

ANNEXE 10

**Cas d'application :
Echantillonnage temporel des HAP pour l'estimation
d'une moyenne annuelle**

Note de synthèse

Echantillonnage temporel des HAP pour l'estimation d'une moyenne annuelle

Synthèse d'une étude de données de HAP réalisée par le GT *Plans d'échantillonnage et reconstitution de données*

Auteurs :

Michel BOBBIA, AIR Normand
Jean-Luc HOUDRET, LCSQA / EMD
Laure MALHERBE, LCSQA / INERIS
Vladislav NAVEL, ATMO Poitou Charentes
Frédéric PENVEN, AIR Pays de la Loire
Céline PHILLIPS, ADEME
Pierre-Yves ROBIC, ORAMIP
Lionel ROSSET, ATMO Auvergne

SOMMAIRE

1. RESUME.....	4
2. INTRODUCTION.....	6
2.1 OBJECTIFS	6
2.2 CONTRAINTES REGLEMENTAIRES	6
2.3 DONNEES UTILISEES	6
3. ETUDE DE LA VARIABILITE TEMPORELLE.....	7
3.1 PROFILS ANNUELS DE CONCENTRATION.....	7
3.2 PROFILS HEBDOMADAIRES.....	10
4. COMPARAISON DE DIFFERENTS PLANS D'ECHANTILLONNAGE TEMPOREL.....	12
4.1 ECHANTILLONNAGE SYSTEMATIQUE	13
4.1.1 Comparaison de trois stratégies sur les deuxième et quatrième trimestres	13
4.1.2 Test sur l'année entière.....	16
4.1.3 Conclusions.....	18
4.2 ECHANTILLONNAGE ALEATOIRE STRATIFIE	19
4.2.1 Comparaison de deux plans d'échantillonnage	19
4.2.2 Conclusions.....	21
5. ECHANTILLONNAGE DIMENSIONNE EN FONCTION DE LA VARIABILITE TEMPORELLE DES DONNEES.....	22
5.1.1 Elaboration d'un plan d'échantillonnage pour une précision d'estimation choisie	22
5.1.2 Evaluation de l'incertitude associée à un plan prédéfini	23
5.1.3 Echantillonnage systématique / échantillonnage aléatoire stratifié. Comparaison de différents plans	24
6. CONCLUSIONS - DISCUSSION	27
6.1 CONCLUSIONS SUR L'ECHANTILLONNAGE	27
6.2 REMARQUES SUR L'ESTIMATION DE LA MOYENNE ANNUELLE	27

1. RESUME

Au moment où se développe la surveillance des HAP dans l'air ambiant, il convient de s'interroger sur les stratégies d'échantillonnage les plus efficaces pour estimer une concentration moyenne annuelle avec une bonne précision. L'historique de données disponibles pour une telle analyse est à ce jour limité. De février à décembre 1998 cependant, des prélèvements quotidiens de HAP ont été réalisés en un site urbain de Paris (sous influence du trafic). AIRPARIF a mis les séries de mesures à la disposition du GT *Plans d'échantillonnage et reconstitution de données*. La présente note synthétise les résultats et recommandations d'échantillonnage issus de l'exploitation de ces données. Le cas du BaP est traité plus particulièrement.

Les profils annuels des concentrations de HAP montrent une saisonnalité marquée, avec des valeurs et une variabilité sensiblement plus élevées en hiver (mois de février, mars, novembre et décembre). Les profils hebdomadaires moyens présentent un maximum en semaine (le mercredi pour le BaP) et un minimum le samedi.

Différents plans d'échantillonnage sont simulés et comparés : des plans d'échantillonnage systématiques, qui consistent à effectuer des mesures à intervalles de temps réguliers ; des plans d'échantillonnage aléatoires stratifiés, qui consistent à découper l'année en grandes périodes, les strates, puis à y sélectionner aléatoirement un nombre approprié de mesures individuelles (*i.e.* de prélèvements journaliers) ou de groupes de mesures consécutives (les grappes).

D'après les statistiques d'erreur, et pour une même couverture annuelle, la réalisation de prélèvements isolés permet d'estimer plus précisément la moyenne annuelle que des prélèvements groupés par semaines. Pour ce qui est du BaP, un échantillonnage systématique d'un jour sur six, soit une soixantaine de mesures également réparties sur l'année, conduit à une erreur d'estimation inférieure à 12% pour cinq des six échantillons possibles et voisine de 33% pour le sixième échantillon. Un plan aléatoire stratifié constitué de quinze mesures individuelles dans chaque trimestre conduit à une précision comparable : 200 échantillons ont été simulés selon un tel plan ; sur ces 200 tirages, la moyenne, la médiane et le centile 95 de l'erreur relative sont respectivement de 15%, 13% et 35%. Ces valeurs deviennent respectivement 24%, 22% et 52% lorsque les mesures sont effectuées par grappes hebdomadaires.

Dans les tests décrits ci-dessus, l'année est découpée en strates de même taille, les trimestres, et le nombre de mesures par trimestre est identique. La théorie des sondages offre un cadre rigoureux pour concevoir d'autres plans d'échantillonnage dans lesquels les strates et/ou les échantillons extraits de ces dernières sont de taille différente. Les strates sont de préférence des parties de l'année relativement homogènes en ce qui concerne les valeurs et la variabilité des concentrations (pour le BaP on a considéré deux strates hivernales, une strate estivale et une strate de transition) ; le nombre, s'il n'est pas fixé par des contraintes externes, et la répartition des grappes entre les strates sont calculés en tenant compte de cette variabilité.

D'autre part, il s'avère que les concentrations de NO mesurées par une station urbaine voisine ont une variabilité temporelle très similaire à celle du BaP. Dans ce

cas, l'élaboration d'un plan d'échantillonnage du BaP peut s'appuyer sur les mesures de NO. A ce jour, les données disponibles ne permettent pas de généraliser ce résultat.

Du point de vue de la réglementation, la mesure fixe ou indicative des HAP n'admet qu'un échantillonnage régulièrement réparti sur l'année et une moyenne annuelle estimée comme la moyenne arithmétique des mesures. Lorsque le suivi des HAP peut s'effectuer par modélisation ou estimation objective, ou dans tout contexte autre que la surveillance réglementaire, l'utilisateur a le choix de la stratégie d'échantillonnage (systématique, aléatoire stratifié). Selon les données dont il dispose, il peut alors appliquer une ou plusieurs méthodes de reconstitution afin d'estimer une concentration moyenne annuelle et son incertitude associée.

2. INTRODUCTION

2.1 OBJECTIFS

En exploitant une série annuelle de prélèvements de HAP, le GT *Plans d'échantillonnage et reconstitution de données* a cherché à déterminer, indépendamment de toute contrainte technique, quels plans d'échantillonnage convenaient le mieux à l'estimation d'une concentration moyenne annuelle. Ce document synthétise les principaux résultats de cette étude.

2.2 CONTRAINTES REGLEMENTAIRES

Afin que les résultats de cette étude soient utilisés à bon escient, il convient de rappeler que :

- dans le cadre de la mesure fixe et indicative, l'échantillonnage des HAP doit couvrir respectivement 33% et 14% de l'année au minimum et se répartir régulièrement sur l'année et les jours de semaine. Seule la moyenne arithmétique est autorisée pour estimer la moyenne annuelle. Il est demandé aux AASQA de pratiquer ce type de suivi pour l'évaluation préliminaire actuellement en cours. L'incertitude sur la moyenne annuelle ne doit pas dépasser 50%.
- dans le cadre de l'estimation objective et de la modélisation, la réglementation n'impose pas de contrainte sur la stratégie d'échantillonnage. L'incertitude sur la moyenne annuelle ne doit pas dépasser 60% en modélisation.

2.3 DONNEES UTILISEES

Les données traitées ont été fournies par AIRPARIF. Elles sont issues d'une campagne de mesure des HAP conduite du 10 février au 31 décembre 1998 selon un **pas de temps journalier**.

Le site de mesure est un site urbain du 13^e arrondissement de Paris soumis à l'influence du trafic. **La série de mesure couvre au total 75% de l'année 1998** ; elle comprend deux grandes plages de données manquantes : du 1^{er} janvier au 9 février et du 1^{er} juillet au 27 juillet.

L'étude se limite aux HAP cités dans la quatrième Directive fille et pour lesquels il existe des données (Tableau 1):

- benzo(a)anthracène (B[a]A),
- benzo(b)fluoranthène (B[b]F),
- benzo(k)fluoranthène (B[k]F),
- benzo(a)pyrène (B[a]P),
- indénopyrène (IP).

La base ne contient pas de données de benzo(j)fluoranthène ni de données de dibenzo(ah)anthracène.

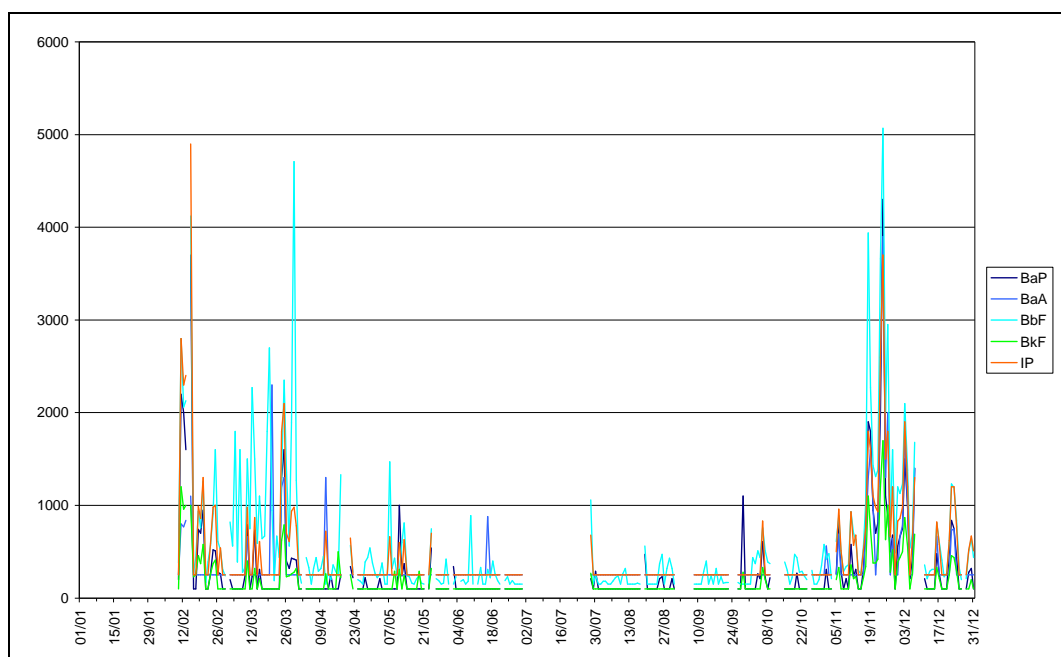
Tableau 1 – Caractéristiques des données de HAP étudiées

	Effectif (nombre de prélèvements journaliers valides)	Couverture annuelle [%]	Moyenne (m) [pg/m ³]	Ecart-type (σ) [pg/m ³]	Coefficient de variation = σ/m [%]
BaP	273	74.8	290.5	481.4	165.7
BaA	273	74.8	354.5	377.7	106.5
BbF	271	74.8	586.3	735.9	125.5
BkF	273	74.8	189.1	215.2	113.9
IP	273	74.8	458.9	535.5	116.7

3. ETUDE DE LA VARIABILITE TEMPORELLE

3.1 PROFILS ANNUELS DE CONCENTRATION

Les séries des mesures journalières (Figure 1) montrent des pics de concentration relativement élevés en février (début des courbes) et en novembre.



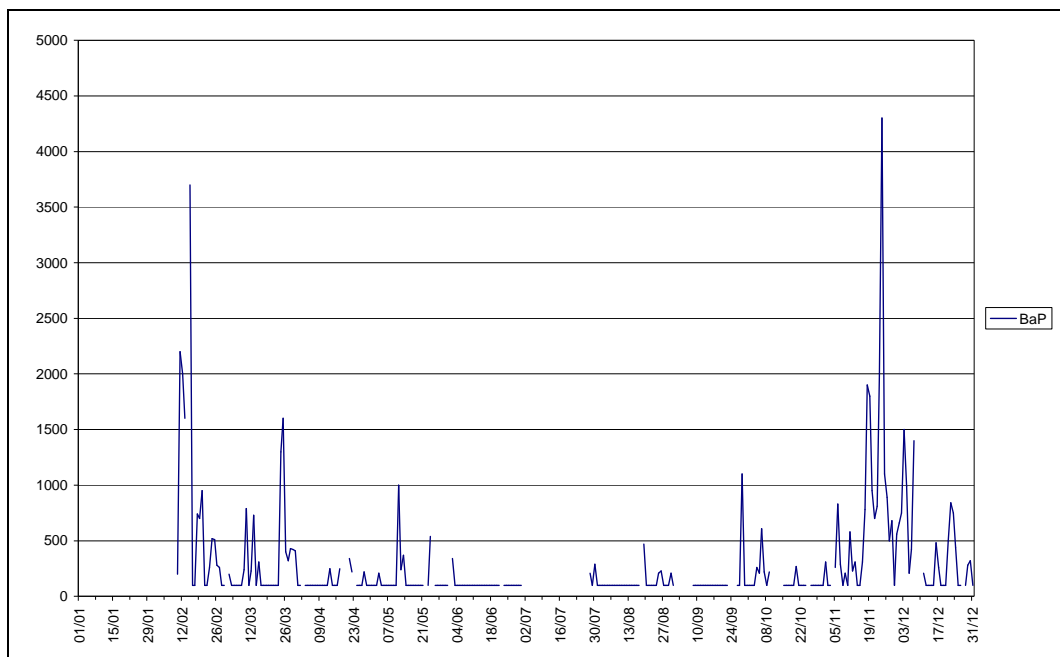
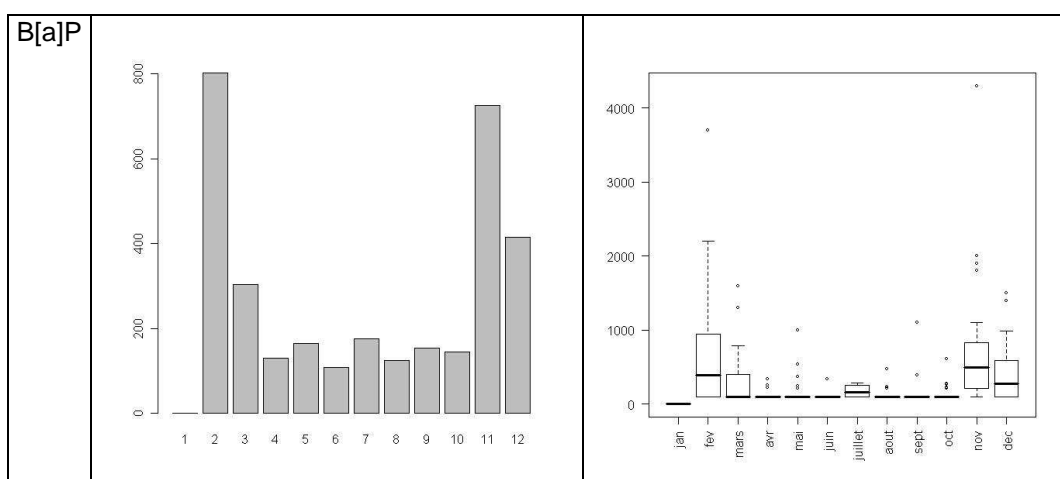


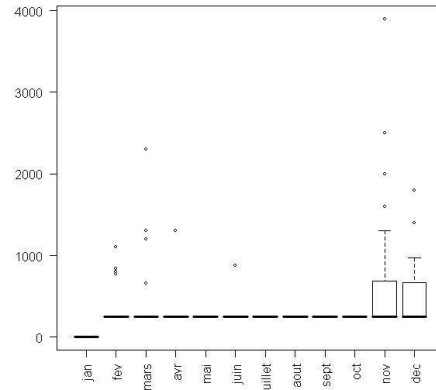
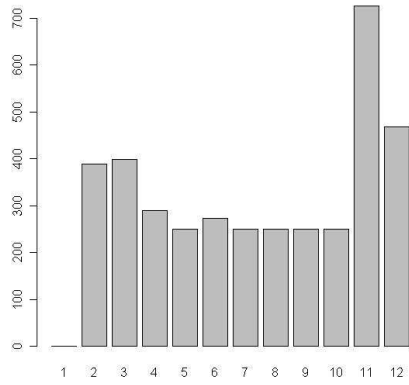
Figure 1 – Séries temporelles des concentrations de 5 HAP (figure du haut), du BaP (figure du bas), sur la période allant du 10 février au 31 décembre 1998. Concentrations en pg/m^3 .

Pour chacun des HAP, les figures suivantes représentent la moyenne et la dispersion mensuelle des données.

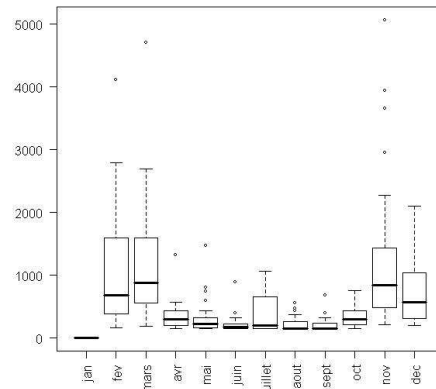
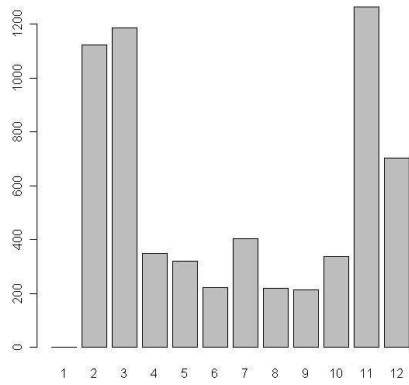
Elles font voir clairement la saisonnalité des concentrations, avec des **valeurs et une variabilité sensiblement plus élevées en hiver** (mois de février, mars, novembre et décembre ; on peut supposer qu'il en est de même en janvier bien qu'aucune mesure ne soit disponible pour l'attester). L'augmentation des concentrations de HAP en juillet n'est pas significative compte tenu du petit nombre de données valides (moins de 5 valeurs).



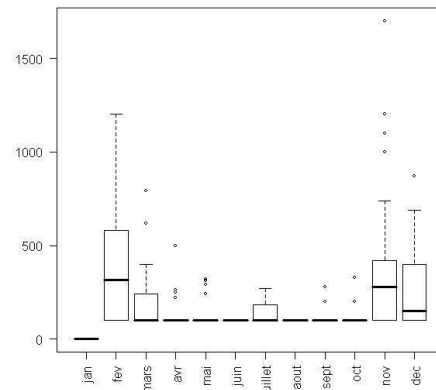
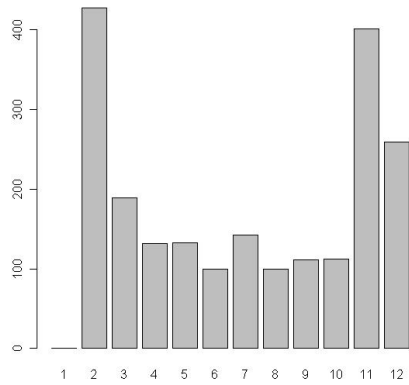
B[a]A



B[b]F



B[k]F



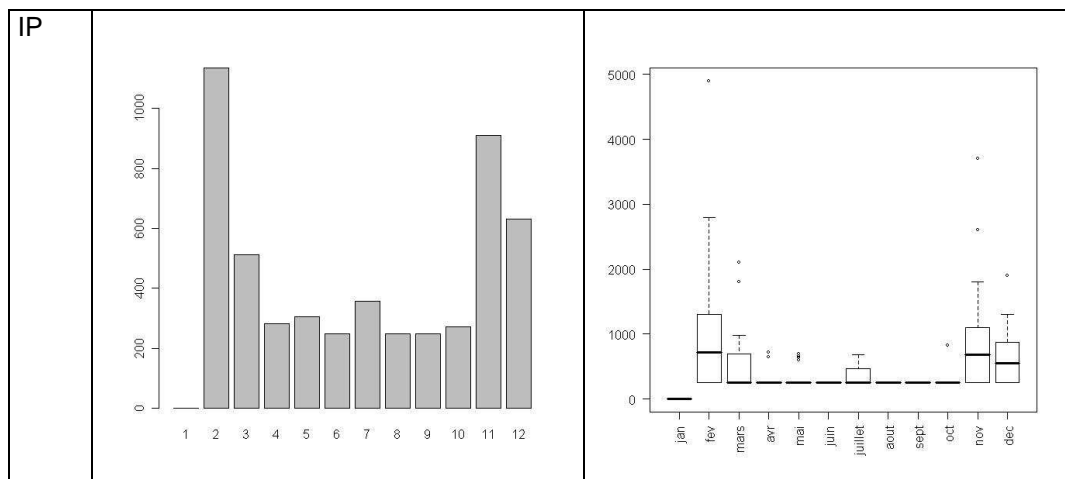


Figure 2 – Colonne de gauche : moyennes mensuelles [pg/m^3] des cinq HAP étudiés. Colonne de droite : représentation sous forme de boîtes à moustaches de la dispersion mensuelle des concentrations. Les traits horizontaux d'une boîte à moustaches matérialisent le minimum (hors valeurs atypiques), le centile 25, la médiane, le centile 75, le maximum (hors valeurs atypiques) ; les points isolés représentent les données atypiques.

3.2 PROFILS HEBDOMADAIRES

- B[k]F, B[a]P et IP montrent des profils hebdomadaires similaires, avec, en moyenne sur l'année, une concentration maximale le mercredi et minimale le samedi (Figure 3). Les valeurs mesurées le dimanche sont en moyenne supérieures à celles du samedi ; pour l'IP, elles sont même comparables aux concentrations du lundi.
- Le profil hebdomadaire du B[b]F ressemble à celui de l'IP à ceci près que le maximum de concentration s'observe le jeudi et que la décroissance du jeudi au vendredi est plus brutale.
- En ce qui concerne le B[a]A, les écarts entre jours ouvrés sont moindres. La concentration est maximale le mardi. Les valeurs mesurées le week end sont plus faibles qu'en semaine ; on ne note pas de différence entre le samedi et le dimanche.

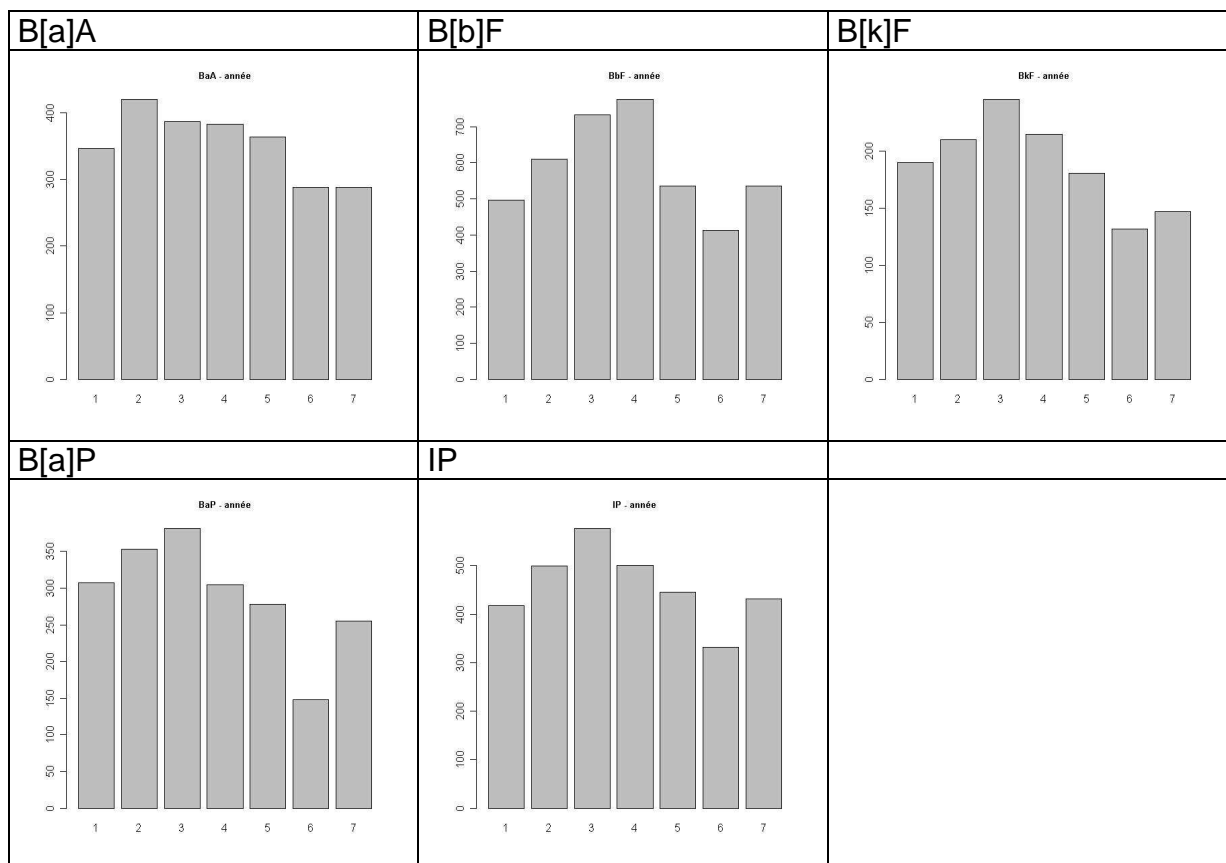


Figure 3 – Profils hebdomadaires de concentration pour les cinq HAP étudiés [pg/m³]. 1 : lundi...7 : dimanche.

4. COMPARAISON DE DIFFERENTS PLANS D'ÉCHANTILLONNAGE TEMPOREL

Dans cette partie, nous comparons plusieurs plans d'échantillonnage qui couvrent environ **14% de l'année** et comptent le même nombre de mesures dans chaque trimestre.

Nous examinons quelles stratégies d'échantillonnage conduisent aux estimations les plus précises.

Les principaux termes de statistique relatifs à l'échantillonnage sont décrits dans le *guide Elaboration d'un plan d'échantillonnage et reconstitution de données* rédigé par le GT. Sont rappelées ci-après quelques définitions utiles à la compréhension de cette note.

*Dans un échantillonnage **systematique**, les mesures sont réalisées à intervalles de temps réguliers tandis que dans un échantillonnage **aléatoire**, les dates de mesure sont sélectionnées au hasard.*

Un plan d'échantillonnage systematique ou aléatoire est dit :

- **par grappes**, si au lieu de mesures isolées, on effectue des groupes de mesures consécutives, par exemple des groupes de sept prélèvements successifs ;
- **stratifié**, si l'année est préalablement découpée en grandes périodes, les strates (de taille égale, comme les saisons, ou inégale), dans chacune desquelles est tiré un échantillon convenablement dimensionné.

Les plans d'échantillonnage systematiques suivants sont simulés sur les données d'AIIRPARIF :

- Mesures **un jour sur six** : à cause de longues périodes de données manquantes (janvier, partie de février, juillet), stratégie testée séparément sur le deuxième et le quatrième trimestres et, à titre indicatif, sur l'année entière ;
- Mesures **un jour sur trois dans chaque première moitié de trimestre** : pour la même raison que précédemment (peu de données disponibles dans les premières moitiés du premier et du troisième trimestre), stratégie testée uniquement pour le deuxième et le quatrième trimestres ;
- Mesures **un jour sur trois dans chaque seconde moitié de trimestre** : stratégie testée pour le deuxième et le quatrième trimestres et pour l'année entière.

En étendant la notion d'*égale répartition des mesures sur l'année* à celle, plus souple, d'*effectif égal par saison*, deux plans d'échantillonnage aléatoires stratifiés sont également considérés :

- **15 mesures journalières sélectionnées aléatoirement dans chaque trimestre**: cette stratégie a été testée en réalisant 200 simulations d'échantillonnage ;

- **deux grappes de 7 prélèvements journaliers sélectionnées aléatoirement dans chaque trimestre** : cette stratégie a été testée en réalisant 200 simulations d'échantillonnage.

Avertissement : A cause des données manquantes, certains échantillons peuvent comporter moins de mesures ou de grappes que ce qui a été planifié. Si individuellement, les résultats de certains tirages doivent être considérés prudemment, les statistiques sur les 200 simulations fournissent des informations exploitables.

Dans ces différents plans, qu'ils reposent sur un échantillonnage systématique ou aléatoire, les trimestres jouent le rôle de strates. Celles-ci étant toutes de même taille (trois mois), l'estimation de la concentration moyenne annuelle d'un HAP est la moyenne arithmétique des concentrations moyennes par strate.

Comme en outre chaque strate compte le même nombre de prélèvements (une quinzaine), cette estimation correspond (sans considération des données manquantes) à la moyenne arithmétique de l'ensemble des mesures.

4.1 ECHANTILLONNAGE SYSTEMATIQUE

4.1.1 Comparaison de trois stratégies sur les deuxième et quatrième trimestres

Pour chaque stratégie testée:

1. mesures 1 jour sur 6 pendant tout le trimestre
2. mesures 1 jour sur 3 pendant chaque 1^{ère} moitié de trimestre
3. mesures 1 jour sur 3 pendant chaque 2^{nde} moitié de trimestre

on effectue respectivement six, trois et trois simulations d'échantillonnage, décalées chacune d'une journée. A partir de chaque échantillon, on estime la concentration du trimestre.

Comme on pouvait s'y attendre au vu de la variabilité temporelle des concentrations (Figure 1, Figure 2), les estimations du quatrième trimestre sont sensibles à l'échantillonnage, en particulier si la stratégie d'échantillonnage consiste à effectuer une mesure tous les trois jours pendant une moitié de trimestre. Entre la première et la seconde moitié du trimestre, les estimations peuvent différer d'un facteur 3 à 4 à cause des pics de novembre (Tableau 4, Tableau 5). En ce qui concerne le BaP, l'écart relatif à la moyenne mesurée est compris entre 20% et 100%.

Les estimations relatives à la stratégie « une mesure tous les six jours » sont moins variables, bien qu'elles dépendent de la séquence de mesure. Un décalage d'une journée peut modifier sensiblement les résultats (Tableau 2 à Tableau 5). Pour le B[a]P au 4^e trimestre, les moyennes estimées s'écartent de 4 à 50% de la moyenne mesurée. Cette erreur relative est moindre pour le B[a]A : de 2 à 22%.

Tableau 2 – Simulation de différents plans d'échantillonnage sur le 2^e trimestre. Concentration moyenne de **BaP** [pg/m³] pour chaque échantillon tiré.

Fréquence d'échantillonnage (freq)	Numéro d'échantillon (n°)	Nombre de prélèvements valides (N)	Moyenne des prélèvements (m) [pg/m ³]	Erreur relative* (err.r.) [%]
Tous les jours		83	134.7	
Mesures un jour sur 6 pendant tout le trimestre				
1 jour/6	1	15	128.0	-5
1 jour/6	2	12	100.0	-25.8
1 jour/6	3	15	147.3	9.4
1 jour/6	4	14	108.6	-19.4
1 jour/6	5	13	180.8	34.2
1 jour/6	6	14	141.4	5
Mesures un jour sur 3 pendant la 1 ^{ère} moitié du trimestre				
1 jour/3	1	15	136.0	1
1 jour/3	2	12	187.5	39.2
1 jour/3	3	14	143.6	6.6
Mesures un jour sur 3 pendant la 2 ^{ème} moitié du trimestre				
1 jour/3	1	14	100.0	-25.8
1 jour/3	2	13	100.0	-25.8
1 jour/3	3	15	145.3	7.9

* L'erreur relative désigne l'écart relatif entre la moyenne de l'échantillon et la moyenne de la série entière.

Tableau 3 – Simulation de différents plans d'échantillonnage sur le 2^e trimestre. Résultats pour les **autres HAP**.

freq	n°	B[a]A			B[b]F			B[k]F			IP		
		N	moy	err.r.	N	moy	err.r.	N	moy	err.r.	N	moy	err.r.
tij		83	270.2		81	298.8		83	121.4		83	279.6	
Mesures un jour sur 6 pendant tout le trimestre													
1/6	1	15	250	-7.5	15	351.3	17.6	15	122	0.5	15	302.7	8.3
1/6	2	12	250	-7.5	12	236.7	-20.8	12	100	-17.6	12	250	-10.6
1/6	3	15	250	-7.5	15	275.3	-7.9	15	122.7	1.1	15	276.7	-1
1/6	4	14	250	-7.5	14	302.9	1.4	14	128.6	5.9	14	250	-10.6
1/6	5	13	379.2	40.3	12	305.8	2.3	13	137.7	13.4	13	313.1	12
1/6	6	14	250	-7.5	13	311.5	4.3	14	115.7	-4.7	14	282.1	0.9
Mesures un jour sur 3 pendant la 1 ^{ère} moitié du trimestre													
1/3	1	15	250	-7.5	15	407.3	36.3	15	136	12	15	302.7	8.3
1/3	2	12	337.5	24.9	12	347.5	16.3	12	140.8	16	12	318.3	13.8
1/3	3	14	250	-7.5	14	347.1	16.2	14	124.3	2.4	14	278.6	-0.4
Mesures un jour sur 3 pendant la 2 ^{ème} moitié du trimestre													
1/3	1	14	250	-7.5	14	242.9	-18.7	14	113.6	-6.4	14	250	-10.6
1/3	2	13	298.5	10.5	12	195	-34.7	13	100	-17.6	13	250	-10.6
1/3	3	15	250	-7.5	14	237.1	-20.6	15	14	-5.5	15	280	0.1

Tableau 4 – Simulation de différents plans d'échantillonnage sur le 4^e trimestre. Concentration moyenne de BaP [pg/m³] pour chaque échantillon tiré.

Fréquence d'échantillonnage (freq)	Numéro d'échantillon (n°)	Nombre de prélèvements valides (N)	Moyenne (m) [pg/m ³]	Erreur relative (err.r.) [%]
Tous les jours		84	434.6	
Mesures un jour sur 6 pendant tout le trimestre				
1 jour/6	1	15	654.7	50.6
1 jour/6	2	15	456	4.9
1 jour/6	3	14	272.1	-37.4
1 jour/6	4	13	367.7	-15.4
1 jour/6	5	13	378.5	-12.9
1 jour/6	6	14	452.9	4.2
Mesures un jour sur 3 pendant la 1 ^{ère} moitié du trimestre				
1 jour/3	1	14	189.3	-56.4
1 jour/3	2	13	168.5	-61.2
1 jour/3	3	14	190.7	-56.1
Mesures un jour sur 3 pendant la 2 ^{ème} moitié du trimestre				
1 jour/3	1	15	638	46.8
1 jour/3	2	14	534.3	22.9
1 jour/3	3	14	853.6	96.4

Tableau 5 – Simulation de différents plans d'échantillonnage sur le 4^e trimestre. Résultats pour les **autres HAP**.

freq	n°	B[a]A			B[b]F			B[k]F			IP		
		N	moy	err.r.	N	moy	err.r.	N	moy	err.r.	N	moy	err.r.
tj		84	486.7		84	779.6		84	261.0		84	611.4	
Mesures un jour sur 6 pendant tout le trimestre													
1/6	1	15	592.7	21.8	15	1100	41.1	15	352.7	35.1	15	779.3	27.5
1/6	2	15	535.3	10.0	15	766	-1.7	15	248.7	-4.7	15	597.3	-2.3
1/6	3	14	466.4	-4.2	14	644.3	-17.4	14	212.9	-18.4	14	473.6	-22.5
1/6	4	13	393.8	-19.1	13	628.5	-19.4	13	226.2	-13.3	13	541.5	-11.4
1/6	5	13	433.1	-11.0	13	681.5	-12.6	13	246.2	-5.7	13	586.2	-4.1
1/6	6	14	477.1	-2.0	14	817.9	4.9	14	270	3.4	14	672.9	10.1
Mesures un jour sur 3 pendant la 1 ^{ère} moitié du trimestre													
1/3	1	14	281.4	-42.2	14	397.1	-49.1	14	131.4	-49.7	14	323.6	-47.1
1/3	2	13	273.8	-43.7	13	369.2	-52.6	13	110	-57.9	13	303.1	-50.4
1/3	3	14	250	-48.6	14	407.1	-47.8	14	142.1	-45.6	14	357.9	-41.5
Mesures un jour sur 3 pendant la 2 ^{ème} moitié du trimestre													
1/3	1	15	673.3	38.3	15	1036.7	33.0	15	366.7	40.5	15	842.7	37.8
1/3	2	14	693.6	42.5	14	1055	35.3	14	340.7	30.5	14	788.6	29.0
1/3	3	14	719.3	47.8	14	1365	75.1	14	456.4	74.9	14	1014.3	65.9

4.1.2 Test sur l'année entière

Comme il a été mentionné, à cause des données manquantes, seule la troisième stratégie, à savoir un jour de mesure sur trois dans chaque seconde moitié de trimestre, peut être testée sur l'année entière de manière satisfaisante. Trois simulations d'échantillonnage sont réalisées, décalées d'une journée l'une par rapport à l'autre.

Les moyennes estimées à partir des deuxième et troisième séquences de mesure sont proches des moyennes annuelles mesurées. L'erreur relative, qui indique une légère surestimation de ces moyennes, est inférieure à 13%. En revanche, la première séquence conduit à des écarts plus importants (de 20 à 45% environ). Les surestimations constatées sont dues principalement à l'influence de deux valeurs extrêmes aux premier et dernier trimestres (quel que soit le HAP).

Tableau 6 – Estimation de la moyenne annuelle à partir de prélèvements journaliers réalisés un jour sur trois pendant chaque seconde moitié de trimestre

Séquence1

	Effectif	Couverture annuelle (%)	Moyenne de l'échantillon (pg/m ³)	Ecart-type	Biais (pg/m ³)	Erreur relative (%)	Moyenne de la série entière (pg/m ³)
BaP	54	14.8	422.4	794.1	131.9	45.4	290.5
BaA	54	14.8	425.4	617.3	70.9	20.0	354.5
BbF	54	14.8	815.7	1134.7	229.4	39.1	586.3
BkF	54	14.8	235.2	296.7	46.1	24.4	189.1
IP	54	14.8	588.7	830.2	129.8	28.3	458.9

Séquence2

	Effectif	Couverture annuelle (%)	Moyenne de l'échantillon (pg/m ³)	Ecart-type	Biais (pg/m ³)	Erreur relative (%)	Moyenne de la série entière (pg/m ³)
BaP	55	15.1	309.1	386.7	18.6	6.4	290.5
BaA	55	15.1	394.2	327.6	39.7	11.2	354.5
BbF	54	15.1	620.6	621.3	34.3	5.8	586.3
BkF	55	15.1	203.6	181.5	14.6	7.7	189.1
IP	55	15.1	487.3	414.5	28.4	6.2	458.9

Séquence3

	Effectif	Couverture annuelle (%)	Moyenne de l'échantillon (pg/m ³)	Ecart-type	Biais (pg/m ³)	Erreur relative (%)	Moyenne de la série entière (pg/m ³)
BaP	59	16.2	296.3	382.5	5.8	2.0	290.5
BaA	59	16.2	380.0	408.5	25.5	7.2	354.5
BbF	58	15.9	661.6	767.8	75.2	12.8	586.3
BkF	59	16.2	197.5	221.7	8.4	4.5	189.1
IP	59	16.2	474.9	487.4	16.0	3.5	458.9

A titre indicatif, nous présentons les résultats obtenus pour la stratégie « 1 jour sur 6 ».

Tableau 7 – Estimation de la moyenne annuelle à partir de prélèvements journaliers réalisés un jour sur six pendant toute l'année.

Séquence1

	Effectif	Couverture annuelle (%)	Moyenne de l'échantillon (pg/m ³)	Ecart-type	Biais (pg/m ³)	Erreur relative (%)	Moyenne de la série entière (pg/m ³)
BaP	46	12.6	287.8	386.6	-2.7	-0.9	290.5
BaA	46	12.6	346.5	381.5	-8.0	-2.3	354.5
BbF	46	12.6	507.2	492.1	-79.1	-13.5	586.3
BkF	46	12.6	184.1	184.1	-4.9	-2.6	189.1
IP	46	12.6	435.9	426.0	-23.0	-5.0	458.9

Séquence2

	Effectif	Couverture annuelle (%)	Moyenne de l'échantillon (pg/m ³)	Ecart-type	Biais (pg/m ³)	Erreur relative (%)	Moyenne de la série entière (pg/m ³)
BaP	41	11.2	268.5	321.3	-21.9	-7.6	290.5
BaA	41	11.2	322.4	183.5	-32.1	-9.0	354.5
BbF	41	11.2	506.3	500.5	-80.0	-13.6	586.3
BkF	41	11.2	188.5	184.9	-0.5	-0.3	189.1
IP	41	11.2	459.8	435.3	0.9	0.2	458.9

Séquence3

	Effectif	Couverture annuelle (%)	Moyenne de l'échantillon (pg/m ³)	Ecart-type	Biais (pg/m ³)	Erreur relative (%)	Moyenne de la série entière (pg/m ³)
BaP	47	12.9	274.0	339.4	-16.4	-5.7	290.5
BaA	47	12.9	326.4	342.3	-28.1	-7.9	354.5
BbF	47	12.9	539.1	601.5	-47.2	-8.0	586.3
BkF	47	12.9	177.9	189.1	-11.2	-5.9	189.1
IP	47	12.9	437.0	426.2	-21.9	-4.8	458.9

Séquence 4

	Effectif	Couverture annuelle (%)	Moyenne de l'échantillon (pg/m ³)	Ecart-type	Biais (pg/m ³)	Erreur relative (%)	Moyenne de la série entière (pg/m ³)
BaP	44	12.1	387.5	854.6	97.0	33.4	290.5
BaA	44	12.1	386.1	581.2	31.6	8.9	354.5
BbF	44	12.1	812.3	1218.8	226.0	38.5	586.3
BkF	44	12.1	219.8	311.2	30.7	16.3	189.1
IP	44	12.1	559.3	887.6	100.4	21.9	458.9

Séquence 5

	Effectif	Couverture annuelle (%)	Moyenne de l'échantillon (pg/m ³)	Ecart-type	Biais (pg/m ³)	Erreur relative (%)	Moyenne de la série entière (pg/m ³)
BaP	48	13.2	271.7	387.0	-18.8	-6.5	290.5
BaA	48	13.2	394.0	355.1	39.5	11.1	354.5
BbF	47	12.9	558.5	601.6	-27.8	-4.7	586.3
BkF	48	13.2	178.3	167.9	-10.7	-5.7	189.1
IP	48	13.2	427.9	382.4	-31.0	-6.8	458.9

Séquence 6

	Effectif	Couverture annuelle (%)	Moyenne de l'échantillon (pg/m ³)	Ecart-type	Biais (pg/m ³)	Erreur relative (%)	Moyenne de la série entière (pg/m ³)
BaP	47	12.9	257.0	414.9	-33.5	-11.5	290.5
BaA	47	12.9	348.5	324.9	-6.0	-1.7	354.5
BbF	46	12.6	597.2	748.2	10.9	1.9	586.3
BkF	47	12.9	187.7	234.0	-1.4	-0.7	189.1
IP	47	12.9	440.2	523.1	-18.7	-4.1	458.9

A l'exception de la quatrième séquence pour le BaP et le BbF, l'erreur sur la moyenne annuelle en révèle presque toujours une sous-estimation mais relativement faible.

4.1.3 Conclusions

- Dans les trimestres de plus grande variabilité interjournalière, l'erreur d'estimation commise avec un échantillonnage d'un jour sur six, uniformément réparti sur cette période, est généralement plus faible qu'avec un échantillonnage d'un jour sur trois, restreint à une moitié de trimestre. En effet, si comme l'automne 1998, le trimestre contient des plages de valeurs contrastées, cette seconde stratégie augmente le risque de biaiser l'estimation de la moyenne trimestrielle.
- En ce qui concerne le BaP, un échantillonnage d'un jour sur six sur l'année entière conduit à une erreur d'estimation sur la moyenne annuelle inférieure à 12% dans cinq cas sur six et de 33% pour l'un des échantillons.

4.2 ECHANTILLONNAGE ALEATOIRE STRATIFIE

4.2.1 Comparaison de deux plans d'échantillonnage

Deux plans d'échantillonnage sont évalués :

- 15 prélèvements individuels effectués aléatoirement dans chaque trimestre ;
- 2 grappes hebdomadaires de mesures sélectionnées aléatoirement dans chaque trimestre.

200 tirages d'échantillons sont simulés pour chacun. A chaque tirage et pour chaque HAP, on calcule la différence entre la moyenne de l'échantillon et la vraie moyenne annuelle. Les tableaux ci-après fournissent des statistiques sur ces écarts.

Signification des variables des tableaux :

moy_vraie : moyenne des mesures journalières sur l'année.

moymoyest, etmoyest : moyenne et écart-type sur les 200 simulations de la moyenne annuelle estimée.

minmoyest, P25moyest, P50moyest, P75moyest, maxmoyest : valeur minimale, centile 25, médiane, centile 75 et valeur maximale des moyennes annuelles estimées pour les 200 simulations.

biais : moyenne annuelle estimée – moyenne annuelle réelle

erreur relative : $100 \times (\text{moyenne annuelle estimée} - \text{moyenne annuelle réelle}) / \text{moyenne annuelle réelle}$

biaismoy, biaiset : moyenne et écart-type du biais sur les 200 simulations .

errelmoy, errelet : moyenne et écart-type de l'erreur relative sur les 200 simulations.

P25errel, P50errel, P75errel, P90errel, P95errel, maxerrel : autres statistiques de l'erreur relative sur les 200 simulations (centile 25, médiane, centile 75 et valeur maximale.)

4.2.1.1 Quinze prélèvements journaliers répartis aléatoirement dans chaque trimestre

Les moyennes annuelles estimées s'écartent en moyenne de **10 à 15%** environ de la moyenne annuelle mesurée (Tableau 8). Pour 95% des simulations, cet écart est inférieur à 36% (BaP), 22% (BaA), 34% (BbF), 26% (BkF), 29% (IP).

Tableau 8 – Estimation de la moyenne annuelle à partir de 15 prélèvements sélectionnés aléatoirement dans chaque trimestre. (valeurs de concentration en $\mu\text{g}/\text{m}^3$.)

			moy_vr aie	moy moye st	etmo yest	minm oyest	P25m oyest	P50m oyest	P75m oyest	max moye st	biais moy	biais et	errel moy	errel et	P25 errel	P50 errel	P75 errel	P90 errel	P95 errel	max errel
BaP	1	15x4	290.5	309.2	53.4	168.9	272.4	308.1	349.7	475.1	18.7	53.4	15.4	11.9	6.2	13.1	23.1	29.7	35.3	63.6
BaA	1	15x4	354.5	354.6	41.7	265.0	326.0	353.1	381.3	477.0	0.1	41.7	9.5	6.9	4.1	7.9	13.8	19.6	21.1	34.5
BbF	1	15x4	586.3	637.1	79.5	465.3	574.4	635.8	690.9	858.5	50.8	79.5	12.8	9.8	4.9	10.9	18.6	25.3	33.1	46.4
BkF	1	15x4	189.0	199.6	23.7	149.3	182.9	196.2	217.1	264.7	10.6	23.7	10.9	8.3	3.5	9.8	16.2	23.5	25.9	40.0
IP	1	15x4	458.9	484.1	64.2	343.3	437.8	481.0	523.5	747.2	25.2	64.2	11.7	9.4	4.7	9.7	16.0	23.8	28.3	62.8

4.2.1.2 Deux grappes de sept prélèvements journaliers consécutifs tirées aléatoirement dans chaque trimestre

Les estimations sont moins précises qu'avec 15 mesures journalières par trimestre. Les concentrations moyennes annuelles estimées s'écartent en moyenne de **15% à 25%** des valeurs mesurées. Par rapport au cas précédent, les centiles et l'écart-type des erreurs relatives ainsi que l'écart-type du biais augmentent.

Tableau 9 – Estimation de la moyenne annuelle à partir de 2 grappes hebdomadaires sélectionnées aléatoirement dans chaque trimestre (valeurs de concentration en $\mu\text{g}/\text{m}^3$).

			moy_ vraie	moy moye st	etmo yest	minm oyest	P25m oyest	P50m oyest	P75m oyest	maxmo yest	biais moy	biais et	errel moy	errel et	P25 errel	P50 errel	P75 errel	P90 errel	P95 errel	max errel
BaP	7	2x4	290.5	286.1	84.1	127.5	219.1	279.4	350.2	536.1	-4.4	84.1	24.0	16.1	10.4	22.3	34.1	44.3	51.9	84.6
BaA	7	2x4	354.5	353.0	70.3	256.1	303.3	334.8	379.8	593.9	-1.5	70.3	15.5	12.3	6.4	12.9	21.9	30.1	41.2	67.5
BbF	7	2x4	586.3	622.9	132.1	404.5	523.0	595.6	702.3	1070.4	36.6	132.1	17.6	15.3	6.1	12.8	24.2	40.0	50.9	82.6
BkF	7	2x4	189.0	186.7	43.4	118.6	154.4	176.7	216.8	306.0	-2.3	43.4	18.7	13.2	8.0	17.1	25.8	36.2	41.7	61.8
IP	7	2x4	458.9	467.6	105.7	258.3	384.5	457.9	543.2	765.3	8.7	105.7	19.2	12.8	9.4	17.5	26.3	35.4	43.9	66.8

4.2.2 Conclusions

D'après les statistiques d'erreur, et pour un même taux d'échantillonnage, la réalisation de prélèvements isolés fournit une estimation plus précise de la moyenne annuelle que des prélèvements groupés par semaines.

S'agissant du BaP, la moyenne et le centile 95 des erreurs relatives sont respectivement de 15% et 35% pour 15 prélèvements aléatoires par trimestre, et de 24% et 52% pour 2 grappes de 7 jours par trimestre.

5. ECHANTILLONNAGE DIMENSIONNE EN FONCTION DE LA VARIABILITE TEMPORELLE DES DONNEES

L'analyse des séries temporelles a révélé un net contraste entre la saison hivernale (concentrations et variabilité plus élevées) et la saison estivale. Celui-ci n'est pas pris en compte lorsqu'on tire un nombre égal de mesures par saison.

Or, comme les concentrations sont peu variables en été, la précision d'estimation de la concentration estivale ne devrait pas se dégrader si le nombre de prélèvements réalisés pendant cette saison est réduit dans une certaine limite. A l'inverse, en augmentant le nombre de mesures en hiver, on peut espérer obtenir une meilleure représentation de la variabilité des données sur cette saison et estimer plus précisément la concentration hivernale.

Cette démarche, qui consiste à adapter le dimensionnement de l'échantillonnage à la variabilité temporelle des données est l'approche développée par le GT dans le guide *Elaboration d'un plans d'échantillonnage et reconstitution de données* (et brièvement appelé *guide* dans la suite).

Elle n'est pas applicable dans le cadre de la mesure fixe ou indicative, puisque la moyenne arithmétique des données, seule formule autorisée dans ce contexte pour le calcul de la moyenne annuelle, risque alors de fournir une estimation biaisée (si le nombre de mesures est supérieur en hiver, la moyenne annuelle risque d'être surestimée).

En revanche, complétée par l'utilisation de méthodes de reconstitution, elle constitue une stratégie pertinente pour l'estimation objective ou la modélisation.

Une illustration en est présentée ci-après.

5.1.1 Elaboration d'un plan d'échantillonnage pour une précision d'estimation choisie

Dans cet exemple d'application, nous cherchons à atteindre une **précision d'estimation sur la moyenne annuelle de 25%**.

L'examen du profil annuel et de la variabilité mensuelle des HAP, et notamment du BaP, nous conduit à découper l'année en quatre strates temporelles :

1. une première strate hivernale : (janvier)-février-mars (concentrations et variabilité relativement grandes) ;
2. une strate regroupant les périodes de transition : avril-mai et septembre-octobre (concentrations relativement basses avec quelques) ;
3. une strate estivale : juin-juillet-août (concentrations et variabilité faibles) ;
4. une seconde strate hivernale (concentrations et variabilité relativement grandes).

Trois tailles de grappes sont considérées :

- des grappes de taille unitaire, autrement dit des prélèvements journaliers isolés;
- des grappes de trois prélèvements journaliers consécutifs ;
- des grappes de sept prélèvements journaliers consécutifs.

Pour atteindre la précision souhaitée, **l'application de la théorie des sondages** (voir le chapitre 1 du guide) fournit le nombre total de grappes à tirer dans l'année et la répartition optimale des grappes entre les strates. Elle aboutit aux dimensionnements suivants (on se limite pour cet exemple au BaP et au BbF) :

Tableau 10 – Nombre de grappes à prélever dans chaque strate pour une précision attendue de 25%

Précision fixée : 25%		Nombre de grappes par strate	Nombre total de grappes à prélever dans l'année	Nombre total de jours de mesure dans l'année
Mesures individuelles	B[a]P	32 9 3 21	65	65
	B[b]F	19 5 3 13	40	40
Grappes de 3 jours	B[a]P	15 4 1 10	30	90
	B[b]F	7 3 1 6	17	51
Grappes de 7 jours	B[a]P	8 1 1 4	14	98
	B[b]F	3 1 1 4	9	63

Plus les grappes sont longues, plus la couverture temporelle requise est grande. D'autre part, pour une même précision souhaitée, le plan d'échantillonnage du BaP est plus contraignant que celui du BbF.

5.1.2 Evaluation de l'incertitude associée à un plan prédéfini

Reprenons l'exemple précédent. Supposons que des considérations pratiques conduisent au choix de grappes de 3 jours mais que les ressources disponibles ne permettent pas la réalisation du plan 15, 4, 1, 10 proposé pour le BaP. Supposons alors que le compromis trouvé entre les contraintes de ressources et l'intérêt d'augmenter le nombre de mesures en hiver aboutisse au plan 7, 4, 3, 7. D'après les formules de la théorie des sondages, l'incertitude d'estimation que l'on peut attendre de ce nouveau plan d'échantillonnage s'élève à 36% pour le BaP et 22% pour le BbF.

En ce qui concerne le BaP, cette incertitude est supérieure aux 25% fixés initialement, ce qui est logique vu qu'on a réduit le nombre de grappes et modifié leur répartition. S'agissant du BbF, elle diminue puisque dans chaque strate, le plan d'échantillonnage comprend plus de grappes qu'il n'était requis pour ce polluant.

5.1.3 Echantillonnage systématique / échantillonnage aléatoire stratifié. Comparaison de différents plans

Dans cette partie, nous comparons plusieurs plans d'échantillonnage qui totalisent chacun une soixantaine de jours de mesure.

Les plans aléatoires stratifiés ont été dimensionnés spécifiquement pour le **BaP** au moyen de la théorie des sondages et réajustés le cas échéant pour tenir compte de certaines contraintes (cf. commentaires du Tableau 12). Pour les cinq HAP, nous avons calculé l'incertitude d'estimation de la concentration moyenne annuelle associée a priori à chacun de ces plans. Les résultats relatifs au BaP et au BbF sont ici mentionnés.

A titre d'information, ce même calcul a été effectué pour le NO et le NO₂ mesurés en un site fixe proche d'AIRPARIF (station Paris 13^e). La variabilité temporelle des concentrations de NO se révèle tout à fait comparable à celle du BaP alors que le NO₂ est nettement moins variable (Tableau 11, Figure 4).

Tableau 11 – Caractéristiques des données de BaP étudiées et des données de NOx issues d'une station fixe AIRPARIF proche du site de prélèvement (Paris 13^e).

	Effectif	Couverture annuelle (%)	Moyenne annuelle	Ecart-type	Coefficient de variation (%)
BaP [$\mu\text{g}/\text{m}^3$]	273	74.8	290.5	481.4	165.7
NO [$\mu\text{g}/\text{m}^3$]	358	98.0	24.5	52.6	165.1
NO ₂ [$\mu\text{g}/\text{m}^3$]	358	98.0	40.5	20.5	38.9

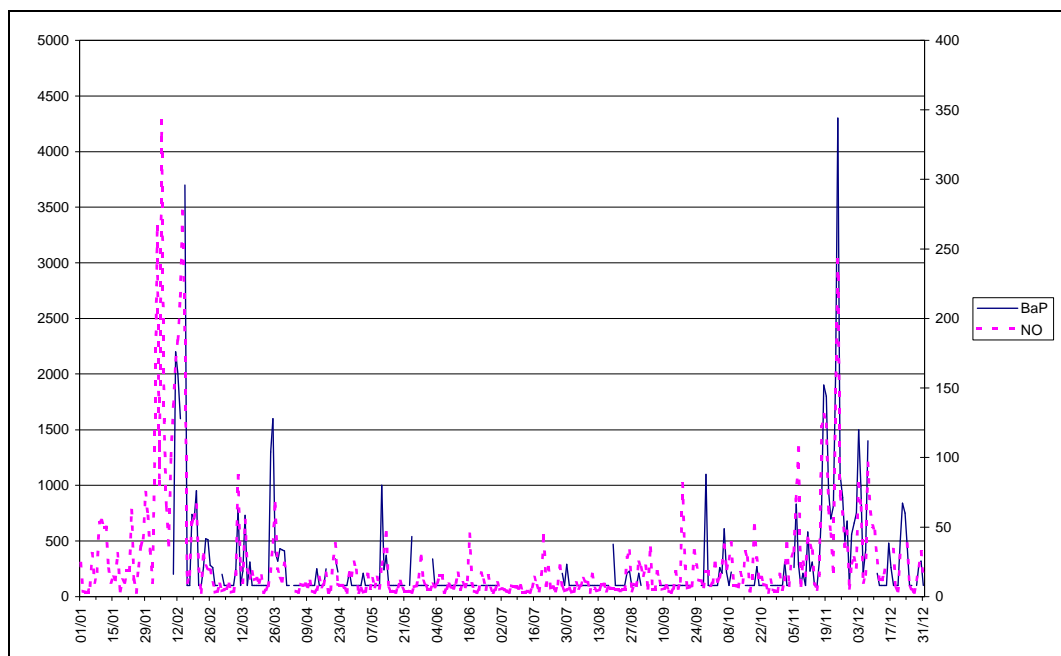


Figure 4 – Séries temporelles des concentrations de BaP [$\mu\text{g}/\text{m}^3$] et de NO [$\mu\text{g}/\text{m}^3$] sur la période allant du 10 février au 31 décembre 1998.

Tableau 12 – Evaluation à priori de l'incertitude associée à des plans d'échantillonnage systématiques ou aléatoires stratifiés qui totalisent une soixantaine de jours de mesure.

Type d'échantillonnage	Stratification	Taille des grappes	Nombre de grappes par strate	Nombre total de jours de mesure	Couverture annuelle	Evaluation à priori de l'incertitude
Systématique	/	1 jour	60 (1 jour sur 6)	60	16,4%	BaP : 28% BbF : 35%** NO : 32% NO2 : 8,7%
Aléatoire stratifié	1.jan-mars 2.avr.-mai + sept.-oct., 3.juin-août, 4.nov.-déc	1 jour	29, 9, 3, 19	60	16,4%	BaP : 26% BbF : 20% NO : 26% NO2 : 14%
Aléatoire stratifié	1.jan-mars 2.avr.-mai + sept.-oct., 3.juin-août, 4.nov.-déc	3 jours	(10,2,1,7)* 9,2,2,7	60	16,4%	BaP : 34% BbF : 23% NO : 37% NO2 : 13%
Aléatoire stratifié	1.jan-mars 2.avr.-mai + sept.-oct., 3.juin-août, 4.nov.-déc	7 jours	(5,1,(0.3),3)* 3,2,2,2	63	17,2%	BaP : 49% BbF : 29% NO : 49% NO2 : 12%

*N.B. : La répartition théorique est légèrement modifiée afin d'avoir un minimum de 2 à 3 grappes par strate. En effet, la variance théorique, qui permet de déterminer la répartition optimale des grappes, est calculée sur la série de référence complète. En revanche, après que le plan d'échantillonnage aura été mis en œuvre, la variance qui lui correspondra réellement ne pourra qu'être qu'estimée sur les données recueillies. Pour que l'estimation de la variabilité dans chaque strate soit possible, il est préférable que chacune d'elles contienne plus d'une grappe.

**L'incertitude associée à un plan d'échantillonnage systématique se déduit de la variabilité, évaluée sur la série de référence, entre les différentes séquences d'échantillonnage possibles. La valeur obtenue pour le BbF, supérieure à celle du BaP, est due au fait que l'une des 6 séquences inclut les deux mesures de BbF les plus élevées, conduisant à une nette surestimation de la moyenne.

En ce qui concerne le BaP, **plus la taille de grappe augmente** et donc plus le nombre de grappes par strate diminue, **plus grande est l'incertitude associée au plan d'échantillonnage** (Tableau 12). Cette observation rejoint les résultats des simulations (Tableau 8 / Tableau 9). Les plans d'échantillonnage aléatoires stratifiés ayant été spécifiquement dimensionnés pour le BaP, cette remarque ne se vérifie pas nécessairement pour les autres composés. On constate cependant que les incertitudes calculées pour le BaP et pour le NO sont proches, ce qui s'accorde avec la variabilité temporelle similaire de ces deux polluants.

L'incertitude calculée pour le NO₂, polluant dont la concentration est sensiblement moins variable, est logiquement plus faible.

Pour des grappes unitaires, les incertitudes attendues d'un plan systématique et d'un plan aléatoire stratifié sont comparables. C'est bien ce que suggéraient les erreurs relatives obtenues dans les simulations d'échantillonnage (4.1.2 et 4.2.1.1). Sachant que les incertitudes évaluées à priori ont une valeur indicative (voir la Figure 5 ci-après et les commentaires correspondants), les chiffres du Tableau 12 ne permettent pas de recommander une stratégie plutôt qu'une autre.

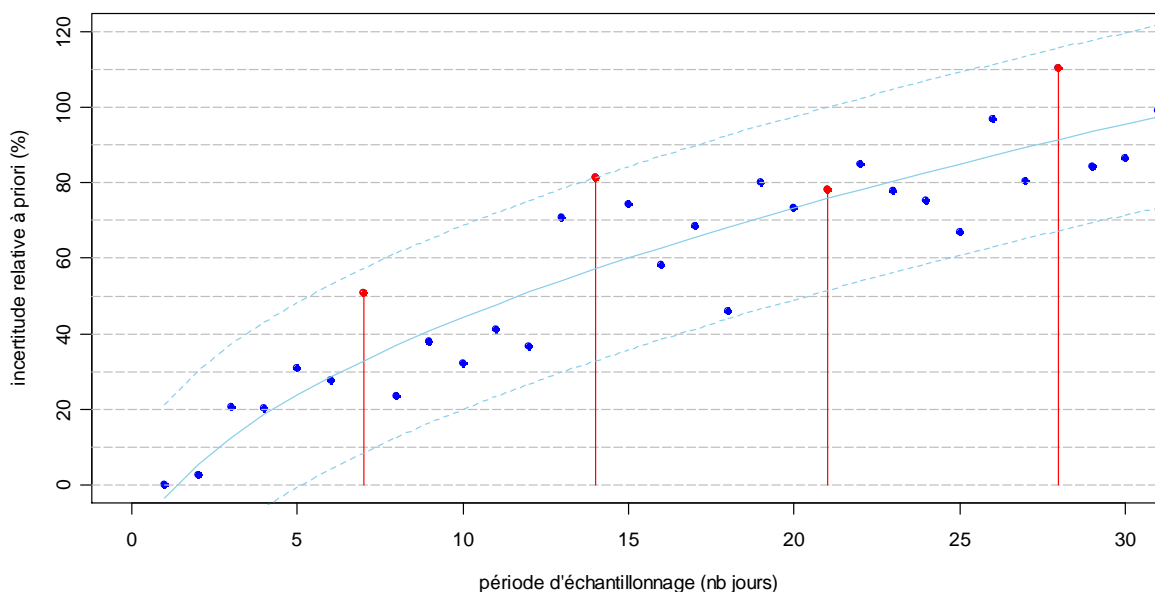


Figure 5 – Evaluation à priori de l'incertitude associée à un plan d'échantillonnage systématique. Evolution de l'incertitude en fonction de la période d'échantillonnage (période de 6 jours = une mesure tous les six jours). Courbe en trait continu bleu : tendance (résultat de la régression linéaire entre l'incertitude à priori et la racine carrée de la période d'échantillonnage).

Courbes en traits pointillés : intervalle de confiance à 95% autour de la courbe de régression.

Points rouges : valeurs d'incertitude correspondant à des périodes multiples de 7 jours.

La figure ci-dessus représente l'incertitude associée à priori à un plan d'échantillonnage systématique. Celle-ci a tendance à croître avec la période d'échantillonnage. Toutefois, la variabilité associée au calcul d'incertitude fait qu'une petite augmentation de la période d'échantillonnage peut s'accompagner d'une légère diminution de l'incertitude. D'autre part, il convient de noter l'augmentation quasi systématique de l'incertitude pour des périodes d'échantillonnage multiples de sept jours. Dans ce cas, l'incertitude exprime la variabilité entre les jours de semaine : si l'on considère les profils hebdomadaires, il est évident que la moyenne d'un échantillon constitué uniquement de lundis sera sensiblement différente de la moyenne d'un échantillon de même taille constitué uniquement de samedis.

6. CONCLUSIONS - DISCUSSION

6.1 CONCLUSIONS SUR L'ÉCHANTILLONNAGE

La base de données de HAP est incomplète mais offre un bon aperçu de la variabilité temporelle des concentrations pour les différents composés. Les nombres de grappes et les incertitudes calculés sont cohérents avec cette variabilité. Ainsi, bien que les données manquantes puissent affecter les valeurs numériques obtenues, elles ne semblent pas mettre en cause les indications générales que l'on peut tirer des résultats.

Pour **ce site urbain de mesure des HAP soumis à l'influence du trafic**, l'analyse des données fait ressortir les points suivants :

- Pour une meilleure précision d'estimation de la moyenne annuelle, des grappes courtes (1 à 3 jours) et nombreuses sont préférables à des grappes plus longues (hebdomadaires) et en effectif réduit.
- Dans le cas de prélèvements isolés (grappes de 1 jour), un échantillonnage systématique d'un jour sur six et un échantillonnage aléatoire stratifié conduisent à des précisions d'estimation comparables.
- Un échantillonnage d'un jour sur six uniformément réparti sur l'année est préférable à un échantillonnage d'un jour sur trois par moitié de trimestre. Certains trimestres sont en effet assez contrastés du point de vue de la variabilité des concentrations, notamment le quatrième trimestre qui oppose le mois d'octobre (faible variabilité) à novembre et décembre. Avec un échantillonnage d'un jour sur trois par moitié de trimestre, le risque de biaiser l'estimation de la moyenne annuelle est accru.
- Du fait que ces deux polluants présentent une variabilité temporelle similaire, l'élaboration d'un plan d'échantillonnage du BaP peut s'appuyer sur une série de NO issue d'un site proche et de mêmes caractéristiques.

Il conviendrait d'examiner si cette corrélation entre le BaP et le NO se vérifie pour des sites de typologie différente et, si ce n'est pas le cas, déterminer quelles autres variables de pollution pourraient être éventuellement exploitées.

Notons que toute connaissance supplémentaire sur la variabilité temporelle des HAP et l'existence de variables corrélées avec ces HAP constituera une information utile pour étendre ou non à d'autres sites les conclusions précédentes.

6.2 REMARQUES SUR L'ESTIMATION DE LA MOYENNE ANNUELLE

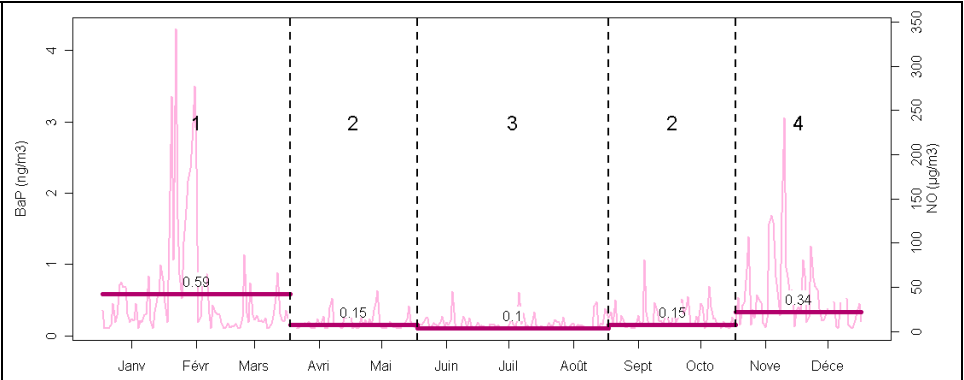
Dans le cas d'un échantillonnage systématique, l'estimation de la moyenne annuelle est la moyenne arithmétique des mesures. Il en va de même dans celui d'un plan d'échantillonnage aléatoire stratifié comportant des **strates de même taille et un**

nombre identique de grappes par strate : dans un tel plan, toutes les données ont le même poids.

Dans le cas d'un plan d'échantillonnage aléatoire stratifié, avec des **strates de taille différente et/ou des nombres différents de grappes par strate**, la **moyenne arithmétique est une estimation biaisée de la moyenne annuelle**. Il convient alors de mettre en œuvre une ou plusieurs **techniques de reconstitution**, comme celles qui sont décrites dans le guide publié par le GT. L'application de la théorie des sondages est illustrée dans les figures suivantes (Figure 6).

<p>Plan d'échantillonnage aléatoire stratifié</p>	<p>4 strates : 1. jan-mars 2. avr.-mai + sept.-oct., 3. juin-août, 4. nov.-déc</p> <p>21 grappes de 3 jours réparties dans les strates de la façon suivante : 7 - 4 - 3 - 7</p>
<p>Représentation des données d'échantillonnage</p>	
<p>Représentation des moyennes par grappe</p>	<p>Estimation de la moyenne annuelle et de son incertitude associée sans tenir compte de la stratification :</p> <p>Moyenne estimée = 0.35 ng/m3 +/- 49% Moyenne réelle = 0.29 ng/m3</p>
<p>Prise en compte des strates</p>	<p>Estimation de la moyenne annuelle et de son incertitude associée en tenant compte de la stratification :</p> <p>Moyenne estimée = 0.28 ng/m3 +/- 42% Moyenne réelle = 0.29 ng/m3</p>

Prise en compte des strates et d'une série auxiliaire (ici, la série de NO)



Estimation de la moyenne annuelle et de son incertitude associée en tenant compte de la stratification et d'une série auxiliaire :

Moyenne estimée = 0.30 ng/m³ +/- 19%

Moyenne réelle = 0.29 ng/m³

Figure 6 – Application de la théorie des sondages à l'estimation de la moyenne annuelle du BaP (exemple extrait des résultats présentés lors du séminaire LCSQA Surveillance dans l'air ambiant des HAP & des métaux, Paris, 30 novembre 2007).