

Indice de la pollution de l'air

Méthodologie

Avril 2012

Présent
pour
l'avenir



Table des matières

Liste des tableaux	3
Liste des figures	3
1 – L'indice de la pollution de l'air	4
1.1 – L'origine des données	4
1.2 – Les types de polluants	4
1.2.1 – Le dioxyde de soufre (SO ₂)	4
1.2.2 – Le dioxyde d'azote (NO ₂)	4
1.2.3 – L'ozone (O ₃)	4
1.2.4 – Les particules de diamètre inférieur à 10 µm (PM ₁₀)	4
1.3 – Les types de stations de mesure	4
2 – Le choix des stations à prendre en compte	5
2.1 – La préselection des stations	5
2.1.1 – Les stations françaises hors Corse et DOM	5
2.1.2 – Les stations des communes urbaines et périurbaines non isolées	5
2.2 – Le filtrage de stations	5
2.2.1 – Les stations ayant fonctionné moins de 90 % de l'année	5
2.2.2 – Les stations ayant connu des périodes d'interruption de plus de 720h consécutives	5
2.3 – La sélection finale des stations	5
2.4 – L'évolution du réseau des stations entre 2000 et 2010	6
3 – Le traitement des données brutes	6
3.1 – Les données source	6
3.2 – Le passage à des données journalières	6
3.3 – L'estimation des valeurs manquantes	6
3.3.1 – La méthodologie	7
3.3.2 – Quelques statistiques	11
3.4 – Le passage à des données mensuelles	13
4 – Le calcul des indices	13
4.1 – Concentrations mensuelles par polluant et par unité urbaine	13
4.2 – Concentrations mensuelles par polluant et par strate	13
4.3 – Indices mensuels par polluant et par strate	13
4.4 – Indices annuels par polluant et par strate	14
4.5 – Série des indices annuels par polluant et par strate	15
4.6 – Série des indices annuels par polluant	15

Liste des tableaux

Tableau 1 : Evolution du réseau de mesure par polluant pour la méthode biannuelle et la méthode champ constant complet -----	6
Tableau 2 : Exemple - Calcul de taux d'évolution journaliers moyens à partir des stations « semblables » -----	8
Tableau 3 : Exemple - Application des taux d'évolution journaliers moyens aux valeurs précédent les valeurs manquantes -----	9
Tableau 4 : Exemple - Estimation définitive après calage -----	10
Tableau 5 : Pourcentage de moyennes journalières manquantes par années et par polluants et durée moyenne des périodes sans données journalières par années et par polluants -----	11
Tableau 6 : Pourcentages de moyennes journalières manquantes estimées à partir des stations de la même commune, de la même unité urbaine, de la même AASQA, de la même grande région et de la même strate par années et par polluants -----	12
Tableau 7 : Concentrations moyennes mensuelles et annuelles et indices mensuels par polluant pour les années 2000-2001 et la strate 1 -----	14
Tableau 8 : Indices annuels par polluant pour les années 2000-2001 et 2001-2002 et la strate 1 -----	14
Tableau 9 : Série des indices annuels par polluant pour la période 2000-2002 et la strate 1 -----	15
Tableau 10 : Poids des différentes strates dans la superficie totale de la France couverte par les stations-----	15
Tableau 11 : Série des indices annuels par polluant et par strate pour la période 2000-2002 -----	16
Tableau 12 : Série des indices annuels par polluant pour la période 2000-2002 -----	16

Liste des figures

Figure 1 : Exemple - Valeurs manquantes en cours d'année -----	8
Figure 2 : Exemple - Estimation par application des taux d'évolution journaliers moyens aux valeurs précédent les valeurs manquantes -----	9
Figure 3 : Exemple - Estimation définitive après calage -----	11

1- L'indice de la pollution de l'air

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 (SO₂, NO₂, O₃ et PM₁₀). Un indice global est également calculé.

1.1 – L'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), gérée par l'ADEME (Agence de l'Environnement et de la Maîtrise de l'Energie) jusque fin 2010. Ces données sont issues des mesures des stations fixes effectuées par les associations agréées de surveillance de la qualité de l'air (AASQA). Depuis le 1^{er} janvier 2011, le LCSQA (Laboratoire centrale de surveillance de la qualité de l'air) est en charge de la gestion de la base nationale de qualité de l'air.

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 (SO₂), le dioxyde d'azote (NO₂), l'ozone (O₃) et les particules.

1.2 – Les types de polluants

1.2.1 – Le dioxyde de soufre (SO₂)

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 SO₂ a aussi une origine volcanique.

1.2.2 – Le dioxyde d'azote (NO₂)

La majorité du NO₂ provient du transport routier et des installations de combustions (centrales énergétiques).

1.2.3 – L'ozone (O₃)

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 (PM₁₀)

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

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.

2 – Le choix des stations à prendre en compte

2.1 – La présélection des stations

2.1.1 – Les stations françaises hors Corse et DOM

Dans un premier temps, toutes les stations françaises hors Corse et DOM présentant des mesures sont conservées. Dans la phase de traitement des données, chaque station est rattachée à une « grande région » ou à défaut à une strate d'unité urbaine. Or il paraît délicat de comparer les DOM et la Corse avec la France métropolitaine. Il a donc été choisi d'exclure les stations des DOM et de la Corse.

2.1.2 – Les stations des communes urbaines et périurbaines non isolées

Parmi les stations françaises hors Corse et DOM, seules les stations des communes urbaines et périurbaines non isolées (selon la définition de l'Insee) sont retenues.

2.2 – Le filtrage de stations

2.2.1 – Les stations ayant fonctionné moins de 90 % de l'année

Pour chaque polluant et chaque année, les stations ayant fonctionné moins de 90 % de l'année sont éliminées.

2.2.2 – Les stations ayant connu des périodes d'interruption de plus de 720h consécutives

Pour chaque polluant et chaque année, les stations ayant connu des périodes d'interruption de plus de 720h consécutives sont éliminées.

2.3 – La sélection finale des stations

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.

Pour les PM₁₀, les modalités de mesure ont été modifiées au 1^{er} 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 de la pollution de l'air a été calculé à partir des données de PM₁₀ non ajustées, afin de permettre une comparaison avec l'année 2006. Pour 2008, il a été calculé à partir des données de PM₁₀ ajustées de 2007 afin de permettre une comparaison avec l'année 2008.

Le « chaînage » des évolutions 2006-2007 et 2007-2008 ainsi calculées permet de gommer l'effet du changement de méthode de mesure.

2.4 – L'évolution du réseau des stations entre 2000 et 2010

SO ₂											
Année	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Nombre de stations retenues pour cette année et la précédente (méthode biannuelle)		167	169	156	148	117	125	123	116	107	97
Nombre de stations retenues depuis 2000 (méthode champ constant complet)		167	146	114	89	67	61	54	48	41	37
NO ₂											
Année	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Nombre de stations retenues pour cette année et la précédente (méthode biannuelle)		217	212	213	227	234	260	263	258	258	252
Nombre de stations retenues depuis 2000 (méthode champ constant complet)		217	189	171	147	134	123	115	108	102	94
O ₃											
Année	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Nombre de stations retenues pour cette année et la précédente (méthode biannuelle)		207	223	239	255	251	280	284	281	267	260
Nombre de stations retenues depuis 2000 (méthode champ constant complet)		207	190	174	160	142	136	128	122	113	102
PM ₁₀											
Année	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Nombre de stations retenues pour cette année et la précédente (méthode biannuelle)		78	91	138	156	155	184	169	151	167	165
Nombre de stations retenues depuis 2000 (méthode champ constant complet)		78	69	68	63	54	52	44	38	33	30

Tableau 1 : Evolution du réseau de mesure par polluant pour la méthode biannuelle et la méthode champ constant complet

3 – Le traitement des données brutes

3.1 – Les données source

Les données sources proviennent de la BDQA (base de données sur la qualité de l'air) gérée jusque fin 2010 par l'ADEME et alimentée par les mesures réalisées par les AASQA. Depuis le 1^{er} janvier 2011, le LCSQA (Laboratoire centrale de surveillance de la qualité de l'air) est en charge de la gestion de la base nationale de qualité de l'air. Les données sources sont fournies chaque année. Chaque fichier contient les relevés de mesures heure par heure de chaque station pour un polluant et une année donnée. Les fichiers fournis peuvent contenir plusieurs jeux de données pour une même station, un même polluant et une même année, chaque jeu de données correspondant à une technique de mesure. Pour les stations concernées, une sélection est réalisée pour ne retenir qu'un jeu de données par station.

3.2 – Le passage à des données journalières

Des concentrations journalières sont calculées à l'aide de moyennes simples appliquées sur les mesures horaires. Lorsque le taux de fonctionnement d'une station sur une journée est inférieur à 75 % (18h/24h), la mesure journalière est considérée comme manquante et une estimation doit être réalisée.

3.3 – L'estimation des valeurs manquantes

A cause de défaillances temporaires, des valeurs manquantes peuvent apparaître dans les concentrations journalières. Cependant, la fréquence d'apparition de ces anomalies est raisonnable puisque le filtrage a déjà permis l'exclusion des stations fonctionnant moins de 90 % de l'année ou connaissant une période d'interruption de plus de 720h consécutives.

3.3.1 – La méthodologie

Il existe 3 types de valeurs manquantes et, par conséquent, 3 types d'estimation :

- Les valeurs manquantes en cours d'année (cas « standard ») : l'estimation consiste à déterminer une valeur qui soit cohérente avec les valeurs précédentes et suivantes et qui permette de compléter fidèlement la série. La technique employée est l'interpolation.
- Les valeurs manquantes en fin d'année : l'estimation consiste à déterminer une valeur qui soit cohérente avec les valeurs précédentes et qui permette de terminer fidèlement la série, il s'agit d'une extrapolation.
- Les valeurs manquantes en début d'année : l'estimation consiste à déterminer une valeur qui soit cohérente avec les valeurs suivantes et qui permette de débiter fidèlement la série, il s'agit d'une rétropolation.

Les estimations se font par année et par polluant.

L'interpolation

1^{ère} étape : Calcul de taux d'évolution journaliers moyens à partir des stations « semblables ».

Pour chaque valeur manquante d'une station donnée, un taux d'évolution journalier moyen N/N-1 est calculé. Ce taux permet de connaître « en moyenne » l'évolution de la concentration du polluant considéré entre le jour précédent et le jour pour lequel la valeur est manquante.

Ce taux moyen est obtenu en réalisant une moyenne géométrique sur les taux d'évolution journalier N/N-1 de toutes les stations semblables à la station d'intérêt entre le jour précédent et le jour pour lequel la valeur est manquante.

Moyenne géométrique :

La moyenne géométrique de n nombres est égale à la racine nième du produit de ces nombres. Elle intervient dès que l'on considère des nombres qui se multiplient plus qu'ils ne s'additionnent (ex : taux d'évolution).

Stations semblables :

Des stations sont considérées comme semblables lorsqu'elles sont proches géographiquement. La difficulté réside dans l'hétérogénéité de la répartition spatiale des stations. Ce constat est d'autant plus vrai qu'au moment de l'estimation des concentrations manquantes de nombreuses stations ont déjà été exclues. Il peut donc y avoir, d'un côté, 3 stations dans une même commune, et d'un autre côté, 1 seule station dans un large périmètre.

Pour contourner cet obstacle, le choix a été fait de déterminer plusieurs niveaux d'appartenance de la station (commune, unité urbaine, AASQA, grande région, strate) et de calculer un taux d'évolution pour chacun de ces niveaux à partir de l'ensemble des stations y étant également présentes. En allant du plus fin au plus grossier, le premier taux calculable est retenu comme taux d'évolution journalier moyen.

A noter le fait que les différents niveaux s'englobent et que lorsqu'il est question de "stations de la même grande région", il s'agit en réalité des stations de la même grande région et de la même strate.

Les grandes régions sont construites ainsi :

- **PAR** : Ile de France et Centre
- **EST** : Bourgogne, Lorraine, Alsace et Franche-Comté
- **NPC** : Champagne-Ardenne, Picardie et Nord-Pas-de-Calais
- **N-W** : Haute-Normandie, Basse-Normandie et Bretagne
- **POI** : Pays de la Loire et Poitou-Charentes
- **S-W** : Aquitaine, Midi-Pyrénées et Limousin
- **ALP** : Rhône-Alpes (Sauf communes des aires urbaines de la Loire et du Rhône)

- **LYO** : Communes des aires urbaines de la Loire et du Rhône
- **AUV** : Auvergne
- **L-R** : Languedoc-Roussillon
- **PAC** : Provence-Alpes-Côte d'Azur

Les strates regroupent les unités urbaines par population :

- Population < 100 000 hab → **Strate 1**
- Population < 250 000 hab → **Strate 2**
- Population < 1 000 000 hab → **Strate 3**
- Population ≥ 1 000 000 → **Strate 4**

Exemple :

Prenons l'exemple d'une station (station n°1) pour laquelle les concentrations journalières en ozone ne sont pas connues pour la période allant du 7 au 12 décembre. La station n°2, présente dans la même commune, présente des valeurs manquantes pour la période allant du 10 au 12 décembre. Les stations n°3 et 4, appartenant à la même unité urbaine que les 2 précédentes, ne présentent pas de valeurs manquantes pour le mois de décembre.

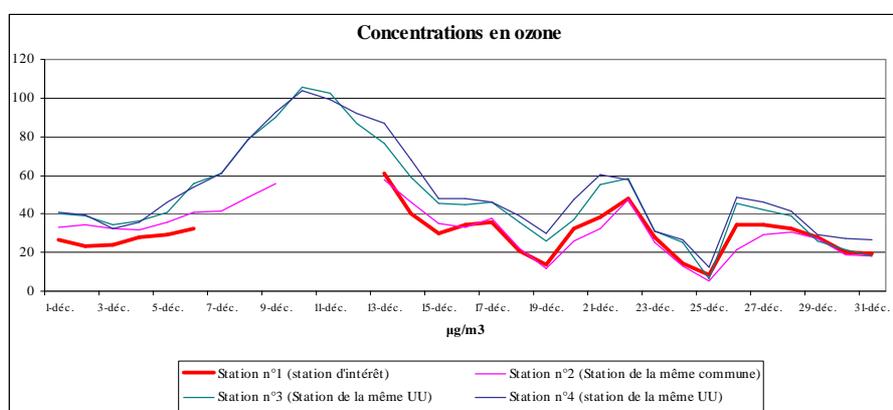


Figure 1 : Exemple -Valeurs manquantes en cours d'année

Le tableau ci-dessous présente la manière dont sont calculés les taux d'évolution journaliers moyens N/N-1.

Date	6-déc.	7-déc.	8-déc.	9-déc.	10-déc.	11-déc.	12-déc.	13-déc.
Concentrations - Station n°1	32,14							61,12
Concentrations - Station n°2	40,65	41,56	48,79	55,46				57,89
Concentrations - Station n°3	55,66	60,84	78,23	90,12	105,41	102,45	87,21	76,53
Concentrations - Station n°4	53,65	61,15	78,25	92,65	103,78	99,25	92,35	86,98
Taux d'évolution journalier N/N-1 - Station n°2		1,0224	1,1740	1,1367				
Taux d'évolution journalier N/N-1 - Station n°3		1,0931	1,2858	1,1520	1,1697	0,9719	0,8512	0,8775
Taux d'évolution journalier N/N-1 - Station n°4		1,1398	1,2796	1,1840	1,1201	0,9563	0,9305	0,9419
Taux d'évolution moyen journalier N/N-1 - Stations de la même commune		1,0224	1,1740	1,1367				
Taux d'évolution moyen journalier N/N-1 - Stations de la même UU		1,1162	1,2827	1,1679	1,1446	0,9641	0,8900	0,9091

Tableau 2 : Exemple - Calcul de taux d'évolution journaliers moyens à partir des stations « semblables ».

Dans un premier temps, des taux d'évolution journaliers N/N-1 sont calculés pour la station de la même commune (station n°2) ainsi que pour les deux autres stations de la même unité urbaine (stations n°3 et 4).

Ex : La concentration en ozone dans la station n°2 passe de 40.65 à 41.56 entre le 6 et le 7 décembre. Le taux d'évolution journalier N/N-1 du 7 décembre est donc $41.56/40.65=1.0224$.

La station n°2 faisant partie de la même commune que la station n°1, c'est le taux d'évolution journalier N/N-1 de cette station qui est privilégié pour la période allant du 7 décembre au 9 décembre.

Pour la période allant du 10 au 12 décembre, il n'y a plus de taux d'évolution journalier N/N-1 au niveau de la commune. Il est nécessaire de passer au niveau supérieur et de calculer des taux d'évolution journaliers moyens N/N-1 sur les stations de la même unité urbaine (les stations n°3 et 4). Ces taux moyens sont obtenus en réalisant une moyenne géométrique des taux d'évolution journaliers N/N-1 des deux stations.

Ex : Le taux d'évolution journalier moyen N/N-1 du 10 décembre est égal à $\sqrt{(1,1697 * 1,1201)} = 1,1446$.

2^{ème} étape : Application des taux d'évolution journaliers moyens aux valeurs précédentes les valeurs manquantes.

En appliquant le taux d'évolution journalier moyen N/N-1 à la valeur précédente N-1, la valeur N est estimée.

La plupart du temps, plusieurs valeurs manquantes se suivent. Dans ce cas, il faut estimer une par une les valeurs manquantes en commençant par le début de manière à construire l'estimation de la seconde valeur manquante à partir de la première (qui aura été précédemment estimée), et ainsi de suite.

Cette estimation n'est que temporaire car un calage doit être effectué par la suite pour permettre une cohérence avec les valeurs suivantes.

Exemple :

Le tableau ci-dessous présente la manière dont sont estimées les valeurs manquantes dans l'exemple.

Date	6-déc.	7-déc.	8-déc.	9-déc.	10-déc.	11-déc.	12-déc.	13-déc.
Concentrations - Station n°1	32,14							61,12
Taux d'évolution moyen journalier N/N-1		1,0224	1,1740	1,1367	1,1446	0,9641	0,8900	0,9091
Station n°1 (estimation)		32,86	38,58	43,85	50,19	48,39	43,07	39,15

Tableau 3 : Exemple - Application des taux d'évolution journaliers moyens aux valeurs précédentes les valeurs manquantes.

Les valeurs manquantes sont estimées en appliquant les taux d'évolution moyens journaliers N/N-1 aux valeurs précédentes (estimées ou non).

Ex : La concentration estimée au 7 décembre pour la station n°1 est égale à $32,14 * 1,0224 = 32,86$. La concentration estimée au 10 décembre pour la station n°1 est égale à $43,85 * 1,1446 = 50,19$.

Le graphique ci-dessous permet de visualiser le résultat de cette estimation.

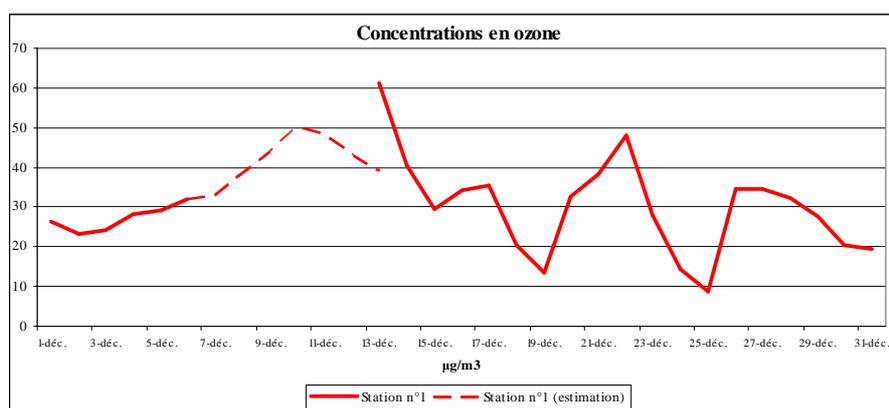


Figure 2 : Exemple - Estimation par application des taux d'évolution journaliers moyens aux valeurs précédentes les valeurs manquantes

La concentration de la journée du 13 décembre a été volontairement estimée alors que la valeur n'était pas manquante pour cette date. Ceci permet d'évaluer la différence entre la valeur estimée et la valeur réelle pour la journée du 13 décembre. La conclusion est que l'estimation sous-évalue la concentration pour cette journée et certainement pour les précédentes. Il est utile de rehausser la courbe en la calant.

3^{ème} étape : Calage

Dans le but de raccorder la série estimée avec les valeurs suivantes, l'estimation est poussée jusqu'à la prochaine valeur non manquante. L'estimation est alors analysée via le ratio « Valeur réelle / Valeur estimée ». En appliquant ce ratio à la dernière valeur estimée (celle correspondant à la prochaine valeur non manquante), la valeur réelle est reconstituée. Il s'agit donc d'un taux de calage global valable pour toute la série manquante.

Pour obtenir le taux de calage journalier, il suffit alors de calculer la racine nième de ce ratio (où n est le nombre de jours entre la dernière valeur non manquante et la prochaine). Chaque estimation est donc corrigée en appliquant le taux de calage journalier à la puissance "nombre de jours écoulés depuis la dernière valeur non manquante".

Exemple :

A chaque étape de l'estimation, la valeur N a été déduite de la valeur N-1 et du taux d'évolution moyen journalier N/N-1. L'objectif est d'appliquer un nouveau taux (créé à partir du taux d'évolution moyen journalier) qui permette d'égaliser la valeur estimée et la valeur réelle pour le 13 décembre.

Schématiquement, l'estimation a été construite ainsi :

$$\begin{aligned} - & VE_7 = V_6 * T_{7/6} \\ - & VE_8 = VE_7 * T_{8/7} = V_6 * T_{7/6} * T_{8/7} \\ - & \dots \\ - & VE_{13} = VE_{12} * T_{13/12} = V_6 * T_{7/6} * T_{8/7} * T_{9/8} * T_{10/9} * T_{11/10} * T_{12/11} * T_{13/12} \end{aligned}$$

avec VE = Valeur estimée temporaire, V = Valeur réelle et T = taux d'évolution journalier.

L'objectif est de créer un nouveau taux TC avec $TC_{N/N-1} = T_{N/N-1} * C$ tel que :

$$\begin{aligned} - & V_6 * TC_{7/6} * TC_{8/7} * TC_{9/8} * TC_{10/9} * TC_{11/10} * TC_{12/11} * TC_{13/12} = V_{13} \\ - & V_6 * T_{7/6} * C * T_{8/7} * C * T_{9/8} * C * T_{10/9} * C * T_{11/10} * C * T_{12/11} * C * T_{13/12} * C = V_{13} \\ - & V_6 * T_{7/6} * T_{8/7} * T_{9/8} * T_{10/9} * T_{11/10} * T_{12/11} * T_{13/12} * C^7 = V_{13} \\ - & VE_{13} * C^7 = V_{13} \end{aligned}$$

Il est donc nécessaire d'avoir : $C = \sqrt[7]{V_{13}/VE_{13}}$.

La série des estimations doit donc être reconstruite ainsi :

$$\begin{aligned} - & VEE_7 = V_6 * T_{7/6} * C \\ - & VEE_8 = VEE_7 * T_{8/7} * C = V_6 * T_{7/6} * C * T_{8/7} * C = V_6 * T_{7/6} * T_{8/7} * C^2 \\ - & \dots \\ - & VEE_{12} = VEE_{11} * T_{12/11} = V_6 * T_{7/6} * T_{8/7} * T_{9/8} * T_{10/9} * T_{11/10} * T_{12/11} * C^6 \end{aligned}$$

avec VEE = Valeur estimée définitive, V = Valeur estimée temporaire, V = Valeur réelle et T = taux d'évolution journalier.

Toutes les nouvelles estimations sont donc obtenues en multipliant l'ancienne estimation par le taux calage à la puissance "nombre de jours écoulés depuis la dernière valeur non manquante".

Date	6-déc.	7-déc.	8-déc.	9-déc.	10-déc.	11-déc.	12-déc.	13-déc.
Station n°1	32,14							61,12
Station n°1 (première estimation)		32,86	38,58	43,85	50,19	48,39	43,07	39,15
Station n°1 (seconde estimation)		35,02	43,81	53,07	64,74	66,51	63,09	61,12

Tableau 4 : Exemple - Estimation définitive après calage

Ex : La concentration estimée au 7 décembre pour la station n°1 est égale à $32,86 * \sqrt[7]{61,12/39,15} = 35,02$. La concentration estimée au 10 décembre pour la station n°1 est égale à $50,19 * (\sqrt[7]{61,12/39,15})^4 = 64,74$.

Le graphique ci-dessous permet de visualiser le résultat de l'estimation définitive.

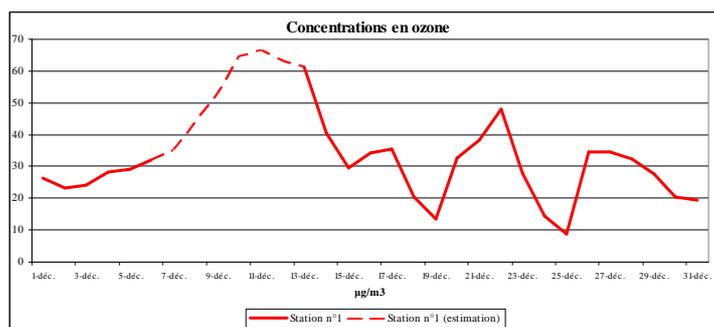


Figure 3 : Exemple - Estimation définitive après calage

L'extrapolation

La méthode utilisée pour extrapoler reprend les étapes 1 et 2 de l'interpolation. Il n'y a pas de calage car pas de valeurs suivantes non manquantes.

La rétropolation

La méthode utilisée fonctionne de manière inverse à l'extrapolation.

1^{ère} étape : Calcul de taux d'évolution journaliers moyens à partir des stations « semblables ».

Pour chaque valeur manquante d'une station donnée, un taux d'évolution journalier moyen N+1/N est calculé. Ce taux permet de connaître « en moyenne » l'évolution de la concentration du polluant considéré entre le jour pour lequel la valeur est manquante et le suivant.

Ce taux moyen est obtenu en réalisant une moyenne géométrique sur les taux d'évolution journalier N+1/N de toutes les stations semblables à la station d'intérêt entre le jour pour lequel la valeur est manquante et le suivant.

2^{ème} étape : Application des taux d'évolution journaliers moyens aux valeurs suivant les valeurs manquantes.

En appliquant le taux d'évolution journalier moyen N+1/N à la valeur suivante N+1, la valeur N est estimée.

La plupart du temps, plusieurs valeurs manquantes se suivent. Dans ce cas, il faut estimer une par une les valeurs manquantes en commençant par la fin de manière à construire l'estimation de l'avant dernière valeur manquante à partir de la dernière (qui aura été précédemment estimée), et ainsi de suite.

3.3.2 – Quelques statistiques

Année	Pourcentage de concentrations journalières manquantes				Durée moyenne des périodes sans concentrations journalières (en jour)			
	SO ₂	NO ₂	O ₃	PM ₁₀	SO ₂	NO ₂	O ₃	PM ₁₀
2000	3,1	3,2	2,3	3,1	3	4	3	4
2001	3,0	3,3	2,1	3,0	3	4	3	4
2002	3,0	3,2	2,0	3,1	3	4	4	4
2003	3,4	3,4	2,1	3,2	3	4	4	4
2004	3,3	3,6	2,3	3,4	3	4	4	4
2005	3,8	3,1	2,2	2,8	3	4	3	3
2006	3,6	3,4	2,3	3,0	3	4	3	3
2007 classique	3,1	2,8	1,9	4,2	3	4	3	4
2007 ajusté	nc	nc	nc	2,9	nc	nc	nc	3
2008	3,4	3,1	2,1	3,1	3	4	3	3
2009	3,9	3,3	2,2	3,9	2	4	4	3
2010	3,5	2,6	2,0	3,9	3	4	3	3

Note : nc = non concerné

Tableau 5 : Pourcentage de moyennes journalières manquantes par années et par polluants et durée moyenne des périodes sans données journalières par années et par polluants

Polluant	Année	avec les stations de la même commune	avec les stations de la même unité urbaine	avec les stations de la même AASQA	avec les stations de la même grande région	avec les stations de la même strate
SO ₂	2000	33%	42%	10%	12%	2%
	2001	28%	43%	15%	13%	1%
	2002	24%	47%	16%	11%	1%
	2003	26%	39%	19%	12%	4%
	2004	30%	28%	22%	15%	5%
	2005	17%	40%	19%	17%	7%
	2006	18%	32%	24%	21%	6%
	2007	11%	43%	17%	20%	9%
	2008	14%	32%	25%	24%	6%
	2009	18%	27%	18%	25%	11%
2010	13%	31%	17%	23%	17%	
NO ₂	2000	32%	42%	13%	12%	2%
	2001	35%	40%	14%	10%	1%
	2002	28%	43%	19%	10%	0%
	2003	35%	40%	14%	10%	1%
	2004	31%	33%	22%	12%	2%
	2005	29%	45%	17%	9%	1%
	2006	23%	42%	23%	11%	1%
	2007	28%	42%	23%	4%	3%
	2008	26%	38%	22%	14%	0%
	2009	23%	42%	22%	11%	1%
2010	25%	44%	20%	10%	1%	
O ₃	2000	33%	42%	11%	13%	2%
	2001	30%	47%	12%	9%	2%
	2002	28%	40%	21%	10%	1%
	2003	29%	38%	21%	12%	1%
	2004	30%	32%	23%	13%	2%
	2005	25%	44%	22%	8%	1%
	2006	22%	38%	26%	13%	1%
	2007	19%	44%	30%	5%	2%
	2008	24%	37%	20%	19%	0%
	2009	23%	34%	23%	17%	2%
2010	19%	43%	23%	14%	1%	
PM ₁₀	2000	25%	41%	16%	14%	5%
	2001	27%	24%	20%	17%	11%
	2002	21%	44%	15%	11%	8%
	2003	26%	41%	20%	11%	2%
	2004	31%	37%	13%	13%	6%
	2005	19%	46%	18%	14%	2%
	2006	19%	42%	19%	16%	3%
	2007 ajustée	11%	34%	21%	25%	9%
	2007 classique	15%	46%	26%	8%	4%
	2008	17%	34%	29%	19%	2%
	2009	19%	28%	26%	23%	4%
2010	15%	28%	32%	21%	5%	

Tableau 6 : Pourcentages de moyennes journalières manquantes estimées à partir des stations de la même commune, de la même unité urbaine, de la même AASQA, de la même grande région et de la même strate par années et par polluants

3.4 – Le passage à des données mensuelles

Des concentrations mensuelles sont ensuite calculées à l'aide de moyennes simples appliquées sur les mesures journalières par année et par polluant.

4 – Le calcul des indices

L'objectif est de calculer un indice annuel par strate d'unité urbaine et par polluant à partir des concentrations mensuelles par polluant et par station. Quatre tailles de strates d'unités urbaine ont été choisies en fonction du nombre d'habitants : strate 1 (< 100 000 habitants), strate 2 (\geq 100 000 et < 250 000 habitants), strate 3 (\geq 250 000 et < 1 000 000 habitants) et strate 4 (\geq 1 000 000 habitants). Un indice au niveau France entière est également calculé. Les concentrations mensuelles sont regroupées par couple biannuel pour chaque polluant.

4.1 – Concentrations mensuelles par polluant et par unité urbaine

Pour chaque polluant et chaque année, des concentrations mensuelles par unité urbaine sont calculées à l'aide de moyennes simples appliquées sur les concentrations mensuelles des stations de chaque unité urbaine.

4.2 – Concentrations mensuelles par polluant et par strate

Pour chaque polluant et chaque année, des concentrations mensuelles par strate sont calculées à l'aide de moyennes pondérées appliquées sur les concentrations mensuelles des unités urbaines de chaque strate. Chaque unité urbaine est pondérée par sa superficie.

4.3 – Indices mensuels par polluant et par strate

Des indices mensuels sont ensuite calculés pour chaque polluant, chaque année et chaque strate en appliquant la formule suivante :

$$\text{Indice mensuel du mois } M \text{ de l'année } N = 100 * \frac{\text{Concentration mensuelle du mois } M \text{ de l'année } N}{\text{Concentration annuelle de l'année } N-1}$$

Exemple :

Le tableau ci-dessous présente, pour la strate 1 et les années 2000 et 2001, les concentrations mensuelles, les concentrations annuelles de 2000 ainsi que les indices calculés pour chacun des polluants.

Mois	Moyenne mensuelle SO ₂	Moyenne mensuelle NO ₂	Moyenne mensuelle O ₃	Moyenne mensuelle PM ₁₀	Moyenne 2000 SO ₂	Moyenne 2000 NO ₂	Moyenne 2000 O ₃	Moyenne 2000 PM ₁₀	Indice mensuel SO ₂	Indice mensuel NO ₂	Indice mensuel O ₃	Indice mensuel PM ₁₀
01/2000	10,7466	38,9524	19,5878	32,6266	5,1111	25,7653	42,7887	20,8339	210,26	151,18	45,78	156,60
02/2000	7,5970	33,0074	33,9913	23,6294	5,1111	25,7653	42,7887	20,8339	148,64	128,11	79,44	113,42
03/2000	6,3779	29,8385	43,4578	24,1571	5,1111	25,7653	42,7887	20,8339	124,79	115,81	101,56	115,95
04/2000	5,5462	24,5132	54,6403	16,3207	5,1111	25,7653	42,7887	20,8339	108,51	95,14	127,70	78,34
05/2000	3,6333	21,1586	54,5662	19,3894	5,1111	25,7653	42,7887	20,8339	71,09	82,12	127,52	93,07
06/2000	3,2115	17,5369	64,0165	19,5249	5,1111	25,7653	42,7887	20,8339	62,83	68,06	149,61	93,72
07/2000	2,7663	16,6409	55,7059	16,5228	5,1111	25,7653	42,7887	20,8339	54,12	64,59	130,19	79,31
08/2000	2,8946	20,4077	56,1810	21,1970	5,1111	25,7653	42,7887	20,8339	56,63	79,21	131,30	101,74
09/2000	3,6012	20,9345	44,8811	19,9899	5,1111	25,7653	42,7887	20,8339	70,46	81,25	104,89	95,95
10/2000	3,8452	25,3185	31,3284	19,1275	5,1111	25,7653	42,7887	20,8339	75,23	98,27	73,22	91,81
11/2000	4,4148	28,2498	30,6735	16,2336	5,1111	25,7653	42,7887	20,8339	86,38	109,64	71,69	77,92
12/2000	6,6981	32,6251	24,4345	21,2882	5,1111	25,7653	42,7887	20,8339	131,05	126,62	57,10	102,18
01/2001	6,7290	33,7383	27,5540	19,8028	5,1111	25,7653	42,7887	20,8339	131,65	130,94	64,40	95,05
02/2001	6,0113	31,2890	32,2437	22,6902	5,1111	25,7653	42,7887	20,8339	117,61	121,44	75,36	108,91
03/2001	5,2993	27,9163	44,5690	15,4730	5,1111	25,7653	42,7887	20,8339	103,68	108,35	104,16	74,27
04/2001	4,0997	20,9185	60,7059	14,2475	5,1111	25,7653	42,7887	20,8339	80,21	81,19	141,87	68,39
05/2001	3,7391	21,5961	64,0653	18,5809	5,1111	25,7653	42,7887	20,8339	73,16	83,82	149,72	89,19
06/2001	3,2547	19,2485	67,9298	16,9105	5,1111	25,7653	42,7887	20,8339	63,68	74,71	158,76	81,17
07/2001	2,7046	18,0443	65,4948	18,0250	5,1111	25,7653	42,7887	20,8339	52,92	70,03	153,07	86,52
08/2001	2,8377	19,4082	60,2988	19,1807	5,1111	25,7653	42,7887	20,8339	55,52	75,33	140,92	92,06
09/2001	2,6463	20,4146	42,4522	16,3148	5,1111	25,7653	42,7887	20,8339	51,78	79,23	99,21	78,31
10/2001	4,0222	26,5927	32,2143	24,1014	5,1111	25,7653	42,7887	20,8339	78,70	103,21	75,29	115,68
11/2001	5,4675	30,3885	21,8122	22,3358	5,1111	25,7653	42,7887	20,8339	106,97	117,94	50,98	107,21
12/2001	7,9428	34,5143	21,3276	26,2029	5,1111	25,7653	42,7887	20,8339	155,40	133,96	49,84	125,77

Tableau 7 : Concentrations moyennes mensuelles et annuelles et indices mensuels par polluant pour les années 2000-2001 et la strate 1

NB : Des indices mensuels intermédiaires sont calculés avant le passage aux indices annuels. Cette étape n'est pas nécessaire mais permet de visualiser les évolutions plus précisément.

Ex : La concentration annuelle en SO₂ de l'année 2000 est égale à :
 $(10,7466+7,5970+6,3779+5,5462+3,6333+3,2115+2,7663+2,8946+3,6012+3,8452+4,4148+6,6881)/12=5,1111$.

L'indice mensuel pour le SO₂ et le mois de janvier 2000 est égal à $100*(10,7466/5,1111)=210,26$

4.4 – Indices annuels par polluant et par strate

Des indices annuels par polluant et par strate sont calculés à l'aide de moyennes simples appliquées sur les indices mensuels par polluant et par strate. A noter que pour tous les couples d'années, les indices de la première année sont égaux à 100.

Exemple :

Le tableau ci-dessous présente, pour la strate 1 et les années 2000-2001 et 2001-2002, les indices annuels pour chacun des polluants.

Année	Indice annuel SO ₂	Indice annuel NO ₂	Indice annuel O ₃	Indice annuel PM ₁₀
2000	100,00	100,00	100,00	100,00
2001	89,27	98,35	105,30	93,54
2001	100,00	100,00	100,00	100,00
2002	92,56	97,96	102,94	103,52

Tableau 8 : Indices annuels par polluant pour les années 2000-2001 et 2001-2002 et la strate 1

Ex : L'indice annuel pour le SO₂ et l'année 2001 est égal à :
 $(131.65+117.61+103.68+80.21+73.16+63.68+52.92+55.52+51.78+78.70+106.97+155.40)/12=89.27$.

4.5 – Série des indices annuels par polluant et par strate

Il s'agit de raccorder les indices entre eux en appliquant l'évolution observée entre deux années à la valeur de l'indice précédent. Ce « chaînage » des indices est réalisé en multipliant l'indice de l'année N+2 à la valeur de l'indice calculé pour l'année N+1 et à diviser par 100.

Un indice global par année est calculé en appliquant une moyenne simple aux indices des 4 polluants.

Exemple :

Les séries d'indices sont constituées comme suit :

Année	Indice SO ₂	Indice NO ₂	Indice O ₃	Indice PM ₁₀	Indice global
2000	100,00	100,00	100,00	100,00	100,00
2001	89,27	98,35	105,30	93,54	96,62
2002	82,63	96,34	108,40	96,84	96,05

Tableau 9 : Série des indices annuels par polluant pour la période 2000-2002 et la strate 1

Ex : L'indice SO₂ est égal à 100 en 2000, $100*89.27/100=89.27$ en 2001, $89.27*92.56/100=82.63$ en 2002.

L'indice global est égal à $(82.63+96.34+108.40+96.84)/4=384.21/4=96.05$.

4.6 – Série des indices annuels par polluant

Les indices "France" sont calculés à l'aide de moyennes pondérées appliquées sur les indices annuels par strate et par polluant. La pondération est de type surfacique : chaque strate est pondérée par la superficie totale de l'ensemble des unités urbaine présentes dans chaque strate.

Année	Strate 1	Strate 2	Strate 3	Strate 4
2001	19,27	22,86	37,61	20,25
2002	21,64	22,19	36,51	19,66
2003	24,76	20,89	34,38	19,96
2004	28,13	17,84	33,36	20,67
2005	28,93	17,64	32,99	20,44
2006	30,25	18,94	32,70	18,10
2007	31,30	18,66	32,21	17,83
2008	31,62	18,57	32,06	17,75
2009	30,87	18,78	32,41	17,94
2010	30,96	18,75	32,37	17,92

Tableau 10 : Poids des différentes strates dans la superficie totale de la France couverte par les stations

Exemple :

Les indices par strates pour la période 2000-2002 sont présentés ci-dessous.

Strate	Année	Indice SO ₂	Indice NO ₂	Indice O ₃	Indice PM ₁₀	Indice global
1	2000	100,00	100,00	100,00	100,00	100,00
1	2001	89,27	98,35	105,30	93,54	96,62
1	2002	82,63	96,34	108,40	96,84	96,05
2	2000	100,00	100,00	100,00	100,00	100,00
2	2001	85,70	100,72	106,18	98,35	97,74
2	2002	78,33	98,15	108,82	101,17	96,62
3	2000	100,00	100,00	100,00	100,00	100,00
3	2001	83,51	101,07	105,15	101,92	97,91
3	2002	78,79	98,73	105,90	102,85	96,57
4	2000	100,00	100,00	100,00	100,00	100,00
4	2001	92,10	99,43	104,14	101,45	99,28
4	2002	92,03	94,58	106,94	102,03	98,89

Tableau 11 : Série des indices annuels par polluant et par strate pour la période 2000-2002

Les indices "France" pour la période 2000-2002 sont constitués comme suit :

Année	Indice SO ₂	Indice NO ₂	Indice O ₃	Indice PM ₁₀	Indice global
2000	100,00	100,00	100,00	100,00	100,00
2001	86,86	100,13	105,21	99,39	97,90
2002	82,12	97,27	107,29	101,02	96,92

Tableau 12 : Série des indices annuels par polluant pour la période 2000-2002

Ex : En 2001, l'indice "France" SO₂ est égal à :
 $(89.27*19.27+85.70*22.86+83.51*37.61+92.10*20.25)/100=86.86$