

ANNEXE 3
Fiche descriptive de la méthode des plans de sondage

Fiche de description méthodologique

Méthode des plans de sondages (PS)

Description

Cette méthode statistique, fondée sur la théorie des sondages, comprend deux aspects :

- la planification de la collecte des données en fonction de la précision d'estimation souhaitée ;
- l'estimation des caractères d'intérêt et de leur incertitude par un ou plusieurs estimateurs.

La mise en œuvre du premier point conduit à la définition d'un plan d'échantillonnage aléatoire stratifié dans le temps. A cause des contraintes de mesure, il s'agit généralement d'un plan par grappes.

Le second point concerne la reconstitution proprement dite, après que l'échantillon de données a été collecté. L'opération est la suivante :

- Estimation des moyennes ou des nombres de dépassements de seuil dans les strates temporelles ;
- Pondération de ces estimations par la fréquence d'occurrence de chaque strate et reconstitution des moyennes annuelles ou des nombres annuels de dépassements.

Type d'échantillonnage requis : échantillonnage aléatoire stratifié constitué de nombreuses mesures indépendantes ou de multiples grappes de mesures.

Indicateur recherché		Pas de temps maximal des mesures
<input checked="" type="checkbox"/>	Moyenne annuelle	1 mois
<input checked="" type="checkbox"/>	Incertitude associée	
<input checked="" type="checkbox"/>	Nombre de dépassements d'une moyenne horaire	1 heure
~	Incertitude associée	
<input checked="" type="checkbox"/>	Nombre de dépassements d'une moyenne journalière	1 jour
~	Incertitude associée	
<input checked="" type="checkbox"/>	Nombre de dépassements d'une moyenne sur 8 heures (glissante ou non)	1 heure (la taille minimale des grappes est 8 heures)
~	Incertitude associée	
<input type="checkbox"/>	Reconstitution d'une série chronologique	
<input type="checkbox"/>	Incertitude associée	

Variables complémentaires

Pour une simple reconstitution : **pas de variable complémentaire requise**

Pour effectuer un redressement sur les indicateurs estimés : la mesure du même polluant sur un site permanent continu (une station fixe par exemple). Le pas de temps des données explicatives doit être un diviseur de celui de la mesure.

Estimation d'une incertitude d'estimation

L'estimation d'indicateurs (moyenne, nombre de dépassements) sur une série temporelle incomplète induit une incertitude. La méthode PS fournit une estimation de l'incertitude liée au plan de sondage sous la forme d'intervalles de confiance à 95%. Ceux-ci sont d'autant mieux estimés que le nombre de grappes est plus grand (au minimum une dizaine de grappes environ). La fiabilité des intervalles de confiance peut être contrôlée en simulant un grand nombre de fois le plan de sondage sur une station fixe annexe.

Outil nécessaire

Excel	<input type="checkbox"/>	Logiciel statistique : R (freeware) disponible à l'adresse http://cran.cict.fr	Autres :
-------	--------------------------	---	----------

Références bibliographiques

LAVANCIER F., CAÏNI F., GAZEAU A., 2003. Plan de sondage pour mesures mobiles de la pollution atmosphérique. Pollution atmosphérique, N°180, octobre-décembre.
TILLÉ Yves. Théorie des sondages. Échantillonnage et estimation en populations finies. Cours et exercices. Dunod, Paris, 2001.
DROESBEKE JJ, FICHET B, TASSI P. Les Sondages. Economica 1987.
HOUDRET J-L., MALHERBE L., 2005. Méthodes de reconstitution temporelle de moyennes et de dépassements de seuils à partir de données de campagnes. Rapport LCSQA, www.lcsqa.org.

ANNEXE 4
Fiche descriptive de la méthode « ISO »

Fiche de description méthodologique

Méthode ISO (issue de la norme ISO 9359)

Description

La méthode ISO s'appuie sur la norme « *ISO 9359 – Qualité de l'air – Echantillonnage aléatoire stratifié pour l'évaluation de la qualité de l'air ambiant* » qui considère l'influence des variables météorologiques et/ou temporelles sur les concentrations de polluants.

Son application comprend plusieurs étapes :

a) Sur des séries de données issues d'un ou plusieurs sites de référence :

- Etude des influences des paramètres météorologiques et temporels sur les niveaux de pollution mesurés ; en général, ces influences peuvent être considérées comme reproductibles d'une année sur l'autre et représentatives d'une zone géographique relativement vaste.
- Définition de strates paramétriques : chaque strate paramétrique correspond à une ou plusieurs combinaisons de valeurs de paramètres concourant à des situations de pollution homogènes.

b) Sur les données d'échantillonnage :

- Répartition des données d'échantillonnage dans les strates paramétrique selon les valeurs correspondantes des paramètres influents ;
- Estimation des moyennes ou des nombres de dépassements de seuil dans les strates paramétriques ;
- Pondération de ces estimations par la fréquence d'occurrence de chaque strate et reconstitution des moyennes annuelles ou des nombres annuels de dépassements.

Type d'échantillonnage requis : échantillonnage aléatoire ou systématique constitué de nombreuses mesures indépendantes ou de plusieurs grappes de mesure réparties sur l'année.

Indicateur recherché		Pas de temps maximal des mesures
<input checked="" type="checkbox"/>	Moyenne annuelle	1 mois
<input checked="" type="checkbox"/>	Incertitude associée	
<input checked="" type="checkbox"/>	Nombre de dépassements d'une moyenne horaire	1 heure
~	Incertitude associée	
<input checked="" type="checkbox"/>	Nombre de dépassements d'une moyenne journalière	1 jour
~	Incertitude associée	
<input checked="" type="checkbox"/>	Nombre de dépassements d'une moyenne sur 8 heures (glissante ou non)	1 heure (la taille minimale des grappes est 8 heures)
~	Incertitude associée	
<input type="checkbox"/>	Reconstitution d'une série chronologique	
<input type="checkbox"/>	Incertitude associée	

Variables complémentaires

Données météorologiques permettant la définition des strates paramétriques (données relatives aux vents et à la température lorsque la mesure est horaire, ensoleillement total sur une journée, etc.). Le pas de temps des données individuelles doit être égal à celui de la mesure.

Pour effectuer un redressement sur les indicateurs estimés : la mesure du même polluant sur un site permanent continu (une station fixe par exemple). Le pas de temps des données auxiliaires doit être un diviseur de celui de la mesure.

Incertitude d'estimation

L'estimation d'indicateurs (moyenne, nombre de dépassements) sur une série temporelle incomplète induit une incertitude. La méthode ISO fournit une **estimation** de cette incertitude sous la forme d'intervalles de confiance à 95%. Ceux-ci sont d'autant mieux estimés que le nombre de mesurages par strate est plus grand et que l'estimation prend en compte l'éventuelle dépendance temporelle entre les données. La fiabilité des intervalles de confiance peut être contrôlée en simulant un grand nombre de fois l'échantillonnage sur une station fixe annexe.

Outil nécessaire

Excel <input checked="" type="checkbox"/> Disponible sur le site LCSQA.	Logiciel statistique : R (freeware) disponible à l'adresse http://cran.cict.fr	Autres :
--	---	----------

Références bibliographiques

Norme ISO 9359 – Qualité de l'air – Echantillonnage aléatoire stratifié pour l'évaluation de la qualité de l'air ambiant.

HOUDRET J-L., MALHERBE L., 2005. Méthodes de reconstitution temporelle de moyennes et de dépassements de seuils à partir de données de campagnes. Rapport LCSQA, www.lcsqa.org.

ANNEXE 5
Fiche descriptive de la régression

Fiche de description méthodologique

Reconstitution par modèle de régression

Description

Le modèle de régression consiste à déterminer, au moyen d'un échantillon d'apprentissage, la relation qui existe entre une variable « à expliquer » Y et n variables « explicatives » (X_1, X_2, \dots, X_n) qui peuvent être les concentrations d'autres polluants ou des paramètres météorologiques. Dans le cas de la mesure non permanente, l'échantillon d'apprentissage est constitué par les données recueillies au cours des périodes de mesure.

Il existe une multitude de modèles. Dans le cas d'un modèle linéaire, la relation s'écrit :

$$Y = a_1.X_1 + a_2.X_2 + \dots + a_n.X_n + \varepsilon$$

où les coefficients (a_1, a_2, \dots, a_n) sont calculés sur l'échantillon d'apprentissage par la méthode des moindres carrés.

En utilisant cette relation, il est ensuite possible de prédire, à partir des valeurs des variables explicatives (X_1, X_2, \dots, X_n), les valeurs individuelles de la variable Y sur les périodes pendant lesquelles elle n'a pas été mesurée, et ainsi, de reconstituer différents indicateurs annuels.

Hypothèses à vérifier :

- Indépendance des variables explicatives ;
- Indépendance et normalité des résidus ;
- Homoscédasticité (la variance des résidus est la même pour toutes les valeurs des variables explicatives) ;
- Diagnostic du modèle (coefficient de détermination).

Type d'échantillonnage requis : n'importe quel type d'échantillonnage. Cependant, pour une reconstitution plus précise, un minimum de deux périodes de mesure est préférable à une unique période plus longue.

Indicateur recherché		Pas de temps maximal des mesures
<input checked="" type="checkbox"/>	Moyenne annuelle	1 mois
<input checked="" type="checkbox"/>	Incertitude associée	
<input checked="" type="checkbox"/>	Nombre de dépassements d'une moyenne horaire	1 heure
~	Incertitude associée	
<input checked="" type="checkbox"/>	Nombre de dépassements d'une moyenne journalière	1 jour
~	Incertitude associée	
<input checked="" type="checkbox"/>	Nombre de dépassements d'une moyenne sur 8 heures (glissante ou non)	1 heure (la taille minimale des grappes est 8 heures)
~	Incertitude associée	
<input checked="" type="checkbox"/>	Reconstitution d'une série chronologique	Identique à celui de la série à reconstituer
<input checked="" type="checkbox"/>	Incertitude associée	

