



Application de l'analyse de la variance à la mise en évidence de l'influence des conditions météorologiques sur la propagation acoustique

F. Junker^a, M. Alayrac^b et M. Soulier^b

^aEDF R&D, 1 Avenue du General de Gaulle, 92141 Clamart Cedex, France

^bEDF-DPIH-DTG, 21 Avenue de l'Europe, 38040 Grenoble Cedex 9, France

fabrice.junker@edf.fr

Les conditions météorologiques ont une influence déterminante sur l'impact acoustique des sources de bruit dans l'environnement. Ainsi, dès que la source de bruit est stable, la variabilité observée dépend essentiellement des effets météorologiques : de la vitesse et la direction du vent le plus fréquemment et des effets thermiques dans certaines situations particulières plus rares. Les pratiques normatives actuelles s'appuient sur une grille d'analyse nommée $U_i T_i$. L'expérience montre que, très souvent, la grille $U_i T_i$ ne permet pas de caractériser les conditions météorologiques de manière satisfaisante, principalement car les conditions d'application de cette grille sont assez éloignées des conditions opérationnelles. Pour pallier cela, on peut étudier de manière statistique la relation entre les observables acoustiques (la plupart du temps des niveaux sonores) et les facteurs d'influence (vent, température). La mise en œuvre de méthodes d'analyse statistiques peut s'avérer fastidieuse pour une personne non familière du domaine. L'objectif de cet article est d'illustrer l'intérêt d'utiliser une analyse de la variance non paramétrique dite de « Kruskal-Wallis » pour repérer les conditions météorologiques plus ou moins favorables à la propagation. L'exemple concret présenté porte sur l'analyse d'un jeu de données expérimentales et met en évidence l'aspect opérationnel de la méthode.

1 Introduction

L'impact acoustique d'une source dans l'environnement est fortement dépendant des caractéristiques du milieu de propagation. Parmi ces dernières, l'influence des conditions météorologiques est considérée comme déterminante dès que la source et le récepteur sont éloignés d'une centaine de mètres.

Dans un contexte opérationnel, il est souvent difficile de caractériser les conditions météorologiques avec une représentativité spatiale adaptée au champ d'étude, soit parce qu'il existe souvent des effets locaux, soit parce que l'on ne peut se procurer que des données météorologiques mesurées sur un site assez éloigné (*ex* : station synoptique de Météo-France).

Les méthodes normatives en vigueur en France proposent d'utiliser la grille $U_i T_i$ qui est une classification des données météorologiques basée sur des travaux relativement anciens [1]. Les limites de cette méthode de classification sont bien connues dans le domaine de l'acoustique environnementale et la nécessité de redéfinir cette grille a été identifiée depuis plusieurs années dans les groupes de travail en normalisation [2].

La méthodologie proposée ici a été établie dans le but de proposer une alternative à l'utilisation de la classification $U_i T_i$. L'objectif est d'associer à des conditions météorologiques observées sur un site donné, une statistique sur des niveaux sonores.

2 Principe

Faire le lien entre des données météorologiques mesurées en un point précis de l'espace et des mesures acoustiques réparties dans un domaine étendu peut s'avérer difficile. Pourtant, disposer de mesures météorologiques en un point de l'espace, c'est disposer d'une certaine information plus ou moins corrélée aux changements des conditions de propagation observées sur le site étudié. Dans ce sens, les mesures météorologiques peuvent être traitées d'un point de vue statistique comme un des facteurs d'influence sur les observables que sont les mesures acoustiques.

Il est ainsi envisageable de classer les niveaux sonores observés en différents groupes, en fonction des mêmes facteurs que ceux de la grille $U_i T_i$: la direction et la vitesse du vent, la température de l'air, la période de la journée (jour, nuit, coucher et lever de soleil), le rayonnement solaire et l'humidité du sol (pluviométrie des dernières 24h et le type de sol).

La méthode exposée ci-dessous se limite à étudier l'effet des facteurs les plus influents, à savoir : la vitesse du vent et sa direction ("*facteur vent*"). Mais elle est généralisable à d'autres facteurs d'influence qui seraient identifiés comme plus pertinents dans une situation spécifique.

Dans tous les cas étudiés, on partira du principe que les données disponibles explorent une gamme suffisamment large de vitesse et de direction de vent (typiquement des vitesses de vent allant de "faibles" jusqu'à 3 à 5 m/s et des secteurs de direction de vent différents). Le nombre de données disponibles doit aussi être suffisant. L'ordre de grandeur de 1000 échantillons peut-être retenu. Si l'on considère des échantillons moyennés sur 1 minute, ceci représente un peu moins de 17 heures d'enregistrement, ce qui est largement accessible avec les matériels disponibles aujourd'hui.

Le jeu de données utilisé à titre d'exemple dans ce document provient de campagnes de mesure dans l'environnement de sites industriels. Il est présenté au §3.

Une fois le facteur d'influence défini (cf. §4), on lui définit des plages de valeurs appelées *niveaux de facteurs*. On réalise ensuite un *test de Kruskal-Wallis* (cf. §5). L'effet du facteur est alors confirmé ou infirmé par le test. Ensuite, les groupes de données acoustiques correspondant à chaque niveau du *facteur vent* sont testés afin de savoir s'ils sont significativement différents (cf. §6). Le cas échéant, ils sont regroupés pour aboutir à un nombre de groupes significativement différents inférieur ou égal au nombre de niveaux de facteur.

La méthodologie a été pensée dans un contexte opérationnel, c'est à dire qu'elle peut être appliquée par un opérateur non familier des analyses statistiques. Pratiquement, l'ensemble des traitements sont réalisables dans le logiciel libre *R*, en quelques lignes d'instruction. Les fonctions correspondantes à chaque traitement sont citées et sont documentées sur le site du logiciel *R* [3].

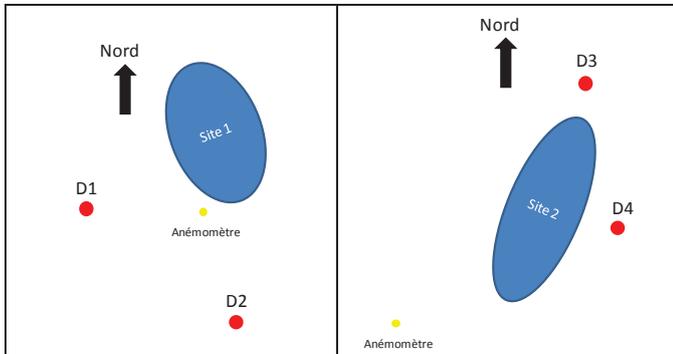
3 Jeu de données expérimentales

Les données expérimentales utilisées pour illustrer la démarche proviennent de mesures en des points situés à plusieurs centaines de mètres de sites industriels (cf. Tableau 1 et Tableau 2). Plusieurs nuits de mesure de $L_{A50,1\text{minute}}$ ont été enregistrées et couplées à des mesures de vitesse et de direction de vent faites par un anémomètre ultrasonique placé à une hauteur de 3m et lui-même localisé à distance importante (de plusieurs centaines de mètres à un kilomètre) de l'axe source-récepteur.

Tableau 1: Jeu de données expérimentales

Données	Nombre d'échantillons 1 minute	Distance au site industriel
D1 (Site 1)	2700	1200m
D2 (Site 1)	7020	1500m
D3 (Site 2)	2160	300m
D4 (Site 2)	2160	200m

Tableau 2: Orientation des points de mesure.



Les données D1 sont présentées sur la Figure 1.

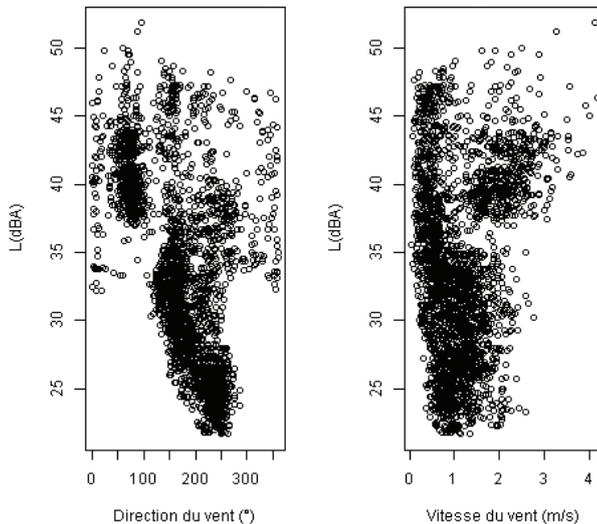


Figure 1 : Jeu de données expérimentales D1.

4 Définition du facteur d'influence

Le facteur d'influence choisi est composé de l'information de vitesse et de direction du vent. On élimine dans un premier temps les échantillons dont la vitesse de vent est supérieure à 5 m/s, conformément aux normes de mesurages en vigueur [2]. On introduit ensuite trois classes de vitesse de vent (faible, moyen, fort) et quatre classes de direction de vent correspondant chacune à un secteur de direction (A, B, C, D).

Les bornes de ces classes de vitesses et de direction sont déterminées à partir d'un algorithme des *k-means* ou des *centres mobiles* afin de s'ajuster aux données disponibles. Le principe de cet algorithme est le suivant : on définit un nombre « K » de points égal au nombre de partitions souhaitées (3 pour les vitesses de vent et 4 pour les directions). Les échantillons sont ensuite regroupés à la valeur initiale sélectionnée par l'algorithme qui leur est le plus proche. Pour chacun des groupes de valeurs formés, un nouveau centre de gravité est calculé à partir de leurs

échantillons respectifs. Toutes les valeurs sont ensuite regroupées à l'un des nouveaux centres de la même manière que précédemment (d'où l'appellation méthode des centres mobiles). L'algorithme est répété jusqu'à convergence qui se définit par deux classifications consécutives identiques.

L'intérêt d'un tel algorithme est qu'il ne présuppose pas un arrangement spécifique des données. Pour peu que les données disponibles couvrent des plages de valeurs assez larges (c'est souvent l'objectif lorsque l'on souhaite pouvoir caractériser une situation donnée), l'algorithme fournit en sortie trois ensembles de données de vitesse de vent et quatre secteurs de direction bien répartis. Si par contre les données sont concentrées sur certaines directions, l'algorithme pourra procéder à un découpage plus fin des directions.

Il est aussi possible d'indiquer à l'algorithme recherchant les classes de direction de vent, un axe privilégié sur lequel il recherchera à ajuster l'un des secteurs (pour, par exemple, faire ressortir une direction favorable à la propagation).

À titre d'exemple, on illustre sur la Figure 2 le résultat obtenu sur les données D1 avec une direction privilégiée à 70° (fonction *kmeans* du package *stats* de R). On observe bien 3 gammes de vitesses de vent et quatre secteurs de directions dont 1 (bleu) qui est centré sur 70° la direction privilégiée indiquée.

Les valeurs numériques correspondant au découpage sont les suivantes :

Directions :

A : 115° -> 199°	Moyenne angle A: 163°
B : 198° -> 291°	Moyenne angle B: 233°
C : 294° -> 28°	Moyenne angle C: 349°
D : 31° -> 115°	Moyenne angle D: 72°

Vitesses :

Faible : 0,044 m/s-> 0,951 m/s
Moyen : 0,953 m/s-> 1,822 m/s
Fort : 1,825 m/s-> 4,171 m/s

La combinaison de ces classes de vitesse et de direction définissent les niveaux du facteur vent (Nul, 1A,1B,...,2D), conformément au Tableau 3.

Tableau 3: Niveaux du facteur vent

Vitesse de vent	Direction	Facteur Vent
Faible	Non définie	Nul
Moyen	A	1A
	B	1B
	C	1C
	D	1D
Fort	A	2A
	B	2B
	C	2C
	D	2D

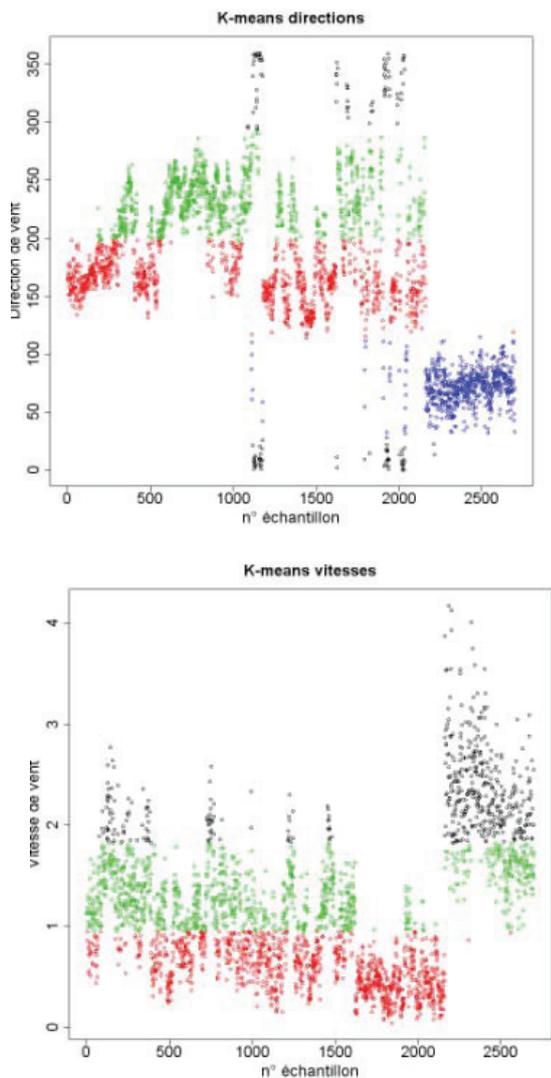


Figure 2 : Résultat de l’algorithme des k-means sur les données D1. En haut, les secteurs de directions A (rouge), B (vert), C (noir), D (bleu). En bas vent faibles (rouge), vent moyens (vert) et vent forts (noir).

Remarque : Il est normal qu’aucune direction ne soit associée aux faibles vitesses de vent puisque la direction du vent n’est pas stable pour les vents faibles.

À chaque échantillon est donc associé un des niveaux de facteur ci-dessus. Le résultat est illustré sur la Figure 3.

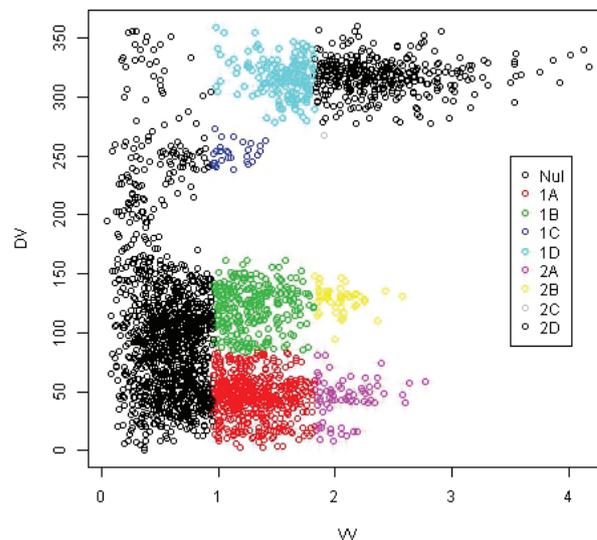


Figure 3 : Application des niveaux de facteur (Nul, 1A, 1B., 2D) aux données D1. VV : Vitesse de vent en m/s. DV : Direction du vent en °.

Le résultat peut aussi être illustré sous la forme d’un graphe à moustaches donné Figure 4.

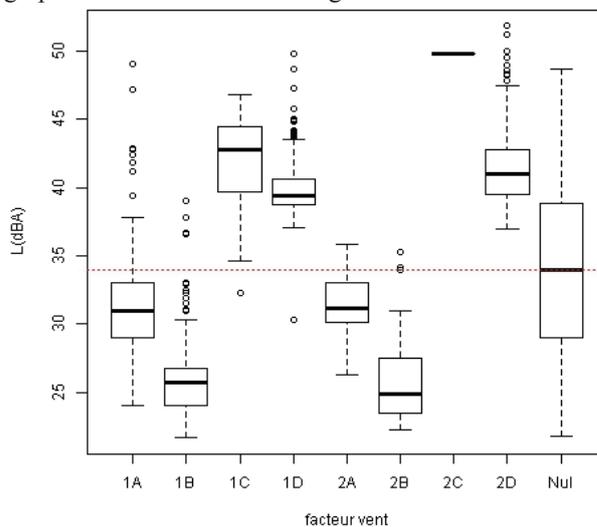


Figure 4 : Populations correspondant à chaque niveau du facteur vent (données D1). Le trait rouge correspond à la valeur moyenne de l’ensemble des données.

5 Test de Kruskal-Wallis

Le test de Kruskal-Wallis [4] permet de tester l’hypothèse « les groupes de données correspondant à chaque niveau de facteur ne sont pas significativement différents ». On rejette l’hypothèse (c’est à dire que l’on considère qu’au moins deux groupes sont différents) lorsque le test de significativité (*p-value*) est inférieur à un seuil (classiquement 5%). Ce test est la version non paramétrique de l’ANOVA classique à un facteur [5]. Comme l’ANOVA classique compare des variances intergroupes à des variances intragroupes, le test de Kruskal-Wallis est construit à partir du rang des données ce qui a pour principal avantage de ne pas requérir la normalité et l’égalité des variances des populations testées, hypothèses souvent non satisfaites pour les données que l’on souhaite analyser.

Dans le cas du jeu de données, on utilise le test *kruskal.test* du package *stats* de R on obtient une *p-value* très faible (très inférieure à 0.01) indiquant qu'au moins deux groupes de données sont significativement différents.

6 Test de comparaison multiple

Le package *pgirmess* de R propose une fonction appelée *kruskalmc* qui implémente un test de comparaison multiple basé sur le test de Kruskal-Wallis [6]. Le résultat se présente sous la forme d'un tableau qui précise si chaque paire de niveau de facteur sont significativement différentes avec un seuil à préciser (on choisit ici 5%).

On choisit de conserver la classe « Nul » séparée des autres car ce groupe de données présente une grande dispersion. On supprime aussi les classes à faible effectif, typiquement inférieur à 10 (2C ne contenait qu'un seul échantillon). On teste ensuite les différences entre les classes restantes.

Finalement, en agrégeant ces résultats, on trouve sur les données D1 les 4 groupes de données suivants :

- Groupe 1 : 1A 2A
- Groupe 2 : 1B 2B
- Groupe 3 : 1C 1D 2D
- Groupe 4 : Nul

Ce qui peut se représenter par les nuages de points colorés de la Figure 5.

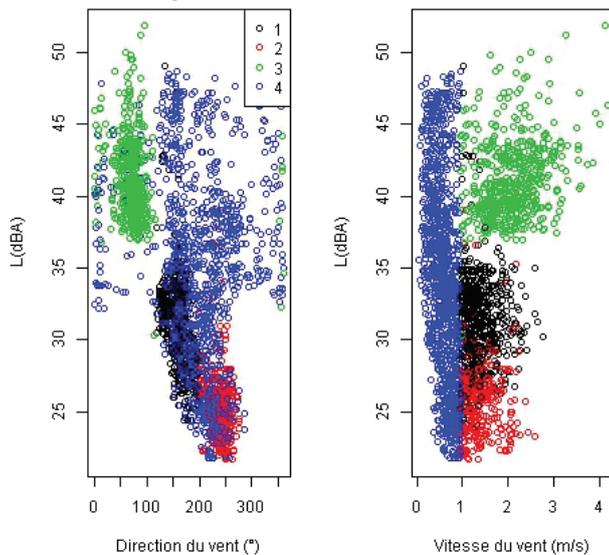


Figure 5 : Mise en évidence de 4 groupes sur les données D1.

On exprime ensuite la valeur des indicateurs statistiques sur le L_{A50} correspondant à chaque groupe.

Tableau 4 : Statistiques sur les valeurs de L_{A50} en dBA pour chacun des 4 groupes pour les données D1.

Groupe	Moyenne	Médiane	Écart-type
1	31,1	31,0	3,1
2	25,9	25,7	2,7
3	41,0	40,4	2,7
4	34,2	34,0	6,7

7 Analyses du jeu de données

Les résultats de l'analyse des données D1 montrent bien une grande dispersion des niveaux pour les vents faibles : on peut d'ailleurs utiliser ce groupe de données pour caractériser le niveau moyen observé sur le site (34 dBA) et sa dispersion (6,7 dBA d'écart-type). Les niveaux les plus forts sont rencontrés pour les directions C et D soit pour des vents peu contraires, de travers du nord et portants (70°). Les niveaux les plus faibles (26 dBA) pour les directions A (vents contraires). Pratiquement, on peut représenter les résultats obtenus avec les données D1 par un diagramme du type de la Figure 6.

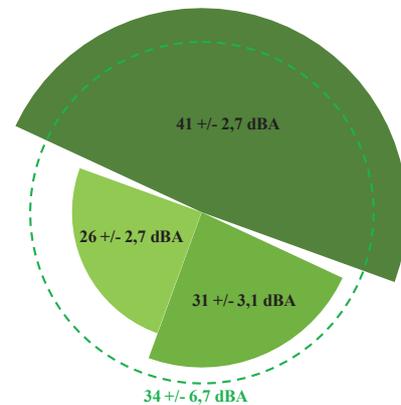


Figure 6 : Synthèse des résultats obtenus avec les données D1 sous la forme de niveaux sonores associés à leur écart-type, pour chaque secteur de direction au point de mesure.

On peut enfin calculer un niveau sonore de long-terme à partir de ces résultats si l'on dispose de statistiques sur les occurrences des données de vent correspondant à chacun des 4 groupes.

La Figure 7 illustre le résultat obtenu sur les données D2 qui est assez semblable au résultat obtenu sur les données D1 concernant l'influence des vitesses et direction de vent. Pourtant l'orientation du point de mesure D2 par rapport aux sources est un peu différente. On pourrait de ce fait s'attendre à ce que les vents portants viennent plutôt du Nord. Cet exemple montre donc que le fait de ne pas avoir d'a priori sur la classification permet de dégager des tendances qui ne seraient pas intuitivement obtenues.

La direction privilégiée a été choisie dans ce cas « au hasard » égale à 180° pour tester la sensibilité de la méthode à ce choix. Ceci a juste pour conséquence de modifier la classification qui est la suivante :

Directions pour les données D2 :

A : $225^\circ \rightarrow 337^\circ$	Moyenne angle A: 252°
B : $338^\circ \rightarrow 107^\circ$	Moyenne angle B: 62°
C : $108^\circ \rightarrow 178^\circ$	Moyenne angle C: 154°
D : $178^\circ \rightarrow 225^\circ$	Moyenne angle D: 203°

Vitesses pour les données D2:

Faible : 0,044 m/s-> 0,926 m/s

Moyen : 0,928 m/s-> 1,799 m/s

Fort : 1,802 m/s-> 4,171 m/s

Groupes de niveaux de facteur pour les données D2 :

1 : 1A 1C 1D 2A 2C 2D

2 : 1B 2B

3 : Nul

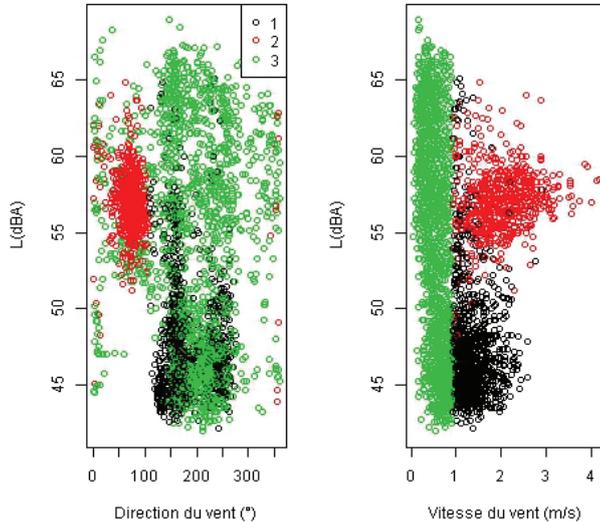


Figure 7 : Mise en évidence des groupes 1, 2 et 3 sur les données D2.

Malgré ce choix de direction privilégiée, on retrouve à nouveau des niveaux forts pour une direction semblable (62° ici pour 72° dans le cas du jeu de données D1). La méthode semble donc peu sensible au choix d'orientation privilégiée. Les niveaux (48, 57 et 54 dBA respectivement pour les groupes 1, 2 et 3) sont nettement supérieurs à cause de la plus grande visibilité des sources. La dispersion est toujours importante pour les vents faibles et couvre toute la gamme de variation observée en ce point.

Les données D3 présentent un caractère singulier. En effet, leur observation (cf. Figure 8) ne montre pas de groupe de données se détachant particulièrement ni d'effet de vitesse et de direction de vent : la dispersion est semblable sur tout le domaine. La méthodologie appliquée sur ces données D3 conduit à ne définir que 2 groupes : le groupe « Nul », intégrant toute la dispersion des mesures, ainsi que un groupe rassemblant l'ensemble des autres classes. L'écart de niveau moyen entre les deux groupes est très faible (1,5 dBA). De plus, si l'on intègre le groupe « Nul » au test de comparaison multiple (*kruskalmc*), on montre que l'écart entre le groupe « Nul » et le groupe 2 n'est pas significatif.

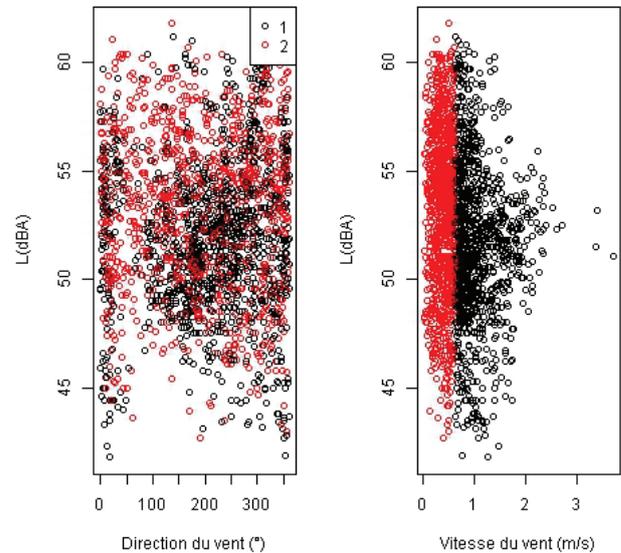


Figure 8 : Données D3. Outre le groupe « Nul » (Groupe 2) volontairement isolé des autres données, tous les autres niveaux sont regroupés (Groupe 1).

La démarche ne produit donc apparemment pas de découpage artificiel des données et recherche bien des groupes de données distincts les uns des autres.

Les données D4 présentent notamment un groupe de valeurs correspondant à des vents forts du nord (Groupe 3) (cf. Figure 9). Là encore, ce résultat n'est pas forcément intuitif, vue l'orientation de ce point par rapport au site.

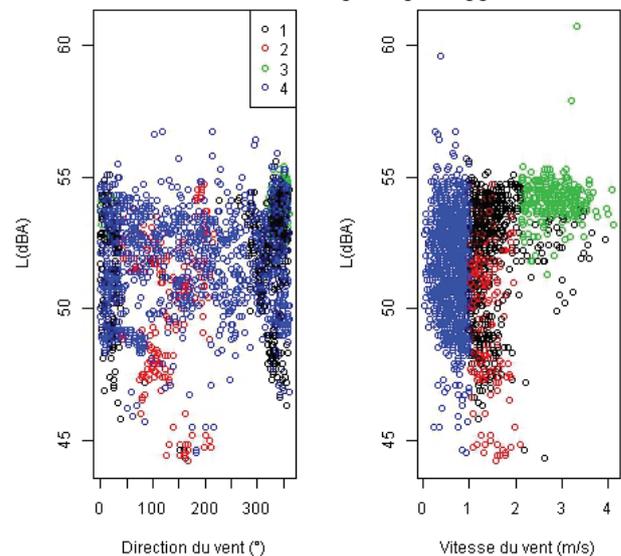


Figure 9 : Les groupes obtenus sur les données D4.

Le graphe à moustaches donné Figure 10 illustre le fait que le groupe 3 est considéré comme distinct grâce à sa faible variance.

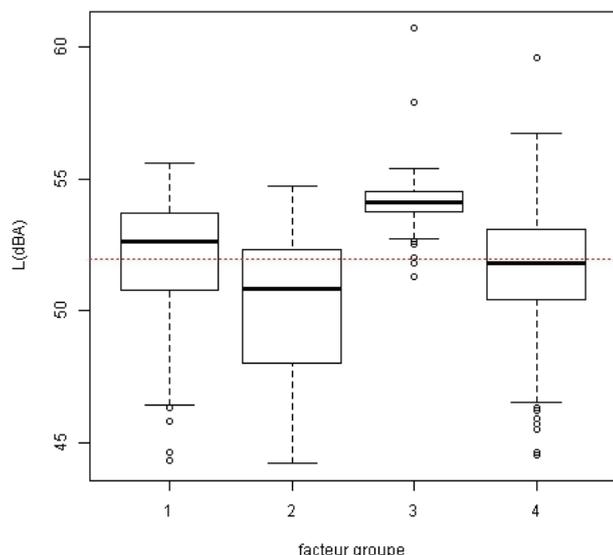


Figure 10 : Statistiques des 4 groupes obtenus sur les données D4.

8 Conclusion

Une méthode de traitement statistique basée sur une classification des conditions de vent par la méthode *k-means* et une analyse de type *Kruskal-Wallis* a été présentée. Elle permet de définir, sur la base de méthodes statistiques rigoureuses, des groupes de données acoustiques significativement différents les uns des autres et caractéristiques de certaines conditions de vitesse et de direction de vent. Dans certains cas, des directions de vent privilégiées sont mises en évidence alors qu'elles n'étaient pas intuitivement anticipées comme les plus favorables à la propagation.

Il est intéressant d'autre part de noter la constance de l'observation faite pour les faibles vitesses de vent : la moyenne et la dispersion observée pour ce type de vent est un bon indicateur de la moyenne et la dispersion observée sur l'intégralité des mesures.

L'application pratique de cette méthode est envisageable, notamment pour déterminer des indicateurs de niveaux acoustiques en conditions favorables et homogènes nécessaires pour le calcul des indicateurs long-terme. Il suffit de disposer d'un nombre de données d'effectif suffisant couvrant si possible une large gamme de vitesse et de directions de vent.

Les traitements nécessaires sont facilement réalisables avec le logiciel libre *R* : la routine de traitement correspondant est disponible sur simple demande auprès des auteurs.

Cette méthodologie sera proposée dans le cadre du groupe de travail sur la partie météorologie de la norme NF S 31.110 [2].

Références

- [1] Zouboff V., Brunet Y., Bérengier M., Séchet E., *A qualitative approach of atmospheric effects on long range sound propagation*, 6th International Symposium on Long Range Sound Propagation, Ottawa, Canada, pp 251-269 (1994).
- [2] Norme NF S 31-110 : Caractérisation et mesurage des bruits de l'environnement (Novembre 2005).
- [3] <http://www.r-project.org/>
- [4] Myles Hollander and Douglas A. Wolfe, *Nonparametric Statistical Methods*, New York: John Wiley & Sons, Pp 115–120 (1973).
- [5] Yadolah Dodge, *Statistiques, Dictionnaire Encyclopédique*, Springer-Verlag France, Paris, 2004, ISBN-2-287-21325-2
- [6] Siegel and Castellan, *Non parametric statistics for the behavioural sciences*, MacGraw Hill Int., New York. pp 213-214 (1988).