



**HAL**  
open science

# Étude des indicateurs acoustiques dynamiques chez les enfants lors d'un trajet école-domicile

François Vieille

► **To cite this version:**

François Vieille. Étude des indicateurs acoustiques dynamiques chez les enfants lors d'un trajet école-domicile. Méthodologie [stat.ME]. 2014. dumas-01059868

**HAL Id: dumas-01059868**

**<https://dumas.ccsd.cnrs.fr/dumas-01059868>**

Submitted on 2 Sep 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université De Strasbourg  
UFR Mathématiques-Informatique

Stage de Master 1 Biostatistiques et Statistiques Industrielles  
CHU de Besançon - Service CMC  
Maître de stage: F. Mauny  
du 02.06.14 au 25.07.14

Etude des indicateurs acoustiques dynamiques chez les enfants lors d'un trajet  
école - domicile

François VIEILLE

Strasbourg, le 15 août 2014  
vieille.francois@gmail.com

# Table des matières

<b>Remerciements</b>	<b>4</b>
<b>Introduction</b>	<b>5</b>
<b>1 Quelques notions d'acoustique</b>	<b>6</b>
1.1 Le son . . . . .	6
1.2 Courbe de pondération A . . . . .	7
1.3 Appareils de mesure . . . . .	7
1.3.1 Le sonomètre . . . . .	8
1.3.2 Le dosimètre . . . . .	8
1.4 Indicateurs sonores énergétiques . . . . .	8
1.4.1 Le niveau équivalent . . . . .	8
1.4.2 Le niveau d'exposition . . . . .	8
1.5 Indicateurs sonores statistiques . . . . .	8
<b>2 L'étude</b>	<b>10</b>
2.1 Introduction . . . . .	10
2.2 Population . . . . .	10
2.3 Les campagnes de mesures . . . . .	11
2.3.1 Au sonomètre . . . . .	11
2.3.2 Au dosimètre . . . . .	11
2.4 Premiers résultats . . . . .	11
<b>3 Les indicateurs acoustiques dynamiques</b>	<b>13</b>
3.1 Définitions . . . . .	13
3.1.1 Indicateur d'étalement sonore . . . . .	13
3.1.2 Indicateur de rugosité sonore . . . . .	14
3.1.3 Indicateur du degré de l'augmentation sonore . . . . .	14
3.1.4 Indicateur de bruyance . . . . .	14
3.1.5 Indicateur de tranquillité . . . . .	15
3.1.6 Indicateur Harmonica . . . . .	16
3.2 Calculs des indices . . . . .	17
3.3 Descriptions des indices . . . . .	17
3.3.1 Mesures sonomètres . . . . .	17
3.3.2 Mesures dosimètres . . . . .	18
3.3.3 Quelques exemples illustratifs . . . . .	19

<b>4</b>	<b>Analyse spatiale</b>	<b>23</b>
4.1	Outils utilisés . . . . .	23
4.1.1	Le S.I.G. et QGIS . . . . .	23
4.1.2	Classement sonore des infrastructures routières . . . . .	23
4.2	Méthode d'analyse . . . . .	24
4.3	Description des résultats . . . . .	25
<b>5</b>	<b>Les tests non paramétriques</b>	<b>27</b>
5.1	Test de Wilcoxon - Mann - Whitney . . . . .	28
5.1.1	Echantillon indépendant . . . . .	28
5.1.2	Echantillon apparié . . . . .	30
5.2	Test exact de Fisher . . . . .	31
5.3	Test du coefficient de corrélation de Spearman . . . . .	32
<b>6</b>	<b>Résultats des analyses statistiques</b>	<b>34</b>
6.1	Heures de pointe contre heures creuses . . . . .	34
6.2	Les éléments cartographiques . . . . .	34
6.2.1	Les voies catégorisées . . . . .	35
6.2.2	Traversées de voies, carrefours . . . . .	35
6.2.3	Parcs, zones piétonnes et voies sans issue . . . . .	36
6.2.4	Le nombre d'éléments cartographiques . . . . .	36
6.3	Axes catégorisées contre hétérogénéité . . . . .	37
6.4	Sonomètre contre dosimètre . . . . .	37
6.5	Synthèse des résultats . . . . .	37
	<b>Conclusion</b>	<b>38</b>
	<b>Bibliographie</b>	<b>39</b>
<b>A</b>	<b>Tableaux</b>	<b>40</b>
<b>B</b>	<b>Graphiques et cartes</b>	<b>46</b>

# Remerciements

Je remercie en premier lieu mon maître de stage, Mr Mauny, pour m'avoir accueilli au sein de son service, pour sa patience et son enseignement.

Je remercie également Mme Pujol pour son aide et son soutien qui ont été d'une grande utilité.

Je remercie bien évidemment l'ensemble du service CMC du CHRU de Besançon pour leur accueil très convivial.

Je remercie Capucine Py pour la relecture attentive du présent document.

Enfin, je remercie l'ensemble de mes collègues du Master de Strasbourg pour leur soutien.

# Introduction

Dans le cadre de la première année de Master Mathématiques et applications : Biostatistiques et statistiques industrielles de Strasbourg, il nous est proposé d'effectuer une immersion professionnelle au travers d'un stage. C'est dans l'optique de pouvoir élargir mes connaissances, d'appliquer mes compétences et de découvrir une certaine réalité professionnelle que j'ai choisi le CHU de Besançon comme lieu de stage.

Le centre hospitalier régional universitaire (CHRU) de Besançon se compose de deux établissements : l'hôpital Saint-Jacques situé au cœur de la ville, et l'hôpital Jean Minjot implanté au cœur du pôle santé. Le CHRU possède une capacité d'accueil de 1 335 lits et places d'hospitalisation au 14 avril 2014. Il compte en son sein près de 6 250 médecins et personnels hospitaliers accueillant plus de 570 000 patients. Le centre hospitalier régional et universitaire de Besançon assure des missions de soins, de prévention, de recherche et d'enseignement.

Le CHRU de Besançon est doté de plusieurs unités de méthodologie transversales indépendantes qui participent activement à la recherche. Au-delà de l'appui apporté à la conception, la formalisation, la conduite et la valorisation des projets de recherche, elles sont elles-mêmes porteuses de projets de recherche et à ce titre rattachés à des laboratoires de recherche universitaire.

Le centre de méthodologie clinique (CMC) fait partie de ces unités, ce service intervient dans les domaines de la recherche clinique, épidémiologique et de la recherche en soins. Les chercheurs du CMC sont associés au laboratoire de recherche Chrono-environnement. L'équipe du CMC est notamment composé de M. Frédéric Mauny, professeur des universités et responsable du service, deux ingénieurs statisticiens, un technicien - gestionnaire de données, une médecin et une ingénier d'études et de recherche clinique (Sophie Pujol). Le CMC accueille régulièrement des stagiaires, des doctorants ou des internes de santé publique.

Depuis quelques années, une partie de la recherche menée au CMC est centrée sur les nuisances sonores et l'exposition de la population au bruit. En 2006, une vaste étude acoustique est lancée dans la ville de Besançon afin de mesurer et quantifier l'impact du bruit environnemental chez les enfants de huit et neuf ans. Au cours de cette étude, il a été mis en évidence un lien entre les performances scolaires des enfants et leur exposition au bruit à l'école et au domicile [1]. Afin de compléter le tableau, les chercheurs portant ce projet s'intéressent également à l'exposition au bruit durant le trajet reliant l'école et le domicile chez les enfants bisontins.

*Comment rendre compte de la réalité acoustique d'un trajet domicile - école ?*

Il sera présenté dans un premier temps des notions d'acoustiques et l'étude globale menée depuis 2006. Sera introduit dans un second temps les outils nécessaires et les propositions mises en avant pour répondre au problème posé. Enfin, une analyse statistique adéquate portera une conclusion scientifique sur notre approche.

# Chapitre 1

## Quelques notions d'acoustique

### 1.1 Le son

Le phénomène sonore est le résultat de la perturbation produite par un objet, une source sonore, vibrant dans un milieu de propagation, dans notre cas l'air. L'onde sonore se propage de proche en proche par mise en vibration de l'air, à une vitesse approximative de  $340 \text{ m.s}^{-1}$ .

La propagation de l'onde sonore provoque une petite variation de la pression atmosphérique notée  $P_a$  autour de sa valeur moyenne ( $P_a = 1,01325 \cdot 10^5$  Pascal). Dans le but de mesurer un niveau sonore, on cherche à quantifier ces variations de pression. On définit alors une pression acoustique efficace, communément appelée pression sonore.

**Définition 1.1.** *La pression acoustique efficace (RMS)*

*On note les variations de la pression atmosphérique  $p'$ , elles sont usuellement calculées sur un intervalle de temps  $T$  dit « fast » de 125ms, ou « slow » de 1s. On note  $P$  la pression acoustique, et on la définit comme suit :*

$$P = \sqrt{\frac{1}{T} \int_0^T p'^2(t) dt} \text{ N.m}^{-2} \quad (1.1)$$

On appelle fréquence d'une onde sonore le nombre de vibrations effectuées par l'onde en une seconde, exprimées en Hertz (Hz). Il est démontré que théoriquement l'oreille humaine est capable de percevoir des fréquences comprises entre 20 Hz et 20 000 Hz. A noter que l'on considère les sons moyens (ni aigus, ni graves) autour de 1 kHz.

L'échelle des pressions sonores auxquelles est sensible l'oreille humaine est très étendue. Par convention, on utilise un son pur (composé d'une seule fréquence) de 1000 Hz pour évaluer cette échelle, qui s'étend de  $2 \cdot 10^{-5}$  Pa pour le seuil d'audibilité à 20 Pa pour le seuil de douleur.

Pour mesurer un niveau sonore, également appelé niveau de pression acoustique, on compare la pression sonore à une valeur de référence. Cette valeur de référence est fixée à  $2 \cdot 10^{-5}$  Pa.

**Définition 1.2.** *Le niveau sonore (SPL)*

*Soit  $P_{eff}$  la pression acoustique efficace mesurée,  $P_{ref}$  la pression sonore de référence égale à  $2 \cdot 10^{-5}$  Pa, et  $T$  le temps de mesure. On définit le niveau sonore  $L$  tel que :*

$$L = 10 \log_{10} \left( \frac{P_{eff}^2}{P_{ref}^2} \right) \quad (1.2)$$

*$L$  est exprimé en décibels (dB)*

## Remarques.

1. Le seuil de la douleur se situe entre 120 dB SPL et 130 dB SPL
2. Un rapport de 2 entre les deux pressions engendre une augmentation de 6 dB SPL.
3. Dans la littérature, le niveau sonore est défini comme le logarithme d'un rapport de deux puissances, cette définition est strictement équivalente.
4. Si deux objets émettent à un niveau sonore de 50 dB, mis côte à côte le niveau sonore est de 53 dB.

## 1.2 Courbe de pondération A

Il est à noter que, d'après la définition du décibel SPL, deux sons donnent la même valeur en décibels s'ils correspondent à la même pression acoustique, qu'ils soient graves ou aigus. Cependant, l'oreille humaine a une sensibilité différente suivant les fréquences d'une source sonore. Pour rendre compte de cette sensibilité il existe plusieurs normes de pondération ou correction fréquentielle, la pondération A de la norme CEI 61672-1 est exigée pour la mesure et la gestion des bruits environnementaux.

**Remarque.** En appliquant cette correction, le niveau sonore est alors exprimé en  $dB(A)$ . Ce niveau est noté  $L_A$  (en dB) ou  $L$  (en  $dB(A)$ ).

Le tableau 1.1 présente à titre d'exemple certaines sources sonores associées à un niveau sonore.

$L_A$ (dB)	Exemple de source
	Tic-tac d'une montre à plusieurs mètres
0	Seuil d'audition
10	Chambre sourde
20	Bruit de fond naturel dans le silence
30	Chambre à coucher
40	Campagne calme
60	Voix normale
70	Centre d'appel téléphonique
80	Orchestre symphonique
90	Voix criée
110	Sirène des pompiers
120	Seuil de douleur
140	Formule 1
150	Avion à réaction
170	Forte Explosion

TABLE 1.1 – Échelle des niveaux sonores

## 1.3 Appareils de mesure

Il est présenté dans cette section deux instruments mesurant la pression acoustique, et indirectement le niveau sonore. Les deux appareils permettent une pondération fréquentielle et une constante d'intégration « slow » ou « fast ».



### 1.3.1 Le sonomètre

Le sonomètre s'utilise dans les études de pollution sonore et d'acoustique environnementale pour quantifier le bruit et les nuisances sonores, principalement les bruits industriels et de transports routiers, ferroviaires et aériens. C'est donc un appareil qui mesure l'ambiance sonore.

### 1.3.2 Le dosimètre

Le dosimètre, ou exposimètre, a été initialement conçu afin de mesurer l'exposition au bruit des salariés travaillant dans une ambiance bruyante. Le boîtier de mesure se porte sur l'épaule, au plus près de l'oreille du porteur. Par extension, le dosimètre évalue une exposition au bruit chez un individu.

## 1.4 Indicateurs sonores énergétiques

### 1.4.1 Le niveau équivalent

Le niveau sonore équivalent est le niveau de pression acoustique d'un bruit stable qui donnerait la même énergie acoustique qu'un bruit à caractère fluctuant, pendant une durée  $T$  donnée. La plupart des études concernant la gêne se réfèrent au niveau de pression sonore équivalent pondéré A noté  $L_{Aeq}$ . Il est exprimé comme suit :

$$L_{Aeq} = 10 \log_{10} \left( \frac{1}{T} \int_0^T 10^{0,1L_A(t)} dt \right) \text{ dB} \quad (1.3)$$

Dans la réalité, on utilise plus facilement cette formule :

$$L_{Aeq} = 10 \log_{10} \left( \frac{1}{T} \sum_{i=0}^n T_i 10^{\frac{L_i}{10}} \right) \text{ dB} \quad (1.4)$$

où  $T$  est le temps de mesure,  $n$  le nombre de mesure,  $i$  la  $i^{\text{ème}}$  mesure,  $\forall i \in \{1, \dots, n\}$ ,  $L_i$  représente le  $i^{\text{ème}}$  niveau sonore instantané en dB(A) pendant une durée  $T_i$ .

### 1.4.2 Le niveau d'exposition

Le niveau d'exposition sonore, noté SEL, est défini comme le niveau de pression sonore pondéré A d'un bruit fictif qui, maintenu constant pendant 1 seconde, aurait la même énergie acoustique qu'un événement acoustique dont le niveau sonore est fluctuant. Il est exprimé comme suit :

$$SEL = 10 \log_{10} \left( \frac{1}{T_0} \int_0^T 10^{0,1L_A(t)} dt \right) \text{ dB} \quad (1.5)$$

où  $T$  est la durée de la mesure,  $T_0$  est la durée du bruit fictif prise égale à 1s.

## 1.5 Indicateurs sonores statistiques

Les indices statistiques, notés  $L_n$  représentent le niveau sonore pondéré A atteint ou dépassé pendant  $n\%$  du temps total d'une séquence de mesure du niveau sonore. Par exemple si on considère une constante d'intégration de 1 seconde, lors d'une mesure de 60 secondes, on obtient 60 mesures, notées  $L_{A,1sec}$ . On considère habituellement trois indices statistiques :

- le  $L_{10}$ , qui caractérise le bruit de crête. Statistiquement, il représente le neuvième décile d'une séquence de mesure.
- le  $L_{50}$ , qui caractérise le niveau de bruit moyen. Statistiquement, il représente la médiane d'une séquence de mesure.
- le  $L_{90}$ , qui caractérise le niveau du bruit de fond. Statistiquement, il représente le premier décile d'une séquence de mesure.

# Chapitre 2

## L'étude

### 2.1 Introduction

Les recherches réalisées depuis une trentaine d'années ont mis en évidence de nombreux effets sanitaires liés à l'exposition sonore de la population générale. L'appareil auditif peut évidemment être directement endommagé, cependant il existe aussi des effets extra auditifs à l'instar des troubles de l'apprentissage et des problèmes cardio-vasculaires.

Les enfants forment un groupe à risque particulier comparé aux adultes. Cela vient notamment du fait que le contrôle de l'enfant sur son environnement est beaucoup plus limité. Leurs capacités d'adaptation perceptive sont moins développées et leur perception du risque est plus basse.

Chez les enfants, les journées peuvent se décomposer en périodes de vie à la maison, à l'école, pendant les trajets et enfin lors d'autres activités (principalement de loisirs). En 2006, dans une première campagne de mesures et d'analyse, l'équipe du CMC parvient à établir une représentation cartographique du bruit environnemental donnant lieu à une estimation valide du niveau d'exposition de l'enfant à l'école et au domicile via un niveau sonore en façade [2], [3].

Le temps d'un trajet constitue une période passée dans un espace particulier (extérieur, confiné dans un transport, variant...). La variabilité du niveau sonore lors d'un trajet, ainsi que l'influence des espaces traversés sur ce niveau sont mal connues. En France, du fait de la carte scolaire, les distances entre le domicile et l'école sont généralement situées entre 500 et 1000 mètres pour les établissements du cycle élémentaire. Ainsi, la majorité des enfants se rendent à l'école à pied.

Les trajets domicile - école apparaissent, pour les enfants scolarisés, parmi les périodes d'exposition au bruit les moins bien renseignées. L'équipe de recherche composée d'unités mixtes de recherche du CNRS et de l'université de Franche-Comte s'est donc donnée pour objectif de quantifier le niveau d'exposition sonore d'enfants scolarisés à l'école primaire lors du trajet domicile - école, et d'estimer la variabilité interindividuelle de ce niveau d'exposition.

### 2.2 Population

Cette étude s'est déroulée à Besançon (120 000 habitants). La population de l'étude est constituée d'enfants scolarisés à l'école primaire dans le cycle d'approfondissements (CE2, CM1, CM2) dans l'une des 35 écoles primaires publiques de la ville. La première étude, en 2006, destinée à évaluer le niveau d'exposition sonore intra et extra-domiciliaire des 900 élèves en classe de CE2 a permis de rassembler de larges informations concernant le domicile, la famille, la situation vis-à-vis des

nuisances sonores intra-domiciliaires [3]. Une campagne de mesures au domicile avait été conduite auprès d'un échantillon de 50 logements. L'échantillon a été constitué à partir d'une liste d'enfants tirés au sort, stratifiés par école (30 écoles finalement impliquées). La figure B.1 page 46 représente la distribution des 900 élèves, des 50 familles contactées et les écoles de l'étude.

Pour les besoins de la présente étude, les 50 familles de la première étude ont été recontactées. Un rendez-vous téléphonique leur a été proposé afin de compléter un questionnaire standardisé permettant de recueillir les caractéristiques des trajets domicile - école. La collecte d'informations a eu lieu en janvier 2008. Deux familles n'ont pas pu être contactées pour cause de déménagement.

Deux campagnes de mesures distinctes ont été réalisées.

## 2.3 Les campagnes de mesures

### 2.3.1 Au sonomètre

Il a été confié à trois opérateurs adultes d'effectuer deux enregistrements sonores pour chaque trajet parmi les 50 enfants. Une mesure était effectuée en période dite « heure creuse » (11h ou 13h), une deuxième en période dite « heure de pointe » (8h ou 16h). Les sonomètres ont été calibrés quotidiennement et la durée d'intégration a été fixée à une seconde.

Finalement, 92 trajets exploitables ont été retenus pour 47 enfants. Pour deux enfants, il n'existe qu'une unique mesure de trajet, pour l'un durant la première période, pour l'autre durant la seconde période. Par conséquent, pour 45 enfants, il existe deux mesures de trajet pour chaque période précédemment décrite.

### 2.3.2 Au dosimètre

En 2009, une deuxième campagne de mesures a été lancée consistant à faire porter un dosimètre aux enfants pendant leurs différents trajets. Les dosimètres, calibrés quotidiennement, étaient programmés sur quatre plages d'enregistrement : 7h45-9h00, 11h00-12h30, 12h45-14h00, 15h45-18h30. Un questionnaire était également rempli afin de décrire les trajets effectués (horaires précis de départs et d'arrivées, mode de déplacement, conditions météorologiques, ambiance sonore).

Au final, 93 trajets exploitables ont été retenus pour 32 enfants dont l'âge moyen est de 10,9 ans (écart type 0,44). Parmi ces 93 trajets, 83 sont piétons, neuf effectués en voiture et un en vélo.

Il est à noter que seulement 63 trajets possèdent leur « homologue » parmi les mesures effectuées avec le sonomètre. Cela signifie que le sonomètre a enregistré l'ambiance sonore sur le même tracé géographique que lors des mesures avec dosimètre.

## 2.4 Premiers résultats

Une première analyse des mesures a donné lieu à des premières conclusions concernant le niveau sonore  $L_{Aeq}$  [4]. Le but de ces analyses statistiques étant de tester l'impact sur le  $L_{Aeq}$  de différentes variables temporelles, météorologiques, typologiques et inhérentes au contexte de mesure telles que le type de sonomètre, le porteur du sonomètre ou la date de la mesure.

Pour les données mesurées au sonomètre, il a été conclu que les facteurs liés aux conditions de mesures n'avaient aucun impact sur le niveau sonore équivalent. Cependant il a été mis en évidence que le niveau d'exposition au bruit varie selon la zone d'habitation : les personnes les plus exposées étant celles qui vivent au centre ville ou dans la première couronne d'habitat dense de Besançon.

Pour les mesures au dosimètre, une analyse équivalente a été menée. Cette dernière a mis en évidence trois facteurs impactant le niveau sonore : l'âge de l'enfant, le sens du trajet et le fait qu'un enfant soit accompagné par d'autres camarades. Il est donc constaté que les enfants de plus de 11 ans, accompagnés par d'autres enfants et rentrant à leur domicile sont soumis à un niveau sonore plus élevé que les autres. Cependant ces facteurs n'expliquent que 20% des variations du modèle.

Comme dit précédemment un trajet à pied est sujet à une multitude d'ambiances sonores très différentes les unes des autres. Il est clair qu'un niveau sonore équivalent ne peut rendre compte du caractère fluctuant de l'ambiance sonore. C'est pourquoi nous voulons étudier ces trajets sous un angle moins statique avec l'utilisation d'indicateurs sonores dynamiques.

## Chapitre 3

# Les indicateurs acoustiques dynamiques

Un enfant, au cours de son trajet est confronté à des niveaux acoustiques qui varient selon des événements sonores auxquels il peut être sensible. Ces variations produisent une gêne ou un confort acoustique difficilement retranscrit par des indicateurs classiques présentés au préalable. En effet, ces derniers permettent de renseigner sur le niveau global perçu par l'enfant. Les indicateurs présentés par la suite ont pour objectif de rendre compte de l'évolution dynamique du bruit [5], [6].

### 3.1 Définitions

#### 3.1.1 Indicateur d'étalement sonore

L'étalement du bruit est particulièrement lié à la forme de la distribution statistique des  $L_{A,1sec}$ . On distingue deux indicateurs permettant de mesurer cette distribution.

##### **ILE**

La largeur de l'étalement, notée  $\Delta L$ , est calculée à partir de la distribution statistique, dB par dB, des  $L_{A,1sec}$  arrondis au dBA (notés pour chaque classe  $i$  :  $L_{A,i}$ ) mais en ne conservant dans la distribution que les classes  $i$  dont le nombre correspondant de  $L_{A,i}$  est supérieur ou égal à 1% du nombre total d'échantillons de  $L_{A,1sec}$  sur l'ensemble de la distribution. La valeur de l'étalement est alors égale à la « distance » entre les deux classes  $i$  extrêmes ainsi déterminées, soit en dBA :

$$\Delta L = \max(i) - \min(i) \quad (3.1)$$

L'indice de largeur d'étalement,  $ILE$ , est alors calculé en référant la valeur de  $\Delta L$ , à une valeur forfaitaire de 30 dBA (considérée comme une valeur élevée de l'étalement sonore), soit :

$$ILE = \frac{\Delta L}{30} \quad (3.2)$$

Quand  $ILE$  tend vers 0, l'étalement devient faible traduisant un niveau de bruit très constant. Autour de la valeur 1, l' $ILE$  est élevé ce que traduit une large plage de niveaux sonores sur la période étudiée. Au-delà de 1, le bruit devient très fluctuant, c'est à dire que l'amplitude de la plage des niveaux sonores est large.

##### **IHE**

L'homogénéité de l'étalement traduit le caractère plus ou moins constant de la distribution. Elle est estimée au travers de l'indice d'homogénéité de l'étalement,  $IHE$ , calculé à partir de l'émergence du pic principal de la distribution, correspondant à la classe  $i^*$  et dont le nombre de réalisations

associées est noté  $N_{i^*}$  par rapport à la moyenne des  $N_i$  effectuée sur les classes  $i$  dont le nombre correspondant de  $L_{A,1sec}$  est supérieur ou égale à 1% du nombre total d'échantillons de  $L_{A,1sec}$  mais en excluant la classe  $i^*$ , cette moyenne étant notée  $moy(N_{i,\geq 1\%, -i^*})$ . L'*IHE* est alors défini par :

$$IHE = \left( 1 - \frac{N_{i^*} - moy(N_{i,\geq 1\%, -i^*})}{N_{i^*}} \right) \quad (3.3)$$

Plus la valeur de *IHE* est proche de 1, plus l'étalement est homogène. Une valeur proche de 0 traduit une forte inhomogénéité de l'étalement.

### 3.1.2 Indicateur de rugosité sonore

La rugosité du bruit traduit le caractère plus ou moins continu de l'évolution des  $L_{A,1sec}$  sur une période considérée. Elle permet de mettre en évidence et de quantifier les sauts brusques du niveau sonore. Elle est jugée dans la dimension temporelle à partir de la distribution statistique de la valeur absolue des écarts  $\delta L_A$ , arrondis au dB près, entre deux valeurs successives du  $L_{A,1sec}$  définis par :

$$\delta L_A = L_{A,1sec}(k) - L_{A,1sec}(k-1) \quad (3.4)$$

$k$  étant l'indice du  $k$ -ième échantillon du  $L_{A,1sec}$ .

On dira qu'un bruit est fortement rugueux si la distribution statistique des  $\delta L_A$ , arrondis au dB près, est assez étalée. Si, au contraire, la distribution est très étroite et centrée autour de 0 dBA, on dira du bruit qu'il est faiblement rugueux, c'est-à-dire sans variation brusque de son niveau au cours de la période d'évaluation. On décide de mesurer cette rugosité via la moyenne de la distribution des  $|\delta L_A|$ .

### 3.1.3 Indicateur du degré de l'augmentation sonore

Il s'agit ici de caractériser le pourcentage du temps où le niveau sonore augmente rapidement, par rapport aux moments où il diminue rapidement. L'augmentation rapide du niveau sonore est une cause de la gêne [Rosinger *et al.*, 1970]. On considère qu'une augmentation ou diminution rapide du bruit est caractérisée par une valeur absolue de  $\delta L_A$  supérieure à 3 dB. L'indice de degré d'augmentation sonore, *IDAS*, s'écrit :

$$IDAS = 10 \log_{10} \left( \frac{N_{\delta L_A < -3}}{N_{\delta L_A > 3}} \right) \quad (3.5)$$

Avec  $N_{\delta L_A > -3}$  le nombre d'échantillons de  $\delta L_A$  supérieur à 3dB. Avec  $N_{\delta L_A < -3}$  le nombre d'échantillons de  $\delta L_A$  inférieur à 3dB.

La formule indique que si *IDAS* est proche de 0, on considère qu'il y a un équilibre entre pourcentage du temps où les niveaux sonores augmentent et diminuent rapidement, lors de la période étudiée. Dans le cas où *IDAS* est négatif, les moments d'augmentation sont plus nombreux. Inversement, si *IDAS* est positif, le nombre de diminution sonore est d'avantage présent.

### 3.1.4 Indicateur de bruyance

Il s'agit de quantifier le nombre et la taille des « séquences de bruit », c'est-à-dire des épisodes où  $L_{A,1sec}$  dépassera une valeur seuil. Pour cela on propose deux seuils :

- Le  $L_{10}$ , le niveau de crête.

- Une valeur forfaitaire de 70 dB, indépendante de la position du récepteur. Ce niveau est considéré comme le début d'une gêne importante selon l'OMS [7].

Afin de bien distinguer la notion de « séquence » et celle du « pic de bruit » (caractère impulsif), nous considérons uniquement les séquences de bruit d'une durée totale d'au moins 4 secondes.

On obtient alors quatre indicateurs. On note  $X$  la durée de la période sur laquelle ces indicateurs ont été calculés.

$N_{>L_{10},X}$

Nombre des séquences d'au moins 4 secondes dont le niveau sonore dépasse le  $L_{10}$ , évalué sur une période de  $X$  minutes de  $L_{A,1sec}$ .

$T_{>L_{10},X}$

La durée moyenne en secondes des séquences calculées dans l'indice précédent.

$N_{>70,X}$

Nombre des séquences d'au moins 4 secondes dont le niveau sonore dépasse 70 dB, évalué sur une période de  $X$  minutes de  $L_{A,1sec}$ .

$T_{>70,X}$

La durée moyenne en secondes des séquences calculées dans l'indice précédent

**Remarque.** Afin de pouvoir comparer à l'avenir les nombres de séquences on introduit le pourcentage du nombre de séquence de quatre secondes défini ainsi :

$$PN_{>L} = \frac{N_{>L,X}}{X} \times 100$$

avec  $L$  un niveau sonore. Cela correspond à rapporter le nombre de séquences sur 100 secondes.

### 3.1.5 Indicateur de tranquillité

Il s'agit de quantifier le nombre et la taille des « séquences de calme », c'est-à-dire d'épisodes où  $L_{A,1sec}$  restera en dessous d'une valeur seuil. Pour cela on propose deux seuils :

- Le  $L_{90}$ , le niveau sonore caractérisant le bruit de fond.
- Une valeur forfaitaire de 60 dB, indépendante de la position du récepteur.

Comme précédemment, nous considérons uniquement les séquences de bruit d'une durée totale d'au moins 4 secondes.

$N_{<L_{90},X}$

Nombre de séquences d'au moins 4 secondes dont le niveau sonore est inférieur à  $L_{90}$ , évalué sur une période de  $X$  minutes de  $L_{A,1sec}$ .

$T_{<L_{90},X}$

La durée moyenne en secondes des séquences calculées dans l'indice précédent.



$N_{<60,X}$

Nombre de séquences d'au moins 4 secondes dont le niveau sonore est inférieur à 70, évalué sur une période de  $X$  minutes de  $L_{A,1sec}$ .

$T_{<60,X}$

La durée moyenne en secondes des séquences calculées dans l'indice précédant.

**Remarque.** Afin de pouvoir comparer à l'avenir les nombres de séquences on introduit le pourcentage du nombre de séquence de quatre secondes défini ainsi :

$$PN_{<L} = \frac{N_{<L,X}}{X} \times 100$$

avec  $L$  un niveau sonore. Cela correspond à rapporter le nombre de séquences sur 100 secondes.

### 3.1.6 Indicateur Harmonica

En 2011, le projet HARMONICA est retenu par la commission européenne dans le cadre du programme européen Life+ environnement [8]. Le projet part du constat selon lequel la pollution sonore et les moyens pour y remédier sont insuffisamment connus du grand public et des autorités. L'objectif des acteurs du projet HARMONICA est la construction et la validation d'un indice de bruit grand public.

Cet indice est innovant car il permet d'informer sur les deux composantes majeures qui impactent l'environnement sonore, à savoir le bruit de fond ambiant et les événements sonores qui émergent de ce bruit de fond. On le définit ainsi :

$$\text{Harmonica} = BGN + EVT \tag{3.6}$$

$BGN$

Une composante associée au bruit de fond : sous-indice  $BGN$

$$BGN = 0,2 \times (L_{Aeq,95} - 30)$$

où  $L_{Aeq,95}$  est le niveau équivalent du bruit de fond au cours de l'heure considérée, le bruit de fond étant évalué chaque seconde à partir du niveau dépassé pendant 95% du temps au cours des 10 minutes précédentes.

$EVT$

Une composante événementielle qui tient compte de l'énergie acoustique des événements qui émergent du bruit de fond : sous-indice  $EVT$

$$EVT = 0,25 \times (L_{Aeq} - L_{Aeq,95})$$

où  $L_{Aeq}$  est le niveau équivalent du bruit au cours de l'heure considérée.

L'indice Harmonica délivre directement une note comprise entre 0 et 10, qui est donnée avec une décimale. Plus la note est élevée et plus l'environnement sonore est dégradé. Cette échelle permet de situer l'environnement sonore par rapport aux objectifs de qualité de l'OMS et aux valeurs reconnues comme critiques pour le bruit [7].

Couleur	Sensation	periode diurne (6h-22h)	periode nocturne (22h-6h)
Vert	Calme	de 0 à 4	de 0 à 3
Orange	Bruyant	de 4 à 8	de 3 à 7
Rouge	Très Bruyant	8 et +	7 et +

TABLE 3.1 – Échelle Harmonica

## 3.2 Calculs des indices

Afin de calculer l'ensemble des indices précédemment introduit sur les mesures dosimètres et sonomètres, les chargés de l'étude et moi-même avons procédé à l'importation des données sous le logiciel *R*. A l'aide de la table des mesures seconde par seconde (noté  $L_{A,1sec}$ ) du niveau sonore de chaque trajet, j'ai procédé à plusieurs vérifications qui m'ont permis de mettre en évidence quelques irrégularités.

- Certain  $L_{Aeq}$  avaient été calculés à partir du mauvais fichier de mesure seconde par seconde. Les corrections ont été apportées en recalculant les bons niveaux sonores
- Pour certaines mesures, la durée de l'enregistrement inscrite dans la base était erronée. Pour palier à cela, la nouvelle durée a été calculée en comptant le nombre d'entrée de la table seconde par seconde concernant le trajet erroné.
- Il est apparu à quelques reprises que certaines mesures isolées de  $L_{A,1sec}$  étaient manquantes ou fausses (saturation du sonomètre). La correction à ce problème fut apportée en lissant les mesures autour de la valeur manquante.
- Quelques indicateurs dynamiques avaient été préalablement calculés à l'aide de Matlab, cependant les formules utilisées n'étaient pas les bonnes.

Après s'être donc assuré d'avoir des jeux de données correctes, j'ai entrepris de calculer l'ensemble des indices en développant quelques fonctions sous *R*. Enfin, j'ai pu dresser deux tables prêtes à l'étude :

- La table *sonometre* comportant 92 observations.
- La table *dosimetre* comportant 93 observations. Huit d'entre elles ont été réalisées en voiture, une mesure a été réalisée en vélo et enfin 83 mesures ont été enregistrées à pied.

## 3.3 Descriptions des indices

### 3.3.1 Mesures sonomètres

Une description détaillée est fournie en annexe, A.1 page 40. On constate que le niveau sonore s'étend de 53,4 dB(A) à 72,6 dB(A), avec une moyenne à 63,3 dB(A). En moyenne le bruit de crête s'élève à 66 dB(A), alors que le bruit de fond a pour valeur 50,4 dB(A). Selon l'OMS, le niveau du bruit de fond provoque une gêne sonore modérée.

Concernant les indicateurs dynamiques nous pouvons noter que l'IDAS possède des valeurs assez proches de sa moyenne située très proche de zéro : l'augmentation sonore semble autant présente

que la diminution sonore.

Avec un maximum à 0,69 et une médiane à 0,41, l'IHE laisse penser que le niveau sonore est inhomogène, c'est à dire qu'il existe des niveaux sonores qui sont beaucoup plus présents que d'autres.

L'ILE ne dépasse la valeur 1 que de très peu, de plus sa moyenne est égale à 0,76 (très proche de sa médiane). Ce résultat suggère qu'en moyenne un enregistrement possède un étalement de 22,8 dB(A).

En moyenne, les trajets semblent faiblement rugueux, car l'indice de rugosité est proche de 2 ce qui ne constitue pas une variation brusque.

Les indices de bruyance sont plus difficilement interprétables dans ce contexte, cependant en moyenne sur deux minutes de trajet il y a une séquence bruyante durant près de quatre secondes, le minimum possible pour ces séquences. L'indice de tranquillité  $PN_{<60,X}$  nous permet de constater qu'il y a 5 séquences de 23 secondes toutes les 3 minutes. Ce résultat paraît aller dans le sens de la moyenne du niveau sonore (63,25dB(A)), et du niveau de bruit de fond qui est à 10 dB en dessous de 60 dB.

Cinquante-neuf pour cent des trajets possèdent des séquences à la fois au delà de 70 dB(A) et inférieures à 60 dB(A). Aucun trajet ne possède aucune séquence supérieure à  $L_{10}$  et inférieure à  $L_{90}$ . Seulement deux mesures (2.2%) ne possèdent pas de séquence inférieure à 60 dB(A).

L'indice Harmonica est légèrement élevé puisqu'il est égal en moyenne à 7,3. Vingt-huit mesures possèdent un indice supérieur à 8, qui est considéré comme très bruyant. Et aucune mesure ne présente des valeurs inférieures à 4.

### 3.3.2 Mesures dosimètres

Les mesures du dosimètres ont été séparées en trois tables différentes suite aux différents moyens de transport utilisés le jour des enregistrements. En annexe est fournie une table pour les 83 mesures à pied (annexe A.2, p. 41), une table pour les neuf observations en voiture (annexe A.3, p. 42) et une dernière pour le seul trajet en vélo (annexe A.4, p. 42).

Le niveau sonore moyen des trajets réalisés à pied a pour valeur  $L_{Aeq} = 74$  dB(A), celui des trajets réalisés en voiture est de  $L_{Aeq} = 70,8$  dB(A).

Le niveau moyen du bruit de crête ( $L_{10}$ ) est égal à 77 dB(A) pour les mesures réalisées à pied, et égal à 73,2 dB(A) pour les mesures réalisées en voiture.

Le niveau moyen du bruit de fond ( $L_{90}$ ) a pour valeur 54,8 dB(A) dans le cas des mesures à pied et 56,2 dB(A) pour celles réalisées en voiture.

L'indice de rugosité dans le cas des mesures réalisées à pied et en voiture est élevé, respectivement 4,5 et 4,1 dB(A), avec des maximums proche de 7 dB(A) : cela signifie qu'il existe des trajet avec des variations sonores importantes.

Néanmoins, l'IDAS semble laisser penser qu'il existe autant d'augmentation que de diminution sonore. Les moyennes sont proches de 0.

Dans le cas des trajets à pied et en voiture, l'ILE est élevé, respectivement 1,07 et 0,96 ce qui signifie que l'étalement est proche de 30 dB(A). Ce résultat est à mettre en perspective avec la durée

moyenne du trajet qui est de 9min 30s pour les enfants à pied et de 5min pour les trajets en voiture.

Concernant les trajets à pied, 86% des trajets possèdent des séquences à la fois supérieures à 70 dB(A) et inférieures à 60 dB(A). 6% des mesures ne possède aucune séquence supérieures à  $L_{10}$  et inférieures à  $L_{90}$ .

L'indice Harmonica semble déraisonnablement très élevé puisque au-delà de la valeur seuil prévue. En moyenne l'indice Harmonica est de valeur 10,1 pour les mesures réalisées à pied, 9 pour les mesures réalisées en voiture, et de valeur 9,6 pour la mesure réalisée en vélo. Cependant, après avoir pris contact avec un responsable du projet, les valeurs trouvées sont valides. L'échelle n'est pas obligatoirement borné par 10 et peut dans certains cas bruyants être dépassée.

### 3.3.3 Quelques exemples illustratifs

#### Remarques.

- Les périodes 1 et 4 correspondent aux « heures de pointe ».
  - la période 1 est aux alentours de 8h30.
  - la période 4 est aux alentours de 16h30.
- Les périodes 2 et 3 correspondent aux « heures creuses ».
  - la période 2 est aux alentours de 11h30.
  - la période 3 est aux alentours de 13h30.

#### IDAS

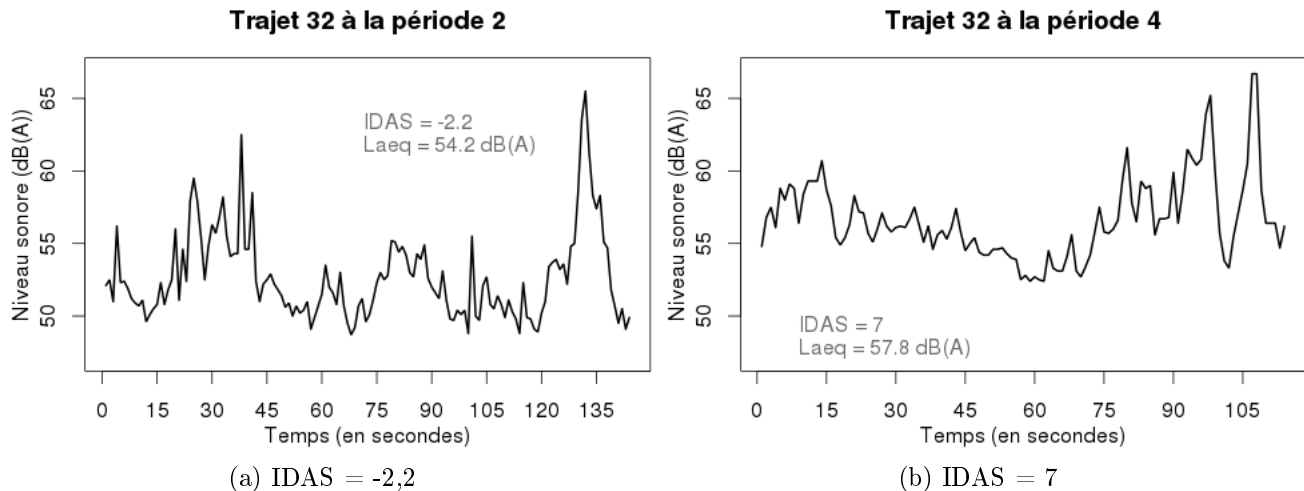


FIGURE 3.1 – Exemples pour l'IDAS

## IHE

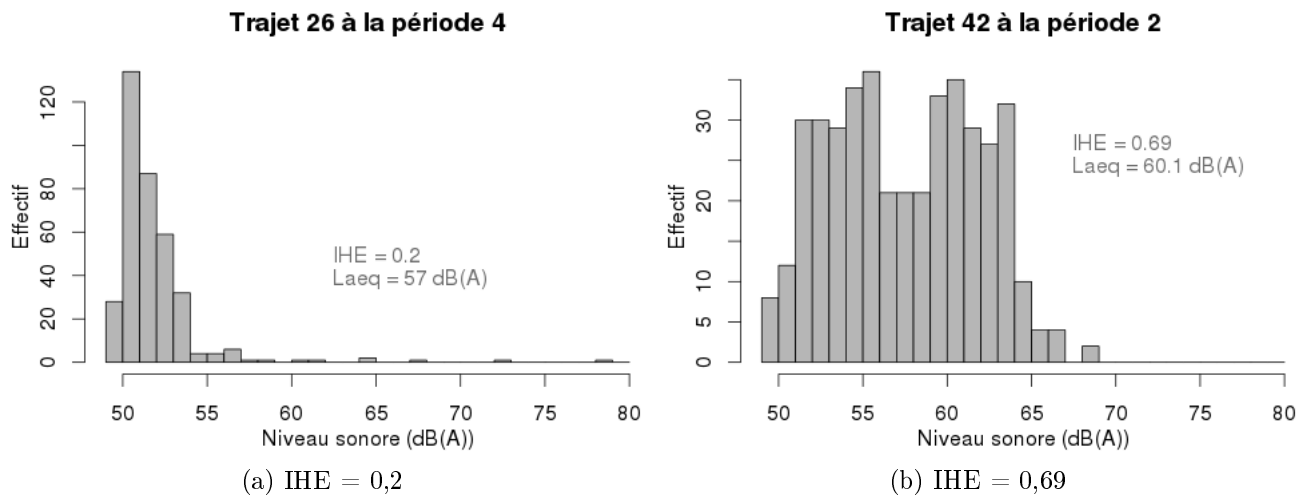


FIGURE 3.2 – Exemples pour l’IHE

## ILE

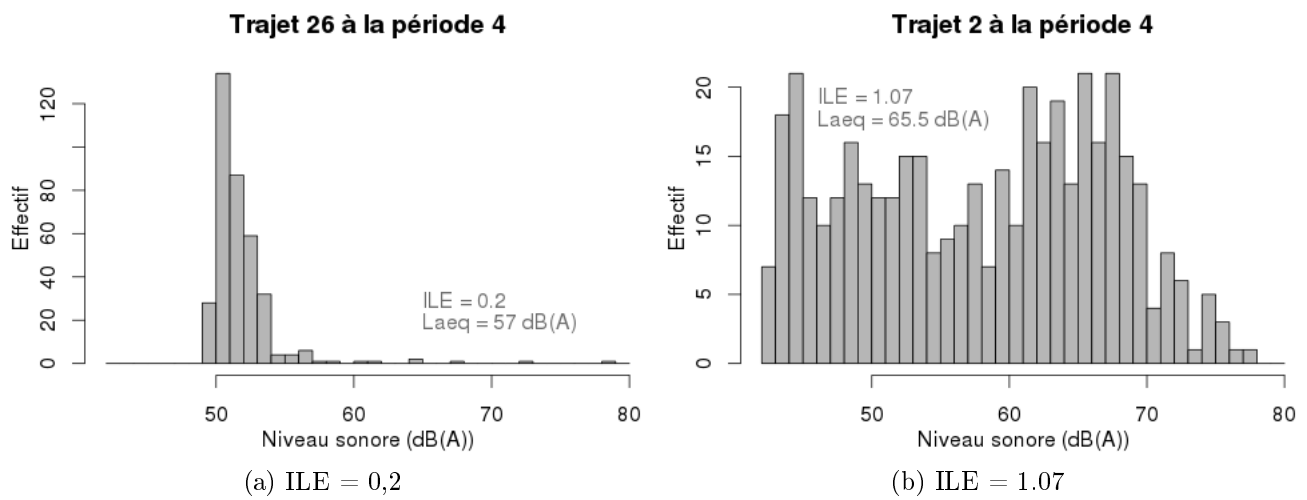


FIGURE 3.3 – Exemples pour l’ILE

## Indice de rugosité

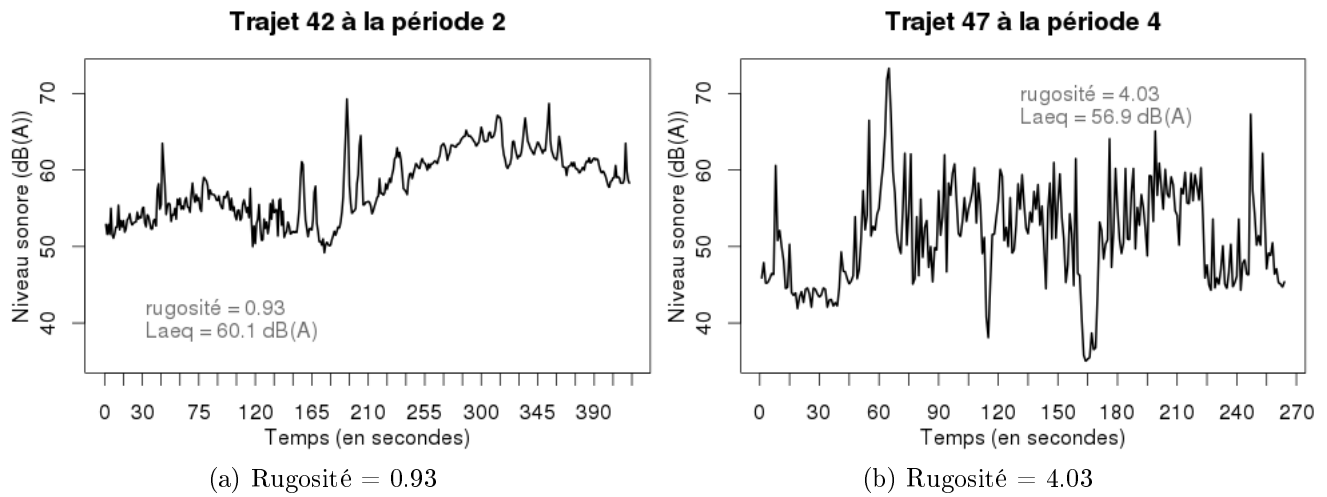


FIGURE 3.4 – Exemples pour l’ILE

$PN_{>70}$

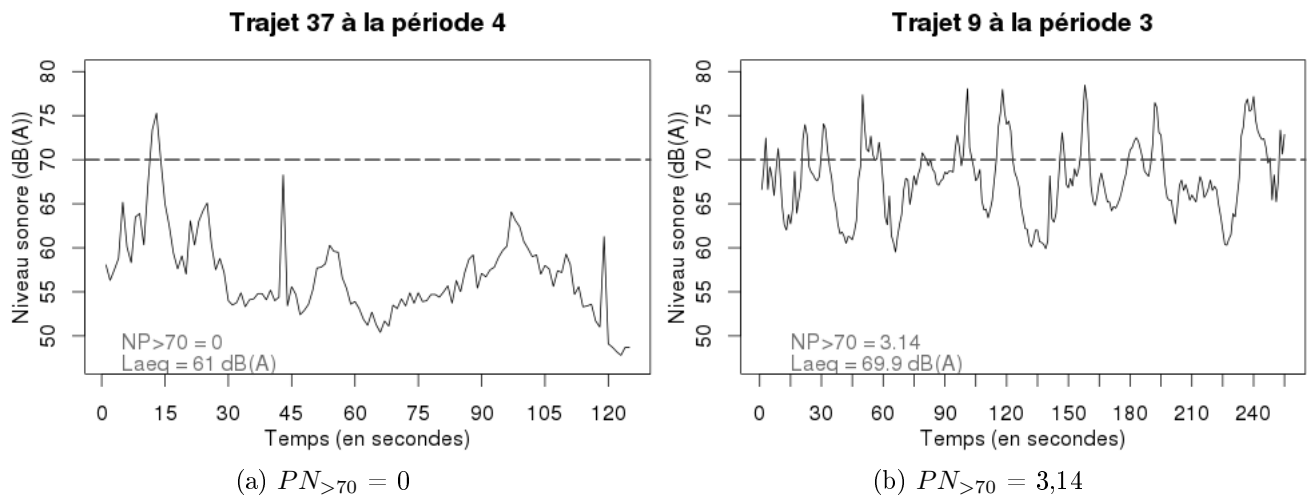


FIGURE 3.5 – Exemples pour la buyance

$PN_{<60}$

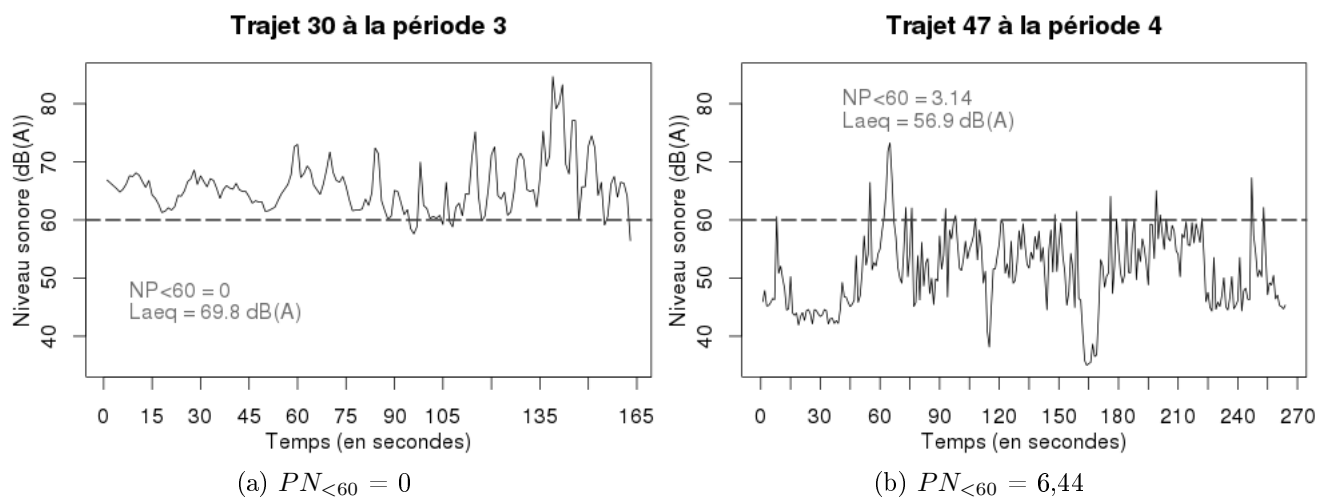


FIGURE 3.6 – Exemples pour la buyance

### Harmonica

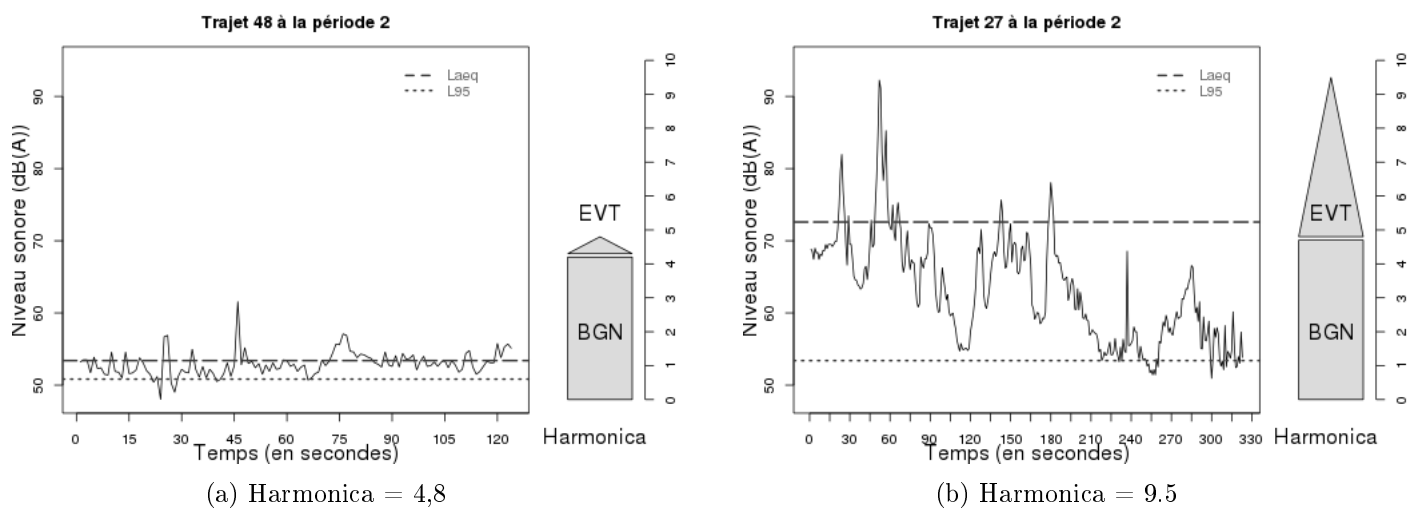


FIGURE 3.7 – Exemples pour l'indice Harmonica

A l'exception de l'indice Harmonica, les indicateurs sonores précédemment introduits et définis permettent donc de décrire et quantifier un caractère précis d'un enregistrement sonore. Ce caractère peut être l'étalement sonore, la rugosité sonore, la bruyance, la tranquillité ou encore l'augmentation sonore. En réalité ces caractères décrivent et traduisent un ou des événements sonores de la vie urbaine lors des mesures. Il serait donc enrichissant de mettre en évidence et d'analyser les différents types d'événements sonores propres à un trajet domicile - école.

# Chapitre 4

## Analyse spatiale

L'analyse des événements sonores lors d'un trajet est fortement liée à la configuration du milieu traversé [6]. Il est clair que les événements sonores seront de natures différentes dans un parc plutôt que près d'un grand axe routier, ou encore dans une zone piétonne.

Une étude des aménagements cyclables en milieu urbain propose dans son approche de découper les trajets en tronçons ayant la même configuration acoustique et le même type d'aménagement urbain [9]. Ce découpage leur permet de calculer des indices sonores sur chaque tronçon. Par la suite il leur est donc possible de mettre en évidence l'influence d'un type de tronçon sur tel ou tel indicateur.

Dans notre présente étude, la découpe n'est plus envisageable, c'est pourquoi nous présentons par la suite une approche exploratrice basée sur une analyse spatiale des trajets. Cette analyse s'est appuyée sur deux outils majeurs.

### 4.1 Outils utilisés

#### 4.1.1 Le S.I.G. et QGIS

Dans son acception courante, un système d'information géographique (S.I.G.) est un outil informatique permettant d'organiser et présenter des données alphanumériques spatialement localisées, dont l'objectif final est habituellement de produire des cartes. Il existe également un concept plus général qui englobe l'ensemble constitué par les logiciels, mais aussi les données, le matériel et les savoir-faire liés à l'utilisation de ces derniers.

Le principe de base d'un S.I.G. est d'utiliser des objets (points, lignes, polygones) repérés géographiquement dans un système de coordonnées (référentiel) et de leurs associer des données dites « attributaires » soit pour décrire plus précisément ces objets géographiques, soit pour localiser des informations : nom d'une route, type d'un bâtiment localisé par son adresse, nombre d'habitants d'un immeuble localisé par ses coordonnées Lambert, etc ...

Pour l'analyse des trajets, nous utiliserons le logiciel QGIS 1.8.0 Lisboa, il est libre et multi plate-forme.

#### 4.1.2 Classement sonore des infrastructures routières

La loi numéro 92-1444 du 31 décembre 1992, dite loi « bruit » oblige les préfets à effectuer un classement acoustique des voies du département comportant un trafic supérieur à 5000 véhicules par jour. Ce classement présente cinq catégories de voies en fonction du niveau de bruit émis mesuré en niveaux  $L_{Aeq,6h-22h}$ .



La détermination de la catégorie sonore est réalisée compte tenu du niveau de bruit calculé selon une méthode réglementaire ou mesuré selon les normes en vigueur. La vitesse et le nombre de véhicules circulant sur les infrastructures concernés sont pris en compte dans le calcul de la catégorie. Le tableau 4.1 présente les niveaux sonores de chaque catégorie.

Catégorie de classement de l'infrastructure	Niveau sonore de référence $L_{Aeq,6h-22h}$ en dB(A)	Largeur maximale des secteurs affectés par le bruit de part et d'autre de l'infrastructure en mètre
1	$L > 81$	300
2	$76 < L < 81$	250
3	$70 < L < 76$	100
4	$65 < L < 70$	30
5	$60 < L < 65$	10

TABLE 4.1 – Classement sonore des infrastructures routières

## 4.2 Méthode d'analyse

La majorité des enfants présents dans l'étude allant à l'école à pied, les analyses ont été restreintes aux trajets dont les mesures ont été réalisées à pied.

Afin d'effectuer une analyse spatiale des trajets, nous avons donc eu recours à différentes cartes géoréférencées (voiries de Besançon, passage piétons, espace verts, classement sonore des infrastructures routières, ...). Ces données ont été exploitées à l'aide du logiciel *QGIS*.

L'objectif de l'analyse a été d'identifier les éléments cartographiques présents au court d'un trajet susceptibles d'impacter l'ambiance sonore. On distingue les éléments cartographiques qui ont été énumérés et ceux pour lesquels il a été mesuré une distance traversée.

En accord avec les chargés de l'étude et parmi la diversité des trajets, il a été retenu de prendre en compte les éléments cartographiques suivants :

### Les éléments cartographiques énumérés

- Les passages piétons traversés lors du trajet. Les passages piétons avec des voies sans issues, des chemins ou des passages privés n'ont pas été comptés.
- Les carrefours ou rond-points. Ils sont comptés lorsqu'un trajet emprunte consécutivement au moins deux traversées de voies. Un trajet a été compté différemment à cause d'une configuration particulière.
- Les parkings traversés. Cet élément prend essentiellement en compte les parkings présents à la sortie d'un immeuble.
- Les ponts traversés.

### les éléments cartographiques mesurés

- Les parcs au sens large (champs, zones vertes, ...).
- Les zones piétonnes. Essentiellement les rues et chemins piétons.

- Les voies sans issues. Principalement des voies menant à un quartier résidentiel.
- Les cours intérieures. Situées en majorité au centre ville.
- Les voies répertoriées au classement sonore des infrastructures routières. La distance sous *QGIS* pour effectuer les calculs d'intersection est de 15 mètres de part et d'autre du centre de la voie. Nous voulions mesurer la distance parcourue sur une voie catégorisée lorsque l'enfant marche le long de cette voie, ou à proximité immédiate. La distance de 15 mètres assure de prendre en compte uniquement les trajets respectant notre volonté.

**Remarque.** *Au cours de l'énumération de ces différents éléments, il est clair qu'il est apparu certains litiges dus à des configurations uniques ou très particulières, cependant des choix ont été faits. Bien que discutables, ces choix reposent sur une connaissance du terrain et de l'étude.*

Par la suite, nous nous sommes intéressés au nombre d'éléments cartographiques présents pour chaque trajet. On définit ce nombre d'éléments comme la somme des effectifs pour chaque élément énuméré et le nombre d'éléments mesurés.

**Exemple.** *Un trajet passant par quatre passages piétons, traversant un parc et deux parkings compte sept éléments cartographiques.*

On définit alors un trajet homogène comme étant un trajet possédant un ou aucun élément cartographiques.

La figure B.2 p. 47 représente l'interface *QGIS*, on y voit entre autre le classement sonore des voies terrestres de Besançon. La figure B.3 p. 47 présente une portion d'un trajet étudié sous *QGIS*.

### 4.3 Description des résultats

Dorénavant, nous allons nous concentrer sur les résultats des mesures réalisées au sonomètre. Pour rappel, les mesures de l'ambiance sonore correspondent à 92 trajets pour 47 enfants. Nous avons donc 47 trajets uniques à décrire. Une description détaillée est fournie en annexe A.5 et A.6 page 43.

Il est notable de voir que huit trajets ne franchissent aucune voie routière, et environ 36% n'en franchissent qu'une. Un quart des trajets traversent un carrefour ou un rond point important.

Seuls neuf trajets sur 47 traversent un parc, un champs ou une zone verte sur une longueur moyenne de 143 mètres. Il est intéressant de noter qu'il existe un individu qui traverse un parc pendant plus de 50% de son trajet.

Il y a seulement un trajet qui emprunte une cours intérieure. Huit trajets empruntent une zone piétonne. Sur ces huit trajets, la distance parcourue en zone piétonne représente en moyenne 28% de la longueur totale.

Seuls 25% des trajets possèdent une partie de leur itinéraire sur une voie de catégorie trois, tandis que 47% des trajets évoluent sur une voie de catégorie quatre, en moyenne sur 44% de leur distance. Il est surtout intéressant de noter que seul 40% des trajets ne rencontrent aucune voie ayant été catégorisée. Enfin, il n'existe aucun trajet concerné par la traversée d'une voie de catégorie une, deux ou cinq, ce qui laisse penser que la distribution des écoles évite à ces écoliers de traverser

de grands axes routiers pour s'y rendre.

Le nombre d'éléments cartographiques (figure 4.1) s'étend de un à neuf. La médiane est à 3 et le troisième quartile est égal à cinq, cela montre que la plupart des trajets ne sont pas sujets à une multitude d'éléments différents. Seulement sept trajets ne comptent qu'un seul éléments cartographiques, cela représente près de 15% des trajets.

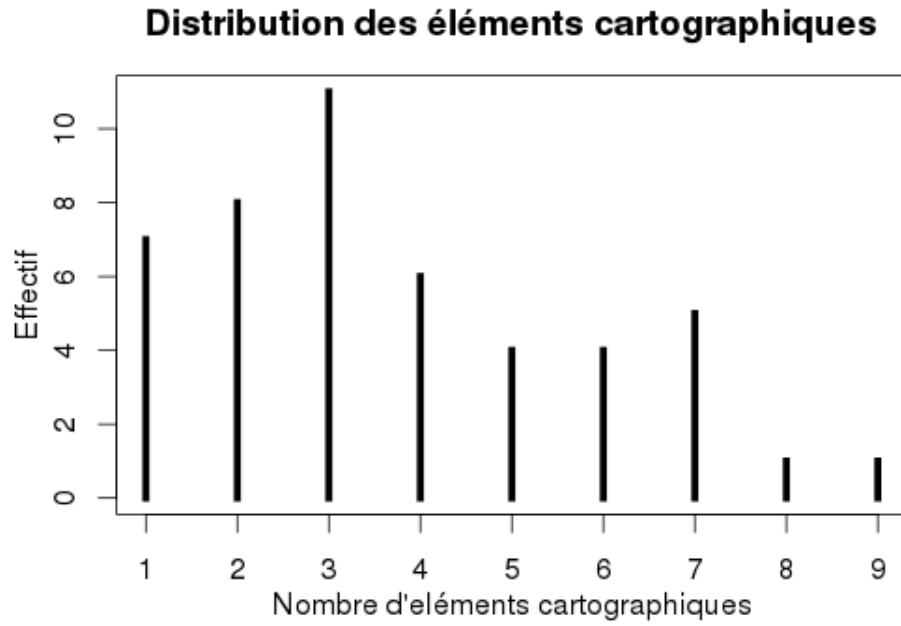


FIGURE 4.1 – Distribution du nombre d'éléments cartographiques pour les 47 trajets réalisés au sonomètre

## Chapitre 5

# Les tests non paramétriques

Certains tests statistiques ne sont valables que sous certaines hypothèses concernant la distribution de la ou les variable(s) observée(s) dans un échantillon d'une population. Par exemple, le test de Student impose une distribution normale.

On parle de tests non paramétriques lorsqu'on ne fait aucune hypothèse sur la distribution des variables observées. Ils sont également appelés *distribution free tests* (tests libres de distribution). Les tests non paramétriques permettent de tester des hypothèses assez générales telles une égalité de lois ou l'indépendance entre deux variables qualitatives.

Les tests paramétriques sont-ils meilleurs que leurs homologues non paramétriques ? Cette question, bien trop souvent posée, ne possède aucune réponse bien tranchée. L'utilisation de tel ou tel test est généralement imposée par le contexte des mesures et de l'analyse.

Cependant, les tests libres de distribution ont tout de même quelques avantages notables. Comme leur nom l'indique, il n'est pas nécessaire de faire des hypothèses sur la forme des distributions, il n'est pas nécessaire non plus d'estimer les paramètres associés. Le champ d'application des techniques est par conséquent plus large.

Lorsque les conditions d'application ne sont pas respectées, les tests paramétriques restent tout de même très robustes pour peu que les effectifs soient assez élevés.

Dans le contexte où les conditions des tests paramétriques sont réunies, l'avantage de ce dernier sur les test non paramétriques n'est pas fondamentalement très grand.

De fait, lors d'une phase exploratrice où l'on recherche à distinguer les informations que recèlent les données, s'affranchir d'hypothèses contraignantes est un atout supplémentaire.

Lorsque les effectifs sont faibles, à moins que l'hypothèse de normalité ne soit établie, les test non paramétriques sont sans concurrence. Des tables statistiques sont disponibles pour réaliser des tests exacts. De plus, la convergence vers les lois asymptotiques est très rapide.

La présentation des différents tests qui suivent se veut plus pratique que théorique et est largement inspirée des ressources en ligne [10], [11], [12].

## 5.1 Test de Wilcoxon - Mann - Whitney

### 5.1.1 Echantillon indépendant

Le test de Wilcoxon - Mann - Whitney recouvre en réalité deux formulations, totalement équivalentes, le test de Wilcoxon d'une part, le test de Mann-Whitney d'autre part. Il sera présenté dans cette partie le dernier de ces tests.

Nous observons, de manière indépendante, une variable aléatoire  $X$  de loi continue sur une population divisée en deux sous populations d'effectifs  $n_1$  et  $n_2$ . Nous obtenons ainsi deux séries d'observations notée  $(x_1, \dots, x_{n_1})$  pour la première et  $(y_1, \dots, y_{n_2})$  pour la seconde. Nous notons  $\mathcal{L}_k$  la loi de la variable aléatoire  $X$  sur la sous population d'ordre  $k$ , et nous notons  $F_k$  sa fonction de répartition.

Le test de Mann-Whitney teste l'hypothèse suivante :

$$\begin{aligned}\mathcal{H}_0 : \mathcal{L}_1(X) &= \mathcal{L}_2(X) \\ \mathcal{H}_1 : \mathcal{L}_1(X) &\neq \mathcal{L}_2(X)\end{aligned}$$

Non paramétriques ne veut pas dire absence total de l'étude de paramètres des distributions lors de la mise en place des tests. Cela indique surtout que le test ne repose pas sur les paramètres des présupposées distributions des données. Nous pouvons donc tester la différence brute entre les fonctions de répartitions, comparer les caractéristiques de localisation ou le décalage entre les fonctions de répartitions.

C'est pourquoi nous pourrions reformuler le test de Mann-Whitney de façon strictement équivalente comme un test comparant les caractéristiques de tendance centrale, à ne surtout pas confondre avec la moyenne. On teste alors :

$$\begin{aligned}\mathcal{H}_0 : F_1(x) &= F_2(x + \theta), \theta = 0 \\ \mathcal{H}_1 : F_1(x) &\neq F_2(x + \theta), \theta \neq 0\end{aligned}$$

### Rangs et somme des rangs

Comme la grande majorité des tests non paramétriques, le test de Mann-Whitney repose sur la notion de rangs. L'idée est de substituer aux valeurs leur numéro d'ordre dans l'ensemble des données.

Pour cela on classe par ordre croissant l'ensemble des observations des deux échantillons  $(x_1, \dots, x_{n_1})$  et  $(y_1, \dots, y_{n_2})$ . Nous affectons à chaque valeur  $x_i$  (respectivement  $y_i$ ) la valeur de son rang, noté  $r_{i1}$  (respectivement  $r_{i2}$ ). Puis, on calcule la somme des rangs cette fois sous chaque échantillon, notée  $S_1$  et  $S_2$  et définie ainsi :

$$S_k = \sum_{i=1}^{n_k} r_{ik}$$

On définit aussi la somme des rangs moyens notée  $\bar{r}_k$  ainsi :

$$\bar{r}_k = \frac{S_k}{n_k}$$

**Remarque.** Si le test de Mann-Whitney présenté plus loin est significatif à un seuil fixé, les rangs moyens donnent une idée de la localisation d'une distribution par rapport à l'autre.

## Traitements des ex-aequos

Il arrive parfois qu'il y ait des individus ayant les mêmes valeurs dans les données observées. Afin d'attribuer un rang à ces ex-aequos, il est possible de procéder par deux méthodes.

La première assigne au hasard un rang aux ex-aequos, le bon fonctionnement de cette technique ne sera pas détaillée ici.

La deuxième fait appelle à la notion de rang moyen. Cette méthode attribue aux individus ayant la même valeur la moyenne de leur rang respectif. Cette méthode semble plus puissante que la première mais nécessite une correction de la statistique de test.

## Statistique de test et région critique

La statistique de Mann-Whitney utilise donc la somme des rangs, à travers cela on retrouve bien l'idée de décalage entre les distributions basées sur leur localisation respective. On définit deux quantités  $U_1$  et  $U_2$  telles que

$$U_1 = S_1 - \frac{n_1(n_1 + 1)}{2} \quad (5.1)$$

$$U_2 = S_2 - \frac{n_2(n_2 + 1)}{2} \quad (5.2)$$

Par convention la statistique de Mann-Whitney correspond à la plus petite quantité, soit

$$U = \min(U_1, U_2)$$

Sous  $\mathcal{H}_0$ , l'espérance et la variance de  $U$  s'écrivent :

$$(i) \mathbb{E}(U) = \frac{n_1 n_2}{2} \quad (5.3)$$

$$(ii) \mathbb{V}(U) = \frac{1}{12}(n_1 + n_2 + 1)n_1 n_2 \quad (5.4)$$

**Démonstration.** *La démonstration est basée sur l'approche initiale de Mann-Whitney : le nombre total d'inversions relativement à l'échantillon  $Y$  (respectivement  $X$ ). La plus petite de ces sommes d'inversions est appelée statistique  $U$  de Mann-Whitney. Elle est identique à celle exposée plus haut. On démontre ensuite que sous  $\mathcal{H}_0$   $U$  est symétrique, vient ensuite (i) et (ii) [12].*

Lorsque la taille des échantillons  $n_1$  ou  $n_2$  est inférieure à 20 avec un seuil donné  $\alpha$ , les tables de Mann-Whitney nous fournissent une valeur critique  $c$ . Et nous décidons :

$$\mathcal{H}_1 \text{ vraie si } U_{n_1, n_2, obs} \leq c$$

$$\mathcal{H}_0 \text{ vraie si } U_{n_1, n_2, obs} > c$$

## Loi asymptotique : l'approximation normale

Pour des grandes valeurs de  $n_1$  et  $n_2$  la loi  $U$  n'est plus tabulée. La raison est la suivante : posons

$$\tilde{U} = \frac{U - \mathbb{E}(U)}{\sqrt{\mathbb{V}(U)}} = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{1}{12}(n_1 + n_2 + 1)n_1 n_2}} \quad (5.5)$$

On a alors :

$$\tilde{U} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1), \text{ avec } n_1 \rightarrow \infty \text{ et } n_2 \rightarrow \infty \quad (5.6)$$

Dans la pratique l'approximation par la loi  $\tilde{U}$  semble appropriée pour des échantillons  $n_1 > 8$  et  $n_2 > 8$ .

La région critique du test au niveau de signification  $\alpha$  est  $|\tilde{U}| \geq u_{1-\frac{\alpha}{2}}$ , avec  $u_{1-\frac{\alpha}{2}}$  le quantile d'ordre  $1 - \frac{\alpha}{2}$  de la loi normale centrée réduite.

## La correction de continuité

Lorsque les effectifs sont de taille modérée, nous pouvons améliorer l'approximation normale en introduisant la correction de continuité. Notre statistique de test s'écrit pour un test bilatéral :

$$|\tilde{U}| = \frac{|U - \mathbb{E}(U)| - 0.5}{\sqrt{\mathbb{V}(U)}} \quad (5.7)$$

## Correction pour les ex-aequos

Lorsque les ex-aequos sont associés à des individus de la même sous population, la statistique du test n'est pas modifiée. En revanche, dans le cas contraire, la statistique du test est modifiée par rapport à la méthode des rangs aléatoires. Néanmoins la différence est négligeable.

En revanche, la variance de la statistique doit être corrigée. La formule corrigée de la variance à utiliser  $\tilde{V}(U)$  est telle que

$$\tilde{V}(U) = V(U) \times \left( 1 - \frac{\sum_{g=1}^G t_g(t_g^2 - 1)}{n^3 - n} \right) \quad (5.8)$$

Où  $n = n_1 + n_2$ , l'effectif total,  $G$  est le nombre de valeurs distinctes dans l'échantillon total,  $t_g$  est le nombre d'observations associées à la valeur numéro  $g$ .

Cette correction peut être assez importante lorsque le nombre de valeurs différentes est faible. C'est inmanquablement le cas lorsque les variables sont ordinales.

### 5.1.2 Echantillon apparié

L'appariement est une procédure très populaire en statistique. Elle permet une analyse fine des différences entre les populations. Le test des rangs signés de Wilcoxon traite la comparaison d'échantillon apparié.

Nous observons, de manière indépendante, une variable aléatoire en deux temps : on observe donc un couple  $(X_1, X_2)$  sur une population d'effectif total  $n$ . Nous obtenons alors une série de couples appariés  $\{(x_{11}, x_{12}), \dots, (x_{n1}, x_{n2})\}$ . On définit une nouvelle variable aléatoire  $D$  telle que  $D = X_1 - X_2$ .

Le test de Wilcoxon travaille sur les  $n$  réalisations de la variable  $D$ , notées  $d_i$  telles que  $d_i = x_{i1} - x_{i2}$ . Enfin le test teste l'égalité des fonctions de répartition de la variable  $X_1$  et  $X_2$ , notées  $F_{X_1}$  et  $F_{X_2}$ .

Pour cela, nous construisons la valeur absolue des écarts  $|d_i|$ . A partir de  $|d_i|$ , nous définissons les rangs des observations  $r_i$ . Ainsi la plus petite valeur des écarts, en valeur absolue, reçoit le rang 1 : le plus grand écart, reçoit le rang  $n$ .

Compte tenu des rangs, nous pouvons calculer la statistique  $T^+$ , elle correspond à la somme des rangs pour les observations présentant un écart positif ( $d_i > 0$ ).

$$T^+ = \sum_{i, d_i > 0} r_i \quad (5.9)$$

Nous pourrions définir  $T^-$  pour les écarts négatifs de la même manière. Mais sachant que la somme totale des rangs est égale à  $\frac{n(n+1)}{2}$ , nous pouvons déduire  $T^-$  à partir de la relation suivante

$$T^- = \frac{n(n+1)}{2} - T^+$$

Sous  $\mathcal{H}_0$ , on a  $\forall x F_{X_1}(x) = F_{X_2}(x)$ , nous aurons  $\mathbb{E}(T^+) = \mathbb{E}(T^-) = \frac{1}{2} \frac{n(n+1)}{2}$ . Ainsi plus  $T^+$  sera grand (resp. petit) par rapport à  $T^-$ , plus nous serons amenés à conclure que les valeurs prises par  $X_1$  sont stochastiquement plus élevées (resp. plus faibles) par rapport à celles de  $X_2$ .

Pour un test bilatéral, la région critique du test au risque  $\alpha$  s'écrit  $(T^+ \geq T_\alpha)$  où le seuil critique  $T_\alpha$  est lue dans une table spécifique due à Wilcoxon.

### Traitement des écarts nuls

Les écarts nuls  $d_i = 0$  posent problèmes car ils ne rentrent pas dans le canevas de notre modélisation. La solution usuelle consiste tout simplement à supprimer les observations correspondantes. De fait, la « vraie » taille d'échantillon  $n$  correspond au nombre d'observations pour lesquelles  $d_i \neq 0$ . Les rangs sont calculés uniquement sur ces individus.

### Les rangs moyens pour les ex-aequos

La situation est différente lorsqu'il existe des ex-aequos. Dans ce cas, plusieurs observations présentent une valeur identique de  $|d_i|$ , nous devons leur attribuer un rang identique. Nous utilisons le principe des rangs moyens déjà vu au paragraphe 5.1.1 page 29.

### Approximation normale

Lorsque  $n$  est assez grand, ( $n > 15$ ), en pratique, on peut approximer la distribution de  $T^+$  sous  $\mathcal{H}_0$  par une loi normale de paramètres

$$\begin{aligned}\mathbb{E}(T^+) &= \frac{1}{4}n(n+1) \\ \mathbb{V}(T^+) &= \frac{1}{24}n(n+1)(2n+1)\end{aligned}$$

On a alors

$$Z = \frac{T^+ - \frac{1}{4}n(n+1)}{\sqrt{\frac{1}{24}n(n+1)(2n+1)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \text{ avec } n \rightarrow \infty$$

La région critique pour un test bilatéral au risque  $\alpha$  s'écrit  $|Z| \geq u_{1-\frac{\alpha}{2}}$ . Pour les effectifs modérés, il est possible d'utiliser la correction de continuité comme vu au paragraphe 5.1.1 à la page 30.

### Correction de la variance pour les ex-aequo

Lorsque les  $|d_i|$  comportent des ex-aequos, il est nécessaire de corriger la variance de la statistique de test.

Soit  $G$  le nombre de valeurs distinctes de  $|d_i|$ , pour la valeur numéro  $g$ , nous observons  $t_g$  observations, la variance corrigée s'écrit alors

$$\tilde{\mathbb{V}}(T^+) = \frac{1}{24}n(n+1)(2n+1) - \frac{1}{48} \sum_{g=1}^G t_g(t_g-1)(t_g+1) \quad (5.10)$$

## 5.2 Test exact de Fisher

Le test exact de Fisher, à ne pas confondre avec le test de Fisher pour les variances, est un test statistique utilisé pour l'analyse des tableaux de contingence, initialement  $2 \times 2$ . Ce test qui doit son nom à son inventeur Ronald Fisher est utilisé en général pour les faibles effectifs, il reste



néanmoins valide pour toute taille d'échantillon. A noter que pour les grandes tailles d'échantillons, le test rejette trop souvent à tort l'hypothèse nulle. Ce test est qualifié d'exact car les probabilités peuvent être calculées exactement, à l'inverse d'un test comme le  $\chi^2$ .

Nous observons, de manière indépendante, une variable aléatoire  $X$  sur une population divisée en deux sous populations. On note  $X_k$  la variable aléatoire de la sous population d'ordre  $k$ . On obtient alors deux échantillons notés  $x_1$  et  $x_2$ . On introduit également un facteur à deux modalités notées  $A$  et  $B$ .

Le test exact de Fisher permet de tester si les deux échantillons  $x_1$  et  $x_2$  sont issus d'une seule et même population ou de deux populations différentes.

Pour reformuler d'une autre façon, si on note  $p_{X_1}$  (resp.  $p_{X_2}$ ) la proportion de la modalité  $A$  dans la sous population 1 (resp. 2) on teste alors :

$$\begin{aligned}\mathcal{H}_0 &: p_{X_1} = p_{X_2} \\ \mathcal{H}_1 &: p_{X_1} \neq p_{X_2}\end{aligned}$$

Afin de continuer la suite des explications, il est nécessaire de disposer d'un tableau de contingence à quatre cases présentant les effectifs des deux échantillons suivant les deux modalités  $A$  et  $B$ .

Echantillon \ Modalité	$x_1$	$x_2$	Sommes
$A$	$a$	$b$	$a + b$
$B$	$c$	$d$	$c + d$
Sommes	$a + c$	$b + d$	$n$

On note  $n = a + b + c + d$

Sous  $\mathcal{H}_0$ , Fisher a montré que la probabilité d'obtenir un tel ensemble de valeurs était donnée par la loi hypergéométrique de paramètre :  $\mathcal{H}(a + c, \frac{a+b}{n}, n)$  et la probabilité est notée  $p_1$  telle que

$$p_1 = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{n}{a+c}} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!} \quad (5.11)$$

La suite du test consiste à calculer cette probabilité dans un cas plus extrême que dans l'échantillon, tout en gardant les sommes marginales fixes (i.e. la loi  $\mathcal{H}$  ne change pas). On note ces probabilités  $p_k$ ,  $k \geq 2$ . Une technique employée par la fonction `fisher.test` du logiciel *R* est de calculer toute les probabilités  $p_k$  inférieures à  $p_1$ .

Après cela on obtient  $p = \sum_{k \geq 1} p_k$ , qui est la probabilité de rejeter à tort  $\mathcal{H}_0$ . Cette valeur est comparée à la valeur seuil  $\alpha$  fixée. Si  $p > \alpha$  on ne rejette pas  $\mathcal{H}_0$ . Sinon, si  $p \leq \alpha$ , on rejette l'hypothèse nulle, et on conclut que les échantillons proviennent de deux populations différentes.

### 5.3 Test du coefficient de corrélation de Spearman

Pour valider l'existence d'un lien entre deux variables, on réalise ordinairement une régression linéaire simple ou non linéaire. La qualité du lien supposée est mesurée par le coefficient de corrélation de Pearson. Cependant, il existe des situations pour lesquelles une mesure de la corrélation est inadaptée. Si les valeurs sont ordinales, discrètes, si les valeurs elles même n'ont que peu d'importance ou si elles ne suivent pas une loi normale, nous sommes amenées à calculer la corrélation des rangs. Le coefficient de corrélation de Spearman utilise cette méthode.

Nous observons, de manière indépendante, un couple de variables aléatoires  $(X, Y)$  sur une population d'effectif total  $n$ . Nous obtenons alors une série bivariée  $\{(x_1, y_1), \dots, (x_n, y_n)\}$ .

On commence par classer par ordre croissant les valeurs observées de  $X$  séparément en leur attribuant leur rang  $r_{x_i}$ . On fait de même avec les valeurs observées de  $Y$ , on note  $r_{y_i}$  leur rang.

On définit alors le coefficient de corrélation de Spearman  $r_s$  comme le coefficient de corrélation de Pearson sur les rangs calculés

$$r_s = \frac{\frac{1}{n} \sum_{i=1}^n (r_{x_i} - \bar{r}_x)(r_{y_i} - \bar{r}_y)}{\sqrt{\frac{1}{n} \sum_{i=1}^n (r_{x_i} - \bar{r}_x)^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (r_{y_i} - \bar{r}_y)^2}} \quad (5.12)$$

Où  $\bar{r}_x$  et  $\bar{r}_y$  sont les moyennes respectives des rangs  $r_{x_i}$  et  $r_{y_i}$ .

**Remarque.**

- *Lorsqu'il existe des ex-aequos dans les séries, on fait appel à la moyenne des rangs vue au paragraphe 5.1.1, à la page 29.*
- $-1 \leq r_s \leq 1$

Lorsque aucune des séries ne possède d'ex-aequos, on a

$$\bar{r}_x = \bar{r}_y = \frac{1}{n} \frac{n(n+1)}{2} = \frac{(n+1)}{2}$$

De plus

$$r_s = 1 - \frac{6 \sum_{i=1}^n (r_{x_i} - r_{y_i})^2}{n(n^2 - 1)} \quad (5.13)$$

Le test du coefficient de corrélation de Spearman a pour hypothèse nulle la nullité de  $r_s$ , et pour hypothèse alternative pour le test bilatéral  $r_s \neq 0$ .

La statistique de test  $r_s$  est tabulée pour des effectifs inférieurs à 10. Au delà, il est possible d'approximer la table en utilisant la statistique de test suivante

$$t = r_s \sqrt{\frac{n-2}{1-r_s}} \quad (5.14)$$

Cette statistique suit une loi de Student à  $n - 2$  degrés de liberté.

# Chapitre 6

## Résultats des analyses statistiques

Notre collecte de données étant terminée, il est nécessaire à ce stade d'analyser nos indicateurs afin de mieux comprendre leurs fluctuations. Nous nous concentrerons dans un premier temps sur les mesures réalisées au sonomètre et essayerons de déceler les informations que possèdent ces données.

Pour cela, nous utiliserons les tests statistiques précédemment introduits. Nous sommes dans une phase très exploratoire, ainsi pouvoir s'affranchir d'hypothèses contraignantes est un avantage au vue du nombre de tests effectués. De plus, les effectifs parfois très réduits de notre échantillon nous poussent à utiliser des tests non paramétriques. A ce propos, nous fixons désormais le seuil  $\alpha$  de la significativité à 5%.

### 6.1 Heures de pointe contre heures creuses

Pour rappel, nous possédons 92 mesures de trajets au sonomètre. 45 enfants ont fait l'objet de deux mesures : une en heures de pointe, et une en heure creuses. Nous faisons face à des mesures appariées, pour traiter l'existence d'une différence entre ces deux sous-populations au seuil  $\alpha$ , nous procédons au test des rangs signés de Wilcoxon.

Le résultat du test rapporte que seul le *SEL* est significatif (p-value = 0.045). Nous notons aussi que malgré les différences observées sur l'indice *Harmonica*, le test n'est pas significatif (p-value = 0.06).

Ainsi, il est possible de conclure avec un degré de confiance à 95% que l'heure des mesures n'influe pas significativement les indicateurs acoustiques, excepté le *SEL*.

### 6.2 Les éléments cartographiques

Comme nous avons un jeu de données contenant plusieurs mesures pour un individu. Nous décidons de ne retenir qu'une mesure unique par individu afin de conserver l'indépendance des mesures. Bien que nous avons montré que l'heure des mesures n'influe pas significativement les indicateurs, on choisit de créer les trois jeux de données ci-dessous afin de mieux connaître le comportement de notre analyse spatiale.

- 46 mesures pour 46 enfants pour les heures de pointe → Jeu de données 1 (J1)
- 46 mesures pour 46 enfants pour les heures creuses → Jeu de données 2 (J2)
- 47 mesures pour 47 enfants. Ces 47 mesures ont été choisies au hasard. → Jeu de donnée 3 (J3) *Ce troisième jeu de données est composé de 22 mesures en heures de pointe et 35 mesures en heures creuses.*

Nous avons procédé dans un premier temps au test de Mann-Whitney sur les différents jeux de données afin de tester la différence entre l'absence et la présence de carrefour, parking, parc, zone piétonne, voie sans issue, voie de catégorie 3, voie de catégorie 4, voie de catégorie. Le test se base sur la somme des rangs, ainsi il est possible de positionner un groupe par rapport à l'autre. Certains des résultats sont représentés en annexe B.4 page 48 et A.7 page 43.

### 6.2.1 Les voies catégorisées

Concernant l'absence / présence d'une voie catégorisée durant le trajet, nous constatons que la longueur et la durée sont inférieures pour les trajets n'ayant pas de voie (p-value  $< 0,05$ ). De plus, on confirme bien une corrélation forte entre ces deux variables métriques et temporelles ( $r = 0,8$  et p-value  $< 0,05$ ). En heures de pointe le test est significatif pour le niveau  $L_{Aeq}$ ,  $IIE$ ,  $IHE$ . L'étalement ( $IIE$ ) et le niveau sonore ( $L_{Aeq}$ ) est plus fort en présence de voie catégorisée. Cependant,  $IHE$  indique une homogénéité plus forte en présence de voie catégorisée.

L'indice harmonica est inférieur pour J1 et J2 en l'absence de voie catégorisée (p-value  $< 0,05$ ). Enfin, peu importe le jeu de données, les indicateurs de bruyance, en particulier  $PN_{<70}$  répondent significativement aux tests. Cela est toujours rassurant de constater que  $PN_{<70}$  est plus élevé en présence de voies catégorisées. De plus les durées de ces séquences sont plus longues en présence des voies.

Les tests vont dans le sens de l'information que donne le classement sonore des infrastructures routières : les trajets qui interceptent ces voies catégorisées montrent un étalement plus large, un niveau sonore élevé, ainsi que des séquences gênantes longues.

### 6.2.2 Traversées de voies, carrefours

On teste la présence et l'absence de carrefours (ou ronds point), notre intuition aurait tendance à croire que le niveau sonore serait plus fort en présence d'intersections traversées, mais aucune des J1, J2 et J3 ne relève une significativité (resp. p-value = 0,062, p-value = 0,35 et p-value = 0,25). Cependant durant les heures de pointe, les tests sur l'indice *Harmonica* et le nombre de séquences supérieures à 70 dB(A) sont significatifs :  $NP_{>70}$  est supérieur en présence de carrefours ou rond points, et l'indice *Harmonica* également.

Nous avons également testé les trajets n'ayant aucune ou une traversée de voie contre plusieurs. Cette fois, dans le cas du J1 le niveau sonore  $L_{Aeq}$  est plus faible dans le premier groupe (p-value  $< 0,05$ ). L'indice *Harmonica* est plus élevé dans le J1 et le J3 pour le second groupe.

Dans le cas du J1, J2 et J3,  $IIE$  est plus fort en présence de plusieurs voies traversées. Cela est conforme à notre intuition : un trajet est soumis à d'avantage de niveaux sonores dans des lieux supposés très hétérogènes. De plus la corrélation entre  $IIE$  et la longueur du trajet est faible ( $r < 0,5$  et p-value  $< 0,05$ ).

On décide de calculer le coefficient de corrélation de Spearman entre les indicateurs et le nombre de voies traversées pour J3. Les corrélations sont faibles mais significatives pour  $IIE$  ( $r = 0,54$ ),  $NP_{>70}$  et  $T_{>70}$  ( $r \approx 0,5$ ),  $L_{Aeq}$  ( $r = 0,3$ ), et *Harmonica* ( $r = 0,37$ ).

Les tests de corrélation tendent à prouver que les rencontres avec des voies à traverser ont pour conséquence d'augmenter la gêne créée par le bruit. Aucune corrélation négative n'est significative.

### 6.2.3 Parcs, zones piétonnes et voies sans issue

Ces éléments sont considérés comme des lieux calmes, cependant les tests face aux indicateurs ne sont pas très révélateurs : aucun résultats n'est identique pour les trois jeux de données.

Pour J2, l'IDAS est significativement différent en l'absence ou en présence de parc, l'augmentation sonore rapide est moins présente en présence d'un parc. A titre informatif la médiane se situe à 0,4 en présence de parc, et à -0,2 en l'absence de parc. L'*IHE* est aussi significativement différent : en présence de parc la distribution des  $L_{A,1sec}$  est plus homogène.

Pour J1 et J3, on note que  $T_{<L_{90}}$  est plus élevé en présence de parc.

Les indicateurs étudiés sont très peu parlant pour ces zones. On note tout de même que si peu parlant soient-ils, ils vont dans le sens de la tranquillité sonore.

### 6.2.4 Le nombre d'éléments cartographiques

Nous nous sommes aussi intéressés à la façon dont se comportaient les indicateurs face au nombre d'éléments présents pour un trajet. Pour cela nous utilisons le coefficient de corrélation de Spearman.

Tout d'abord, la durée et la longueur du trajet sont bien corrélées avec le nombre d'éléments ( $r \approx 0,6$ ).

Aucune corrélation n'est très forte cependant il existe quelques corrélations significatives qu'il faut noter. Pour J1, J2 et J3 le niveau de crête ( $L_{10}$ ) est corrélé avec le nombre d'éléments (resp.  $r = 0,44$ ,  $r = 0,3$  et  $r = 0,3$ ). Le nombre de séquences supérieures à 70 dB(A) est aussi très corrélé pour les trois jeux de données (resp.  $r = 0,53$ ,  $r = 0,45$  et  $r = 0,48$ ). Il en est de même avec la durée moyenne ces séquences ( $r \approx 0,4$ ).

Pour les tables J1 et J3, l'indice *Harmonica* est également positivement corrélé (resp.  $r = 0,42$  et  $r = 0,35$ ). Dans une moindre mesure l'*ILE* et l'indice de rugosité sont positivement corrélés ( $r < 0,4$ ).

Nous sommes aussi intéressé par le comportement des trajets définis comme homogène. Pour cela, nous procédons à un test opposant les trajets homogènes contre les trajets hétérogènes.

Concernant les trois jeux de données, la proportion de séquences de plus de quatre secondes supérieures à 70 dB(A) est plus élevée dans le cas d'un trajet hétérogène que dans le cas d'un trajet homogène.

Concernant les jeux de données J2 et J3, le  $L_{Aeq}$ ,  $L_{10}$ , l'*ILE* et l'indice *Harmonica* sont significativement inférieurs dans le cas des trajets homogènes.

La durée moyenne des séquences de quatre secondes considérées comme tranquille dans le cas de J2 sont, avec une confiance de 95%, plus longues pour les trajets homogènes que les trajets hétérogènes.

Cependant, pour le jeu J3, la durée moyenne des séquences tranquilles inférieures au bruit de fond sont plus courtes dans le cas des trajets homogènes que hétérogènes.

Cette distinction entre trajets homogènes et hétérogènes semble appropriée car les indicateurs dynamiques ont tendance à donner des informations bien différentes suivant le type de trajet.

### 6.3 Axes catégorisées contre hétérogénéité

On établit un tableau de contingence entre les trajets interceptant une voie de catégorie 3 ou 4 et les trajets homogènes. On obtient le tableau 6.1 dans le cas du jeu J3.

	Absence d'axe cat.	Présence d'axe cat.
homogène	6	1
hétérogène	13	27

TABLE 6.1 – Répartitions des trajets selon les deux variables

On applique le test exact de Fisher, adéquat pour ce type d'effectif, afin de qualifier le lien entre l'homogénéité et la présence d'axe. Le test donne une p-value  $< 0,05$ , ce qui implique que l'on rejette l'indépendance entre absence d'axe catégorisée et l'homogénéité. Le test est aussi significatif pour les deux autres jeux de données.

### 6.4 Sonomètre contre dosimètre

Il a été fait la remarque que nous possédions 63 trajets réalisés avec le dosimètre dont le tracé géographique est le même que les mesures réalisées au sonomètre. Parmi ces trajets, seulement 20 d'entre eux correspondent exactement à une mesure au dosimètre pour un unique enfant : Le sens du trajet et la période sont les mêmes.

Afin de connaître le comportement des indicateurs dans ces deux contextes, nous appliquons le test de Wilcoxon pour mesures appariées. On se base sur la somme des rangs pour donner une localisation de notre distribution.

Le niveau  $L_{Aeq}$  est significativement supérieure dans le cas des mesures réalisées au dosimètre. L'*ILE*, l'indice d'étalement, est également supérieure pour les mesures dosimètre. L'*IDAS*, le degré d'augmentation sonore, n'est pas significatif (p-value = 0,89), il en est de même pour l'*IHE* (p-value = 0,4). L'indice *Harmonica* est bien supérieure dans le cadre des mesures dosimètre (p-value  $< 0,05$ ).

Le nombre de séquence de plus de quatre secondes supérieures à 70 dB(A) semble significativement plus élevées dans le cas du dosimètre que du sonomètre. De plus ces séquences sont également plus longues. Concernant le nombre de séquence pour les indicateurs de tranquillité, le test n'est pas significatif. Néanmoins, la durée des séquences inférieures à 60 dB(A) est plus longue pour les mesures réalisées au sonomètre (p-value  $< 0,05$ ).

La différence déjà notée de la rugosité par rapport aux deux types de mesures est significative : les enregistrements au dosimètre sont plus rugueux.

### 6.5 Synthèse des résultats

Nous souhaitons étudier les indicateurs dynamiques lors des trajets école - domicile via une analyse spatiale post-mesures, les résultats sont présentés dans les sections précédentes et en annexe. Cette approche semble être une bonne alternative à une analyse pré-mesures, où nous aurions pu, avec une grande précision, connaître le comportement exact des indicateurs. Ici, l'idée globale de découper les trajets en éléments cartographiques est à retenir. Il est particulièrement judicieux de prendre en compte le nombre d'éléments cartographiques par trajet. Cependant, après cette première réflexion, il serait nécessaire de mieux définir ou proposer d'autres éléments cartographiques.

# Conclusion

Durant deux mois, j'ai participé activement aux travaux de recherches concernant le bruit environnemental chez les enfants scolarisés en cycle élémentaire. Pour cela, il m'a été confié d'étudier les indicateurs dynamiques lors des trajets école - domicile de ces enfants.

A travers cet objectif, il a fallu que je me familiarise avec la thématique acoustique et particulièrement avec les indicateurs dynamiques. Dans cette première étape, il m'a été nécessaire de savoir accéder à l'information adéquate, la sélectionner et l'utiliser.

Dans un second temps, pour répondre à la problématique fixée, gérer les données collectées durant les campagnes de mesures sous forme de base de données informatique a été essentiel. L'apprentissage de nouvelles technologies informatiques s'est avéré aussi important.

Enfin, l'apprentissage et la mise en pratique de nouvelles techniques statistiques a été un point important, sans quoi les analyses n'auraient pu aboutir.

Ce stage s'est déroulé dans un cadre qui a permis une large autonomie, sans oublier pour le travail de fond une collaboration avec les chargés de l'étude. Ce travail en équipe est, selon moi et pour moi, une force dans ce type de démarche scientifique afin de garder une efficacité dans le travail.

Ce stage m'a également permis de me confronter à la recherche en santé publique et à ses contraintes. Mon avis est double : j'ai apprécié la thématique, l'objectif, ainsi que la collaboration avec les experts du milieu. Cependant, je pense avoir été déstabilisé par la liberté même que donne la recherche.

D'un point de vue statistique, cette expérience m'a donné la possibilité d'enrichir mes connaissances et de les appliquer. Malgré cela, je suis légèrement déçu : selon moi, les études en filière scientifique et particulièrement mathématiques sont souvent trop théoriques et manquent tellement de pratique. Pour cette première occasion de mise en pratique dans le milieu professionnel mes attentes étaient sûrement très élevées. De plus, il est évident que mes connaissances et mon expérience encore limitées ont pu être un frein dans mes analyses statistiques au cours de ce stage.

En conclusion, ce premier stage a été une bonne expérience et m'encourage à prendre connaissance de nouvelles méthodes statistiques à des fins plus pratiques.

A nouveau, je remercie l'ensemble du personnel du service CMC du CHRU de Besançon.

# Bibliographie

- [1] S. Pujol, J-P. Levain, H. Houot, and al. Association between ambient noise exposure and school performance of children living in an urban area : A cross-sectional population-based study. *J Urban Health*, 2013.
- [2] S. Pujol, H. Houot, M. Berthillier, and al. Modélisation de l'exposition au bruit en milieu urbain et études épidémiologiques : quelles sources sonores ? In : Actes des Neuvièmes rencontres de Théo Quant. Besançon : Foltête J.-C ; 2009 :10.
- [3] S. Pujol, M. Berthillier, J. Defrance, and al. Urban ambient outdoor and indoot noise exposure at home : a population-based study on schoolchildren. *Applied Acoustics*, 2012.
- [4] E. Ksiazek. Etude du niveau d'exposition sonore des enfants lors du trajet école-domicile. Master's thesis, Université de France-Comté, 2012.
- [5] L. Leclercq, A. Can, P. Crepeaux, and al. Estimation dynamique du bruit de circulation en milieu urbain. 2008. Rapport final, étude CSTB num. ER-712-060004-712-ENV, convention ADEME num. 0504C0063.
- [6] S. Marry and J. Defrance. Analysis of perception and representation of sonic public spaces through on site survey, acoustic indicators and in-depth interviews. *Applied Acoustics*, 2012.
- [7] Recommandations de l'OMS.  
[www.bruitparif.fr/bruit-et-sante/recommandations-de-loms](http://www.bruitparif.fr/bruit-et-sante/recommandations-de-loms).
- [8] Harmonica project. [www.noiseineu.eu](http://www.noiseineu.eu).
- [9] J. Defrance, J. Palacino, and B. Marine. Auscultation acoustique des aménagements cyclables en milieu urbain. Avril 2010. 10ème Congrès Français d'Acoustique, Lyon.
- [10] Ricco Rakotomalala. Comparaison de populations, tests non paramétriques, 2008. Université Lumière Lyon 2.
- [11] Frédéric Bertrand and Myriam Maumy. Tests non paramétriques.  
[www-irma.u-strasbg.fr/~fbertran/enseignement/DUS2\\_2011/DUS2\\_CoursNonPara\\_1.pdf](http://www-irma.u-strasbg.fr/~fbertran/enseignement/DUS2_2011/DUS2_CoursNonPara_1.pdf), 2011.
- [12] Léonard Gallardo. Notions sur les statistiques d'ordre et les distributions d'échantillonnage.  
[www.lmpt.univ-tours.fr/~gallardo/Stat2008-4.pdf](http://www.lmpt.univ-tours.fr/~gallardo/Stat2008-4.pdf).



# Annexe A

## Tableaux

Paramètre	Minimum	Maximum	Médiane	Moyenne	Écart-type
Indicateurs acoustiques					
$L_{Aeq}$	53,4	72,6	62,9	63,3	3,9
$SEL$	74,3	97,7	88,9	88,4	4,6
Indicateurs statistiques					
$L_{10}$	53,9	73,3	66,7	66	4,5
$L_{50}$	48,5	67,7	56,8	56,9	4,3
$L_{90}$	41,5	61,9	50,4	50,4	4,4
Indicateurs dynamiques					
$IDAS$	-2,2	7	-0,1	-0,01	1,05
$IHE$	0,2	0,69	0,41	0,42	0,09
$ILE$	0,2	1,07	0,79	0,76	0,19
$BGN$	2,2	6,2	3,8	3,9	0,9
$EVT$	0,6	5,7	3,4	3,5	1,1
Harmonica	4,8	9,5	7,4	7,3	0,9
$PN_{>70}$	0	3,14	0,23	0,46	0,6
$PN_{>L_{10}}$	0	2,34	0,81	0,82	0,33
$PN_{<60}$	0	6,44	2,83	2,9	1,19
$PN_{<L_{90}}$	0	1,57	0,76	0,73	0,33
$T_{>70,X}$	0	18	4,17	3,54	3,64
$T_{>L_{10},X}$	0	18	6,5	6,78	2,74
$T_{<60,X}$	0	76,25	20,98	23,34	14,16
$T_{<L_{90},X}$	0	19	6	6,47	3,21
$moy(\delta L_{Aeq})$	0,93	4,03	2,07	2,08	0,48
$\sigma_{L_{Aeq}}$	1,71	9,65	6,02	6,17	1,69
Indicateurs temporel (seconde) - métrique (mètre)					
Durée	89	769	357,5	366,8	157,2
Longueur	38	994	505,5	483,2	213

TABLE A.1 – Indicateurs - sonomètre - 92 obs.

Paramètre	Minimum	Maximum	Médiane	Moyenne	Écart-type
Indicateurs acoustiques					
$L_{Aeq}$	49,9	90,7	75,3	75	6,9
$SEL$	67,7	120,2	101,1	101,8	8,3
Indicateurs statistiques					
$L_{10}$	53,1	90,5	76,5	77	6,4
$L_{50}$	45,9	77,7	64,6	65,1	6,3
$L_{90}$	38,1	69,7	54,9	54,8	6,5
Indicateurs dynamiques					
$IDAS$	-1,2	1,2	0,1	0,112	0,39
$IHE$	0,27	0,67	0,45	0,46	0,09
$ILE$	0,67	1,6	1,07	1,07	0,21
$BGN$	1,4	7,4	4,5	4,5	1,2
$EVT$	1,9	8,9	5,7	5,6	1,5
Harmonica	4,4	13,9	10	10,1	1,6
$PN_{>70}$	0	6	2,5	2,28	1,45
$PN_{>L_{10}}$	0	1	0,52	0,5	0,26
$PN_{<60}$	0	5	1,67	1,91	1,34
$PN_{<L_{90}}$	0	1,67	0,67	0,62	0,32
$T_{>70,X}$	0	26,92	8,58	9,24	5,43
$T_{>L_{10},X}$	0	26	5,22	5,72	3,77
$T_{<60,X}$	0	35,77	9,62	10,4	6,51
$T_{<L_{90},X}$	0	16,83	5,83	6,16	3,06
$moy(\delta L_{Aeq})$	1,98	7,39	4,22	4,46	1,25
$\sigma_{L_{Aeq}}$	4,25	13,13	8,61	8,64	1,78
Indicateurs temporel (seconde) - métrique (mètre)					
Durée	60	1260	600	570	280,55
Longueur	45	853	511	501,2	198,04

TABLE A.2 – Indicateurs - dosimètre - à pied - 83obs.

Paramètre	Minimum	Maximum	Médiane	Moyenne	Écart-type
Indicateurs acoustiques					
$L_{Aeq}$	64,8	78,1	70,8	70,8	4,1
$SEL$	87,4	105,9	93,3	94,8	6,1
Indicateurs statistiques					
$L_{10}$	67,5	79,2	73,1	73,2	3,9
$L_{50}$	58,5	71,8	61,5	63,9	4,7
$L_{90}$	43,1	64,4	59	56,2	7,3
Indicateurs dynamiques					
$IDAS$	-0,9	0,6	0	-0,04	0,5
$IHE$	0,09	0,51	0,41	0,35	0,13
$ILE$	0,5	1,43	0,9	0,96	0,3
$BGN$	2,4	6,5	5,1	4,8	1,6
$EVT$	2,1	7,2	4	4,19	1,8
Harmonica	7,4	10,5	9	9	0,9
$PN_{>70}$	0	5,05	1,5	1,46	1,51
$PN_{>L_{10}}$	0	0,83	0,33	0,36	0,37
$PN_{<60}$	0	4,17	0,83	1,78	1,7
$PN_{<L_{90}}$	0	1,11	0,5	0,47	0,4
$T_{>70,X}$	0	17,83	5	6,15	5,52
$T_{>L_{10},X}$	0	18	4	4,11	5,75
$T_{<60,X}$	0	33	8,75	9,54	9,69
$T_{<L_{90},X}$	0	14	4,5	4,87	4,9
$moy(\delta L_{Aeq})$	2,01	7,11	3,47	4,03	1,73
$\sigma_{L_{Aeq}}$	4,78	11,46	5,74	6,93	2,32
Indicateurs temporel (seconde) - métrique (mètre)					
Durée	60	600	240	310,7	194,74
Longueur	356	2000	853	997,3	597,61

TABLE A.3 – Indicateurs - dosimètre - en voiture - 9obs.

$L_{Aeq}$	$SEL$	$L_{10}$	$L_{50}$	$L_{90}$	$IDAS$	$IHE$
74,5	99,3	76,5	66,7	61,2	0,6	0,34
$ILE$	$BGN$	$EVT$	Harmonica	$PN_{>70}$	$PN_{>L_{10}}$	$PN_{<60}$
0,87	6	3,6	9,6	2,33	0,66	0
$PN_{<L_{90}}$	$T_{>70,X}$	$T_{>L_{10},X}$	$T_{<60,X}$	$T_{<L_{90},X}$	$moy(\delta L_{Aeq})$	$\sigma_{L_{Aeq}}$
0	0,66	6,43	5,5	0	6	3,16
Durée en secondes				Longueur en mètres		
300				615		

TABLE A.4 – Indicateurs - dosimètre - en vélo - 1obs.

Variable	Effectif	Pourcentage	Variable	Effectif	Pourcentage
Passage piétons			Parking		
0	8	17,02	0	35	74,47
1	17	36,17	1	10	21,28
2	6	12,77	Ponts		
3	9	19,15	0	45	95,74
4	4	8,51	1	2	4,26
5	2	4,25	Carrefour - rond point		
6	1	2,13	0	34	72,34
			1	13	27,66

TABLE A.5 – Elément des 47 trajets uniques au sonomètres (1)

<i>Les distances sont exprimées en mètres</i>						
Variable	Effectif	Pourcentage	min	max	médiane	moyenne
<b>Parc - zone verte</b>						
Absence	38	80,85				
Présence	9	19,15	47	270	26,5	26,64
		<i>Pourcentage du trajet</i>	8,3	52,5	26,5	29,64
<b>Zone piétonne</b>						
Absence	39	82,98				
Présence	8	17,02	72	503	128	223,2
		<i>Pourcentage du trajet</i>	16,1	72,5	28,4	37,99
<b>Voie sans issue</b>						
Absence	40	85,11				
Présence	7	14,89	59	215	88	115
		<i>Pourcentage du trajet</i>	15,2	70	24,6	32,4
<b>Cours intérieure</b>						
Absence	46	97,87				
Présence	1	2,13	123			
		<i>Pourcentage du trajet</i>	17,9			
<b>Voie de catégorie 3</b>						
Absence	35	74,47				
Présence	12	25,53	37	347	152,5	172,9
		<i>Pourcentage du trajet</i>	7,3	63,1	31,7	32,83
<b>Voie de catégorie 4</b>						
Absence	25	53,19				
Présence	22	46,81	45	496	139,5	186,8
		<i>Pourcentage du trajet</i>	9	100	33,9	39,77
<b>Voie de catégorie 3 ou 4</b>						
Absence	19	40,43				
Présence	28	59,57	37	496	220,9	187,5
		<i>Pourcentage du trajet</i>	9	100	38,25	45,31

TABLE A.6 – Elément des 47 trajets uniques au sonomètres (2)



	zone piétonne		voie sans issue		élément cartographique		
	absence	présence	absence	présence	$\leq 1$	$\geq 2$	p
$L_{Aeq}$	62,8	64,4	63,3	62,6	59,7	64,3	$< 10^{-3}$
$SEL$	88,3	89,5	89,2	87,4	83,8	89,4	$< 10^{-3}$
$L_{10}$	66,9	66,9	66,9	66,9	63	67,5	$< 10^{-3}$
$L_{50}$	56,4	57,5	56,4	57,9	54,6	57,7	0,120
$L_{90}$	49,8	51,2	49,8	51,3	50,4	49,9	0,777
$IDAS$	-0,1	-0,25	-0,1	-0,3	-0,3	-0,1	0,427
$ILE$	0,8	0,75	0,84	0,63	0,7	0,87	0,022
$IHE$	0,43	0,42	0,43	0,43	0,42	0,43	0,765
<i>Harmonica</i>	7,4	7,3	7,4	7,1	6,5	7,6	$< 10^{-3}$
$PN_{>70}$	0	0,2	0,2	0	0	0,2	0,010
$PN_{>L_{10}}$	0,8	1	0,8	0,8	0,8	0,8	0,483
$PN_{<60}$	2,7	3,6	2,8	2,1	3,3	2,7	0,483
$PN_{<L_{90}}$	0,8	0,8	0,8	0,5	0,7	0,8	0,151
$T_{>70,X}$	0	4,2	4,3	0	0	4,4	0,010
$T_{>L_{10},X}$	6	6,6	6	4,7	5	6,1	0,330
$T_{<60,X}$	22,7	16,5	22,2	13,8	24,7	20,8	0,232
$T_{<L_{90},X}$	6	5,8	6,2	4,3	5	6,5	0,016
rugosité	2,2	2	2,1	2,3	2	2,2	0,437
durée (s)	363	427	379	335	246	394	0,014
longueur (m)	505	542	508	358	281	517	0,017

TABLE A.8 – Valeurs de la médiane dans le cas de sous population pour le jeu de donnée numéro 3 avec degré de signification (p). (2)

# Annexe B

## Graphiques et cartes

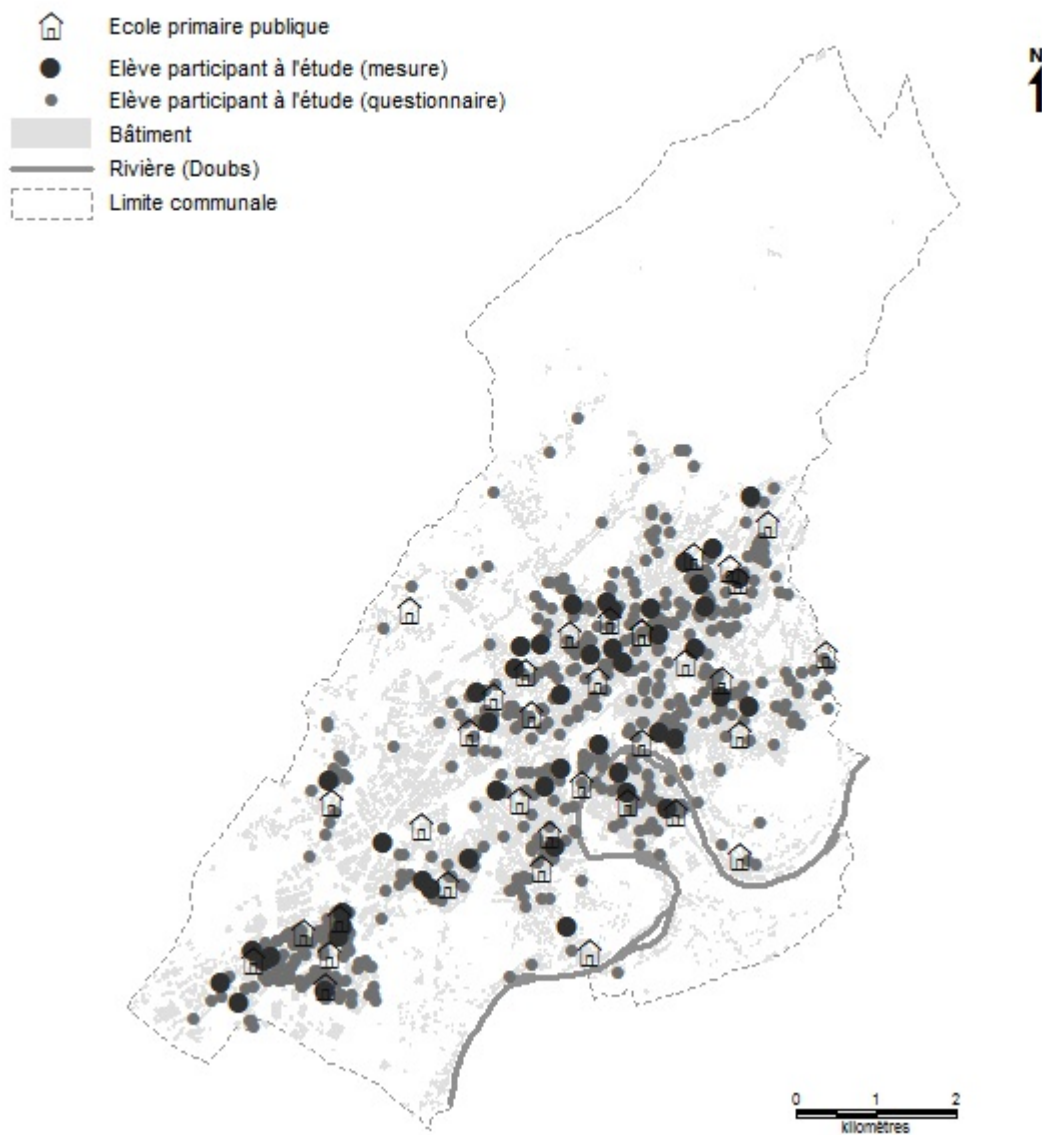


FIGURE B.1 – Répartition des élèves et des écoles de l'étude sur Besançon

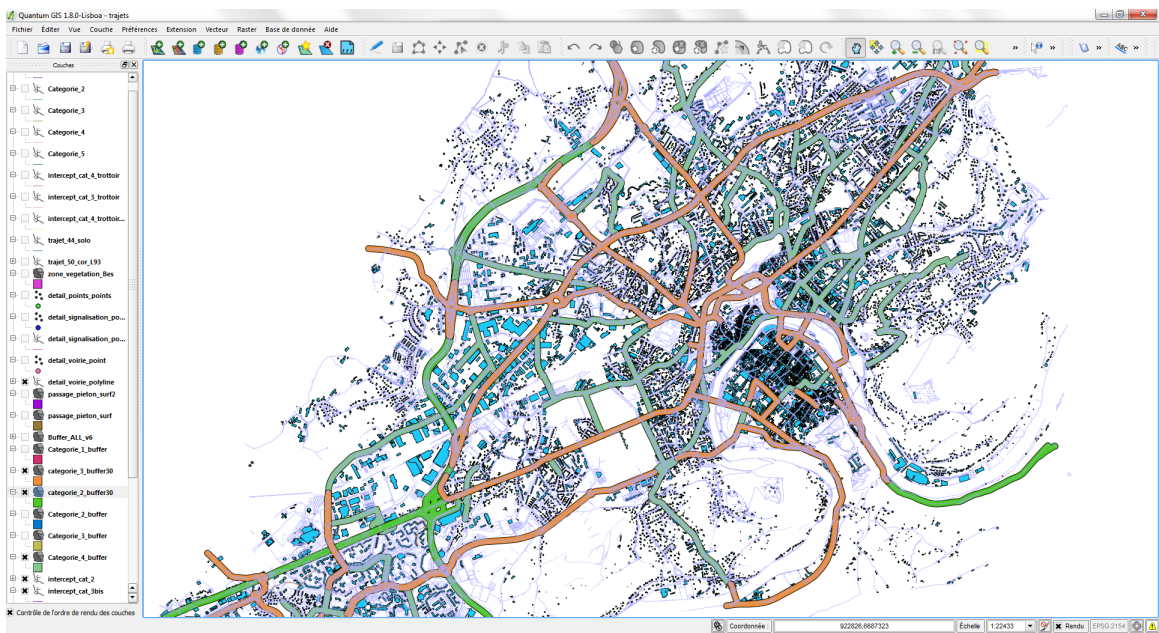


FIGURE B.2 – Interface de QGIS

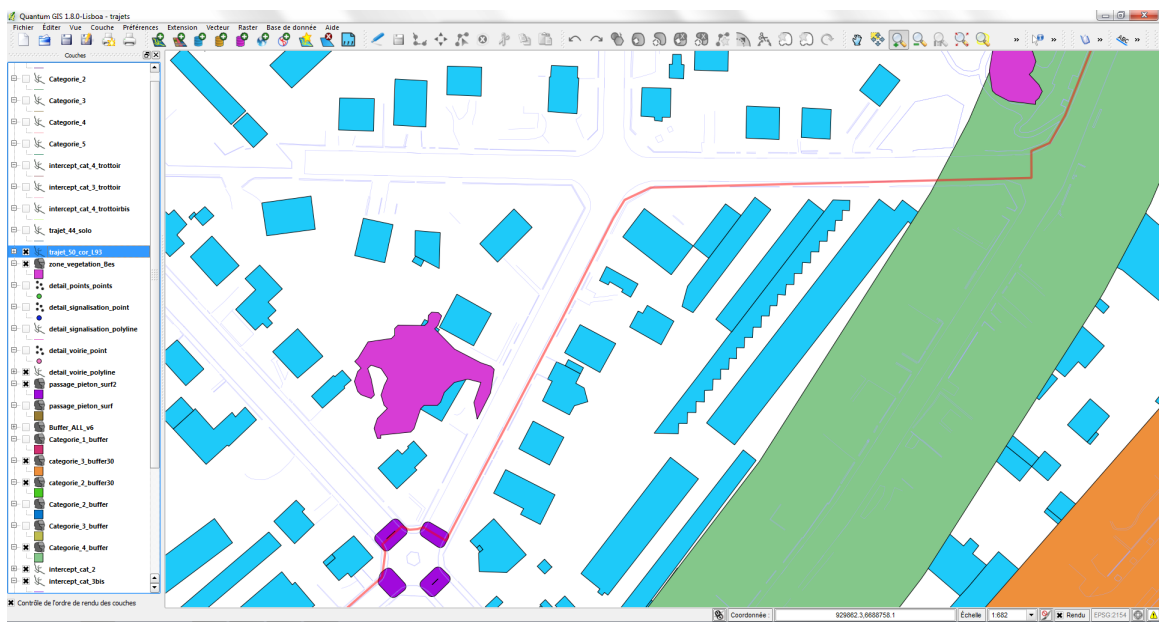


FIGURE B.3 – Un trajet sur QGIS



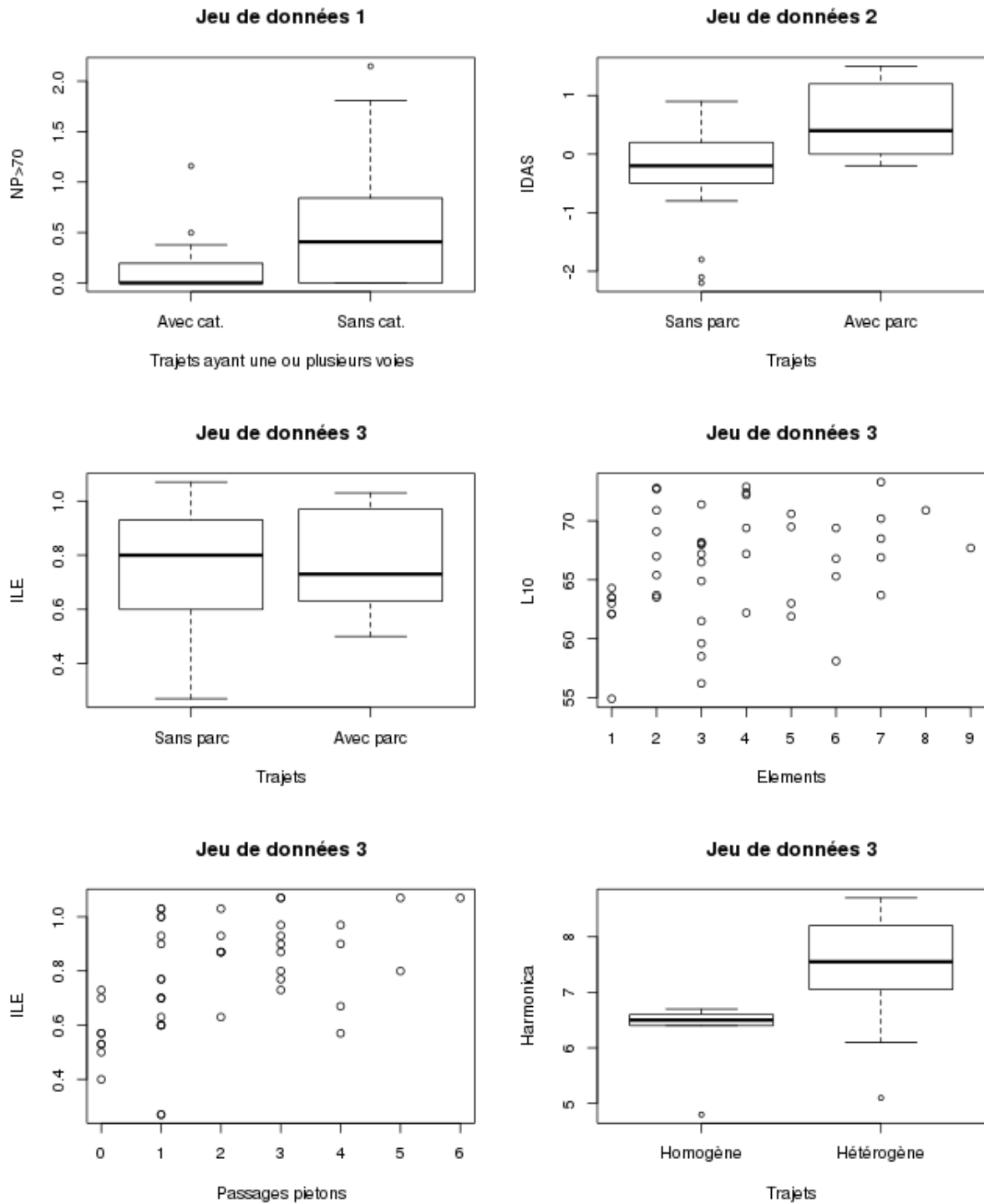


FIGURE B.4 – Représentation des données de l'analyse spatiale