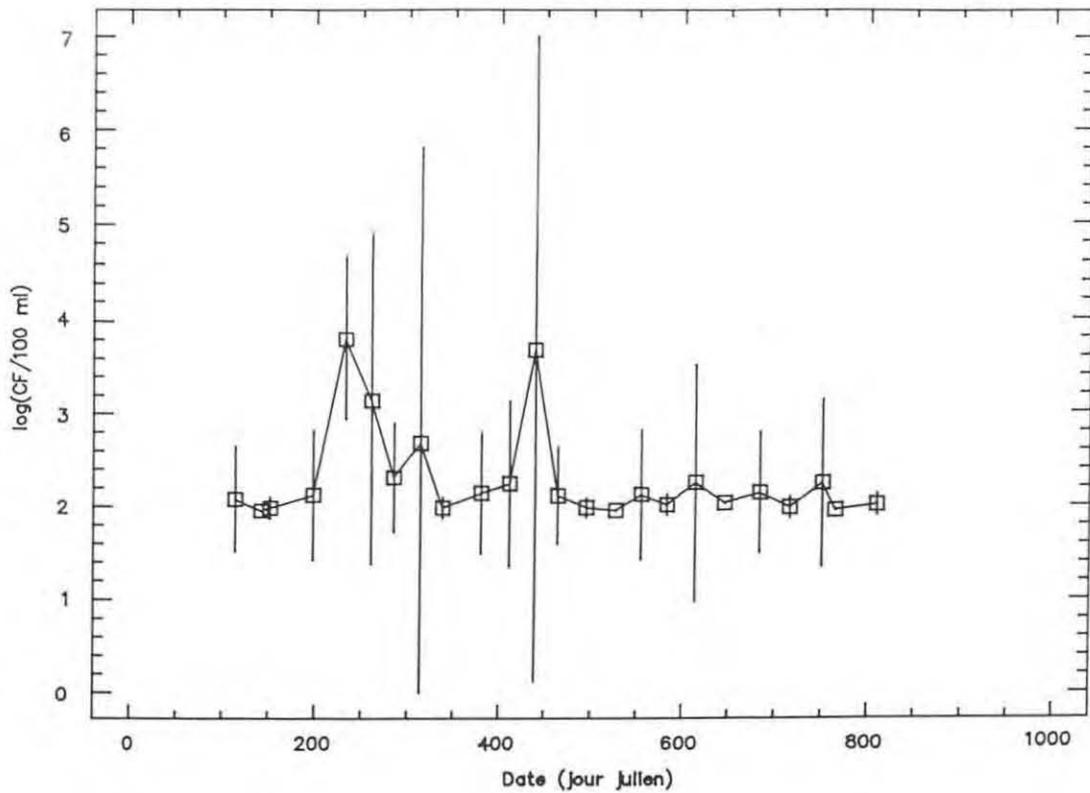
Chaque observation de la série chronologique représentative de la zone ou sous-zone identifiée préalablement est une estimation de la contamination moyenne dans l'espace et pour la période considérée (ici, une seule valeur par mois). Cette estimation sera d'autant plus réaliste que la partition aura été judicieuse, et ceci doit avoir pour conséquence une diminution de la variance à l'intérieur de chaque sous-zone et corrélativement y induire une augmentation de la représentativité de l'échantillon.

#### Choix d'un estimateur

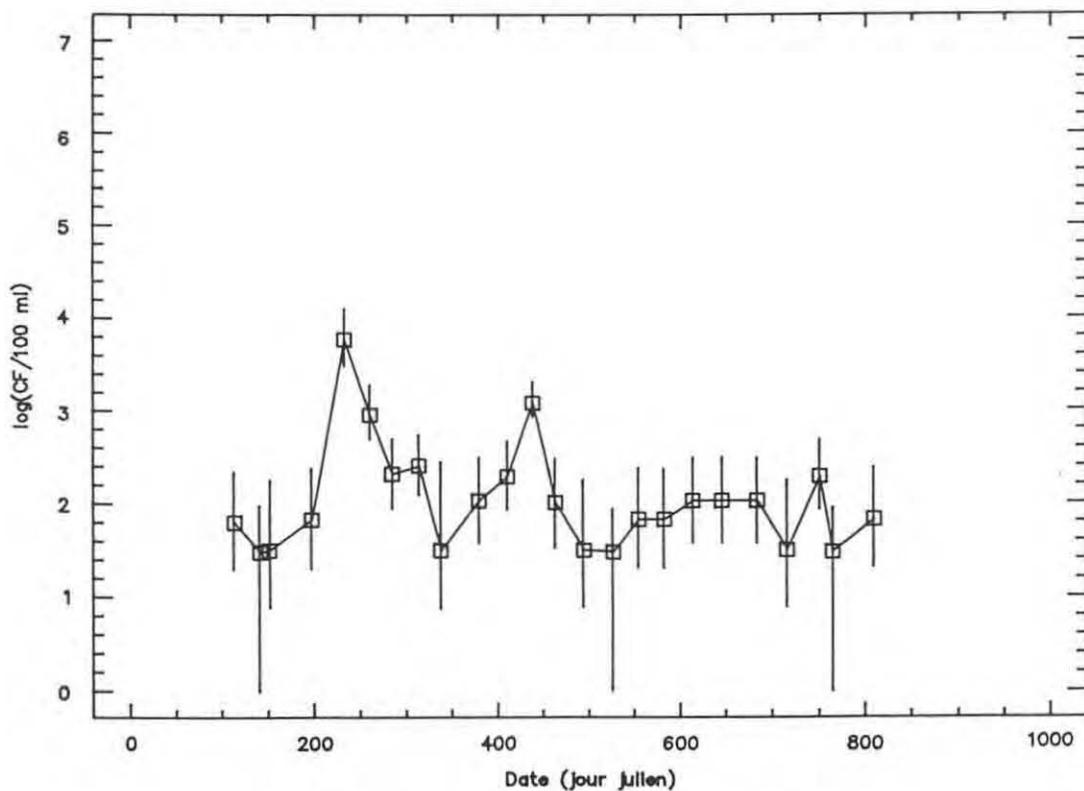
Deux estimateurs de valeur centrale ont été utilisés :

La moyenne arithmétique des logarithmes décimaux des concentrations observées aux stations (Fig. III.12), ou, ce qui est rigoureusement identique, le logarithme décimal de la moyenne géométrique, est compatible avec une hypothèse de distribution lognormale des dénombrements microbiens. Cette hypothèse est fréquemment rencontrée dans la littérature (Thomas, 1952; Pipes *et al.*, 1977; El-Shaarawi *et al.*, 1981; Sørensen *et al.*, 1991). Dans le cas des dénombrements indirects par la méthode de dilution des tubes avec un faible nombre de tubes, elle peut être remise en question par le caractère fortement discret de la distribution de l'estimation NPP.

Disposant d'estimations en nombre limité, on s'expose à obtenir une variance nulle dès lors que les résultats des prélèvements à moyenner présentent le même résultat caractéristique; compte tenu de la forte imprécision de l'estimation ceci paraît paradoxal. De plus, quels



**Fig. III.12** : Profil de la contamination moyenne de la rivière d'Etel (Stations 1, 6 et 12) : moyenne arithmétique des logarithmes décimaux de la concentration en coliformes fécaux et intervalle de confiance à 95%



**Fig. III.13** : Profil de la contamination moyenne de la rivière d'Etel (Stations 1, 6 et 12) : estimation du "Nombre le Plus Probable" de coliformes fécaux dans 100 ml de l'échantillon initial (3 dilutions et 9 tubes par dilution) et intervalle bayésien à 95%

nombre utiliser en deçà du seuil de détection et pour un mauvais choix de gamme de dilution (tous les tubes sont positifs), cas pour lesquels il n'y a pas d'estimations NPP ?

L'estimation NPP correspondant au "nombre caractéristique somme" (Fig. III.13), obtenu en sommant les résultats caractéristiques individuels pour les dilutions respectives (§ I.2.3.5), pallie ces deux problèmes. L'estimation NPP résultante et l'intervalle bayésien à 95% correspondant sont obtenues par lecture dans la table NPP pour trois dilutions et neuf tubes par dilution et un ensemencement de 1 ml pour la suspension mère (annexe II.1).

Une hypothèse forte d'utilisation de l'estimateur NPP est que les bactéries sont réparties aléatoirement dans l'échantillon, ou en d'autres termes que la distribution des abondances en germes suit une loi de Poisson. En pratique, ceci est difficilement généralisable. Hormis la distribution requise, se pose le problème de résultats caractéristiques fortement invraisemblables, résultant de la "somme" de résultats eux très vraisemblables. Considérons l'exemple suivant :

Le 15 mars 1990 (jour julien 439), les résultats caractéristiques obtenus pour les stations 1, 6 et 12 ont été respectivement 3/3/2 (33000 CF/100 ml), 3/3/2 et 1/0/0 (108 CF/100 ml), donnant le résultat somme 7/6/4. La probabilité d'observer un tel résultat pour une concentration de 1190 CF/100 ml (estimation NPP correspondante) est égale à  $8 \cdot 10^{-8}$ . On remarque au passage que la concentration la plus vraisemblable pour ce résultat n'en est pas moins extrêmement improbable. Dans ce cas, l'intervalle bayésien sous-estime la forte hétérogénéité des résultats obtenus sur la zone pour la date considérée, qui est exprimée fortement dans le cas de la moyenne (Fig. III.12). On préférera alors la moyenne arithmétique des concentrations (après transformation logarithmique) à l'estimateur NPP.

Dans le cas de la rivière d'Etel, les prélèvements sont effectués aux mêmes dates. Si tel n'est pas le cas, il y a lieu de régulariser les différentes séries à moyenner. Ibanez (1991) propose la méthode de régularisation de Fox & Brown (1964). Si maintenant on veut obtenir

une série moyenne à partir d'un groupe de séries dont au moins deux présentent des profils significativement différents, on effectue une moyenne pondérée : pour une série donnée de ce groupe, le poids correspond au nombre de stations dont les profils respectifs ne lui sont pas significativement différents (Ibanez, 1991). Cette procédure permet de tenir compte de l'apport relatif de chaque station dans la construction de la série unique et a un intérêt opérationnel certain : soit un groupe de séries non significativement distinctes et soient deux stations d'une part significativement différentes entre elles et d'autre part non significativement différentes de chacune des stations de l'autre groupe. D'un côté on ressent une certaine gêne à effectuer une moyenne sur l'ensemble des stations, de l'autre se pose le problème du choix de la station éventuellement à écarter. La moyenne pondérée apporte une solution intellectuellement confortable pour ce cas : les deux stations sont conservées mais interviennent de façon moindre dans le calcul de la série par rapport au groupe "homogène".

Le problème de la détermination d'un niveau moyen dans le temps est tout à fait analogue : une moyenne n'a de sens que dans le cas où elle est effectuée sur un échantillon tiré d'une même distribution statistique, c'est-à-dire si on peut considérer que les observations successives se répartissent aléatoirement autour d'un niveau moyen ( $\mu$ ), avec une même variance ( $\sigma^2$ ) quelle que soit la fenêtre temporelle considérée (stationnarité d'ordre 2).

Le problème de la recherche de la stationnarité d'une série est équivalent à celui de l'homogénéité spatiale dans un secteur. El-Shaarawi (1981) propose des outils statistiques adaptés à l'optimisation de la fréquence d'échantillonnage dans le temps, disposant de deux stations de prélèvement.

#### Optimisation de l'effort d'échantillonnage spatial

Compte tenu de la couverture nationale réalisée pour le réseau de surveillance, l'allocation de l'effort dans une zone ou sous-zone particulière est fonction du budget disponible, c'est-à-

dire du nombre total de prélèvements et corrélativement d'analyses possiblement réalisables, pour une période annuelle.

Idéalement, c'est-à-dire sans contraintes de coût ou opérationnelles, on veut rechercher le nombre d'échantillons requis pour un degré de précision voulu. Classiquement, la détermination de cet effort passe par la connaissance des variances aux différentes échelles spatio-temporelles considérées. El-Shaarawi (1981), Baleux & Troussellier (1982) et Esterby (1982) ont montré la variation des caractéristiques des distributions de variables bactériologiques mesurées dans l'eau d'une campagne à l'autre. La contamination fécale sur une grande étendue résulte de l'apport de différentes sources à différents moments. Ceci est un inconvénient majeur pour une procédure d'optimisation fondée sur l'ajustement d'une distribution théorique à des données bactériologiques (*e.g.* Baleux & Troussellier, 1982). Nous reviendrons sur ce point pour la conformité de résultats bactériologiques vis-à-vis d'une norme (§ 3.4.3).

Une conséquence importante est que la valeur de la variance est dépendante du niveau de contamination, ces paramètres étant calculés avant transformation. Après avoir examiné une grande série d'échantillons, Taylor (1961) a proposé de formaliser la relation entre la moyenne ( $\hat{\mu}$ ) et la variance ( $\hat{\sigma}^2$ ) observées sous la forme :

$$\hat{\sigma}^2 = a\hat{\mu}^b$$

Troussellier *et al.* (1984) indiquent que cette relation, appelée loi de puissance de Taylor, traduite en échelle logarithmique permet d'apprécier :

- le type de distribution des abondances,
- 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,

- le degré de représentativité en fonction de la concentration moyenne en bactéries et du nombre de mesures.

La connaissance de cette relation permet, par exemple, de déterminer le nombre de stations (n) à échantillonner dans un site et un secteur donné. Soit P la précision, d'une manière générale P peut être définie comme le pourcentage d'erreur dans l'estimation de la moyenne de la variable (Troussellier *et al.*, 1989), ici la concentration en coliformes fécaux. Nous choisissons d'exprimer P en terme d'erreur standard :

$$P = \frac{\widehat{SE}(\hat{\mu})}{\hat{\mu}} \text{ où } \widehat{SE}(\hat{\mu}) = \frac{\hat{\sigma}}{\sqrt{n}}$$

Il vient :

$$n = \frac{\hat{\sigma}^2}{P^2 \hat{\mu}^2}$$

A partir de la loi de puissance, on peut écrire :

$$\hat{\sigma}^2 = 10^{\log(a)+b \log(\hat{\mu})}$$

Et ainsi, pour  $\hat{\mu}$  observé, on obtient la relation :

$$n = \frac{10^{\log(a)+b \log(\hat{\mu})}}{P^2 \hat{\mu}^2}$$

Comme le précise les auteurs, l'application de cette relation à l'optimisation de l'échantillonnage n'est valable que pour un niveau de variabilité. Ainsi, seule la variabilité intra-site est prise en compte ici, dans la mesure où l'on ne dispose pas des variabilités intra-mensuelles. Une optimisation mettant en jeu plusieurs niveaux de variabilité serait résolue par la connaissance des composantes de la variance totale (Laurec *et al.*, 1979; Troussellier *et al.*, 1982), telle qu'utilisée pour l'enquête pilote (§ 1.2.4).

Les stations 2, 3 et 4 du secteur 021 (Fig. III.7b) sont considérées comme représentant la même zone homogène, et les observations à ces stations pour une date donnée constituent l'échantillon à partir duquel sont calculées moyenne et variance. Pour la période 04/89 à 04/91, et au moyen d'une régression au sens des moindres carrés (Fig. III.14;  $r = 0,78$ , pour 17 couples considérés), on obtient la relation,

$$n = \frac{10^{[2,816 \log(\hat{\mu}) - 2,049]}}{P^2 \hat{\mu}^2}$$

qui donne la taille d'échantillon requise pour une précision  $P$  et un niveau *a priori* moyen  $\hat{\mu}$  pour la zone.

Par exemple, 4 stations permettent d'obtenir une précision de 30% pour une contamination moyenne de 100 CF/100 ml, ce qui situe la moyenne théorique inconnue entre 40 et 160 CF/100 ml.

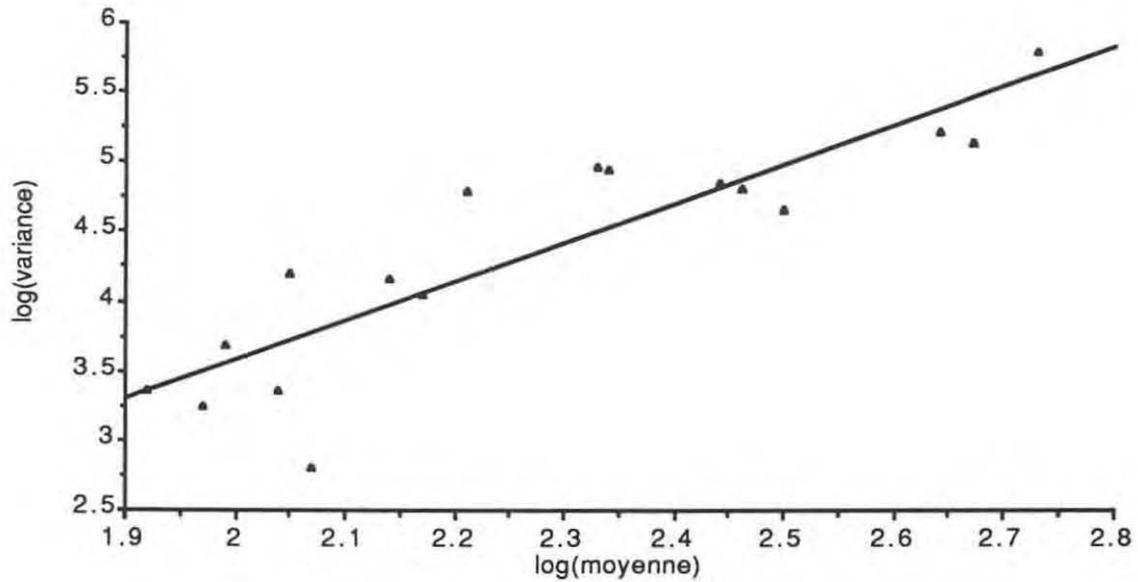
**Tab. III.10 :** Taille d'échantillon nécessaire pour une précision  $P$  et pour un niveau moyen  $\hat{\mu}$  donnés (en CF/100 ml). Secteur de Hauteville-Granville (21), stations 2, 3 et 4.

	$\hat{\mu} = 100$	300	500	1000
$P = 10 \%$	38	94	142	251
30 %	4	10	16	28
50 %	2	4	6	10
100 %		1	1	3

Cette optimisation ne concerne que des niveaux moyens ne s'écartant pas de la gamme ayant permis la détermination des paramètres de la loi de puissance de Taylor (a et b).

Disposant d'une valeur de  $b$ , la transformation est (Taylor, 1961)  $Z = X^p$  avec  $p = 1 - \frac{b}{2}$ .

Pour  $b = 1$  et  $b = 2$ ,  $p = 0$ , les transformations racine carrée et logarithmique sont



**Fig. III.14** : Régression linéaire de la variance en moyenne [ $\log(\hat{\sigma}^2) = 2,82 \log(\hat{\mu}) - 2,05$ ] en application de la loi de puissance de Taylor. Les variances et moyennes sont calculées à chaque date pour les données brutes (en CF/100 ml), obtenues aux stations 2, 3 et 4 du secteur de Hauteville-Granville, secteur 021 du réseau de surveillance microbiologique, pour la période 04/89-04/91

respectivement indiquées. La valeur de  $b$  obtenue ici (2,82) se situe sensiblement au delà de 2. Ainsi la transformation idoine serait :

$$Z = X^{-0,59} \text{ proche de } Z = \frac{1}{\sqrt{X}}$$

Néanmoins par commodité, la transformation logarithmique sera effectuée sur les données bactériologiques brutes.

## 2.2. Structure spatiale de la contamination fécale et cartographie

### 2.2.1. Introduction

Le paragraphe précédent (2.1) montre que l'échantillonnage pratiqué à l'échelle d'un secteur peut conduire à un niveau de contamination moyen non représentatif de la vraie distribution spatio-temporelle du descripteur. L'effort d'échantillonnage alloué devrait être fonction de la complexité du site, qui dépend en particulier du nombre et de la variété des émissaires, mais également des phénomènes courantologiques locaux.

Des travaux antérieurs (Legendre *et al.*, 1984; Legendre & Fortin, 1989; Lebaron *et al.*, 1990) ont montré l'intérêt de la connaissance de la structuration spatio-temporelle des abondances en coliformes thermo-tolérants pour la définition de stratégies d'échantillonnage optimales, en milieu lagunaire méditerranéen.

Les concepts et outils de la géostatistique semblent particulièrement adaptés à la description et à la caractérisation de la structure spatiale dans le cas d'un matériel bactérien : le caractère à la fois aléatoire et globalement structuré de la distribution de la variable bactériologique est prise en compte au travers des fonctions aléatoires (*Random function*) définies précédemment (§ 1.2.4.5).

Le but de la présente étude est, en utilisant les techniques de la géostatistique, de décrire puis de caractériser la structure spatiale de la contamination fécale des coquillages sur une zone de production conchylicole (Havre de la Vanlée, Manche, France), en mer macrotidale, à l'échelle d'un secteur du réseau de surveillance microbiologique (REMI).

### 2.2.2. Matériel et méthodes

#### Site et période d'échantillonnage

Le site d'échantillonnage correspond à la zone de bouchots de moules s'étendant face au

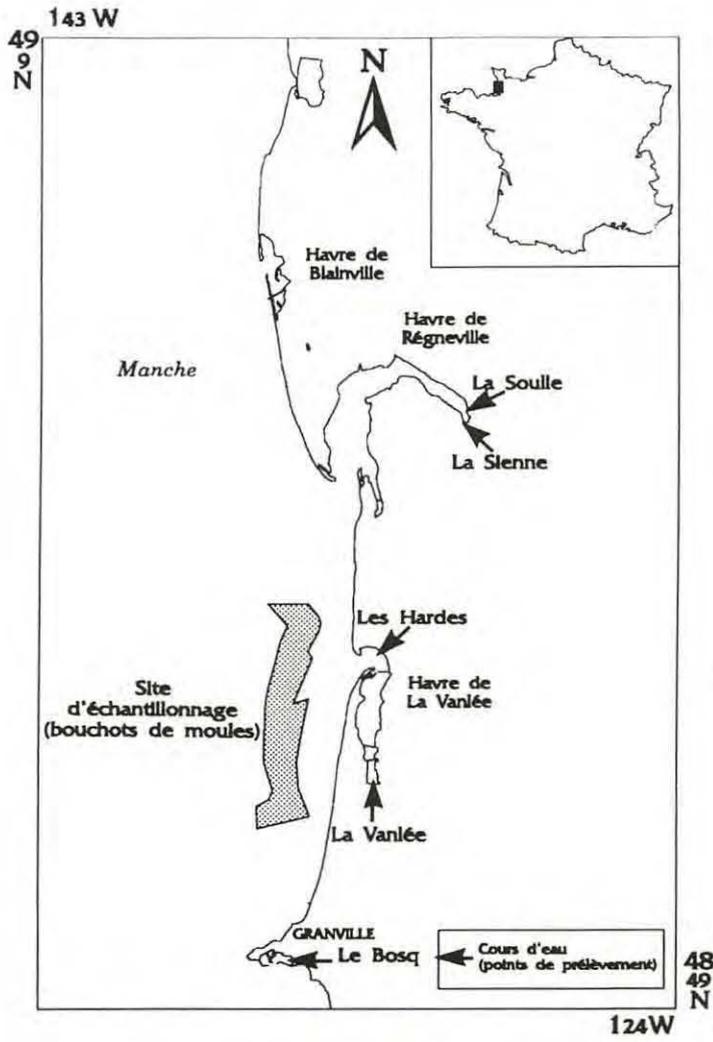
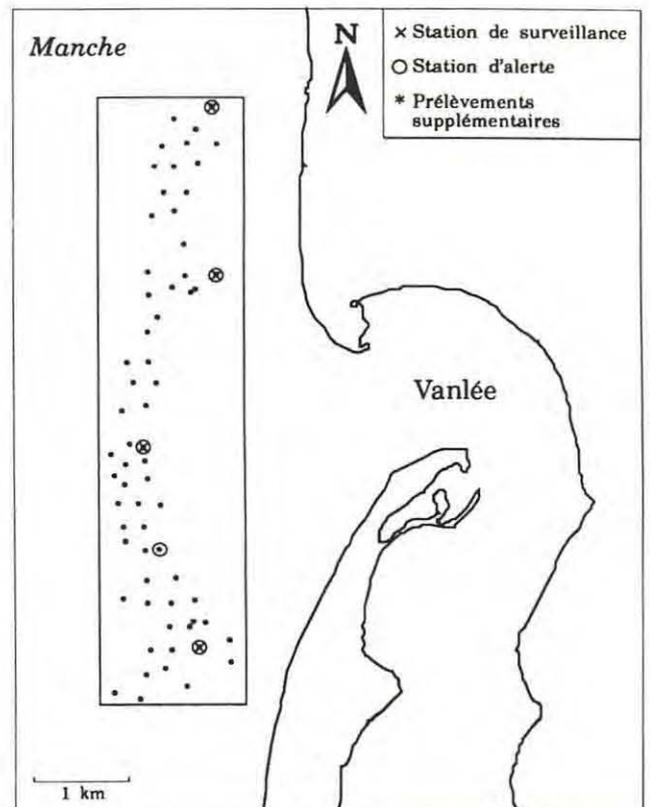


Fig. III.15 : Description de la zone d'étude

Fig. III.16 : Grille d'échantillonnage



Havre de la Vanlée (Fig. III.15), entre le Havre de Règneville au nord et de Bréville/Mer au sud. A la suite des grands froids de 1962, les pêcheurs des communes côtières se reconvertirent dans la mytiliculture et les premiers bouchots furent implantés en 1964. On y dénombre actuellement 108 parcelles concédées représentant 86,3 km de bouchots, soit 6% des longueurs attribués en France.

Les sources potentielles de pollution sur le site sont les élevages, le maraîchage, et pour la pollution d'origine anthropique essentiellement le tourisme (augmentation jusqu'à 600% de la population du bassin versant en été).

La zone d'étude correspond grossièrement à un rectangle d'une longueur de 6,5 km et d'une largeur moyenne de 0,9 km (max. 1,5 km; min. 0,4 km), soit une surface d'environ 6 km<sup>2</sup>. L'échantillonnage du site n'a été réalisé que lors de périodes où les parcs sont découverts (étale de basse mer, coefficients de marée voisins de 100). Cette situation autorise un accès aisé aux stations d'échantillonnage et permet de diminuer le temps nécessaire à la réalisation de l'ensemble des prélèvements. Enfin l'étendue de la couverture spatiale de la zone est tributaire des conditions météorologiques du jour; en particulier, une dépression accompagnée de vent d'ouest interdit l'accès aux parcs situés le plus au large, même avec de forts coefficients. Deux campagnes ont été menées : le 15 avril 1991 et le 7 octobre 1991. Les conditions météorologiques sur le site étaient les suivantes (données fournies par le sémaphore du Roc à Granville) :

- 15 avril : temps sec et ensoleillé, vent de nord-nord-est ( $< 10 \text{ m.s}^{-1}$ ), mer agitée à peu agitée. Coefficient de marée : 102, basse mer : 15 h 25 mn.

- 7 octobre : temps sec et ensoleillé succédant à une quinzaine comportant de nombreux épisodes pluvieux, notamment les 5 et 6 octobre, vent de sud tournant sud-sud-ouest en début d'après-midi ( $< 12 \text{ m.s}^{-1}$ ), mer peu agitée à agitée. Coefficient de marée : 100, basse mer : 14 h.

### Protocole d'échantillonnage

La répartition des stations correspond à une couverture systématique de la zone de bouchots. La maille retenue d'environ 250 m permet d'obtenir finalement 63 stations réparties autant que possible en quinconce (Fig. III.16). Un repérage préalable sur le terrain avait révélé que certaines stations étaient localisées dans un secteur dépourvu de moules. Un repositionnement de ces stations a été effectué, ainsi que des stations avoisinantes pour assurer le maintien de la régularité de la couverture. Les stations du réseau de surveillance (REMI) sur le site sont également représentées (Fig. III.16).

Deux stations (Fig. III.16) ont par ailleurs été choisies lors de la campagne du 15 avril, dans deux secteurs *a priori*, et respectivement, peu et plus contaminés. A ces stations, 5 prélèvements supplémentaires sont effectués dans le but d'obtenir une mesure de la variabilité locale. Seule la station située au sud a fait l'objet d'un semblable extra-échantillonnage le 7 octobre.

D'autre part, des prélèvements d'eau sont réalisés dans les principaux émissaires (Fig. III.15) susceptibles d'être impliqués dans la contamination de la zone.

### Protocole analytique

La concentration en coliformes fécaux (CF) dans les coquillages est obtenue indirectement à partir de la mesure conductance-métrique (§ II.3). Pour les moules, l'équation de la droite de régression<sup>1</sup> s'écrit :

$$\log(\text{CF}) = 10,875 - \text{TD}$$

Le descripteur retenu est alors le logarithme décimal de la concentration en coliformes fécaux pour 100 ml de broyat, directement donné par la droite de régression. L'extrapolation de

<sup>1</sup> Equation très légèrement différente de l'équation finale obtenue par ailleurs (§ II.3.3.1), d'autres couples (estimations NPP, temps de détection) étant venus s'ajouter par la suite.

la droite fournit le temps de détection tout à fait théorique pour lequel le logarithme est nul, c'est-à-dire correspondant à un unique coliforme. Cette limite est estimée à 10,875 h. Le pas horaire de la mesure conductimétrique étant égal à 0,1 h, tout temps supérieur à 10,8 h conduit à une valeur nulle du descripteur.

Ne disposant pas d'une régression pour les eaux douces, les dénombrements dans les émissaires sont des estimations du "Nombre le Plus Probable" de coliformes fécaux pour 100 ml d'échantillon, après ensemencement en milieu liquide sélectif (§ I.2.2.2).

### Traitement des données

L'estimateur du krigeage, outil fondamental de la géostatistique, présente, par comparaison aux méthodes usuelles d'interpolation, l'avantage majeur de fournir des estimations non biaisées et de variances connues et minimales (Burgess & Webster, 1980). Développée initialement pour la géologie minière par Krige, géologue sud-africain, puis par G. Matheron, cette méthode a fourni quelques applications récentes dans le domaine océanographique (Conan & Wade, 1989; Armstrong *et al.*, 1989; Gros & Hamon, 1989; voir également la liste de références fournies par Ibanez, 1990), et même, plus précisément, en microbiologie marine (Legendre *et al.*, 1989; Lebaron *et al.*, 1990).

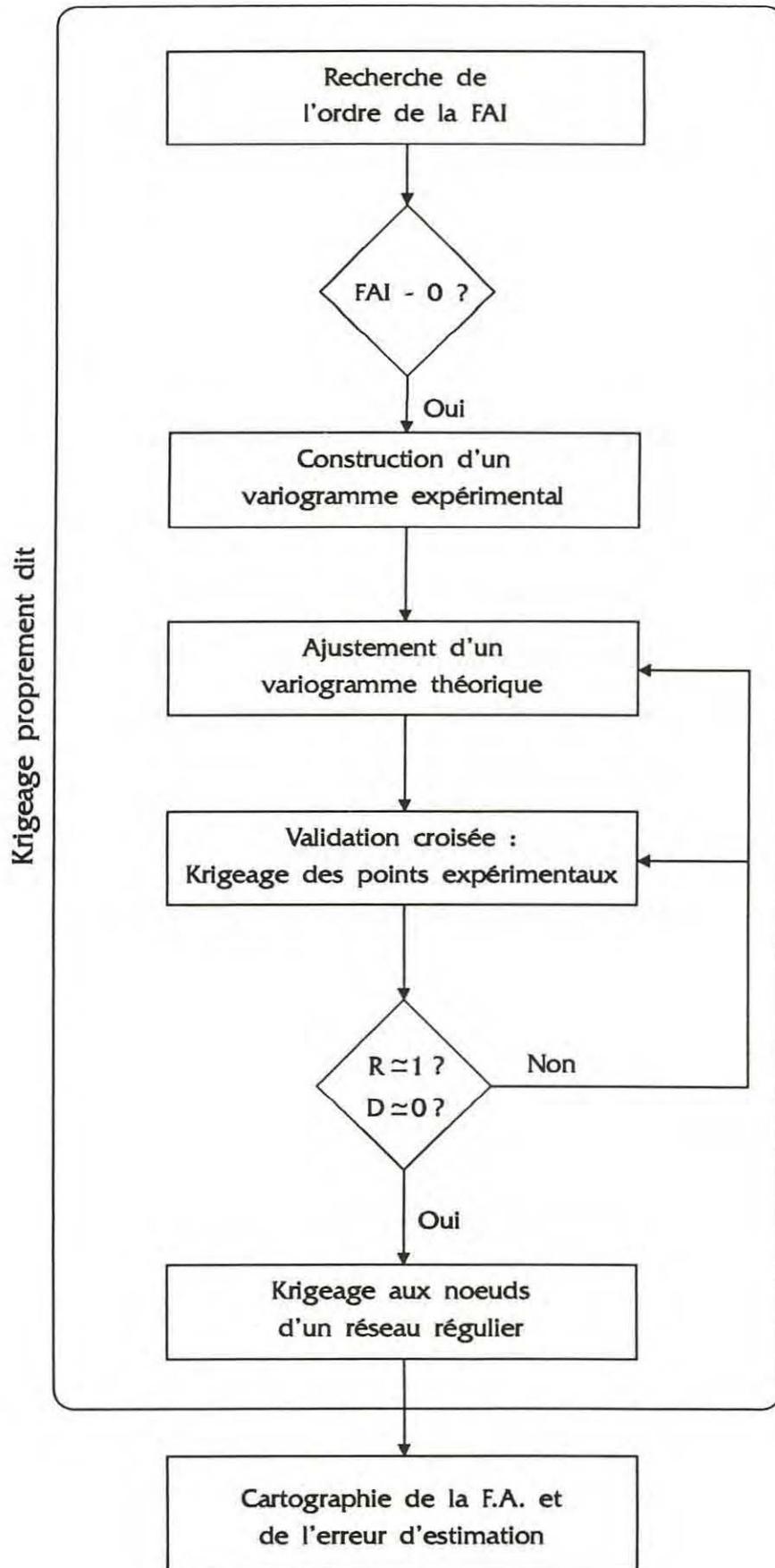
Des généralités sur le krigeage et les résultats théoriques utiles à notre propos ont été décrits au paragraphe I.2.4.5.

Le logiciel MAGIK<sup>1</sup> (Le Chauve & Gohin, Centre ORSTOM de Brest) est utilisé pour les différentes étapes du krigeage ponctuel (Fig. III.17), ainsi que la cartographie finale des valeurs krigées et de leurs erreurs associées, sous forme d'isoplèthes.

Dans le but de la recherche d'un échantillonnage optimal, dans le contexte de la surveillance bactériologique, un programme informatique permet, à partir du positionnement des 63 stations

---

<sup>1</sup> Méthode d'Application de la Géostatistique et de l'Interpolateur du Krigeage



**Fig. III.17** : Représentation schématique de la démarche de traitement des données conduisant à la cartographie de la structure spatiale de la contamination

et des valeurs observées correspondantes pour les deux campagnes, de simuler *a posteriori* l'échantillonnage, approximativement systématique linéaire, tel que pratiqué en surveillance microbiologique. L'effort d'échantillonnage correspond au nombre de stations allouées pour la surveillance bactériologique du secteur considéré. Pour un effort  $n$  donné et à chaque simulation,  $n$  stations sont tirées au sort parmi les 63, de telle façon que la distance séparant chaque station soit comprise dans un certain intervalle permettant de respecter approximativement l'équidistance propre à l'échantillonnage systématique. On dispose alors d'un échantillon de  $n$  valeurs du descripteur, à partir duquel est estimée la variance  $\sigma^2$  de la population; soit  $\hat{\sigma}^2$  cette estimation.

Pour chaque effort, celui-ci variant de  $n = 3$  à  $n = 6$  stations, 250 simulations sont effectuées. Pour une distribution normale du descripteur et dans le cas de grands échantillons ( $n$  grand),  $\hat{\sigma}^2$  a une loi de probabilité normale de moyenne  $\sigma^2$  et de variance  $\frac{2\sigma^4}{n}$  (Schwartz, 1963). Nous estimerons alors la variance moyenne comme la moyenne des variances et nous considèrerons l'"erreur standard"  $\sqrt{\frac{2\hat{\sigma}^4/n}{250}}$  comme un indice de la dispersion des valeurs de variance pour un effort donné, les conditions requises d'utilisation des estimateurs n'étant pas remplies.

### 2.2.3. Résultats

#### Statistiques élémentaires

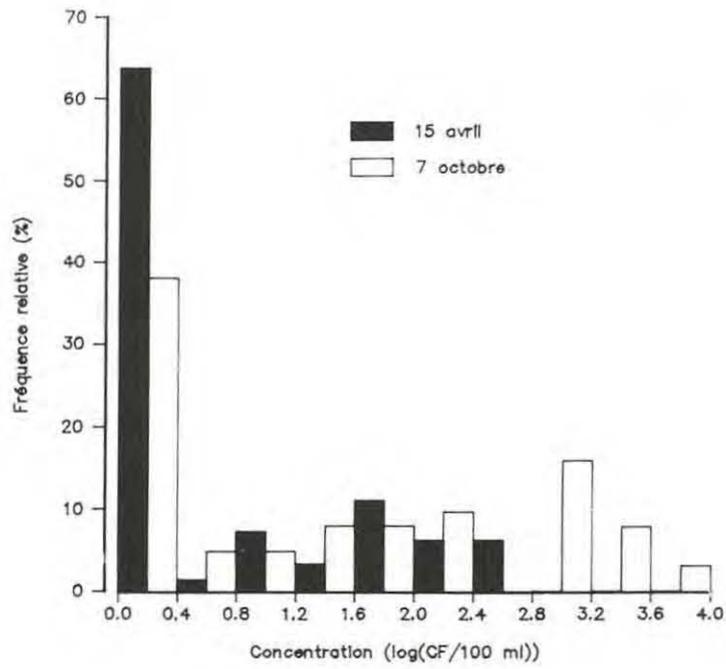
Le tableau III.11 présente les statistiques élémentaires calculées sur les données brutes.

Tab. III.11 : Statistiques élémentaires

	15/04	07/10
Moyenne *	0,66	1,49
Médiane *	0	1,31
Minimum *	0	0
Maximum *	2,8	3,7
Ecart-type *	0,93	1,26
Variance **	0,86	1,58
Moyenne locale *(Station sud)	1,13	0,55
Variance locale **	0,43	0,29

\* : en  $\log(\text{CF}/100 \text{ ml})$ \*\* : en  $\log^2(\text{CF}/100 \text{ ml})$ 

Pour la campagne du 15 avril, la première classe de concentration, de 0 à 0,4  $\log(\text{CF}/100 \text{ ml})$ , représente près de 64% des observations contre 38% pour la campagne du 7 octobre (Fig. III.18); les pourcentages de valeurs nulles sont respectivement de 57% et 38%. Ces pourcentages mais surtout les différences de concentration moyenne (Tab. III.11) entre les deux campagnes sont à rapprocher des différences de concentration (Tab. III.12) dans les apports continentaux. On constate que pour la deuxième campagne le niveau de contamination des eaux est plus élevé, pour des débits plus faibles et donc des flux comparables.



**Fig. III.18** : Histogrammes des fréquences relatives des concentrations en coliformes fécaux en log(CF/100 ml)

**Tab III.12** : Concentrations en coliformes fécaux (CF/100 ml de broyat) , intervalles de confiance [IC (95%)] des concentrations, débits ( $Q$ ) en  $m^3.s^{-1}$  et flux correspondant ( $\Phi$ ) en  $CF.s^{-1}$  pour les cours d'eau de la zone étudiée.

	15/04/1991			07/10/1991		
	CF/100 ml IC (95%)	$Q$ $m^3.s^{-1}$	$\Phi$ $CF.s^{-1}$	CF/100 ml IC (95%)	$Q$ $m^3.s^{-1}$	$\Phi$ $CF.s^{-1}$
La Soulle	720 300 - 3000	0,50	$3,5 \cdot 10^6$	3700 1500 - 14000	0,115	$4,3 \cdot 10^6$
La Sienne	430 190 - 1200	2,50	$1,1 \cdot 10^7$	5600 2200 - 17500	0,915	$5,1 \cdot 10^7$
Les Hardes	230 90 - 760	$Q_h^*$	$2,3 \cdot 10^5$	330 140 - 1200	0,006	$2,0 \cdot 10^4$
La Vanlée	720 300 - 3000	$Q_v^*$	$7,2 \cdot 10^5$	6900 2700 - 23300	0,026	$1,8 \cdot 10^6$
Le Bosq	$1,6 \cdot 10^5$ $6 \cdot 10^4 - 5,5 \cdot 10^5$	0,20	$3,2 \cdot 10^8$	$1,4 \cdot 10^6$ $6 \cdot 10^5 - 6,6 \cdot 10^6$	0,079	$1,1 \cdot 10^9$

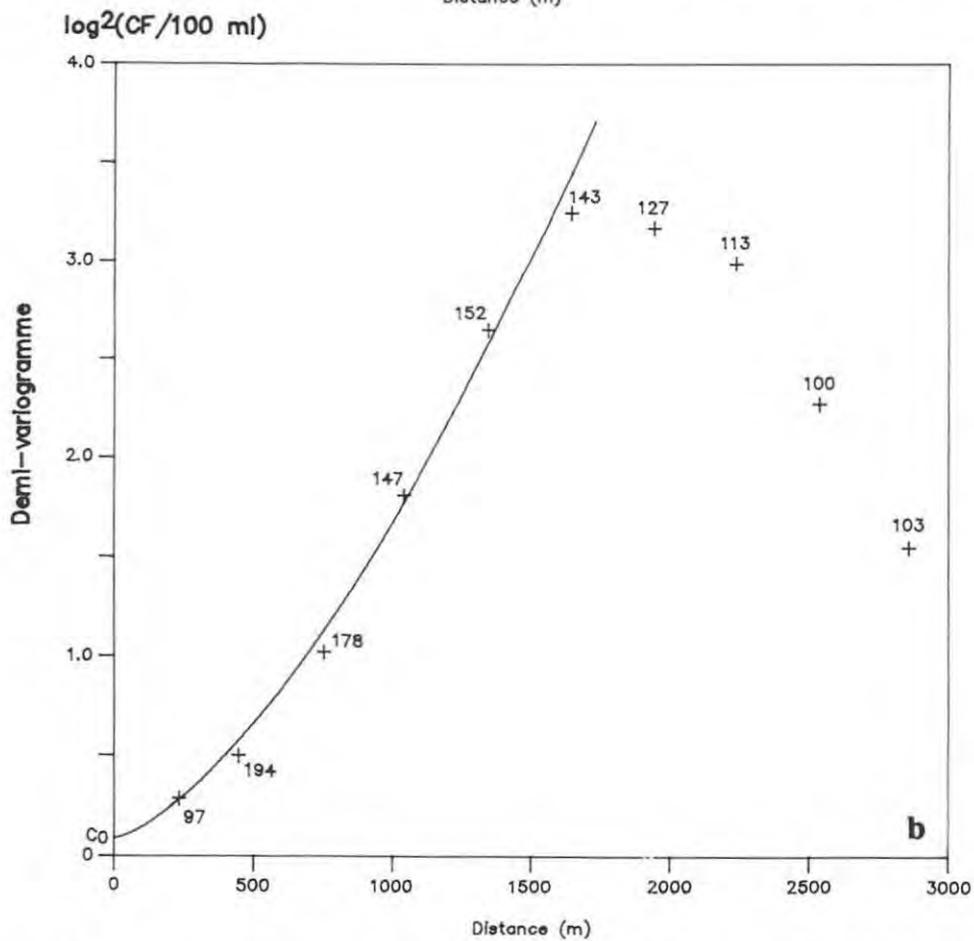
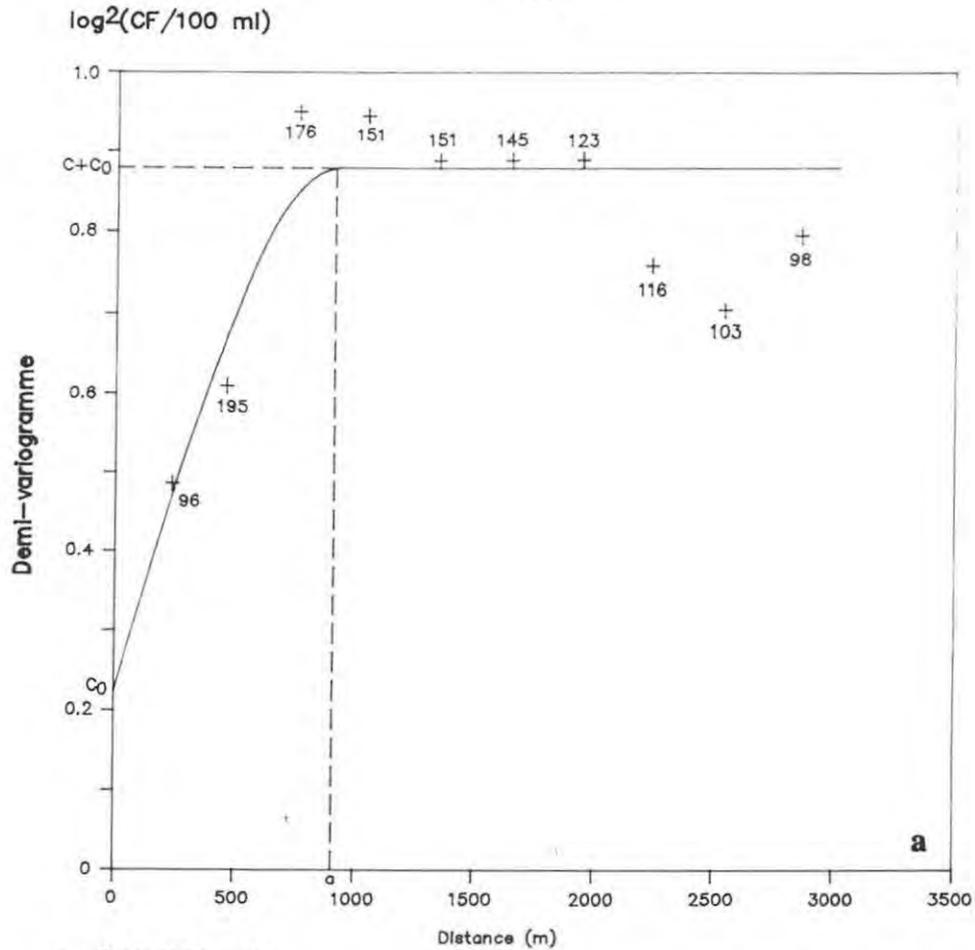
\* :  $Q_h + Q_v = 0,20 m^3.s^{-1}$

### Résultats du krigeage

L'ordre de la fonction aléatoire intrinsèque n'a pu être déterminé. Le faible nombre de stations et la "morphologie" quasi-unidimensionnelle de la zone étudiée empêchent la définition des anneaux ou voisinages nécessaires à la recherche de l'ordre dont l'algorithme a, par ailleurs, été décrit par Gros et Hamon (1989).

Comme indiqué par Journel et Huijbregts (1978), l'abscisse maximale du variogramme doit représenter environ la moitié de la plus grande dimension de la zone. L'ajustement utilise le critère des moindres carrés avec une pondération en  $1/h$ . Ceci confère aux faibles distances, et par conséquent à l'estimation des caractéristiques de la structure, un poids plus important.

Un modèle sphérique de variogramme isotrope a été ajusté au variogramme expérimental obtenu pour la première campagne (Fig. III.19a). Un modèle en  $h^\alpha$  ( $\alpha = 1.51 m^{-1}$ ) avec adjonction de pépite, a été retenu pour la deuxième campagne (Fig. III.19b).



**Fig. III.19** : Demi-variogrammes (a : 15 avril - modèle sphérique avec  $c_0 = 0,22 \log^2(\text{CF}/100 \text{ ml})$ ,  $c = 0,66 \log^2(\text{CF}/100 \text{ ml})$  et  $a = 920 \text{ m}$ ); b : 7 octobre - modèle puissance avec  $c_0 = 0,087 \log^2(\text{CF}/100 \text{ ml})$ ,  $w = 0,0015 \log^2(\text{CF}/100 \text{ ml})$  et  $a = 1,51 \text{ m}^{-1}$ )

L'abscisse d'un point du variogramme expérimental est la moyenne des distances incluses dans la classe de distance correspondante. La distance maximale est choisie égale à la moitié de la plus grande dimension de la zone étudiée, comme suggéré par Journel & Huijbregts (1978). L'ajustement du second variogramme a été réalisé à partir des seuls six premiers points du variogramme expérimental, du fait d'une décroissance importante du variogramme au-delà de la sixième classe de distance (Fig. III.19b). Ceci peut être attribué à l'alternance de zones à faibles et fortes concentrations à l'échelle spatiale du secteur, conduisant à un effet de trou (Legendre *et al.*, 1989).

Clark (1982) recommande comme largeur de classe de distance 10% de la distance moyenne entre points, calculée pour toutes les paires possibles de points, à condition que les classes ainsi formées soient environ équiréquentes. Ceci conduit ici à un pas de distance de 230 m, mais le nombre de paires de points a été jugé alors insuffisant pour la première classe [0, 230]. Une largeur de 300 m a été retenue pour les deux variogrammes.

Les nombres jouxtant les points des variogrammes expérimentaux (Fig. III.19a et b) représentent le nombre d'incrémentes utilisés pour le calcul des variogrammes, *i.e.* les valeurs  $n(h)$  (§ I.2.4.5).

Un critère d'ajustement des moindres carrés avec un facteur de pondération en  $h^{-1}$  a été utilisé pour les deux campagnes.

Le krigeage des observations permet la validation croisée du modèle utilisé : après avoir oté une observation  $z(x_i)$  de l'échantillon, sa valeur  $\hat{z}(x_i)$  est estimée par krigeage à partir des points avoisinants. Nous disposons ainsi pour chaque observation de l'écart  $z(x_i) - \hat{z}(x_i)$ . Le modèle sera considéré comme satisfaisant si la somme des écarts,  $D$ , n'est pas trop éloignée de 0 (erreur sur la portée) et si la variance des écarts n'est pas trop éloignée de la variance de krigeage ponctuel  $\sigma_k^2$  (Gros & Hamon, 1989), c'est-à-dire si  $R$  n'est pas trop éloigné de 1 (mauvaise estimation du palier), avec :

$$D = \sum_i (z(x_i) - \hat{z}(x_i)) \approx 0$$

$$R = \frac{\sum_i (z(x_i) - \hat{z}(x_i))^2}{\sigma_k^2} \approx 1$$

Dans MAGIK,  $R$  se calcule comme :

$$R = \frac{1}{n} \sum_{i=1}^n \left( \frac{z(x_i) - \hat{z}(x_i)}{\sigma_{ik}} \right)^2$$

où  $\sigma_{ik}$  est l'écart-type de krigeage ponctuel pour la station  $i$ . Gohin (1990) propose un critère objectif pour la détermination de valeurs observées aberrantes : le rapport, calculé par MAGIK, de la valeur absolue de l'écart entre la valeur observée et la valeur interpolée, à l'écart-type ne doit pas dépasser 2,5. Un tel rapport n'a pas été observé pour la première campagne, la moyenne se situant à 0,74 log(CF/ 100 ml). Par contre, un rapport de 3,24 a été calculé pour la deuxième campagne à une station. Il s'agit d'une forte valeur de concentration se situant dans une plage de résultats nuls; Après remplacement par une valeur nulle et réajustement d'un variogramme, la somme des écarts quadratiques au numérateur de  $R$  diminue. Une valeur convenable de  $R$  est alors obtenu en réduisant l'erreur de mesure, paramètre du logiciel MAGIK, ce qui est cohérent avec le retrait d'une valeur semblant une erreur. L'écart-type des résidus varie alors de 0,64 à 0,54 log(CF/100 ml).

Les deux critères,  $R$  et  $D$ , sont reportés dans le tableau III.13 pour les deux campagnes. Les modèles préconisés apparaissent très satisfaisants en regard des valeurs obtenus.

**Tab. III.13** : Paramètres du krigeage des points expérimentaux (validation croisée).

	15/04	07/10
$R$	1,01	1,00
$D$ *	0,01	-0,02
Ecart-type des résidus *	0,72	0,56
Erreur théorique moyenne *	0,74	0,54

\* en log(CF/100 ml)

Le krigeage aux noeuds d'un réseau régulier permet la cartographie de la pollution fécale dans les limites du polygone de krigeage, circonscrivant la zone de bouchots dans un rectangle de 6,5 km de longueur sur 1,5 km dans la dimension perpendiculaire à la côte (Fig. III.20a et b). Le maillage de couverture de la zone envisagé *a priori* (250 m - 300 m) semble cohérent avec la structure observée, après l'essai d'autres pas de distance lors de la détermination du variogramme expérimental, ce qui nous a conduit à envisager un maillage carré de 300 m de côté, dont nous avons krigé les noeuds.

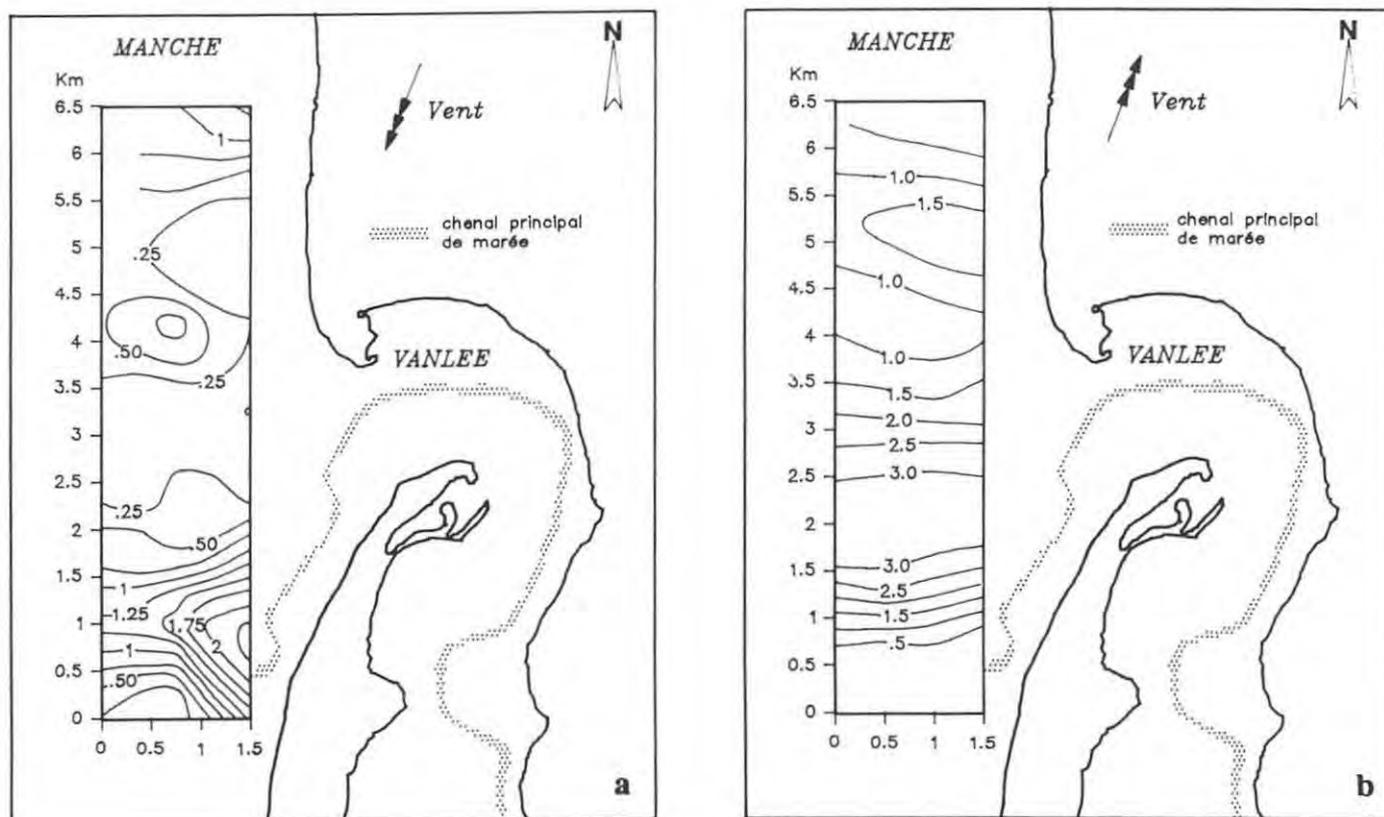
L'erreur de krigeage se situe respectivement pour les deux campagnes entre 0,57 et 1,16  $\log(\text{CF}/100 \text{ ml})$ , et entre 0,41 et 1,26  $\log(\text{CF}/100 \text{ ml})$ , avec les erreurs d'estimation les plus importantes situées aux noeuds périphériques du polygone de krigeage (Fig. III.21a et b).

Concernant les résultats des simulations de l'échantillonnage pour la deuxième campagne, la variance croît pour 4 stations, puis diminue pour se stabiliser à partir de 5 stations (Fig. III.22). L'évolution est beaucoup plus chaotique pour la campagne du 15 avril et n'est pas représentée ici. Aucune tendance n'est observable; ceci pourrait être lié au fort pourcentage de valeurs nulles observées.

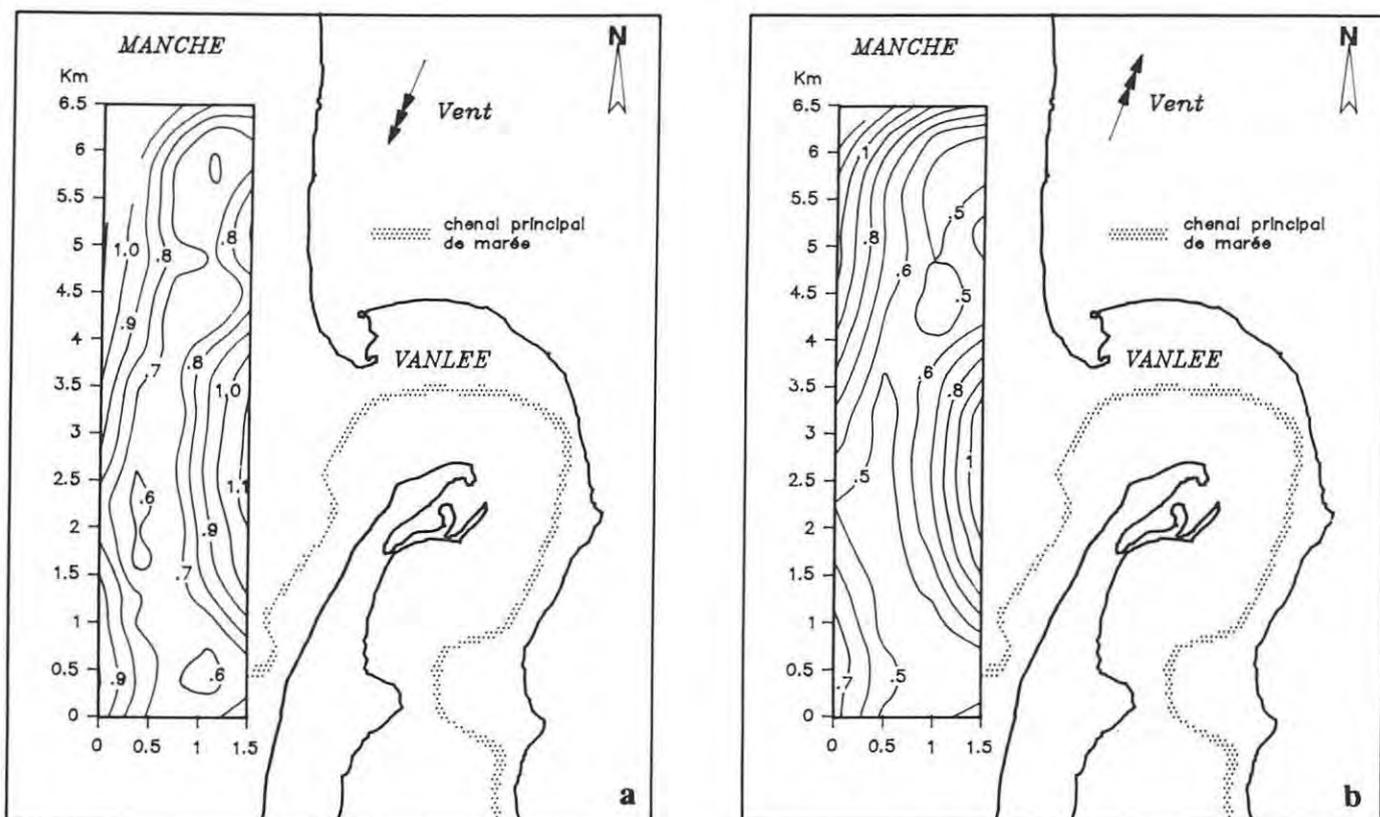
#### 2.2.4. Discussion

##### Ajustement du variogramme

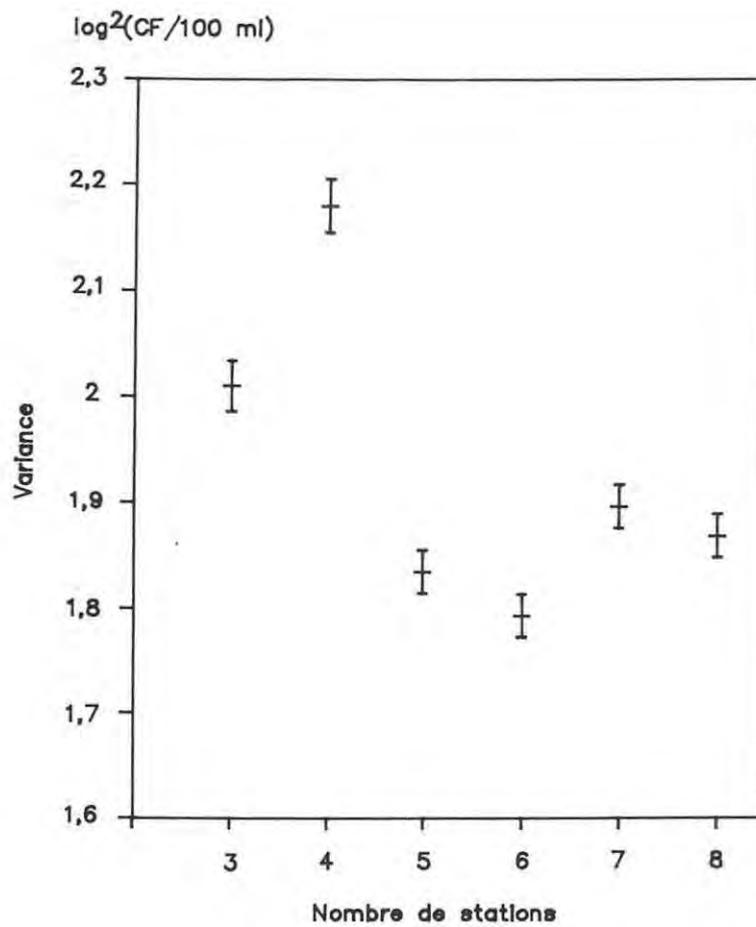
La qualité des estimations krigées est fonction en premier lieu du choix judicieux du variogramme théorique. Concernant l'allure du variogramme pour la première campagne (Fig. III.19a), la décroissance observée en début de palier correspond environ aux premières abscisses supérieures à la largeur moyenne de la zone. Ainsi, le poids des couples de stations, ou bipoints, à direction plutôt parallèle à la côte sur l'ensemble des couples est de plus en plus important jusqu'à être rapidement essentiel. Ceci induit un changement dans la nature du variogramme. Par ailleurs, la forme de la zone, alliée au faible nombre de stations, nous interdit la recherche d'une anisotropie géométrique, pouvant être préjudiciable à la cartographie finale.



**Fig. III.20** : Cartographie de la pollution fécale exprimée en  $\log(\text{CF}/100 \text{ ml})$  - a : 15 avril; b : 7 octobre)



**Fig. III.21** : Cartographie des erreurs d'estimation exprimées en  $\log(\text{CF}/100 \text{ ml})$  - a : 15 avril;  
b : 7 octobre)



**Fig. III.22** : Variance moyenne et erreur standard afférente calculées à partir de 250 simulations d'un plan systématique linéaire sur le site de la Vanlée, en fonction du nombre de stations tirées au sort par simulation (taille de l'échantillon)

L'ajustement d'un modèle aux valeurs du variogramme expérimental est une étape importante et délicate. En effet, il convient de critiquer les paramètres du modèle obtenus et de s'interroger sur leur signification.

Pour la première campagne, la pépite ( $0,22 \log^2(\text{CF}/100 \text{ ml})$ ) est très proche de la demi-variance locale ( $0,215 \log^2(\text{CF}/100 \text{ ml})$ ); la comparaison reste satisfaisante pour la deuxième campagne (respectivement  $0,087$  et  $0,145 \log^2(\text{CF}/100 \text{ ml})$ ) Compte-tenu de l'hypothèse d'intégration de la variabilité hydrique réalisée par les mollusques filtreurs (*e.g.* Turick *et al.*, 1988), l'effet de pépite serait plus imputable à l'erreur de mesure, notoirement importante en bactériologie, qu'à des phénomènes de micro-régionalisation. Ceux-ci évoquent des micro-distributions d'abondance en coliformes fécaux, à des échelles spatiales inférieures au pas d'échantillonnage (Le Baron *et al.*, 1990). Nous avons montré qu'à l'échelle d'une station de prélèvement du réseau de surveillance, la variabilité analytique, ou erreur de mesure, était prépondérante vis -à-vis de la variabilité spatiale au voisinage immédiat d'une station, ce qui accèderait l'hypothèse d'intégration spatiale réalisée par les mollusques.

La détermination de la portée est dépendante, à la fois, de la justesse de l'ordonnée à l'origine (pépite) extrapolée à partir du variogramme théorique et de la précision de l'estimation du palier. Pour la première campagne, celui-ci ( $0,92 \log^2(\text{CF}/100 \text{ ml})$ ) est légèrement supérieur à la variance expérimentale ( $0,86 \log^2(\text{CF}/100 \text{ ml})$ ). Ceci satisferait à l'hypothèse de stationnarité requise. Pour la deuxième campagne, le modèle théorique est non borné - aucun palier n'est atteint avant la demie-longueur de la zone - et ne fait pas intervenir ces deux derniers paramètres. La puissance déterminée ( $\alpha = 1,51 \text{ m}^{-1}$ ) peut être interprétée soit comme résultant d'un effet de dérive parabolique rendant caduque l'hypothèse de stationnarité, soit comme la puissance d'un modèle stationnaire (Journel & Huijbregts, 1978), ce que nous avons admis, compte tenu de la robustesse de la représentation cartographique vis-à-vis de l'hypothèse de stationnarité. En conclusion, les paramètres obtenus semblent réalistes.

### Evolution temporelle de la structure

Une forte différence de structure spatiale est observée entre les campagnes du 15 avril et du 7 octobre. Ceci est caractérisé par des variogrammes de nature différente; en particulier, pour le 7 octobre, le variogramme puissance ajusté aux valeurs du variogramme expérimental dénote une très forte structuration spatiale de la contamination bactériologique mesurée dans les coquillages.

Comment interpréter une telle différence, pour des flux de pollution comparables et des coefficients de marée identiques ? Divers auteurs ont montré que les bactéries entériques peuvent survivre plusieurs semaines dans les sédiments (voir § I.1.4.3). Ainsi, on peut supposer une remise en suspension de sédiments plus chargés en germes fécaux en fin de période estivale qu'en fin d'hiver. Ces sédiments seraient alors entraînés par le courant de jusant et filtrés puis accumulés dans les moules.

### Variances de krigeage et choix du polygone de krigeage

Un des attraits majeurs du krigeage réside dans l'obtention des variances de krigeage, c'est-à-dire ici des variances d'estimation locale, ou variance d'extension (Journel & Huijbregts, 1978). La cartographie des écart-types correspondants (Fig. III.21a et b) montre l'homogénéité des erreurs sur le domaine circonscrit à la zone de bouchots, ce qui paraît naturel compte-tenu de l'uniformité de la densité d'échantillonnage sur la zone, mis à part à sa périphérie, où les points informants sont naturellement moins uniformément répartis et où, par conséquent, les erreurs sont les plus importantes. Il s'agit de l'"effet de bord", décrit par Conan et Wade (1989), utilisé pour définir les limites de la zone où les estimations peuvent être jugées fiables. Ici se pose les problèmes du nombre de zéros ou non-réponses ainsi que celui, lié, du choix d'un polygone de krigeage. Dans leur étude sur l'estimation d'une population de bivalves sur la côte atlantique, Armstrong *et al.* (1989), obtenant 27 valeurs nulles sur 67 observations, s'interrogent sur la nécessité ou non d'inclure les zéros dans le calcul des estimations krigées.

Ils préconisent deux solutions. Le choix est fonction de la bonne connaissance du phénomène biologique ou physique étudié. La première possibilité consiste à kriger à l'intérieur d'un polygone circonscrit à la ou aux zones ne contenant que les valeurs non nulles. Ceci est envisageable si l'on considère que des facteurs biologiques ont constitué une barrière infranchissable pour la population étudiée.

La répartition spatiale des coliformes fécaux ne rentre pas dans ce cas : l'abondance décroît progressivement jusqu'à devenir nulle, conformément aux lois de la dilution, nuancées par l'hydrodynamisme local. La nullité observée constitue dans notre étude une information de même poids que des valeurs non nulles. La deuxième solution est alors préférable; elle consiste à kriger dans un polygone englobant l'ensemble des valeurs, une discontinuité brutale valeurs non-nulles/valeurs nulles ne pouvant alors exister. Pour des raisons de commodité, nous avons élargi notre polygone à un rectangle (Fig. III.20a et b), sachant que de nombreuses mailles n'y seront pas informées, en supposant qu'il n'existe aucune contrainte pour que des bouchots n'aient pu être implantés dans l'aire rectangulaire choisie.

### Positionnement des stations

Un positionnement optimal *a posteriori* des stations consisterait, afin de diminuer les variances d'estimation du krigeage ponctuel, à accroître l'effort d'échantillonnage sur la façade est de la zone de bouchots et plus précisément dans le secteur de débouché de la Vanlée.

Cependant, la cartographie de la pollution fécale mesurée dans les coquillages semble dépendante des conditions météorologiques, et en particulier de la direction du vent, comme il a pu être noté par ailleurs (Quetin, 1991). Ainsi, on observe que par vent de sud-sud-ouest (campagne du 7 octobre) l'influence de la Vanlée se situe sensiblement plus au nord sur la zone de bouchots que pour la première campagne et l'aire polluée n'est pas circonscrite au voisinage du chenal principal de marée (Fig. III.20b). Cette mouvance du panache au gré des caprices d'Eole remettrait en cause un effort accru dont le positionnement serait fonction des résultats

d'une campagne. Notons par ailleurs qu'il n'est pas à rejeter une très forte contamination sur le site provenant soit de l'émissaire du Bosq, particulièrement insalubre (Tab. III.12), dans le cas de forts vents de secteur sud établis, soit du Havre de Règneville pour des vents de secteur nord et des flux plus importants. Les fortes valeurs de concentration obtenues dans les émissaires montrent que malgré l'effort d'assainissement réalisé, des problèmes subsistent dans ce secteur : dysfonctionnement des ouvrages d'épuration lié à la nature humide des terrains, raccordements partiels aux réseaux, et rejets illicites d'eaux usées.

### 2.3. Conclusion

Les dimensions des secteurs de surveillance bactériologique sont telles qu'on pouvait suspecter une forte hétérogénéité spatiale, préjudiciable à la signification des séries chronologiques obtenues *in fine*. L'examen de certains secteurs du réseau de surveillance a permis de confirmer ceci.

Ainsi, il est nécessaire de définir des zones homogènes où l'on puisse considérer que l'on dispose d'une distribution statistique dans l'espace. Conjointement, on peut rejeter des stations jugées non informatives. La loi de Taylor permet ensuite d'évaluer le nombre de stations requises dans ces zones pour un degré de précision fixé.

L'étude fine de la structure spatiale de la contamination fécale à l'échelle d'une zone de production mytilicole a permis de visualiser l'hétérogénéité spatiale, par l'observation de macrostructures. La remise en suspension des sédiments, pièges à bactéries, serait une explication à l'évolution remarquable des caractéristiques de la structure spatiale d'une campagne à l'autre. Ceci, allié à l'action des facteurs météorologiques sur le panache de pollution, induit un changement important de la configuration de l'hétérogénéité à l'échelle du secteur, rendant complexe la démarche de détermination de zones homogènes.

### **3. Salubrité des zones conchylicoles. Un exemple : la plage Benoît (La Baule, Loire-Atlantique)**

#### **3.1. Problématique**

L'arrêté ministériel du 12 octobre 1976 (annexe I.3), en application du décret du 20 août 1939, précise les critères autorisant le classement de salubrité bactériologique des zones conchylicoles. Le texte fixe la méthode analytique, le nombre de prélèvements annuels ainsi que les concentrations critiques en coliformes fécaux mesurés dans les coquillages.

L'étude de salubrité du gisement naturel de coques de la plage Benoît (Le Pouliguen, Loire-Atlantique) a été menée par le laboratoire côtier de Nantes en suivant rigoureusement les directives de l'arrêté. La zone est classée insalubre depuis le 13 juin 1988, suite à un suivi bactériologique effectué en 1985-1986 (Catherine *et al.*, 1991). Le maintien du classement en zone insalubre passe par l'analyse des données recueillies au cours d'un nouveau suivi. A la lumière des résultats obtenus, un jugement critique est porté sur les termes de l'arrêté; en particulier, le fondement probabiliste du texte est examiné.

Le "risque consommateur" est l'erreur que le décideur commet en décrétant la zone salubre alors qu'elle ne l'est pas. La quantification de cette erreur est primordiale pour la protection de la santé humaine; une analyse du risque reposant sur les données de la plage Benoit est proposée.

#### **3.2. Classement de salubrité de la plage Benoît**

##### *3.2.1. Description de la zone et protocole*

##### Bassin versant et sources de pollution

Le bassin versant concerné par l'étude de la qualité bactériologique de la plage Benoît est très fortement urbanisé. L'affluence touristique en juillet et en août est très importante, le nombre d'habitants passant de 15000 hors-saison à 110000 en moyenne pendant cette période.

La capacité de traitement de la station d'épuration de la région bauloise et du SIVOM de Piriac-sur-Mer (218000 équivalents habitants) est alors très largement dépassée (345000 habitants), conduisant à une augmentation importante de la charge bactériologique lors de l'arrivée des rejets en mer, dans la baie du Pouliguen (Fig. III.23).

Par ailleurs, le taux de raccordement au réseau d'assainissement est insuffisant en milieu urbain, et une charge polluante importante est drainée par le réseau hydrographique de l'étier du Pouliguen qui se déverse dans la baie du Pouliguen en contaminant les eaux de mer et consécutivement les gisements naturels de coques de la plage Benoit.

#### Protocole d'échantillonnage et recueil des données météorologiques

Le gisement naturel de coques s'étend sur une surface de 160 ha. 11 stations y sont positionnées, selon une grille systématique, avec une maille d'environ 400 m (Fig III.24). A chaque station une vingtaine de coques sont prélevées, dans le but d'obtenir un poids de chair suffisant à l'analyse (minimum 50 g).

26 sorties ont été effectuées pendant un an du 20 juillet 1989 au 12 juillet 1990, soit en moyenne une sortie bimensuelle, fréquence parfois non respectée (durée inter-sorties variant d'une semaine à presque un mois en novembre-décembre 1989 et mai-juin 1990).

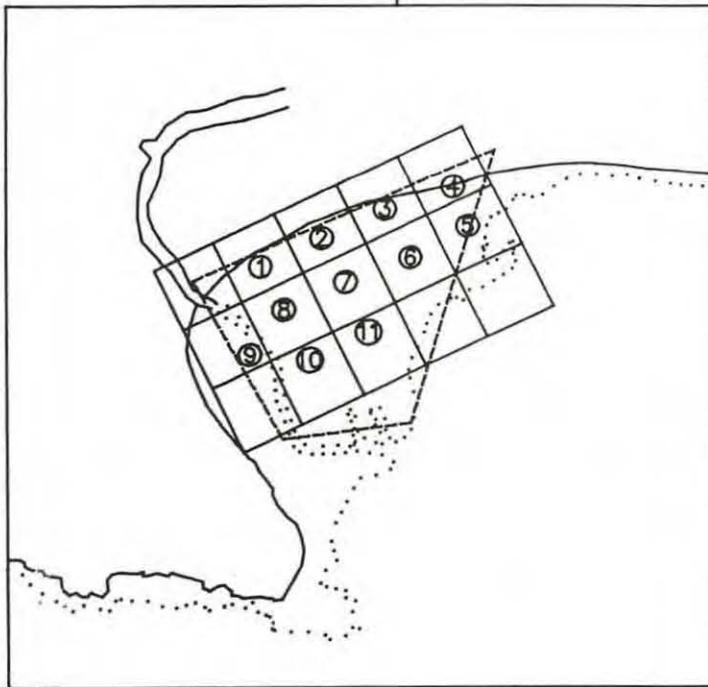
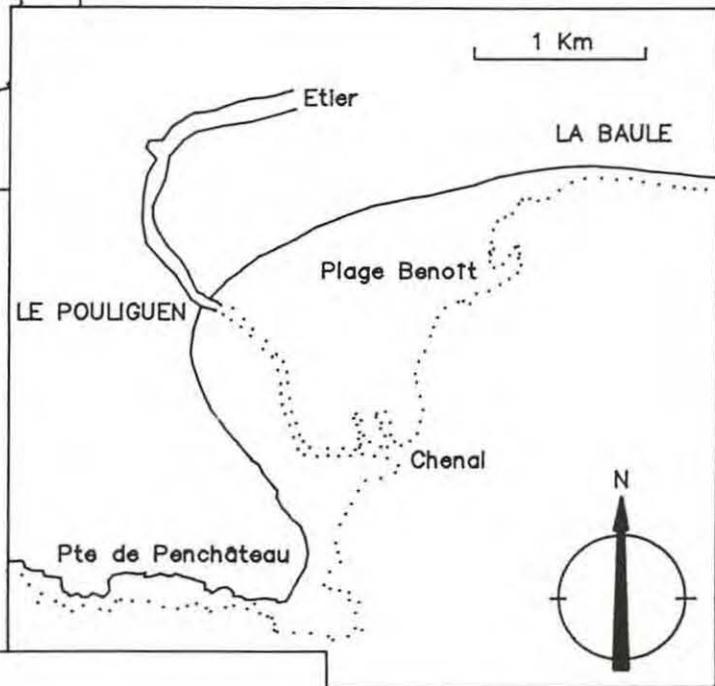
Des données météorologiques journalières pour l'année d'étude ont été obtenues par l'intermédiaire de la station de Saint-Nazaire: pluviométrie, force et direction du vent moyen maximum (moyenne calculée sur 6 minutes).

#### Protocole analytique

"La contamination de la zone est déterminée", conformément à l'arrêté de 1976, "sur la base d'isolement de germes tests de contamination fécale, présents dans les coquillages vivant au lieu considéré".



**Fig. III.23 :** Schéma de la Baie du Pouliguen  
(Loire-Atlantique)



**Fig. III.24 :** Grille d'échantillonnage (d'après Catherine *et al.*, 1991)

Le dénombrement des germes indicateurs est indirect : la mesure de la concentration est une estimation du "Nombre le Plus Probable", après utilisation de la méthode de dilution des tubes (§ I.2.2.2), pour trois dilutions et trois tubes par dilution.

### 3.2.2. Méthodes de traitement des données

Pour chaque station échantillonnée selon le protocole défini plus haut est effectué le calcul des fréquences, et des pourcentages correspondants, de mesures appartenant aux classes définies par l'arrêté de 1976. L'évolution de la contamination par station est également représentée. Des liaisons entre la colimétrie et la pluviométrie sont recherchées à l'aide du coefficient de corrélation de rangs de Spearman  $r_s$  (Lebart *et al.*, 1982). Outre sa robustesse vis-à-vis des méthodes paramétriques, il présente l'avantage de ne pas recourir à l'inévitable et arbitraire découpage inhérent à l'utilisation des tableaux de contingence. On notera néanmoins son manque de puissance et le fait qu'il ne prenne pas en compte la chronologie des observations. Une analyse factorielle des correspondances a néanmoins été tentée par ailleurs (Catherine *et al.*, 1991).

La détermination d'éventuelles sous-zones est effectuée au moyen de l'analyse des lignes polygonales, décrite par ailleurs (§ III.2.1.2). En outre, l'hétérogénéité fixe de la colimétrie aux onze stations est testée au moyen du test non paramétrique de Friedman (Siegel, 1956). Ce test constitue l'analogue non paramétrique d'une analyse de variance à deux facteurs croisés, ici les facteurs "date de prélèvement" et "numéro de station"; il repose sur la comparaison des rangs de la colimétrie obtenue à chaque station, date par date.

La moyenne géométrique des résultats bruts est calculée, par date, pour la zone ou les zones homogènes déterminées plus haut. Les distributions des résultats bactériologiques sur le terrain étant le plus souvent dissymétriques (§ 2.1.4), cet estimateur est alors préféré à la moyenne arithmétique, dans ce cas biaisée (El-Shaarawi & Pipes, 1982).

La distribution de fréquence des valeurs ainsi obtenues est comparée aux proportions admissibles des valeurs guides de l'arrêté de 1976. La moyenne géométrique est confrontée à d'autres estimateurs de valeur centrale, la médiane et l'estimateur du "Nombre le Plus Probable", celui-ci permettant une estimation à partir du résultat caractéristique "somme" (voir § I.2.3.5).

### 3.2.3. Résultats

#### La contamination bactériologique par station

Considérant les données brutes (Tab. III.14), seules les stations 5 et 6 présentent des résultats conformes aux normes (Fig. III.25) et les plus mauvais résultats sont obtenus aux stations 9 et 10, bordant l'étier du Pouliguen, et à la station 11, située plus à l'ouest (Fig. III.24).

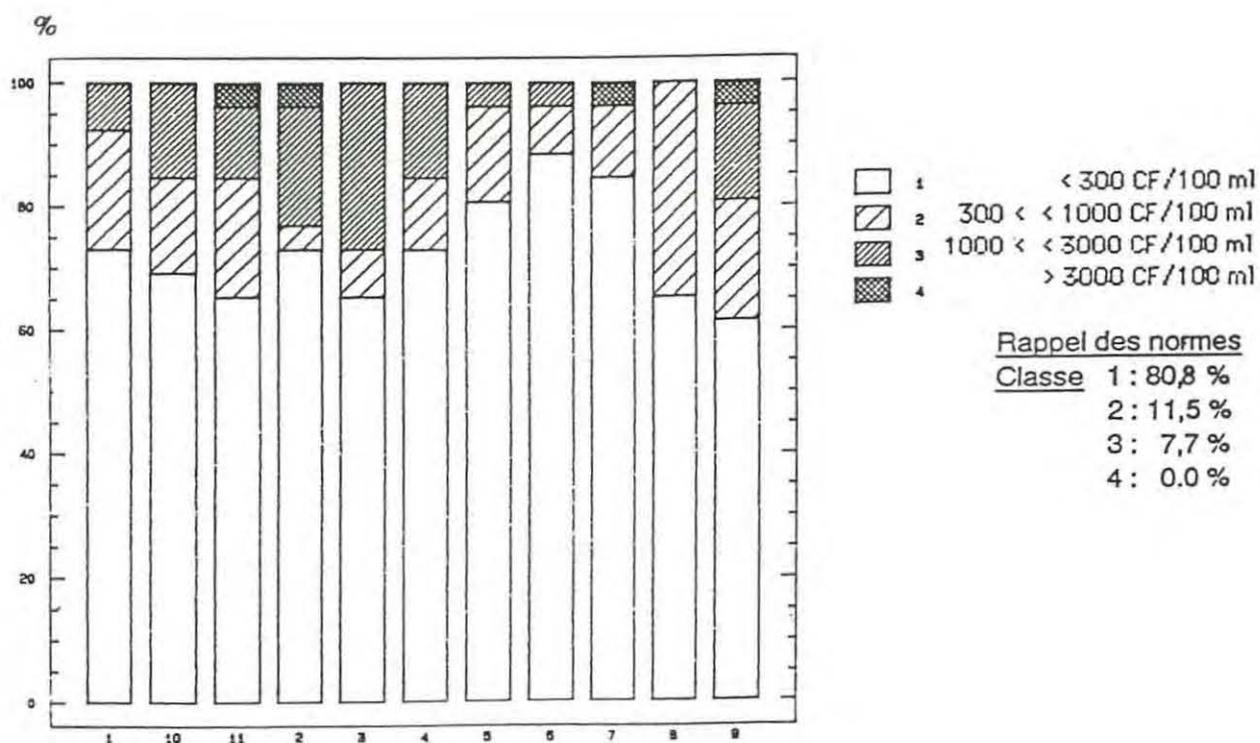
#### Recherche de sous-zones homogènes

L'analyse des lignes polygonales montre des différences significatives entre les profils de contamination des différentes stations (Tab. III.15).

Tab. III.15 : Résultats des comparaisons de profil

	2	3	4	5	6	7	8	9	10	11
1	xx*									
2			xx	xx			xx			
3				xx		xx	xx			
4										
5								xx	xx	xx
6										
7										
8								xx	xx	xx
9										
10										

\*: Au moins un des deux tests utilisés est significatif au niveau de 5%



**Fig. III.25** : Histogramme de fréquences relatives des classes de contamination définies par l'arrêté du 12 octobre 1976, pour chaque station

Tab. III.14 : Résultats bruts du suivi bactériologique de la plage Benoît (07/89 - 07/90) aux 11 stations de prélèvement; les concentrations sont exprimées en CF/100 ml.

Date	1	2	3	4	5	6	7	8	9	10	11
20/07/89	558	1440	558	2580	22	258	138	138	22	138	14400
04/08/89	138	440	258	258	258	90	138	440	2580	258	138
17/08/89	1440	14400	1440	900	90	558	5580	440	2580	1260	1440
31/08/89	120	138	18	22	18	22	90	258	126	258	90
14/09/89	558	2580	1440	138	258	258	558	258	258	138	558
28/09/89	138	55	138	18	55	22	< 18	258	58	< 18	18
12/10/89	558	55	44	55	138	138	258	55	558	90	558
19/10/89	55	258	< 18	18	18	55	< 18	22	55	90	55
02/11/89	138	138	138	258	138	55	22	22	55	558	55
16/11/89	22	55	55	90	18	90	44	22	558	55	55
14/12/89	558	2580	2580	2580	558	2580	55	558	4500	1440	2580
27/12/89	258	138	258	1440	258	138	138	558	900	258	558
11/01/90	138	138	1440	55	558	55	138	258	1440	55	138
25/01/90	168	1440	44	55	258	44	120	55	558	1440	258
01/02/90	55	22	< 18	18	138	< 18	22	22	55	22	22
08/02/90	55	22	22	22	90	< 18	55	90	1440	55	55
01/03/90	258	258	90	558	126	90	558	558	558	558	558
15/03/90	258	138	1440	558	138	258	138	558	258	558	1440
29/03/90	22	22	55	22	258	55	138	55	228	22	22
12/04/90	138	55	126	258	258	22	22	< 18	138	138	55
26/04/90	558	1440	1440	1440	1440	138	138	< 18	< 18	< 18	138
10/05/90	1440	90	1440	258	558	900	258	558	55	558	258
07/06/90	258	44	37	138	18	55	22	558	37	37	37
21/06/90	168	90	558	55	450	258	558	558	138	2580	450
28/06/90	< 18	< 18	< 18	< 18	< 18	< 18	22	< 18	< 18	< 18	< 18
12/07/90	55	< 18	< 18	< 18	< 18	22	22	< 18	55	< 18	< 18

Au vu de ces résultats la partition suivante en sous-ensembles de stations pourrait être imaginée: {2,3}, {5,8}, {1,4,6,7,9,10,11}. Ceci pose néanmoins le problème de la contiguïté spatiale, non résolu pour la station 4 en particulier.

Le test de Friedman se révèle non significatif ( $p = 0,27$ ) : on ne met pas en évidence de différence de colimétrie entre les différentes stations. Un manque de puissance du test pourrait en être à l'origine. Siegel (1956) montre qu'elle est similaire à celle obtenue par l'analyse de variance paramétrique à deux facteurs croisés. On doit suspecter une forte interaction entre la date de prélèvement et la station considérée vis-à-vis des résultats bactériologiques : la différence de concentrations observée entre deux stations est sensiblement fonction de la date de prélèvement. Ceci serait imputable à l'influence des conditions météorologiques.

Cette interaction constitue certainement un inconvénient à l'utilisation de la méthode de l'analyse des lignes polygonales dans l'objectif de déterminer des sous-zones homogènes en matière de bactériologie de l'environnement. Ceci est sans doute beaucoup plus gênant ici que lors de la détermination de sous-secteurs dans le contexte de la surveillance microbiologique de l'environnement marin (§ 2.1), compte tenu des échelles spatiales différentes : la distance inter-stations est ici de 250 m alors que celle séparant deux stations du réseau de surveillance est souvent de l'ordre de quelques km.

Pour la suite des calculs, l'ensemble des stations est considéré.

#### Liaisons avec les variables météorologiques

Une concordance très nette a pu être observée entre les pics de contamination et les fortes précipitations (pluviométrie sur cinq jours), pour la station 10 (Fig. III.26), située au voisinage de l'étier. La faible valeur du coefficient de Spearman correspondant ( $r_s = 0,35$ ;  $p = 0,08$ ) est imputable à l'utilisation des rangs dans son calcul : les pics ont autant de poids que les faibles valeurs et la concordance est alors masquée par le bruit inter-pics. La concordance n'est pas

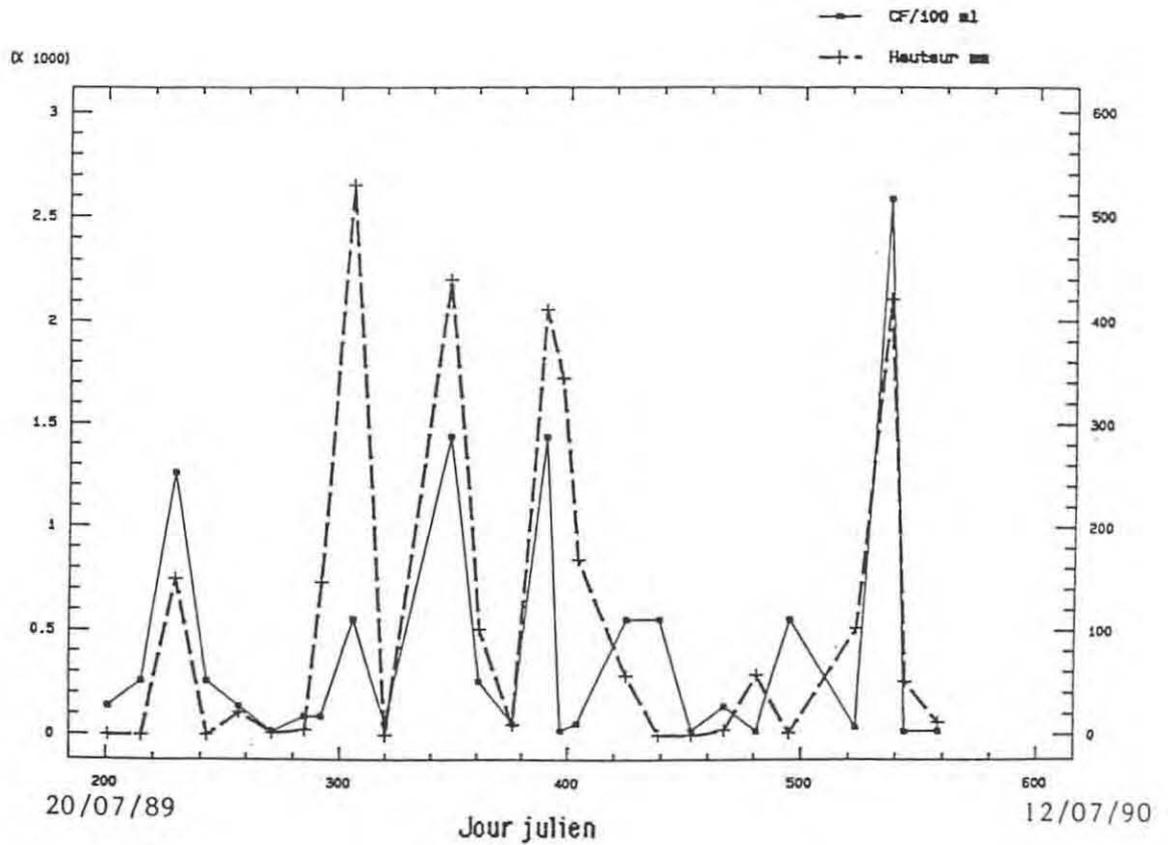
aussi visible et la liaison encore bien moins bonne aux autres stations et en particulier à la station 9 ( $r_s = 0,06$ ;  $p = 0,77$ ) bien que directement située sous les apports en eaux usées de l'étier. Cette station est la seule à être située dans une zone à sédiments sablo-vaseux, à l'ouest de l'étier, la colimétrie y serait alors plus liée à la remise en suspension des sédiments, processus dont l'importance du rôle est une hypothèse déjà émise (§ 2.2.4). Compte tenu des faibles courants de marée sur la zone, le vent serait alors le facteur déterminant de la contamination de la station n° 9.

L'étude de la structure spatiale de la contamination sur le site de la Vanlée a montré l'influence probable de la direction du vent sur la structure spatiale de la contamination bactériologique. La direction du vent conditionne ici l'éloignement ou *a contrario* le plaquage des rejets d'eau douce à la côte. Une analyse factorielle des correspondances a révélé pour cette étude l'importance de ce facteur dans la contamination de la station 9 : le maintien de la pollution à la côte est imputable aux vents de secteur sud-sud-ouest (Catherine *et al.*, 1991). Des mesures de salinité en caractérisant l'impact du rejet permettraient sans doute de mieux appréhender les différences de contamination entre les stations.

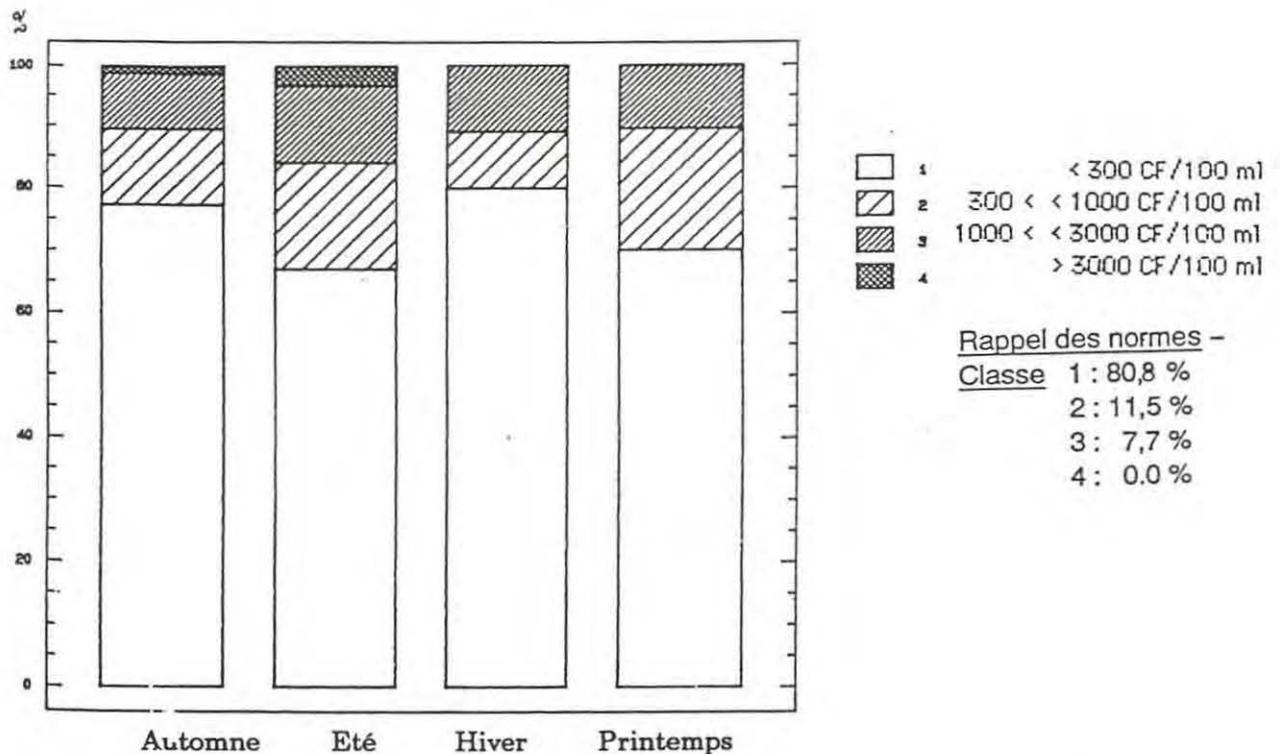
### Résultats synthétiques

Le tableau III.16 présente les estimations des concentrations en indicateurs à chaque date pour l'ensemble des stations et ceci pour trois estimateurs différents, moyenne géométrique, médiane et estimateur du maximum de vraisemblance (NPP). En se référant strictement aux valeurs seuils de l'arrêté, on constate que seule la moyenne géométrique conduit à classer le gisement de coques en zone insalubre.

Le regroupement des résultats par date et par saison montre que l'hiver est la saison présentant la moins forte contamination, alors que la période estivale est la moins "salubre": 34% des concentrations sont supérieures à 300 CF/100 ml (Fig. III.27).



**Fig. III.26 :** Evolutions de la contamination fécale (concentration en coliformes fécaux) et de la pluviométrie (précipitations cumulées sur cinq jours) à la station 10



**Fig. III.27 :** Histogramme de fréquences relatives des classes de contamination définies par l'arrêté du 12 octobre 1976, par saison

### 3.2.4. Formulation de l'avis

Suivant l'estimateur de valeur centrale considéré (Tab. III.16), l'avis diffère : d'une non-conformité à l'arrêté pour la moyenne géométrique, les résultats redeviennent conformes si l'on considère la médiane ou l'estimateur NPP. Ces paramètres statistiques ont en commun le fait de gommer les pics de contamination et ainsi de lisser l'évolution de la contamination pour l'ensemble de la zone.

La variabilité analytique est importante : par exemple, l'intervalle bayésien correspondant à une estimation ponctuelle égale de 1440 CF/100 ml (Résultat caractéristique 3/3/0) donne une limite inférieure de 240 CF/100 ml et une limite supérieure de 5940 CF/100 ml (d'après De Man, 1983), supérieure à la valeur impérative de l'arrêté. Cependant la variabilité analytique ne peut à elle seule expliquer les pics significatifs observés et l'objectif de protection du consommateur nécessite la prise en compte des maxima de concentration. Il est alors plus contraignant, mais plus "sûr" de tenir compte des observations obtenues à chaque station, qui, toutes, hormis les stations 5 et 6, conduisent à une non-conformité vis-à-vis de la réglementation.

L'interaction des facteurs environnementaux sur la colimétrie aux différentes stations, caractérisée par la non-existence d'une hétérogénéité fixe, empêche de définir d'éventuelles sous-zones homogènes. On ne peut raisonnablement scinder le gisement naturel en une partie "salubre" et une partie "insalubre". La totalité du gisement est ainsi concernée par l'avis de maintien en zone insalubre. Cependant, si un tel suivi devait être reconduit, l'échantillonnage des stations 9, 10 et 11 (les plus contaminées) avec un échantillonnage temporel orienté si possible vers des conditions météorologiques propices aux contaminations maximales (vents de sud-sud-ouest et fortes précipitations) serait particulièrement indiqué : protocole allégé et information équivalente ou supérieure.

Tab. III.16 : Contamination "moyenne" de la plage Benoît (07/89 - 07/90), en CF/100 ml

Date de prélèvement	Moyenne géométrique	Médiane	Estimation NPP
20/07/89	332	258	199
04/08/89	269	258	243
17/08/89	1283	1440	780
31/08/89	69	90	70
14/09/89	410	258	380
28/09/89	46	55	39
12/10/89	144	138	116
19/10/89	41	55	33
02/11/89	92	138	85
16/11/89	56	55	53
14/12/89	1195	2580	676
27/12/89	328	258	308
11/01/90	198	138	152
25/01/90	186	168	132
01/02/90	29	22	26
08/02/90	57	55	53
01/03/90	304	558	221
15/03/90	367	258	335
29/03/90	52	55	53
12/04/90	77	126	71
26/04/90	211	138	112
10/05/90	385	558	266
07/06/90	60	37	55
21/06/90	312	450	217
28/06/90	18	18	2
12/07/90	23	18	11

La formulation de l'avis ne doit pas uniquement être fondée sur les résultats de colimétrie obtenus et ceci d'autant plus que les observations se situent au voisinage des valeurs seuils édictées par le texte réglementaire. Ainsi, il convient de resituer l'étude dans un cadre temporel élargi, en particulier pour les paramètres climatiques :

- La pluviométrie des années 1989 et 1990 a été sensiblement inférieure à la pluviométrie annuelle moyenne. Or, les précipitations importantes perturbent le fonctionnement du couple réseau d'assainissement/station d'épuration par des intrusions, dans le réseau, d'eaux de ruissellement ou en provenance de la nappe phréatique. Une année "sèche" peut ainsi entraîner une sous-estimation de la contamination bactériologique du secteur.

- Les vents sud-sud-ouest, responsables du maintien des rejets à la côte, ont été moins fréquents que pour les années précédentes, en moyenne, et inversement pour les vents entraînant ces rejets. Des vents orientés plus ouest, correspondant à une année plus typique pour ce facteur, auraient probablement contribué à affecter encore la qualité bactériologique du milieu.

Les études de zones dans l'objectif d'un classement de salubrité ou du maintien en zone insalubre doivent, au delà de toute considération trop mathématique, révéler l'état de santé d'un bassin versant et ainsi constituer une mesure de l'efficacité de l'assainissement. Catherine *et al.* (1991) soulignent que malgré les aménagements réalisés sur le réseau d'assainissement du SIVOM de la Baule, des efforts restent à fournir et en particulier pour la pollution provenant du ruisseau de la Torre, se déversant dans l'étier du Pouliguen. Rendant compte de cette source de pollution, la DDASS de Loire-Atlantique (1990) avait également émis en 1989 un avis de maintien en zone insalubre.

L'avis formulé est un élément d'appréciation dans la prise de décision finale. Cet élément est fondé sur la distance séparant les mesures observées des exigences de la réglementation, fondées sur des valeurs-guide. Le lecteur trouvera d'autres exemples d'évaluation de la qualité

bactériologique du milieu dans Boeddu *et al.* (1976), Volterra *et al.* (1980), Tosti & Volterra (1981) et plus récemment Ramteke *et al.* (1992).

### 3.3. Critique de l'arrêté

#### 3.3.1. Couverture spatiale

L'arrêté ne donne pas d'indications concernant les modalités d'échantillonnage, en particulier le nombre de stations de prélèvement et le type de plan d'échantillonnage à pratiquer. Ceux-ci sont bien sûr fonction des caractéristiques de la zone. Le quadrillage total de la zone où s'étend la zone conchylicole à classer semble le plan d'échantillonnage spatial le plus raisonnable. Le positionnement des stations doit permettre de connaître les maxima de concentration sur le site, mais également d'avoir une idée de l'extension du phénomène dans l'espace.

La définition d'éventuelles sous-zones homogènes passe, *a posteriori*, par la connaissance de l'évolution de la structure spatiale de la contamination bactériologique. L'identification de la structure est rendue difficile par l'intervention des facteurs météorologiques, eux-mêmes influant sur les paramètres hydrologiques. Suivant le positionnement de la station, l'effet d'un certain facteur météorologique n'induera pas les mêmes modifications du niveau de contamination à une date donnée. Nous retrouvons ici la notion d'interaction entre le facteur "date" et le facteur "station" explicitée dans les résultats précédents. Lewis (1978) a déjà noté le rapport causal entre les phénomènes météorologiques et consécutivement hydrologiques, et l'interaction spatio-temporelle pour les populations phyto-planctoniques.

#### 3.3.2. Fréquence d'échantillonnage

Le texte fait mention de 26 prélèvements échantillonnés régulièrement, soit une fréquence de 14 jours. Le choix d'un plan systématique peut présenter un inconvénient majeur si le descripteur observé, ici le niveau de contamination, est **en phase** avec un autre phénomène (Scherrer, 1983), ce qui est le cas avec la périodicité lunaire. En effet, les coques sont alors

systématiquement prélevées soit en mortes-eaux, soit en vives-eaux. Cette concordance peut introduire un biais dans la mesure de la salubrité. Une étude sur l'Elorn (Pommepuy *et al*, 1987) montre que de forts coefficients induisent une remise en suspension des sédiments, pièges à bactéries, entraînant une élévation importante de la densité bactérienne dans les eaux. Par ailleurs, Catherine (1992) a observé une remarquable concordance entre la colimétrie et le coefficient de marée. Sans informations préalables sur la zone, et si les contraintes opérationnelles le permettent (accessibilité sur le site aux faibles coefficients), une solution serait d'échantillonner à une semaine d'intervalle dans chaque mois. (exemple: 09/01, 16/01, 09/02, 16/02, ..., 09/12, 16/12).

Les mois seraient alors échantillonnés régulièrement, avec simplement une répartition intramensuelle différente des dates de prélèvements. Dans cette étude, la fréquence n'a pas été scrupuleusement de 14 jours. Plus que dans l'espace où il est difficile de préjuger de l'action des facteurs environnementaux sur la répartition de la pollution, l'utilisation d'un plan systématique dans le temps est critiquable, dans la mesure où dans le domaine de la santé des consommateurs, on devrait s'orienter vers une quête raisonnée des maxima de concentration. Une telle idée a déjà été avancé par Mazurié (1988), à propos de la mise en place d'un réseau microbiologique. Il conviendrait alors de connaître l'impact des accidents météorologiques sur la contamination du secteur, comme ont pu le réaliser Teste-Cristofoli & Baleux (1979) pour tenter d'expliquer la contamination bactériologique de l'étang de Thau.

### *3.3.3. Fondement probabiliste du texte*

On suppose que la distribution des concentrations bactériennes suit une loi lognormale. Cette hypothèse faite, on peut obtenir des estimations des deux paramètres de la distribution gaussienne des logarithmes décimaux des concentrations compte tenu des fréquences admissibles indiquées par la norme (Fig. III.28).

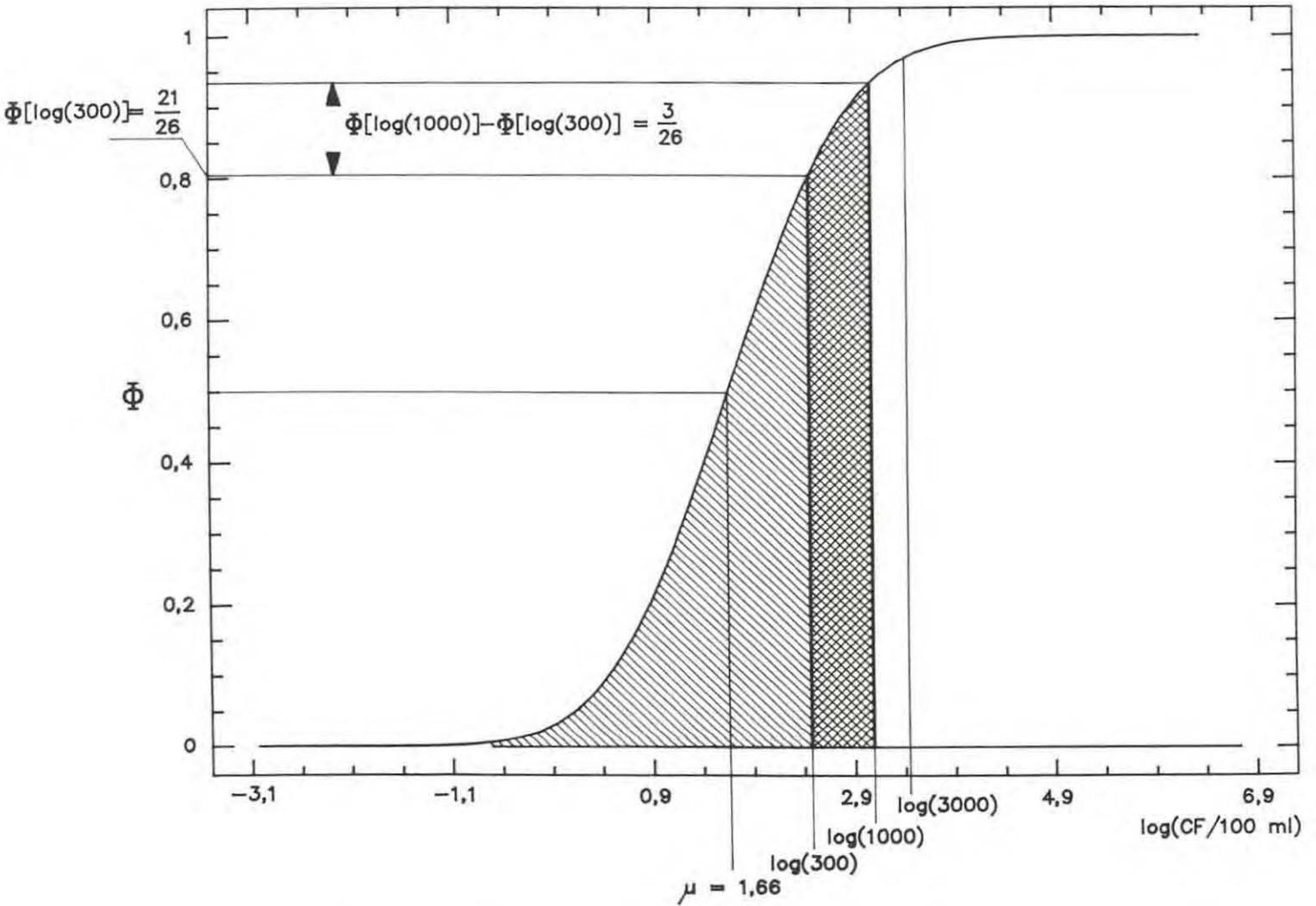


Fig. III.28 : Distribution normale  $N(1,66; 0,94)$  ajustée aux fréquences relatives obtenues à partir de l'arrêté du 12 octobre 1976

Soient,

- $\mu$  = la valeur moyenne théorique de la distribution normale
- $\sigma$  = l'écart-type théorique de la distribution normale
- $\Phi$  la fonction de répartition de la distribution  $N(0, 1)$

Alors, la détermination de  $\mu$  et  $\sigma$  passent par la résolution des équations, en utilisant la table normale centrée réduite:

$$\begin{cases} \Phi\left[\frac{\log(300) - \mu}{\sigma}\right] = 1 - \frac{5}{26} \\ \Phi\left[\frac{\log(1000) - \mu}{\sigma}\right] - \Phi\left[\frac{\log(300) - \mu}{\sigma}\right] = \frac{3}{26} \\ \Phi\left[\frac{\log(3000) - \mu}{\sigma}\right] - \Phi\left[\frac{\log(1000) - \mu}{\sigma}\right] = \frac{2}{26} \end{cases}$$

Le déterminant du système constitué par les deux premières équations n'étant pas nul, il y a une solution unique:

$$\begin{cases} \mu = 1,66 \log(\text{CF}/100 \text{ ml}) \\ \sigma = 0,94 \log(\text{CF}/100 \text{ ml}) \end{cases}$$

Ces valeurs des paramètres de la loi doivent vérifier également la troisième relation; après calcul, et sous l'hypothèse de validité des paramètres de la distribution, on obtient un nombre de mesures comprises entre 1000 et 3000 CF/100 ml compris entre 1 et 2 (1,3...), sur les 26 requises par l'arrêté.

Concernant la valeur de 3000 CF/100 ml, la valeur correspondante de la fonction de répartition est égale à 1,93... très proche de 1,96, c'est-à-dire la valeur de la loi normale réduite n'ayant que 2,5% de chances d'être dépassée. Cette quasi-coïncidence laisse à penser que l'ajustement effectué n'est pas totalement irréaliste et qu'un raisonnement probabiliste serait à la base des valeurs figurant dans l'arrêté. Tel que formulé, l'arrêté fixe les 3000 CF/100 ml

comme une valeur impérative, ce qui est plus contraignant, mais du côté de la protection de la santé humaine.

### 3.4. Conformité de résultats bactériologiques à une norme

Le problème de la conformité à une norme n'a pas été traité ici de façon approfondie et nous ne présentons pas de véritables études dans ce domaine. Ce chapitre a pour but de présenter quelques notions et outils qu'il conviendrait de rattacher à la récolte d'un échantillon dans l'objectif du respect de la conformité. En outre, il nous paraît nécessaire de cerner l'ensemble des problématiques dans le contexte de la surveillance bactériologique en milieu marin.

#### 3.4.1. Problématique et quelques définitions

Pour une interprétation satisfaisante d'une campagne de suivi bactériologique, on doit accepter l'existence d'un risque d'erreur lié à toute décision ou toute formulation d'un avis. Ainsi les résultats doivent-ils se situer dans un contexte probabiliste. La décision va porter sur le rejet ou non d'hypothèses définies *a priori*. Considérons le corps d'hypothèse suivant :

- $H_0$  (hypothèse nulle) : la zone considérée est bactériologiquement salubre
- $H_1$  (hypothèse alternative) : la zone considérée est bactériologiquement insalubre

L'acte d'échantillonnage va générer, de par les résultats obtenus, deux types d'erreur :

- L'erreur de première espèce qui correspond à la probabilité  $\alpha$  de rejeter  $H_0$  alors que  $H_0$  est vraie.
- L'erreur de seconde espèce qui correspond à la probabilité  $\beta$  de ne pas rejeter  $H_0$  quand  $H_0$  est fausse.

Cette dualité s'exprime dans notre cas en des termes moins formels; supposons que la zone étudiée soit salubre mais que les aléas de l'échantillonnage fassent que les résultats obtenus

témoignent d'une contamination significative du site et conduisent alors à un classement insalubre. On commet alors l'erreur de première espèce qui se produira dans  $100\alpha$  % des cas. On parlera du risque "producteur" car la décision ou l'avis vont pénaliser le producteur ou le pêcheur à pied.

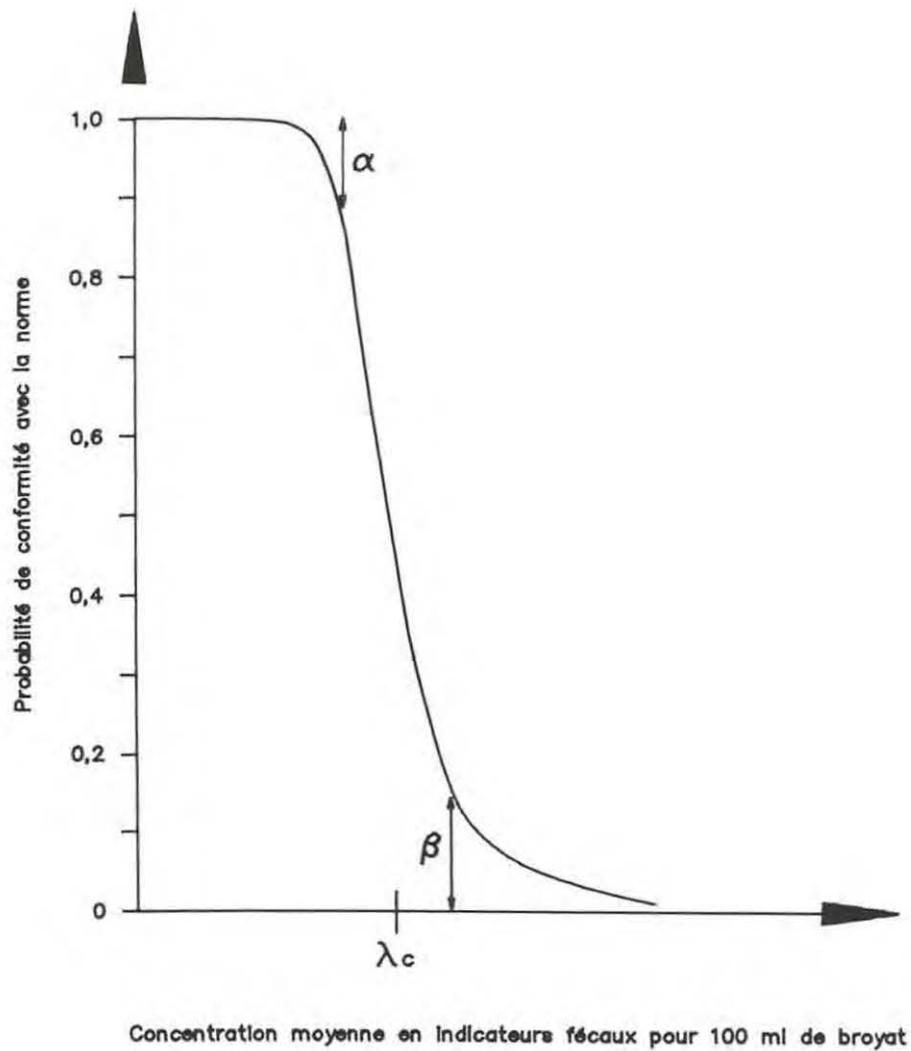
A l'inverse, l'échantillonnage peut conduire à ce que la zone soit déclarée salubre alors qu'elle ne l'est pas. Compte tenu de la ou des distributions en abondances en germes indicateurs fécaux, ceci adviendra dans  $100\beta$  % des cas. Il s'agit du risque "consommateur", dans la mesure où la décision erronée est dangereuse pour la santé humaine.

Le choix des valeurs de  $\alpha$  et  $\beta$  est fonction du coût des erreurs correspondantes. On peut considérer ingénument que la santé humaine n'a pas de prix et ainsi vouloir  $\beta = 0$ . On peut par contre considérer que l'impact socio-économique d'une fermeture de zone soit le coût prépondérant face au coût dû à la poignée d'estivants atteints de gastro-entérites. Dans ce cas, on privilégiera l'erreur de première espèce.

A effort d'échantillonnage fixé, la diminution de l'erreur de première espèce entraîne corrélativement l'augmentation de l'erreur de seconde espèce (Gros, 1988) : moins il est risqué de conclure à tort à l'insalubrité d'une zone conchylicole alors qu'elle est salubre (diminution du risque "producteur") et plus on augmente nos chances de ne pas classer une zone insalubre alors qu'elle l'est (augmentation du risque "consommateur").

#### *3.4.2. Procédures d'échantillonnage relatives à la conformité à une norme*

Idéalement les procédures d'échantillonnage relatives à la problématique de conformité à une norme doivent permettre de connaître la taille de l'échantillon requis pour des risques  $\alpha$  et  $\beta$  donnés. La courbe de puissance fournit l'interprétation graphique ces deux risques : elle représente la probabilité de conformité à une norme en fonction d'une concentration bactérienne moyenne (Fig. III.29).



**Fig. III.29** : Probabilité d'observer la conformité d'un échantillon en fonction de la concentration bactérienne moyenne, lorsque la concentration maximale admissible est fixée à  $\lambda_c$  CF/100 ml (d'après Maul *et al.*, 1989)

La configuration de cette courbe dépend de deux paramètres : le nombre d'échantillons analysés et le degré d'hétérogénéité du système étudié. Le nombre de prélèvements requis pour un risque "consommateur" de 5% lorsque le niveau moyen de contamination est égal à la concentration maximale admissible par une norme est d'autant plus important que la zone à classer est hétérogène.

On peut construire une courbe de puissance à partir d'une hypothèse de distribution théorique ajustée aux données expérimentales. Dès lors, l'hétérogénéité du système est caractérisée par le paramètre de dispersion de cette distribution.

La littérature fournit des exemples de procédures, rattachées à la construction d'une courbe de puissance, après ajustement d'une distribution théorique aux données expérimentales. Ainsi, Maul *et al.* (1989) proposent un bilan ponctuel de la qualité de l'eau dans un système simple de distribution d'eau potable, fondé sur l'ajustement d'une distribution binomiale négative. Ils généralisent leur procédure à un réseau complexe par décomposition en plusieurs réseaux simples avec une allocation optimale des stations par réseau selon que le risque "consommateur" ou l'effort total ait préalablement été fixé.

S'intéressant à la qualité des eaux de baignade en mer, Sørensen *et al.* (1991) élaborent un contrôle statistique sous l'hypothèse d'une distribution lognormale des concentrations en *E. coli*, mesurées dans l'eau de mer.

Citons encore Bonde (1977) qui établit une courbe sigmoïde de référence, similairement à l'arrêté du 12 octobre 1976. Plusquellec (1984) a utilisé ce type de courbe pour classer des zones en fonction de la position des distributions de fréquence cumulées correspondantes par rapport à la sigmoïde de référence. Breittmayer et Gauthier (1979) proposent également une répartition des concentrations dans les 4 classes de fréquence cumulée de limites correspondant aux percentiles 0,25, 0,5 et 0,75. Mais il s'agit plus dans ces deux derniers cas de caractériser le niveau moyen de la contamination sur une zone que de véritables procédures de contrôle.

On trouvera également des procédures de contrôle fondées sur des hypothèses de distribution binomiale dans Mazurié (1988). De tels protocoles sont facilement adaptables à la problématique de décision de réouverture d'une zone préalablement contaminée par des germes pathogènes. La zone pourrait être l'étang de Thau et les germes pathogènes des salmonelles...

### *3.4.3. Les limites des procédures*

#### Ajustement d'une loi théorique et simulation

Nous avons déjà souligné la difficulté à ajuster des distributions théoriques aux données bactériologiques. Ceci rend arbitraire l'utilisation de tels protocoles à moins, comme Sørensen *et al.* (1991), de se satisfaire (et pourquoi pas ?) d'ajustements significatifs à des niveaux de signification supérieurs à 5%.

Nous n'avons pu réaliser un ajustement correct de loi théorique sur les données du suivi bactériologique de la plage Benoît. En particulier, la distribution des valeurs observées s'écarte très sensiblement d'une loi lognormale. Par conséquent, on doit recourir à d'autres techniques pour évaluer le risque "consommateur", telle que le rééchantillonnage dont nous présentons ci-dessous une application possible.

Une fréquente formulation de norme consiste à fixer une valeur à ne pas dépasser pour un certain pourcentage  $p$  des observations. Ainsi, la directive européenne du 30 octobre 1979 (1979) précise que la qualité des eaux conchylicoles est vérifiée dès lors qu'au plus 25% des échantillons prélevés présentent une contamination supérieure à 300 CF/100 ml de broyat de chair de mollusques filtreurs.

La méthode de rééchantillonnage du bootstrap (Efron *in* Lebart *et al.*, 1982) a été appliquée à l'échantillon constitué des données brutes du suivi bactériologique de la plage Benoît. Cela consiste à former des échantillons de même taille ( $n = 286$ )

par tirage aléatoire simple avec remise dans l'échantillon initial. Pour chaque échantillon ainsi simulé, on détermine le percentile  $X_p$  correspondant à la probabilité  $p$  prévue par la norme, c'est-à-dire la valeur de la concentration en indicateurs fécaux telle qu'elle ne soit dépassée que dans  $100p$  %. A l'issue de  $N$  simulations on construit la courbe de fréquence cumulée de la variable aléatoire  $X_p$  et on dénombre les échantillons donnant un percentile inférieur à la valeur prévue par la norme; soit  $n_\beta$  ce nombre, le risque "consommateur" est alors donné par  $n_\beta/N$ .

En illustration, pour une concentration de 730 CF/100 ml, correspondant à  $X_{0,75}$  pour la fonction de répartition de la distribution ajustée à partir des valeurs de l'arrêté de 1976, on obtient  $\beta = 4,1\%$  : 41 valeurs de  $X_{0,75}$  sur les 1000 calculées sont inférieures à 730 CF/100 ml.

#### Imprécision de la mesure

Si l'on ne dispose que d'un faible nombre de prélèvements, la variabilité de l'estimateur NPP peut entraîner une augmentation de l'erreur de seconde espèce. Pour illustrer notre propos, considérons le cas extrême d'un unique prélèvement et supposons que la concentration moyenne en coliformes fécaux soit exactement égale à 3000 CF/100 ml. Alors, le calcul montre que l'analyse conduira à des estimations NPP inférieures à cette valeur deux fois sur trois (Tab. III.17) et donc  $\beta = 0,66$ . Dès lors qu'on multiplie les échantillons,  $\beta$  diminue mais l'influence de l'imprécision se fera d'autant plus sentir à effort fixé que la composante de variabilité analytique, incluse dans la variance résiduelle, est forte devant les autres composantes de la variance totale, qu'elles soient spatiales ou temporelles.

Tab. III.17 : Estimation NPP (en CF/100 ml), probabilité (f) et probabilité cumulée (F) pour une concentration vraie de 3000 CF/100 ml (avec  $n_1 = n_2 = n_3 = 3$ ;  $q_3 = q_2/10 = q_1/100 = 0,01$  ml et pour une suspension diluée préalablement au 1/3)

Résultat caractéristique	NPP	f	F
Autres résultats	...	...	$10^{-4}$
320	558	0,0045	0,0046
321	900	0,0087	0,0133
322	1320	0,0056	0,0189
323	1440	0,0012	0,0201
330	1440	0,2187	0,2388
331	2760	0,4255	0,6643
332	6600	0,2761	0,9404
332	/	0,0597	1,000

Elaboration d'une norme : choix d'un percentile

Les normes en matière de qualité bactériologique des eaux sont fondées sur la comparaison de la distribution des concentrations à une loi de référence, comme pour l'arrêté de 1976, ou bien, et ceci semble le cas le plus fréquent, font référence à un percentile, comme pour la directive de 1979. Dans ce dernier cas, quel percentile choisir ? Berthouex et Hau (1991) montrent qu'en choisissant un percentile élevé, en l'occurrence le 99<sup>ième</sup>, on s'expose à un risque "consommateur" élevé du fait de la forte variabilité de cet estimateur. La médiane représente alors le meilleur choix. Les auteurs définissent un certain intervalle ("*fuzzy region*") à l'intérieur duquel l'importance des erreurs de décision est rédhitoire : si le percentile réglementaire se situe dans cet intervalle centré sur le percentile observé, les risques d'erreur, définis à l'aide de la courbe puissance, sont jugés trop importants.

Sélectionner la médiane lors de l'élaboration est le bon choix si cette même norme fixe des objectifs pour les risques "consommateur" et "producteur", par exemple  $\alpha = \beta = 5\%$ . Dans le

cas où ces objectifs ne sont pas fixés, à cause par exemple de problèmes d'ajustement, la protection de la santé humaine requièrerait que soit fixée une valeur impérative suffisamment contraignante pour garantir une sécurité de presque 100% (100% ne sont jamais atteints). On conçoit qu'en matière de concentrations en coliformes fécaux, un dépassement de 50% d'une certaine concentration, le cas de la médiane, autorise la possibilité d'obtenir un contingent de fortes valeurs, alors que fixer un fractile élevé,  $X_{0,95}$  ou  $X_{0,99}$ , avec un seuil de tolérance faible constitue une sécurité. Chiffrer cette sécurité est impossible sans connaissance préalable du risque sanitaire réel.

Reprenant l'idée de Beder (1991), les consommateurs de coquillages sont moins intéressés par la connaissance de la contamination moyenne d'un secteur que par celle de la fréquence de dépassement d'une valeur maximale.

#### Connaissance du risque sanitaire

"Quelle est la probabilité d'être malade quand une analyse effectuée sur un broyat de coquillages révèle une certaine concentration en coliformes fécaux?"

Il n'est malheureusement pas établi, dans le domaine marin, de liens entre la concentration en coliformes fécaux, germes-test empruntés aux eaux-douces, et le risque sanitaire. Maul *et al.* (1990) précisent que "le manque de données épidémiologiques fiables et la connaissance relativement imparfaite de l'incidence pathologique sur la population, (...), ont contraint les autorités responsables en santé publique à fixer les normes de salubrité sur des bases caractérisées par un manque d'objectivité" et ajoutent que "toutefois, la justification des normes actuellement en vigueur est pour une large part fondée sur le pragmatisme des législateurs".

De ce point de vue, l'arrêté de 1976 semble suffisamment contraignant pour que le niveau moyen de contamination, obtenu à partir du modèle probabiliste sous-jacent, garantisse une salubrité bactériologique des coquillages satisfaisante, et par la même un risque sanitaire acceptable, sans qu'on soit en mesure de chiffrer ce degré d'acceptabilité.

Des facteurs environnementaux interagissent vraisemblablement sur les relations coliformes fécaux/germes pathogènes (voir § I.1.4). Ainsi, on peut penser qu'une même concentration en coliformes fécaux puisse être accompagnée de concentrations en germes pathogènes très différentes selon, par exemple, l'intensité lumineuse ou bien la turbidité de l'eau de mer. Ceci est valable dans le temps et dans l'espace : une concentration de 600 CF/100 ml mesurée dans les coquillages ne correspond sans doute pas au même risque sanitaire selon que la mesure est effectuée en été ou en hiver, mais également selon que l'on se trouve à Sète ou à Saint-Malo. Ceci mérite d'être signalé, même si la nécessité de l'unicité d'une réglementation, au moins nationale sinon européenne, paraît devoir ne pas être remise en question.

### **3.5. Conclusion**

A partir d'un cas concret, a été réalisé un traitement des données bactériologiques dans l'objectif du respect de la réglementation en vigueur en matière de classement de salubrité des zones conchylicoles.

Outre l'examen des résultats obtenus au cours de l'année de suivi, le remplacement du problème dans un contexte temporel plus élargi nous a conduit à émettre un avis de maintien en zone insalubre.

La mise en évidence de liaisons entre les facteurs météorologiques (direction, force du vent et pluviométrie) et la colimétrie permet de détecter les causes éventuelles de la contamination du secteur et à terme d'orienter l'échantillonnage pour un futur suivi du site considéré.

Si les fréquences retenues par l'arrêté de 1976 semblent avoir une base probabiliste, sa formulation apparaît floue et incomplète : toute réglementation visant à protéger la santé humaine devrait impérativement garantir un risque "consommateur". La non-connaissance du risque sanitaire lié à la concentration en indicateurs fécaux autorise un certain pragmatisme dans l'élaboration des normes : il convient alors de sélectionner une valeur impérative contraignante, sans pour autant gréver trop sensiblement la production conchylicole locale.

Enfin, un suivi bactériologique motivé par une réglementation fournit un moyen objectif d'attirer l'attention des collectivités lorsque l'environnement littoral et consécutivement les consommateurs de coquillages sont menacés par la pollution bactériologique. Les collectivités sont les demandeurs d'enquête de salubrité bactériologique de leurs zones conchylicoles naturelles, alors interdites à la pêche de loisir et dont ils apprécieraient la réouverture, compte tenu de l'attrait touristique exercé (*a contrario* une zone fermée a une connotation négative vis-à-vis de l'estivant). Un avis de maintien en zone insalubre doit amener la prise de conscience par les collectivités de l'imperfection de leur système d'assainissement, visualisée au travers des résultats bactériologiques de l'enquête.

RECOMMANDATIONS ET CONCLUSION GENERALE

Un réseau de surveillance est évolutif. L'obtention des premiers résultats du réseau, ainsi que les études exploratoires qui ont pu être menées au cours de ce travail rendent compte des difficultés à satisfaire les objectifs définis initialement.

Ceci nous amène à envisager une amélioration des stratégies. En pratique, les nouvelles options proposées doivent tenir compte des contraintes afférentes à l'existence du réseau microbiologique.

Il n'est pas dans mon intention de proposer ici la nouvelle physionomie du réseau microbiologique, mais plutôt de tenter d'apporter quelques éléments de réflexion utiles à ce propos, après avoir au préalable rappelé les principaux résultats obtenus.

#### *Précision de la mesure du descripteur bactériologique*

Dans le contexte du réseau microbiologique de l'IFREMER (REMI), la concentration en coliformes fécaux mesurée dans les coquillages a été retenue comme indice de la qualité bactériologique du milieu.

La mesure de cet indice au laboratoire est une estimation du "Nombre le plus Probable" de bactéries initialement présentes dans l'échantillon. Telle qu'obtenue en routine cette estimation est particulièrement imprécise, comme en témoigne la largeur des intervalles bayésiens. De plus, sa distribution est très éloignée d'une loi normale, ce qui nécessite un recours fréquent aux statistiques non paramétriques lors de l'exploitation des données. Par ailleurs, le test "exact" se substitue au test classique dans l'objectif de la comparaison de deux estimations NPP indépendantes.

La technique conductance-métrique fournit un signal quantitatif plus continu, la prédiction de la concentration en coliformes fécaux restant imprécise. Elle est sensiblement moins onéreuse à l'usage et permet un grand nombre d'analyses simultanées.

### ***Représentativité de l'échantillonnage et risque sanitaire***

Des études exploratoires ont permis d'évaluer la représentativité d'un échantillon pour diverses échelles spatio-temporelles. L'étude réalisée à l'échelle d'une station de surveillance (de l'ordre de la centaine de m<sup>2</sup>) permet d'étayer l'hypothèse d'intégration de la variabilité spatiale, réalisée par les mollusques. Ainsi, à cette échelle, la variabilité analytique est-elle prépondérante.

Cette homogénéité à petite échelle réalisée par les mollusques rend possible l'observation de macro-structures à l'échelle d'un secteur de surveillance (de l'ordre de la dizaine de km<sup>2</sup>). L'hétérogénéité ainsi révélée doit nous inciter à déterminer des zones homogènes pour l'obtention de séries représentatives d'un secteur donné. La recherche des niveaux moyens et des tendances passe également par une augmentation de la fréquence d'échantillonnage.

L'alerte bactériologique ou le classement de salubrité des zones conchylicoles répondent au même objectif de protection de la santé humaine. Pour probabiliste qu'ait été le raisonnement à la base de l'élaboration de certaines réglementations, il n'en reste pas moins que le risque sanitaire réel n'est pas connu ce qui autorise *in fine* un certain pragmatisme dans l'examen de la conformité à une norme. Dans ce contexte et dans un souci d'efficacité de la procédure d'échantillonnage ressort naturellement la nécessité d'une quête des maxima de concentrations en temps et lieux choisis. La contribution de l'étude de l'influence des facteurs environnementaux et en particulier météorologiques y est prépondérante.

### ***Perspectives d'évolution des différentes stratégies***

On peut schématiser le concept d'évolution des stratégies d'un réseau ainsi : si les résultats obtenus ne satisfont pas aux objectifs initiaux, les stratégies doivent être améliorées. Le coût induit par les nécessaires modifications ou toutes autres contraintes peut être prohibitif. Ce constat effectué, il reste la possibilité de limiter l'ambition du réseau à des objectifs réalisables et

par voie de conséquence réalistes. Le tableau III.18 résume cette démarche appliquée au réseau microbiologique.

Les améliorations prévues, telle que par exemple l'augmentation de la fréquence d'échantillonnage, entraînent sans doute un surcoût de fonctionnement prohibitif pour le réseau dans son état actuel. L'optimisation de la stratégie doit alors se situer au plus haut niveau : à coût total constant, on ne peut envisager une surveillance bactériologique efficace sur l'ensemble du littoral pour la détermination des niveaux et tendances.

Un argument supplémentaire en faveur d'une surveillance plus "locale" vient de l'impossibilité de comparer les différents secteurs entre eux et ceci pour plusieurs raisons :

- Les coquillages sont d'espèces différentes. Nous avons vu que les vitesses de filtration à la base de la concentration différaient sensiblement (§ I.1.4.4). Les différences les plus flagrantes opposent les mollusques filtreurs suspensivores (huîtres, moules,...) aux déposivores (coques, palourdes,..).

- Les modes d'élevage des coquillages sont différents. Par exemple, dans les mêmes conditions, des huîtres cultivées à plat seront plus contaminées que des huîtres en surélevé (Catherine, comm. pers.).

- Un niveau de contamination biaisé : considérant le maillage spatial lâche dans certains cas et l'existence de macro-structures, le niveau de contamination d'une zone est alors totalement dépendant du positionnement des stations. Ainsi, suivant la localisation des stations, tel secteur présentant une unique source de contamination très localisée apparaîtra particulièrement insalubre, alors que tel autre secteur subissant de nombreux rejets polluants pourra se révéler "propre", en raison de la fixité des stations.

- L'interaction des facteurs environnementaux (météorologiques, hydrodynamiques) sur les différents facteurs régissant la survie des bactéries et la contamination des coquillages.

- Enfin, les analyses sont naturellement effectuées par des laboratoires différents d'une région à l'autre. Des différences significatives entre laboratoires ont pu être mises en évidence, pour ce qui est de la technique de dilution des tubes (Beliaeff & Miossec, non publ.). Notons également qu'un laboratoire doit veiller à la reproductibilité des mesures obtenues.

**Tab. III.18** : Tableau résumé des modifications de stratégies à apporter pour satisfaire aux objectifs compte tenu des problèmes et contraintes existant

OBJECTIFS	Détermination des niveaux et tendances de la contamination <i>Coliformes fécaux = indice de la qualité bactériologique du milieu</i>	Protection de la santé humaine <i>Coliformes fécaux = germes indicateurs d'un risque sanitaire</i>
PROBLEMES	<ul style="list-style-type: none"> <li>- Imprécision de la mesure</li> <li>- Maillage spatio-temporel trop important</li> <li>- Existence de stations non informatives</li> </ul>	<ul style="list-style-type: none"> <li>- Non connaissance du risque sanitaire</li> </ul>
STRATEGIES POSSIBLES	<ul style="list-style-type: none"> <li>- Duplication de la mesure</li> <li>- Sélection de zones "intéressantes"</li> <li>- Augmentation ou allègement du nombre de stations de prélèvement</li> <li>- Redéfinition de zones homogènes</li> <li>- Doublement de la fréquence d'échantillonnage</li> </ul>	<ul style="list-style-type: none"> <li>- Recherche des maxima de concentrations en temps et en lieux</li> <li>- Etude des liaisons colimétrie-facteurs environnementaux : recherche de modèles d'alerte et compréhension des mécanismes de la contamination</li> </ul>
CONTRAINTES	<ul style="list-style-type: none"> <li>- Coût total</li> <li>- Choix des stations tributaire de la               <ul style="list-style-type: none"> <li>- Accessibilité aux sites d'échantillonnage</li> <li>- Disponibilité des agents</li> <li>- Disponibilité du matériel d'analyse</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>- Courte durée de vie des indicateurs et longueur de l'analyse</li> <li>- présence de coquillages filtreurs</li> </ul>

Circonscrire un secteur à une zone considérée comme homogène doit permettre, sans forcément contraindre à une augmentation de l'effort d'échantillonnage spatial, l'obtention d'une série chronologique représentative de la qualité bactériologique du milieu à l'échelle correspondante. Selon les cas, un secteur sera scindé en plusieurs zones homogènes, ou bien des stations jugées peu ou non informatives seront abandonnées avec un report d'effort sur une zone sous-échantillonnée.

### *Conclusion*

Les réseaux de surveillance et d'intervention répondent à deux objectifs bien distincts. A problématiques différentes, stratégies d'échantillonnage différentes. D'une part dans le temps, où une fréquence régulière d'échantillonnage s'oppose à un déclenchement suivi d'une fréquence plus importante dans le but d'assister à un retour à la normale. D'autre part dans l'espace, où une stratégie aléatoire devrait s'opposer à une recherche raisonnée des maxima de concentrations. Dans ce dernier cas, la connaissance *a priori* de l'influence des conditions environnementales sur la structure spatiale de la contamination joue un rôle déterminant dans le positionnement optimal des stations de prélèvement. Ceci a pu être mis en évidence lors de l'étude sur le site de la Vanlée. Citons également Chalmer et Edmonds (1986) qui préconisent un critère de positionnement basé sur la direction du vent.

Comment ? Où ? Quand ? sont des questions importantes en matière de stratégies d'échantillonnage. Pourquoi ? est une question fondamentale. Si l'objectif de protection de la santé humaine apparaît clair, il est légitime de s'interroger sur le bien fondé de l'acquisition de séries bactériologiques à long terme dans une approche "environnementale". En d'autres termes, si on peut attribuer une signification statistique à une série après optimisation de la procédure de recueil des données, quel sens lui accorder ? Quelle est la portée de l'information obtenue ? " *Whether changes in the environment are statistically significant has no bearing on the extent to which the changes may be either meaningful or important (i.e., have ecological or human consequences)*" (National Research Council, 1990).

Un peu de prospective : la discussion ne tournera vraisemblablement pas autour de la linéarité de la tendance, mais plutôt sur le caractère discontinu, en paliers, de la série obtenue. Une évolution significative sera plus le fait d'un dysfonctionnement local et sporadique du réseau d'assainissement de telle agglomération - dont on pourrait avoir connaissance par ailleurs - que d'une aggravation régulière de l'état bactériologique du milieu marin.

D'autres pays ont-ils eu ce même objectif de recherche des niveaux et tendances de la contamination bactériologique ? On peut citer le programme national américain de surveillance bactériologique mis en place par la NOAA, où la mesure des spores de *Clostridium perfringens* dans les sédiments est un indice de la qualité bactériologique du milieu (National Research Council, 1990); les stations sont distantes de 80 km en mer ouverte, et de 20 km dans les baies, et le prélèvement est annuel...

Tous les processus liés à la contamination des coquillages n'ont pas encore été élucidés : la remise en suspension des sédiments semble notamment jouer un rôle clef pour la compréhension et nécessiterait des études particulières poussées.

Pour nécessaires qu'elles soient (Frontier, 1983; Clarke & Green, 1988; Camacho & Vascotto, 1991), les conclusions que l'on peut tirer de ces études restent limitées aux conditions pour lesquelles elles ont été obtenues, et notamment les conditions environnementales dont nous avons vu l'importance dans la compréhension des phénomènes. Les modèles numériques hydrodynamiques (voir la revue bibliographique dans Mauvais, 1990; Salomon & Pommepuy, 1991) et plus spécifiquement hydrosédimentaires (Pommepuy *et al.*, 1986; Le Hir *et al.*, 1990) sont susceptibles d'apporter leur concours dans cette problématique, de par la multiplicité des situations qu'ils sont capables de simuler.

BIBLIOGRAPHIE

- Anonyme, 1972. Bactériologie à l'usage des étudiants en médecine. Grouau & Roques, 5ème édition, Lille, 382 p.
- ARMSTRONG M., RENARD D. & BERTHOU P., 1989. Applying geostatistics to the estimation of a population of bivalves. International Council for the Exploration of the Sea, C. M. 1989/K : 37 Sess U.
- AYRES P.A., BURTON H.W. & CULLUM M.L., 1978. Sewage pollution and shellfish. In : Techniques for the study of mixed populations, Society for Applied Microbiology (ed.) Technical series 11, pp. 51-62, BC Press.
- BALEUX B. & TROUSSELLIER M., 1982. Distribution spatiale et échantillonnage des bactéries hétérotrophes dans des sédiments lagunaires superficiels. J. Fr. Hydrol. 13 : 125-139.
- BALEUX B., TROUSSELLIER M., GOT P., MONFORT P., ALIBOU J. & MEZRIOUI N., 1988. Devenir des bactéries "témoins de contamination" et des germes pathogènes d'origine continentale dans les eaux, les sédiments et les productions conchylicoles d'un étang saumâtre. Océanis 14 : 61-70.
- BARCINA I., GONZÁLEZ J.M., IRIBERRI J. & EGEA L., 1989. Effect of visible light on progressive dormancy of *Escherichia coli* cells during the survival process in natural fresh water. Appl. Environ. Microbiol. 55 : 246-251.
- BEAUCHAMP J. J. & OLSON J. S., 1973. Corrections for bias in regression estimates after logarithmic transformation. Ecology 54 : 1403-1407.
- BEDER S., 1991. Statistics and environmental regulation. The many meanings of means. Search 22 : 88-90.
- BELLAIR J. T., PARR-SMITH G.A. & WALLIS I.G., 1977. Significance of diurnal variations in fecal coliform die-off rates in the design of ocean outfalls. J. WPCF 49 : 2022-2030.
- BELLAN G., 1985. Chapitre III. Bioaccumulateurs et bioévaluation. Bull. Ecol. 4 : 215-307.
- BERTHOUX P.M. & HAU I., 1991. Difficulties related to using extreme percentiles for water quality regulations. Research Journal WPCF 63 : 873-879.

- BISSONNETTE G.K., JESESKI J.J., McFETERS G.A. & STUART D.G., 1975. Influence of environmental stress on enumeration of indicator bacteria from natural waters. *Appl. Microbiol.* 29 : 186-194.
- BLOCK J.C., MAUL A. & DOLLARD M.A., 1982. Estimation du nombre le plus probable (NPP) des coliformes en rivière en réalisant 96 ensemencements par niveau de dilution. *Rev. Fr. Sci. Eau* 1 : 387-399.
- BLOCK J.C. & PRIEUR D., 1987. Microbiologie et virologie marines. Synthèse. *Oceanis* 14 : 113-116.
- BOEDDU M, VILLA L. & VOLTERRA L., 1977. State of pollution on italian coast (Summer 1976). *Ann. Sclavo* 19 : 451-454.
- BONADONNA L., VOLTERRA L., AULICINO F.A. & MANCINI L., 1990. Accumulation power of some bivalve molluscs. *Mar. Poll. Bull.* 21 : 81-84.
- BONDE G.J., 1977. Bacterial indication of water pollution. *Adv. Aquat. Microbiol.* 1 : 273-364.
- BONNEFONT J.L., LELONG P. & MARTIN Y., 1986. Etude expérimentale de la disparition de coliformes fécaux d'un effluent urbain en milieu marin. In : GERBAM (ed.) - Deuxième Colloque International de Bactériologie marine - CNRS/IFREMER, Actes de Colloques n° 3, pp. 567-572.
- BOUSQUET G. & DEVOUCOUX J., 1982. Prévision et interprétation des résultats en bactériologie des eaux : facteurs de variabilité et stratégies. *J. Fr. Hydrol.* 13 : 157-169.
- BREITTMAYER J.P., 1978. Fiabilité des résultats. *Rev. Int. Océanogr. Méd.* Tome L pp.,65-68.
- BREITTMAYER J.P. & GAUTHIER M.J., 1979. Dénombrement des bactéries en milieu marin : facteurs de variation et d'incertitude. *Ann. Microbiol.(Inst. Pasteur)* 130 A : 245-256.
- BROWN G.H. & FISHER N.I., 1972. Subsampling a mixture of sampled material. *Technometrics* 14 : 663-668.

- BURGESS T.M. & WEBSTER R., 1980. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. *J. Soil Science* 31 : 315-331.
- CABELLI V.J., DUFOUR A.P., MCCABE L.J. & LEVIN M.A., 1982. Swimming-associated gastroenteritis and water quality. *Am. J. Epidemiol.* 115 : 606-616.
- CAMACHO F. & VASCOTTO G.L., 1990. Framework for enhancing the statistical design of aquatic environmental studies. *Environmental Monitoring and Assessment.* 17 : 303-314.
- CATHERINE M., 1992. Méthodologie des études bactériologiques de zones conchylicoles. IFREMER, RI.DEL/92.01, 96 p.
- CATHERINE M., BELIAEFF B. & PEZERON A., 1991. Etude de salubrité du gisement naturel de coques (*Cerastoderma edule*) de la plage Benoît en baie du Pouliguen - Loire-Atlantique (1989-1990). IFREMER RI-DEL-91.04 - Nantes. 85 p.
- CHALMER P.L. & EDMONDS L.W., 1986. Cape Peron environmental monitoring. *Water* 13 : 20-25.
- CLARK I., 1982. Practical geostatistics. Applied Science Publishers Ltd, London.
- CLARKE K.R. & GREEN R.H., 1988. Statistical design and analysis for a "biological" effects study. *Mar. Ecol. Prog. Ser.* 46 : 213-226.
- COCHRAN W.G., 1950. Estimation of bacterial densities by means of the "Most Probable Number". *Biometrics* 6 : 105-116.
- COCHRAN W.G., 1977. Sampling techniques. John Wiley & Sons, 428 p.
- COHEN J. & SHUVAL H.I., 1972. Coliforms, fecal coliforms, and fecal streptococci as indicators of water pollution.

- COLWELL R.R., BRAYTON P.R., GRIMES D.J., ROSZAK D.B., HUQ S.A. & PALMER L.M., 1985. Viable but non-culturable *Vibrio cholerae* and related pathogens in the environment : implications for release of genetically engineered microorganisms. *Biotechnology* 3 : 817-820.
- CONAN G.Y. & WADE E., 1989. Geostatistical analysis, mapping and global estimation of harvestable resources in a fishery of northern shrimp (*Pandalus borealis*). International Council for the Exploration of the Sea, Statistics Committee C.M. 1989/D : 1.
- COWELL N.D. & MORISETTI M.D., 1969. Microbiological techniques - some statistical aspects. *J. Sci. Fd Agric.* 20 : 573-579.
- CORNAX R., MORIÑIGO M.A., BALEBONA M.C., CASTRO D. & BORREGO J.J., 1991. Significance of several bacteriophage groups as indicators of sewage pollution in marine waters. *Wat. Res.* 25 : 673-678.
- CRESSIE N., 1988. Spatial prediction and ordinary kriging. *Mathematical Geology* 20 : 405-421.
- DALLA VALLE J.M., 1941. Note on the "Most Probable Number" index as used in bacteriology. *Pub. Health Rep.* 56 : 229-234.
- DAN T.B.-B. & STONE L., 1991. The distribution of fecal pollution indicator bacteria in lake Kinneret. *Wat. Res.* 3 : 263-270.
- DAWE L.L. & PENROSE W.R., 1978. "Bactericidal" property of seawater : death or debilitation? *Appl. Environ. Microbiol.* 35 : 829-833.
- DDASS de Loire-Atlantique, 1990. Qualité bactériologique des gisements naturels de coquillages. Littoral de la Loire-Atlantique, 76 p.
- DELHOMME J.-P., 1976. Application de la théorie des variables régionalisées dans les Sciences de l'Eau. *Bull. BRGM (2) III, 4* : 341-375.

- DELATTRE J.M., 1988. Indicateurs fécaux et stress en milieu marin : étude des variations à court terme. *Océanis* 14 : 89-95.
- DELATTRE J.M. & DELESMONT R., 1981. L'analyse des coquillages peut-elle servir au contrôle microbiologique du littoral? *Rev. Int. Océanogr. Méd. Tomes LXIII-LXIV* : 11-16.
- DE MAN J.C., 1975. The probability of Most Probable Numbers. *European J. Appl. Microbiol.* 1 : 67-78.
- DE MAN J.C., 1977. MPN Tables for more than one test. *European J. Appl. Microbiol.* 4 : 307-316.
- DE MAN J.C., 1983. MPN tables, corrected. *Eur. J. Appl. Microbiol. Biotechnol.* 17 : 301-305.
- DE NADAILLAC G. & BRETON M., 1986. Les courants en baie de Vilaine. Synthèse des résultats. IFREMER RI-DERO-86.02 - EL, Brest.
- DE ROUVILLE M. & QUETIN B., 1983. Rejets en mer. Disparition des bactéries. *T. S. M.-l'Eau* 1-2 : 43-48.
- DESLOUS-PAOLI J.M., HERAL M., GOULLETQUER P., BOROMTHANARAT W., RAZET D., GARNIER J., PROU J. & BARILLET L., 1987. Evolution saisonnière de la filtration des bivalves intertidaux dans les conditions naturelles. *Oceanis* 13 : 575-579.
- Directive du Conseil du 30 octobre 1979, 1979. Qualité requise des eaux conchylicoles. *Journal officiel des Communautés européennes* du 10 novembre 1979, n° L 281, pp 47-52.
- DRAPER N.R. & SMITH H., 1966. *Applied regression analysis*. John WILEY & Sons, New-York.

- DUPONT J., MENARD D., HERVE C., CHEVALIER F. & MINIER B., 1989. Application de la mesure de la conductance à la colimétrie des coquillages marins vivants. Premiers résultats. Rapport préliminaire IFREMER.
- DUTKA B.J., 1973. Coliforms are an inadequate index of water quality. *J. Environ. Health* 36 : 39-46.
- EISENHART C. & WILSON P.W., 1943. Statistical methods and control in bacteriology. *Bacteriological reviews* 7 : 57-138.
- ELLIOT E.L. & COLWELL R.R., 1985. Indicator organisms for estuarine and marine waters. *FEMS Microbiology Reviews* 32 : 61-79.
- EL-SHAARAWI A.H. & PIPES W.O., 1982. III. Enumeration and statistical inferences. In : *Bacterial indicators of pollution*, Pipes W.O. (ed.), pp. 43-66. CRC Press, Boca Raton, Florida.
- EL-SHAARAWI A.H., 1987. Frequency of sampling required for monitoring the Niagara river. *Can. J. Fish. Aquat. Sci.* 44 : 1315-1319.
- EL-SHAARAWI A.H., ESTERBY S.R. & DUTKA B.J., 1981. Bacterial density in water determined by Poisson or negative binomial distributions. *Appl. Environ. Microbiol.* 41 : 107-116.
- EL-SHARKAWI F., EL-ATTAR L., ABDEL GAWAD A. & MOLAZEM S., 1991. Some environmental factors affecting survival of fecal pathogens and indicator organisms. In : *PNUE/OMS : Cycles biogéochimiques de polluants spécifiques (Activité K) : survie des pathogènes*, pp.21-32. Rapports finaux sur les projets de recherche. MAP Technical Reports Series N° 49. UNEP, Athens, 1991.
- ENZINGER R.M. & COOPER R.C., 1976. Role of bacteria and protozoa in the removal of *Escherichia coli* from estuarine waters. *Appl. Environ. Microbiol.* 31 : 758-763.
- ERCKENBRECHER C.W. JR, 1981. Sediment bacterial indicators in an urban shellfishing subestuary of the lower Chesapeake Bay. *Appl. Environ. Microbiol.* 42 : 484-492.
- ESTERBY S.R., 1982. Fitting probability distributions to bacterial data considerations for surveys and for water quality guidelines. *J. Fr. Hydrol.* 13 : 189-203.

- FATTAL B., PELEG-OLEVSKY E., YOSHPE-PURER Y. & SHUVAL H.I., 1986. The association between morbidity among bathers and microbial quality of seawater. *Wat. Sci. Tech.* 18 : 59-69.
- FIRSTENBERG-EDEN R., 1983. Rapid estimation of the number of microorganisms in raw meat by impedance measurement. *Food. Technol.* 37 : 64-70.
- FIRSTENBERG-EDEN R. & TRICARICO M.K., 1983. Impedimetric determination of total, mesophilic and psychrotrophic counts in raw milk. *J. Food Sci.* 48 : 1750-1754.
- FONTVIEILLE D. & MAUL A., 1985. 8. Dénombrement et mesure d'activité des bactéries. In : *Point sur l'épuration et le traitement des effluents. 2.2. Bactériologie des milieux aquatiques*, Martin G. (ed.), pp. 79-131. Collection Technique et Documentation Lavoisier, Paris.
- FRONTIER S., 1983. Choix et contraintes de l'échantillonnage écologique. In : *Stratégies d'échantillonnage en écologie*, Frontier S. (ed.), pp. 15-62, Masson & Les Presses de l'Université Laval, Québec.
- GALLAGHER T.P. & SPINO D.F., 1968. The significance of numbers of coliform bacteria as an indicator of enteric pathogens. *Wat. Res.* 2 : 169-175.
- GARANS M. & COURTOIS G., 1991. État sanitaire des zones de baignades en mer. *La revue de l'Agence de l'Eau Rhône-Méditerranée-Corse*, 39 : 62-64.
- GARCIA-LARA J., MENON P., SERVAIS P. & BILLEN G., 1991. Mortality of fecal bacteria in seawater. *Appl. Environ. Microbiol.* 57 : 885-888.
- GAUTHIER M.J., MUNRO P.M. & BREITTMAYER V.A., 1991. Adaptation des entérobactéries pathogènes à l'eau de mer (modèle *Escherichia coli*). In : *PNUE/OMS : Cycles biogéochimiques de polluants spécifiques (Activité K) : survie des pathogènes*, pp 31-42. Rapports finaux sur les projets de recherche. MAP Technical Reports Series N° 49. UNEP, Athens, 1991.

- GAUTHIER M. & PIETRI C., 1989. Devenir des bactéries et virus entériques en mer. In : Microorganismes dans les systèmes océaniques, Bianchi M, Marty D., Bertrand J.C., Caumette P., Gauthier M. *et al.* (eds.), pp. 319-342, Masson, Paris.
- GELDREICH E.E., 1966. Sanitary significance of fecal coliforms in the environment. U.S. Department of the Interior. FWPCA. 111 p.
- GERBA C.P. & MAC LEOD J.S., 1976. Effect of sediments on the survival of *Escherichia coli* in marine waters. *Appl. Environ. Microbiol.* 32 : 114-120.
- GIBSON D.M., OGDEN I.D. & HOBBS G., 1984. Estimation of the bacteriological quality of fish by automated conductance measurements. *Int. J. Food Microbiol.* 1 : 127-134.
- GILBERT R.O., 1987. Statistical methods for environmental pollution monitoring. Van Nostrand Reinhold Company, New York. 320 p.
- GOHIN. F, 1990. Analyse géostatistique et cartographie des champs thermiques à partir d'observations obtenues par bateaux et par télédétection spatiale. In : Biométrie et océanographie - Société de Biométrie (ed.), 6, Lille - 26-28 mai 1986. IFREMER - Actes de colloque n° 10, pp. 65-86.
- GOLDBERG E.D., BOWEN V.T., FARRINGTON J.W., HARVEY G., MARTIN J.H., PARKER P.L., RISEBROUGH R.W., ROBERTSON W., SCHNEIDER E. & GAMBLE E., 1978. The mussel watch. *Environmental conservation* 2 : 101-125.
- GREENWOOD M. JR & YULE G.U., 1917. On the statistical interpretation of some bacteriological methods employed in water analysis. *J. Hyg.* 16 : 36-54.
- GROS P., 1988. Bases statistiques de la stratégie de surveillance du milieu marin. IFREMER Rapports Scientifiques et techniques. N° 12, 150 p.
- GROS P. & HAMON D., 1989. Estimation de la biomasse des bivalves intertidaux (moule, coque) exploités en baie de Saint-Brieuc (Manche Ouest). Rapport Interne IFREMER DFO-EL/89 - 25, 137 p.

- HALVORSON H.O. & ZIEGLER N.R., 1933a. Application of statistics to problems in bacteriology. I. A means of determining bacterial population by the dilution method. *J. Bacteriol.* 25 : 101-121.
- HALVORSON H.O. & ZIEGLER N.R., 1933b. Application of statistics to problems in bacteriology. II. A consideration of the accuracy of dilution data obtained by using a single dilution. *J. Bacteriol.* 26 : 331-339.
- HALVORSON H.O. & ZIEGLER N.R., 1933c. Application of statistics to problems in bacteriology. III. A consideration of the accuracy of dilution data obtained by using several dilutions. *J. Bacteriol.* 26 : 559-567.
- HALVORSON H.O. & ZIEGLER N.R., 1934. Application of statistics to problems in bacteriology. IV. Experimental comparison of the dilution method, the plate count, and the direct count for the determination of bacterial populations. *J. Bacteriol.* 29 : 609-634.
- HERAL M., 1989. L'ostréiculture française traditionnelle. In : *Aquaculture*, Vol. 1, Barnabé G. (ed.), pp 347-397. Techniques et Documentation-Lavoisier, 565 p.
- HOOD M.A., NESS G.E. & BLAKE N.J., 1983. Relationship among fecal coliforms, *Escherichia coli*, and *Salmonella* spp. in shellfish. *Appl. Environ. Microbiol.* 45 : 122-126.
- HOSKINS J.K., 1933. The Most Probable Number of *B. coli* in water analysis. *J. Amer. Water Works Ass.* 25 : 867-877.
- HUGUES B., 1981. Nouvelle utilisation de la méthode du nombre le plus probable en virologie - Application à la mise en évidence et à la quantification des virus dans le milieu hydrique. Thèse de Doctorat d'Etat ès Sciences, Université de Metz.
- HUSSONG D., COLWELL R.R. & WEINER R.M., 1981. Seasonal concentration of coliform bacteria by *Crassostrea virginica*, the eastern oyster, in Chesapeake bay. *J. Food Prot.* 44 : 201-203.
- IBANEZ F., 1990. Interprétation des structures spatio-temporelles en écologie planctonique. In : *Biométrie et océanographie - Société de Biométrie* (ed.), 6, Lille - 26-28 mai 1986. IFREMER - Actes de colloque n° 10, pp. 33-64.

- IBANEZ F., 1991. Treatment of the data deriving from the COST 647 project on coastal benthic ecology : The within-site analysis. In : Space and time series. Data analysis in coastal benthic ecology, Keegan F. (ed.), pp. 5-42, CEE.
- ICES, 1989. Report of the ICES Advisory Committee on Marine Pollution, 1988. Co-operative Research Report n° 160, pp. 18-23.
- IFREMER, 1990. Environnement Littoral. Equinoxe Numéro Spécial n°32, 94 p.
- International Organisation for Standardization, 1984. Microbiology - General guidance for enumeration of presumptive *Escherichia coli* - Most probable number technique. First ed., Ref. ISO 7251-1984 (E).
- JOURNEL A.G. & HUIJBREGTS Ch.J.. 1978. Mining geostatistics. Academic Press, London.
- KENDALL M. & STUART A., 1961. The advanced theory of statistics. Vol. 2, Inference and relationship. Charles Griffin, London.
- KWIATKOWSKI R.E., 1991. Statistical needs in national water quality monitoring programs. Environmental Monitoring and Assessment 17 : 253-271.
- LABELLE R.L., GERBA C.P., GOYAL S.M., MELNICK J.L., CECH I. & BOGDAN G.F., 1980. Relationships between environmental factors, bacterial indicators, and the occurrence of enteric viruses in estuarine sediments. Appl. Environ. Microbiol. 39 : 588-596.
- LARKIN E.P. & HUNT D.A., 1982. Bivalve mollusks : control of microbiological contaminants. Bioscience 32 : 193-197.
- LAUREC A., CHARDY P., LEFEVRE G. & TOULARASTEL F., 1979. Définition d'un état de référence écologique : problèmes d'inférences statistiques. 2<sup>es</sup> Journées de la Thermo-écologie. ISTPM, 14-15 novembre 1979, pp. 158-191.

- LEBARON P., TROUSSELLIER M., GOT P. & BALEUX P., 1990. Distribution spatio-temporelle d'une population bactérienne allochtone (coliformes thermotolérants) dans un écosystème marin côtier (Bassin de Thau, France). *Can. J. Microbiol.* 36 : 307-317.
- LEBART L., MORINEAU A. & FENELON J.P., 1982. *Traitement des données statistiques. Méthodes et programmes.* Dunod.
- LEGENDRE P., BALEUX B. & TROUSSELLIER M., 1984. Dynamics of pollution-indicator and heterotrophic bacteria in sewage treatment lagoons. *Appl. Environ. Microbiol.* 1984. 48 : 586-593.
- LEGENDRE P. & FORTIN M.-J., 1989. Spatial pattern and ecological analysis. *Vegetatio* 80 : 107-138.
- LEGENDRE P., TROUSSELLIER M., JARRY V. & FORTIN M.-J., 1989. Design for simultaneous sampling of ecological variables : from concepts to numerical solutions. *Oikos* 55 : 30-42.
- LE GUYADER F., POMMEPUY M. & CORMIER M., 1991. Implantation d'*Escherichia coli* en pilote expérimental et influence des compétitions de flore. *Can. J. Microbiol.* 36 : 116-121.
- LE HIR P., GUILLAUD J.F., POMMEPUY M., LE GUYADER F. & SALOMON J.C., 1990. Modélisation du transport dissous et particulaire dans l'estuaire de Morlaix. Application au devenir des bactéries entériques. *La Houille Blanche* 3 : 273-278.
- LEROY C., CADIOU Y. et DELAPORTE F., 1977. *Mesure densimétrique de la salinité.* ISTPM, Nantes.
- LEWIS W.M. JR., 1978. Comparison of temporal and spatial variation in the zooplankton of a lake by means of variance components. *Ecology* 59 : 666-671.

- LOYER M.W. & HAMILTON M.A., 1984. Interval estimation of the density of organisms using a serial-dilution experiment. *Biometrics* 40 : 907-916.
- MAC CARTHY J.A., THOMAS H.A. JR & DELANEY J.E., 1958. Evaluation of the reliability of coliform density tests. *Am. J. Pub. Health* 48 : 1628-1635.
- MAC CRADY M.H., 1915. The numerical interpretation of fermentation-tube results. *J. Infect. Dis.* 17 : 183-212.
- MARTEIL L., 1976. La conchyliculture française. Deuxième partie. Biologie de l'huître et de la moule. *Rev. Trav. Inst. Pêches marit.* 2 : 125-320.
- MAUL A., BLOCK J.C. & EL-SHAARAWI A.H., 1985a. Statistical approach for comparison between methods of bacterial enumeration in drinking water. *J. Microbiol. Meth.* 4 : 67-77.
- MAUL A., DOLLARD M.A. & BLOCK J.C., 1982. Etude de l'hétérogénéité spatio-temporelle des bactéries coliformes en rivière. *J. Fr. Hydrol.* 13 : 141-156.
- MAUL A., EL-SHAARAWI A.H. & BLOCK J.C., 1985b. Heterotrophic bacteria in water distribution systems. I. Spatial and temporal variation. *Sc. Tot. Environ.* 44 : 201-214.
- MAUL A., VAGOST D. & BLOCK J.C., 1989. Stratégies d'échantillonnage pour analyse microbiologique sur réseaux de distribution d'eau. *Techniques et documentation* Lavoisier, Paris.
- MAUVAIS J.L., 1991. The role of hydrodynamics in environmental studies. The case of tidal seas. In : *Actes du colloque Environnement des mers épicontinentales*, Chamley H. (ed.), pp. 29-45, Lille, 20-21-22 mars 1990. *Oceanologica acta* vol. sp. n° 11.
- MAZURIE J., 1988. Stratégies d'échantillonnage en conchyliculture. IFREMER. RI.DRV-88-001-RA/La Tremblade, 77 p.

- METCALF T.G., SLANETZ L.W. & BARTLEY C.H., 1973. Enteric pathogens in estuary waters and shellfish. In : Microbiological safety of fishery products, Chichester C.O. & Graham H.D. (eds.), pp. 215-234. Academic Press.
- MITCHELL R., 1968. Factors affecting the decline of non-marine micro-organisms in seawater. *Wat. Res.* 2 : 535-543.
- MIOSSEC L., 1990. Mise en place d'un réseau de surveillance microbiologique du littoral français. IFREMER. RI-DRV - 90.03-CSRU/Nantes. 100 p.
- MIOSSEC L. & RAFFIN B., 1990. Réseau de surveillance microbiologique (REMI). Cartographie des points de surveillance et d'intervention. IFREMER RI-DRV -90.52, Nantes.
- MITCHELL R. & CHAMBERLIN C., 1974. Factors influencing the survival of enteric micro-organisms in the sea. An overview. In : Discharge sewage from sea outfall - Pergamon Press.
- MUNRO P., 1988. Modifications somatiques et adaptation d'*E. coli* en milieu marin. Thèse de 3<sup>ème</sup> cycle, spécialité océanographie. Aix-Marseille II. 155 p.
- National Research Council, 1990. Managing troubled waters. The role of environmental monitoring. National Academy Press, 125 p.
- OGDEN I.D., 1986. Use of conductance methods to predict bacterial counts in fish. *J. Appl. Bacteriol.* 61 : 263-268.
- PAERL H.W., 1975. Microbial attachment to particles in marine and freshwater ecosystems. *Microb. Ecol.* 2 : 73-83.
- PARNOW R.J., 1972. Computer program estimate bacterial densities by means of most probable numbers. *Food Technol.* 26 : 56-62.
- PETITGAS P. & POULARD J.Ch., 1989. Applying stationary geostatistics to fisheries : a study on hake in the bay of Biscay. International Council for the Exploration of the Sea, Demersal fish committee. C.M. 1989/G : 62 Sess U.

- PIPES W.O., WARD P. & AHN S.H., 1977. Frequency distributions for coliform bacteria in water. J. Am. Water Works Assoc. 69 : 664-668.
- PLUSQUELLEC A., 1984. Contribution à l'étude de la pollution bactérienne des eaux littorales. Cas particulier de la Baie de Concarneau-La Forêt. Thèse de 3<sup>ème</sup> cycle, Université de Bretagne Occidentale, Brest, 202 p.
- PLUSQUELLEC A., BEUCHER M., PRIEUR D. & LE GAL Y., 1990. Contamination of the mussel, *Mytilus edulis* Linnaeus, 1758, by enteric bacteria. J. Shell. Res. 9 : 95-101.
- POGGI R., 1986. Epidémiologie en zone littorale conchylicole. Océanis 12 : 439-448.
- POGGI R., 1990. Impacts sanitaires des contaminations microbiologiques. Bendor, 13-15 juin 1990. IFREMER. Actes de colloques, 11, pp. 115-132.
- POMMEPUY M., CORMIER M., BRUNEL L. & BRETON M., 1987. Etude de la flore bactérienne d'un estuaire breton (Elorn, rade de Brest, France). Oceanologica Acta 10 : 187-195.
- POMMEPUY M. & GUILLAUD J.F., 1992. Devenir des bactéries entériques rejetées en mer. TSM L'eau 1 : 49-53.
- POMMEPUY M., GUILLAUD J.F., LE GUYADER F., DUPRAY E. & CORMIER M., 1989. Le devenir de la charge bactériologique des sédiments dragués. Actes du séminaire international sur les aspects environnementaux liés aux activités de dragage. Nantes, 27 novembre- 1<sup>er</sup> décembre 1989.
- POMMEPUY M., GUILLAUD J.F., MARTIN Y., DUPRAY E., DERRIEN A., L'YAVANC J. & CORMIER M., 1990. Le devenir des bactéries en zone littorale. La Mer et les Rejets Urbains. Bendor, 13-15 juin 1990. IFREMER, Actes de colloques 11, 89-100.
- POMMEPUY M., LE HIR P., CORMIER M. & BASSOULET P., 1986. Etude de la dispersion des rejets en rade de Brest. Relations bactéries-particules. T. S. M. L'eau. 11 : 539-543.

- PNUE/OMS, 1991. Etudes épidémiologiques relatives aux critères de la qualité de l'environnement pour les eaux servant à la baignade, à la culture des coquillages et à l'élevage d'autres organismes comestibles (Activité D). Rapport final sur l'étude épidémiologique menée parmi les baigneurs de certaines plages à Malaga, Espagne (1988-1989). MAP Technical Report Series N° 53. UNEP, Athens.
- PRESS W.H., FLANNERY B.P., TEUKOLSKY S.A. & VETTERLING W.T., 1986. Numerical recipes. The art of scientific computing. Cambridge University Press, Cambridge.
- PRIEUR D., 1981. Nouvelles données sur les relations entre bactéries et bivalves marins. *Haliotis*. 11 : 251-260.
- PRIEUR D., MEVEL G., NICOLAS J.L., PLUSQUELLEC A. & VIGNEULLE M., 1990. Interactions between bivalve molluscs and bacteria in the marine environment. *Oceanogr. Mar. Biol. Annu. Rev.* 28 : 277 - 352.
- PROST H. & HUGUES B., 1982. Dénombrement des microorganismes par la technique du nombre le plus probable (indice NPP) : emploi statistique de cet indice. *Ann. Fals. Exp. Chim.* 75 : 185-207.
- QUETIN B., 1991. Hydrodynamique côtière et dispersion des eaux usées. Influence sur l'implantation et la conception des émissaires. *T.S.M. L'eau* 6 : 303-311.
- RAMTEKE P.W., BHATTACHARJEE J.W., PATHAK S.P. & KALRA N., 1992. Evaluation of coliforms as indicators of water quality in India. *J. Appl. Bacteriol.* 72 : 352-356.
- REED J.L., 1925. *B. coli* densities as determined from various types of samples. *Pub. Health Rep.* 40 : 704-716.
- RHODES M.W. & KATOR H., 1988. Survival of *Escherichia coli* spp. in estuarine environments. *Appl. Environ. Microbiol.* 54 : 2902-2907.
- RICHARDS G.P., 1988. Microbial purification of shellfish : a review of depuration and relaying. *J. Food Prot.* 51 : 218-251.

- RIVILLA R. & GONZÁLEZ C.C., 1989. Seasonal variations of pollution indicators in a wildfowl : reserve (Doñana National Park, Spain). *J. Appl. Bacteriol.* 67 : 219-223.
- RNO, 1988. Réseau National d'Observation de la Qualité du Milieu Marin. Dix années de surveillance 1974-1984. Document technique, volume I, présentation générale. IFREMER et Secrétariat d'État auprès du Premier Ministre chargé de l'Environnement, 63 p.
- ROHDE C.A., 1976. Composite sampling. *Biometrics* 32 : 273-282.
- ROSZAK D.B. & COLWELL R.R., 1987. Survival strategies of bacteria in the natural environment. *Microbiol. Rev.* 51 : 365-379.
- RUSSEK E. & COLWELL R.R., 1983. Computation of Most Probable Numbers. *Appl. Environ. Microbiol.* 5 : 1646-1650.
- SALOMON J.C. & POMMEPUY M., 1990. Mathematical model of bacterial contamination of the Morlaix estuary (France). *Wat. Res.* 24 : 983-994.
- SCHERRER B., 1983. Techniques de sondage en écologie. In : *Stratégies d'échantillonnage en écologie*, Frontier S. (ed.), pp. 63-162, Masson & Les Presses de l'Université Laval, Québec.
- SCHWARTZ D., 1963. *Méthodes statistiques à l'usage des médecins et des biologistes*. Flammarion Médecine Sciences, 318 p.
- SCHWARTZBROD L., 1990. Les contaminations par les virus. La Mer et les Rejets Urbains. Bendor, 13-15 juin 1990. IFREMER, Actes de colloques, 11, pp. 101-114.
- SEGAR D.A. & STAMMAN E., 1986. Fundamentals of marine pollution monitoring programme design. *Mar. Poll. Bull.* 17 : 194-200.
- SIEGEL S., 1956. *Non parametric statistics for the behavioral sciences*. Mac Graw-Hill. 312 p.

- SNEDECOR G.W. & COCHRAN W.G., 1957. Statistical methods, 6<sup>th</sup> ed. Iowa State University Press.
- SØRENSEN P.S., JANSEN J.L.C. & SPLIID H., 1991. Statistical control of hygienic quality of bathing water. Environ. Monit. Assess. 17 : 217-226.
- STEIN M.F., 1919. The interpretation of *B. coli* test results on a numerical and comparative basis as applied to the investigation of water and sewage. 4 : 243-265.
- SWAROOP S., 1938. Numerical estimation of *B. coli* by dilution method. Ind. Jour. Med. Res. 26 : 353-377.
- TASSI P., 1989. Méthodes statistiques. Economica, Paris. 474 p.
- TASWELL C., 1981. Limiting dilution assays for the determination of immunocompetent cell frequencies. I. Data analysis. J. Immunol. 126 : 1614-1619.
- TAYLOR L.R., 1961. Aggregation, variance and the mean. Nature 189 : 732-735.
- TESTE-CRISTOFOLI A. & BALEUX B., 1979. Influence du lessivage des sols des bassins versants nord sur la pollution bactérienne des eaux de l'étang de Thau. Bull. Ecol. t. 10, 1 : 55-67.
- THOMAS H.A., 1952. On averaging the results of coliform tests. Jour. Boston Soc. Civil Engrs. 39 : 253.
- TOSTI E. & VOLTERRA L., 1981. Water quality along the napolitean coast. Water, Air, and Soil Pollution 16 : 433-444.
- TROLLOPE D.R., 1984. Use of molluscs to monitor bacteria in water. In : Microbiological methods for environmental biotechnology, Grainger & Lynch (eds.), pp. 393-409, The Society for Applied Bacteriology, Technical Series n° 19, Academic Press.
- TROUSSELLIER M., ALBAT M., ANDRE P. & BALEUX B., 1985. Dénombrements directs des bactéries dans les milieux aquatiques par microscopie en épifluorescence : distribution et précision des mesures. Rev. Fr. Sci. Eau. 4 : 35-49.

- TROUSSELLIER M., BALEUX B. & ANDRE P., 1986. Echantillonnage de variables bactériologiques dans les milieux aquatiques. In : GERBAM (ed.) - Deuxième Colloque International de Bactériologie marine - Actes de Colloques n° 3, CNRS/IFREMER, Brest, pp. 23-33.
- TROUSSELLIER M., MAUL A. & BALEUX B., 1989. Stratégies d'échantillonnage. In : Microorganismes dans les systèmes océaniques, Bianchi M, Marty D., Bertrand J.C., Caumette P., Gauthier M. *et al.* (eds.), pp. 27-62, Masson, Paris.
- TURICK C.E., SEXSTONE A.J. & BISSONNETTE G.K., 1988. Freshwater mussels as monitors of bacteriological water quality. *Water, Air, and Soil Pollution* 40 : 449-460.
- VOLTERRA L. & AULICINO F. A., 1980. Microbiological analyses on a beach of the roman coast. *Ig. Mod.* 74 : 926-937.
- VOLTERRA L., AULICINO F. A., TOSTI E. & ZICARELLI M., 1980. Bacteriological monitoring of pollution in shellfish : methodological evaluation. *Water, Air, and Soil Pollution* 13 : 399-410.
- VON MISES R., 1942. On the correct use of Bayes formula. *Ann. Math. Stat.* 13 : 156-165.
- WALNE P.R., 1972. The influence of current speed, body size and water temperature on the filtration rate of five species of bivalves. *J. Mar. Biol. Ass. U. K.* 52 : 345-374.
- WHITE H., 1989. Relationship of growing water quality to incidence of gastroenteritis in consumers of raw shellfish. Interstate Seafood Quality Seminar Ocean City, Maryland, October 18, 1989.
- WOLMAN A. & WEAVER H.L., 1917. A modification of the McCrady method of the numerical interpretation of fermentation-tube results. *J. Infect. Dis.* 21 : 287-291.
- WOOD P.C., 1979. Public Health aspects of shellfish from polluted waters. In : Biological indicators of water quality, James A. & Evison L. (eds.), pp. 13-1 - 13-15, John Wiley & Sons.
- WOODWARD R.L., 1957. How probable is the Most Probable Number ? *J. Am. Water Works Ass.* 49 : 1060-1068.

WYSHAK G. & DETRE K., 1972. Estimating the number of organisms in quantal assays. *Appl. Microbiol.* 23 : 784-790.

XU H.-S., ROBERTS N., SINGLETON F.L., ATTWELL R.W., GRIMES D.J. & COLWELL R.R., 1982. Survival and viability of nonculturable *Escherichia coli* and *Vibrio cholerae* in the estuarine and marine environment. *Microb. Ecol.* 8 : 313-323.

ANNEXES

## Annexe I.1

Tableau d'allocation de l'effort d'échantillonnage lors de la mise en place du réseau  
microbiologique (REMI). D'après Miossec (1990)

## SURVEILLANCE

Indice  $I_s$ 

$$I_s = S_s \times P$$

 $S_s$  = moyenne géométrique de CF $P$  = production relative  
(valeurs centrées, réduites)

## INTERVENTION

Indice  $I_i$ 

$$I_i = S_i \times P$$

 $S_i$  = % > 1 000 CF $P$  = production relative  
(valeurs centrées, réduites)

Allocation de l'effort en fonction de  $I_s$  et  $I_i$   
sur un total de 7 000 analyses/an au CSRU

60 %

2 points minimum par secteur (92)  
reliquat réparti sur la base

 $I_{s1}$ 

$$\frac{I_{s1}}{\sum I_s} \times 1992 = n_1$$

56

 $\sum I_s$ 

1

n56

 $n_1$ 

— = nombre de points (en + du  
12 minimum) pour le secteur 1

.

.

n56

12

40 %

Répartition proportionnelle à  $I_i$  $I_{i1}$ 

$$\frac{I_{i1}}{\sum I_i} \times 2\ 800 = M_1$$

56

 $\sum I_i$ 

1

=> nombre de points minimum $M_1$ 

— = nombre de points pour le secteur  
12 1

.

.

M56

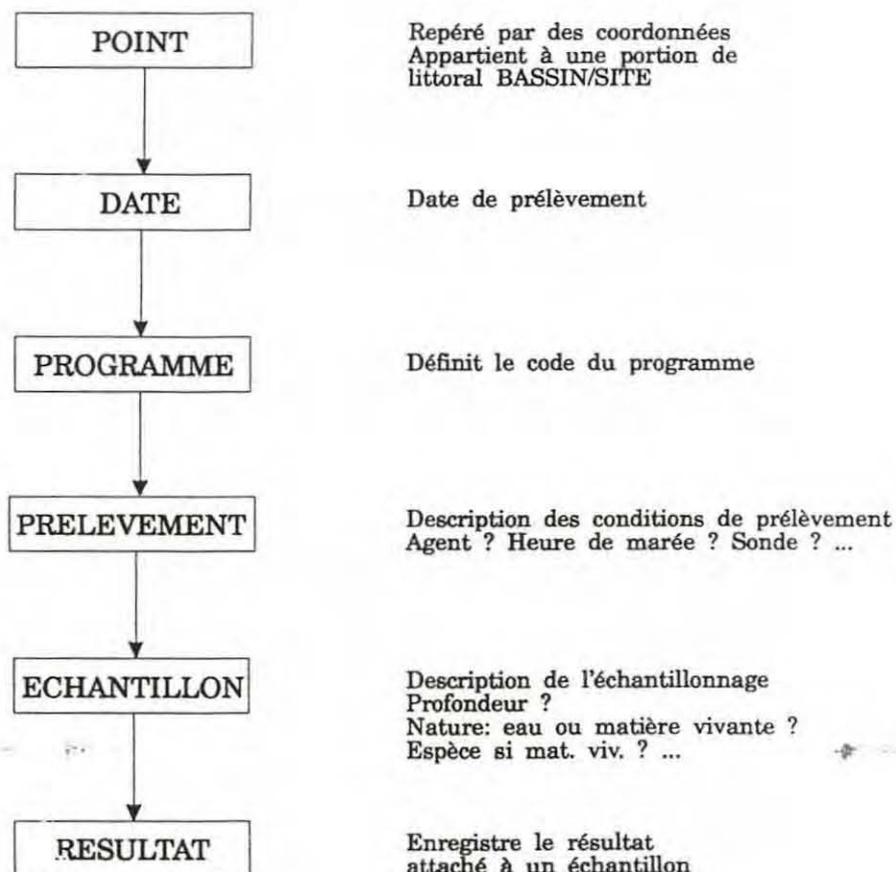
12

## Annexe I.2

## DESCRIPTIF DE LA BASE DE DONNEES

Les données du réseau microbiologique sont stockées sur une base IDS II (30 Mo après 5 ans d'exploitation)  
 La base est implantée sur un DPS 7000 (Bull)  
 à Brest (centre de calcul)  
 Extraction : par programmes COBOL ou requêtes IQS

### DESCRIPTION HIERARCHIQUE des DONNEES



## Règles:

- Un prélèvement peut contenir plusieurs échantillons
- Un échantillon peut donner lieu à plusieurs sous-échantillons
- A un sous-échantillon peut correspondre plusieurs résultats

## Annexe I.3

## ARRETE DU 12 OCTOBRE 1976

fixant les normes de salubrité des zones conchylicoles.

*(Journal officiel du 23 novembre 1976.)*

Le ministre de la santé et le secrétaire d'Etat auprès du ministre de l'équipement (Transports),

Vu le décret du 20 août 1939, modifié par les décrets n° 48-1324 du 25 août 1948 et n° 69-578 du 12 juin 1969 ;

Vu l'avis émis par le conseil supérieur d'hygiène publique de France au cours de sa séance du 26 avril 1976,

Arrêtent :

Article 1<sup>er</sup>.

La salubrité des eaux conchylicoles est déterminée sur la base d'isolement des germes tests de contamination fécale présents dans les coquillages vivant au lieu considéré.

## Article 2.

L'évaluation de la contamination est exprimée par les nombres les plus probables de coliformes fécaux trouvés dans 100 millilitres de chair de coquillages broyée et diluée dans les conditions fixées à l'annexe technique au présent arrêté.

Pour tenir compte des fluctuations naturelles dans la charge microbienne des eaux marines, l'évaluation s'effectue sur vingt-six prélèvements échelonnés sur douze mois consécutifs.

## Article 3.

Remplissent les conditions nécessaires pour être classées salubres les zones dans lesquelles le nombre de coliformes fécaux par 100 millilitres de chair de coquillages ainsi déterminé est inférieur ou égal à 300.

Les normes sont considérées comme respectées si le nombre des résultats en dépassement n'excède pas cinq en douze mois consécutifs, les teneurs en coliformes pour 100 millilitres de chair restant dans ce cas inférieures à 1 000 pour trois des prélèvements et à 3 000 pour les deux autres.

## Article 4.

Les zones ne répondant pas aux conditions fixées dans l'article 3 font l'objet de la procédure de classement en zone insalubre. La récolte des coquillages y est interdite, sauf autorisations données dans les conditions fixées à l'article 5.

## Article 5.

Dans les zones classées insalubres, seule la récolte des coquillages, qui doivent faire ensuite l'objet d'une épuration ou d'un repavage, peut être autorisée par le directeur des affaires maritimes après avis conforme du directeur de l'institut scientifique et technique des pêches maritimes.

Toutefois, lorsque la teneur en coliformes fécaux dépasse 10 000 par 100 millilitres de chair de coquillages dans 25 p. 100 des échantillons, l'autorisation requiert en outre l'avis conforme du directeur départemental de l'action sanitaire et sociale.

## Article 6.

Le directeur général de la santé, le directeur des pêches maritimes et le directeur de l'institut scientifique et technique des pêches maritimes sont chargés, chacun en ce qui le concerne, de l'application du présent arrêté, qui sera publié au *Journal officiel* de la République française et au *Bulletin officiel* de la marine marchande.

Fait à Paris, le 12 octobre 1976.

*Le ministre de la santé,*

Pour le ministre et par délégation :

*Le directeur général de la santé,*  
PIERRE DENOIX.

*Le secrétaire d'Etat*  
auprès du ministre de l'équipement (Transports),

Pour le secrétaire d'Etat et par délégation :  
*Le secrétaire général de la marine marchande,*  
JEAN CHAPON.

## Annexe I.4

Descriptif des programmes informatiques

*Nom.f* : Fortran V

- NPPFIC.f Construction d'un fichier NPP (accès direct, utilisable par d'autres programmes) pour une combinaison de nombres de tubes et de dilutions donnée
- NPPTAB.f Construction des tables NPP pour une combinaison de nombres de tubes et de dilutions donnée (à partir des résultats de NPPFIC.f)
- NPPSIMU.f Simulation de résultats caractéristiques à partir d'une hypothèse de répartition aléatoire des bactéries dans le broyat (Distribution de Poisson ou binomiale)
- DISTRDIF.f Test de comparaison entre deux estimations NPP indépendantes
- NPPBIAIS.f Calcul du biais de l'estimation NPP
- NPPPUIS.f Elaboration de courbes puissance du test "exact" (cf DISTRDIF.f)
- NPPVRAI.f Test du rapport de vraisemblance entre deux estimations NPP
- SIMCHI2.f Vérification de l'adéquation de la distribution simulée des rapports de vraisemblance à une distribution du  $\chi^2$ .

*Nom.pas* : Turbo-Pascal 5.0

- NPPFIC.pas Construction d'un fichier NPP (accès séquentiel) pour une combinaison de nombres de tubes et de dilutions donnée.
- EQVR.pas Calcul des limites de confiance bayésiennes pour un intervalle non symétrique en probabilité.

## Annexe II.1

Table NPP ( $n_1 = 2, n_2 = 5, n_3 = 2; q_3 = q_2/10 = q_1/100 = 0,05/3 \text{ ml}$ ) pour  $p > 10^{-3}$ 

! Resultat !	NPP	!	p	!	int.conf. (95%)	!	int.conf. (99%)	!
! 0 0 1 !	24	!	0.0029	!	6 - 133	!	2 - 177	!
! 0 1 0 !	24	!	0.0745	!	6 - 135	!	3 - 181	!
! 0 2 0 !	50	!	0.0092	!	15 - 179	!	8 - 230	!
! 0 3 0 !	76	!	0.0010	!	28 - 223	!	17 - 279	!
! 1 0 0 !	30	!	0.3680	!	7 - 181	!	3 - 245	!
! 1 0 1 !	62	!	0.0056	!	19 - 241	!	11 - 314	!
! 1 1 0 !	64	!	0.1459	!	20 - 248	!	11 - 324	!
! 1 1 1 !	98	!	0.0039	!	36 - 309	!	22 - 393	!
! 1 2 0 !	101	!	0.0424	!	37 - 319	!	22 - 406	!
! 1 2 1 !	138	!	0.0017	!	56 - 381	!	37 - 476	!
! 1 3 0 !	143	!	0.0095	!	57 - 395	!	38 - 493	!
! 1 4 0 !	190	!	0.0015	!	82 - 477	!	57 - 587	!
! 2 0 0 !	95	!	0.2772	!	30 - 501	!	16 - 686	!
! 2 0 1 !	162	!	0.0117	!	59 - 678	!	35 - 893	!
! 2 1 0 !	172	!	0.3325	!	63 - 751	!	38 - 995	!
! 2 1 1 !	273	!	0.0246	!	104 - 963	!	67 - 1237	!
! 2 2 0 !	301	!	0.3084	!	114 - 1100	!	72 - 1423	!
! 2 2 1 !	448	!	0.0393	!	176 - 1362	!	116 - 1722	!
! 2 2 2 !	611	!	0.0018	!	253 - 1632	!	173 - 2028	!
! 2 3 0 !	517	!	0.2893	!	200 - 1641	!	131 - 2100	!
! 2 3 1 !	715	!	0.0621	!	294 - 2005	!	199 - 2523	!
! 2 3 2 !	930	!	0.0045	!	407 - 2394	!	284 - 2973	!
! 2 4 0 !	880	!	0.3013	!	359 - 2763	!	239 - 3575	!
! 2 4 1 !	1186	!	0.1123	!	515 - 3472	!	355 - 4432	!
! 2 4 2 !	1555	!	0.0143	!	707 - 4305	!	501 - 5422	!
! 2 5 0 !	1955	!	0.4284	!	867 - 12357	!	570 - 17186	!
! 2 5 1 !	4282	!	0.4978	!	1672 - 26549	!	1067 - 36234	!

Table NPP ( $n_1 = 9, n_2 = 9, n_3 = 9; q_3 = q_2/10 = q_1/100 = 0,01/3 \text{ ml}$ ) pour  $p > 5 \cdot 10^{-2}$ 

! Resultat !	NPP	!	p	!	int.conf. (95%)		!	int.conf. (99%)		!		
! 1 0 0 !	32	!	0.3490	!	8	-	177	!	3	-	236	!
! 2 0 0 !	67	!	0.2421	!	21	-	244	!	11	-	314	!
! 2 1 0 !	101	!	0.0615	!	37	-	299	!	23	-	375	!
! 3 0 0 !	107	!	0.1874	!	39	-	317	!	24	-	399	!
! 3 1 0 !	144	!	0.0719	!	59	-	375	!	39	-	463	!
! 4 0 0 !	153	!	0.1513	!	62	-	402	!	41	-	497	!
! 4 1 0 !	194	!	0.0810	!	85	-	464	!	59	-	567	!
! 5 0 0 !	208	!	0.1239	!	91	-	504	!	64	-	617	!
! 5 1 0 !	254	!	0.0891	!	119	-	574	!	86	-	696	!
! 6 0 0 !	275	!	0.1013	!	128	-	633	!	92	-	772	!
! 6 1 0 !	328	!	0.0962	!	160	-	717	!	119	-	866	!
! 7 0 0 !	362	!	0.0810	!	176	-	814	!	130	-	991	!
! 7 1 0 !	426	!	0.1021	!	216	-	922	!	163	-	1116	!
! 7 2 0 !	494	!	0.0675	!	258	-	1037	!	198	-	1247	!
! 8 0 0 !	484	!	0.0614	!	243	-	1108	!	183	-	1362	!
! 8 1 0 !	569	!	0.1057	!	295	-	1273	!	225	-	1557	!
! 8 2 0 !	664	!	0.0962	!	352	-	1457	!	272	-	1774	!
! 8 3 0 !	770	!	0.0605	!	415	-	1663	!	324	-	2015	!
! 9 1 0 !	841	!	0.1034	!	436	-	2209	!	332	-	2794	!
! 9 2 0 !	1029	!	0.1502	!	536	-	2714	!	411	-	3408	!
! 9 2 1 !	1235	!	0.0517	!	643	-	3159	!	495	-	3919	!
! 9 3 0 !	1282	!	0.1632	!	663	-	3346	!	509	-	4163	!
! 9 3 1 !	1556	!	0.0708	!	797	-	3864	!	613	-	4749	!
! 9 4 0 !	1637	!	0.1516	!	830	-	4136	!	635	-	5099	!
! 9 4 1 !	1996	!	0.0848	!	1002	-	4735	!	767	-	5773	!
! 9 5 0 !	2133	!	0.1309	!	1055	-	5136	!	804	-	6285	!
! 9 5 1 !	2579	!	0.0960	!	1279	-	5837	!	975	-	7073	!
! 9 6 0 !	2798	!	0.1100	!	1371	-	6450	!	1037	-	7858	!
! 9 6 1 !	3331	!	0.1062	!	1664	-	7299	!	1265	-	8820	!
! 9 6 2 !	3892	!	0.0543	!	1995	-	8185	!	1528	-	9821	!
! 9 7 0 !	3684	!	0.0904	!	1822	-	8316	!	1375	-	10137	!
! 9 7 1 !	4337	!	0.1161	!	2206	-	9429	!	1679	-	11420	!
! 9 7 2 !	5036	!	0.0784	!	2633	-	10618	!	2027	-	12789	!
! 9 8 0 !	4952	!	0.0711	!	2496	-	11424	!	1885	-	14073	!
! 9 8 1 !	5836	!	0.1257	!	3020	-	13164	!	2308	-	16141	!
! 9 8 2 !	6824	!	0.1178	!	3609	-	15120	!	2789	-	18454	!
! 9 8 3 !	7941	!	0.0765	!	4273	-	17323	!	3333	-	21039	!
! 9 9 1 !	8797	!	0.1338	!	4542	-	23917	!	3453	-	30415	!
! 9 9 2 !	10906	!	0.2064	!	5640	-	29832	!	4304	-	37612	!
! 9 9 3 !	13865	!	0.2434	!	7069	-	37382	!	5391	-	46663	!
! 9 9 4 !	18209	!	0.2543	!	9011	-	47070	!	6840	-	58214	!
! 9 9 5 !	24437	!	0.2595	!	11764	-	59840	!	8857	-	73522	!
! 9 9 6 !	32968	!	0.2731	!	15812	-	77832	!	11812	-	95462	!
! 9 9 7 !	45123	!	0.3061	!	22012	-	107128	!	16395	-	132394	!
! 9 9 8 !	65917	!	0.3897	!	32719	-	176249	!	24357	-	224802	!

## Annexe II.2

Résultats théoriques relatifs à la régression linéaire(Snedecor & Cochran, 1957)

Soient X le temps de détection, Y le logarithme décimal de la concentration en coliformes fécaux pour 100 ml de broyat (chair et intervalvaire), n le nombre de couples (X,Y) et  $\hat{Y}$  l'estimation de Y par la régression, pour un X donné. Posons également  $x = X - \bar{X}$ ,  $y = Y - \bar{Y}$ ,  $d_{y \cdot x} = Y - \hat{Y}$ , et  $\hat{y} = \hat{Y} - \bar{Y}$ , où  $\bar{X}$  et  $\bar{Y}$  sont les moyennes respectives des n valeurs de X et de Y. Les résultats afférents au calcul de la droite de régression de l'échantillon,  $\hat{Y} = a + bX$ , sont :

- le coefficient de régression de Y en X :  $b = \frac{\sum_{i=1}^n xy}{\sum_{i=1}^n x^2}$

- la somme des carrés dûs à la régression :  $\left( \sum_{i=1}^n xy \right)^2 / \sum_{i=1}^n x^2$

- la somme des carrés des écarts à la régression :  $\sum_{i=1}^n y^2 - \left( \sum_{i=1}^n xy \right)^2 / \sum_{i=1}^n x^2$

- le carré moyen des écarts à la droite de régression :  $\sum_{i=1}^n d_{y \cdot x}^2 / (n-2) = s_{y \cdot x}^2$

- l'écart-type observé du coefficient de régression :  $s_b = s_{y \cdot x} / \sqrt{\sum_{i=1}^n x^2}$  ; l'écart-type

observé de l'ordonnée à l'origine :  $s_a = s_{y \cdot x} / \sqrt{\frac{1}{n} \left( \sum_{i=1}^n X^2 / \sum_{i=1}^n x^2 \right)}$

- l'écart-type observé de  $\hat{Y}$ , considéré comme estimation de  $\mu$ , avec  $\mu = \alpha + \beta x$  l'équation

de la droite de régression de la population :  $s_{\hat{\mu}} = s_{y \cdot x} / \sqrt{1/n + x^2 / \sum_{i=1}^n x_i^2}$

- l'écart-type de  $\hat{Y}$  :  $s_{\hat{Y}} = s_{y \cdot x} \sqrt{1 + 1/n + x^2 / \sum_{i=1}^n x_i^2}$ . Les limites de confiance d'une valeur Y prédite sont données par la formule  $Y = \hat{Y} \pm t_{n-2} s_{\hat{Y}}$ , où  $t_{n-2}$  est le t de Student pour n-2 degrés de liberté, définissant ainsi l'enveloppe de confiance de la prédiction.

Préalablement à l'utilisation de ces résultats, les hypothèses suivantes doivent être émises, en toute rigueur: pour chaque X choisi, il existe une distribution normale de Y de laquelle la valeur observée de Y a été tirée au hasard; la population des Y qui correspond à une valeur choisie de X a pour moyenne  $\mu$ ; enfin, dans chaque population, l'écart-type de la distribution de Y autour de sa moyenne a la même valeur quelque soit x et on le désigne par  $\sigma_{y \cdot x}$ , dont une estimation sans biais est donnée par  $s_{y \cdot x}$ .

Tableau d'analyse de la variance S.C.E.: somme des carrés des écarts; C.M.: carré moyen.

Source de var.	S.C.E.	Degrés de liberté	C. M.	F
Régression	$\sum \hat{y}^2$	1	$\sum \hat{y}^2$	$(\sum \hat{y}^2) / (\sum d_{y \cdot x}^2) / (n-2)$
Écarts	$\sum d_{y \cdot x}^2$	n - 2	$(\sum d_{y \cdot x}^2) / (n-2)$	
Total	$\sum Y^2$	n - 1		

Le complément à 1 du rapport entre le carré moyen résiduel et le carré moyen "total" exprime la part de variance expliquée par le modèle : cette quantité est également le carré du coefficient de corrélation et mesure l'intensité de la liaison entre les deux variables.

#### *Test de comparaison des droites*

A partir d'un exemple portant sur deux droites, Snedecor & Cochran (1957) présentent le test de comparaison des droites de régression; il s'agit de comparer les variances résiduelles,

puis les pentes et enfin les ordonnées à l'origine. Nous présentons ici le cas de la comparaison des pentes de trois droites, utile pour notre propos, après avoir rappelé le modèle mathématique sous-jacent. Les notations précédentes sont toujours effectives.

Soit le modèle,

$$Y_{ij} = \alpha_i + \beta_i X_{ij} + \epsilon_{ij}$$

où  $i = 1, 2$  et  $3$  désigne les trois mollusques bivalves considérés (huître, moule et coque).

Le tableau permet la comparaison des variances résiduelles et des pentes.

Tableau de comparaison des trois droites de régression

Source de variation	d.l.	Coef rég.	Ecart à la régression		
			d.l.	SCE	CM
Intra					
Huître	$n_1-1$	$b_1$	$n_1-2$	$\left(\sum d_{y.x}^2\right)_i$	$\left(\sum d_{y.x}^2\right)_i / (n_i-2)$
Moule	$n_2-1$	$b_2$	$n_2-2$		
Coque	$n_3-1$	$b_3$	$n_3-2$		
			$N-6$	$S = \sum_{i=1}^3 \left[ \left(\sum d_{y.x}^2\right)_i \right]$	$S/(N-6)$
Différence entre pentes		$\bar{b}$	2	$D = \sum_i w_i (b_i - \bar{b})^2$	$D/2$

avec  $N = n_1 + n_2 + n_3$

où  $w_i = \frac{1}{\sum_{i=1}^3 x_i^2}$  et  $\bar{b}$  est le coefficient de régression pondéré :  $\bar{b} = \frac{\sum_i w_i b_i}{\sum_i w_i}$

Le calcul de  $F = \frac{D/2}{S/(N-6)}$  permet la comparaison des pentes; il est à comparer au  $(F_{N-6}^2)_{0,05}$

donné par la table. La comparaison des ordonnées à l'origine a habituellement moins d'importance à moins que les droites ne soient parallèles.

Tableaux des analyses de variance (huître, moule, coque, tous coquillages)

**Huître**

Source de variation	S.C.E.	Degrés de liberté	C. M.	F
Régression	446,2	1	446,2	1762,6
Écarts (résidus)	32,9	130	0,253	
Total	479,1	131		

**Moule**

Source de variation	S.C.E.	Degrés de liberté	C. M.	F
Régression	391,9	1	391,9	2019,9
Écarts (résidus)	21,9	113	0,194	
Total	413,8	114		

**Coque**

Source de variation	S.C.E.	Degrés de liberté	C. M.	F
Régression	207,2	1	207,2	1077,2
Écarts (résidus)	15,8	82	0,192	
Total	223,1	83		

**Total des coquillages**

Source de variation	S.C.E.	Degrés de liberté	C. M.	F
Régression	1068,0	1	1068,0	1762,6
Écarts (résidus)	72,0	329	0,219	
Total	1140,0	330		

Résidus des régressions linéaires