



Laboratoire Central de Surveillance de la Qualité de l'Air



EXPLOITATION DE DONNEES PM

Etude de l'impact du trafic routier sur les teneurs en PM
à partir de la base de données PM10 et PM2.5

Jean-Luc HOUDRET

Novembre 2008

Version finale



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'Énergie, de l'Écologie, du Développement Durable et de l'Aménagement du Territoire 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

EXPLOITATION DE DONNEES PM

**Etude de l'impact du trafic routier
sur les teneurs en PM à partir de la base de données
PM10 et PM2.5**

Jean Luc HOUDRET

Convention : 0001190

Novembre 2008

Sommaire

RESUME DE L'ETUDE EMD 2008.....	7
1. OBJECTIF DE L'ETUDE.....	9
2. VALIDATION DE LA BASE DE DONNEES HORAIRES 2007.....	9
2.1 BILAN GENERAL.....	9
2.2 AJUSTEMENT DES DONNEES	10
2.2.1 Rappel du principe de l'ajustement	10
2.2.2 Modalités de calcul du « delta »	11
2.2.3 Incidence de l'ajustement sur 4 heures.....	11
2.2.4 Influence des distances entre référence et autres stations.....	13
2.3 LES STATIONS UTILISEES	13
2.3.1 Les stations trafic.....	13
2.3.2 Les stations urbaines associées.....	14
3. METHODE DE COMPARAISON	14
4. IMPACTS DES AJUSTEMENTS SUR LES MOYENNES 24 H	15
4.1 PENTES DE REGRESSION ESTIMATEURS DE FACTEURS DE CORRECTION.....	15
4.2 IMPACT SUR LES COMPARAISONS DES AJUSTEMENTS PAR « DELTA 4H »	17
4.3 IMPACT DES AJUSTEMENTS PAR « DELTA » 1H / 4H / 24H SUR LES REGRESSIONS	19
4.4 IMPACT DES AJUSTEMENTS SUR LES DEPASSEMENTS DU SEUIL JOURNALIER REGLEMENTAIRE DE 50 µG.M ⁻³	22
4.4.1 Impact du « delta 4h » pour l'ensemble des stations	22
4.4.2 Impact du choix de « delta » 1h / 4h / 24h.....	22
5. DIFFERENCES ENTRE LES TYPOLOGIES.....	23
5.1 LES STATIONS TRAFIC ET URBAINES SEMBLABLES	23
5.2 LES STATIONS TRAFIC ET STATIONS URBAINES DIFFERENTES.....	24
5.3 LES STATIONS EQUIPEES DE MATERIELS DIFFERENTS.....	26
6. AUTRES COMPARAISONS	27
6.1 DIFFERENCES ENTRE SAISONS CHAUDE ET FROIDE	27
6.2 DIFFERENCES ENTRE JOURS OUVRES ET NON OUVRES	29
7. CONCLUSIONS	30
ANNEXE 1 - DOCUMENT DE REFERENCE DE L'ETUDE.....	34
ANNEXE 2 – MOYENS STATISTIQUES.....	36

RESUME DE L'ETUDE EMD 2008

EXPLOITATION DE DONNEES PM Etude de l'impact du trafic routier sur les teneurs en PM à partir de la base de données PM10 et PM2.5

Jean-Luc HOUDRET
houdret@ensm-douai.fr - tél : 03 27 71 26 13

L'objet de cette étude est la détermination de la part prise par le trafic automobile par rapport aux stations urbaines dans les évolutions spatio-temporelles des PM10 à partir des données d'un peu moins de 400 stations de mesure, dont seulement 46 sont de typologie « trafic ». Pour 8 AASQA, il n'existe pas dans la BDQA de données PM10 issues de stations de typologie trafic, ce qui ne signifie pas forcément l'inexistence de stations. Chacune de ces 46 stations a été comparée à la station urbaine la plus représentative dans le voisinage le plus proche possible, à l'aide de l'étude des corrélations et des régressions linéaires orthogonales.

Une question préliminaire s'est posée d'office : « quelle base de données brutes ou ajustées faut-il utiliser? ». Aussi, l'impact de l'ajustement des données a été considéré dans la première phase de l'étude. Il apparaît d'emblée un manque d'homogénéité dans la base ajustée : différences de modes de calcul, différences de types de données communiquées.

La première partie de l'étude a montré que les ajustements des mesures de toutes les stations, modifient d'une manière significative les droites de régression : pentes des droites nettement abaissées et corrélations augmentées. Plusieurs résultats ont conduit à s'interroger sur le bien-fondé d'un ajustement basé sur une heure ou sur 24 heures. On a constaté avec les ajustements à l'aide des écarts entre FDMS et TEOM moyennés sur 4 heures « delta 4h » que le nombre total de dépassements du seuil $50 \mu\text{g.m}^{-3}$ a augmenté de 723 à 1955.

Plusieurs conséquences sont apparues : simplifier et harmoniser la procédure d'ajustement, ne l'appliquer qu'aux données journalières, à l'aide d'une station de référence représentative de l'ensemble des autres stations qu'elle est censée ajuster, de mêmes typologies et de niveaux d'exposition assez proches, car on sous-ajuste les stations trafic quand on utilise un « delta » issu d'une station urbaine, comme c'est en pratique presque toujours le cas. La suite de l'étude n'a porté que sur les données brutes, d'autant plus que plusieurs stations sont équipées de jauges Bêta classiques ou RST pour lesquelles les ajustements étaient bien-entendu sans objet.

Chaque station trafic a été comparée à une seule station urbaine soigneusement choisie. Il est apparu que 15 d'entre elles semblent ne pas subir d'influence marquée d'origine automobile et pourraient être considérées comme urbaines.

Pour 22 stations de typologie trafic plus marquée, on a constaté que :

- des pentes élevées des régressions linéaires ne sont pas forcément liées à des niveaux de PM10 importants,
- les niveaux vont du simple au double par rapport aux niveaux urbains,
- les corrélations supérieures à 0,90 indiquent une homogénéité temporelle des variations des niveaux entre trafic et urbain (cf notions de zones d'homogénéité – rapport LCSQA - J.-L. HOUDRET, Equivalence et représentativité des méthodes de surveillance des particules – Partie 2 : Représentativité, étude LCSQA-EMD, 2005),
- les dépassements du seuil $50 \mu\text{g.m}^{-3}$ sont compris de 1 à 134 en situation trafic, et de 0 à 9 en situation urbaine.

Enfin, dans plusieurs villes où les deux stations ne sont pas équipées d'appareils de même technologie, les comparaisons étaient difficiles à interpréter. Cela laisse à penser qu'il est plus cohérent de disposer au sein d'une même ville (ou zone) de groupes homogènes d'appareils en terme de technologie analytique.

On a constaté que la proximité automobile primant sur le fond urbain n'est pas un phénomène essentiellement hivernal, et n'est pas un phénomène uniquement lié aux jours ouvrés

La trop grande diversité de situations de proximité automobile, affecte à chacune des stations concernées une carte d'identité peu ou pas comparable avec d'autres stations de même typologie, alors que les situations urbaines sont globalement plus homogènes.

Il devient de plus en plus intéressant par voie de conséquence, de rendre plus précise la définition du réseau de surveillance de la proximité trafic en resserrant les diverses variantes décrites dans le Guide Ademe, qu'il est prévu de réviser dans les prochains travaux 2009.

Exploitation des données PM

Etude de l'impact du trafic routier sur les teneurs en PM à partir de la base de données PM10

1. OBJECTIF DE L'ETUDE

La détermination de la part prise par le trafic automobile dans les évolutions spatio-temporelles des PM10 proposée dans cette étude comporte :

- La validation de la base de données horaires issue du parc de stations existant en 2007,
- Les réflexions liées aux ajustements des données pratiquées à l'aide des appareils FDMS,
- La comparaison pour l'ensemble du parc national des concentrations moyennes journalières entre les stations de proximité automobile et les stations de fond à l'aide des régressions linéaires et corrélations,
- La mise en évidence des écarts rencontrés entre ces deux typologies de stations,
- L'étude des écarts observés entre les saisons froides et chaudes.
- L'étude des écarts observés entre les jours ouvrés et non ouvrés.

Par extension, on vérifiera la réalité de l'appellation « trafic » attribuée à certaines stations à l'aide des écarts observés entre les deux typologies « urbaine » et « trafic ».

2. VALIDATION DE LA BASE DE DONNEES HORAIRES 2007

2.1 BILAN GENERAL

Les données utilisées proviennent de la base obtenue auprès de l'ADEME dépendant des données fournies par les AASQA ; aussi, elle peut ne pas refléter totalement le parc existant en 2007 si des données n'ont pas été communiquées.

Ces données horaires fournies sur CD au format texte, ont due être converties au format Excel, triées par station puis par ordre chronologique.

Les données concernent un peu moins de 400 stations de mesure des PM10, soit :

- 297 stations équipées de TEOM50 tous ajustés à l'aide de FDMS, dont :
 - 267 pour lesquels les données brutes et ajustées à l'aide de FDMS situés dans d'autres stations, ont été communiquées,
 - 30 (22+8) pour lesquels les données brutes et ajustées à l'aide de FDMS dans la même station dite de référence, ont été communiquées,

- 34 autres stations de référence équipées d'un FDMS, dont les données ont été communiquées avec seulement les données brutes des TEOM50 associés, pas les données ajustées,
- 21 FDMS ou Bêta-RST de référence dont les données ont été communiquées seules, sans celles d'autres appareils associés.

La majorité des AASQA a donc communiqué les données TEOM brutes et ajustées, certaines n'ont communiqué que les données ajustées sans les données brutes, soit avec ou sans les données des FDMS associés quand ils existent, et d'autres enfin n'ont communiqué que les données FDMS de référence sans celles des TEOM associés. Ainsi, il n'a pas été possible de vérifier pour toutes les AASQA le mode d'ajustement pratiqué. Le Tableau 1 donne le détail des modalités des données obtenues selon les stations ; on peut donc noter **un manque d'homogénéité**.

Tableau 1 : détail des listes de données obtenues

	Nombre de Stations	Appareils PM10	Données Brutes	Données ajustées	Données de référence
297	266	TEOM50	+	+	Dans autres stations
	<u>22</u>	TEOM50	+	+	+
	<u>8</u>	TEOM50	nc	+	+
	1	TEOM50	nc	+	Dans autre station
	34	TEOM50	+	nc	+
21	10	FDMS seuls	nc	nc	+
	11	Bêta-RST seuls	Sans objet	Sans objet	+
	44	Bêta	+	Sans objet	Sans objet

D'autre part, plusieurs appareils ont eu un taux de fonctionnement inférieur à 50% pour diverses raisons : TEOM et / ou FDMS installés en cours d'année, appareils déplacés ou arrêtés ; il s'agit de :

- 15 appareils classiques TEOM ou Bêta,
- 31 TEOM ajustés,
- références FDMS ou RST.

On ne dénombre que 43 stations PM10 où les PM2.5 sont également mesurées.

2.2 AJUSTEMENT DES DONNEES

Le thème de l'ajustement des données n'entraîne pas à priori dans le domaine de cette étude, mais la question suivante s'est posée d'office : « **sur quelle base de données cette étude doit-elle porter, ajustée ou non ?** ». Aussi, l'impact de l'ajustement des données a été considéré.

2.2.1 Rappel du principe de l'ajustement

On rappelle que le FDMS est un TEOM30 (chauffé à 30°C) complété par un élément de séchage (nafion) dont la fonction est de pouvoir abaisser la température de 50°C à 30°C afin de limiter la perte d'éléments volatils, et d'un filtre total permettant d'évaluer le reste des pertes en éléments volatils non encore compensé par cet élément. Ainsi, le FDMS ne mesure réellement que 50% du temps, soit 6 minutes sur 12, en alternance avec le passage sur filtre total.

Le FDMS est placé dans une station déjà équipée d'un TEOM50 ; cette station dite de référence permet de calculer heure par heure les écarts entre les deux appareils ; ces écarts appelés « **delta** » sont ensuite appliqués aux autres appareils TEOM50 voisins afin d'ajuster leurs valeurs mesurées sur celles fournies par l'appareil FDMS de référence de la zone.

2.2.2 Modalités de calcul du « delta »

Il avait été acté en Commission de suivi PM que le « **delta** » d'ajustement serait la moyenne glissante sur 4 heures de 4 écarts horaires « FDMS – TEOM50 » rafraîchie toutes les heures ; notons les : « **delta 4h** » = moyenne de 4 « **delta 1h** ».

Il est apparu à la lecture des données que plusieurs modalités d'ajustement ont été utilisées. En effet, lorsque qu'un « **delta 1h** » (valeur horaire) est négatif, (ie lorsque la mesure brute du TEOM50 est restée supérieure à celle du FDMS associé) :

- certaines AASQA calculent le « delta 4h » avec les « delta 1h » tels qu'ils sont apparus, et d'autres AASQA avec des « deltas 1/4h » ;
- d'autres forcent à zéro ces « delta 1h » négatifs avant de calculer le « delta 4h » ;
- d'autres calculent le « delta 4h » avec les « delta 1h » tels qu'ils sont mais forcent à zéro le « delta 4h » s'il est négatif ;
- d'autres enfin n'avaient pas communiqué l'ensemble des informations « TEOM50 brutes + TEOM ajustées + FDMS de référence » ne permettant pas de déterminer le mode d'ajustement.

Comme il n'est pas rare de trouver des « delta 1h » négatifs, ces diversités d'ajustement posent problème.

2.2.3 Incidence de l'ajustement sur 4 heures

Chaque mesure horaire issue d'un TEOM50 est additionnée d'un « delta 4h » qui est par construction influencé par les 3 heures de mesure précédentes.

Il s'agit dans le premier exemple présenté dans la Figure 1 d'une séquence de 48 heures à la station urbaine 7008 de Clermont-Ferrand ajustée par la station de référence 7004. La figure fait apparaître les variations temporelles des « deltas 1h et 4h », du FDMS de référence, et des valeurs brutes, ajustées sur 1 heure et sur 4 heures.

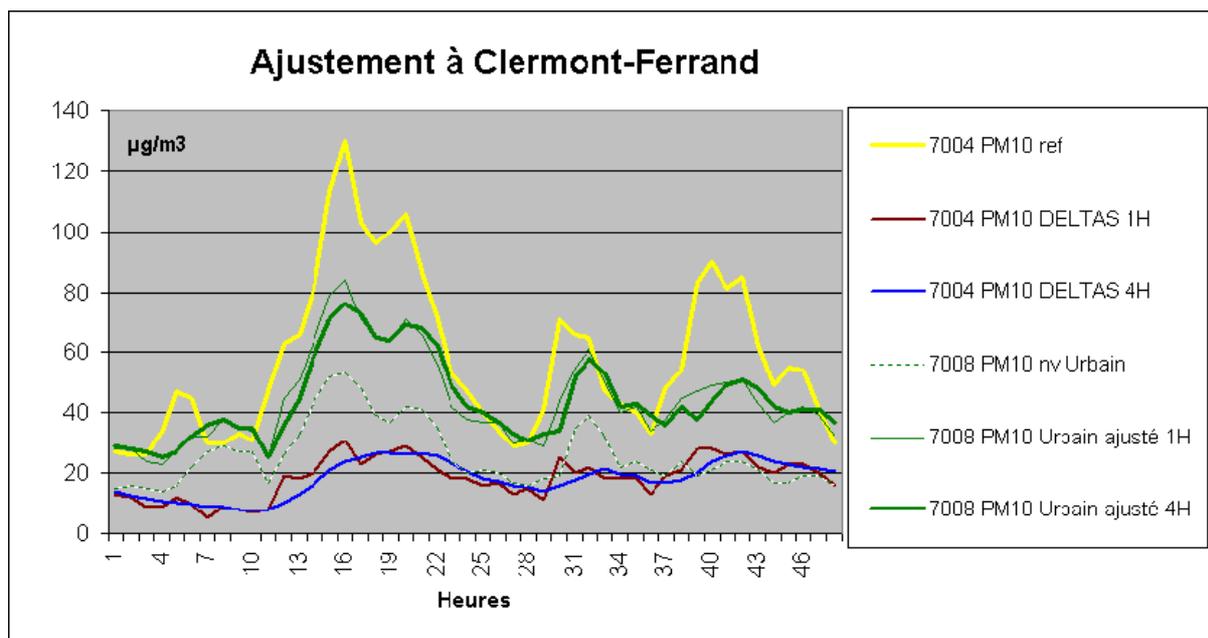


Figure 1 : Incidence de l'ajustement sur 4 h ou 1 h à Clermont-Ferrand (U)

Le second exemple présenté dans la Figure 2 présente la même problématique lors d'une séquence de 106 heures dans la station trafic 4031 à Paris.

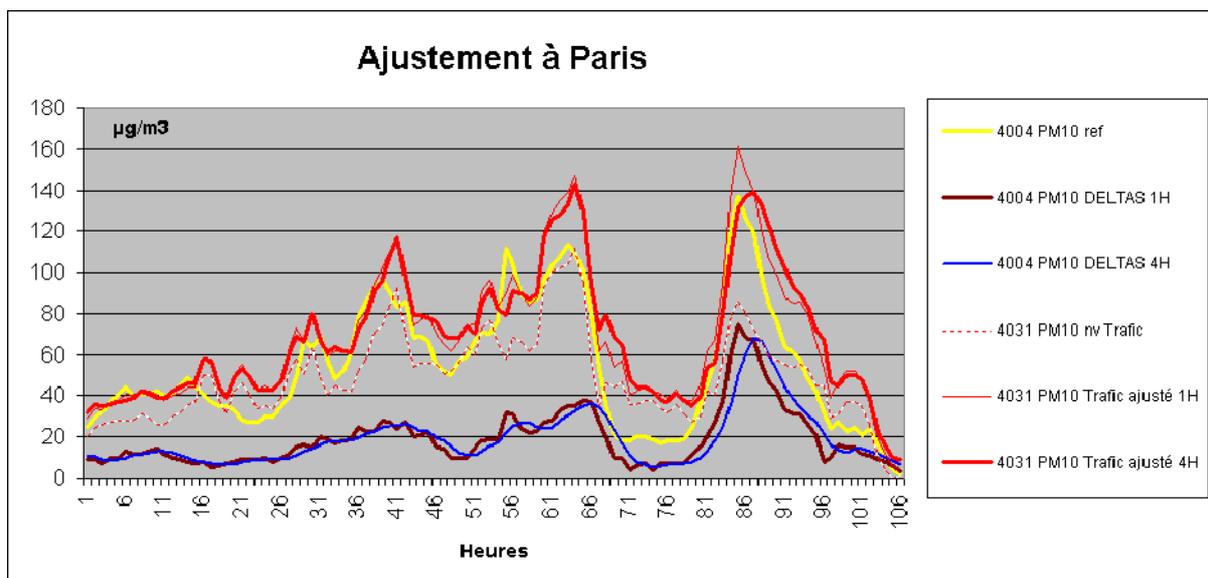


Figure 2 : Incidence de l'ajustement sur 4 h ou 1 h à Paris (T)

On distingue dans les deux cas un décalage temporel qui est différent selon que l'ajustement est réalisé par « delta 4h » et par « delta 1h », et que les données TEOM50 sont ainsi sous-corrigées lors des montées en pointe et pendant les sommets des pointes (effet des niveaux précédents plus faibles), puis sur-corrigées en descente et au retour en bruit de fond (effet des niveaux précédents plus élevés).

La conséquence immédiate est une **surestimation des valeurs horaires faibles et une sous-estimation des valeurs horaires fortes** et qui seront détaillées en 4.2 et 4.3 lors des comparaisons des données journalières entre les stations trafic et urbaines. Par contre, cela n'a pas d'incidence sur les moyennes 24h finales.

2.2.4 Influence des distances entre référence et autres stations

Bien que ce ne soit pas l'objet de ce rapport, il est nécessaire aussi de mentionner l'importance du choix de la station de référence qui servira à l'ajustement d'autres stations dans une même zone, ou dans d'autres zones éloignées ; il est préférable que **la référence soit représentative** de ces autres stations, autrement dit que toutes soient non seulement de même typologie mais qu'elles mesurent des niveaux assez proches de ceux de la référence.

Il est évident que plus les niveaux en composés semi-volatils sont élevés, et plus les pertes seront élevées ; si la représentativité évoquée ci-dessus n'est pas effective, cela peut entraîner aussi des **sur- ou sous-corrrections** des autres TEOM50 selon les niveaux mesurés par l'appareil FDMS de référence.

En effet, une station de référence exposée à des niveaux conséquents, fournira un « delta » trop élevé à une autre station plus faiblement exposée, et réciproquement. Ce point est encore plus crucial quand on sous-ajuste une station trafic à l'aide d'un « delta » issu d'une station urbaine, comme c'est **en pratique presque toujours le cas**.

2.3 LES STATIONS UTILISEES

2.3.1 Les stations trafic

Sur le parc d'environ 400 stations dont les données PM10 de 2007 sont disponibles dans la BDQA, seulement **46** sont de typologie trafic, et dans seulement **9** d'entre elles, les PM2.5 sont aussi mesurées. Il apparaît aussi qu'il n'existe pas dans la BDQA de données PM10 issues de stations de typologie trafic dans les AASQA suivantes :

Air c.o.m., ATMO PACA (2),
 ATMO Lorraine (Nancy), Atmo Picardie,
 Limair, QualitairCorse,
 et les AASQA - DOM.

Etant donné ce nombre restreint de stations trafic, elles ont toutes été prises en compte. Elles sont réparties dans 24 réseaux ; les équipements sont précisés dans le Tableau 2.

Tableau 2 : parc de stations trafic

Nombre	Appareil
37	TEOM50 PM10
dont 9	Avec TEOM50 PM2.5
1	FDMS
6	Bêta MP101M PM10
2	Bêta-RST PM10

Les données des 6 stations trafic équipées de jauges Bêta n'ont pas pu être correctement comparées avec des données issues de stations urbaines, car aucune de ces dernières n'utilise cette même technologie de mesure.

Les 2 stations Bêta-RST ne pouvaient être comparées qu'à des FDMS, mais la présence en une même ville de ces deux techniques n'existe pas.

2.3.2 Les stations urbaines associées

La détermination de l'influence du trafic sur les mesures, est effectuée par comparaison avec des stations de typologie urbaine.

Les stations urbaines exploitables ont été recensées, sur la base de la représentativité refaite selon la technique des bassins d'air, avec des données en nombre suffisant. Puis chaque station trafic a ensuite été comparée avec toutes les stations urbaines proches possibles, lorsqu'il y a plusieurs possibilités. Toutes les comparaisons réalisées apparaissent dans les tableaux de résultats du chapitre 3.

Il apparaît que les stations trafic ne peuvent être associées qu'aux stations urbaines avec lesquelles des corrélations temporelles fortes apparaissent pour que les résultats des comparaisons puissent être exploitables.

Pour la partie de l'étude du chapitre 4, chaque station trafic ne sera comparée qu'à la station urbaine avec laquelle les statistiques sont les plus proches des critères définis en 3, et considérées comme prioritaires par rapport aux distances inter stations qui n'ont donc pas été considérées.

3. METHODE DE COMPARAISON

La méthode de comparaison consiste après avoir assemblé des séries de données deux par deux, d'estimer les paramètres statistiques tels que définis dans l'annexe 2, la corrélation, et la pente de la droite de régression orthogonale. Les valeurs trouvées chacun des trois paramètres pris individuellement sont réparties en 4 classes comme indiqué dans le Tableau 3.

L'ordonnée à l'origine ne sera utilisée que dans quelques cas particuliers.

La méthode s'appliquera selon les thèmes abordés ci-après, aux comparaisons suivantes :

- FDMS vs TEOM-brut ,
- TEOM-ajusté vs TEOM-brut ,

entre stations trafic, puis entre stations urbaines,

- puis TEOM-Trafic vs TEOM-urbain , pour les données brutes, puis ajustées.

Le même outil s'appliquant à des comparaisons diverses, les interprétations seront adaptées à chaque type de comparaison.

Tableau 3 : classification des paramètres statistiques

	Classe	Corrélation	Pente	Ordonnée origine
LEGENDE	1	> 0,90	0,95 à 1,05	< 1
	2	0,90 à 0,85	1,05 à 1,10 et 0,90 à 0,95	1 à 2,5
	3	0,85 à 0,80	1,10 à 1,20 et 0,80 à 0,90	2,5 à 5
	4	< 0,80	< 0,80 et > 1,20	> 5

Les comparaisons les plus significatives sont obtenues lorsque les pentes sont élevées (classes 3 et 4) avec des corrélations élevées (classes 1 et 2) ; cela correspond à des situations où par exemple, les stations trafic mesurent des niveaux significativement plus élevés que les stations urbaines, mais avec des variations temporelles simultanées.

Des pentes de classe 1 indiquent que les stations trafic mesurent pratiquement les mêmes niveaux que les stations urbaines, surtout si les corrélations sont élevées. Des pentes de classe 2 ne permettent pas d'affirmer que les stations sont significativement différentes, et sont considérées comme « équivalentes » selon le concept décrit dans la norme EN 12341.

Des corrélations faibles de classes 3 et 4, correspondent à des dispersions élevées des points qui rendent difficiles les conclusions, et qui laissent supposer que d'autres influences sur les stations existent.

4. IMPACTS DES AJUSTEMENTS SUR LES MOYENNES 24 H

4.1 Pentés de régression estimateurs de facteurs de correction

En premier lieu, le Tableau 4 présente sous la forme de régressions linéaires orthogonales « **FDMS vs TEOM-brut** », à quelles corrections globales sont soumises les données brutes des TEOM dans les stations urbaines de référence où sont présents les deux types d'appareils, et à quels ajustements globaux par transfert de deltas sont soumises les données brutes des TEOM des stations trafic dans les mêmes zones, soit : **TEOM-ajusté vs TEOM-brut**.

Note : cela n'a été réalisé seulement que dans les villes où existent des stations de référence ainsi équipées et dont les données nécessaires ont été communiquées, soit 32 sites, excluant ainsi les stations équipées de jauges Bêta.

Tableau 4 : pentes de régression ou facteurs de correction

Sta Ref Urbaine / Sta Trafic	FDMS vs TEOM-brut en station Urbaine			TEOM-ajusté vs TEOM-brut en station trafic		
	Corrél	Pente	Ord Orig	Corrél	Pente	Ord Orig
Thionville 1020 / Thionville 1003	0.91	1.10	1.9	1.00	1.04	1.0
Marseille 3043 / Aix 3021	0.93	1.18	1.6	0.92	1.08	4.1
Marseille 3043 / Toulon 3068	0.93	1.18	1.6	0.97	1.08	3.2
Marseille 3043 / Marseille 3002	0.93	1.18	1.6	0.93	1.17	1.0
Nimes 8614 / Nimes 8616	0.87	1.21	-4.3	0.88	1.26	-0.8
Biarriz / Anglet	0.90	1.28	-2.9	0.91	1.22	-3.4
Montferrand / Clermont.F. 7034	0.90	1.33	-2.9	0.90	1.59	-6.0
Toulouse Berth / Toulouse.Ccit	0.91	1.48	-5.8	0.93	1.39	-5.1
St-Etienne C. / St-Etienne RdP	0.96	1.59	-3.9	0.84	2.00	-7.3
Salaise 27008 / Salaise 27008	0.95	1.60	-1.0	0.93	1.60	-1.8
Lyon Centre / Grenoble	0.94	1.61	-6.8	0.95	1.39	-1.5
Lyon Centre / Villeurbanne	0.94	1.61	-6.8	0.95	1.40	-0.4
Lyon Centre / Lyon La Mul.	0.94	1.61	-6.8	0.93	1.47	-5.3
Lyon Centre / Lyon Garib.	0.94	1.61	-6.8	0.91	1.64	-7.1
Lyon Centre / Lyon Berth.	0.94	1.61	-6.8	0.90	1.72	-5.1
Montbéliard / Audincourt	0.94	1.66	-5.3	0.95	1.53	-3.0
Montbéliard / Belfort	0.94	1.66	-5.3	0.93	1.56	-4.0
Gennevilliers / St-Denis	0.95	1.67	-6.1	0.93	1.37	-10.5
Paris 18 / Paris 8	0.95	1.70	-6.7	0.95	1.41	-6.6
Paris 18 / Paris 16	0.95	1.70	-6.7	0.96	1.48	-11.7
Paris 18 / Paris 14	0.95	1.70	-6.7	0.95	1.58	-13.2
Vaulx-en-Velin / Echirolles	0.89	1.70	-5.8	0.94	1.48	-3.5
Sotteville / Rouen LC	0.95	1.73	-6.6	0.93	1.67	-10.7
Gravelines / Boulogne	0.88	1.73	-3.4	0.87	1.82	-10.1
Gravelines / Dunkerque M.	0.88	1.73	-3.4	0.89	1.87	-9.4
Le Havre EH / Le Havre Rep	0.92	1.75	-9.8	0.94	1.67	-12.1
Strasbourg N / Strasbourg CI	0.93	1.77	-6.7	0.94	1.55	-2.8
Tourcoing / Valenciennes	0.94	1.82	-8.5	0.93	1.68	-7.0
Tourcoing / Douai	0.94	1.82	-8.5	0.92	1.80	-6.8
St Brieux / Rennes	0.86	1.82	-8.9	0.80	1.87	-13.5
Joué-les-Tours / Tours	0.91	1.93	-8.9	0.91	1.83	-7.2
Rezé / Nantes	0.89	2.15	-13.8	0.88	1.69	-11.4
MEDIANES						
	0.94	1.66	-6.3	0.93	1.55	-5.6

Il apparaît dans ces 32 sites après ajustement par « delta4h » que :

- Les corrélations entre données brutes et ajustées sont élevées et de manière identique dans les stations de référence urbaine (sources des delta) et dans les stations trafic (transfert de delta) ; ce critère statistique est significatif ;
- Les **pent**es sont en majorité **inférieures** dans les stations trafic par rapport aux stations de référence associées, (médianes respectives 1,55 et 1,66), laissant à penser que l'ajustement est peut être sous-évalué en station trafic ;
- Les **ordonnées** à l'origine sont presque toutes **né**gatives dans les deux cas, cela est en partie dû à l'offset programmé d'origine sur le TEOM : $1.03x + 3$.

4.2 Impact sur les comparaisons des ajustements par « delta 4h »

Ce paragraphe est destiné à montrer que les **ajustements** des mesures issues de toutes les stations trafic et urbaines, **modifient fortement les pentes** des droites de régression « trafic vs urbaine », pour les raisons déjà évoquées en 2.2.3.

L'impact des ajustements a été testé sur tous les couples de stations « trafic vs urbaine » possibles, mais le Tableau 5 ne présente que les comparaisons qui ont donné à la fois les corrélations et les pentes les plus élevées, donc les plus significatives, soit :

- Colonnes 1 à 4 : les couples de stations «Trafic vs Urbain »,
- Colonnes 5 et 6 : les corrélations et les pentes des droites de régression établies à partir des données brutes,
- Colonnes 7 et 8 : les corrélations et les pentes des droites de régression établies à partir des données ajustées avec les delta établis sur 4 heures glissantes.

Note : le tri dans ce tableau est effectué sur les pentes croissantes brutes en colonne 6, et les codes BDQA **en gras** correspondent aux TEOM50 situés dans des stations de référence avec FDMS.

La Figure 3 et la Figure 4 présentent ces résultats sous une forme graphique où l'on voit qu'avec l'ajustement, les corrélations augmentent et passent presque toutes en classe 1 (selon les codes du Tableau 3) et les pentes des droites sont nettement abaissées, mais d'une manière variable selon les sites, le même s'appliquant à des comparaisons diverses, les interprétations sont adaptées à chaque mode de comparaison.

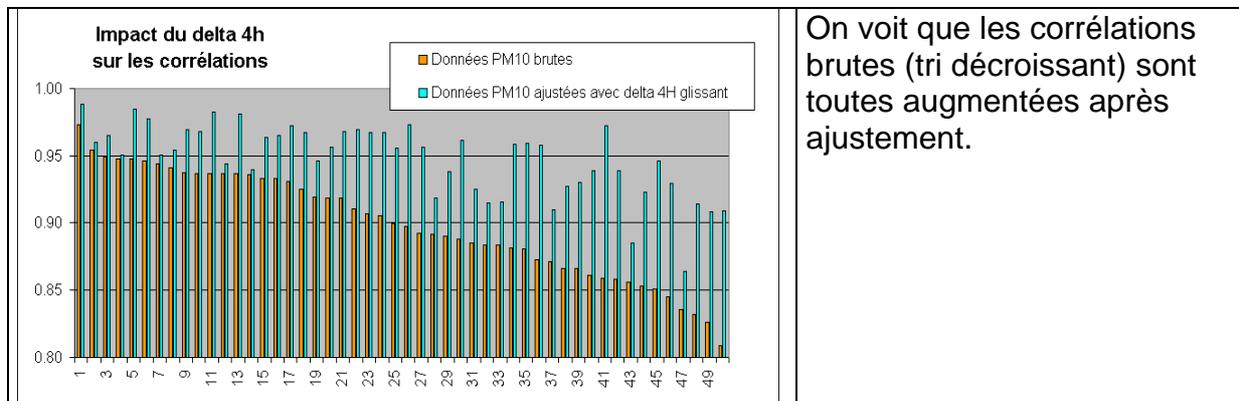


Figure 3 : Impact sur les corrélations «Trafic vs Urbain »

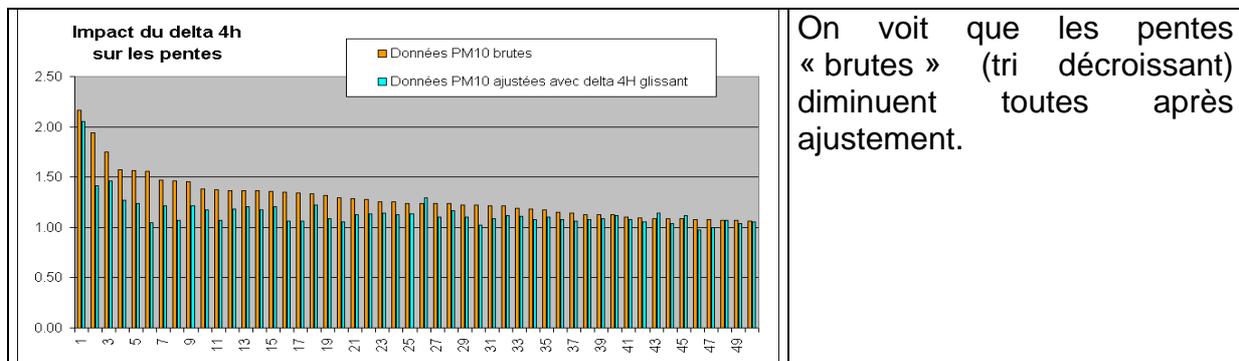


Figure 4 : Impact sur les pentes «Trafic vs Urbain »

Tableau 5 : Impact des ajustements par delta moyen sur 4 heures

Ville <small>Indique si sta de Réf</small>	BDQ A Trafic <small>(si Réf)</small>	Ville <small>Indique si sta de Réf</small>	BDQA Urbain <small>(si Réf)</small>	Données PM10 brutes		Données PM10 ajustées avec delta 4h glissant	
				Corr.	Pente	Corr.	Pente
MARSEILLE	3002	MARSEILLE	3047	0.89	1.12	0.92	1.12
LE HAVRE	5077	LE HAVRE	5074	0.91	1.13	0.97	1.08
NIMES	8616	NIMES	8614	0.95	1.13	0.96	1.08
AUDINCOURT	13005	MONTBELIARD	13007	0.94	1.14	0.97	1.06
TOURS	34022	TOURS	34025	0.95	1.15	0.98	1.08
PARIS-14E	4012	BOBIGNY	4156	0.92	1.18	0.97	1.10
BOULOGNE-SUR-MER	10031	BOULOGNE-SUR-MER	10032	0.90	1.18	0.97	1.07
LYON	20013	VAULX-EN-VELIN	20046	0.83	1.19	0.91	1.11
BORDEAUX	31002	FLOIRAC	31004	0.93	1.21	0.96	1.12
LOMME	11029	LILLE	11025	0.94	1.21	0.98	1.09
RENNES	19002	RENNES	19005	0.87	1.22	0.96	1.03
PARIS-14E	4012	PARIS- 1ER	4055	0.93	1.22	0.97	1.10
FORBACH	22012	FREYMING-MERLEBACH	22004	0.94	1.23	0.94	1.17
ECHIROLLES	15039	GRENOBLE	15043	0.88	1.24	0.92	1.10
TOULON	3068	TOULON	3062	0.84	1.24	0.86	1.30
STRASBOURG	16034	STRASBOURG	16029	0.95	1.24	0.98	1.13
LE HAVRE	5077	LE HAVRE	5064	0.91	1.25	0.97	1.13
REIMS	14009	REIMS	14002	0.97	1.25	0.99	1.14
PARIS-16E	4053	BOBIGNY	4156	0.86	1.28	0.94	1.14
CLERMONT-FERRAND	7034	CLERMONT-FERRAND	7009	0.89	1.28	0.96	1.13
NANTES	23140	NANTES	23107	0.85	1.29	0.95	1.06
SAINT-DENIS	4058	BOBIGNY	4156	0.83	1.32	0.91	1.08
ANGLET	31017	BIARRITZ	31018	0.92	1.33	0.95	1.22
SAINT-DENIS	4058	PARIS- 1ER	4055	0.81	1.34	0.91	1.06
NIORT	9019	NIORT	9020	0.90	1.35	0.96	1.06
LYON	20013	LYON	20062	0.86	1.36	0.88	1.21
VALENCIENNES	6006	DENAIN	6011	0.89	1.36	0.96	1.17
ECHIROLLES	15039	FONTAINE	15017	0.89	1.37	0.92	1.20
VALENCIENNES	6006	VALENCIENNES	6001	0.88	1.37	0.96	1.18
CLERMONT-FERRAND	7034	CLERMONT-FERRAND	7008	0.91	1.38	0.97	1.07
GRENOBLE	15046	SAINT-MARTIN-D'HERES	15038	0.88	1.38	0.92	1.18
PARIS-16E	4053	PARIS- 1ER	4055	0.86	1.45	0.94	1.22
SOTTEVILLE-LES-ROUEN	25043	ROUEN	25036	0.94	1.46	0.98	1.07
PARIS- 8E	4031	BOBIGNY	4156	0.87	1.47	0.93	1.21
ROUEN	25037	ROUEN	25036	0.86	1.56	0.97	1.05
PARIS- 8E	4031	PARIS- 1ER	4055	0.87	1.57	0.93	1.23
GRENOBLE	15046	GRENOBLE	15043	0.87	1.57	0.91	1.27
VILLEURBANNE	20019	LYON	20062	0.94	1.75	0.94	1.46
NANTES	23140	REZE	23179	0.68	1.75	0.85	1.12
VILLEURBANNE	20019	VAULX-EN-VELIN	20046	0.85	1.94	0.92	1.42
THONVILLE	1003	THONVILLE	1019	0.95	2.16	0.95	2.05

Exemple d'application : la Figure 5 présente un exemple de l'influence de l'ajustement sur un des couples de stations pris dans le Tableau 5 :
LE HAVRE (5077 T) vs LE HAVRE (5064 U).

On constate une réduction de la pente de la droite de régression, une augmentation de l'ordonnée à l'origine et de la corrélation. Ces influences sont spécifiques de chaque site, mais sont toutes dans le même sens.

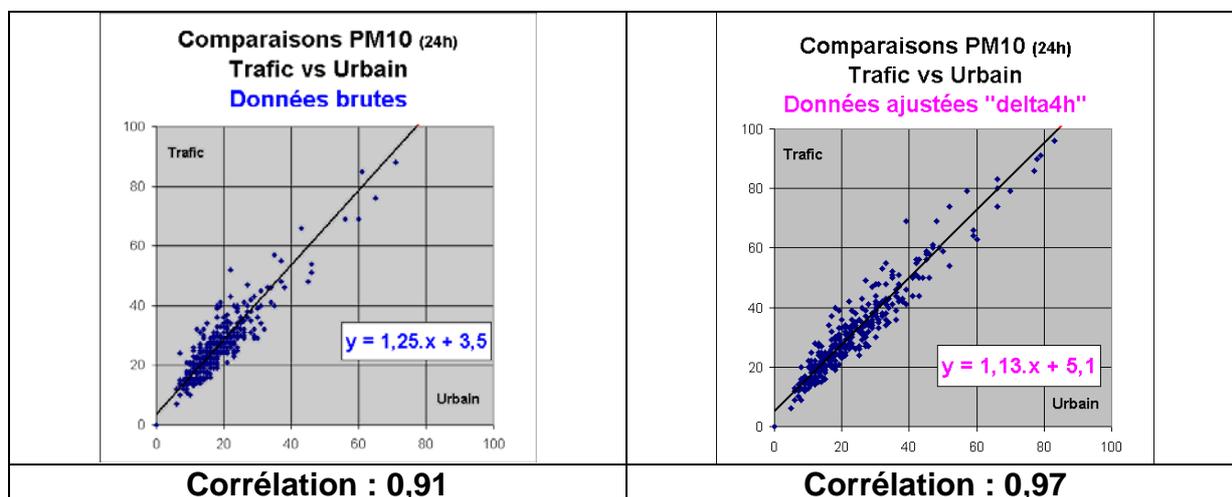


Figure 5 : impact de l'ajustement sur les comparaisons des données PM10 « Trafic vs Urbain » en $\mu\text{g}/\text{m}^3$

4.3 Impact des ajustements par « delta » 1h / 4h / 24h sur les régressions

Le problème évoqué en 2.2.3 (influence des 3 heures précédentes) a conduit à s'interroger sur le bien-fondé d'un ajustement différent, soit par delta horaire, ou par delta global sur 24 heures.

NOTE : il convient d'éviter un calcul de deltas à partir des données quart-horaires car ils sont calculés en fait toutes les 12 minutes (voir principe du FDMS), soit 5 fois par heure et donc en déphasage par rapports aux 4 quarts d'heure correspondants.

Le Tableau 6 et les figures associées (6 à 8) tirées de ce tableau montrent que l'ajustement :

- Quel qu'il soit, augmente pour chaque station les corrélations d'une manière identique, les données sembleraient donc améliorées ;
- Quel qu'il soit, réduit pour chaque station les pentes, surtout pour des delta 4h, **un peu moins** pour les delta calculés sur **1 ou 24 heures**, lesquels ont le même effet, car il semblerait que les données trafic soient sous-ajustées, pas les données urbaines, ce qui mathématiquement fait baisser la pente $T = f(U)$.

Tableau 6 : impact du type de calcul du delta (1h, 4h, 24h)

				A partir des moyennes 24h				
Ville	Ville	BDQA Trafic	BDQA Urbain		brutes	ajustées delta 4h	ajustées delta 1h	ajustées delta 24h
PARIS	PARIS	4031	4055	Corrélations	0.87	0.93	0.93	0.92
ROUEN	ROUEN	25037	25036		0.86	0.97	0.95	0.95
PARIS	PARIS	4053	4055		0.86	0.94	0.94	0.94
CLERMONT-F	CLERMONT-F	7034	7008		0.91	0.97	0.96	0.96
NIORT	NIORT	9019	9020		0.90	0.96	0.96	0.97
LE HAVRE	LE HAVRE	5077	5064		0.91	0.97	0.97	0.97
STRASBOURG	STRASBOURG	16034	16029		0.95	0.98	0.98	0.98
PARIS	PARIS	4031	4055	Pentes	1.57	1.23	1.22	1.23
ROUEN	ROUEN	25037	25036		1.56	1.05	1.17	1.18
PARIS	PARIS	4053	4055		1.45	1.22	1.22	1.22
CLERMONT-F	CLERMONT-F	7034	7008		1.38	1.07	1.14	1.14
NIORT	NIORT	9019	9020		1.35	1.06	1.15	1.15
LE HAVRE	LE HAVRE	5077	5064		1.25	1.13	1.12	1.13
STRASBOURG	STRASBOURG	16034	16029		1.24	1.13	1.13	1.13
PARIS	PARIS	4031	4055	Ord.Orig.	2.1	7.1	7.3	7.2
ROUEN	ROUEN	25037	25036		6.7	5.0	10.5	10.3
PARIS	PARIS	4053	4055		10.3	13.6	13.7	13.7
CLERMONT-F	CLERMONT-F	7034	7008		-0.5	2.8	2.0	2.1
NIORT	NIORT	9019	9020		-0.5	3.6	1.7	1.6
LE HAVRE	LE HAVRE	5077	5064		3.5	5.1	5.2	5.1
STRASBOURG	STRASBOURG	16034	16029		1.0	1.7	2.0	2.1

Note : les codes BDQA **en gras** correspondent aux TEOM50 situés dans des stations de référence avec FDMS.

- La première conséquence est que l'on peut simplifier la procédure d'ajustement en ne **l'appliquant qu'aux données journalières** pour le respect de la réglementation européenne.
- La seconde est que l'ajustement horaire n'est utile que pour des études réalisées sur cette base horaire, par exemple pour des comparaisons avec les mesures de gaz.
- La troisième conséquence est que l'on peut s'interroger sur la **pertinence de l'ajustement** des données horaires (et de l'usage qui en est fait) avec un delta sur 4 heures.
- La quatrième conséquence est que la suite de cette étude ne peut porter que sur les **données brutes** à cause des disparités et des hétérogénéités dans les ajustements pratiqués.

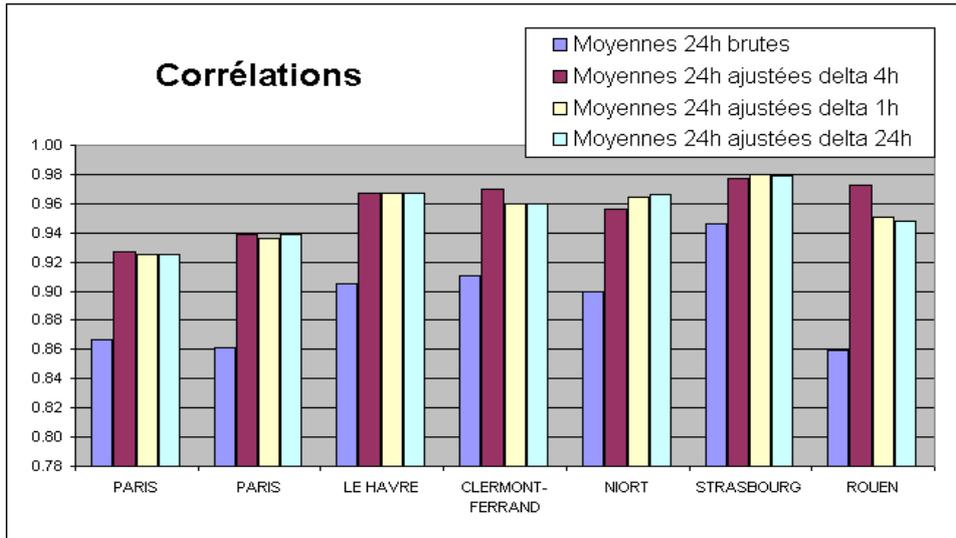


Figure 6 : impact sur les corrélations du calcul temporel du delta

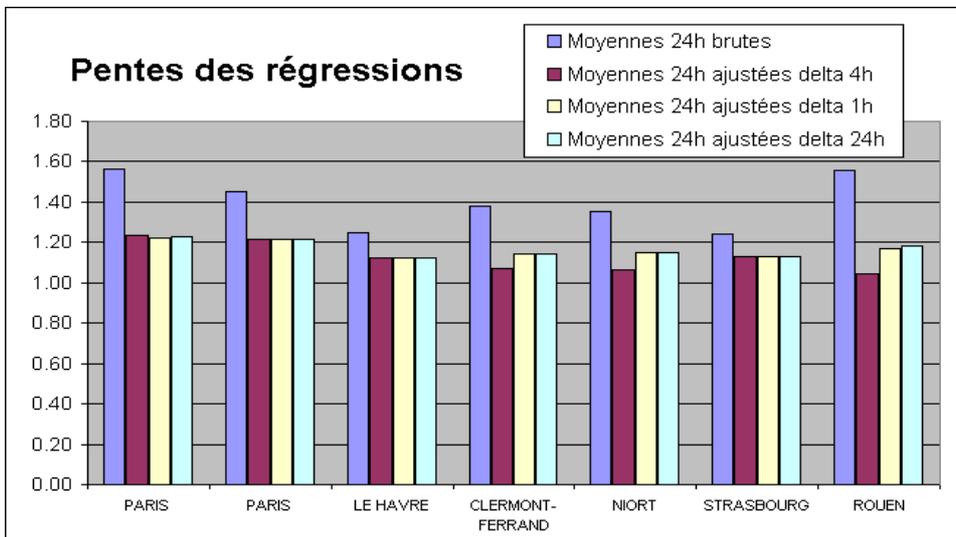


Figure 7 : impact sur les pentes du calcul temporel du delta

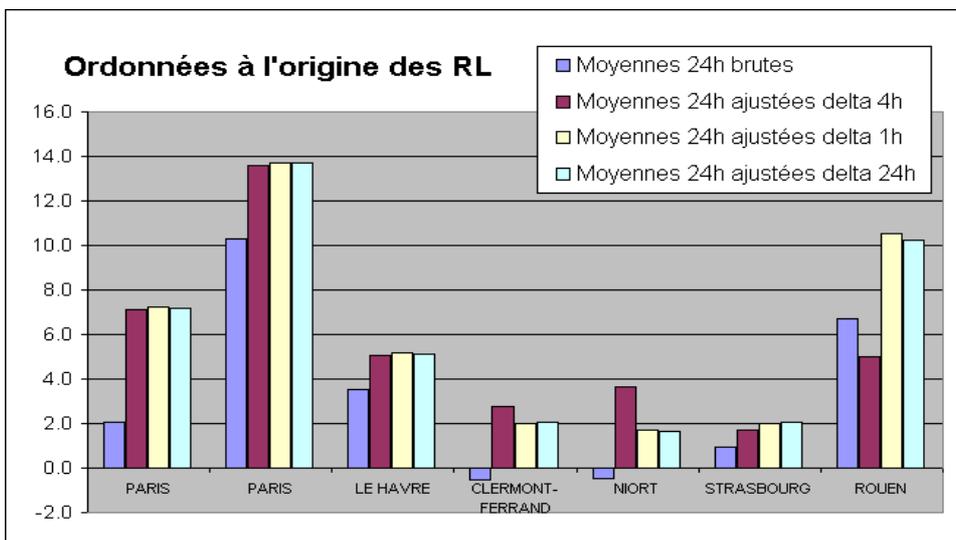


Figure 8 : impact du calcul temporel du delta

4.4 Impact des ajustements sur les dépassements du seuil journalier réglementaire de $50 \mu\text{g.m}^{-3}$

4.4.1 Impact du « delta 4h » pour l'ensemble des stations

Les stations équipées de jauges Bêta (6) et celle d'une jauge-RST n'ont pas pu être prises en compte ici. Le Tableau 7 montre que :

- Les majorations de nombres de dépassements sont bien plus élevées pour les stations de typologies rurale, urbaine et périurbaine, (de **x6** à **x5,5**) ;
- Pour l'ensemble des 39 stations « trafic » qui ont enregistré en données brutes au moins 1 (et jusqu'à 136) dépassement du seuil $50 \mu\text{g.m}^{-3}$, l'ajustement à l'aide des « delta 4h » a augmenté le nombre total de dépassements de 723 à 1955 (soit **x 2,7**) ;
- cela confirme que les ajustements déterminés en stations urbaines et appliqués aux stations trafic, **sous-estiment les dépassements** observés dans ces dernières.

Tableau 7 : Impact du « delta 4h » sur les nombres de dépassement de seuil

	nb de stations	nb dépassements bruts	nb dépassements ajustés	soit majoration
Rural	10	20	120	x 6.0
Urbain	177	624	3693	x 5.9
Périurbain	41	161	886	x 5.5
Observ Spécif.	2	11	37	x 3.4
Industriel	38	389	1175	x 3.0
Trafic	39	723	1955	x 2.7
GLOBAL	307	1928	7866	x 4.1

4.4.2 Impact du choix de « delta » 1h / 4h / 24h

Pour les mêmes stations déjà sélectionnées en 4.3 dans le Tableau 6 , le Tableau 8 ci-après donne les différentes modifications des nombres de dépassements en fonction du mode d'ajustement utilisé. On constate que l'usage des « delta 1h ou 24h » produit le même effet ; l'usage du « delta 4h » semblerait majorer un peu plus les nombres de dépassements.

Tableau 8 : impact des ajustements sur les dépassements de seuil

Données 24h	nb de dépassements
brutes	140
ajustées par delta 4h glissantes	363 (x 2,6)
ajustées par delta 1h	340 (x 2,4)
ajustées par delta 24h	342 (x 2,4)

5. DIFFERENCES ENTRE LES TYPOLOGIES

La suite de l'étude sur les différences entre les stations trafic et urbaines ne prend en considération que les données brutes (non ajustées).

Le choix de la station urbaine la mieux placée pour être associée à chaque station trafic a été effectué dans l'ordre des deux critères suivants lorsque plusieurs possibilités se présentent :

- une station urbaine de référence de la ville considérée dont les données FDMS et TEOM brutes sont disponibles,
- la station urbaine TEOM de la ville considérée la mieux corrélée avec les autres stations, et de niveau mesuré le plus faible pour s'affranchir d'influences d'autres proximités éventuelles.

En cas d'égalité de critères de choix, ce qui a été souvent le cas, la préférence a été donnée à la station la plus proche.

5.1 Les stations trafic et urbaines semblables

Dans le cadre des comparaisons, il convient en premier lieu de traiter les couples de stations où les niveaux mesurés sont faibles et très proches au sein des deux stations de chaque couple ; pour cela, avec une clé de tri décroissant des moyennes des stations trafic, la Figure 9 montre les moyennes urbaines associées ; on voit que pour des valeurs inférieures à environ $25 \mu\text{g}\cdot\text{m}^{-3}$, les stations trafic sont presque au même niveau que les stations urbaines dont les moyennes sont dans ce cas presque toutes inférieures à $20 \mu\text{g}\cdot\text{m}^{-3}$.

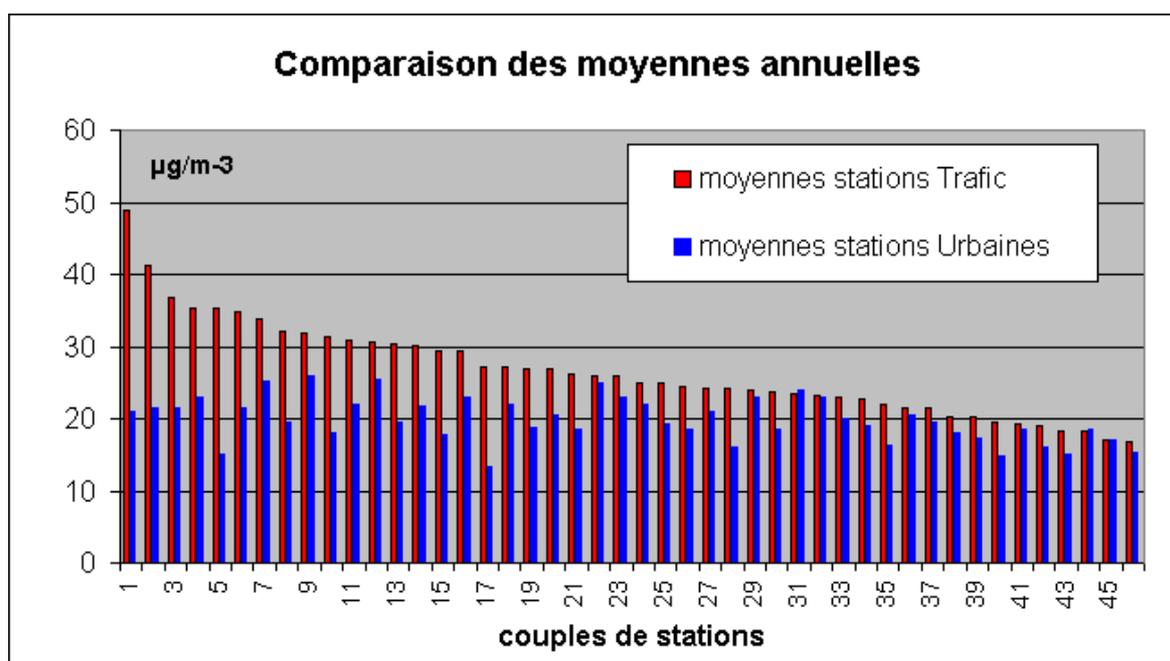


Figure 9 : comparaison des moyennes annuelles en PM10

Le Tableau 9 dresse la liste par ordre croissant des moyennes annuelles des stations classifiées trafic inférieures à $25 \mu\text{g.m}^{-3}$, seuil au-dessus duquel elles prédominent toujours. Elles sont assez peu (ou pas) exposées à une proximité sensible du trafic automobile, car elles présentent des moyennes du même ordre de grandeur pour les deux typologies trafic et urbaine.

D'autre part, les nombres de dépassements du seuil $50 \mu\text{g.m}^{-3}$ en site trafic n'excèdent pas **13**; le nombre de **16** dépassements à la station de St-Jean de Maurienne (classé urbaine) provient d'une situation estivale locale qui n'a pas fait l'objet d'investigation.

Les pentes des droites de régression comprises entre 0,92 et 1,15, associées à des corrélations presque toutes supérieures à 0,85, confirment les similarités entre ces stations dites trafic et les stations urbaines.

La comparaison entre une jauge Bêta et un TEOM (Bligny vs Dijon) est à considérer avec précaution étant donné les différences de technologies et le caractère autoroutier de la station de Bligny. La comparaison entre deux jauges Bêta surlignées dans le tableau (Chalon vs Champforgeuil) est par contre prise en compte.

Tableau 9 : les stations trafic et urbaines semblables

STATIONS TRAFIC			STATIONS URBAINES			Rég. Lin.		Moyennes annuelles		Nombre dépass	
Ville	BDQA	App.	Ville	BDQA U	App.	Corrél	Pente	Sta Traf	Sta Urb	Sta T	Sta U
TOURS	34022	Teom	TOURS	34025	Teom	0.95	1.15	17	15	1	1
AUDINCOURT	13005	Teom	MONTBELIARD	13007	Teom	0.94	1.14	18	18	6	3
BLIGNY-LES-BEA	26018	Bêta	DIJON	26005	Teom	0.87	1.03	18	15	0	0
SAINT-ETIENNE	29428	Teom	SAINT-ETIENNE	29424	Teom	0.92	0.95	19	16	1	1
BELFORT	13009	Teom	MONTBELIARD	13007	Teom	0.94	1.05	19	18	3	3
CHALON-SUR-S	32001	Bêta	CHAMPFORGEUIL	32002	Bêta	0.89	1.09	20	17	6	4
SALAISE-SUR-S	27008	Teom	ROUSSILLON	27002	Teom	0.94	0.98	22	20	6	7
DOUAI	6016	Teom	WAZIERS	6010	Teom	0.92	1.05	22	20	4	6
TOULOUSE	12032	Teom	TOULOUSE	12030	Teom	0.93	1.07	23	19	3	1
LYON	20002	Teom	LYON	20062	Teom	0.97	0.92	23	23	6	9
ST-JULIEN-MT-DE	33112	Teom	ST-JEAN-DE-MAUR	33111	Teom	0.89	1.09	23	24	10	16
LYON	20003	Teom	LYON	20062	Teom	0.96	1.02	24	23	13	9
DUNKERQUE	10002	Teom	Malo-les-Bains	10004	Teom	0.83	1.10	24	21	10	2
NIMES	8616	Teom	NIMES	8614	Teom	0.95	1.13	24	19	6	2
BOURG-LES-VAL	36003	Teom	PORTES-LES-VAL	36001	Teom	0.95	1.08	25	19	7	5

On peut considérer que ces **15 stations** de typologie trafic ne subissent pas d'influence marquée d'origine automobile et pourraient être **considérées comme urbaines**.

5.2 Les stations trafic et stations urbaines différentes

Le Tableau 10 présente le classement par ordre des pentes croissantes des régressions linéaires, les 22 comparaisons entre les deux typologies de stations, où sont mesurés des niveaux plus conséquents, et où apparaissent des différences des teneurs en PM10 significatives entre les deux typologies.

Tableau 10 : les stations trafic significativement différentes des stations urbaines

STATIONS TRAFIC			STATIONS URBAINES			Rég. Lin.		Moyennes annuelles		Nombre dépass	
Ville	BDQA	App.	Ville	BDQA	App.	Corrél	Pente	Sta Traf	Sta Urb	Sta T	Sta U
AIX-EN-PROV	3021	Teom	AIX-EN-PROV	3029	Teom	0.89	1.00	31	25	13	1
LYON	20063	Teom	LYON	20062	Teom	0.94	1.08	26	23	11	9
MARSEILLE	3002	Teom	MARSEILLE	3047	Teom	0.89	1.12	32	26	21	7
BOULOGNE / M	10031	Teom	BOULOGNE / MER	10032	Teom	0.90	1.18	27	20	12	4
RENNES	19002	Teom	RENNES	19005	Teom	0.87	1.22	22	16	1	1
PARIS-14E	4012	Teom	PARIS- 1ER	4055	Teom	0.93	1.22	37	22	37	3
ECHIROLLES	15039	Teom	GRENOBLE	15043	Teom	0.88	1.24	30	19	26	1
TOULON	3068	Teom	TOULON	3062	Teom	0.84	1.24	34	25	68	6
STRASBOURG	16034	Teom	STRASBOURG	16029	Teom	0.95	1.24	24	18	8	2
LE HAVRE	5077	Teom	LE HAVRE	5064	Teom	0.91	1.25	27	19	11	5
NANTES	23140	Teom	NANTES	23107	Teom	0.85	1.29	24	16	6	0
ROUEN	25037	Teom	ROUEN	25036	Teom	0.88	1.29	27	13	10	2
SAINT-DENIS	4058	Teom	BOBIGNY	4156	Teom	0.83	1.32	49	21	134	5
ANGLET	31017	Teom	BIARRITZ	31018	Teom	0.92	1.33	31	22	19	1
LYON	20013	Teom	LYON	20062	Teom	0.86	1.36	35	23	39	9
VALENCIENNES	6006	Teom	VALENCIENNES	6001	Teom	0.88	1.37	26	19	13	4
CLERMONT-F.	7034	Teom	CLERMONT-F.	7008	Teom	0.91	1.38	20	15	4	1
PARIS-16E	4053	Teom	PARIS- 1ER	4055	Teom	0.86	1.45	41	22	81	3
PARIS- 8E	4031	Teom	PARIS- 1ER	4055	Teom	0.87	1.57	35	22	26	3
GRENOBLE	15046	Teom	GRENOBLE	15043	Teom	0.87	1.57	32	19	42	1
VILLEURBANNE	20019	Teom	LYON	20062	Teom	0.94	1.75	29	23	43	9
THIONVILLE	1003	Teom	THIONVILLE	1019	Teom	0.95	2.16	31	18	10	15

On peut constater que :

- Des **pent**es élevées des régressions ne sont **pas forcément liées** à des niveaux de PM10 conséquents, comme par exemple le couple de stations de Airparif (moyennes les plus élevées) et le couple de ATMO-Auvergne (moyennes les plus faibles), de pentes voisines respectives **1,32** et **1,38** ;
- Les pentes des régressions s'étendent de **1** à **2,16** , c'est-à-dire que les stations trafic ont mesuré des niveaux pouvant aller du simple au double par rapport aux niveaux urbains,
- Presque toutes les corrélations sont dans les classes 1 ou 2 ; cela signifie qu'il y a une **homogénéité temporelle** des variations des niveaux entre les stations trafic et urbaines, due aux évolutions spatio-temporelles à l'échelle régionale (cf notions de zones d'homogénéité – rapport LCSQA - J.-L. HOUDRET, Equivalence et représentativité des méthodes de surveillance des particules – Partie 2 : Représentativité, étude LCSQA-EMD, 2005).
- Les nombres de dépassements du seuil $50 \mu\text{g.m}^{-3}$ s'étendent de 1 à 134 en situation trafic et de 0 à 9 en situation urbaine, exception faite du cas de Thionville, site « urbain ? » chargé (15 dépassements).

5.3 Les stations équipées de matériels différents

Le Tableau 11 présente les couples de stations où les méthodes de mesure sont différentes, ce qui rend difficile d'effectuer des comparaisons et de tirer des conclusions.

Tableau 11 : méthodologies différentes entre les stations

STATIONS TRAFIC			STATIONS URBAINES			Rég. Lin.		Moyennes annuelles		Nombre dépass	
Ville	BDQA	App.	Ville	BDQA	App.	Corrél	Pente	Sta Traf	Sta Urb	Sta T	Sta U
BESANCON	17011	Teom ajust	BESANCON	17004	FDMS	0.96	1.02	25	22	17	9
BREST	19014	FDMS	BREST	19012	Teom ajust	0.94	1.03	27	22	20	13
BORDEAUX	31003	Bêta RST	FLOIRAC	31004	Teom ajust	0.79	1.10	17	17	0	6
ORLEANS	34013	Bêta RST	ORLEANS	34012	Teom ajust	0.94	1.01	26	25	21	19
BETHUNE	28025	Bêta	BETHUNE	28028	Bêta RST	0.89	0.70	20	18	3	12
BORDEAUX	31005	Bêta	BORDEAUX	31001	FDMS	0.86	0.99	29	18	23	4
PAU	31015	Bêta	PAU	31013	Teom	0.85	1.45	30	22	22	3
DIJON	26014	Bêta	DIJON	26005	Teom	0.79	2.81	35	15	39	0
PASSY	33220	Teom	CHAMONIX	33232	Teom	0.39	1.75	23	20	a : 19/1	b : 3/8

a : 19 en hiver et 1 en été

b : 3 en hiver et 8 en été

A Besançon et à Brest, une des stations est équipée d'un FDMS, mais aucune donnée TEOM brute n'a été communiquée ; les comparaisons avec l'autre station de chaque couple ont donc été réalisées avec les données TEOM ajustées ; les écarts entre trafic et urbain sont significatifs en terme de moyennes et de nombres de dépassements, tout en présentant des pentes de régressions proches de l'unité avec en plus de très bonnes corrélations.

A Bordeaux et à Orléans, les stations trafic équipées de jauges Bêta-RST n'ont pu être comparées qu'avec des TEOM ajustés ; les deux stations de Orléans semblent donner les mêmes résultats ; les stations de Bordeaux (31003) et Floirac sont assez mal corrélées et sont difficiles à comparer.

Il en est de même pour les jauges Bêta de Béthune, Bordeaux (31005), Pau et Dijon, que l'on ne peut comparer qu'avec des appareils de technologies différentes.

Le dernier cas de « Chamonix vs Passy », n'est pas exploitable car les deux stations semblent être exposées très différemment et ne peuvent pas être soumises à comparaison.

Cela laisse à penser qu'il est plus aisé de disposer dans un même ville (ou zone) de groupes d'appareils **homogènes en terme de technologie analytique**.

6. AUTRES COMPARAISONS

6.1 Différences entre saisons chaude et froide

Sur l'ensemble des 36 comparaisons, les régressions linéaires ont été calculées en séparant les données estivales (avril à septembre) et hivernales (octobre à mars) afin de déterminer s'il existe des influences saisonnières sur les écarts entre les stations trafic et les stations urbaines, et dans le respect des choix des stations exprimés en tête du chapitre 5.

La Figure 10 présente deux groupes de villes, celles où :

- les pentes des régressions linéaires (stations trafic vs urbaine) en période estivale priment sur les pentes trouvées en période hivernale, c'est-à-dire là où le trafic prime sur l'urbain en été, (partie gauche de la figure : classement sur les pentes estivales décroissantes), ,
- les pentes hivernales priment sur les pentes estivales, là où le trafic prime sur l'urbain en hiver (partie droite de la figure : classement sur les pentes hivernales croissantes).

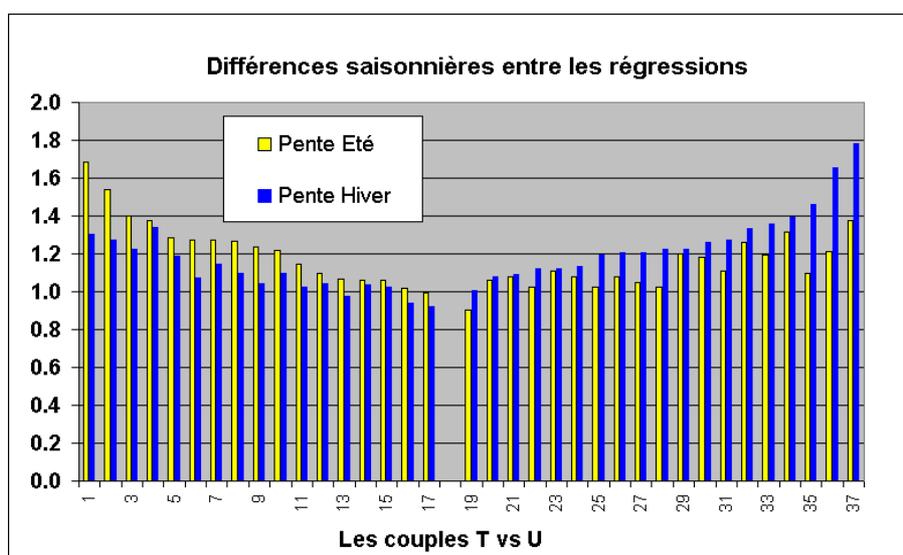


Figure 10 : différences saisonnières entre les régressions

On peut noter que les deux groupes (parties « gauche » et « droite » de cette figure) sont du même ordre de grandeur en nombre de villes (ici égaux par hasard), ce qui montre que la **proximité automobile** primant sur le fond urbain n'est **pas un phénomène essentiellement hivernal**. En effet, on a calculé sur l'ensemble de ces stations, les moyennes globales dans chacun des deux groupes de villes « gauche » et « droite » de la figure, en situation trafic et urbain, et en été et hiver.

Exprimé en résultats chiffrés sur le Tableau 12, il apparaît que :

- les moyennes globales des stations trafic entre les groupes « g » et « d » sont plus différenciées en été « 31 et 22 $\mu\text{g.m}^{-3}$ » qu'en hiver « 34 et 30 $\mu\text{g.m}^{-3}$ » ; cela peut expliquer les différences montrées dans la figure ;
- Alors que les stations urbaines ont des moyennes annuelles du même ordre de grandeur dans chaque groupe par saison : soit 19 et 17 $\mu\text{g.m}^{-3}$ en été, 23 et 20 $\mu\text{g.m}^{-3}$ en hiver.

Tableau 12 : les moyennes des teneurs en $\mu\text{g}/\text{m}^3$ de PM10 des groupes de stations selon les saisons

	« g » en été	« g » en hiver	« d » en été	« d » en hiver
Moyennes stations Trafic	31	34	22	30
Moyennes stations Urbaines	19	23	17	20
Rapports moyennes T / U	1,63	1,48	1,29	1,50

Le Tableau 13 suivant présente les détails chiffrés ayant permis le tracé de la figure précédente, et surtout les villes et stations concernées dans chaque groupe.

Tableau 13 : différences saisonnières entre les stations trafic et urbaines (régressions et moyennes en $\mu\text{g}/\text{m}^3$ de PM10)

Stations Trafic			Stations Urbaines			ETE			HIVER		
Ville	BDQA	App	Ville	BDQA	App	Pente	moy Trafic	moy Urb	Pente	moy Trafic	Moy Urb
COUPLES DE STATIONS OU L'EFFET DU TRAFIC PRIME EN ETE											
PARIS-16E	4053	Teom	PARIS- 1ER	4055	Teom	1.68	40	20	1.30	42	22
ROUEN	25037	Teom	ROUEN	25036	Teom	1.54	24	11	1.27	25	13
SAINT-DENIS	4058	Teom	BOBIGNY	4156	Teom	1.40	45	19	1.22	47	21
PARIS- 8E	4031	Teom	PARIS- 1ER	4055	Teom	1.38	32	18	1.34	33	22
PARIS-14E	4012	Teom	PARIS- 1ER	4055	Teom	1.28	35	20	1.19	38	22
LYON	20063	Teom	LYON	20062	Teom	1.27	20	17	1.07	32	29
BOULOGNE / M	10031	Teom	BOULOGNE / M	10032	Teom	1.27	24	19	1.14	26	22
LYON	20013	Teom	LYON	20062	Teom	1.27	30	19	1.10	35	24
ECHIROLLES	15039	Teom	GRENOBLE	15043	Teom	1.23	24	15	1.04	33	22
TOULON	3068	Teom	TOULON	3062	Teom	1.22	35	25	1.10	33	25
CHALON-SUR-S	32001	Bêta	CHAMPFORGEUIL	32002	Bêta	1.15	16	13	1.03	23	20
BELFORT	13009	Teom	MONTBELIARD	13007	Teom	1.10	18	16	1.04	20	20
AIX-EN-PROV	3021	Teom	AIX-EN-PROV	3029	Teom	1.07	28	23	0.97	30	26
DOUAI	6016	Teom	WAZIERS	6010	Teom	1.06	20	19	1.04	22	20
LYON	20003	Teom	LYON	20062	Teom	1.06	22	20	1.02	26	26
SALAISE / SANNE	27008	Teom	ROUSSILLON	27002	Teom	1.02	19	16	0.94	23	22
LYON	20002	Teom	LYON	20062	Teom	0.99	21	20	0.92	25	26
COUPLES DE STATIONS OU L'EFFET DU TRAFIC PRIME EN HIVER											
SAINT-ETIENNE	29428	Teom	SAINT-ETIENNE	29424	Teom	0.90	18	16	1.00	19	16
TOULOUSE	12032	Teom	TOULOUSE	12030	Teom	1.06	21	17	1.08	24	20
DUNKERQUE	10002	Teom	Malo-les-Bains	10004	Teom	1.08	18	16	1.09	25	21
BOURG-LES-VALENCE	36003	Teom	PORTES-LES-VALENCE	36001	Teom	1.02	23	18	1.12	26	20
BLIGNY-LES-B	26018	Bêta	DIJON	26005	Teom	1.11	18	15	1.12	18	16
AUDINCOURT	13005	Teom	MONTBELIARD	13007	Teom	1.08	15	16	1.13	21	20
ST-JULIEN-M-D	33112	Teom	ST-JEAN-DE-MAU	33111	Teom	1.02	19	20	1.19	27	27
NIMES	8616	Teom	NIMES	8614	Teom	1.08	25	18	1.21	23	17
MARSEILLE	3002	Teom	MARSEILLE	3047	Teom	1.05	30	25	1.21	33	26
TOURS	34022	Teom	TOURS	34025	Teom	1.02	15	14	1.22	17	16
RENNES	19002	Teom	RENNES	19005	Teom	1.20	20	15	1.22	24	18
LE HAVRE	5077	Teom	LE HAVRE	5064	Teom	1.18	24	17	1.26	29	19
STRASBOURG	16034	Teom	STRASBOURG	16029	Teom	1.11	21	17	1.27	25	19
NANTES	23140	Teom	NANTES	23107	Teom	1.26	21	15	1.33	22	16
ANGLET	31017	Teom	BIARRITZ	31018	Teom	1.20	27	19	1.36	34	24
VALENCIENNES	6006	Teom	VALENCIENNES	6001	Teom	1.32	24	17	1.40	27	19
CLERMONT-F	7034	Teom	CLERMONT-F	7008	Teom	1.10	17	13	1.46	22	15
GRENOBLE	15046	Teom	GRENOBLE	15043	Teom	1.21	26	16	1.65	38	22
VILLEURBANNE	20019	Teom	LYON	20062	Teom	1.37	21	20	1.78	37	26

Ces constatations sont néanmoins à prendre **avec prudence**, car les résultats des comparaisons sont très tributaires des implantations des stations ; en effet, on prend en compte au sein de chaque ville des stations situées en des sites trafic dont les niveaux d'exposition sont les reflets de situations locales variées ; on les compare ensuite avec des stations urbaines qui par contre sont un peu plus homogènes quelles que soient les villes, mais avec néanmoins des spécificités locales.

6.2 Différences entre jours ouvrés et non ouvrés

Ce paragraphe s'attache à déterminer si les différences entre les stations trafic et les stations urbaines sont plus marquées pendant les jours ouvrés que pendant les jours non ouvrés.

Pour les 36 couples de stations équipées d'une manière homogène en terme d'appareillage (cf 5.1 et 5.2), le Tableau 14 présente les moyennes annuelles en PM10 des mesures réalisées pour les deux typologies de station dans les deux types de jour.

Tableau 14 : incidence des jours ouvrés ou non ouvrés

	Moyennes annuelles des stations Trafic	Moyennes annuelles des stations urbaines
Jours ouvrés	27 $\mu\text{g.m}^{-3}$	20 $\mu\text{g.m}^{-3}$
Jours non ouvrés	24 $\mu\text{g.m}^{-3}$	19 $\mu\text{g.m}^{-3}$

Il apparaît que les stations urbaines mesurent pratiquement les mêmes niveaux de PM10 pendant les jours ouvrés et non ouvrés, alors que les stations trafic apparaissent plus exposées pendant les jours ouvrés. Les différences entre les deux typologies de stations sont donc amplifiées en semaine par rapport aux week-ends, à l'exception des stations soumises à des activités touristiques de week-end.

Cependant, pour ce qui concerne les régressions linéaires, à la différence de l'effet saisonnier, le type de jours a moins d'influence sur les pentes des droites de régression. La Figure 11 montre néanmoins qu'environ dans un tiers des villes concernées (les 14 de la partie gauche de la figure), l'influence urbaine prime sur l'influence du trafic pendant les jours non ouvrés, dont 10 en saison froide ; cela montre que la **proximité automobile** primant sur le fond urbain n'est **pas un phénomène uniquement lié aux jours ouvrés**.

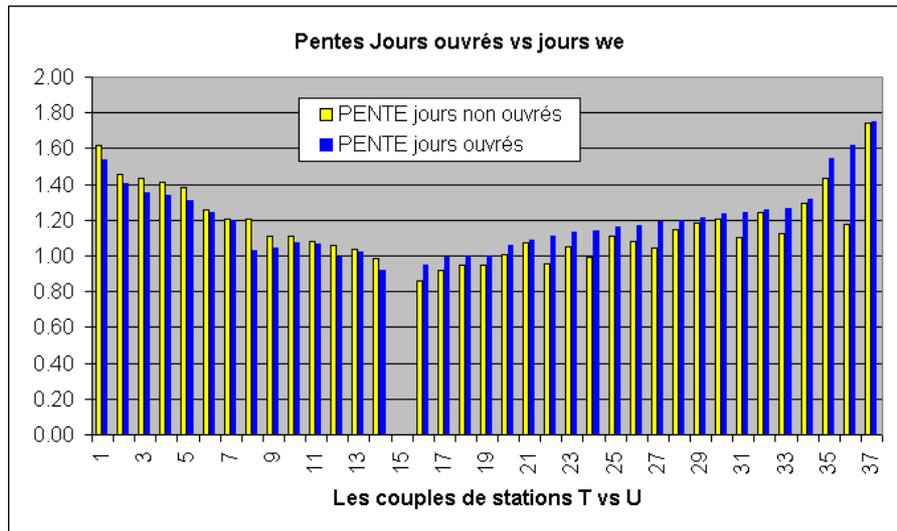


Figure 11 : influence des jours ouvrés ou non ouvrés sur les régressions

7. CONCLUSIONS

L'objet initial de cette étude est la détermination de la part prise par le trafic automobile par rapport aux stations urbaines dans les évolutions spatio-temporelles des PM10.

Les données issue de la BDQA ont concerné un peu moins de 400 stations de mesure des PM10, dont seulement 46 sont de typologie « trafic ». Pour 8 AASQA, il n'existe pas dans la BDQA de données PM10 issues de stations de typologie trafic.

Chacune des 46 stations a été comparée à la station urbaine la plus représentative dans le voisinage le plus proche possible, à l'aide de l'étude des corrélations et des régressions linéaires orthogonales.

Il convient de considérer les conclusions **avec prudence**, car les résultats des comparaisons sont très tributaires des implantations des stations ; dans la pratique, on compare des stations au sein de chaque ville où les niveaux d'exposition des stations peuvent être très disparates en situation de trafic, alors que les stations de typologie urbaine semblent exposées à des niveaux plus proches.

Bien que le thème de l'ajustement des données n'entraîne a priori dans le domaine de cette étude, la question suivante s'est posée d'office : « **sur quelle base** de données cette étude doit-elle porter, **brutes ou ajustées** ? » ; aussi, l'impact de l'ajustement des données a été considéré dans la première phase de l'étude.

Il apparaît d'emblée un **manque d'homogénéité dans la base ajustée** ; en effet,

- si la majorité des AASQA a communiqué les données TEOM brutes et ajustées associées,
- certaines d'entre elles ont communiqué les données ajustées sans les données brutes, mais avec les données des FDMS associés ;

- certaines d'entre elles ont communiqué les données ajustées sans les données brutes et aussi sans les données des FDMS associés ;
- d'autres enfin n'ont communiqué que les données FDMS de référence en lieu et place des TEOM associés ajustés.

Il est apparu également, que plusieurs modes différents de calcul des delta d'ajustement ont été utilisés.

La première partie de l'étude a montré que les **ajustements** des mesures de toutes les stations, trafic et urbaines, **modifient** les pentes des droites de régression ainsi que les ordonnées à l'origine. On voit en général avec l'ajustement que les **corrélations augmentent** et deviennent presque toutes supérieures à 0,90, et que les **pentés des droites sont abaissées**, d'une manière qui varie selon les sites.

Lors des ajustements pratiqués à l'aide d'un delta moyenné sur 4 heures, le problème relatif à l'influence des 3 heures précédentes sur l'ajustement de la quatrième heure entraîne une surestimation des valeurs faibles et une sous-estimation des valeurs fortes ; cela a conduit à s'interroger sur le bien-fondé d'utiliser un ajustement basé sur une autre base temporelle, soit horaire en cas de besoin, ou sur 24 heures. Il faut éviter de calculer des delta quart horaire.

L'étude montre que l'ajustement pour chaque couple de stations (trafic vs urbain) quel qu'il soit, 1h, 4h ou 24h :

- augmente les corrélations d'une manière identique,
- réduit les pentes, surtout avec les delta 4 h, un peu moins pour les delta calculés sur 1 ou 24 heures (lesquels ont le même effet),
- augmente légèrement les ordonnées à l'origine des droites de régression.

On constate avec les ajustements à l'aide de delta 4h que :

- pour les stations de typologie **autre que trafic**, les nombres de dépassements, surtout en milieux urbain peu exposé et rural, sont **jusqu'à 6 fois plus élevés** ;
- pour l'ensemble des 39 stations « trafic » (hors jauges Bêta non traitées) qui ont enregistré en données brutes au moins 1 (et jusqu'à 136) dépassements du seuil $50 \mu\text{g}\cdot\text{m}^{-3}$, l'ajustement à l'aide de la méthode des « delta » a augmenté le nombre total de dépassements de 723 à 1955, soit **seulement 2,7 fois plus**, parce que les ajustements déterminés en milieu urbain de fond sont **sous-estimés** par rapport à la réalité des sites trafic ;
- l'usage du « delta 4h » majore encore plus les nombres de dépassements que l'usage de « delta 1h » ou « delta 24h ».

La première conséquence est que l'on pourrait simplifier et harmoniser la procédure d'ajustement en ne l'appliquant qu'aux données journalières.

La seconde conséquence est que l'ajustement horaire n'a d'intérêt que pour des études de données réalisées sur une base horaire.

La troisième conséquence est qu'il faudrait que la station prise comme référence soit représentative de l'ensemble des autres stations qu'elle est censée ajuster, autrement dit que toutes soient non seulement de même typologie mais qu'elles mesurent des niveaux assez proches les uns des autres.

La quatrième conséquence est qu'on **sous-ajuste les stations trafic** puisqu'on utilise un « delta » issu d'une station urbaine, comme c'est en pratique presque toujours le cas.

La cinquième conséquence est que la suite de cette étude ne peut porter que sur les données brutes, d'autant plus que plusieurs stations sont équipées de jauges Bêta classiques ou de jauges Bêta - RST pour lesquelles les ajustements étaient bien-entendu sans objet.

Chaque station trafic a été comparée à une seule station urbaine soigneusement choisie.

Il apparaît qu'une **quinzaine de stations de typologie trafic ne subissent pas d'influence marquée d'origine automobile** et pourraient être considérées comme urbaines (corrélations supérieures à 0,90 et pentes comprises entre 0,90 et 1,10).

Pour 22 stations de typologie trafic où les niveaux rencontrés sont plus conséquents que ceux des stations urbaines proches, on a constaté que :

- des pentes élevées des régressions linéaires ne sont pas forcément liées à des niveaux de PM10 importants,
- les pentes des régressions s'étendent de 1 à 2, c'est-à-dire que les stations trafic ont mesuré des niveaux allant du simple au double par rapport aux niveaux urbains,
- presque toutes les corrélations sont supérieures à 0,90 ; cela signifie qu'il y a une **homogénéité temporelle des variations des niveaux entre trafic et urbain**, ce qui renforce la notion d'évolutions spatio-temporelles à l'échelle régionale,
- les nombres de dépassements du seuil $50 \mu\text{g}\cdot\text{m}^{-3}$ sont compris de 1 à 134 en situation trafic, et de 0 à 9 en situation urbaine, exception faite du cas de Thionville, où la station urbaine pourrait être influencée par une proximité plutôt industrielle.

Enfin, dans 8 villes où les deux stations ne sont pas équipées d'appareils de même technologie, les comparaisons étaient difficiles à interpréter. Cela laisse à penser qu'il est plus cohérent de disposer au sein d'une même ville (ou zone) de groupes d'appareils homogènes en terme de technologie analytique.

On a constaté pour la moitié des couples de stations exploitables, que les écarts entre les deux typologies de stations étaient plus importants en saison chaude qu'en saison froide, ce qui montre que la **proximité automobile primant sur le fond urbain n'est pas un phénomène essentiellement hivernal**.

On a constaté pour un tiers des couples de stations exploitables, que les écarts entre les deux typologies de stations étaient plus importants pendant les jours non ouvrés

que pendant les jours ouvrés, et cela en saison froide, ce qui montre que la proximité automobile primant sur le fond urbain n'est pas un phénomène uniquement lié aux jours ouvrés.

La trop grande diversité de situations de proximité automobile, affecte à chacune des stations concernées une carte d'identité peu ou pas comparable avec les autres stations de même typologie, alors que les situations urbaines sont globalement plus homogènes.

Il devient de plus en plus intéressant par voie de conséquence, de rendre plus précise la définition du réseau de surveillance de la proximité trafic en resserrant les diverses variantes décrites dans le Guide ADEME, qu'il est prévu de réviser dans les prochains travaux 2009.

ANNEXE 1 - DOCUMENT DE REFERENCE DE L'ETUDE

Etude n° 34

Responsables de l'étude : EMD

Etude de l'impact du trafic routier sur les teneurs en PM à partir de la base de données PM₁₀ et PM_{2,5}

Contexte et objectifs

Dans le cadre de la surveillance de la Qualité de l'Air et du respect de la réglementation, les nombres de dépassements des valeurs seuils fixées par la Directive Européenne doivent être communiqués pour chaque zone administrative, en agglomération et hors agglomération.

Les mesures des PM₁₀ et des PM_{2,5} requises, sont réalisées essentiellement en milieu urbain ; les pointes et donc les dépassements sont dus à des aérosols secondaires, à des phénomènes naturels et aussi aux spécificités des situations de proximité des installations industrielles et du trafic automobile.

Cette étude propose de déterminer la part prise par les transports automobiles dans les évolutions spatio-temporelles de toutes les zones du dispositif national de surveillance des PM, en comparant les données ajustées obtenues entre les stations de fond et les stations de proximité automobile.

Travaux antérieurs

Cette étude s'appuiera sur l'ensemble des connaissances acquises par le LCSQA depuis plusieurs années et tout particulièrement ceux réalisés de 2005 à 2007 concernant :

- les degrés d'homogénéité des stations de mesure des PM₁₀ et des PM_{2,5} ,
- les évolutions spatio-temporelles des teneurs en PM₁₀ et en PM_{2,5} ,
- les ajustements des données à l'aide des stations de référence.

Travaux proposés :

Une présentation des travaux menés sur l'étude des pointes PM₁₀ et PM_{2,5} sera effectuée lors du Comité de suivi Particules du 1^{er} trimestre 2008.

La détermination de la part prise par le trafic automobile dans les évolutions spatio-temporelles des PM₁₀ et des PM_{2,5} proposée dans cette étude sera constituée de :

- La validation de la base de données horaires issue du parc de stations existant en 2007, et éventuellement en 2006 ;
- La comparaison pour l'ensemble du parc national des concentrations moyennes entre les stations de proximité automobile et les stations de fond ;
- La mise en évidence des écarts rencontrés entre ces deux typologies de stations ;
- L'étude fine sur les évolutions horaires selon les situations de trafic et d'absence de trafic ;
- L'étude des écarts observés entre les jours ouvrés et non ouvrés, les saisons froides et chaudes.

Par extension, on vérifiera la réalité des appellations « trafic » attribuées à certaines stations en relation avec la densité réelle des transports proches , et les écarts observés entre les deux typologies « urbaine » et « trafic ».

Un tentera dans la limite du possible, un rapprochement entre les données de PM_{2,5} et les données de Fumées Noires si tant est qu'il en reste encore.

Renseignements synthétiques

Titre de l'étude	Evolutions spatio-temporelles des teneurs en PM		
Personne(s) responsable(s) de l'étude	Jean-Luc Houdret (EMD)		
Travaux	Pluriannuels		
Durée des travaux	1 an		
Collaboration AASQA	Oui		
Heures d'ingénieur 2008	EMD : 400	INERIS :	LNE :
Heures de technicien 2008	EMD :	INERIS :	LNE :
Document de sortie attendu	Rapport LCSQA		
Lien avec le tableau de suivi CPT	1) Stratégie de surveillance		
Lien avec un groupe de travail LCSQA	Commission de Suivi « Particules »		
Matériel à acquérir pour l'étude			

ANNEXE 2 – MOYENS STATISTIQUES

A2.1 - Modèle linéaire

Plusieurs modes de régression linéaire peuvent être réalisés. La régression orthogonale a été choisie ; le mode de calcul est présenté ci-après.

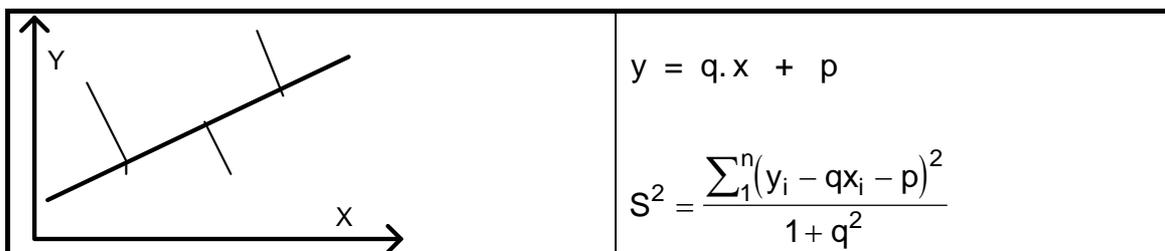
On utilise un modèle linéaire de régression ainsi que le coefficient de corrélation :

- la corrélation,
 - on considère qu'une corrélation élevée entre deux stations est un indice d'homogénéité temporelle puisque les variations sont simultanées, et cela quelles que soient la pente et l'ordonnée à l'origine de la droite de régression ; les valeurs trouvées sont réparties en 4 classes, dont seules les deux premières sont retenues ;
- la pente de la droite de régression orthogonale,
 - on considère qu'une pente proche de « 1 » est un indice d'homogénéité de concentrations entre les deux sites considérés ; les valeurs trouvées sont réparties en 4 classes, dont seules les deux premières sont retenues ;
- l'ordonnée à l'origine de cette droite,
 - ce paramètre est censé être le plus proche de zéro, les valeurs élevées devant être considérées comme des artefacts ; elle n'est en fait pas utilisée dans l'étude, sauf si une valeur est élevée caractérisant une station à problème.

A2.2 - Régression linéaire orthogonale

La régression quadratique ci-dessous semble plus appropriée pour la comparaison de données indépendantes car au contraire de la régression linéaire classique, elle ne dépend pas du sens dans lequel les variables sont prises.

On cherche l'équation de la droite $y = q.x + p$ telle que la somme des carrés des écarts quadratiques S^2 soit minimale :



$$\frac{\partial S^2}{\partial p} = \frac{\sum_1^n (y_i - qx_i - p)}{1 + q^2} = 0 \quad , \quad \text{on en tire } p = \bar{y} - q\bar{x}$$

la droite passe par le point moyen \bar{x}, \bar{y} .

$$\frac{\partial S^2}{\partial q} = \frac{\sum_1^n [(y_i - \bar{y}_i) - q(x_i - \bar{x})]}{1 + q^2} = 0$$

La résolution des dérivées partielles par rapport à « p » puis à « q » conduit à une équation du second degré dont la racine positive est :

$$q = \frac{SCEy - SCEx + \sqrt{(SCEy - SCEx)^2 + 4SPExy^2}}{2SPExy}$$

où :

SPExy est la somme des produits des écarts entre les x_i, y_i et leurs moyennes respectives,

SCEy est la somme des écarts entre les y_i et leur moyenne.

SCEx est la somme des écarts entre les x_i et leur moyenne.

A2.2 - Coefficient de corrélation

Le coefficient de corrélation est égal à :

$$r = \frac{SPExy}{\sqrt{SCEx \cdot SCEy}} = \frac{\sum_1^n [(y_i - \bar{y})(x_i - \bar{x})]}{\sqrt{\sum_1^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_1^n (y_i - \bar{y})^2}}$$

Le coefficient de corrélation est symétrique en x et en y . Son estimation est une variable aléatoire qui permet de déceler l'existence d'une liaison entre les variables x et y. Il faut connaître sa loi de distribution pour évaluer la confiance de cette estimation pour par exemple la comparer à des valeurs théoriques telles que 0 ou 1. En fait, le coefficient de corrélation est surtout efficace pour vérifier l'absence de liaison entre les variables x et y , c'est-à-dire s'il est inférieur à une valeur prédéterminée. Il n'est pas assez précis pour vérifier si la liaison entre x et y est forte. Cependant, il est d'usage de considérer en matière de mesurage qu'un coefficient supérieur à 0,9 montre une bonne dépendance des variables, et qu'il permet d'affecter à une série de comparaisons multiples un ordre de classement.