



## Laboratoire Central de Surveillance de la Qualité de l'Air



**Méthodes de reconstitution temporelle de moyennes et de nombres de dépassements de seuils à partir de données de campagne**

JEAN-LUC HOUDRET ET LAURE MALHERBE

Novembre 2005





Ministère de l'Ecologie  
et du Développement Durable

## PREAMBULE

### **Le Laboratoire Central de Surveillance de la Qualité de l'Air**

**Le Laboratoire Central de Surveillance de la Qualité de l'Air est constitué de laboratoires de l'Ecole des Mines de Douai, de l'INERIS et du LNE. Il mène depuis 1991 des études et des recherches finalisées à la demande du Ministère chargé de l'environnement, sous la coordination technique de l'ADEME et en concertation avec les Associations Agréées de Surveillance de la Qualité de l'Air (AASQA). Ces travaux en matière de pollution atmosphérique supportés financièrement par la Direction des Préventions des Pollutions et des Risques du Ministère de l'Ecologie et du Développement Durable sont réalisés avec le souci constant d'améliorer le dispositif de surveillance de la qualité de l'air en France en apportant un appui scientifique et technique aux AASQA.**

**L'objectif principal du LCSQA est de participer à l'amélioration de la qualité des mesures effectuées dans l'air ambiant, depuis le prélèvement des échantillons jusqu'au traitement des données issues des mesures. Cette action est menée dans le cadre des réglementations nationales et européennes mais aussi dans un cadre plus prospectif destiné à fournir aux AASQA de nouveaux outils permettant d'anticiper les évolutions futures.**

**ECOLE DES MINES DE DOUAI**

**DEPARTEMENT CHIMIE ET ENVIRONNEMENT**

**METHODES DE RECONSTITUTION  
TEMPORELLE DE MOYENNES ET DE  
NOMBRES DE DEPASSEMENTS DE  
SEUILS PREDIFINIS A PARTIR  
DE DONNEES DE CAMPAGNE**

**Jean Luc HOUDRET**

**Novembre 2005**



Ministère de l'Ecologie  
et du Développement Durable

## *Méthodes de reconstitution temporelle de moyennes et de nombres de dépassements de seuils à partir de données de campagnes*

Laboratoire Central de Surveillance  
de la Qualité de l'Air

### **Thème : Traitements numériques**

**Convention 05000051**

**Financée par la Direction des Préventions des Pollutions et des Risques  
(DPPR)**

	<b>Rédaction</b>	<b>Vérification</b>	<b>Approbation</b>
<b>NOM</b>	Laure MALHERBE	Laurence ROUÏL	M.RAMEL
<b>Qualité</b>	Ingénieur d'étude Direction des Risques Chroniques	Responsable de l'unité MECO Direction des Risques Chroniques	Responsable LCSQA/INERIS Direction des Risques Chroniques
<b>Visa</b>			

**Ce document comporte 97 pages (hors couverture et annexes).**

**Novembre 2005**

**Jean-Luc Houdret et Laure Malherbe**

# TABLE DES MATIERES

<b><u>1. RESUMÉ</u></b> .....	<b>8</b>
<b><u>2. INTRODUCTION</u></b> .....	<b>10</b>
<u>2.1 Contexte</u> .....	10
<u>2.2 Principe des deux méthodes étudiées</u> .....	11
<u>2.2.1 Méthode de reconstitution « ISO »</u> .....	11
<u>2.2.2 Méthode des plans de sondages</u> .....	12
<u>2.3 Plan</u> .....	12
<b><u>3. PARTIE I : METHODE ISO</u></b> .....	<b>13</b>
<u>3.1 Introduction</u> .....	14
<u>3.2 Principe</u> .....	15
<u>3.3 Les paramètres de stratification pris en compte</u> .....	16
<u>3.4 Influence des paramètres météorologiques</u> .....	17
<u>3.4.1 Influence de la vitesse du vent sur NO et NO<sub>2</sub> à Toulouse</u> .....	17
<u>3.4.2 Influence de la température</u> .....	17
<u>3.4.3 Influence de la température sur NO et NO<sub>2</sub> à Strasbourg</u> .....	18
<u>3.4.4 Influence de la température sur NO et NO<sub>2</sub> à Annecy</u> .....	18
<u>3.4.5 Influence de la température sur O<sub>3</sub> au HAVRE</u> .....	19
<u>3.4.6 Influence de la température sur les PM2.5 et PM10</u> .....	19
<u>3.5 Influence des paramètres temporels</u> .....	20
<u>3.5.1 Principe</u> .....	20
<u>3.5.2 Influence de l'heure de la journée</u> .....	20
<u>3.5.3 Influence de l'heure de la journée sur NO et NO<sub>2</sub> à Chambéry</u> .....	21
<u>3.5.4 Influence de l'heure de la journée sur O<sub>3</sub> à Toulouse</u> .....	21
<u>3.6 Influence du jour de la semaine</u> .....	22
<u>3.6.1 Influence du jour de la semaine sur NO et NO<sub>2</sub> à Strasbourg</u> .....	22
<u>3.7 Application de la norme ISO 9359</u> .....	23
<u>3.8 Bilan complet des influences des paramètres pour les diverses villes</u> .....	23
<u>3.9 Stratification des séquences horaires de mesure</u> .....	24

3.10	<a href="#">Calculs des moyennes et des nombres de dépassement par strate</a>	25
3.11	<a href="#">Durée des campagnes</a>	26
3.12	<a href="#">Conclusions et perspectives</a>	<b>Erreur! Signet non défini.</b>
3.13	<a href="#">Références</a>	28
<b>4.</b>	<b><a href="#">PARTIE II : METHODE PLANS DE SONDAGE</a></b>	<b>29</b>
4.1	<a href="#">objectifs, théorie, principe de mise en oeuvre</a>	31
4.2	<a href="#">Objectifs de la méthode</a>	31
4.3	<a href="#">Données requises</a>	32
4.4	<a href="#">Eléments théoriques</a>	32
4.4.1	<a href="#">Définition d'un plan de sondage</a>	33
4.4.2	<a href="#">Estimation d'une moyenne et d'un taux de dépassement de seuil</a>	35
4.5	<a href="#">Mise en oeuvre</a>	37
4.6	<a href="#">Vérification des programmes</a>	38
4.7	<a href="#">Principe des simulations</a>	38
4.8	<a href="#">Résultats</a>	39
4.8.1	<a href="#">Grappes d'une heure</a>	41
4.8.1.1	<a href="#">Estimation de la moyenne annuelle</a>	41
4.8.1.2	<a href="#">Estimation du taux de dépassement de seuil</a>	41
4.8.2	<a href="#">Grappes de 24 heures</a>	41
4.8.2.1	<a href="#">Estimation de la moyenne annuelle</a>	41
4.8.2.2	<a href="#">Estimation du taux de dépassement de seuil</a>	42
4.8.3	<a href="#">Grappes d'une semaine</a>	42
4.8.3.1	<a href="#">Estimation de la moyenne annuelle</a>	42
4.8.3.2	<a href="#">Estimation d'un taux de dépassement de seuil</a>	43
4.8.3.3	<a href="#">Amélioration de l'estimation des intervalles de confiance</a>	43
4.8.4	<a href="#">Simulations complémentaires</a>	46
4.8.5	<a href="#">Tests sur une autre année</a>	47
4.8.6	<a href="#">Quelles conclusions préliminaires peut-on tirer sur l'application de la méthode ?</a>	47
4.9	<a href="#">étude de plans d'échantillonnage</a>	48
4.10	<a href="#">Remarques préliminaires. Choix des paramètres de sondage</a>	48
4.11	<a href="#">Resultats des estimations</a>	49
4.12	<a href="#">Application de la méthode des plans de sondage : conclusions, recommandations, mises en garde</a>	49
4.12.1	<a href="#">Définition de l'échantillonnage</a>	49

4.12.2	<u>Estimation</u>	51
4.13	<u>TRavaux futurs</u>	52
4.14	<u>REFERENCES</u>	52
<b>5.</b>	<b><u>PARTIE III : COMPARAISON ENTRE LA MÉTHODE ISO ET LA MÉTHODE PLANS DE SONDAGE</u></b>	<b>54</b>
5.1	<u>Protocole de comparaison</u>	55
5.2	<u>Éléments généraux de comparaison</u>	57
5.3	<u>Théorie</u>	57
5.4	<u>Mise en œuvre des calculs</u>	57
5.4.1	<u>Travail préparatoire</u>	57
5.4.2	<u>Données d'entrée</u>	57
5.4.3	<u>Outils de calcul</u>	58
5.5	<u>Taille des échantillons</u>	58
5.6	<u>Résultats des Estimations</u>	61
5.7	<u>Reconstitution de moyennes</u>	62
5.7.1	<u>TOULOUSE</u>	62
5.7.2	<u>STRASBOURG</u>	65
5.7.3	<u>DUNKERQUE</u>	68
5.7.4	<u>LE HAVRE</u>	71
5.7.5	<u>ANNECY</u>	74
5.7.6	<u>ANNEMASSE</u>	77
5.7.7	<u>CHAMBERY</u>	80
5.8	<u>Reconstitution des nombres de dépassements de seuils</u>	83
5.8.1	<u>TOULOUSE</u>	83
5.8.2	<u>STRASBOURG</u>	87
5.8.3	<u>DUNKERQUE</u>	90
5.8.4	<u>LE HAVRE</u>	93
5.9	<u>Conclusions</u>	96
<b>6.</b>	<b><u>LISTE DES ANNEXES</u></b>	<b>97</b>

## 1. RESUME

Les campagnes de mesure conduites avec des moyens mobiles permettent d'obtenir une information sur les concentrations ou les nombres de dépassements de seuils dans des zones géographiques dépourvues de stations fixes. Cependant les durées des campagnes, de quelques semaines en été et/ou en hiver, ne sont représentatives que de situations temporaires.

Des travaux destinés à satisfaire aux directives - qui imposent une information sur de plus longues périodes, comme la saison ou l'année - ont été réalisés ces dernières années par le LCSQA et les AASQA, notamment dans le cadre du groupe de travail Moyens Mobiles. Ils ont pour but de développer et de mettre à la disposition des AASQA des outils qui les aident à établir une stratégie d'utilisation des moyens de mesure mobiles, de la définition du plan d'échantillonnage à l'estimation de l'indicateur recherché (moyenne annuelle, taux de dépassement d'un seuil déterminé).

Deux méthodes ont été plus particulièrement étudiées. L'EMD a mis au point une méthode fondée sur la norme ISO 9359 qui décrit, à partir des mesures de pollution issues de campagnes mobiles, la reconstitution sur le long terme d'une concentration moyenne assortie de son incertitude, et le dénombrement des dépassements de seuils prédéfinis. Cette méthode repose sur la connaissance et la prise en compte de l'impact des variations de paramètres météorologiques et/ou temporels sur les concentrations mesurées. L'INERIS a évalué, en collaboration avec ATMO Poitou-Charentes, la technique dite des *plans de sondage*.

L'objet de cette étude conjointe EMD-INERIS est de fournir un cadre d'application de ces deux méthodes. Celles-ci sont testées séparément, puis sur des bases de données identiques issues de stations fixes d'ORAMIP (Toulouse, années 2001/2002), de l'ASPA (Strasbourg, années 2003/2004), d'AIR NORMAND (Le Havre, années 2002/2003) et d'AIR APS (Annecy/Annemasse/Chambéry, années 2003/2004). Une démarche commune est adoptée : 1) Des pré-études sur les années (n) de chaque ville permettent de déterminer, en fonction de la précision d'estimation souhaitée, les caractéristiques des campagnes prévues pour les années suivantes ; 2) La réalisation de ces campagnes est simulée en prélevant une partie des bases des années (n+1) ; 3) Les paramètres annuels et leurs intervalles de confiance sont estimés. La comparaison des résultats permet d'identifier les différences et les complémentarités des approches, et d'en définir les limites d'utilisation.



La norme ISO 9359 prescrit une méthode pour prendre en compte l'impact de paramètres météorologiques (vitesse de vent et température) et temporels (jour, heure) sur les mesures. Il convient en premier lieu de connaître les influences des paramètres choisis sur les données horaires sur une année d'un site-test, de créer des strates correspondant à des combinaisons de ces paramètres. Ensuite, l'opérateur choisit le niveau d'incertitude désiré, et les méthodes déterminent le nombre de mesures minimum à réaliser lors de campagnes futures pour reconstituer les moyennes et les écarts types des polluants concernés et les nombres de dépassements en fonction de cette incertitude. On réalise ensuite des campagnes mobiles et on estime les moyennes annuelles et le nombre de dépassements de seuils prédéterminés avec leurs incertitudes respectives au prorata des occurrences des strates pendant la campagne et sur l'année complète.

Fondée sur une théorie statistique reconnue, la méthode des plans de sondage est d'une utilisation relativement aisée et ne requiert qu'une pratique limitée du logiciel « R » avec lequel elle fonctionne actuellement. Son efficacité et sa fiabilité ont été évaluées en simulant un grand nombre d'échantillonnages sur des séries de données de stations fixes. Les résultats obtenus s'accordent avec la théorie. Ces simulations ainsi que les tests effectués par l'INERIS en parallèle avec l'EMD montrent la pertinence de la méthode en situations périurbaine et urbaine (y compris de proximité automobile) ; en revanche, ils donnent moins satisfaction en situation industrielle. D'autre part, si la pratique ne permet pas de respecter strictement les hypothèses de la méthode, il convient de s'en approcher. Un minimum de huit à dix périodes de mesure dans l'année est ainsi nécessaire pour estimer correctement une moyenne annuelle et son intervalle de confiance à 95%. L'estimation précise d'un taux de dépassement de seuil s'avère plus exigeante.

Quelle que soit la méthode, la qualité des résultats est limitée par la durée réduite des campagnes qui peuvent refléter incomplètement les événements qui se déroulent sur une année entière, notamment les pointes élevées de pollution. L'incidence est plus importante sur la reconstitution de nombres de dépassement de seuil que sur celle des moyennes.

## 2. INTRODUCTION

### 2.1 CONTEXTE

Pourvu qu'elles aient été soigneusement planifiées, les campagnes de mesure conduites avec des camions laboratoires ou d'autres moyens mobiles plus légers (remorques) peuvent compléter efficacement les mesures effectuées aux stations fixes. Les objectifs de ces campagnes sont multiples, comme le montre une enquête réalisée auprès des AASQA (*Représentativité des mesures et méthodes statistiques*, rapport INERIS/LCSQA, 2001) :

- exploration de régions peu ou non équipées en stations fixes
- contrôle des mesures obtenues par échantillonnage passif
- choix du point d'implantation d'une station fixe
- validation d'une station fixe existante
- étude d'impact
- réponse à des demandes locales (de riverains par exemple).

Dans une stratégie de surveillance de la qualité de l'air destinée à produire une information sur l'ensemble du territoire français, **l'exploration de régions peu ou non instrumentées** représente l'enjeu majeur de ces campagnes. **Mais comment, de mesures nécessairement limitées dans le temps, est-il possible d'extraire une information moyenne fiable sur la qualité de l'air en un point ?**

Le groupe de travail *Etudes Mobiles*, animé par l'Association ATMO Poitou-Charentes de 2000 à 2004, et le groupe de travail *Plans d'échantillonnage et reconstitution de données*, récemment créé et animé par l'ADEME, ont fait de la représentativité temporelle des mesures un thème d'étude prioritaire. Il s'agit en particulier de répondre aux questions suivantes :

- Quand, pendant combien de temps, et à quelle fréquence faut-il effectuer les mesures afin d'atteindre une précision d'estimation fixée ?
- Comment peut-on estimer une concentration moyenne et un nombre de dépassements de seuil sur une période plus longue que celle des mesures (saison, année) ?

Aucune méthode n'est prescrite en France pour traiter ces deux points. Quelques préconisations portant sur les mesures indicatives sont données par les directives européennes [1]. Celles-ci fournissent, « à titre d'orientation pour les programmes d'assurance de la qualité », des objectifs de qualité en ce qui concerne « l'incertitude admise des méthodes d'évaluation, la période minimale prise en compte et la saisie minimale des données » (Tableau 1).

*Tableau 1 - Objectifs de qualité proposés à titre d'orientation par les directives européennes*

Polluant	Période minimale prise en compte	Saisie minimale de données	Incertitude maximale
SO <sub>2</sub> , NO <sub>2</sub> , NO (1999/30/CE) CO (2000/69/CE)	14%  (une mesure par semaine, au hasard, également répartie sur l'année, ou huit semaines, également réparties sur l'année)	90%	25%
O <sub>3</sub> , NO <sub>2</sub> , NO (2002/3/CE)	>10% en été	90%	30%
Particules, Plomb (1999/30/CE)	14% (idem)	90%	50%
Benzène (2000/69/CE)	14% (idem)	90%	30%

Un guide technique européen (EEA, 1998) recommande d'étendre l'échantillonnage sur une durée égale à **20% de la période de référence** associée à la valeur limite considérée, par exemple deux fois cinq semaines ou cinq fois deux semaines si cette période est l'année.

Dans les AASQA, des techniques empiriques ou statistiques sont parfois appliquées (Malherbe, 2001). Deux méthodes ont été plus particulièrement étudiées au sein du GT Moyens mobiles : la méthode de reconstitution temporelle de l'EMD, fondée sur la prise en compte de la météorologie et sur l'application de la norme ISO 9359 ; la méthode des plans de sondage d'ATMO Poitou-Charentes, évaluée par l'INERIS. Dans la suite du rapport, elles pourront être plus brièvement désignées par les expressions « méthode ISO » et « méthode Plans ».

## **2.2 PRINCIPE DES DEUX METHODES ETUDIEES**

### **2.2.1 METHODE DE RECONSTITUTION « ISO »**

Elle est fondée sur l'application de la norme :

“ ISO 9359 – Qualité de l'air – Echantillonnage aléatoire stratifié pour l'évaluation de la qualité de l'air ambiant ”, nome à l'élaboration de laquelle l'EMD avait en son temps participé au sein de l' ISO TC146/SC4 .

Elle consiste, entre autres possibilités, à faire sur un ou plusieurs sites des regroupements de mesures ponctuelles selon différents critères, tels que :

- des directions de vent sélectionnées par rapport à un émetteur industriel,
- des classes de concentration spécifiées,
- des classes de valeurs de paramètres météorologiques, tels que la température, la pluviométrie, la pression,
- des typologies spécifiques de stations, ... .

Cette norme décrit comment reconstituer sur le long terme une concentration moyenne et comment calculer son incertitude à partir de groupes de mesures partielles.

La méthode ainsi développée repose sur la sélection de groupes de mesures par classe de paramètres météorologiques et sur la prise en compte de l'influence de ces paramètres et de leurs variations sur les concentrations de polluants atmosphériques.

Un outil de traitement statistique des données a été mis au point et validé, à l'aide de séries chronologiques complètes de mesures à poste fixe de SO<sub>2</sub>, NO, NO<sub>2</sub>, O<sub>3</sub>, CO et PM<sub>10</sub>, PM<sub>2.5</sub>, provenant des réseaux de Lille, Marseille, Toulon, Toulouse, Le Havre et Strasbourg.

## **2.2.2 METHODE DES PLANS DE SONDAGES**

Proposée par ATMO Poitou-Charentes [Lavancier et al., 2003], la méthode des plans de sondage offre une réponse pertinente et simple d'usage aux questions de l'échantillonnage et de l'estimation de moyennes annuelles. Fondée sur une théorie statistique rigoureuse, elle a pour fonction :

- de déterminer, en s'appuyant sur des données de stations fixes, les stratégies d'échantillonnage qui permettent d'atteindre une précision donnée. Les plans ainsi définis valent pour les stations et pour la ou les années qui ont servi à les élaborer, mais ils fournissent des indications utiles sur le nombre de mesures qu'il est nécessaire de collecter.
- d'estimer, une fois que l'échantillonnage a été réalisé, la concentration moyenne annuelle et un intervalle de confiance autour de cette moyenne. Ces estimations peuvent être corrigées à l'aide d'une série complète de données issues d'une station fixe auxiliaire.

Si pour être efficace, elle impose des contraintes d'échantillonnage, la méthode des plans de sondage n'exige aucun travail de modélisation ni de données de variables externes. Elle peut être mise en œuvre grâce à des programmes informatiques initialement conçus par ATMO Poitou-Charentes et transposés sous R (logiciel libre de statistique) par l'INERIS.

## **2.3 PLAN**

Le travail a été conduit en deux étapes :

- une étude séparée des méthodes ISO et Plans de sondage, comprenant les développements informatiques et les évaluations nécessaires (parties I et II);
- une étude comparative, dans le but de préciser les conditions d'application, les avantages et les limites de chaque approche. Un protocole de comparaison a été conçu à cette fin (partie III).

# **3. PARTIE I : METHODE ISO**

### 3.1 INTRODUCTION

Les directives européennes imposent des objectifs de surveillance divers selon plusieurs critères :

- évaluation de la qualité de l'air par des estimations objectives des teneurs en polluants dans les zones où elles sont inférieures aux seuils minimaux d'évaluation,
- surveillance représentative de tout le territoire,
- surveillance renforcée dans les zones où les seuils réglementaires risquent d'être dépassés

L'usage de moyens mobiles de mesure peut apporter des informations sur le critère prévalant dans une zone donnée. Cette étude est issue de l'un des besoins exprimés par le groupe de travail sur les moyens mobiles de mesure, à savoir :

*« le nombre de campagnes mobiles sur un site qu'il faudrait mettre en œuvre par an et leur durée, pour avoir une estimation de la moyenne annuelle comparable à celle qui aurait été obtenue avec une station fixe ».*

L'application de la norme :

*« ISO 9359 – Qualité de l'air – Echantillonnage aléatoire stratifié pour l'évaluation de la qualité de l'air ambiant »,*

permet d'étudier et de prendre en compte l'impact de paramètres météorologiques et/ou temporels sur les mesures réalisées lors de campagnes mobiles afin d'estimer les moyennes annuelles et le nombre de dépassements de seuils prédéterminés avec leurs incertitudes respectives, qui auraient été obtenus avec une station fixe :

- pour des directions de vent sélectionnées par rapport à un émetteur industriel,
- pour des classes de concentration spécifiées,
- pour des classes de valeurs de paramètres météorologiques, tels que la température, la pluviométrie, la pression, ... ,
- selon l'heure du jour et le jour dans la semaine.

Cette norme décrit comment à partir de groupes de mesures partielles :

- reconstituer sur un plus long terme, par exemple sur une année, une concentration moyenne assortie de son incertitude,
- dénombrer les dépassements de seuils prédéfinis (ex : valeurs limites).

En revanche, cette norme est succincte sur les modalités de choix des classes de paramètres à prendre en compte et la méthode de construction des strates. Cela est laissé au libre arbitre des utilisateurs selon les contextes locaux. Cette étude tente de combler ce manque. Des études de faisabilité appliquées aux données journalières de pollution atmosphérique de Lille de 1996 à 1999 , aux données tri-horaires de Marseille de 1998 à 2000, de Toulon en 2000, de Toulouse en 2001 et 2002, de Le Havre en 2002 puis de Strasbourg en 2003 ont été réalisées (Houdret, étude LCSQA, 2004).

## 3.2 PRINCIPE

Différentes combinaisons de paramètres météorologiques et/ou temporels entraînent la présence de fortes, moyennes ou faibles concentrations en polluants ; le nombre minimum de mesures nécessaires sur un site pour estimer une moyenne et un nombre de dépassements de seuil avec une incertitude donnée, requiert la connaissance des distributions des données et de leurs écarts types attendus sur un site-test similaire ; les résultats de ces estimations sont plus tributaires des concentrations élevées que des faibles ; aussi il est préférable de regrouper en strates définies par les combinaisons de paramètres, les données de mêmes ordres de grandeur, afin de réduire les écarts types des données dans chaque strate.

Il convient donc :

- en premier lieu de connaître les influences des paramètres météorologiques et/ou temporels sur les données du site-test choisi, et de créer des strates de concentrations correspondant à des combinaisons de ces paramètres,
- de calculer les moyennes et les écarts types des polluants concernés dans chaque strate,
- de déterminer le nombre minimum de mesures qu'il conviendrait de faire pour une incertitude donnée dans chaque strate lors de campagnes ultérieures,
- de réaliser les campagnes de mesure sur les autres sites à prospector, et de ventiler dans chaque strate les résultats obtenus correspondant aux mêmes combinaisons,
- de reconstituer les moyennes annuelles et/ou les nombres de dépassements de seuils au prorata annuel des occurrences des strates pendant les campagnes,
- de tester et valider la méthode sur la base du nombre de mesures trouvés, en simulant des campagnes de mesures prises dans la base de données du site-test.

L'objectif de ce rapport est de rédiger un mode d'emploi applicable dans toutes les AASQA, après l'avoir testé avec les données de plusieurs réseaux, et surtout pour diverses typologies de stations.

Un classeur EXCEL de calcul automatique est disponible sur demande. Les modalités d'utilisation sont décrites en annexe.

### **3.3 LES PARAMETRES DE STRATIFICATION PRIS EN COMPTE**

Pour l'aspect météorologique, ce sont :

- la vitesse du vent,
- la température.

Pour l'aspect temporel, ce sont :

- Les heures des séquences,
- Les jours de la semaine.

La limitation volontaire à ces paramètres de base, disponibles dans toutes les associations, rend cette étude applicable par tous.

**Note** : l'aspect saisonnier n'est pas considéré ici sur un plan temporel mais est pris en compte indirectement par les classes de température qui reflètent les variations saisonnières.

Pour créer les strates, on a au préalable déterminé les degrés d'influence de chaque paramètre météorologique et temporel retenu, sur chacun des polluants considérés et pour chaque station ; cela a permis de définir les classes les plus homogènes possibles pour lesquelles les concentrations mesurées pour chaque polluant sont assez voisines les unes des autres, par exemple de faibles, moyennes, fortes, à très fortes, afin d'obtenir pour chaque classe des écarts types plus faibles que celui de l'ensemble des données. On combine ensuite les paramètres choisis pour créer les strates.



### 3.4 INFLUENCE DES PARAMETRES METEOROLOGIQUES

#### 3.4.1 INFLUENCE DE LA VITESSE DU VENT SUR NO ET NO<sub>2</sub> A TOULOUSE

L'influence de la vitesse du vent est traitée pour les concentrations moyennes et pour les dépassements de seuils prédéfinis.

L'étendue des vitesses de vent de 0 à 11 m.s<sup>-1</sup> a été divisée en 8 classes. Les résultats sont présentés sur la Figure 1.

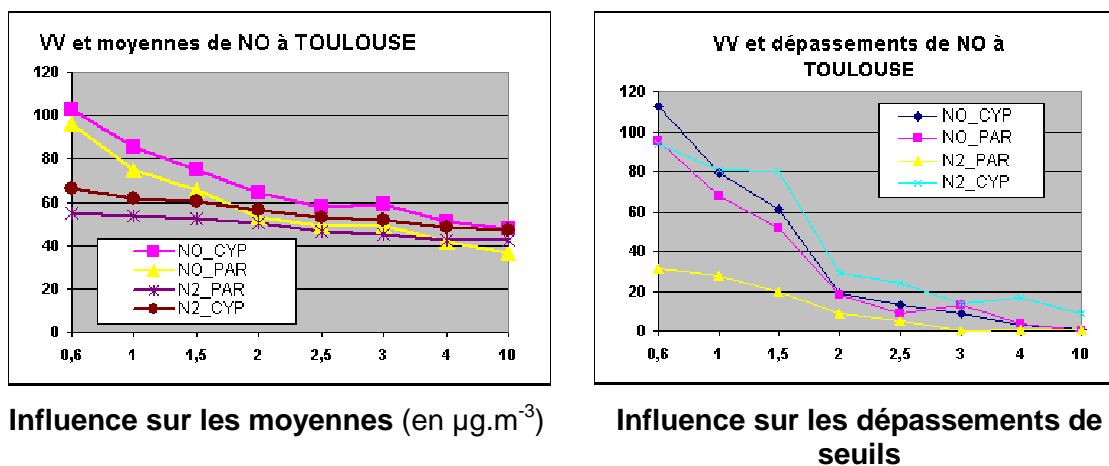


Figure 1: moyennes et pointes de NO et NO<sub>2</sub> en fonction de la vitesse du vent

On constate que les concentrations horaires en NO en fonction de la vitesse de vent diminuent rapidement lorsque la vitesse du vent augmente, surtout de 1 à 2 m.s<sup>-1</sup>, puis plus lentement au-delà.

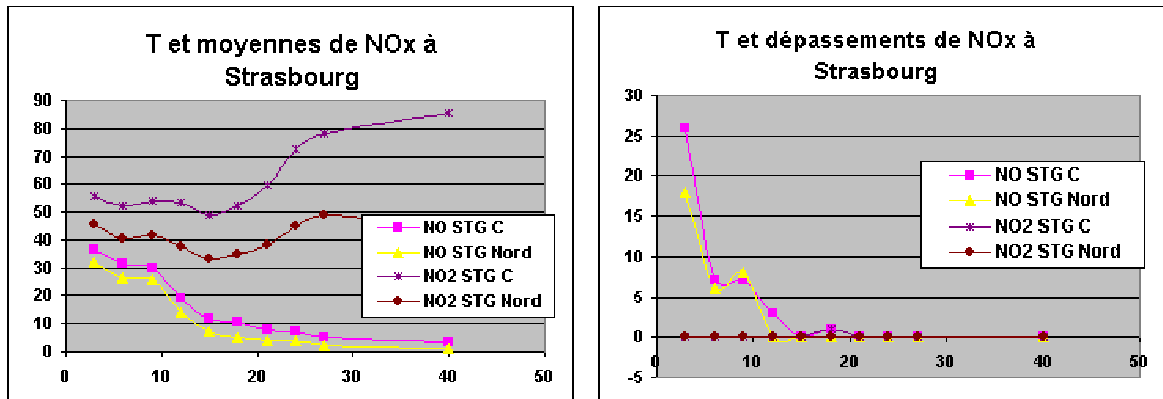
On considère donc 2 classes de vitesse de vent de part et d'autre de 1,5 m.s<sup>-1</sup>.

#### 3.4.2 INFLUENCE DE LA TEMPERATURE

Les températures s'échelonnent de -5 à +40 degrés ; cette étendue a été divisée en 10 classes : T < 0°C, par pas de 3 degrés et T > 24°C.

Les moyennes des valeurs horaires en polluants ont été calculées pour chaque classe, et par polluant. Les influences de la température sur les moyennes et les dépassements de seuils pour divers polluants sont représentées sur les figures suivantes.

### 3.4.3 INFLUENCE DE LA TEMPERATURE SUR NO ET NO<sub>2</sub> A STRASBOURG



Influence sur les moyennes (en  $\mu\text{g.m}^{-3}$ )

Influence sur les dépassements des seuils 100 (NO) et 75 (NO<sub>2</sub>) en  $\mu\text{g.m}^{-3}$

Figure 2 : moyennes et pointes de NO et NO<sub>2</sub> selon la température

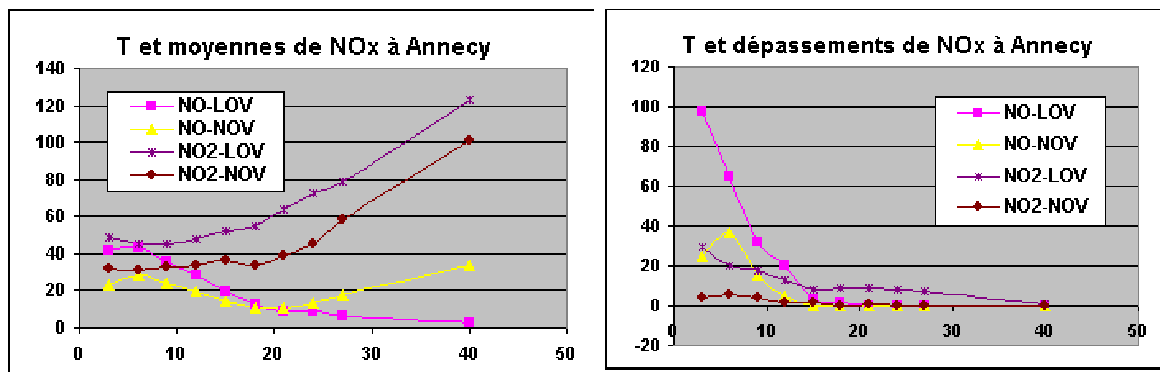
Une élévation de température entraîne une baisse des concentrations de NO et une augmentation de celles de NO<sub>2</sub>. L'influence est assez faible sur les dépassements de seuils pour NO<sub>2</sub>.

On considère 2 classes de température de part et d'autre de :

10°C pour les moyennes et les dépassements de NO.

22°C pour les moyennes de NO<sub>2</sub>, et pas d'influence notable pour les dépassements de NO<sub>2</sub>.

### 3.4.4 INFLUENCE DE LA TEMPERATURE SUR NO ET NO<sub>2</sub> A ANNECY



Influence sur les moyennes (en  $\mu\text{g.m}^{-3}$ )

Influence sur les dépassements du seuil 100 (NO et NO<sub>2</sub>) en  $\mu\text{g.m}^{-3}$

Figure 3 : moyennes et pointes de NO et NO<sub>2</sub> selon la température

Une élévation de température entraîne une baisse des concentrations de NO et une augmentation de celles de NO<sub>2</sub>. L'influence est assez faible sur les dépassements de seuils pour NO<sub>2</sub>. On considère 2 classes de température de part et d'autre de :

12°C pour les moyennes et les dépassements de NO,

22°C pour les moyennes de NO<sub>2</sub>, et pas d'influence notable pour les dépassements de NO<sub>2</sub>.

### 3.4.5 INFLUENCE DE LA TEMPERATURE SUR O<sub>3</sub> AU HAVRE

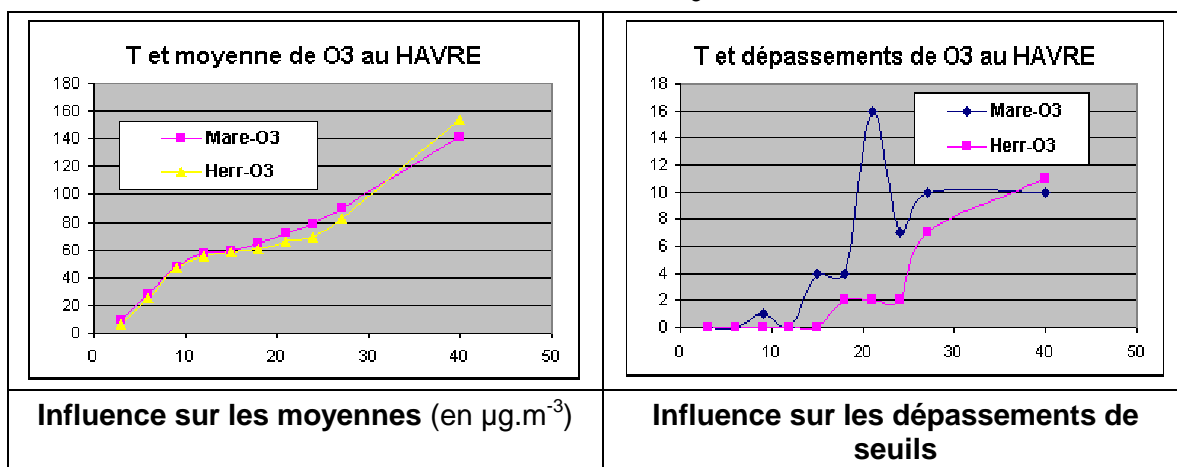


Figure 4 : moyennes et pointes de O<sub>3</sub> en fonction de la température

On constate une très forte similitude entre les moyennes des 2 stations. Une élévation de température entraîne une augmentation continue des moyennes des concentrations en O<sub>3</sub> sur l'ensemble de l'échelle des températures (Figure 4 de gauche). L'apparition des pointes est plus tranchée, au-delà de 20°C comme le montre la Figure 4 de droite.

On considère donc 2 classes de part et d'autre de 20°C.

### 3.4.6 INFLUENCE DE LA TEMPERATURE SUR LES PM2.5 ET PM10 (TEOM)

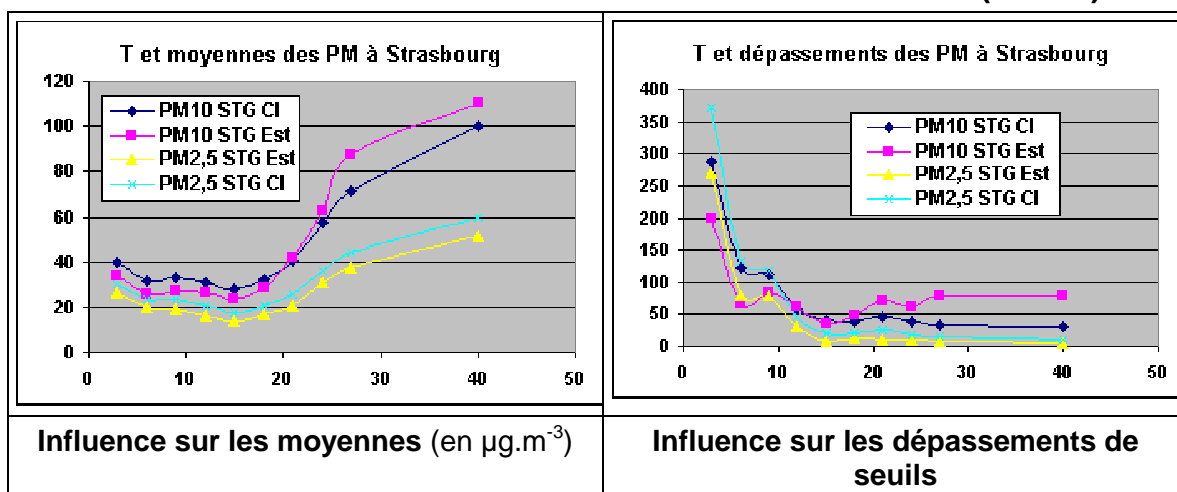


Figure 5 : moyennes et pointes de PM en fonction de la température

On observe une analogie entre les stations de mesure des PM. Une élévation de température entraîne une légère baisse des moyennes des concentrations jusqu'à 14 degrés environ ; puis elles augmentent jusqu'aux températures les plus élevées.

Cela met en évidence une différence entre pollution hivernale et pollution estivale.

L'effet de la température est différent sur l'apparition des pointes accrues entre 0 et 10 °C, et qui sont peu nombreuses pour les températures négatives et inexistantes pour les températures élevées (sauf STG-Est).

On choisit pour les moyennes une séparation en 2 classes, l'une entre 5 et 22 °C et l'autre en dehors de cet intervalle.

Pour les pointes, on choisit une séparation en 2 classes de part et d'autre de 10°C.

### 3.5 INFLUENCE DES PARAMETRES TEMPORELS

#### 3.5.1 PRINCIPE

Les concentrations en polluants sont également fonction des périodes où ont eu lieu les émissions, reflets des activités humaines, **l'heure dans la journée, la journée dans la semaine** et le moment dans l'année.

**Note** : le moment dans l'année est en fait l'influence saisonnière sur les polluants d'origine non industrielle ; il n'est pas traité ici directement mais par l'influence de la température.

#### 3.5.2 INFLUENCE DE L'HEURE DE LA JOURNEE

On calcule pour chaque ensemble de séquences de 1 heure et par station, les moyennes et les nombres de dépassements de seuils de chaque polluant. Cette influence temporelle sur les polluants est représentée sur les graphiques suivants.

### 3.5.3 INFLUENCE DE L'HEURE DE LA JOURNEE SUR NO ET NO<sub>2</sub> A CHAMBERY

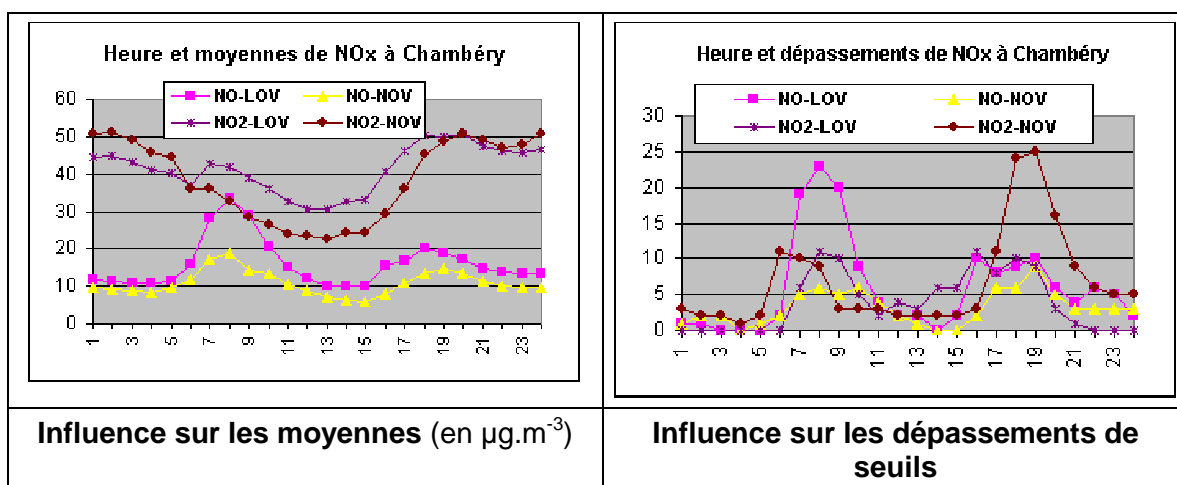


Figure 6 : moyennes et pointes de NO et NO<sub>2</sub> selon l'heure de la journée

On observe une analogie certaine entre les 2 stations pour les moyennes de NO et les pointes de NO et NO<sub>2</sub> de 6 à 9 heures et entre 15 à 21 heures. On définit 2 classes de périodes :

**classe 1 : de 21h à 6h et de 9h à 15h et classe 2: de 6h à 9h et de 15h à 21h.**

Les moyennes de NO<sub>2</sub> par contre augmentent plutôt en fin de journée et jusqu'à la fin de la nuit.

### 3.5.4 INFLUENCE DE L'HEURE DE LA JOURNEE SUR O<sub>3</sub> A TOULOUSE

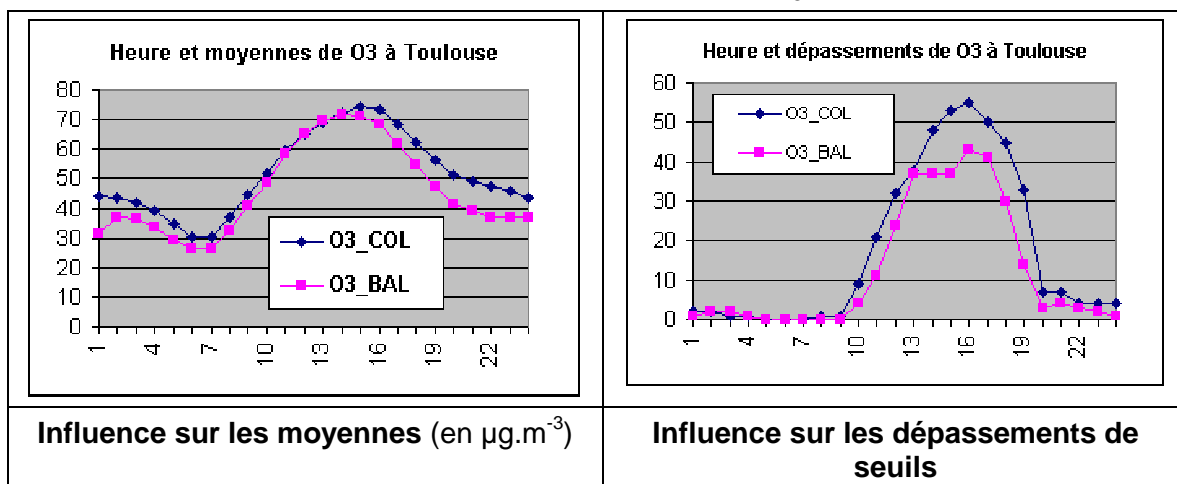


Figure 7 : moyennes de O<sub>3</sub> en fonction de l'heure de la journée

On observe une analogie certaine entre les 2 stations. Les niveaux élevés de 10h à 18 heures en O<sub>3</sub> définissent la séparation en 2 classes :

**Classe 1 : de 18h à 9h et Classe 2 : de 9h à 18 h.**

### 3.6 INFLUENCE DU JOUR DE LA SEMAINE

On calcule pour chacun des 7 jours, les moyennes de chaque polluant par station. Cette influence temporelle sur les polluants est représentée sur les graphiques suivants. L'influence sur les apparitions des pointes est similaire et n'est pas reportée ici.

#### 3.6.1 INFLUENCE DU JOUR DE LA SEMAINE SUR NO ET NO<sub>2</sub> A STRASBOURG

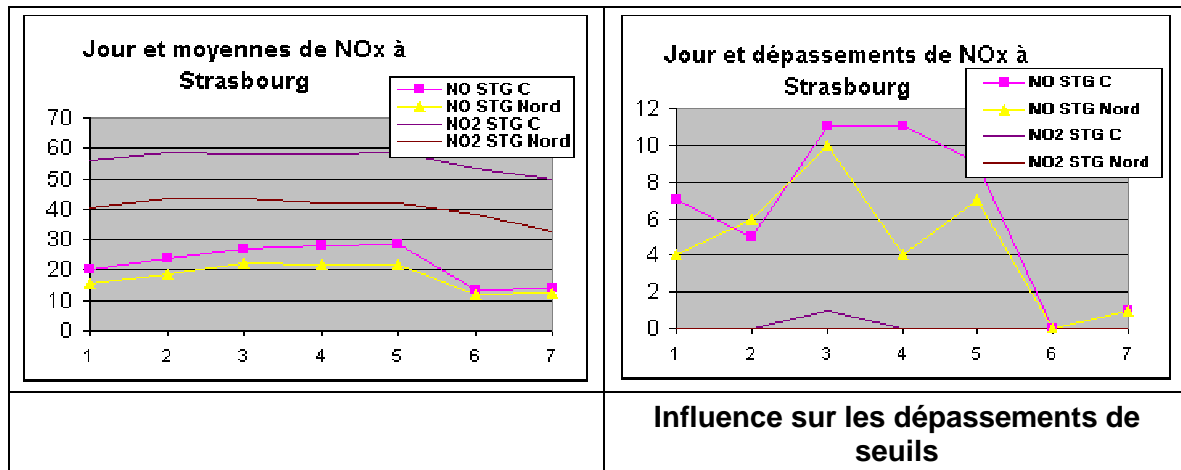


Figure 8 : NO et NO<sub>2</sub> en fonction du jour de la semaine

On distingue les 5 jours ouvrés de la semaine et les deux jours de fin de semaine qui constituent le choix de deux classes communes pour NO et NO<sub>2</sub> :

**classe 1 : jours S et D (6 et 7) et classe 2 : jours L à V (1 à 5).**

### 3.7 APPLICATION DE LA NORME ISO 9359

### 3.8 BILAN COMPLET DES INFLUENCES DES PARAMETRES POUR LES DIVERSES VILLES

Des études similaires à celles qui sont données comme exemple au paragraphe 3.3 ont été réalisées pour l'ensemble des villes et des paramètres météorologiques et temporels. Elles conduisent aux stratifications suivantes :

*Tableau 2 : récapitulatif des classes de paramètres d'influence favorisant des valeurs horaires élevées dans plusieurs villes*

		MOYENNES				DEPASSEMENTS DE SEUILS			
VILLES	Polluants	VV m.s <sup>-1</sup>	T en °C	Périodes en heures	Jours	VV m.s <sup>-1</sup>	T en °C	Périodes en heures	Jours
Toulouse	NO	< 2	< 12	7-11,16-21	1-5	< 2	> 18	7-11,16-21	2-4
	NO <sub>2</sub>	< 2	< 12	7-11,16-21	1-5	< 2	> 18	7-11,16-21	2-4
	O <sub>3</sub>	> 2	> 20	10-19	1-7	< 2	> 27	11-19	1-3
	CO	< 1,5	< 12	7-11,16-21	2-6	< 1,5	< 18	7-10,16-21	1-4
	PM	< 2	<6,>20	7-11,19-22	1-5	< 2	< 6	7-11,19-22	1-4
Strasbourg	NO	< 1,5	< 12	7-11,16-21	2-5	< 1,5	< 6	7-11,16-21	3-5
	NO <sub>2</sub>	< 1,5	> 24	7-11	2-5	< 1,5	< 6	7-21	3-5
	O <sub>3</sub>	< 1,5	> 24	10-21	2-5	> 2,5	> 24	10-20	1-7
	PM	< 2	> 24	14-22	2-5	< 1,5	< 6	7-11,19-22	2-5
Le Havre	SO <sub>2</sub>	< 1,5	< 10	1-12		< 6	< 18	4-11	
	O <sub>3</sub>	> 4	> 20	11-24	2-5	> 2,5	> 18	11-20	4-5
	PM	< 3	<10,>20	6-12	2-3	< 6	< 24	6-12	3-5
Annecy	NO	< 1,5	< 12	7-11,16-21	2-5	< 1,5	< 6	7-11,16-21	3-5
	NO <sub>2</sub>	< 1,5	> 24	7-11	2-5	< 1,5	< 6	7-21	3-5
	O <sub>3</sub>	< 1,5	> 24	10-21	2-5	> 2,5	> 24	10-20	1-7
	PM	< 2	> 24	14-22	2-5	< 1,5	< 6	7-11,19-22	2-5
Annemasse	NO	< 1,5	< 12	7-11,16-21	2-5	< 1,5	< 6	7-11,16-21	3-5
	NO <sub>2</sub>	< 1,5	> 24	7-11	2-5	< 1,5	< 6	7-21	3-5
	O <sub>3</sub>	< 1,5	> 24	10-21	2-5	> 2,5	> 24	10-20	1-7
	PM	< 2	> 24	14-22	2-5	< 1,5	< 6	7-11,19-22	2-5
Chambéry	NO	< 1,5	< 12	7-11,16-21	2-5	< 1,5	< 6	7-11,16-21	3-5
	NO <sub>2</sub>	< 1,5	> 24	7-11	2-5	< 1,5	< 6	7-21	3-5
	O <sub>3</sub>	< 1,5	> 24	10-21	2-5	> 2,5	> 24	10-20	1-7
	PM	< 2	> 24	14-22	2-5	< 1,5	< 6	7-11,19-22	2-5
Dunkerque	SO <sub>2</sub>	< 1,5	< 10	1-12		< 6	< 18	4-11	
	NO	> 4	> 20	11-24	2-5	> 2,5	> 18	11-20	4-5
	NO <sub>2</sub>	< 3	<10,>20	6-12	2-3	< 6	< 24	6-12	3-5

Ce Tableau 2 indique aussi que la stratification est différente pour les **moyennes** et pour les **dépassements** de seuils.

Les paramètres temporels (heures et jours) sont à peu près les mêmes dans chaque ville et ne sont pas reportés ici et cela montre que l'on pourrait pratiquer à peu près :

**la même stratification dans chacune des villes considérées.**

Cela montre également que la pré-étude ne revêt pas un caractère obligatoire mais est néanmoins conseillée pour approfondir la connaissance de la pollution locale.

### **3.9 STRATIFICATION DES SEQUENCES HORAIRES DE MESURE**

**CETTE ETAPE EST INDISPENSABLE ET OBLIGATOIRE**

Il convient de disposer pour l'ensemble de l'année incluant les campagnes de mesure, du maximum de données de vitesse de vent et de température.

On complète cette base de données en affectant de la valeur de classe « **2** », les séquences horaires correspondant pour chaque polluant aux conditions fixées dans le Tableau 2, et de la valeur « **1** » pour toutes les autres.

Ensuite, pour chaque séquence horaire pour chaque polluant, on calcule la somme «  $\Sigma$  » des nombres de classes « **2** » simultanées correspondant aux paramètres générateurs de concentrations élevées, soit «  $\Sigma = 0 \text{ à } 4$  ».

**Note 1** : pour simplifier la méthode, on a attribué une importance égale à chacun des paramètres d'influence pris en compte ; on pourrait affiner la stratification en pondérant les influences des divers paramètres. On peut en effet imaginer une hiérarchie des influences dans cet ordre décroissant : *la température, la vitesse du vent, les heures de pointe et les jours ouvrés.*

On définit ensuite 4 strates dans lesquelles on ventile les séquences selon le modèle suivant :

Strate 1 si  $\Sigma = 0$  // Strate 2 si  $\Sigma = 1$  // Strate 3 si  $\Sigma = 2$  // Strate 4 si  $\Sigma = 3$  ou **4**.

**Note 2** : ce choix permet d'avoir des strates contenant des nombres de séquences les plus voisins possibles ; la norme ISO 9359 préconise de limiter à 3 ou 4 le nombre de strates, car un nombre plus élevé n'est plus efficace pour la réduction des écarts types, et réduirait trop les nombres de mesures par strate, à fortiori encore plus lors de campagnes de trop courtes durées.



### **Exemple d'application :**

Le Tableau 2 a montré que les concentrations en NO sont plus élevées lorsque :

- la vitesse du vent est  $< 1,5$  m/s,
- ou la température est  $< 10^{\circ}\text{C}$ ,
- ou de 6 à 9h et de 15 à 21h,
- ou du lundi au vendredi.

Dans le cas d'une séquence de mesure où ces 4 paramètres sont simultanément dans cet état (classe « 2 »), cette séquence sera ventilée dans la strate 4. Si cette séquence était intervenue un dimanche, elle serait dans la strate 3. Mais si le vent avait en plus une vitesse  $> 1,5$  m.s<sup>-1</sup>, la séquence serait alors ventilée dans la strate 2. Si en plus, la température était de  $> 10^{\circ}\text{C}$ , elle passerait en strate 1.

### **3.10 CALCULS DES MOYENNES ET DES NOMBRES DE DEPASSEMENT PAR STRATE**

On calcule ensuite pour chaque strate, les moyennes des données de pollution et les nombres de dépassement de seuils prédéfinis, par polluant et par station. Grâce à la stratification, on obtient, des groupes de données plus centrées autour de leurs moyennes respectives, et donc affectées d'écart type réduits par rapport à celui de la population totale des données.

Ces calculs sont réalisés dans 3 cas de figure :

- sur une base annuelle complète de données lors de tests de la méthode de reconstitution, en simulant des campagnes en prélevant une partie de la base,

#### **C'EST LE CAS DANS CETTE ETUDE,**

- sur une base annuelle antérieure dans une zone de mêmes caractéristiques météorologiques que celle où est prévue une campagne, cela permet d'estimer la durée totale des campagnes à réaliser pour une incertitude prédéterminée, et les nombres de séquences de mesure par strate,

#### **UTILE MAIS NON OBLIGATOIRE**

- sur les données issues des campagnes,

#### **EVIDEMMENT OBLIGATOIRE.**

Si l'utilisateur admet que la méthode est applicable à l'ensemble d'une zone élargie qu'il juge homogène, et à d'autres années que celle qu'il a étudiée, il n'aura pas à refaire la pré-étude à l'avenir. Les connaissances dont il dispose sur ses propres conditions locales sont peut-être suffisantes pour qu'il n'ait pas à faire de pré-étude des influences pour créer ses strates.

Ces calculs sont automatisés sur un classeur EXCEL créé pour cet usage, disponible sur le site web du LCSQA.

### 3.11 INCERTITUDES SOUHAITEES ET DUREE DES CAMPAGNES

Les **incertitudes** sur les reconstitutions des moyennes et des nombres de dépassements de seuils sont **inversement proportionnelles aux nombres de séquences de mesure réalisées**.

La méthode proposée calcule automatiquement les nombres minima de mesures nécessaires pour la reconstitution des moyennes avec une **incertitude pré-définie** et qui a été choisie égale à **+/-10%**.

Les durées des campagnes définies par la méthode pour la reconstitution des moyennes sont les mêmes pour les dépassements de seuils. La distinction dans la méthode est effectuée au niveau de la différence de stratification.

Les calculs (classeur EXCEL sur site web du LCSQA) fournissent les nombres de séquences horaires à réaliser.

L'annexe 6 de cette étude reprend un extrait du rapport EMD 2004 qui donne le détail des calculs effectués [2].

Le Tableau 3 indique **pour chaque ville traitée, selon les stations** considérées et **par polluant**, les nombres « **mini à maxi** » de semaines de campagnes à réaliser. Le choix de la semaine comme base de travail se justifie par les influences variées des 7 jours hebdomadaires, et par le fait qu'une campagne de mesure pour des raisons logistiques est rarement inférieure à cette durée.

*Tableau 3 : Nombres de semaines requises pour les campagnes*

	SO <sub>2</sub>	NO	NO <sub>2</sub>	O <sub>3</sub>	PM 10
<b>Toulouse 2001</b>	4 à 6	6 à 7	2 à 3	2 à 3	
<b>Strasbourg 2003</b>	4 à 5	8 à 17	2 à 4	1 à 2	2 à 3
<b>Le Havre 2002</b>	4 à 8	3 à 15	1 à 2	1 à 2	2 à 3
<b>Annecy 2003</b>	3 à 5	3 à 15	1 à 2	1 à 2	2 à 3
<b>Annemasse 2003</b>	3 à 5	3 à 15	1 à 2	1 à 2	2 à 3
<b>Chambéry 2003</b>	13 à 15	10 à 12	2 à 3	1 à 2	2 à 3
<b>Dunkerque 2002</b>	2 à 10	10 à 12	1 à 2	3 à 4	2 à 4

La disparité entre les nombres de semaines requises dépend des étendues de mesure de chaque polluant ainsi que des écarts types associés, liés à la variabilité temporelle des concentrations.

Les nombres de semaines les plus élevés requis pour SO<sub>2</sub> (au Havre et à Strasbourg) correspondent à des stations de proximité industrielle ; ceux requis pour NO correspondent à des stations de proximité automobile ; dans les deux cas, l'étendue de mesure est très large.

Les nombres de semaines requis pour NO<sub>2</sub>, O<sub>3</sub> et PM sont en général plus faibles pour des raisons inverses, c'est-à-dire correspondant à des étendues de mesure étroites liées à des pollutions de fond de variabilité temporelle plus faible.

Si l'on considère que lors des campagnes, on mesure plusieurs polluants en même temps, on va considérer que :

**LA DUREE MINIMALE CUMULEE DES CAMPAGNES SERA DONC DE 4 SEMAINES MINIMUM EN SITUATION DE FOND,**

**ET D'AU MOINS 12 SEMAINES EN SITUATION DE PROXIMITE POUR SO<sub>2</sub> ET NO.**

Ces nombres théoriques sont destinés à suggérer à l'utilisateur un ordre de grandeur du nombre de mesures à effectuer par strate et donc par type de conditions, ce qui signifie que le planning doit être préparé selon des prévisions météorologiques :

- Soit propices aux hautes concentrations en polluant considéré, ozone en été, SO<sub>2</sub> en hiver, etc ..., périodes de vent faible, directions de vent spécifiques, pour la reconstitution de nombres de dépassements,
- Ou représentatives des situations multiples que l'on peut rencontrer sur l'année pour la reconstitution des moyennes annuelles.

Le choix entre les deux options ci-dessus dépendra du but recherché par les campagnes.

*En réalité, les campagnes de mesure se programment en général trop longtemps à l'avance pour que des prévisions météorologiques fiables soient disponibles ; il semble alors judicieux de faire des campagnes de 4 à plus de 12 semaines étalées dans l'année, de durée suffisante pour espérer inclure les périodes les plus propices à la stratégie souhaitée.*

### 3.12 DISCUSSION

La pré-étude sur les liens entre la météorologie et les données de pollution atmosphériques présente déjà en elle-même un intérêt qui s'étend au-delà des campagnes mobiles de mesure. Elle peut constituer, après quelques développements, une étude à part entière sur les caractéristiques des concentrations de polluants mesurés dans une ville, une zone ou une région.

Il ressort des pré-études réalisées sur des données de plusieurs années consécutives à Lille, Marseille, Toulon et Toulouse, que les roses des vents et les roses de pollution sont dans l'ensemble reproductibles d'une année sur l'autre, et par suite, il n'y a pas nécessité de les refaire. Elles sont néanmoins très utiles lorsque l'on ne connaît pas exactement les limites des classes de vitesse de vent et de température. Il apparaît utile également de connaître les étendues géographiques où ces limites de classe prévalent.

L'application de la norme ISO 9359 aux données des villes mentionnées ci-dessus apporte des éléments de réponse sur la mise en œuvre de campagnes de mesures ponctuelles et sur le traitement possible des données qui en sont issues.

Afin de s'assurer de la robustesse statistique de la méthode, l'étude avait été appliquée à d'autres simulations de campagnes de durées différentes et à d'autres moments de l'année au Havre et à Strasbourg [2]. La qualité des reconstitutions réalisées dépendait beaucoup des niveaux de pollution rencontrés lors des campagnes, soit trop faibles ou trop élevés, qui conduisent respectivement à des sous ou sur-estimations des reconstitutions.

En effet, la qualité des résultats, quelle que soit la méthode utilisée, est limitée par la durée réduite des campagnes qui peut ne pas refléter totalement les événements se déroulant sur une année entière. Ainsi des pointes élevées de pollution peuvent échapper à la mesure, surtout en proximité industrielle et automobile. L'incidence est plus importante sur les reconstitutions de nombres de dépassement de seuil que sur celles des moyennes.

Cette méthodologie a été comparée avec la méthode des plans de sondage élaborée par l'INERIS et Atmo-Poitou-Charentes, afin de déterminer leurs différences et surtout leurs complémentarités pour améliorer in fine la stratégie de mesure à l'aide de moyens mobiles.

### 3.13 REFERENCES

[1] - « ISO 9359 – Qualité de l'air – Echantillonnage aléatoire stratifié pour l'évaluation de la qualité de l'air ambiant »,

[2] - Jean-Luc HOUDRET – Influence de paramètres météorologiques sur la stratégie de mesure à l'aide de moyens mobiles - Rapport LCSQA 2004

# **4. PARTIE II : METHODE PLANS** **DE SONDAGE**

**Cette deuxième partie porte exclusivement sur la méthode des plans de sondage.** Elle se compose de trois chapitres :

**1.** Le premier décrit l'application que l'on peut faire de la théorie des sondages afin :

- **d'élaborer une stratégie d'échantillonnage dans le temps** à l'aide d'un moyen de mesure mobile ;
- **d'estimer une moyenne annuelle avec son intervalle de confiance** ou d'autres indicateurs utiles à l'évaluation de la qualité de l'air (dépassement de seuil, centile).

**2.** Dans un deuxième chapitre, des échantillonnages sont simulés sur des séries annuelles de données, selon des plans préétablis. Le but est de s'assurer que les programmes fournissent des résultats conformes à la théorie. Il s'agit également :

- d'évaluer les conditions d'échantillonnage les plus favorables à l'utilisation de la méthode, autrement dit les échantillonnages qui aboutissent à des intervalles de confiance estimés fiables et resserrés ;
- d'examiner la qualité d'estimation atteinte dans des circonstances plus réalistes et les précautions qui s'imposent alors pour interpréter les résultats.

**3.** Enfin, les tests accomplis en parallèle avec l'EMD complètent cette évaluation et permettent de préciser le cadre d'utilisation de la méthode.

Tous les calculs, de la définition d'un plan d'échantillonnage à l'estimation de moyennes et d'intervalles de confiance, sont effectués au moyen de programmes informatiques développés avec le logiciel R (logiciel de statistique en libre accès sur Internet, <http://cran.cict.fr/>). Une notice explicative et un mode d'emploi sont joints en annexe.

## 4.1 OBJECTIFS, THEORIE, PRINCIPE DE MISE EN OEUVRE

### 4.2 OBJECTIFS DE LA METHODE

La méthode des plans de sondage a un double rôle : avant la collecte des données, elle sert à définir une stratégie d'échantillonnage ; à l'issue des campagnes de mesure, elle permet d'en exploiter les résultats. On distingue ainsi deux catégories de fonctions :

**a) Celles qui, pour une précision souhaitée, assistent l'utilisateur dans l'élaboration d'un plan d'échantillonnage.** Dans notre cas, cette précision s'exprime comme la demi-largeur de l'intervalle de confiance à 95% qui encadre la moyenne. Elle est donnée de façon relative (en % de la moyenne) ou absolue (en  $\mu\text{g}/\text{m}^3$ ).

Les données d'entrée sont celles d'un site fixe supposé de mêmes caractéristiques que le point d'échantillonnage. Elles sont issues de l'année qui précède ou résultent de la moyenne de plusieurs années afin de s'affranchir de conditions singulières. Le résultat obtenu est :

- pour une précision fixée, un ensemble de configurations possibles d'échantillonnage. Chacune d'elle se définit par le nombre total annuel de jours de mesures, le nombre de strates, la taille des grappes.

*Le terme **strates** désigne les périodes de temps qui composent l'année, par exemple les saisons, et au sein desquelles la variabilité des concentrations est moindre que dans l'année entière. Chaque strate fait l'objet d'un échantillonnage.*

*Le terme **grappe** désigne la succession de mesures effectuées à chaque passage du moyen mobile, pendant une durée déterminée, par exemple sept jours de données quart-horaires.*

- pour une configuration donnée, le nombre optimal de passages dans chaque strate.

Par souci de simplification nous ne considérons que des strates et des grappes de même durée.

**b) Celles qui, après que l'échantillonnage a été réalisé, permettent d'estimer la moyenne annuelle et l'intervalle de confiance à 95% autour de cette dernière.** Si un seuil de concentration a été préalablement défini, le taux de dépassement de ce seuil et l'intervalle de confiance à 95% autour de ce taux sont estimés de la même manière. Différents estimateurs peuvent être utilisés : l'estimateur classique  $\pi$  ou l'estimateur redressé par régression linéaire, par la différence ou par le quotient (cf. description au §2.2). Dans ces trois derniers cas, les calculs font intervenir les données d'un site auxiliaire qui fonctionne en continu pendant la saison ou l'année étudiée.

Ces deux groupes de fonctions peuvent être mis en œuvre de manière indépendante.

### 4.3 DONNEES REQUISES

Seules des données de campagnes de mesure et de stations fixes sont nécessaires (Tableau 4). Comparée à une estimation par modélisation, la méthode des plans de sondage a l'avantage de ne pas requérir de variables externes, comme des paramètres de trafic ou de météorologie. Mais la connaissance de l'influence de ces variables sur les concentrations peut guider le choix d'une stratification.

Tableau 4 : Données nécessaires à l'application de la méthode des plans de sondage

Étape	Données requises	Données optionnelles
Définition d'un plan d'échantillonnage	Données d'une ou plusieurs stations fixes dans la zone d'étude, sur une ou plusieurs années précédant l'année de l'échantillonnage	
Estimation d'une moyenne annuelle (ou saisonnière)	Données d'échantillonnage avec un pas de temps quart-horaire, horaire, journalier, voire hebdomadaire ou bihebdomadaire (camions laboratoires ou tubes à échantillonnage passif)	Données d'une ou plusieurs stations fixes dans la zone d'étude, pour l'année de l'échantillonnage
Estimation d'un taux de dépassement de seuil	Données d'échantillonnage avec un pas de temps quart-horaire, horaire ou journalier, selon le support temporel du seuil (camion laboratoire)	Données d'une ou plusieurs stations fixes dans la zone d'étude, pour l'année de l'échantillonnage

### 4.4 ELEMENTS THEORIQUES

Quelques éléments théoriques, que nous présentons ici, sont nécessaires à la bonne compréhension de la méthode des plans de sondage. Pour plus de détails, on pourra se référer à l'ouvrage d'Y.Tillé (2001) et à l'article de F. Lavancier et al. (2002).



Comme il a été évoqué précédemment, l'objectif est double : il est d'établir un plan d'échantillonnage permettant d'**estimer une concentration moyenne annuelle avec une précision acceptable et de fournir une estimation assortie d'un intervalle de confiance**. Ce contrôle de l'erreur est impossible si l'échantillonnage est établi de façon déterministe, suivant la connaissance que l'on a du polluant et du terrain. C'est pourquoi il convient de se placer dans le cadre de **la théorie des sondages aléatoires**. Cela ne minimise en rien l'importance de l'expérience, qui peut aider, sans sortir du cadre probabiliste, à trancher entre divers plans de sondage possibles.

#### 4.4.1 DEFINITION D'UN PLAN DE SONDRAGE

##### Principe :

Combien de données faut-il tirer aléatoirement dans l'année ou la saison pour estimer avec une certaine précision la concentration moyenne annuelle ou saisonnière d'un polluant ? Ce nombre dépend de la variance de l'estimateur de la moyenne, qui, dans le cas idéal (excellente précision), doit être la plus faible possible. En effet, si pour un même plan de sondage, un estimateur varie peu selon l'échantillon tiré, la valeur qu'il fournit à chaque tirage peut être considérée comme fiable.

Soit  $\hat{y}$  l'estimateur de la moyenne  $\bar{y}$ , égal à la moyenne des données de l'échantillon :  $\hat{y} = \sum_{i=1}^n y_i$  (avec  $y$  : variable de concentration)

La variance de cet estimateur peut s'écrire :  $\text{var}(\hat{y}) = \frac{N-n}{n.N} \cdot S^2$  où

$S^2 = \frac{N}{N-1} \text{var}(y)$  est la variance corrigée de  $y$ .

$N$  est le nombre total de données dans l'année ou la saison étudiée, et  $n$ , le nombre de données dans l'échantillon.

L'intervalle de confiance à 95% autour de la moyenne

est  $IC_{95\%} = \left[ \hat{y} - 1,96 \sqrt{\frac{N-n}{n.N} \cdot S^2} ; \hat{y} + 1,96 \sqrt{\frac{N-n}{n.N} \cdot S^2} \right]$ . La largeur  $L_{IC95\%}$  de cet

intervalle vaut donc  $L_{IC95\%} = 2 \cdot 1,96 \cdot \sqrt{\frac{N-n}{n.N} \cdot S^2}$

Pour des raisons de coût et de disponibilité du moyen mobile, le nombre  $n$  de données est nécessairement limité. Si l'on fixe une erreur à ne pas dépasser, erreur exprimée comme la largeur  $L$  ( $\mu\text{g}/\text{m}^3$ ) de l'intervalle de confiance autour de la moyenne, alors  $n$  se déduit de  $L$  et de  $S^2$  par la relation ci-dessus.

S étant inconnue au point d'échantillonnage, on est contraint d'estimer cette variance sur une série auxiliaire de données. Il s'agit par exemple des mesures en continu effectuées l'année précédente à une station fixe que l'on pense être de mêmes caractéristiques que le site à sonder. Comme indiqué en 4.2, il peut être pertinent de prendre une série auxiliaire moyenne sur plusieurs années.

### Pratique :

Les mesures individuelles fournies par les analyseurs des moyens mobiles sont des concentrations quart-horaires. Tirer de façon indépendante  $n$  valeurs quart-horaires dans l'année n'est toutefois pas envisageable en pratique mais il est possible de tirer aléatoirement des groupes de mesures, par exemple des jours ou des semaines de mesures. Ces ensembles sont désignés sous le terme de **grappes**. Une nouvelle formule de la variance de l'estimateur est alors proposée pour ce type d'échantillonnage. Dans cette étude nous considérons uniquement des **grappes de même taille** (i.e. constituées d'un même nombre de jours).

D'autre part, un **échantillonnage aléatoire stratifié** judicieusement défini est préférable à un échantillonnage aléatoire simple. En effet, si les grappes se répartissent entre des sous-ensembles homogènes de la série à sonder, alors deux échantillons tirés selon le même plan fournissent théoriquement des estimations assez proches : la variabilité de l'estimateur est donc moindre. Ces sous-ensembles, appelés **strates**, peuvent être déterminés selon la connaissance qu'on a du terrain et du comportement des polluants. Dans notre cas, il s'agit de divisions de l'année, par exemple des saisons, au sein desquelles les concentrations sont supposées plus homogènes que dans l'année entière. Par simplification, nous ne considérons là encore que des **strates de même taille**. Les grappes sont ensuite tirées aléatoirement dans chacune d'elles.

■ **En résumé, le type de plan choisi dans cette étude est un plan par grappes, de préférence stratifié.** Pour une largeur maximale autorisée de l'intervalle de confiance autour de la moyenne, des formules adéquates permettent d'évaluer, grâce à une série auxiliaire de données, le nombre minimal de grappes à tirer dans l'année et la répartition de ces grappes entre les strates (cf. annexe 1). Notons que **le plan de sondage est établi d'après les données d'un site auxiliaire**. Aussi n'est-il pas nécessairement optimisé pour le point d'échantillonnage si ce dernier ne possède pas les caractéristiques attendues. Néanmoins, le calcul de la moyenne et de son intervalle de confiance reste toujours possible.

### ■ Entre plusieurs plans possibles, comment choisir ?

Pour une même erreur fixée par l'utilisateur, la théorie des sondages offre à celui-ci un choix de plusieurs plans (nombre de strates, taille des grappes et nombre total de grappes). Divers aspects entrent en jeu dans la sélection de l'un d'entre eux :

- la connaissance du polluant : on choisira de préférence une stratification qui correspond à la physique du phénomène, de façon que les strates soient homogènes.
- les exigences de la théorie des sondages :
  - \* Il faut plus d'une grappe par strate pour estimer la variance (cf § ci-après).
  - \* Les formules d'estimation de l'intervalle de confiance à 95% reposent sur l'hypothèse qu'une infinité de grappes a été tirée. Il est généralement admis qu'avec un minimum de 30 grappes, on se situe dans le cadre de cette hypothèse. Un échantillonnage aussi contraignant n'est pourtant guère envisageable. En pratique, on s'efforcera de prélever un nombre suffisant de grappes ; celui-ci ne devrait pas descendre en dessous de la dizaine. Les simulations faites aux chapitres suivants apportent à ce sujet des éléments d'appréciation supplémentaires.
- les contraintes de personnel et de coût, la disponibilité du matériel.

S'il est difficile de respecter strictement les hypothèses de la méthode, on peut tenter de s'en approcher. Les calculs effectués par la suite montrent que dans de nombreuses situations, l'application de la méthode fournit des résultats satisfaisants. Ils permettent également de mieux cerner les limites de l'approche.

#### 4.4.2 ESTIMATION D'UNE MOYENNE ET D'UN TAUX DE DEPASSEMENT DE SEUIL

Une fois l'échantillonnage réalisé, comment sont estimés la moyenne et son intervalle de confiance ?

##### ***a) Estimation de la moyenne à partir des données de l'échantillon :***

L'estimateur le plus simple de la moyenne, le  $\pi$ -estimateur, est la moyenne des données de l'échantillon. (Notons que pour le plan par grappes, **une donnée correspond à la moyenne d'une grappe.**) Si la série était mesurée continûment sur l'année, sa variance serait parfaitement connue et l'on pourrait en déduire l'**intervalle de confiance théorique** autour de la moyenne. Or l'échantillon ne fournit qu'une information partielle sur la série. La variance de l'estimateur doit donc être estimée d'après les données dont il se compose (cf. annexe 1, formules (3) et (4)). **L'intervalle de confiance qui en découle est un intervalle estimé. La qualité de cette estimation dépend du nombre de données dans l'échantillon, c'est-à-dire dans notre cas du nombre de grappes.**

##### ***b) Prise en compte d'une information auxiliaire :***

L'information fournie en continu par des stations de mesure fixes (information auxiliaire) peut être mise à profit pour corriger (*redresser*) les estimations. Il existe trois types d'estimateurs redressés :

- **l'estimateur par la différence** : le  $\pi$ -estimateur est corrigé par un facteur additif.

$$\hat{y}_D = \hat{y} + (\bar{y}_{aux} - \hat{y}_{aux})$$

- **l'estimateur par le quotient** : le  $\pi$ -estimateur est corrigé par un facteur multiplicatif.

$$\hat{y}_Q = \hat{y} \frac{\bar{y}_{aux}}{\hat{y}_{aux}}$$

- **l'estimateur par la régression** : le  $\pi$ -estimateur est corrigé par régression linéaire.

$$\hat{y}_R = \hat{y} + \hat{b} \cdot (\bar{y}_{aux} - \hat{y}_{aux})$$

A chacun est associée une expression de l'intervalle de confiance à 95% de la moyenne, qui tient compte du degré de proximité entre la série échantillonnée et la série auxiliaire. Celui-ci est quantifié par la covariance calculée dans l'échantillon :

$$\hat{S}_{xy} = \frac{1}{n-1} \sum_{i=1}^N (x_i - \hat{x}) \cdot (y_i - \hat{y}) \quad \text{où } x_i \text{ et } y_i, \text{ respectivement } \hat{x} \text{ et } \hat{y}, \text{ sont les données,}$$

resp. la moyenne des données, de la série d'intérêt et de la série auxiliaire pendant la période d'échantillonnage.

Comme précédemment, l'intervalle de confiance autour de la moyenne n'est qu'un **intervalle estimé**.

### **c) Estimation d'un taux de dépassement de seuil :**

L'estimation de centiles est particulièrement délicate car elle exige des échantillons de grande taille, c'est-à-dire constitués d'un très grand nombre de grappes, ce qui est rarement le cas. Mais afin de s'assurer du bon respect des valeurs réglementaires, on peut chercher à estimer des **taux de dépassement de seuil**. Un taux de dépassement de seuil est le pourcentage de données supérieures à un seuil fixé. Plusieurs valeurs réglementaires s'expriment en effet comme une concentration limite à ne pas dépasser plus d'un certain nombre de fois dans l'année.

En remplaçant les concentrations mesurées par 1 si le seuil considéré est dépassé et par 0 s'il ne l'est pas, l'estimation du taux de dépassement s'identifie à l'estimation d'une moyenne. **Il faut cependant veiller à ce que dans chaque grappe, le pas de temps des données coïncide avec la base temporelle de la valeur seuil (1h, 8h, 1 jour...).**

Notons qu'en procédant par essais successifs, par exemple par dichotomie, on peut approcher une valeur de centile. Il suffit de fixer à chaque essai une valeur seuil différente et d'estimer le taux de dépassement associé. Lorsque ce pourcentage correspond au centile recherché (ex : 2% des valeurs horaires si le centile recherché est le 98<sup>e</sup> centile des concentrations horaires), alors l'estimateur du centile est la dernière valeur seuil fixée. Cette procédure ne permet cependant pas d'estimer un intervalle de confiance autour du centile. Elle n'est pas testée dans cette étude.

#### d) Remarque sur l'échantillonnage

**Dans l'estimation de la moyenne et de son intervalle de confiance, seules interviennent les valeurs moyennes des grappes.** Si les grappes sont de durée hebdomadaire ou bihebdomadaire, **des mesures par tube à échantillonnage passif conviennent.** En revanche, comme on l'a fait remarquer, **l'estimation d'un taux de dépassement de seuil exige d'effectuer des mesures avec une résolution au moins aussi fine que le support temporel de ce seuil (horaire, journalier).**

### 4.5 MISE EN ŒUVRE

Toutes les équations nécessaires à l'application de la méthode ont été codées avec R, logiciel de statistique en libre accès sur Internet. Les programmes informatiques sont adaptés des programmes initialement développés sous MATLAB ATMO Poitou-Charentes. Ils seront mis à disposition sur le site Internet du LCSQA ([www.lcsqa.org](http://www.lcsqa.org)).

Chaque étape fait appel à une ou plusieurs fonctions, comme l'indique la notice descriptive en annexe. Un mode d'emploi de ces fonctions est également joint au rapport. Le Tableau 5 indique le déroulement de la méthode.

*Tableau 5 : application de la méthode des plans de sondage – étapes*

D : définition d'un plan d'échantillonnage

E : estimation d'une moyenne annuelle et de son intervalle de confiance

	<b>Etapes</b>	<b>Commentaire</b>
D	Choix d'un site fixe auxiliaire de caractéristiques similaires à celles du site étudié.	Fait par l'utilisateur
D	On fixe une incertitude maximale (demi-largeur de d'intervalle de confiance à 95% (en % de la moyenne ou en $\mu\text{g}/\text{m}^3$ ).	Fait par l'utilisateur
D	Sur la série auxiliaire de données, on recherche le nombre total de grappes qui permet de respecter cette précision.	Programme R
D	Choix d'une taille de grappe, d'un nombre de strates et d'un nombre total de grappes.	Fait par l'utilisateur
D	Sur la même série auxiliaire, on recherche la répartition optimale des grappes entre les strates.  (Rem : Pour certains polluants, l'évolution temporelle des concentrations rend pertinent un découpage en deux strates : été et hiver. Afin de définir le nombre optimal de grappes dans chacune de	Programme R

	ces strates, il convient de travailler sur une série auxiliaire allant d'octobre à fin septembre.)	
E	Recherche d'un site fixe auxiliaire corrélé avec le site d'échantillonnage.	Fait par l'utilisateur
E	Disposant d'un échantillon de valeurs mesurées par le moyen mobile, on lui associe les données aux mêmes dates du site fixe auxiliaire, si celui-ci existe	Fait par l'utilisateur en situation réelle Fait par le programme de simulation d'échantillonnage
E	Estimation de la concentration moyenne annuelle et de son intervalle de confiance associé <ul style="list-style-type: none"> <li>- par l'estimateur PI, si aucune station auxiliaire n'est disponible,</li> <li>- par l'estimateur PI et par l'estimateur redressé par la régression, la différence ou le quotient si une station auxiliaire adéquate est présente dans le domaine d'étude.</li> </ul>	Programme R
E	Si nécessaire, estimation d'un taux de dépassement de seuil et de son intervalle de confiance associé <ul style="list-style-type: none"> <li>- par l'estimateur PI, si aucune station auxiliaire n'est disponible,</li> <li>- par l'estimateur PI et par l'estimateur redressé par la régression, la différence ou le quotient si une station auxiliaire adéquate est présente dans le domaine d'étude.</li> </ul>	Programme R

Les parties D et E sont indépendantes. **La partie D est facultative** : on peut ainsi procéder à l'estimation même si l'échantillonnage n'a pas été établi d'après les données d'un site auxiliaire. **Cet échantillonnage doit néanmoins respecter les conditions d'un échantillonnage aléatoire (ou aléatoire stratifié).**

#### 4.6 VERIFICATION DES PROGRAMMES

Ce chapitre constitue un préalable nécessaire à l'application de la méthode. Son but est de s'assurer de la cohérence entre les résultats attendus en théorie et ceux que l'on obtient en pratique. Cette vérification repose sur la réalisation de simulations à l'aide des programmes R.

#### 4.7 PRINCIPE DES SIMULATIONS

Les simulations sont fondées sur l'exploitation de séries annuelles de mesures issues de stations fixes. Dans chaque situation étudiée, les données de deux stations sont utilisées : l'un des deux sites est considéré comme le point à sonder et l'autre, comme un site auxiliaire.

**Pour un plan de sondage préétabli**, les simulations consistent à tirer un très grand nombre d'échantillons conformes à ce plan et à calculer à chaque tirage :

- l'estimation de la moyenne annuelle, avec et sans redressement ;
- l'intervalle de confiance à 95% qui encadre cette estimation ;  
NB : deux intervalles sont en fait calculés, l'un, à partir de la variance estimée -intervalle dit «estimé »-, l'autre, à partir de la variance théorique -intervalle dit «théorique »-
- l'estimation du pourcentage de dépassements de seuil, avec ou sans redressement, et les intervalles de confiance à 95% estimé et théorique qui lui sont associés.

En raison des temps de calcul, le nombre de tirages est limité à 2000.

Le résultat final (moyenne estimée de la variable, borne inférieure, borne supérieure et largeur de l'intervalle de confiance à 95%) est une moyenne sur tous les tirages. **Le taux de couverture indique le pourcentage de tirages pour lesquels l'intervalle de confiance estimé ou théorique contient effectivement la vraie moyenne (qui, ici, est connue). Si le plan de sondage satisfait aux exigences de la théorie, soit plus de 30 mesures indépendantes, il devrait avoisiner 95%.**

Remarque : Pour calculer la variance intergrappe théorique, la fonction programmée dans R découpe préalablement la série annuelle en grappes successives de la taille considérée. Ainsi, s'il s'agit de grappes hebdomadaires, les débuts de grappes sont invariablement les suivants : 1<sup>er</sup>, 8, 15, 22, 29 janvier, 5 février.... Or, dans les simulations, n'importe quelle date peut être choisie comme début de grappe afin de multiplier les possibilités d'échantillonnage, par exemple le 2, le 14, le 24 janvier... Calculer la variance théorique rigoureusement associée à ce tirage représente un problème complexe. Néanmoins, on peut considérer que la variance précédemment calculée en constitue une bonne approximation.

## 4.8 RESULTATS

Les simulations ont été réalisées pour le dioxyde d'azote à l'aide de données de concentration horaires fournies par AIRMARAIX (Tableau 6).

Tableau 6 – Vérification des programmes – échantillonnages simulés

Polluant	Station supposée mobile	Station fixe de référence	Taille des grappes (durée de passage du moyen mobile)	Nombre de grappes dans l'année ***	Durée cumulée relative (%)	
NO <sub>2</sub>	Cinq Avenues (U)	Sainte-Marguerite (U)	1 h	[1752]	20*	1
			1 h	[876 ;876]		2
			1 h	[438 ;438 ;438 ;438]		3
			24 h	[73]		4
			24 h	[37 ;36]		5
			24 h	[19 ;18 ;18 ;18]		6

		168 h (1 sem.)	[11]		7
		168 h (1 sem.)	[6 ;5]		8
		1 h	[1227]	14**	9
		1 h	[614 ;613]		10
		1 h	[307 ;307 ;306 ;307]		11
		24 h	[52]		12
		24 h	[26 ;26]		13
		24 h	[13 ;13 ;13 ;13]		14
		168 h (1 sem.)	[8]		15
		168 h (1 sem.)	[4 ;4]		16

\* durée minimale préconisée dans le Guide Technique Européen (EEU, 1998)

\*\* durée minimale préconisée pour le SO<sub>2</sub>, le NO<sub>2</sub> et les NO<sub>x</sub> dans la Directive 1999/30/CE

\*\*\* [n] : n grappes prélevées aléatoirement dans l'année

[n1 ; n2] : n1 grappes prélevées aléatoirement en hiver (janvier-mars + octobre-décembre) ;

n2 grappes prélevées aléatoirement en été (avril-septembre)

[n1 ;n2 ;n3 ;n4] : n1, n2, n3, n4 grappes prélevées respectivement dans chaque trimestre.

La moyenne annuelle réelle à la station Cinq Avenues est de 38,2 µg/m<sup>3</sup> en 1999 et de 33,5 µg/m<sup>3</sup> en 2000. Le centile 90, i.e. la valeur de concentration dépassée pendant 10% de l'année (en durée cumulée), est de 70 µg/m<sup>3</sup> en 1999 et de 64 µg/m<sup>3</sup> en 2000 (Tableau 7).

Tableau 7 – Statistiques du NO<sub>2</sub> mesuré en 1999 et 2000 aux stations urbaines Cinq Avenues et Sainte-Marguerite

Station	Cinq Avenues		Sainte-Marguerite	
	1999	2000	1999	2000
% de données valides	98,6	99,1	98,6	97,4
Moyenne (µg/m <sup>3</sup> )	38,2	33,5	30,2	26,3
Variance (µg/m <sup>3</sup> ) <sup>2</sup>	546,1	444,8	566,7	452,3
Min (µg/m <sup>3</sup> )	1	1	0	0
P90 (µg/m <sup>3</sup> )	70	64	64	57
Max (µg/m <sup>3</sup> )	168	156	180	170
Corrélation entre les moyennes mesurées par les deux stations	0,785 en 1999 0,787 en 2000			



## 4.8.1 GRAPPES D'UNE HEURE

### 4.8.1.1 ESTIMATION DE LA MOYENNE ANNUELLE

- **En moyenne sur les 2000 tirages**, la concentration moyenne annuelle estimée est très proche de la concentration moyenne annuelle mesurée ; l'écart est inférieur à 0,1%.
- Le taux de couverture de l'intervalle de confiance à 95% estimé avoisine, voire excède 95%, quels que soient le plan et l'estimateur. Il diffère peu du taux de couverture de l'intervalle de confiance théorique.
- La précision de l'estimation est excellente. En moyenne sur les simulations, l'incertitude relative est inférieure à 3%, quel que soit l'estimateur considéré. Le redressement agit cependant : il réduit légèrement la largeur de l'intervalle de confiance. Par ordre croissant de précision, le classement des estimateurs est le suivant :  $\pi$ , quot, diff, reg.
- La précision augmente avec le nombre de grappes, bien qu'ici, la différence soit peu sensible. A taux de confiance voisins, l'incertitude relative est de 3,2% pour 1227 heures de mesure tirées aléatoirement dans l'année et de l'ordre de 2,5% pour 1752 jours de mesure.
- Compte tenu du grand nombre de grappes et de la qualité de l'estimation, la stratification n'a pas d'effet notable.

### 4.8.1.2 ESTIMATION DU TAUX DE DEPASSEMENT DE SEUIL

Les remarques faites pour l'estimation de la moyenne s'appliquent également à l'estimation du taux de dépassement de seuil.

## 4.8.2 GRAPPES DE 24 HEURES

### 4.8.2.1 ESTIMATION DE LA MOYENNE ANNUELLE

- **En moyenne sur les 2000 tirages**, la concentration moyenne annuelle estimée est très proche de la concentration moyenne annuelle mesurée ; l'écart est inférieur à 0,3%.
- Le taux de couverture de l'intervalle de confiance à 95% estimé avoisine 95%, quel que soit le plan et l'estimateur. Celui de l'intervalle de confiance théorique, compris entre 95% et 97% environ, est légèrement supérieur.
- Le redressement améliore la précision de l'estimation. Par ordre croissant de précision, le classement des estimateurs est encore le suivant :  $\pi$ , quot, diff, reg. Ainsi, quand le nombre de grappes est élevé, l'estimateur reg se révèle le plus performant. Lorsqu'en revanche ce nombre diminue (moins de 15 grappes par strate), le taux de couverture associé décroît et l'usage des autres estimateurs est préférable.
- La précision augmente avec le nombre de grappes. A taux de confiance voisins, l'incertitude relative est de 10,3% pour 52 jours de mesure tirés aléatoirement dans l'année et de 8,4% pour 73 jours de mesure.

- La précision s'accroît lorsque le plan de sondage est stratifié. A taux de confiance voisins, l'incertitude relative moyenne est de 8,4% pour 73 jours de mesure tirés aléatoirement dans l'année et de 4,0% pour 37 et 36 jours de mesure tirés respectivement dans les strates hivernale et estivale. Les résultats pour deux ou quatre strates sont très proches.

#### 4.8.2.2 ESTIMATION DU TAUX DE DEPASSEMENT DE SEUIL

- **En moyenne sur les 2000 tirages**, le taux de dépassement de seuil estimé est proche du taux de dépassement réel ; l'écart est inférieur à 1,0% pour les estimateurs  $\pi$ , reg et diff et à 2,3% pour l'estimateur quot.

- Le taux de couverture de l'intervalle de confiance à 95% estimé avoisine 95% pour les estimateurs  $\pi$  et diff. Celui de l'intervalle de confiance théorique, compris entre 95% et 97%, est légèrement supérieur. **Les résultats obtenus pour les estimateurs reg et quot sont de moindre qualité**, avec un taux de confiance estimé compris entre 89 et 93,5% selon les cas.

- Le redressement accroît la précision de l'estimation. L'incertitude relative moyenne diminue presque de moitié lorsqu'on redresse la moyenne par la différence. Notons qu'en moyenne sur les simulations, l'estimateur  $\pi$  ne permet pas de conclure sur le dépassement du seuil annuel de  $40 \mu\text{g}/\text{m}^3$ . Le redressement par la différence, quant à lui, fournit une borne supérieure d'intervalle comprise entre 39,7 et  $40,1 \mu\text{g}/\text{m}^3$  selon les plans.

- La précision augmente avec le nombre de grappes. A taux de confiance voisins, l'incertitude relative moyenne est de 37,3% pour 52 jours de mesure tirés aléatoirement dans l'année et de 30,5% pour 73 jours de mesure. La stratification fait peu gagner en précision.

- En moyenne sur les 2000 simulations, l'incertitude relative est 2,5 à 4,5 fois plus élevée que l'incertitude associée à l'estimation de la moyenne.

### 4.8.3 GRAPPES D'UNE SEMAINE

#### 4.8.3.1 ESTIMATION DE LA MOYENNE ANNUELLE

- En moyenne sur les 2000 tirages, la concentration moyenne annuelle estimée est très proche de la concentration moyenne annuelle mesurée ; l'écart est inférieur à 1,3%.

- A durée totale de mesure identique, le taux de couverture de l'intervalle de confiance est plus faible qu'avec des grappes d'une journée. Ce résultat n'est pas étonnant puisque les grappes d'une semaine sont nettement moins nombreuses.

L'estimateur redressé par la régression est le moins fiable, ce qui concorde avec la remarque faite en 4.8.2.1. Le taux de couverture de l'intervalle de confiance à 95% est inférieur à 92% ; il chute à 81% pour un plan 4-4. Pour les autres estimateurs, ce taux varie entre 90 et 94% selon les plans.

Les taux de couverture des intervalles théoriques, compris entre 94% et 97%, montrent que si la vraie variance intergrappe était connue, alors les intervalles de confiance à 95% seraient correctement estimés.

- Le redressement par la différence ou le quotient améliore le taux de couverture et la précision de l'estimation. Par ordre croissant de taux de couverture, le classement des estimateurs est généralement le suivant : reg,  $\pi$ , quot, diff. Par ordre croissant de précision il devient :  $\pi$ , quot, diff, reg. Le croisement de ces deux classements indique ici la plus grande efficacité d'un redressement par la différence.

- La précision augmente légèrement avec le nombre de grappes, ainsi qu'avec la stratification (pourvu que le nombre de grappes par strate ne soit pas trop faible).

#### 4.8.3.2 ESTIMATION D'UN TAUX DE DEPASSEMENT DE SEUIL

- En moyenne sur les 2000 tirages, le taux de dépassement estimé est très proche du taux mesuré (écart inférieur à 3,6%).

- A durée totale de mesure identique, le taux de couverture de l'intervalle de confiance est sensiblement plus faible qu'avec des grappes journalières. Dans le cas de l'estimateur  $\pi$ , il ne dépasse 90% que pour un plan de 11 grappes. L'estimation est alors assez peu précise (incertitude relative de 38 à 47%). Seul l'estimateur redressé par la différence affiche des taux de couverture supérieurs à 90% pour l'ensemble des plans testés (90-91% pour un total de 8 grappes, 93-94% pour un total de 11 grappes). L'incertitude relative associée varie de 22% pour un plan de 11 grappes à 27% pour un plan de 8 grappes.

- La stratification améliore la précision de l'estimateur  $\pi$ . Avec les autres estimateurs, elle n'a pas d'effet notable.

#### 4.8.3.3 AMELIORATION DE L'ESTIMATION DES INTERVALLES DE CONFIANCE

**Le tirage d'un faible nombre de grappes ne favorise pas l'estimation correcte de l'intervalle de confiance à 95%.** Plus adapté cependant aux réalités du terrain, ce type de stratégie est couramment appliqué dans les AASQA. Afin d'accroître le taux de confiance des intervalles estimés, deux procédures sont testées.

La première, valable pour un plan stratifié, consiste à regrouper les strates au moment de l'estimation. Ainsi, étant donné un plan [4,4], tout se passe comme si les huit grappes appartenaient à la même strate.

La seconde consiste à estimer des intervalles à des taux de confiance supérieurs (96%, 97%, 98% ou 99%) de façon que le taux de confiance effectivement atteint s'approche des 95% souhaités.

Dans les deux approches, l'intervalle estimé n'est pas le véritable intervalle associé au plan d'échantillonnage. Il est théoriquement plus large mais il est plus fiable.

#### **Première approche**

##### *Estimation de la moyenne*

Le regroupement des strates agit favorablement sur le taux de couverture.

Pour un plan [6-5], celui-ci devient voisin de 95%, exception faite de l'estimateur reg. En contre-partie, l'incertitude moyenne augmente mais de façon limitée (elle demeure inférieure à 14%).

Pour un plan [4,4], le gain est également significatif. Après regroupement, le taux de couverture associé à l'estimateur  $\pi$  est de 95,5% pour une incertitude de 16,1%. Toutefois la précision de l'estimation ne permet pas d'établir s'il y a ou non dépassement du seuil annuel de 40  $\mu\text{g}/\text{m}^3$ .

### *Estimation du dépassement de seuil*

Là encore, le regroupement des strates augmente le taux de couverture, en particulier pour l'estimateur  $\pi$ . Dans le cas d'un plan [6-5], il permet d'approcher un taux de 95% pour les estimateurs  $\pi$  et diff, avec une incertitude moyenne de 46% et 22%. En revanche, dans le cas d'un plan [4-4], il ne permet pas d'accroître le taux de couverture au-delà de 91%.

*Tableau 8 – Estimation de la moyenne annuelle. Effet d'un regroupement des strates sur l'estimation de l'intervalle de confiance*

Taille des grappes (h)	Nombre de grappes par strate	Estimateur	Taux de couverture de l'intervalle estimé (%)		Incertitude relative (%) (moyenne sur les 2000 simulations)	
			Sans regrouper	En regroupant	Sans regrouper	En regroupant
168	6-5	PI	90.5	96.8	10.9	13.5
168	6-5	REG	90.4	92.3	5.9	6.5
168	6-5	DIFF	93.7	94.6	6.9	7.4
168	6-5	QUOT	93.8	95.9	8.4	9.7
168	4-4	PI	90.1	95.5	13.5	16.1
168	4-4	REG	82.3	90.4	6.7	7.4
168	4-4	DIFF	91.0	94.1	8.7	8.8
168	4-4	QUOT	91.5	93.9	10.5	11.4

### **Seconde approche**

Des simulations complémentaires ont été réalisées en estimant des intervalles à différents degrés de confiance (de 95% à 99%). Les résultats obtenus dans cet exemple indiquent qu'on peut espérer atteindre un taux de couverture de 95% en estimant des intervalles à 96%, 97%, 98% voire 99% selon les cas (Tableau 9).

Pour l'estimation du taux de dépassement de seuil, seuls l'estimateur  $\pi$  et l'estimateur redressé par la différence conduisent à des taux de couverture satisfaisants. L'incertitude est cependant élevée, surtout en l'absence de redressement.

*Tableau 9 – Recherche d'un taux de couverture de 95%. Estimation de l'IC à différents degrés de confiance*

**Estimation de la moyenne**

Taille des grappes (h)	Nombre de grappes par strate	Estimateur	Degré de confiance théorique de l'intervalle estimé	Taux de couverture de l'intervalle estimé	Incertitude relative (moyenne sur les 2000 simulations)
168	11	PI	96	94.4	14.0
			97	95.3	14.8
168	11	REG	96	94.5	6.6
			97	95.2	7.0
168	11	DIFF	96	95.3	7.2
168	11	QUOT	96	94.6	9.7
			97	95.6	10.2
168	8	PI	99	95.7	21.1
168	8	REG	98	94.8	8.6
			99	96.5	9.6
168	8	DIFF	97	94.9	9.4
			98	96.1	10.1
168	8	QUOT	98	94.5	13.0
			99	96.1	14.4
168	6-5	PI	97	94.2	12.3
			98	95.6	13.2
168	6-5	REG	98	94.1	7.0
			99	95.6	7.7
168	6-5	DIFF	96	94.8	7.3
			97	95.7	7.7
168	6-5	QUOT	95	94.8	8.3
			96	95.5	8.8
168	4-4	PI	98	95.0	16.3
168	4-4	REG		<i>Taux&lt;91</i>	
168	4-4	DIFF	99	95.4	11.3
168	4-4	QUOT	99	95.3	13.6

**Estimation du taux de dépassement de seuil**

Taille des grappes (h)	Nombre de grappes par strate	Estimateur	Degré de confiance théorique de l'intervalle estimé	Taux de couverture de l'intervalle estimé	Incertitude relative (moyenne sur les 2000 simulations)
168	11	PI	99	95.3	30.2
168	11	REG		<i>Taux&lt;91</i>	
168	11	DIFF	97	94.3	24.8
			98	95.1	26.6

168	11	QUOT	99	94.3	30.2
168	8	PI		Taux<92	
168	8	REG		Taux<87	
168	8	DIFF	98	95.1	31.6
168	8	QUOT		Taux<93	
168	6-5	PI	98 99	94.4 95.7	46.3 51.3
168	6-5	REG		Taux<91	
168	6-5	DIFF	98	95.7	25.1
168	6-5	QUOT	99	94	29.5
168	4-4	PI	99	94.1	65.7
168	4-4	REG		Taux<85	
168	4-4	DIFF	99	95.7	34.3
168	4-4	QUOT		Taux<92	

### Comparaison des deux approches

Dans les exemples étudiés, le regroupement des strates et l'estimation d'intervalles de confiance à 97% ou 98% fournissent des résultats comparables. Le regroupement des strates est néanmoins préféré pour sa plus grande robustesse car il permet une estimation plus fiable de la variance (l'estimation se passe comme si l'on avait tiré dans une seule strate un plus grand nombre de grappes).

#### 4.8.4 SIMULATIONS COMPLEMENTAIRES

Les simulations décrites aux paragraphes précédents ont mis en évidence :

- pour une même durée cumulée de mesure, **l'effet conjoint du nombre et de la taille des grappes** ;
- **pour une même taille de grappe, l'effet du nombre de grappes.**

Des simulations complémentaires sont effectuées afin d'évaluer isolément **l'effet de la taille des grappes** : on applique à des grappes d'une journée les plans d'échantillonnage définis pour des grappes d'une semaine.

L'estimation de la moyenne annuelle est assez satisfaisante. La réduction du temps de passage du moyen mobile fait perdre en précision mais l'incertitude relative reste inférieure à 30% pour l'estimateur  $\pi$  et 15% pour l'estimateur diff. Toutefois, les intervalles de confiance moyens (sur les 2000 simulations) sont trop larges pour déterminer s'il y a ou non dépassement du seuil annuel. L'estimation du pourcentage de dépassement de seuil est d'une qualité très inférieure.

#### 4.8.5 TESTS SUR UNE AUTRE ANNEE

Des simulations identiques ont été réalisées sur les données de l'année 2000, pour des grappes hebdomadaires ou journalières. Dans l'ensemble, elles confirment les observations faites pour l'année 1999 mais le regroupement des strates a un effet plus marqué pour l'estimateur quot que pour l'estimateur diff.

Tableau 10 – Estimation de la moyenne annuelle. Effet d'un regroupement des strates sur l'estimation de l'intervalle de confiance

Taille des grappes (h)	Nombre de grappes par strate	Estimateur	Taux de couverture de l'intervalle estimé (%)		Incertitude relative (%) (moyenne sur les 2000 simulations)	
			Sans regrouper	En regroupant	Sans regrouper	En regroupant
168	6-5	PI	92.2	<b>98.8</b>	10.4	14.5
168	6-5	REG	87.1	90.5	4.7	5.0
168	6-5	DIFF	92.7	93.0	5.6	5.8
168	6-5	QUOT	93.3	<b>96.0</b>	7.2	8.4
168	4-4	PI	90.4	<b>97.6</b>	12.9	17.7
168	4-4	REG	83.2	89.6	5.1	5.9
168	4-4	DIFF	93.2	93.6	6.8	6.9
168	4-4	QUOT	92.5	<b>95.6</b>	8.8	10.1

#### 4.8.6 QUELLES CONCLUSIONS PRELIMINAIRES PEUT-ON TIRER SUR L'APPLICATION DE LA METHODE ?

- Les résultats obtenus s'accordent avec la théorie.
- Lorsque l'échantillon se compose d'un très grand nombre de grappes, la moyenne annuelle et le taux de dépassement de seuil sont estimés avec précision ; de plus, l'estimation des intervalles de confiance associés est fiable : il y a bien 95% de chances pour que ceux-ci contiennent la vraie moyenne.
- La méthode s'applique également à des configurations d'échantillonnage plus conformes à la pratique, c'est-à-dire à des grappes moins nombreuses et plus longues. **Toutefois, si le nombre annuel de grappes est faible, en particulier s'il est inférieur à la dizaine, le risque de mal estimer l'intervalle de confiance augmente.** Une façon de garantir un niveau de confiance plus proche de 95% est de grouper les strates avant l'estimation ;
- Pour un même nombre de passages (une douzaine par année dans les tests réalisés), des temps de séjour hebdomadaires assurent une meilleure précision d'estimation que des mesures d'une journée.

- Le redressement à l'aide des données d'une station fixe auxiliaire améliore la qualité de l'estimation : l'intervalle de confiance autour de la moyenne ou du taux de dépassement de seuil est mieux estimé et plus resserré. Dans les gammes de concentrations voisines de la valeur limite annuelle, ce gain de précision peut être utile pour conclure quant à un éventuel dépassement. **Le redressement par la régression n'est pas recommandé car pour être efficace, il requiert un grand nombre de grappes.** Le redressement par la différence ou par le quotient est plus satisfaisant.
- Avec des grappes en petit effectif (de l'ordre de la douzaine) et de brève durée (~1 jour), l'estimation de la moyenne annuelle reste acceptable. En revanche, l'estimation du taux de dépassement d'un seuil horaire est entachée d'une grande incertitude.

#### 4.9 ETUDE DE PLANS D'ÉCHANTILLONNAGE

Des campagnes d'échantillonnage sont simulées sur des séries annuelles complètes de mesures automatiques, qui correspondent à différentes conditions géographiques et différents polluants. La méthode de l'EMD et la méthode des plans de sondage sont appliquées en parallèle afin d'évaluer les avantages et limites de chacune ainsi que leur complémentarité. Nous présentons dans ce chapitre les résultats propres aux plans de sondage.

Dans tous les résultats de tests, la nomenclature suivante est adoptée :

IC-	Borne inférieure de l'intervalle de confiance
IC+	Borne supérieure de l'intervalle de confiance
PI 95 (resp. DIFF 95, QUOT 95)	Estimateur $\pi$ (resp. DIFF, QUOT) et intervalle de confiance à 95%
PI 95g (res. DIFF 95g, QUOT 95g)	Estimateur $\pi$ (resp. DIFF, QUOT) et intervalle de confiance à 95%, après regroupement des strates

#### 4.10 REMARQUES PRELIMINAIRES. CHOIX DES PARAMETRES DE SONDRAGE

Le choix des paramètres de sondage est guidé par les exigences suivantes :

- conformité au protocole de comparaison avec l'EMD (cf. partie III) ;
- réalisme des campagnes d'échantillonnage simulées et possibilité d'appliquer les résultats à des campagnes par tubes ;
- adéquation avec les hypothèses de la théorie (dans la mesure du possible).

On s'est ainsi limité à des grappes hebdomadaires ou bihebdomadaires.



Les plans d'échantillonnage proposés par les fonctions *choixn* et *choixnh* de la théorie des sondages (cf. annexe 3) n'ont qu'un caractère indicatif. L'utilisateur peut appliquer les programmes d'estimation à n'importe quel autre plan, pourvu que celui-ci respecte les hypothèses de la théorie des sondages (tirage aléatoire des grappes dans les strates, minimum d'une dizaine de grappes).

Par simplification, dans la mesure où l'élaboration d'un plan de sondage peut être soumise à des contraintes matérielles et à une connaissance du terrain dont il n'est pas ici question, on conserve les résultats de *choixn* et *choixnh*.

Lorsque le nombre de grappes requis en théorie est inférieur à 8, le risque de mal estimer la variance s'accroît. De ce fait, le tirage d'un échantillon de huit grappes est également simulé.

A l'inverse, un nombre de grappes élevé (plus d'une douzaine) est favorable à l'application de la méthode mais moins adapté à la pratique. Là encore, le tirage d'un échantillon de huit grappes est également simulé.

Les plans retenus et des dates d'échantillonnage tirées aléatoirement sont indiqués en annexe 3.

#### **4.11 RESULTATS DES ESTIMATIONS**

Les conclusions, recommandations et mises en garde qui découlent des résultats obtenus<sup>1</sup> sont présentées sous forme de questions/réponses. Elles reprennent et complètent les remarques faites au paragraphe 4.6.

#### **4.12 APPLICATION DE LA METHODE DES PLANS DE SONDRAGE : CONCLUSIONS, RECOMMANDATIONS, MISES EN GARDE**

##### **4.12.1 DEFINITION DE L'ECHANTILLONNAGE**

***Faut-il nécessairement disposer des données d'un site fixe auxiliaire ? Doit-on prendre le résultat des fonctions *choixn* et *choixnh* ?***

Non, ce n'est pas obligatoire. Les résultats de ces fonctions n'ont qu'un **caractère indicatif**. Ils sont d'autant plus pertinents que la station auxiliaire considérée a les mêmes caractéristiques que le site d'échantillonnage. D'autre part, il est préférable que la série de mesures utilisée ne corresponde pas à une année singulière (le cas échéant, une série moyenne sur plusieurs années peut être plus adaptée).

Pourvu qu'il respecte au mieux les exigences de la théorie des sondages, n'importe quel plan convient, indépendamment de tout site auxiliaire.

***Quelles sont les contraintes de la théorie des sondages ?***

- Les dates de début des grappes doivent être tirées de manière **aléatoire** (dans l'année ou dans chaque strate).

---

<sup>1</sup> Un extrait des résultats numériques est présenté à titre d'exemple en annexe 4. Les fichiers complets pourront être mis à disposition sur le site du LCSQA.

- **Quelle que soit la taille des grappes, celles-ci doivent être en nombre suffisant afin d'estimer correctement la variance et par suite, l'intervalle de confiance.**

Les résultats obtenus pour des durées de mesure hebdomadaires ou bihebdomadaires montrent qu'**en dessous de huit grappes, la qualité de l'estimation dépend de la présence ou non d'un site fixe corrélé**. Sans un tel site, l'intervalle de confiance est mal estimé (faible taux de couverture) ou relativement large (précision médiocre). Ainsi, lorsque ses bornes encadrent la valeur limite annuelle, il est impossible de conclure sur le dépassement ou non dépassement de cette valeur (*voir par exemple en annexe le cas du NO<sub>2</sub> mesuré à Strasbourg : avec 6 grappes de deux semaines, le taux de couverture de l'intervalle de confiance associé à l'estimateur PI est inférieur à 80%; quand les strates sont regroupées au moment de l'estimation, il augmente jusqu'à 99% environ mais la borne supérieure de l'intervalle de confiance dépasse alors 40 µg/m<sup>3</sup>*).

**Pour une bonne estimation de l'incertitude, un minimum de 10 grappes** est préconisé. Ce minimum peut être augmenté en fonction du polluant et de l'indicateur à estimer.

### ***Vaut-il mieux un échantillonnage aléatoire pur ou un échantillonnage aléatoire stratifié ?***

**De manière générale, une stratification été/hiver est recommandée** puisque les concentrations de nombreux polluants sont soumises à des influences saisonnières.

Dans la plupart des tests relatifs aux oxydes d'azote et aux particules, la fonction de répartition *choixnh* attribue une majorité de grappes à la strate hivernale. Dans le cas du monoxyde d'azote cependant, une répartition trop déséquilibrée des grappes (suréchantillonnage de l'hiver) peut conduire à une surestimation de la moyenne annuelle ; le redressement permet de corriger cet effet.

### ***L'échantillonnage dépend-il du polluant ?***

Oui, **le nombre de grappes à tirer est d'autant plus grand que la variabilité temporelle des concentrations est plus élevée**. Cette observation vaut notamment pour le monoxyde d'azote (Le nombre de grappes requis est de l'ordre de la trentaine, ce qui, en pratique, semble irréalisable) ou pour le dioxyde de soufre en situation industrielle (exemple de la station Mare Rouge à Rouen).

### ***L'échantillonnage dépend-il de l'indicateur recherché ?***

Oui, **l'estimation d'un taux de dépassement de seuil est plus contraignante que celle d'une moyenne annuelle** : pour être précise, elle requiert des grappes plus nombreuses et de plus longue durée.

#### 4.12.2 ESTIMATION

##### ***Quel est le meilleur estimateur ?***

Lorsqu'il existe un **site fixe bien corrélé au site d'échantillonnage** le redressement améliore la précision de l'estimation. Plus ou moins important pour le NO<sub>2</sub>, ce gain de précision est très net pour l'ozone et dans la plupart des cas, pour le monoxyde d'azote.

Le **redressement par la différence ou le quotient** doit être préféré au redressement par la régression, beaucoup plus exigeant en nombre de grappes, donc peu adapté à la pratique. Concernant la moyenne annuelle, les estimateurs redressés par la différence et le quotient fournissent des résultats comparables. Pour le dépassement de seuil (voir les tests sur le monoxyde d'azote), l'estimateur DIFF produit des intervalles de confiance plus larges mais plus fiables.

En revanche l'efficacité du redressement est peu visible pour les particules. Lorsque la corrélation entre site d'échantillonnage et site fixe est faible (0,4-0,5, voir les tests sur Strasbourg), il dégrade même les résultats.

De même, si la variabilité des concentrations au site auxiliaire est grande, il arrive que la correction apportée par le redressement soit trop importante. (En effet, sur la période d'échantillonnage, la moyenne des mesures au site auxiliaire est très supérieure ou inférieure à la moyenne annuelle réelle.) L'estimateur PI est alors préférable.

Pour des sites industriels représentatifs de situations localisées, le redressement n'est pas non plus conseillé.

##### ***Quelle est la fiabilité de l'intervalle de confiance ?***

L'intervalle de confiance à 95% qui encadre l'estimation est lui aussi **estimé. Il est fiable, autrement dit, son taux de confiance est effectivement proche de 95%, si son calcul repose sur une variance correctement estimée.** Quand le nombre de grappes par strate est faible, la variance peut être sur ou sous-estimée et l'intervalle de confiance qui en découle n'est pas exploitable.

Une possibilité est de **regrouper les strates au moment de l'estimation** et de faire comme si la somme des grappes avait été tirée dans une unique strate. Ce procédé accroît systématiquement le taux de couverture -évalué par simulations-, celui-ci pouvant atteindre ou dépasser 95% (voir notamment, l'estimation du NO<sub>2</sub> à Strasbourg). En contrepartie, il s'accompagne généralement d'une augmentation de l'incertitude estimée (1/2 largeur de l'IC). Cette augmentation reste faible pour les estimateurs redressés ; elle est plus ou moins importante pour l'estimateur PI : ponctuellement, on observe un facteur 2 (voir l'estimation du NO ou du NO<sub>2</sub> à Strasbourg).

**Dans tous les cas, seul un nombre suffisant de grappes (au moins 10/an si possible) garantit une estimation correcte de l'incertitude.**

### **Quelles sont les limites de la méthode ?**

La méthode des plans de sondage offre un cadre à la fois rigoureux et pratique à l'élaboration de plans d'échantillonnage et à l'estimation d'indicateurs annuels. Mais des précautions d'usage s'imposent. En effet, si les fonctions développées sont relativement rapides et aisées à mettre en œuvre, elles n'en reposent pas moins sur des hypothèses qu'il est impératif de contrôler avant toute estimation.

En particulier, la méthode des plans de sondage ne s'applique pas aux stratégies d'échantillonnage qui consistent à effectuer des mesures sur de longues périodes (ex : plusieurs semaines consécutives) mais à faible fréquence (ex : deux à quatre fois par an.)

Dans cette étude, les plans d'échantillonnage testés conduisent à des résultats tout à fait satisfaisants pour l'estimation de concentrations moyennes annuelles en situation urbaine de fond et de proximité, en particulier pour le NO<sub>2</sub>. D'après les résultats obtenus, une étude plus approfondie des conditions d'échantillonnage adaptées :

- à la proximité industrielle,
- à une estimation précise des taux de dépassement de seuil,

est nécessaire afin de déterminer la pertinence et l'efficacité de la méthode dans ces situations.

### **4.13 TRAVAUX FUTURS**

Ce rapport d'évaluation pourra être exploité par le nouveau groupe de travail « Plans d'échantillonnage et reconstitution de données ». La facilité de prise en main et d'application de la méthode par les AASQA et la pertinence de cette approche pour les besoins de la surveillance de la qualité de l'air seront éprouvées à cette occasion.

Les tests et développements envisageables à ce stade de l'étude sont notamment :

- l'application à d'autres polluants, comme les HAP ;
- l'adaptation des programmes à des grappes de taille variable ;
- la recherche de conditions d'échantillonnage plus propices à l'estimation de taux de dépassements de seuils ;
- la mise au point d'une interface, pour un usage plus rapide et aisé des programmes.

### **4.14 REFERENCES**

Directive du Conseil n°1999/30/CE du 22 avril 1999 relative à la fixation de valeurs limites pour l'anhydride sulfureux, le dioxyde d'azote et les oxydes d'azote, les particules et le plomb dans l'air ambiant.

Directive 2000/69/CE du Parlement européen et du Conseil du 16 novembre 2000 concernant les valeurs limites pour le benzène et le monoxyde de carbone dans l'air ambiant.

Directive 2002/3/CE du Parlement européen et du conseil du 12 février 2002 relative à l'ozone dans l'air ambiant.

European Environment Agency (EEA), 1998. Guidance report on preliminary assessment under EC air quality directives. Technical report.

MALHERBE L., 2001. Représentativité des mesures et méthodes statistiques, rapport LCSQA.

TILLÉ Yves. Théorie des sondages. Échantillonnage et estimation en populations finies. Cours et exercices. Dunod, Paris, 2001

LAVANCIER F., CAÏNI F., GAZEAU A., 2003. Plan de sondage pour mesures mobiles de la pollution atmosphérique. Pollution atmosphérique, N°180, octobre-décembre.

# **5. PARTIE III : COMPARAISON** **ENTRE LA METHODE ISO ET LA** **METHODE PLANS DE SONDAGE**

## 5.1 PROTOCOLE DE COMPARAISON

Afin de comparer les méthodes, des jeux de données communs, correspondant à plusieurs zones géographiques et à plusieurs polluants, sont constitués. Une présentation des sites de mesure est fournie en annexe 5.

Chaque test repose sur l'exploitation de deux années de données issues de deux stations de mesure : l'une de ces stations joue le rôle du moyen mobile, et l'autre, celle d'un site fixe auxiliaire ; les rôles sont ensuite intervertis.

Pour chaque polluant on définit l'incertitude maximale autorisée sur la moyenne et une valeur seuil de concentration. L'incertitude peut s'exprimer soit de façon absolue, dans les mêmes unités que la variable d'intérêt (demi-largeur de l'intervalle de confiance), soit en pourcentage, relativement à la moyenne. Nous avons choisi des incertitudes absolues ( $\mu\text{g}/\text{m}^3$ ). D'autre part, le nombre de dépassements étant nul ou très faible, les valeurs limites réglementaires, initialement choisies comme seuils, ont été abaissées.

La démarche adoptée de commun accord est la suivante :

- 1) Les données de la station auxiliaire de l'année (n) servent à optimiser les strates météo (ISO) et à définir le plan d'échantillonnage. Chacune des méthodes fournit le nombre total de mesures horaires (ISO) ou de grappes de mesures horaires (*Plans*) à prélever et débouche sur un plan d'échantillonnage qui lui est propre. Pour la méthode *Plans*, la durée de stationnement du moyen mobile à chacun de ses passages, i.e. la taille des grappes, est fixée à une semaine.
- 2) L'échantillonnage est simulé sur les données de la station « mobile » de l'année (n+1), conformément au plan défini.
- 3) On estime la moyenne et le taux de dépassement de seuil assortis de leurs intervalles de confiance. Les données de la station auxiliaire de l'année (n+1) sont utilisées, le cas échéant, pour corriger les estimations (*Plans*).

Remarque - L'estimation par la méthode ISO est faite :

- avec un échantillon de mesures conforme aux spécifications de la méthode (*dates ISO et durées ISO étendues à 672 heures, soit 4x1 semaines*) ;
- avec l'échantillon de grappes hebdomadaires de la méthode *Plans* (*dates et durées Plans*) ;
- avec l'échantillon de la méthode *Plans*, en ne conservant que les débuts de grappe (*durées 672h / nb de grappes*); l'échantillon ainsi obtenu a la même taille que le premier.

Le tableau ci-dessous indique, pour chaque jeu de données, les paramètres retenus dans l'élaboration des plans d'échantillonnage et l'estimation. Pour l'ozone, il est décidé de limiter l'échantillonnage à la saison estivale. Pour les polluants CO et NO, on utilise les plans d'échantillonnage et les échantillons associés au NO<sub>2</sub>.

Ville	Année de référence (n)	Année test (n+1)	Polluant	a) Station mobile b) Station fixe	a) Station fixe b) Station mobile	Incertitude Prise égale à ( $\mu\text{g}/\text{m}^3$ )	Valeur seuil considérée (seuil horaire, $\mu\text{g}/\text{m}^3$ )
Toulouse	2001	2002	NO	St Cyprien	Pargaminières	Même plan que NO <sub>2</sub>	100
			NO <sub>2</sub>	St Cyprien	Pargaminières	4	75 (StC.), 100 (Parg.)
			CO (trafic)	St Cyprien	Pargaminières	Même plan que NO <sub>2</sub>	3000 (StC) 4000 (Par.)
			O <sub>3</sub>	Colomiers	Balma	5	110
			PM10	Berthelot	CCIT	4	50
			PM2.5	Berthelot	CCIT	3	35
Le Havre	2002	2003	SO <sub>2</sub>	Mare	Herriot	5	200
			O <sub>3</sub>	Mare	Herriot	5	110
			PM10	Mare	Herriot	4	50
Dunkerque	2002	2003	SO <sub>2</sub>	Fort-Mardyck	St Pol	5	50 (FM) 75 (StP)
			NO	Fort-Mardyck	Petite Synthe	Même plan que NO <sub>2</sub>	100
			NO <sub>2</sub>	Fort-Mardyck	Petite Synthe	4	50
Strasbourg	2003	2004	NO	Nord	Centre	Même plan que NO <sub>2</sub>	100
			NO <sub>2</sub>	Nord	Centre	4	75
			O <sub>3</sub>	Nord	Centre	5	110
			PM10	Est	Clémenceau	4	50
			PM2.5	Est	Clémenceau	3	35
Annecy	2003	2004	NO	Loverchy	Novel	Même plan que NO <sub>2</sub>	*
			NO <sub>2</sub>	Loverchy	Novel	4	*
			O <sub>3</sub>	Loverchy	Novel	5	*
			PM10	Loverchy	Novel	4	*
Annemasse	2003	2004	NO	Gaillard	Centre	Même plan que NO <sub>2</sub>	*
			NO <sub>2</sub>	Gaillard	Centre	4	*
			O <sub>3</sub>	Gaillard	Centre	5	*
			PM10	Gaillard	Centre	4	*
Chambéry	2003	2004	NO	Pasteur	Le Haut	Même plan que NO <sub>2</sub>	*
			NO <sub>2</sub>	Pasteur	Le Haut	4	*
			O <sub>3</sub>	Pasteur	Le Haut	5	*
			PM10	Pasteur	Le Haut	4	*

h : seuil horaire, j : seuil journalier



## 5.2 ELEMENTS GENERAUX DE COMPARAISON

### 5.3 THEORIE

La méthode ISO et, pour être plus efficace, la méthode des plans de sondage reposent toutes deux sur une stratification de l'année. Pour une incertitude fixée, elles indiquent, grâce à l'exploitation d'un jeu de données test, le nombre de mesures à effectuer dans chaque strate. Ce nombre dépend de la variance à l'intérieur des strates.

En revanche, le principe de la stratification est différent.

Dans la méthode ISO, celle-ci est fonction de paramètres météorologiques et temporels court terme (jours et heures). Les campagnes de mesure sont réparties librement sur l'année par l'utilisateur. La part d'aléatoire réside dans le fait qu'on ne connaît pas à l'avance les conditions météorologiques précises dans lesquelles auront lieu ces campagnes. En conséquence, le nombre de mesures effectivement collecté dans chaque strate peut s'écarter du nombre prévu.

Dans la méthode des plans de sondage, les strates sont des périodes de temps déterminées par une date de début et une date de fin (par simplification, on s'est limité à des strates de même taille, telles que des trimestres ou des saisons). En l'absence de problèmes techniques, les nombres réel et prévu de mesures par strate sont donc les mêmes. L'aspect aléatoire réside dans le tirage au hasard des dates de début de grappe.

## 5.4 MISE EN ŒUVRE DES CALCULS

### 5.4.1 TRAVAIL PREPARATOIRE

L'application de la méthode ISO requiert une étude préliminaire propre à la zone géographique concernée. L'objet de cette pré-étude est

- d'évaluer les paramètres météorologiques et temporels qui agissent le plus sur les concentrations ;
- pour chaque paramètre, de définir une ou plusieurs valeurs seuils permettant de stratifier les mesures.

Cette analyse peut être facilitée par une bonne connaissance du terrain. Sauf modifications majeures dans les conditions météorologiques et/ou dans les sources de pollution, il n'est pas besoin de la reconduire avant chaque campagne.

L'application de la méthode Plans n'exige aucun travail préalable. Dans les tests effectués, nous nous sommes limités à une stratification été/hiver. Cependant, un examen poussé des données de stations fixes pourrait être utile pour proposer un autre découpage.

### 5.4.2 DONNEES D'ENTREE

La méthode ISO nécessite des données météorologiques. Mais les calculs réalisés montrent que **des paramètres classiques, tels que la vitesse du vent et la température, suffisent**. Ces paramètres sont mesurés en routine près des stations fixes, et aussi pendant les campagnes.

La méthode Plans est peu contraignante ; elle peut se passer de variable auxiliaire. L'utilisation de données d'une station fixe est cependant recommandée dans l'étape d'estimation ; elle est obligatoire si l'on effectue un redressement.

	<b>Méthode ISO</b>	<b>Méthode Plans</b>
<b>Définition d'un plan d'échantillonnage</b>  (données d'années antérieures : n, n-1...)	O : série annuelle ou pluriannuelle d'une station fixe auxiliaire supposée de mêmes caractéristiques que le site d'échantillonnage.  O : séries correspondantes des principaux paramètres météorologiques ou temporels influents	O : série annuelle ou pluriannuelle d'une station fixe auxiliaire supposée de mêmes caractéristiques que le site d'échantillonnage.
<b>Estimation</b>  (année n+1)	Données d'échantillonnage  O : séries correspondantes des principaux paramètres météorologiques ou temporels influents	Données d'échantillonnage  F (O si redressement) : série annuelle d'une station fixe auxiliaire supposée de mêmes caractéristiques que le site d'échantillonnage.  F : série de toute autre variable auxiliaire qui pourrait être utilisée dans le redressement

O : obligatoire pour l'étape considérée ; F : facultatif pour l'étape considérée

### 5.4.3 OUTILS DE CALCUL

Des outils de calcul ont été développés pour chaque méthode. Leur utilisation n'exige pas de compétences particulières, y compris l'exécution des programmes R si l'on s'en tient aux fonctions proposées. L'évaluation de ces outils sera poursuivie en 2006, notamment au sein du GT *Plans d'échantillonnage et reconstitution de données*.

	<b>Méthode ISO</b>	<b>Méthode Plans</b>
Outil de calcul	Excel  (Feuille de calcul préprogrammée)	R  (logiciel de statistique librement téléchargeable sur Internet à l'adresse : <a href="http://cict-cran.fr">http://cict-cran.fr</a> )
Facilité d'utilisation	Calculs aisément exécutables	Calculs facilement exécutables avec un mode d'emploi mais une prise en main du logiciel semble nécessaire

### 5.5 TAILLE DES ECHANTILLONS

Pour un objectif de précision donné, les deux méthodes ne proposent pas le même nombre de mesures. Ce nombre est plus faible pour la méthode ISO.

AASQA	Polluant	Année d'échantillonnage	Station de référence utilisée pour définir l'échantillonnage	Nombre annuel de jours de mesures demandé par la méthode ISO	Nombre annuel de jours de mesures demandé par la méthode plans pour des grappes d'une semaine
ORAMIP	NO2	2002	Pargaminières	19	11x7=77
	NO			86	27x7=189
	NO2	2002	Saint-Cyprien	12	8x7=56
	NO			87	28x7=196
	O3, été	2002	Balma	12	7x7=49
	O3, été	2002	Colomiers	12	9x7=63
	PM10	2002	CCIT	6	9x7=63
	PM10	2002	Berthelot	6	9x7=63
	PM2.5	2002	CCIT	5	9x7=63
	PM2.5	2002	Berthelot	5	9x7=63
ASPA	NO2	2004	Stbg-Centre	16	14x7=98
	NO			37	17x7=119
	NO2	2004	Stbg-Nord	13	10x7=70
	NO			30	15x7=105
	O3, été	2004	Stbg-Centre	24 (année)	16x7=112
	O3, été	2004	Stbg-Nord	23 (année)	14x7=98
	PM10	2004	Stbg-Est	18	15x7=115
	PM10	2004	Stbg-Clémenceau	13	13x7=91
	PM2.5	2004	Stbg-Est	12	11x7=77
	PM2.5	2004	Stbg-Clémenceau	10	14x7=98
OPALAIR	NO2	2003	Petite-Synthe	11	8x7=56
	NO			32	15x7=105
	NO2	2003	Fort-Mardick	10	8x7=56
	NO			13	7x7=49
	SO2	2003	Saint-Pol Nord	9	5x7=35
	SO2	2003	Fort-Mardick	12	7x7=49
AIR NORMAND	SO2	2003	Herriot	31	12x7=84
	SO2			54	21x7=147
	O3	2003	Herriot	11	5x7=35
	O3	2003	Mare-Rouge	13	6x7=42
	PM10	2003	Herriot	10	7x7=49
	PM10	2003	Mare-Rouge	7	7x7=49
AIR APS Anney	NO2	2004	Novel	11	8x7=56
	NO			27	16x7=112

	NO2 NO	2004	Loverchy	21 66	14x7=98 20x7=140
	O3	2004	Novel	23	14x7=98
	O3	2004	Loverchy	21	13x7=91
	PM10	2004	Novel	9	10x7=70
	PM10	2004	Loverchy	11	11x7=77
AIR APS Annemasse	NO2 NO	2004	Centre	17 29	12x7=84 11x7=77
	NO2 NO	2004	Gaillard	15 37	11x7=77 14x7=98
	O3	2004	Centre	22	16x7=112
	O3	2004	Gaillard	24	14x7=98
	PM10	2004	Centre	11	11x7=77
	PM10	2004	Gaillard	9	11x7=77
AIR APS Chambéry	NO2 NO	2004	Le Haut	14 16	12x7=84 9x7=63
	NO2 NO	2004	Pasteur	14 28	11x7=77 13x7=91
	O3	2004	Le Haut	24	16x7=112
	O3	2004	Pasteur	23	16x7=112
	PM10	2004	Le Haut	8	8x7=56
	PM10	2004	Pasteur	8	9x7=63

Pour la méthode des plans de sondage : on s'est limité à présenter les effectifs relatifs à des grappes d'une semaine. Avec une autre taille de grappe, ces effectifs seraient différents, aussi bien en nombre de grappes qu'en nombre cumulé de jours (La relation entre nombre et taille de grappes n'est pas inversement proportionnelle).

Il en est de même pour la méthode ISO : si on travaillait sur des données tri-horaires au lieu des données horaires, les calculs préconiseraient un nombre total de séquences de mesures divisé par un nombre plus faible que 3, entraînant ainsi une durée totale de campagne plus longue.

En fait, pour les deux méthodes, si on modifie la durée des données de base, les variances ne sont pas divisées par le même rapport que ces durées.

### Attention :

**Méthode Plan** : le calcul du nombre de mesures à prélever est une étape optionnelle, effectuée à titre indicatif. Les méthodes d'estimation s'appliquent encore, même si les échantillons ont été prélevés selon d'autres plans. On peut par exemple prendre un nombre supérieur de grappes afin d'améliorer l'estimation de la variance (voir les recommandations sur la méthode des plans de sondage). Il faut cependant respecter l'hypothèse d'un échantillonnage aléatoire (tirage aléatoire des grappes de mesures dans l'année) ou aléatoire stratifié (tirage aléatoire des grappes de mesure dans chaque strate).

**Méthode ISO** : étant donné le choix de la durée minimum de chaque campagne fixée à une semaine, et le besoin de répartition sur les 4 saisons, la méthode ISO a été appliquée avec un minimum de 28 jours par an, soit plus que ne le requiert la méthode (sauf pour NO).

## 5.6 RESULTATS DES ESTIMATIONS

Ce chapitre présente les résultats des calculs effectués selon le protocole de comparaison. Dans la suite, le terme *méthode* désigne la combinaison d'un plan d'échantillonnage (dates ISO, dates plans, dates plans débuts\_de\_grappes) et d'un estimateur (ISO, PI, DIFF, QUOT.)

Pour chaque indicateur, d'abord la moyenne annuelle, et ensuite le taux de dépassement de seuil, pour chaque ville concernée par cette étude et pour chaque polluant pris en compte, deux types de graphique sont réalisés :

- Le premier type de graphique permet de comparer entre elles et avec la moyenne réelle annuelle, les valeurs estimées par les différentes méthodes ; on trouvera dans l'ordre :
  - la moyenne annuelle réelle, puis :
    1. méthode ISO, campagnes pendant chaque première semaine des mois 02, 05, 08 et 11,
    2. méthode ISO, campagnes aux dates et durées de la méthode PLAN,
    3. méthode ISO, même durée totale que 1, mais répartie sur les débuts de chaque grappe,
    4. méthode PLAN-PI,
    5. méthode PLAN-DIFF ;
    6. méthode PLAN-QUOT<sup>2</sup>.
  
- Les 5 (ou 6) graphiques qui suivent indiquent dans l'ordre des méthodes 1 à 5 (ou 6) ci-dessus, l'écart relatif entre l'estimation, encadrée de son intervalle de confiance, et la valeur réelle. Ils permettent d'apprécier l'exactitude et la précision de chaque méthode.

On effectue les comparaisons entre les 5 méthodes et la moyenne annuelle réelle avec le codage : « - , - , 0 , + , ++ » (moyenne annuelle sous-estimée, légèrement sous-estimée, bien approchée, légèrement surestimée, surestimée).

Les méthodes comparées deux à deux ci-après seront jugées équivalentes pour l'estimation de la moyenne si elles sont affectées ensemble du code « 0 » :

- Entre les méthodes ISO 1 et 3, mêmes durées mais dates différentes,

---

<sup>2</sup> Les résultats de cette méthode ne sont présentés que pour l'estimation des dépassements. Pour l'estimation de la moyenne, on a montré l'efficacité généralement plus grande du redressement par la différence (cf. partie II).

- Entre les méthodes 2 et 4, puis 2 et 5, mêmes durées et dates,
- Entre méthodes ISO 2 et 3, mêmes débuts mais durées différentes,
- Entre méthodes PLAN 4 et 5.

## 5.7 RECONSTITUTION DE MOYENNES

**N.B. :** Pour l’ozone, l’estimateur PI (méthode Plans) ne peut estimer que la moyenne estivale, puisque les grappes ont été prélevées en été, conformément au protocole de comparaison. Les résultats sont présentés sur les graphiques à titre d’information mais ils ne doivent pas être considérés comme une estimation de la moyenne annuelle réelle. En revanche, le redressement a été effectué sur l’année complète.

### 5.7.1 TOULOUSE

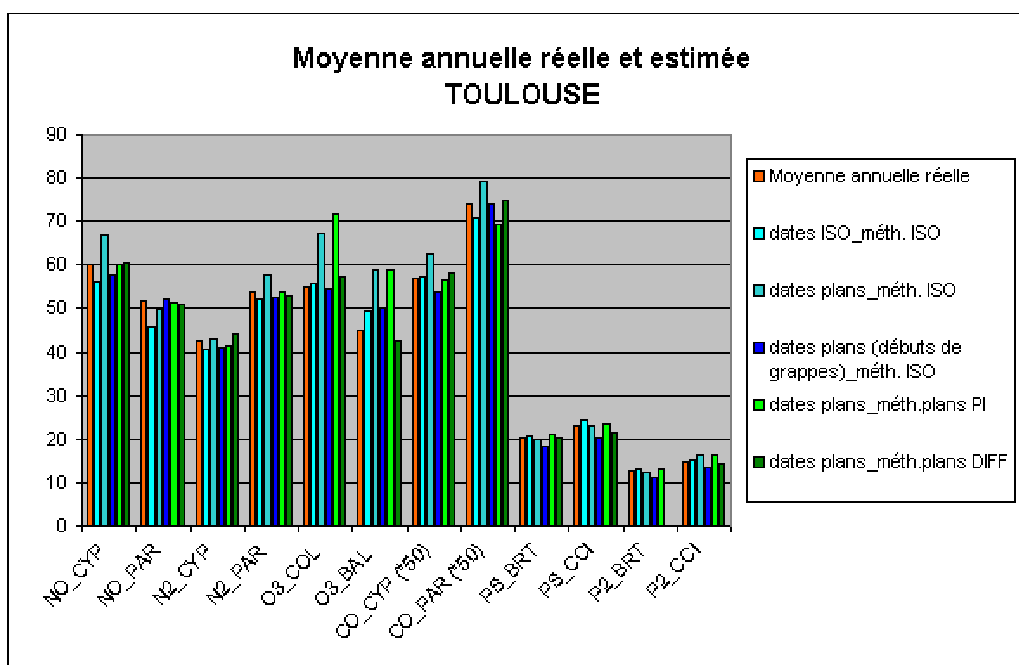
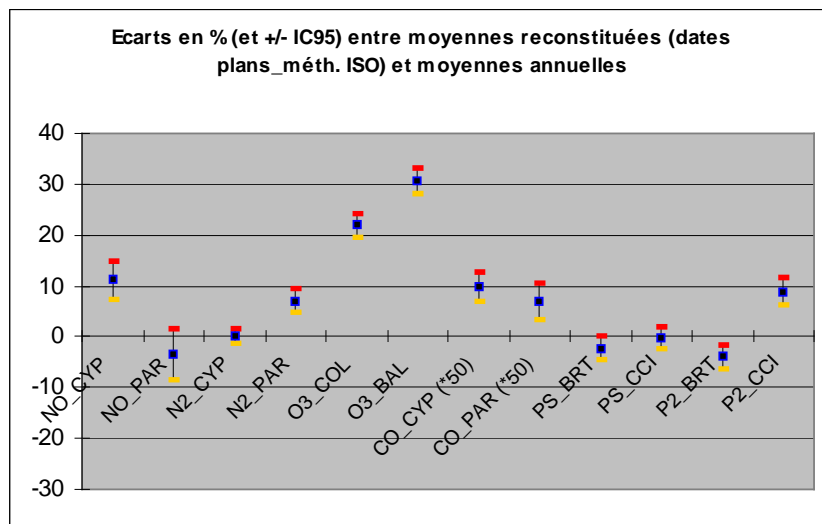
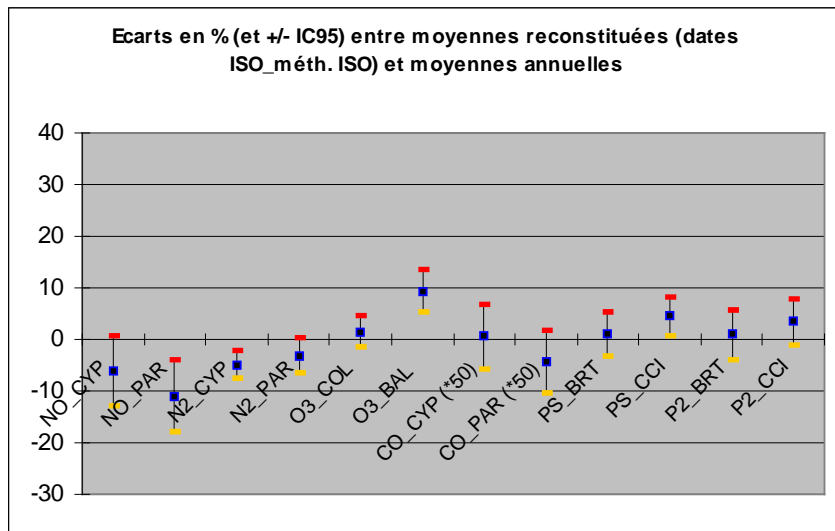
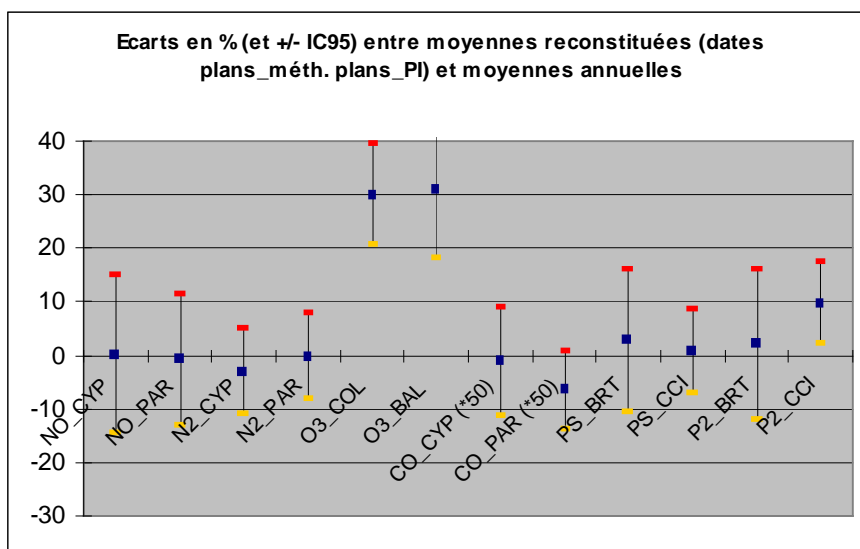
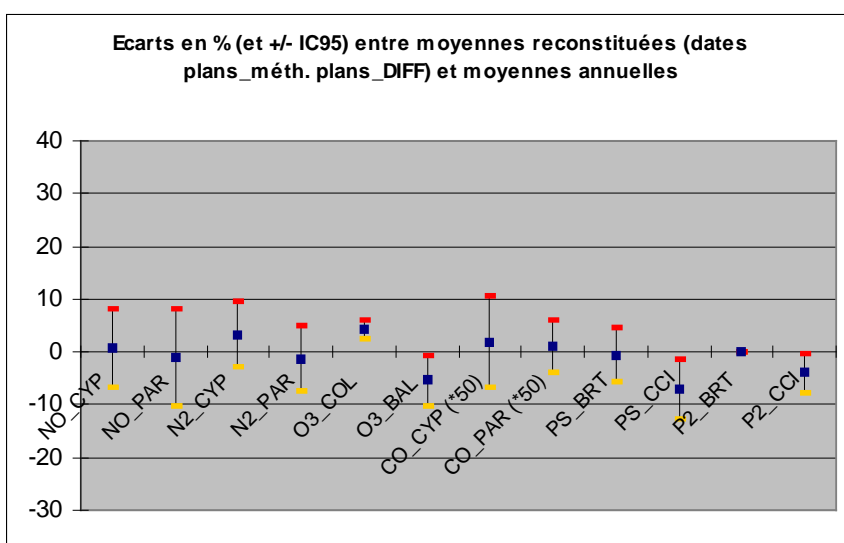
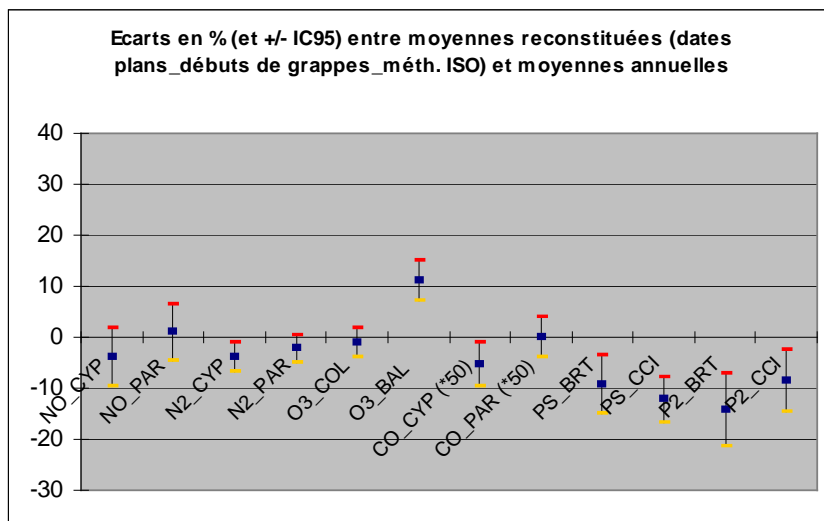


Figure 9 : les moyennes reconstituées à Toulouse (en  $\mu\text{g.m}^{-3}$  sauf CO en  $\text{mg.m}^{-3} * 50$ )

Polluant	NO		NO2		O3		CO		PM <sub>10</sub>		PM <sub>2.5</sub>	
	CYP	PAR	CYP	PAR	COL	BAL	CYP	PAR	BRT	CCI	BRT	CCI
Entre 0 et 1	-	--	-	-	0	+	0	-	0	0	0	0
Entre 0 et 2	++	0	0	++	++	++	+	+	0	0	0	+
Entre 0 et 3	-	0	-	0	0	+	-	0	-	-	-	-
Entre 0 et 4	0	0	-	0	++	++	0	-	0	0	0	+
Entre 0 et 5	0	0	0	0	+	-	0	0	0	-		0

Tableau 11: comparaison des méthodes à Toulouse







## 5.7.2 STRASBOURG

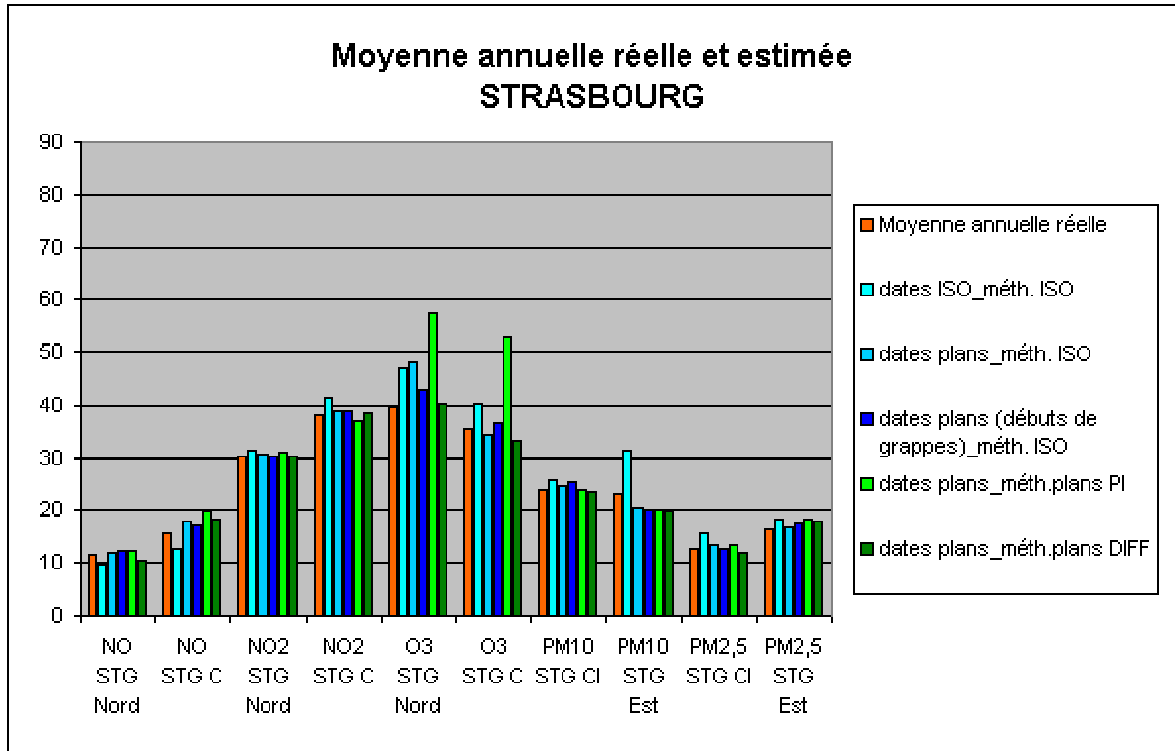
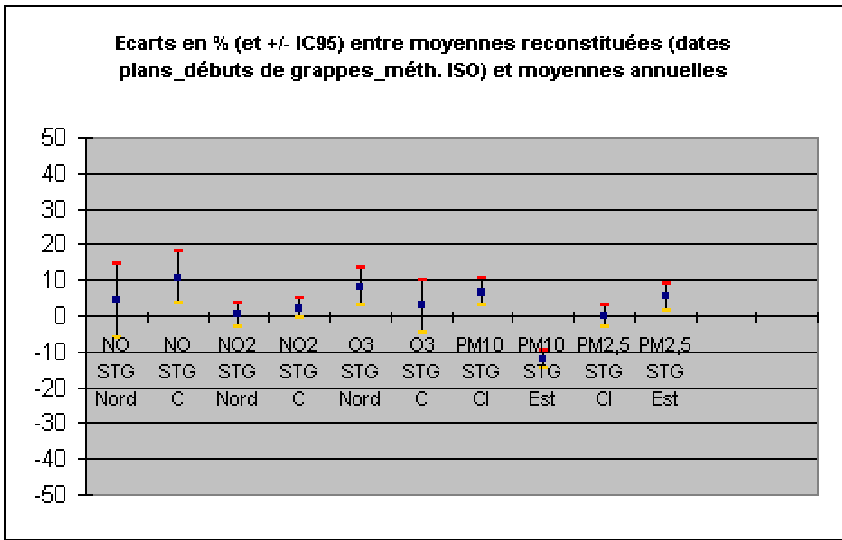
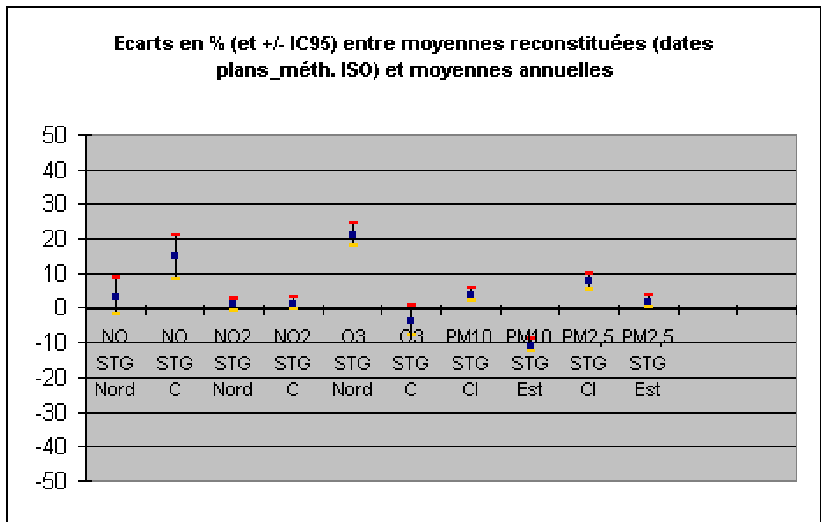
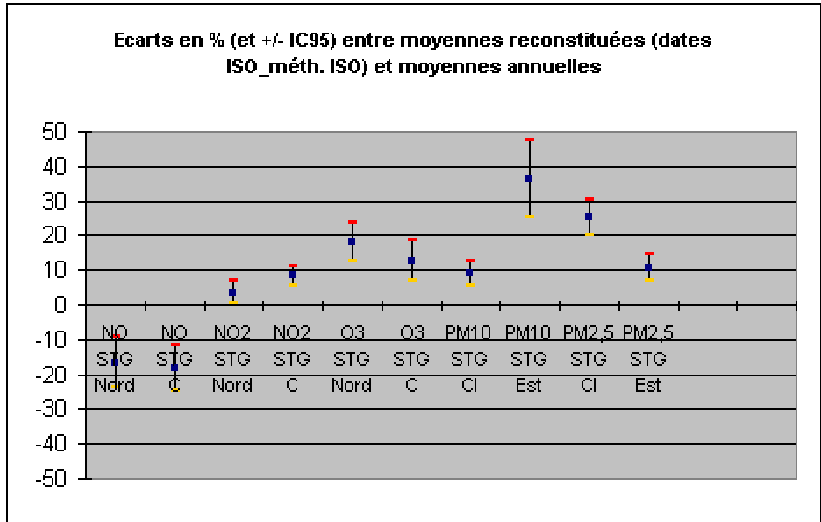
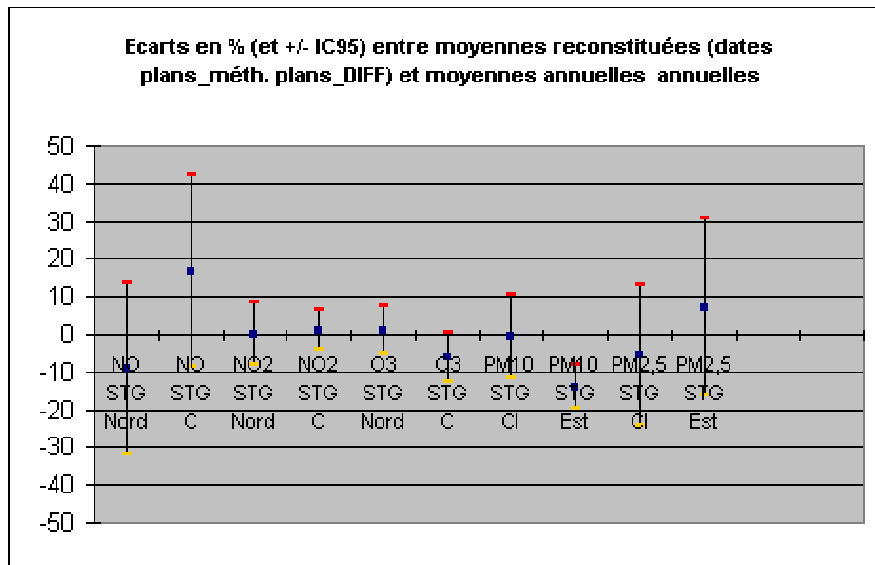
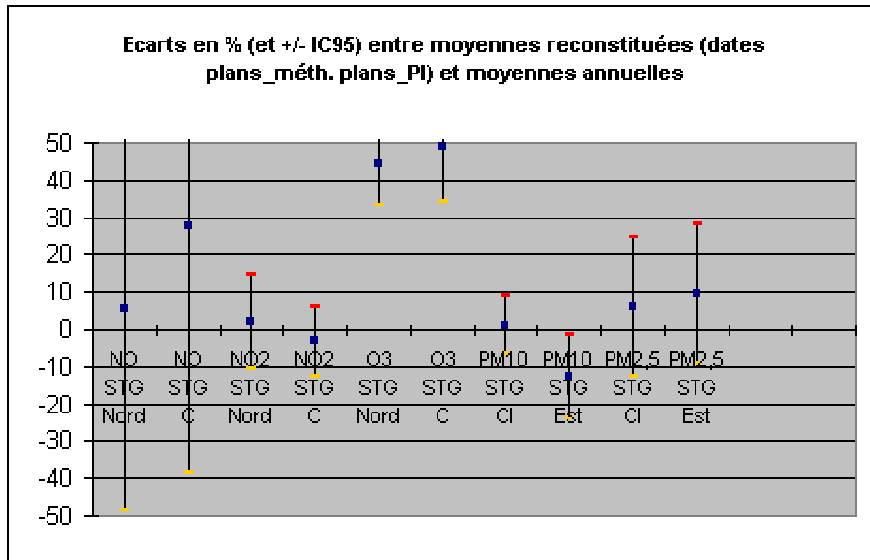


Figure 10: les moyennes reconstituées à Strasbourg (en  $\mu\text{g.m}^{-3}$ )

Polluant	NO		NO2		O3		PM <sub>10</sub>		PM <sub>2.5</sub>	
	STG-N	STG-C	STG-N	STG-C	STG-N	STG-C	STG-CI	STG-E	STG-CI	STG-E
Entre 0 et 1	-	-	+	+	++	+	0	++	+	0
Entre 0 et 2	0	+	0	0	++	0	0	-	0	0
Entre 0 et 3	0	+	0	0	+	0	0	-	0	0
Entre 0 et 4	0	+	0	-	++	++	0	-	0	0
Entre 0 et 5	-	+	0	0	0	-	0	-	-	0

Tableau 12: comparaison des méthodes à Strasbourg





### 5.7.3 DUNKERQUE

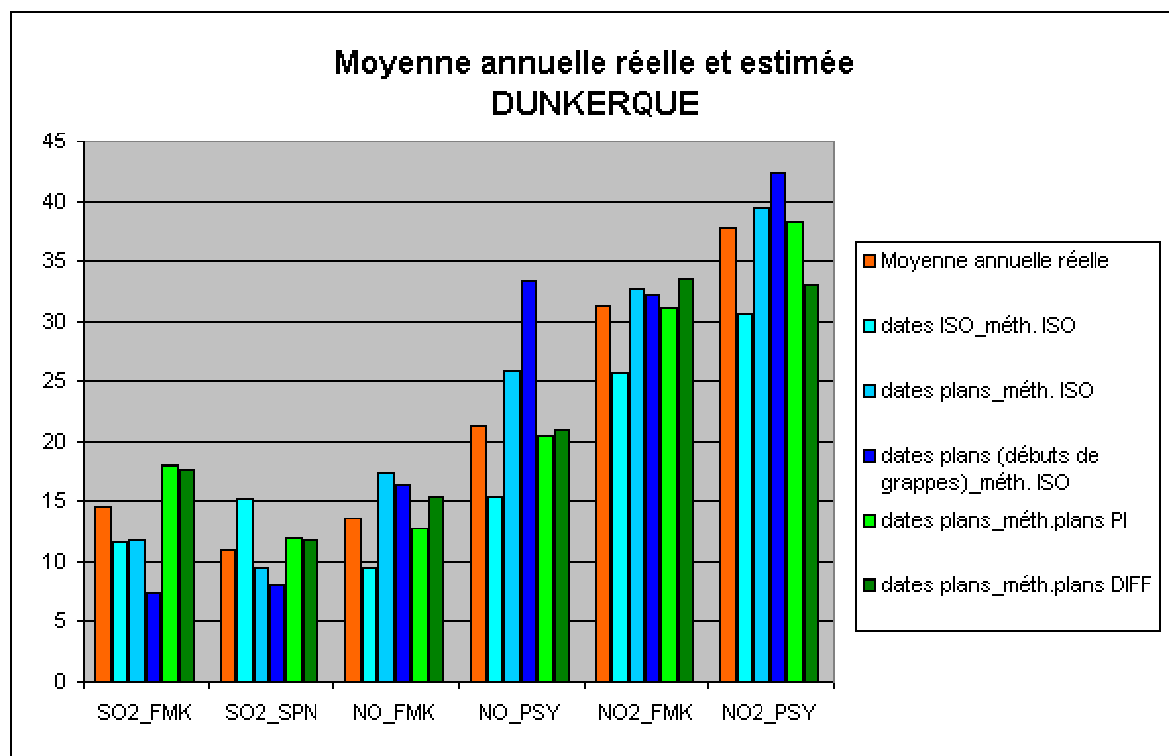
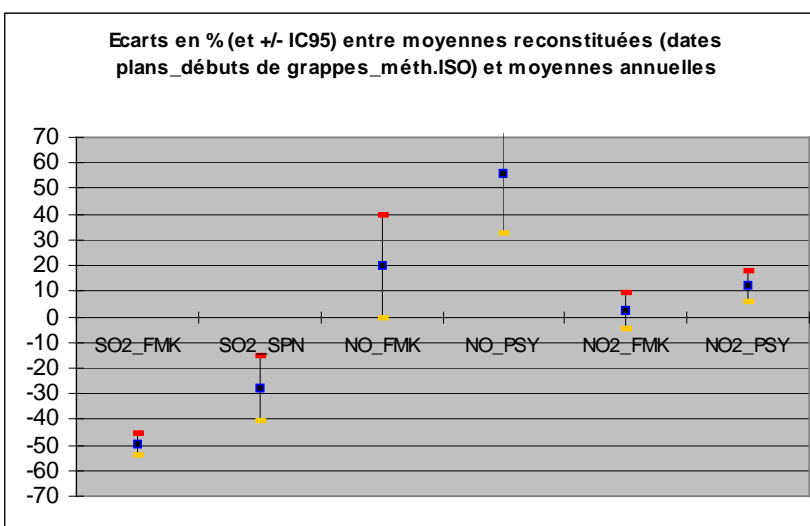
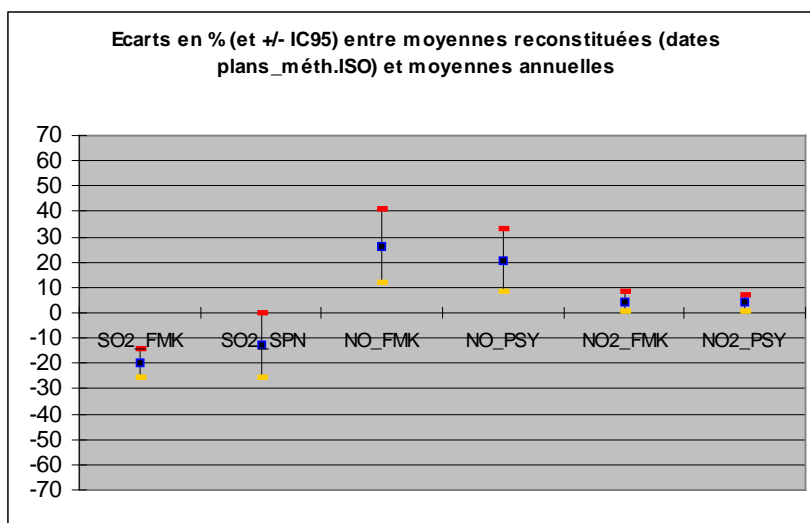
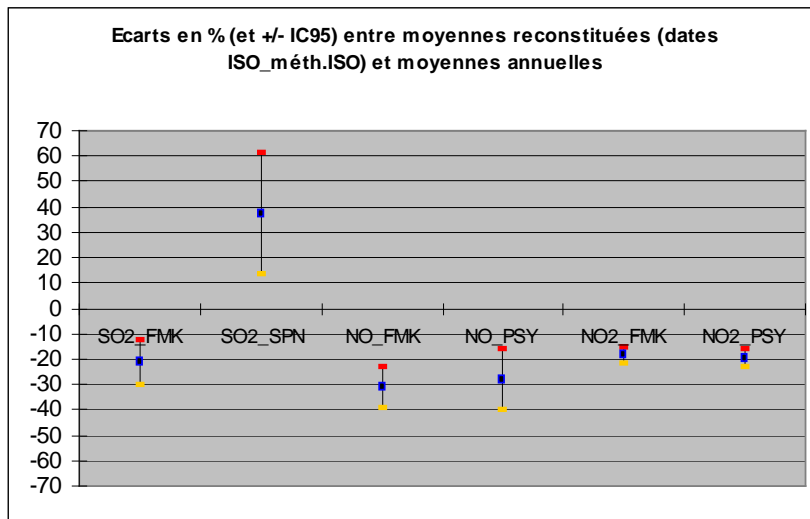
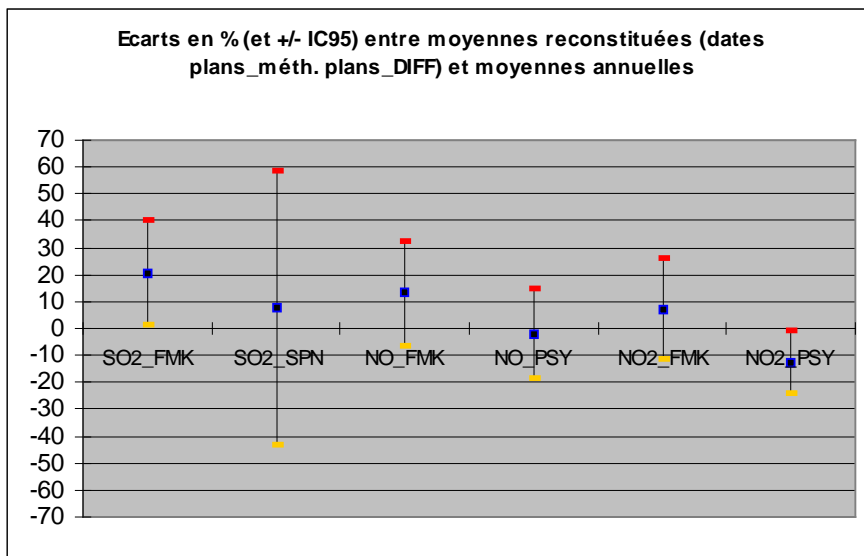
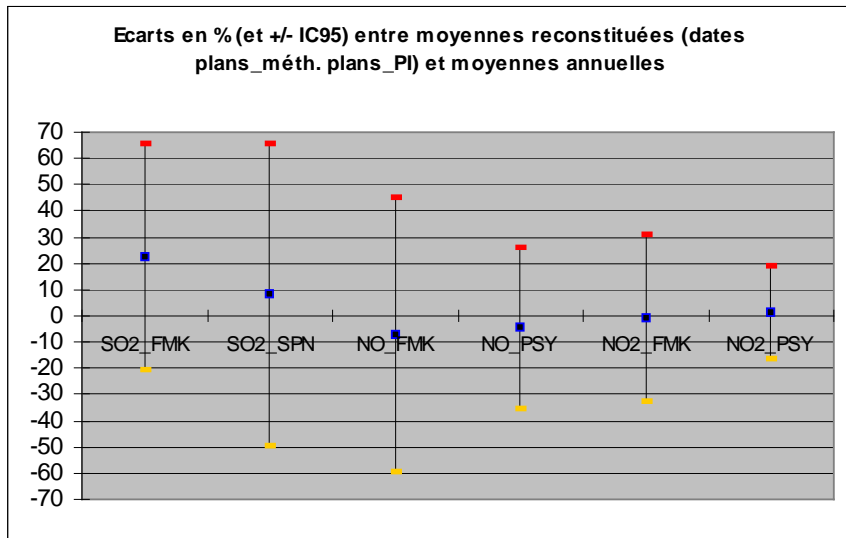


Figure 11: les moyennes reconstituées à Dunkerque (en  $\mu\text{g.m}^{-3}$ )

Polluant	SO <sub>2</sub>		NO		NO <sub>2</sub>	
	FMK	SPN	FMK	PSY	FMK	PSY
Entre 0 et 1	-	++	--	--	--	--
Entre 0 et 2	-	-	++	+	0	0
Entre 0 et 3	--	--	++	++	0	+
Entre 0 et 4	++	+	-	0	0	0
Entre 0 et 5	++	+	+	0	+	--

Tableau 13: comparaison des méthodes à Dunkerque





## 5.7.4 LE HAVRE

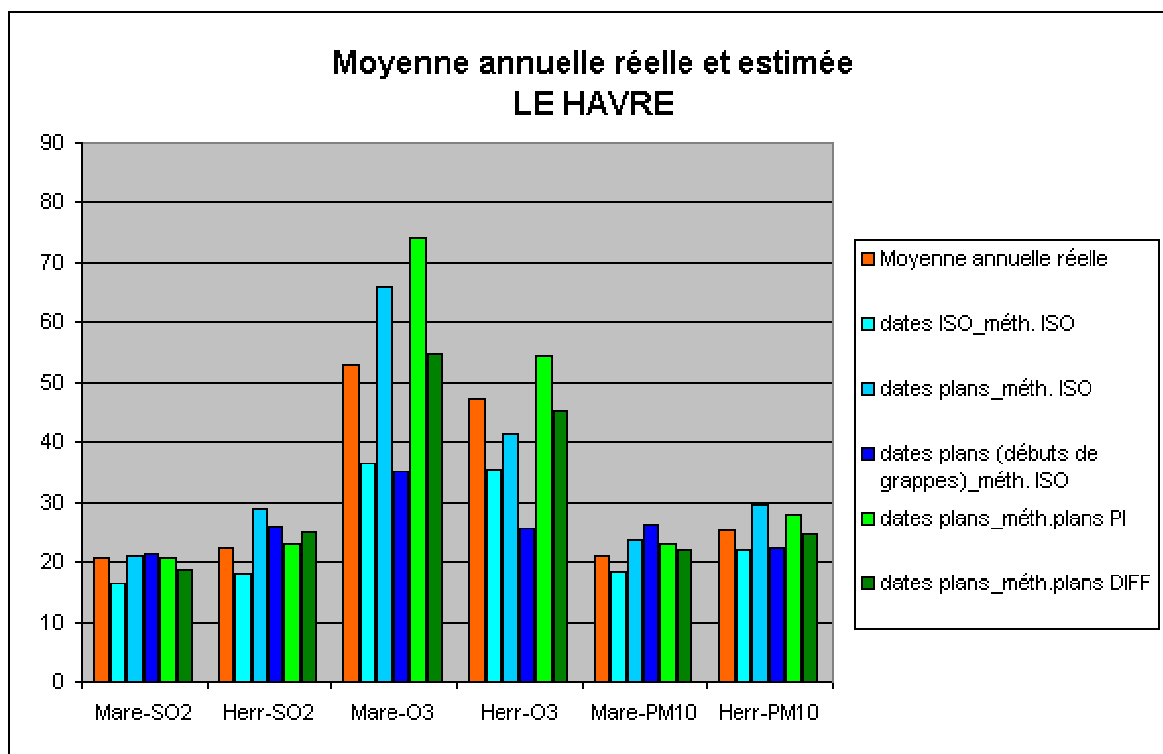
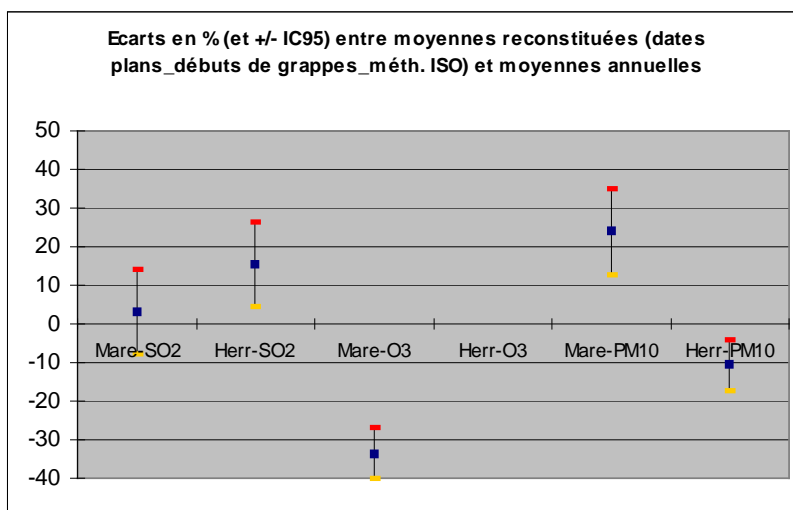
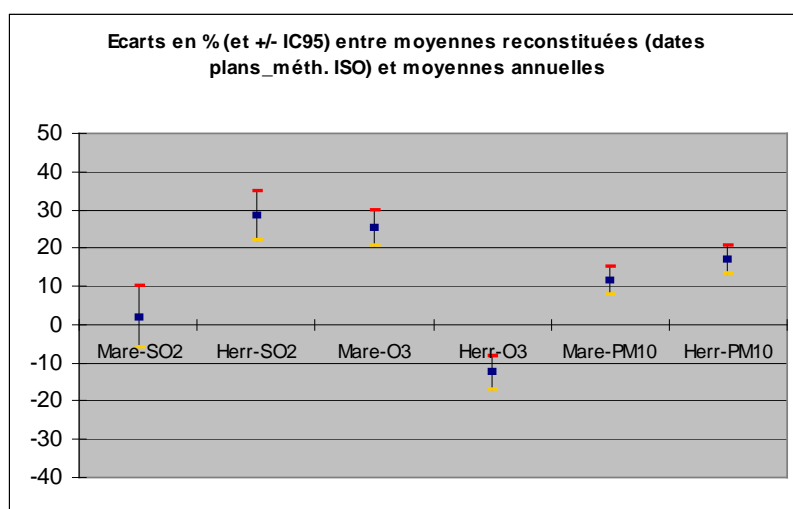
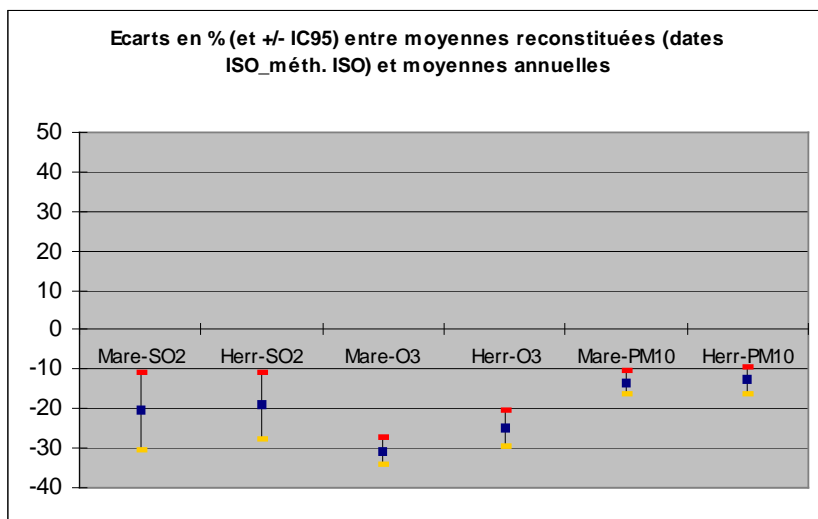


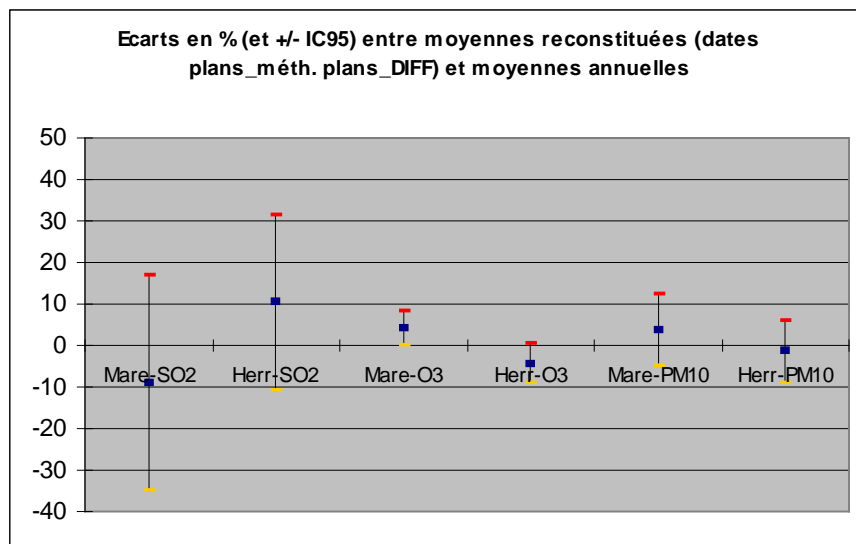
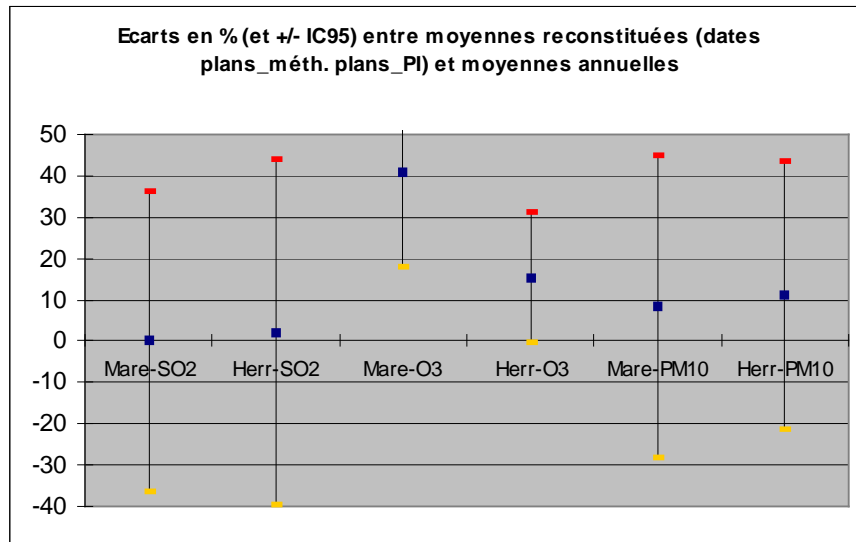
Figure 12: les moyennes reconstituées à Le Havre (en  $\mu\text{g.m}^{-3}$ )

Polluant	SO2		O3		PM10	
	Mare	Herr	Mare	Herr	Mare	Herr
Entre 0 et 1	-	-	--	--	-	-
Entre 0 et 2	0	++	++	-	+	+
Entre 0 et 3	0	+	--	--	++	-
Entre 0 et 4	0	0	++	+	+	+
Entre 0 et 5	-	+	0	0	0	0

Tableau 14: comparaison des méthodes à Le Havre







## 5.7.5 ANNECY

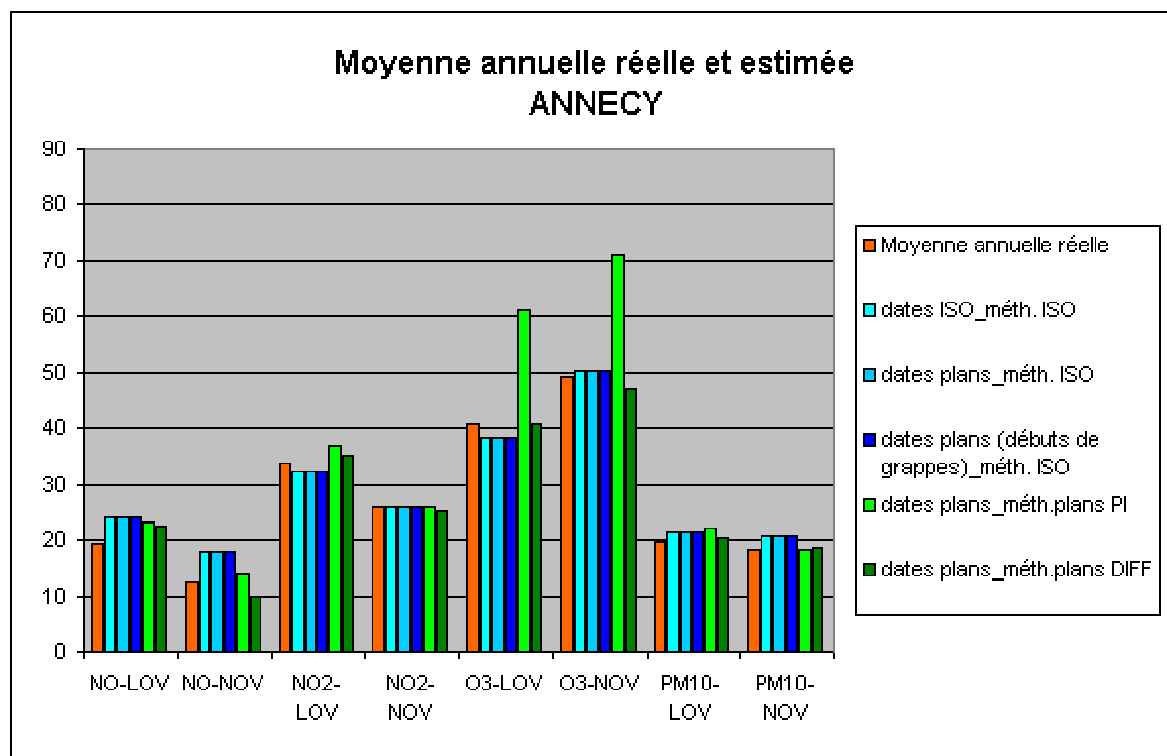
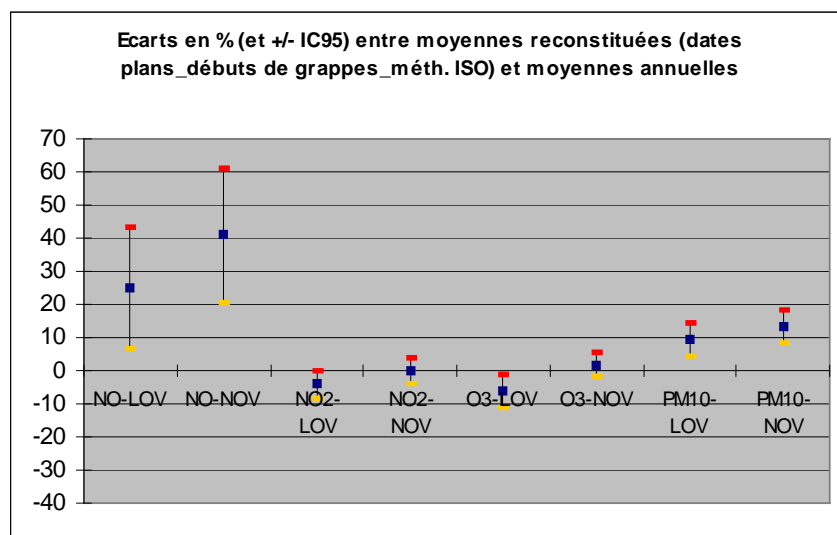
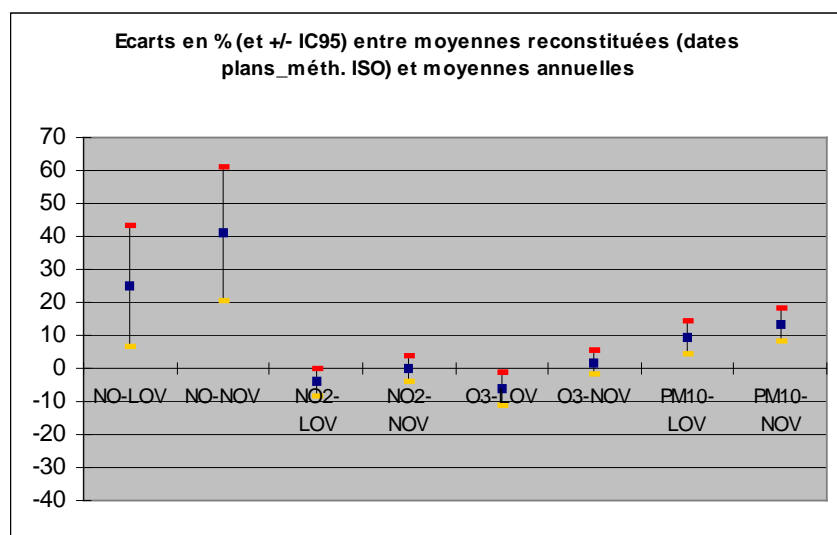
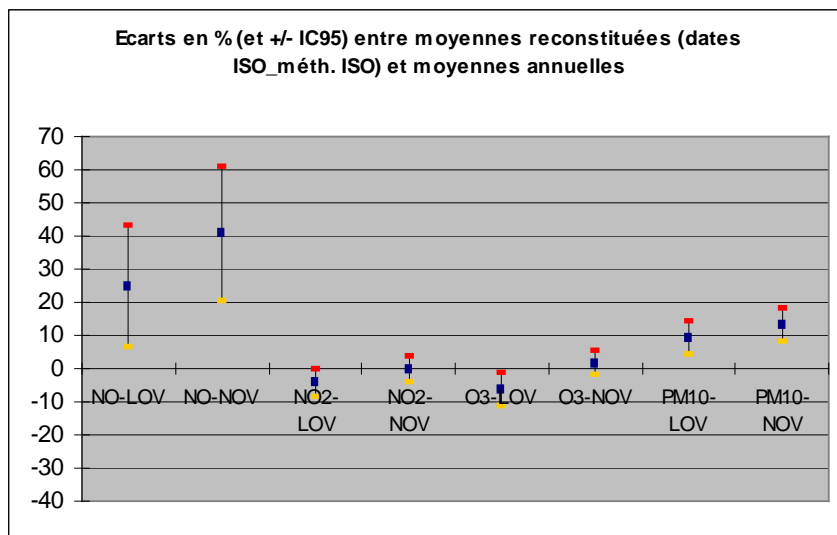
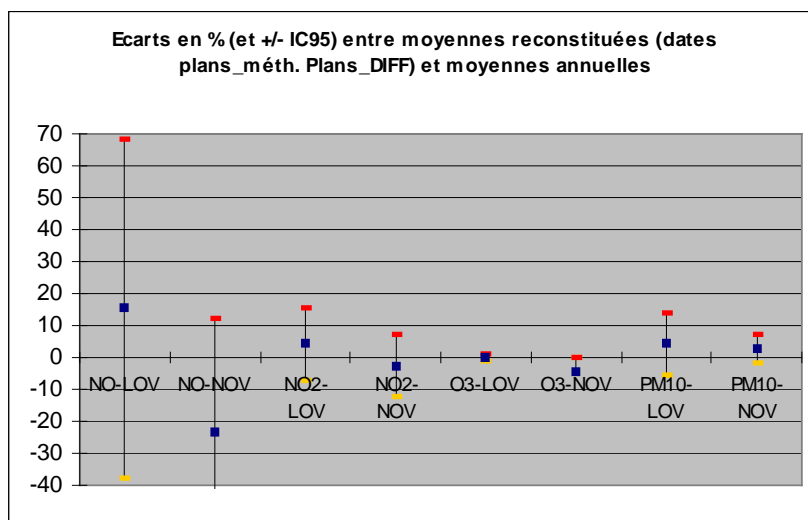
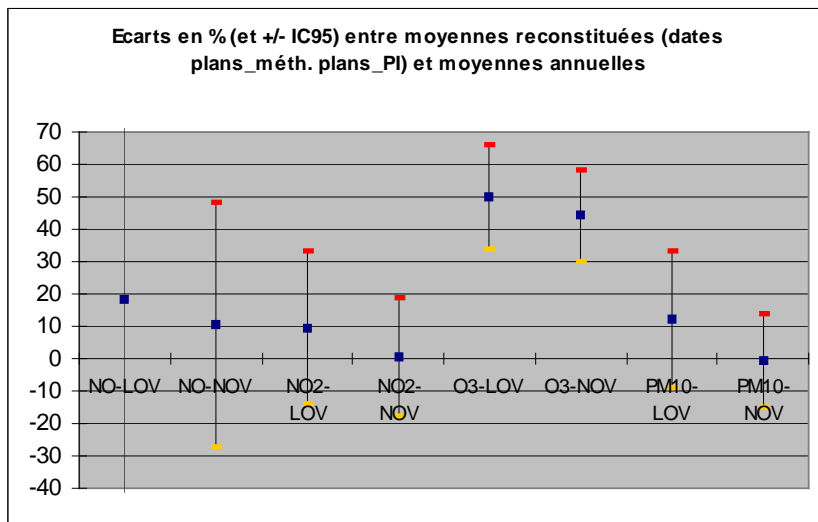


Figure 13: les moyennes reconstituées à Annecy (en  $\mu\text{g.m}^{-3}$ )

Polluant	NO		NO2		O3		PM <sub>10</sub>	
	LOV	NOV	LOV	NOV	LOV	NOV	LOV	NOV
Entre 0 et 1	+	+	0	0	0	0	+	+
Entre 0 et 2	+	+	0	0	0	0	+	+
Entre 0 et 3	+	+	0	0	0	0	+	+
Entre 0 et 4	+	0	+	0	++	++	+	-
Entre 0 et 5	+	-	-	0	00	-	0	-

Tableau 15: comparaison des méthodes à Annecy





## 5.7.6 ANNEMASSE

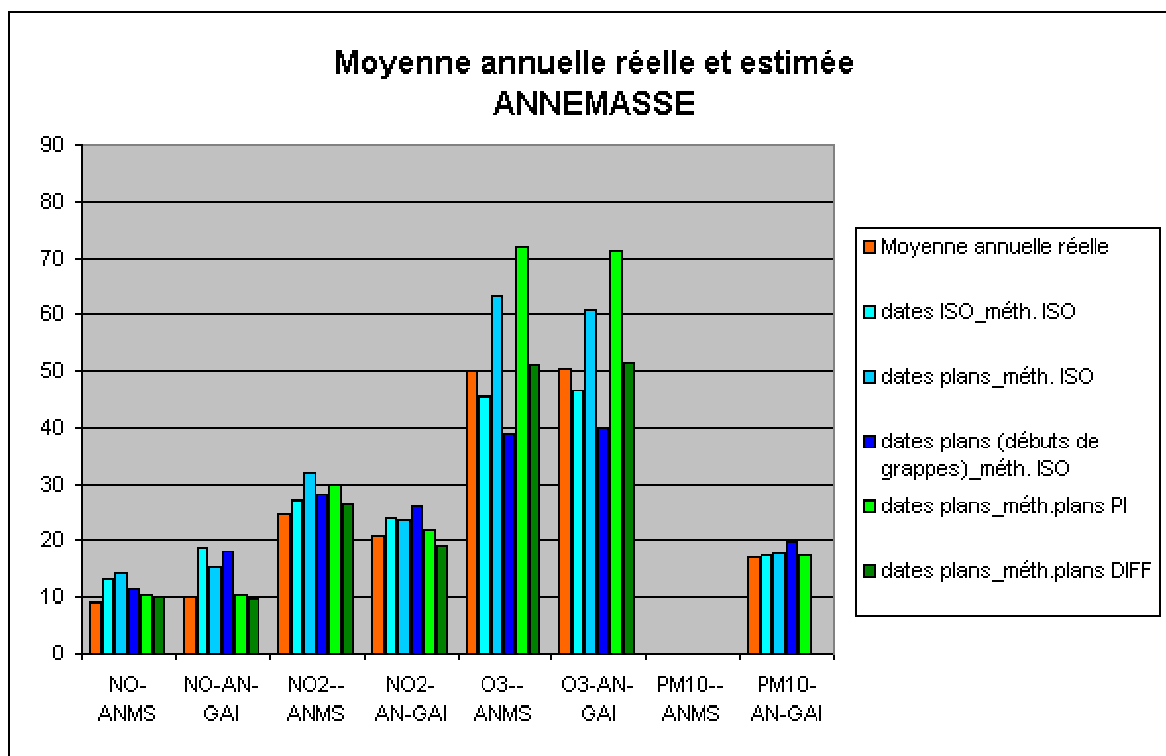
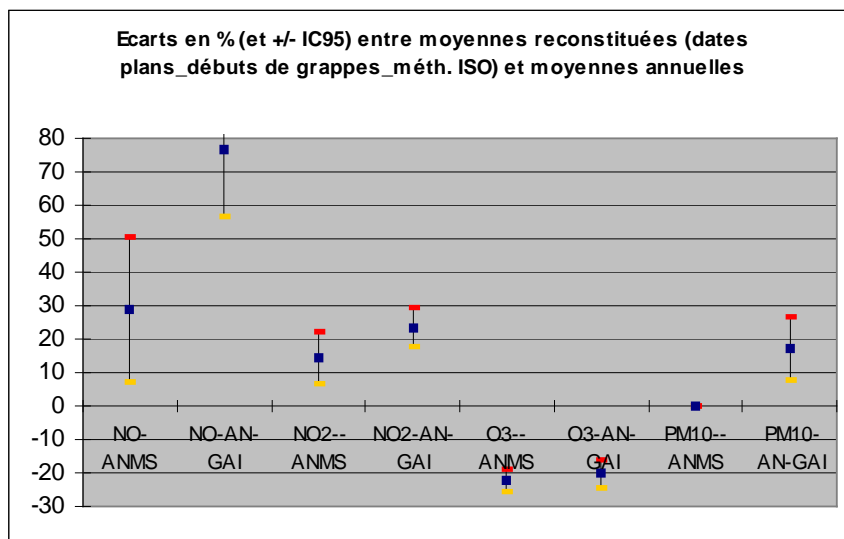
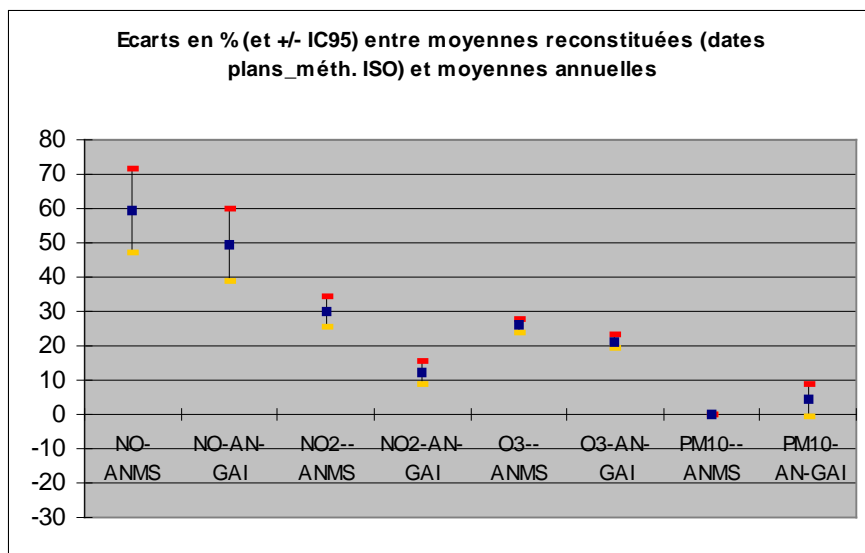
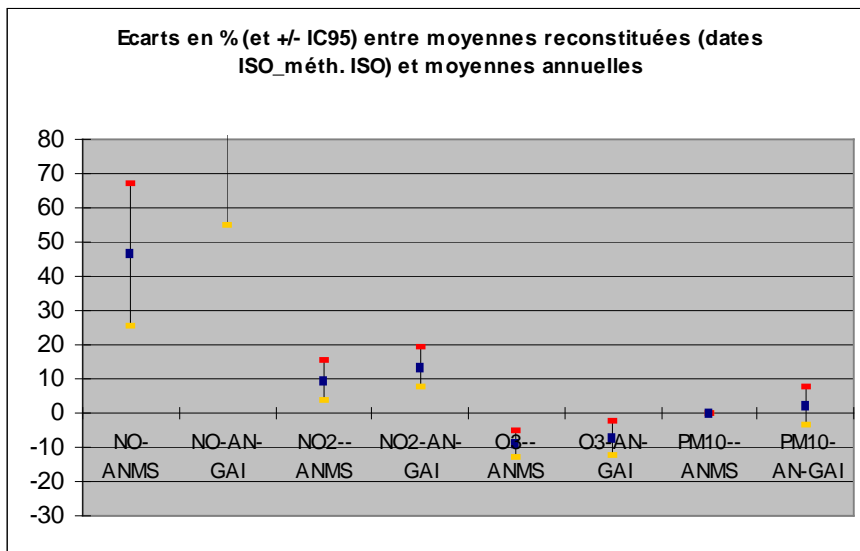
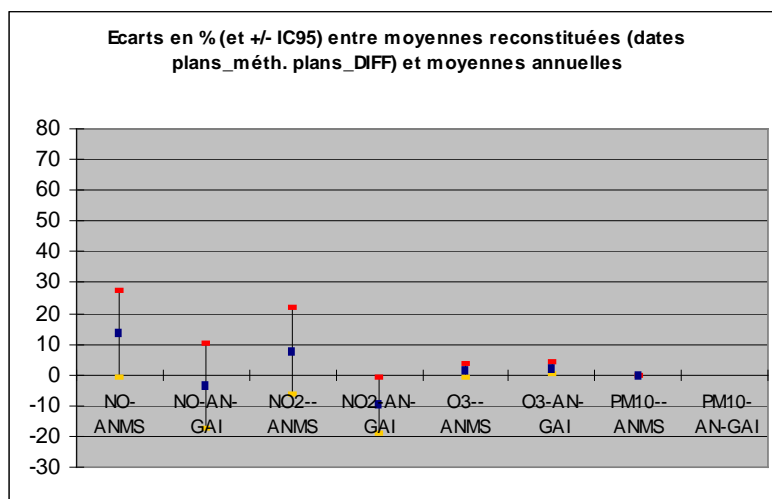
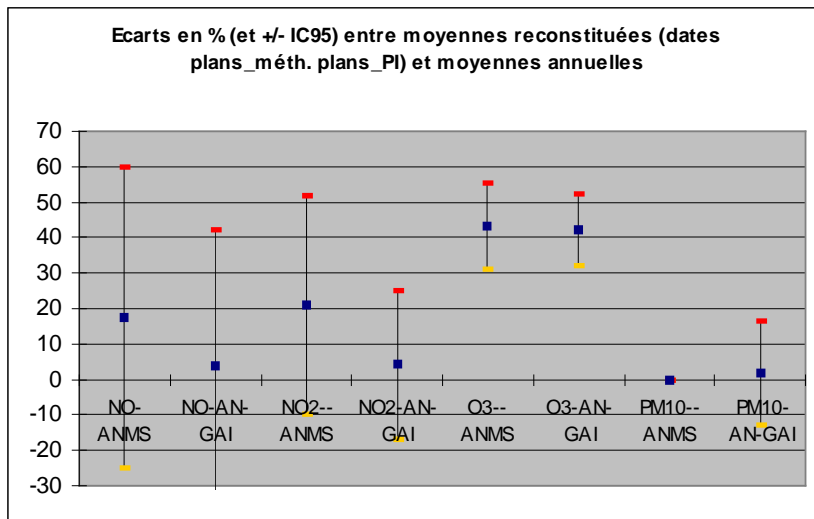


Figure 14: les moyennes reconstituées à Annemasse (en  $\mu\text{g.m}^{-3}$ )

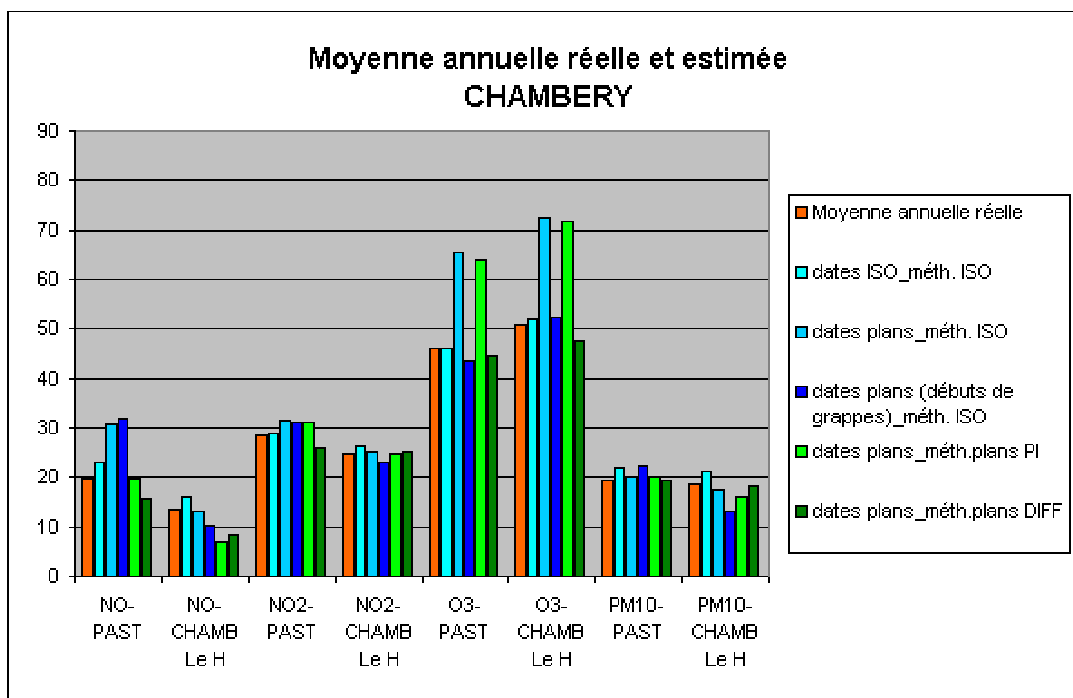
Polluant	NO		NO2		O3		PM <sub>10</sub>	
	ANMS	GAI	ANMS	GAI	ANMS	GAI	ANMS	GAI
Entre 0 et 1	++	++	+	+	-	-		0
Entre 0 et 2	++	++	++	+	++	++		0
Entre 0 et 3	+	++	+	++	--	--		+
Entre 0 et 4	0	0	+	+	++	++		0
Entre 0 et 5	0	0	+	-	0	0		

Tableau 16: comparaison des méthodes à Annemasse



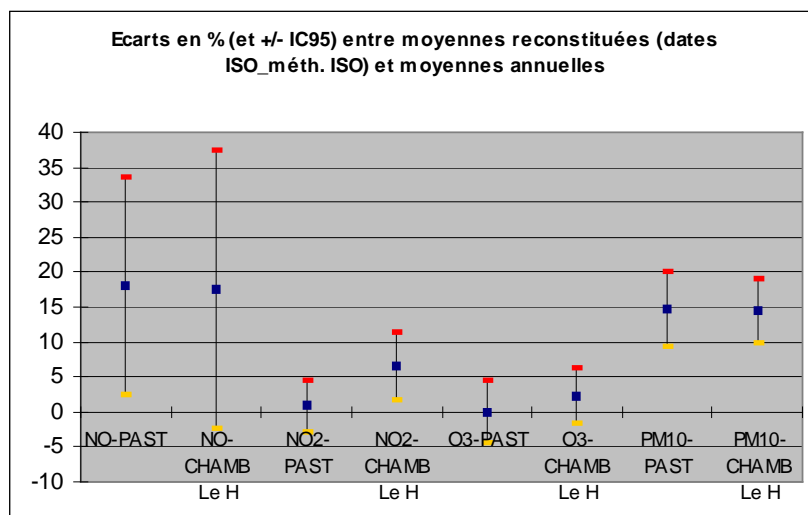


## 5.7.7 CHAMBERY

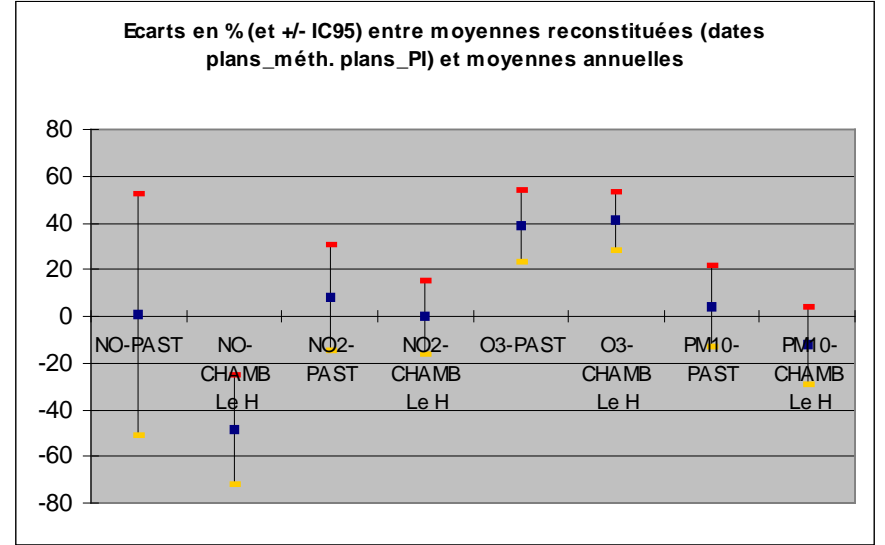
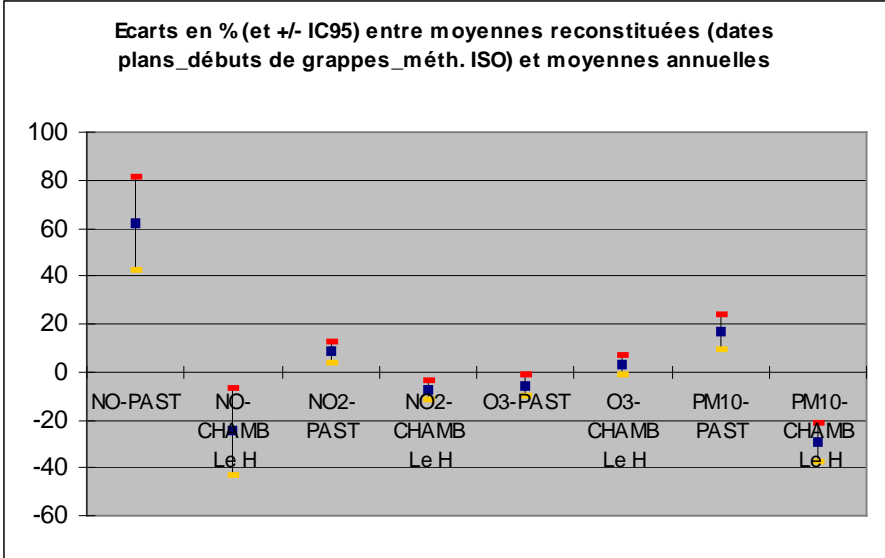
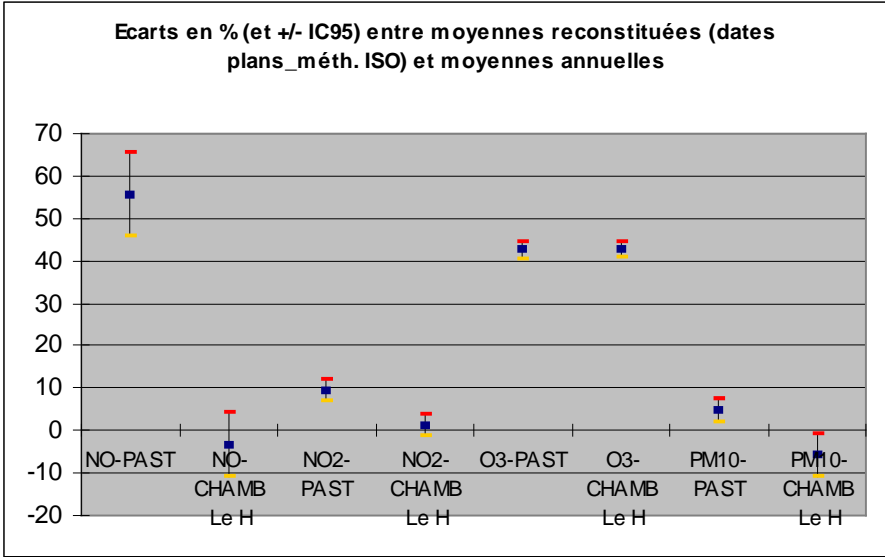


Polluant	NO		NO2		O3		PM <sub>10</sub>	
	PAST	CHAM	PAST	CHAM	PAST	CHAM	PAST	CHAM
Entre 0 et 1	+	+	0	0	0	0	+	+
Entre 0 et 2	++	0	+	0	++	++	0	0
Entre 0 et 3	++	-	+	-	0	0	+	--
Entre 0 et 4	0	--	+	0	++	++	0	-
Entre 0 et 5	-	--	-	0	0	-	0	0

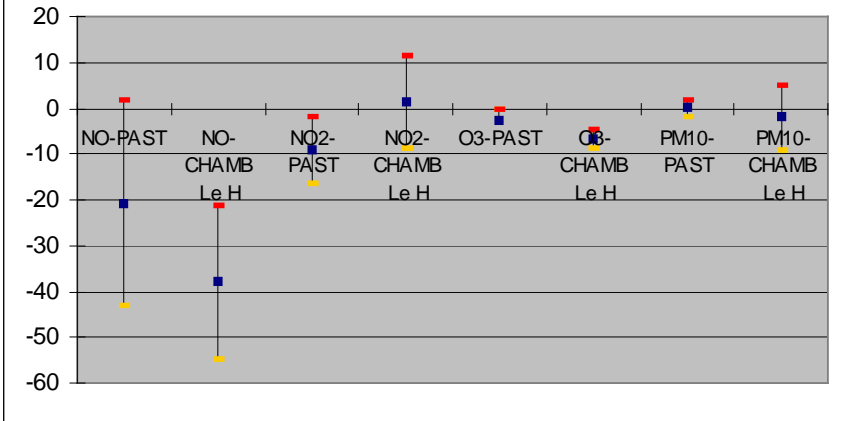
Tableau 17: comparaison des méthodes à Chambéry







**Ecart en % (et +/- IC95) entre moyennes reconstituées (dates plans\_méth. plans\_DIFF) et moyennes annuelles**



## 5.8 RECONSTITUTION DES NOMBRES DE DEPASSEMENTS DE SEUILS

### 5.8.1 TOULOUSE

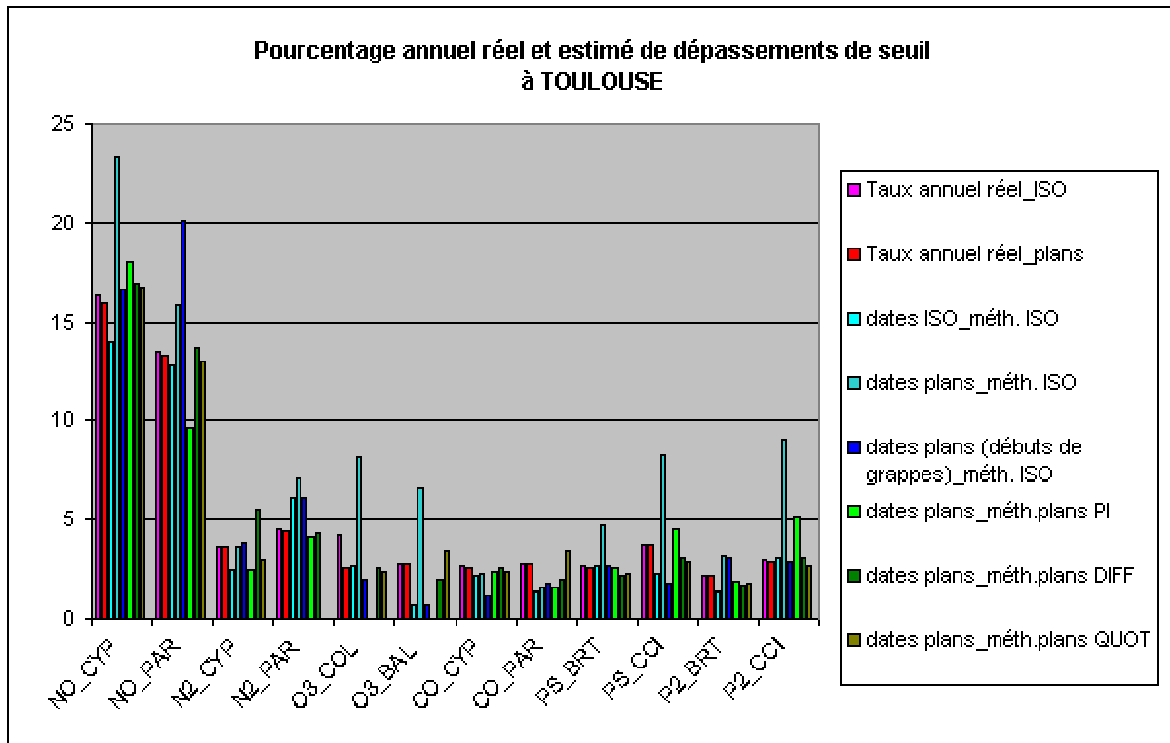
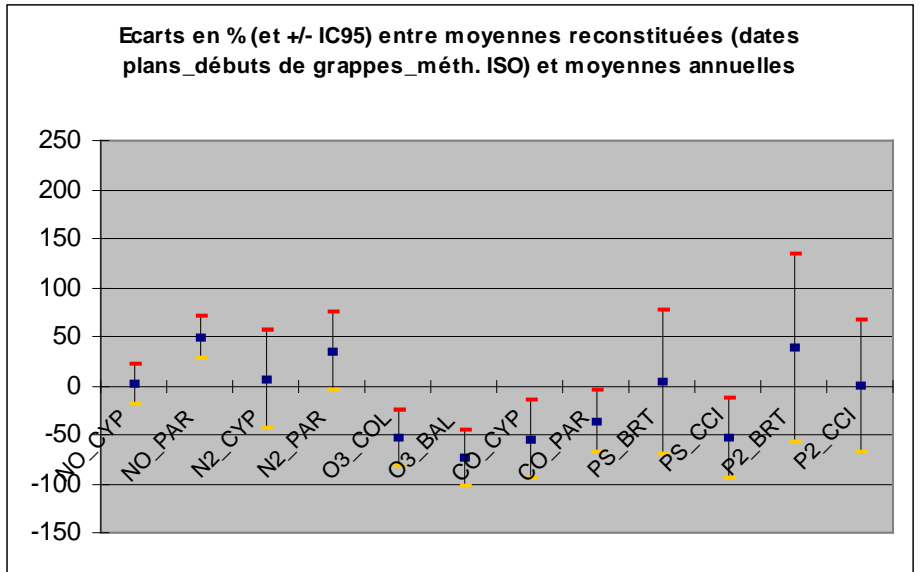
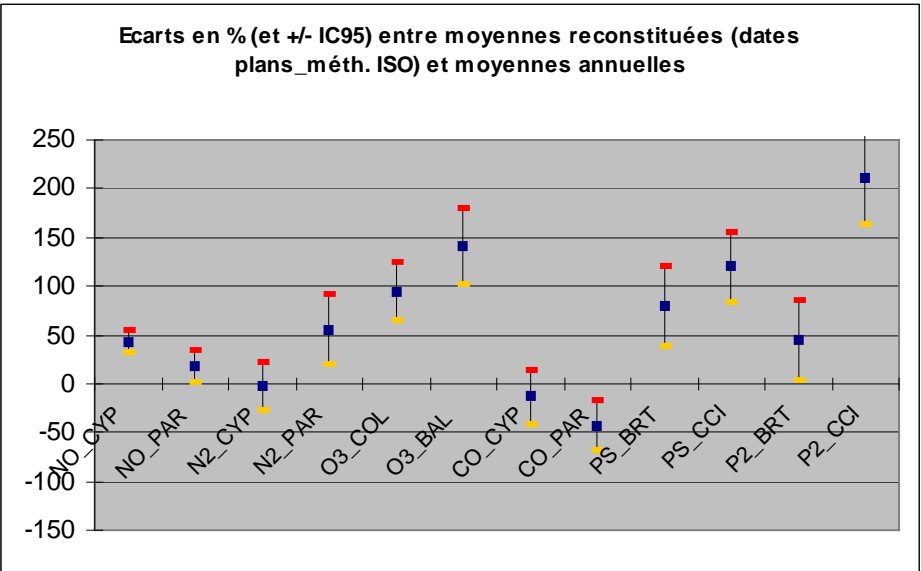
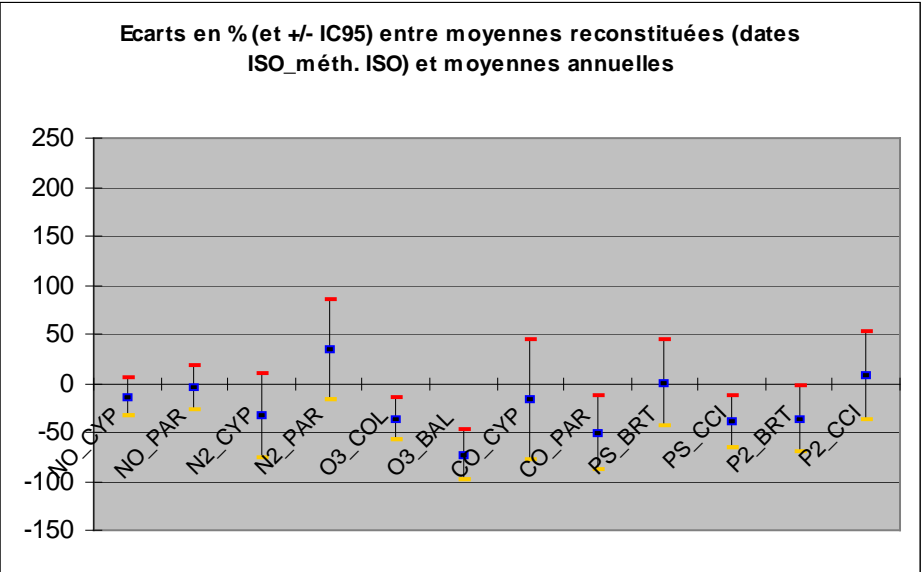
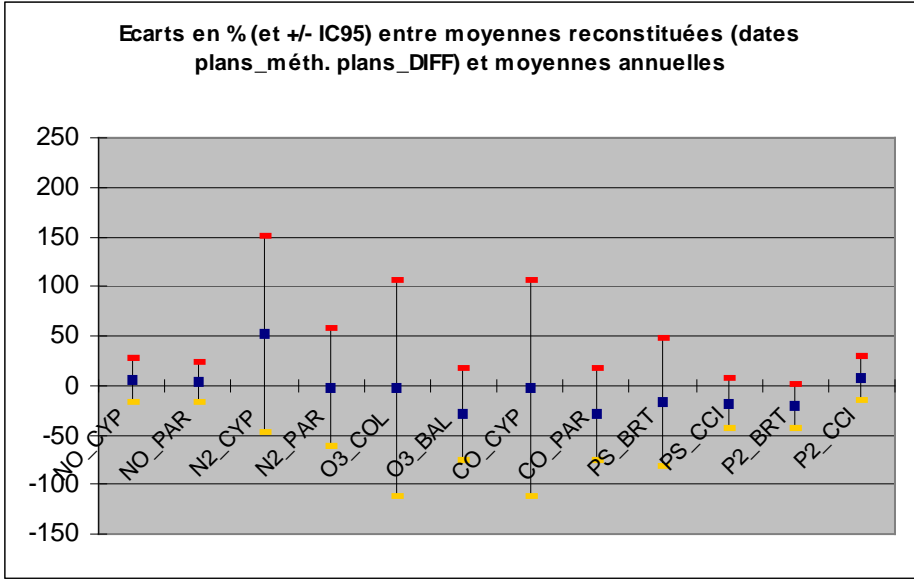


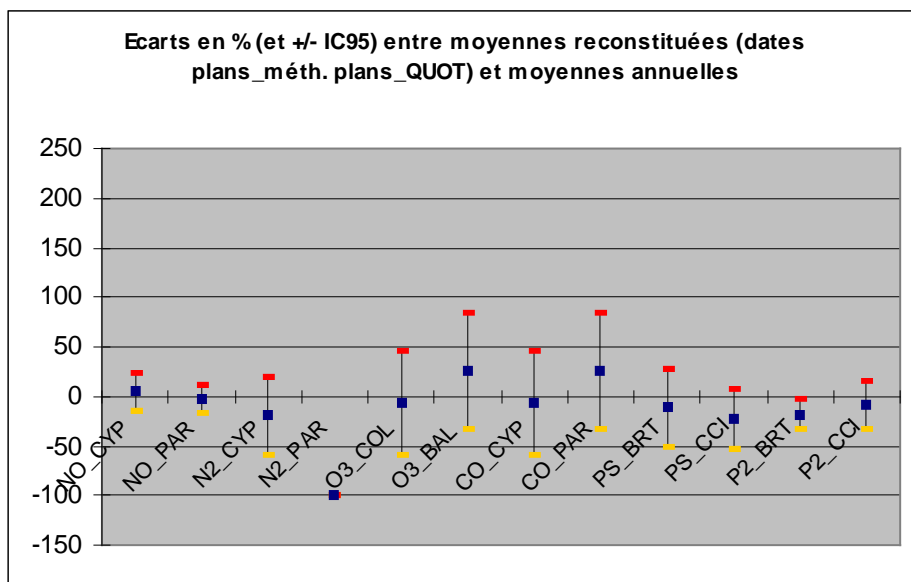
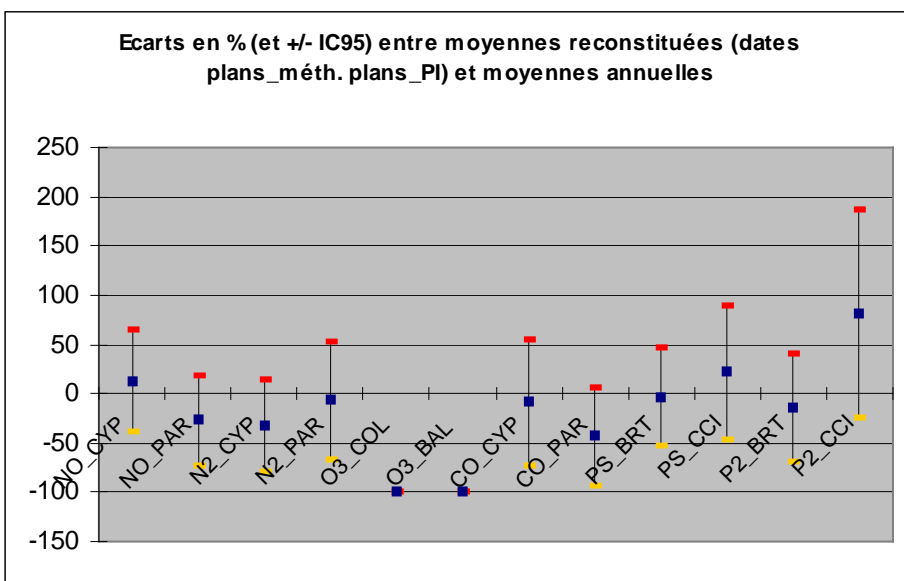
Figure 16 : nombres de dépassements reconstitués à Toulouse

Polluant	NO		NO2		O3		CO		PM <sub>10</sub>		PM <sub>2.5</sub>	
	CYP	PAR	CYP	PAR	COL	BAL	CYP	PAR	BRT	CCI	BRT	CCI
Entre 0 et 1	-	0	--	++	--	--	0	--	0	--	--	0
Entre 0 et 2	++	+	0	++	++	++	0	--	++	++	++	++
Entre 0 et 3	0	++	0	++	--	--	--	--	0	--	++	0
Entre 0 et 4	+	--	--	0			0	--	0	+	-	++
Entre 0 et 5	0	0	++	0	0	-	0	--	-	-	-	0
Entre 0 et 6	0	0	--		0	+	0	++	-	-	-	-

Tableau 18: comparaison des méthodes à Toulouse







## 5.8.2 STRASBOURG

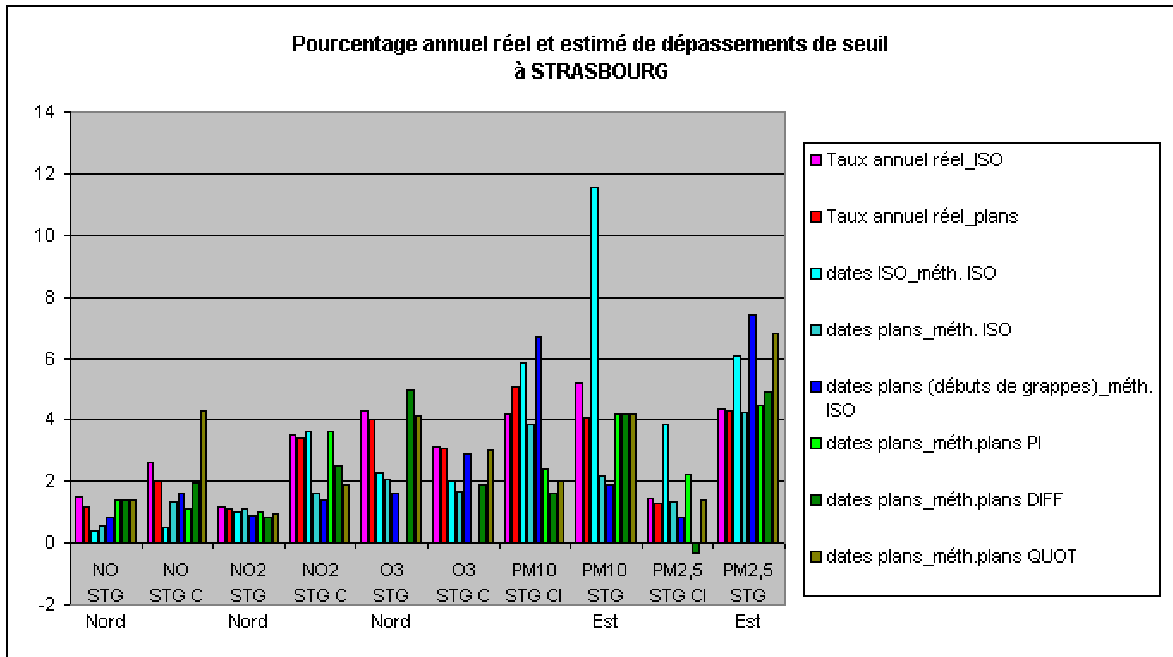
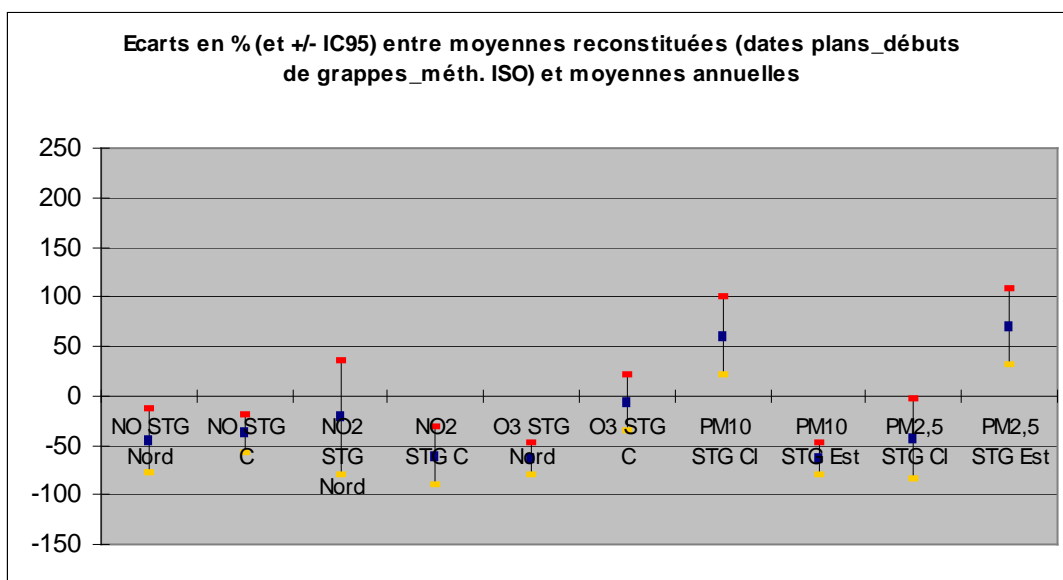
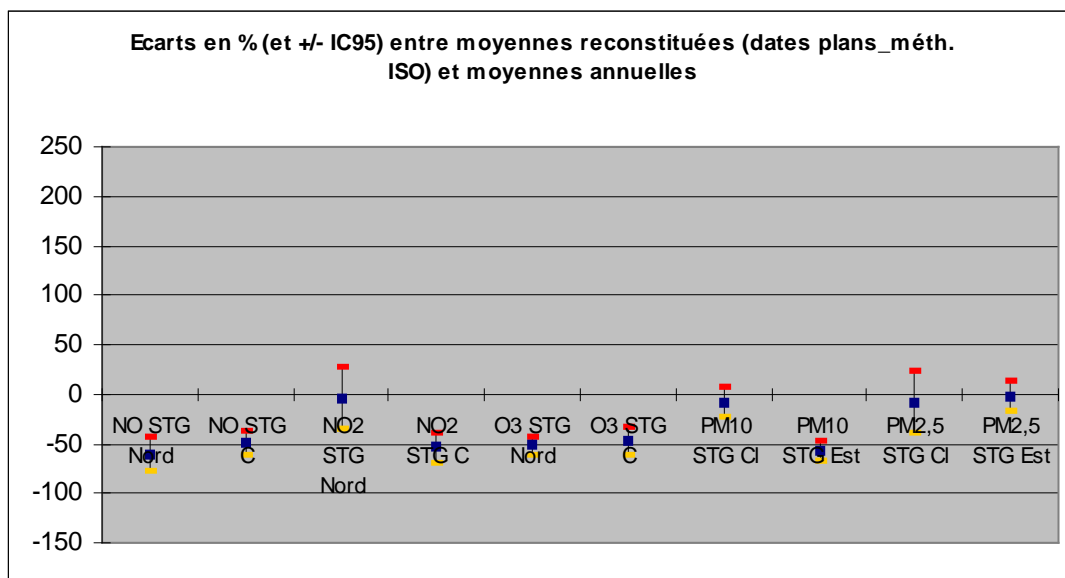
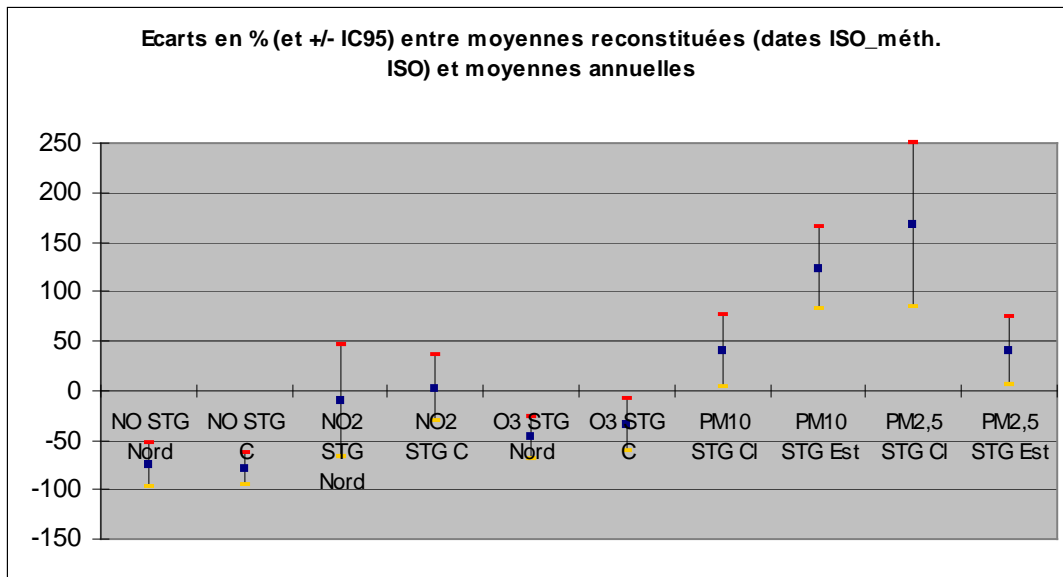


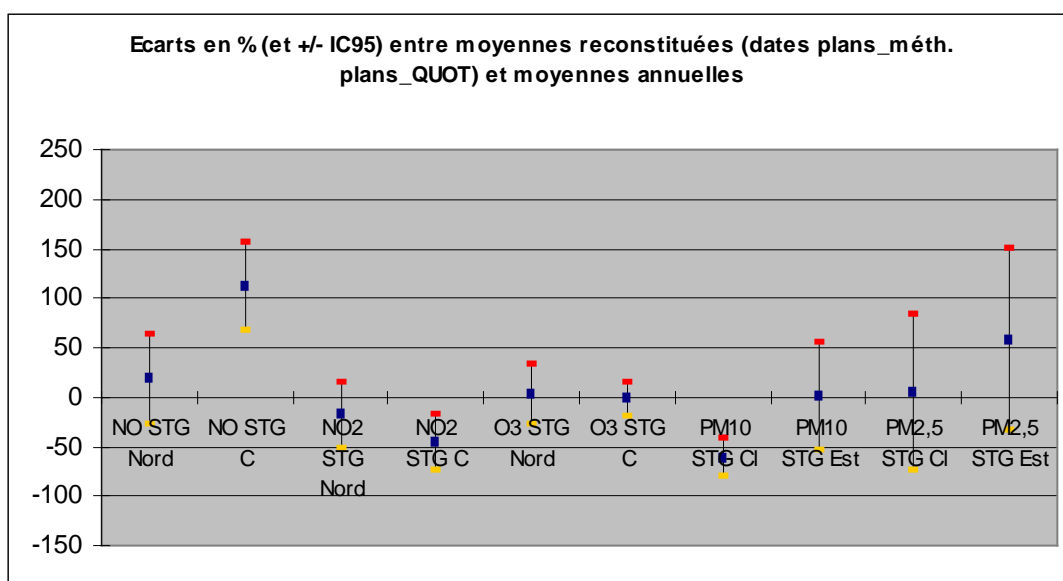
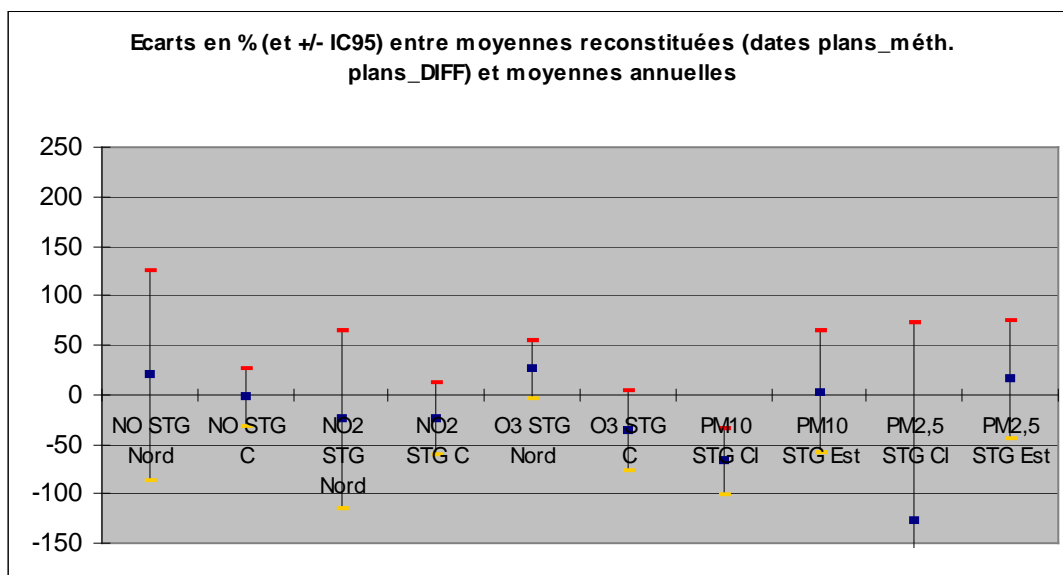
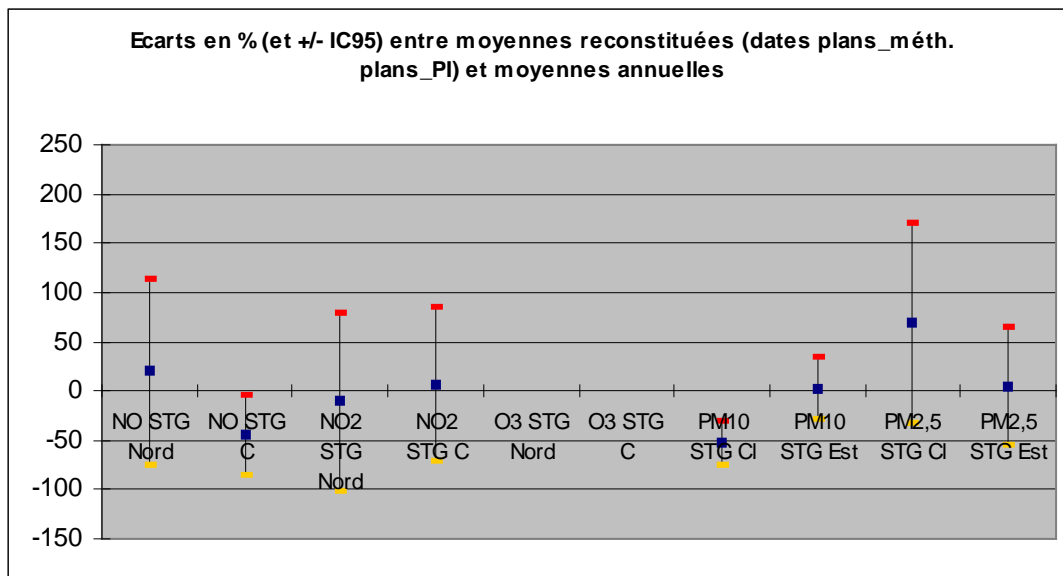
Figure 17: nombres de dépassements reconstitués à Strasbourg

Polluant	NO		NO2		O3		PM <sub>10</sub>		PM <sub>2.5</sub>	
	STG-N	STG-C	STG-N	STG-C	STG-N	STG-C	STG-CI	STG-E	STG-CI	STG-E
Entre 0 et 1	--	--	0	0	--	--	++	++	++	++
Entre 0 et 2	--	--	0	--	--	--	0	--	0	0
Entre 0 et 3	--	--	-	--	--	0	++	--	--	++
Entre 0 et 4	+	--	-	0	--	--	--	0	++	0
Entre 0 et 5	+	0	-	--	+	--	--	0	--	+
Entre 0 et 6	+	++	-	--	0	0	--	0	0	++

Tableau 19: comparaison des méthodes à Strasbourg







### 5.8.3 DUNKERQUE

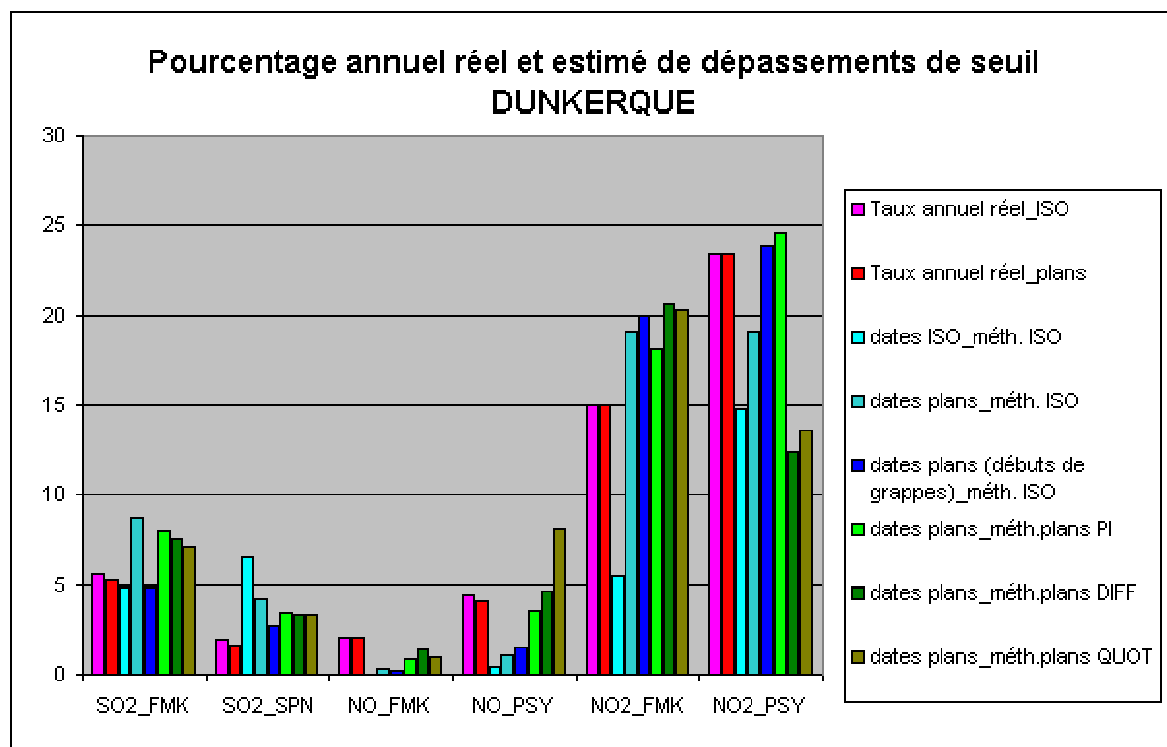
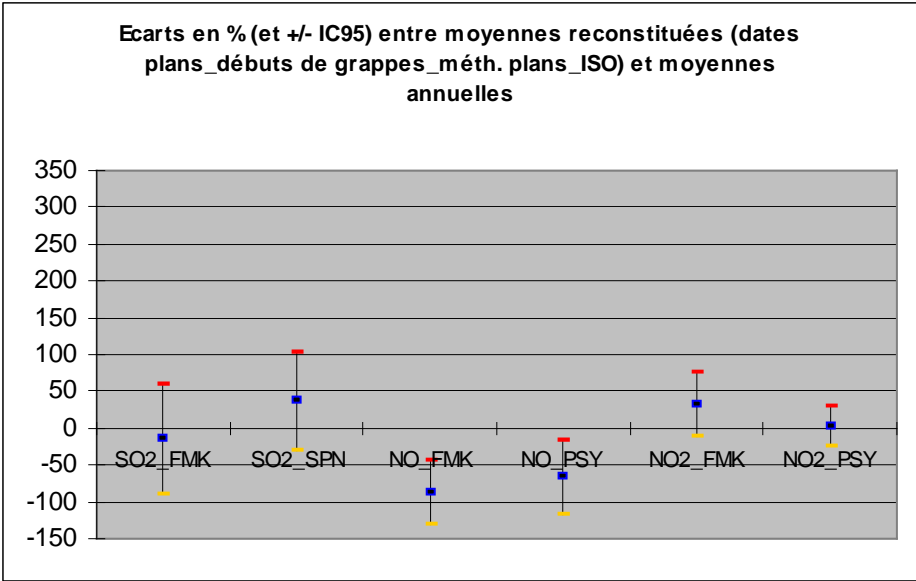
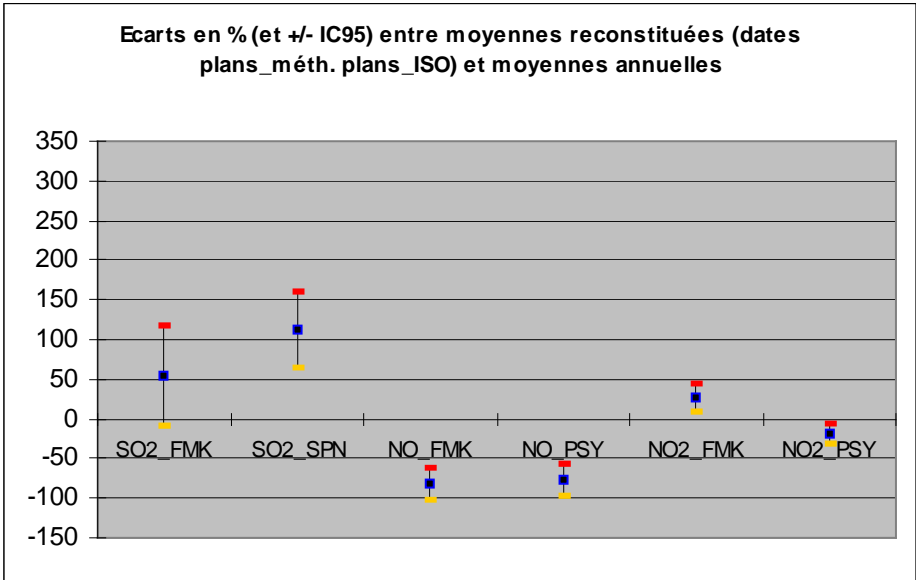
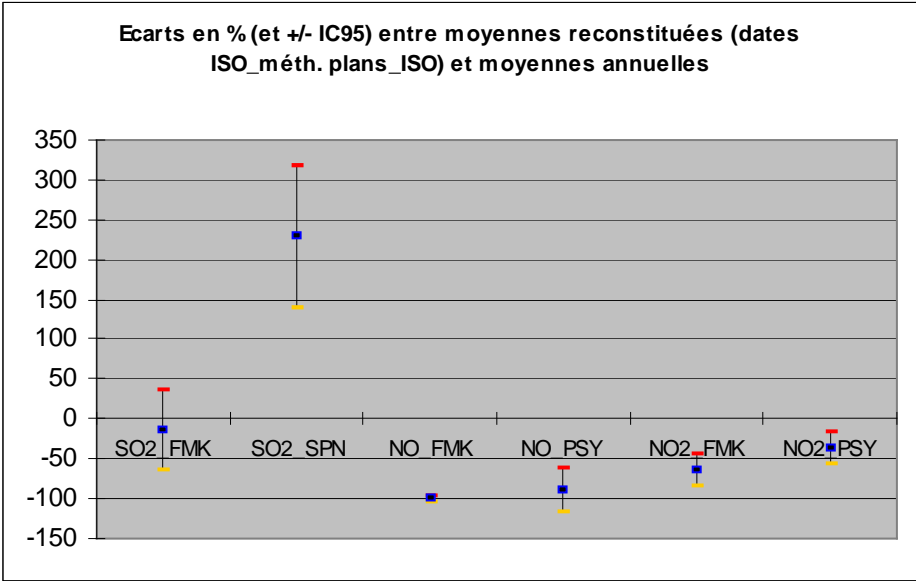


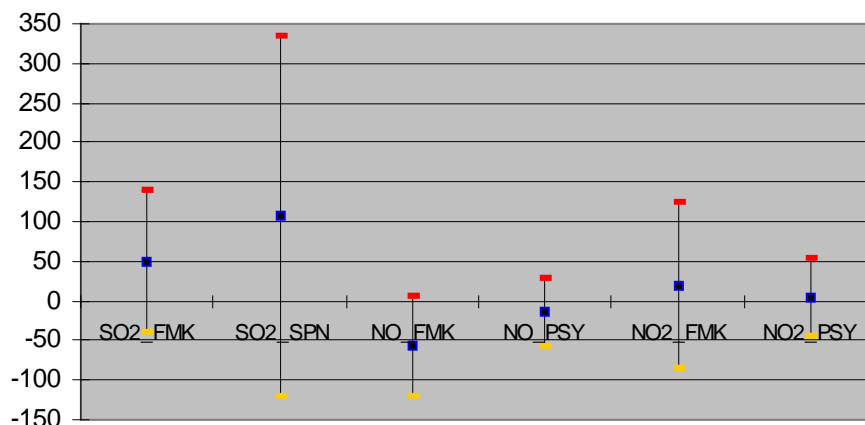
Figure 18: nombres de dépassements reconstitués à Dunkerque

Polluant	SO2		NO		NO2	
	FMK	SPN	FMK	PSY	FMK	PSY
Entre 0 et 1	-	++	--	--	--	--
Entre 0 et 2	++	+	--	--	+	--
Entre 0 et 3	-	+	--	--	+	0
Entre 0 et 4	++	+	--	-	+	0
Entre 0 et 5	++	+	--	0	+	--
Entre 0 et 6	++	+	--	++	+	--

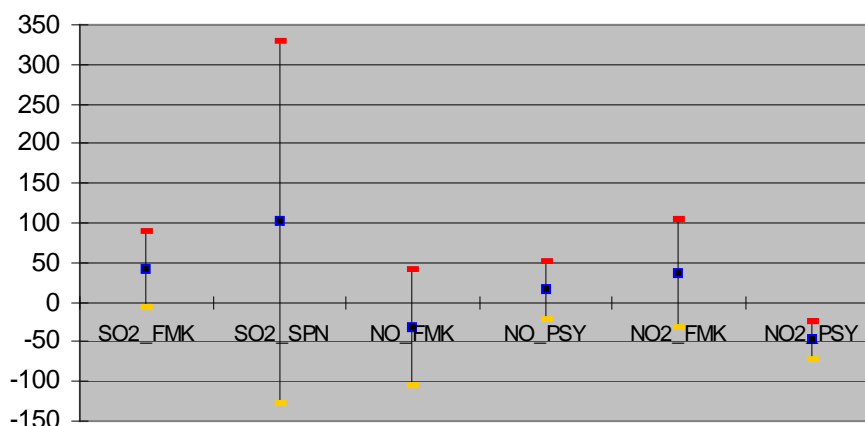
Tableau 20: comparaison des méthodes à Dunkerque



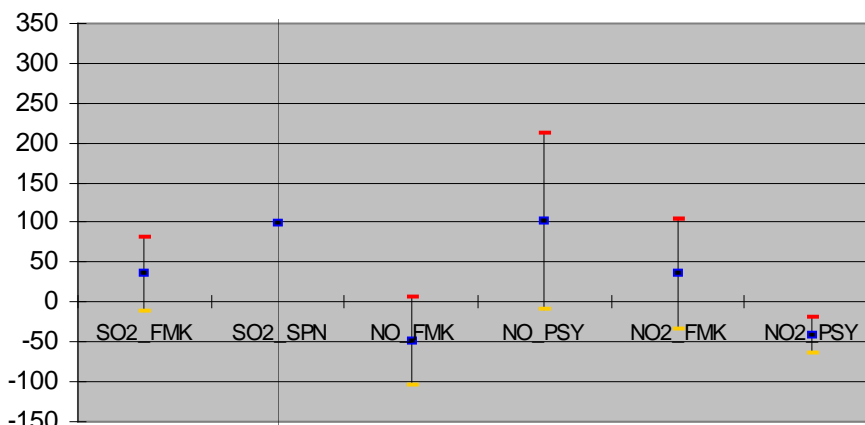
**Ecarts en % (et +/- IC95) entre moyennes reconstituées (dates plans\_méth. plans\_PI) et moyennes annuelles**



**Ecarts en % (et +/- IC95) entre moyennes reconstituées (dates plans\_méth. plans\_DIFF) et moyennes annuelles**



**Ecarts en % (et +/- IC95) entre moyennes reconstituées (dates plans\_méth. plans\_QUOT) et moyennes annuelles**



## 5.8.4 LE HAVRE

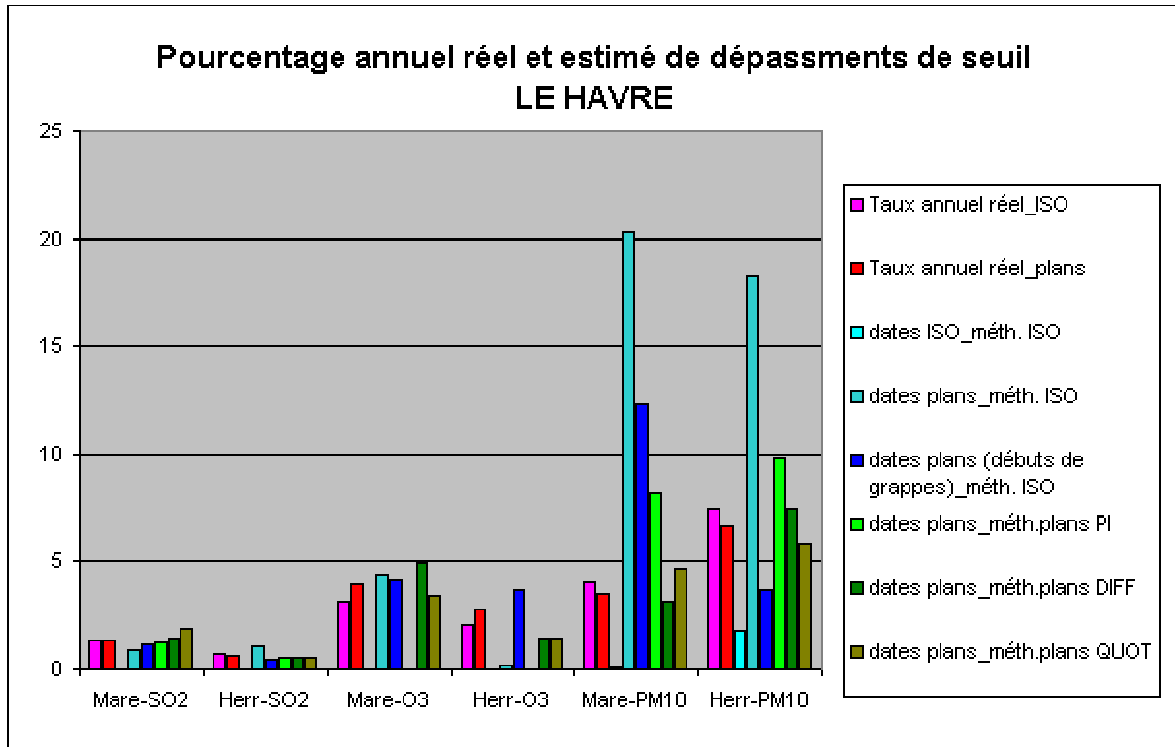
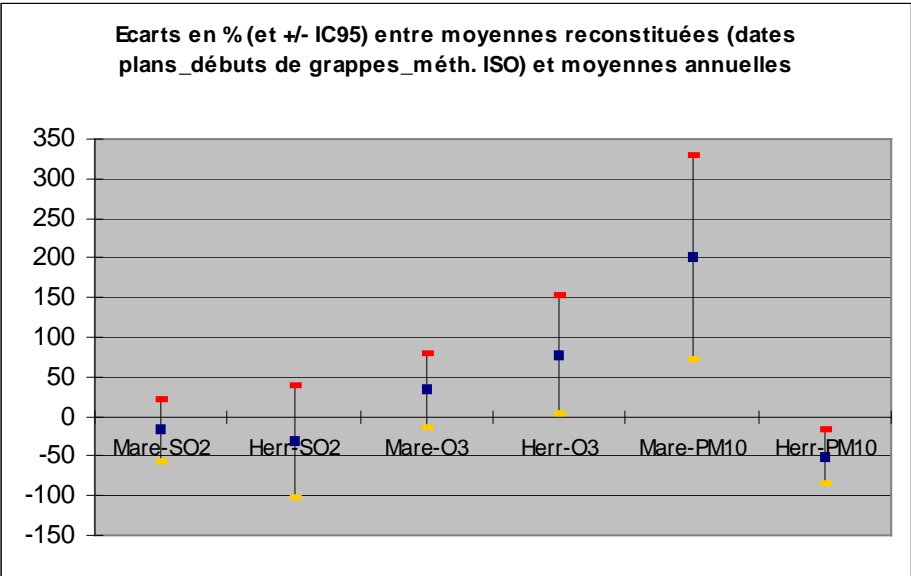
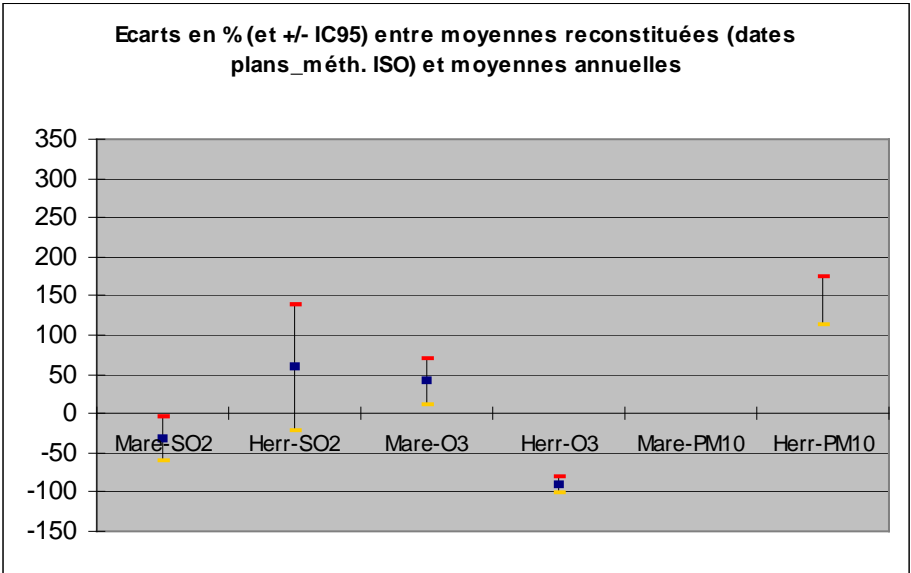
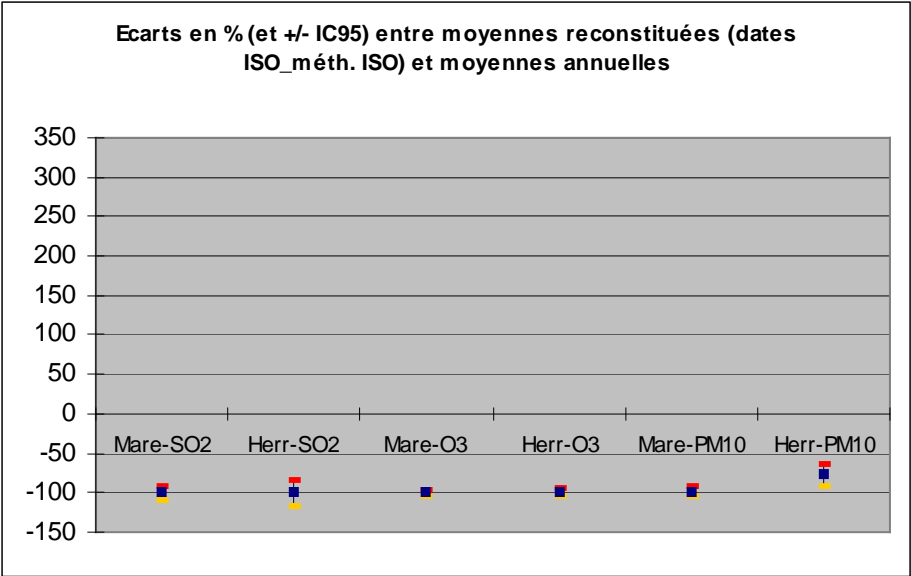
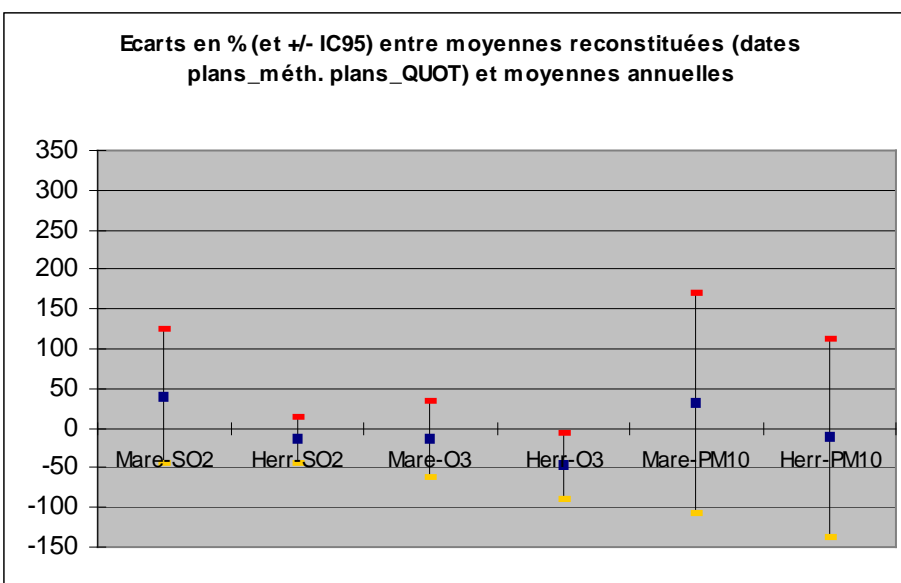
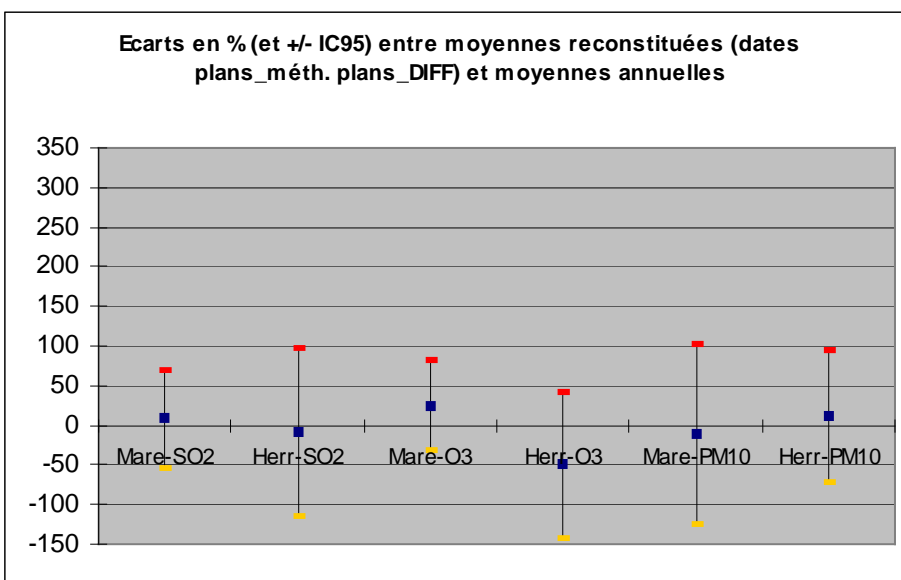
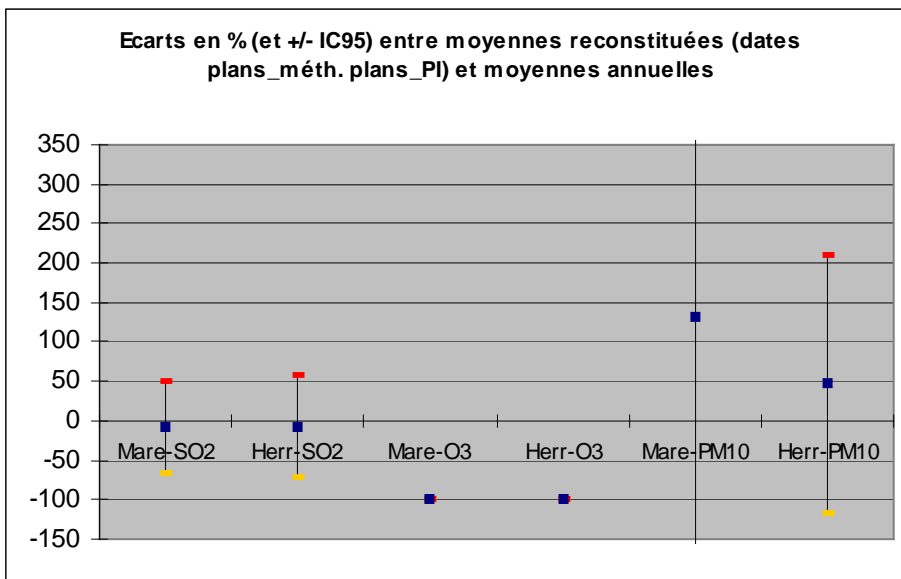


Figure 19: nombres de dépassements reconstitués à Le Havre

Polluant	SO2		O3		PM10	
	Mare	Herr	Mare	Herr	Mare	Herr
Entre 0 et 1					--	--
Entre 0 et 2	-	+	+	--	++	++
Entre 0 et 3	0	0	+	+	++	-
Entre 0 et 4	0	0			++	+
Entre 0 et 5	0	0	+	-	0	0
Entre 0 et 6	+	0	0	-	+	-

Tableau 21: comparaison des méthodes à Le Havre





## 5.9 CONCLUSIONS

Dans l'état actuel de l'étude, les conclusions ne porteront que sur les résultats bruts des reconstitutions de moyennes et de taux de dépassements de seuils. Si les incertitudes associées à ces résultats, présentées sous forme d'intervalles de confiance, ont été également calculées, les différences constatées entre les plans d'échantillonnage et les méthodes montrent la difficulté d'en fournir une estimation. Ce point sera discuté et approfondi dans le cadre du groupe de travail *Plans d'échantillonnage et reconstitution de données*.

Il apparaît en règle générale que, quelles que soient les méthodes d'échantillonnage et de reconstitution pratiquées, la fréquence d'occurrence des pointes de pollution pendant les campagnes de mesure demeure un paramètre non maîtrisable. En effet, selon que cette fréquence est élevée ou non, les reconstitutions des moyennes et des dépassement de seuils sont respectivement sur- ou sous-estimées.

Les principales disparités observées entre les méthodes décrites proviennent plus de ces taux d'occurrence que des méthodologies elle-mêmes, en d'autres termes, plus du hasard dans la sélection des dates de campagnes que du choix de leurs durées. Ces disparités dépendent aussi du contexte géographique ; la convergence entre les méthodes varie en effet d'une ville à l'autre.

Cependant, pour certains polluants dont la variabilité temporelle est peu élevée, comme NO<sub>2</sub>, les PM, les reconstitutions sont meilleures (erreur relative inférieure à 15%). A l'inverse, pour les polluants dits de proximité, comme SO<sub>2</sub>, NO, CO, les reconstitutions sont plus sujettes à caution. Quant à l'ozone, les reconstitutions sont dans tous les cas difficiles à réaliser.

On obtient néanmoins des résultats acceptables pour la majorité des stations choisies dans cette étude, aussi bien pour les reconstitutions des moyennes que pour celles des dépassements.

Il conviendrait de mieux intégrer dans les incertitudes cette part d'aléatoire due au hasard du temps et de l'espace (représentativité des points de mesure). Le groupe de travail prévoit cet élargissement du travail.



## **6. LISTE DES ANNEXES**

<b>Repère</b>	<b>Désignation</b>	<b>Nombre de pages</b>
1	Méthode PLANS : présentation des programmes	6 A4
2	Méthode PLANS : mode opératoire	9 A4
3	Méthode PLANS : étude de plans d'échantillonnage. Liste des échantillonnages simulés	8 A4
4	Méthode PLANS : étude de plans d'échantillonnage. Exemples de résultats numériques	2 A4
5	Présentation des sites de mesure dont sont exploitées les données	6 A4
6	Méthode de calcul ISO 9359	11 A4

## Annexe 1

### Descriptif des programmes

Etape	Programme	Données d'entrée	Résultats
Elaboration d'un plan d'échantillonnage	nopt	<ul style="list-style-type: none"> <li>▪ Série auxiliaire</li> <li>▪ Nombre de strates dans la saison ou l'année étudiée</li> <li>▪ Taille des grappes (durée des périodes de mesures exprimées en nombre d'heures)</li> <li>▪ Incertitude souhaitée</li> <li>▪ Paramètre booléen indiquant si l'incertitude est exprimée en valeur absolue ou relative</li> </ul>	Fonction utilisée dans la fonction choixn ci-après
Elaboration d'un plan d'échantillonnage	choixn	<ul style="list-style-type: none"> <li>▪ Série auxiliaire</li> <li>▪ Incertitude souhaitée</li> <li>▪ Pas de la série temporelle (exprimé en nombre d'heures)</li> <li>▪ Paramètre booléen indiquant si l'incertitude est exprimée en valeur absolue ou relative</li> </ul>	Donne, en fonction du nombre de strates et de la durée des périodes de mesure, le nombre total de jours de mesure nécessaires pour atteindre la précision fixée.
Elaboration d'un plan d'échantillonnage	choixnh	<ul style="list-style-type: none"> <li>▪ Série auxiliaire</li> <li>▪ Nombre de strates dans la saison ou l'année étudiée</li> <li>▪ Nombre total de grappes</li> <li>▪ Taille des grappes (durée des périodes de mesures exprimées en nombre d'heures)</li> </ul>	Fournit le nombre optimal de périodes de mesure dans chacune des strates
Elaboration d'un plan d'échantillonnage	varstratopt	<ul style="list-style-type: none"> <li>▪ Série de données</li> <li>▪ Nombre de strates</li> <li>▪ Nombre total de grappes</li> <li>▪ Taille des grappes</li> </ul>	Renvoie la plus petite variance espérée pour un plan d'au total H strates et n grappes
Simulation d'un échantillonnage	choixdate	<ul style="list-style-type: none"> <li>▪ Série complète de mesures au point étudié</li> <li>▪ Série auxiliaire</li> <li>▪ Nombre nh de grappes (périodes de mesure) dans la strate</li> <li>▪ Taille des grappes</li> </ul>	Tire aléatoirement nh périodes de mesure dans une strate donnée (fonction utilisée dans le programme ech ci-après)

Simulation d'un échantillonnage	ech	<ul style="list-style-type: none"> <li>• Série complète de mesures au point étudié</li> <li>• Série auxiliaire</li> <li>• Vecteur donnant le nombre de grappes dans chaque strate</li> <li>• Taille des grappes</li> </ul>	Tire un échantillon selon un plan de sondage défini
Estimation	varstrat	<ul style="list-style-type: none"> <li>• Série de données</li> <li>• Vecteur donnant le nombre de grappes dans chaque strate</li> <li>• Taille des grappes</li> </ul>	Calcule la variance d'un plan stratifié quelconque
Estimation	varstratest	<ul style="list-style-type: none"> <li>• Taille de la série complète</li> <li>• Echantillon de la série</li> <li>• Vecteur donnant le nombre de grappes dans chaque strate</li> <li>• Taille des grappes</li> </ul>	Calcule l'estimateur de la variance d'un plan stratifié quelconque
Estimation	varstratdiff	<ul style="list-style-type: none"> <li>• Série de données</li> <li>• Série auxiliaire de données</li> <li>• Vecteur donnant le nombre de grappes dans chaque strate</li> <li>• Taille des grappes</li> </ul>	Calcule la variance de l'estimateur de la moyenne redressé par la différence
Estimation	varstratdiffest	<ul style="list-style-type: none"> <li>• Echantillon de la série de données</li> <li>• Série auxiliaire de données</li> <li>• Echantillon de la série auxiliaire</li> <li>• Vecteur donnant le nombre de grappes dans chaque strate</li> <li>• Taille des grappes</li> </ul>	Calcule l'estimateur de la variance d'un plan stratifié quelconque
Estimation	varstratquot	<ul style="list-style-type: none"> <li>• Série de données</li> <li>• Série auxiliaire de données</li> <li>• Vecteur donnant le nombre de grappes dans chaque strate</li> <li>• Taille des grappes</li> </ul>	Calcule la variance de l'estimateur de la moyenne redressé par le quotient
Estimation	varstratquotest	<ul style="list-style-type: none"> <li>• Echantillon de la série de données</li> <li>• Série auxiliaire de données</li> <li>• Echantillon de la série auxiliaire</li> <li>• Vecteur donnant le nombre de grappes dans chaque strate</li> <li>• Taille des grappes</li> </ul>	Calcule l'estimateur de la variance d'un plan stratifié quelconque

Estimation	varstratreg	<ul style="list-style-type: none"> <li>▪ Série de données</li> <li>▪ Série auxiliaire de données</li> <li>▪ Vecteur donnant le nombre de grappes dans chaque strate</li> <li>▪ Taille des grappes</li> </ul>	Calcule la variance de l'estimateur de la moyenne redressé par la régression
Estimation	varstratregest	<ul style="list-style-type: none"> <li>▪ Echantillon de la série de données</li> <li>▪ Série auxiliaire de données</li> <li>▪ Echantillon de la série auxiliaire</li> <li>▪ Vecteur donnant le nombre de grappes dans chaque strate</li> <li>▪ Taille des grappes</li> </ul>	Calcule l'estimateur de la variance d'un plan stratifié quelconque
Estimation	moydiff	<ul style="list-style-type: none"> <li>▪ Echantillon de la série de données</li> <li>▪ Série auxiliaire de données</li> <li>▪ Echantillon de la série auxiliaire</li> <li>▪ Vecteur donnant le nombre de grappes dans chaque strate</li> <li>▪ Taille des grappes</li> </ul>	Calcule l'estimateur de la moyenne redressé par la différence
Estimation	moyquot	<ul style="list-style-type: none"> <li>▪ Echantillon de la série de données</li> <li>▪ Série auxiliaire de données</li> <li>▪ Echantillon de la série auxiliaire</li> <li>▪ Vecteur donnant le nombre de grappes dans chaque strate</li> <li>▪ Taille des grappes</li> </ul>	Calcule l'estimateur de la moyenne redressé par le quotient
Estimation	moyreg	<ul style="list-style-type: none"> <li>▪ Echantillon de la série de données</li> <li>▪ Série auxiliaire de données</li> <li>▪ Echantillon de la série auxiliaire</li> <li>▪ Vecteur donnant le nombre de grappes dans chaque strate</li> <li>▪ Taille des grappes</li> </ul>	Calcule l'estimateur de la moyenne redressé par la régression

Estimation	estmoy	<ul style="list-style-type: none"> <li>• Echantillon de la série de données</li> <li>• Série auxiliaire de données</li> <li>• Echantillon de la série auxiliaire</li> <li>• Vecteur donnant le nombre de grappes dans chaque strate</li> <li>• Taille des grappes</li> </ul>	Renvoie l'estimateur de la moyenne, les bornes inf et sup de l'IC95 <b>estimé</b> et la largeur de l'IC95 estimé, pour l'estimateur PI et l'estimateur redressé par la différence, la regression et le quotient.
Estimation	estmoytheo	<ul style="list-style-type: none"> <li>• Série complète de données</li> <li>• Echantillon de la série de données</li> <li>• Série auxiliaire de données</li> <li>• Echantillon de la série auxiliaire</li> <li>• Vecteur donnant le nombre de grappes dans chaque strate</li> <li>• Taille des grappes</li> </ul>	Renvoie l'estimateur de la moyenne, les bornes inf et sup de l'IC95 <b>théorique</b> et la largeur de l'IC95 théorique, pour l'estimateur PI et l'estimateur redressé par la différence, la regression et le quotient.
	estmoypi	<ul style="list-style-type: none"> <li>• Echantillon de la série de données</li> <li>• Vecteur donnant le nombre de grappes dans chaque strate</li> <li>• Taille des grappes</li> </ul>	idem estmoy mais sans redressement
	estmoypitheo	<ul style="list-style-type: none"> <li>• Série complète de données</li> <li>• Echantillon de la série de données</li> <li>• Vecteur donnant le nombre de grappes dans chaque strate</li> <li>• Taille des grappes</li> </ul>	idem estmoytheo mais sans redressement
Estimation	estprop	<ul style="list-style-type: none"> <li>• Echantillon de la série de données</li> <li>• Série auxiliaire de données</li> <li>• Echantillon de la série auxiliaire</li> <li>• Valeur seuil</li> <li>• Vecteur donnant le nombre de grappes dans chaque strate</li> <li>• Taille des grappes</li> </ul>	Renvoie l'estimateur du taux de dépassement de seuil, les bornes inf et sup de l'IC95 <b>estimé</b> et la largeur de l'IC95 estimé, pour l'estimateur PI et l'estimateur redressé par la différence, la regression et le quotient
Estimation	estprop%	<ul style="list-style-type: none"> <li>• Idem ci-dessus</li> <li>• Taux de confiance de l'IC</li> </ul>	Même programme que précédemment. L'utilisateur peut choisir un taux de confiance différent de 95%.

Estimation	estproptheo	<ul style="list-style-type: none"> <li>▪ Série complète de données</li> <li>▪ Echantillon de la série de données</li> <li>▪ Série auxiliaire de données</li> <li>▪ Echantillon de la série auxiliaire</li> <li>▪ Valeur seuil</li> <li>▪ Vecteur donnant le nombre de grappes dans chaque strate</li> <li>▪ Taille des grappes</li> </ul>	Renvoie l'estimateur du taux de dépassement de seuil, les bornes inf et sup de l'IC95 <b>théorique</b> et la largeur de l'IC95 théorique, pour l'estimateur PI et l'estimateur redressé par la différence, la regression et le quotient
Estimation	Estproptheo%	<ul style="list-style-type: none"> <li>▪ Idem ci-dessus</li> <li>▪ Taux de confiance de l'IC</li> </ul>	Même programme que précédemment. L'utilisateur peut choisir un taux de confiance différent de 95%.
Elaboration d'un plan d'échantillonnage Estimation	varinterg	<ul style="list-style-type: none"> <li>▪ Série de données</li> <li>▪ Taille des grappes</li> </ul>	Renvoie la variance intergrappes corrigée. Fonction utilisée par les fonctions <i>nopt</i> , <i>choixnh</i> et par les différentes fonctions de calcul de variance
Estimation	covinterg	<ul style="list-style-type: none"> <li>▪ Série de données</li> <li>▪ Série auxiliaire</li> <li>▪ Taille des grappes</li> </ul>	Renvoie la covariance intergrappes corrigée
Simulations	simul_ssgroupe retengroupant	<ul style="list-style-type: none"> <li>▪ Tableau des séries de données</li> <li>▪ Résolution temporelle des séries</li> <li>▪ Nom du polluant considéré</li> <li>▪ Vecteur donnant le nombre de grappes dans chaque strate</li> <li>▪ Taille des grappes</li> <li>▪ Valeur du seuil</li> <li>▪ Support temporel du seuil</li> <li>▪ Paramètre booléen indiquant si le support temporel du seuil correspond à une moyenne glissante</li> <li>▪ Paramètre booléen indiquant si l'on découpe la série en deux strates estivale et hivernale</li> <li>▪ Nombre de simulations</li> <li>▪ Numéros des colonnes contenant la série choisie pour l'échantillonnage et la série auxiliaire</li> </ul>	<p>Simule le tirage de très nombreux échantillons selon un plan donné. Pour chaque échantillon, ce programme calcule les estimateurs de la moyenne et du taux de dépassement de seuil (si un seuil a été fixé) ainsi que les IC théoriques et estimés. A la fin des tirages, il renvoie le taux de couverture de la vraie moyenne (et du vrai taux de dépassement) par les intervalles de confiance théoriques et estimés.</p> <p>Quand le plan se compose de plusieurs strates, l'estimation est faite :</p> <ul style="list-style-type: none"> <li>- en tenant compte de la stratification telle qu'elle est réellement,</li> </ul> <p>en considérant que toutes les</p>

		<ul style="list-style-type: none"> <li>Nom du répertoire où sont écrits les résultats</li> </ul>	grappes. appartiennent à une unique strate annuelle
Utilitaires	cleanmat	<ul style="list-style-type: none"> <li>Série de données</li> <li>Série auxiliaire</li> </ul>	Enlève simultanément les lignes de données manquantes dans la série étudiée et dans la série auxiliaire
	grouphiver	<ul style="list-style-type: none"> <li>Série de données</li> <li>Pas de la série (en nombre d'heures)</li> </ul>	Place les mois d'octobre-novembre-décembre avant celui de janvier
	calculmoy	<ul style="list-style-type: none"> <li>Série des dates</li> <li>Série de données</li> <li>Pas sur lequel on calcul la moyenne</li> <li>Paramètre booléen indiquant s'il s'agit ou non d'une moyenne glissante</li> </ul>	Fonction utilisée dans le calcul et l'estimation du taux de dépassement de seuil. Transforme une série de concentrations horaires en une série de concentrations calculées sur un plus grand support temporel plus grand (moyennes glissantes ou non).
	calculmoyech	<ul style="list-style-type: none"> <li>Vecteur donnant le nombre de grappes dans chaque strate</li> <li>Taille des grappes</li> <li>Tableau des séries partielles de mesures avec la date en première colonne</li> <li>Pas sur lequel on calcule la moyenne</li> <li>Paramètre booléen indiquant s'il s'agit ou non d'une moyenne glissante</li> </ul>	Même fonction que précédemment, adaptée au cas d'un échantillon de plusieurs séries, tiré selon un plan (nh, ng) : la transformation est faite grappe par grappe.
	Exploitsimul1	<ul style="list-style-type: none"> <li>Nom du répertoire où est écrit le tableau de résultats</li> <li>Résultats des simulations</li> <li>Paramètres de l'échantillonnage</li> </ul>	Fonction utilisée à l'intérieur du programme de simulation Met sous la forme d'un tableau les résultats des simulations
	PROGestimation	<ul style="list-style-type: none"> <li>Paramètres permettant de récupérer les fichiers de données</li> <li>Paramètres de l'échantillonnage</li> </ul>	Met successivement en œuvre les fonctions d'estimation de la moyenne et du dépassement de seuil A adapter par l'utilisateur

## Annexe 2

### Application de la méthode des plans de sondage

#### Notice d'utilisation des programmes R

##### 1. Installation du logiciel R

Télécharger R à l'adresse <http://cran.cict.fr>. Plusieurs versions sont proposées selon le système d'exploitation.

Le **module de base** suffit.

Par défaut, R est installé dans C://Program Files

##### 2. Création des répertoires de travail

On crée un répertoire contenant les fonctions R, un répertoire contenant les données à traiter et un répertoire de résultats, par exemple :

- C ://sondages/fonctions
- C ://sondages/donnees
- C ://sondages/resultats

*L'application de la méthode est décrite à partir d'un exemple concret : l'étude du dioxyde d'azote urbain à Toulouse. On suppose qu'il est prévu d'effectuer un échantillonnage pendant l'année 2002 au site urbain Saint-Cyprien. Les données mesurées en 2001 par la station urbaine Pargaminières servent à définir le plan de sondage ; les données mesurées en 2002 par cette même station sont utilisées, à la suite de l'échantillonnage, pour corriger les estimations.*

##### 3. Détermination d'un plan d'échantillonnage

1 strate=1 division de l'année.

1 grappe=1 période de mesure

##### Ouverture de R

(Dans Linux, taper la commande R)

##### Chargement de toutes les fonctions utiles

```
> source("C://sondages/chargefct.txt")
```

##### Chargement de la base de données

```
> Tlse2001<-read.table("C://sondages/donnees/TLSE_2001.txt")
```

Tableau de chiffres contenant les dates en première colonne (aaaammjj ou jjmmaaaa) et les séries de données de concentration dans les autres colonnes.



### **Vérification des dimensions de la base**

```
> print(dim(Tlse2001))
```

### **L'utilisateur entre les numéros des colonnes contenant les séries de concentration**

```
> numcol<-c(2:30)
```

```
#colonne 2 à colonne 30
```

### **L'utilisateur indique à R de bien considérer les colonnes des données de concentration comme des séries numériques**

```
> for (j in numcol) Tlse2001[,j]<-as.numeric(Tlse2001[,j])
```

### **Sélection des colonnes de la base correspondant aux polluants et aux stations étudiés**

```
> NO2Par01<-Tlse2001[,12]
```

```
#Les données de NO2 de la station Pargaminières sont dans la 12e colonne du fichier.
```

### **Choix du nombre de mesures**

On fixe préalablement une incertitude. La fonction renvoie, pour la série de données choisie et le niveau de précision recherché, le nombre total optimal de grappes à prélever dans l'année en fonction du nombre de strates et de la taille des grappes.

Si l'incertitude est exprimée en valeur absolue (en µg/m<sup>3</sup> par exemple), elle est égale à la demi-largeur de l'intervalle de confiance à 95% :

$$\varepsilon_a = 1/2 \cdot L = 1,96 \cdot S$$

Si elle est exprimée en relatif, elle est égale à la demi-largeur de l'intervalle de confiance divisée par la moyenne annuelle de la série :

$$\varepsilon_r = 100 \cdot 1,96 \cdot \frac{S}{\bar{x}} \quad (\%)$$

S désigne la variance de l'estimateur de la moyenne.

```
> choixn(NO2Par01,10,1,1,ngmin,ngmax)
```

```
#10 : incertitude fixée à 10%
```

```
#1 : la série a un pas horaire (p=24 si série journalière)
```

```
#1 : l'incertitude est donnée en relatif (0 si incertitude absolue)
```

```
#ngmin : taille minimale de grappes en jours
```

```
#ngmax : taille maximale de grappes en jours
```

Cette fonction affiche une série de possibilités, constituées chacune d'un nombre de strates, d'une taille de grappe (en jours) et d'un nombre total optimal de grappes à tirer. A titre d'indication sont fournis le nombre total de jours de mesures dans l'année et le nombre moyen de grappes par strate.

**Les plans contenant moins d'une grappe par strate doivent être exclus. Pour les plans contenant moins de deux grappes par strate, il conviendra de regrouper les strates au moment de l'estimation.**

Plans proposés dans cet exemple pour des grappes d'une semaine :

Regroupement préalable des mois d'hiver ?	Nombre de strates	Nombre total de grappes
non	1	7,4
non	2	6,6
non	3	7,1
non	4	5,1
oui	2	6,8

La dernière proposition est ici retenue. Le nombre de grappes est arrondi à 7.

**Vérification : on s'assure que la combinaison choisie conduit bien à l'incertitude fixée**

> [epsverif\(NO2Par01,1,2,7,168,1,1\)](#)

#1 : la série a un pas horaire

#2 : 2 strates

# 7 : 7 grappes au total

#168 : taille des grappes en nombre de pas (ici en heures)

#1 : incertitude donnée en relatif

#1 : les mois d'hiver sont regroupés avant découpage en deux strates

L'incertitude affichée est de 9,8%, proche des 10% recherchés.

**Répartition des grappes entre les strates**

Pour la (ou les) combinaison(s) sélectionnée(s) parmi les propositions précédentes (nombre de strates, taille de grappe, nombre total de grappes à tirer), on détermine la meilleure répartition des grappes entre les strates.

> [choixnh\(NO2Par01,1,2,7,168,1\)](#)

#1 : la série a un pas horaire

#2 : 2 strates

# 7 : 7 grappes au total

#168 : taille des grappes en nombre de pas (ici en heures)

#1 : les mois d'hiver sont regroupés avant de découper la série deux strates

Le résultat, dans cet exemple, est un vecteur de longueur 2, indiquant le nombre de grappes à prélever dans la strate hivernale (oct.-déc.+jan-mars) et dans la strate estivale (avril-sept.) :

3,95 3,05

Ces nombres sont arrondis à 4 et 3.

**Les résultats de *choixn* et *choixnh* donnent une idée a priori des plans d'échantillonnage qui, pour une station et une année données, conduisent à une certaine précision d'estimation.**

#### 4. Echantillonnage

Cette étape correspond aux mesures effectuées sur le terrain par le moyen mobile conformément au plan établi.

Les dates de début de campagne sont tirées de manière aléatoire dans chaque strate.

Pour illustrer la suite de la démarche, l'échantillonnage est simulé sur une série complète de données.

La fonction *tireech* tire aléatoirement des dates de campagne selon le plan défini. Elle prélève les données correspondantes dans les séries de mesure de la station supposée mobile et de la (ou des) stations auxiliaires.

```
> Tlse2002<- read.table("D://sondages/donnees/TLSE_2002.txt")
```

```
#Tableau des séries de mesure de 2002
```

```
> NO2Par02<-Tlse2002[,12]
```

```
#Les données de NO2 de la station auxiliaire) Pargaminières sont dans la 12e colonne du fichier.
```

```
> date<- Tlse2002[,1]
```

```
#colonne des dates
```

```
> listmp<-as.logical(date)
```

```
#Ici listmp est entièrement constitué de TRUE. Si pour des raisons techniques ou des questions d'organisation, certaines dates ne pouvaient être des débuts de campagne, il conviendrait de les invalider par FALSE. Par exemple, si du 1er au 3 février au matin, il est impossible de se rendre sur le terrain, la commande à taper est :
```

```
> listmp[which(date==c(2002020100 :2002020308))]<-FALSE
```

```
> tirage<-tireechb(Tlse2002,c(8,12),listmp,1,c(4,3),168,1)
```

```
#c(8,12) : numéros des colonnes du tableau associées aux stations sélectionnées. Ici deux stations sont considérées, Saint-Cyprien (8e colonne) et Pargaminières (12e colonne).
```

```
#1: pas horaire
```

```
#c(4,3): vecteur des nombres de grappes par strate
```

```
#168 : taille des grappes en nombre de pas (ici en heures)
```

```
#1 : les mois d'hiver sont regroupés avant de découper la série deux strates
```

```
> datesdeb<-tirage$deb
```

```
#dates de début de chaque semaine de mesure
```

```
> echant<-tirage$y
```

#échantillon constitué des mesures prises à Saint-Cyprien et des mesures concomitantes de la station Pargaminières. Les dates sont en 1<sup>ère</sup> colonne.

```
> write.table(echant,"C ://sondages/resultats/echant.txt", quote=F,col.names=F,row.names=F)
```

#On enregistre au format .txt le tableau des données d'échantillonnage.

## 5. Estimation

### 5.1. Estimation de la moyenne annuelle

```
> NO2Par02b<-grouphiver(NO2Par02,1)
```

# On groupe les mois d'hiver sur la série de référence

```
> NO2Cyp02ech<-echant[,2]
```

# Données d'échantillonnage au point sondé

```
> NO2Par02ech<-echant[,3]
```

#Données d'échantillonnage à la station de référence

```
> cor(NO2Cyp02ech, NO2Par02ech,use= "complete.obs")
```

#Calcul de la corrélation entre l'échantillon de données au point sondé et l'échantillon de données à la station de référence

Cette corrélation est de 0,815

```
> NO2Cyp02m<-estmoy(NO2Cyp02ech,NO2Par02b,NO2Par02ech,c(4,3),168)
```

#Estimation de la moyenne annuelle et de son intervalle de confiance par les quatre estimateurs

#c(4,3): vecteur des nombres de grappes par strate

#168 : taille des grappes en nombre de pas (ici en heures)

```
> print(NO2Cyp02m)
```

Le résultat suivant s'affiche :

```
[,1] [,2] [,3] [,4]
[1,] 43.013793 42.180094 41.797712 41.98133
[2,] 41.334100 41.318077 40.018009 40.69426
[3,] 44.693486 43.042110 43.577416 43.26839
[4,] 3.359386 1.724032 3.559407 2.57413
```

avec,

- de la ligne 1 à la ligne 4 : la moyenne estimée, la borne inférieure, la borne supérieure et la longueur de l'intervalle de confiance à 95% estimé ;
- de la colonne 1 à la colonne 4, l'estimateur PI et l'estimateur redressé par la régression, la différence, le quotient.

En supposant que l'intervalle de confiance à 95% est correctement estimé – on exclut l'estimateur redressé par la régression, peu performant lorsque le nombre de grappes est faible - la précision relative sur la moyenne est inférieure à 5%.

```
> NO2Cyp02mg<-estmoy(NO2Cyp02ech,NO2Par02b,NO2Par02ech,c(7),168)
```

```
#Les strates sont groupées préalablement à l'estimation.
```

```
> print(NO2Cyp02m)
```

Le résultat suivant s'affiche :

```
[,1] [,2] [,3] [,4]
[1,] 43.013793 42.274125 41.622733 41.928445
[2,] 40.557282 41.445071 40.008321 40.920845
[3,] 45.470304 43.103179 43.237146 42.936044
[4,] 4.913023 1.658108 3.228825 2.015199
```

L'intervalle de confiance pour l'estimateur PI est légèrement plus large mais il est normalement mieux estimé.

## 5.2. Estimation d'un taux de dépassement de seuil

### 5.2.1. 1<sup>er</sup> cas : le seuil est horaire

```
> NO2Cyp02d<- estprop(NO2Cyp02ech,NO2Par02b,NO2Par02ech,80,c(4,3),168)
```

```
#Estimation du taux de dépassement de seuil et de son intervalle de confiance par les quatre estimateurs
```

```
#100 : seuil en  $\mu\text{g}/\text{m}^3$ 
```

```
#c(4,3): vecteur des nombres de grappes par strate
```

```
#168 : taille des grappes en nombre de pas (ici en heures)
```

```
> print(NO2Cyp02d)
```

Le résultat suivant s'affiche :

```
[,1] [,2] [,3] [,4]
[1,] 0.6884682 0.7058489 1.476482 0.7391152
[2,] -0.3258330 0.2619991 0.338395 0.1439112
[3,] 1.7027693 1.1496987 2.614570 1.3343193
[4,] 2.0286022 0.8876996 2.276175 1.1904081
```

```
> NO2Cyp02dg<- estprop(NO2Cyp02ech,NO2Par02b,NO2Par02ech,100,c(7),168)
```

```
#Les strates sont groupées préalablement à l'estimation.
```

```
> print(NO2Cyp02dg)
```

Le résultat suivant s'affiche :

```
[1,] 0.6884682 0.78733752 1.3077358 0.79979185
[2,] -0.1582797 0.01889619 -0.1270705 0.02780639
```

```
[3,] 1.5352161 1.55577884 2.7425421 1.57177731
[4,] 1.6934958 1.53688265 2.8696126 1.54397092
```

### 5.2.2. 2<sup>e</sup> cas : le support temporel du seuil est de plusieurs heures (8h, 24h..., moyenne glissante ou non). Exemple d'un seuil journalier

```
> NO2Par02m<-calculmoy(date, NO2Par02,24,0)
# On transforme la série horaire en série journalière
> NO2Par02mb<-grouphiver(NO2Par02m,24)
# On groupe les mois d'hiver sur la série journalière de référence
> echantm<-calculmoyech(c(4,3),168, echant, 24,0)
# On transforme l'échantillon horaire en échantillon journalier
> NO2Cyp02echm<-echantm[,2]
# Données d'échantillonnage au point sondé
> NO2Par02echm<-echantm[,3]
#Données d'échantillonnage à la station de référence
```

```
> NO2Cyp02dj<- estprop(NO2Cyp02echm,NO2Par02mb,NO2Par02echm,50,c(4,3),7)
#Estimation du taux de dépassement de seuil et de son intervalle de confiance par les
quatre estimateurs
#50 : seuil en  $\mu\text{g}/\text{m}^3$ 
#c(4,3): vecteur des nombres de grappes par strate
#7 : taille des grappes en nombre de pas (ici en jours)
> print(NO2Cyp02dj)
```

Le résultat suivant s'affiche :

```
[,1] [,2] [,3] [,4]
[1,] 16.279070 16.564736 16.41727 16.564736
[2,] 5.011963 9.402095 -7.40305 1.621279
[3,] 27.546176 23.727376 40.23758 31.508193
[4,] 22.534213 14.325281 47.64063 29.886914
```

```
> NO2Cyp02djg<- estprop(NO2Cyp02echm,NO2Par02mb,NO2Par02echm,50,c(7),7)
#Les strates sont groupées préalablement à l'estimation.
> print(NO2Cyp02djg)
```

Le résultat suivant s'affiche :

```
[1,] 16.279070 15.807799 14.218346 15.763889
[2,] 6.322932 6.438503 -2.929661 4.574952
[3,] 26.235208 25.177095 31.366354 26.952826
[4,] 19.912276 18.738592 34.296015 22.377874
```

## 6. Contrôle de la qualité de l'estimation

Cette étape a pour but de contrôler la qualité de l'estimation de l'intervalle de confiance. Elle consiste

- 1) à simuler l'échantillonnage un très grand nombre de fois
  - sur la série complète de mesure du site d'échantillonnage, si cette série est en réalité connue (cas où l'on teste des stratégies sur des données de stations fixes) ;
  - sur la série complète d'une station fixe de la zone d'étude, si l'on est en conditions réelles
  
- 2) à comptabiliser les cas où l'intervalle de confiance estimé contient la vraie moyenne (ou le vrai taux de dépassement de seuil).

La station urbaine Mazades est ici sélectionnée.

	ST-CYPRIEN	MAZADES	Pargaminières
<b>MOYENNE (<math>\mu\text{g}/\text{M}^3</math>)</b>	42,8	25,4	53,7
<b>Variance (<math>\mu\text{g}/\text{m}^3</math>)<sup>2</sup></b>	285,7	340,7	648,7
<b>CORRELATION AVEC PARGAMINIÈRES</b>	0,78	0,65	

>simulscript(Tlse2002,"h","NO2",c(4,3),168,100,1,0,1,2000,c(8,12), C ://sondages/resultats")

>simulscript(Tlse2002,"h","NO2",c(4,3),168,50,24,0,1,2000,c(8,12), C ://sondages/resultats")

>simulscript(Tlse2002,"h","NO2",c(4,3),168,100,1,0,1,2000,c(10,12), C ://sondages/resultats")

>simulscript(Tlse2002,"h","NO2",c(4,3),168,50,24,0,1,2000,c(10,12), C ://sondages/resultats")

# Tlse2002: tableau des séries de mesure

# "h" : série horaire ("j" si série journalière)

# "NO2" : polluant

# c(4,3) : nombre de grappes par strate

# 168 : taille des grappes

# 100 ou 50 : seuil en  $\mu\text{g}/\text{m}^3$

# 1 ou 24 : support temporel du seuil en nombre d'heures

# 0 : le seuil n'est pas une moyenne glissante

# 1 : on découpe l'année en 2 strates (jan-mars+oct-dec: strate1, avril-sept: strate 2) (0 sinon)

# 2000: nombre de simulations

# c(8 ou 10,12): vecteur des numéros de colonnes correspondant aux stations étudiées, dans l'ordre la station supposée mobile ou la station test (St Cyprien ou Mazades) et celle de référence (Pargaminières)

#"C ://sondages/resultats": répertoire dans lequel sont écrits les tableaux de résultats

A l'issue de 2000 tirages, on obtient les résultats suivants :

*Tableau 22 - Pourcentage de tirages pour lesquels l'intervalle de confiance estimé (après regroupement des strates) contient effectivement la vraie moyenne, et longueur moyenne de l'intervalle de confiance*

	Saint-Cyprien			Mazades		
	Moyenne annuelle	Dépassement du seuil horaire de 100 µg/m3	Dépassement du seuil journalier de 50 µg/m3	Moyenne annuelle	Dépassement du seuil horaire de 100 µg/m3	Dépassement du seuil journalier de 50 µg/m3
PI	92,2%	67,7%	87,1%	97,8%	65,7%	81,0%
DIFF	95,3%	95,3%	88,8%	91,7%	93,9%	88,0%
QUOT	93,2%	66,9%	85,5%	96,0%	64,3%	79,1%
PI	8,4	0,8	32,5	11,9	0,6	13,6
DIFF	8,7	5,5	37,4	8,5	5,6	34,6
QUOT	7,2	0,7	29,4	8,4	0,5	12,9

Les résultats obtenus sur la station Mazades suggèrent :

- que pour le plan considéré, la moyenne et son incertitude sont assez bien estimées. Le taux de confiance atteint dans les simulations est d'au moins 91% quel que soit l'estimateur ;
- de rester prudent sur l'estimation du pourcentage de dépassement de seuil. L'intervalle de confiance est assez mal estimé (taux de couverture inférieur à 90% dans les simulations) ou alors il est très large.

Ces résultats sont confirmés par les simulations réalisées au site même de Saint-Cyprien.



## Annexe 3

### Résultats des simulations d'échantillonnage

#### Campagnes de mesure simulées

Les plans sont définis **pour l'estimation de moyennes**, pas pour celle des dépassements. Les dépassements sont donc estimés à partir des tirages effectués pour la moyenne.

Format des dates : aaaammjjhh

*Taille des grappes=168h ; Répartition=[6 5]* signifie 6 semaines de mesure tirées aléatoirement en hiver (jan-mars+oct-déc) et 5 semaines de mesures tirées aléatoirement en été

Pour l'ozone, toutes les grappes sont tirées en été (1<sup>er</sup> avril-30septembre).

Pour NO, le nombre de grappes à tirer en théorie est très élevé et irréalisable en pratique. Pour CO, il n'existe pas de valeur limite annuelle. Pour ces deux polluants, on utilise l'échantillonnage effectué pour le NO2.

Pour le NO2, on teste aussi le tirage de périodes de 2 semaines pour faire coïncider cela avec la durée d'exposition des tubes.

Dans certains cas, le nombre proposé de grappes est faible (<8). Un tirage comprenant 8 grappes est alors effectué en complément afin de voir si la variance, donc l'intervalle de confiance, sont mieux estimés. De la même façon, le nombre de semaines à tirer pour l'ozone est parfois élevé. On teste alors le tirage d'un nombre plus faible de grappes afin de voir l'incidence d'un échantillonnage plus réduit.

AASQA	Polluant	Année d'échantillonnage	Station de référence utilisée pour définir l'échantillonnage	Station supposée mobile	Taille des grappes (en heures)	Répartition des grappes	Dates de début des grappes
ORAMIP	NO2	2002	Pargaminières	Saint-Cyprien	168	6 5	2002010118 2002020402 2002021105 2002031415 2002042605 2002050413 2002051408 2002080506 2002092317 2002102813 2002112923
	NO2	2002	Pargaminières	Saint-Cyprien	336	5 4	2002011605 2002020323 2002030618 2002042923 2002052305 2002070211 2002080509 2002110612 2002112213
	NO2	2002	Saint-Cyprien	Pargaminières	168	3 5	[2002020405 2002031501 2002051113 2002060111 2002062005 2002071520 2002091320 2002100118
	NO2	2002	Saint-Cyprien	Pargaminières	336	2 4	2002031207 2002041101 2002051821 2002072013 2002082620 2002111214
						3 5	2002011321 2002022208 2002040307 2002052215 2002061703 2002070608 2002081421 2002121723

	NO, CO	2002	Pargaminières	Saint-Cyprien			Idem NO2
	NO, CO	2002	Saint-Cyprien	Pargaminières			Idem NO2
	O3, été	2002	Balma	Colomiers	168	7	2002042502 2002062714 2002071916 2002073103 2002081500 2002082517 200209102
	O3, été	2002	Colomiers	Balma	168	9	2002042315 2002050115 2002051716 2002052422 2002060819 2002071621 2002080809 2002081812 2002092301
	PM10	2002	CCIT	Berthelot	168	6 3	2002020510 2002022202 2002061216 2002062305 2002081608 2002100918 2002102916 2002111319 2002112812
	PM10	2002	Berthelot	CCIT	168	6 3	[2002010203 2002012916 2002021821 2002060915 2002082721 2002091213 2002111415 2002121023 2002122113
	PM2.5	2002	CCIT	Berthelot	168	6 3	idem PM10 CCIT
	PM2.5	2002	Berthelot	CCIT	168	6 3	Idem PM10 Berthelot
ASPA	NO2	2004	Stbg-Centre	Stbg-Nord	168	8 6	2004012409 2004020205 2004022304 2004030917 2004042803 2004050807  2004061412 2004071503 2004081315 2004092001 2004100121 2004101212  2004103114 2004120813
	NO2	2004	Stbg-Centre	Stbg-Nord	336	3 2  6 2	2004010605 2004020309 2004022712 2004071616 2004083111 2004112220  2004011312 2004021911 2004031106 2004052820 2004091308 2004103017 2004112921 2004121417
	NO2	2004	Stbg-Nord	Stbg-Centre	168	6 4	2004010523 2004011302 2004032108 2004062108 2004080604 2004082320  2004092214 2004101408 2004112820 2004121418
	NO2	2004	Stbg-Nord	Stbg-Centre	336	4 2  6 2	2004012622 2004081902 2004090419 2004102022 2004111322 2004120600  2004021218 2004030718 2004050100 2004081322 2004100623 2004102117 2004110709 2004112903
	O3, été	2004	Stbg-Centre	Stbg-Nord	168	16	2004041401 2004042316 2004050215 2004051003 2004051812 2004052803  2004061118 2004062009 2004070213 2004071017 2004072513 2004080310  2004081402 2004082507 2004090619 2004091507

	O3, été	2004	Stbg-Nord	Stbg-Centre	168	14	2004040707 2004041902 2004050309 2004051210 2004052616 2004060503  2004061606 2004062909 2004070908 2004071623 2004072822 2004081113  2004081910 2004090719
	O3, été	2004	Les 2 stations		168	12	2004040415 2004041314 2004050700 2004052203 2004062722 2004070509  2004072118 2004081000 2004081822 2004082702 2004090510 2004092108
	O3, été	2004	Les 2 stations		168	8	2004041212 2004051708 2004061602 2004062913 2004071313 2004080204 2004081111 2004082805
	PM10	2004	Stbg-Est	Stbg-Clémenceau	168	9 6	2004012312 2004021203 2004030711 2004032010 2004041502 2004050604  2004060908 2004062805 2004072609 2004091103 2004100209 2004102301 2004110418 2004112006 2004121703
	PM10	2004	Stbg-Est	Stbg-Clémenceau	336	6 3	2004010911 2004012900 2004031019 2004061007 2004072710 2004082820 2004101501 2004112521 2004121419
	PM10	2004	Stbg-Clémenceau	Stbg-Est	168	9 4	2004010823 2004020322 2004022522 2004060311 2004070607 2004072610  2004081112 2004100617 2004102207 2004110605 2004112105 2004120108 2004121410
	PM10	2004	Stbg-Clémenceau	Stbg-Est	336	6 2	2004010907 2004020422 2004030423 2004041100 2004072523 2004101004 2004112014 2004121406
	PM2.5	2004	Stbg-Est	Stbg-Clémenceau	168	8 3	2004010308 2004011204 2004012217 2004021108 2004022407 2004042219  2004090201 2004092208 2004100216 2004111807 2004120608
	PM2.5	2004	Stbg-Est	Stbg-Clémenceau	336	6 2	2004012412 2004021613 2004030209 2004043016 2004063002 2004100715 2004111306 2004121705
	PM2.5	2004	Stbg-Clémenceau	Stbg-Est	168	10 4	2004011704 2004020103 2004021419 2004022713 2004031000 2004031700  2004053023 2004062923 2004070712 2004090815 2004100407 2004102105 2004110100 2004121308

	PM2.5	2004	Stbg-Clémenceau	Stbg-Est	336	6 2	2004011910 2004020708 2004022414 2004052720 2004061122 2004100817 2004102918 2004120120
OPALAIR	NO2	2003	Petite-Synthe	Fort-Mardick	168	4 4	2003010119 2003032404 2003042403 2003070206 2003081517 2003082309 2003110122 2003111223
	NO2	2003	Petite-Synthe	Fort-Mardick	336	5 2  4 4	2003010723 2003052219 2003083016 2003103005 2003120821  2003012412 2003022510 2003040418 2003050421 2003070621 2003082004 2003110613 2003120510
	NO2	2003	Fort-Mardick	Petite-Synthe	168	5 3	2003010202 2003030201 2003032006 2003041211 2003070115 2003090205 2003110211 2003122106
	NO2	2003	Fort-Mardick	Petite-Synthe	336	6 2  5 3	2003012307 2003031202 2003060115 2003071804 2003102115 2003110515  2003010323 2003012209 2003021806 2003030521 2003050603 2003073015 2003091522 2003102722
	SO2	2003	Saint-Pol Nord	Fort-Mardick	168	2 3  3 5	2003021719 2003061303 2003071800 2003091710 2003121602  2003020914 2003032308 2003071910 2003072721 2003081009 2003081907 2003090817 2003120603
	SO2	2003	Fort-Mardick	Saint-Pol Nord	168	3 3  4 4	2003012612 2003050204 2003062223 2003071116 2003100607 2003101618 2003122201  2003040203 2003041418 2003042803 2003071816 2003101311 2003111219 2003120415 2003121918
AIR NORMAN D	SO2	2003	Herriot	Mare-Rouge	168	7 5	2003011016 2003020122 2003021711 2003031100 2003042710 2003052117  2003062500 2003080703 2003092107 2003101620 2003110710 2003120306

	SO2	2003	Mare-Rouge	Herriot	168	16 5	2003010712 2003011723 2003012709 2003020707 2003022017 2003031005  2003032104 2003042911 2003051902 2003062021 2003070102 2003081116  2003100116 2003101321 2003102120 2003110503 2003111218 2003112103  2003120120 2003120901 2003122117  2003010215 2003020623 2003022004 2003040818 2003042217 2003082219  2003100318 2003101511 2003110500 2003111723 2003120409 2003122101
	O3	2003	Herriot	Mare-Rouge	168	5	2003041314 2003050917 2003070606 2003072605 2003080903
	O3	2003	Mare-Rouge	Herriot	168	6	2003041011 2003051622 2003062014 2003072416 2003082311 2003090613
	O3	2003	Les 2 stations		168	8	2003043013 2003052409 2003062312 2003070501 2003080309 2003081321 2003083009 2003091914
	PM10	2003	Herriot	Mare-Rouge	168	3 4	2003011122 2003021915 2003050508 2003080823 2003090406 2003092213 2003111823
	PM10	2003	Mare-Rouge	Herriot	168	4 3	2003011223 2003021700 2003022800 2003060520 2003061418 2003080414 2003100812
	PM10	2003	Les 2 stations		168	4 4	2003041502 2003061207 2003080208 2003083005 2003101000 2003111220 2003120503 2003122106
AIR APS Annecy	NO2	2004	Novel	Loverchy	168	6 2	2004020816 2004022208 2004030300 2004051709 2004060210 2004101219 2004110818 2004112118
	NO2	2004	Novel	Loverchy	336	3 2  6 2	2004011016 2004042302 2004052002 2004100722 2004103000  2004012115 2004022217 2004030916 2004061004 2004081601 2004101409 2004110409 2004121007
	NO2	2004	Loverchy	Novel	168	9 5	2004010715 2004011611 2004020616 2004021821 2004051405 2004061205  2004062914 2004071208 2004080605 2004100113 2004101515 2004111422  2004120117 2004121611

	NO2	2004	Loverchy	Novel	336	7 2	2004011303 2004020317 2004022200 2004050906 2004081303 2004101220 2004110704  2004112114 2004121404
	O3	2004	Novel	Loverchy	168	14	2004040119 2004041016 2004042707 2004050518 2004051307 2004053018  2004061323 2004062502 2004070805 2004072515 2004081103 2004082303  2004083103 2004091217
	O3	2004	Loverchy	Novel	168	13	2004040301 2004041300 2004042008 2004050203 2004052315 2004060115  2004061509 2004071303 2004072207 2004080216 2004082509 2004090320  2004092104
	PM10	2004	Novel	Loverchy	168	6 4	2004013022 2004021122 2004030923 2004052118 2004060306 2004072904  2004080602 2004101119 2004101905 2004103113
	PM10	2004	Loverchy	Novel	168	7 4	2004012509 2004021700 2004022707 2004031021 2004032306 2004050623  2004071222 2004083008 2004092001 2004100522 2004121810
AIR APS Annemasse	NO2	2004	Centre	Gaillard	168	8 4	2004010908 2004012208 2004021921 2004030203 2004040522 2004041807  2004062802 2004082719 2004101813 2004102801 2004110620 2004121518
	NO2	2004	Centre	Gaillard	336	6 2	2004012302 2004020703 2004022204 2004050120 2004082803 2004101815 2004111807 2004121505
	NO2	2004	Gaillard	Centre	168	8 3	2004011707 2004020616 2004021508 2004031221 2004032412 2004070310  2004080504 2004091603 2004100209 2004110319 2004120316
	NO2	2004	Gaillard	Centre	336	6 2	2004011912 2004020723 2004030608 2004041323 2004051918 2004100218 2004110316 2004112323
	O3	2004	Centre	Gaillard	168	16	2004040317 2004041208 2004042820 2004050616 2004051908 2004060120  2004061200 2004062016 2004062823 2004070602 2004071917 2004073019  2004080823 2004082120 2004090906 2004091911

	O3	2004	Gaillard	Centre	168	14	2004040320 2004041200 2004042506 2004051609 2004052620 2004061623  2004063004 2004070709 2004071410 2004072201 2004080405 2004082721  2004090809 2004092223
	PM10	2004	Centre	Gaillard	168	7 4	2004012221 2004021920 2004022911 2004031805 2004052615 2004062912  2004071216 2004081109 2004112023 2004120206 2004122407
	PM10	2004	Gaillard	Centre	168	7 4	2004012421 2004020510 2004030621 2004032213 2004061814 2004072801  2004081404 2004082300 2004110410 2004112616 2004122106
AIR APS Chambéry	NO2	2004	Le Haut	Pasteur	168	8 4	2004010407 2004020207 2004021800 2004022605 2004040314 2004052800  2004061917 2004070503 2004100614 2004101619 2004120323 2004121604
	NO2	2004	Le Haut	Pasteur	336	5 3	2004010515 2004013007 2004031518 2004042320 2004062421 2004091516 2004103106 2004112104
	NO2	2004	Pasteur	Le Haut	168	6 5	2004011512 2004021223 2004040404 2004042316 2004060109 2004062922  2004091011 2004101113 2004103016 2004120517 2004121822
	NO2	2004	Pasteur	Le Haut	336	5 3	2004012412 2004022906 2004042300 2004071922 2004090309 2004100712 2004111906 2004120610
	O3	2004	Le Haut	Pasteur	168	16	2004040819 2004050109 2004051320 2004052817 2004061008 2004061808  2004062912 2004070806 2004072403 2004073105 2004081020 2004081816  2004082723 2004090504 2004091303 2004092123
	O3	2004	Pasteur	Le Haut	168	16	2004040315 2004041514 2004042503 2004051511 2004052314 2004060112  2004061701 2004062614 2004071001 2004072113 2004073116 2004080722  2004081717 2004083012 2004091214 2004092013

	PM10	2004	Le Haut	Pasteur	168	5 3	2004011504 2004030301 2004031021 2004052321 2004072618 2004091102 2004110212 004112503
	PM10	2004	Pasteur	Le Haut	168	6 3	2004011605 2004012722 2004041517 2004050417 2004071102 2004102814 2004111917 2004120202 2004122223
	O3	2004	Les 6 stations O3 d'AIR APS		168	8	2004040713 2004041607 2004051910 2004060616 2004070411 2004080814 2004082323 2004090720



## Annexe 4

### Exemples de résultats numériques obtenus par la méthode des plans de sondage

				Vraie moyenne	Corrélation entre séries annuelles complètes	ng	nh_hiv	nh_été	Corrélation entre échantillons	Estimateur	Estimation sans regrouper les strates					Estimation en regroupant les strates					Taux de couverture obtenu par simulations*	
											Moyenne annuelle estimée	IC95- IC95+	LIC95	Ecart relatif (%) entre moyennes estimée et mesurée	Taux de couverture obtenu par simulations*	Moyenne annuelle estimée	IC95- IC95+	LIC95	Ecart relatif (%) entre moyennes estimée et mesurée	Taux de couverture obtenu par simulations*		
ORAMP 2002	NO2	Saint-Cyprien	Pargaminières	42.77	0.78	168	6	5	0.8	PI	41.559	38.156	44.963	6.808	-2.83	93.9	41.559	38.341	44.778	6.437	-2.83	94.6
										DIFF	44.417	41.930	46.903	4.973	3.85	94.1	44.223	41.567	46.879	5.312	3.40	96.0
										QUOT	43.871	41.786	45.955	4.169	2.57	95.1	43.727	41.593	45.861	4.268	2.24	96.4
	NO2	Saint-Cyprien	Pargaminières	42.77	0.78	336	5	4	0.77	PI	42.888	39.578	46.198	6.620	0.28	96.5	42.888	39.274	46.503	7.228	0.28	96.9
										DIFF	43.247	40.572	45.922	5.350	1.12	95.9	42.913	39.469	46.358	6.889	0.34	97.7
										QUOT	43.226	40.758	45.694	4.936	1.07	95.7	42.908	40.046	45.771	5.725	0.32	96.6
	NO2	Pargaminières	Saint-Cyprien	53.74	0.78	168	3	5	0.78	PI	53.700	49.426	57.974	8.548	-0.07	90.0	53.700	49.081	58.320	9.239	-0.07	95.9
										DIFF	53.589	50.466	56.713	6.247	-0.28	89.2	53.031	49.730	56.332	6.602	-1.32	92.3
										QUOT	53.309	50.031	56.586	6.555	-0.80	87.6	52.873	49.716	56.030	6.313	-1.61	89.6
	NO2	Pargaminières	Saint-Cyprien	53.74	0.78	336	2	4	0.77	PI	55.348	46.707	63.989	17.282	2.99	87.3	55.348	48.890	61.806	12.916	2.99	96.9
										DIFF	55.722	49.001	62.444	13.444	3.69	77.1	54.689	50.417	58.960	8.543	1.77	85.8
										QUOT	55.683	49.476	61.890	12.413	3.62	75.9	54.508	50.477	58.539	8.062	1.43	82.1
	NO2	Pargaminières	Saint-Cyprien	53.74	0.78	336	3	5	0.81	PI	54.412	51.442	57.382	5.940	1.25	90.0	54.412	50.983	57.841	6.858	1.25	98.4
										DIFF	53.166	51.480	54.852	3.372	-1.07	87.0	52.908	51.208	54.609	3.401	-1.55	88.8
										QUOT	52.622	50.523	54.720	4.197	-2.08	86.0	52.564	50.769	54.359	3.590	-2.19	85.6
	NO	Saint-Cyprien	Pargaminières	60.01	0.85	168	6	5	0.86	PI	60.172	51.329	69.015	17.686	0.27	86.6	60.172	46.692	73.652	26.959	0.27	99.9
										DIFF	60.431	55.688	65.175	9.486	0.70	93.4	60.509	56.105	64.914	8.809	0.83	95.7
										QUOT	61.153	56.456	65.851	9.395	1.90	94.5	60.568	55.735	65.400	9.665	0.93	99.0
	NO	Saint-Cyprien	Pargaminières	60.01	0.85	336	5	4	0.86	PI	65.052	56.298	73.806	17.508	8.40	86.3	65.052	49.506	80.598	31.092	8.40	98.7
										DIFF	62.477	59.672	65.282	5.610	4.11	94.3	62.900	60.111	65.688	5.577	4.82	93.9
										QUOT	64.360	61.436	67.284	5.848	7.25	95.3	62.547	59.053	66.041	6.987	4.23	97.6
	NO	Pargaminières	Saint-Cyprien	51.60	0.85	168	3	5	0.86	PI	51.266	44.950	57.581	12.63	-0.65	85.3	41.998	32.414	51.582	19.168	-18.61	91.3
										DIFF	51.025	46.282	55.769	9.49	-1.11	86.7	50.563	47.025	54.101	7.076	-2.01	88.9
										QUOT	50.507	46.767	54.248	7.48	-2.12	88.1	48.990	46.201	51.780	5.580	-5.06	90.0
	NO	Pargaminières	Saint-Cyprien	51.60	0.85	336	2	4	0.82	PI	46.360	30.613	62.107	31.494	-10.15	76.1	46.360	33.850	58.871	25.021	-10.15	89.3
										DIFF	51.515	47.868	55.162	7.294	-0.16	78.4	50.721	47.326	54.115	6.789	-1.70	83.7
										QUOT	51.821	47.152	56.491	9.339	0.43	78.1	49.993	45.201	54.785	9.584	-3.11	83.6
	NO	Pargaminières	Saint-Cyprien	51.60	0.85	336	3	5	0.86	PI	51.552	43.248	59.855	16.606	-0.09	86.2	51.552	37.282	65.821	28.538	-0.09	95.0
										DIFF	52.130	43.904	60.357	16.453	1.03	88.0	51.734	46.393	57.075	10.683	0.26	87.3
										QUOT	52.477	45.207	59.746	14.539	1.70	89.6	51.709	46.744	56.674	9.929	0.21	90.6
	CO	Saint-Cyprien	Pargaminières	1.14	0.76	168	6	5	0.77	PI	1.129	1.013	1.245	0.232	-0.94		1.129	0.951	1.308	0.358	-0.94	
										DIFF	1.165	1.070	1.259	0.189	2.18		1.162	1.063	1.260	0.196	1.90	
										QUOT	1.152	1.074	1.230	0.156	1.02		1.155	1.077	1.232	0.155	1.27	
	CO	Saint-Cyprien	Pargaminières	1.14	0.76	336	5	4	0.78	PI	1.210	1.101	1.319	0.218	6.10		1.210	1.003	1.417	0.414	6.10	
										DIFF	1.178	1.126	1.230	0.104	3.34		1.177	1.112	1.242	0.130	3.26	
										QUOT	1.176	1.141	1.212	0.071	3.18		1.184	1.153	1.214	0.061	3.82	
	CO	Pargaminières	Saint-Cyprien	1.48	0.76	168	3	5	0.78	PI	1.386	1.280	1.492	0.212	-6.35		1.386	1.169	1.603	0.433	-6.35	
										DIFF	1.523	1.483	1.564	0.081	2.93		1.497	1.425	1.568	0.144	1.13	
										QUOT	1.547	1.485	1.608	0.123	4.51		1.535	1.476	1.594	0.118	3.74	
	CO	Pargaminières	Saint-Cyprien	1.48	0.76	336	2	4	0.73	PI	1.387	1.159	1.615	0.456	-6.29		1.387	1.179	1.595	0.417	-6.29	
										DIFF	1.452	1.415	1.489	0.074	-1.89		1.429	1.359	1.500	0.142	-3.42	
										QUOT	1.455	1.402	1.507	0.106	-1.72		1.441	1.381	1.500	0.119	-2.65	
	CO	Pargaminières	Saint-Cyprien	1.48	0.76	336	3	5	0.78	PI	1.531	1.387	1.675	0.288	3.46		1.531	1.256	1.807	0.551	3.46	
										DIFF	1.539	1.431	1.647	0.216	4.01		1.503	1.384	1.622	0.238	1.54	
										QUOT	1.507	1.323	1.690	0.367	1.81		1.494	1.376	1.612	0.235	0.95	

Corrélation entre séries annuelles complètes	ng	nh_hiv	nh_été	Corrélation entre échantillons	Estimateur	Estimation sans regrouper les strates				Ecart relatif (%) entre moyennes estimée et mesurée	Taux de couverture obtenu par simulations*	Estimation en regroupant les strates				Ecart relatif (%) entre moyennes estimée et mesurée	Taux de couverture obtenu par simulations*
						Moyenne annuelle estimée	IC95-	IC95+	LIC95			Moyenne annuelle estimée	IC95-	IC95+	LIC95		
						0.90	168		7			0.91	PI				
					DIFF					70.543	69.639	71.446	1.807	-58.01	89.5		
					QUOT					70.320	68.862	71.778	2.916	-58.14	90.0		
0.90	168		9	0.91	PI					59.082	53.267	64.896	11.630	-64.83	91.7		
					DIFF					58.073	56.168	59.977	3.809	-65.43	91.6		
					QUOT					58.248	56.224	60.272	4.048	-65.33	91.9		
0.92	168	6	3	0.89	PI	20.937	18.236	23.638	5.401	-87.54	90.6	20.937	18.163	23.711	5.547	-87.54	92.6
					DIFF	19.954	18.772	21.137	2.365	-88.12	90.3	20.230	19.176	21.283	2.107	-87.96	90.4
					QUOT	20.226	19.205	21.247	2.041	-87.96	88.3	20.318	19.345	21.291	1.945	-87.91	90.4
0.92	168	6	3	0.88	PI	23.437	21.649	25.225	3.576	-86.05	92.5	23.437	21.641	25.234	3.593	-86.05	93.0
					DIFF	21.630	20.230	23.030	2.800	-87.13	89.7	21.588	20.282	22.893	2.611	-87.15	90.7
					QUOT	21.485	20.045	22.925	2.880	-87.21	88.9	21.484	20.101	22.868	2.767	-87.21	91.2
0.84	168	6	3	0.89	PI	13.195	11.389	15.001	3.612	-92.15	90.8	13.195	11.352	15.038	3.686	-92.15	93.4
					DIFF	Données manquantes à la station CCIT				88.8						92.1	
					QUOT					89.3						93.3	
0.84	168	6	3	0.94	PI	16.283	15.144	17.423	2.279	-90.31	90.9	16.283	14.808	17.758	2.950	-90.31	93.2
					DIFF	14.247	13.672	14.823	1.150	-91.52	89.1	14.218	13.683	14.753	1.070	-91.54	91.7
					QUOT	14.054	13.427	14.680	1.252	-91.63	88.0	14.038	13.420	14.657	1.237	-91.64	92.8

Vrai taux de dépassement (%)	ng	nh_hiv	nh_été	cor	Estimateur	Estimation sans regrouper les strates				Ecart relatif (%) entre taux estimés et mesurés	Taux de couverture obtenu par simulations*	Estimation en regroupant les strates				Ecart relatif (%) entre taux estimés et mesurés	Taux de couverture obtenu par simulations*
						Taux de dépassement de seuil	IC95-	IC95+	LIC95			Taux de dépassement de seuil	IC95-	IC95+	LIC95		
						3.04	168	6	5			0.57	PI	3.288	1.415		
					DIFF	2.763	2.127	3.399	1.271	-9.11	90.5	2.810	2.134	3.486	1.352	-7.57	93.5
					QUOT	2.790	2.227	3.353	1.126	-8.23	74.8	2.797	2.143	3.450	1.307	-8.01	87.0
3.04	336	5	4	0.64	PI	4.187	2.013	6.360	4.347	37.71	90.5	4.187	1.110	7.263	6.154	37.71	97.6
					DIFF	3.350	2.648	4.052	1.404	10.19	93.3	3.494	2.662	4.326	1.664	14.93	96.3
					QUOT	3.246	2.793	3.700	0.907	6.78	71.2	3.337	2.862	3.812	0.951	9.76	85.2
2.72	168	3	5	0.62	PI	1.342	0.426	2.258	1.832	-50.65	73.3	1.342	0.112	2.573	2.462	-50.65	75.4
					DIFF	2.435	2.071	2.800	0.729	-10.46	84.6	2.497	2.098	2.896	0.798	-8.20	88.7
					QUOT	2.263	1.852	2.674	0.822	-16.81	73.8	2.165	1.861	2.469	0.607	-20.41	66.6
2.72	336	2	4	0.55	PI	1.553	0.861	2.246	1.385	-42.90	66.4	1.553	0.453	2.653	2.201	-42.90	71.7
					DIFF	2.609	1.903	3.315	1.412	-4.07	84.9	2.879	2.248	3.510	1.262	5.83	82.6
					QUOT	2.494	2.118	2.870	0.752	-8.31	61.6	2.755	2.271	3.239	0.968	1.29	52.4
2.72	336	3	5	0.68	PI	2.820	0.983	4.656	3.673	3.66	74.3	2.820	0.441	5.198	4.757	3.66	79.9
					DIFF	2.511	0.548	4.475	3.927	-7.67	88.6	2.649	1.386	3.912	2.526	-2.60	86.1
					QUOT	2.730	1.230	4.231	3.001	0.38	73.0	2.670	1.717	3.623	1.906	-1.84	63.5

## Annexe 5

### Présentation des stations de mesure

On applique la méthode aux données de plusieurs réseaux dont on dispose des données sur deux années consécutives.

#### Le réseau de Toulouse - ORAMIP

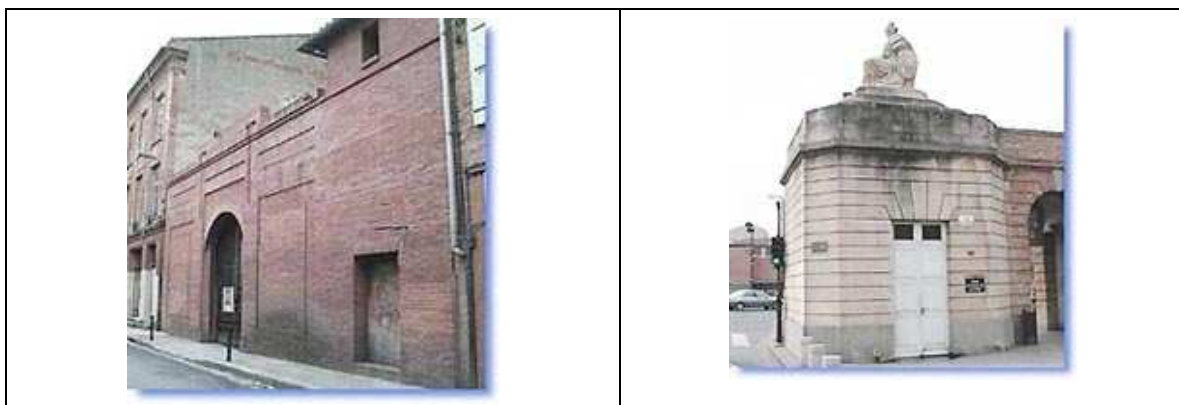
L'association ORAMIP a communiqué à l'EMD les données horaires de 2001 et 2002 pour plusieurs stations de typologies différentes, ainsi que des mesures météorologiques horaires effectuées à la station de Colomiers.

Cette étude s'applique aux stations décrites dans le Tableau 23 et dans la Figure 20.

Tableau 23 : stations et polluants mesurés

Stations de mesure			Polluants mesurés			
Nom	Dim.	Typologie	NOx	O3	CO	PM <sub>2.5</sub> et PM <sub>10</sub>
St cyprien	CYP	TRAFIC	X		X	
Pargaminière	PAR	TRAFIC	X		X	
Colomiers	COL	PERIURBAIN		X		
Balma	BAL	PERIURBAIN		X		
CCIT	CCI	TRAFIC				X
Berthelot	BRT	URBAIN				X

La station météorologique de Colomiers a fourni des données qui ont été appliquées à l'ensemble des stations de mesure ; elles ne sont peut-être pas représentatives de toute la zone d'étude.









Pargaminière	St Cyprien
	
Balma	Colomiers
	
CCIT	Berthelot

Figure 20 : stations de mesure – réseau de Toulouse

### Le réseau de Strasbourg - ASPA

Les stations de Strasbourg prises en compte sont présentées dans la Figure 21 et le Tableau 24.

	
Strasbourg Nord	Strasbourg Centre

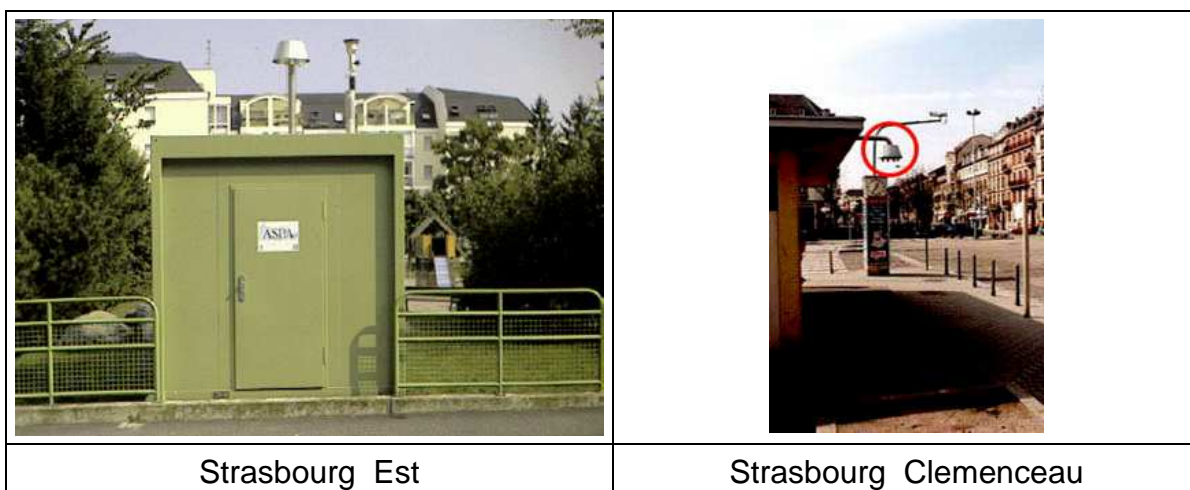


Figure 21: stations de mesure – réseau de Strasbourg

L'association ASPA a communiqué à l'EMD les données horaires des années 2003 et 2004 pour plusieurs stations de typologies différentes, ainsi que des mesures météorologiques horaires effectuées dans plusieurs stations.

Tableau 24: stations et polluants pris en compte

Stations de mesure		Polluants mesurés				
Nom	Typologie	NO	NO <sub>2</sub>	O <sub>3</sub>	PM <sub>10</sub>	PM <sub>2.5</sub>
Strasbourg Centre	Urbaine	X	X	X		
Strasbourg Nord	Urbaine	X	X	X		
Strasbourg Clemenceau	Trafic				X	X
Strasbourg Est	Urbaine				X	X

La station météorologique CRR a été retenue car la base de données était la plus complète ; elle se situe en dehors de l'agglomération et est peut-être la plus représentative pour l'ensemble de la zone d'étude.

#### Le réseau du Havre – ALPA

Le Havre – Mare rouge	Le Havre - Hérriot

Tableau 25: stations et polluants pris en compte

Stations de mesure		Polluants mesurés		
Nom	Typologie	SO <sub>2</sub>	O <sub>3</sub>	PM <sub>10</sub>

Le Havre – Mare rouge	Industriel	X	X	X
Le Havre - Hérriot	Urbain	X	X	X

### Le réseau de Annecy – AIR-APS

	
Annecy - Loverchy	Annecy - Novel

Tableau 26: stations et polluants pris en compte

Stations de mesure		Polluants mesurés			
Nom	Typologie	NO	NO <sub>2</sub>	O <sub>3</sub>	PM <sub>10</sub>
Annecy - Loverchy	Urbain	X	X	X	X
Annecy - Novel	Urbain	X	X	X	X

## Le réseau de Annemasse - AIR-APS

	
Annemasse	Annemasse - Gaillard

Tableau 27: stations et polluants pris en compte

Stations de mesure		Polluants mesurés			
Nom	Typologie	NO	NO <sub>2</sub>	O <sub>3</sub>	PM <sub>10</sub>
Annemasse	Urbain	X	X	X	X
Annemasse - Gaillard	Urbain	X	X	X	X

## Le réseau de Chambéry - AIR-APS



	
Chambéry - Pasteur	Chambéry - Le Haut

Tableau 28: stations et polluants pris en compte

Stations de mesure		Polluants mesurés			
Nom	Typologie	NO	NO <sub>2</sub>	O <sub>3</sub>	PM <sub>10</sub>
Chambéry - Pasteur	Urbain	X	X	X	X
Chambéry – Le Haut	Urbain	X	X	X	X

### Le réseau de Dunkerque – OPALAIR



	
Fort-Mardyck	St Pol - Nord

Tableau 29: stations et polluants pris en compte

Stations de mesure		Polluants mesurés		
Nom	Typologie	SO <sub>2</sub>	NO	NO <sub>2</sub>
Fort-Mardyck	Industriel	X	X	X
St Pol - Nord	Urbain	X	X	X



## Annexe 6

### Méthode de calcul ISO 9359 – extrait du rapport EMD 2004 [2]

Le classeur EXCEL est constitué des feuilles de calcul suivantes assorties de leurs modes d'emploi.

#### 6.1 – Feuille « stratification » (extraits partiels)

##### 6.1.1 – affectation, à chaque paramètre du N° de classe (1 ou 2 selon son degré d'influence), pour chaque séquence de mesure (ici exemple appliqué à SO<sub>2</sub>)

Cette feuille contient la base tri-horaire des paramètres pris en compte, vitesse du vent VV, température T, séquence 3heures dans la journée (1 à 8) et jour de la semaine (1 à 7). En fonction des choix de classe (1 ou 2) de chaque paramètre, le tableau 6-1 relatif à SO<sub>2</sub> dans cet exemple se remplit automatiquement des valeurs 1 ou 2.

**Tableau 6-1**

Date	VV-3h	T-3h	heures	N°jours		classe VV pour SO <sub>2</sub>	classe T pour SO <sub>2</sub>	classe heure pour SO <sub>2</sub>	classe Jour pour SO <sub>2</sub>
01/01/03 03:00	1,32	1,6	1	3		<u>2</u>	<u>2</u>	<u>1</u>	<u>2</u>
01/01/03 06:00	0,32	1,9	2	3		<u>2</u>	<u>2</u>	<u>1</u>	<u>2</u>
01/01/03 09:00	1,04	2,66	3	3		<u>2</u>	<u>2</u>	<u>2</u>	<u>2</u>
01/01/03 12:00	1,86	5,46	4	3		<u>1</u>	<u>2</u>	<u>2</u>	<u>2</u>
01/01/03 15:00	2,51	9,13	5	3		<u>1</u>	<u>2</u>	<u>1</u>	<u>2</u>
01/01/03 18:00	3,36	9,16	6	3		<u>1</u>	<u>2</u>	<u>1</u>	<u>2</u>
01/01/03 21:00	4,04	9,96	7	3		<u>1</u>	<u>2</u>	<u>2</u>	<u>2</u>
02/01/03 00:00	2,93	12,06	8	3		<u>1</u>	<u>1</u>	<u>1</u>	<u>2</u>
02/01/03 03:00	2,64	12,93	1	4		<u>1</u>	<u>1</u>	<u>1</u>	<u>2</u>
02/01/03 06:00	2,41	13,3	2	4		<u>1</u>	<u>1</u>	<u>1</u>	<u>2</u>

### 6.1.2 – Affectation d'un N° de strate (1 à 3) à chaque séquence et pour chaque polluant

Tableau 6-2

Date	classe VV pour SO2	classe T pour SO2	classe heure pour SO2	classe Jour pour SO2	strate SO2	strate NO	strate NO2	strate O3	strate PM
01/01/03 03:00	2	2	1	2	3	3	3	1	3
01/01/03 06:00	2	2	1	2	3	3	3	1	3
01/01/03 09:00	2	2	2	2	3	3	3	1	3
01/01/03 12:00	1	2	2	2	3	2	2	1	1
01/01/03 15:00	1	2	1	2	2	2	2	2	1
01/01/03 18:00	1	2	1	2	2	3	3	2	1
01/01/03 21:00	1	2	2	2	3	3	3	1	2
02/01/03 00:00	1	1	1	2	1	1	1	1	2
02/01/03 03:00	1	1	1	2	1	1	1	1	1
02/01/03 06:00	1	1	1	2	1	1	1	1	1

A partir du tableau 6-1, chaque cellule des colonnes « strate SO<sub>2</sub> à strate PM » du tableau 6-2 se remplit automatiquement du numéro de la strate selon le nombre de paramètres en classe 2.

### 6.1.3 – Nombres de données par strate pour les 5 polluants

A partir du tableau 6-2, le tableau 6-3 dénombre les séquences tri-horaires pour strate et pour chaque polluant. Ces valeurs sont indispensables pour les reconstitutions de moyennes et de nombres de dépassements de seuils prédéfinis; ils sont automatiquement transmis sur les feuilles d'exploitation des campagnes.

Tableau 6-3

N° de la strate	nb données strate SO2	nb données strate NO	nb données strate NO2	nb données strate O3	nb données strate PM
1	742	729	729	1931	466
2	1146	1192	1192	687	1222
3	1032	999	999	302	1232

## 6.2 – Feuille « Calculs » (extrait partiel)

Tableau 6-4

STRATES		Séquences utilisées	SEUILS (P90)	15,6	56,3	65	99,6	48	29
strate PM		Cp01	DATES	SO2 Stg Est	NO STG Est	NO2 STG Est	O3 STG Est	PM10 STG Est	PM2,5 STG Est
3		1	01/01/03 03:00	5	3,3	22	18,6	29	26,6
3		1	01/01/03 06:00	5,3	6,3	32,3	7	19,6	19
3		1	01/01/03 09:00	7,3	4,3	24,6	8,3	21,6	20,3
1			01/01/03 12:00						
1			01/01/03 15:00						
1			01/01/03 18:00						
2			01/01/03 21:00						
2		1	02/01/03 00:00	2,5	1	8,6	60,5	5,6	5
1		1	02/01/03 03:00	3	0	6,3	64,3	3,6	2,6
1		1	02/01/03 06:00	4	0,6	9,6	60	2	1,3
2		1	02/01/03 09:00	4,3	3,3	19,6	43	3,6	3
1		1	02/01/03 12:00	5	5,3	21,6	39,6	4,3	2,6

Cette feuille de calcul des moyennes et des nombres de dépassement de seuil nécessite les copiés-collés manuels vers le tableau 6-4 :

**colonne 1** : l'une des colonnes issues du tableau 6-2 (ici : **strate PM**) ;

**colonne 2** : le marquage des séquences à prendre en compte dans l'étude, soit toute l'année, ou comme les séquences issues de la campagne nommée ici **Cp01** ;

**colonne 3 et suivantes** : la base de données, soit annuelle pour une pré-étude, ou celles issues des campagnes; dans tous les cas, les données invalidées ou inexistantes doivent se traduire par des cellules soit vides ou marquées d'un « **0** ».

## 6.3 – Bilan des calculs automatiques

### 6.3.1 – Etude réalisée sur une base annuelle de données

Tableau 6-5

SO2 Stg Est	NO STG Est	NO2 STG Est	O3 STG Est	PM10 STG Est	PM2,5 STG Est
<b>COPIE MANUELLE DES CALCULS ANNUELS DU TABLEAU DE GAUCHE</b>					
<b>MOYENNES DES DONNEES PAR STRATE</b>					
5,2	5,8	26,9	29,8	21,1	10,7
7,3	13,4	37,5	60,7	24,2	14,8
11,5	44,8	52,1	95,8	31,8	21,0
<b>ECARTS TYPES DES DONNEES PAR STRATE</b>					
3,4	7,5	13,6	26,3	20,2	6,2
4,9	23,1	17,9	40,7	16,7	8,8
8,2	59,4	20,6	42,0	21,9	13,1
<b>NOMBRES DE DONNEES PAR STRATE</b>					
711	674	710	1908	453	453
1114	1118	1164	680	1209	1203
1007	968	975	300	1217	1216
<b>NOMBRES DE DEPASSEMENTS PAR STRATE</b>					
6	2	14	27	27	3
59	41	72	122	87	64
216	231	194	139	173	213

### 6.3.2 – Etude réalisée sur la base de données des campagnes

Tableau 6-6

SO2 Stg Est	NO STG Est	NO2 STG Est	O3 STG Est	PM10 STG Est	PM2,5 STG Est
<b>COPIE MANUELLE DES CALCULS CAMPAGNE DU TABLEAU DE GAUCHE</b>					
<b>MOYENNES DES DONNEES PAR STRATE</b>					
5,3	5,7	27,4	27,6	17,8	11,8
7,6	11,9	36,7	67,0	25,1	16,2
12,7	48,9	53,8	97,7	31,8	20,9
<b>ECARTS TYPES DES DONNEES PAR STRATE</b>					
2,9	7,2	11,5	26,1	8,4	6,0
6,0	20,2	17,1	42,2	16,8	10,6
10,7	60,9	22,2	35,3	18,6	12,6
<b>NOMBRES DE DONNEES PAR STRATE</b>					
121	112	121	297	70	70
161	154	163	97	182	184
165	163	163	52	193	193
<b>NOMBRES DE DEPASSEMENTS PAR STRATE</b>					
0	1	1	4	0	0
11	3	9	22	15	15
41	44	33	28	24	27

Les résultats des tableaux 6.5 et 6.6 sont copiés automatiquement sur les tableaux 6.7 et 6.8 suivants.

## 6.4 – Pré-étude annuelle

### 6.4.1 – Reconstitution de moyennes

Tableau 6-7

STRASBOURG 2003		SO2 Stg Est	NO STG Est	NO2 STG Est	O3 STG Est	PM10 STG Est	PM2,5 STG Est
<b>MOYENNES ANNUELLES ET INCERTITUDES</b>							
Moyenne annuelle	en µg/m3	8,3	22,5	39,9	44,0	27,0	16,8
Ecart-type annuel	en µg/m3	6,6	41,7	20,4	38,8	20,0	11,2
Intervalle de confiance en µg/m3	+/-	0,20	1,31	0,63	1,19	0,61	0,34
Intervalle de confiance en +/-%	Médiane = 2,0	2,4	5,8	1,6	2,7	2,3	2,0
<b>CALCUL DES MOYENNES PAR STRATE</b>							
Moyennes	m1	5,2	5,8	26,9	29,8	21,1	10,7
des données	m2	7,3	13,4	37,5	60,7	24,2	14,8
par strate	m3	11,5	44,8	52,1	95,8	31,8	21,0
<b>ECARTS-TYPE REDUITS PAR STRATES</b>							
Ecart-types des données par strate	S(X)1	3,4	7,5	13,6	26,3	20,2	6,2
	S(X)2	4,9	23,1	17,9	40,7	16,7	8,8
	S(X)3	8,2	59,4	20,6	42,0	21,9	13,1
Ecart-Type pondéré		5,7	32,0	17,7	31,3	19,4	10,2
<b>NOMBRES DE MESURES PAR STRATE</b>							
nombres de mesures ventilées par strate sur l'ensemble de l'année	N1	711	674	710	1908	453	453
	N2	1114	1118	1164	680	1209	1203
	N3	1007	968	975	300	1217	1216
<b>FRACTIONS DES MESURES PAR STRATE</b>							
fractions des mesures par strate sur l'ensemble de l'année	w1	0,25	0,24	0,25	0,66	0,16	0,16
	w2	0,39	0,41	0,41	0,24	0,42	0,42
	w3	0,36	0,35	0,34	0,10	0,42	0,42
<b>INTERVALLES DE CONFIANCE PONDERES DES MOYENNES ANNUELLES</b>							
IC95 en µg/m3		0,19	1,20	0,55	0,98	0,60	0,32
IC95 en %	médiane = 2,0	2,3	5,3	1,4	2,2	2,2	1,9
<b>NOMBRES DE MESURES A FAIRE PAR STRATE POUR UNE INCERTITUDE DONNEE</b>							
Incertitude +/- 10% (CO)	(en µg/m3 sauf CO)	0,83	2,25	3,99	4,40	2,70	1,68
nombre total de mesures à faire	médiane = 134	130	547	54	138	141	100
nombres de mesures à faire pour calcul des moyennes	n(x)1	20	32	11	77	24	10
	n(x)2	45	161	23	43	51	37
	n(x)3	67	356	22	20	68	55
<b>Si on choisit une campagne de 4 semaines (224 triH) avec les mêmes écarts-type pondérés :</b>							
on aurait : IC95 en µg/m3 = +/-	224	0,66	4,21	1,97	3,52	2,15	1,15
IC95 en % = +/-	médiane = 7,0	8,0	18,7	4,9	8,0	8,0	6,9

Le tableau 6.7 calcule après une pré-étude sur une base annuelle, les nombres de mesures qu'il conviendrait de réaliser par strate et pour chaque polluant lors de campagnes pour une reconstitution de moyennes annuelles assorties d'une incertitude choisie par l'opérateur (ici +/- 10%).

## 6.4.2 – Reconstitution de nombres de dépassements des centiles 90

**Tableau 6-8**

STRASBOURG 2003		SO2 Stg Est	NO STG Est	NO2 STG Est	O3 STG Est	PM10 STG Est	PM2,5 STG Est
<b>AUTRE OBJECTIF DE L'ETUDE :</b>							
<b>NOMBRES DE DEPASSEMENTS DE SEUILS PAR STRATE</b>							
<b>P90 choisis comme exemple en µg/m3 (sauf CO) =</b>		15,6	56,3	65	99,6	48	29
nb de dépassements constatés par classe sur l'ensemble	strate 1	6	2	14	27	27	3
	strate 2	59	41	72	122	87	64
	strate 3	216	231	194	139	173	213
<b>nombre total annuel</b>		<b>281</b>	<b>274</b>	<b>280</b>	<b>288</b>	<b>287</b>	<b>280</b>
<b>INTERVALLES DE CONFIANCE PONDERES DES DEPASSEMENTS DE SEUILS</b>							
<b>IC95 des % de dépassements</b>		<b>0,884</b>	<b>0,879</b>	<b>0,888</b>	<b>0,808</b>	<b>0,912</b>	<b>0,886</b>
<b>IC95 des nombres de dépassements</b>		<b>25</b>	<b>24</b>	<b>25</b>	<b>23</b>	<b>26</b>	<b>25</b>
<b>NOMBRES DE MESURES A FAIRE PAR STRATE POUR UNE INCERTITUDE DONNEE</b>							
<b>RAPPEL des nombres totaux à faire</b>		<b>130</b>	<b>547</b>	<b>54</b>	<b>138</b>	<b>141</b>	<b>100</b>
nombres de mesures à faire pour calcul des dépassements	n(f1)	12	31	7	49	18	5
	n(f2)	45	175	20	57	53	36
	n(f3)	74	343	28	33	71	61
<b>Ces nombres sont différents de ceux prédits pour le calcul des moyennes</b>							

Le tableau 6.8 calcule après une pré-étude sur une base annuelle, les nombres de mesures qu'il conviendrait de réaliser par strate et pour chaque polluant lors de campagnes pour une reconstitution de nombres de dépassement de seuil assortis d'une incertitude choisie par l'opérateur (ici +/- 10%).

## 6.5 – Exploitation des campagnes

### 6.5.1 – Reconstitution de moyennes

Le tableau 6.9 calcule après les campagnes annuelle, par strate, les moyennes de chaque polluant et leurs écarts types, puis reconstitue leurs moyennes annuelles assorties de leurs incertitudes réelles.

Les incertitudes ne sont pas de 10% comme souhaité au départ, car les écarts types obtenus dépendent des mesures réalisées lors des campagnes effectuées, et non plus de celles utilisées dans la pré-étude.

**Tableau 6-9**

STRASBOURG 2003		SO2 Stg Est	NO STG Est	NO2 STG Est	O3 STG Est	PM10 STG Est	PM2,5 STG Est
<b>AUTRES STATISTIQUES UTILES</b>							
nombre de données météo sur l'année		2920	2920	2920	2920	2920	2920
<b>NOMBRES DE MESURES PAR STRATE</b>							
nombres de données météo ventilées par strate sur l'ensemble de l'année	N1	742	729	729	1931	466	466
	N2	1146	1192	1192	687	1222	1222
	N3	1032	999	999	302	1232	1232
<b>CAMPAGNES</b>							
Moyennes des x par strate moyennes brutes de la campagne	m'1	5,3	5,7	27,4	27,6	17,8	11,8
	m'2	7,6	11,9	36,7	67,0	25,1	16,2
	m'3	12,7	48,9	53,8	97,7	31,8	20,9
		<b>8,9</b>	<b>24,4</b>	<b>40,4</b>	<b>44,3</b>	<b>26,9</b>	<b>17,5</b>
<b>moyennes annuelles reconstituées</b>		<b>8,8</b>	<b>23,1</b>	<b>40,2</b>	<b>44,1</b>	<b>26,8</b>	<b>17,5</b>
<b>ECARTS-TYPE REDUITS PAR STRATES</b>							
Ecart-types S(x)	S(x)1	2,9	7,2	11,5	26,1	8,4	6,0
	S(x)2	6,0	20,2	17,1	42,2	16,8	10,6
	S(x)3	10,7	60,9	22,2	35,3	18,6	12,6
<b>NOMBRES DE MESURES PAR STRATE</b>							
nb de mesures validées pendant la campagne ventilées par strate	n1	121	112	121	297	70	70
	n2	161	154	163	97	182	184
	n3	165	163	163	52	193	193
nb de valeurs >0 (max 448)	moyenne = 443	<b>447</b>	<b>429</b>	<b>447</b>	<b>446</b>	<b>445</b>	<b>447</b>
<b>INTERVALLES DE CONFIANCE DES MOYENNES ANNUELLES RECONSTITUEES</b>							
IC95 en µg/m3	moyenne = 1,58	<b>0,58</b>	<b>2,91</b>	<b>1,40</b>	<b>2,48</b>	<b>1,29</b>	<b>0,85</b>
IC95 en %	moyenne = 6,3	<b>6,6</b>	<b>12,6</b>	<b>3,5</b>	<b>5,6</b>	<b>4,8</b>	<b>4,9</b>
<b>INTERVALLES DE CONFIANCE DES MOYENNES BRUTES DE LA CAMPAGNE</b>							
Ecart type bruts		6,1	26,0	25,5	36,9	29,5	9,1
IC95 en µg/m3	moyenne = 1,74	<b>0,47</b>	<b>2,07</b>	<b>1,98</b>	<b>2,88</b>	<b>2,30</b>	<b>0,71</b>
IC95 en %	moyenne = 6,3	<b>5,4</b>	<b>8,5</b>	<b>4,9</b>	<b>6,5</b>	<b>8,6</b>	<b>4,0</b>

Le tableau 6.10 calcule après les campagnes annuelle, par strate, les nombres de dépassement de seuils de chaque polluant et leurs écarts types, puis reconstitue ces mêmes nombres sur l'année complète assortis de leurs incertitudes réelles.



## 6.5.2 – Reconstitution de nombres de dépassements des centiles 90

**Tableau 6-10**

STRASBOURG 2003		SO2 Stg Est	NO STG Est	NO2 STG Est	O3 STG Est	PM10 STG Est	PM2,5 STG Est
<b>AUTRE OBJECTIF DE L'ETUDE :</b>							
<b>NOMBRES DE DEPASSEMENTS DE SEUILS PAR STRATE</b>							
<b>P90 choisit comme exemple en µg/m3 (sauf CO) =</b>		15,6	56,3	65	99,6	48	29
nb de dépassements observés par strate pendant la campagne	n1	0	1	1	4	0	0
	n2	11	3	9	22	15	15
	n3	41	44	33	28	24	27
<b>nombres totaux</b>	<b>n</b>	<b>52</b>	<b>48</b>	<b>43</b>	<b>54</b>	<b>39</b>	<b>42</b>
<b>Nb reconstitués de dépassements annuels</b>							
		<b>335</b>	<b>299</b>	<b>274</b>	<b>344</b>	<b>254</b>	<b>272</b>
<b>Intervalle de confiance</b>	<b>+/-</b>	<b>71</b>	<b>65</b>	<b>66</b>	<b>66</b>	<b>66</b>	<b>68</b>

## 6.6 - Formulaires

<b>FORMULAIRE DE L'ETUDE PREALABLE</b>		
$i = 1 \text{ à } k$	Numéro de chaque strate	1
$N$	Nombre total de mesures	1
$N_i$	Nb de mesures dans la strate $i$	2
$M_i$	Nb de dépassements d'un seuil prédéfini dans la strate $i$	3
$w_i = \frac{N_i}{N}$ avec $\sum_{i=1}^k w_i = 1$	Fraction des données dans la strate $i$	4

$\bar{X}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} X_{ij}$	Moyenne arithmétique des $N_i$ valeurs dans chaque strate $i$	5
$s_i^2(X_{ij}) = \frac{1}{(N_i - 1)} \sum_{j=1}^{N_i} (X_{ij} - \bar{X}_i)^2$	Variance des $N_i$ valeurs dans chaque strate $i$	6
$\bar{X} = \sum_1^k w_i \cdot X_{ij}$	Moyenne globale	7

$s^2(\bar{X}) = \left( \sum_{i=1}^k w_i \cdot s_i(x_{ij}) \right)^2$	Variance de la moyenne globale	8
$t = t_{v;1-\alpha}$	Coefficient de Student-Fisher unilatéral	9
$IC95(\bar{X}) = \frac{t \cdot s(\bar{X})}{\sqrt{N}} = \Delta(N)$	Intervalle de confiance à 95% de la moyenne globale	10
$\Delta(n) \geq \Delta(N)$	Marge d'erreur choisie pour une campagne courte à effectuer	11
$f_i = \frac{M_i}{M}$	Fraction de données supérieures à un seuil prédéfini dans la strate i	12
$n = \left( \frac{2t_{v;1-\alpha}}{\Delta(n)} \right)^2 \cdot \left( \sum_{i=1}^k w_i s_i \right)^2$	Nombre théorique total (n) de mesures à réaliser lors d'une campagne courte pour $\Delta$ donné	13
$n_{i(x)} = n \frac{w_i s_i}{\sum_{i=1}^k w_i s_i}$	Nombre théorique de mesures à réaliser par strate pour la reconstitution d'une moyenne sur une plus longue durée pour $\Delta$ donné	14
$n_{i(f)} = n \frac{w_i \sqrt{f_i(1-f_i)}}{\sum_{i=1}^k w_i \sqrt{f_i(1-f_i)}}$	Nombre théorique de mesures à réaliser par strate i pour la reconstitution d'un nombre de dépassements de seuil sur une plus longue durée pour $\Delta$ donné	15

<b>FORMULAIRE DE LA CAMPAGNE DE MESURE</b>		
$k, w_i$	Données issues de l'étude préalable	16
$n, n_{i(x)}, n_{i(f)}$	Nombres de mesure réellement réalisées lors d'une campagne et reventilées dans chaque strate i	17
$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$	Moyenne arithmétique des mesures dans chaque strate i	18
$s_i^2(x_{ij}) = \frac{1}{(n_i - 1)} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$	Variance des mesures dans chaque strate i	19
$\bar{x} = \sum_{i=1}^k w_i \cdot \bar{x}_i$	Moyenne arithmétique pondérée des mesures dans chaque strate, valeur dite reconstituée	20

$s^2(\bar{x}) = \sum_{i=1}^k \frac{w_i^2 \cdot s_i^2(x_{ij})}{n_i}$	Variance de la moyenne pondérée	21
$IC95 = \pm t_{v;1-\alpha} \cdot s(\bar{x})$	Intervalle de confiance de la moyenne pondérée	22

$f_i = \frac{m_i}{n_i}$	Fraction de données supérieures à un seuil prédéfini dans la strate i obtenue lors de la campagne	23
$f = \sum_{i=1}^k w_i \cdot f_i$	Fraction pondérée reconstituée sur le long terme des données supérieures à un seuil prédéfini	24
$s_i^2(f_i) = f_i \cdot (1 - f_i)$	Variance de la fraction $f_i$ par strate	25
$s^2(f) = \sum_{i=1}^k \frac{w_i^2 \cdot s_i^2(f_i)}{n_i}$	Variance de la fraction pondérée	26
$u = u_{1-\alpha}$	Valeur de la loi normale réduite	27
$IC95 = \pm s(f) \cdot u_{1-\alpha} + \frac{1}{2n}$	Intervalle de confiance de la fraction pondérée	28

## **THEME : TRAITEMENTS NUMERIQUES**

### **METHODES DE RECONSTITUTION TEMPORELLE DE MOYENNES ET DE NOMBRES DE DEPASSEMENTS DE SEUILS PREDEFINIS A PARTIR DE DONNEES DE CAMPAGNE**

#### **CONTEXTE ET OBJECTIF**

Les campagnes de mesure organisées à l'aide de moyens mobiles permettent d'accéder à une information sur les concentrations ou les nombres de dépassements de seuils dans des zones géographiques dépourvues de stations fixes. Cependant les durées des campagnes ne sont représentatives que de situations partielles, , soit quelques semaines en été et/ou en hiver.

Des travaux destinés à satisfaire aux directives, qui imposent une information sur de plus longues périodes (typiquement la saison ou l'année), ont été menés depuis deux ans par le LCSQA et les AASQA, notamment dans le cadre des travaux du GT Moyens Mobiles.

Ils ont pour objectif d'élaborer une stratégie d'utilisation des moyens mobiles de mesure, qui permettra de déterminer les durées et le nombre de campagnes mobiles qu'il faudrait mettre en œuvre dans une année sur un site pour estimer la moyenne annuelle et le nombre de dépassements de seuils prédéterminés avec leurs incertitudes respectives qui auraient été obtenus dans une station fixe.

#### **TRAVAUX ANTERIEURS ET EN COURS**

Deux approches ont été privilégiées jusqu'ici. L'EMD a développé en collaboration avec AREMALM, AIRMARAIX, ORAMIP, ALPA et ASPA une méthode basée sur la prise en compte des impacts des variations de paramètres météorologiques sur les mesures réalisées lors de campagnes mobiles. L'INERIS a évalué, en collaboration avec Atmo Poitou-Charentes, la techniques dite des plans de sondage. Les principes de ces méthodes sont rappelés brièvement ci-dessous :

#### **Méthode de reconstitution temporelle de l'EMD :**

Elle est basée sur l'application de la norme :

“ ISO 9359 – Qualité de l'air – Echantillonnage aléatoire stratifié pour l'évaluation de la qualité de l'air ambiant ”, pour laquelle l'EMD avait en son temps participé à l'élaboration au sein de l' ISO TC146/SC4 .

Elle consiste entre autres possibilités, à faire sur un ou plusieurs sites, des regroupements de mesures ponctuelles selon différents critères, par exemple :

- correspondant à des directions de vent sélectionnées par rapport à un émetteur industriel,
- pour des classes de concentration spécifiées,
- pour des classes de valeurs de paramètres météorologiques, tels que la température, la pluviométrie, la pression, ... ,

- pour des typologies spécifiques de stations, ... .

Cette norme décrit comment reconstituer sur un long terme une concentration moyenne et comment calculer son incertitude, par exemple sur une année, à partir de groupes de mesures partielles.

La méthode est basée sur la prise en compte de l'influence des variations de certains paramètres météorologiques sur les mesures de polluants atmosphériques après sélection de groupes de mesures par classe de paramètres météorologiques.

Un outil de traitement statistique des données a ainsi été mis au point et validé, à l'aide de séries chronologiques complètes de mesures à poste fixe de SO<sub>2</sub> , NO , NO<sub>2</sub> , O<sub>3</sub> , CO et PM<sub>10</sub> , PM<sub>2.5</sub>, provenant des réseaux de Lille, Marseille, Toulon, Toulouse, Le Havre et Strasbourg.

### **Méthode des plans de sondages :**

Proposée par ATMO Poitou-Charentes, la méthode des plans de sondage offre une réponse intéressante et simple d'usage aux questions de l'échantillonnage et de l'estimation de moyennes annuelles. Fondée sur une théorie statistique rigoureuse, elle a pour fonction :

- de déterminer, en s'appuyant sur des données de stations fixes, les stratégies d'échantillonnage qui permettent d'atteindre une précision donnée. Les plans ainsi définis valent pour les stations et pour la ou les années qui ont servi à les élaborer, mais ils fournissent des indications utiles sur le nombre de mesures qu'il est nécessaire de collecter.
- d'estimer, une fois que l'échantillonnage a été réalisé, la concentration moyenne annuelle et un intervalle de confiance autour de cette moyenne. Ces estimations peuvent être corrigées à l'aide d'une série complète de données issues d'une station fixe auxiliaire.

Si, pour être efficace, elle impose quelques contraintes d'échantillonnage, la méthode des plans de sondage n'exige aucun travail de modélisation ni de données de variables externes. Elle peut être mise en œuvre grâce à des programmes informatiques conçus par ATMO Poitou-Charentes et transposés sous R (logiciel libre de statistique) par l'INERIS.

### **TRAVAUX PROPOSES POUR 2005**

Le but de cette étude conjointe est de valoriser, sous la forme de guides utilisateurs à destination des AASQA les travaux fait les années précédentes et en cours sur ces méthodes. De plus, une synthèse des travaux de mise au point et des résultats des deux méthodes sera réalisée dans le cadre du GT « Plan d'échantillonnage et reconstitution de données ». Les outils logiciels simples, développés dans le cadre de ces études pourront être mis à disposition des AASQA (macro de calcul Excel, logiciel statistique R pour les plans de sondage par exemple).

Au préalable une comparaison des résultats obtenus avec chacune des méthodes sur des jeux de données identiques sera réalisée, afin d'identifier les différences et complémentarités des approches, et de définir les contexte d'utilisation.

Ainsi, les méthodologies seront appliquées sur des bases de données éventuellement déjà exploitées lors des travaux de mise au point, sur de nouvelles situations à définir en concertation avec les AASQA.

## **COLLABORATIONS**

AASQA , ADEME

## **DUREE DES TRAVAUX**

1 an