



Qualité de l'air



Février 2006

Ouelair est-il ?

Plan d'échantillonnage : Estimation d'une moyenne annuelle à l'aide de mesures discontinues



Association pour la Surveillance de la Qualité de l'Air de la Région de l'Etang de Berre et de l'Ouest des Bouches-du-Rhône

Route de la Vierge - 13 500 Martigues - Tel. 04 42 13 01 20 - Fax. 04 42 13 01 29

Site internet: www.airfobep.org - e-mail : airfobep@airfobep.org

Serveur vocal 04 42 49 35 35 (selon tarification téléphonique en vigueur)



Résumé

Pour estimer une moyenne annuelle à partir de mesures discontinues, on définit un plan d'échantillonnage à partir de simulations réalisées sur le NO₂ et les PM10. Les résultats relatifs à la taille des échantillons et aux périodes de prélèvement nous amènent à fixer à quatre mois la durée annuelle de mesures discontinues, mais ne nous permet pas de définir une période optimale de prélèvement. En effet, pour une période fixée, les erreurs d'estimation de la moyenne annuelle varient fortement.

Cela nous conduit à proposer une méthode de correction permettant de s'affranchir du choix des périodes de mesures. Un groupe de stations fixes est utilisé pour corriger l'estimation de la moyenne annuelle. Il permet d'évaluer l'erreur commise sur le site échantillonné. Cette évaluation faite, un coefficient de correction est appliqué à l'estimation de la moyenne annuelle et atténue la sur-estimation ou sous-estimation potentielle. Cette correction restreint les erreurs d'estimation de la moyenne annuelle du NO₂ et des PM10 dans la gamme des +/- 10 % en considérant quatre mois de mesures discontinues par l'année civile.

Ci-dessous, les estimations de la moyenne annuelle 2005 de NO₂ aux huit stations AIRFOBEP arrêtées en 2006. Ces estimations sont calculées à partir de simulations d'échantillonnages et corrigées à partir du groupe de stations fixes de référence. Pour chaque stations, sur les neuf simulations effectuées (toutes différentes de part leurs périodes d'échantillonnage) nous présentons les cas générant les erreurs les plus importantes.

Moyennes annuelles 2005			
Stations	Réelle	Corrigée	Erreur
BETG	24	26	2.2
FSCB	15	16	0.9
ISTR	21	23	1.9
MNDM	15	17	1.6
PDBE	37	34	-2.5
SMDC	20	21	1.0
SSLP	13	13	-0.6
VTRL	21	20	-0.9

Le cas du NO₂ : Comparaison des moyennes annuelles réelles avec les moyennes estimées par l'échantillonnage et corrigées par un groupe de stations fixes (en µg/m³).

L'utilisation de cette méthode de correction se justifie du fait de la relative homogénéité des conditions météorologiques dans la zone d'étude d'AIRFOBEP et de la représentativité du groupe de stations fixes de référence utilisé pour établir le coefficient de correction.

Le plan d'échantillonnage ainsi défini peut s'appliquer en des lieux où les niveaux de pollution tolèrent une incertitude de 10 % à laquelle s'ajoute celle relative au moyen de mesure employé.

SOMMAIRE

1	STRATEGIE D'ECHANTILLONNAGE.....	4
1.1	TAILLE DE L'ECHANTILLON	4
1.2	PERIODES D'ECHANTILLONNAGE.....	6
2	CORRECTION A PARTIR D'UN GROUPE DE STATIONS FIXES	8
2.1	METHODOLOGIE.....	8
2.2	RESULTATS 2000-2004	10
2.3	RESULTATS 2005.....	13
3	APPLICATION AU PM10.....	15
4	SYNTHESE ET APPLICATION	16
4.1	UTILISATION DE LA MACRO	17
4.2	UTILISATION DE LA FEUILLE DE CALCUL	17
5	ANNEXE 1 : ECHANTILLONNAGE STRATIFIE.....	19
5.1	LA METHODE	19
5.2	ANALYSE DES DONNEES	20
5.2.1	<i>La vitesse du vent</i>	21
5.2.2	<i>La température</i>	22
5.2.3	<i>L'heure du jour</i>	23
5.2.4	<i>Le jour de la semaine</i>	24
5.3	DEFINITION DES STRATES DE CONCENTRATIONS.....	25
5.4	L'ECHANTILLONNAGE	26
5.5	ESTIMATION DES MOYENNES ANNUELLES	26
5.6	BILAN DES SIMULATIONS.....	28
6	ANNEXE 2 : COMPARAISON DES METHODES DE CORRECTION	29
6.1	ECHANTILLONNAGE STRATIFIE.....	29
6.2	CORRECTION A PARTIR DE STATIONS DE REFERENCES	29
6.3	COMPARAISON	30

Introduction

L'objectif de cette étude est de définir une stratégie de mesures discontinues permettant d'estimer au mieux la concentration moyenne annuelle d'un polluant. Dans une première partie nous traitons de la taille des échantillons à prélever et des périodes durant lesquelles sont effectuées les campagnes de mesures discontinues. Nous fixons cependant les conditions suivantes : les échantillons sont également répartis sur quatre trimestres de l'année civile, et chacun d'une durée fixe de 2 semaines (durée optimale de mesures pour les tubes passifs) et 1 mois (durée minimale de mesures pour les stations fixes). Nous proposons ensuite une méthode de correction sur l'estimation de la moyenne annuelle. Le plan d'échantillonnage et la méthode de correction seront testés pour le cas de la mesure du NO₂ et des PM10. Pour finir, le dernier paragraphe indique comment utiliser la macro Excel permettant de réaliser cette correction.

Remarque : les résultats présentés dans cette étude sont issus de simulations d'échantillonnage réalisées sur l'ensemble des stations du réseau AIRFOBEP. Ce point de vue empirique permet d'approcher au plus près les erreurs d'estimation de la concentration moyenne annuelle.

1 Stratégie d'échantillonnage

1.1 Taille de l'échantillon

Afin d'estimer la taille nécessaire de nos échantillons, considérons le cas du NO₂ avec un échantillonnage également réparti sur les quatre trimestres de l'année civile. Dans chaque trimestre, nous prélevons par groupe de deux semaines plusieurs échantillons (2 semaines par trimestre, puis 4 semaines par trimestre, etc...). Nous constituons donc une série d'échantillonnage annuel différent par leur durée. Pour chacun d'entre eux, nous estimons la moyenne annuelle. Nous pouvons donc observer la qualité d'estimation en fonction de la taille des échantillons recueillis. Pour chaque taille d'échantillon nous avons simulé 1400 échantillonnages aléatoires répartis sur 5 ans (de 2000 à 2004) et sur les 14 stations de mesure.

Regardons d'abord les moyennes des erreurs en valeur absolue des estimations (figure 1). (Celles-ci sont calculées en pourcentage par rapport à la moyenne annuelle réelle de même que l'ensemble des erreurs en pourcentage présentées dans ce rapport). Nous pouvons remarquer une amélioration significative entre les tailles de 12 à 16 semaines, alors qu'entre 16 et 20 l'écart des erreurs est faible. Le passage à la taille suivante de 20 à 24 semaines apporte également une amélioration significative. Nous retiendrons donc les tailles d'échantillonnages suivantes : 16 semaines et 24 semaines par année civile.

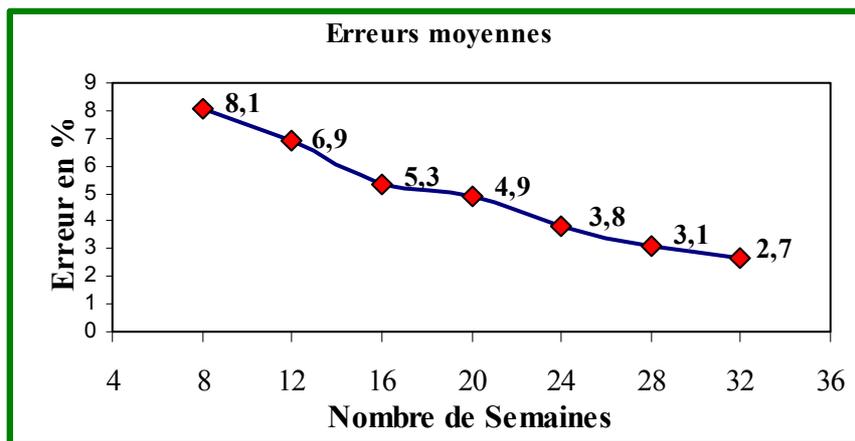


Figure 1 : La moyenne des erreurs en valeurs absolues (en pourcentage) selon la taille de l'échantillonnage (calculée sur l'ensemble des stations du réseau de 2000 à 2004).

Le « boxplot » qui suit (figure 2) permet de visualiser l'ensemble des erreurs commises pour chacune des tailles d'échantillonnage. Un échantillonnage de 16 semaines génère des erreurs entre -20 et 20 %, alors qu'un échantillonnage de 24 semaines génèrent des erreurs comprises entre -15 et 15 % par rapport à la moyenne annuelle réelle. Les quelques valeurs hors de ces intervalles seraient relativement rares compte tenu du nombre de simulations effectuées.

Erreurs d'estimation de la moyenne annuelle selon la taille de l'échantillonnage

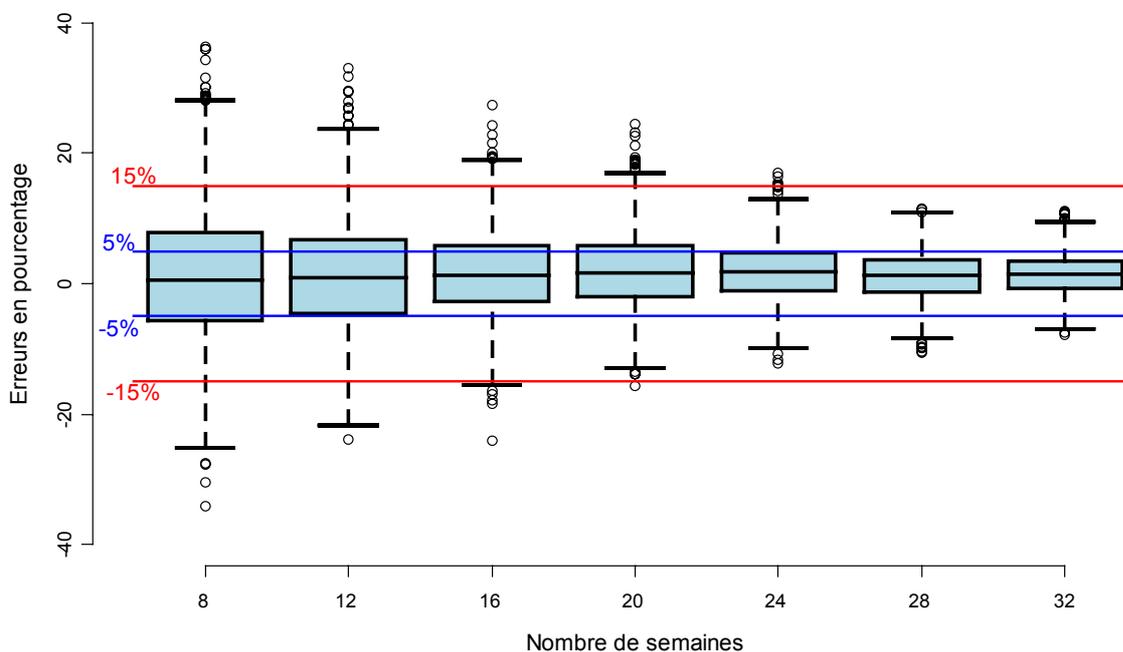


Figure 2 : Les « Boxplot » des erreurs (en pourcentage) selon la taille de l'échantillonnage. (1400 simulations pour chaque taille d'échantillon sur l'ensemble des stations du réseau de 2000 à 2004)

1.2 Périodes d'échantillonnage

Les périodes d'échantillonnage ont-elles une influence sur la qualité d'estimation de la moyenne annuelle ? Pour répondre à cette question nous avons simulé 9 échantillonnages de 16 semaines (1 mois consécutif par trimestre) différents par les périodes de mesures (tableau1).

ech1	ech2	ech3	ech4	ech5	ech6	ech7	ech8	ech9
1 ^{er}	8	15	22	29	5	12	19	26
janvier	janvier	janvier	janvier	janvier	février	février	février	février
1 ^{er}	8	15	22	29	6	13	20	27
avril	avril	avril	avril	avril	mai	mai	mai	mai
1 ^{er}	8	15	22	29	5	12	19	26
juillet	juillet	juillet	juillet	juillet	août	août	août	août
1 ^{er}	8	15	22	29	5	12	19	26
octobre	octobre	octobre	octobre	octobre	novembre	novembre	novembre	novembre

Tableau 1 : Date de début des 4 périodes de 1 mois constituant les échantillonnages simulés

Les marges d'erreurs varient selon la période échantillonnée, mais aussi selon l'année échantillonnée. Tandis que l'ensemble des erreurs sur 5 années est centré sur zéro (figure 3), les « boxplot » annuels montrent la variabilité des erreurs d'un même échantillonnage d'une année sur l'autre (figure 4).

Erreurs d'estimation de la moyenne annuelle selon la période d'échantillonnage

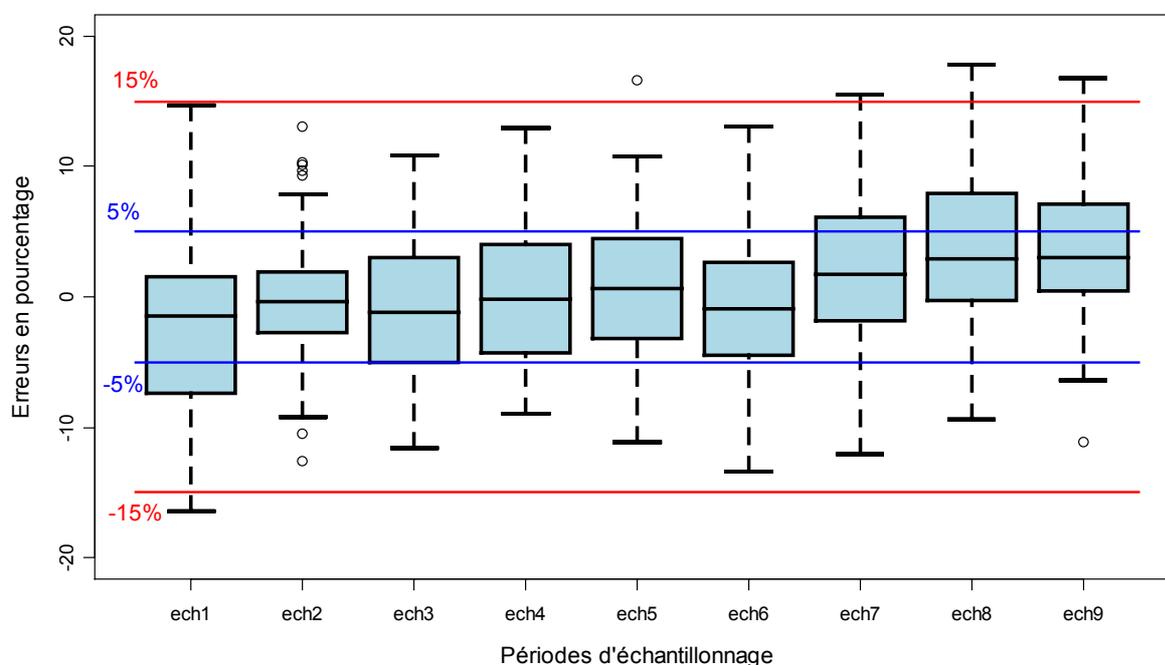


Figure 3 : Les « Boxplot » des erreurs (en pourcentage) selon la période d'échantillonnage. Se référer au tableau 1 pour avoir le détail sur les périodes échantillonnées. (calculées sur l'ensemble des stations du réseau de 2000 à 2004)

Les périodes d'échantillonnage ont donc une influence sur la qualité d'estimation de la moyenne annuelle. Cependant cette influence varie en fonction des années, il n'est donc pas possible de sélectionner une période qui serait systématiquement meilleure que les autres. Nous pouvons faire deux remarques intéressantes :

- pour une même année, un décalage d'une semaine dans l'échantillonnage peut changer significativement la qualité d'estimation de la moyenne annuelle.
- d'une année à l'autre un même échantillonnage n'a pas la même qualité d'estimation de la moyenne annuelle.

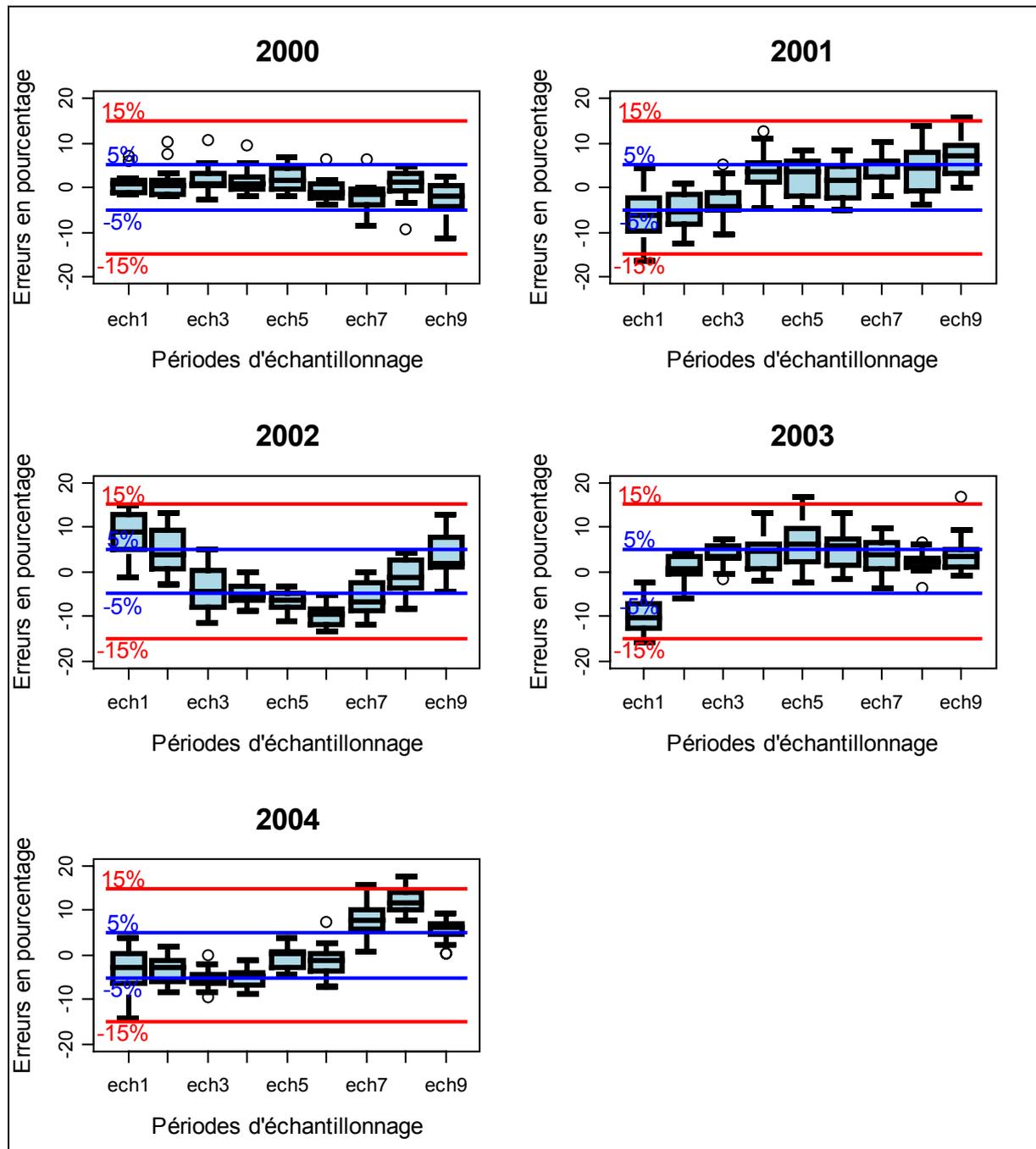


Figure 4 : Les « Boxplot » des erreurs (en pourcentage) par année selon la période d'échantillonnage. Se référer au tableau 1 pour avoir le détail sur les périodes échantillonnées (calculés sur l'ensemble des stations du réseau).

2 Correction à partir d'un groupe de stations fixes

Le but de cette correction est de limiter les grosses erreurs d'estimation dans le cas où l'échantillon recueilli ne serait pas représentatif de l'année entière, et de réduire la variabilité des erreurs dans l'estimation de la moyenne annuelle. Le principe de cette méthode est de corriger la moyenne estimée sur la campagne de mesures discontinues à partir d'un groupe de stations fixes de référence. Pour ce faire, nous effectuons le même échantillonnage que celui de la campagne de mesures sur ces stations de référence. Nous pouvons alors calculer l'erreur que l'on aurait fait en estimant la moyenne annuelle à l'aide d'un tel échantillonnage sur ce groupe de stations. Cette erreur d'estimation nous permet de corriger la moyenne estimée lors de la campagne de mesures discontinues.

2.1 Méthodologie

Dans le cas d'un échantillonnage dans un cœur de ville disposant d'un appareil fixe, la correction s'effectue uniquement à l'aide de cette station fixe (figure 5). Dans le cas d'un échantillonnage dans un lieu où il n'y a pas de station fixe à proximité, la correction s'effectue à l'aide de toutes les stations fixes du réseau AIRFOBEP (figure 6). Dans les deux cas, elle se déroule de la façon suivante :

- Sur le site où l'on effectue la campagne de mesures discontinues :
 - On recueille les échantillons
 - On calcule leur moyenne brute
- Sur les ou la station fixe de référence :
 - On simule cet échantillonnage, c'est à dire, on rassemble les valeurs des stations fixes correspondantes aux jours de la campagne de mesures discontinues
 - On estime la moyenne annuelle à partir de cet échantillonnage simulé pour chacune des stations fixes
 - On calcule la moyenne annuelle réelle pour chacune des stations fixes
 - On calcule l'erreur en pourcentage (notée « E ») entre la moyenne estimée par l'échantillonnage simulé (notée « ME ») et la moyenne réelle (notée « MR ») pour chacune des stations fixes :

$$E = \frac{(ME - MR)}{MR}$$

- On calcule la moyenne de ces erreurs
- On utilise cette erreur moyenne calculée sur les stations fixes pour corriger la moyenne obtenue lors de la campagne de mesure discontinue. Si on note cette erreur moyenne « EM » et la moyenne brute obtenue lors de la campagne de mesures discontinues « MB », alors la moyenne corrigée « MC » se calcule de la façon suivante :

$$MC = \frac{MB}{(EM + 1)}$$

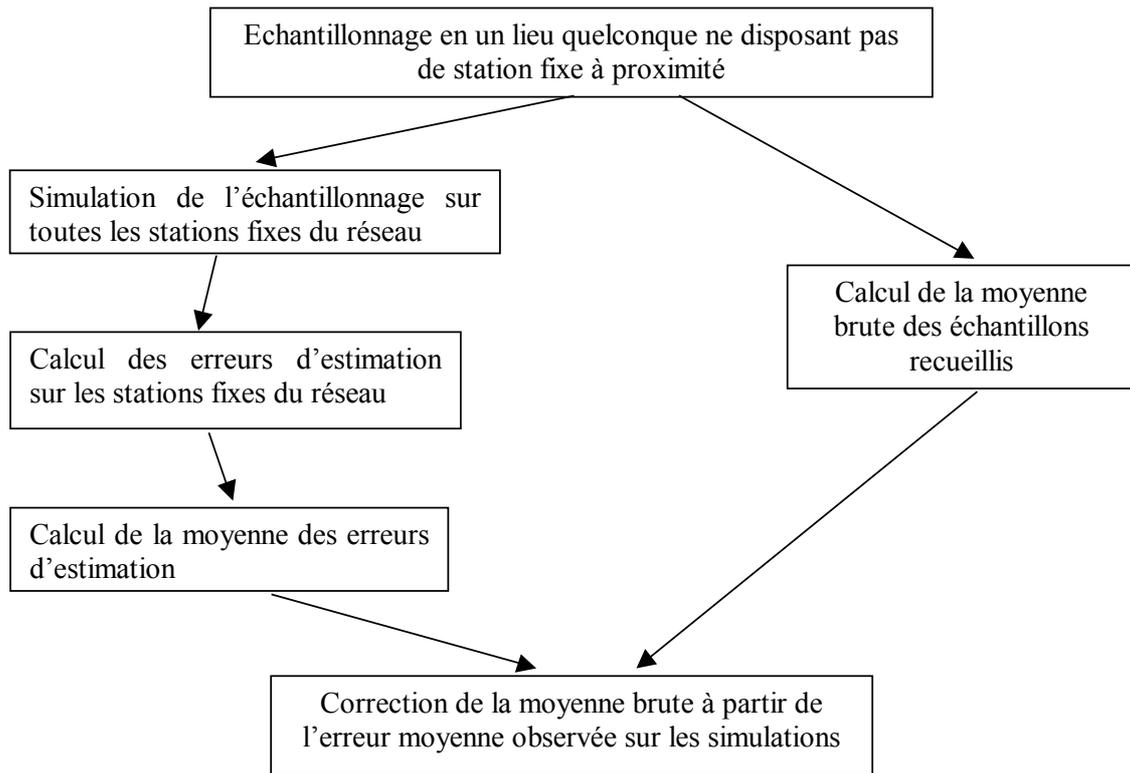


Figure 5 : Cas d'un échantillonnage dans un lieu quelconque ne disposant pas d'appareil fixe à proximité.

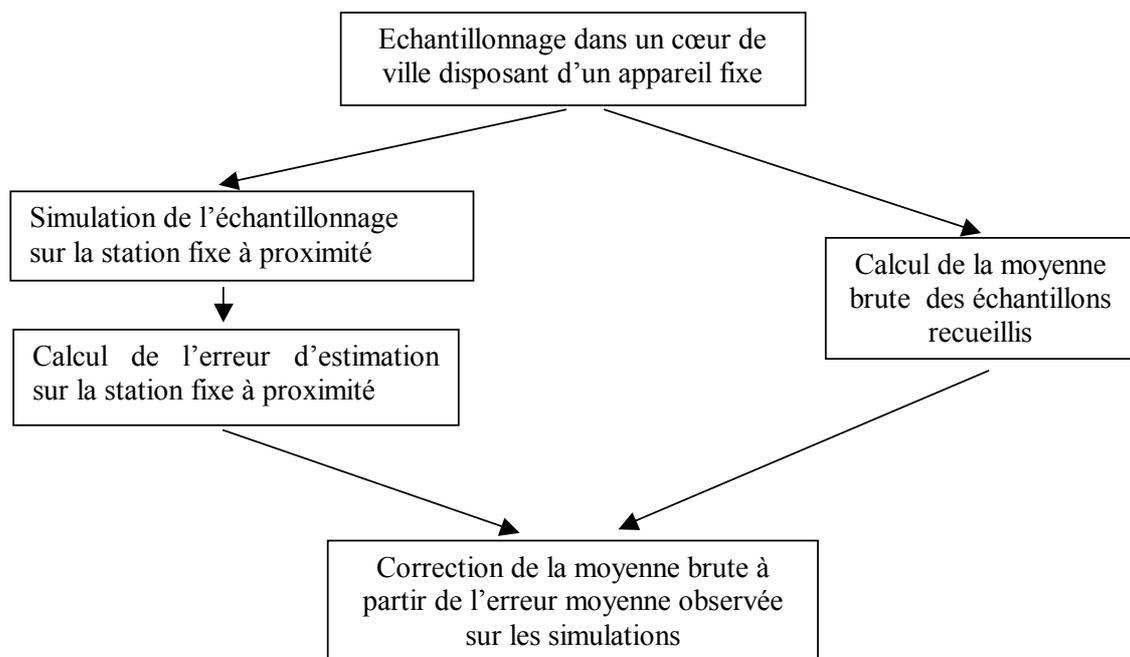


Figure 6: Cas d'un échantillonnage dans un cœur de ville disposant d'un appareil fixe à proximité.

2.2 Résultats 2000-2004

Observons sur la figure 7, les erreurs commises sur les stations de Berre, Fos-sur-Mer, Istres, Martigues Notre Dame, Port-de-Bouc, St Martin-de-Crau, Sausset-les-Pins et Vitrolles, de 2000 à 2004 et corrigées par les autres stations fixes : Arles, Marignane ville, Martigues Île, Miramas, Rognac-les-brets et Salon. Afin de se situer au plus proche de la situation réelle, les stations utilisées pour corrigées sont celles qui continueront d'exister en 2006. Les échantillonnages simulés sont ceux décrits dans le tableau 1.(4 fois 1 mois réparti sur chaque trimestre).

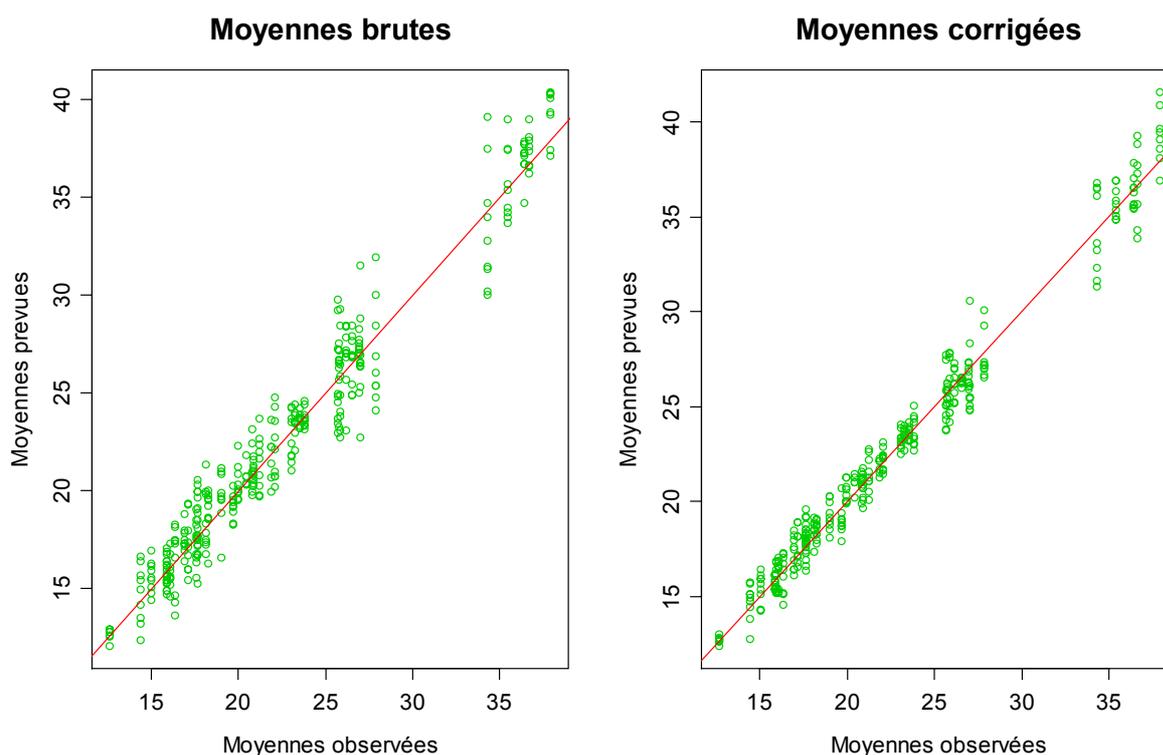


Figure 7 : Graphiques des moyennes estimées brutes et corrigées en fonction des moyennes observées (en $\mu\text{g}/\text{m}^3$), sur l'ensemble des stations de Berre, Fos-sur-Mer, Istres, Martigues Notre Dame, Port-de-Bouc, St Martin-de-Crau, Sausset-les-Pins et Vitrolles, de 2000 à 2004.

La correction effectuée diminue les erreurs dans toute la gamme des valeurs de moyenne annuelle observée (figure 7). La moyenne des erreurs en valeur absolue des estimations corrigées diminue de 1.7 points par rapport à la moyenne des erreurs brutes (tableau 2).

	Moyenne Brute	Moyenne corrigée
Erreurs en %	4.9	3.2

Tableau 2 : Moyenne des erreurs en valeur absolue.

Cette correction permet également de rendre homogène la variabilité selon les périodes d'échantillonnage. La figure 8 montre son effet sur chacune des périodes échantillonnées. Ces périodes sont comparables dans leur gamme d'erreurs sur les moyennes corrigées.

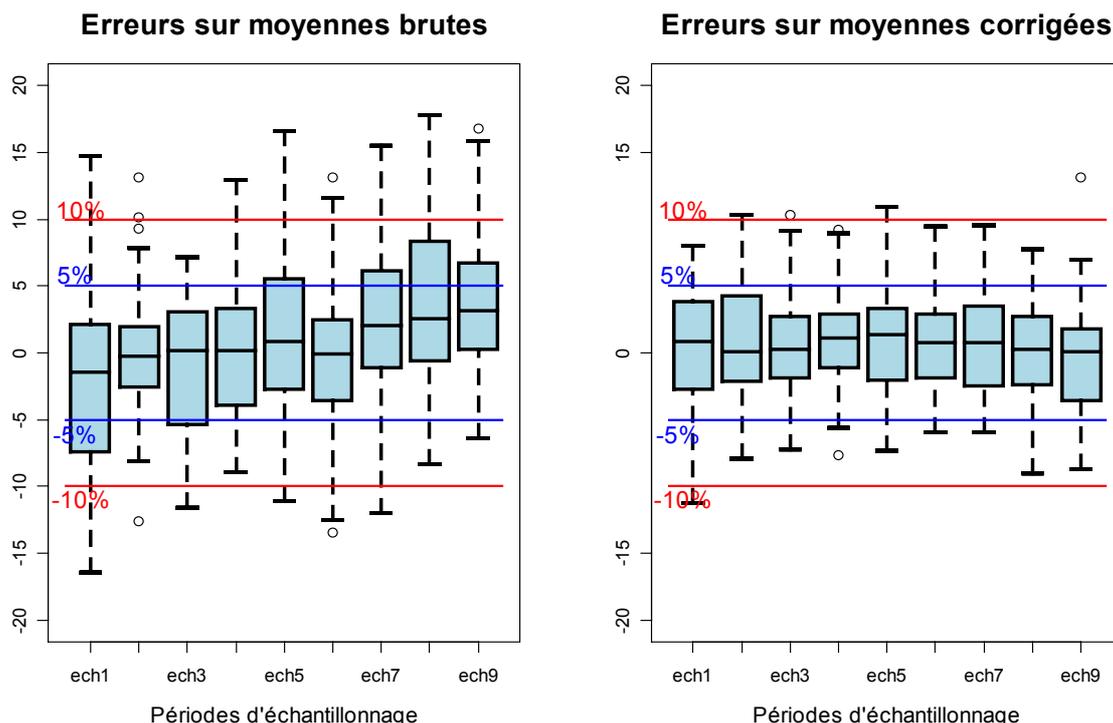


Figure 8 : Comparaison des erreurs brutes et corrigées (en pourcentage) par période d'échantillonnage, sur l'ensemble des stations de Berre, Fos-sur-Mer, Istres, Martigues Notre Dame, Port-de-Bouc, St Martin-de-Crau, Sausset-les-Pins et Vitrolles, de 2000 à 2004.

Le groupe de stations fixes restantes en 2006 corrige donc relativement bien les stations qui vont être arrêtées. On peut remarquer tout de même que les stations de Saint-Martin-de-Crau et de Sausset-les-Pins sont les seules ayant des erreurs sur moyenne corrigées parfois supérieures à 10 % et inférieure à -10 % (figure 9).

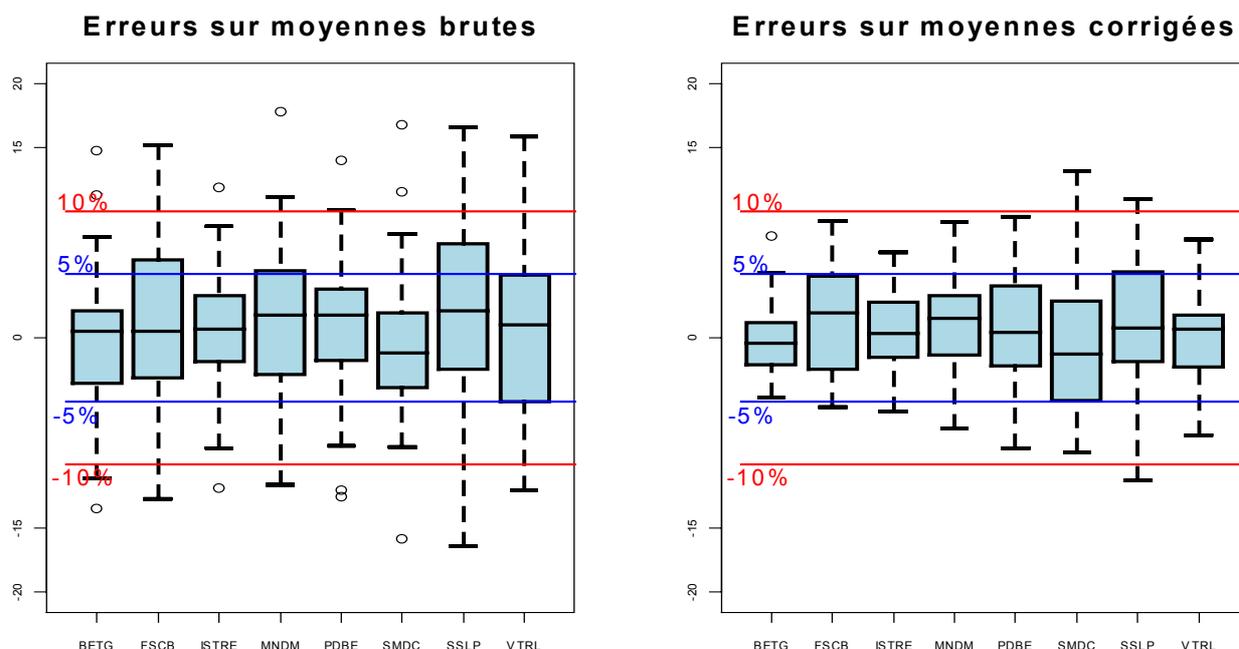


Figure 9 : Comparaison des erreurs brutes et corrigées (en pourcentage) des stations de Berre, Fos-sur-Mer, Istres, Martigues Notre Dame, Port-de-Bouc, St Martin-de-Crau, Sausset-les-Pins et Vitrolles, de 2000 à 2004.

Le tableau 3 permet de visualiser les moyennes annuelles réelles ainsi que leurs estimations brutes et corrigées générant les plus grosses erreurs par année et par station (celles qui vont être arrêtées en 2006). L'écart maximal pour les moyennes brutes est de $4.8 \mu\text{g}/\text{m}^3$, pour les moyennes corrigées est de $3.7 \mu\text{g}/\text{m}^3$. On peut noter que dans 13 % des cas la correction dégrade l'estimation (la station de PDBE en 2000 par exemple). Cependant l'erreur après correction reste dans des proportions restreintes : +/-10 %.

2000					2001				
Stations	Réelle	Brute	Corrigée	Ecart	Stations	Réelle	Brute	Corrigée	Ecart
BETG	24	24	24	0.6	BETG	27	28	26	-0.6
FSCB	16	17	17	1.0	FSCB	15	17	16	1.4
ISTRE	24	25	25	1.2	ISTRE	21	22	20	-1.2
MNDM	16	15	15	-0.5	MNDM	17	19	18	1.5
PDBE	38	40	42	3.6	PDBE	37	39	34	-2.8
SMDC	20	22	21	1.0	SMDC	20	18	18	-1.8
SSLP	13	12	13	0.4	SSLP	16	14	15	-1.8
VTRL	23	22	24	0.7	VTRL	26	30	28	2.0

2002					2003				
Stations	Réelle	Brute	Corrigée	Ecart	Stations	Réelle	Brute	Corrigée	Ecart
BETG	28	32	30	2.2	BETG	27	25	26	-1.0
FSCB	16	15	17	1.0	FSCB	19	17	20	1.3
ISTRE	23	21	24	1.0	ISTRE	26	23	27	1.1
MNDM	18	16	16	-1.3	MNDM	18	16	19	0.8
PDBE	34	39	31	-3.0	PDBE	36	35	38	1.4
SMDC	22	20	21	-0.6	SMDC	27	32	31	3.5
SSLP	17	19	19	1.8	SSLP	18	21	20	1.9
VTRL	26	29	28	2.0	VTRL	26	23	25	-0.6

2004				
Stations	Réelle	Brute	Corrigée	Ecart
BETG	21	23	20	-0.9
FSCB	18	20	19	0.9
ISTRE	20	22	21	1.3
MNDM	18	21	19	1.1
PDBE	35	39	37	1.5
SMDC	21	24	23	1.5
SSLP	14	17	13	-1.6
VTRL	22	25	23	1.0

Tableau 3 : comparaison des moyennes annuelles réelles observées avec les moyennes brutes et corrigées estimées par l'échantillonnage (en $\mu\text{g}/\text{m}^3$) sur les stations qui vont être arrêtées en 2006. Ces dernières correspondent aux estimations produisant les plus grosses erreurs en pourcentage. L'écart est calculé entre la moyenne réelle et la moyenne corrigée.

2.3 Résultats 2005

Sur la période du 20 novembre 2004 au 20 novembre 2005, nous avons simulé des échantillonnages basés sur ceux décrits dans le tableau 1 (le premier échantillon débute le 20 novembre 2004, les suivants sont décalés successivement de 7 jours). Les erreurs commises sur les stations de Berre, Fos-sur-Mer, Istres, Martigues Notre Dame, Port-de-Bouc, St Martin-de-Crau, Sausset-les-Pins et Vitrolles, de 2000 à 2004 sont corrigées par les autres stations fixes du réseau : Arles, Marignane ville, Martigues Île, Miramas, Rognac-les-brets et Salon. Les stations utilisées pour corrigées sont donc celles qui continueront d'exister en 2006.

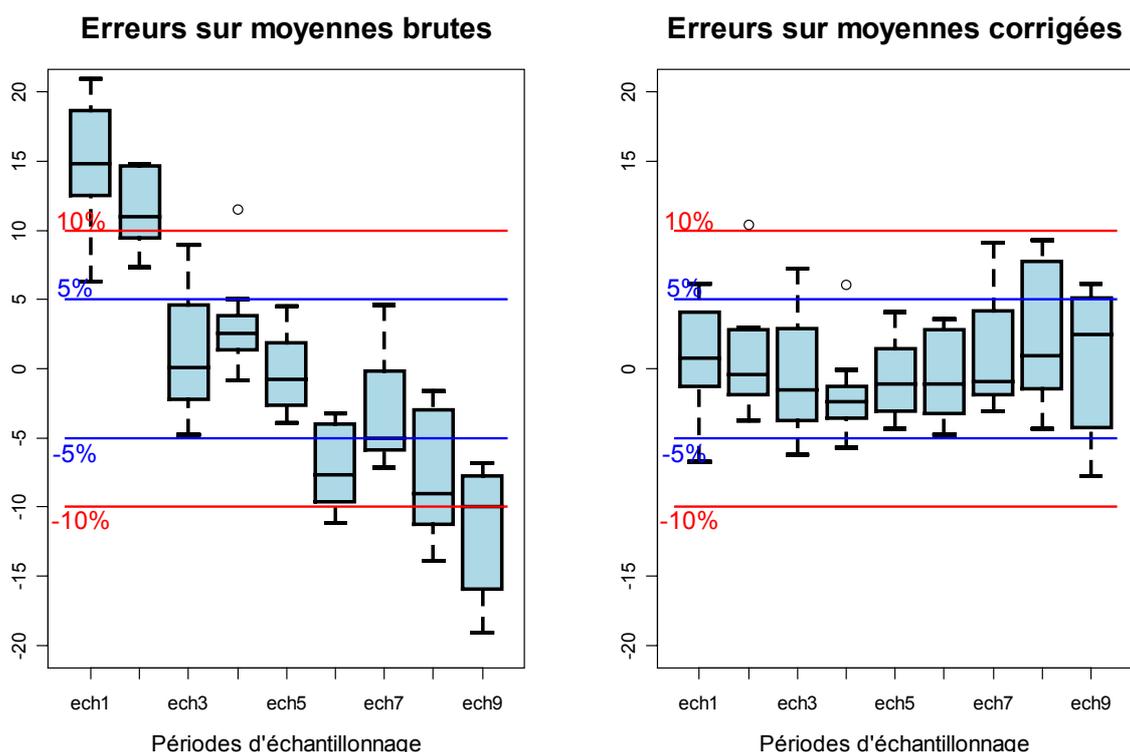


Figure 10 : Comparaison des erreurs brutes et corrigées (en pourcentage) des stations de Berre, Fos-sur-Mer, Istres, Martigues Notre Dame, Port-de-bouc, St Martin-de-Crau, Sausset-les-Pins et Vitrolles, en 2005.

La figure 10 nous permet de voir l'effet de la correction par échantillon. Les biais dans l'estimation sont très variables selon les échantillons sur cette période 2005. Cependant, après correction, la moyenne des erreurs par échantillon est ramenée dans la gamme des +/-5 %. La méthode de correction a été efficace pour tous les échantillonnages testés.

La figure 11, compare les erreurs brutes et corrigées par stations. On passe d'une marge d'erreur de +/-20 % à +/-10 %. L'erreur maximale est de $3.6 \mu\text{g}/\text{m}^3$ pour la moyenne brute contre $2.5 \mu\text{g}/\text{m}^3$ pour la moyenne corrigée qui est systématiquement meilleure sur cette période (tableau 4).

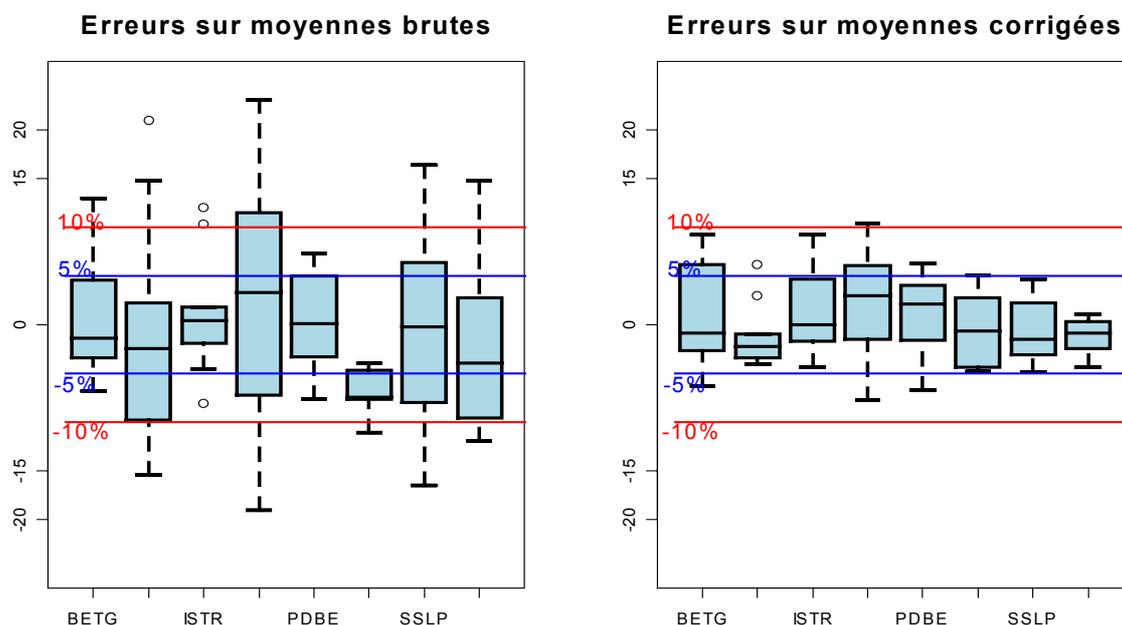


Figure 11: Comparaison des erreurs brutes et corrigées (en pourcentage) des stations de Berre, Fos-sur-Mer, Istres, Martigues Notre Dame, Port-de-bouc, St Martin-de-Crau, Sausset-les-Pins et Vitrolles, en 2005.

Moyennes annuelles 2005				
Stations	Réelle	Brute	Corrigée	Ecart
BETG	24	27	26	2.2
FSCB	15	18	16	0.9
ISTR	21	23	23	1.9
MNDM	15	19	17	1.6
PDBE	37	34	34	-2.5
SMDC	20	17	21	1.0
SSLP	13	11	13	-0.6
VTRL	21	24	20	-0.9

Tableau 4 : comparaison des moyennes annuelles réelles observées avec les moyennes brutes et corrigées estimées par l'échantillonnage (en $\mu\text{g}/\text{m}^3$) sur les stations qui vont être arrêtées en 2006. Ces dernières correspondent aux estimations produisant les plus grosses erreurs en pourcentage. L'écart est calculé entre la moyenne réelle et la moyenne corrigée.

3 Application au PM10

Afin d'illustrer sur un autre polluant la méthode de correction basée sur un ensemble de station fixe de référence, nous avons simulé des échantillonnages (de 2001 à 2004) pour les stations de mesures des PM10 (voir tableau 1 pour les périodes échantillonnées). Les erreurs d'estimations brutes et corrigées de la moyenne annuelle sont représentées sur la figure 12.

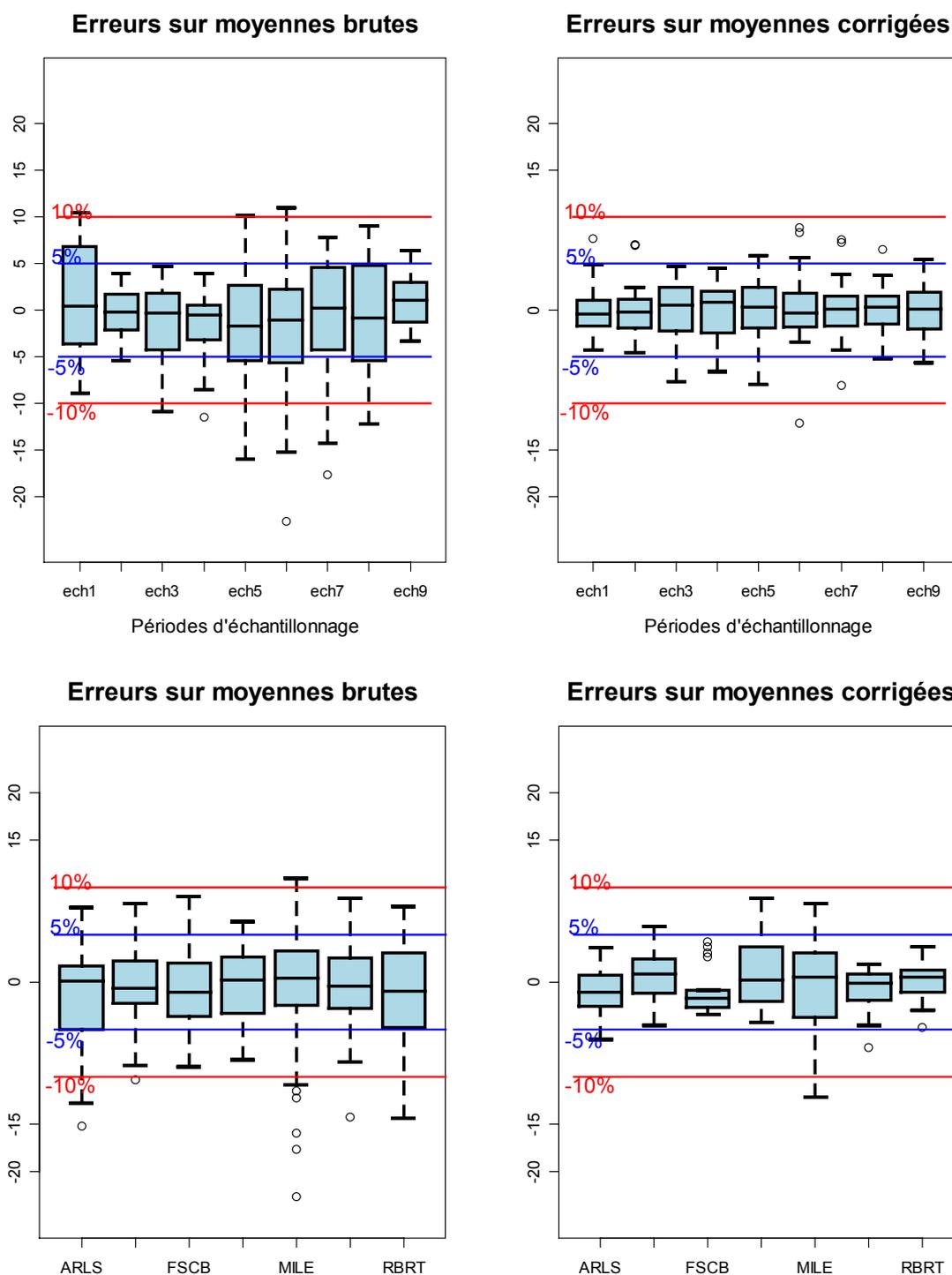


Figure 12 : Les erreurs d'estimations brutes et corrigées (en pourcentage) de la moyenne annuelle des PM10 sur les stations de : Arles, Port-de-Bouc, Fos Carabin, Carry-le-Rouet, Martigues Île, la Mède, Rognac-les-brets et pour chacun des échantillonnages du tableau 1.

Pour chacune des stations considérées, la correction est calculée à partir des autres stations. L'ensemble des erreurs sur moyenne corrigées est compris entre -10 % et + 10 %, hormis une erreur égale à -12 %.

Année 2001					Année 2002				
Stations	Réelle	Brute	Corrigée	Ecart	Stations	Réelle	Brute	Corrigée	Ecart
ARLS	24	25	25	1	ARLS	25	21	25	-1
PDBL	28	26	27	-1	PDBL	28	25	29	2
FSCB					FSCB				
CLRT	25	27	27	2	CLRT	26	24	28	2
MILE	16	17	17	1	MILE	13	10	12	-2
MEDE					MEDE	31	26	30	-1
RBRT					RBRT	24	20	24	1

Année 2003					Année 2004				
Stations	Réelle	Brute	Corrigée	Ecart	Stations	Réelle	Brute	Corrigée	Ecart
ARLS	29	27	27	-1	ARLS	26	24	24	-2
PDBL	30	31	31	1	PDBL	29	31	30	1
FSCB	32	29	30	-1	FSCB	25	27	26	1
CLRT	27	26	28	1	CLRT	26	24	27	2
MILE	24	27	26	2	MILE	25	26	26	1
MEDE	32	30	31	-1	MEDE	28	26	26	-2
RBRT	26	25	25	-1	RBRT	25	26	26	1

Tableau 5 : Comparaison des moyennes annuelles réelles observées avec les moyennes estimées (en $\mu\text{g}/\text{m}^3$). Ces dernières correspondent aux estimations produisant les plus grosses erreurs en pourcentage. L'écart est calculé entre la moyenne réelle et la moyenne corrigée.

4 Synthèse et application

Un échantillonnage destiné à estimer la moyenne annuelle des concentrations de NO_2 pourra donc durer 16 ou 24 semaines ; sachant que 20 semaines n'apportent pas beaucoup de précision par rapport à 16 semaines. Les périodes d'échantillonnage ont peu d'importance (pourvu qu'elles soient également réparties sur les 4 trimestres de l'année civile) car la méthode de correction élimine les biais systématiques potentiels. En effet, chaque année et pour une même période échantillonnée, la plupart des stations du réseau se comportent de la même façon : lorsqu'il y a un biais dans l'estimation de la moyenne annuelle, celui-ci se retrouve souvent dans l'ensemble des stations étudiées (figure 4 et 10). On estime donc que le groupe de stations permettant d'établir la correction nous livre une erreur moyenne qui a de forte chance de se retrouver dans l'estimation sur le site à échantillonner. Quand ce n'est pas le cas, l'erreur après correction peut être plus importante que l'erreur brute : cependant, cette erreur ne dépasse pas les 10 % dans les simulations effectuées entre 2000 et 2005. On suppose donc que l'erreur moyenne du groupe de stations qui permet d'établir la correction est l'erreur à laquelle on peut s'attendre sur tout autre site sur lequel on effectuera un échantillonnage.

Concrètement, dans le cas d'un échantillonnage dans un cœur de ville disposant d'un appareil fixe, la correction s'effectue uniquement à l'aide de cette station fixe. Dans le cas d'un échantillonnage dans un lieu où il n'y a pas de station fixe à proximité, la correction s'effectue à l'aide de toutes les stations fixes du réseau AIRFOBEP (cf. figures 5 et 6).

Nous avons réalisé une macro Excel permettant de calculer la correction et une feuille de calcul permettant de corriger la moyenne brute. La macro calcule de coefficient de correction à appliquer à la moyenne brute de la campagne de mesures discontinues. Elle utilise en entrée:

- La date des périodes échantillonnées
- Les concentrations moyennes journalières des stations fixes de référence

La feuille de calcul donne la moyenne corrigée ainsi que son intervalle de confiance à 95 % (cf. norme ISO 11222). Elle utilise en entrée :

- Le coefficient de correction
- Les concentrations échantillonnées durant la campagne de mesures discontinues

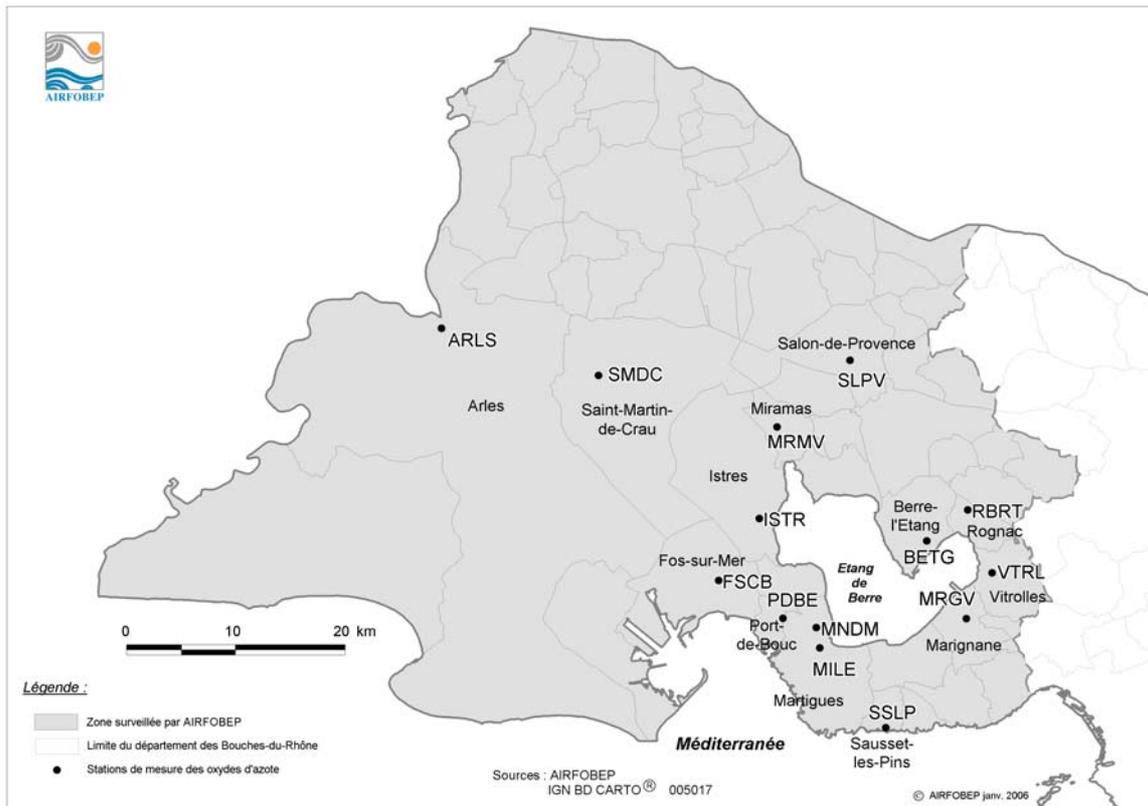
4.1 Utilisation de la macro

- Faire une extraction XAIR des concentrations moyennes journalières de l'ensemble des stations fixes du réseau (10 maximums) ou de la station à proximité du site de mesures discontinues sur l'année échantillonnée
- Copier/coller ces données dans la feuille « données » du classeur à partir de la colonne B.
- Dans la colonne « Périodes d'échantillonnage » de la feuille « données » du classeur, noter « 1 » les jours échantillonnés et rien les autres jours.
- Lancer la macro.
- Dans la feuille « résultat », la première ligne est le nom des stations utilisées, la deuxième la moyenne annuelle réelle, la troisième la moyenne annuelle échantillonné brute, la quatrième l'erreur en % entre la moyenne brute et réelle, la cinquième est le coefficient de correction associé à l'erreur.
- Noter le coefficient de correction moyen (dans la colonne moyenne, la valeur encadrée)

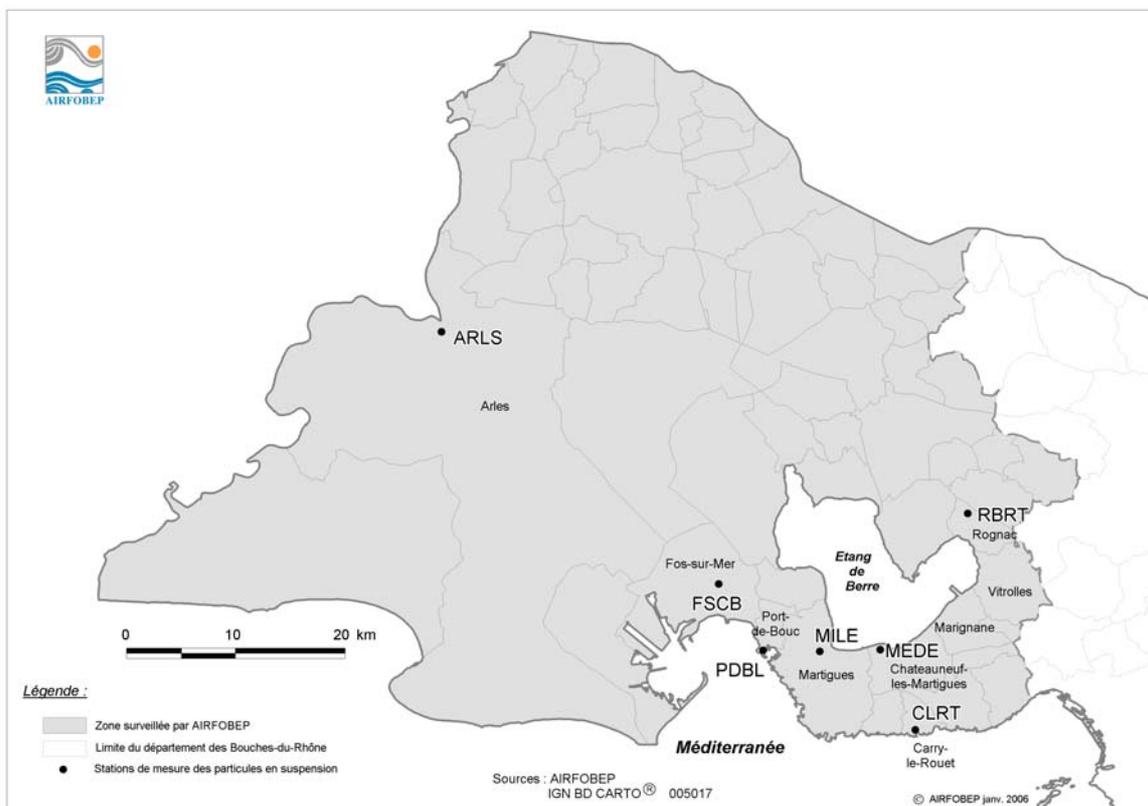
Remarque : la quatrième ligne permet de voir l'erreur de l'échantillonnage sur chaque station fixe de références, ainsi que l'erreur moyenne à partir de laquelle le coefficient de correction est calculé.

4.2 Utilisation de la feuille de calcul

- Sur la deuxième ligne « nom » de la feuille « données » du classeur, entrer les noms des sites de mesures discontinues (10 maximum).
- Dans les colonnes correspondantes entrer les valeurs mesurées lors des campagnes, ces valeurs doivent être des moyennes journalières ou sur des moyennes sur 15 jours (365 données maximums)
- Sur la ligne 10, de la feuille « résultat » de la macro 2, entrer le coefficient issu de la première macro, autant de fois qu'il y a de site de mesures discontinues.
- Les lignes 3 à 6 donnent les informations suivantes sur les échantillons : nom, nombre de données, moyenne brute et écart type.
- Les lignes 15 à 17 calculent la moyenne corrigée ainsi que les bornes de l'intervalle de confiance à 95 % par chaque site échantillonné.



Carte 1 : localisation des stations de mesures du NO₂



Carte 2 : localisation des stations de mesures des PM₁₀.

5 Annexe 1 : échantillonnage stratifié

Dans l'étude « Influence des paramètres météorologiques dans la stratégie de mesure à l'aide de moyens mobiles » J.L. Houdret propose une méthode permettant d'estimer une moyenne annuelle à partir de campagne de mesures discontinues. Nous proposons d'illustrer la méthode à l'aide d'une simulation de campagnes de mesures discontinues du NO₂ sur les stations d'Istres, Marignane et Salon au cours de l'année 2004. Le paragraphe suivant synthétise la méthode. Nous évaluons ensuite l'effet que les paramètres météorologiques et temporels peuvent avoir sur les concentrations. Une troisième partie explique comment en regroupant ces informations on peut construire les strates de concentration. Nous exposerons alors les résultats obtenus en estimant la moyenne annuelle sur les sites étudiés.

5.1 La méthode

Sur le site test ou de référence :

- Analyser les données pour connaître l'influences des paramètres météo et temporels sur les mesures de concentration.
- Elaborer de « strates » de concentration correspondant à des combinaisons de ces paramètres.
- Déterminer le nombre de mesures qu'il conviendrait de faire dans chaque strate pour estimer la moyenne à un niveau d'incertitude donné (à partir du calcul des moyennes et écarts types des polluants dans chacune des strates).

Sur le site ou la concentration est estimée :

- Effectuer des mesures couvrant une période supérieure à celle estimée par strate (pour être sûr d'avoir au moins le nombre de mesures nécessaires).
- Calculer la moyenne brute obtenue par les mesures discontinues (ou la somme des moyennes de chaque strate pondérées par les proportions de mesures par strate obtenues pendant la campagne discontinue, c'est pareil !)
- Calculer l'intervalle de confiance de cette moyenne, en utilisant les proportions de mesures discontinues obtenues dans chaque strate (la moyenne des écarts types de chaque strate, pondérés par la proportion de mesures discontinues associée à chacune d'entre elles ; on pourrait utiliser l'écart type de moyenne standard, les résultats seraient différents)
- Calculer la moyenne reconstituée obtenue par les mesures discontinues (la somme des moyennes de chaque strate pondérées par les proportions théoriques de mesures par strate obtenues sur le site test)
- Calculer l'intervalle de confiance de cette moyenne en corrigeant les proportions de mesures discontinues obtenues dans chaque strate par les proportions « idéales » obtenues sur le site de référence.

Moyenne brute :

Elle ne tient pas compte de la représentativité météorologique de l'échantillon :

La moyenne brute est indépendante des conditions météo rencontrées durant la campagne de mesure discontinue. Son intervalle de confiance dépend de la proportion de mesure dans

chaque strate, mais ne tient pas compte de leur représentativité par rapport aux proportions idéales du site de référence.

Moyenne reconstituée :

Elle tient compte de la représentativité météorologique de l'échantillon :

La moyenne reconstituée est fonction des proportions théoriques obtenues sur le site de référence. (Si les proportions de mesures dans les strates ne sont pas respectées cela n'a pas d'importance car les moyennes des strates sont pondérées par les proportions théoriques). Ces proportions n'étant pas forcément respectées par la campagne de mesure discontinue, son intervalle de confiance tient compte de leur représentativité par rapport aux proportions théoriques. Cet intervalle sera d'autant plus grand que les mesures discontinues seront effectuées durant des périodes météo non représentatives de la météo annuelle.

Remarques :

Plus les campagnes de mesure sont longues et plus on a de chance d'avoir un échantillon représentatif, donc plus on peut s'affranchir sans trop de risque de la stratification.

Il est nécessaire d'avoir deux paramètres météo durant les campagnes de mesures discontinues : il est possible de les prendre sur les stations de mesures du réseau AIRFOBEP (si celles-ci couvrent la zone météo considérée) ou de les commander à la climathèque.

5.2 Analyse des données

Il nous faut définir les relations entre les variables météorologiques et temporelles, et les concentrations en NO₂. Les paramètres météorologiques et temporels pris en compte sont:

- la vitesse du vent,
- la température,
- l'heure des séquences tri-horaires,
- le jour de la semaine.

Pour chacun d'entre eux, nous allons déterminer le seuil à partir duquel les valeurs prises par le paramètre considéré favorisent les fortes concentrations de NO₂. Les graphiques présentés dans les sous paragraphes permettent d'avoir une représentation visuelle de l'effet d'une variable sur les concentrations. Ils représentent la moyenne des valeurs de NO₂ (en ordonnée) pour les différentes classes ou modalités de la variable étudiée (en abscisse). Le seuil choisi est celui qui scinde les données de concentrations en deux groupes les plus homogènes possibles. Il est calculé par arbre de régression. Celui-ci définit deux sous-groupes de concentration optimaux au sens de la déviance. Il sélectionne donc la coupure qui minimise la somme des carrés des écarts à la moyenne de chaque sous-groupes. Les sous-groupes ainsi créés sont le plus centrés possible autour de leur moyenne respective. Cela est particulièrement utile par exemple, pour la vitesse de vent à la station de Salon pour laquelle le graphique ne présente pas de coupure franche permettant de créer les zones favorables et défavorables aux fortes concentrations de NO₂.

5.2.1 La vitesse du vent

La figure 13 montre l'effet de la vitesse du vent sur les concentrations en NO₂. La concentration moyenne par classe diminue avec la force du vent. Le seuil optimal séparant les classes des fortes et faibles concentrations est entre 2 et 5 m/s selon les stations (tableau 5).

Stations	Istres	Marignane	Salon
Seuils	4.5	5.2	2.6

Tableau 6 : les seuils sélectionnés en mètres par seconde.

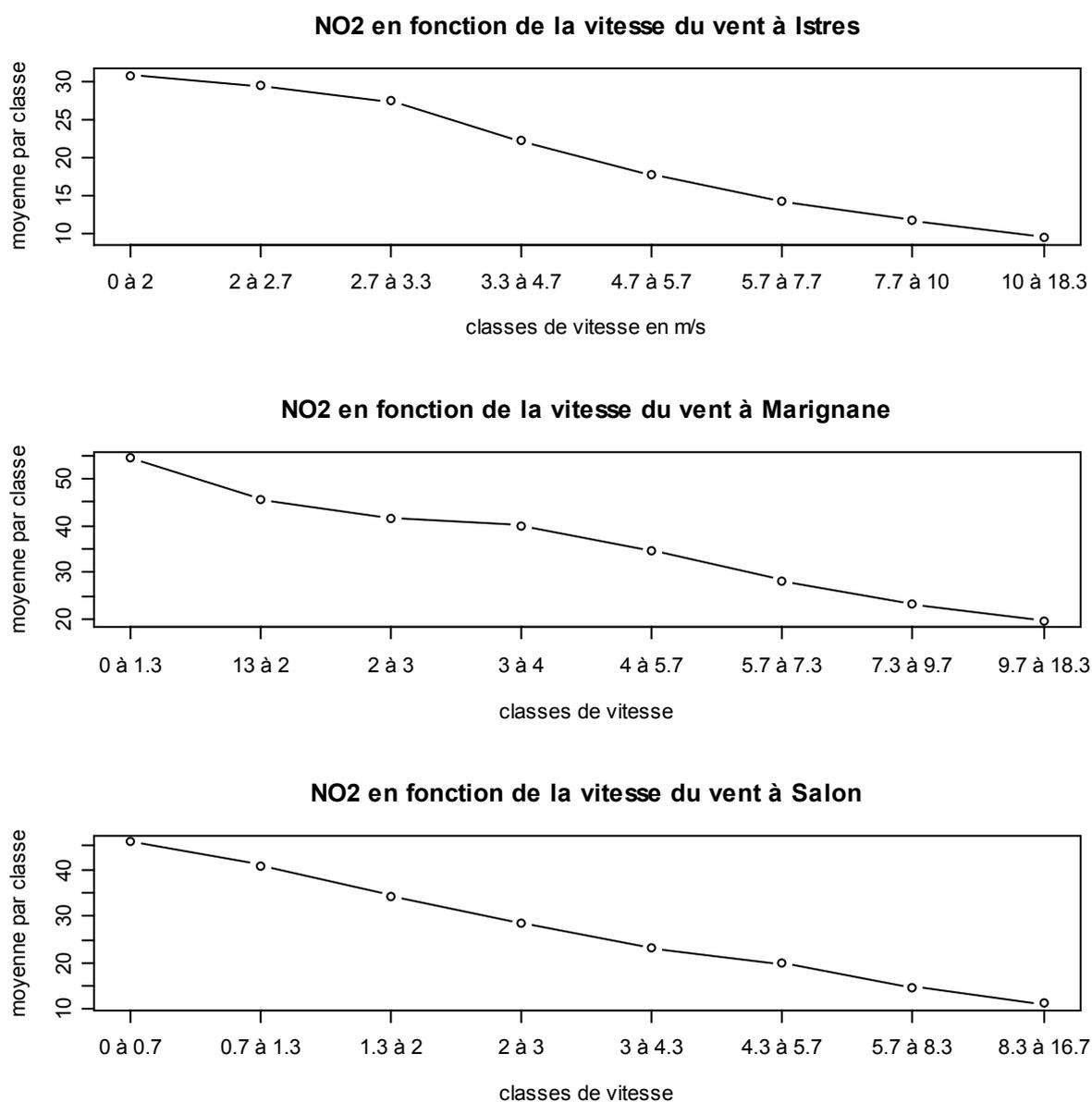


Figure 13 : la concentration moyenne de NO₂ (en µg/m³) par classe de vitesse de vent. Celles-ci sont définies à partir des percentiles suivants: 12.5, 25, 37.5, 50, 62.5, 75, 87.5, 100.

5.2.2 La température

La figure 14 montre qu'une élévation de température entraîne une baisse des concentrations. On note que c'est entre 8 et 12 degrés que la concentration moyenne par classe diminue significativement. Le seuil optimal séparant les classes des fortes et faibles concentrations est entre 10 et 12 degrés selon les stations (tableau 7).

Stations	Istres	Marignane	Salon
Seuils	12.4	10.6	11.2

Tableau 7 : les seuils sélectionnés en degrés Celsius

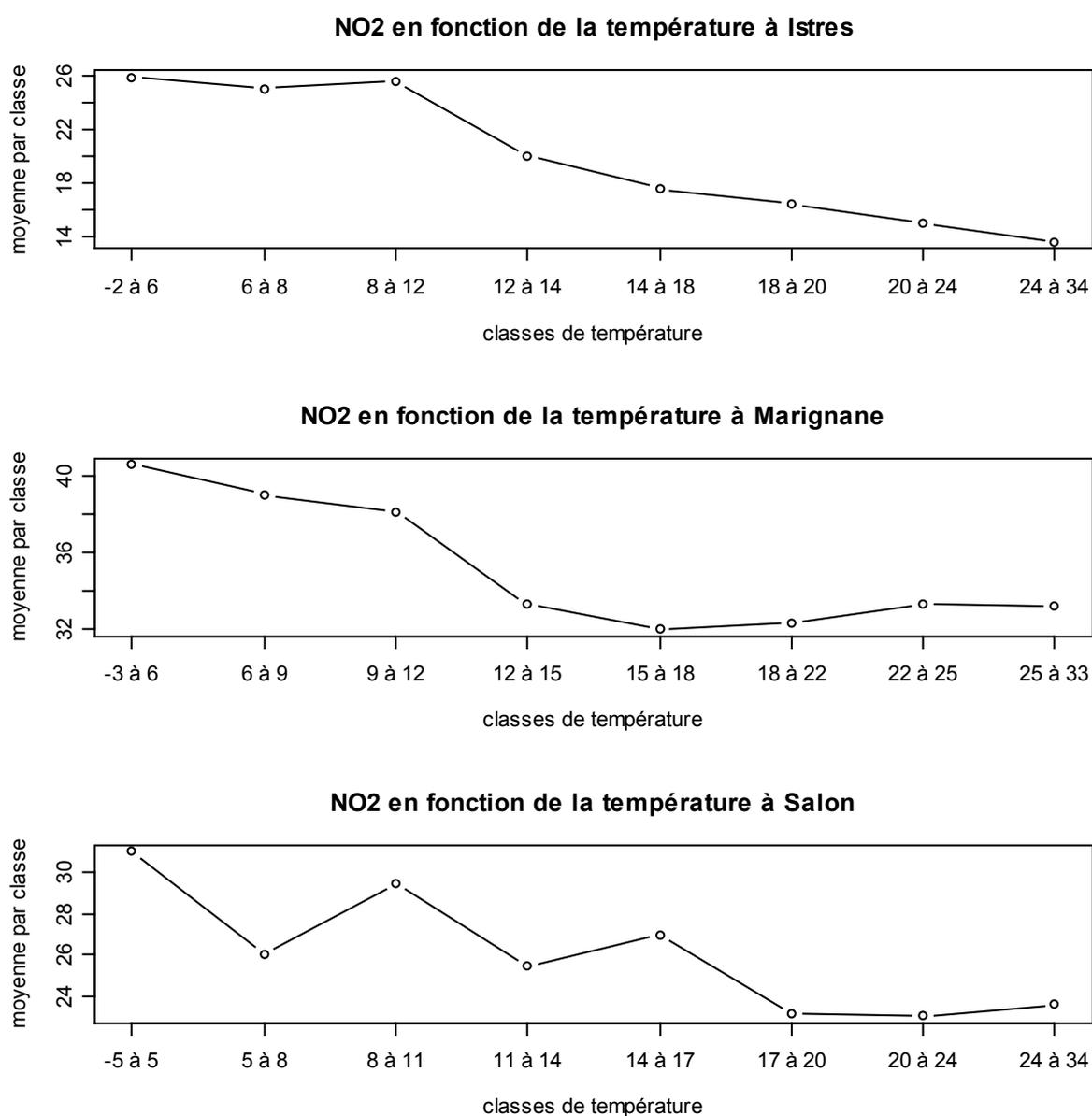


Figure 14 : la concentration moyenne de NO₂ (en µg/m³) par classe de température. Celles-ci sont définies à partir des percentiles suivants : 12.5, 25, 37.5, 50, 62.5, 75, 87.5, 100.

5.2.3 L'heure du jour

Les profils journaliers sont sensiblement les mêmes pour les trois stations étudiées : les tranches tri-horaires de 9 heures et de 18 heures ont les concentrations les plus élevées (figure 15). Cependant, les tranches tri-horaires favorables et défavorables sélectionnées sont différentes selon les stations (tableau 8).

Stations	Istres	Marignane	Salon
Heures peu favorables	3, 15, 24	3, 6, 24	3, 6, 12, 15, 24

Tableau 8 : les heures peu favorables aux fortes concentrations de NO₂.

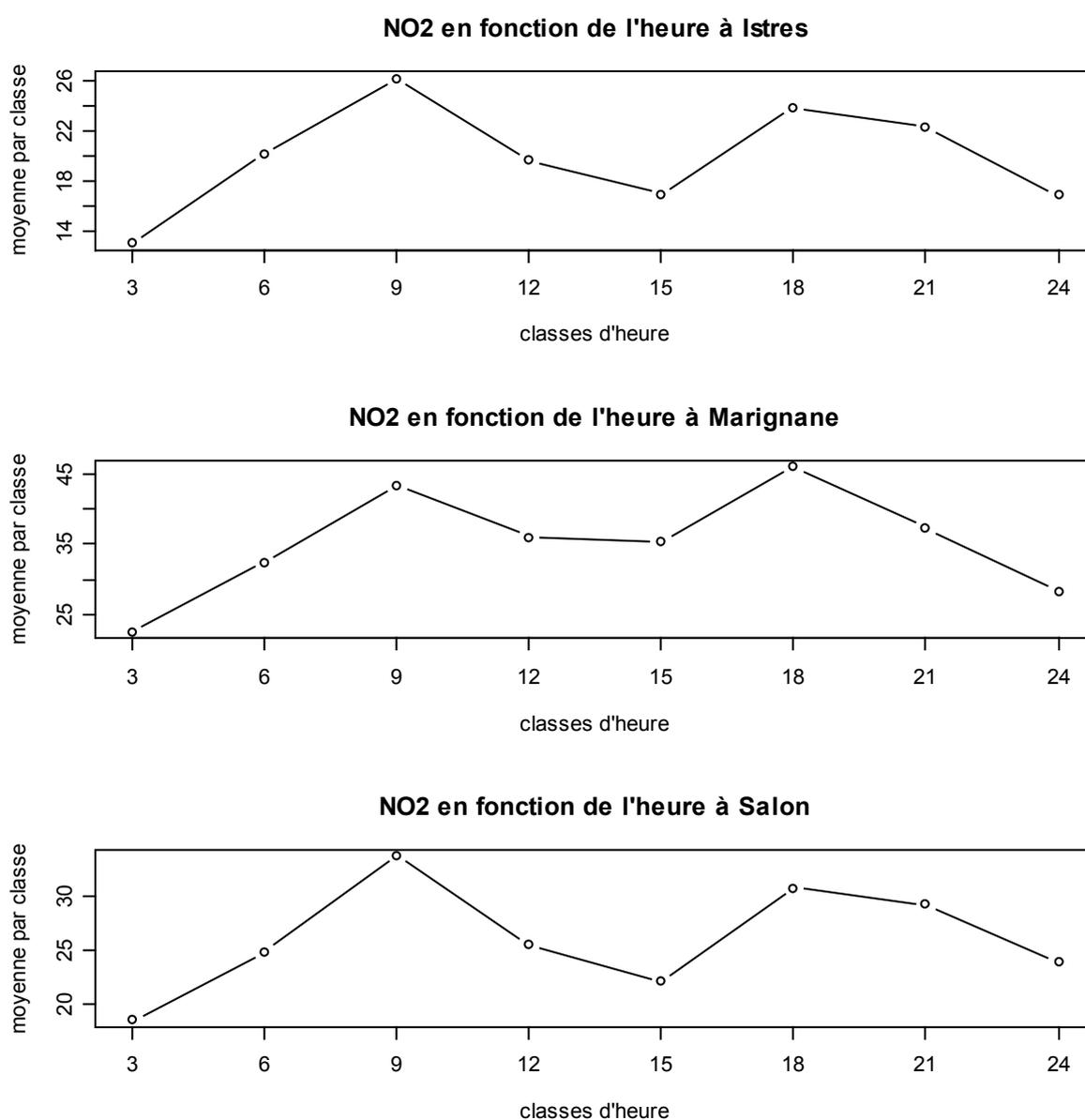


Figure 15 : la concentration moyenne de NO₂ (en µg/m³) par tranche de 3 heures dans une journée.

5.2.4 Le jour de la semaine

On observe sur la figure 16 que les profils hebdomadaires sont similaires pour les trois stations : une baisse sensible des concentrations en fin de semaine. Le samedi et le dimanche sont donc considérés comme des jours peu favorables aux fortes concentrations alors que les autres jours de la semaine le sont.

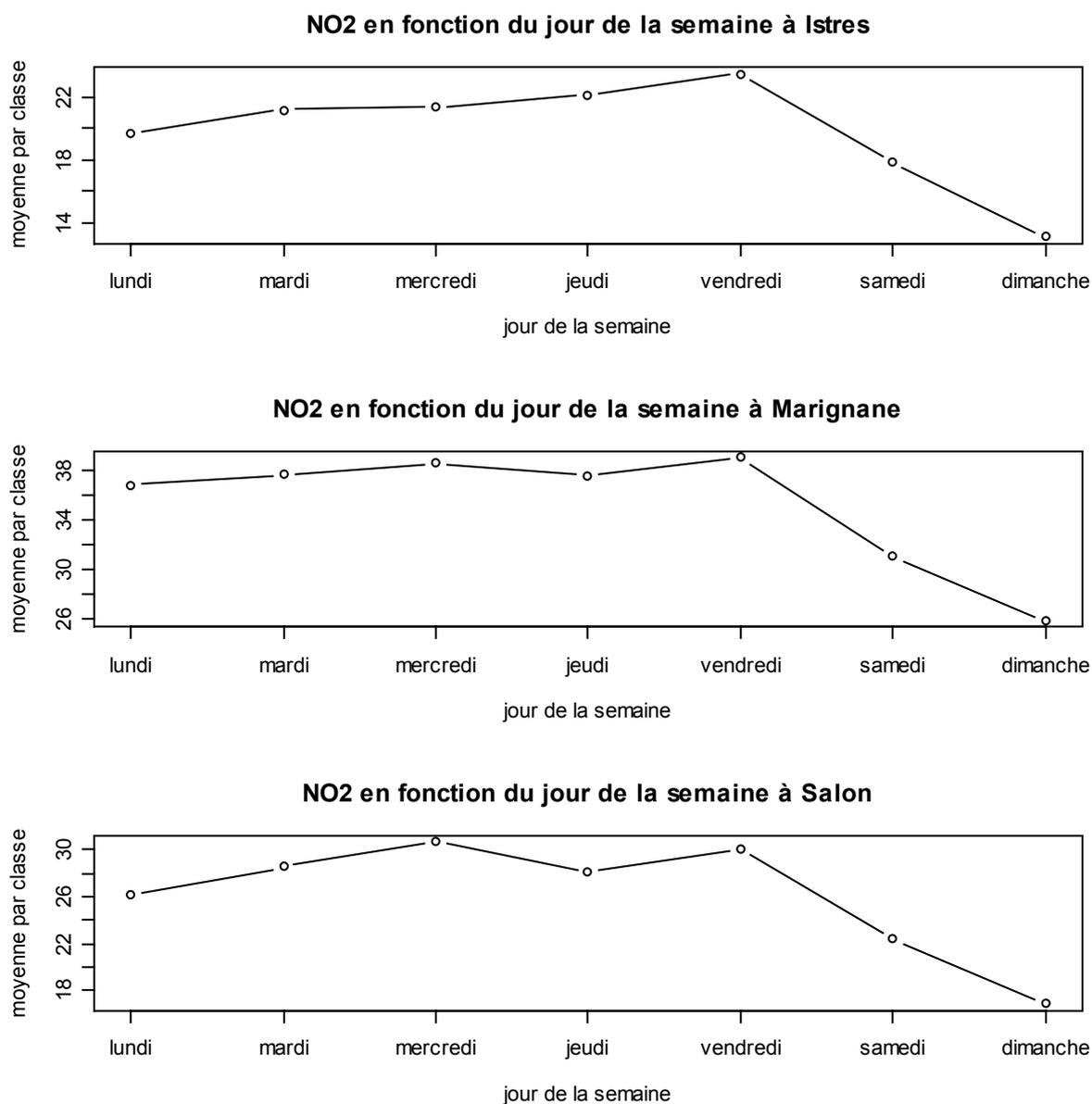


Figure 16 : la concentration moyenne de NO2 (en $\mu\text{g}/\text{m}^3$) par jour de la semaine.

5.3 Définition des strates de concentrations

Connaissant pour chaque paramètre les classes favorisant les fortes concentrations en NO₂, nous construisons les strates de la façon suivante. Pour chaque séquence tri-horaire :

- On compte le nombre de fois où la séquence se situe dans une classe de paramètre favorisant les fortes concentrations (tableau 9),
- On définit ensuite la strate à laquelle appartient cette séquence :
 - A la strate 1 si on compte 0 ou 1,
 - A la strate 2 si l'on compte 2,
 - A la strate 3 si l'on compte 3 ou 4.

Considérons un exemple sur la station d'Istres : une séquence tri-horaire de 15 heures, le lundi enregistre un vent à 3 m/s, une température de 10 degrés. Cette séquence se situe par trois fois dans une classe favorisant les fortes concentrations de NO₂ (tableau 9):

- la vitesse du vent est inférieure à 4.5 m/s,
- la température est inférieure à 12.4 degrés,
- le jour de la semaine est un lundi.

On attribue donc cette séquence à la strate 3.

	Vitesse de vent (m/s)	Température (degrés)	Heure du jour	Jour de la semaine
Istres	<4.5	<12.4	6,9,12,18,21	Du lundi au vendredi
Marignane	<5.2	<10.6	9,12,15,18,21	Du lundi au vendredi
Salon	<2.6	<11.2	9,18,21	Du lundi au vendredi

Tableau 9 : les classes de paramètres favorisant des concentrations tri-horaires élevées.

En ventilant toutes les données de chaque station, dans les strates on obtient les résultats suivants :

Strates\stations	Istres	Marignane	Salon
1	644	557	1013
2	1118	1231	1086
3	1165	1139	828

Tableau 10 : le nombre de données par strate de concentration.

5.4 L'échantillonnage

La maintenance des stations de mesures s'effectuant environ toutes les quatre semaines, et la durée optimale de mesure d'un tube étant de 2 semaines, on teste deux campagnes de mesures distinctes. Elles sont composées respectivement de 4 fois 1 mois et 4 fois 2 semaines réparties sur 4 trimestres. On testera la robustesse de la méthode en répétant ces campagnes simulées en les décalant autant de fois 1 semaine qu'il est possible.

ech1&10	ech2&11	ech3&12	ech4&13	ech5&14	ech6&15	ech7&16	ech8&17	ech9&18
1 ^{er}	8	15	22	29	5	12	19	26
janvier	janvier	janvier	janvier	janvier	février	février	février	février
1 ^{er}	8	15	22	29	6	13	20	27
avril	avril	avril	avril	avril	mai	mai	mai	mai
1 ^{er}	8	15	22	29	5	12	19	26
juillet	juillet	juillet	juillet	juillet	août	août	août	août
1 ^{er}	8	15	22	29	5	12	19	26
octobre	octobre	octobre	octobre	octobre	novembre	novembre	novembre	novembre

Tableau 11 : les dates de début des 4 périodes constituant les campagnes simulées : de 1 à 9 celles de type 4 fois 1 mois, de 10 à 18 celles de type 4 fois 2 semaines. Concernant la campagne de 4 fois 2 semaine, on a simulé un 19^{ème} échantillonnage : 3 février, 3 juin, 4 septembre, 3 décembre.

Il nous est possible d'estimer à partir d'une marge d'erreur fixée à 10 % le nombre de données à échantillonner (tableau 10). Cependant, les échantillonnages simulés sont systématiquement plus grands que le nombre minimal de données nécessaires à une marge d'erreur de 10%.

Istres	Marignane	salon
97	94	93

Tableau 10 : le nombre de valeur tri-horaires nécessaires à une marge d'erreur : +/- 10 % pour l'estimation de la moyenne.

Nous avons donc calculé la marge d'erreur théorique à laquelle nous pouvons nous attendre à partir de la taille de nos échantillons (tableau 11). Celle-ci varie entre 3 % et 5 %.

	Istres	Marignane	salon
4 fois 1 mois	3.4 %	3.3 %	3.3 %
4 fois 2 semaines	4.8 %	4.7 %	4.6 %

Tableau 11 : la marge d'erreur théorique en pourcentage de la moyenne annuelle en fonction de la taille de l'échantillon.

5.5 Estimation des moyennes annuelles

Pour chacun des 19 échantillonnages simulés, nous avons calculé la moyenne brute (moyenne simple des valeurs recueillies) et la moyenne reconstituée (corrigée par la méthode de stratification). Les tableaux 12 et 13 nous permettent de comparer ces deux estimateurs de la moyenne annuelle. Les moyennes reconstituées donnent les meilleurs résultats (voir les deux dernières lignes des tableaux : moyenne et écart type). Principalement pour la campagne de 4 fois 2 semaines (tableau 13) où les écarts sont plus significatifs.

Stations	Istres		Marignane		Salon	
Moyenne annuelle	20		35.2		26.1	
Estimations	brute	reconstituée	brute	reconstituée	brute	reconstituée
ech01	20	20.2	33.2	33.9	26.5	26.7
ech02	20.6	20.5	33.1	33.4	26.1	25.9
ech03	20.3	20	33.2	33.3	24.8	24.7
ech04	19.6	19.4	33.9	33.9	24.1	23.9
ech05	19.9	19.5	34.8	34.4	25.1	24.8
ech06	19.5	19.2	35.7	35.2	24.6	24.4
ech07	20.9	20.4	37.3	36.4	26.9	26.6
ech08	22.4	21.8	38.9	37.9	29.5	29
ech09	21.5	21.2	37.6	37.3	28.2	28.3
Moyenne	20.5	20.2	35.3	35.1	26.2	26.0
Ecart type	0.95	0.85	2.19	1.73	1.79	1.78

Tableau 12 : les estimations de la moyenne annuelle (en $\mu\text{g}/\text{m}^3$) obtenues lors de la simulation de 4 fois 1 mois. L'estimation brute est simplement la moyenne de l'échantillon recueilli, la moyenne reconstituée est corrigée à l'aide de la stratification.

Stations	Istres		Marignane		Salon	
Moyenne annuelle	20		35.2		26.1	
Estimations	brute	reconstituée	brute	reconstituée	brute	reconstituée
ech10	20.6	20.6	35.5	35.8	28.1	28.1
ech11	20.1	19.8	32.8	33.1	26.8	26.4
ech12	19.6	19.8	31.9	32.3	24.9	25
ech13	21	20.9	33.8	33.7	26	25.8
ech14	20.8	20	33.7	33.4	24.6	24.2
ech15	19.2	18.4	35.5	34.5	23.5	22.9
ech16	19	18.6	36.3	35	25	24.3
ech17	19.4	19.3	35.2	34.9	25	25
ech18	23.1	22.1	39.3	37.9	29.5	28.9
ech19	25.9	24.4	42.7	40.4	33.9	32.7
Moyenne	20.9	20.4	35.7	35.1	26.7	26.3
Ecart type	2.13	1.78	3.22	2.44	3.09	2.88

Tableau 13 : les estimations de la moyenne annuelle (en $\mu\text{g}/\text{m}^3$) obtenues lors de la simulation de 4 fois 2 semaines. L'estimation brute est simplement la moyenne de l'échantillon recueilli, la moyenne reconstituée est corrigée à l'aide de la stratification.

La marge d'erreur en pourcentage par rapport à la moyenne réelle est supérieure à la moyenne théorique donnée précédemment. Elle se situe entre 3% et 8 % (tableau 14 et 15). L'erreur faite par la moyenne reconstituée est plus faible de 0.5 point à 1 point par rapport à la moyenne brute. Cet écart n'est donc pas très important. Cependant, les écarts entre les erreurs maximales pour les deux estimateurs vont jusqu'à 7 points. C'est à dire que la correction par la stratification limite les grosses erreurs d'estimation.

stations	Istres		Marignane		Salon	
	brute	reconstituée	brute	reconstituée	brute	reconstituée
Erreur en % moyenne	3.7	3.3	5.2	4.1	5.3	5.5
max	12	9	10.5	7.7	13.0	11.1

Tableau 14 : la moyenne et le maximum des erreurs (en pourcentage par rapport à la moyenne annuelle) de la campagne 4 fois 1 mois (calculés à partir du tableau 9).

stations	Istres		Marignane		Salon	
	brute	reconstituée	brute	reconstituée	brute	reconstituée
Erreur en % moyenne	7.2	6.1	6.2	5.1	8.2	8.1
max	29.5	22	21.3	14.8	29.9	25.3

Tableau 15 : la moyenne et le maximum des erreurs (en pourcentage par rapport à la moyenne annuelle) de la campagne 4 fois 2 semaines (calculés à partir du tableau 10).

5.6 Bilan des simulations

Les résultats obtenus nous permettent d'évaluer la marge d'erreur à laquelle nous pouvons nous attendre lors de campagnes réelles, d'une part sur un site pour lequel nous avons des années de mesures complètes, d'autre part, sur les sites où il n'y a jamais eu de mesures effectuées. Dans le premier cas, en situation d'échantillonnage réel, la construction des strates pourra se faire avec l'historique des années disponibles pour lesquelles il existait une station fixe (on a besoin des paramètres météorologiques et des valeurs de concentration sur l'année complète pour définir les strates). Il nous faudra émettre l'hypothèse que les années disponibles (en polluant et météo) pour un site donné, nous permettent de définir une année météorologique type. Dans le deuxième cas (où l'on estime la moyenne annuelle dans un lieu où il n'y a jamais eu de station fixe) la construction des strates pourra se faire sur une station située dans une même zone météorologique que le site à échantillonner. On pourra utiliser alors uniquement l'année concernée par l'échantillonnage pour effectuer la stratification.

La simulation présentée ne correspond donc pas exactement au deux cas de figure réels. Elle consiste à construire des strates de concentration à partir de l'année 2004 et à simuler un échantillonnage sur cette même année. Elle représente le cas théorique le plus favorable à une bonne estimation de la moyenne annuelle par la méthode des strates : c'est sur la même station et sur la même année que nous construisons les strates et que nous simulons les campagnes de mesure discontinue. Au regard des résultats obtenus, on peut dire que cette méthode génère des erreurs moins importantes que le calcul d'une simple moyenne. Le bénéfice se trouve plus dans le contrôle des grandes erreurs, que dans une amélioration de l'erreur moyenne.

6 Annexe 2 : Comparaison des méthodes de correction

Pour diminuer les erreurs d'estimation de la moyenne annuelle, nous allons comparer les deux méthodes de correction décrites

6.1 Echantillonnage stratifié

Dans l'étude « Influence des paramètres météorologiques dans la stratégie de mesure à l'aide de moyens mobiles » J.L. Houdret propose une méthode permettant d'estimer une moyenne annuelle à partir de campagne de mesures discontinues. Cette méthode peut se résumer en quatre points :

- Sur le site de référence :
 - Analyser les données pour connaître l'influence des paramètres météorologiques et temporels sur les mesures de concentration,
 - Elaborer de « strates » de concentration correspondant à des combinaisons de ces paramètres,
- Sur le site où l'on estime la concentration annuelle :
 - Effectuer des mesures discontinues,
 - Reconstituer la moyenne annuelle à partir de ces mesures, en considérant la stratification du site de référence

L'annexe au paragraphe 6 synthétise la méthode et en donne une illustration à l'aide d'une simulation de campagnes de mesures discontinues du NO₂ sur les stations d'Istres, Marignane et Salon au cours de l'année 2004.

6.2 Correction à partir de stations de références

- Sur le site où l'on effectue la campagne de mesures discontinues :
 - On recueille les échantillons
 - On calcule leur moyenne brute
- Sur les ou la station fixe de référence :
 - On simule cet échantillonnage,
 - On estime la moyenne annuelle à partir de cet échantillonnage simulé pour chacune des stations fixes
 - On calcule l'erreur entre la moyenne estimée par l'échantillonnage simulé et la moyenne pour chacune des stations fixes :
 - On calcule la moyenne de ces erreurs
- On utilise cette erreur moyenne calculée sur les stations fixes pour corriger la moyenne obtenue lors de la campagne de mesure discontinue.

6.3 Comparaison

Les résultats suivants (figure 17, tableau 16) permettent de comparer les deux méthodes sur les stations d'Istres, Marignane, et Salon de 2000 à 2004.

	Moyenne Brute	Moyenne reconstituée	Moyenne corrigée
Erreurs en %	3.88	3.66	2.76

Tableau 16 : La moyenne des erreurs en valeur absolue sur les moyennes brutes, reconstituées et corrigées. (sur les stations d'Istres, Marignane et Salon en de 2000 à 2004)

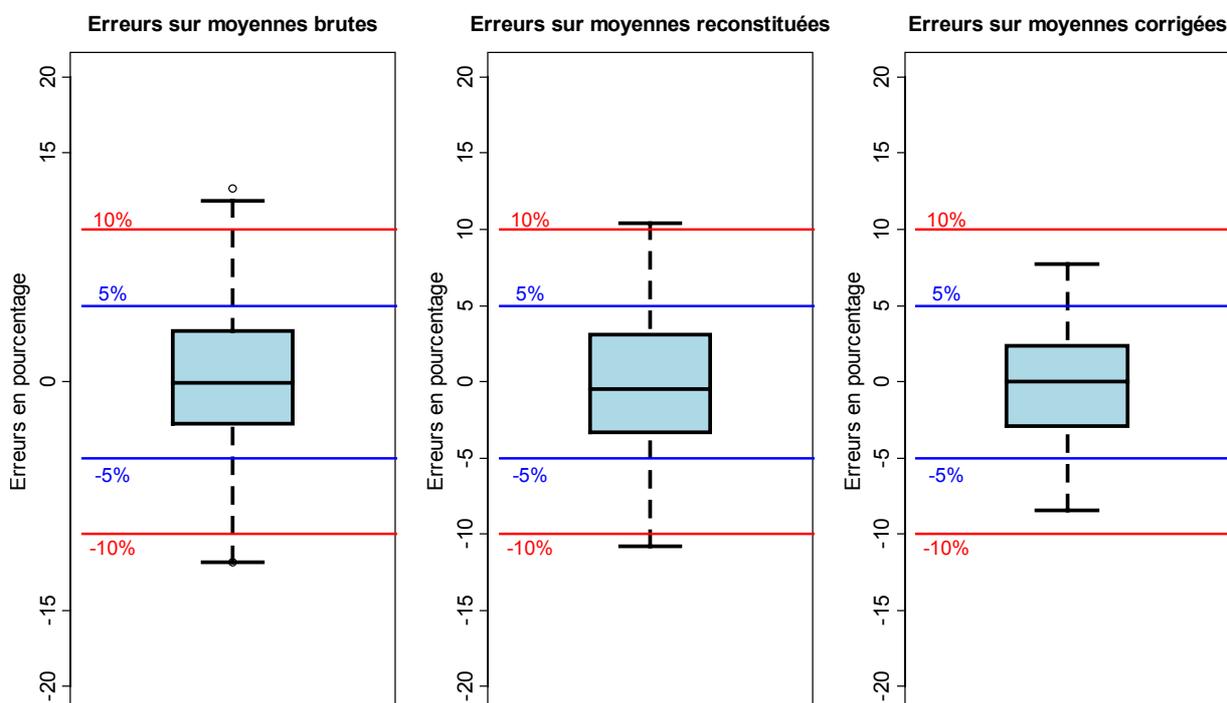


Figure 17 : Les "boxplot" des erreurs (en %) commises en estimant les moyennes annuelles de 2000 à 2004 pour les stations d'Istres, Marignane et Salon. Les trois graphiques correspondent aux erreurs sur les moyennes : brutes (le 1^{er}), reconstituées par la stratification (le 2^{ème}) et corrigées à l'aide des autres stations du réseau (le 3^{ème}).

Les erreurs calculées sur les moyennes reconstituées (échantillonnages stratifiés) n'améliore que très peu les estimations par rapport aux moyennes brutes (sans aucune correction). Les erreurs calculées sur les moyennes corrigées par l'ensemble des stations du réseau diminuent l'erreur moyenne de 1 point (tableau 16) et restreint l'ensemble des erreurs dans la gamme de -10 à 10 % (figure 17).

La méthode retenue est donc celle utilisant le groupe de stations fixes pour corriger les moyennes brutes. Cette méthode ne nécessite pas d'étude météorologique préalable ni de site de référence effectuant des mesures météo et polluant.