

# *Méthodologie de l'indice de la pollution de l'air*

*Janvier 2011*



Présent  
pour  
l'avenir



## SOMMAIRE

<b>Liste des tableaux .....</b>	<b>3</b>
<b>Liste des figures .....</b>	<b>3</b>
<b>Préambule .....</b>	<b>4</b>
<b>1- Étude préliminaire .....</b>	<b>4</b>
1.1- Origine des données.....	4
1.2- Les types de polluants .....	4
1.2.1- Le dioxyde de soufre (SO <sub>2</sub> ).....	4
1.2.2- Le dioxyde d'azote (NO <sub>2</sub> ).....	4
1.2.3- L'ozone (O <sub>3</sub> ) .....	4
1.2.4- Les particules de diamètre inférieur à 10 µm(PM <sub>10</sub> ) .....	4
1.2.5- Le monoxyde de carbone (CO).....	4
1.3- Les types de stations de mesure .....	4
1.4- Étude statistique .....	4
1.4.1- Statistiques de base.....	4
1.4.2- Le comportement spatial des polluants .....	5
1.5- Le choix de l'indicateur de tendance centrale .....	6
<b>2- Traitement des données brutes .....</b>	<b>6</b>
2.1- Données sources.....	6
2.2- Passage à des valeurs journalières.....	6
2.3- Calcul des valeurs manquantes .....	6
2.3.1- Méthodologie .....	6
2.3.2- Statistiques.....	7
2.3.2.1- Fonctionnement des stations urbaines et périurbaines .....	7
2.3.2.2- Valeurs manquantes .....	9
2.3.3- Méthodologie .....	10
2.3.3.1- Critères .....	10
2.3.3.2- Statistiques .....	10
<b>3- Détermination des stations à prendre en compte .....</b>	<b>11</b>
3.1- Méthodologie .....	11
3.2- Evolution du réseau .....	12
<b>4- Calcul des indices .....</b>	<b>12</b>
4.1- Station.....	13
4.2- Agglomération.....	13
4.3- Strate .....	13
4.4- Indice global.....	13
4.5- Agrégation annuelle.....	13
4.6- Indice France .....	13
4.7- Changement de la mesure des particules .....	13
4.8- Indices mensuels .....	13

## Liste des tableaux

<i>Tableau 1 : Statistiques des moyennes mensuelles en <math>\mu\text{g}/\text{m}^3</math> sur la période échantillonnée : 1998-2004.....</i>	<i>5</i>
<i>Tableau 2 : Statistiques du dioxyde de soufre par type de stations.....</i>	<i>5</i>
<i>Tableau 3 : Statistiques du dioxyde d'azote par type de stations.....</i>	<i>5</i>
<i>Tableau 4 : Statistiques de l'ozone par type de stations.....</i>	<i>5</i>
<i>Tableau 5 : Statistiques des particules de diamètre inférieur à <math>10\ \mu\text{m}</math> par type de stations.....</i>	<i>6</i>
<i>Tableau 6 : Statistiques de fonctionnement des stations urbaines et périurbaines par polluant et par année.....</i>	<i>8</i>
<i>Tableau 7 : Nombre moyen de jours manquants pour les stations urbaines et périurbaines ayant fonctionnées plus de 90% de l'année (résultats par année et par polluant).....</i>	<i>9</i>
<i>Tableau 8 : Découpage des grandes régions.....</i>	<i>10</i>
<i>Tableau 9 : Nombre de jours redressés et nombre de jours redressés avec des stations de la même agglomération par polluant et par année.....</i>	<i>11</i>
<i>Tableau 10 : Evolution du réseau de mesure par polluant.....</i>	<i>12</i>
<i>Tableau 11 : Nombre d'agglomérations par strate.....</i>	<i>13</i>

## Liste des figures

<i>Figure 1 : Exemple d'interpolation.....</i>	<i>7</i>
--	----------

## **Préambule**

Le SOeS produit chaque année depuis 2005 un indice de la pollution de l'air dans les villes avec des données depuis 2000. Cet indice est décliné par strate de taille d'agglomération (quatre classes) et par polluant (SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub> et PM<sub>10</sub>). Un indice global est également diffusé. Ce document décrit la méthodologie appliquée.

## **1- Étude préliminaire**

### **1.1- Origine des données**

Toutes les données concernant les concentrations de polluants dans l'air sont extraites de la BDQA (base de données sur la qualité de l'air), jusqu'à présent gérée par l'ADEME (Agence de l'Environnement et de la Maîtrise de l'Energie). Ces données sont issues des mesures des stations fixes effectuées par les associations agréées de surveillance de la qualité de l'air (AASQA).

Les AASQA gèrent des réseaux permettant à la fois de surveiller des polluants réglementaires (directives européennes notamment) et de tenir compte des besoins locaux.

Les polluants les plus suivis à ce jour sont le dioxyde de soufre (SO<sub>2</sub>), le dioxyde d'azote (NO<sub>2</sub>), l'ozone (O<sub>3</sub>), les particules et le monoxyde de carbone (CO).

### **1.2- Les types de polluants**

#### **1.2.1- Le dioxyde de soufre (SO<sub>2</sub>)**

Il provient de la présence d'impuretés dans les combustibles fossiles (charbon, pétrole, fuel) qui réagissent avec l'oxygène de l'air pour donner des composés soufrés. Le SO<sub>2</sub> a aussi une origine volcanique.

#### **1.2.2- Le dioxyde d'azote (NO<sub>2</sub>)**

La majorité du NO<sub>2</sub> provient du transport routier et des installations de combustions (centrales énergétiques).

#### **1.2.3- L'ozone (O<sub>3</sub>)**

C'est un polluant secondaire issu de réactions photochimiques à partir de polluants précurseurs dans certaines conditions d'ensoleillement.

#### **1.2.4- Les particules de diamètre inférieur à 10 µm (PM<sub>10</sub>)**

Elles ont une origine anthropique (chauffage urbain, centrales électriques, industries, véhicules à moteur, agriculture...) et naturelle (poussières désertiques, volcaniques, feux de forêt).

#### **1.2.5- Le monoxyde de carbone (CO)**

Il provient de la combustion incomplète de combustibles fossiles et des carburants. La majorité est émise par la circulation automobile, le chauffage, l'industrie ainsi que les feux de biomasse.

### **1.3- Les types de stations de mesure**

Il existe plusieurs types de stations de mesures de la qualité de l'air :

- Les stations dites de « fond », réalisant un suivi de l'exposition moyenne des personnes et de l'environnement. Selon leur lieu d'implantation, il peut s'agir de stations urbaines, périurbaines ou rurales.

- Les stations dites de « proximité ». Placées à proximité de sources d'émissions importantes, elles fournissent des informations sur les concentrations mesurées dans des zones représentatives du niveau maximum auquel la population riveraine est susceptible d'être exposée. Il peut s'agir de stations industrielles ou de trafic automobile.

### **1.4- Étude statistique**

#### **1.4.1- Statistiques de base**

La première remarque porte sur la grande disparité des mesures. En effet, quand les échantillons de SO<sub>2</sub>, de NO<sub>2</sub> ou de O<sub>3</sub> présentent plus de 16 000 valeurs mensuelles, les particules n'en présentent que 9 000 et le CO 2 000.

Les moyennes et médianes sont différentes, sauf pour l'ozone. La médiane est toujours inférieure (de 10 à 30%) à la moyenne indiquant la présence de fortes valeurs dans les échantillons. Ceci est confirmé par les valeurs des maxima, très forts notamment pour le SO<sub>2</sub> et le CO.

Statistiques	SO <sub>2</sub>	NO <sub>2</sub>	O <sub>3</sub>	PM <sub>10</sub>	CO
Moyenne	9	29.7	47.8	20.2	772.8
Médiane	6.5	27.9	47.8	19.8	601.3
Ecart-type	8.4	14.3	20.8	6.4	573
Min	0.1	0.1	0.9	1.8	42.3
Max	100	99.2	134.2	96.2	5723.9

*Tableau 1 : Statistiques des moyennes mensuelles en µg/m<sup>3</sup> sur la période échantillonnée : 1998-2004*

De ces principaux polluants, le CO est le moins observé par les réseaux de mesure.

#### 1.4.2- Le comportement spatial des polluants

Le traitement porte sur les séries de base relevées dans trois agglomérations test. L'hypothèse de travail testée concerne la pondération des différentes stations d'une agglomération.

- SO<sub>2</sub> :

Type	Stations urbaines	Stations périurbaines	Stations de proximité
Moyenne	9.7	7.6	5.5
Médiane	7.1	5.8	3.7
Ecart-type	8.9	7	5.1
Max	100	82.6	36.4

*Tableau 2 : Statistiques du dioxyde de soufre par type de stations*

Nous sommes en présence de trois grands échantillons, de taille inégale. Les médianes sont très différentes et inférieures aux moyennes, en raison de la contribution des fortes valeurs isolées. Les écart-types sont élevés. Les stations urbaines et périurbaines sont plus proches.

- Le NO<sub>2</sub>

Type	Stations urbaines	Stations périurbaines	Stations de proximité
Moyenne	32.8	24.9	10.7
Médiane	30.9	23.9	8.3
Ecart-type	14.2	10.5	8.4
Max	99.2	98.3	125.4

*Tableau 3 : Statistiques du dioxyde d'azote par type de stations*

Les moyennes sont assez différentes des médianes, à l'exception des stations périurbaines.

- O<sub>3</sub>

Type	Stations urbaines	Stations périurbaines	Stations rurales
Moyenne	44.1	47.5	61.9
Médiane	43.5	47.7	61.1
Ecart-type	19.7	18.8	22.1
Max	131.7	114	134.2

*Tableau 4 : Statistiques de l'ozone par type de stations*

Les moyennes sont très proches des médianes. Les trois types de stations présentent des plages de valeurs comparables avec des ordres de grandeur des concentrations différents. Les moyennes sont statistiquement différentes deux à deux.

- Les PM<sub>10</sub>

Type	Stations urbaines	Stations périurbaines	Stations de proximité
Moyenne	21.0	17.3	9.5
Médiane	20.4	17.3	15.1
Ecart-type	5.6	8.1	22.1
Max	68.5	96.2	45.2

*Tableau 5 : Statistiques des particules de diamètre inférieur à 10 µm par type de stations*

Les stations urbaines et périurbaines sont plus homogènes que les stations de proximité.

### 1.5- Le choix de l'indicateur de tendance centrale

Il s'agit de vérifier la pertinence du choix technique de l'indicateur de tendance centrale. En fait, il faut vérifier si le choix de la moyenne ou de la médiane peut avoir une incidence sur le classement des stations.

Les données disponibles laissent présager que le choix de l'indicateur est important.

Par exemple, pour les stations urbaines mesurant le SO<sub>2</sub>, les quartiles du rapport D9/D1 valent respectivement 6.3, 8.5 et 13 avec une moyenne à 12.6, ce qui montre à la fois une dispersion importante du polluant dans la station et entre les stations (déjà vu plus haut).

Les grands écarts entre les moyennes n'auraient pas trop d'incidence sur l'indice si l'objectif était de suivre les évolutions et non les comparaisons entre les agglomérations. On peut imaginer que les évolutions sont parallèles.

L'étude des données annuelles et mensuelles permet de déterminer quelques contraintes :

- les polluants les plus suivis par le dispositif de surveillance en France sont le SO<sub>2</sub>, le NO<sub>2</sub>, l'O<sub>3</sub> et les PM<sub>10</sub>,
- la moyenne mensuelle est la statistique qui permet de mieux respecter le cycle saisonnier qui caractérise la plupart des polluants,
- enfin, seules les stations périurbaines et urbaines sont à prendre en compte, les autres étant très hétérogènes et trop peu représentées.

## 2- Traitement des données brutes

### 2.1- Données sources

Les données sources proviennent de la BDQA (base de données sur la qualité de l'air) gérée jusqu'à par l'ADEME et alimentée par les mesures réalisées par les AASQA. Elles sont fournies chaque année. Les fichiers contiennent les relevés de mesures heure par heure de chaque station.

### 2.2- Passage à des valeurs journalières

Des valeurs moyennes journalières sont calculées pour chaque station. Quelques valeurs manquantes peuvent apparaître à cause de défaillances temporaires.

### 2.3- Calcul des valeurs manquantes

#### 2.3.1- Méthodologie

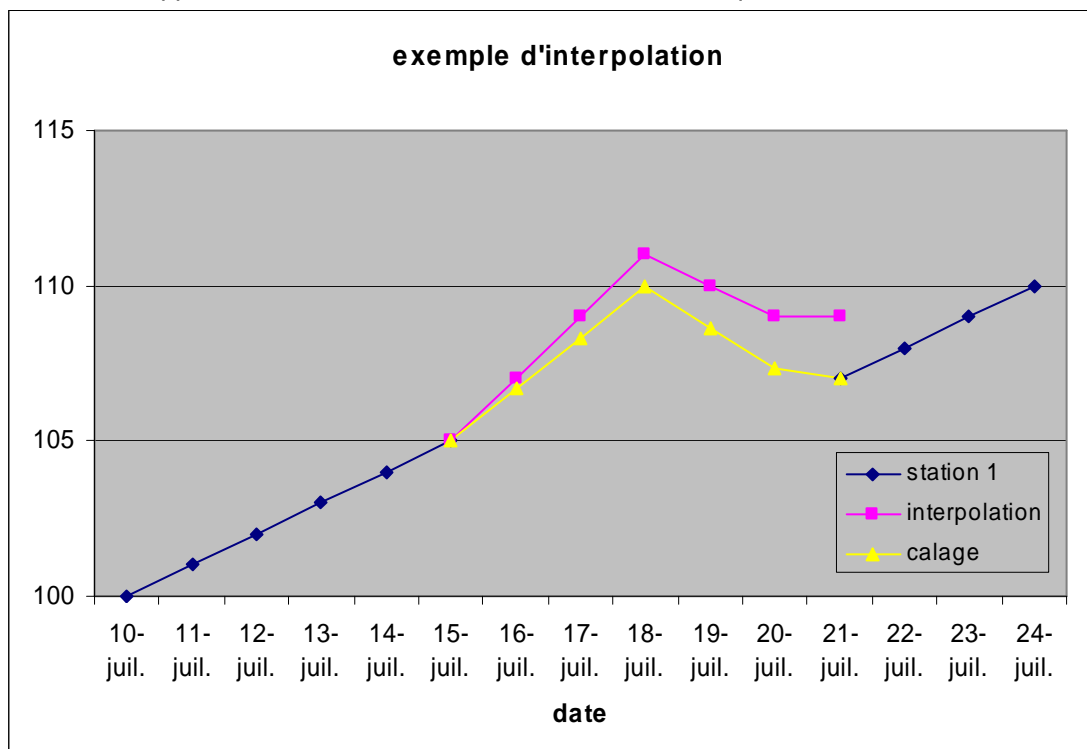
Les stations ayant fonctionné moins de 90% du temps (calculé sur la base des données horaires) sont rejetées. De plus, les stations ayant fonctionnées plus de 90% de l'année mais présentant plus de 30 jours consécutifs sans mesures ne sont pas sélectionnées.

Pour les autres, les valeurs manquantes sont calculées par :

- Interpolation : cas « standard »
  - Détermination des taux d'évolutions journaliers moyens sur la période pour les stations « semblables » : le taux d'évolution moyen est calculé avec une moyenne géométrique
  - Application des taux journaliers sur la période
  - Comparaison avec la première valeur réelle
  - Calage
    - Calcul d'un taux de calage journalier avec la première valeur non manquante

$$t = \sqrt[n]{\frac{V_{exacte}}{V_{int\ erpolée}}}$$

- Application du taux sur l'intervalle de valeurs manquantes



*Figure 1 : Exemple d'interpolation*

- Extrapolation : valeurs manquantes en fin de période
  - Même méthodologie que précédemment mais sans calage
- Rétropolation : valeurs manquantes en début de période
  - Même méthodologie que l'extrapolation mais en sens inverse

Le critère de 90% de fonctionnement a été choisi de façon arbitraire et peut masquer de grandes disparités dans le fonctionnement des stations, l'interpolation étant d'autant plus risquée que la plage de non fonctionnement est longue.

### 2.3.2- Statistiques

#### 2.3.2.1- Fonctionnement des stations urbaines et périurbaines

Polluant	Année	Nombre de stations	Nombre de stations >90%	%
SO <sub>2</sub>	2000	256	199	78
	2001	259	208	80
	2002	258	200	78
	2003	246	187	76
	2004	234	167	71
	2005	218	157	72
	2006	205	150	73
	2007	177	141	80
	2008	172	138	80
	2009	163	122	75
NO <sub>2</sub>	2000	295	247	84
	2001	313	251	80
	2002	318	249	78
	2003	316	265	84
	2004	341	269	79
	2005	345	293	85
	2006	348	307	88
	2007	339	289	85
	2008	331	296	89
	2009	320	282	88
O <sub>3</sub>	2000	289	235	81
	2001	308	249	81
	2002	318	265	83
	2003	330	289	88
	2004	347	286	82
	2005	358	309	86
	2006	359	326	91
	2007	353	317	90
	2008	350	311	89
	2009	339	296	87
PM <sub>10</sub>	2000	157	106	68
	2001	162	103	64
	2002	201	147	73
	2003	209	181	87
	2004	231	178	77
	2005	237	199	84
	2006	246	212	86
	2007 (classiques)	235	201	74
	2007 (ajustées)	233	174	75
	2008	246	210	85
	2009	237	188	79

*Tableau 6 : Statistiques de fonctionnement des stations urbaines et périurbaines par polluant et par année*

Le pourcentage de stations urbaines et périurbaines ayant fonctionné plus de 90% de l'année varie assez fortement suivant l'année et le polluant. Mais il est toujours supérieur à 70% excepté pour les PM<sub>10</sub> en 2000 et 2001.



## 2.3.2.2- Valeurs manquantes

Polluant	Année	Nombre moyen de jours manquants
SO <sub>2</sub>	2000	5
	2001	5
	2002	3
	2003	4
	2004	5
	2005	5
	2006	4
	2007	3
	2008	4
	2009	4
NO <sub>2</sub>	2000	5
	2001	6
	2002	6
	2003	6
	2004	7
	2005	6
	2006	6
	2007	5
	2008	6
	2009	7
O <sub>3</sub>	2000	4
	2001	4
	2002	3
	2003	3
	2004	4
	2005	4
	2006	4
	2007	3
	2008	3
	2009	4
PM <sub>10</sub>	2000	5
	2001	5
	2002	5
	2003	5
	2004	6
	2005	4
	2006	5
	2007 (ajustées)	7
	2008	4
	2009	6

*Tableau 7 : Nombre moyen de jours manquants pour les stations urbaines et périurbaines ayant fonctionnées plus de 90% de l'année (résultats par année et par polluant)*

Le nombre moyen de jours manquants varie entre 3 et 7. Il est donc assez éloigné du seuil à 90% (36 jours).

## 2.3.3- Méthodologie

### 2.3.3.1- Critères

Le redressement est effectué à partir de taux d'évolutions moyens sur les stations considérées comme proches.

Les critères sont par ordre décroissant :

- Même agglomération
- Même « grande région » et même tranche de taille d'agglomération

Les grandes régions ne correspondent pas à un découpage administratif, c'est un zonage particulier :

Grande région	Zones
1	Ile-de-France
	Centre
2	Nord-Pas-de-Calais
	Picardie
	Champagne-Ardenne
3	Lorraine
	Alsace
	Franche-Comté
	Bourgogne
4	Auvergne
5	Rhône
	Loire
6	Alpes (autres stations de Rhône-Alpes)
7	PACA
8	Languedoc-Roussillon
9	Midi-Pyrénées
	Aquitaine
	Limousin
10	Poitou-Charentes
	Pays de la Loire
11	Bretagne
	Basse-Normandie
	Haute-Normandie

*Tableau 8 : Découpage des grandes régions*

### 2.3.3.2- Statistiques

Polluant	Année	Nombre de jours redressés	Nombre de jours redressés avec des stations de la même agglomération	%
SO <sub>2</sub>	2000	897	323	36
	2001	992	360	36
	2002	682	255	37
	2003	734	225	31
	2004	776	242	31
	2005	835	240	29
	2006	632	196	31
	2007	484	185	38
	2008	572	179	31
	2009	538	210	39

Polluant	Année	Nombre de jours redressés	Nombre de jours redressés avec des stations de la même agglomération	%
NO <sub>2</sub>	2000	1343	387	29
	2001	1539	408	27
	2002	1427	355	25
	2003	1550	399	26
	2004	1816	450	25
	2005	1630	429	26
	2006	1779	459	26
	2007	1337	371	28
	2008	1643	405	25
O <sub>3</sub>	2000	912	266	29
	2001	893	300	34
	2002	824	203	25
	2003	967	260	27
	2004	1262	309	24
	2005	1117	342	31
	2006	1184	336	28
	2007	904	279	31
	2008	987	293	30
PM <sub>10</sub>	2000	559	135	24
	2001	514	161	31
	2002	662	142	21
	2003	836	218	26
	2004	995	274	28
	2005	772	237	31
	2006	965	259	27
	2007 (ajustées)	1254	254	20
	2008	854	247	29
2009	1099	392	36	

*Tableau 9 : Nombre de jours redressés et nombre de jours redressés avec des stations de la même agglomération par polluant et par année*

Le redressement se fait assez peu avec des stations de la même agglomération (entre 20 et 40%) en raison de leur faible nombre dans la plupart des cas.

### 3- Détermination des stations à prendre en compte

#### 3.1- Méthodologie

La détermination des stations à prendre en compte est un facteur important dans le calcul d'un indice. Idéalement le panier doit être à la fois constant mais toujours représentatif de la réalité. Dans le cas présent, le champ des stations est en constante évolution (plutôt en progression) mais il faut également tenir compte des pannes.

Il a été décidé d'utiliser une méthode avec panier évolutif. Elle consiste à utiliser pour le calcul de l'indice de l'année n, les stations ayant fonctionné cette même année et la précédente (méthode biannuelle). Cette méthode présente l'avantage de tenir compte de l'évolution du réseau tout en restant sur un champ constant qui permet de mesurer l'évolution réelle entre deux années.

D'autres méthodes étaient envisageables :

- Méthode à champ constant complet :  
Prise en compte des stations qui fonctionnent depuis 2000 : le nombre de stations éligibles ne peut que baisser suite à l'arrêt ou à des pannes. Il s'agit d'une méthode qui chaque année perd en fiabilité avec la réduction de la taille de l'échantillon. Par ailleurs elle impose de rediffuser chaque année les séries depuis 2000.
- Méthode à champ évolutif complet :  
Prise en compte de toutes les stations ayant fonctionné chaque année : on utilise toute la richesse du réseau, les résultats par strate seront fiables en revanche au niveau de l'agglomération l'ajout ou la suppression d'une station peut avoir un impact fort sur les résultats et rend beaucoup moins pertinent, l'analyse des tendances.

### 3.2- Evolution du réseau

Polluant	SO <sub>2</sub>										
	Année	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Nb stations	256	259	258	246	234	218	205		177	172	163
Nb stations constantes depuis 2000	256	243	234	219	197	173	157		129	125	117
Nb stations constantes depuis 2000 avec fonctionnement >90%	199	168	145	111	90	71	62		54	48	41
Nb stations biannuelles avec fonctionnement >90%	.	167	169	156	150	122	127		124	117	108
Polluant	NO <sub>2</sub>										
Année	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	
Nb stations	295	313	318	316	341	345	348		339	331	320
Nb stations constantes depuis 2000	295	285	278	268	267	257	252		237	226	216
Nb stations constantes depuis 2000 avec fonctionnement >90%	247	216	188	171	148	137	126		117	110	104
Nb stations biannuelles avec fonctionnement >90%	.	216	213	214	230	239	266		269	264	263
Polluant	O <sub>3</sub>										
Année	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	
Nb stations	289	308	318	330	347	358	359		353	350	339
Nb stations constantes depuis 2000	289	276	265	263	254	245	237		229	220	206
Nb stations constantes depuis 2000 avec fonctionnement >90%	235	208	191	177	164	146	140		131	124	114
Nb stations biannuelles avec fonctionnement >90%	.	207	223	244	259	255	291		298	294	278
Polluant	PM <sub>10</sub>										
Année	2000	2001	2002	2003	2004	2005	2006	2007 classiques	2007 ajustées	2008	2009
Nb stations	157	162	201	209	231	237	246	235	233	246	237
Nb stations constantes depuis 2000	157	130	126	125	124	118	112		103	100	94
Nb stations constantes depuis 2000 avec fonctionnement >90%	106	78	69	68	64	56	54		41	38	31
Nb stations biannuelles avec fonctionnement >90%	.	78	91	138	157	157	188	174		157	166

Tableau 10 : Evolution du réseau de mesure par polluant

## 4- Calcul des indices

Une fois que les données sont prêtes, le calcul des indices peut débuter. Il consiste à calculer des résultats par station et polluant et ensuite à les agréger.

L'indice calculé est de type « surfacique » (ou « territoire »), il ne tient pas compte de la population vivant à proximité des stations.

#### 4.1- Station

Les valeurs mensuelles moyennes sont calculées pour chaque polluant de chaque station.

#### 4.2- Agglomération

Les valeurs mensuelles moyennes pour chaque polluant sont calculées par une moyenne simple des valeurs de chaque station.

#### 4.3- Strate

Les agglomérations sont réparties en quatre strates.

Strate	Population	Nombre d'agglomérations
1	< 100 000	129
2	de 100 000 à 250 000	30
3	de 250 000 à 1 000 000	20
4	> 1 000 000	4

*Tableau 11 : Nombre d'agglomérations par strate*

Une valeur moyenne mensuelle est calculée pour chaque polluant en pondérant chaque agglomération par sa superficie, on en déduit un indice par rapport à la valeur moyenne de l'année précédente.

#### 4.4- Indice global

Un indice global « multipolluants » est estimé en calculant la moyenne simple des indices par polluant. Il n'y a pas de pondération affectée à chaque polluant car l'indice est de type « territoire » et non de type « population ».

#### 4.5- Agrégation annuelle

Les indices annuels sont calculés avec une moyenne simple des indices mensuels.

#### 4.6- Indice France

L'indice France métropolitaine se déduit avec une moyenne des indices par strate pondérée par la superficie des agglomérations.

#### 4.7- Changement de la mesure des particules

Pour les PM<sub>10</sub>, les modalités de mesure ont été modifiées au 1<sup>er</sup> janvier 2007, afin de rendre les résultats équivalents à ceux obtenus par la méthode de référence fixée par la réglementation européenne. Pour 2007, l'indice d'évolution de la pollution de l'air a été calculé à partir des données de PM<sub>10</sub> non ajustées, afin de permettre une comparaison avec l'année 2006. Pour 2008, il a été calculé à partir des données de PM<sub>10</sub> ajustées de 2007 afin de permettre une comparaison avec l'année 2008.

#### 4.8- Indices mensuels

Les indices mensuels sont calculés dans le processus actuel mais non diffusés.