

INRS

INSTITUT NATIONAL DE RECHERCHE ET DE SECURITE
pour la prévention des accidents du travail et des maladies professionnelles

MAV - NT-373/LT

Juin 1998

Note Technique :

**Evaluation de l'exposition sonore en milieu professionnel :
application de méthodes
géostatistiques à l'estimation du Leq et conséquences sur les
stratégies de mesurage.**

Hans WACKERNAGEL, Christian LAJAUNIE,
Centre de Géostatistique - Ecole des Mines de Paris

Léon THIERY, Michel GRZEBYK
Institut National de Recherche et de Sécurité (INRS)

SIÈGE SOCIAL :
30, rue Olivier-Noyer

75680 PARIS CEDEX 14
01.40.44.30.00

CENTRE DE RECHERCHE :
avenue de Bourgogne - B.P. n° 27
54501 VANDOEUVRE CEDEX
03.83.50.20.00

Avant-propos

Une étude statistique de la variabilité temporelle de séries de mesures d'exposition au bruit professionnel a été réalisée pour améliorer la fiabilité des résultats et proposer des stratégies de mesurage adaptées aux caractéristiques de variabilité en temps des données.

Ce travail a été réalisé dans le cadre d'une convention de recherche passée entre l'INRS et le Centre de Géostatistique de l'École des Mines de Paris, intitulée "Application des méthodes de géostatistique à l'étude de l'environnement du travail (bruit, produits chimiques)".

L'application de méthodes géostatistiques à l'estimation de l'exposition sonore en milieu professionnel est une nouveauté. C'est pourquoi cette note rappelle succinctement les méthodes géostatistiques employées, mais le lecteur intéressé pourra trouver des compléments en consultant la bibliographie.

Les bases de la géostatistique et plusieurs de ses applications dans des domaines proches de l'hygiène industrielle ont été présentées lors d'un séminaire, dont les actes sont regroupés dans l'ouvrage suivant :

Estimation des nuisances, approches statistiques — Séminaire interne INRS, 17 mars 1993. Paris, INRS (Ed.), Notes Scientifiques et Techniques, N° 104, 1993, 108p.

Pour faciliter la lecture de cette note, l'introduction et la dernière section, "applications pratiques", de chaque chapitre sont en caractères gras : ils regroupent les points clés de la problématique et les acquis.

La progression de l'analyse, d'un chapitre au suivant, est donnée dans le résumé ci-contre.

Résumé

Après rappel de la problématique de l'évaluation de l'exposition sonore en milieu professionnel (Chapitre 1), les divers outils employés ici sont présentés succinctement (Chapitre 2).

Trois séries de données d'exposition homogènes sont décrites (Chapitre 3) selon les méthodes classiques puis à l'aide d'indicateurs quantifiant la typologie des distributions des L_{Aeq} courts et l'importance de l'effet des "pointes de bruit" dans l'exposition sonore quotidienne. Ces descriptions dépendent de la durée d'intégration élémentaire des valeurs mesurées et une autocorrélation a été mise en évidence dans les séries étudiées (Chapitre 4), indiquant l'existence de deux échelles de durée dans les données d'exposition.

Dans ces conditions, l'estimation de l'exposition sonore et de sa précision doit prendre en compte la structure temporelle des données, reflétée par le variogramme. Ceci passe par le choix d'une modélisation bien ajustée du variogramme expérimental. Dans ce but, les performances de trois modélisations différentes sont discutées (Chapitre 5). Cette discussion amène à exclure l'un des modèles testés et à montrer qu'entre les deux modèles validés, le choix peut être effectué en fonction de la tendance de l'exposition sonore à long terme que décrivent les modèles, fournissant de ce fait une méthode d'extrapolation au-delà de la période observée.

Après clarification des hypothèses et des limites d'emploi de deux normes de mesurage de l'exposition sonore, plusieurs estimateurs de l'exposition sonore et de sa variance ont été élaborés pour tenir compte de l'autocorrélation des données (Chapitre 6); ils ont permis de quantifier les biais sur la précision des résultats quand les données sont analysées sous hypothèse d'indépendance. Ces résultats ont été appliqués pour étudier une journée de travail comprenant plusieurs phases d'exposition très distinctes (Chapitre 7) puis en déduire des stratégies d'échantillonnage optimisées, applicables lorsque les mesurages sont renouvelés périodiquement. Dans ce cas, il a été montré qu'une réduction très sensible du nombre de mesures était possible, sans perte de précision d'estimation, en employant une stratégie de mesurage et des estimateurs de l'exposition adaptés à la structure des données.

Table des matières

1	Introduction	1
2	Présentation méthodologique	3
2.1	Cadre probabiliste : notion de fonction aléatoire	3
2.2	Variabiles acoustiques : dB et exposition sonore	3
2.3	Sélectivité : $T(z)$, $Q(z)$, Gini, U	5
2.4	Mesure de l'autocorrélation : le variogramme	7
2.5	Variogrammes de divers types : interprétation	9
2.6	Applications pratiques	10
3	Description de l'exposition sonore	13
3.1	Description des postes de travail observés	13
3.2	Description acoustique classique en dB	14
3.3	Transformation en exposition sonore	16
3.4	Courbes de répartition $T(z)$ et $Q(z)$	16
3.5	Courbes de sélectivité $Q(T)$	16
3.6	Applications pratiques	20
4	Influence du temps d'intégration	23
4.1	Les trois séries intégrées sur 10s, 1mn et 5mn	23
4.2	Histogrammes pour trois supports	23
4.3	Courbes $Q(z)$ et $Q(T)$ pour trois supports	24
4.4	Variogrammes expérimentaux pour trois supports	24
4.5	Applications pratiques	24
5	Structure temporelle	39
5.1	Première interprétation : le modèle lognormal autohomothétique	40
5.2	Deuxième interprétation : un modèle lognormal stationnaire	49
5.3	Troisième interprétation : un modèle lognormal localement stationnaire	53
5.4	Applications pratiques	55
6	Estimation du L_{EQ}	61
6.1	Analyse de deux textes normatifs	61
6.2	Méthode de comparaison des estimateurs	65
6.3	Estimation lognormale sous hypothèse d'indépendance	65
6.4	L_{EQ} arithmétique ou lognormal ?	68
6.5	Estimation en présence d'autocorrélation	70
6.6	Krigeage lognormal de la moyenne de la journée	71
6.7	Krigeage lognormal de la moyenne à long terme	72

6.8	Krigeage lognormal des trois séries	75
6.9	Applications pratiques	77
7	L'exposition à des phases multiples. Méthodes d'échantillonnage stratifié	81
7.1	Terminologie employée	81
7.2	L'échantillonnage	87
7.3	Application à une série multiphase	92
7.4	Applications pratiques	114
8	Conclusions générales	117
8.1	Synthèse des conclusions partielles	117
8.2	Prolongements	119
	Bibliographie	121
A	Optimisation de l'échantillonnage	123
A.1	Introduction	123
A.2	Echantillonnage libre, et fonction état I inconnue ou ignorée	125
A.3	Echantillonnage libre sur un sous ensemble $V' \subset V$	128
A.4	Echantillonnage avec état I connu	130
A.5	Discussion	132

Chapitre 1

Introduction

Les stratégies de mesurage de l'exposition professionnelle au bruit ont généralement pour objectif principal de savoir si les niveaux d'exposition sonore quotidienne des travailleurs sont inférieurs ou supérieurs aux seuils d'actions réglementaires, fixés à 85 et 90 dB(A). Cette question s'applique à des journées habituelles de travail et vise donc l'exposition à long terme. Or les mesures sont nécessairement limitées en nombre et en durée, réparties dans des périodes d'observations particulières. En évoquant les *stratégies de mesurage*, le problème qui est posé est de savoir comment obtenir une estimation de l'exposition suffisamment précise avec un nombre aussi réduit que possible de mesures.

En milieu professionnel, l'exposition sonore d'une personne varie durant le temps de travail pour différentes raisons : l'alternance des tâches et celle des modes de fonctionnement des machines, les trajets dans l'atelier, le bruit généré par l'activité des autres opérateurs. . . Des *événements acoustiques rares*, mais intenses, peuvent survenir et majorer fortement l'exposition sonore globale d'une journée [31]. Parmi les facteurs susceptibles d'expliquer des variations d'exposition sonore, certains sont déterminés par le contenu et l'organisation du travail. Dans ce cas, ils peuvent être identifiés et contrôlés a priori en employant une méthode d'analyse des circonstances de l'exposition sonore adaptée, conduisant notamment à la constitution de groupes d'exposition homogène [15]. D'autres facteurs entraînent des variations aléatoires et provoquent des incertitudes de mesure qu'il faut pouvoir estimer avant d'interpréter le résultat en terme de dépassement — ou de non dépassement — des seuils d'actions réglementaires.

L'estimation de la précision des mesures d'exposition au bruit est évoquée dans plusieurs normes [2, 6, 9, 33], mais plusieurs questions restent posées :

- d'une norme à l'autre, les estimateurs proposés sont différents. Comment comprendre ces différences et quelle méthode adopter quand les domaines de validité des estimateurs proposés ne sont pas mentionnés ?
- avec le progrès des techniques des appareils, l'usage s'est considérablement développé du mesurage en continu durant une journée de travail de séries de L_{Aeq} courts. Mais les méthodes d'estimation de la précision n'ont pas progressé simultanément et ne considèrent que le cas de données supposées indépendantes. Que faire si les séries mesurées sont autocorrélées ?

- la durée d'intégration élémentaire des séries de L_{Aeq} courts est un paramètre de mesure dont la valeur est fixée arbitrairement, par exemple à 1 s, à 1 mn ou plus... Alors que le choix de ce paramètre influence la variabilité des niveaux sonores mesurés, comment intégrer l'effet de ce paramètre arbitraire dans l'estimation de la précision du niveau sonore global ?

Afin d'apporter des éléments de réponse à ces questions, une étude portant sur la variation en temps de données d'exposition au bruit professionnel a été réalisée. L'analyse des structures de variation en temps a été effectuée avec les objectifs suivants :

- compléter les moyens permettant de caractériser ces structures,
- proposer des estimateurs de l'exposition moyenne adaptés à ces structures,
- en dégager les conséquences en terme de stratégies de mesurage.

Cette étude a été effectuée à l'aide des méthodes de géostatistique, qui ont été préférées aux méthodes statistiques plus courantes en hygiène industrielle pour deux raisons essentielles :

- elles offrent un cadre plus large que les méthodes classiques pour étudier l'autocorrélation des données, avec l'avantage de pouvoir étudier la variabilité en temps des données d'exposition en fonction de sa dépendance à l'égard de la durée d'intégration élémentaire des séries mesurées [12].
- dans le cas d'un mesurage renouvelé périodiquement, elles permettent de proposer des stratégies d'échantillonnage optimisées, basées sur la variabilité des données déjà observées lors d'une première campagne de mesurage.

Compte tenu de la nouveauté de cette approche pour étudier l'exposition sonore, la durée d'observation des données analysées ici a été limitée à une journée. Mais cette limitation pourrait être repoussée afin d'étudier ultérieurement la variabilité inter-journées.

Chapitre 2

Présentation méthodologique

L'analyse de la structure temporelle de données d'exposition sonore demande la prise en compte des particularités liées à l'expression des résultats des mesures en décibels et à la présence éventuelle de pointes d'exposition. Ce chapitre présente les méthodes développées dans ce but et rappelle succinctement la définition du variogramme, qui est l'outil de base de l'analyse géostatistique pour décrire l'autocorrélation et la structure temporelle des données. Il présente les principaux types de variogramme rencontrés dans cette étude ainsi que l'interprétation de leurs caractéristiques dominantes.

2.1 Cadre probabiliste : notion de fonction aléatoire

La théorie des probabilités nous livre un cadre mathématique utile pour analyser les séries de bruit.

Une série de mesures d'exposition sonore évolue dans le temps de manière beaucoup trop irrégulière et imprévisible pour pouvoir être modélisée par une fonction déterministe simple. D'où l'idée de considérer qu'une mesure z en un instant donné est la réalisation d'une variable aléatoire Z . Comme si l'on effectuait un tirage au sort d'une valeur z dans un ensemble de valeurs Z .

Si l'on prend un ensemble de mesures $z(t_\alpha)$ en différents instants t_α ($\alpha = 1, \dots, n$) on considérera que chaque mesure est la réalisation d'une variable aléatoire $Z(t_\alpha)$ correspondante. L'ensemble des valeurs mesurées est donc la réalisation d'un ensemble de variables aléatoires. En statistique classique on fait parfois l'hypothèse que les variables aléatoires sont indépendantes et de même nature (identiquement distribuées). En géostatistique ces variables aléatoires seront en général corrélées dans le temps et l'on se servira des données pour tenter de caractériser cette autocorrélation.

On peut également s'intéresser à tous les instants t dans un intervalle \mathcal{D} . Le bruit $z(t)$ en chaque instant est alors considéré comme la réalisation d'une *fonction aléatoire* $Z(t)$, qui n'est autre qu'une famille infinie de variables aléatoires.

2.2 Variables acoustiques : dB et exposition sonore

La réalité physique du bruit est donnée par les expositions sonores que nous noterons $Z(t)$. L'acousticien a pris l'habitude d'exprimer les mesures en termes de logarithmes des

expositions sonores, puisque la sensibilité de l'oreille est réputée être grossièrement logarithmique. Nous noterons $L(t)$ les logarithmes des expositions sonores, comme c'est l'usage en acoustique industrielle. Celles-ci sont exprimées en décibels (dB) et la convention veut que les décibels soient obtenus à partir des expositions sonores avec des logarithmes de base 10. En statistique, par contre, il est plus courant d'effectuer les calculs avec des logarithmes naturels (notés "log"). Nous introduisons un coefficient $a = \log(10)/10$ qui permet de passer des logarithmes de base 10 aux logarithmes naturels et un coefficient $b = \log(10^9)$ qui ramène les expositions sonores à une échelle numérique de taille réduite (en les divisant par 10^9).

Nous appellerons donc expositions sonores $Z(t)$ les données obtenues par la transformation suivante à partir de mesures exprimées en dB :

$$Z(t) = \frac{1}{10^9} 10^{L(t)/10} = \exp(a L(t) - b) \quad (2.1)$$

Les logarithmes naturels de $Z(t)$ seront désignés par $Y(t)$:

$$Y(t) = \log(Z(t)) \quad (2.2)$$

que l'on peut aisément exprimer en dB par la transformation linéaire :

$$L(t) = \frac{b + Y(t)}{a} \quad (2.3)$$

Les mesures ne sont en général pas instantanées, mais effectuées sur un *support* (temps d'intégration) Δt , de sorte que l'on a souvent affaire à des expositions sonores moyennées sur le support temporel :

$$Z_{\Delta t}(t) = \frac{1}{\Delta t} \int_{t-\Delta t/2}^{t+\Delta t/2} Z(x) dx \quad (2.4)$$

Cette opération étant linéaire, les changements de variables sur support Δt sont identiques à ceux sur support ponctuel.

La norme ISO/DIS 9612.2 [9] définit *l'exposition au bruit quotidienne*, $E_{A,Te}$ et sa relation avec le niveau acoustique continu équivalent, $L_{Aeq,Te}$, quand Te représente la durée quotidienne d'exposition sonore. Si Te est exprimé en heures, $E_{A,e}$ est exprimé en pascals carrés-heures. On a

$$E_{A,e} = p_0^2 \cdot Te \cdot 10^{(L_{Aeq,Te}/10)} \quad (2.5)$$

où $p_0 = 2 \cdot 10^{-5} Pa$.

Quand le temps d'exposition Te est subdivisé en N intervalles de durée élémentaire dT , la norme spécifie que l'exposition au bruit quotidienne s'écrit :

$$E_{A,Te} = \sum E_{A,dT} \quad (2.6)$$

Avec ces relations, on montre que :

$$Z_{dT}(t) = E_{A,dT}(t)/0.4 dT \quad (2.7)$$

$E_{A,dT}(t)$ exprime l'intégrale de la pression quadratique durant l'intervalle dT alors que $Z_{dT}(t)$ exprime, au facteur 0.4 près, la moyenne de cette intégrale.

Pratiquement, on retiendra que $Z_{dT}(t)$, qui s'exprime en pascals carrés, vaut 1 pour un $L_{Aeq,dT}$ de 90 dB(A), 2 pour 93 dB(A), 10 pour 100 dB(A), et ainsi de suite.

2.3 Sélectivité : $T(z)$, $Q(z)$, Gini, U

Examinons “les pointes d'exposition”. Elles incitent à faire le bilan de la partie d'une série temporelle d'expositions sonores dépassant une valeur limite z_c donnée.

On appelle $T(z_c)$ la proportion de valeurs supérieures à une valeur donnée z_c (voir Figure 2.1), c'est-à-dire la probabilité :

$$T(z_c) = P(Z \geq z_c) = 1 - F(z_c) \quad (2.8)$$

La fonction $T(z)$ est donc le complémentaire de la fonction de répartition $F(z)$. Etant donné que $F(z)$ est une fonction croissante, $T(z)$ est une fonction décroissante.

La contribution, dans l'exposition sonore moyenne, des valeurs excédant le seuil z_c donné est représentée par l'intégrale :

$$Q(z_c) = \int_{z_c}^{+\infty} u dF(u) \quad (2.9)$$

Comme $T(z_c)$, la fonction $Q(z_c)$ est décroissante. La valeur de $Q(z_c)$ pour $z=0$ est la contribution totale, c'est-à-dire l'exposition sonore moyenne :

$$Q(0) = \int_0^{+\infty} u dF(u) = E[Z] = m \quad (2.10)$$

où $E[Z]$ symbolise l'espérance de Z .

Notons aussi que le rapport de $Q(z_c)$ sur $T(z_c)$ n'est rien d'autre que la moyenne des valeurs au dessus du seuil d'exposition :

$$m(z_c) = E[Z|Z \geq z_c] = Q(z_c)/T(z_c) \quad (2.11)$$

Un autre paramètre intéressant est la sélectivité. La sélectivité S d'une loi F est définie comme :

$$S = \frac{1}{2} E[|Z - Z'|] \quad (2.12)$$

où Z et Z' sont deux variables indépendantes de même loi F . La sélectivité, dont la définition rappelle celle de la variance :

$$\sigma^2 = \frac{1}{2} E[(Z - Z')^2] \quad (2.13)$$

est donc également un indicateur de dispersion, qui est cependant plus robuste.

Le rapport de la sélectivité à la moyenne est connu depuis longtemps en économétrie sous le nom de *coefficient de Gini* (cf KENDALL & STUART [11], pp47-51) :

$$\text{Gini} = \frac{S}{m} \quad (2.14)$$

et constitue une alternative au coefficient de variation σ/m . Sachant que $S \leq m$, le coefficient de Gini est toujours inférieur à l'unité :

$$0 \leq \text{Gini} \leq 1 \quad (2.15)$$

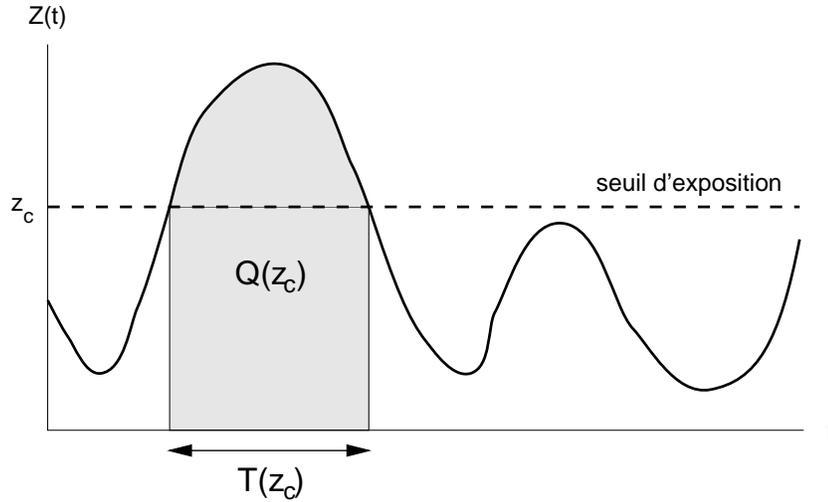


FIGURE 2.1 – Dessin représentant schématiquement une série temporelle de mesures d'exposition sonore dépassant un seuil donné d'exposition. La fonction $T(z_c)$ donne la proportion du temps total pour laquelle les expositions sont au-dessus du seuil. La fonction $Q(z_c)$ exprime la contribution, dans l'exposition sonore moyenne, des valeurs excédant le seuil z_c .

On peut démontrer l'inégalité suivante entre la sélectivité et l'écart-type :

$$S \leq \frac{\sigma}{\sqrt{3}} \quad (2.16)$$

l'égalité n'étant réalisée que dans le cas d'une distribution uniforme (MATHERON [22]). Cela permet de définir un coefficient d'uniformité :

$$U = \frac{\sqrt{3} S}{\sigma} \quad (2.17)$$

avec $0 \leq U \leq 1$.

Du point de vue pratique, on peut présenter les choses en se servant d'une fonction indicatrice exprimant par les deux valeurs 1 ou 0 le résultat du test d'une condition

$$\mathbf{1}_{\text{condition}} = \begin{cases} 1 & \text{si le test de la condition est positif} \\ 0 & \text{s'il est négatif} \end{cases} \quad (2.18)$$

L'indicateur du fait qu'une valeur mesurée Z se trouve, ou non, au-dessus d'un seuil d'exposition z_c s'écrit alors par la fonction indicatrice

$$\mathbf{1}_{Z \geq z_c} = \begin{cases} 1 & \text{si } Z \geq z_c \\ 0 & \text{si } Z < z_c \end{cases} \quad (2.19)$$

Pour une série temporelle, le temps cumulé au-dessus d'un seuil (avec des intervalles de longueur unité) est le nombre de mesures dont les valeurs vérifient $Z > z_c$

$$\sum_{\alpha=1}^n \mathbf{1}_{Z_{\alpha} \geq z_c} \quad (2.20)$$

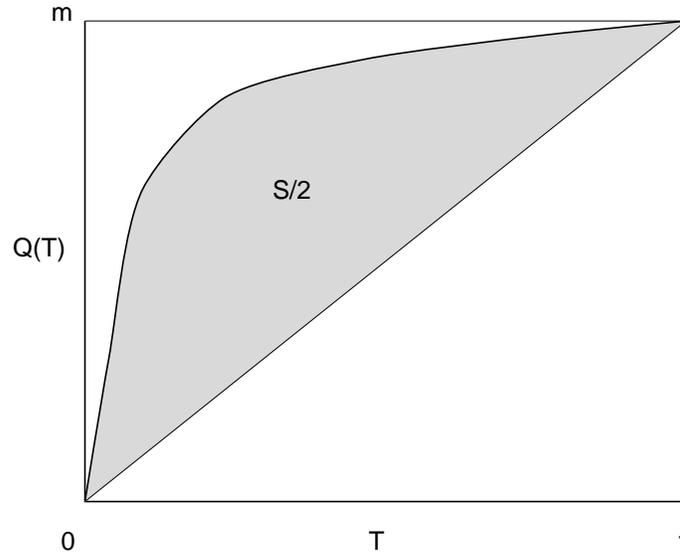


FIGURE 2.2 – Fonction $Q(T)$. La surface teintée, entre la courbe $Q(T)$ et la ligne diagonale, correspond à la moitié de la valeur de la sélectivité S .

En divisant par le temps total (nombre d'intervalles), on obtient la proportion de temps $T(z_c)$ au-dessus du seuil z_c :

$$T(z_c) = \frac{1}{n} \sum_{\alpha=1}^n \mathbf{1}_{Z_\alpha \geq z_c} \quad (2.21)$$

L'acousticien s'intéresse à la quantité Q d'énergie accumulée au-dessus d'un seuil z_c d'exposition sonore. Il est judicieux de diviser cette énergie par le nombre n d'échantillons et d'exprimer ainsi par $Q(z_c)$ l'exposition sonore moyenne au-dessus d'un seuil donné :

$$Q(z_c) = \frac{1}{n} \sum_{\alpha=1}^n Z_\alpha \cdot \mathbf{1}_{Z_\alpha \geq z_c} \quad (2.22)$$

Les fonctions $T(z_c)$ et $Q(z_c)$ permettent de faire le bilan des dépassements de seuil pour une série de mesures d'exposition sonores comme cela est montré sur la Figure 2.1.

Les deux courbes $T(z_c)$ et $Q(z_c)$ peuvent être représentées sur un même graphique en normant la courbe $Q(z_c)$ par sa valeur à l'origine, qui est la moyenne m .

Un autre graphique consiste à dessiner la courbe de Q (normé par m) en fonction de T . Cette courbe est concave. Une courbe de $Q(T)$ est montrée sur la Figure 2.2. L'aire (teintée en gris) en dessous de la courbe $Q(T)$ et au-dessus de la diagonale mT (entre l'origine et le point $Q(1) = m$) vaut la moitié du coefficient de sélectivité S .

2.4 Mesure de l'autocorrélation : le variogramme

Soit h le vecteur séparant deux instants de mesure à l'intérieur du domaine temporel \mathcal{D} que l'on étudie.

On peut alors comparer les valeurs z de deux mesures en fonction de leur éloignement h dans le temps :

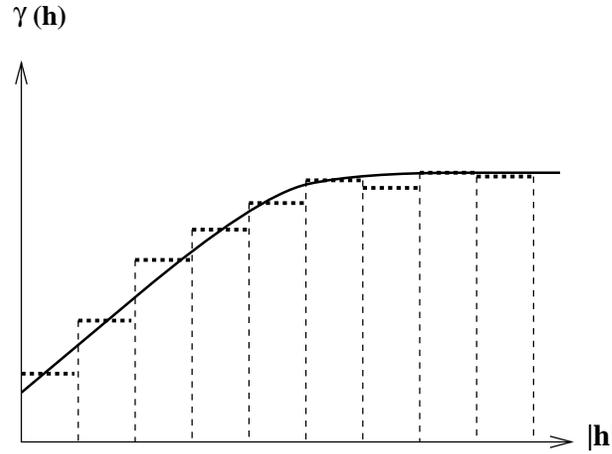


FIGURE 2.3 – Un modèle de variogramme théorique est ajusté au variogramme expérimental.

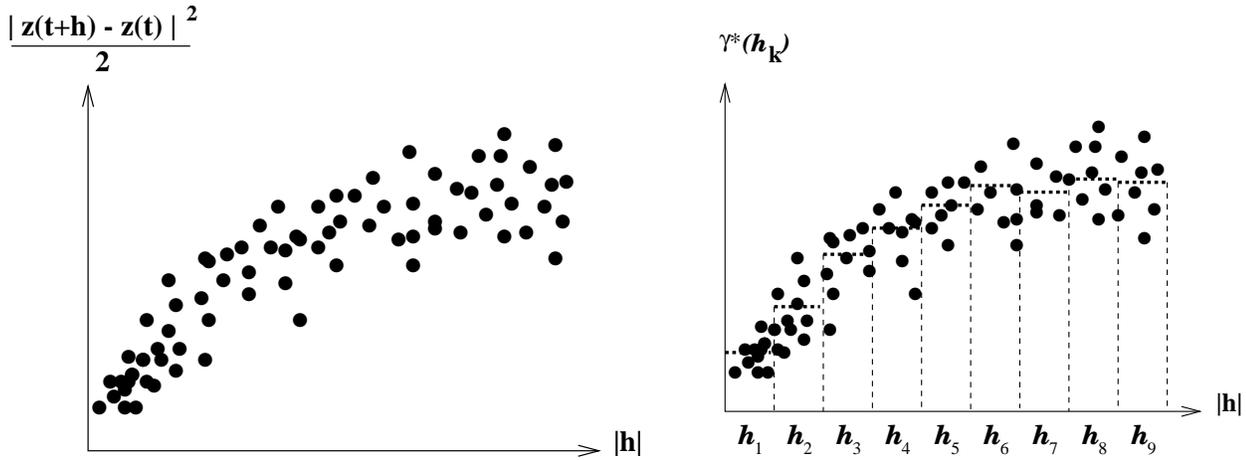


FIGURE 2.4 – A gauche : Nuée des écarts (quadratiques) des mesures, en fonction de leur distance $|h|$ dans le temps. A droite : La nuée variographique est subdivisée à l'aide de classes de distance \mathbf{h}_k . Des moyennes $\gamma^*(\mathbf{h}_k)$ sont calculées pour chaque classe, ce qui donne le variogramme expérimental (en pointillés).

$$\frac{(z(t+h) - z(t))^2}{2} \quad (2.23)$$

En reportant ces demi-écarts quadratiques pour toutes les paires de points de mesure (jusqu'à une distance inférieure à la moitié du diamètre du domaine \mathcal{D}), sur un diagramme les opposant aux écarts h dans le temps, on obtient un nuage de points appelé *nuée variographique*. La Figure 2.4 (gauche) en montre un exemple : il est commun que les écarts soient plus faibles près de l'origine, ce qui signifie que des valeurs à courte distance ont tendance à se ressembler.

Dans une seconde étape (si ce n'est d'emblée) on forme des moyennes γ^* pour des classes de distance \mathbf{h}_k , tel que le suggère la Figure 2.4 (droite), et l'on appelle *variogramme expérimental* l'ensemble des valeurs $\gamma(\mathbf{h}_k)$ pour les différentes classes de distance.

L'interprétation du variogramme expérimental dans le cadre d'un modèle de fonction

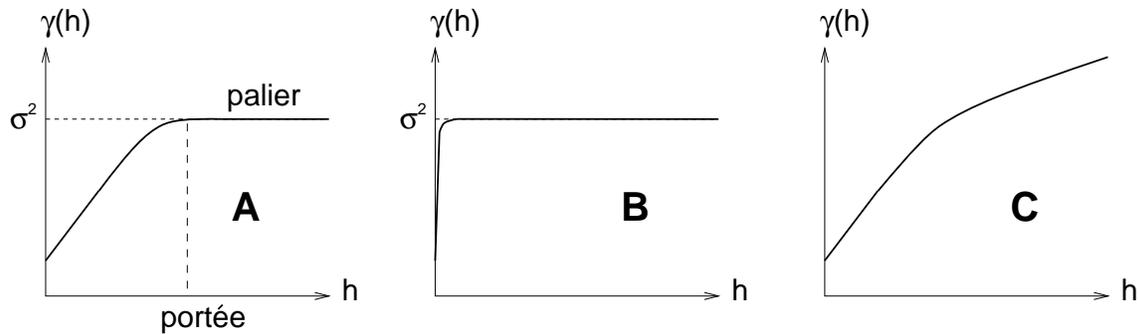


FIGURE 2.5 – Différents comportements aux grandes distances de l'origine.

aléatoire se fait en ajustant un modèle théorique $\gamma(h)$ aux valeurs expérimentales $\gamma(\mathbf{h}_k)$, comme on peut le voir sur la Figure 2.3.

Des modèles courants sont les schémas *exponentiel*, *sphérique* ou à *effet de pépîte*, ainsi que des combinaisons de ceux-ci. Dans le contexte d'un modèle lognormal il peut être question d'un schéma *de Wijsien*.

2.5 Variogrammes de divers types : interprétation

Deux questions sont importantes dans le choix d'un modèle de variogramme : le comportement au voisinage de l'origine et le comportement aux grandes distances.

A l'origine le variogramme vaut zéro par définition. Au voisinage de l'origine trois types de comportements peuvent être distingués :

- le variogramme est continu et dérivable à l'origine ; cela correspond à une variable régionalisée lisse.
- le variogramme est continu, mais non dérivable ; la variable régionalisée est rugueuse.
- le variogramme est discontinu à l'origine ; la variable régionalisée présente des discontinuités.

Aux grandes distances de l'origine le variogramme peut croître à l'infini ou être borné. La Figure 2.5 illustre quelques comportements typiques. Le variogramme A est borné et atteint un palier à une distance appelée la portée du variogramme. Le palier peut correspondre à la valeur de la variance σ^2 . Le variogramme B correspond à un phénomène avec une autocorrélation temporelle de très courte portée. Le variogramme C n'atteint pas de palier à l'intérieur de la fenêtre d'observation ; deux hypothèses sont possibles dans ce genre de situation :

- soit on considère que le variogramme C atteint un palier à des distances non accessibles expérimentalement ; il faut alors fixer relativement arbitrairement les valeurs du palier et de la portée correspondantes.
- soit on juge qu'il n'y a pas de limite dans la croissance du variogramme C et l'on adopte un modèle sans palier.

La somme de variogrammes élémentaires est à nouveau un variogramme : on est en droit d'additionner des modèles de variogramme et le résultat s'appelle un *variogramme gigogne*. Le variogramme C de la Figure 2.5 en est un exemple : on peut considérer qu'il est composé de 3 structures élémentaires, à savoir

- une discontinuité à l'origine, qui s'appelle *effet de pépîte* en géostatistique ou *bruit blanc* en électronique, et qui correspond expérimentalement, soit à de la variabilité

- à très petite échelle (inférieure à l'espacement des mesures dans le temps), soit à une erreur de mesure liée à l'appareillage de mesure,
- une structure de courte portée (comparable à celle du variogramme A de la Figure 2.5), qui peut par exemple refléter la durée d'un cycle de travail,
- une structure aux grandes distances, qui représente la variabilité sur le long terme.

On privilégie l'utilisation du **variogramme plutôt que la fonction de covariance**, plus classique, pour la bonne raison qu'avec le variogramme on est capable de décrire une classe de phénomènes plus large que celle des phénomènes pour lesquels la fonction de covariance est adéquate. Comme il y a inclusion d'une classe de phénomènes dans l'autre, on peut toujours construire un variogramme à partir d'une fonction de covariance $C(h)$ en la soustrayant de sa valeur à l'origine, ce qui donne la formule

$$\gamma(h) = C(0) - C(h) \quad (2.24)$$

A l'inverse, il n'existe pas toujours une fonction de covariance pour un variogramme donné. C'est seulement vrai lorsque le variogramme est borné.

2.6 Applications pratiques

Dans le domaine de l'acoustique, l'usage des décibels s'est imposé avec le temps, mais reste délicat car le décibel exprime un rapport entre deux grandeurs. Dans le mesurage du bruit reçu durant le travail, l'utilisation de la variable *exposition sonore* est récent. Aussi les données mesurées analysées ici seront exprimées en employant les deux variables : l'exposition sonore, en pascals carrés et le niveau en décibels.

L'application d'indices de sélectivité à l'acoustique permet de quantifier la contribution dans l'exposition sonore globale des valeurs supérieures à un seuil spécifié. De tels indices sont donc particulièrement utiles si la valeur limite d'exposition est indiquée par une probabilité de dépassement d'un seuil, ce qui n'est pas le cas en France actuellement. Toutefois, on verra dans le Chapitre 3 l'intérêt de ces indices pour définir une typologie des situations d'exposition.

Trois types schématiques de variogrammes ont été présentés. Leur intérêt pratique apparaît notamment quand on examine le variogramme de type C, non borné dans la fenêtre d'observation employée. Dans cette situation, alors que la fonction de covariance donnerait une fausse idée de l'autocorrélation, le variogramme indique la présence d'une structure temporelle sur une distance à *long terme*, supérieure à celle de la fenêtre d'observation. Dans le domaine de l'exposition au bruit professionnel, il est peu vraisemblable de concevoir que la variance des données ne soit pas bornée. En d'autres termes, le variogramme du processus que l'on veut estimer doit atteindre un palier, après un temps suffisant. Si l'observation aboutit à un variogramme de type C, ceci signifie que l'observation est insuffisante pour estimer l'exposition à long terme.

Pratiquement comment faire en présence d'un variogramme de type C, non borné dans l'intervalle observé T ? - S'il est possible de stratifier l'intervalle T en plusieurs phases spécifiques d'exposition, on calculera les variogrammes de chaque phase. Si chacun d'eux est borné (de type A ou B), on considérera comme acceptable la description de la structure de chaque phase (un exemple de ce type sera traité au Chapitre 7). Sinon, il est souhaitable de reprendre

le mesurage et de le prolonger au-delà de la fenêtre d'observation initiale T , car l'exposition comprend des cycles dont la durée est supérieure à T . En accroissant suffisamment la durée de mesurage, on visera à obtenir un variogramme borné, de type A.

La relation qui apparaît entre l'interprétation du type de variogramme, le plan de mesurage et l'analyse des situations d'expositions permet d'esquisser une méthode de mesurage nouvelle, adaptée à l'évaluation de l'exposition sonore à long terme. Ce point sera précisé notamment dans les Chapitres 5, 6 et 7.

Chapitre 3

Description de l'exposition sonore

Les données d'exposition sonore sont constituées de séries chronologiques de $L_{Aeq,dT}$ *court*, présentées ici avec une durée d'intégration élémentaire dT de 10s. Elles ont été collectées à l'aide d'exposimètres de marque Aclan SIE 95, appareils à mémoire portés par les travailleurs durant plusieurs heures et dotés d'un microphone de mesure placé près de l'oreille. Dans les séries de données analysées ici, l'appareil de mesure a été porté durant un demi poste ou un poste de travail. Seule la variabilité *intra-journée* fait l'objet de cette étude.

Les situations d'exposition choisies pour cette étude sont homogènes. "Homogène" signifie qu'il n'était pas possible d'identifier *a priori*, sur la base de l'analyse du travail, des périodes d'activités spécifiques et de durée suffisante pour envisager une stratification du temps d'exposition en plusieurs phases homogènes. Le cas de l'exposition à des phases multiples est traité au Chapitre 7. Les données décrites dans ce chapitre sont analysées et modélisées dans les Chapitres 4, 5 et 6.

3.1 Description des postes de travail observés

Trois séries de données d'exposition ont été collectées à des postes de travail dans lesquels l'exposition est considérée comme homogène, non stratifiable a priori en plusieurs phases. A ces postes, les variations d'exposition sonore dépendent notamment des facteurs suivants : la localisation et la nature des tâches effectuées, la nécessité pour l'opérateur d'intervenir épisodiquement dans une machine bruyante en fonctionnement, la possibilité d'être exposé à des événements acoustiques épisodiques.

Le poste "jointag" : Ce poste de travail est fixe. L'opérateur pilote et contrôle le fonctionnement automatisé d'une petite presse qui réalise l'emboîtement d'éléments en plastique. Le système d'alimentation de la presse est la source de bruit principale. Il comprend un bol vibrant et un système d'avance des pièces par jet d'air comprimé. Les tâches de l'opérateur comprennent l'approvisionnement du dispositif d'alimentation, le contrôle de fabrication, des interventions manuelles dans le pot vibrant en cas de bourrage, le nettoyage épisodique de la machine par soufflette. Le poste de travail est situé à proximité immédiate du système d'alimentation.

Le poste "regleur" : Dans un atelier sont regroupées une douzaine de presses à injection, utilisées selon les besoins pour la production de diverses petites pièces de

plastique. Leur fonctionnement en mode normal est automatique. Le régleur est un opérateur qui est mobile dans tout l'atelier, avec une présence plus fréquente auprès des vieilles injecteuses. Ses tâches sont très variées : changements des moules ; réglages sur toutes les presses à injecter ; démarrage de production ; interventions en cas d'incidents ; mise en place et contrôle des dispositifs d'alimentation ; contrôle de l'évacuation des produits et pesée des cartons pleins ; contrôles de production ; nettoyage de moules. Des interventions sur machine en fonctionnement sont possibles mais peu fréquentes.

Le poste "saranex" : Dans un atelier fonctionnent plusieurs lignes d'extrusion qui produisent en continu des bandes de quelques mm d'épaisseur, enroulées en bobine en sortie de fabrication. Les tâches de l'opérateur comprennent le pilotage d'une ligne, le réglage de la machine, le contrôle de la fabrication et l'évacuation des bobines pleines. Ce poste est situé dans une zone d'étendue limitée, le long de la ligne d'extrusion (environ 10 m). De plus les trois opérateurs de l'atelier se répartissent une tâche épisodique plus bruyante : l'alimentation d'un broyeur de déchets, situé dans le local voisin. Cette opération dure environ 2 mn et se renouvelle 3 ou 4 fois par poste, selon les besoins.

3.2 Description acoustique classique en dB

Les trois séries de mesures exprimées en dB sont montrées sur la Figure 3.1. Elles ont toutes les trois été effectuées l'après-midi. Les séries sont d'une durée inégale : la série jointag1 couvre 2h41mn, tandis que regleur2 et saranex1 ont été mesurées pendant respectivement 5h59mn et 5h36mn.

Les histogrammes des trois séries, calculés en utilisant des classes de largeur 0.5 dB, sont visibles sur la Figure 3.2. Ils sont raisonnablement symétriques et ne comportent pas de multimodalités flagrantes.

Le Tableau 3.1 donne un résumé de la description statistique des trois séries acoustiques : le L_{EQ} , l'écart-type s , le nombre de valeurs N et les quantiles principaux (5%, 10%, 50%, 90%). Les L_{EQ} sont tous trois supérieurs à la médiane (L50) des histogrammes.

Série	L_{EQ}	s	N	L5	L10	L50	L90
jointag1	89.2	3.1	966	93.8	92.4	87.6	84.6
regleur2	86.9	4.9	2153	92.6	89.8	82.8	79.2
saranex1	90.1	6.3	2017	95.8	92.3	86.0	78.4

TABLE 3.1 – Description acoustique résumée des trois séries.

Note : La pondération fréquentielle "A" a été employée lors des mesurages d'exposition sonore et l'unité de tous les résultats est le $dB(A)$. Ici, la mention d'un résultat *en dB* n'est qu'un raccourci, utilisé par opposition à un résultat *en exposition sonore*.

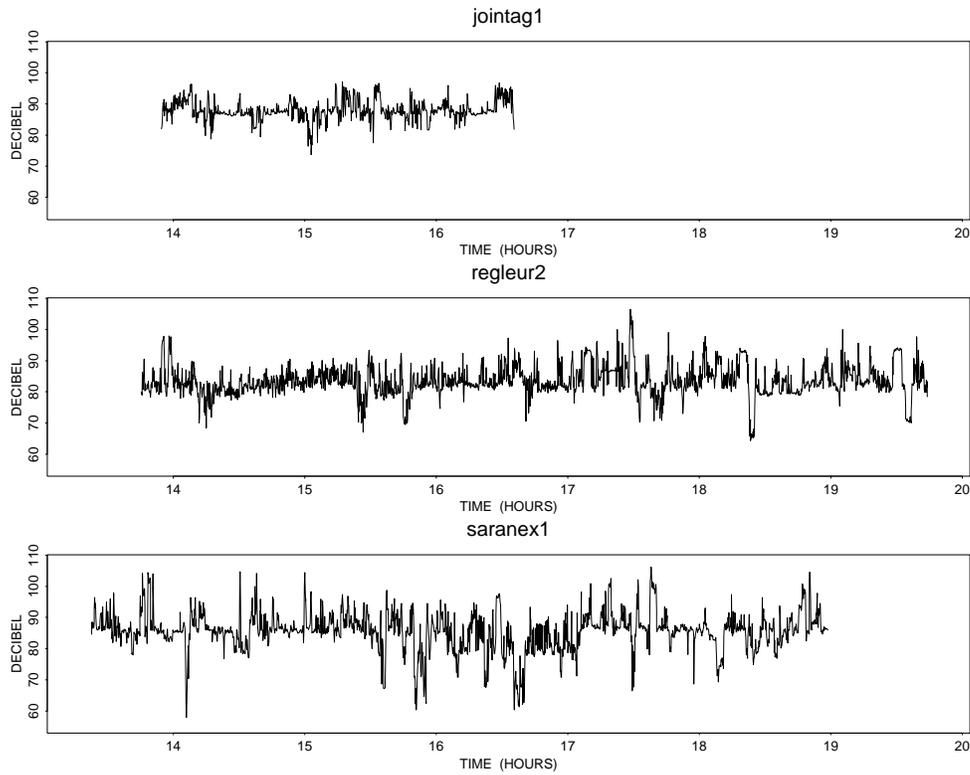


FIGURE 3.1 – Les trois séries en dB.

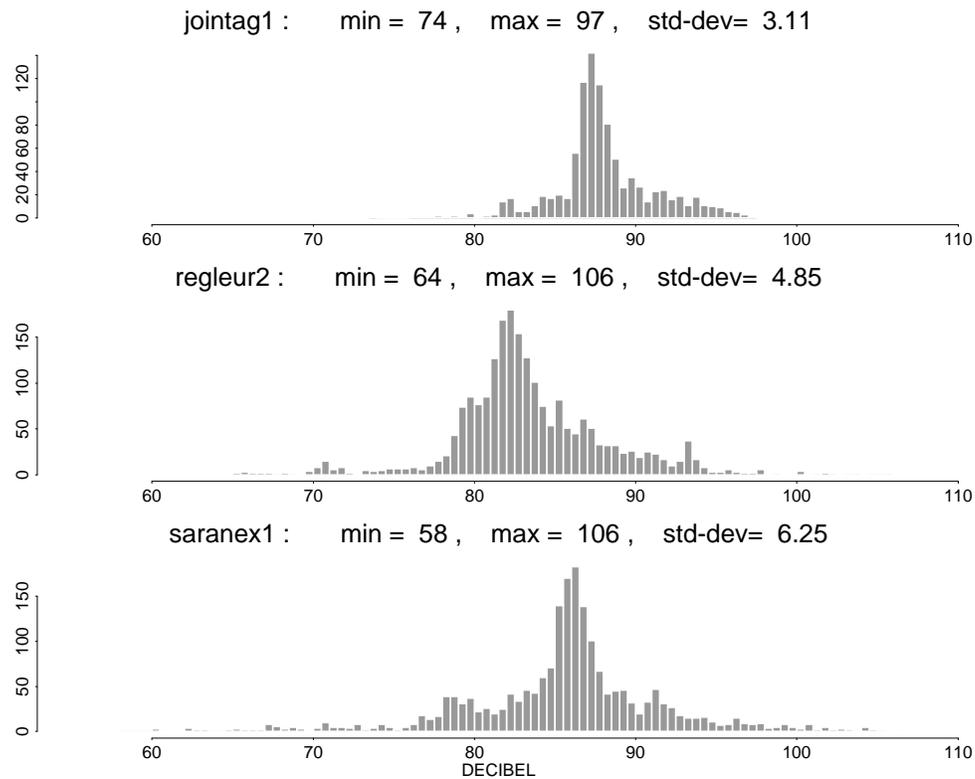


FIGURE 3.2 – Les histogrammes en dB.

3.3 Transformation en exposition sonore

Les séries ont également été transformées en expositions sonores $Z(t)$ selon la formule (2.1) ; elles sont montrées sur la Figure 3.3. On voit d'un coup d'oeil à quels moments de la journée ont eu lieu les expositions maximales. On note que les expositions sont plus fortes sur les deux dernières séries que sur la première, et que les fortes expositions sont les plus nombreuses sur `saranex1`.

Les histogrammes correspondants sont visibles sur la Figure 3.4. Vu la disymétrie extrême de la répartition des expositions sonores, ils ne sont pas très parlants. Pour chaque série 25 classes ont été utilisées : pour `jointag1` le maximum est de 5 seulement, contre 44 pour `regleur2` et 41 pour `saranex1` ; de ce fait, les classes de `jointag1` sont bien plus étroites en utilisant la même subdivision de l'abscisse pour les représenter.

Nous n'utiliserons plus par la suite des histogrammes pour représenter des expositions sonores.

3.4 Courbes de répartition $T(z)$ et $Q(z)$

Les diagrammes des courbes de répartition $T(z)$ et $Q(z)$ sont montrés sur la Figure 3.5. Ces courbes indiquent la contribution des valeurs excédant un seuil z dans le temps d'exposition pour $T(z)$ et dans l'exposition sonore moyenne pour $Q(z)$. Sur ces diagrammes le seuil z est exprimé en dB sur l'abscisse et $Q(z)$ est normé par $Q(0)$. De ce fait, les courbes Q et T évoluent entre 0 et 1 et peuvent être superposées. De plus, l'équation 2.11 (Chapitre 2, p 5) montre que dans tous les cas, la courbe Q normée (représentée en trait gras sur la figure) est située au-dessus de la courbe T (représentée en trait fin sur la figure).

Considérons le seuil d'exposition $L_c = 95\text{dB}$ représenté en tiretés sur les graphiques. On voit, par exemple, avec la courbe $T(z)$ que pour `jointag1` seule une faible proportion des valeurs est au dessus du seuil 95dB. La courbe $Q(z)$ montre que les valeurs au dessus de 95dB totalisent moins de 10% de l'énergie acoustique totale de `jointag1`. Pour `regleur2`, la courbe $T(z)$ montre que le seuil 95dB est dépassé pour une très faible proportion du temps total, tandis que la courbe $Q(z)$ montre que les valeurs correspondantes totalisent un quart de l'énergie contenue dans la série. En ce qui concerne `saranex1`, si les valeurs au dessus de 95dB ne représentent que moins d'un dixième de la série, elles accumulent cependant plus de la moitié de l'énergie acoustique de cette série !

3.5 Courbes de sélectivité $Q(T)$

Les courbes $Q(T)$, qui pour chaque valeur de z représentent Q en fonction de T , sont données sur la Figure 3.6. Sur ces graphiques, la courbe $Q(T)$ est dessinée en trait gras et la bissectrice en tiretés.

La fonction $T(z)$ est décroissante : T diminue lorsque z augmente. $Q(T)$ est croissante : lorsque la proportion de temps T diminue, la proportion d'énergie Q diminue également. Une courbe $Q(T)$ montre à quelle vitesse cette diminution s'effectue. L'intérêt de la courbe $Q(T)$ est donc de visualiser cet effet : plus il est important, plus la courbe $Q(T)$ s'éloigne par rapport à la bissectrice.

La surface entre la courbe $Q(T)$ et la bissectrice représente la moitié du coefficient de sélectivité S pour une fonction $Q(T)$ non normée par la moyenne. Dans notre cas, `jointag1`

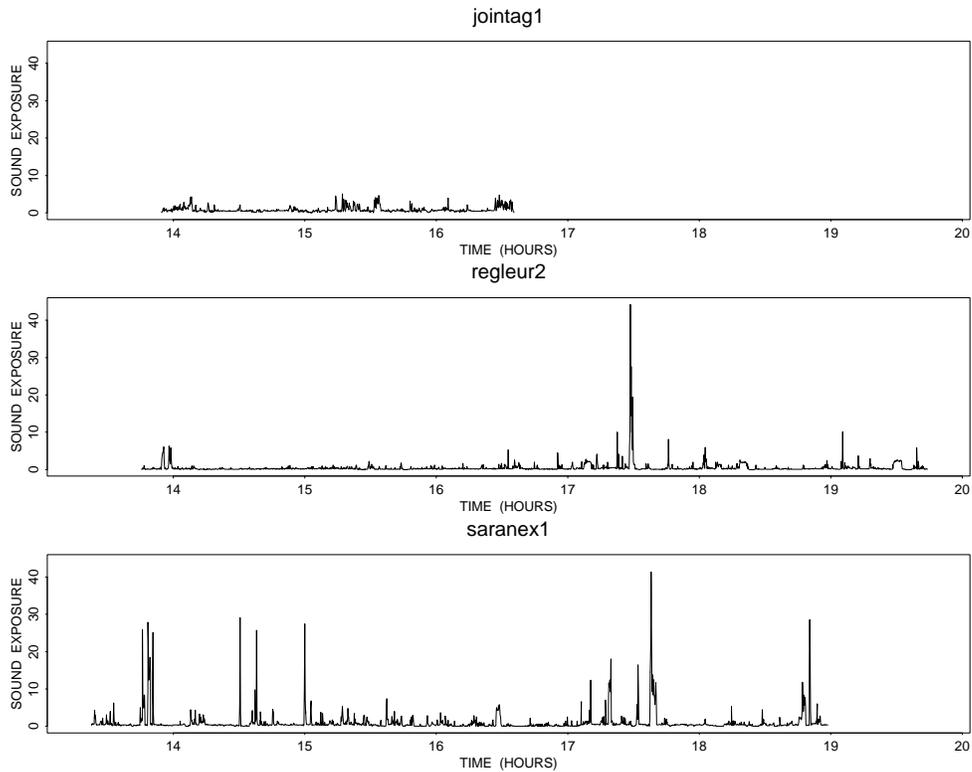
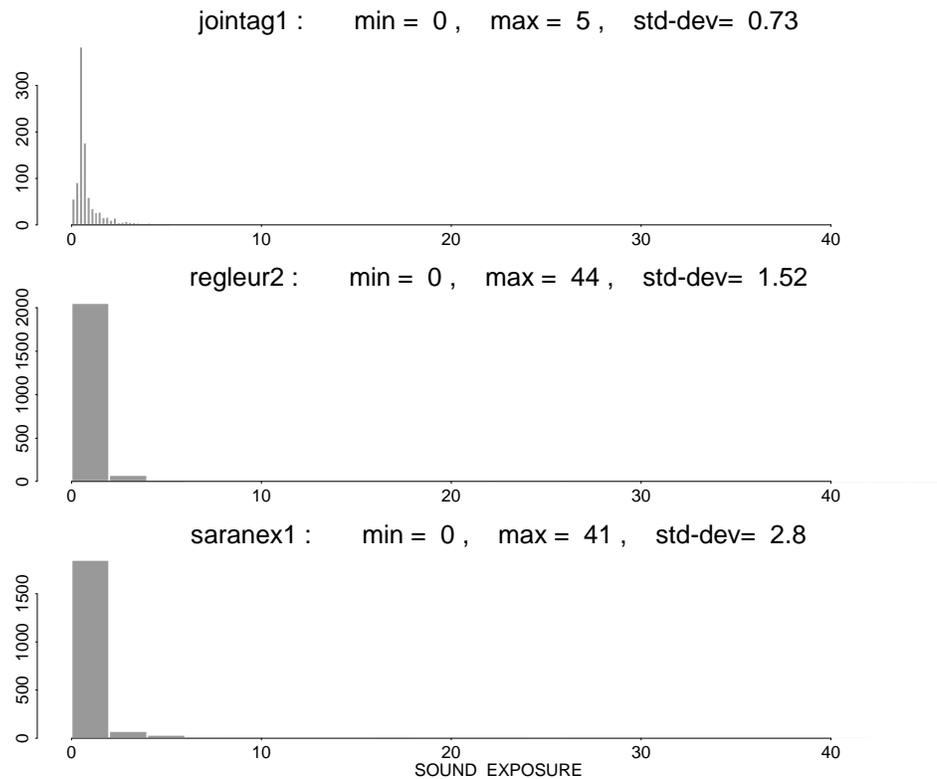
FIGURE 3.3 – Les trois séries transformées en expositions sonores $Z(t)$.

FIGURE 3.4 – Les histogrammes des expositions sonores.

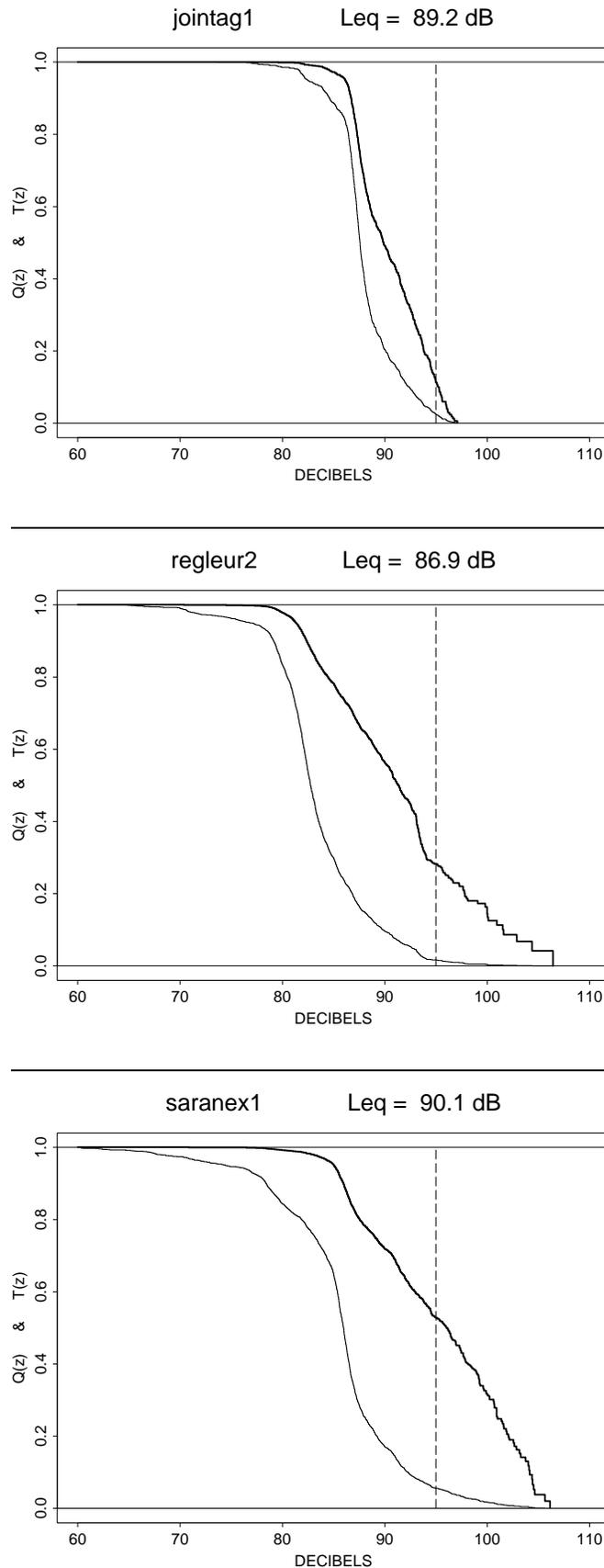


FIGURE 3.5 – Courbes $Q(z)$ (en trait gras) et $T(z)$ (en trait fin) des trois séries. Un seuil d'exposition de 95dB est dessiné en tiretés.

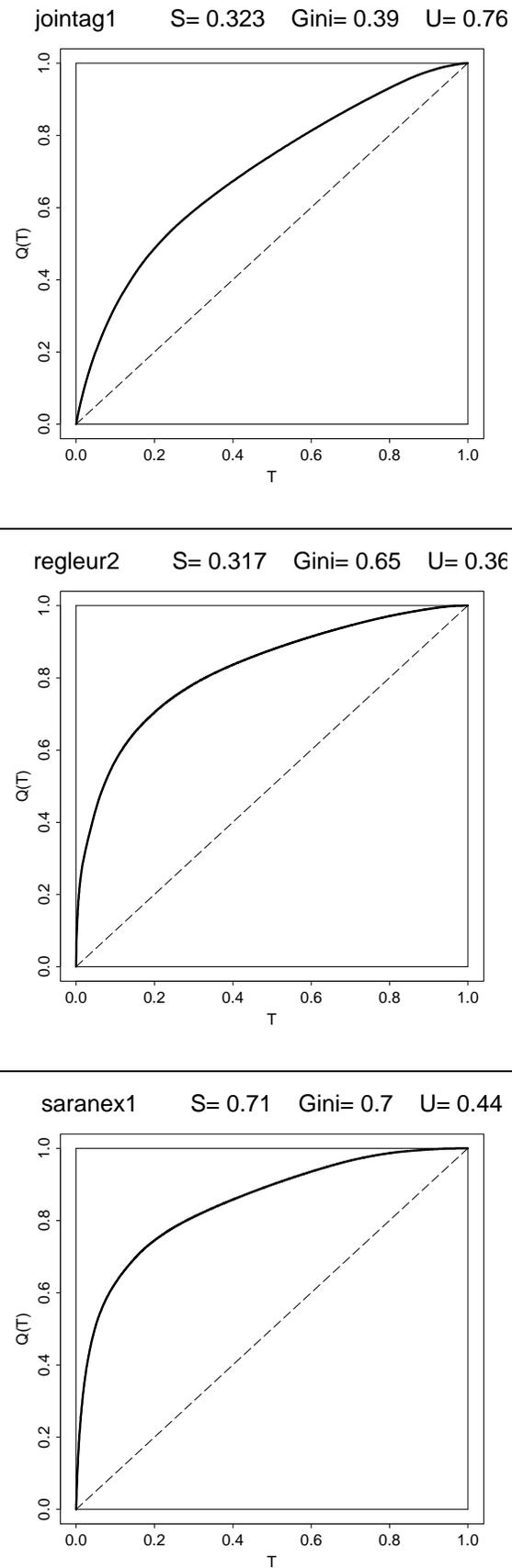


FIGURE 3.6 – Courbes de sélectivité $Q(T)$ (en trait gras) des trois séries. La bissectrice est dessinée en tiretés.

et `regleur2` ont un coefficient S du même ordre de grandeur, tandis que celui `saranex1` vaut le double. Lorsque l'on norme par la moyenne, le coefficient de sélectivité devient le coefficient de Gini, qui décrit la proportion de surface entre $Q(T)$ et la bissectrice, par rapport à celle du triangle au dessus de la bissectrice. Dans notre cas, le coefficient de Gini pour `jointag1` est inférieur à celui de `regleur2` (vu que la moyenne des expositions sonores de la première série est supérieure à celle de la seconde). Le coefficient de Gini de `regleur2` est du même ordre que celui de `saranex1`.

Le coefficient d'uniformité U est le rapport de la sélectivité S à l'écart type σ . Ces derniers étant tous les deux des indicateurs de dispersion, il s'agit du rapport de deux mesures de la dispersion. S est basé sur la norme de la valeur absolue, tandis que σ utilise la norme euclidienne, cf les formules (2.12) et (2.13). Le premier étant par ce fait plus robuste à la présence de valeurs extrêmes que le second, le coefficient d'uniformité indique à quel point on est proche d'une loi uniforme ou gaussienne (pour lesquelles il est égal à l'unité, ou proche de l'unité). Dans notre cas, la série `jointag1` a un coefficient U plus proche de l'unité que `saranex1` ou `regleur2`.

3.6 Applications pratiques

Examinons tout d'abord les valeurs exprimées en décibels. Les séries mesurées et leurs histogrammes font apparaître des variations du niveau sonore dont l'amplitude peut être quantifiée à l'aide de différents indicateurs : les écarts types (compris entre 3.1 dB et 6.3 dB), les écarts entre percentiles (L10–L90 est compris entre 7.8 dB et 13.9 dB), l'étendue des distributions (Lmax–Lmin est compris entre 23 dB et 48 dB). Les 3 histogrammes sont d'allure relativement symétrique.

L'insuffisance de cette représentation vient du fait que l'histogramme des valeurs en dB ne fournit pas d'indication sur la contribution des valeurs dans l'exposition sonore elle-même. De ce fait, elle incite à considérer de la même façon des valeurs qui ont des poids très différents en terme d'exposition sonore. La transformation en exposition sonore des valeurs mesurées offre un moyen simple de bien faire ressortir les pointes d'exposition, mais elle fait disparaître les périodes de pause.

La distribution cumulée est employée depuis longtemps en acoustique, par exemple pour le calcul des percentiles des valeurs exprimées en dB. La courbe de répartition $T(z)$, avec une abscisse en dB, représente exactement cette distribution cumulée. La nouveauté introduite ici consiste à représenter sur le même graphique que $T(z)$ la variable $Q(z)$ normée, qui donne la contribution dans l'exposition sonore moyenne des valeurs excédant un seuil spécifié.

L'intérêt de la représentation simultanée de $T(z)$ et $Q(z)$ a été illustré ici avec le seuil de 95 dB. Pour les 3 séries, les valeurs qui excèdent 95 dB représentent respectivement environ 3%, 2% et 6% du temps d'observation, alors que leur contribution dans l'exposition sonore globale est d'environ 10%, 25% et 55%. L'examen (même très minutieux) de l'histogramme de la série `saranex1` ne permet pas d'entrevoir que les quelques valeurs repérées comme excédant le seuil de 95 dB peuvent expliquer 55% de l'exposition sonore à ce poste, que seule la représentation $T(z)$ – $Q(z)$ fait apparaître.

Cette représentation nouvelle paraît donc particulièrement indispensable

dès que les distributions ont une étendue qui atteint ou dépasse 20 dB. De plus, les courbes de sélectivité $Q(T)$ prolongent cette représentation en fournissant deux indicateurs numériques, les coefficients d'uniformité et de sélectivité, U et Gini. Ils diffèrent sensiblement dans les 3 séries étudiées et pourront donc être utilisés pour quantifier la typologie des distributions des mesures d'exposition selon l'importance de l'effet des pointes d'exposition.

Les distributions et les courbes $T(z)$, $Q(z)$ et $Q(T)$ dépendent du nombre et de l'amplitude des pointes de bruit, mais restent invariantes si seule la répartition des pointes est modifiée. Or la structure temporelle des données dépend aussi, bien évidemment, de cette répartition des pointes dans la période observée. Pour étudier cette répartition, un autre outil est nécessaire. La géostatistique utilise dans ce but des variogrammes. Ce point sera abordé aux Chapitres 4 et 5.

Chapitre 4

Influence du temps d'intégration

Quand les niveaux sonores varient dans le temps, les séries de valeurs mesurées par intégration de $L_{Aeq,dT}$ *court* dépendent de la durée d'intégration élémentaire dT employée. Compte tenu des nombreux choix possibles pour ce paramètre dT dans la pratique du mesurage du bruit, l'impact du choix du temps d'intégration est étudié ici tout d'abord sur les descripteurs employés au Chapitre 3 (séries chronologiques, histogrammes, indicateurs de sélectivité), puis sur les variogrammes expérimentaux.

L'effet de dépendance entre la structure temporelle des données et la durée élémentaire des échantillons est classique en géostatistique et s'appelle *effet de support*. Le support dont il sera question ici est donc la durée d'intégration élémentaire dT employée pour calculer les séries de $L_{Aeq,dT}$ *court*.

4.1 Les trois séries intégrées sur 10s, 1mn et 5mn

Les séries avec temps d'intégration de 10s, présentées au chapitre précédent, ont été recomposées sur un support de 1 minute et de 5 minutes. Elles sont représentées pour chaque support à la fois en exposition sonore et en décibels sur les Figures 4.1 à 4.6. A chaque fois on a utilisé les mêmes échelles sur l'abscisse et l'ordonnée, afin de bien faire apparaître la réduction de la variabilité et de la résolution temporelle lorsque le support augmente.

4.2 Histogrammes pour trois supports

Les histogrammes des séries exprimées en décibels sont montrés pour les supports 10s, 1mn et 5mn sur les Figures 4.7 à 4.9. On a utilisé les mêmes unités sur les abscisses pour les trois séries et une largeur de classe de 1dB. Avec l'augmentation du support, on constate le resserrement des histogrammes, la réduction des valeurs extrêmes (minimum, maximum) et de l'écart-type s . Seule augmente la moyenne m des décibels (m , moyenne arithmétique qui n'est évidemment pas équivalente au L_{EQ}), et elle tend à s'approcher du L_{EQ} quand le support croît.

4.3 Courbes $Q(z)$ et $Q(T)$ pour trois supports

Les courbes $Q(z)$ pour les trois supports sont montrées sur les Figures 4.10, 4.12 et 4.14. Elles mettent bien en évidence l'effet d'une augmentation du support sur les fortes valeurs, qui représentent l'essentiel de l'énergie contenue dans une série. L'abscisse des trois graphiques commence à chaque fois à 80dB et une ligne en tiretés a été tracée à 95dB, afin de faciliter les comparaisons entre les séries.

Les courbes $Q(T)$ sont représentées sur les Figures 4.11, 4.13 et 4.15. On remarquera qu'elles sont bombées de manière différente d'une série à l'autre, et que l'écart par rapport à la bisectrice ne diminue pas de la même manière lorsque le support augmente.

4.4 Variogrammes expérimentaux pour trois supports

Des variogrammes expérimentaux ont été calculés pour les trois séries, à la fois en exposition sonore et sur leurs logarithmes naturels. Ils ont représentés sur les Figures 4.16 à 4.18. Sur chaque graphique, on voit trois lignes horizontales en tiretés, qui représentent les variances, qui diminuent avec l'augmentation du support.

Pour *jointag1* (Figure 4.16), les variogrammes des expositions sonores diffèrent peu des variogrammes calculés sur les logarithmes (même l'échelle sur les ordonnées est d'un ordre de grandeur analogue, ce qui est tout-à-fait fortuit).

Pour *regleur2* (Figure 4.17), les variogrammes des logarithmes croissent rapidement depuis l'origine (sauf au support 5mn) et oscillent ensuite autour de la valeur de la variance. Les variogrammes des expositions sonores croissent au-delà de la valeur de la variance et subissent un décrochement (brutal pour les supports 10s et 1mn) au bout de 2h30. Ce décrochement correspond au pic d'exposition sonore vers 17h30 dans la série *regleur2* et sera discuté au Chapitre 5, p55.

Pour *saranex1* (Figure 4.18), on constate que les variogrammes des logarithmes croissent nettement au dessus de la valeur de la variance pour les supports de 10s et 1mn, ce qui n'est plus vrai sur un support de 5mn. Les variogrammes des expositions sonores sont relativement chahutés, surtout au support 10s, à cause de la présence de pics d'exposition multiples dans la série *saranex1*.

Au support de 5mn, le variogramme des expositions sonores de *saranex1* ne montre pratiquement pas de structure au voisinage de l'origine et oscille faiblement autour de la valeur de la variance. Si l'on ne considérait que ce variogramme, on pourrait croire à un phénomène sans autocorrélation. Le variogramme correspondant des logarithmes n'ammène cependant pas la même conclusion.

4.5 Applications pratiques

Pour les 3 séries de $L_{Aeq,dT}$ *court* analysées, le passage d'un temps d'intégration de 10s à 1mn et 5mn modifie toutes les représentations employées ici et provoque un effet de lissage des données.

L'hypothèse d'une distribution gaussienne n'est vérifiée dans aucun des cas étudiés et plusieurs histogrammes présentent des multimodalités. L'accroissement de la durée d'intégration modifie aussi les courbes de répartition cumulée de l'exposition sonore $Q(z)$ et de sélectivité $Q(T)$. Un examen minutieux des courbes $Q(T)$ montre que l'effet de support est différent d'une série à l'autre.

Sur les variogrammes expérimentaux, l'accroissement de la durée d'intégration se traduit tout d'abord par une réduction de la variance expérimentale, très apparente dans tous les cas étudiés. Entre 10s et 1mn, la forme générale des variogrammes est conservée, mais ce n'est plus le cas quand on passe au support de 5mn. Les variogrammes expérimentaux montrent que les données étudiées sont autocorrélées et que l'autocorrélation dépend du support ou durée d'intégration des données.

Ces différents constats sont liés entre eux par la structure temporelle des données. Pour prendre en compte et quantifier ces effets dans l'estimation du niveau d'exposition sonore quotidien, l'analyse géostatistique demande une étape supplémentaire, objet du Chapitre 5 : la modélisation des variogrammes expérimentaux.

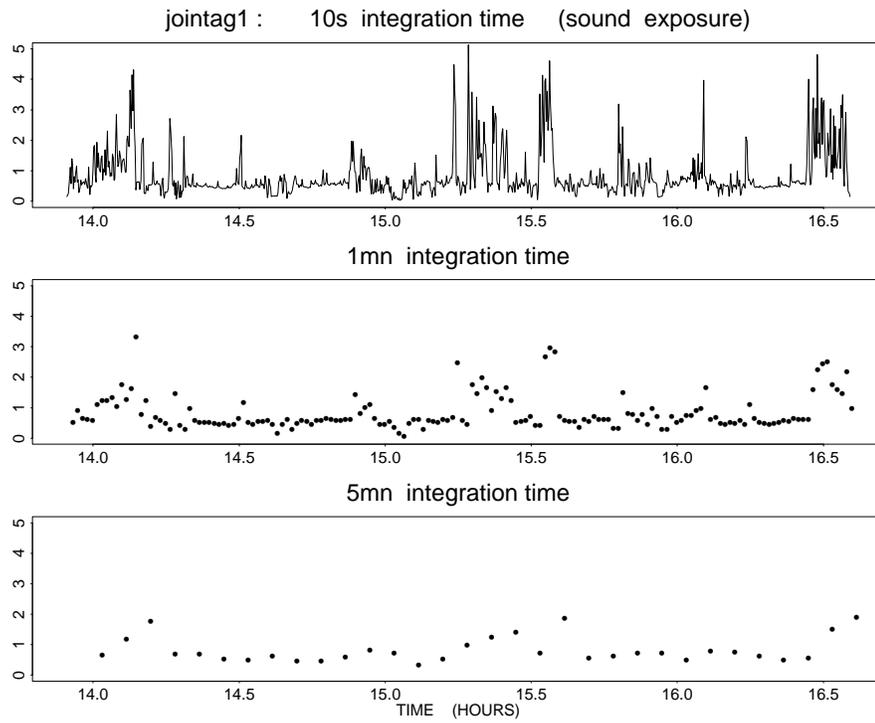


FIGURE 4.1 – Les expositions sonores jointag1 aux trois supports de 10s, 1mn et 5mn.

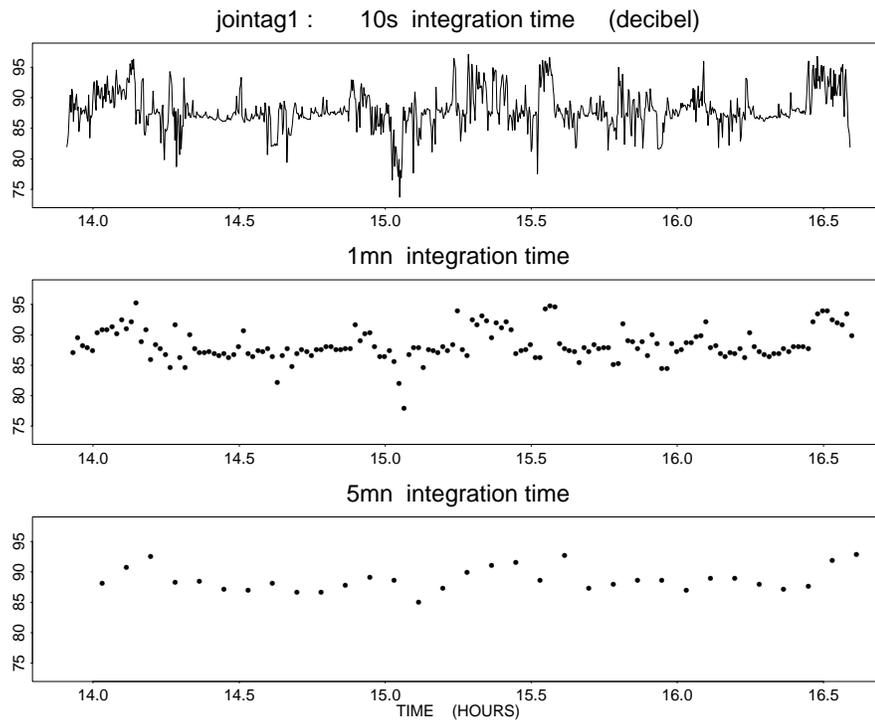


FIGURE 4.2 – Les valeurs de jointag1 exprimées en dB.

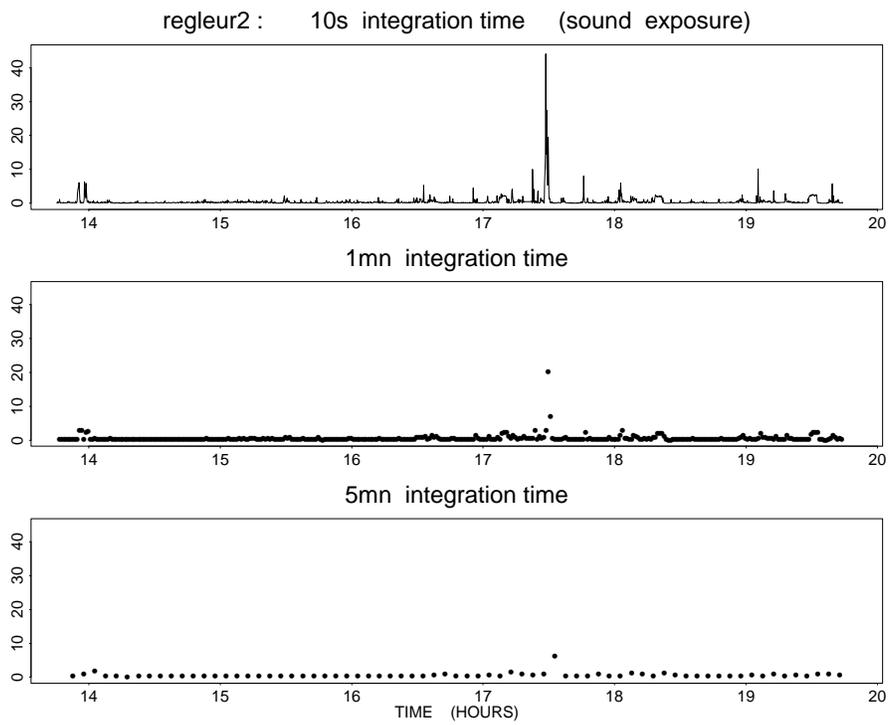


FIGURE 4.3 – Les expositions sonores regleur2 aux trois supports de 10s, 1mn et 5mn.

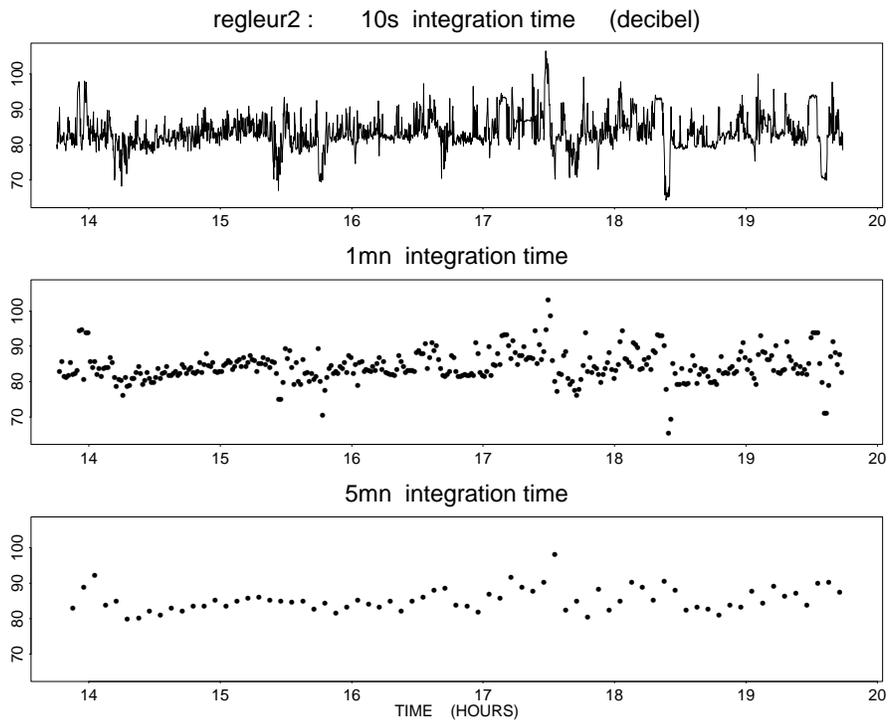


FIGURE 4.4 – Les valeurs de regleur2 exprimées en dB.

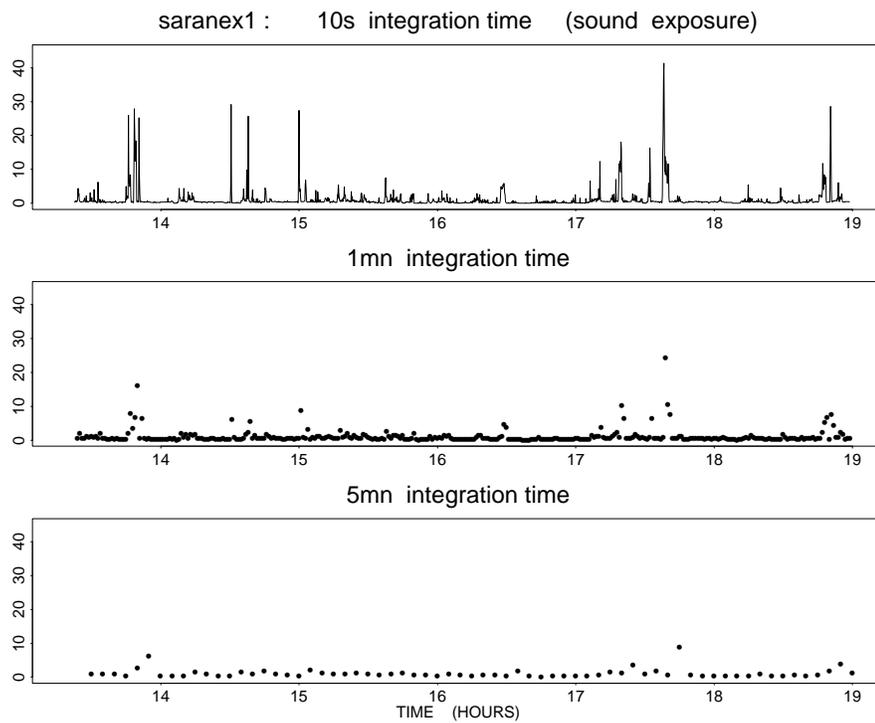


FIGURE 4.5 – Les expositions sonores saranex1 aux trois supports de 10s, 1mn et 5mn.

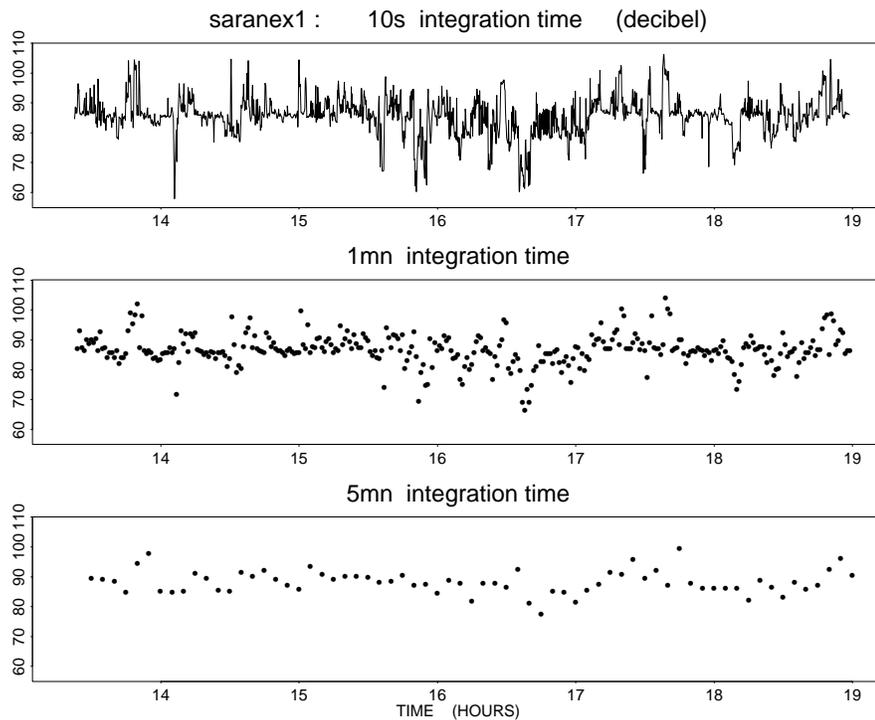


FIGURE 4.6 – Les valeurs de saranex1 exprimées en dB.

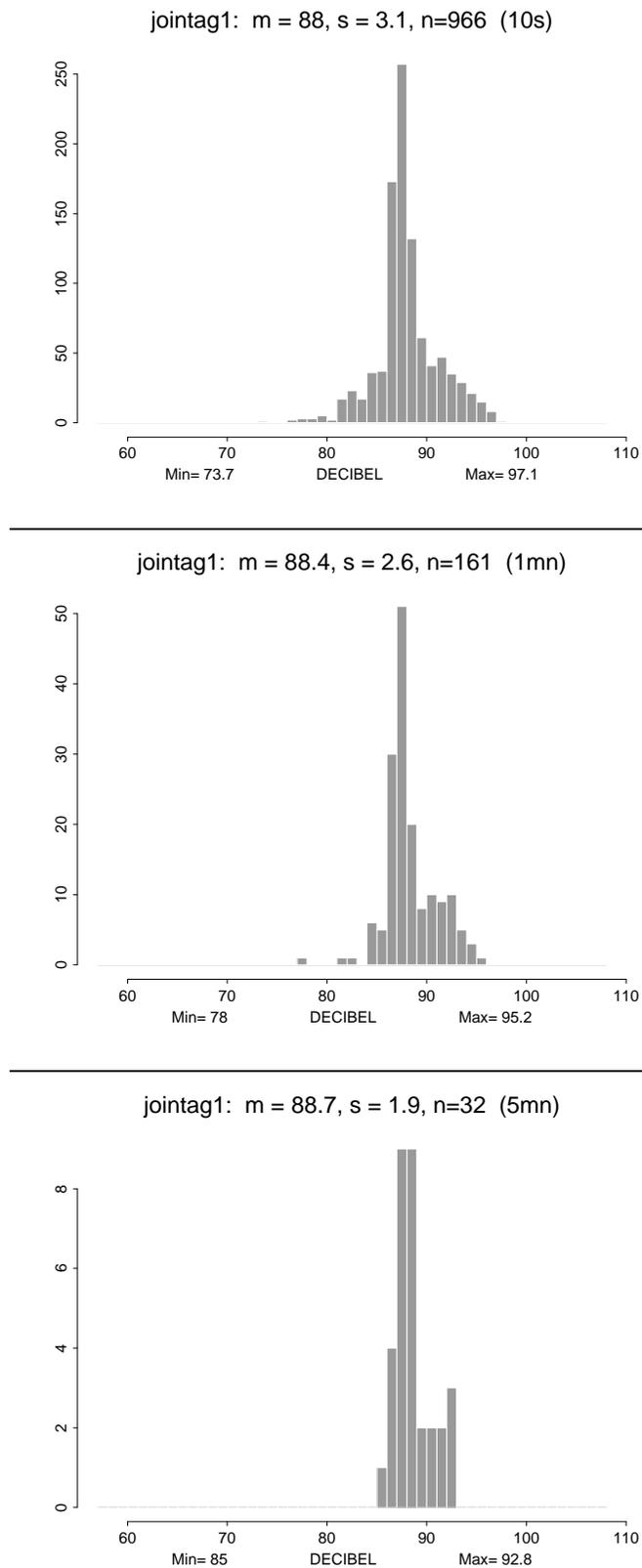


FIGURE 4.7 – Histogrammes en décibels de jointag1 pour des temps d'intégration de 10s, 1mn et 5mn.

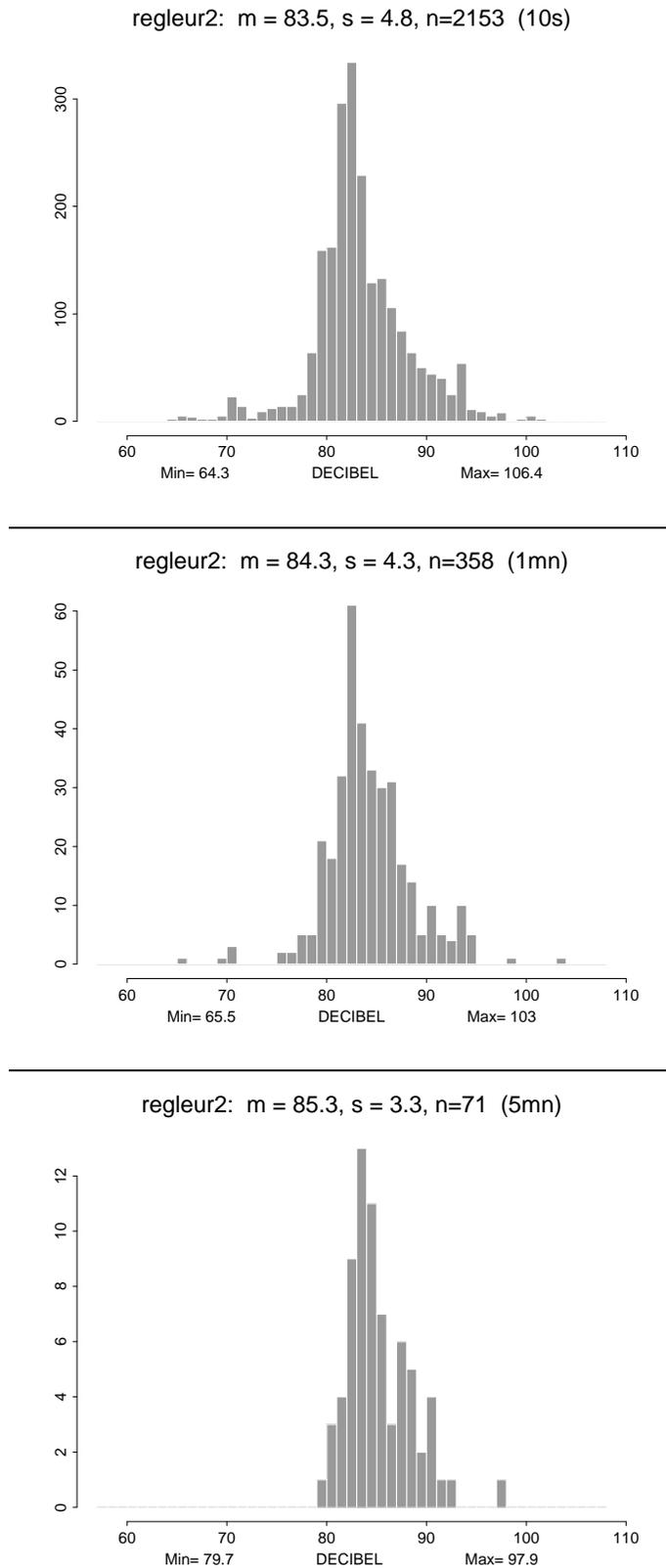


FIGURE 4.8 – Histogrammes en décibels de regleur2 pour des temps d'intégration de 10s, 1mn et 5mn.

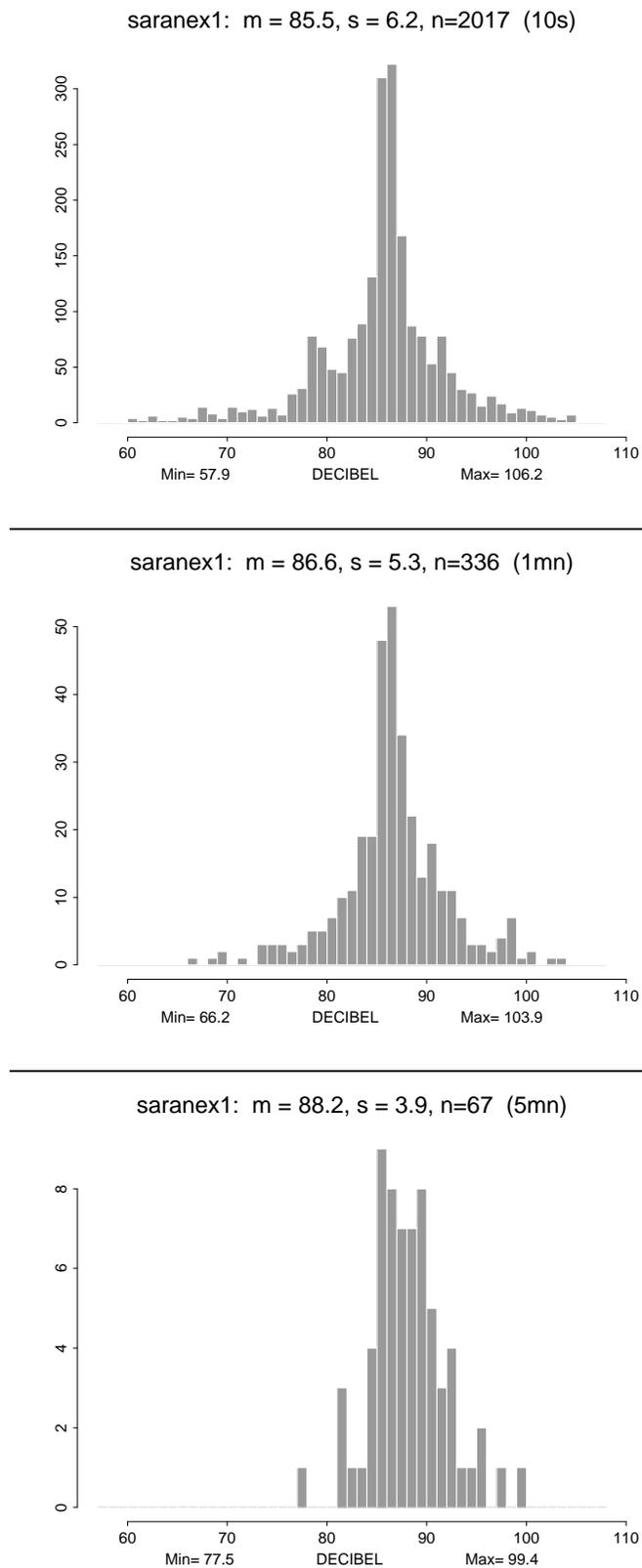


FIGURE 4.9 – Histogrammes en décibels de *saranex1* pour des temps d'intégration de 10s, 1mn et 5mn.

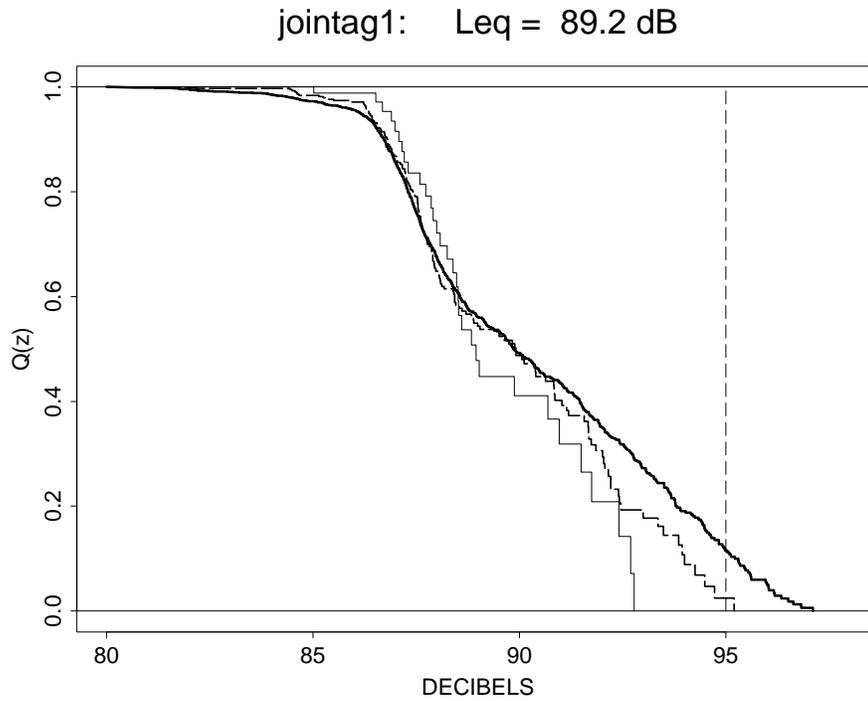


FIGURE 4.10 – Courbes $Q(z)$ de jointag1 aux supports de 10s (trait gras), 1mn (tiretés) et 5mn (trait fin).

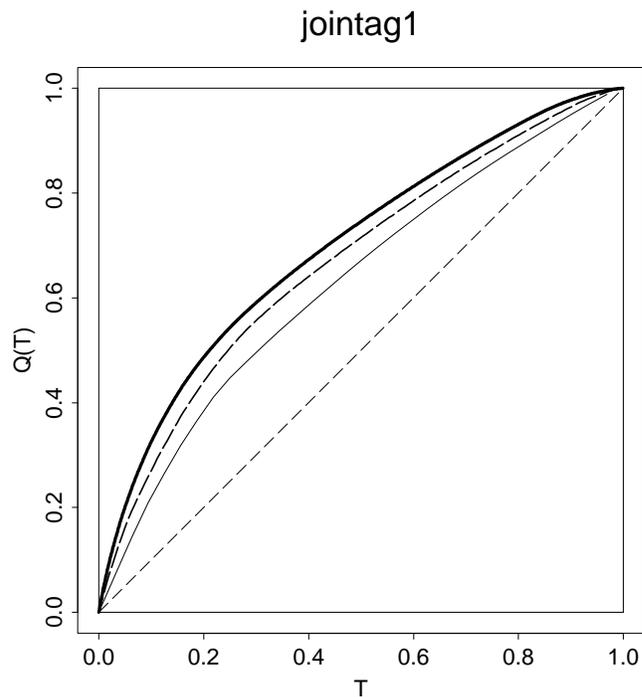


FIGURE 4.11 – Courbes $Q(T)$ de jointag1 aux supports de 10s (trait gras), 1mn (tiretés) et 5mn (trait fin).

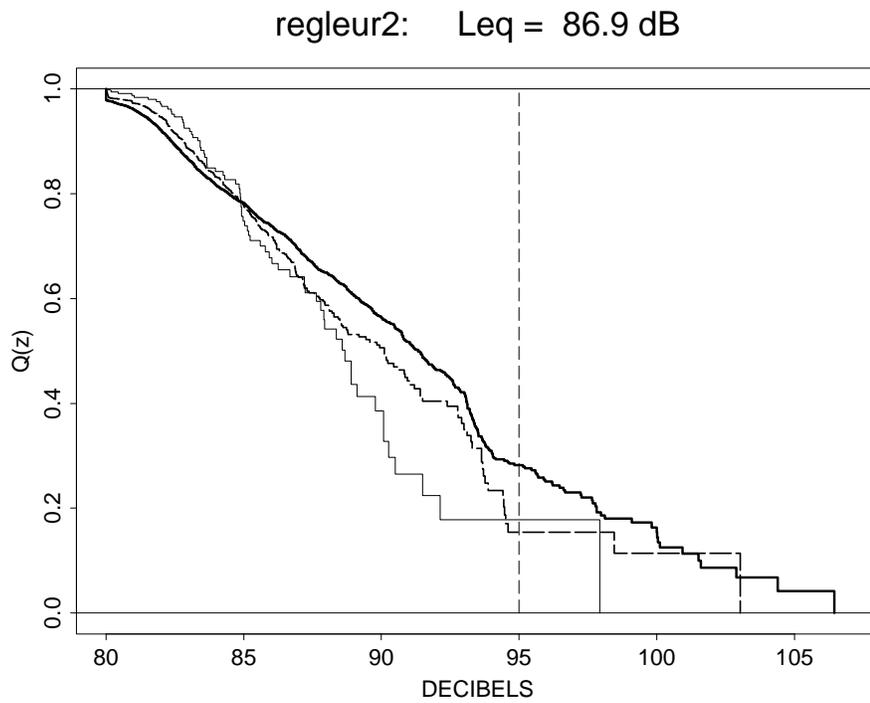


FIGURE 4.12 – Courbes $Q(z)$ de regleur2 aux supports de 10s (trait gras), 1mn (tiretés) et 5mn (trait fin).

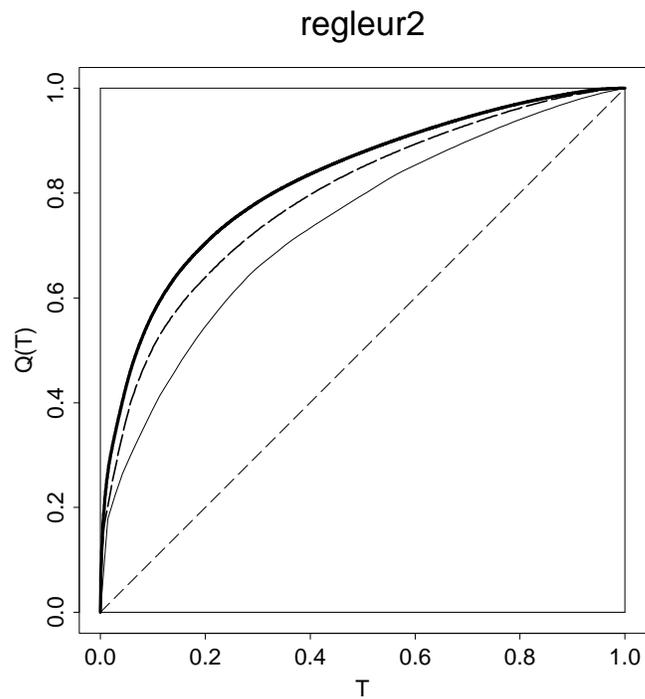


FIGURE 4.13 – Courbes $Q(T)$ de regleur2 aux supports de 10s (trait gras), 1mn (tiretés) et 5mn (trait fin).

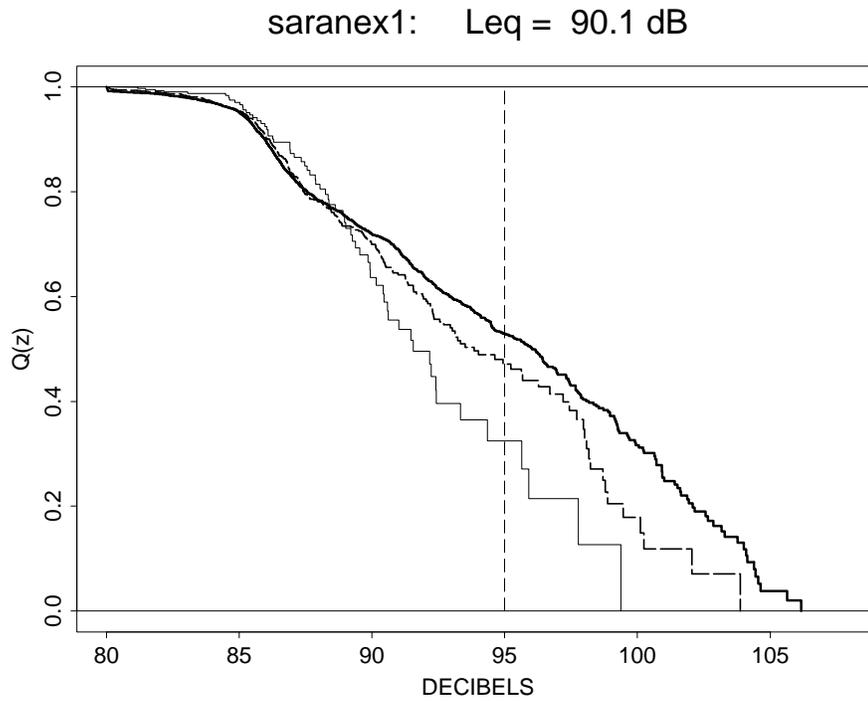


FIGURE 4.14 – Courbes $Q(z)$ de saranex1 aux supports de 10s (trait gras), 1mn (tiretés) et 5mn (trait fin).

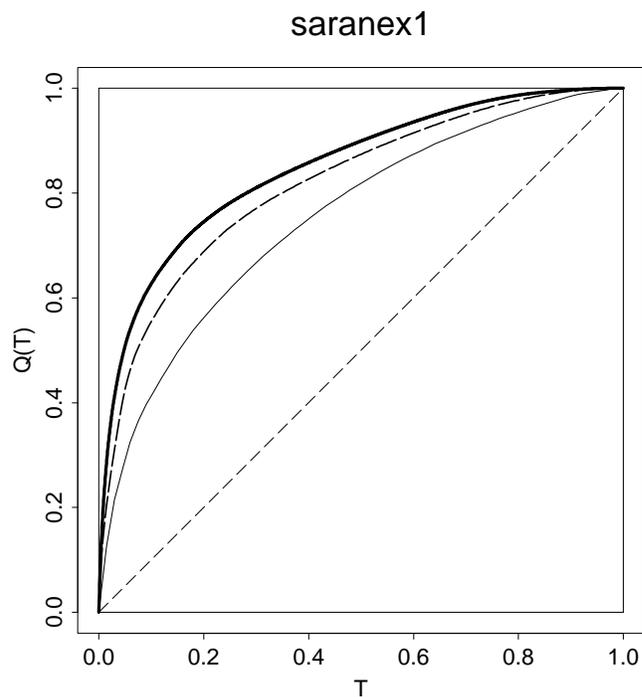


FIGURE 4.15 – Courbes $Q(T)$ de saranex1 aux supports de 10s (trait gras), 1mn (tiretés) et 5mn (trait fin).

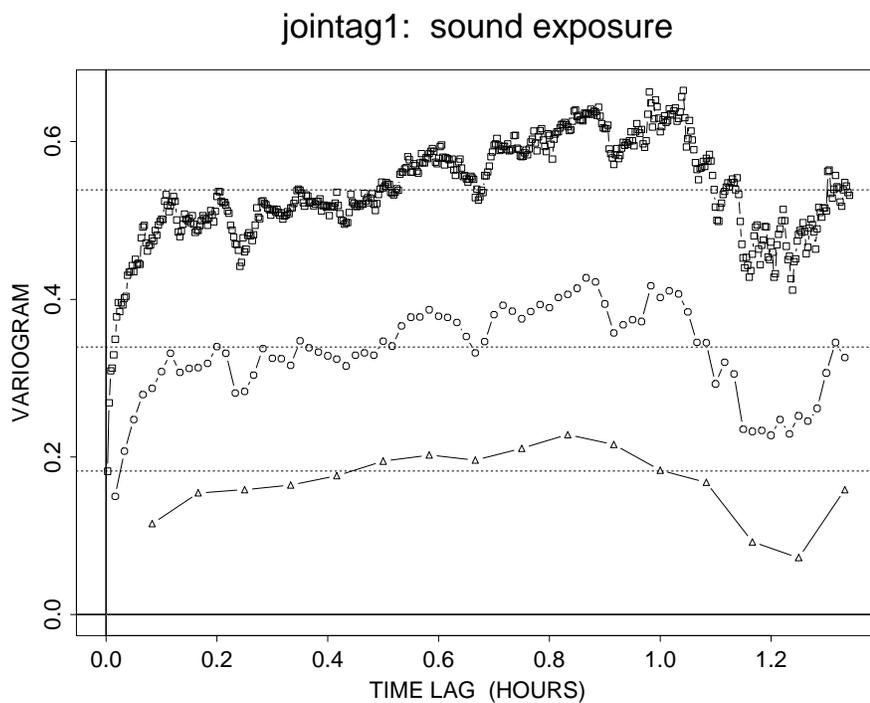
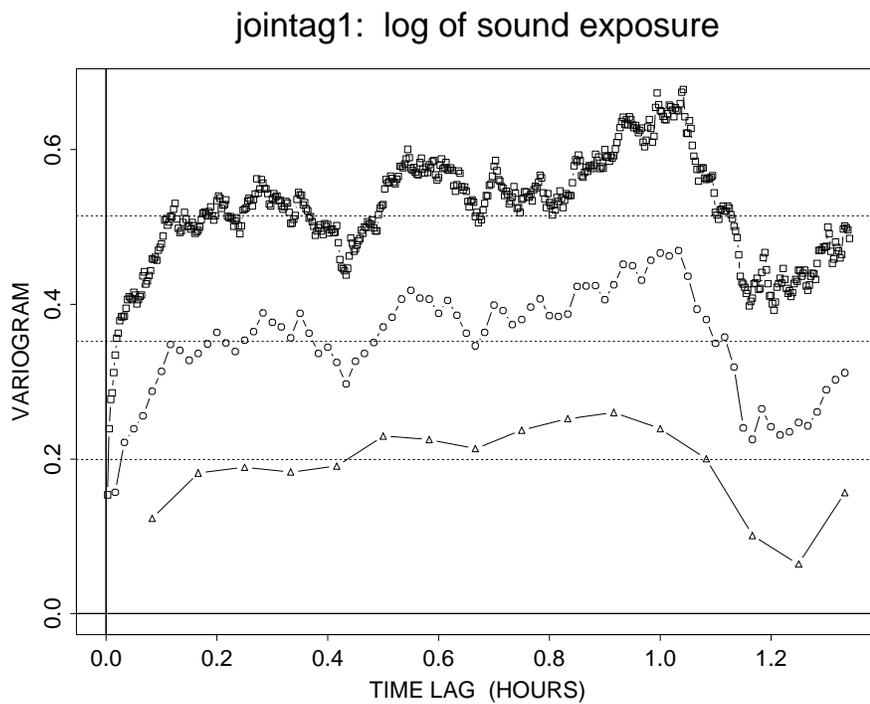


FIGURE 4.16 – Variogrammes expérimentaux de `jointag1` des expositions sonores et de leurs logarithmes pour les supports 10s (haut), 1mn (milieu), 5mn (bas). Les lignes horizontales en tiretées représentent les variances expérimentales pour chacun des trois supports.

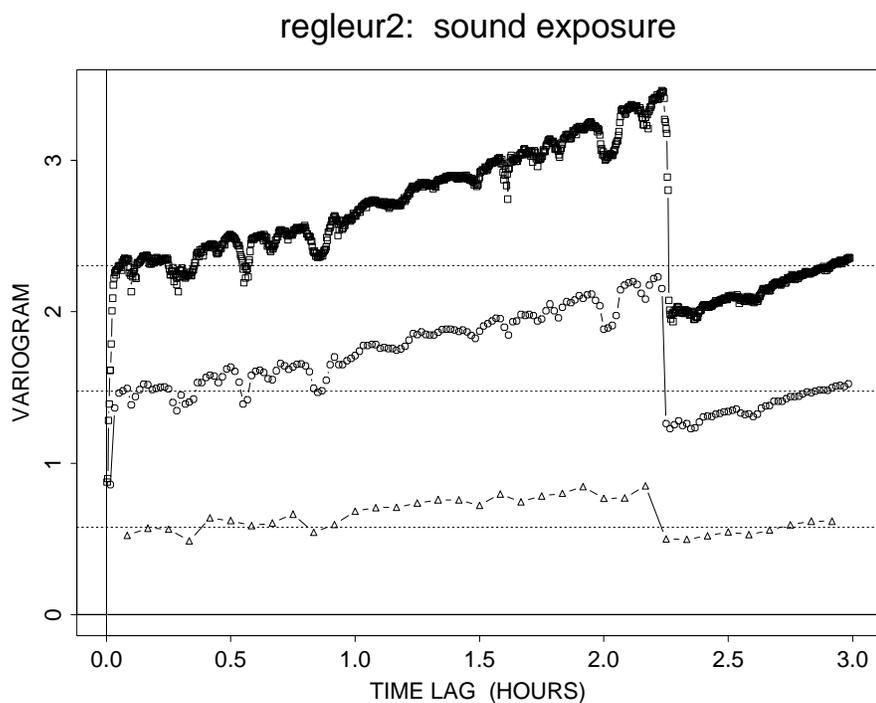
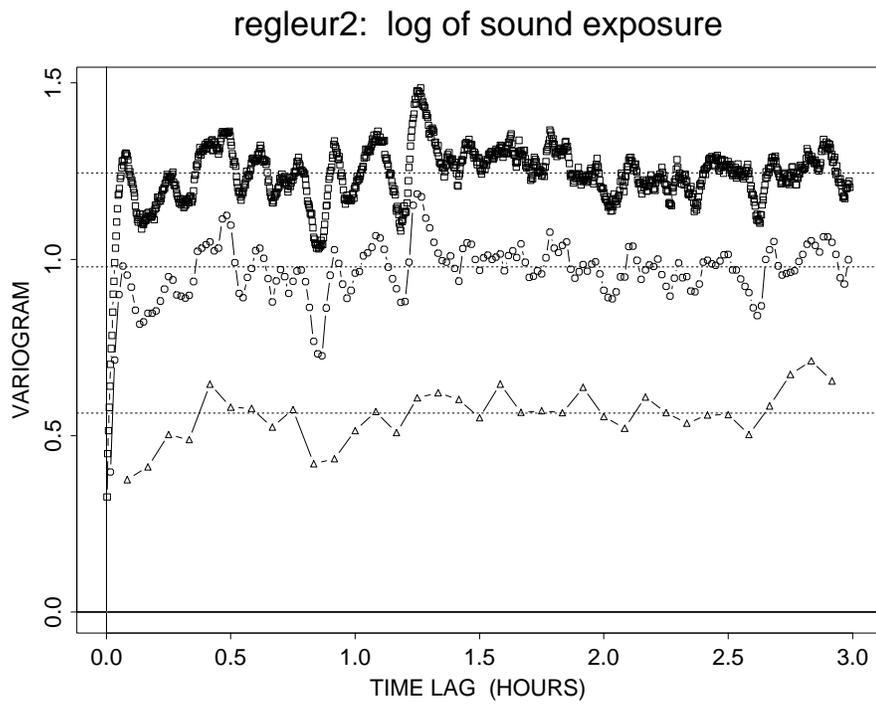


FIGURE 4.17 – Variogrammes expérimentaux de regleur2 des expositions sonores et de leurs logarithmes pour les supports 10s (haut), 1mn (milieu), 5mn (bas). Les lignes horizontales en tiretés représentent les variances expérimentales pour chacun des trois supports.

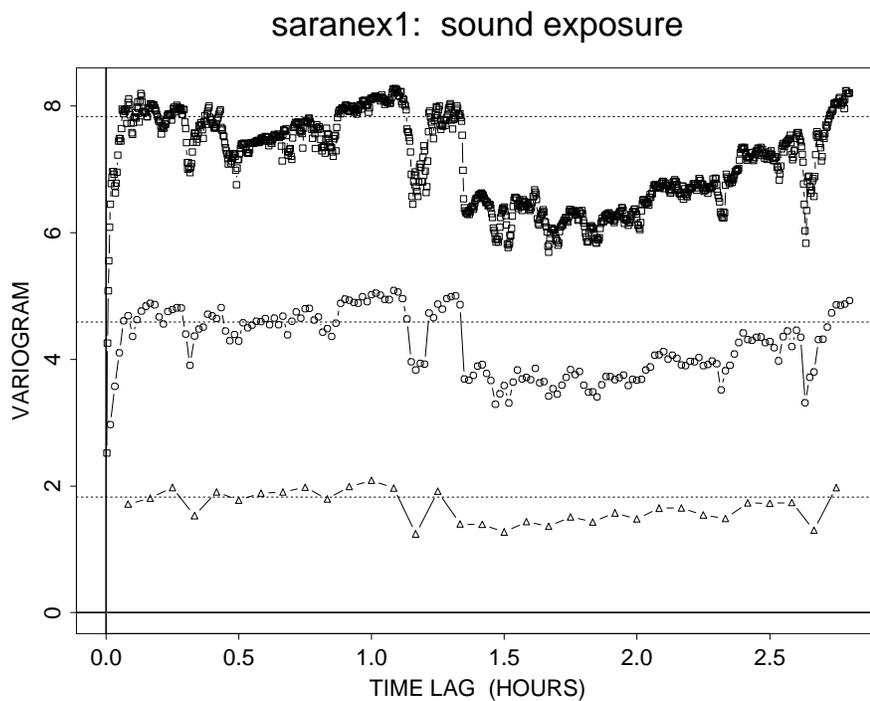
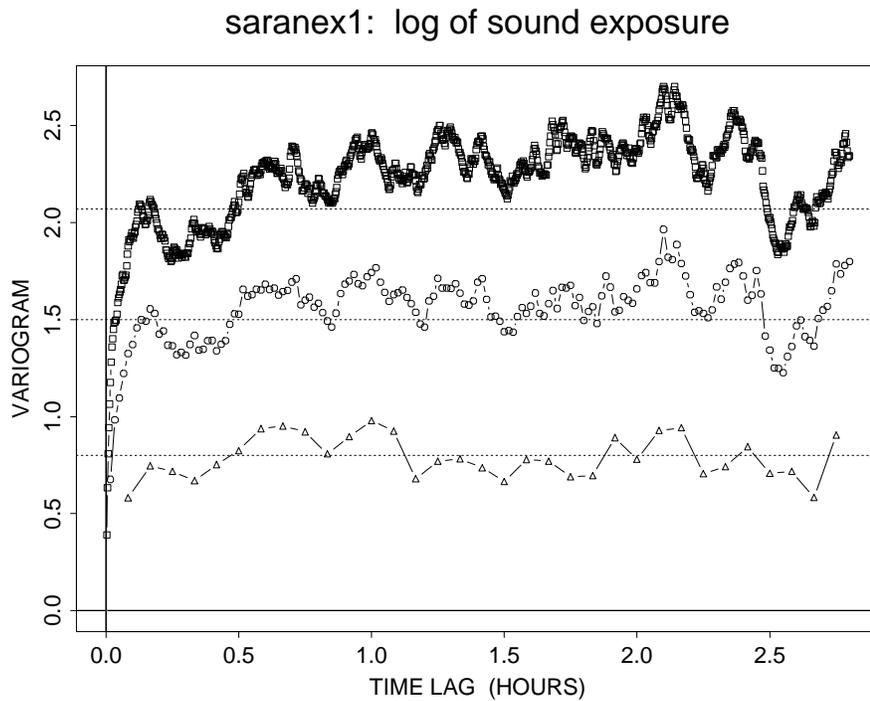


FIGURE 4.18 – Variogrammes expérimentaux de saranex1 des expositions sonores et de leurs logarithmes pour les supports 10s (haut), 1mn (milieu), 5mn (bas). Les lignes horizontales en tiretées représentent les variances expérimentales pour chacun des trois supports.

Chapitre 5

Structure temporelle

Il a été montré précédemment que les séries de données d'exposition sonore sont autocorrélées et dépendantes de l'effet de support, qui est ici la durée d'intégration employée pour le mesurage. Ces effets sont liés à la structure temporelle des données et leur prise en compte dans l'estimation de l'exposition sonore quotidienne et de sa variance est nécessaire pour éviter des biais systématiques d'estimation.

Les variogrammes expérimentaux font apparaître la structure temporelle des données, mais la description qu'ils fournissent ne se prête pas à la construction d'estimateurs de la moyenne et de la variance des données qui tiennent compte de cette structure temporelle.

Pour construire de tels estimateurs, il faut transformer les variogrammes expérimentaux en variogrammes modélisés. Compte tenu du grand nombre de modèles de variogrammes disponibles, il convient d'examiner quels sont les modèles les plus adaptés aux caractéristiques des données étudiées ici, puis de discuter la pertinence et l'impact du choix d'un modèle spécifique.

Dans ce chapitre, trois approches différentes de modélisation de la structure temporelle de séries acoustiques, telle qu'elle apparaît sur le variogramme des logarithmes de l'exposition sonore, sont décrites et appliquées.

La première approche est basée sur un principe de similitude donnant lieu à une théorie lognormale qui rappelle celle de LARSEN [14], souvent citée en liaison avec la modélisation de phénomènes de pollution. Il s'avère cependant que le modèle de variogramme théorique découlant de cette théorie n'est en général pas compatible avec les variogrammes expérimentaux de séries d'exposition sonore et qu'en conséquence cette approche ne pourra pas être retenue.

La deuxième approche utilise un modèle gigogne composé de variogrammes exponentiels et d'un modèle d'effet de pépite. Il comporte suffisamment de paramètres pour être aisément et simultanément ajusté aux variogrammes des trois supports (10s, 1mn et 5mn). C'est ce modèle qui sera utilisé au Chapitre 6 pour l'estimation géostatistique du L_{EQ} .

La troisième approche se propose d'ajuster simultanément les variogrammes des expositions sonores et de leurs logarithmes en utilisant un modèle de variogramme caractérisé par trois paramètres. Cette approche reproduit bien l'allure générale des variogrammes expérimentaux des trois séries jointag1, regleur2 et saranex1.

5.1 Première interprétation : le modèle lognormal auto-homothétique

“La chance a voulu que, d’entrée de jeu, deux circonstances imposent à la géostatistique d’utiliser des *variogrammes* (plutôt que des covariances) c’est-à-dire des *F.A. intrinsèques* (plutôt que des F.A. stationnaires).

La première circonstance, ce sont les résultats expérimentaux de l’Ecole d’Afrique du Sud, en particulier cette courbe d’allure logarithmique où D.G. KRIGE (1952) présentait la variance expérimentale d’échantillons de taille fixée dans des zones de plus en plus grandes du gisement (panneaux, quartiers, concessions, le Rand tout entier, etc. . .).

La seconde circonstance, c’était le modèle à homothétie interne élaboré par DE WIJS (1951), et qui permettait de retrouver cette loi logarithmique de la variance.”

G. MATHERON, 1976 [21]

Effet proportionnel et lognormalité

Dans leur livre classique sur le modèle lognormal AITCHISON ET BROWN [1] définissent l’*effet proportionnel*¹ en stipulant qu’une variable obéit à la loi de l’effet proportionnel, si la variation de cette variable est à tout moment une proportion aléatoire de l’état précédant de la variable. Autrement dit, la variation d’une grandeur z (que l’on supposera positive) obéissant à la loi de l’effet proportionnel est plus adéquatement décrite par le rapport dz/z plutôt que par le seul élément différentiel dz . L’addition d’un grand nombre de petits effets dz/z à peu près indépendants générant une variable y normale :

$$y = \int \frac{dz}{z} = \log z \quad (5.1)$$

il en découle que la variable z est lognormale.

Par analogie on parle d’effet proportionnel lorsque la variance ou le variogramme de z sont proportionnels au carré de sa moyenne. Ce phénomène est souvent un argument pour postuler la lognormalité de z .

Principe de similitude et permanence de la lognormalité

DE WIJS [36] a publié en 1951 un modèle dans lequel les variances de dispersion de valeurs $Z(\Delta t)$ dans un domaine sont proportionnelles à une fonction du rapport de la taille T du domaine à la taille Δt du support de mesure :

$$\sigma^2(\Delta t|T) = \alpha f\left(\frac{T}{\Delta t}\right) \quad (5.2)$$

D’autre part KRIGE a démontré de manière générale qu’une variance de dispersion peut s’exprimer comme la différence entre les valeurs d’une fonction appliquée aux deux supports :

$$\sigma^2(\Delta t|T) = F(T) - F(\Delta t) \quad (5.3)$$

1. Notion qui semble remonter à WEBER H (1834) *De pulsa resorptione auditu et tactu* (cf [1], p3).

MATHERON [16] [17] a montré que le modèle de DE WIJS implique un *principe de similitude* indépendant de la forme de la distribution des valeurs. Seule une fonction logarithmique peut satisfaire aux deux formules et les variances de dispersion dans le modèle de DE WIJS sont donc de la forme :

$$\sigma^2(\Delta t|T) = 3\alpha \log \frac{T}{\Delta t} \quad (5.4)$$

En ce qui concerne la distribution lognormale, on a constaté expérimentalement dans diverses applications² une *permanence de la lognormalité* pour toute une gamme de supports, c'est-à-dire que pour différents supports Δt donné, les distributions avaient toutes un caractère apparemment lognormal. Un principe de similitude a alors servi à étayer ce fait expérimental et a conduit à la formule (5.4) de DE WIJS pour les variances de dispersion. Cette formule n'implique nullement une distribution lognormale, mais seulement un variogramme logarithmique, tel qu'il sera présenté plus bas.

Il faut noter au passage que la permanence de la lognormalité est incompatible avec un modèle de variables aléatoires lognormales *indépendantes* (c'est-à-dire avec une hypothèse d'indépendance effectuée souvent en appliquant des méthodes statistiques). Elle implique en particulier que des valeurs voisines dans le temps, pour un support donné, ne sont pas indépendantes (cf [16], p58).

Variogramme de Wijsien et dispersion absolue

Le variogramme associé au modèle de DE WIJS est tout simplement de la forme :

$$\gamma(h) = 3\alpha \log |h| \quad \text{pour } |h| \neq 0 \quad \text{et} \quad 0 \leq \alpha < 1/3 \quad (5.5)$$

Le paramètre α de la formule (5.4) de DE WIJS caractérise ce que MATHERON a appelé la *dispersion absolue*, tandis que le terme $\log(T/\Delta t)$ est de nature purement géométrique. Dans un contexte minier, en 1955, MATHERON ([16], p63) envisageait la possibilité "qu'à chaque type de minéralisation corresponde une valeur déterminée de la dispersion absolue"³. Si, quarante ans plus tard, un objectif de classification aussi ambitieux n'a toujours pas été réalisé en géologie, on pourrait espérer, en acoustique industrielle, pouvoir comparer des séries de mesures du point de vue de leur dispersion absolue, pour autant qu'elles soient un peu conformes au principe de similitude. C'est ce que nous tenterons sur nos séries.

La dispersion absolue est nécessairement inférieure à l'unité pour des données spatiales à 3D. A une dimension, elle doit même être inférieure à 1/3, puisque la variance D^2 des expositions sonores s'écrit en modèle lognormal, en insérant (5.4) :

$$D^2 = M^2 (e^{\sigma^2} - 1) = M^2 \left(\left(\frac{T}{\Delta t} \right)^{3\alpha} - 1 \right) \quad (5.6)$$

Une valeur de $\alpha = 1/3$ impliquerait une variance arithmétique en $1/\Delta t$, ce qui équivaudrait à une hypothèse d'indépendance des expositions sonores des segments Δt et serait contradictoire avec la permanence de la lognormalité.

2. Pour les polluants atmosphériques, par exemple, LARSEN [14] a redécouvert en 1969 le modèle de Wijs. En acoustique ce modèle a été repris par MALCHAIRE [15], pp88-89. L'article de LARSEN a d'ailleurs suscité un certain nombre de réaction critiques : voir [24], [30].

3. En rajoutant : "Il ne sera permis, cependant, de l'affirmer qu'après un énorme travail de compilation statistico-géologique, portant sur les types métallogéniques les plus variés. On conçoit l'intérêt que présenterait une classification des gisements, fondée sur un critère aussi précis". Une idée analogue est la tentative de caractériser des phénomènes à l'aide de leur *dimension fractale*.

Dispersion absolue par ajustement du variogramme

La dispersion absolue 3α est le paramètre de pente du modèle (5.5) de variogramme de Wijsien. En prenant une échelle logarithmique pour l'abscisse, le variogramme décrit une droite de pente 3α . La valeur à l'origine du variogramme de Wijsien n'est pas définie, ce qui n'est pas gênant dans les applications.

La fonction F de la formule de KRIGE (5.3) dans le cas du variogramme de Wisjien s'exprime :

$$F(L) = 3\alpha \left(\log L - \frac{3}{2} \right) \quad (5.7)$$

et l'on retrouve bien la variance de dispersion (5.4) associée au modèle de Wisjien.

Pour l'ajustement à des variogrammes expérimentaux de données sur support Δt , on a besoin du variogramme théorique régularisé (cf [10], p86) :

$$\begin{aligned} \gamma_{\Delta t}(h) = 3\alpha & \left(\frac{(|h| + \Delta t)^2}{2(\Delta t)^2} \log(|h| + \Delta t) + \frac{(|h| - \Delta t)^2}{2(\Delta t)^2} \log(|h| - \Delta t) \right. \\ & \left. - \frac{|h|^2}{(\Delta t)^2} \log |h| - \log \Delta t \right) \quad \text{pour } |h| \geq \Delta t \end{aligned} \quad (5.8)$$

Pour des distances multiples du support $|h| = k \Delta t$, qui sont les seules intéressantes dans des applications à maille régulière la formule devient (cf [17], p84) :

$$\begin{aligned} \gamma_{\Delta t}(k) = 3\alpha & \left(\frac{(k+1)^2}{2} \log(k+1) + \frac{(k-1)^2}{2} \log(k-1) \right. \\ & \left. - k^2 \log k \right) \quad \text{pour } k = 2, 3, \dots \end{aligned} \quad (5.9)$$

On remarque que la valeur pour $k = 1$ (premier pas) n'est pas définie.

Dès que k est supérieur à 2 ou 3, l'expression se réduit pratiquement à :

$$\gamma_{\Delta t}(k) \cong 6\alpha \left(\log k + \frac{3}{2} \right) \quad \text{pour } k = 3, 4, \dots \quad (5.10)$$

Pour $k = 1, 2$ MATHERON ([17], p85) donne les valeurs suivantes :

$$\gamma_{\Delta t}(1)/6\alpha = 1.386 \quad \gamma_{\Delta t}(2)/6\alpha = 2.171 \quad (5.11)$$

Pour la pratique cette formule signifie que pour déterminer la dispersion absolue de données à maille régulière et à supports adjacents, il suffit d'ajuster une droite au variogramme expérimental en se servant sur l'abscisse du logarithme du numéro des classes de distance. On ne cherchera pas à faire absolument passer la droite par les deux premiers points du variogramme expérimental, puisque la formule (5.10) n'est pas exacte pour les deux premiers pas.

Dispersion absolue par la méthode des équivalents

La *méthode des équivalents* ([17], p163) porte ce nom parce qu'elle utilise dans les applications 3D des échantillons à support 1D linéaires, considérés comme équivalents à des parallélépipèdes. Dans les applications temporelles qui nous intéressent, les supports 1D sont évidemment d'emblée linéaires.

La dispersion absolue peut être déterminée par une formule calquée sur (5.4) n'utilisant que la variance s^2 logarithmique expérimentale :

$$s^2 = 3\alpha \log \frac{T}{\Delta t} \quad (5.12)$$

On estime alors α par :

$$\alpha = \frac{s^2}{3 \log \frac{T}{\Delta t}} \quad (5.13)$$

Cette formule peut donner en métallogénie une bonne approximation de la dispersion absolue *sans nécessiter le calcul d'un variogramme*, comme MATHERON ([17], ch. IX) a pu le constater sur un certain nombre d'exemples miniers⁴. Nous allons examiner cette question dans le cas de quatre séries acoustiques dans la prochaine section.

Application : calcul de la dispersion absolue

Nous utilisons deux méthodes pour calculer le coefficient de dispersion absolue. D'une part, nous le calculons par la formule 5.13 de la méthode des équivalents et le notons *eqv* sur les graphiques. D'autre part, nous ajustons le coefficient par moindres carrés en nous basant sur les expressions (5.10) et (5.11).

Pour chaque série on montre sur les Figures 5.1 à 5.4 (pp 45 à 48) l'ajustement du variogramme des logarithmes des expositions sonores en utilisant deux représentations différentes. D'abord, avec une abscisse en terme de logarithme des distances (plus la constante 3/2), ce qui permet de représenter le modèle de Wijsien par une droite. Puis, en dessous, nous donnons le graphique habituel du variogramme expérimental, sans transformation logarithmique des distances.

Le critère d'ajustement de ce modèle simple (voir simpliste !) est que l'on va viser à reproduire au mieux le comportement du variogramme à proximité de l'origine, sans se préoccuper du comportement aux grandes distances. Dans un deuxième temps nous commentons le comportement aux grandes distances vis-à-vis du modèle ajusté et tirons les conséquences du point de vue de l'interprétation.

Il est intéressant de comparer la valeur du coefficient de dispersion absolue obtenu par ajustement du variogramme des logarithmes avec la valeur calculée par la méthode de équivalents (notée *eqv*).

La Figure 5.1 nous montre pour que *jointag1* le coefficient de dispersion ajusté vaut la moitié de celui calculé par la formule. Pour *regleur2* la Figure 5.2 nous indique que la valeur calculée par la formule correspond, grosso modo, à celle d'un ajustement du variogramme près de l'origine. Pour *saranex1* la Figure 5.3 donne un coefficient ajusté d'un quart inférieur au coefficient calculé par la formule.

Nous montrons également, sur la Figure 5.4, le variogramme d'une nouvelle série, *cyclo41*, qui ne sera pas discutée en dehors du présent chapitre. Il s'agit d'une série non-homogène (contrairement aux trois séries vues jusqu'ici) qui est subdivisible en distinguant plusieurs phases de travail (telle la série *amenag31* qui sera traitée en détail dans le Chapitre 7).

4. Lorsqu'il n'y a pas d'anisotropie, c'est-à-dire quand le variogramme est le même pour toutes les directions de l'espace 2D ou 3D. Cette restriction ne s'applique évidemment pas pour nos applications unidimensionnelles, le long de l'axe temporel.

La série `cyclo41` nous donne un exemple de série se laissant relativement bien ajuster à toutes les distances du domaine par une droite. Le coefficient ajusté vaut cependant la moitié de la valeur calculée par la méthode des équivalents.

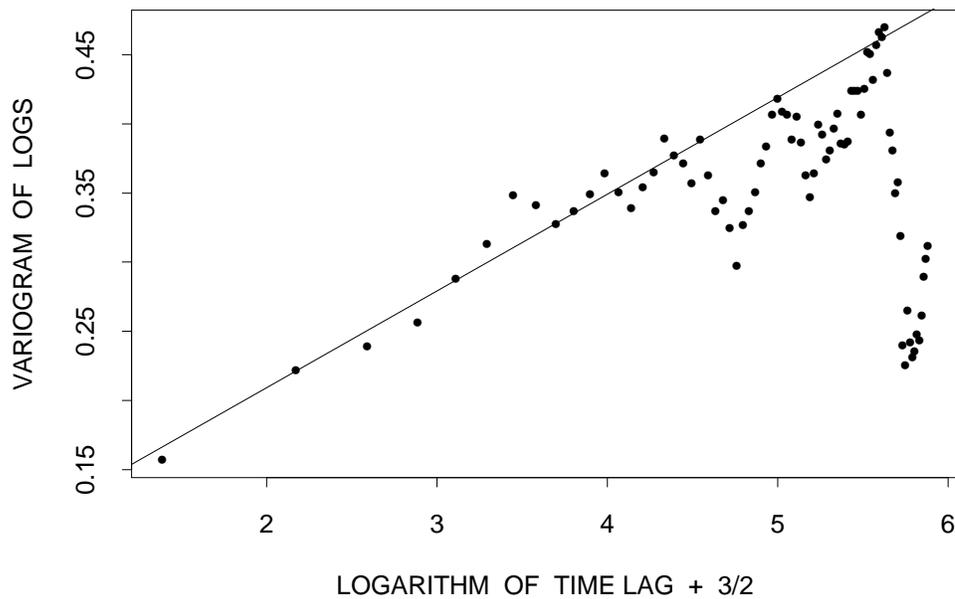
Ce qui distingue la série `cyclo41` des trois autres, c'est qu'elle n'a pas de palier, et c'est pour cela que le modèle de Wijsien peut "marcher" dans ce cas-ci ! En effet, un palier est contradictoire avec le principe de similitude, puisque la formule (5.4) suppose que la variance de dispersion croît indéfiniment en fonction du support Δt . Elle suppose également une croissance sans changement de pente, puisque le modèle repose entièrement sur la dispersion absolue.

Parmi les quatre exemples traités, nous avons deux séries (`regleur2` et `saranex1`) avec un palier bien marqué, ce qui n'est pas compatible avec le modèle Wijsien. Nous avons une série (`cyclo41`) pour laquelle le modèle semble pertinent du point de vue du variogramme et une série (`jointag1`) exhibant un décrochement aux grandes distances, dont on ne peut dire s'il s'agit de l'amorce d'un palier ou non.

La méthode des équivalents ne donne en général pas une valeur identique à celle obtenue par ajustement du variogramme au voisinage de l'origine. C'est étrangement justement pour la série `regleur2`, pour laquelle le palier est le plus affirmé, que la méthode des équivalents donne une valeur coïncidant avec la pente du variogramme au voisinage de l'origine.

En conclusion, le modèle de Wijsien est généralement inadapté à la modélisation log-normale de séries d'exposition sonores. La théorie séduisante d'autohomothétie des variances de dispersion ne s'applique pas dans ce contexte.

jointag1: 0.012 absolute dispersion (eqv: 0.023)



De Wijsian variogram model

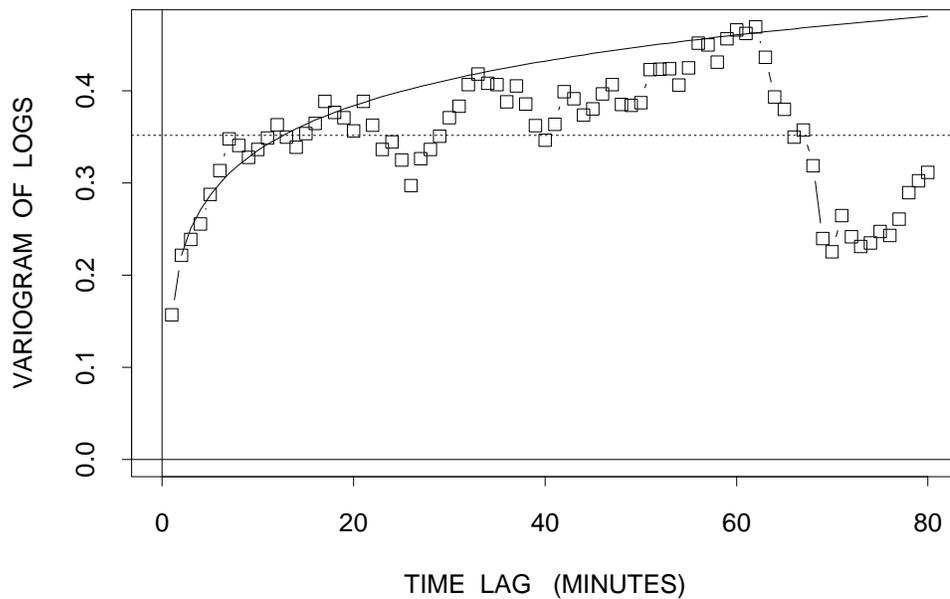
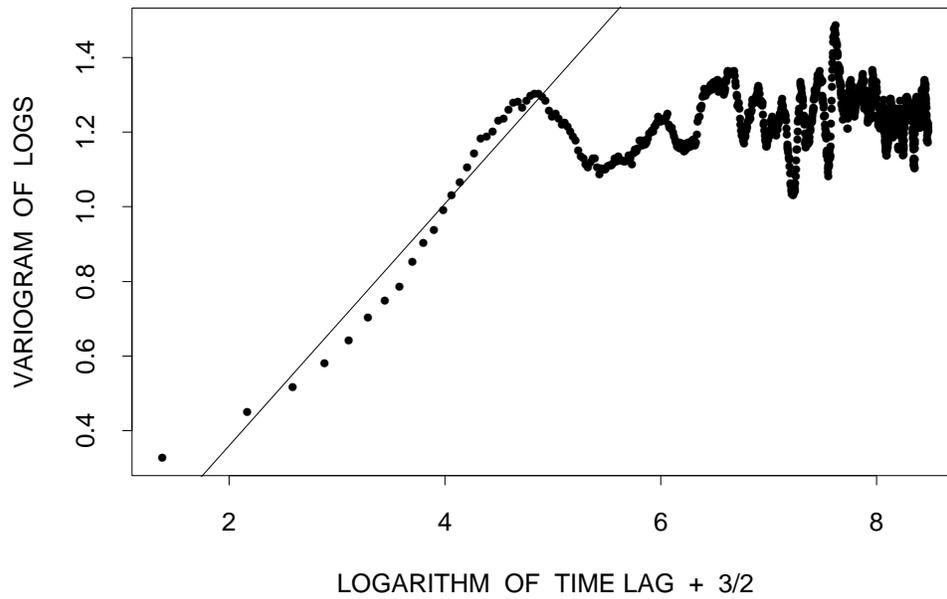


FIGURE 5.1 – Variogramme expérimental au pas (et support) de $\Delta t = 1$ minute, ajusté avec un modèle de Wijsien.

regleur2: 0.054 absolute dispersion (eqv: 0.054)



De Wijnsian variogram model

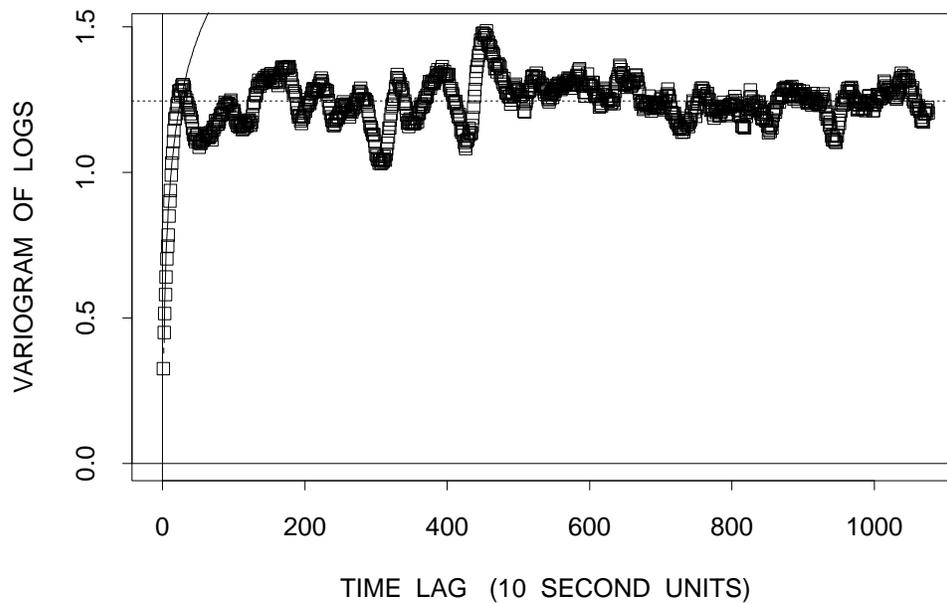
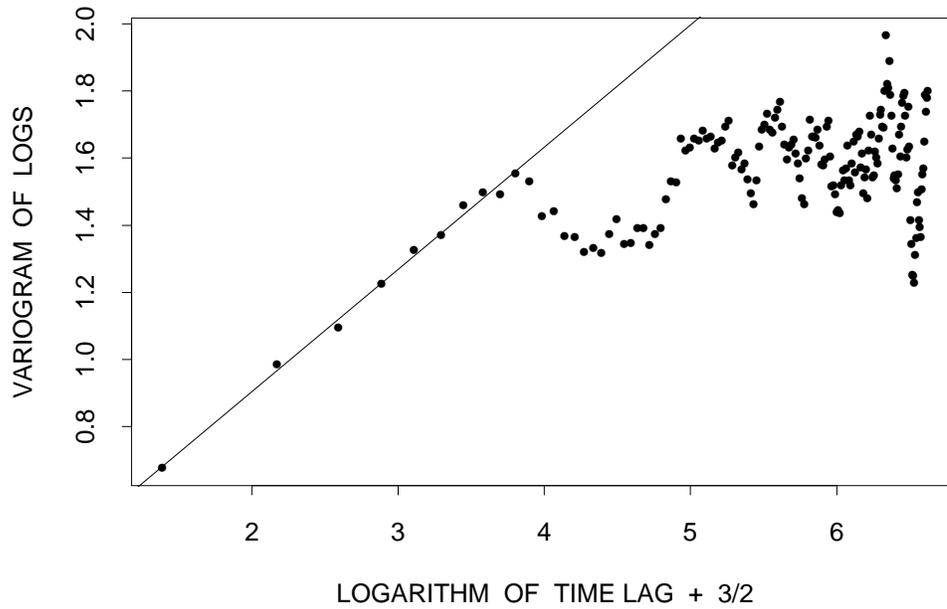


FIGURE 5.2 – Variogramme expérimental au pas (et support) de $\Delta t = 10$ secondes, ajusté avec un modèle de Wijnsien.

saranex1: 0.061 absolute dispersion (eqv: 0.086)



De Wijsian variogram model

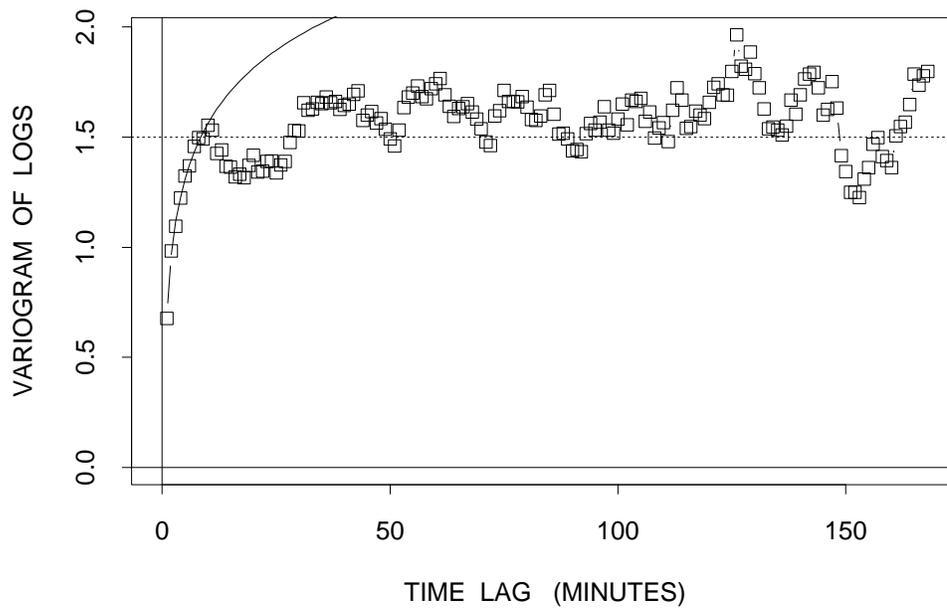
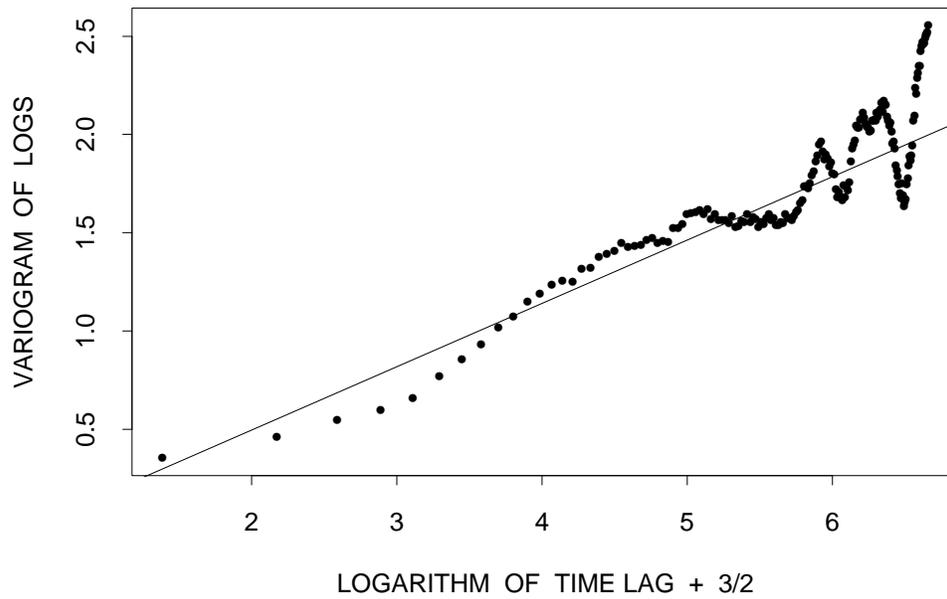


FIGURE 5.3 – Variogramme expérimental au pas (et support) de $\Delta t = 1$ minute, ajusté avec un modèle de Wijsien.

cyclo41: 0.054 absolute dispersion (eqv: 0.112)



De Wijsian variogram model

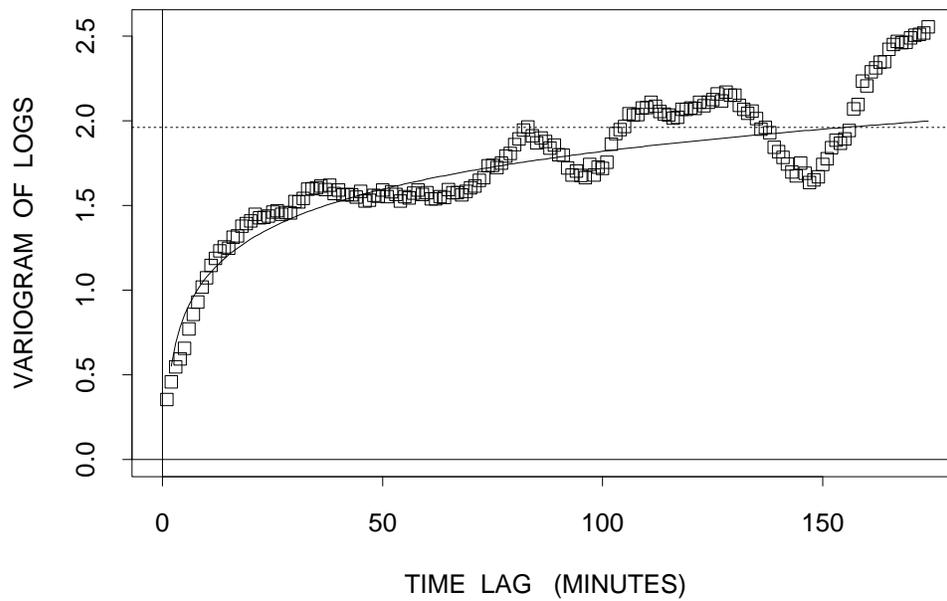


FIGURE 5.4 – Variogramme expérimental au pas (et support) de $\Delta t = 1$ minute, ajusté avec un modèle de Wijsien.

5.2 Deuxième interprétation : un modèle lognormal stationnaire

Modèle exponentiel régularisé

Nous utilisons maintenant une approche différente et considérons que les variogrammes des logarithmes de l'exposition sonore présentent tous des paliers. On suppose que sur un support ponctuel le variogramme est de type exponentiel :

$$\gamma(h) = b \left(1 - \exp\left(-\frac{|h|}{a}\right) \right) \quad \text{avec } a, b > 0 \quad (5.14)$$

où a représente le paramètre de portée du modèle de variogramme et b le palier. Pour le modèle exponentiel, qui n'atteint qu'asymptotiquement son palier (pour $|h| \rightarrow \infty$) il est coutumier de parler d'une *portée pratique* de trois fois le paramètre de portée, puisque pour une distance de $3a$ le variogramme vaut 95% du palier b .

Pour ce modèle il existe une formule analytique donnant le variogramme pour différents supports à une dimension (cf [8] ; [10], p84 ; [34], p57) :

$$\gamma_{\Delta t}(h) = \begin{cases} \frac{b a^2}{(\Delta t)^2} \left(2e^{-\Delta t/a} - 2 + \frac{2h}{a} + e^{-h/a} (2 - e^{-\Delta t/a}) - e^{(h-\Delta t)/a} \right) & \text{pour } 0 \leq h \leq \Delta t, \\ \frac{b a^2}{(\Delta t)^2} \left(e^{-\Delta t/a} - e^{\Delta t/a} + (e^{-\Delta t/a} + e^{\Delta t/a} - 2) \cdot (1 - e^{-h/a}) \right) & \text{pour } h > \Delta t. \end{cases} \quad (5.15)$$

Nous tenons à souligner que c'est l'interprétation statistique par un modèle lognormal qui nous autorise à moyenner les logarithmes des expositions sonores, chose qui a priori, d'un point de vue physique, ne présente aucun sens.

Ajustement de trois séries par un modèle exponentiel gigogne

Pour donner plus de flexibilité à l'ajustement nous utilisons un modèle gigogne composé de S structures exponentielles et d'un effet de pépite, que nous régularisons sur le support Δt correspondant à un variogramme expérimental :

$$\gamma_{\Delta t}(h) = \sum_{u=0}^S \gamma_{\Delta t}^u(h) \quad (5.16)$$

où chaque structure a un paramètre de portée et un palier différent.

Chacune des trois séries `jointag1`, `regleur2`, `saranex1` et `cyclo41` a été ajustée avec un modèle gigogne à deux structures exponentielles dont les paramètres des modèles ponctuels de référence sont donnés sur le Tableau 5.1. Le modèle ponctuel a été choisi de telle manière qu'en le régularisant, il se superpose à chacun des trois variogrammes expérimentaux. On y trouve également l'effet de pépite associé au variogramme sur support de 10s (étant donné que ce concept n'a pas de sens à l'échelle ponctuelle). L'effet de pépite correspond à la microvariabilité à une échelle inférieure à 10s et il est linéairement régularisé pour les variogrammes aux supports de 1mn et 5mn.

Série	Pépité (10s)	Petite échelle		Grande échelle	
		portée pratique	palier	portée pratique	palier
jointag1	.14	3mn	.30	1h 19mn	.18
regleur2	.17	5mn	1.01	1h 11mn	.11
saranex1	.07	2mn	1.91	2h 52mn	.52

TABLE 5.1 – Paramètres des modèles exponentiels ponctuels, ainsi que de l’effet de pépité au support de 10s.

Les ajustements sont montrés sur la Figure 5.5 (page 51). La structure prépondérante pour `regleur2` est celle de courte portée (5mn pour le modèle ponctuel de référence) auquel vient s’ajouter une structure d’une heure de portée. La structure de grande portée prend plus d’importance sur les variogrammes de `jointag1` et `saranex1`. Dans ce dernier cas elle est de l’ordre de trois heures.

Les modèles gigognes exponentiels ajustés seront utilisés pour une estimation lognormale du L_{EQ} par krigeage au Chapitre 6.

Variances de dispersion en fonction du support

La variance de dispersion d’échantillons de support Δt dans un domaine temporel \mathcal{D} peut être calculée à partir du modèle de variogramme. On a la formule

$$\sigma^2(\Delta t|\mathcal{D}) = \bar{\gamma}(\mathcal{D}, \mathcal{D}) - \bar{\gamma}(\Delta t, \Delta t) \quad (5.17)$$

$$= F(\mathcal{D}) - F(\Delta t) \quad (5.18)$$

exprimée en termes de variogrammes moyens $\bar{\gamma}$, que l’on peut également réécrire sous la forme de fonctions auxiliaires F ne dépendant que d’un seul paramètre.

Pour un modèle exponentiel la fonction auxiliaire est de la forme

$$F(L) = b \left(1 + \frac{2a}{L} \left(\frac{a}{L} - 1 \right) - \frac{2a^2}{L^2} \exp \left(-\frac{L}{a} \right) \right) \quad (5.19)$$

où L représente soit Δt , soit \mathcal{D} .

Pour l’effet de pépité (bruit blanc) on calcule la variance de dispersion à partir de la valeur $b_{\Delta t}^0$ de l’effet de pépité d’un support Δt donné

$$\sigma^2(\Delta t|\mathcal{D}) = b_{\Delta t}^0 \left(1 - \frac{\Delta t}{\mathcal{D}} \right) \quad (5.20)$$

Une courbe des variances de dispersion en fonction du temps d’intégration Δt en prenant $\mathcal{D}=8\text{h}$ est montrée sur la Figure 5.6 (page 52) pour chacune des séries `jointag1`, `regleur2` et `saranex1`. La courbe décrit la réduction de la variance de dispersion lorsque l’on augmente le temps d’intégration. La décroissance est très rapide près de l’origine et ralentit lorsque l’on passe à l’échelle de l’heure. On remarque une allure particulière de la courbe de `saranex1` due à la présence d’une structure exponentielle de l’ordre de 3h de portée pratique, au lieu de 1h pour `jointag1` et `regleur2` (Tableau 5.1).

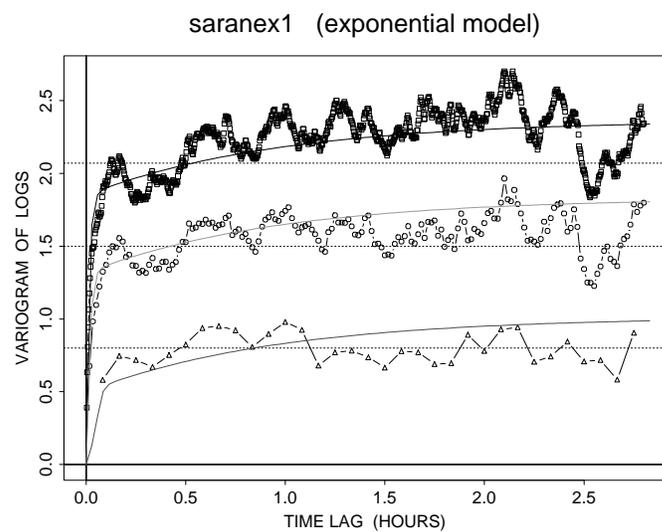
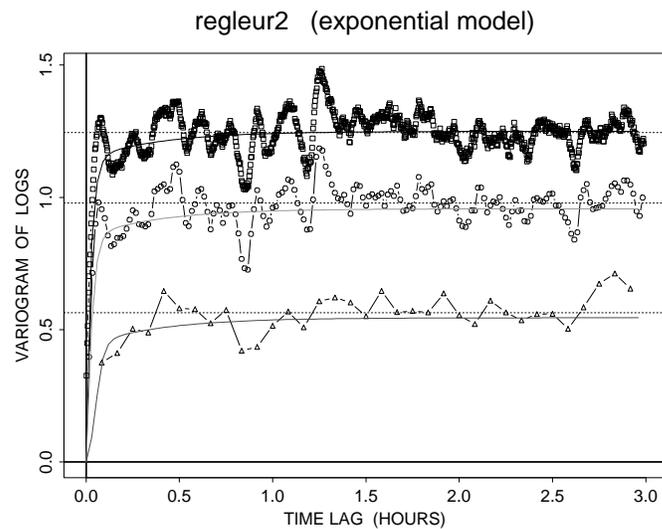
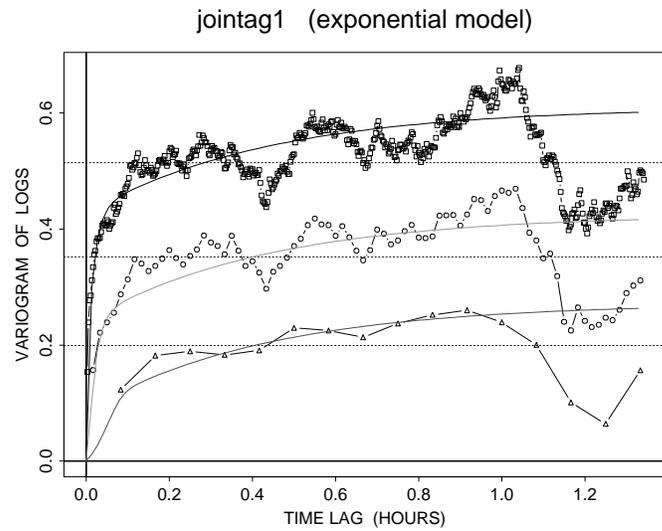


FIGURE 5.5 – Ajustement pour les séries jointag1, regleur2 et saranex1 de variogrammes des logarithmes avec un modèle exponentiel gignone pour les supports 10s (carrés), 1mn (cercles), 5mn (triangles).

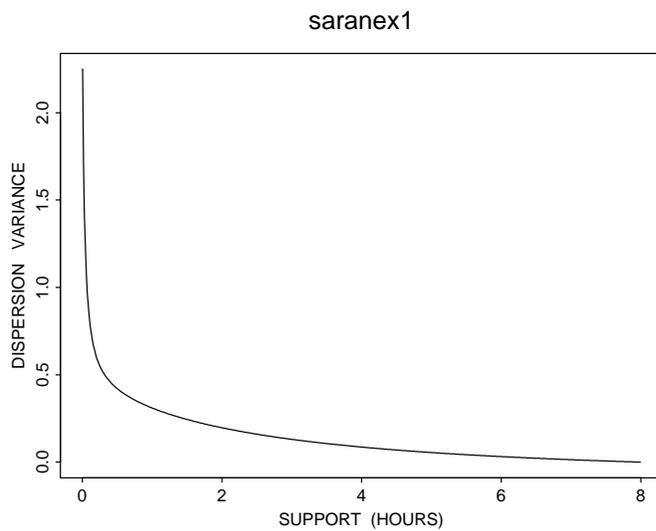
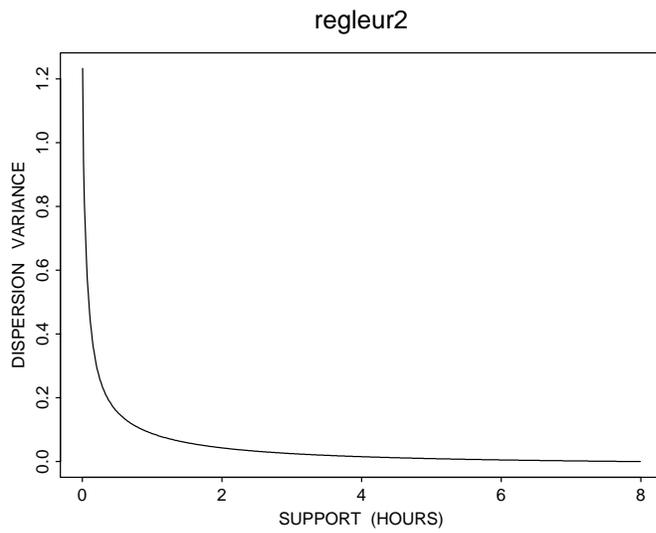
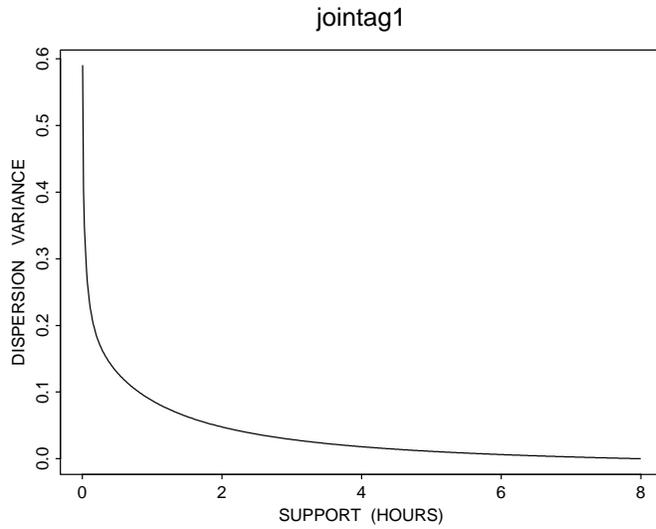


FIGURE 5.6 – Courbes de variance de dispersion (dans une journée de 8h) en fonction du support.

5.3 Troisième interprétation : un modèle lognormal localement stationnaire

Rappel sur les FASL

MATHERON [20] définit $Z(t)$ comme étant une fonction aléatoire *stationnaire lognormale* (FASL) de covariance $C(h)$ et de moyenne M , si $Y(t) = \log Z(t)$ est une fonction aléatoire stationnaire gaussienne d'espérance m et de covariance $\sigma(h)$. On a alors :

$$M = e^{m + \sigma^2/2} \quad (5.21)$$

$$C(h) = M^2 \left(e^{\sigma(h)} - 1 \right) \quad (5.22)$$

Si $\gamma(h)$ est le variogramme de $Y(t)$ et A une constante telle que :

$$\sigma(h) = A - \gamma(h) \quad (5.23)$$

il vient pour le variogramme de $Z(t)$:

$$\Gamma(h) = M^2 e^A \left(1 - e^{-\gamma(h)} \right) \quad (5.24)$$

Théorie des FALLS

Pour couvrir des phénomènes dont le variogramme logarithmique $\gamma(h)$ ne présente pas de palier MATHERON [20] contourne la difficulté qu'il ne peut exister de fonction aléatoire intrinsèque lognormale, en considérant une fonction aléatoire $Y(t)$ *localement stationnaire sur T* (par le biais d'une représentation privilégiée d'une fonction aléatoire intrinsèque), ayant une espérance inconnue m_T et une covariance du type $\sigma(h) = A - \gamma(h)$, où le variogramme est cependant autorisé à prendre la forme non bornée

$$\gamma(h) = b|h|^p \quad \text{avec} \quad 0 < p < 2 \quad (5.25)$$

(en sus, évidemment, des modèles habituels à palier du type $\gamma(h) = \sigma^2 - \sigma(h)$, dérivés de fonctions de covariance, qui caractérisent les FASL).

La fonction aléatoire $Z(t) = \exp(Y(t))$, pour laquelle $Y(t)$ est gaussienne et localement stationnaire sur T , est dite *lognormale et localement stationnaire* (FALLS). Le variogramme d'une FALLS est toujours de la forme :

$$\Gamma(h) = B \left(1 - e^{-\gamma(h)} \right) \quad (5.26)$$

Par exemple, si le variogramme des logarithmes $Y(t)$ est linéaire de la forme $\gamma(h) = b|h|$, le variogramme correspondant des expositions sonores $Z(t)$ est le modèle exponentiel :

$$\Gamma(h) = B \left(1 - e^{-b|h|} \right) \quad (5.27)$$

où l'on voit que la pente b de $\gamma(h)$ devient l'inverse du paramètre de portée de $\Gamma(h)$.

Plus généralement, à un variogramme des logarithmes en puissance p de la forme (5.25) correspond le modèle dit *stable* :

$$\Gamma(h) = B \left(1 - e^{-b|h|^p} \right) \quad \text{avec} \quad 0 < p < 2 \quad (5.28)$$

Le variogramme d'une FALLS vérifie $\Gamma(h) \leq B$ pour tout h et présente donc un palier asymptotique et une portée pratique. Si le variogramme des expositions sonores exhibe expérimentalement une portée apparente, il est toujours intéressant de calculer le variogramme des logarithmes afin de vérifier s'il s'agit d'un effet de FALLS ou, si ce dernier présente également un palier, s'il convient plutôt de ranger le phénomène dans la classe plus étroite des FASL.

Ajustement de séries avec un modèle stable

Afin de voir s'il est possible de mener une analyse structurale lognormale dans l'esprit d'une FALLS nous avons utilisé un seul modèle en puissance p , selon la formule (5.25), sans effet de pépite, pour ajuster les variogrammes des logarithmes des expositions sonores. Ce modèle est ensuite injecté dans la formule (5.28) du modèle stable, pour lequel il ne reste alors plus qu'à déterminer le paramètre B lors de l'ajustement (effectué également sans effet de pépite). On a une économie substantielle de paramètres, puisque l'on n'utilise que deux paramètres pour le variogramme logarithmique, plus un paramètre pour le variogramme des expositions sonores.

Cet ajustement avec une seule structure se fait bien sur les trois séries **jointag1**, **regleur2** et **saranex1**. Les Figures 5.8, 5.9 et 5.10 montrent les ajustements et indiquent les valeurs des trois paramètres (puissance p , pente b ou *slope*, palier B ou *sill*). Sur ces figures, le variogramme expérimental au support de 1mn est représenté par des ronds. Si l'on ordonne ces trois séries par puissances p croissantes de l'ajustement : **regleur2** ($p = .05$), **saranex1** ($p = .1$), **jointag1** ($p = .2$), on a un classement des graphiques des variogrammes de ces trois séries par ordre de déviation croissante du modèle FASL.

Pour le variogramme logarithmique de la série **regleur2** (Figure 5.8) le palier expérimental concorde à peu près avec la variance expérimentale σ^{*2} et, au lieu de la structure non bornée (mais de coefficient p très faible) utilisée ici, l'on pourrait tout aussi bien adopter un modèle FASL en ajustant $\gamma(h)$ par un variogramme à palier. Le modèle stable correspondant a été ajusté avec un palier B (certes asymptotique) de 80% supérieur à la variance expérimentale D^{*2} des expositions sonores.

Le décrochement brutal du variogramme arithmétique expérimental de **regleur2** au bout de 140mn, est dû à une impulsion acoustique vers 17h30, soit un peu plus de deux heures avant la fin de la série. Ce décrochement n'existe pas sur le variogramme expérimental des logarithmes, vu que l'ampleur du pic acoustique exprimé en dB est atténué par la prise de logarithme. Si l'on regarde la série **regleur2**, représentée sur la Figure 5.7 à la fois en dB et en exposition sonore, le pic d'exposition sonore à 17h30 ne représente plus qu'une fluctuation d'un ordre de grandeur comparable à plusieurs autres dans la série exprimée en dB.

Le variogramme expérimental des logarithmes de **saranex1** (Figure 5.9) croît au-dessus de la variance expérimentale et a donc été ajusté avec un coefficient $p = .1$, ce qui se traduit au niveau du modèle stable par un palier d'environ 30% supérieur à la variance arithmétique expérimentale.

La série **jointag1** présente des variogrammes (Figure 5.10) pour lesquels le modèle FALLS semble bien adapté, car, à la fois pour les expositions sonores et pour leurs logarithmes, les variogrammes expérimentaux ne se stabilisent pas autour des valeurs des variances expérimentales, mais croissent au-delà. Une valeur de $p = .2$ a été adoptée pour l'ajustement de variogramme des logarithmes et on a trouvé un palier pour le modèle

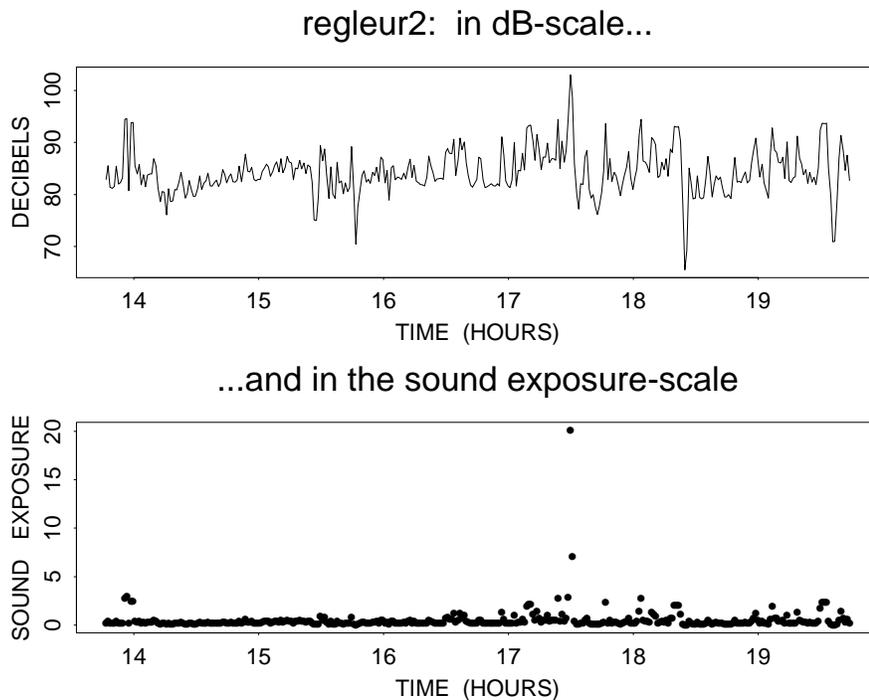


FIGURE 5.7 – La série **regleur2** exprimée en dB et en expositions sonores (au support de $\Delta t = 1\text{mn}$).

stable 3 fois supérieur à la variance expérimentale.

En conclusion, le modèle stable permet une modélisation simultanée des variogrammes des expositions sonores et de leurs logarithmes n'utilisant que trois paramètres. L'ajustement aux trois séries jointag1, regleur2 et saranex1 reproduit bien l'allure générale de la structure temporelle.

5.4 Applications pratiques

La modélisation de la structure temporelle des données d'exposition sonore peut être réalisée à l'aide d'approches très différentes. Les trois approches employées ici mettent bien en évidence que la préférence d'un modèle de variogramme parmi plusieurs est guidée par deux aspects majeurs. Ces deux aspects sont : a) le comportement à court terme du modèle, autrement dit sa capacité à reproduire la variabilité de la période observée, b) le comportement du modèle à long terme, au-delà des limites de la fenêtre d'observation.

Dans la première approche, celle du modèle de Wijsien, l'autocorrélation est entièrement caractérisée par le modèle lui-même, à un coefficient d'échelle près. Ce modèle est séduisant par sa simplicité et évite tout calcul de variogramme. De plus, il généralise le modèle de Larsen, qui est fréquemment employé en hygiène industrielle dans le domaine des polluants atmosphériques. Il était donc intéressant de tester son applicabilité à l'acoustique. Dans le cas des données analysées ici, cette approche ne permet pas de modéliser correctement la variabilité constatée durant les périodes observées. En outre, il a été montré qu'elle implique un variogramme non borné. Or cette hypothèse qui

a déjà été discutée (Chapitre 2) ne semble pas acceptable quand il s'agit de modéliser des données relatives à une situation d'exposition homogène. Cette approche est donc nettement inadaptée à la modélisation des données étudiées ici.

Dans les deux autres approches, le comportement à court terme des données est reproduit de façon acceptable par l'un et l'autre des modèles de variogramme. Les modèles employés, lognormal stationnaire et lognormal localement stationnaire, imposent un variogramme borné. De plus la qualité de la modélisation a été éprouvée par la prise en compte simultanée de deux effets : celui de 3 durées d'intégration dans la seconde approche, celui des valeurs exprimées en exposition sonore et après transformation logarithmique dans la troisième approche.

Examinons maintenant le comportement à long terme des 2 modèles qui ont été jugés acceptables par leur comportement à court terme. Si on prolonge les modèles au-delà de la fenêtre d'observation, on s'aperçoit que leur comportement est différent.

Le choix du modèle de variogramme (et donc celui de son comportement au delà du domaine observé) constitue en soi une hypothèse d'extrapolation. Alors que de nombreux modèles sont disponibles, ce choix constitue la seule hypothèse qu'utilise la géostatistique pour estimer l'exposition à long terme. L'incidence du choix du modèle de variogramme sera illustré de façon très éclairante dans le Chapitre 7 (p110).

Sur un plan pratique, pour les données étudiées ici, on retiendra que le choix d'une hypothèse portant sur le comportement à long terme de l'exposition sonore est le critère qui permettrait de préférer l'un ou l'autre des modèles lognormal stationnaire ou lognormal localement stationnaire. Ce choix peut rester arbitraire ou au contraire être validé par des données expérimentales complémentaires, collectées au-delà du domaine d'observation initial, mais il ne dépend pas de la méthode d'analyse statistique. En l'absence de données complémentaires sur l'exposition à long terme, les 2 modèles sont acceptables.

L'approche géostatistique de l'estimation de l'exposition à long terme renvoie donc à des choix qui relèvent de l'analyse du travail et des circonstances de l'exposition à long terme. Elle ne requiert pas d'hypothèse sur la distribution des valeurs mesurées et prend en compte simultanément l'autocorrélation des données et l'effet du temps d'intégration choisi lors du mesurage.

Le modèle géostatistique permet de prévoir aisément la variance de dispersion d'échantillons dans un domaine temporel, sur un support autre que celui utilisé pour les mesures. Les courbes de variance de dispersion décrivent la décroissance de la variance en fonction du temps d'intégration des mesures.

La structure temporelle des données qui apparaît au terme de cette analyse montre que ces données présentent une autocorrélation sur une petite échelle, de durée comprise entre 2mn et 5mn, et sur une grande échelle, de durée supérieure à une heure. Les données mesurées ne sont donc pas indépendantes et nous verrons au Chapitre 6 les conséquences de ce fait dans l'estimation du niveau d'exposition sonore quotidien.

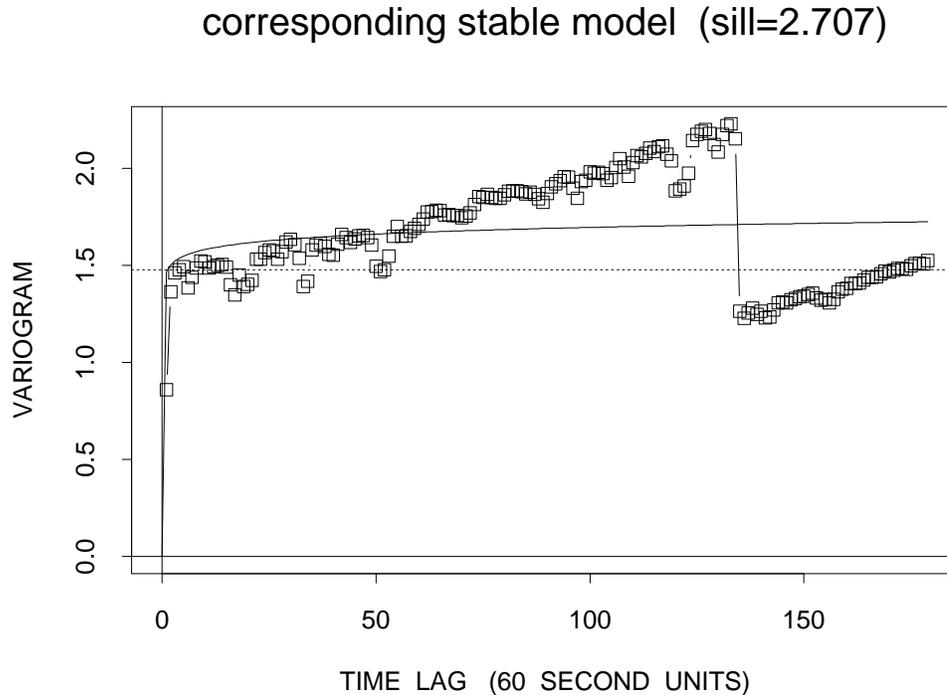
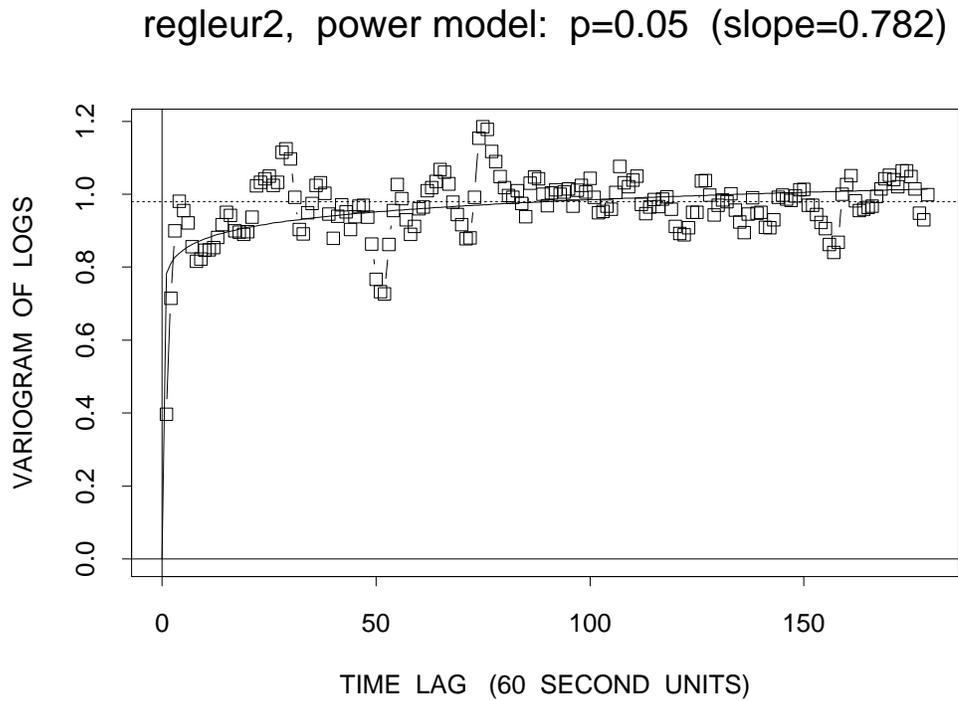
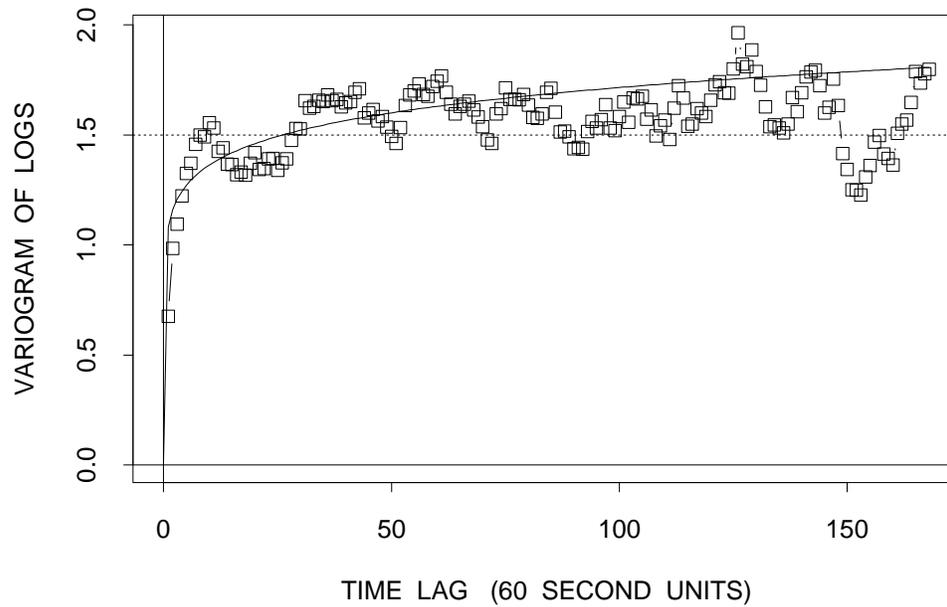


FIGURE 5.8 – Variogramme expérimentaux au pas (et support) de $\Delta t = 1$ minute, ajustés avec un modèle puissance (pour les logarithmes) et un modèle stable correspondant (pour les expositions sonores).

saranex1, power model: $p=0.1$ (slope=1.082)



corresponding stable model (sill=5.908)

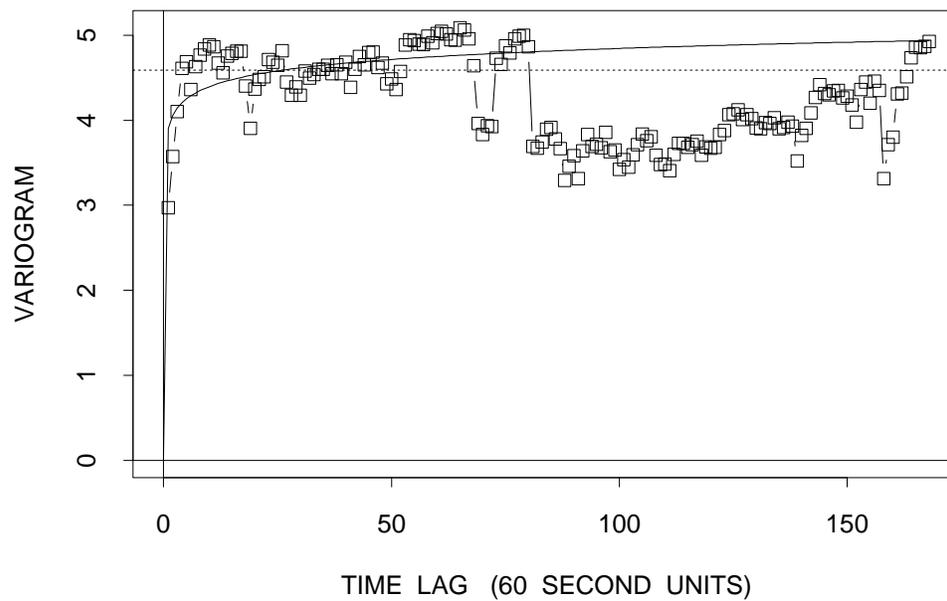


FIGURE 5.9 – Variogramme expérimentaux au pas (et support) de $\Delta t = 1$ minute, ajustés avec un modèle puissance (pour les logarithmes) et un modèle stable correspondant (pour les expositions sonores).

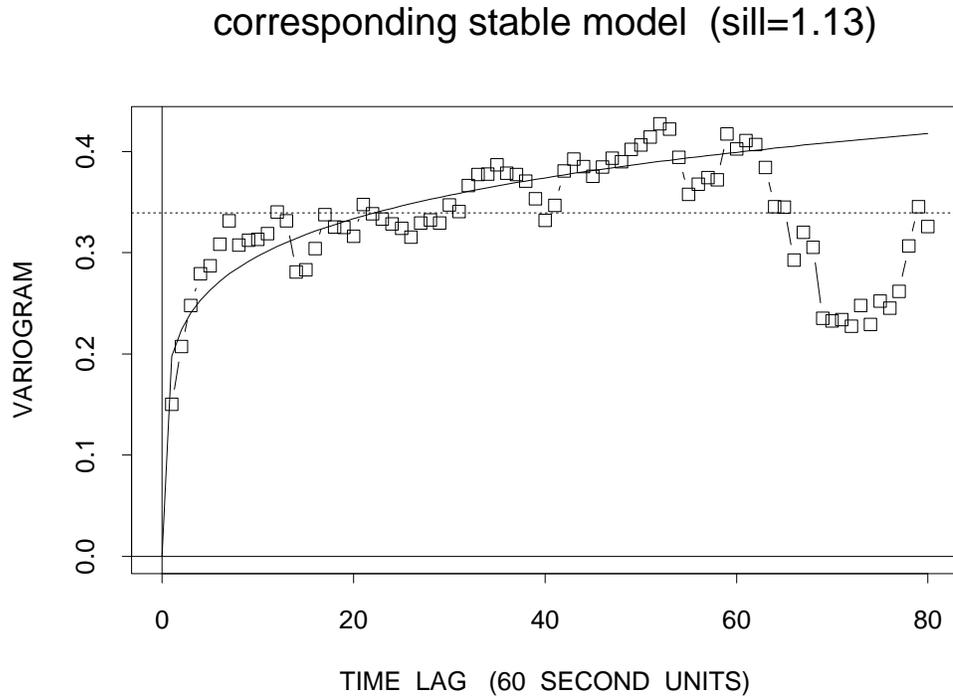
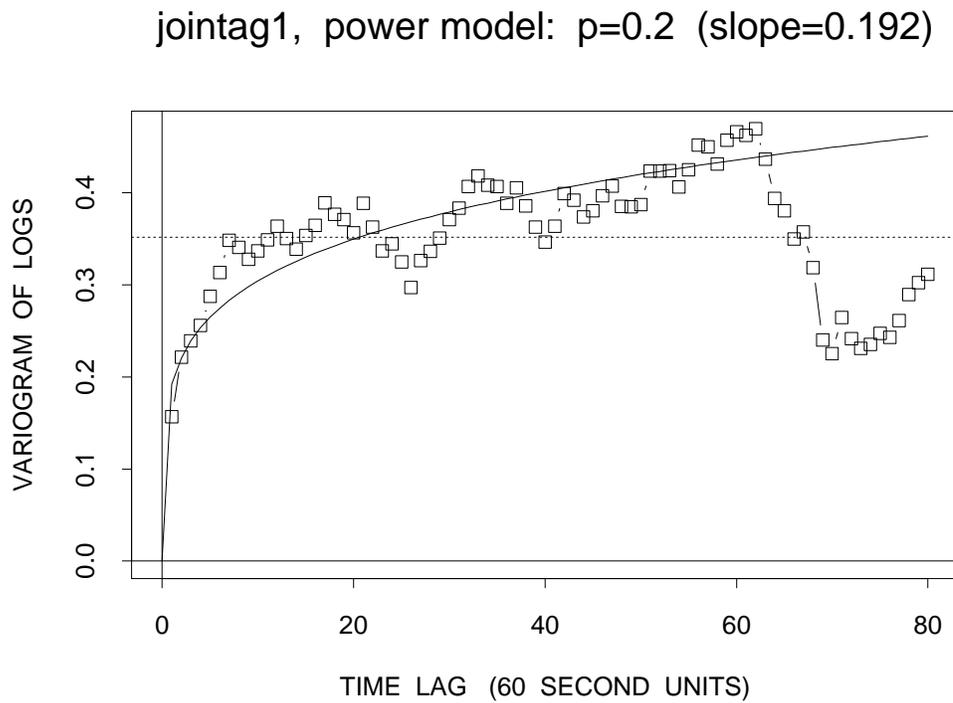


FIGURE 5.10 – Variogramme expérimentaux au pas (et support) de $\Delta t = 1$ minute, ajustés avec un modèle puissance (pour les logarithmes) et un modèle stable correspondant (pour les expositions sonores).

Chapitre 6

Estimation du L_{EQ}

Ce chapitre présente tout d’abord deux normes d’acoustique, qui emploient des approches statistiques fort contrastées de l’estimation du L_{EQ} . Après avoir montré dans les chapitres précédents l’intérêt d’une modélisation lognormale de l’exposition sonore, la théorie de l’estimation lognormale sous hypothèse d’indépendance des données est rappelée puis appliquée aux 3 séries de données d’exposition sonore déjà étudiées.

L’estimation lognormale est développée ensuite afin de prendre en compte l’autocorrélation des données. Après avoir montré les limites de la théorie classique dans le cas de données autocorrélées, on présente la méthode d’estimation géostatistique adaptée à des données de ce type, nommée krigeage lognormal, puis les estimateurs issus de cette méthode.

Les performances de ces estimateurs sont comparées à celles d’estimations classiques, normalisées. Finalement, on compare les largeurs des intervalles de confiance du L_{EQ} des 3 séries homogènes, estimées selon deux approches : a) quand on prend en compte l’autocorrélation, b) quand on fait l’hypothèse – non vérifiée ici – que les données sont indépendantes.

6.1 Analyse de deux textes normatifs

Nous présentons deux textes normatifs qui ont été choisis en fonction des approches radicalement différentes du calcul de la variance d’estimation du L_{EQ} . La variance d’estimation est utilisée pour construire des intervalles de confiance pour l’estimateur du L_{EQ} . Dans les deux cas le L_{EQ} est estimé à partir de la moyenne arithmétique de données d’exposition sonore.

La norme NF S 31-084

Une méthode statistique (inspirée de l’approche préconisée dans un article de BASTIDE [4]) est décrite dans l’Annexe B (p13) de la norme NF S 31-084 [2]. Elle permet “d’estimer, à partir d’un calcul réalisé sur *un nombre limité d’échantillons prélevés au hasard*, au cours d’un intervalle de temps T , la valeur probable du niveau de pression acoustique continu équivalent pondéré A d’une ambiance sonore sur cet intervalle de temps, ainsi que *l’intervalle de confiance* autour de cette valeur”. Le but est de “diminuer le coût de chaque intervention *en limitant la durée de mesurage* et en utilisant du matériel simple d’emploi”.

Les échantillons L_α ($\alpha = 1, \dots, n$) sont, soit des niveaux “instantanés mesurés avec la caractéristique temporelle du sonomètre”, soit des niveaux “continus équivalents mesurés sur un court intervalle de temps Δt compris, par exemple, entre 1mn et 15mn”.

Le niveau de pression acoustique continu équivalent moyen et l'écart-type sont calculés selon deux principes différents :

- Le niveau de pression acoustique continu équivalent $L_{Aeq,T}$ est estimé en formant la moyenne des mesures transformées en expositions sonores, puis en prenant le logarithme de cette moyenne, afin de l'exprimer en dB :

$$L_{Aeq,T}^* = 10 \log_{10} \left(\frac{1}{n} \sum_{\alpha=1}^n 10^{L_\alpha/10} \right) \quad (6.1)$$

- L'écart-type σ_L^* des niveaux L_α , par contre, n'est pas calculé sur la base des expositions sonores, mais directement sur les données exprimées en dB :

$$\sigma^* = \sqrt{\frac{1}{n-1} \sum_{\alpha=1}^n (L_\alpha - m^*)^2} \quad (6.2)$$

où intervient la moyenne m^* des niveaux en dB :

$$m^* = \frac{1}{n} \sum_{\alpha=1}^n L_\alpha \quad (6.3)$$

On suppose ensuite que la variable

$$\frac{m^* - m}{\sigma^*} \sqrt{n-1} \quad (6.4)$$

suit une loi de Student à $n-1$ degrés de liberté (ce qui est équivalent à dire que les niveaux de pression L_α sont des variables gaussiennes indépendantes) pour construire un intervalle de confiance bilatéral dans lequel la moyenne inconnue m se trouve avec une probabilité de 95%.

Les deux principes sont alors imbriqués en appliquant l'intervalle de confiance calculé pour m à $L_{Aeq,T}$. On identifie ainsi m à $L_{Aeq,T}$, bien que les valeurs numériques des deux quantités et leur signification physique, de surcroît, ne soient pas les mêmes.

Le premier principe est de nature physique, du type “seules les expositions sonores sont additives”. Le deuxième principe repose sur un point de vue statistique du genre “les expositions sonores ont un histogramme beaucoup plus penché à gauche que les mesures exprimées en dB” qui se traduit par une hypothèse sur la lognormalité des expositions sonores, puisque l'on choisit d'appliquer la méthode de Student au logarithme des expositions sonores. C'est naturellement contradictoire avec le fait que la moyenne à laquelle on se réfère en fin de compte n'est pas celles des logarithmes, mais celle des expositions sonores elle-mêmes.

Exemple numérique

L'exemple numérique suivant est donné en p14 de la norme [2] et est basé sur dix mesures exprimées en dB :

$$84, 86, 87.5, 88.5, 90, 91, 92, 93, 95, 97 \quad (6.5)$$

Le estimations sont les suivantes :

$$L_{Aeq,T}^* = 92.02 \quad (6.6)$$

$$m^* = 90.4 \quad (6.7)$$

$$\sigma^* = 4.04 \quad (6.8)$$

$$P(87.51 < m < 93.29) = 95\% \quad (6.9)$$

Par la soustraction de la moyenne estimée de la borne supérieure de l'intervalle, (la loi de Student étant symétrique) on trouve la demi-longueur de l'intervalle :

$$93.29 - m^* = 2.89 \quad (6.10)$$

et on utilise ce chiffre pour définir une fourchette (purement) conventionnelle pour le niveau continu équivalent :

$$L_{Aeq,T}^* \pm 2.89 \quad (6.11)$$

On notera au passage que la moyenne m^* utilisée pour construire l'intervalle de confiance est inférieure de 1.6 dB à la moyenne $L_{Aeq,T}^*$ vers laquelle cet intervalle a été translaté (sans la moindre justification).

La directive VDI 3723 (Blatt 1)

La directive VDI 3723 décrit les méthodes statistiques à employer, en Allemagne, pour la caractérisation des expositions sonores fluctuantes. Elle définit le niveau de pression acoustique continu équivalent de la même manière que la norme NF S 31-084 décrite ci-dessus :

$$L_{Aeq,T}^* = 10 \log_{10} M^* \quad (6.12)$$

où M^* est la moyenne des expositions sonores :

$$M^* = \frac{1}{n} \sum_{\alpha=1}^n 10^{L_{\alpha}/10} \quad (6.13)$$

L'écart-type de référence est par contre calculé sur les expositions sonores elles-mêmes (au lieu des valeurs en décibels) :

$$D^* = \sqrt{\frac{1}{n-1} \sum_{\alpha=1}^n (10^{L_{\alpha}/10} - M^*)^2} \quad (6.14)$$

On suppose (en travaillant à variance inconnue) que la variable :

$$\frac{M^* - M}{D^*} \sqrt{n-1} \quad (6.15)$$

est distribuée selon une loi de Student à $n - 1$ degrés de liberté, c'est-à-dire que les expositions sonores sont supposées être gaussiennes. Si τ_{n-1} est une borne pour un niveau de confiance donné, l'intervalle de confiance en dB s'écrit (équations 5 à 7 dans [33]) :

$$L_{Aeq,T}^* + \log_{10} \left(1 \pm \tau_{n-1} \frac{D^*}{M^*} \right) \quad (6.16)$$

ce qui équivaut en termes d'expositions sonores à :

$$M^* \pm \tau_{n-1} D^* \quad (6.17)$$

Dans la directive VDI on fait tous les calculs sur les expositions sonores et l'on ne fait que convertir les résultats en décibels, ce qui apparaît comme plus cohérent tant du point physique que statistique. L'hypothèse de normalité faite sur les expositions sonores est cependant très critiquable : dans tous les exemples qui nous ont été soumis, les expositions sonores exhibent des histogrammes fortement penchés à gauche, de sorte que l'hypothèse de lognormalité de la norme NF est plus réaliste dans ces cas-là, que l'hypothèse de normalité proposée dans la directive VDI.

La directive VDI insiste à plusieurs reprises sur l'hypothèse d'indépendance des échantillons et préconise de les prélever à des intervalles de temps suffisamment grands pour éviter une corrélation temporelle entre ceux-ci.

Exemple numérique (suite)

Nous reprenons l'exemple numérique (6.5). Nous obtenons :

$$L_{Aeq,T}^* = 92.02 \quad (6.18)$$

$$P(87.25 < L_{Aeq,T} < 94.24) = 95\% \quad (6.19)$$

c'est-à-dire une fourchette conventionnelle :

$$L_{Aeq,T}^* \begin{matrix} +2.22 \\ -4.77 \end{matrix} \quad (6.20)$$

On aboutit au résultat (certes un peu troublant pour celui qui a une vision lognormale, ou même seulement logarithmique, des données acoustiques) que la borne supérieure de l'intervalle de confiance est deux fois plus proche de $L_{Aeq,T}^*$ que la borne inférieure.

Conclusions sur les deux textes normatifs

Les deux textes préconisent un estimateur du L_{EQ} basé sur la moyenne (arithmétique) des mesures d'exposition sonore et considèrent ces mesures comme des réalisations de variables aléatoires indépendantes de même moyenne M . Les deux textes cherchent à construire un intervalle de confiance autour du logarithme (en base 10) de M , c'est-à-dire le L_{EQ} , mais utilisent des approches très différentes qui conduisent à des résultats fort dissemblables.

La norme NF se sert de la variance d'estimation de la moyenne m^* des logarithmes népériens des expositions sonores et en déduit par simple transformation linéaire les bornes de l'intervalle de confiance de l'estimateur du L_{EQ} . Les intervalles de confiance de la norme NF sont ainsi basés sur une théorie *lognormale*, puisque l'on suppose que le logarithme des mesures d'exposition sonore suit une loi normale (gaussienne).

La directive VDI construit directement un intervalle de confiance pour la moyenne M des expositions sonores et effectue ensuite une transformation logarithmique des bornes pour les exprimer en dB. Les intervalles de confiance de la directive VDI sont basés sur une théorie *normale*, puisqu'elle fait l'hypothèse que les mesures d'exposition sonore obéissent à une loi normale.

Les deux textes font une hypothèse d'indépendance des mesures. Or, par le calcul de variogrammes on peut aisément se convaincre que cette hypothèse est fautive dans le cas des séries chronologiques analysées ici. On verra par la suite (section 6.5, p70) que la dispersion de la variance d'estimation est forte lorsqu'il y a ne serait-ce qu'un peu d'autocorrélation dans les données. Les intervalles de confiance deviennent ainsi très peu fiables lorsque l'hypothèse d'indépendance n'est pas applicable et l'on est contraint d'en abandonner l'usage.

En ce qui concerne le choix entre une théorie normale ou lognormale des expositions sonores, les histogrammes des valeurs mesurées exprimées en exposition sonore (présentés Chapitre 3, Figure 3.4, p17) amènent à préférer sans hésitation la théorie lognormale.

Se pose alors la question de savoir si l'on ne peut pas définir de meilleurs estimateurs du L_{EQ} que celui obtenu à partir de la simple moyenne arithmétique des expositions sonores. Il nous sera permis de répondre par l'affirmative dans le cas des trois séries `jointag1`, `regleur2` et `saranex1`.

6.2 Méthode de comparaison des estimateurs

Dans la suite de ce chapitre nous présentons différents estimateurs du L_{EQ} . Les performances de chaque estimateur seront étudiées sur l'exemple des trois séries `jointag1`, `regleur2` et `saranex1`.

On procédera par rééchantillonnage. On fera comme si 1000 préleveurs pouvaient prendre chacun, indépendamment les uns des autres, un lot de 36 échantillons prélevés au hasard dans une série de mesures donnée. L'estimateur auquel on s'intéresse est alors appliqué à chacun des 1000 lots, ce qui permet d'étudier la dispersion des valeurs estimées. En particulier, l'écart-type des 1000 lots d'un estimateur donné sera comparée à l'écart-type du L_{EQ} classique. Plus cet écart-type d'estimation sera faible, plus l'estimation sera en moyenne proche de la vraie valeur. On notera que, suivant l'objectif d'estimation que se fixe la méthode employée, la vraie valeur sera la moyenne de la journée ou la moyenne à long terme.

6.3 Estimation lognormale sous hypothèse d'indépendance

Nous appelons *moyenne logarithmique* la moyenne m des logarithmes des expositions sonores. Voici les formules classiques d'estimation de la moyenne et de la variance de variables normales dans ce contexte :

$$m^* = \frac{1}{n} \sum_{\alpha=1}^n \log Z_{\alpha} \quad (6.21)$$

$$\sigma^{*2} = \frac{1}{n-1} \sum_{\alpha=1}^n (\log Z_{\alpha} - m^*)^2 \quad (6.22)$$

Les formules classiques des variances d'estimation relatives à ces deux estimateurs sont, dans le cas où il y a indépendance :

$$\text{var}(m^*) = \frac{\sigma^2}{n} \quad (6.23)$$

$$\text{var}(\sigma^{*2}) = \frac{2\sigma^4}{n-1} \quad (6.24)$$

On peut distinguer entre un *estimateur lognormal* et un *estimateur arithmétique* de la moyenne M des expositions sonores Z .

L'estimateur lognormal de la moyenne est construit en se servant des estimateurs de la moyenne et de la variance logarithmique :

$$M_L^* = e^{m^* + \sigma^{*2}/2} \quad (6.25)$$

La variance logarithmique de l'estimateur lognormal de la moyenne est :

$$\sigma^2 [\log M_L^*] = \frac{\sigma^2}{n} + \frac{\sigma^4}{2(n-1)} \quad (6.26)$$

tandis que sa variance arithmétique vaut :

$$D^2 [M_L^*] = M^2 \left(e^{\sigma^2 [\log M_L^*]} - 1 \right) \quad (6.27)$$

$$\cong \frac{M^2}{n} \left(\sigma^2 + \frac{\sigma^4}{2} \right) \quad (6.28)$$

L'estimateur arithmétique consiste simplement à prendre la moyenne arithmétique des expositions sonores :

$$M_A^* = \frac{1}{n} \sum_{\alpha=1}^n Z_\alpha \quad (6.29)$$

et il a pour variance :

$$D^2 [M_A^*] = \frac{M^2}{n} (e^{\sigma^2} - 1) \quad (6.30)$$

$$= \frac{M^2}{n} \left(\sigma^2 + \frac{\sigma^4}{2} + \frac{\sigma^6}{3!} + \frac{\sigma^8}{4!} + \dots \right) \quad (6.31)$$

Il s'avère donc que l'estimateur lognormal a une variance (6.28) qui est toujours *inférieure* à celle (6.31) de l'estimateur arithmétique de M . L'estimateur classique des normes acoustiques :

$$L_{Aeq}^* = \frac{b + \log M_A^*}{a} \quad (6.32)$$

apparaît donc comme *moins optimal* qu'un estimateur analogue basé sur l'estimateur lognormal M_L^* .

L'estimateur lognormal (6.25) est cependant entaché d'un biais :

$$E[M_L^*] = M_L \exp \left(\frac{\sigma^2}{2n} + \frac{\sigma^4}{4(n-1)} \right) \quad (6.33)$$

Le facteur de biais est d'autant plus voisin de 1 que n est grand et devient donc négligeable pour un grand nombre d'échantillons. Pour un nombre d'échantillons de l'ordre de plusieurs dizaines, MATHERON ([17], p273) suggère une correction du premier ordre :

$$M_{L1}^* = \nu^* \exp \left(\frac{n-1}{n} \frac{\sigma^{*2}}{2} - \frac{\sigma^{*4}}{4(n-1)} \right) \quad (6.34)$$

où ν^* est la médiane des expositions sonores.

Signalons que par ailleurs SICHEL a proposé des estimateurs lognormaux sans biais dans ses travaux [29] [5].

Estimateur	classique			lognormal		
	36	18	6	36	18	6
jointag1	.45	.66	1.25	.44	.65	1.22
regleur2	1.46	1.74	2.48	1.02	1.41	2.06
saranex1	1.45	1.95	2.93	1.24	1.68	2.33

TABLE 6.1 – Écarts-types (en dB) du L_{Aeq}^* au support de 1mn pour l'estimateur classique et lognormal.

Estimateur	classique			lognormal		
	Support	2mn	1mn	10s	2mn	1mn
jointag1	.34	.45	.61	.34	.44	.59
regleur2	1.18	1.46	1.68	.86	1.02	1.22
saranex1	1.04	1.45	1.86	.94	1.24	1.52

TABLE 6.2 – Écarts-types (en dB) du L_{Aeq}^* avec 36 échantillons pour l'estimateur classique et lognormal.

Application aux trois séries

Nous allons comparer les performances de l'estimateur classique, formule (6.32), avec celles d'un estimateur obtenu à partir de la moyenne lognormale avec correction de biais au premier ordre :

$$L_{Aeq}^* = \frac{b + \log M_{L1}^*}{a} \quad (6.35)$$

Sur chaque série on a répété 1000 fois un échantillonnage avec 36, 18 ou 6 mesures distinctes prises en des temps aléatoires. Sur chacun des 1000 lots de mesures ainsi obtenus, on a calculé la valeur des deux estimateurs (6.32) et (6.35). A titre d'exemple, les histogrammes correspondants pour la série *regleur2* sur un support de 1mn sont montrés sur la Figure 6.1 (p69).

Le Tableau 6.1 résume les écarts-types d'estimation (en dB) que l'on trouve reportés (entre parenthèses) au-dessus de chacun des histogrammes. L'estimateur lognormal est nettement plus performant que l'estimateur arithmétique pour *regleur2* et *saranex1*, alors qu'il n'apporte pas d'amélioration sensible pour *jointag1*.

L'histogramme de l'estimateur classique de *regleur2* (Figure 6.1, p 69) est bimodal et exhibe un petit second mode centré sur 90, 92 ou 96 dB (suivant que l'on prend 36, 18 ou 6 échantillons), qui montre la sensibilité de la moyenne arithmétique à la présence du pic sonore vers 17h30. L'estimateur lognormal s'avère bien plus robuste. Cela veut dire pour la pratique, que le préleveur diagnostiquera par exemple un $L_{EQ} > 90$ dB pour une série de type *regleur2* dans environ 4% de cas, s'il utilise l'estimateur classique au lieu de l'estimateur lognormal, avec lequel il aurait généralement obtenu un $L_{EQ} < 89$ dB et jamais en dessus de 90dB.

Le Tableau 6.2 présente une comparaison des écarts-types d'estimation (en dB) pour

les trois supports 10s, 1mn, 2mn, sur la base de 36 échantillons. On remarque que la meilleure performance de l'estimateur lognormal devient plus marquée lorsque le support est plus fin.

6.4 L_{EQ} arithmétique ou lognormal ?

Pour des données conformes au modèle lognormal, la question du choix d'un estimateur arithmétique ou lognormal dépend aussi de la densité de l'échantillonnage et de l'intensité de l'autocorrélation.

Or les normes de l'acoustique présentées ci-dessus calculent l'exposition sonore moyenne en appliquant l'estimateur arithmétique, quel que soit la stratégie d'échantillonnage de l'exposition.

Dans ce chapitre, deux stratégies d'échantillonnage sont évoquées :

- L'échantillonnage régulier continu, qui correspond à la pratique du mesurage de séries continues de $L_{Aeq,dT}$ "court". Dans ce cas l'intervalle de temps séparant deux observations successives est égal au temps d'intégration dT ou support.
- L'échantillonnage régulier fragmentaire, qui consisterait, par exemple, à mesurer le bruit durant les 5 premières minutes de chaque heure en renouvelant ce mesurage d'heure en heure. Dans ce cas, l'intervalle de temps séparant deux observations successives est un multiple du temps d'intégration dT .

Ces deux stratégies concernent l'échantillonnage réalisé avec un pas régulier. Mais d'autres stratégies d'échantillonnage sont employées ailleurs, notamment l'échantillonnage aléatoire, souvent préconisé et employé en hygiène industrielle (voir le Chapitre 7 et l'Annexe portant sur l'optimisation de l'échantillonnage).

Echantillonnage régulier continu

Dans un échantillonnage régulier continu d'une journée de travail T partitionnée en intervalles de mesure Δt , la moyenne des valeurs des expositions sonores $z_{\Delta t}(t_\alpha)$ nous donne sans équivoque la valeur de la moyenne journalière :

$$\frac{1}{n} \sum_{\alpha=1}^n z_{\Delta t}(t_\alpha) = \frac{1}{T} \int_0^T z(t) dt = z_T \quad (6.36)$$

Il n'y a ici pas de problème d'estimation et il n'y a ainsi pas de place pour une compétition entre l'estimateur arithmétique et l'estimateur lognormal de la moyenne arithmétique journalière. La divergence éventuelle des deux estimateurs, comme le montre un raisonnement de MATHERON, ne nous donnerait guère qu'un indice d'écart à la normalité. Pour un échantillonnage continu l'estimateur L_{EQ} des normes acoustiques, au sens d'une moyenne journalière, est donc parfaitement adéquat même dans le cadre d'un modèle lognormal.

Echantillonnage régulier fragmentaire

Partitionnons une journée T en n intervalles réguliers T' et plaçons un échantillon de support Δt au centre de chaque intervalle T' . Sur la base d'un modèle simple, MATHERON ([17], p319) aboutit à la constatation que si la variance d'extension $\sigma_E^2(\Delta t, T')$ de l'échantillon à l'intervalle le contenant est quelque peu significative, l'estimateur lognormal est

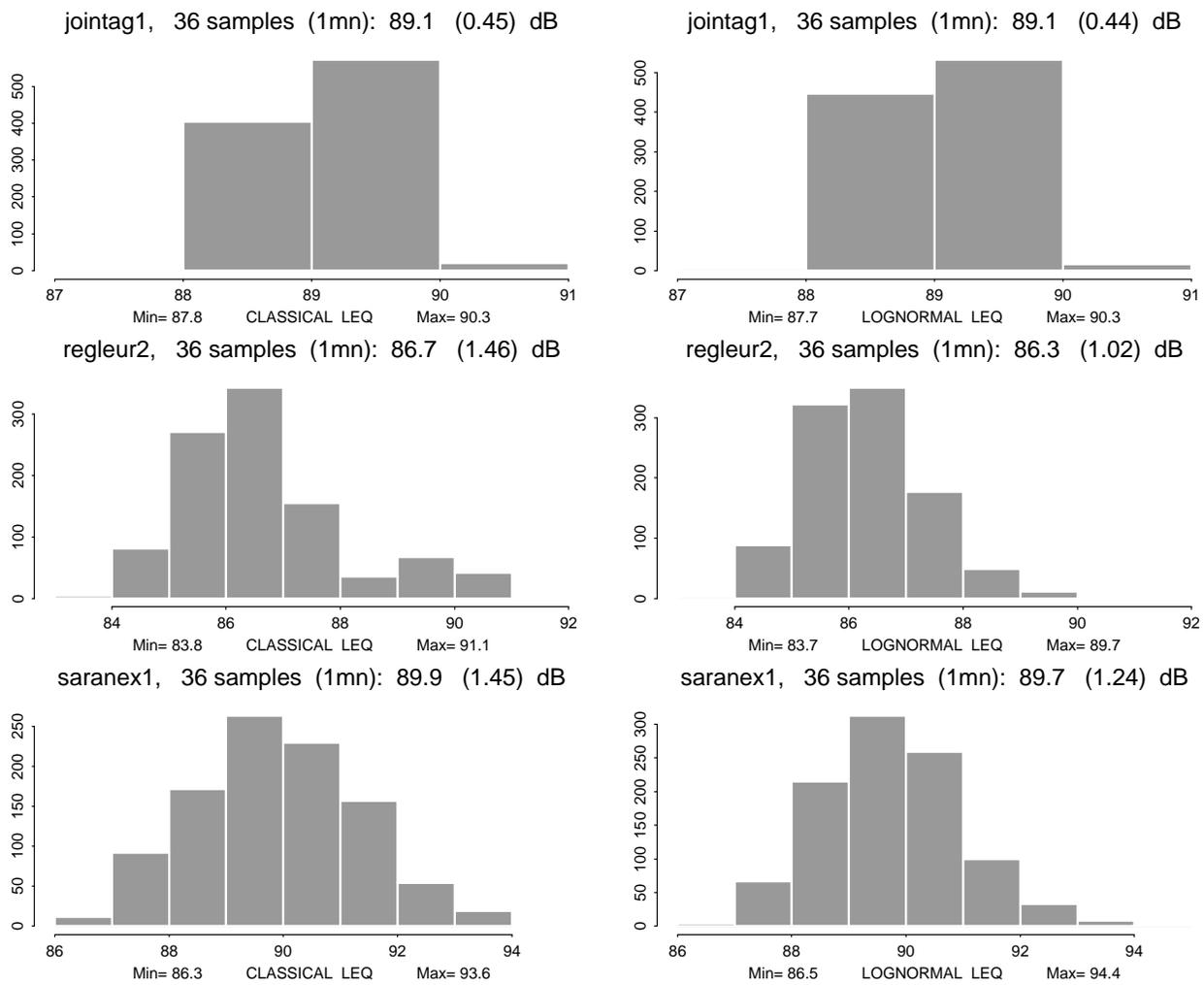


FIGURE 6.1 – Estimation des L_{EQ} de jointag1, regleur2 et saranex1 à partir d’une moyenne arithmétique ou d’une moyenne lognormale des expositions sonores, sur la base de 1000 tirages de 36 échantillons (support 1mn). La largeur des classes d’histogramme est de 1dB.

supérieur à l'estimateur arithmétique de l'exposition sonore moyenne. Dans le cas d'un échantillonnage régulier fragmentaire, le calcul du L_{EQ} global que préconisent les normes de mesure actuelles n'est donc pas optimum. Il faut lui préférer un estimateur lognormal.

En outre, pour des données autocorrélées, la variance d'estimation est d'autant plus grande que le support Δt est petit par rapport au pas d'échantillonnage T' et l'estimateur lognormal du L_{EQ} devient d'autant plus intéressant à utiliser.

6.5 Estimation en présence d'autocorrélation

L'abandon de l'hypothèse d'indépendance rend la formulation d'estimateurs lognormaux et des variances d'estimation associées extrêmement complexe.

L'estimateur de la moyenne logarithmique est le même que dans le cas de l'indépendance :

$$m^* = \frac{1}{n} \sum_{\alpha=1}^n \log Z_{\alpha}$$

L'espérance de l'estimateur est égale à la vraie valeur (non biais) :

$$E[m^*] = m \quad (6.37)$$

tandis que sa variance s'exprime en terme de covariances entre les $\log Z_{\alpha}$:

$$\text{var}(m^*) = \frac{1}{n^2} \sum_{\alpha=1}^n \sum_{\beta=1}^n \text{cov}(\log Z_{\alpha}, \log Z_{\beta}) \quad (6.38)$$

En dissociant dans cette expression les variances des covariances, on peut aisément la comparer à l'expression (6.23) du cas de l'indépendance :

$$\text{var}(m^*) = \frac{\sigma^2}{n} + \frac{1}{n^2} \sum_{\alpha \neq \beta} \sigma_{\alpha\beta} \quad (6.39)$$

On prend comme estimateur de la variance logarithmique :

$$\sigma^{*2} = \frac{1}{n} \sum_{\alpha=1}^n (\log Z_{\alpha} - m^*)^2 \quad (6.40)$$

$$= \frac{1}{n} \sum_{\alpha=1}^n (\log Z_{\alpha})^2 - (m^*)^2 \quad (6.41)$$

Par rapport au cas de l'indépendance, le dénominateur $n-1$ est remplacé par n , car l'utilisation de $n-1$ en présence d'autocorrélation ne garantit pas un estimateur non biaisé et ne se justifie donc pas.

La variance de cet estimateur est :

$$\begin{aligned} \text{var}(\sigma^{*2}) &= \frac{2}{n^2} \sum_{\alpha=1}^n \sum_{\beta=1}^n (\sigma_{\alpha\beta})^2 - \frac{4}{n^3} \sum_{\alpha=1}^n \left(\sum_{\beta=1}^n \sigma_{\alpha\beta} \right)^2 \\ &\quad + \frac{2}{n^4} \left(\sum_{\alpha=1}^n \sum_{\beta=1}^n \sigma_{\alpha\beta} \right)^2 \end{aligned} \quad (6.42)$$

Dans le cas de l'indépendance, seules les variances sont non nulles et l'on retombe sur l'expression (6.24) :

$$\text{var}(\sigma^{*2}) = \frac{2}{n}\sigma^4 - \frac{4}{n^2}\sigma^4 + \frac{2}{n^2}\sigma^4 \quad (6.43)$$

$$= \frac{2(n-1)}{n^2}\sigma^4 \quad (6.44)$$

En comparant (6.42) avec (6.43) on remarque que dans les deux cas le premier terme est largement prépondérant. Dans le cas de données (même faiblement) corrélées, ce premier terme sera aisément bien supérieur à celui du cas de l'indépendance, puisqu'aux variances viennent s'ajouter les carrés de toutes les covariances, autant positives que négatives.

La variance de l'estimateur de la variance logarithmique étant en général élevée dans le cas de données corrélées, l'élaboration d'un estimateur incorporant cette variance logarithmique, tel l'estimateur lognormal de Sichel, se présente sous de mauvais auspices. MATHERON [17] se borne donc à examiner des cas particuliers en employant des hypothèses d'approximation.

De même, le calcul d'intervalles d'un intervalle de confiance sera peu fiable, puisque la dispersion de la variance d'estimation est en général élevée dès qu'il y a un peu d'autocorrélation.

6.6 Krigeage lognormal de la moyenne de la journée

Le mot *krigeage*, dérivé du nom d'un pionnier de la géostatistique, D.G. KRIGE, concerne la méthode géostatistique d'estimation qui prend en compte l'autocorrélation des données et l'effet de support. Lorsque les données sont conformes au modèle lognormal on parle de *krigeage lognormal*.

Il existe plusieurs manières de formuler un krigeage lognormal de la valeur moyenne Z_T sur un domaine T , suivant la manière de formuler le critère (variance d'estimation) à optimiser [20][26]. Nous ne développerons ici que la forme qui était la plus simple à implémenter.

Dans un premier temps on procède à un *krigeage ordinaire* (KO) de valeurs de $Y^*(t_k)$ en un certain nombre N_T de points t_k de discrétisation de l'intervalle T . Puis on transforme chacune de ces valeurs logarithmiques en expositions sonores, dont on fait la moyenne sur T .

L'estimateur d'une valeur logarithmique s'écrit :

$$Y^*(t_k) = \sum_{\alpha=1}^n w_{\alpha}(t_k) Y(t_{\alpha}) \quad (6.45)$$

Les pondérateurs $w_{\alpha}(t_k)$ sont, à t_k fixé, solution d'un système :

$$\begin{cases} \sum_{\beta=1}^n w_{\beta}(t_k) \gamma(t_{\alpha}-t_{\beta}) + \mu(t_k) & = \gamma(t_{\alpha}-t_k) & (\alpha=1, \dots, n) \\ \sum_{\beta=1}^n w_{\beta}(t_k) & = & 1 \end{cases} \quad (6.46)$$

et la variance (logarithmique) d'estimation associée est :

$$\sigma_{KO}^2(t_k) = 2 \sum_{\alpha=1}^n w_{\alpha}(t_k) \gamma(t_{\alpha} - t_k) - \sum_{\alpha=1}^n \sum_{\beta=1}^n w_{\alpha}(t_k) w_{\beta}(t_k) \gamma(t_{\alpha} - t_{\beta}) \quad (6.47)$$

$$= \sum_{\alpha=1}^n \sum_{\beta=1}^n w_{\alpha}(t_k) w_{\beta}(t_k) \gamma(t_{\alpha} - t_{\beta}) + 2 \mu(t_k) \quad (6.48)$$

L'estimateur lognormal en chaque point de discrétisation est :

$$Z^*(t_k) = \exp(Y^*(t_k) - \mu(t_k) + \sigma_{KO}^2(t_k)/2) \quad (6.49)$$

Pour l'estimateur de la moyenne journalière il vient finalement :

$$Z_T^* = \frac{1}{N_T} \sum_{k=1}^{N_T} Z^*(t_k) \quad (6.50)$$

6.7 Krigeage lognormal de la moyenne à long terme

Dans le cadre du modèle de FALLS (dont il a été question au Chap. 5, p53) nous cherchons d'abord à estimer la moyenne locale m_0 du processus $Y(t)$ des logarithmes des expositions sonores, dont nous connaissons le variogramme $\gamma(h)$. L'estimateur est une moyenne pondérée des échantillons :

$$m_0^* = \sum_{\alpha=1}^n w_{\alpha} Y(t_{\alpha}) \quad (6.51)$$

où les poids w_{α} sont choisis de manière à minimiser la variance d'estimation sous contrainte de non biais. Ils sont solution du système de krigeage lognormal (LKM) :

$$\begin{cases} \sum_{\beta=1}^n w_{\beta}^{LKM} \gamma(t_{\alpha} - t_{\beta}) = \mu_{LKM} & (\alpha = 1, \dots, n) \\ \sum_{\beta=1}^n w_{\beta}^{LKM} = 1 \end{cases} \quad (6.52)$$

avec un paramètre de Lagrange μ_{LKM} .

L'estimateur *lognormal* de la moyenne (conditionnelle) du processus correspondant $Z(t)$ de expositions sonores s'écrit alors :

$$M_T^* = E[Z(t) | m_T^*] = e^{m^* + \mu_{LKM}/2} \quad (6.53)$$

Dans le cadre d'un modèle stationnaire (FASL, voir p53), c'est-à-dire avec des variogrammes à palier (comme dans le modèle exponentiel gigogne ajusté aux trois séries), l'estimateur de la moyenne locale (6.51) donne une estimation de la moyenne du processus stationnaire (aussi appelée *moyenne à long terme*). Le paramètre de Lagrange donne sous ces conditions la variance d'estimation de la moyenne à long terme :

$$\sigma_E^2 = \mu_{LKM} \quad (6.54)$$

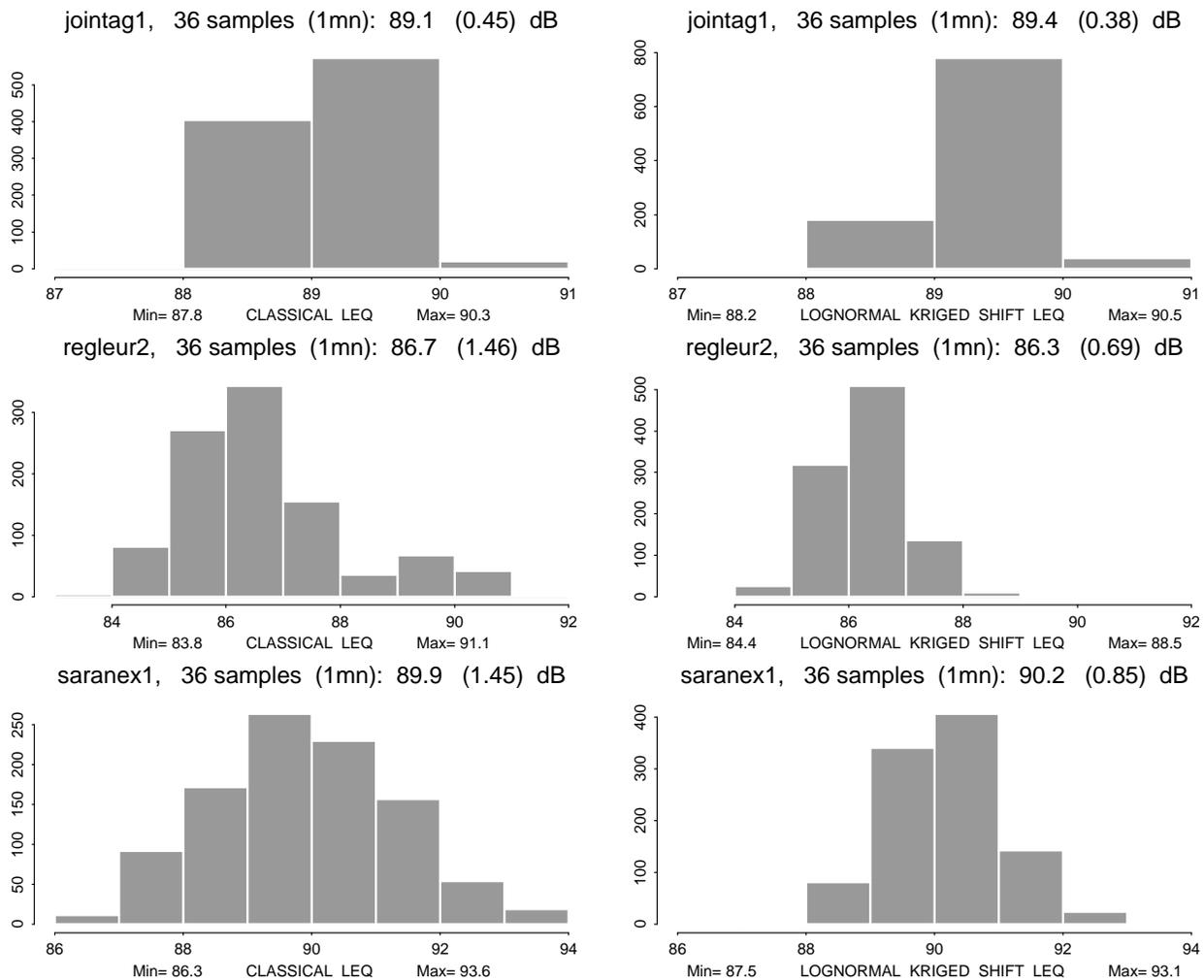


FIGURE 6.2 – Estimation des L_{EQ} de jointag1, regleur2 et saranex1 à partir d’une moyenne arithmétique ou d’un krigeage de la moyenne journalière (LKB), sur la base de 1000 tirages de 36 échantillons (support 1mn). La largeur des classes d’histogramme est de 1dB.

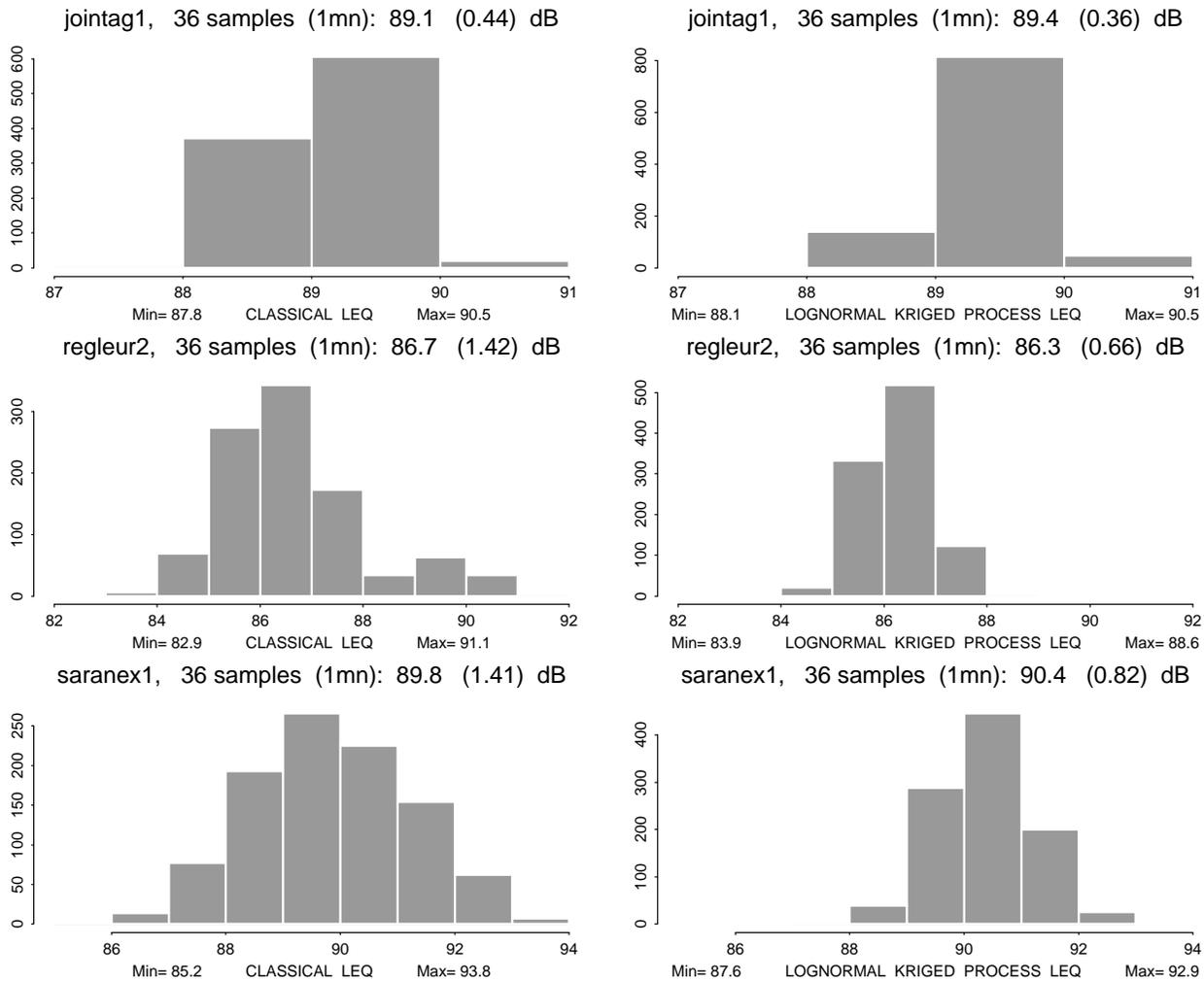


FIGURE 6.3 – Estimation des L_{EQ} de jointag1, regleur2 et saranex1 à partir d'une moyenne arithmétique ou d'un krigeage de la moyenne à long terme (LKM), sur la base de 1000 tirages de 36 échantillons (support 1mn). La largeur des classes d'historgramme est de 1dB.

Estimateur	jointag1	regleur2	saranex1
Classique	.45	1.46	1.45
LKB	.38	.69	.85

TABLE 6.3 – Ecart-types (en dB) de l'estimation du L_{EQ} avec 36 échantillons (tirés 1000 fois au hasard) en utilisant l'estimateur classique et le krigeage lognormal de la moyenne journalière Z_T (LKB).

qui nous permet de calculer une fourchette d'erreur pour la moyenne logarithmique à long terme, que l'on pourrait comparer à celle de la norme NF basée sur une hypothèse d'indépendance. On a montré par ailleurs [35], sur des mesures d'exposition au trichloréthylène, que les intervalles de confiance ainsi obtenus sont beaucoup plus larges que dans le cas de l'indépendance. Cela provient du fait que l'autocorrélation crée une certaine redondance des mesures, qui explique une plus grande incertitude dans l'estimation de la moyenne à long terme.

6.8 Krigeage lognormal des trois séries

On utilise l'ajustement des variogrammes expérimentaux des logarithmes des expositions sonores par un modèle exponentiel gigogne tel qu'il a été représenté sur la Figure 5.5(p51).

Krigeage de la moyenne journalière (LKB)

Sur la Figure 6.2 (p 73) on peut voir les histogrammes relatifs au krigeage de la moyenne journalière (sur 1000 tirages de 36 échantillons) en comparaison avec ceux de l'estimateur classique.

Les moyennes des estimations par ce krigeage ne diffèrent pas de plus de .3 ou .4 dB de la valeur de l'estimateur classique.

Le Tableau 6.3 montre que les estimations du LKB sont de 16% (jointag1), 53% (regleur2) et 41% (saranex1) meilleures que l'estimateur classique. Si l'on compare avec l'estimateur lognormal sous hypothèse d'indépendance (Tableau 6.1, p67) on remarque que le fait de tenir compte de l'autocorrélation des données conduit à des performances nettement améliorées de l'estimateur lognormal.

Krigeage de la moyenne à long terme (LKM)

La Figure 6.3 (p 74) montre les histogrammes relatifs au krigeage de la moyenne à long terme en comparaison avec l'estimateur classique, sur 1000 tirages de 36 échantillons pour chacune des trois séries.

L'estimateur classique donne en moyenne une valeur très proche de la moyennes exhaustive des séries complètes. La moyenne des estimations lognormales par LKM est proche (à moins d'un demi dB) de la valeur de l'estimateur classique.

Le Tableau 6.4 donne les écarts-types des histogrammes. Le LKM est bien plus performant que l'estimateur classique et que l'estimateur lognormal sous hypothèse d'indépen-

Estimateur	jointag1	regleur2	saranex1
Classique	.44	1.42	1.41
LKM	.36	.66	.82

TABLE 6.4 – Ecarts-types (en dB) de l'estimation du L_{EQ} avec 36 échantillons (tirés 1000 fois au hasard) en utilisant l'estimateur classique et par krigeage logarithmique de la moyenne (LKM).

Série	Leq	Borne supérieure	
		Indépendance	Autocorrélation
jointag1	89.2	90	94.6
regleur2	86.9	88.1	94.5
saranex1	90.1	91.7	101.5

TABLE 6.5 – L_{EQ} exhaustif et bornes supérieures (en dB) de la fourchette conventionnelle estimée avec 36 échantillons sous hypothèse d'indépendance (classique) ou en tenant compte de l'autocorrélation dans l'estimation de la moyenne à long terme (LKM).

dance (voir Tableau 6.1, p 67) pour les trois séries. Les résultats donnent des ordres de grandeur très proches du LKB, pour un effort de calcul bien moindre.¹

Note : En comparant les Tableaux 6.3 et 6.4 il faut prendre en considération le fait qu'on n'a pas utilisé les mêmes 1000 tirages de 36 échantillons : c'est pourquoi l'estimateur classique ne donne pas les mêmes valeurs dans les deux cas ; par contre, cela nous donne une idée de la (faible) fluctuation des résultats des simulations d'échantillonnage.

Estimation de la borne supérieure de l'intervalle 95%

Afin de quantifier un peu l'impact de l'autocorrélation sur l'estimation des bornes d'un intervalle de confiance, nous avons construit une estimation de la borne supérieure de la fourchette conventionnelle de la manière suivante (en nous inspirant de l'approche de la norme NF discutée en début de ce chapitre) :

- *Indépendance.* Nous calculons la variance logarithmique σ^2 de la série exhaustive et construisons la borne supérieure conventionnelle (à variance connue) en additionnant au L_{EQ} classique estimé avec 36 échantillons deux fois l'écart-type d'estimation σ/\sqrt{n} , où $n = 36$.
- *Autocorrélation.* Nous additionnons au L_{EQ} , estimé par krigeage lognormal de la moyenne à long terme, deux fois l'erreur d'estimation logarithmique convertie en dB, c'est-à-dire $\sigma_E = \sqrt{\mu_{LKM}}/a$, où $a = 10 \log(10)$.

Nous avons effectué les deux estimations sur 1000 jeux de 36 échantillons prélevés sur chacune des trois séries. La Figure 6.4 montre les histogrammes correspondants. Le

1. Pour le LKB, tel qu'il a été défini ici, il est nécessaire de faire un krigeage par point de discrétisation, tandis que le LKM se résume à un seul krigeage.

Tableau 6.5 donne, d'une part le L_{EQ} exhaustif de chaque série et la borne supérieure moyenne obtenue dans les 1000 simulations pour les deux estimations.

Sous hypothèse d'indépendance, l'estimation de la borne supérieure (à variance connue) nous donne une borne supérieure moyenne qui ne diffère du L_{EQ} exhaustif que de l'ordre de 1 à 2dB. Si l'on examine la dispersion de ces estimations sur la Figure 6.4 on s'aperçoit que la borne supérieure estimée est carrément inférieure dans un cinquième des cas pour `regleur2` et dans un dixième des cas pour `saranex1` au L_{EQ} exhaustif.

En se basant sur un modèle de l'autocorrélation, l'estimation de la borne supérieure de la fourchette conventionnelle pour la moyenne du processus nous donne en moyenne des bornes qui diffèrent de 5 à 11 dB de la valeur du L_{EQ} exhaustif correspondant (cf le Tableau 6.5). Vu que les 1000 estimations sont assez faiblement dispersées, comme on peut s'en convaincre sur la Figure 6.4, la borne supérieure estimée est toujours bien plus élevée que le L_{EQ} exhaustif et la fourchette conventionnelle le contient donc toujours.

Dans les exemples traités, l'hypothèse d'indépendance conduit donc à une fourchette conventionnelle trop étroite, qui ne contient pas la moyenne de la journée dans une proportion allant jusqu'à un cinquième des cas (alors que cela n'est sensé se produire théoriquement, pour la borne supérieure, que dans 2.5% des cas!). Le krigeage de la moyenne à long terme nous donne un intervalle bien plus large à cause de l'autocorrélation (redondance) des données, qui inclut toujours la moyenne journalière dans les trois exemples traités. Il faut bien comprendre que cet intervalle est sensé contenir également, avec une probabilité de 95%, la moyenne du processus. En particulier, la moyenne annuelle devrait être contenue dans cet intervalle avec au minimum la probabilité indiquée (si le modèle est encore valable à cette échelle de temps, bien sûr).

6.9 Applications pratiques

L'analyse des estimateurs employés dans 2 normes de mesurage de l'exposition sonore montre que des contradictions subsistent pour appliquer ces estimateurs à des données du type de celles qui sont analysées ici. Lorsqu'elles sont exprimées en terme d'exposition sonore (en pascal carrés seconde et non en décibels), les données ont des caractéristiques qui montrent qu'il faut éviter de supposer leur distribution normale. Il est préférable d'employer la théorie lognormale pour analyser ces données. En conséquence, des estimateurs lognormaux de l'exposition sonore moyenne et de sa variance ont été proposés, à la fois pour des données supposées indépendantes et pour des données autocorrélées.

Après avoir clarifié les bases théoriques des divers estimateurs utilisables, les performances des estimateurs ont été comparées par application aux 3 séries de mesures "homogènes". Cette comparaison confirme les résultats de l'approche théorique. En outre, une analyse détaillée des performances des estimateurs montre que les écarts-types d'estimations "classiques" sont plus détériorés dans le cas des séries `regleur2` et `saranex1` que dans le cas de `jointag1`. Ce résultat s'explique notamment par la typologie des distributions des 3 séries considérées (résumée au Chapitre 4 par les courbes $Q(T)$, au support de 1 mn) : la contribution des pointes de bruit dans l'exposition totale est nettement plus importante pour les séries `regleur2` et `saranex1` que pour `jointag1`. En d'autres termes, plus la contribution des pointes de bruit est prépondérante

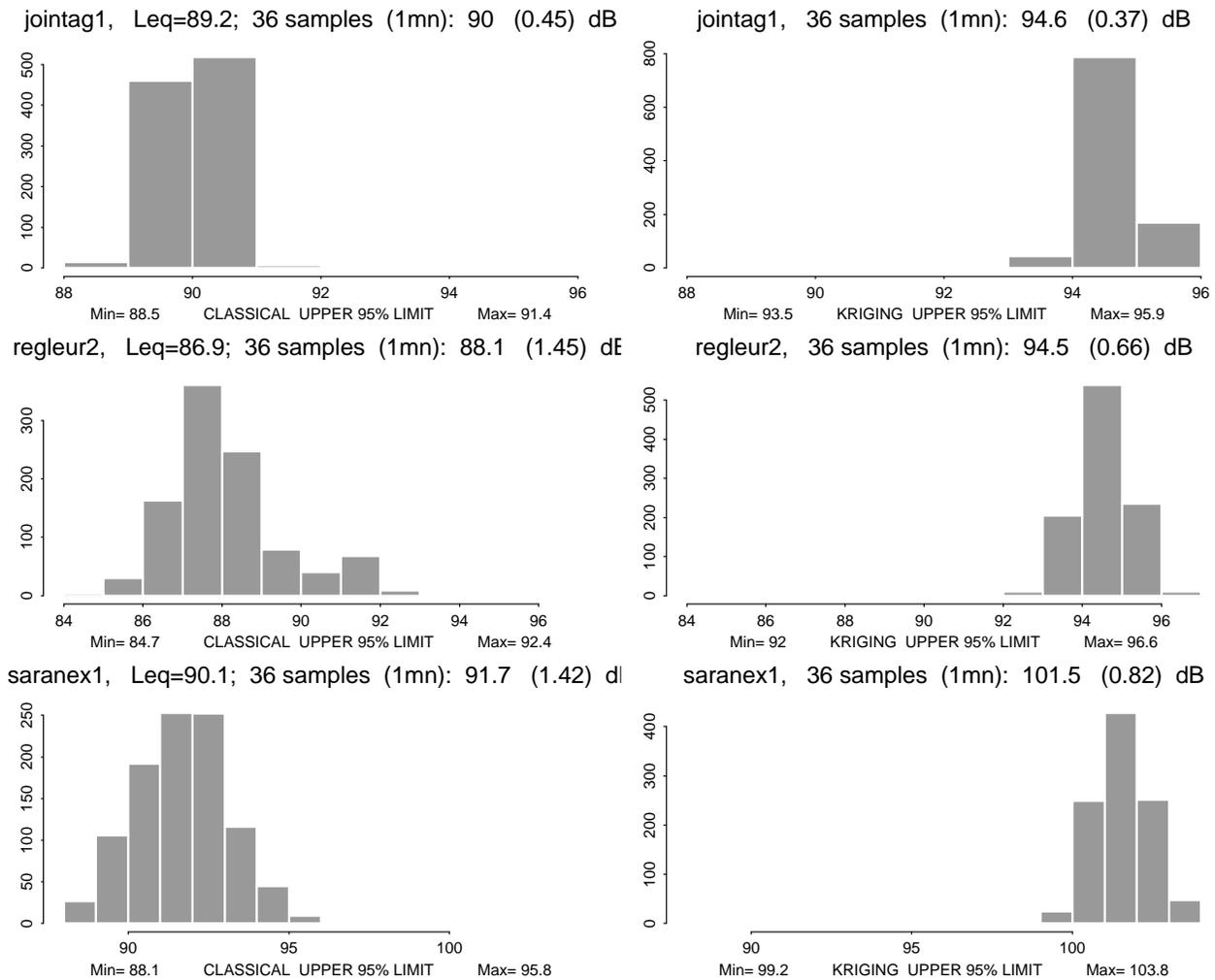


FIGURE 6.4 – Estimation de la borne supérieure de la fourchette conventionnelle. La largeur des classes d'histogramme est de 1dB.

dans l'exposition sonore totale, plus l'estimateur classique est à éviter.

Dans le cadre du modèle lognormal, il a été montré (section 6.4) que la méthode d'estimation du L_{EQ} dépend de la stratégie d'échantillonnage employée. Ceci concerne l'hypothèse d'indépendance des données et la répartition des échantillons.

La prise en compte de l'autocorrélation des données est source de difficultés car la fiabilité des estimateurs ne peut être assurée dans le cadre classique de l'estimation lognormale. Dans ce cas, la géostatistique propose une méthode d'estimation de l'exposition sonore moyenne et de sa variance, moyennant la connaissance du variogramme modélisé. Bien entendu l'effort de mesurage nécessaire à l'estimation du variogramme peut être important, mais c'est le même que celui qu'imposerait l'estimation d'une fonction d'autocorrélation.

L'impact de l'autocorrélation sur l'estimation des bornes d'un intervalle de confiance de l'exposition sonore moyenne est important dans le cas des séries de données analysées ici : l'exposition moyenne est majorée de 5 à 11 dB selon la série considérée. L'ampleur de ces majorations provient de l'autocorrélation à long terme des données (cf Chapitre 5, résultats du Tableau 5.1). Si la composante à court terme (de l'ordre de la minute) peut être bien estimée par des mesures de plusieurs heures, l'estimation de la composante à long terme (de l'ordre de l'heure pour ces données) est relativement incertaine. Ce point très important en pratique sera repris (Chapitre 7) dans le cas de l'échantillonnage stratifié, qui offre une solution alternative économe en mesurage en vue d'estimer l'exposition à long terme.

Par contre, en supposant indépendantes ces 3 séries de données alors qu'elles ne le sont pas, on pourrait estimer que la borne supérieure de l'intervalle de confiance de l'exposition sonore moyenne majore de 1 à 2 dB cette moyenne ! Au lieu de 5 à 11 dB, compte tenu de l'autocorrélation... Entre les estimations "normalisées" et ces résultats, les biais d'estimation de la borne supérieure de l'intervalle de confiance sont d'autant plus importants que les séries contiennent des "pointes" déterminant fortement l'exposition globale.

Ces résultats demandent d'être confirmés par d'autres travaux et par l'analyse d'autres séries chronologiques. Mais dorénavant, le contrôle et la quantification du type d'exposition est possible par emploi des outils déjà élaborés dans cette étude. Ces outils permettent notamment la quantification de l'effet de pointes par les indices de sélectivité, la prise en compte de l'autocorrélation par les variogrammes modélisés, et l'emploi d'estimateurs lognormaux, préférables à ceux des normes actuelles.

Pour exploiter cette possibilité dans la pratique courante du mesurage de l'exposition sonore, une étape préalable est nécessaire : connaître la typologie des situations d'exposition et inclure dans leur description un variogramme. Ainsi, en disposant à terme de ces connaissances, on peut concevoir de définir a priori, une stratégie de mesurage optimisée. Dans ce cadre, l'analyse du travail et des circonstances de l'exposition sonore est absolument indispensable. De plus elle peut conduire à une stratification des périodes d'exposition en phases homogènes. L'intérêt de cette méthode fait l'objet du Chapitre 7.

Chapitre 7

L'exposition à des phases multiples. Méthodes d'échantillonnage stratifié

Dans ce chapitre, nous rappelons d'abord quelques notions de base, et précisons quelques définitions et notations ayant trait au problème de l'estimation d'exposition, et de l'échantillonnage. Cette partie traite de variables additives, c'est à dire de variables dont la somme arithmétique a un sens physique. En acoustique, il en est ainsi des puissances moyennes. Il est bien sur commode de présenter les valeurs en échelle logarithmique (décibels), mais les compositions se font toujours en sommant les énergies, et donc en moyennant les puissances.

7.1 Terminologie employée

Variable régionalisée

On distingue habituellement les variables régionalisées, de leur modélisation par fonctions aléatoires en utilisant les minuscules dans le premier cas, et en réservant les majuscules au second. Les premières sont les quantités effectivement accessibles expérimentalement, et la modélisation probabiliste est là essentiellement pour offrir un cadre propre à une généralisation.

On notera t un point de l'espace dans lequel la variable régionalisée est définie, qui peut être à une ou plusieurs dimensions. Si l'application que l'on discutera concerne des séries temporelles, il est souvent plus commode de représenter graphiquement les choses à deux dimensions.

Soit V une fenêtre d'observation d'une variable régionalisée z . On note donc :

$$z(V) = \frac{1}{|V|} \int_V z(t) dt \quad (7.1)$$

la moyenne sur V (l'intégrale ci-dessus peut être à une, deux, trois, et même quatre dimensions).

Estimation locale et estimation à long terme

Il semble essentiel dans toute discussion ayant trait à l'échantillonnage de distinguer deux objectifs d'estimation :

- L'estimation de la moyenne sur la fenêtre d'observation, $z(V)$. Lorsqu'il s'agira de cet objectif d'estimation, nous parlerons d'estimation de moyenne locale, ou d'estimation à court terme. La difficulté de cette estimation provient seulement du caractère fragmentaire de l'échantillonnage.
- L'estimation de la moyenne à long terme, c'est à dire de :

$$\lim_{V \rightarrow \infty} z(V) \quad (7.2)$$

par $V \rightarrow \infty$, il est sous-entendu que la fenêtre d'observation augmente indéfiniment dans toutes les directions. Dans le cas unidimensionnel qui nous occupe, cela signifie seulement que la durée de V est grande, par exemple une année. Cet objectif d'estimation est beaucoup plus ambitieux que le précédent, puisqu'il nécessite une extrapolation d'une plage reconnue vers des plages non observées. Cette extrapolation n'est possible que par le recours à des hypothèses permettant la transposition de l'observé au non observé. Ces hypothèses sont guidées par la connaissance des processus industriels, et ne sont pas du ressort d'un statisticien. C'est donc aussi un objectif plus risqué, car toute hypothèse non réfutable par des moyens statistiques (par une confrontation à des données) implique un risque d'erreur radicale.

En matière d'exposition, le second objectif d'estimation est probablement le plus intéressant.

Trois concepts de variogramme

Plusieurs variogrammes sont susceptibles d'intervenir dans un problème d'estimation. On distinguera :

- le *variogramme régional* associé à une variable régionalisée,
- le *variogramme expérimental* obtenu à partir de l'échantillonnage de cette variable régionalisée,
- le *modèle de variogramme* associé à la fonction aléatoire dont la variable régionalisée est une réalisation.

Ces trois conceptualisations de la notion de variogramme sont présentées.

Le variogramme régional

Le variogramme régional est en quelque sorte un variogramme expérimental obtenu à partir de l'ensemble des points de la fenêtre d'observation. Sa valeur pour une distance h est donc la moyenne de :

$$\frac{1}{2} [z(t+h) - z(t)]^2 \quad (7.3)$$

lorsque les deux points t et $t+h$ appartiennent à V . Le variogramme régional s'écrit, si $|V|$ désigne la longueur de l'intervalle V :

$$\gamma_V(h) = \frac{1}{2 |V \cap V_{-h}|} \int_{V \cap V_{-h}} [z(t+h) - z(t)]^2 dt \quad (7.4)$$

puisque $V \cap V_{-h}$ est l'ensemble pour lequel t et $t+h$ sont dans V ¹. L'aire de cet ensemble est appelé covariogramme géométrique de V . Avec la notation des indicatrices, $\mathbf{1}_{t \in V}$ ou,

1. V_{-h} est le translaté de V par le vecteur $-h$, c'est à dire précisément l'ensembles des points t tels que $t+h \in V$

plus simplement, $1_V(t)$ (fonction qui vaut 1 si le point t est dans V , et 0 sinon), le covariogramme géométrique vaut :

$$K_V(h) = |V \cap V_{-h}| = \int 1_V(t) 1_V(t+h) dt \quad (7.5)$$

où l'intégrale est étendue à tout l'espace.

Si maintenant une donnée est prélevée dans V , selon un schéma aléatoire pur, et que l'on note \underline{t} la position de cette donnée, le résultat est une variable aléatoire, notée $z(\underline{t})$. La loi de cette variable est appelée loi régionale. Explicitement, elle est donnée par :

$$F_V(u) = \text{Prob}[z(\underline{t}) < u] = \frac{1}{|V|} \int_V 1_{z(t) < u} dt \quad (7.6)$$

(la probabilité de $\{z(\underline{t}) < u\}$ est la proportion de la fenêtre V pour laquelle le niveau de z est en dessous de la coupure u). L'espérance de $z(\underline{t})$ est bien sûr la moyenne de z dans V :

$$E[z(\underline{t})] = z(V) \quad (7.7)$$

et sa variance est donc :

$$\text{Var}(z(\underline{t})) = \frac{1}{|V|} \int_V (z(t) - z(V))^2 dt \quad (7.8)$$

Nous allons voir que cette variance est reliée de manière simple au variogramme régional.

Plus précisément, notons $\bar{\gamma}_V(V, V)$ la valeur moyenne du variogramme régional $\gamma_V(y - y')$ lorsque l'origine et l'extrémité du vecteur décrivent la fenêtre V indépendamment :

$$\bar{\gamma}_V(V, V) = \frac{1}{|V|^2} \int_V \int_V \gamma_V(y - y') dy dy' \quad (7.9)$$

Un résultat que l'on démontre aisément est alors le suivant :

$$\boxed{\text{Var}\{z(\underline{t})\} = \bar{\gamma}_V(V, V)}$$

La valeur moyenne du variogramme régional dans V est donc la variance d'un échantillon d'implantation aléatoire uniforme prélevé dans V . Ce résultat est obtenu sans aucune hypothèse sur la variable régionalisée z et aussi sans recours à un modèle de fonction aléatoire. Il sera utilisé plus loin lorsqu'il sera question d'échantillonnage aléatoire.

Le variogramme expérimental

En pratique on ne connaît qu'un nombre limité de points de la fenêtre d'observation. Dans ces conditions, on ne peut former qu'une approximation du variogramme régional. La forme suivante :

$$\hat{\gamma}_V(h) = \frac{1}{2} \sum_{i,j} [z(t_i) - z(t_j)]^2 w(t_i - t_j; h) \quad (7.10)$$

donne une telle approximation, lorsque la fonction $w(d; h)$ est choisie de manière appropriée. Elle doit donner un poids plus élevé aux distances d proches de h , et avoir une somme unité :

$$\sum_{i,j} w(t_i - t_j; h) = 1 \quad (7.11)$$

Le plus souvent on utilise pour w simplement le quotient de l'indicatrice d'une classe de distances, et du nombre de couples t_i, t_j dans cette classe.

$$\hat{\gamma}_V(h) = \frac{1}{2 N(h)} \sum_{|t_i - t_j| \in [h-dh, h+dh]} [z(t_i) - z(t_j)]^2 \quad (7.12)$$

Cette quantité est donc une approximation du variogramme régional.

Le modèle de variogramme

La variable régionalisée est interprétée ici comme réalisation particulière d'un modèle de fonction aléatoire Z . Le variogramme de cette fonction est défini par :

$$\gamma(h) = \frac{1}{2} E[Z(t+h) - Z(t)]^2 \quad (7.13)$$

Pour les modèles ayant de bonnes propriétés d'ergodicité, le variogramme régional calculé sur une réalisation particulière tend vers le variogramme de la fonction aléatoire lorsque la fenêtre d'observation augmente indéfiniment dans toutes les directions :

$$\gamma_V(h) \rightarrow \gamma(h) \quad (7.14)$$

Cela signifie que pour des séries chronologiques, le variogramme à long terme coïncide avec le variogramme du modèle. C'est cette circonstance qui permet de réaliser l'inférence du variogramme sur la base d'une réalisation unique.

En pratique on a toujours une fenêtre d'observation limitée, ce qui soulève les questions suivantes :

- Le variogramme expérimental estime-t-il bien le variogramme régional ?
- Le variogramme régional est-il proche du variogramme à long terme ?
- Quel variogramme est important pour tel ou tel problème d'estimation ?

Il existe des éléments de réponse à ces questions. D'abord, si la portée du phénomène n'est pas petite par rapport à la dimension de la fenêtre d'observation, le variogramme régional peut être sérieusement différent du variogramme à long terme. La Figure 7.1 montre par exemple une simulation de 1000 valeurs d'une série temporelle de variogramme exponentiel. La Figure 7.2 montre le variogramme régional associé à cette simulation, en superposition du variogramme du modèle. Le palier est pratiquement atteint pour une distance $h = 60$, et de fait, on constate que sur une fenêtre d'observation de 1000 points, le variogramme régional est proche du modèle.

La Figure 7.3 montre les quatre premiers variogrammes régionaux calculés sur quatre fenêtres successives de longueur 100 extraites de la série de 1000 valeurs. On constate que, bien que ces fenêtres soient plus grandes que les "portées pratiques" du variogramme du modèle, les estimations sont de qualité peu satisfaisante (noter les valeurs très différentes des variances expérimentales). Ce fait, qui répond à la seconde question, a été à l'origine de certains malentendus : puisque l'estimation du variogramme n'est pas possible dans de bonnes conditions, la géostatistique ne devrait pas marcher. Nous allons voir comment ce problème se résout.

Pour cela, nous abordons la question suivante, et nous distinguons les deux circonstances évoquées plus haut : estimation de la moyenne sur la fenêtre d'observation, et estimation de la moyenne à long terme. Il se trouve que, en ce qui concerne la première

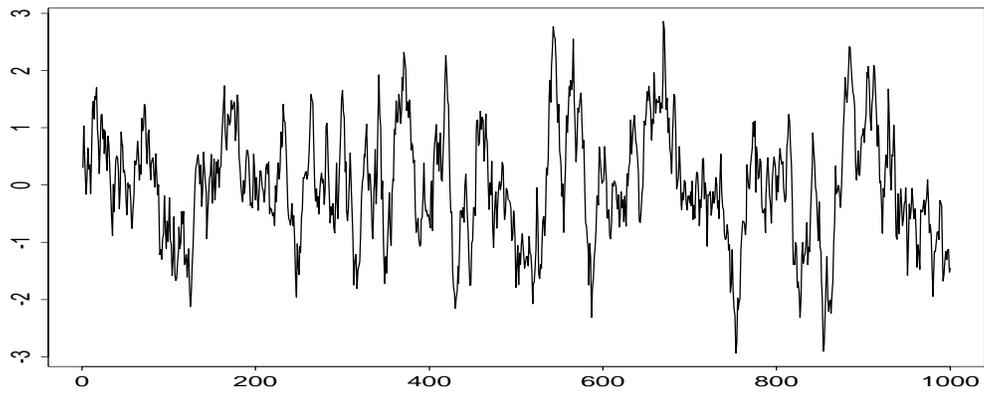


FIGURE 7.1 – Simulation de 1000 valeurs à variogramme exponentiel $\gamma(h) = 1 - 0.9^{|h|}$.

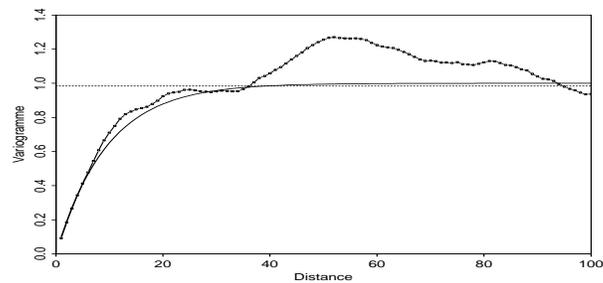


FIGURE 7.2 – Variogramme régional et variogramme théorique de la simulation.

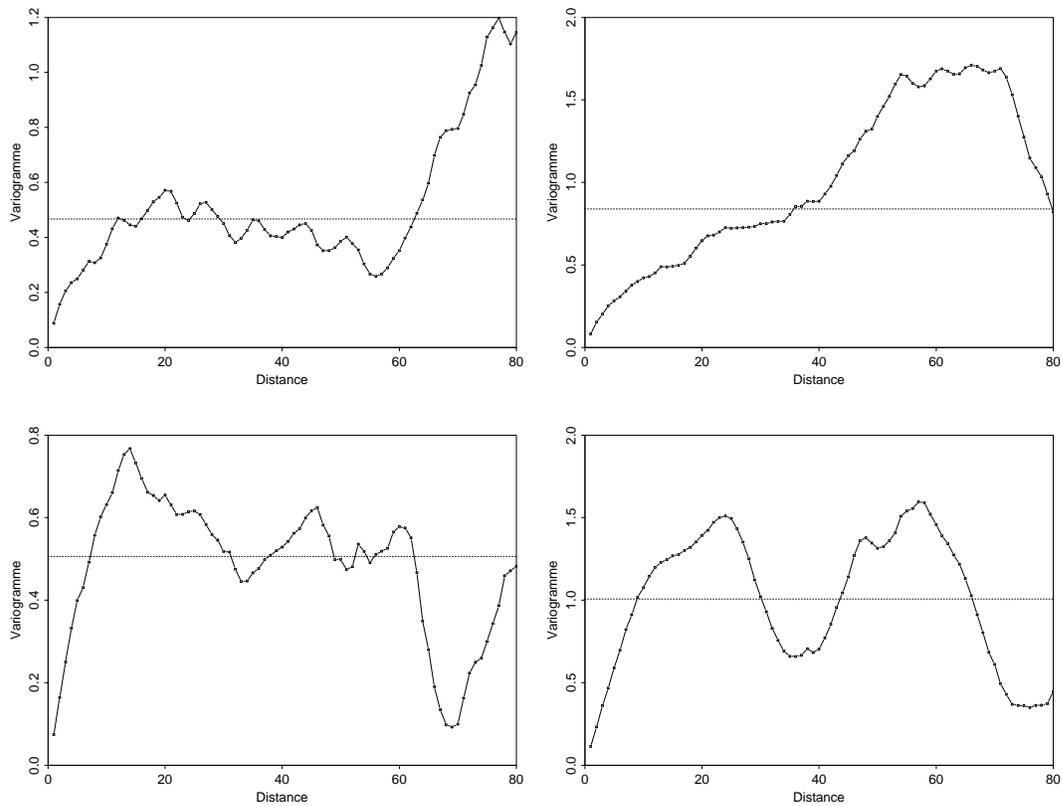


FIGURE 7.3 – Quatre variogrammes régionaux basés sur des fenêtres de 100 points (extraits parmi les 1000 valeurs de la série simulée).

estimation, et donc pour résoudre un problème d'estimation posé dans une fenêtre particulière, à partir de données prélevées dans cette fenêtre, les variogrammes régionaux, et les modèles de variogrammes que nous pouvons en déduire, sont les plus pertinents. C'est seulement dans le cas d'une estimation de moyenne à long terme, qu'il est souhaitable d'utiliser un variogramme à long terme. Cela fait que, *tant que l'on ne cherche pas à extrapoler* au delà de la fenêtre d'observation, la réponse négative à la seconde question est sans incidence sur l'efficacité de la méthode.

7.2 L'échantillonnage

On envisagera essentiellement deux schémas d'échantillonnage : le schéma aléatoire dans lequel les implantations des données sont uniformes et indépendantes, et le schéma régulier dans lequel elles sont régulièrement espacées. Ces deux schémas peuvent être utilisés pour estimer la moyenne locale, ou pour estimer la moyenne à long terme. Enfin ils peuvent être associés à une stratification du domaine. Les données peuvent alors être réparties en nombre égal d'une phase à l'autre, ou cette répartition peut être optimisée en tenant compte des variabilités, et des durées de chaque phase. Commençons par dire quelques mots sur le schéma aléatoire pur et voyons ce en quoi il simplifie l'analyse.

Schéma aléatoire pur

Du point de vue de la variable régionalisée

Nous considérons ici une variable régionalisée (déterministe) définie à l'intérieur d'une fenêtre d'observation. L'objectif est donc d'estimer la moyenne sur cette fenêtre, notée $z(V)$. Si les points d'échantillonnage sont n points tirés suivant une loi uniforme dans V , indépendamment les uns des autres, les valeurs associées :

$$z(\underline{t}_1), z(\underline{t}_2), \dots, z(\underline{t}_n) \quad (7.15)$$

sont n variables indépendantes suivant la loi régionale dans V définie plus haut :

$$F_V(u) = \frac{1}{|V|} \int_V 1_{z(t) < u} dt \quad (7.16)$$

dont l'espérance est $z(V)$ et la variance $\bar{\gamma}_V(V, V)$. L'estimateur :

$$z(V)^* = \frac{1}{n} \sum_i z(\underline{t}_i) \quad (7.17)$$

est donc sans biais, et de variance :

$$\text{Var}[z(V)^* - z(V)] = \frac{\bar{\gamma}_V(V, V)}{n} \quad (7.18)$$

Cette variance n'est connue exactement que si le variogramme régional est lui même connu. En pratique, il est nécessaire d'utiliser l'approximation :

$$\text{Var}[z(\underline{t})] \approx \frac{1}{n-1} \sum_i [z(\underline{t}_i) - z(V)^*]^2 \quad (7.19)$$

et donc :

$$\text{Var}(z(V)^* - z(V)) \approx \frac{1}{n(n-1)} \sum [z(\underline{t}_i) - z(V)^*]^2 \quad (7.20)$$

Ce schéma permet donc d'estimer la *moyenne locale* sur V , et la variance d'estimation de cette moyenne *sans recours à un modèle probabiliste, et sans calcul de variogramme*. En revanche, on ne peut pas aller beaucoup plus loin, et en particulier on ne peut pas estimer la moyenne à long terme, sans modélisation probabiliste.

Du point de vue probabiliste pour une estimation locale

Si la variable régionalisée est interprétée comme réalisation d'un modèle probabiliste Z , il convient de distinguer les variances suivantes :

- $\text{Var}[Z(\underline{t})] = \text{Var}[Z(t)] = \sigma^2$ variance au support ponctuel, qui intègre la variabilité à très long terme, en même temps que la variabilité d'une réalisation à l'autre de Z (la randomisation de t ne modifie pas la loi pour un modèle stationnaire).
- $\text{Var}[Z(\underline{t}) - Z(V)]$, qui est la variance d'une donnée dans la fenêtre d'observation lorsque l'on considère toutes les réalisations possibles de Z . C'est donc la moyenne de $\text{Var}[z(\underline{t}) - z(V)]$ sur l'ensemble des réalisations. Cette quantité vaut :

$$\text{Var}[Z(\underline{t}) - Z(V)] = \bar{\gamma}(V, V) = S^2[o|V] \quad (7.21)$$

en utilisant cette fois le variogramme du modèle, puisque nous sommes passés à l'espérance, et que le variogramme du modèle est l'espérance du variogramme régional. Cette quantité est appelée *variance de dispersion* (du support ponctuel) dans V .

Pour l'estimation de $Z(V)$ par échantillonnage aléatoire pur à partir de n données, on a :

$$\text{Var}\{Z(V)^* - Z(V)\} = \frac{1}{n} S^2[o|V] \quad (7.22)$$

Du point de vue de l'estimation de la moyenne à long terme

L'estimateur $Z(V)^* = \sum Z(\underline{t}_i)/n$ est aussi un estimateur sans biais de la moyenne à long terme $m = E[Z(t)]$. Mais la variance attachée à cet estimateur n'est plus $S^2[o|V]/n$. Pour chiffrer cette variance, il est commode d'introduire une définition plus générale de la variance de dispersion.

Soient v et V des volumes compatibles en ce sens que V puisse être découpé en volumes égaux à v . Nous pouvons alors considérer un tel découpage $V = \cup_i v_i$, et lui associer la variable aléatoire $Z(\underline{v})$ dans lequel le volume \underline{v} est choisi par tirage aléatoire uniforme dans v_1, v_2, \dots . La variance de dispersion de v dans V est par définition la variance de $Z(\underline{v}) - Z(V)$. Elle se calcule à partir du variogramme selon la formule suivante :

$$S^2[v|V] = \text{Var}[Z(\underline{v}) - Z(V)] = \bar{\gamma}(V, V) - \bar{\gamma}(v, v) \quad (7.23)$$

Cette formule englobe la précédente, puisque si v est réduit à un point, $\bar{\gamma}(v, v) = 0$. D'autre part, si le volume V devient infini, que le variogramme est borné (fonction stationnaire pour laquelle il existe une variance), et pour simplifier, à une portée finie, le variogramme moyenné sur V tend vers la variance :

$$\bar{\gamma}(V, V) \rightarrow \sigma^2 \quad \text{pour } V \rightarrow \infty \quad (7.24)$$

La variance de dispersion de v dans un champ infini est la variance de $Z(v)$, et se calcule d'après la relation précédente par :

$$\text{Var}[Z(v)] = S^2[v|\infty] = \sigma^2 - \bar{\gamma}(v, v) \quad (7.25)$$

Enfin, si $v_1 \subset v_2 \subset v_3$ sont des volumes compatibles, on a immédiatement, d'après 7.23 :

$$S^2[v_1|v_3] = S^2[v_1|v_2] + S^2[v_2|v_3] \quad (7.26)$$

Appliqué à $v_1 = t$, $v_2 = V$ et $v_3 = \infty$, cela donne :

$$\text{Var}[Z(t)] = \text{Var}[Z(\underline{t}) - Z(V)] + \text{Var}[Z(V)] \quad (7.27)$$

ce qui signifie que la variance totale est la somme de la variance observée dans V et de la variance de $Z(V)$. En ce qui concerne la variance de l'estimateur en échantillonnage aléatoire pur, $Z(V)^*$, on remarquera que deux échantillons $Z(\underline{t}_1)$ et $Z(\underline{t}_2)$ ne sont plus indépendants puisqu'ils ont été prélevés dans la même fenêtre V . Les relations correctes sont maintenant :

$$\text{Cov}\{Z(\underline{t}_1), Z(\underline{t}_2)\} = \sigma^2 - \bar{\gamma}(V, V) \quad (7.28)$$

$$\text{Cov}\{Z(\underline{t}_1) - Z(V), Z(\underline{t}_2) - Z(V)\} = 0 \quad (7.29)$$

Ce sont donc les erreurs d'estimation de $Z(V)$ par $Z(\underline{t}_1)$ et $Z(\underline{t}_2)$ qui sont décorréelées, et non les données d'échantillons. Quand à l'estimation de cette moyenne à long terme, $m^* = Z(V)^*$, on a :

$$\boxed{\text{Var}[m^* - m] = \text{Var}[m^* - Z(V)] + \text{Var}[Z(V)]} \quad (7.30)$$

Il faut donc rajouter à la variance d'estimation de la moyenne sur la fenêtre d'observation, la variance $\text{Var}[Z(V)] = \sigma^2 - \bar{\gamma}(V, V)$.

En résumé, le *variogramme régional* permet d'obtenir la variance d'estimation de la moyenne sur la fenêtre d'observation à partir de données prises dans cette fenêtre. Si l'on cherche à estimer une moyenne à long terme, il sera nécessaire d'utiliser le variogramme régional sur une fenêtre infinie, pour tenir compte de la variance de la moyenne de la fenêtre. Ce dernier variogramme est en fait égal au modèle de variogramme.

Schéma d'implantation déterministe

Ici, on suppose que l'implantation des données obéit à un schéma déterministe, par exemple n points espacés régulièrement, les points les plus extrêmes occupant les limites de la fenêtre. En ce cas un estimateur faisant intervenir une pondération des données peut être plus efficace que la moyenne arithmétique. La discussion se fera nécessairement en référence à un modèle de fonction aléatoire. On distinguera l'estimation locale, c'est-à-dire celle de la valeur moyenne sur la fenêtre V

$$Z(V)^* = \sum_i \lambda_i Z(t_i) \quad (7.31)$$

et l'estimation de la moyenne à long terme :

$$m^* = \sum_i \lambda_i^m Z(t_i) \quad (7.32)$$

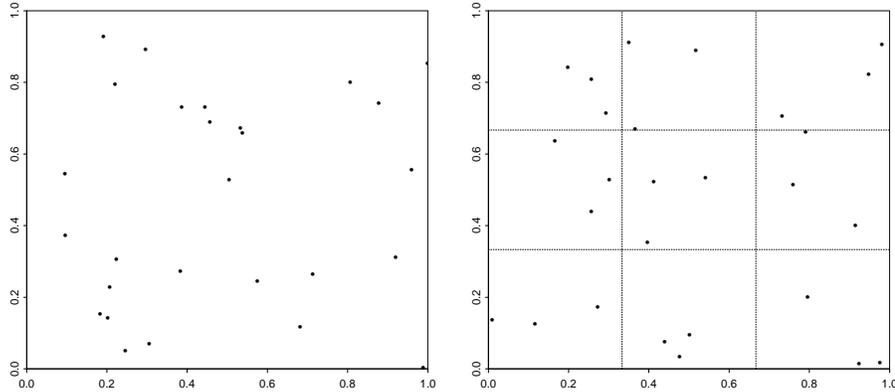


FIGURE 7.4 – Semis de 27 points aléatoires répartis de manière uniforme sur un domaine géographique (à gauche) ou à l'intérieur de strates (à droite).

Ces deux estimateurs sont sans biais si $\sum \lambda_i = 1$ et $\sum \lambda_i^m = 1$, et leurs variances sont de :

$$\text{Var}\{Z(V) - \sum_i \lambda_i Z(t_i)\} = \sum_i \lambda_i \bar{\gamma}(t_i, V) - \sum_{ij} \lambda_i \lambda_j \gamma(t_i - t_j) - \bar{\gamma}(V, V) \quad (7.33)$$

et

$$\text{Var}\{m - \sum_i \lambda_i^m Z(t_i)\} = \sigma^2 - \sum_{ij} \lambda_i^m \lambda_j^m \gamma(t_i - t_j) \quad (7.34)$$

L'optimisation de ces quantités par rapport aux pondérateurs, sous contrainte de non biais et pour une position donnée des t_i conduit respectivement au krigeage de $Z(V)$ et au krigeage de la moyenne. Les pondérateurs et les variances associées ne sont pas les mêmes dans les deux cas ci-dessus. Par exemple les points extrêmes reçoivent un poids plus élevé dans le krigeage de la moyenne à long terme. Enfin, si l'optimisation rigoureuse de la position des points de données est un problème ardu, il est en revanche facile de comparer différentes implantations.

Schémas d'échantillonnage stratifiés

Estimation locale par schéma aléatoire stratifié

L'échantillonnage aléatoire pur n'est pas très efficace, car contrairement à ce que l'intuition suggère, il ne couvre pas le domaine exploré régulièrement. La Figure 7.4, qui montre un semis de points aléatoires uniformes sur un domaine géographique, illustre cela. On constate que les points paraissent se regrouper ici et là, alors que des zones non couvertes subsistent. Nous pouvons chercher à rendre ce semis plus régulier en stratifiant le domaine V , comme il est montré à droite de la Figure 7.4 pour une stratification en 9 parcelles avec trois points aléatoires par strate. Conceptuellement on a alors, en restant dans le point de vue déterministe, autant de variogrammes et de lois régionales que nous avons introduit de strates.

Plus généralement, il est intéressant de stratifier lorsqu'on suspecte une inhomogénéité. En ce cas, les strates, que l'on notera $v_j; j = 1, 2, \dots, k$ peuvent être d'importance inégale.

La moyenne sur $V = \cup v_j$ est obtenue par pondération :

$$z(V) = \frac{1}{|V|} \sum_j |v_j| z(v_j) \quad (7.35)$$

Cette pondération est appliquée aux estimateurs :

$$z(V)^* = \frac{1}{|V|} \sum_j |v_j| z(v_j)^* \quad (7.36)$$

l'estimateur par strate $z(v_j)^*$ pouvant être simplement la moyenne arithmétique des données dans la strate. Nous savons que la variance de chacun de ces estimateurs est donné par les variogrammes régionaux γ_{v_j} associés :

$$\text{Var}\{z(v_j)^* - z(v_j)\} = \frac{1}{n_j} \bar{\gamma}_{v_j}(v_j, v_j) \quad (7.37)$$

Comme les données de deux strates différentes sont indépendantes (il s'agit de deux réalisations indépendantes des lois régionales associées), les estimateurs $z(v_j)^*$ et $z(v_{j'})^*$ sont aussi deux à deux indépendants. Cela conduit à la variance globale :

$$\text{Var}\{z(V)^* - z(V)\} = \frac{1}{|V|^2} \sum_j |v_j|^2 \text{Var}\{z(v_j)^* - z(v_j)\} \quad (7.38)$$

Là encore, les variogrammes régionaux étant inconnus, les variances d'estimation dans chaque strate sont estimées par :

$$\text{Var}\{z(v_j)^* - z(v_j)\} \approx \frac{1}{n_j(n_j - 1)} \sum_i \{z(t_i^j) - z(v_j)^*\}^2 \quad (7.39)$$

Estimation à long terme avec stratification

Le recours à un modèle est inévitable, puisque le variogramme régional ne reflète pas nécessairement la variabilité à plus long terme. Si p_j^* et m_i^* désignent l'estimation de la proportion et de la moyenne de la phase j , toujours à long terme, l'estimation de la moyenne globale est :

$$m^* = \sum_j p_j^* m_j^* \quad (7.40)$$

L'estimation des proportions p_j^* est effectuée à base d'une analyse des tâches et des processus industriels, et l'erreur dont elle est entachée est donc indépendante de l'erreur $m_j - m_j^*$. C'est l'hypothèse de base dans les calculs qui suivent. Elle est à réviser si les estimations de proportions p^* sont effectuées sur la base de l'examen du fichier de données. Si on a une idée de la précision avec laquelle cette estimation est réalisée, et si on suppose cette estimation non biaisée, cela pourra être traduit en terme de variance-covariance :

$$\Sigma^p = \{\sigma_{jj'}^p\} = \{\text{Cov}(p_j^* - p_j, p_{j'}^* - p_{j'})\} \quad (7.41)$$

On note aussi Σ^m la matrice des covariances des m_i^* . Cela conduit à l'expression de la variance d'estimation :

$$\text{Var}(m^*) = \sum_{i,j} \sigma_{ij}^p \sigma_{ij}^m + \sum_{i,j} m_i m_j \sigma_{ij}^p + \sum_{i,j} p_i p_j \sigma_{ij}^m \quad (7.42)$$

ou, matriciellement :

$$\boxed{\text{Var}(m^*) = \text{Trace}\{\Sigma^p \Sigma^m\} + M^t \Sigma^p M + P^t \Sigma^m P} \quad (7.43)$$

avec bien sur $M^t = (m_1, m_2, \dots)$ et $P^t = (p_1, p_2, \dots)$. Moyennant l'hypothèse d'indépendance des estimations de M et P , cette formule est générale. Nous la retrouverons à plusieurs reprises.

Elle peut être simplifiée moyennant quelques hypothèses supplémentaires. Si l'on cherche seulement à optimiser un échantillonnage, il n'est pas nécessaire de prendre en compte le second terme de cette expression. Si nous supposons de plus les estimations des moyennes dans chaque phase décorréelées², la matrice Σ^m est diagonale, et la quantité à optimiser est réduite à :

$$S^2 = \sum_i \text{Var}(m_i^*) \{ \text{Var}(p_i^*) + p_i^2 \} \quad (7.44)$$

Cette quantité fait intervenir p_i qui est inconnu, et qui est en pratique remplacé par son estimateur dans l'optimisation.

7.3 Application à une série multiphase

Avertissement : Le facteur normatif pour calibrer les expositions sonores par rapport aux décibels est de 10^{10} dans ce chapitre, au lieu de 10^9 prédominant. L'exposition sonore vaut 1 pour 100 dB dans ce chapitre, alors qu'elle valait 1 pour 90 dB dans les chapitres précédents.

Description du poste de travail observé

Le poste de travail `amenag31` est un poste de commande d'une presse à découper automatique, qui fonctionne à une cadence voisine de 3 coups par seconde. L'alimentation de la presse et l'évacuation des pièces sont automatisées. Les tâches de l'opérateur comprennent la mise en place de la bobine d'alimentation et le réglage de la presse en cas de changement de production, la surveillance de la fabrication, le contrôle dimensionnel. L'opérateur est mobile dans une zone limitée autour de la presse. Cette presse est relativement isolée dans l'atelier et l'opérateur est exposé surtout au bruit de cette machine.

La nature de la production varie dans une journée. Dans la journée observée, on note 3 phases de travail différentes (repérées sur les Figures 7.5 et 7.15) : la phase A correspond à une découpe de tôle d'épaisseur $18/10^0$, la phase B la découpe de tôle de $10/10^0$, la phase C des réglages entre 2 productions différentes.

2. Le cas le plus simple dans lequel cette condition est remplie est l'indépendance des phases, mais plus généralement il suffit pour cela que les estimations soient conditionnellement non biaisées, c'est à dire que pour $i \neq j$, on ait $E[m_i^* - m_i | m_j^*] = 0$. Différents facteurs, par exemple le facteur climatique en chimie, peuvent être source de corrélation entre les estimations de moyennes par phase, et invalider le calcul. Il faut y prendre garde, car négliger cette corrélation si elle existe conduit à un optimisme exagéré.

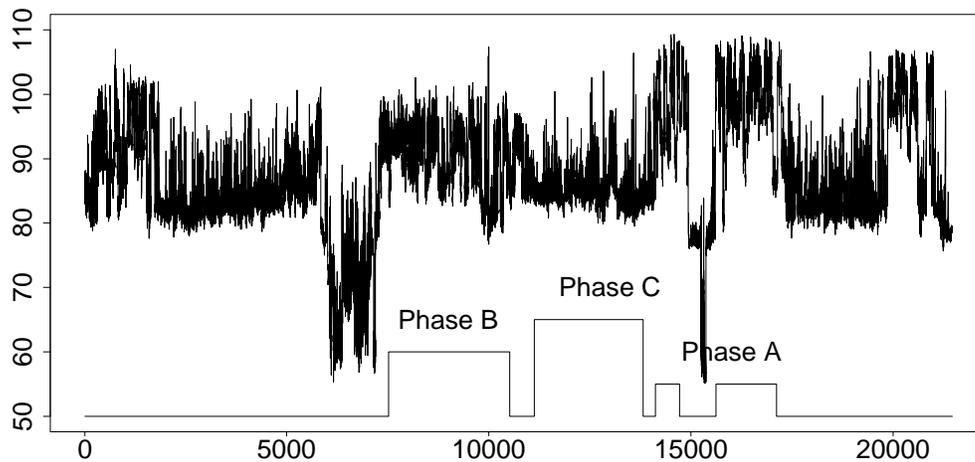


FIGURE 7.5 – Série amenag31, en décibels, pas de 1 seconde.

Présentation des estimations réalisées

Nous détaillons les résultats obtenus pour la série amenag31 qui a été stratifiée en phases déterminées par une analyse préalable du travail. Le traitement statistique commence par une validation de la segmentation proposée en terme de distribution et de variogramme. On considère ensuite les estimations suivantes :

1. *Estimation locale* de la moyenne de l'ensemble des phases identifiées en se basant sur un échantillonnage aléatoire stratifié. Cette moyenne diffère de la moyenne de la journée, puisque seul un sous-ensemble est identifié. On considèrera le cas de l'échantillonnage à répartition égale, pour lequel chaque strate reçoit le même nombre de donnée, et un échantillonnage optimisé en tenant compte des durées et des variances des différentes phases.
2. *Estimation de la moyenne à long terme.* Cette opération, qui demande une extrapolation n'est possible qu'en ayant recours à un modèle de fonction aléatoire. On envisagera différents cas de figure en ce qui concerne les durées des différentes phases à long terme, et en ce qui concerne la précision avec laquelle ces valeurs sont connues. Enfin, les estimations, et surtout la précision que nous leur attribuons, sont basées sur l'hypothèse que la variabilité propre à chaque phase est connue à partir des données de la fenêtre d'observation. Nous illustrerons cela en introduisant à titre d'alternative, une variabilité supplémentaire à grande portée, qui conduit à une variance inter journées.

Variogrammes des trois phases

La Figure 7.5 montre la série des valeurs mesurées durant environ 6 heures, représentée ici au pas de une seconde. Trois phases d'activité définies ci-dessus sont mentionnées. Elles se distinguent bien par des niveaux sonores différents et par des durées facilement identifiables dans la période observée. Les trois phases identifiées sont indiquées sur le graphique. Les différences de distributions, au support de la seconde, et toujours en décibels, sont mises en évidence par les histogrammes de la Figure 7.6. La phase repérée par le code "0" correspond à ce qui n'est pas identifié comme représentatif des phases A, B ou

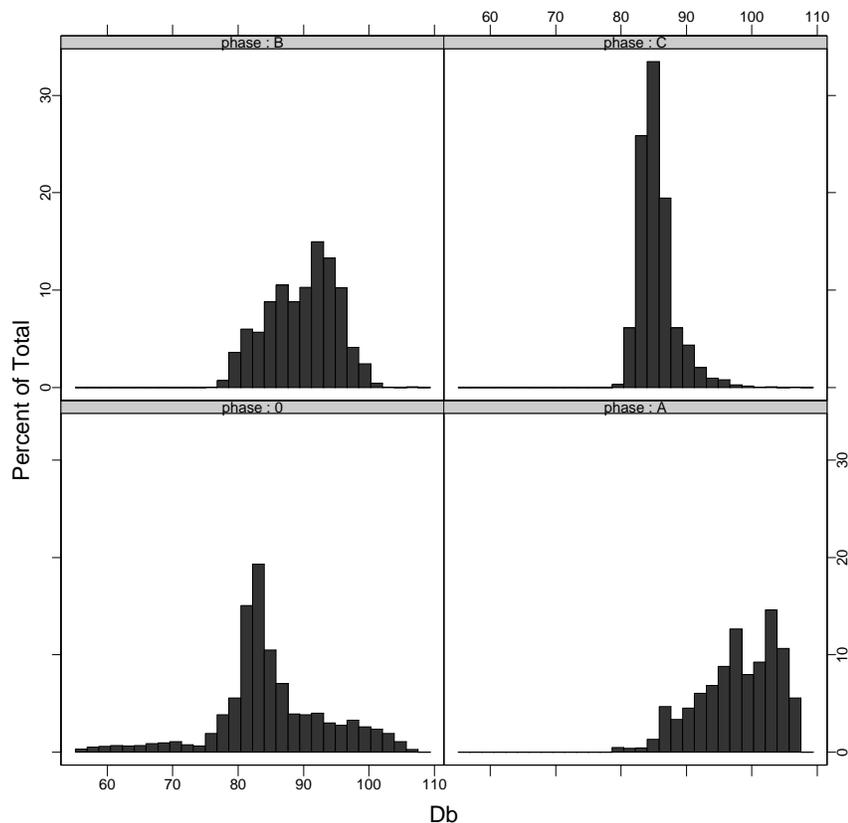


FIGURE 7.6 – Histogrammes des différentes phases de `amenag31`. De haut en bas et de gauche à droite, phase B, C, phase “O”, et phase A. Les échelles de ces histogrammes sont les mêmes

C. Elle comprend bien sur des parties qui auraient pu être rattachées à l’une ou l’autre de ces phases.

La Figure 7.7 représente le variogramme de la série `amenag31` avant segmentation. Il est représenté ici au pas de 10 secondes et non de 1 seconde, afin de rendre possible son calcul jusqu’à la valeur maximum de l’écart de temps entre 2 valeurs (soit 3 heures dans ce cas). Ce variogramme est difficile à interpréter. A court terme, il présente une singularité à l’échelle de la minute ; vers 45 minutes, une portée apparaît, qui pourrait être validée si le variogramme était limité à un intervalle de durée de 1 heure, alors que la convergence vers la variance expérimentale n’est plus nette au-delà de 1 heure et demi. En fait, ce variogramme est certainement sensible aux transitions entre les phases A, B et C. Autrement dit, si les mêmes tâches étaient effectuées par l’opérateur dans un ordre différent durant la période observée, le variogramme serait certainement modifié.

Après segmentation en 3 phases, les variogrammes des données d’exposition sonore au support de 1 seconde sont montrés dans la Figure 7.8. Les caractéristiques des variogrammes de chaque phase sont très différentes de celles du variogramme de la série avant segmentation. La phase A, la plus bruyante, présente une portée voisine de deux minutes. Cette phase est largement dominante en terme de variance. La phase B a une structure de portée 30 secondes et une composante à faible variance et grande portée. Le variogramme de la phase C est dominé par la présence de quelques valeurs fortes. La Figure 7.9 illustre

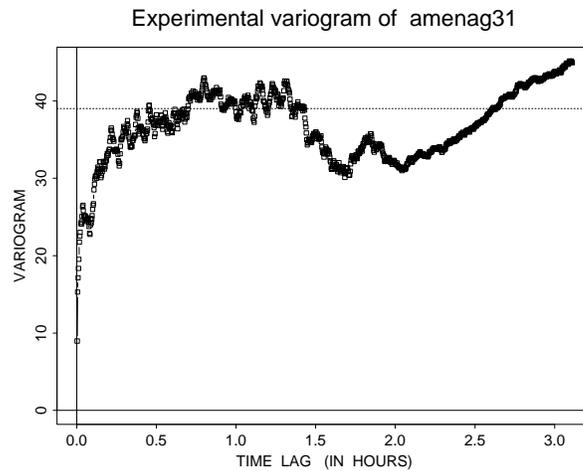


FIGURE 7.7 – Variogramme des données `amenag31`, avant segmentation, en expositions sonores au support de 10s.

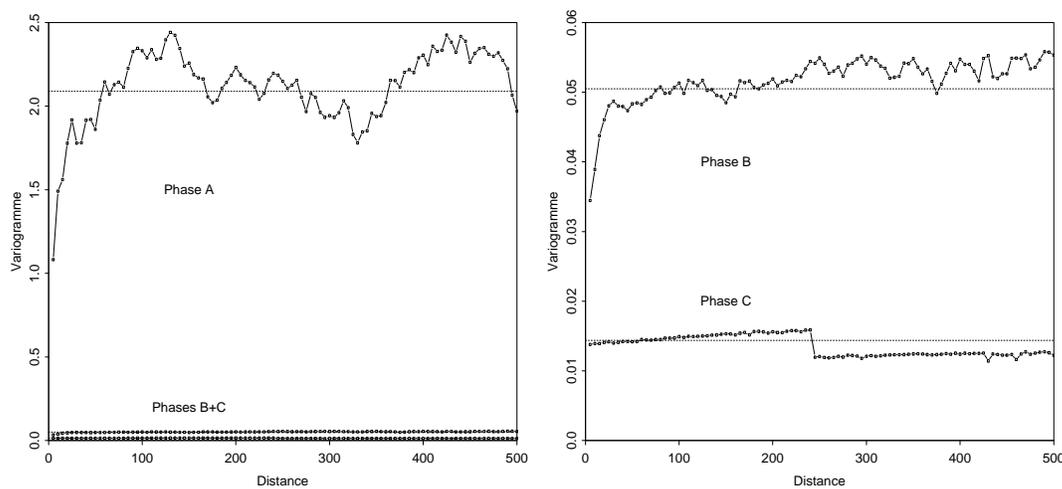


FIGURE 7.8 – Variogrammes des expositions sonores des 3 phases de `amenag31` (toutes les variances sont à multiplier par 10^{20} pour obtenir les valeurs effectives).

cela en comparant le variogramme de cette phase jusqu'à une distance de 2000 secondes, avec toutes les données d'une part, et en masquant seulement les 3 plus fortes données d'autre part.

Les statistiques d'ordre deux par phase sont données dans le Tableau 7.1. Notons les variances calculées sur toutes les données de chacune de ces phases. D'après ce qui a été vu précédemment, ces variances sont les valeurs moyennes des variogrammes régionaux dans les sous fenêtres. Ces valeurs contrastées soulignent l'importance de la phase A. Notons aussi que la moyenne sur l'ensemble des phases identifiées (0,444) est presque deux fois supérieure à la moyenne sur l'ensemble du fichier (0,288). Or c'est cette première valeur que l'échantillonnage stratifié estimera. Pour estimer la seconde, une extrapolation est nécessaire, dans laquelle il faudra spécifier les durées de chaque phase dans la journée, ainsi que la durée et le niveau de la phase non représentée dans l'échantillonnage (pauses).

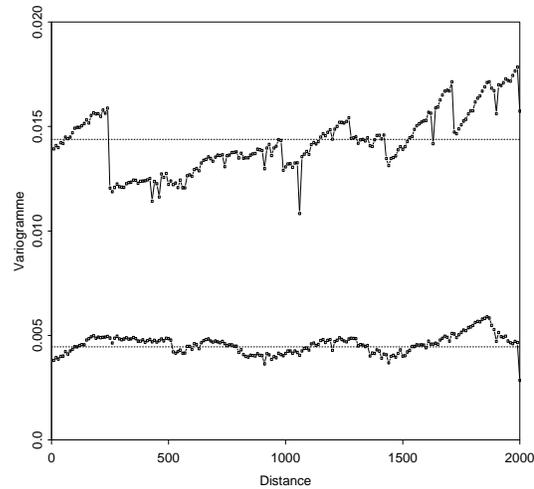


FIGURE 7.9 – Variogramme des données d'exposition sonore au support 1 seconde, de la phase C de amenag31. La courbe donnant les plus fortes variances est obtenue avec toutes les données de la phase, alors que l'autre est obtenue en masquant les 3 plus fortes données.

Ensemble	Variance $\sigma^2/10^{20}$	moyenne $m/10^{10}$	Nombre de données
A	2,089	1,334	2 099
B	0,050	0,176	3 000
C	0,014	0,050	2 699
A+B+C	0,881	0,444	7 798
$(A+B+C)^C$	0,266	0,199	13 669
Fichier complet	0,503	0,288	21 467

TABLE 7.1 – Moyennes variances et effectifs des différentes phases de amenag31 (données en exposition sonore) au support de 1s.

Estimation de la moyenne locale

Nous pouvons, à partir de ces valeurs prévoir l'efficacité en terme de variance, des estimateurs de la moyenne sur l'ensemble des phases identifiées (7798 secondes) pour la journée en cours, soit de la valeur 0,444, à partir de N_A, N_B, N_C échantillons prélevés dans les trois phases. En effet, si :

$$z(V)^* = \frac{1}{|v_A| + |v_B| + |v_C|} \left\{ \frac{|v_A|}{N_A} \sum z_A^i + \frac{|v_B|}{N_B} \sum z_B^i + \frac{|v_C|}{N_C} \sum z_C^i \right\} \quad (7.45)$$

en raison de l'indépendance des termes d'erreurs³, la variance d'estimation est :

$$\begin{aligned} \text{Var}\{z(V) - z(V)^*\} &= \frac{1}{|v_A + v_B + v_C|^2} \\ &\times \left\{ \frac{|v_A|^2}{N_A} \bar{\gamma}_A(v_A, v_A) + \frac{|v_B|^2}{N_B} \bar{\gamma}_B(v_B, v_B) + \frac{|v_C|^2}{N_C} \bar{\gamma}_C(v_C, v_C) \right\} \end{aligned}$$

Avec les valeurs numériques observées, cela donne :

$$z(V)^* = 0,27 m_A^* + 0,38 m_B^* + 0,35 m_C^* \quad (7.46)$$

et

$$\text{Var}\{z(V) - z(V)^*\} = \frac{0,151}{N_A} + \frac{7 \cdot 10^{-3}}{N_B} + \frac{1,7 \cdot 10^{-3}}{N_C} \quad (7.47)$$

Ces valeurs très contrastées proviennent des fortes différences portant sur les variances de l'énergie. Une procédure pour optimiser la répartition d'un nombre donné d'échantillons est décrite en Annexe. Son application ici conduit par exemple, pour 60 échantillons d'une seconde, à la répartition $N_A = 45, N_B = 10$, et $N_C = 5$. La variance théorique calculée pour cet échantillonnage est de $\text{Var}\{z(V) - z(V)^*\} = 4,4 \cdot 10^{-3}$, tandis que la répartition 20, 20, 20 conduit à⁴ $\text{Var}\{z(V) - z(V)^*\} = 8 \cdot 10^{-3}$. La connaissance exhaustive sur la fenêtre d'observation permet de vérifier ces valeurs par simulation d'échantillonnage aléatoire, et aussi de comparer les distributions associées. Dans le cas de ce fichier, à partir de 1000 simulations d'estimateur, nous avons obtenu les valeurs suivantes, qui sont donc très proches des prévisions :

Échantillonnage N_A, N_B, N_C	Moyenne de l'estimateur $z(V)^*$	Variance de l'estimateur
(20, 20, 20)	0,446	$8,2 \cdot 10^{-3}$
(45, 10, 5)	0,442	$4,6 \cdot 10^{-3}$

TABLE 7.2 – Moyennes et variances obtenues expérimentalement par échantillonnage aléatoire stratifié basé sur 60 données d'une seconde, pour `amenag31`.

Les histogrammes des deux estimateurs sont montrés dans la Figure 7.10. L'avantage de l'optimisation d'échantillonnage est évidente sur ces histogrammes qui ont les mêmes échelles et les mêmes seuils de classes pour faciliter la comparaison. Nous traitons dans

3. Cette indépendance est acquise par construction, puisque le seul élément aléatoire dans cette analyse est la position des échantillons dans chaque phase. C'est seulement lorsque nous raisonnons sur l'échantillonnage d'une fonction aléatoire que cette indépendance peut être remise en question. Ce cadre est nécessaire si nous estimons une moyenne à long terme.

4. Ces variances sont toujours relatives aux énergies "standardisées" $10^{Leq/10}/10^{10}$.

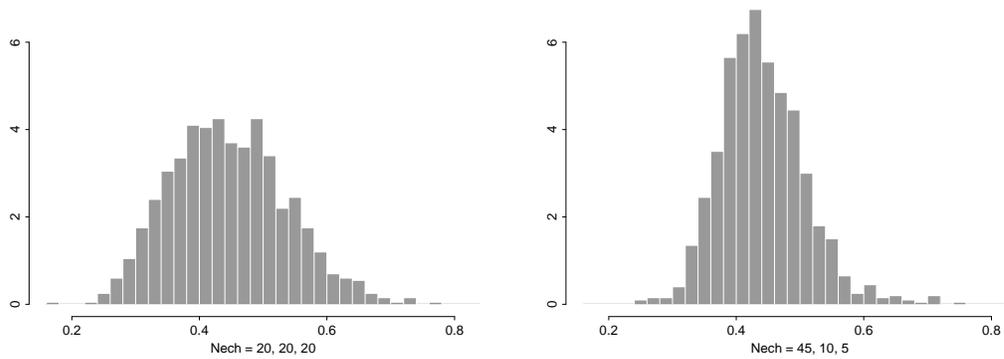


FIGURE 7.10 – Histogrammes des estimateurs de la moyenne des phases identifiées de *amenag31*. A gauche, estimation basée sur 3 fois 20 échantillons d'une seconde, et à droite, basée sur $N_A = 45$, $N_B = 10$, et $N_C = 5$ données d'une seconde.

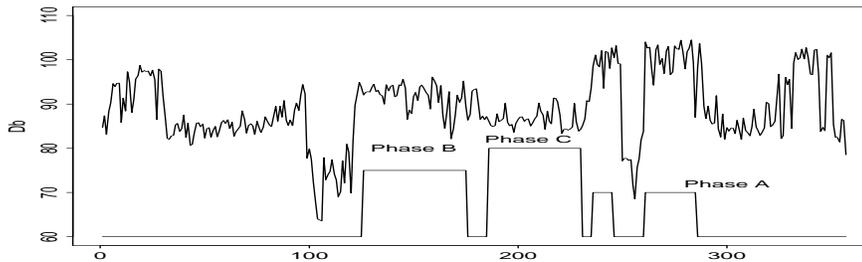


FIGURE 7.11 – Données de *amenag31*, et phases reconnues, au support une minute.

le paragraphe suivant du cas de mesures au support 1 et 2 minutes, avant de passer au problème plus complexe de l'extrapolation.

Estimation de la moyenne locale à partir de données au support de la minute

Il serait possible de régulariser les variogrammes pour obtenir les variances des données au support d'une minute dans chaque fenêtre pour laquelle la phase a été repérée. Outre le fait que cette approche aurait été plus lourde, elle aurait compliqué l'interprétation en raison d'effets de bords. On a préféré ici régulariser les données elles mêmes, en attribuant à chaque valeur au support de la minute la phase et la date du point milieu. On a ainsi obtenu la série montrée en Figure 7.11. Les statistiques par phases sont données dans le Tableau 7.3.

Phase	Moyenne	Variance	Nombre
A	1,32	0,545	35
B	0,178	$0,46 \cdot 10^{-3}$	50
C	0,050	$0,85 \cdot 10^{-4}$	45
A+B+C	0,442	0,438	130

TABLE 7.3 – Premiers moments empiriques pour les trois phases de *amenag31*, au support 1 minute.

Durée	N_A	N_B	N_C	$\hat{\sigma}_1^2$	$N_A = N_B = N_C$	$\hat{\sigma}_2^2$
6	4	1	1	0,011	6	0,020
12	9	2	1	0,005	12	0,010
18	14	3	1	0,0033	18	0,0068
24	19	4	1	0,0025	24	0,0051
30	25	4	1	0,002	30	0,0041
36	29	5	2	0,0017	36	0,0034

TABLE 7.4 – Echantillonnage optimum en schéma aléatoire stratifié des trois phases de `amenag31`, au support 1 minute. La première variance correspond à la répartition optimale, la seconde à une répartition égale (même nombre d'échantillons dans chaque phase). L'optimisation divise approximativement par deux les variances.

Durée (minutes)	Nombre d'échantillons	Variance	Nombre d'échantillons	Variance
6	4, 1, 1	0,01	2, 2, 2,	0,018
18	14, 3, 1	$2,3 \cdot 10^{-3}$	6, 6, 6	$5,6 \cdot 10^{-3}$
36	29, 5, 2	$5,3 \cdot 10^{-4}$	12, 12, 12	$2,2 \cdot 10^{-3}$

TABLE 7.5 – Comparaison expérimentale des échantillonnages stratifiés sans remise, pour une répartition optimisée et une répartition égale pour `amenag31` au support 1 minute.

Les valeurs de variance sont encore assez contrastées, ce qui donnera un échantillonnage optimum favorisant nettement la phase A. Les comparaisons des différents échantillonnages obtenus pour les durées totales fixées a priori sont montrées dans le Tableau 7.4. Une réserve doit être faite à propos des chiffres portés dans ce tableau : tous les calculs sont effectués sur la base d'un échantillonnage aléatoire pur, et donc sur un tirage avec remise en jeu. Or les populations devenant assez petites à ce support (35 données seulement dans la phase A), cet échantillonnage conduit à tirer plusieurs fois la même valeur. Les calculs sont donc basés sur cette démarche peu réaliste, et il faudrait plutôt considérer un tirage sans remise. C'est ce qui a été fait dans la simulation d'échantillonnage qui a conduit aux variances du Tableau 7.5, et aux histogrammes de la Figure 7.12. Par la suite lorsqu'il sera de nouveau question d'échantillonnage aléatoire, il s'agira sauf mention explicite du contraire, d'échantillonnage avec remise.

On notera que *l'échantillonnage optimisé basé sur 18 données est équivalent à l'échantillonnage à répartition égale en nombre comportant 36 données.*

Estimation de la moyenne locale à partir de données au support deux minutes

La procédure de régularisation décrite au paragraphe précédent a été appliquée au support de deux minutes, ce qui a donné la série chronologique de la Figure 7.13, et les statistiques du Tableau 7.6. Ces valeurs conduisent pour un échantillonnage stratifié aux variances théoriques du Tableau 7.7. Par échantillonnage sans remise en jeu, les valeurs expérimentales obtenues sont indiquées dans le Tableau 7.8. Noter que l'échantillonnage aléatoire stratifié optimum pour 24 données par exemple, conduit à 19 données dans la phase A, ce qui est incompatible avec un échantillonnage sans remise, car supérieur au nombre de données dans la phase. De ce fait, les échantillonnages optimaux des Tableaux 7.7 et 7.8 ne sont pas les mêmes. On constate que *l'échantillonnage opti-*

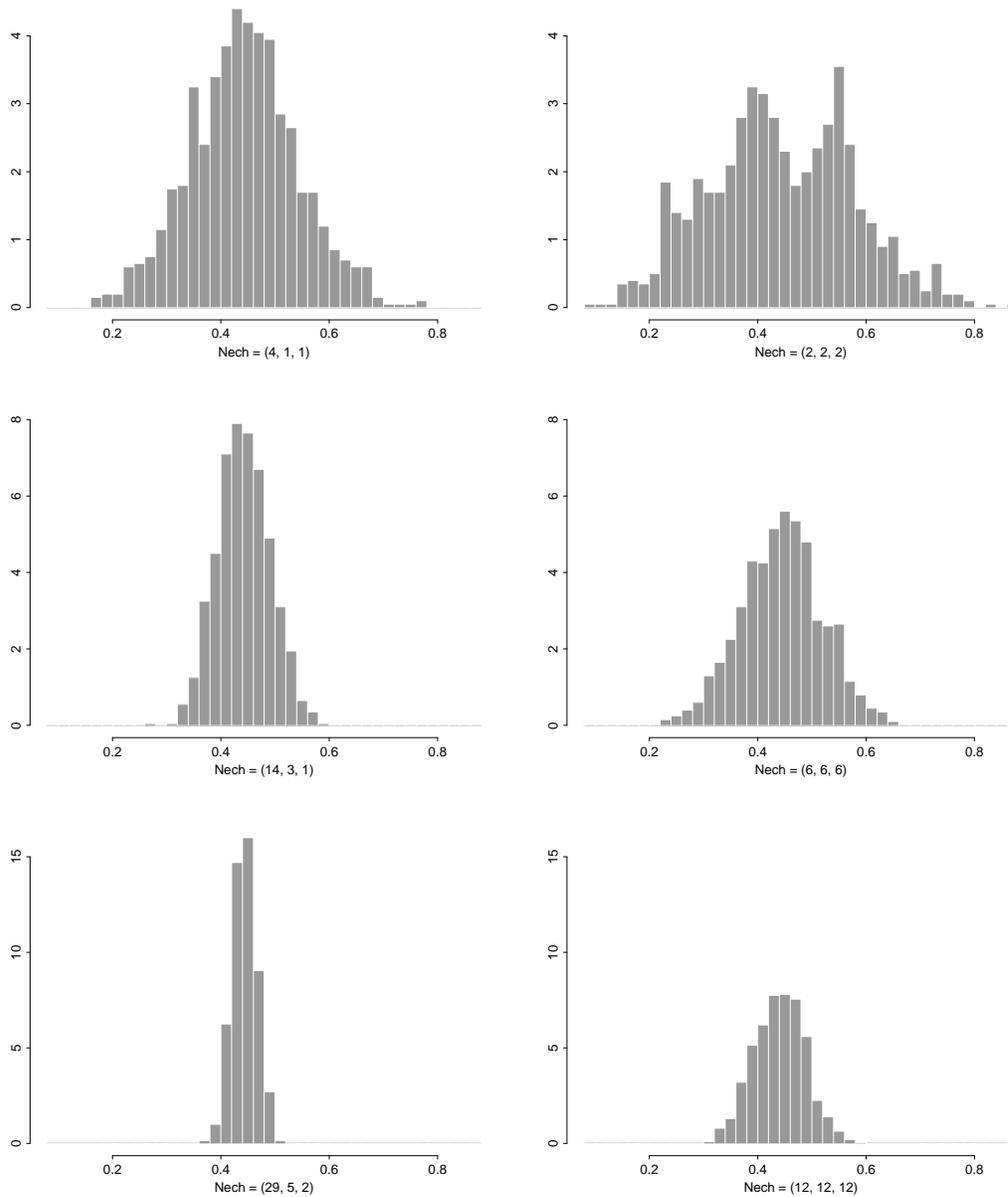


FIGURE 7.12 – Histogrammes obtenus pour les estimateurs basés sur différents échantillonnages stratifiés sans remise de `amenag31` au support 1 minute. La première ligne du graphique est relative aux estimateurs basés sur 6 données, la seconde sur 18 données, et la troisième sur 36. La première colonne est relative à l'échantillonnage optimisé, la seconde à l'échantillonnage à répartition égale.

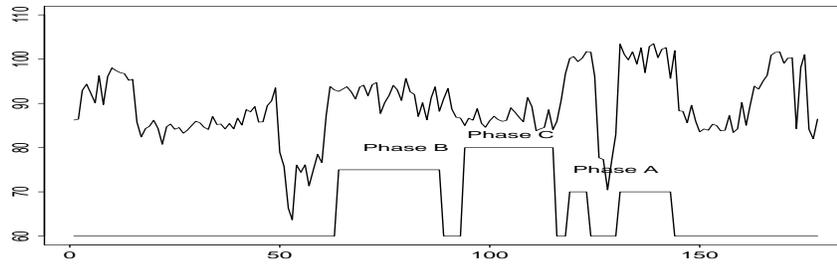


FIGURE 7.13 – Données de amenag31 au support 2 minutes.

Phase	m	σ^2	Effectif
A	1,32	0,30	18
B	0,175	$6,9 \cdot 10^{-3}$	25
C	0,05	$7,0 \cdot 10^{-4}$	22

TABLE 7.6 – Statistiques par phase des données au support 2 minutes.

mum basé sur 18 données est pratiquement équivalent à l'échantillonnage à répartition égale basé sur 36 données, ce que les histogrammes associés (Figure 7.14) permettent de mieux apprécier.

Estimation de la moyenne à long terme

Le problème est, on l'a vu, nettement plus complexe. En outre, il est illusoire de procéder par rééchantillonnage dans une fenêtre d'observation fixe, puisque seule la variabilité à l'intérieur de cette fenêtre particulière peut être prise en compte par ce procédé, alors qu'il est nécessaire de prévoir la variabilité à plus long terme. On procédera par étapes.

Estimation des durées des phases

Cette estimation n'est pas du ressort du statisticien, puisqu'elle résulte de l'analyse des tâches et des processus de fabrication. Pour illustration, nous avons choisi d'estimer la journée complète à partir du seul échantillonnage des phases reconnues. Cela implique

Échantillonnage optimisé				Échantillonnage à répartition égale	
N_A	N_B	N_C	σ^2	$N_A = N_B = N_C$	σ^2
4	1	2	$6,8 \cdot 10^{-3}$	2	0,012
9	2	1	$3,1 \cdot 10^{-3}$	4	$6 \cdot 10^{-3}$
14	3	1	$2,1 \cdot 10^{-3}$	6	$4 \cdot 10^{-3}$
19	4	1	$1,5 \cdot 10^{-3}$	8	$3 \cdot 10^{-3}$
24	5	1	$1,2 \cdot 10^{-3}$	10	$2,4 \cdot 10^{-3}$
28	6	2	$1 \cdot 10^{-3}$	12	$2 \cdot 10^{-3}$

TABLE 7.7 – Variances théoriques pour des échantillonnages aléatoires stratifiés dans le cas de amenag31 au support 2 minutes.

Échantillonnage optimisé				Échantillonnage régulier	
N_A	N_B	N_C	σ^2	$N_A = N_B = N_C$	σ^2
4	1	2	$5,8 \cdot 10^{-3}$	2	0,011
9	2	1	$1,8 \cdot 10^{-3}$	4	$4,7 \cdot 10^{-3}$
14	3	1	$7,7 \cdot 10^{-4}$	6	$2,8 \cdot 10^{-3}$
18	5	1	$2,4 \cdot 10^{-4}$	8	$1,6 \cdot 10^{-3}$
18	9	3	$1,0 \cdot 10^{-4}$	10	$1,0 \cdot 10^{-3}$
18	14	4	$5 \cdot 10^{-5}$	12	$6,7 \cdot 10^{-4}$

TABLE 7.8 – Variances expérimentales d'estimateurs pour des échantillonnages aléatoires sans remise dans le cas de amenag31 au support 2 minutes.

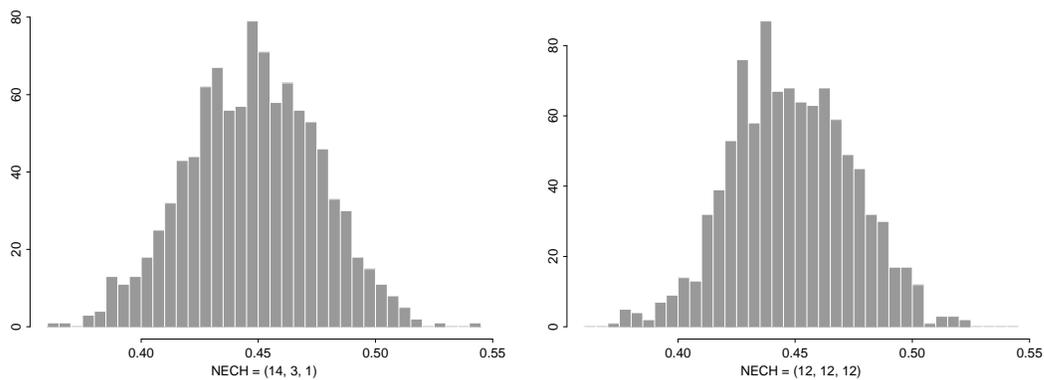


FIGURE 7.14 – Les histogrammes des estimateurs basés sur l'échantillonnage $N = (14, 3, 1)$ et $N = (12, 12, 12)$ pour des données au support 2 minutes sont presque équivalents.

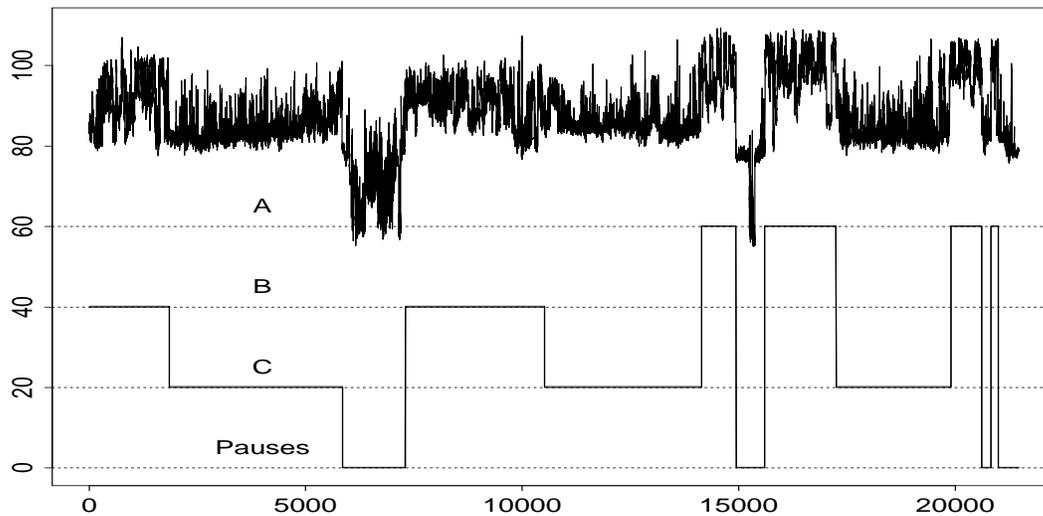


FIGURE 7.15 – Segmentation complète du fichier amenag31 en 3 phases plus pauses.

que l'on puisse identifier les différentes phases sur l'ensemble de la journée, ce que nous avons fait de manière approximative. On distinguera quatre phases, à savoir les trois phases identifiées, et une quatrième phase non échantillonnée, les pauses, auxquelles on affectera une exposition nulle. Elle ne contribue donc pas à l'énergie totale, mais abaisse l'exposition sonore moyenne.

La segmentation complète du jeu de données est illustrée en Figure 7.15. Les moyennes et variances associées sont données par le Tableau 7.9. En se basant sur les fréquences résultant de cette segmentation, p_i^* , et sur les moyennes calculées sur la fenêtre restreinte considérée au paragraphe précédent, on obtient l'estimation suivante de l'exposition sonore moyenne par seconde :

$$\begin{aligned} m^* &= p_A^* m_A^* + p_B^* m_B^* + p_C^* m_C^* \\ &= 0,155 \times 1,334 + 0,236 \times 0,176 + 0,479 \times 0,050 = 0,272 \end{aligned}$$

valeur à rapprocher de la moyenne de la journée, qui est de 0,288. Notons que la prise en compte de la contribution effective de la phase *pauses* (tenue pour nulle dans le calcul précédent) ne rend pas compte de l'écart entre ces deux valeurs, puisque l'on obtient alors $m^* = 0,273$. Cela montre bien les limites de la segmentation effectuée. Ces limites sont encore plus manifestes si on compare les moyennes sur la phase B au niveau de la journée et de la sous-fenêtre (0,224 et 0,176 respectivement).

Il faut en outre spécifier la covariance des estimations des p_i . Comme $\sum_i p_i = 1$, il ne peut s'agir d'estimations décorrélatées. En fait,

$$\forall i \quad \sum_j (p_i^* - p_i)(p_j^* - p_j) = 0 \quad (7.48)$$

et donc :

$$\forall i \quad \sum_j C_{ij}^p = 0 \quad (7.49)$$

Phase	Nombre	Moyenne	Variance
A	3320	1,31	1,73
B	5070	0,224	0,10
C	10 280	0,064	$3 \cdot 10^{-2}$
Pauses	2797	0,012	$2,5 \cdot 10^{-3}$

TABLE 7.9 – Statistiques des énergies par seconde pour la segmentation complète de amenag31.

Si les durées par phase sont estimées avec la même précision, le modèle suivant :

$$\Sigma^p = \sigma_p^2 \begin{pmatrix} 1 & -\frac{1}{n-1} & \dots & \dots \\ -\frac{1}{n-1} & 1 & -\frac{1}{n-1} & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & -\frac{1}{n-1} & 1 \end{pmatrix} = \sigma_p^2 I - \frac{\sigma_p^2}{n-1}(E - I) \quad (7.50)$$

ou E est une matrice à coefficients unité, pourra être adopté. On obtient avec ce modèle :

$$M^t \Sigma^p M = \sigma_p^2 \frac{n}{n-1} \sum_i (m_i - \bar{m})^2 \quad (7.51)$$

Il reste à fixer la variance σ_p^2 . Nous l'avons fait en prenant $\sigma_p^2 = 8,3 \cdot 10^{-4}$, variance que l'on obtiendrait par exemple si l'estimation de la durée de la phase B était une variable uniforme entre 4000 et 6000 secondes.

Variogrammes

Les variogrammes par phase ont été ajustés sur les données au support de la seconde, en éliminant pour la phase C les trois données les plus élevées qui perturbaient trop l'estimation. Ces ajustements qui sont montrés dans la Figure 7.16 correspondent aux modèles suivants :

$$\begin{aligned} \gamma_A(h) &= 0,838 \mathbf{1}_{|h|>0} + 0,914 \text{Sph}\left(\frac{h}{23}\right) + 0,423 \text{Sph}\left(\frac{h}{140}\right) \\ \gamma_B(h) &= 0,031 \mathbf{1}_{|h|>0} + 0,0172 \text{Sph}\left(\frac{h}{33}\right) + 6 \cdot 10^{-3} \text{Sph}\left(\frac{h}{480}\right) \\ \gamma_C(h) &= 3,7 \cdot 10^{-3} \mathbf{1}_{|h|>0} + 1,2 \cdot 10^{-3} \text{Sph}\left(\frac{h}{230}\right) \end{aligned}$$

Ces modèles ont été régularisés au support de la minute, et les variogrammes ainsi obtenus ont été comparés aux variogrammes des données régularisées au même support. Les résultats de cette comparaison sont montrés dans la Figure 7.17. L'accord est bon aux distances accessibles expérimentalement à partir des données régularisées (une minute). Le comportement aux petites distances est pour une large part du au recouvrement des plages. La variance de la phase B semble un peu surestimée, ce qui provient d'un mélange de données provenant d'autres phases au bord du domaine lors de la régularisation (l'attribution d'un segment d'une minute à l'une ou l'autre des phases a été faite, on l'a vu, sur la base du point milieu).

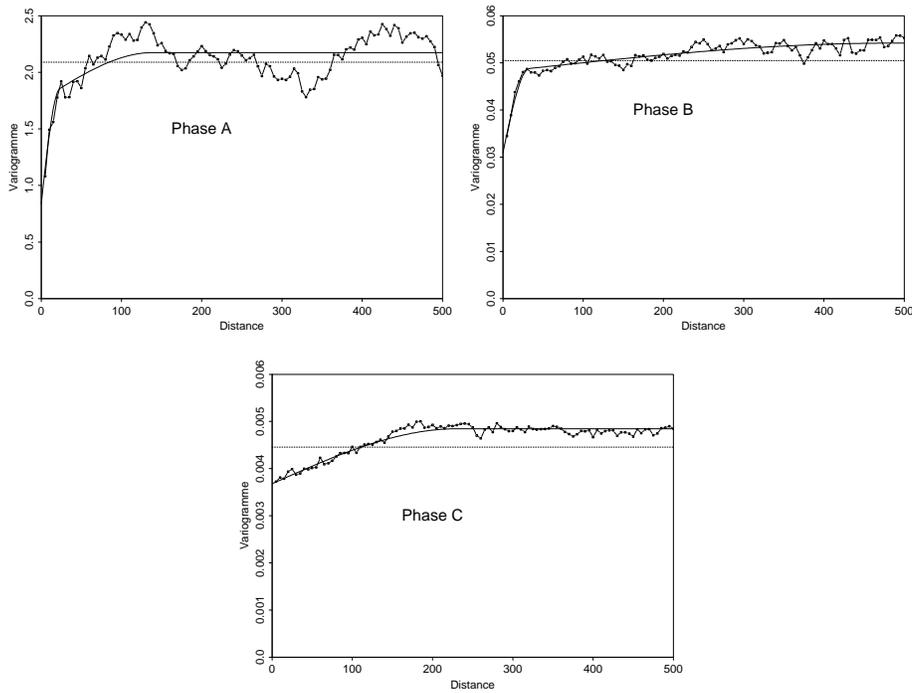


FIGURE 7.16 – Ajustements de variogrammes au support 1 seconde pour amenag31.

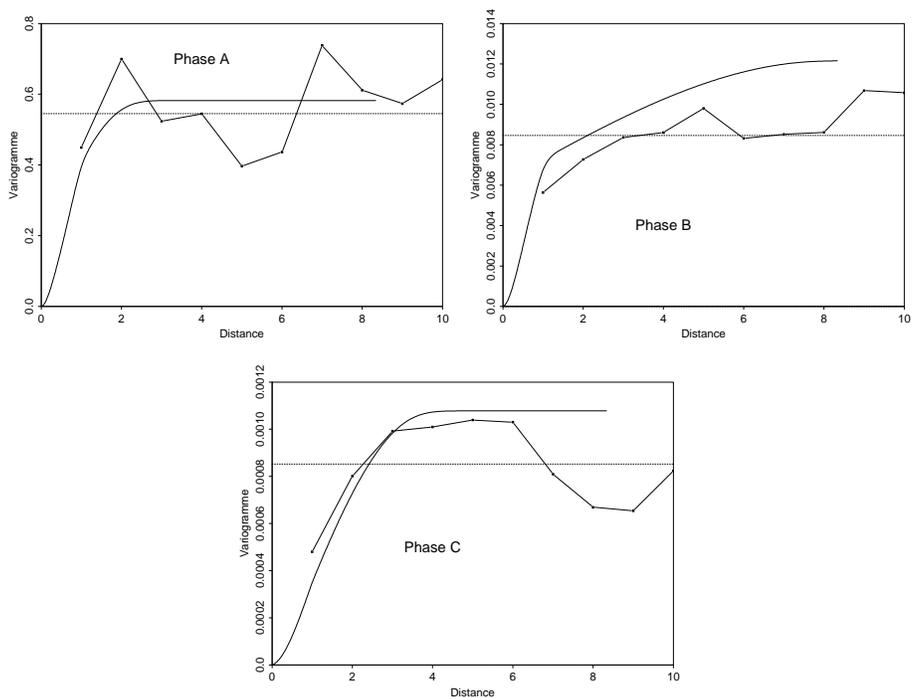


FIGURE 7.17 – Régularisation des variogrammes au support 1 minute, et variogrammes expérimentaux des données régularisées au même support.

Variances d'estimation

Sur la base des variogrammes modélisés, il est possible de calculer les variances d'erreur des estimations des moyennes dans chaque phase, pour des échantillonnages aléatoires, et pour des échantillonnages à points équidistants. Le Tableau 7.10 donne les formules pour les estimations des moyennes à long terme, m_A , m_B et m_C , à partir de données aléatoires de support 1 minute prises dans les sous-fenêtres identifiées initialement. Ces variances se calculent à partir de la covariance, de la taille de fenêtre V , et du support de données v selon les formules :

$$\text{Var}[m_i^* - m_i] = \frac{1}{N_i} \{ \bar{\gamma}(V, V) - \bar{\gamma}(v, v) \} + \sigma^2 - \bar{\gamma}(V, V) \quad (7.52)$$

Le fait que la fenêtre de la phase A soit en deux parties a été pris en compte dans les calculs du Tableau 7.10.

Phase	$\text{Var}\{Z(v_A) - Z(v_A)^*\}$	$\text{Var}\{Z(v_A)\}$
A	$\frac{1}{N_A} 0,554$	$2,8 \cdot 10^{-2}$
B	$\frac{1}{N_B} 1,13 \cdot 10^{-2}$	$8,42 \cdot 10^{-4}$
C	$\frac{1}{N_C} 10^{-3}$	$7,5 \cdot 10^{-5}$

TABLE 7.10 – Variance d'estimation de la moyenne à long terme pour chaque phase, pour un échantillonnage aléatoire pur concentré dans les sous-fenêtres identifiées, et pour un support de 1 minute, en fonction du nombre de données. La variance totale est la somme de la colonne 2 (variance d'estimation de la moyenne sur la sous-fenêtre), et de la colonne 3 (variance de cette moyenne).

A propos de ces valeurs, il convient de formuler la réserve suivante : les valeurs reportées dans la troisième colonne (variance de $Z(v_A) - m_A$ par exemple), dépendent fortement des valeurs du variogramme aux grandes distances. Or, dans la mesure où le variogramme est basé sur les seules données des sous fenêtres, le comportement aux grandes distances résulte de *notre choix de modélisation* (on admet que le bruit sur une autre plage d'une même phase se comporte de manière analogue à ce que l'on observe ici) et non sur une évidence statistique. Il se pourrait fort bien que la variance en question soit en réalité beaucoup plus élevée que ce que l'on prévoit, en raison de causes externes de variabilité non identifiables dans la fenêtre restreinte d'observation. L'hypothèse sur laquelle nous nous basons est optimiste. Seule l'expérimentation pourrait trancher cette question en définitive.

On constate néanmoins que si le terme dû à la variance de la fenêtre (terme $Z(v_i)$) est négligeable lorsque le nombre de données est petit, il ne sert à rien d'augmenter ce nombre au delà de certaines valeurs, le terme de variance de fenêtre restant irréductible.

Nous pouvons aussi comparer ces valeurs avec le résultat d'un échantillonnage à pas régulier dans chaque phase⁵, et à poids optimisés (krigeage de la moyenne à long terme). Cette comparaison est montrée dans le Tableau 7.11.

5. Dans le cas de la phase A, dont la fenêtre d'observation n'est pas contigue, cette régularité est approximative. Cela a été pris en compte dans le calcul.

N	Phase A		Phase B		Phase C	
	aléatoire	pas régulier	aléatoire	pas régulier	aléatoire	pas régulier
2	0,305	0,292	$6,65 \cdot 10^{-3}$	$6,1 \cdot 10^{-3}$	$5,7 \cdot 10^{-4}$	$5,38 \cdot 10^{-4}$
4	0,167	0,146	$3,68 \cdot 10^{-3}$	$3,05 \cdot 10^{-3}$	$3,26 \cdot 10^{-4}$	$2,69 \cdot 10^{-4}$
6	0,120	0,097	$2,73 \cdot 10^{-3}$	$2,03 \cdot 10^{-3}$	$2,43 \cdot 10^{-4}$	$1,79 \cdot 10^{-4}$
8	0,097	0,073	$2,26 \cdot 10^{-3}$	$1,56 \cdot 10^{-3}$	$2,0 \cdot 10^{-4}$	$1,35 \cdot 10^{-4}$
10	0,083	0,058	$1,98 \cdot 10^{-3}$	$1,37 \cdot 10^{-3}$	$1,76 \cdot 10^{-4}$	$1,08 \cdot 10^{-4}$
12	0,074	0,048	$1,79 \cdot 10^{-3}$	$1,25 \cdot 10^{-3}$	$1,59 \cdot 10^{-4}$	$8,97 \cdot 10^{-5}$

TABLE 7.11 – Variances d'estimation des moyennes par phase, à partir des données au support de 1 minute prises dans les phases identifiées initialement, pour un échantillonnage aléatoire et pour un échantillonnage régulier.

Ces valeurs peuvent être utilisées pour calculer les variances d'estimation de la moyenne globale à long terme, en tenant compte des proportions d'occupation de chaque phase (à long terme), et de la variance de l'estimation de cette quantité.

Les résultats que nous donnons à ce sujet sont obtenus sous les conditions suivantes :

- pas de variabilité additionnelle non identifiée sur la fenêtre d'observation. Cela implique que les variances des moyennes sur les fenêtres $Z(v_i)$ sont correctement chiffrées à partir des variogrammes modélisés.
- absence de corrélation des estimations de moyennes d'une phase à l'autre (m_i^* et m_j^* non corrélés). Cela est possible par exemple s'il y a perte de mémoire à la suite d'une transition de phase. Là encore la présence d'un facteur externe de variabilité invaliderait cette hypothèse, et là encore nous privilégions l'alternative optimiste.
- proportions d'occupation des phases à long terme conformément à l'estimation du paragraphe précédent :

$$P^* = (p_A \ p_B \ p_C \ p_0)^t = (0,155 \ 0,236 \ 0,479 \ 0,13)^t \quad (7.53)$$

et variance d'estimation de ces quantités de $\sigma_p^2 = 8,3 \cdot 10^{-4}$.

- moyenne de la phase *pauses* nulle, et connue exactement. Nous négligeons donc délibérément l'exposition qui pourrait être associée à cette phase. On a vu que l'impact de cette approximation était négligeable.

Les variances d'estimation de la moyenne à long terme, pour différentes stratégies d'échantillonnage sont données dans le Tableau 7.12. Les variances obtenues sont plus faibles que dans le cas de l'estimation de la moyenne locale, ce qui peut sembler surprenant, puisque que les estimations des moyennes par phase sont majorées des variances de fenêtres. Cela provient du fait que la part temporelle que nous attribuons à la phase A, est plus faible à long terme que dans la fenêtre d'observation. On verra plus loin l'incidence d'une modification de cette hypothèse d'extrapolation.

On constate que l'échantillonnage aléatoire comportant 3×10 données est un peu moins bon qu'un échantillonnage aléatoire optimisé basé sur $(13, 3, 2)$ données, lui même étant inférieur à un échantillonnage à pas régulier basé sur $(9, 2, 1)$ données et estimation des moyennes par krigeage.

Valeurs estimées

Nous pouvons évaluer les valeurs d'estimation associées à ces échantillonnages, pour fixer les idées. Dans le cas de l'échantillonnage à pas régulier, ces valeurs sont définies

Aléatoire à répartition égale		Aléatoire optimisé		Régulier optimisé	
Nombre	σ_m^*	N_A, N_B, N_C	σ_m^*	N_A, N_B, N_C	σ_m^2
3 × 2	9,4 10 ⁻³	4 1 1	6,4 10 ⁻³	4 1 1	5,3 10 ⁻³
3 × 4	5,7 10 ⁻³	9 2 1	4,2 10 ⁻³	9 2 1	3,4 10 ⁻³
3 × 6	4,5 10 ⁻³	13 3 2	3,5 10 ⁻³	13 3 2	2,6 10 ⁻³
3 × 8	3,9 10 ⁻³	18 4 2	3,1 10 ⁻³	16 5 3	2,3 10 ⁻³
3 × 10	3,6 10 ⁻³	22 5 3	2,9 10 ⁻³	20 6 4	2,2 10 ⁻³
3 × 12	3,3 10 ⁻³	27 6 3	2,8 10 ⁻³	25 7 4	2,1 10 ⁻³

TABLE 7.12 – Variance de l'estimation de la moyenne à long terme pour différents schémas d'échantillonnages stratifiés (Données de amenag31, support 1 minute).

N_A, N_B, N_C	m_A^*	m_B^*	m_C^*	m^*
4 1 1	1,72 / 102,4	0,182 / 92,6	0,045 / 86,5	0,333 / 95,2
9 2 1	1,35 / 101,3	0,15 / 91,8	0,045 / 86,5	0,266 / 94,2
13 3 2	1,21 / 100,8	0,134 / 91,3	0,035 / 85,4	0,236 / 93,7
16 5 3	1,42 / 101,5	0,186 / 92,7	0,058 / 87,6	0,292 / 94,7
20 6 4	1,32 / 101,2	0,180 / 92,6	0,034 / 85,3	0,263 / 94,2
25 7 4	1,24 / 100,9	0,149 / 91,7	0,034 / 85,3	0,244 / 93,9

TABLE 7.13 – Moyennes à long terme estimées par échantillonnage à pas régulier et nombre optimisé. Les premières valeurs sont les expositions sonores moyennes normées, les secondes sont en Décibels

univoquement, et sont consignées dans le Tableau 7.13. Elles sont à rapprocher des valeurs sur la journée, qui sont de $(m_A m_B m_C) = (1,31 0,224 0,064)$ et $m = 0,272$ pour les expositions sonores moyennes, et de $(m_A m_B m_C) = (101,2 93,5 88,1)$ et $m = 94,3$ en Leq.

Dans le cas de l'échantillonnage aléatoire stratifié, l'estimateur n'est pas défini univoquement. Le résultat est donc une distribution conditionnelle, et nous la montrons dans la Figure 7.18 pour des échantillonnages basés sur 6, 18 et 36 données au support 1 minute.

Tous ces résultats sont obtenus avec stratification du jeu de données en 3 phases plus pauses, et prise de mesure dans les trois sous-fenêtres pour lesquelles les phases ont été identifiées par l'INRS.

Comparaison avec un échantillonnage non stratifié

Maintenant nous nous posons la question

“Que gagne t-on à stratifier l'échantillonnage ?”

compte tenu du fait que la stratification que nous utilisons élimine les données dont la phase n'est pas reconnue. Les histogrammes de la Figure 7.19 répondent partiellement à cette question en opposant les histogrammes d'estimateurs en échantillonnage aléatoire pur sans remise, et en échantillonnage aléatoire stratifié (les histogrammes de droite sont les mêmes que dans la figure précédente, à une échelle différente pour faciliter la comparaison). Cette comparaison, pour éloquente qu'elle soit ne répond pas entièrement à la question, puisque seule la variabilité propre à la fenêtre considérée est reflétée par le

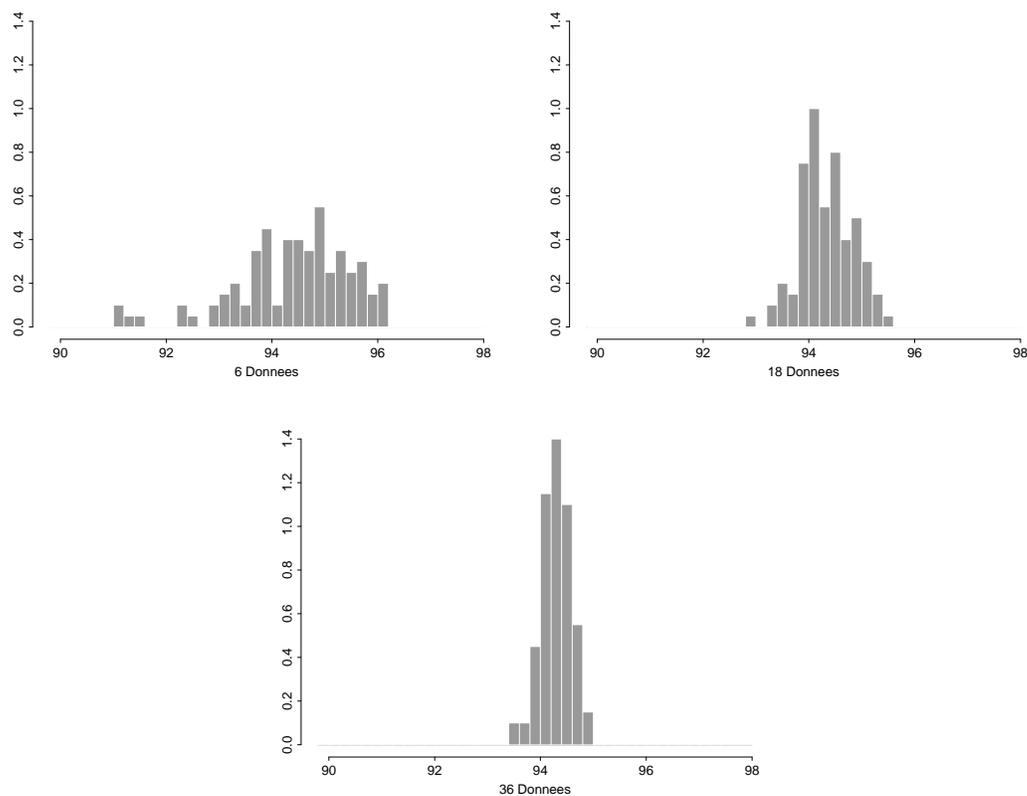


FIGURE 7.18 – Histogrammes des estimateurs de la moyenne à long terme, basés sur un échantillonnage aléatoire stratifié de respectivement 6, 18 et 36 données de support 1 minute.

rééchantillonnage.

Nous pouvons néanmoins, toujours en ayant recours à un modèle d'extrapolation, essayer de donner l'ordre de grandeur de la variance associée à l'estimateur sans stratification. Puisqu'il s'agit d'un schéma aléatoire pur, nous avons toujours :

$$\text{var}(m^*) = \text{Var}[m^* - Z(J)] + \text{Var}[Z(J)] = \frac{1}{N} S^2[1|J] + \sigma^2 - \bar{\gamma}(J, J) \quad (7.54)$$

où J est la journée de travail, et γ est cette fois le variogramme (à long terme) lorsque nous mélangeons les strates. La variance expérimentale des données au support 1 minute peut être utilisé pour avoir l'ordre de grandeur de la variance de dispersion $S^2[1|J]$. Compte tenu de la simplification qui néglige la contribution de la phase *pause* la variance de la moyenne de la journée est :

$$\text{Var}(Z(J)) = \text{Var}\{\bar{p}_A Z(A) + \bar{p}_B Z(B) + \bar{p}_C Z(C)\} \quad (7.55)$$

où nous avons noté \bar{p}_A la proportion (aléatoire) de la journée occupée par la phase A . En notant aussi $P^t = (p_A, p_B, p_C)$ et $M^t = (m_A, m_B, m_C)$, cela conduit encore à une formule de variance du type :

$$\text{Var}(Z(J)) = P^t \Sigma_{Z(V_i)} P + M^t \Sigma_P M + \text{Trace}\{\Sigma_{Z(V_i)} \cdot \Sigma_P\} \quad (7.56)$$

Dans cette expression la contribution due à la variance des moyennes par phase sur la journée peut être négligée en première approximation⁶. Cela conduit, si nous utilisons pour la covariance des \bar{p}_i la forme matricielle $\Sigma_p = \sigma_p^2 \{I + (E - I)/(n - 1)\}$ avec⁷ $\sigma_p^2 = 5 \cdot 10^{-3}$ à la variance d'estimation pour l'échantillonnage aléatoire pur :

$$\text{Var}(m^*) = \frac{1}{N} 0,27 + 2,5 \cdot 10^{-3} \quad (7.57)$$

et donc, pour respectivement 6, 18, et 36 données à des variances de $4,8 \cdot 10^{-2}$, de $1,8 \cdot 10^{-2}$ et de 10^{-2} . Ces chiffres sont pratiquement 10 fois plus élevés que ceux obtenus avec le meilleur échantillonnage stratifié basé sur le même nombre de données.

En définitive, l'intérêt de la stratification est double :

- concentrer l'effort d'échantillonnage sur les phases donnant la plus forte contribution à l'exposition.
- réduire ou corriger l'effet de la variabilité des durées de chaque phase d'une journée à l'autre. Cela suppose que l'analyse des tâches soit capable de donner une meilleure estimation de la durée des phases de travail que la simple observation d'une journée.

Incidence du choix du modèle d'extrapolation

Nous abordons maintenant un autre aspect soulevé par ces estimations. Le passage des valeurs régionales aux valeurs à long terme n'a été possible que par le recours à des hypothèses permettant l'extrapolation. Des alternatives à certaines de ces hypothèses peuvent être introduites à titre d'illustration. On a supposé dans les calculs précédents :

6. La principale contribution, $p_A^2 \text{Var}\{Z(A)\}$ est de l'ordre de 10^{-3}

7. Cette variance correspond par exemple à une durée de phase fluctuant de ± 1 heure sur 8 heures de travail, avec une loi uniforme.

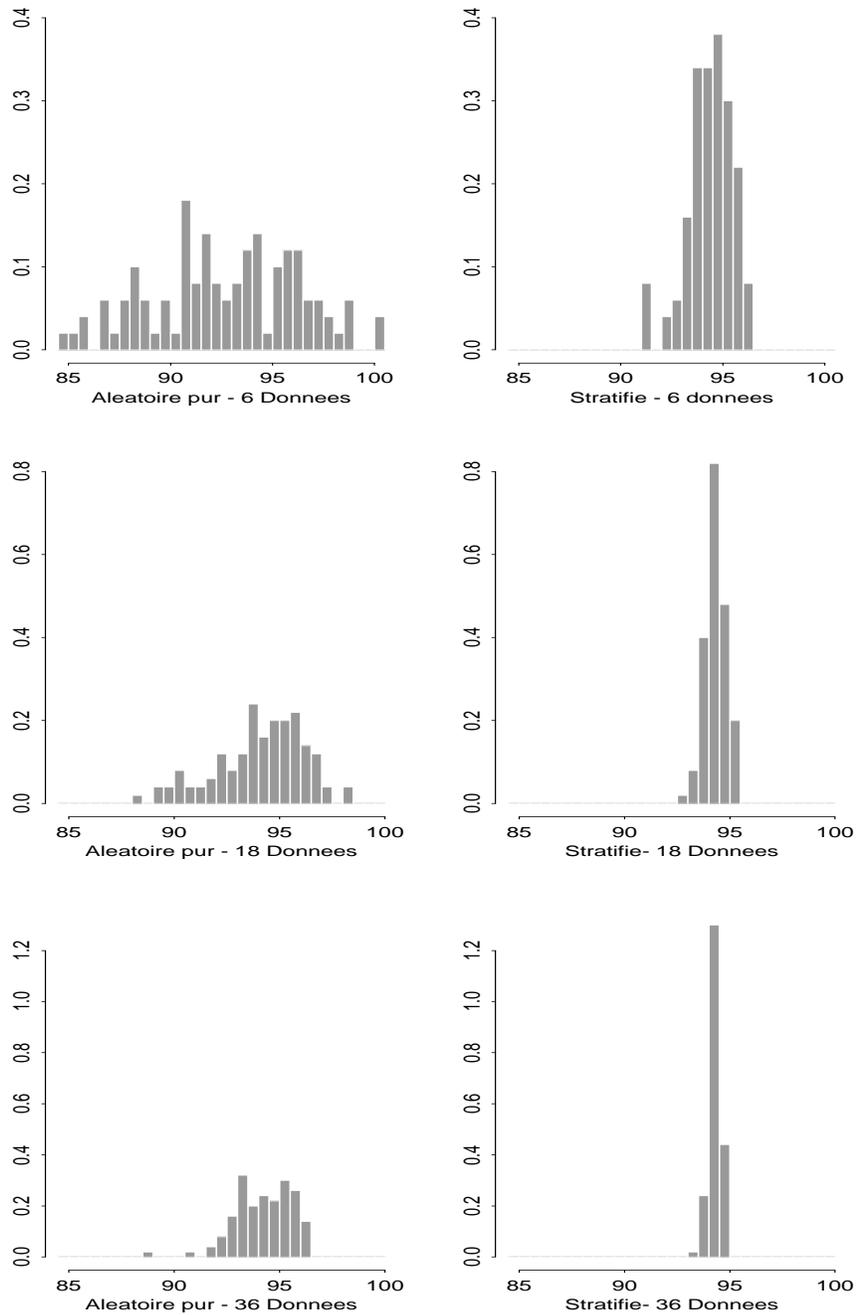


FIGURE 7.19 – Comparaison des estimations par échantillonnage aléatoire pur (à gauche) et aléatoire stratifié optimisé (à droite). Ces histogrammes sont obtenus par rééchantillonnage dans le jeu de données de la journée, et ne reflètent donc pas toute la variabilité à long terme.

- la variabilité à long terme pour une phase donnée entièrement observée sur les sous-fenêtres, ce qui a conduit à une certaine contribution des variances des $Z(v_i)$ à la variance d'estimation des moyennes m_i . La contribution ainsi calculée est minimale, c'est à dire qu'elle se place dans l'alternative optimiste.
- les proportions occupées par les différentes phases sont estimées par $p_A^* = 0,155$, $p_B^* = 0,236$ et $p_C^* = 0,479$.
- la précision de ces chiffres est reflétée par la variance commune $\text{Var}(p_i^*) = 8,3 \cdot 10^{-4}$.
- les fluctuations sont indépendantes d'une phase à l'autre.

A titre d'alternative, nous apporterons maintenant les modifications suivantes au modèle :

- une structure supplémentaire à très longue portée est introduite dans chacune des phases, avec diverses variances. Cette portée est grande par rapport à la journée, par exemple de l'ordre du mois. La composante aléatoire associée est pratiquement constante dans chaque phase à l'échelle de la journée de travail, et sa variabilité n'est donc pas perceptible au niveau de la fenêtre d'observation. Les variances que nous associons sont respectivement de 0,5 de $5 \cdot 10^{-3}$ et de $8 \cdot 10^{-4}$ dans les phases A , B et C . Pour fixer les idées, disons que ces variances autorisent la moyenne locale de la phase A à passer de 101 à 104 décibels d'un jour à l'autre. Pour la phase B , cela permet de passer de 93 à 95 décibels, et de 88 à 90 pour la phase C . De telles fluctuations pourraient par exemple être dues à un défaut de réglage d'une machine, qui aurait des répercussions différentes suivant le type de travail.

La présence de cette nouvelle composante aux variogrammes ne modifiera pas les variances de dispersion, puisque pour cette composante $\bar{\gamma}(v, v) \approx 0$, mais les variances des moyennes locales $Z(v_i)$ reçoivent la totalité de sa variance. Il en est donc ainsi également des variances des estimations des moyennes. La formule pour l'estimation de m_A par échantillonnage aléatoire devient :

$$\text{Var}(m_A^*) = \frac{1}{N_A} 0,554 + 0,53 \quad (7.58)$$

La décroissance de cette variance avec l'effort d'échantillonnage est donc vite limitée par la variance de fenêtre, qui vaut maintenant 0,53. La seule manière de faire chuter efficacement la variance de m_A^* est de procéder à une seconde campagne de mesures à une date suffisamment éloignée pour éliminer les effets de la grande portée. A titre de comparaison, 10 échantillons dans la fenêtre A donnent une variance de :

$$\text{Var}(m_A^*) = 0,59 \quad (7.59)$$

tandis le même nombre d'échantillons pris en deux campagnes donnent :

$$\text{Var}(m_A^*) = 0,32 \quad (7.60)$$

Un autre effet de la structure supplémentaire est que les estimations de moyennes sont maintenant corrélées. Si la matrice de variance-covariance initiale est notée Σ_m^0 , elle devient :

$$\Sigma^m = \Sigma_m^0 + S S^t \quad \text{avec} \quad S^t = (\sqrt{0,5} \quad \sqrt{5 \cdot 10^{-3}} \quad \sqrt{8 \cdot 10^{-4}}) \quad (7.61)$$

L'estimation de la moyenne globale, obtenue par composition des estimations sur chaque phase a une variance donnée par la relation déjà rencontrée :

$$\text{Var}(m^*) = M^t \Sigma^p M + P^t \Sigma^m P + \text{Trace}(\Sigma^p \Sigma^m) \quad (7.62)$$

Aléatoire à répartition égale		Aléatoire optimisé		Régulier	
N	σ_m^{*2}	N_A, N_B, N_C	σ_m^{*2}	N_A, N_B, N_C	σ_m^{*2}
3×2	$7,18 \cdot 10^{-2}$	4, 1, 1	$6,35 \cdot 10^{-2}$	4, 1, 1	$6,21 \cdot 10^{-2}$
3×4	$6,23 \cdot 10^{-2}$	9, 2, 1	$5,77 \cdot 10^{-2}$	9, 2, 1	$5,6 \cdot 10^{-2}$
3×6	$5,92 \cdot 10^{-2}$	14, 3, 1	$5,6 \cdot 10^{-2}$	14, 3, 1	$5,4 \cdot 10^{-2}$
3×8	$5,76 \cdot 10^{-2}$	19, 4, 1	$5,5 \cdot 10^{-2}$	18, 5, 1	$5,35 \cdot 10^{-2}$
3×10	$5,66 \cdot 10^{-2}$	24, 5, 1	$5,47 \cdot 10^{-2}$	21, 7, 2	$5,33 \cdot 10^{-2}$
3×12	$5,60 \cdot 10^{-2}$	28, 6, 2	$5,44 \cdot 10^{-2}$	27, 7, 2	$5,31 \cdot 10^{-2}$
$3 \times (3 \times 3)$	$2,18 \cdot 10^{-2}$	$3 \times (7, 1, 1)$	$1,99 \cdot 10^{-2}$	$3 \times (7, 1, 1)$	$1,93 \cdot 10^{-2}$

TABLE 7.14 – Variances pour les trois types d'échantillonnages stratifiés envisagés en fonction du nombre de données au support 1 minute, pour le modèle alternatif. Ce modèle comporte une composante de variance à grande portée, et attribue un poids plus élevé à la phase A que le modèle de référence. La dernière ligne du tableau correspond à un échantillonnage en trois campagnes indépendantes.

L'augmentation de variance d'estimation associée à l'introduction de la nouvelle structure est donc :

$$\text{Var}(m^*) - \text{Var}(m_0^*) = (P^t S)^2 + \text{Trace}(S S^t \Sigma^p) \quad (7.63)$$

$$= \left(\sum_i p_i S_i \right)^2 + \sigma_p^2 \frac{n}{n-1} \sum_i (S_i - \bar{S})^2 \quad (7.64)$$

n étant ici le nombre total de phases. On remarque que les termes supplémentaires ne dépendent pas de la répartition des échantillons, et donc que les stratégies optimales ne sont pas modifiées ainsi.

- nous modifions aussi les proportions des différentes phases, en prenant maintenant :

$$(p_A \ p_B \ p_C \ p_0) = (0,25 \ 0,36 \ 0,32 \ 0,07) \quad (7.65)$$

avec une variance de $\sigma_p^2 = 3 \cdot 10^{-3}$. Ces proportions correspondent à peu près aux valeurs données par la fenêtre d'observation, si nous supposons que les pauses comptent pour une demi heure par journée de travail.

Avec un tel modèle on doit s'attendre à une sérieuse dégradation des résultats, puisque

1. Les moyennes par phases sont considérées comme moins bien estimées,
2. Les estimateurs associés sont maintenant corrélés.
3. L'incidence de la phase A , qui est dominante en terme de variance est augmentée.
4. La précision de l'estimation de ces proportions est dégradée.

Compte tenu de ce qui vient d'être dit, on s'attend à ce que les différentes stratégies se démarquent assez peu en terme de variance. Nous avons en fait obtenu les valeurs des Tableaux 7.14 d'où il ressort que le bénéfice lié à l'optimisation est assez marginal. De plus la décroissance de la variance avec l'effort d'échantillonnage est rapidement insignifiante, toujours à cause de la composante de variance à long terme. Par contre, si au lieu d'une seule campagne comportant 30 échantillons d'une minute, nous organisons l'échantillonnage en trois campagnes indépendantes de 9 données, nous obtenons les variances de la dernière ligne du tableau, avec un gain substantiel.

En conclusion à cette partie traitant de l'estimation de la moyenne à long terme, nous ferons les remarques suivantes :

- contrairement à l'estimation de la moyenne locale, l'estimation de la moyenne à long terme demande de faire des hypothèses extrapolatoires. Ces hypothèses concernent la variabilité à long terme, et, dans le cas d'une stratification, les pourcentages d'occupation des différentes phases.
- en présence d'une variabilité à long terme, il est illusoire de chercher à optimiser un échantillonnage concentré sur une plage trop petite. Seul le fractionnement dans le temps des campagnes est capable de conduire à une amélioration des estimations.
- dans le cas d'une stratification, la proportion d'occupation à long terme doit être estimé avec précision, car elle est facteur de biais, et conduit à l'inefficacité d'une optimisation.
- si la variabilité à long terme est expliquée par la seule stratification (et non par une composante de variance à grande portée), l'optimisation de la répartition des échantillons permet en gros de diviser par deux le nombre de données nécessaires à l'obtention d'un certain niveau de précision, compte tenu des contrastes de variance observés.

7.4 Applications pratiques

Les conclusions de ce chapitre tiennent en quelques points que nous énumérons :

- la stratification de l'échantillonnage est très intéressante dans le cas traité, les contrastes de variance étant très marqués.
- cette stratification présente l'intérêt de permettre de concentrer l'effort sur les phases possédant les variances les plus fortes, donc les phases les plus difficiles à estimer, et aussi sur les phases ayant la plus grande contribution.
- en outre, les durées des différentes phases sur le long terme peuvent être différentes des durées pendant la journée mesurée. Cela est une cause de biais et d'augmentation des variances d'estimation que la stratification permet d'éliminer, dans la mesure où l'analyse des tâches est capable de conduire à une bonne estimation des durées à long terme.
- l'utilisation de krigeage avec un échantillonnage à pas régulier, en complément de la stratification permet encore une amélioration supplémentaire.
- nous avons examiné l'effet d'une corrélation à long terme. Cet effet est double, puisque les estimations des moyennes par phase se trouvent dégradés, et que ces estimations deviennent corrélées. Le gain apporté par une optimisation d'échantillonnage, dans le cadre d'une campagne limitée à une seule journée, est alors marginal. La seule solution efficace en définitive est la multiplication des campagnes de mesure.
- nous insistons encore sur la différence fondamentale entre l'estimation de la moyenne de la journée mesurée, et l'estimation de la moyenne à long terme. La première estimation est possible sans recours à un modèle probabiliste à travers des schémas aléatoires purs ou stratifiés, et il est facile de construire des intervalles de confiance non paramétriques par rééchantillonnage. La seconde estimation nécessite une modélisation probabiliste, et l'introduction d'hypothèses permettant l'extrapolation. On a illustré l'incidence de ces hypothèses, ce qui montre que la prudence est de mise en la matière, et une confirmation expérimentale est toujours

–pour le moins– souhaitable.

Chapitre 8

Conclusions générales

8.1 Synthèse des conclusions partielles

Les séries de mesure de L_{Aeq} courts étudiées ici présentent une variabilité intra-journée importante. Elle a été quantifiée notamment par l'étendue des distributions, comprise entre 23 et 48 dB(A) quand le temps d'intégration employé dans le mesurage est de 10 s. S'agissant de mesures d'exposition réalisées par technique dosimétrique, les résultats du Chapitre 3 montrent que cette dispersion des valeurs s'explique notamment par le fait que les opérateurs subissent des *pointes d'exposition*, dès qu'ils doivent intervenir à proximité immédiate d'une machine et réaliser des tâches brèves mais plus bruyantes qu'habituellement. De telles circonstances d'exposition sont relativement courantes en milieu professionnel. Les résultats montrent que l'hypothèse d'une distribution gaussienne des séries mesurées n'est jamais vérifiée dans les séries étudiées. En présence de *pointes d'exposition*, des indicateurs de sélectivité ont été proposés pour quantifier cet effet de pointe dans l'exposition sonore quotidienne. Ils permettront dorénavant de distinguer divers types de situations d'exposition sonore par des valeurs quantifiées et non plus seulement par des mentions purement qualitatives.

Une autocorrélation a été mise en évidence indiquant l'existence de deux échelles de temps, la minute et l'heure. L'analyse des circonstances de l'exposition permet d'interpréter partiellement ce constat : l'autocorrélation à l'échelle de la minute semble liée au cycle de base du processus de fabrication alors qu'à l'échelle de l'heure, l'alternance des tâches et de la présence épisodique de pointes de bruit pourrait l'expliquer. Un aspect original des résultats du Chapitre 4 est d'avoir montré que dans toutes les séries étudiées, les distances d'autocorrélation dépendaient de la durée d'intégration élémentaire des L_{Aeq} courts, paramètre de mesure fixé ici respectivement à 10 s, 1 mn et 5 mn. De même, il a été montré que l'allure des histogrammes des distributions dépendait de la valeur de ce paramètre. La dépendance de l'autocorrélation à l'égard de la durée d'intégration ne peut pas être prise en compte par les méthodes statistiques classiques alors que la géostatistique l'intègre dans ce qu'elle nomme l'*effet de support*. Ces caractéristiques reflètent la structure de variation en temps des données expérimentales et leur prise en compte dans

les estimations de l'exposition sonore moyenne et de sa variance passe par le choix d'un modèle de variogramme.

Dans le Chapitre 5, trois modélisations différentes des variogrammes expérimentaux ont été employées afin de comparer leurs aptitudes respectives à décrire la structure temporelle des données. Le premier modèle testé, un modèle *lognormal autohomothétique* dit de Wijsien, est séduisant par sa simplicité et intéressant parce que les études d'exposition aux substances chimiques emploient couramment un modèle proche (le modèle de Larsen [14]). Il a été montré que ce modèle était inadapté à la modélisation des données d'exposition au bruit. Par contre, les deux autres modèles employés, un modèle *lognormal stationnaire* et un modèle *lognormal localement stationnaire* modélisent correctement la variabilité des données expérimentales.

Cette modélisation est obtenue sans formuler d'hypothèse sur les lois de distribution des valeurs mesurées, ce qui constitue une extension importante des méthodes actuelles d'estimation de l'exposition sonore. Après validation de deux modèles par les données expérimentales observées durant une journée, on s'est intéressé à la tendance des modèles au-delà de la journée. Il a été montré que ces tendances sont différentes, ce qui pose la question de leur choix. Il relève d'une analyse des circonstances de l'exposition sonore et des possibilités de variation à long terme, au-delà de la journée observée.

Ce point pourra être précisé en disposant de données observées durant plusieurs journées. Mais dorénavant une méthode d'estimation de l'exposition sonore à long terme est disponible, qui évite de supposer la permanence de l'hypothèse de normalité des distributions des valeurs mesurées, hypothèse inapplicable aux données étudiées ici.

Le Chapitre 6 a eu pour objet de bâtir des estimateurs de l'exposition sonore moyenne et de sa précision qui soient adaptés à la structure en temps des données. Une analyse critique des estimateurs proposés dans deux normes de mesurage de l'exposition sonore en milieu professionnel a montré leurs limites et leur inadéquation aux caractéristiques des données étudiées ici. Plusieurs estimateurs ont été élaborés, sur la base du modèle lognormal appliqué aux valeurs d'exposition sonore,

- en distinguant les estimations à court terme, dans la période observée, des estimations à long terme, au-delà du domaine observé,
- en précisant les relations liant le mode de répartition des échantillons et les estimateurs de moyenne et de variance employés,
- en traitant le cas de données indépendantes et celui de données autocorrélées.

Ce travail a permis de confirmer le caractère optimum de l'estimateur de l'exposition sonore quotidienne normalisé lorsque l'échantillonnage est régulier et continu, mais il a été montré ici que cet estimateur n'était plus optimum lorsque l'échantillonnage devient fragmentaire. A l'aide de ces estimateurs, l'impact de l'autocorrélation des données expérimentales a été quantifié sur les bornes d'un intervalle de confiance de l'exposition sonore estimée à long terme : pour les 3 séries étudiées ici, la borne supérieure de cet intervalle dif-

fére de plus de 5 dB(A) du niveau d'exposition sonore moyen. . . Ceci confirme une observation générale : l'emploi de méthodes ne tenant pas compte de l'autocorrélation des données sous-estime systématiquement l'étendue des intervalles de confiance. Toutefois l'ampleur du biais d'estimation mis en évidence sur les données étudiées ici impose une confirmation par d'autres données.

La méthode géostatistique d'analyse a été appliquée à une série de données comportant 3 phases d'activités bien distinctes durant la journée observée. Après étude des caractéristiques temporelles des phases d'activités, 4 stratégies d'échantillonnage présentées au Chapitre 7 ont été appliquées à ces données. Distinctes par le nombre et la répartition des échantillons, ces stratégies induisent des variances d'estimation de l'exposition quotidienne différentes. En comparant les variances d'estimation liées à chacune des stratégies, il a été montré qu'il était possible d'obtenir la même précision d'estimation en réduisant de 30 à 11 le nombre des échantillons. En outre, l'observation d'une journée singulière peut être employée pour estimer l'exposition lors d'une journée "moyenne", durant laquelle les durées relatives des phases d'exposition sont spécifiées indépendamment de la journée observée.

8.2 Prolongements

Des prolongements à ce travail méthodologique peuvent être envisagés dans diverses directions.

1. En études et recherches :

- Application à diverses situations d'exposition des méthodes, estimateurs et logiciels élaborés ici.
- Etude des variations inter-journées et estimations de l'exposition sonore à long terme. Méthodes de validation de la tendance à long terme des modèles de variogramme
- Constitution progressive d'une banque de données d'exposition sonore, regroupant pour chaque situation étudiée : analyse des circonstances d'exposition, caractérisation quantifiée par les indices de sélectivité, variogrammes expérimentaux.
- Application à des postes de travail faisant l'objet de mesurage répétés périodiquement de stratégies d'échantillonnage optimisées et comparaison des précisions d'estimations avec celles des stratégies classiques.

2. Valorisation des résultats acquis :

- Proposition d'indicateurs de sélectivité, pour quantifier l'effet des pointes de bruit dans l'exposition sonore globale.
- Clarification de normes de mesurage à l'aide du modèle log-normal appliqué à l'exposition sonore.
- Méthode de prise en compte de la dépendance entre la durée d'intégration du mesurage et les indicateurs classiques de dispersion des distributions (histogrammes, percentiles, écart-type, . . .)
- Estimation de la précision de mesures d'exposition sonore quotidienne compte tenu de l'autocorrélation des valeurs mesurées.

Cette liste ne prétend pas être exhaustive. Ce rapport a été rédigé notamment dans le but de constituer un support aux prolongements évoqués ici.

Bibliographie

- [1] AITCHISON J & BROWN JAC (1957) *The Lognormal Distribution : with special reference to its use in economics*. Cambridge University Press, Cambridge, 176p.
- [2] AFNOR (1987) *Méthode de mesurage des niveaux sonores en milieu de travail en vue de l'évaluation du niveau d'exposition sonore quotidienne des travailleurs*. Norme Française, NF S 31-084, Paris.
- [3] AFNOR (1995) *Workplace atmospheres – Guidance for the assessment of exposures by inhalation to chemical agents for comparison with limit values and measurement strategy*, Norme Française, NF EN 689, Paris.
- [4] BASTIDE MJC (1988) *Estimation et mesure du niveau acoustique continu équivalent*. Revue de Statistique Appliquée, 36, 5–14.
- [5] BROWN FH (1995) *ANSI-C routines for the estimation of the lognormal mean*. Computers & Geosciences, 21, 1191–1199.
- [6] DIN (1990) *Mittelung von Schallpegeln*. Deutsche Norm, DIN 45 641, Berlin.
- [7] EFRON B & TIBSHIRANI RJ (1993) *An Introduction to the Bootstrap*. Chapman & Hall, New York, 436p.
- [8] HUIJBREGTS CJ (1971) *Reconstitution du variogramme ponctuel à partir d'un variogramme expérimental régularisé*. Publication N-244, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, 25p.
- [9] ISO (1995) *Acoustics – Guidelines for the measurement and assessment of exposure to noise in a working environment*. International Organization for Standardization, ISO/DIS 9612.2, Geneva.
- [10] JOURNAL AG & HUIJBREGTS CJ (1978) *Mining Geostatistics*. Academic Press, London, 600p.
- [11] KENDALL M & STUART A (1977) *The Advanced Theory of Statistics*. Volume 1, 4th edition, Griffin, London, 472p.
- [12] LAJAUNIE C (1993) *Les bases de la modélisation géostatistique de l'effet de support*. Les Notes Scientifiques et Techniques de l'INRS, 104, 15–28.
- [13] LAJAUNIE C (1993) *L'estimation Géostatistique Non Linéaire*. Notes de cours C-152, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.
- [14] LARSEN RI (1969) *A new mathematical model of air pollutant concentration averaging time and frequency*. Journal of the Air Pollution Control Association, 19, 24–30.
- [15] MALCHAIRE J (1994) *Programmes de Conservation de l'Audition : Organisation en Milieu Industriel*. Masson, Paris.
- [16] MATHERON G (1955) *Application des méthodes statistiques à l'évaluation des gisements*. Annales des Mines, 144 (12), 50–75.
- [17] MATHERON G (1962) *Traité de Géostatistique Appliquée*. Technip, Paris, volume 1, 333p.
- [18] MATHERON G (1965) *Les Variables Régionalisées et leur Estimation*. Masson, Paris, 305p.

- [19] MATHERON G (1970) *La Théorie des Variables Régionalisées et ses Applications*. Fascicule 5, Les Cahiers du Centre de Morphologie Mathématique, Ecole des Mines de Paris, Fontainebleau, 211p.
- [20] MATHERON G (1974) *Effet proportionnel et lognormalité ou : le retour du serpent de mer*. Publication N-374, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, 43p.
- [21] MATHERON G (1976) *Les concepts de base et l'évolution de la géostatistique minière*. In : Guarascio M et al. (eds) *Advanced Geostatistics in the Mining Industry*, 3–10, NATO ASI Series C 24, Reidel, Dordrecht.
- [22] MATHERON G (1982) *La sélectivité des distributions*. Publication N-686, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, 40p.
- [23] ORFEUIL JP (1973) *Interprétation statistique du modèle de Larsen*. Publication N-413, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, 8p.
- [24] PATEL NR (1973) *Comment on : a new mathematical model of air pollutant concentration*. *Journal of the Air Pollution Control Association*, 23, 291–292.
- [25] PRÉAT B (1987) *Application of geostatistical methods for estimation of the dispersion variance of occupational exposures*. *American Industrial Hygiene Association Journal*, 48, 877–884.
- [26] RIVOIRARD J (1990) *A review of lognormal estimators for in situ reserves*. *Mathematical Geology*, 22, 213–221.
- [27] RIVOIRARD J (1994) *Introduction to Disjunctive Kriging and Non-Linear Geostatistics*. Oxford University Press, Oxford, 181p.
- [28] SAPORTA G (1990) *Probabilités, Analyse des Données et Statistique*. Technip, Paris, 493p.
- [29] SICHEL HS (1966) *The estimation of means and associated confidence limits for small samples from lognormal populations*. *Journal of the South African Institute of Mining and Metallurgy*, March 1966, 106–122.
- [30] SINGPURWALLA ND (1972) *Extreme values from a lognormal law with applications to air pollution problems*. *Technometrics*, 14, 703–711.
- [31] THIÉRY L, LOUIT P, LOVAT G, LUCARELLI D, RAYMOND F, SERVANT JP & SIGNORELLI C (1994) *Exposition des Travailleurs au Bruit : Méthode de Mesurage*, Editions INRS, ED 772, Paris.
- [32] THIÉRY L, WACKERNAGEL H & LAJAUNIE C (1996) *Geostatistical analysis of time series of short LAeq values*. *Proceedings of Inter-Noise96*, Liverpool, 2065–2068.
- [33] VDI (1993) *Anwendung statistischer Methoden bei der Kennzeichnung schwankender Geräuschmissionen*. VDI-Richtlinien, VDI 3723 Blatt 1, Verein Deutscher Ingenieure, Düsseldorf.
- [34] WACKERNAGEL H (1995) *Multivariate Geostatistics*, Springer-Verlag, Berlin, 256p.
- [35] WACKERNAGEL H, LAJAUNIE C, THIÉRY L, VINCENT R & GRZEBYK M (1996) *Applying geostatistics to exposure monitoring data in industrial hygiene*. In : Soares A, Gomez-Hernandez J & Froidevaux R (eds) *geoENV I : Geostatistics for Environmental Applications*, Kluwer, Amsterdam.
- [36] WIJS HJ DE (1951) *Statistics of ore distribution, part I : frequency distribution of assay values*. *Geologie en Mijnbouw*, 13, 365–375.
- [37] WIJS HJ DE (1953) *Statistics of ore distribution, part II : theory of the binomial distribution applied to sampling and engineering problems*. *Geologie en Mijnbouw*, 15, 12–24.

Annexe A

Optimisation de l'échantillonnage

A.1 Introduction

Les séries temporelles associées à une nuisance sont souvent liées à une variable discrète prenant un petit nombre de modalités. Il s'agira par exemple des différentes phases de fonctionnement d'un atelier ou des différentes ambiances auxquelles un travailleur peut être exposé durant les diverses tâches qu'il réalise au cours d'une journée, s'il s'agit de mesures d'exposition. La répartition des mesures dans les différentes phases d'exposition peut avoir une influence sur l'estimation de l'exposition sonore totale et de sa variance. Les lignes qui suivent sont destinées à ouvrir une discussion sur le problème de l'optimisation de l'échantillonnage dans un tel contexte.

On considère un processus temporel continu, $Z(t)$, qui représentera une nuisance, et un processus discret $I(t)$, associé comme suggéré ci dessus à l'état du système dans lequel les mesures Z sont prises. Le processus d'état I ne prendra qu'un nombre fini de valeurs $I \in \{1, \dots, N\}$. La spécification d'un tel modèle peut être limitée à l'ordre deux pour les problèmes abordés ici. Par exemple, les moments d'ordre deux conditionnels suivants peuvent être supposés connus :

$$E[Z(t) | I(t) = i] = \mu_i \quad (\text{A.1})$$

$$\text{Cov}(Z(t), Z(t') | I(t'') = i; \forall t'' \in [t, t']) = C_i(t, t') \quad (\text{A.2})$$

On ne cherche pas ici à préciser quels modèles pourraient être utilisés pour $C_i(t, t')$. Ouvrons cependant une parenthèse pour citer au moins un exemple possible qui a le mérite de la simplicité :

Soit une fonction stationnaire gaussienne réduite, Y , de covariance $\rho(h)$, indépendante de $I(t)$, et soient N variances logarithmiques $\sigma_i^2; i = 1, \dots, N$. Le modèle conditionnellement lognormal :

$$Z(t) = \mu_{I(t)} \exp \left\{ \sigma_{I(t)} Y(t) - \frac{\sigma_{I(t)}^2}{2} \right\}$$

donne les covariances conditionnelles stationnaires suivantes :

$$C_i(t, t') = (\mu_{I(t)})^2 \left(e^{\sigma_i^2 \rho(t'-t)} - 1 \right)$$

Ce modèle peut évidemment être généralisé à peu de frais, en considérant à la place de la transformation exponentielle ci-dessus, une anamorphose générale, indexée par un paramètre discret.

Cela étant, la *nuisance à estimer* peut prendre différentes formes :

- Il peut s'agir simplement de la moyenne arithmétique de Z sur un intervalle fixé à l'avance :

$$Z([0, T]) = \frac{1}{T} \int_0^T Z(t) dt$$

Ce qui suppose bien sur la variable Z additive, par exemple une quantité de substance chimique récupérée sur un filtre par unité de temps, ou en acoustique, une puissance instantannée, alors la moyenne représente une puissance moyenne (l'intégrale \int_0^T représente l'énergie accumulée), etc..

- Cela peut être la fraction du temps total au cours de laquelle le dépassement d'un certain seuil critique a eu lieu¹ :

$$\mathcal{T}([0, T]) = \frac{1}{T} \int_0^T 1_{Z(t) \geq z_c} dt$$

- Plus généralement, une nuisance peut être représentée par la valeur moyenne sur $[0, T]$ d'une transformation non linéaire donnée h , c'est à dire par :

$$h([0, T]) = \frac{1}{T} \int_0^T h(Z(t)) dt$$

Ainsi, si l'effet pathogène de Z est fortement non linéaire, par exemple parce que la croissance de l'effet est beaucoup plus rapide que celle du niveau Z , la nuisance globale est mieux représentée par une quantité telle que $h([0, T])$ que par $Z([0, T])$.

On envisagera ici l'estimation de la première quantité, les autres étant du ressort de la géostatistique non linéaire.

L'échantillonnage de Z sur $[0, T]$ peut être réalisé pratiquement dans des conditions très différentes.

- L'échantillonnage peut être quasiment ponctuel, et réalisé à des instants librement choisis par l'échantillonneur dans l'intervalle $[0, T]$, par exemple à maille régulière sur tout l'intervalle.
- Les autres conditions étant les mêmes que précédemment, l'échantillonnage peut être librement réalisé sur un sous-domaine imposé de $[0, T]$ (par exemple échantillonnage pendant deux jours, alors que l'intervalle à estimer représente une semaine).
- La variable $I(t)$ peut être connue seulement à posteriori. Dans ce cas les temps d'échantillonnage t_α sont fixés indépendamment des valeurs de I .
- I peut au contraire être connue avant mesure, par exemple s'il s'agit des différentes phases d'un processus cyclique. Le choix des instants d'échantillonnage peut alors être optimisé en tenant compte de cela.
- L'échantillonnage peut avoir un support (temps de chaque prélèvement) fixe -ce qui est hautement souhaitable pour l'analyse statistique- ou au contraire variable.

Les exemples ci dessus n'épuisent évidemment pas la combinatoire. On envisagera seulement quelques cas de figure ici. Bien que la théorie n'exige pas une telle restriction, on se référera par la suite à des modèles stationnaires (correspondants à des variogrammes présentant un palier), des phénomènes intrinsèques stricts (correspondants à des variogrammes sans paliers) étant assez peu vraisemblables dans ce contexte (absence de dérive, et retour au niveau 0 en l'absence d'action extérieure).

1. Rappelons la définition d'une indicatrice, $1_{\{A\}}$, fonction qui vaut 1 lorsque la condition $\{A\}$ est réalisée et 0 sinon. Il s'ensuit que dans l'intégrale ci dessus, seules contribuent les zones telles que $Z(t) \geq z_c$

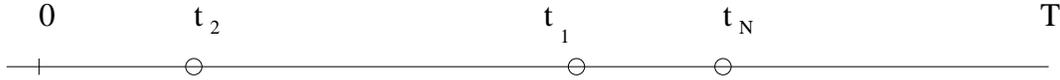


FIGURE A.1 – Echantillonnage aléatoire pur.

A.2 Echantillonnage libre, et fonction état I inconnue ou ignorée

Ce cas est peu réaliste, et on l'envisage ici pour permettre quelques rappels. La simplification résultant de la non prise en compte de l'état I nous ramène à une situation habituelle en géostatistique linéaire. Les outils structuraux considérés ici sont donc non conditionnels, puisque I est ignoré. On considère classiquement trois schémas d'échantillonnage dans ce contexte :

Echantillonnage aléatoire pur

Dans ce schéma, les implantations des échantillons sont uniformes dans le domaine étudié, et mutuellement indépendantes (Figure A.1). On note par la suite t_α la position du α -ième échantillon, et plus généralement \underline{t} une implantation uniforme dans $[0, T]$.

Fixons ensuite, pour un instant, la fonction aléatoire Z , qui devient ainsi une fonction déterministe, ce que l'on indiquera par l'usage de minuscules². Cela revient à raisonner conditionnellement à l'ensemble de la réalisation de Z .

La loi de probabilité de chaque échantillon est ainsi obtenue par l'intégrale suivante :

$$P[z(\underline{t}) < u] = F_{[0,T]}(u) = \frac{1}{T} \int_0^T 1_{z(x) < u} dx$$

Les données prélevées constituent alors un N -échantillon au sens habituel pour la loi ci-dessus³, et les estimateurs standards peuvent être utilisés :

$$(E[z(\underline{t})])^* = \frac{1}{N} \sum z(t_\alpha) = \bar{z} \quad (\text{A.3})$$

$$\text{Var}(z(\underline{t}))^* = \frac{1}{N-1} \sum (z(t_\alpha) - \bar{z})^2 \quad (\text{A.4})$$

Comme on a aussi :

$$\begin{aligned} E[z(\underline{t})] &= z[0, T] \\ \text{Var}(z(\underline{t})) &= \frac{1}{T} \int_0^T (z(x) - z[0, T])^2 dx \end{aligned}$$

il en résulte que la moyenne arithmétique \bar{z} peut être prise comme estimateur de $z[0, T]$ et que la variance d'une tel estimateur est obtenue par :

$$\text{Var}(z[0, T] - \bar{z}) = \frac{1}{N} \text{Var}(z(\underline{t})) \approx \frac{1}{N(N-1)} \sum (z(t_\alpha) - \bar{z})^2$$

2. La géostatistique transitive (MATHERON [18]) se place dans cette perspective d'échantillonnage de fonction déterministe, le seul aléas provenant en définitive de l'échantillonnage lui-même.

3. C'est-à dire N variables indépendantes, suivant la loi $F_{[0,T]}$



FIGURE A.2 – Echantillonnage centré pour l'estimation de l'intervalle $[0, T]$.

L'interprétation en géostatistique intrinsèque de cette démarche est réalisée en rendant à z son caractère aléatoire, et en passant à l'espérance pour chiffrer la précision :

$$\text{Var}(Z[0, T] - \bar{Z}) = \frac{1}{N} \text{Var}(Z(\underline{t}) - Z[0, T])$$

Or cette dernière quantité est la *variance de dispersion* de mesures ponctuelles dans le champ $[0, T]$, quantité que l'on note $D^2[o|T]$, et qui s'obtient facilement à partir du variogramme. Ainsi :

$$\text{Var}(Z[0, T] - \bar{Z}) = \frac{1}{N} D^2[o|T]$$

Les deux mesures d'erreur quadratique ci dessus, la conditionnelle et la non conditionnelle, diffèrent en pratique assez peu si le champ est assez grand vis à vis de la portée.

Mentionons le fait que si le point de vue déterministe suffit à chiffrer la précision d'estimation sans nécessiter d'analyse structurale ni de modélisation variographique, il présente l'inconvénient de ne pas permettre d'aller beaucoup plus loin. En particulier :

- Il ne permet pas d'analyser des schémas d'échantillonnage plus généraux, et donc d'optimiser les prélèvements.
- Il ne permet pas d'estimation locale (estimation de sous-domaines), ni de traiter de l'extrapolation à des durées plus grandes que $[0, T]$.
- Il ne permet pas la prise en compte d'effet de support.

Cela, sans parler des possibilités offertes par les méthodes de simulation, ni de la géostatistique non linéaire.

Echantillonnage centré

Si l'effort d'échantillonnage conduit à N mesures, ces mesures seront disposées au centre de segments de longueur T/N (Figure A.2).

Cela conduit à l'estimateur :

$$Z([0, T])^* = \frac{1}{N} \sum_{i=1}^N Z(t_\alpha) \quad \text{pour } t_\alpha = \frac{\alpha - 0.5}{N} \quad \alpha = 1, \dots, N$$

La variance d'estimation, définie dans le modèle probabiliste comme :

$$(\sigma^*)^2 = \text{Var}[Z([0, T]) - Z([0, T])^*]$$

se calcule aisément en terme de covariance (non conditionnelle) de Z , ou de variogramme, par la formule :

$$(\sigma^*)^2 = \bar{C}(V, V) - 2 \sum_{\alpha} \frac{1}{N} \bar{C}(V, t_\alpha) + \frac{1}{N^2} \sum_{\alpha, \beta} C(t_\alpha, t_\beta) \quad (\text{A.5})$$

expression dans laquelle on a noté V l'intervalle $[0, T]$ pour alléger quelque peu la notation, et :

$$\begin{aligned}\overline{C}(V, t_\alpha) &= \frac{1}{T} \int_0^T C(u, t_\alpha) du \\ \overline{C}(V, V) &= \frac{1}{T^2} \int_0^T \int_0^T C(u, v) dudv\end{aligned}$$

La relation A.5, en dépit de son apparente trivialité, se prête mal à un calcul numérique. En effet, le dernier terme, c'est à dire celui comportant une double sommation, est une approximation numérique d'assez bonne qualité du premier. Il en est de même du second terme. Donc, si le calcul des \overline{C} est réalisé par intégration numérique, on obtient pour résultat la différence entre deux schémas d'approximation d'une intégrale double. Cela montre que, à moins d'utiliser un schéma de très grande précision dans le calcul des \overline{C} , on obtient à peu près n'importe quoi.

Heureusement, à une dimension, l'intégration est très souvent réalisable exactement. Plus généralement (en particulier en dimension deux et trois), on a recours à un "principe d'approximation" qui donne de bons résultats. Ce principe revient, en notant $Z(v_\alpha)$ la moyenne de Z sur le segment $[\tau_{\alpha-1}, \tau_\alpha]$ contenant t_α , à considérer les erreurs d'approximations élémentaires suivantes :

$$Z(v_\alpha) - Z(t_\alpha)$$

comme orthogonales deux à deux. Il en résulte alors que les variances d'erreurs correspondantes peuvent être composées :

$$\begin{aligned}\text{Var}(Z(V) - Z(V)^*) &= \text{Var} \left\{ \frac{1}{N} \sum_{\alpha} (Z(v_\alpha) - Z(t_\alpha)) \right\} \\ &= \frac{1}{N^2} \sum_{\alpha} \text{Var}(Z(v_\alpha) - Z(t_\alpha))\end{aligned}$$

Chacune de ces variances élémentaires porte le nom de variance d'extension du point dans le segment, et est notée $\sigma^2[o|v]$ ⁴. Ces variances d'extension sont d'un calcul élémentaire, et finalement la relation de variance approchée est :

$$\text{Var}(Z(V) - Z(V)^*) = \frac{1}{N} \sigma^2[o|v]$$

Echantillonnage aléatoire stratifié

Ce type d'échantillonnage consiste à partitionner comme précédemment l'intervalle $[0, T]$ en segments égaux (Figure A.3). Ensuite un échantillon est prélevé dans chaque segment, d'implantation aléatoire uniforme, avec indépendance mutuelle. On note ainsi t_α l'implantation du α -ème point.

Fixons ensuite, comme dans le cas de l'échantillonnage aléatoire pur, la fonction aléatoire Z . Les erreurs d'estimations élémentaires :

$$z(v_\alpha) - z(t_\alpha)$$

4. Plus généralement, la variance d'extension d'un volume v dans un volume v' , notée $\sigma^2[v|v']$, est la variance de l'erreur commise en estimant $Z(v')$ par $Z(v)$.



FIGURE A.3 – Echantillonnage aléatoire stratifié.

sont ainsi indépendantes du fait de l'indépendance même des implantations. Les variances d'estimation se composent alors comme précédemment :

$$\text{Var}(z(V) - z(V)^* | Z) = \frac{1}{N} \sum \text{Var}(z(v_\alpha) - z(t_\alpha) | Z)$$

En prenant l'espérance en Z , on a finalement :

$$\text{Var}(Z(V) - Z(V)^*) = \frac{1}{N} D^2[o|v]$$

expression dans laquelle on a noté $D^2[o|v]$ la variance de dispersion d'un point dans un segment élémentaire v ⁵.

On a donc dans les deux cas les variances d'estimation des moyennes sur $[0, T]$ en terme de variogramme.

A.3 Echantillonnage libre sur un sous ensemble $V' \subset V$

On envisage le cas où l'échantillonnage doit nécessairement être réalisé pendant une durée totale plus petite que T , de manière à limiter la durée de l'intervention (par exemple quelques jours consécutifs, alors que l'objectif est l'estimation de moyenne sur plusieurs semaines). On ne fait toujours pas intervenir de processus d'état.

Le problème peut être abordé avec la décomposition :

$$Z(V) - Z^* = Z(V) - Z(V') + Z(V') - Z^*$$

Z^* étant encore simplement la moyenne arithmétique des échantillons. Là encore, le calcul de variance peut être simplifié par une approximation analogue à celle du paragraphe précédent. Les échantillons étant tous prélevés dans V' , on supposera orthogonalité des deux termes d'erreur ci-dessus. Il vient :

$$\text{Var}(Z(V) - Z^*) = \sigma^2[V'|V] + \text{Var}(Z(V') - Z^*)$$

le second terme se calculant comme précédemment. Ainsi, il n'y a pas de difficulté particulière, tout au moins en principe, à chiffrer la précision de l'estimation.

Le problème vient en fait de l'inférence du variogramme, puisque l'échantillonnage ne nous permet d'observer la variabilité du phénomène que sur les distances inférieures

5. Plus généralement, lorsque deux supports v et V sont compatibles, au sens où V peut être décrit comme l'union disjointe de translats de v , la variance de dispersion de v dans V , notée $D^2[v|V]$ est la variance de $Z(\underline{v}) - Z(V)$ où \underline{v} désigne l'un des translats de v réalisant la partition de V , choisi suivant une loi uniforme. Cette variance se calcule à partir du variogramme par la relation : $D^2[v|V] = \bar{\gamma}(V, V) - \bar{\gamma}(v, v)$. La variance de dispersion ponctuelle $D^2[o|V]$ est l'espérance de la variance expérimentale que l'on trouverait en échantillonnant V par prélèvements uniformes et indépendants.

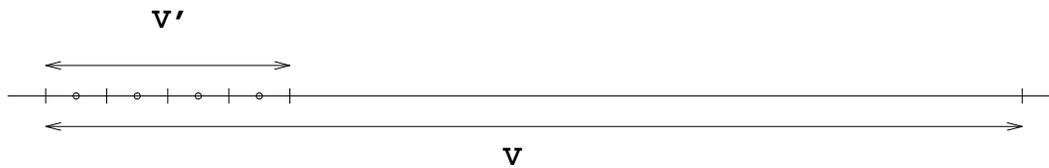


FIGURE A.4 – Echantillonnage d'un sous-domaine.

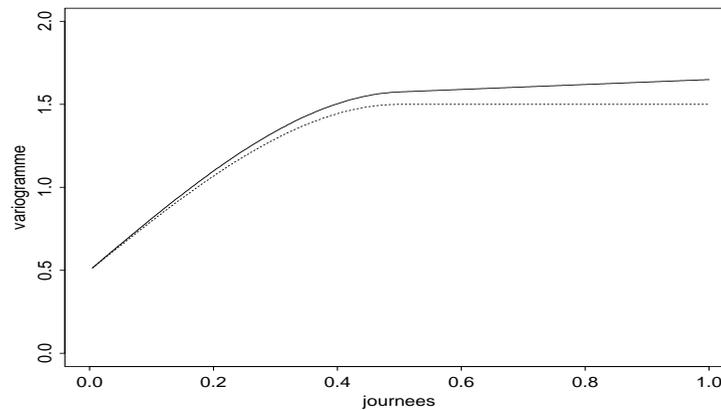


FIGURE A.5 – Variogramme pour des intervalles inférieurs à la journée.

aux dimensions de V' ⁶. Il en résulte que le calcul du terme $\sigma^2[V'|V]$, lequel sollicite la modélisation variographique aux distances supérieures à celles de V' , sera largement tributaire d'hypothèses extrapolatoires.

En effet, lorsque nous concluons sur la base d'un variogramme expérimental reposant par exemple sur des données collectées pendant une demi-journée, à l'existence d'un palier pour un intervalle d'une heure, par exemple, cela ne fait que traduire la variabilité propre à la demi-journée particulière étudiée. Il se pourrait fort bien que des données prises pendant plusieurs journées mettent en évidence de nouvelles plages de variabilité, par exemple parce que les données de l'après-midi correspondraient à une autre phase de travail, à d'autres conditions d'environnement, etc. Pour employer une terminologie plus proche de celle des praticiens de l'analyse des tableaux de variances, la variabilité inter demi-journées n'est pas prise en compte par l'échantillonnage. Elle est donc conjecturée, et c'est ce que l'on fait en extrapolant le variogramme à des distances supérieures à V' . Cela n'est bien sûr pas sans risques, et en pratique, il est recommandé d'utiliser des alternatives très tranchées, de manière à savoir à quoi l'on peut s'attendre.

A titre d'illustration de la remarque précédente, considérons le cas d'un phénomène observé pendant une journée, et pour lequel on souhaite une estimation de moyenne sur 10 jours. La Figure A.5 montre deux modèles différant assez peu sur des distances de l'ordre de la demi-journée, et donc tous deux compatibles avec un même jeu de données portant sur une seule journée.

L'examen des deux modèles aux distances supérieures (Figure A.6 permet de mieux

6. En réalité, l'analyse conduite dans MATHERON [19] montre que l'estimation de variogramme n'est même possible dans de bonnes conditions que jusqu'aux distances de l'ordre de la moitié du domaine échantillonné.

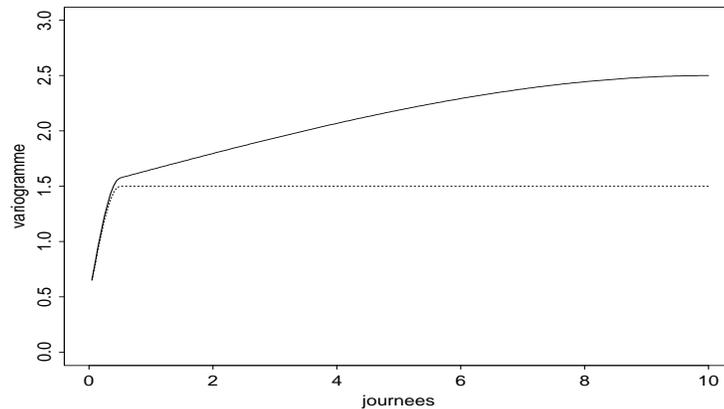


FIGURE A.6 – Variogramme sur 10 jours.

apprécier les différences. Ces deux modèles sont en fait :

$$\gamma_1(h) = 0.5\text{Pep} + 1.\text{Sph}\left(\frac{h}{0.5}\right)$$

avec h compté en journées, et :

$$\gamma_2(h) = 0.5\text{Pep} + 1.\text{Sph}\left(\frac{h}{0.5}\right) + 1.\text{Sph}\left(\frac{h}{10}\right)$$

Ils donnent des variances d'estimation comparables pour le terme $Z(V') - Z^*$, estimation de la moyenne de la journée, puisque le calcul des termes $\sigma^2[o|v_\alpha]$ ne fait intervenir que les valeurs du variogramme aux petites distances. Par contre, le second terme, variance d'extension de V' dans V dépendra beaucoup du modèle, puisqu'un calcul effectif donne pour les deux modèles ci-dessus :

$$\sigma_1^2[V'|V] = 0.50 \quad \text{et} \quad \sigma_2^2[V'|V] = 0.88$$

La variabilité supplémentaire introduite par la structure de portée 10 jours, laquelle n'est pas accessible expérimentalement avec un échantillonnage limité à la journée, est donc supérieure à 65 %.

En pratique, on a donc le choix entre :

- Donner deux variances d'estimation correspondant à deux modèles extrêmes compatibles avec les données.
- Revenir 10 jours après la première série de mesures, et prendre quelques mesures supplémentaires pour préciser la variabilité à cette échelle.

A.4 Echantillonnage avec état I connu

On suppose connus :

- L'état I au moment t_α de chaque prise d'échantillon.
- Les durées totales de chaque état dans l'intervalle $[0, T]$:

$$T_i = \int_0^T 1_{\{I(t)=i\}} dt$$

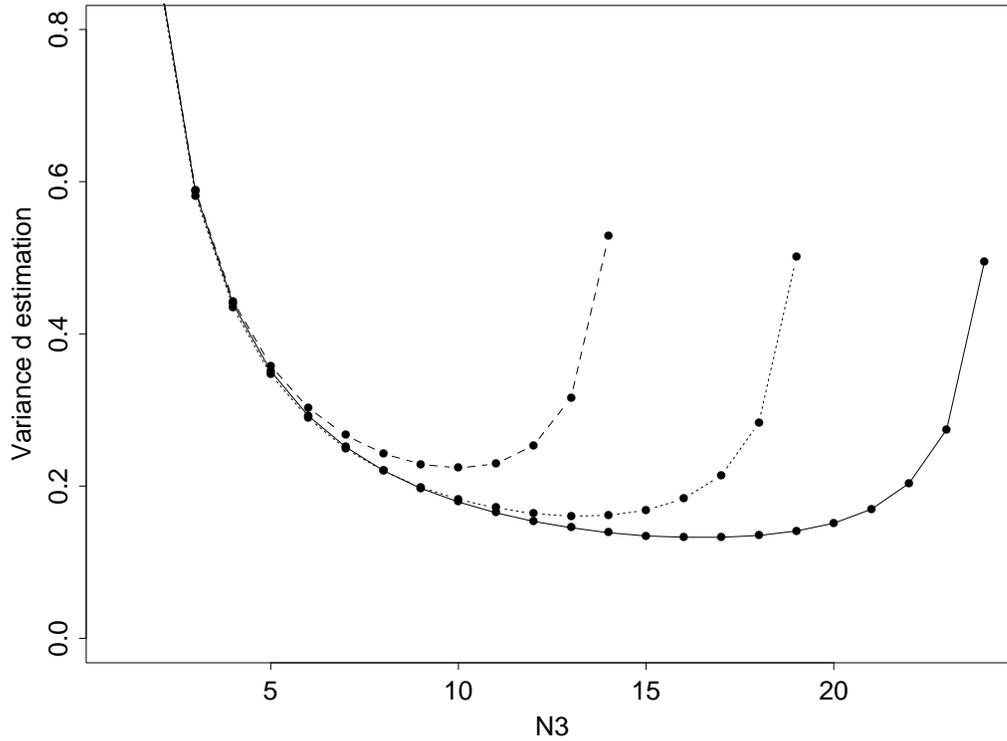


FIGURE A.7 – Variances d'estimations pour trois phases, de covariances exponentielles (de facteur d'échelle unité), de variances $S_i = 1, 2$ et 4 , et de longueur 100 . Un nombre total de 30 données sont prélevées suivant un schéma centré dans chaque plage. Les trois courbes correspondent à un nombre dans la première plage de $N_1 = 15, 10$ et 5 données, les plus faibles variances étant obtenues dans le premier cas.

Dans ces conditions, on pourra procéder par estimations séparées des quantités :

$$Z_i(V) = \frac{1}{T_i} \int_0^T Z(t) 1_{\{I(t)=i\}} dt$$

à l'aide des estimateurs usuels :

$$Z_i^* = \frac{1}{N_i} \sum_{\alpha} Z(t_{\alpha}) 1_{\{I(t_{\alpha})=i\}}$$

Puis par reconstruction :

$$Z(v)^* = \sum_i \frac{T_i}{T} Z_i^*$$

Le calcul des variances se fera avec l'approximation d'orthogonalité des termes d'erreur :

$$Z_i(V) - Z_i^*$$

La variance de chacun de ces termes dépend de la covariance conditionnelle propre à chaque situation $I = i$. Par exemple, dans le cas d'un échantillonnage centré, si le domaine $\{t : I(i) = i\}$ est partitionné en N_i segments de longueur $l_i = T_i/N_i$, et si $\sigma_i^2[0|l_i]$ est la variance d'extension dans un tel segment, pour la covariance conditionnelle C_i ,

$$\text{Var}(Z_i(V) - Z_i^*) = \frac{1}{N_i} \sigma_i^2[0|l_i]$$

et, globalement :

$$\text{Var}(Z(V) - Z(V)^*) = \sum_i \left(\frac{T_i}{T}\right)^2 \frac{1}{N_i} \sigma_i^2[0|l_i]$$

A titre d'exemple, si les covariances conditionnelles sont toutes exponentielles de même facteur d'échelle, et ne diffèrent que par leur variance, c'est à dire si elles sont de la forme :

$$C_i(t, t') = S_i^2 e^{|t-t'|}$$

on trouve les variances d'extension :

$$\sigma_i^2[0|l_i] = S_i^2 \left\{ 1 - \frac{2}{l_i} - \frac{2}{l_i^2}(1 - e^{-l_i}) + \frac{4}{l_i}e^{-l_i/2} \right\}$$

et finalement :

$$\text{Var}(Z(V) - Z(V)^*) = \sum_i \left(\frac{T_i}{T}\right)^2 \frac{S_i^2}{N_i} \left\{ 1 - 2\frac{N_i}{T_i} - 2\left(\frac{N_i}{T_i}\right)^2(1 - e^{-T_i/N_i}) + 4\frac{N_i}{T_i}e^{-T_i/(2N_i)} \right\}$$

Cette formule permet, si les variances S_i^2 et les durées T_i sont connues, de comparer différentes stratégies d'échantillonnage en terme de variance d'estimation de $Z(V)$.

La Figure A.7 montre les résultats dans le cas de trois modes, pour lesquels les variances conditionnelles sont respectivement de $S_1^2 = 1$, $S_2^2 = 2$ et $S_3^2 = 4$. La longueur de chaque plage associée est $T_i = 100$, quelque soit le mode. En outre, on décide de prendre un total de 30 échantillons, à répartir dans les trois phases. La figure montre trois courbes, correspondant, de gauche à droite, à $N_1 = 15$, $N_1 = 10$ et $N_1 = 5$ valeurs prises dans la phase 1, la valeur portée en abscisse est le nombre affecté à la phase 3. En examinant de manière plus systématique l'ensemble des configurations possibles, on trouve que la répartition optimale des 30 échantillons est obtenue pour :

$$N_1 = 4 \quad N_2 = 9 \quad N_3 = 17$$

Pour conclure ce paragraphe, mentionons le fait que si l'on a supposé connue exactement la durée totale de chaque état dans le domaine à échantillonner, rien n'empêche, et c'est certainement une situation plus réaliste, de considérer cette valeur comme entachée d'une certaine erreur.

A.5 Discussion

Soulignons que la relation utilisée dans le paragraphe précédent fait usage des variances par phase, mais n'utilise pas les moyennes. Elle conduit à intensifier l'effort d'échantillonnage dans les plages où la variance est la plus élevée, mais non nécessairement dans les plages de plus haut niveau. Si très souvent les deux évoluent dans le même sens, c'est à dire que à moyenne élevée est associée une variance élevée, cela n'est en aucune manière toujours le cas.

Cela appelle deux remarques :

- Le critère quadratique d'optimalité retenu, présente deux défauts. D'une part il donne un poids égal à une erreur par excès et à une erreur par défaut. Et d'autre part il pénalise autant l'erreur d'estimation d'un niveau faible, que l'erreur d'un niveau

élevé. Cette dernière remarque conduit à introduire une pondération, et donc à remplacer le critère :

$$\text{Var}(Z(V) - Z^*) = E[(Z(V) - Z^*)^2] = \int E[(Z(V) - Z^*)^2 \mid Z(V) = z] F(dz)$$

par :

$$\int E[(Z(V) - Z^*)^2 \mid Z(V) = z] W(z) F(dz)$$

où W est librement choisie par l'utilisateur. Un tel critère est malheureusement sensiblement moins maniable que le précédent.

- La moyenne arithmétique sur le domaine V n'est sans doute pas non plus l'indicateur de nuisance le plus approprié. Les valeurs maximales, les pourcentages de dépassement de seuils critiques par exemple, sont des indicateurs d'exposition probablement plus pertinents. Cela conduit à l'estimation de l'histogramme dans V , c'est à dire à l'estimation de quantités telles que :

$$1_z(V) = \frac{1}{|V|} \int_V 1_{\{Z(x) < z\}} dx$$