



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

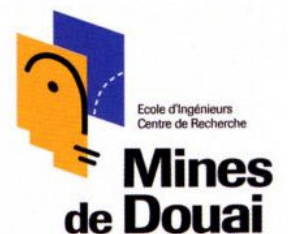


**Etude n° 11 - 2004
Cartographie des concentrations
en dioxyde d'azote
à partir des mesures en stations fixes**

Décembre 2004 - version définitive

Convention : CV 04 000 088

**Esperanza PERDRIX, André WROBLEWSKI, Benoît FOURCHE
et Hervé PLAISANCE**





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 en liaison avec les Associations Agréées de Surveillance de la Qualité de l'Air (AASQA), le Ministère de l'Ecologie et du Développement Durable et l'ADEME. Ces travaux supportés financièrement par la Direction des Préventions des Pollutions et des Risques du Ministère de l'Ecologie et du Développement Durable sont réalisés avec le souci constant d'apporter l'appui scientifique et technique indispensable aux AASQA et d'assurer la liaison entre la recherche en matière de pollution atmosphérique et son application sur le terrain.

L'objectif principal du LCSQA est de participer à l'amélioration de la qualité des mesures effectuées en France, depuis le prélèvement des échantillons jusqu'au traitement des données issues des mesures. Cette action est menée dans le cadre notamment des Directives européennes mais aussi dans un cadre prospectif pour fournir aux AASQA de nouveaux outils permettant d'anticiper sur les évolutions futures.

ECOLE DES MINES DE DOUAI

DEPARTEMENT CHIMIE ET ENVIRONNEMENT

ETUDE n°11

**CARTOGRAPHIE DES CONCENTRATIONS
EN DIOXYDE D'AZOTE
A PARTIR DES MESURES EN STATIONS FIXES**

Esperanza PERDRIX, André WROBLEWSKI, Benoît FOURCHE, Hervé PLAISANCE
Décembre 2004

SOMMAIRE

Contexte et principe de l'étude	3
▪ Contexte	3
▪ Principe de l'étude	4
▪ Structure de l'étude	4
Première partie : Recherche des corrélations entre sites de mesure temporaires (échantillonneurs passifs) et stations permanentes (analyseurs automatiques)	5
Rappel des objectifs	5
Présentation du jeu de données	5
▪ Campagnes de mesure par échantillonnage passif	5
▪ Les analyseurs fixes automatiques	6
▪ Relation entre un analyseur fixe automatique et un échantillonneur passif placé à côté	7
Recherche des régressions linéaires entre un analyseur fixe automatique donné et différents sites d'échantillonnage passif de la zone d'étude	9
▪ Méthodologie	9
▪ Résultats	10
Etablissement des cartes de pollution à partir des sites corrélés	15
Deuxième partie : Etude des corrélations entre stations permanentes (analyseurs automatiques) en fonction du pas de temps choisi	20
Rappel des objectifs et de la méthodologie	20
Présentation du jeu de données	21
Calcul des régressions linéaires entre stations automatiques, sur une base temporelle de 14 jours, 7 jours et 24 heures	21
▪ Corrélations pour des valeurs bihebdomadaires	21
▪ Corrélations pour des valeurs hebdomadaires ou journalières	25
▪ Validation absolue des corrélations bihebdomadaires et journalières	28
▪ Estimation de valeurs journalières, à partir de corrélations établies pour des valeurs bi-hebdomadaires	30
Conclusion	36
Remerciements	37
Bibliographie	38
Annexe 1 : corrélations avec la station urbaine « Roubaix château »	
Annexe 2 : corrélations avec la station périurbaine « Halluin cailloux »	
Annexe 3 : recherche des corrélations entre stations automatiques	

RESUME de l'étude n°11 du rapport d'activités de l'école des Mines de Douai 2004

Etude suivie par Esperanza PERDRIX et André WROBLEWSKI
☎ : 03 27 71 26 15 et 03 27 71 26 11

CARTOGRAPHIE DES CONCENTRATIONS EN DIOXYDE D'AZOTE A PARTIR DES MESURES EN STATIONS FIXES

Les stations de mesure automatiques permettent de suivre, en continu, les variations de la concentration en NO₂. Néanmoins, leur nombre limité ne permet pas de donner une information, en continu, en tout point d'un territoire, sous forme de cartographie.

Pour fournir des cartes de répartition spatiale du polluant, on a souvent recours soit à de la modélisation déterministe, soit à des calculs d'interpolation effectués à partir de mesures réalisées par échantillonnage passif en de nombreux points. Contrairement à la modélisation déterministe, qui nécessite la connaissance d'un grand nombre de phénomènes et paramètres (mécanismes physico-chimiques, identification et quantification des émissions, structure de l'atmosphère, météorologie), l'interpolation peut être menée à partir des seules concentrations en polluant dans l'air ambiant.

Cependant, la réalisation des campagnes de mesure par échantillonnage passif, sur un grand nombre de sites, n'est pas aisée, en raison des importants moyens humains et financiers qu'elle requiert. Les cartes dressées par interpolation, à partir de telles campagnes, sont donc établies avec une fréquence peu élevée (par exemple, quinquennale).

De plus, la période nécessaire à l'échantillonnage passif du NO₂ est de 14 jours. Les cartes peuvent donc, au mieux, être établies avec une période bi-hebdomadaire.

Pour cette raison, nous cherchons à mettre au point la réalisation de cartographies du dioxyde d'azote, avec une fréquence plus élevée (hebdomaire ou journalière), à partir des mesures d'un petit nombre de stations automatiques et en tirant profit des campagnes de mesure par échantillonnage passif existantes.

L'étude se scinde en deux parties.

Dans la première partie, nous avons étudié un jeu de données issues de campagnes de mesure du NO₂ par échantillonnage passif, pour tenter de répondre aux questions suivantes :

- Peut-on établir une corrélation linéaire entre les mesures d'une station fixe automatique et celles obtenues par échantillonnage passif en un site donné ? Quelle est la précision des estimations fournies par la corrélation ?
- Le nombre de sites passifs corrélés à une ou deux stations fixes permet-il de réaliser une cartographie estimative de la pollution avec une bonne précision ?

Notre étude montre que, pour le jeu de données étudié (AREMA-LM), plus de la moitié des sites utilisés pour l'échantillonnage passif présentent des valeurs corrélées à celles de stations automatiques. Leur nombre est suffisant pour qu'une carte estimative des teneurs en NO₂ soit réalisée par interpolation des données (utilisation du krigeage possible).

De plus, nous avons trouvé que seules deux stations automatiques suffisent pour permettre la couverture d'une large zone, à condition de choisir des stations de typologie distincte : l'une urbaine, l'autre périurbaine.

Néanmoins, peu de sites corrélés ont été trouvés dans les centres-villes, à proximité des grands axes routiers et des industries émettrices. Les cartes établies à partir des sites corrélés sont donc des cartes de « pollution de fond », où les pollutions locales ou liées à la proximité automobile ne sont pas bien visibles. Pour y remédier, il serait judicieux d'effectuer l'interpolation des données par krigeage, en utilisant des variables externes (telle que le cadastre des émissions de NO_x).

Dans la seconde partie de l'étude, nous avons étudié un jeu de données, constitué des mesures de stations fixes automatiques sur plusieurs années, de 1998 à 2003, pour la même zone d'étude (AREMA-LM).

Notre but était de savoir si la corrélation entre un site passif et une station automatique, établie pour des valeurs moyennées sur 14 jours (durée d'exposition des échantillonneurs passifs pour le NO₂), pouvait s'appliquer, avec une bonne précision, pour des valeurs moyennées sur des périodes plus courtes (7 jours ou 24 heures).

Dans ce cas, il pourrait être possible d'utiliser les corrélations entre des sites passifs et une station automatique, pour estimer des valeurs hebdomadaires ou journalières et réaliser ainsi, par interpolation de ces valeurs, des cartes hebdomadaires ou journalières de la répartition du polluant.

Nous avons montré que, dans le cas du jeu de données étudié, toutes les corrélations établies pour un pas de temps de 14 jours existent pour des pas de temps plus courts, de 7 jours ou 24h.

De plus, les paramètres des corrélations subissent des variations limitées en fonction du pas de temps. On peut donc estimer, dans environ 90% des cas, les concentrations moyennes journalières en un site à partir de la régression établie pour des valeurs bi-hebdomadaires entre ce site et une station automatique, avec une précision suffisante au regard de la directive européenne 1999/30/CE.

Cependant, dans 10% des cas environ, la précision se révèle insuffisante. Ces cas semblent relever de situations de pollution peu fréquentes, vraisemblablement liées à des situations météorologiques particulières, qu'il conviendrait d'étudier ultérieurement. De même, l'application de la méthode à un territoire présentant un relief plus contrasté serait nécessaire.

L'étude a donné lieu à la réalisation d'un fichier de procédures sous EXCEL permettant le calcul des corrélations entre stations automatiques (paramètres statistiques, exactitude des corrélations).

Contexte et principe de l'étude

Contexte

Depuis plusieurs années, des campagnes de mesure de NO₂ par échantillonnage passif sont effectuées, afin d'évaluer l'exposition de la population au NO₂ de fond. Les échantillonneurs passifs sont exposés pendant une période de 14 jours, et sont ensuite envoyés dans un laboratoire où le polluant est extrait et sa teneur quantifiée. Les teneurs en NO₂ sont exprimées en µg/m³ et correspondent à une valeur moyenne sur 14 jours d'exposition. L'expérience acquise par les AASQA et le LCSQA dans ce domaine a été consignée dans un guide publié en 2002 [1].

Le coût relativement réduit des échantillonneurs et de leur analyse permettent leur installation en de nombreux sites (plusieurs dizaines à quelques centaines) couvrant toute la zone à surveiller. Il est alors possible de dresser a posteriori, par interpolation spatiale des données, une carte de répartition de la concentration moyenne en NO₂, sur toute la zone surveillée, pour la période de 14 jours couverte par la campagne. Lorsque plusieurs campagnes de mesure ont été menées, sur des périodes représentatives d'une année, on peut fournir des cartes estimatives de la teneur moyenne annuelle en polluant sur une zone. Cette étape d'exploitation cartographique des données issues des campagnes fait l'objet de plusieurs travaux du LCSQA, visant à mieux appliquer les techniques d'interpolation spatiale [2], à mieux estimer les incertitudes associées [3] et à mieux choisir les périodes d'échantillonnage représentatives [4].

La méthode d'échantillonnage passif du NO₂ étant employée depuis plusieurs années, certaines AASQA disposent désormais de plusieurs campagnes de mesure menées (de façon discontinue) sur plusieurs années.

Dans une première étude [5], nous avons montré qu'il était possible d'exploiter les données issues de campagnes répétées pour étudier la représentativité spatiale des stations de mesure fixes automatiques, sur la base temporelle de valeurs moyennées sur 14 jours.

Nous avons observé que certains sites d'échantillonnage passif présentaient des valeurs semblables à celles mesurées par des analyseurs fixes automatiques, de typologie donnée (urbaine ou périurbaine), quelle que soit la campagne de mesure considérée. Autrement dit, les valeurs mesurées en certains sites sont du même ordre de grandeur et présentent des variations similaires à celles données par des stations fixes automatiques, en moyenne, sur une base de 14 jours. Ces sites semblent relever d'une typologie, urbaine ou périurbaine, identique à celle de l'analyseur automatique considéré.

Nous avons pu montrer également qu'en considérant l'ensemble des sites similaires à une station donnée, on pouvait définir une zone, appelé zone de « représentativité spatiale de la station (pour des valeurs moyennées sur 14 jours) ». Nous avons constaté que les zones relatives à des stations de même typologie étaient superposables. En outre, les zones relatives aux stations périurbaines et celles relatives aux stations urbaines étaient complémentaires.

Les zones de représentativité des stations péri-urbaines correspondaient à la périphérie des urbanisations ; celles des stations urbaines aux centres villes, nœuds autoroutiers et zones industrielles.

Principe de l'étude

Dans ce travail, nous proposons d'examiner, plus en détail, les relations mathématiques existant entre des sites d'échantillonnage passif et des stations fixes automatiques.

La méthodologie proposée repose donc sur l'exploitation des données issues de campagnes de mesure multiples. Lorsque le nombre de campagnes cumulées devient important (de l'ordre d'une dizaine) et que les emplacements des sites de mesure par échantillonnage passif varient peu d'une campagne à l'autre, il devient possible d'examiner l'existence de corrélations linéaires entre les valeurs obtenues sur un site de mesure donné et celles mesurées par un analyseur fixe.

Si de telles corrélations existent, on peut alors considérer les sites corrélés comme des stations fixes « virtuelles », auxquelles on peut affecter une valeur calculée, à partir de la corrélation existant avec la station automatique réelle. Cela permet de multiplier « virtuellement » le nombre de stations de mesure et donc de pouvoir utiliser des méthodes d'interpolation spatiale des données (telles que la géostatistique) pour établir une cartographie de la pollution.

Structure de l'étude

En première partie de ce rapport, nous proposons d'étudier un jeu de données mesurées comprenant :

- des mesures de NO₂ effectuées par échantillonnage passif sur plusieurs sites, pendant plusieurs campagnes bi-hebdomadaires ;
- des mesures de NO₂ obtenues au moyen d'analyseurs fixes automatiques.

Nous mettrons en évidence les sites corrélés à des analyseurs automatiques et détaillerons les paramètres des régressions linéaires trouvées.

Nous comparerons des cartes de pollution établies

- soit à partir de l'ensemble des sites ;
- soit à partir des seuls sites corrélés aux stations automatiques.

Dans une seconde partie de l'étude, nous proposons d'examiner la possibilité d'établir des cartes de répartition de la pollution, à partir des « stations virtuelles », pour une période de temps inférieure à 14 jours: 7 jours ou 24 heures.

A cette fin, nous considérerons le jeu de données constitué des mesures journalières effectuées par des analyseurs automatiques sur plusieurs années.

Nous comparerons les corrélations entre analyseurs automatiques, obtenues pour des valeurs moyennées sur des pas de temps différents : 14 jours, 7 jours ou 1 journée. Nous saurons ainsi si des corrélations établies pour des valeurs moyennées sur 14 jours restent valables sur 7 jours ou 24 heures. Nous étudierons aussi la précision des estimations pour les différents pas de temps.

PREMIERE PARTIE : RECHERCHE DES CORRELATIONS ENTRE SITES DE MESURE TEMPORAIRES (ECHANTILLONNEURS PASSIFS) ET STATIONS PERMANENTES (ANALYSEURS AUTOMATIQUES)

Rappel des Objectifs

Les objectifs de l'étude sont :

- d'une part, de montrer l'existence de sites de mesure (parmi les sites utilisés pour les campagnes de mesure par échantillonnage passif) corrélés à des stations fixes automatiques. Ces sites corrélés seraient nommés des « stations virtuelles » ;
- d'autre part, d'examiner la possibilité d'utiliser ces « stations virtuelles » pour dresser des cartes estimatives de la pollution moyenne sur 24 heures ou 1 semaine, à partir des mesures des stations fixes automatiques.

Présentation du jeu de données

Campagnes de mesure par échantillonnage passif

Le jeu de données étudié a été mis à notre disposition par l'AREMA-LM. Il est constitué par les résultats issues de 15 campagnes de mesure du dioxyde d'azote, effectuées par échantillonnage passif au cours des années 1998 à 2003¹.

Le tableau 1 présente les dates des campagnes et le type d'échantillonneur passif utilisé. La durée du prélèvement est de 14 jours pour toutes les campagnes. Les informations complètes sur chaque campagne sont consignées dans trois rapports d'étude de l'AREMALM [6-8].

Les campagnes les plus récentes du jeu de données ont été menées sur une zone élargie, par rapport aux campagnes les plus anciennes. Afin d'avoir un maximum de mesures par site, nous avons choisi de restreindre l'étude à la zone commune à toutes les campagnes, d'une étendue d'environ 20km×23km, soit 460 km².

Le tableau 2 présente les principales caractéristiques statistiques des mesures effectuées aux sites situés dans la zone commune.

¹ la numérotation des campagnes s'étend de 1 à 18, car certaines campagnes ont été invalidées.

Tableau 1 – Jeu de données : dates des campagnes de mesure et type d'échantillonneur passif utilisé.

Saison	N°	Date début	Date fin	Echantillonneur passif
Eté 1998	1	27 mai 98	10 juin 98	Gradko
	2	10 juin 98	24 juin 98	Gradko
	3	10 juin 98	8 juillet 98	Gradko
	4	8 juillet 98	22 juillet 98	Gradko
	5	8 juillet 98	5 août 98	Gradko
	6	5 août 98	19 août 98	Gradko
Hiver 98-99	7	3 décembre 98	16 décembre 98	Gradko
	10	20 janvier 99	3 février 99	Gradko
	11	3 mars 99	17 mars 99	EMD Palmes NO2/SO2
Eté 2003	12	22 mai 2003	5 juin 2003	Passam
	13	5 juin 2003	19 juin 2003	Passam
	14	5 juin 2003	3 juillet 2003	Passam
Hiver 2004	16	8 janvier 2004	22 janvier 2004	Passam
	17	22 janvier 2004	5 février 2004	Passam
	18	5 février 2004	19 février 2004	Passam

Tableau 2 – Jeu de données : principales statistiques des mesures effectuées pendant les campagnes ([NO₂] en µg/m³).

N° campagne	1	2	3	4	5	6	7	10	11	12	13	14	16	17	18
minimum	15,1	9,3	9,3	8,0	8,0	11,0	11,2	17,6	26,6	8,0	15,1	14,4	17,5	19,0	22,9
25 ^{ème} percentile	22,8	14,2	14,0	13,0	13,2	16,9	33,7	32,7	38,6	22,6	21,3	21,1	27,5	28,4	33,1
médiane	25,9	17,9	17,9	16,3	16,7	19,6	39,1	36,6	43,0	26,3	23,8	23,8	31,2	33,6	38,2
moyenne	26,5	18,1	18,8	16,9	16,8	20,5	38,9	37,4	43,1	27,5	25,0	25,2	31,1	33,0	37,7
75 ^{ème} percentile	29,1	21,4	21,9	19,9	19,6	23,7	44,6	40,6	47,6	31,5	28,7	28,4	34,7	36,6	42,6
maximum	40,7	32,0	52,6	40,6	32,3	36,0	68,6	68,3	61,7	63,4	53,9	60,7	52,3	47,7	52,0
variance	28,1	26,6	39,6	25,8	22,0	28,8	65,2	61,2	40,5	68,6	40,0	49,2	34,3	35,4	41,2
nombre de points	120	117	117	127	129	125	130	132	125	134	129	136	134	134	137

Les analyseurs fixes automatiques

La zone commune d'étude comprend plusieurs analyseurs fixes automatiques de mesure des oxydes d'azote, de typologies différentes (Tab. 3).

Nous avons sélectionné les stations automatiques qui existaient déjà lors de la première campagne de mesure, c'est-à-dire celles dont la création est antérieure à mai 1998. Cela doit permettre d'obtenir des corrélations entre station automatique et site d'échantillonnage passif avec un maximum de points.

Nous avons exclu les stations de typologie « trafic » et gardé celles de typologie « urbaine » ou « périurbaine ». En effet, les mesures intégrées sur 14 jours obtenues par échantillonnage passif donnent une indication sur la pollution de fond et non sur celle de proximité ; on ne peut donc les comparer qu'avec des stations urbaines ou périurbaines.

Tableau 3 – Dénomination, date de création et typologie des stations fixes de mesure situées dans la zone d'étude

(sur fond blanc, les stations retenues pour l'étude ; sur fond gris, les stations non retenues).

Station automatique	N°BDQA	Date de création	Typologie
Armentières Beauvais	11022	antérieure à mai 1998	périurbaine
Halluin Cailloux	11026	antérieure à mai 1998	périurbaine
Lesquin Charlemagne	11033	Avril 2000	périurbaine
Villeneuve d'Ascq Rameau	11032	Juillet 1999	périurbaine
Haubourdin Pompiers	11036	Janvier 2003	urbaine
Lille Lakanal Fives	11025	antérieure à mai 1998	urbaine
Lomme Victor Hugo	11029	Mai 1999	urbaine
Marcq-en-Baroeul CTM	11016	antérieure à mai 1998	urbaine
Roubaix Château	11002	antérieure à mai 1998	urbaine
Tourcoing bienfaisance	11027	antérieure à mai 1998	urbaine
Lille Faidherbe	11024	antérieure à mai 1998	trafic
Lille Liberté	11005	antérieure à mai 1998	trafic
Lille Pasteur	11019	antérieure à mai 1998	trafic
Roubaix Serres	11034	Mars 2001	trafic

Relation entre un analyseur fixe automatique et un échantillonneur passif placé à côté.

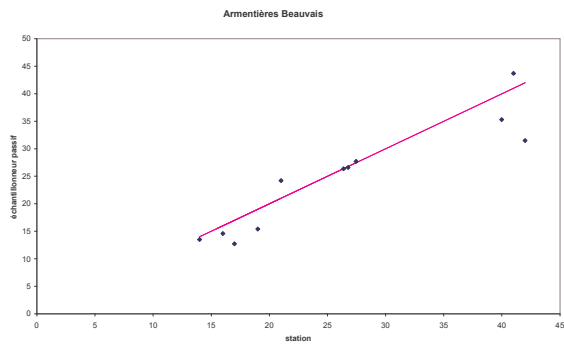
Dans un premier temps, nous avons étudié la relation entre un analyseur automatique et un échantillonneur passif placé à côté.

Dans le cas idéal, les couples de points de mesure (analyseur ; échantillonneur) doivent se situer sur la première bissectrice (Fig. 1).

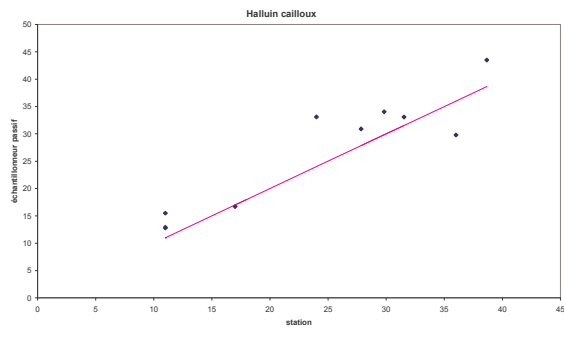
C'est le cas pour la plupart des stations sauf pour la station urbaine « Tourcoing bienfaisance ». Les points relatifs à cette station sont situés au-dessus de la première bissectrice.

Etant donné ce mauvais accord, nous avons décidé de ne pas prendre en compte la station « Tourcoing bienfaisance » dans la suite de cette partie.

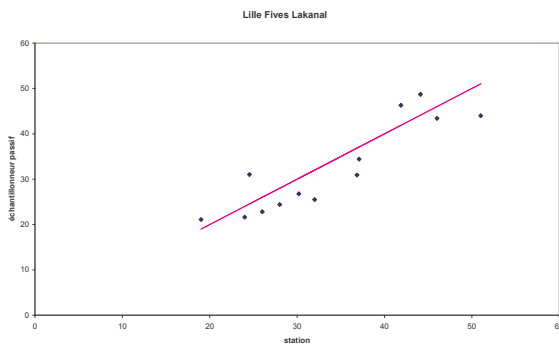
Le tableau 4 présente les équations des regressions linéaires obtenues. Pour toutes les stations, le coefficient de détermination (r^2), qui est un indicateur de la qualité de la régression, est supérieur à 0,8 . L'écart-type résiduel, qui est une estimation de l'erreur commise sur la mesure de Y c'est-à-dire sur la valeur mesurée par échantillonnage passif, est en moyenne de 3,9 $\mu\text{g}/\text{m}^3$.



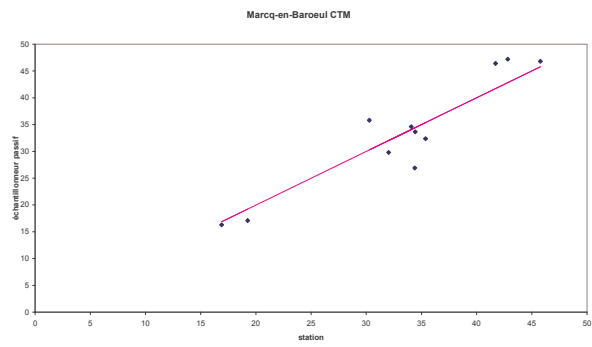
(a) Armentières Beauvais (Périurbaine)



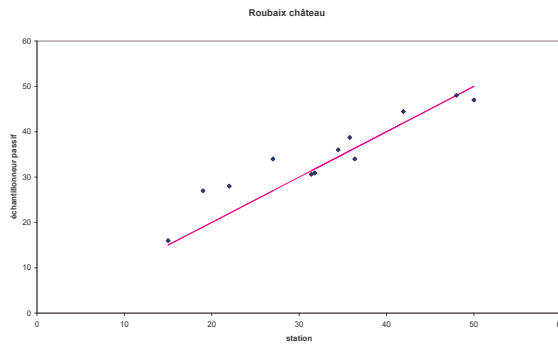
(b) Halluin cailloux (Périurbaine)



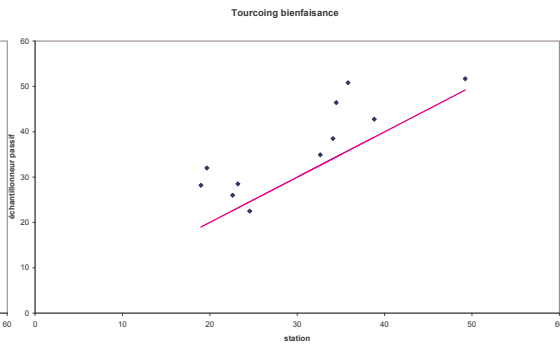
(c) Lille Fives Lakanal (urbaine)



(d) Marcq-en-Baroeul CTM (urbaine)



(e) Roubaix château (urbaine)



(f) Tourcoing bienfaisance (urbaine)

Figure 1 – Relation ($[\text{NO}_2]$ en $\mu\text{g}/\text{m}^3$) entre une station automatique (en abscisse) et un échantillonneur passif (en ordonnée) placé à côté (la droite en trait plein est la 1^{ère} bissectrice).

Tableau 4 – Equation de la régression linéaire entre une station automatique et un échantillonneur passif placé à côté (P : périurbaine ; U : urbaine).

Station automatique	Equation de la régression linéaire entre échantillonneur passif (Y) et station (x)	Coefficient de détermination (r ²)	Nombre de points	Ecart-type résiduel $\mu\text{g}/\text{m}^3$
Armentières Beauvais (P)	$Y = 0,8914x + 1,1367$	0,8613	11	3,9
Halluin Cailloux (P)	$Y = 0,9464x + 3,7221$	0,8681	10	4,2
Lille Fives Lakanal (U)	$Y = 0,9201x + 1,1844$	0,8084	13	4,6
Marcq-en-Baroeul CTM (U)	$Y = 1,1316x - 4,4059$	0,8838	11	3,9
Roubaix château (U)	$Y = 0,8028x + 8,2878$	0,9	12	3,0

Recherche des régressions linéaires entre un analyseur fixe automatique donné et différents sites d'échantillonnage passif de la zone d'étude.

Methodologie

Pour chacune des cinq stations automatiques retenues, nous avons constitué une matrice de données, sous EXCEL, de dimension (252×15) comprenant :

- en ligne, la station automatique considérée et les 251 sites différents d'échantillonnage passif de la zone d'étude ;
- en colonne, les mesures effectuées durant les 15 campagnes de mesure.

Tableau 5 – Matrice des données pour la recherche de la régression linéaire entre une station automatique donnée et les 251 sites d'échantillonnage passif de la zone d'étude.

	1 ^{ere} Campagne	2 ^{eme} Campagne	...	j ^{eme} campagne	...	15 ^{eme} Campagne
Station automatique	X _{st,1}	X _{st,2}	...	X _{st,j}	...	X _{st,18}
Site passif 1	X _{1,1}	X _{1,2}	X _{1,15}
Site passif 2	X _{2,1}	X _{2,2}	X _{2,15}
...
Site passif i	X _{i,j}
...
Site passif 251	X _{251,1}	X _{251,2}	X _{251,15}

Afin d'éviter l'obtention de droites de régression trop peu renseignées, nous avons calculé pour chaque site passif le nombre de points disponibles, trié les sites par ordre décroissant selon le nombre de points, et éliminé les sites dont le nombre de points était inférieur ou égal à 9. Nous avons ainsi obtenu une matrice de taille (125×15) plus réduite.

Nous avons appliqué à cette matrice (125×15) la procédure suivante, en trois étapes :

- Etape 1
Nous avons calculé, pour chacun des 124 sites passifs retenus, le coefficient de détermination (r^2) de la régression linéaire ($y = a.x + b$) entre la station automatique en abscisse (x) et le site passif en ordonnée (y). Nous avons trié les données selon le coefficient de détermination (r^2) par ordre décroissant, et éliminé les sites pour lesquels le coefficient de détermination était strictement inférieur à 0,80. Autrement dit, nous avons gardé les sites passifs pour lesquels la régression linéaire expliquait au moins 80% de la variance totale des données.
- Etape 2
Nous avons tracé, pour chaque site passif, la régression linéaire trouvée et testé visuellement la présence de points sur toute la gamme de concentration. Les sites pour lesquels la droite de régression ne comportait que deux classes de valeurs, par exemple une classe de valeurs faibles et une classe de valeurs hautes sans classe de valeurs intermédiaires, ont été éliminés.
- Etape 3
La présence de valeurs aberrantes a été recherchée. Celles-ci appartenaient le plus souvent à la campagne n°10 (hiver 1999), dont les valeurs par échantillonnage passif semblent sous-estimées.

Résultats

Après application de cette procédure aux 5 stations du tableau 4, nous avons remarqué que la plupart des sites corrélés à une station de typologie donnée étaient aussi corrélés à une autre station de même typologie.

Autrement dit, de nombreux sites passifs corrélés sont communs aux deux stations périurbaines « Armentières Beauvais » et « Halluin Cailloux ». De même, de nombreux sites sont communs aux deux stations urbaines « Roubaix Château » et « Lille Fives Lakanal ».

Comme nous le verrons en seconde partie de ce rapport, cela peut s'expliquer par l'existence d'une forte corrélation entre les valeurs mesurées par des stations de même typologie, en particulier lorsque les valeurs comparées sont moyennées sur de longues périodes (telle qu'une période de 14 jours).

Etant donné cette similarité des résultats, nous avons donc sélectionné deux stations de typologie différente, celles qui étaient corrélées au plus grand nombre de sites passifs, et les avons nommées « stations principales ».

Il s'agit de la station urbaine « Roubaix château » et de la station périurbaine « Halluin cailloux ».

Les stations « Lille Fives Lakanal » (urbaine) et « Armentières Beauvais » (périurbaine) ont été nommées « stations secondaires ».

La station urbaine « Marcq-en-Baroeul CTM » s'est révélée mal corrélée aux différents sites passifs, en raison d'une trop forte dispersion des points de corrélation. Nous ne l'avons donc pas prise en considération pour la suite du travail.

Pour chaque station principale, nous avons recherché les sites passifs corrélés, en distinguant trois types de sites, selon la position des points de corrélations par rapport à la 1^{ère} bissectrice : proche, aud-dessus ou en-dessous de la 1^{ère} bissectrice.

Nous supposons qu'un site passif, pour lequel la régression linéaire avec la station automatique est proche de la 1^{ère} bissectrice, relève de la même typologie que la station.

Dans le cas où un même site passif est corrélé aux deux stations principales, nous avons retenu la régression la plus proche de la 1^{ère} bissectrice, ceci afin que chaque site passif ne puisse être corrélé qu'à une seule station.

Le tableau 6 présente le nombre de sites passifs corrélés à chaque station principale.

La station urbaine « Roubaix château » est corrélée à 36 sites passifs, dont 19 proches de la 1^{ère} bissectrice.

La station périurbaine « Halluin Cailloux » est corrélée à 30 sites passifs (distincts des précédents), tous de typologie périurbaine (proches de la 1^{ère} bissectrice).

C'est donc un total de 66 sites passifs qui sont corrélés aux deux stations principales.

Les annexes 1 et 2 présentent les graphes des corrélations pour chaque site. Les figures 2, 3 et 4 montrent la localisation des sites corrélés.

Le nombre total de sites passifs étant de 122², nous avons donc trouvé que 54% des sites utilisés pour l'échantillonnage passif étaient corrélés aux deux stations automatiques principales, l'une urbaine, l'autre périurbaine.

A ces 66 sites corrélés aux deux stations principales, on peut ajouter 19 autres sites, corrélés de façon plus satisfaisante avec les stations secondaires (Tab. 7).

Selon le nombre de stations automatiques considérées (de 2 à 4), on peut donc disposer de 66 à 85 sites corrélés, soit 54% à 70% des sites disponibles.

² Le nombre de sites initial est de 124, mais il faut retrancher 2 sites correspondant aux échantillonneurs placés à côté de chacune des deux stations.

Tableau 6 – Nombre de sites passifs corrélés aux stations principales « Roubaix château » et « Halluin Cailloux » (chaque site n'est corrélé qu'à une seule station)

Station automatique typologie	Roubaix château Urbaine (U)	Halluin cailloux Périurbaine (P)
sites proches de la 1 ^{ère} bissectrice	19	30
sites en-dessous de la 1 ^{ère} bissectrice	16	0
sites au-dessus de la 1 ^{ère} bissectrice	1	0
Nombre total de sites corrélés	36	30

Tableau 7 – Nombre de sites passifs additionnels, mieux corrélés aux stations secondaires « Lille Fives Lakanal » et « Armentières Beauvais » qu'aux stations principales.

Station automatique typologie	Lille Fives Lakanal urbaine	Armentières Beauvais périurbaine
sites proches de la 1 ^{ère} bissectrice	5	4
sites en-dessous de la 1 ^{ère} bissectrice	6	0
sites au-dessus de la 1 ^{ère} bissectrice	1	3
Nombre total de sites corrélés	12	7

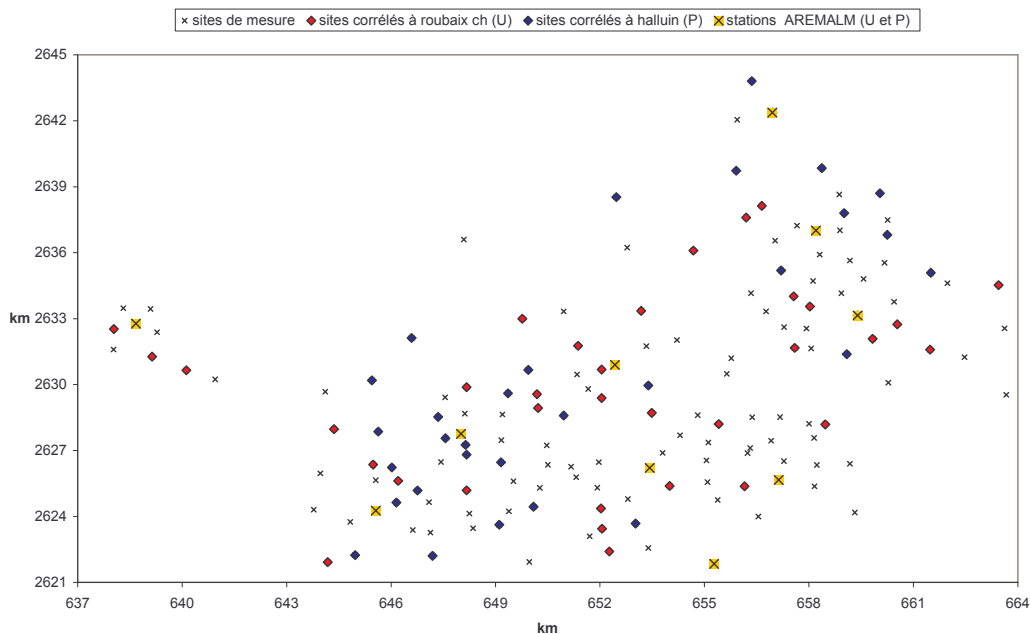


Figure 2 – Localisation des sites d'échantillonnage passif (corrélés ou non) et des stations automatiques de l'AREMA Lille Métropole pour le dioxyde d'azote.

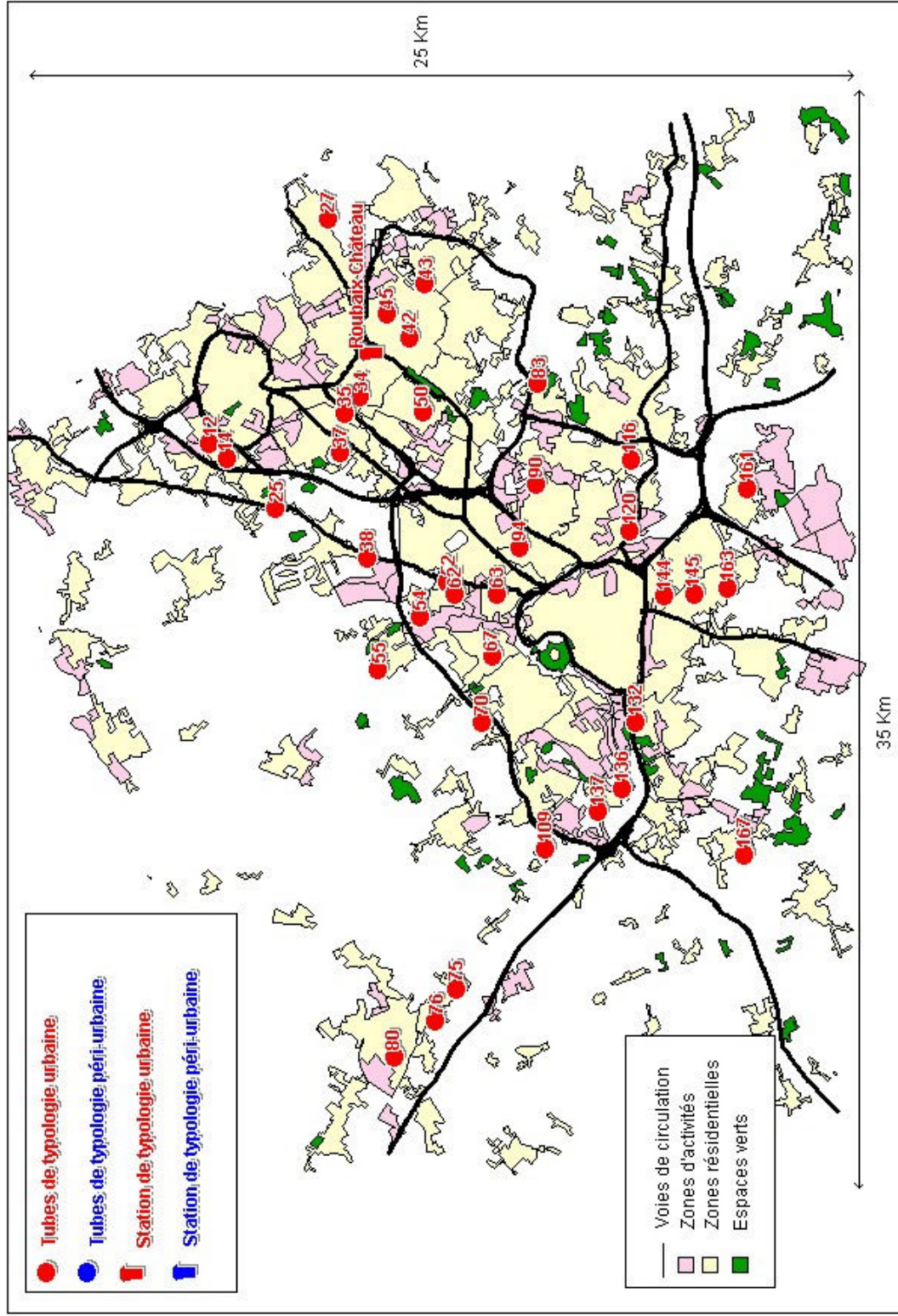


Figure 3 – Localisation des sites d'échantillonnage passif corrélés à la station urbaine « Roubaix château »

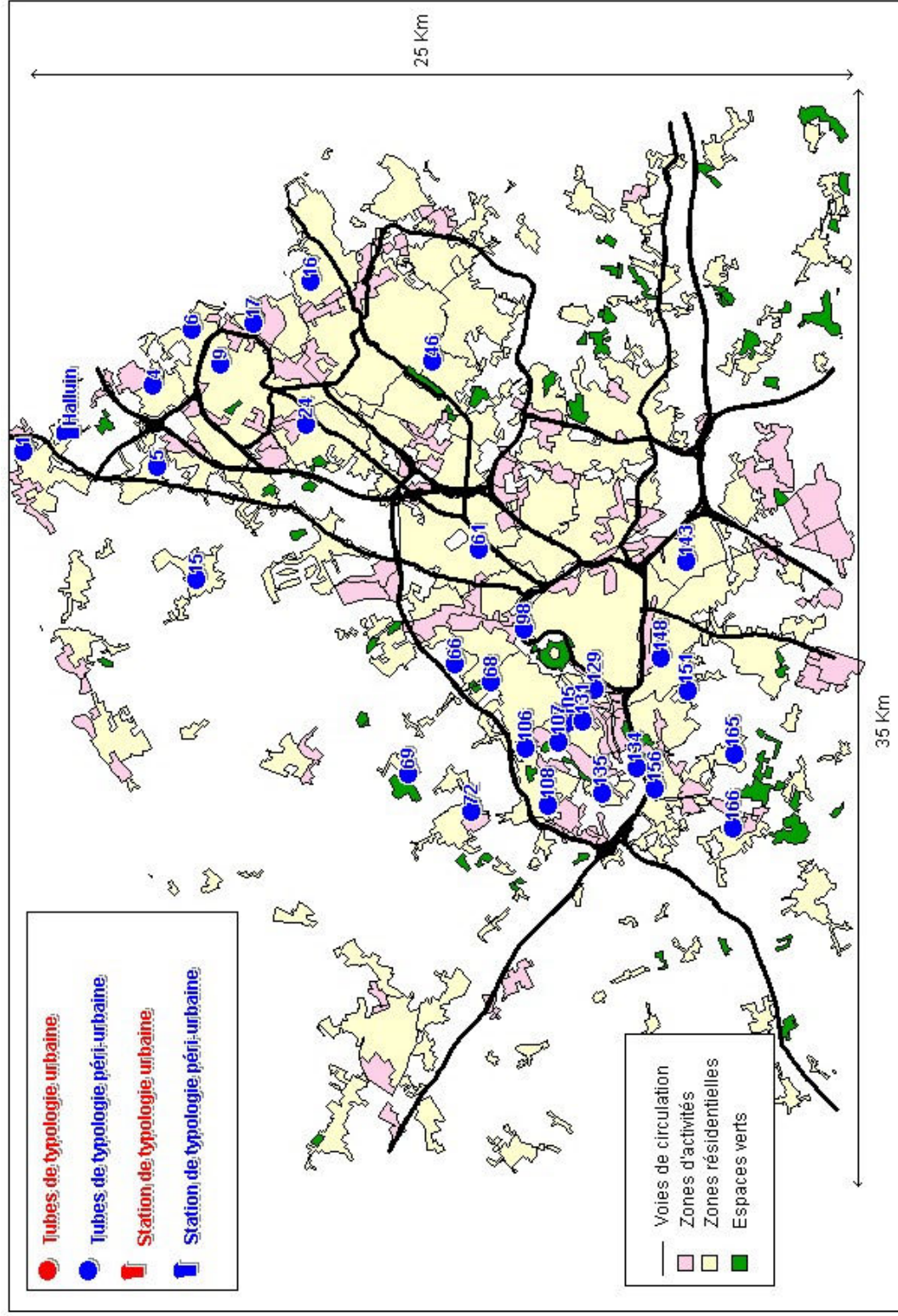


Figure 4 – Localisation des sites d'échantillonnage passif corrélés à la station périurbaine « Halluin cailloux »

Etablissement des cartes de pollution à partir des sites corrélés

Nous souhaitons savoir s'il est possible d'établir des cartes estimatives de la pollution, à partir des valeurs calculées aux sites corrélés.

On doit pour cela comparer différentes cartes :

- Il faut d'une part examiner l'effet de la diminution du nombre de valeurs connues sur la qualité de la carte obtenue ;
- il faut d'autre part tester l'effet de l'imprécision des valeurs sur la qualité des cartes, lorsque l'on passe d'une carte établie à partir de valeurs mesurées à une carte établie à partir de valeurs calculées par régression.

A titre d'illustration, nous allons présenter les résultats relatifs à la campagne de mesure n°3, menée du 24 juin au 8 juillet 1998. Les différentes cartes présentées sont établies par interpolation des données, en utilisant le logiciel ISATIS (Geovariances).

Pour cette campagne, nous disposons d'un ensemble de 116 mesures obtenues par échantillonnage passif et de 6 mesures fournies par des stations automatiques urbaines ou périurbaines (les stations automatiques de typologie « trafic » étant exclues), soit un total de 122 mesures. Le tableau 8 résume les principales caractéristiques statistiques de la campagne : les valeurs des concentrations en dioxyde d'azote sont typiques d'une période estivale, dans cette zone.

Nous avons extrait, de cette base de mesures, les données correspondant aux 2 stations automatiques « Halluin cailloux » et « Roubaix château » et aux 66 sites passifs qui leur sont corrélés. Sur ces 66 sites passifs, seuls 57 d'entre eux sont renseignés pour la campagne 3 (les 9 sites restant correspondent à des valeurs manquantes). La sélection comprend donc un total de 59 valeurs mesurées (2 stations + 57 sites passifs).

Tableau 8 – Principales caractéristiques des valeurs ($[\text{NO}_2]$ en $\mu\text{g}/\text{m}^3$) mesurées ou estimées pour la campagne n°3.

Nombre total de valeurs	Nombre de sites passifs	Nombre de stations	minimum	25 ^{ème} percentile	médiane	75 ^{ème} percentile	maximum	moyenne	écart-type
Mesures									
122	116	6	9,3	14	17,95	21,8	35,1	18,47	5,45
59	57	2	9,9	13,63	15,60	19,27	24,20	16,19	3,73
Estimations par corrélation									
68	66	2	8,63	13,81	15,93	19,38	27,08	16,47	3,68

La comparaison (Tab. 8) des caractéristiques statistiques obtenues avec l'ensemble des mesures de la campagne n°3 (122 valeurs) ou avec une sélection de mesures (59 valeurs) montre que la réduction du nombre de points de mesure s'accompagne d'un raccourcissement de l'étendue des valeurs (minimum plus

élevé, maximum plus faible), avec une diminution particulièrement importante du nombre des valeurs élevées (diminution de la médiane, du 75^{ème} percentile et de la moyenne). La sélection a aussi pour effet de réduire l'écart-type, ce qui montre une homogénéité plus importante des valeurs, par rapport au jeu de données complet. Nous verrons par la suite comment l'ensemble de ces effets peuvent s'expliquer.

Nous avons, d'autre part, estimé les valeurs aux 66 sites passifs corrélés aux stations automatiques « Halluin cailloux » et « Roubaix château », à partir des régressions linéaires. Les caractéristiques de ces valeurs estimées (Tab.8 : 68 points) sont semblables à celles correspondant à la sélection de valeurs mesurées (Tab.8 : 59 points). La figure 5 montre, en effet, que les valeurs estimées sont en bon accord avec les valeurs réellement mesurées.

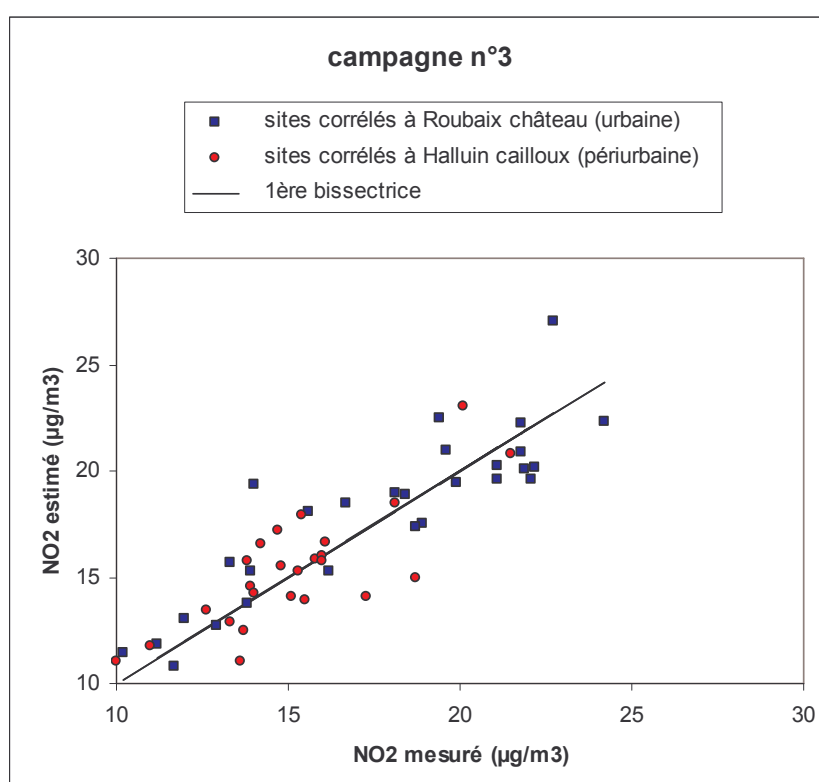


Figure 5 – Comparaison des valeurs (teneurs bi-hebdomadaires) estimées et mesurées, aux sites corrélés, pour la campagne 3.

Nous avons établi des cartes de répartition des teneurs bi-hebdomadaires en dioxyde d'azote lors de la campagne n°3, par géostatistique, pour chacun des trois cas considérés (Tab. 8) :

1. intégralité des 116 mesures;
2. sélection de 59 mesures ;
3. utilisation de 66 valeurs (estimations) calculées par corrélation.

L'interpolation des données a été effectuée par krigeage ordinaire, sur une zone de 20km×23km découpée en blocs de 1km×1km. Le variogramme, omnidirectionnel, est modélisé pour chaque jeu de données, en conservant les

mêmes de structure de base (pépitique et sphérique). Le voisinage (glissant, circulaire) est le même dans les trois cas.

Les figures 6 et 7 permettent d'apprécier l'effet de la diminution du nombre de mesures sur la qualité de la carte des teneurs bi-hebdomadaires en NO_2 : les niveaux de fond restent semblables mais certains « points chauds » disparaissent.

Pour la campagne n°3, ces points chauds sont au nombre de 4 sur un total de 123 données, soit 3,25% des points. Ils correspondent à des valeurs, mesurées par échantillonnage passif, anormalement élevées, traduisant l'existence, dans le jeu de données complet, de sites passifs de typologie « trafic » ou « industriel » (centres-ville, nœud autoroutier, émetteurs industriels importants).

Le jeu de données complet comprend, en particulier, des mesures effectuées au centre-ville de Lille, influencées par le trafic. Or, comme on peut le vérifier sur les figures 3 et 4, les sites passifs du centre-ville de Lille ne sont pas corrélés aux deux stations automatiques « Roubaix château » et « Halluin cailloux », utilisées pour les corrélations, qui sont respectivement de typologie « urbaine » et « périurbaine ».

La disparition de ces « points chauds » explique ainsi les différences, constatées précédemment (Tab.8), entre les caractéristiques statistiques du jeu de données complet et celles de la sélection des données corrélées aux stations « Roubaix château » et « Halluin cailloux ».

Nous avons finalement comparé les cartes de pollution obtenues à partir des mesures (Fig. 6 et 7) à celle obtenue à partir des valeurs calculées à l'aide des régressions linéaires (Fig. 8).

Comme le laissait prévoir la bonne adéquation entre valeurs estimées et mesurées (Fig. 5), ces deux cartes sont très semblables.

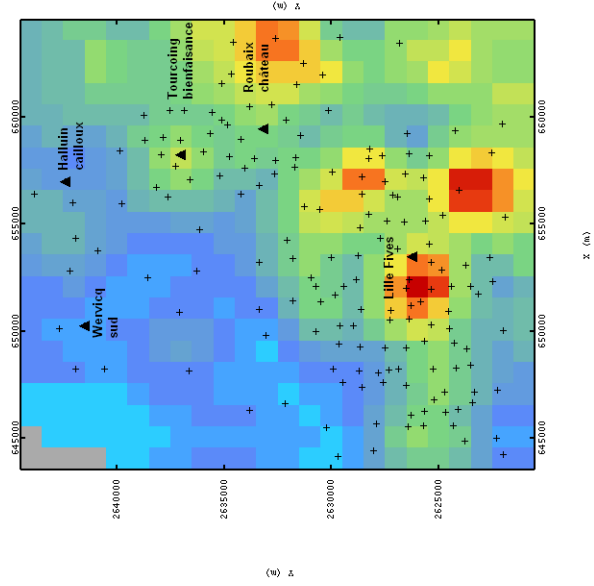


Figure 6 – Carte des concentrations bi-hebdomadaires en NO₂ pendant la campagne n°3, établie par krigeage à partir de 6 mesures de stations automatiques³ urbaines ou périurbaines et de 116 mesures par échantillonnage passif.

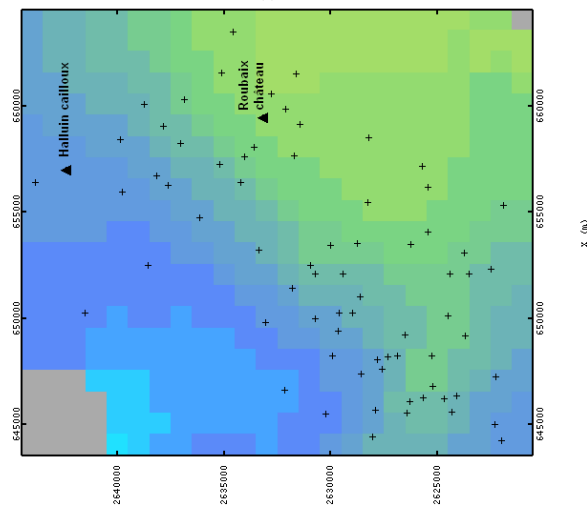


Figure 7 – Carte des concentrations bi-hebdomadaires en NO₂ pendant la campagne n°3, établie par krigeage à partir de 2 mesures provenant des stations automatiques « Halluin cailloux » et « Roubaix chateau » et de 57 mesures relatives aux sites passifs qui leur sont corrélés.

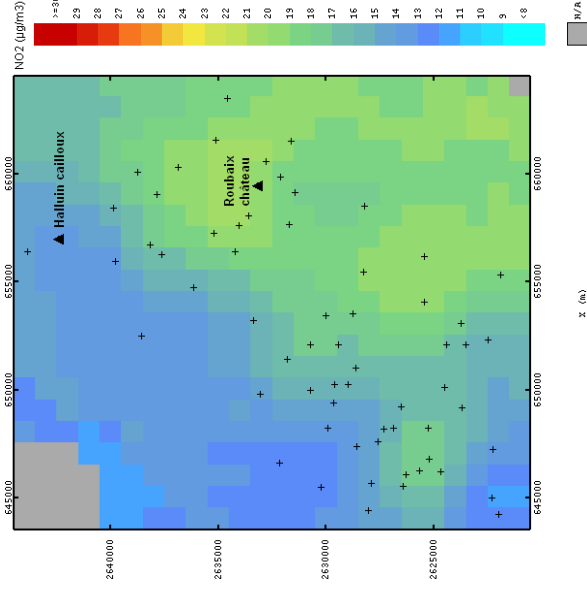


Figure 8 – Carte des concentrations bi-hebdomadaires en NO₂ pendant la campagne n°3, établie par krigeage à partir des 2 mesures des stations automatiques « Roubaix chateau » et de 66 valeurs calculées à partir des régressions linéaires aux sites passifs corrélés à ces deux stations.

³ l'une des stations automatiques se trouve en-dehors de la carte.

En conclusion de cette première partie, il semble possible d'établir, avec une assez bonne précision, des cartes bi-hebdomadaires estimatives de la pollution de fond en NO_2 , à partir des valeurs mesurées par deux stations automatiques, l'une urbaine et l'autre péri-urbaine, et des valeurs calculées par corrélation avec ces deux stations automatiques.

Le nombre de stations automatiques, utilisées pour les corrélations, peut être assez restreint : uniquement 2 stations dans le cas de la zone de 460 km^2 considérée dans cette étude. Mais il semble important de choisir des stations de typologie différente, l'une urbaine et l'autre périurbaine, afin d'assurer une bonne complémentarité des sites corrélés.

Les sites d'échantillonnage passif, corrélés aux stations automatiques urbaine ou péri-urbaine, sont des sites de fond (fond urbain, péri-urbain ou rural). En effet, les sites d'échantillonnage, placés en centre-ville, à proximité des nœuds autoroutiers ou des émetteurs industriels importants, ne sont pas corrélés de façon satisfaisante aux stations automatiques (urbaine ou périurbaine).

Cela se traduit, au niveau des cartes de pollution, par une disparition des zones de fortes concentrations en NO_2 , lorsque seules les valeurs aux sites corrélés sont considérées. On obtient ainsi des cartes de la pollution de fond, qui excluent les phénomènes de pollution locale (« points chauds »), de faible portée spatiale.

L'obtention d'un grand nombre de sites corrélés et, par conséquent, l'application de la méthode seront donc plus aisées dans des zones urbaines de fond, périurbaines ou rurales. Ces zones présentent généralement un gradient de pollution faible et continu, de grande portée spatiale, et ne font souvent pas l'objet d'une surveillance en continu.

Notons également que la zone étudiée dans ce rapport ne présente pas de relief notable, ce qui favorise la dispersion des polluants sur de grandes distances. Il pourrait en être autrement pour des zones présentant du relief. L'étude d'un jeu de données correspondant à une telle situation serait intéressant.

L'une des façons de pallier le manque d'informations concernant les pollutions très localisées (trafic, industrie) serait d'utiliser des variables auxiliaires, telles que le cadastre des émissions de NO_x , lors de l'établissement des cartes par krigeage.

En effet, dans le cadre de cette étude, nous n'avons pas étudié spécifiquement la phase de constitution des cartes de pollution par géostatistique. Cette phase devrait être approfondie lors d'une étude ultérieure, au cours de laquelle les données du cadastre des émissions, disponibles pour cette zone, pourraient être utilisées afin d'améliorer la précision des cartes.

DEUXIEME PARTIE : ETUDE DES CORRELATIONS ENTRE STATIONS PERMANENTES (ANALYSEURS AUTOMATIQUES) EN FONCTION DU PAS DE TEMPS CHOISI

Rappel des objectifs et de la méthodologie

Dans cette deuxième partie, nous étudierons les régressions linéaires entre stations automatiques, dans le but de transposer les conclusions tirées de l'étude des régressions entre stations au cas des régressions entre une station et un site passif.

Nous souhaiterions savoir si des corrélations linéaires significatives, calculées pour des valeurs moyennées sur 14 jours, permettent d'estimer des valeurs moyennées sur des durées plus courtes (7 jours ou 24h), avec une précision suffisante au regard des objectifs de qualité des données de la directive européenne 1999/30/CE.

Celle-ci demande que l'exactitude des données obtenues par modélisation soit au minimum de 50%, pour des moyennes journalières.

S'il s'agit d'estimation objective, l'objectif d'exactitude n'est plus que de 75%.

Si les corrélations obtenues sur une base temporelle de 14 jours permettent d'estimer convenablement des valeurs sur des bases temporelles de 7 jours ou de 24h, alors il sera possible d'utiliser les corrélations établies avec les sites passifs, sur une base de 14 jours, pour estimer les valeurs hebdomadaires ou journalières en ces sites et établir, par interpolation spatiale, des cartes de pollution hebdomadaires ou journalières.

Notre démarche comporte trois étapes :

- 1) Calcul des régressions linéaires entre stations automatiques, sur une base temporelle de 14 jours, à partir de données établies sur plusieurs années.
- 2) Vérification de la validité des régressions pour une année différente.
- 3) Estimation des valeurs moyennées sur 7 jours ou 24h à partir des régressions obtenues sur une base de 14 jours, et calcul de la précision des estimations.

Les différents calculs (paramètres et exactitude des régressions ; estimations) ont été menés sous EXCEL. Un classeur EXCEL comportant des liaisons entre les différentes feuilles de calcul⁴ et différentes macros a été développé. Il est présenté à l'Annexe 3 et est disponible sur demande (contact : André WROBLEWSKI, Tel : 03 27 71 26 11, Email : wroblewski@ensm-douai.fr).

⁴ traitement des données issues de la BDQA, calcul des corrélations entre stations prises 2 à 2, synthèse des résultats des paramètres et de l'exactitude des régressions pour tous les couples de stations.

Présentation du jeu de données

Les stations automatiques considérées sont les cinq stations qui ont servi, en première partie de ce rapport, à étudier les corrélations entre station et sites passifs. Il s'agit de trois stations urbaines (Lille Fives Lakanal, Marcq-en-Baroeul CTM et Roubaix château) et de deux stations périurbaines (Armentières Beauvais et Halluin cailloux).

Nous disposons des données moyennées sur 24h de chacune de ces stations pour les années 1998 à 2003 (correspondant aux années où ont eu lieu les campagnes de mesure par échantillonnage passif).

Calcul des régressions linéaires entre stations automatiques, sur une base temporelle de 14 jours, 7 jours et 24h

Corrélations pour des valeurs bihebdomadaires

Nous avons calculé les régressions linéaires entre stations automatiques, à partir des données des années 1998 à 2003, à l'exclusion de l'année 2002 qui est gardée pour validation. Ces données sont moyennées sur 14 jours, afin de simuler des données obtenues par échantillonnage passif.

Le tableau 9 présente les paramètres des régressions linéaires obtenues :

- le coefficient de détermination (r^2), égal au rapport de la variance expliquée par le modèle sur la variance totale des données;
- la pente de la droite de régression (a) et son ordonnée à l'origine (b) ;
- la moyenne des valeurs de la station de référence (en abscisse) ;
- la moyenne de l'erreur d'estimation des valeurs de la station en ordonnée à partir des valeurs de la station de référence ;
- la variance de l'erreur d'estimation, calculée selon l'équation (1) :

$$\sigma_{\text{erreur}}^2 = \frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2 \quad (\text{équation 1})$$

avec e_j : erreur commise lors de l'estimation de la $i^{\text{ème}}$ valeur ;
 \bar{e} : moyenne des erreurs ;
 n : nombre total de valeurs.

- la précision⁵ moyenne, calculée selon l'équation (2) :

$$\text{precis_moy} = \frac{\sqrt{\sigma_{\text{erreur}}^2}}{\bar{y}} \quad (\text{équation 2})$$

avec \bar{y} : moyenne des valeurs mesurées de la station en ordonnée.

⁵ Les termes « précision » et « exactitude » sont utilisés indifféremment dans ce rapport, tout comme dans la version française de l'Annexe VIII (objectifs de qualité des données) de la directive 1999/30/CE. Cependant, selon le VIM (Vocabulaire International des termes fondamentaux et généraux de Métrologie, AFNOR NFX 07-001, déc. 1994), le terme adéquat est « exactitude ».

- la moyenne de la précision absolue, calculée selon l'équation (3) :

$$\text{moy_precis} = \frac{1}{n} \sum_{i=1}^n \frac{|e_i|}{y_i} \quad (\text{équation 3})$$

avec y_j : i^{ème} valeur mesurée de la station en ordonnée.

Tableau 9 – Caractéristiques des régressions linéaires entre stations automatiques, calculées pour des valeurs moyennées sur une base de 14 jours.

(chaque régression est établie pour un nombre de données compris entre 887 et 1310).

U : urbaine

P : péri-urbaine

Tableau récapitulatif pour les moyennes bi-hebdomadaires

Stations sti/stj sti = f(stj)	r ²	a a sti(x)	b b sti(x)	moy sti moy sti	moy_err l_stj-(a.val_sti+b)	var_err racine(var_err)/moy_sti	precis_moy	moy precis
Marcq / Roubaix	0,64	0,61	15,35	36,07	0,00	40,91	0,18	0,15
Roubaix / Marcq	0,64	0,68	9,44	33,91	0,00	45,39	0,20	0,17
Marcq / Fives	0,69	0,49	16,24	35,87	0,00	40,59	0,18	0,16
Fives / Marcq	0,69	0,97	5,20	39,82	0,00	79,52	0,22	0,21
Roubaix / Fives	0,82	0,68	7,05	33,93	0,00	28,92	0,16	0,13
Fives / Roubaix	0,82	1,00	6,00	39,78	0,00	42,61	0,16	0,13
MOYENNE U / U	0,72	0,74	9,88	36,56	0,00	46,32	0,18	0,16
Marcq / Armentières	0,75	0,83	11,13	35,12	0,00	30,90	0,16	0,13
Armentières / Marcq	0,75	0,68	5,10	28,85	0,00	25,12	0,17	0,14
Marcq / Halluin	0,71	0,73	17,19	36,49	0,00	38,97	0,17	0,15
Halluin / Marcq	0,71	0,69	1,18	26,49	0,00	37,12	0,23	0,20
Roubaix / Armentières	0,84	1,01	3,26	32,86	0,00	25,23	0,15	0,14
Armentières / Roubaix	0,84	0,70	6,33	29,33	0,00	17,50	0,14	0,13
Roubaix / Halluin	0,64	0,67	15,46	32,50	0,00	44,04	0,20	0,18
Halluin / Roubaix	0,64	0,61	5,73	25,61	0,00	40,50	0,25	0,22
Fives / Armentières	0,77	1,10	5,66	37,31	0,00	51,25	0,19	0,17
Armentières / Fives	0,77	0,54	8,67	28,84	0,00	25,26	0,17	0,15
Fives / Halluin	0,72	0,99	13,86	39,52	0,00	68,55	0,21	0,19
Halluin / Fives	0,72	0,52	5,35	25,92	0,00	36,02	0,23	0,21
MOYENNE U / P	0,74	0,76	8,24	31,57	0,00	36,71	0,19	0,17
Armentières / Halluin	0,79	0,70	11,30	29,17	0,00	24,18	0,17	0,13
Halluin / Armentières	0,79	0,88	-0,21	25,43	0,00	30,24	0,22	0,15
MOYENNE P / P	0,79	0,79	5,54	27,30	0,00	27,21	0,19	0,14

On constate que les stations automatiques sont corrélées linéairement entre elles, avec un coefficient de détermination moyen de 0,75 (toutes typologies de stations confondues). Cela signifie que ces modèles linéaires permettent, en moyenne, d'expliquer 75% de la variance totale des données. Autrement dit, les phénomènes qui régissent à 75% les variations de la concentration bi-hebdomadaire en NO₂ sont les mêmes pour les stations urbaines et périurbaines étudiées (on peut penser, par exemple, aux conditions météorologiques). Précisons que la zone étudiée (au nord de la France) ne présente pratiquement pas de relief.

On remarque que certaines corrélations (telle que la régression entre les deux stations urbaines Fives=f(Roubaix), Tab. 9) ont une pente égale à 1, ce qui indique que, pour des moyennes bi-hebdomadaires, les valeurs de l'une des stations se déduisent des valeurs mesurées par l'autre station, à une constante près (égale à $6 \mu\text{g}/\text{m}^3$ dans le cas de Fives=f(Roubaix) avec Fives>Roubaix).

Dans le cas d'une relation entre des stations de typologie différente (par exemple Fives=f(Halluin) où Fives est une station urbaine et Halluin une station périurbaine), on trouve de telles corrélations mais avec une valeur de constante plus élevée ($13,9 \mu\text{g}/\text{m}^3$ pour Fives en fonction de Halluin).

L'erreur moyenne des estimations est nulle, pour toutes les régressions. Néanmoins, la dispersion des erreurs autour de cette moyenne varie d'une régression à l'autre. La variance de l'erreur, qui traduit cette dispersion, est plus faible en moyenne pour une régression entre stations périurbaines (Fig. 9) que pour une régression entre stations urbaines (Fig. 10).

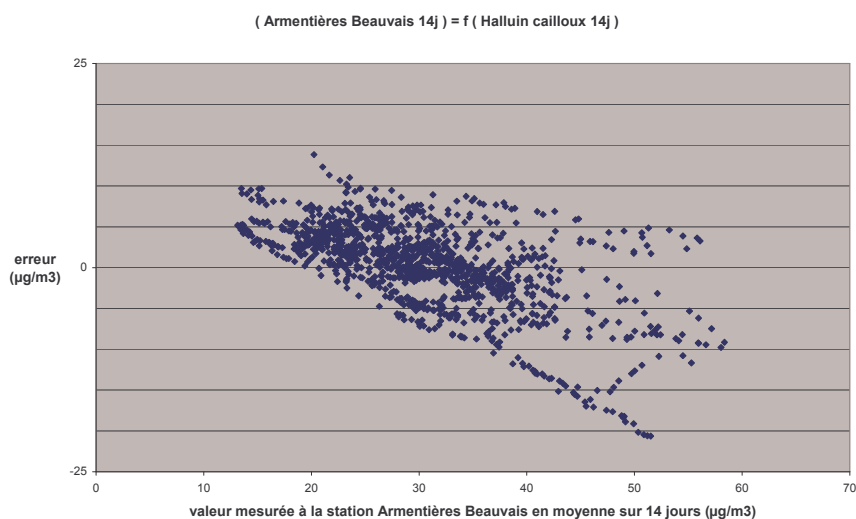


Figure 9 – Dispersion de l'erreur d'estimation dans le cas d'une corrélation entre deux stations périurbaines.

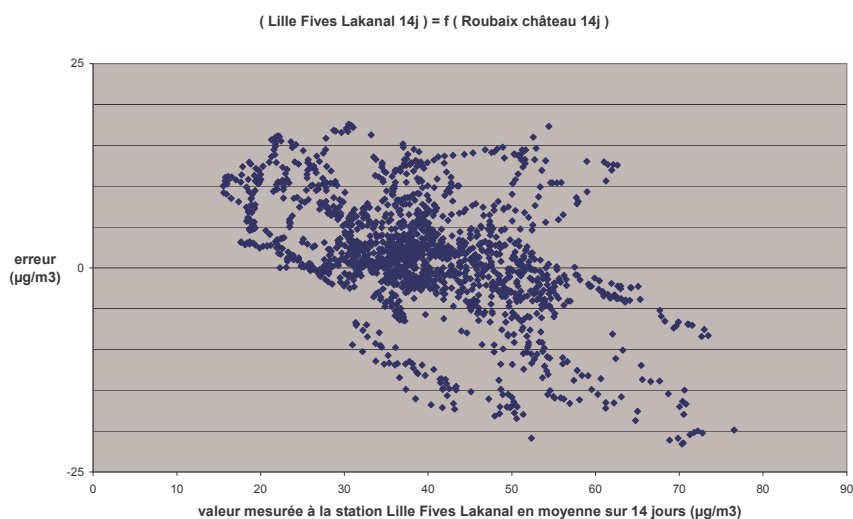


Figure 10 – Dispersion de l'erreur d'estimation dans le cas d'une corrélation entre deux stations urbaines.

L'exactitude moyenne des estimations est comprise entre 14% et 25% avec une valeur moyenne de 19%. La moyenne des exactitudes est moins élevée, comprise entre 13% et 22%, avec une valeur moyenne de 16 %.

On peut donc dire qu'en moyenne l'exactitude est bien inférieure à 50% (valeur maximale admissible pour la modélisation de moyennes journalières en NO₂, directive 1999/30/CE).

Néanmoins, les exactitudes absolues ponctuelles peuvent parfois dépasser 50% :

- pour la corrélation entre les deux stations périurbaines, Armentières en fonction de Halluin, cela représente 1,1% des cas (Fig. 11) ;
- pour la corrélation entre les deux stations urbaines, Lille Fives en fonction de Roubaix, cela représente 4,6% des cas (Fig. 12) ;

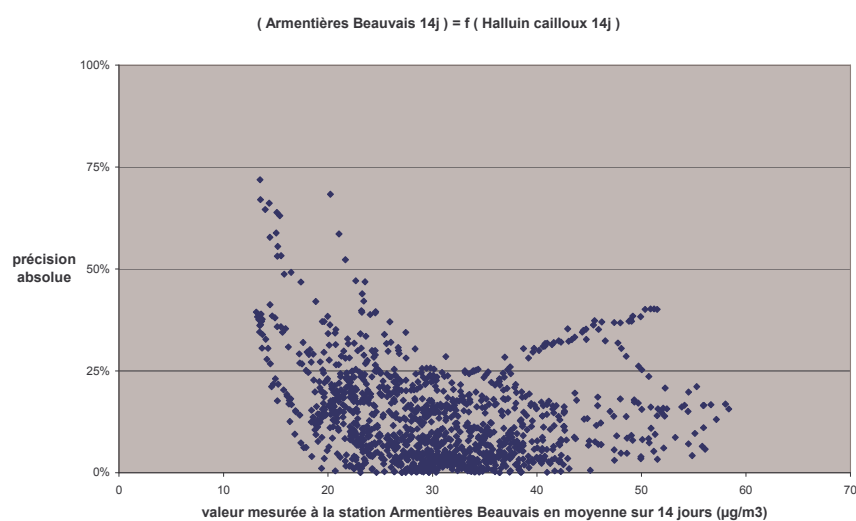


Figure 11 – Exactitude(précision) absolue dans le cas d'une corrélation entre deux stations périurbaines.

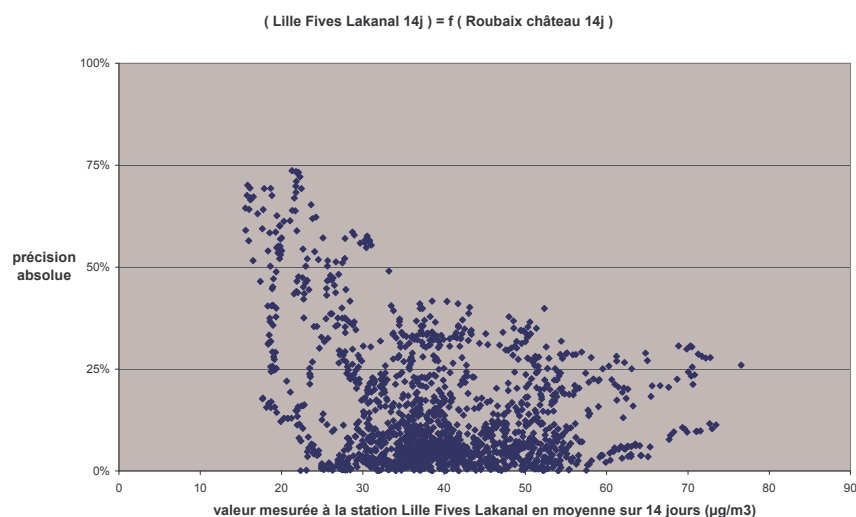


Figure 12 – Exactitude (précision) absolue dans le cas d'une corrélation entre deux stations urbaines.

Corrélations pour des valeurs hebdomadaires ou journalières.

Comme précédemment, nous avons calculé les régressions linéaires entre stations automatiques, à partir des données des années 1998 à 2003, à l'exclusion de l'année 2002 qui est gardée pour validation.

Afin de savoir si des stations corrélées sur une base de 14 jours le restent pour des pas de temps plus courts, nous avons recherché les corrélations entre stations pour des valeurs moyennées sur 7 jours ou 24h (Tab. 10 et 11).

D'après les tableaux 10 et 11, il est clair que des stations corrélées entre elles, pour des valeurs moyennées sur 14 jours, le sont aussi pour des valeurs moyennées sur 7 jours ou 24h. Néanmoins les paramètres des corrélations évoluent.

Tableau 10 – Caractéristiques des régressions linéaires entre stations automatiques, calculées pour des valeurs moyennées sur une base de 7 jours.

(chaque régression est établie pour un nombre de données compris entre 1000 et 1346).

U : urbaine

P : péri-urbaine

Tableau récapitulatif pour les moyennes hebdomadaires

Stations sti/stj sti = f(stj)	r ²	a a sti(x)	b b sti(x)	moy sti moy sti	moy_err l_stj-(a.val_sti+b)	var_err	precis_moy racine(var_err)/moy_sti	moy precis
Marcq / Roubaix	0,69	0,67	13,05	35,70	0,00	45,46	0,19	0,16
Roubaix / Marcq	0,69	0,72	8,09	33,87	0,00	49,11	0,21	0,18
Marcq / Fives	0,73	0,55	13,60	35,84	0,00	47,45	0,19	0,17
Fives / Marcq	0,73	0,96	6,11	40,58	0,00	83,31	0,22	0,21
Roubaix / Fives	0,83	0,70	6,07	33,77	0,00	33,47	0,17	0,14
Fives / Roubaix	0,83	0,98	6,65	39,61	0,00	46,71	0,17	0,14
MOYENNE U / U	0,75	0,76	8,93	36,56	0,00	50,92	0,19	0,17
Marcq / Armentières	0,79	0,87	9,60	35,17	0,00	36,05	0,17	0,15
Armentières / Marcq	0,79	0,71	4,30	29,34	0,00	29,47	0,18	0,16
Marcq / Halluin	0,76	0,81	14,53	35,61	0,00	44,24	0,19	0,16
Halluin / Marcq	0,76	0,71	0,78	26,05	0,00	38,79	0,24	0,22
Roubaix / Armentières	0,84	0,96	4,68	33,19	0,00	31,53	0,17	0,16
Armentières / Roubaix	0,84	0,74	5,19	29,72	0,00	24,29	0,17	0,15
Roubaix / Halluin	0,73	0,78	12,90	33,26	0,00	46,18	0,20	0,18
Halluin / Roubaix	0,73	0,68	3,32	26,08	0,00	40,48	0,24	0,22
Fives / Armentières	0,82	1,09	6,09	38,24	0,00	52,53	0,19	0,17
Armentières / Fives	0,82	0,61	6,14	29,57	0,00	29,60	0,18	0,16
Fives / Halluin	0,78	1,02	13,18	40,08	0,00	66,31	0,20	0,19
Halluin / Fives	0,78	0,60	2,32	26,33	0,00	38,90	0,24	0,22
MOYENNE U / P	0,79	0,80	6,92	31,89	0,00	39,86	0,20	0,18
Armentières / Halluin	0,82	0,75	9,94	28,79	0,00	23,19	0,17	0,13
Halluin / Armentières	0,82	0,91	-0,94	25,16	0,00	28,06	0,21	0,16
MOYENNE P / P	0,82	0,83	4,50	26,98	0,00	25,62	0,19	0,15

Tableau 11 – Caractéristiques des régressions linéaires entre stations automatiques, calculées pour des valeurs journalières.

(chaque régression est établie pour un nombre de données compris entre 956 et 1335).

U : urbaine

P : péri-urbaine

Tableau récapitulatif pour les moyennes 24h

Stations sti/stj sti = f(stj)	r ²	a a sti(x)	b b sti(x)	moy sti moy sti	moy_err l_stj-(a.val_sti+b)	var_err racine(var_err)/moy_sti	precis_moy	moy precis
Marcq / Roubaix	0,79	0,79	8,44	35,12	0,00	62,02	0,22	0,22
Roubaix / Marcq	0,79	0,80	5,68	33,63	0,00	62,21	0,23	0,22
Marcq / Fives	0,81	0,70	7,06	34,90	0,00	60,85	0,22	0,22
Fives / Marcq	0,81	0,93	7,13	39,60	0,00	80,55	0,23	0,22
Roubaix / Fives	0,79	0,77	3,32	33,62	0,00	42,59	0,19	0,17
Fives / Roubaix	0,79	0,98	6,18	39,27	0,00	54,33	0,19	0,16
MOYENNE U / U	0,80	0,83	6,30	36,02	0,00	60,43	0,22	0,20
Marcq / Armentières	0,87	0,98	6,19	34,84	0,00	46,79	0,20	0,18
Armentières / Marcq	0,87	0,77	2,47	29,24	0,00	36,67	0,21	0,18
Marcq / Halluin	0,85	0,90	11,67	35,03	0,00	52,50	0,21	0,20
Halluin / Marcq	0,85	0,81	-2,27	25,99	0,00	47,11	0,26	0,26
Roubaix / Armentières	0,79	0,96	4,78	33,09	0,00	51,82	0,22	0,20
Armentières / Roubaix	0,79	0,75	4,89	29,56	0,00	40,35	0,21	0,19
Roubaix / Halluin	0,79	0,76	13,65	33,84	0,00	85,20	0,27	0,25
Halluin / Roubaix	0,79	0,73	1,87	26,55	0,00	81,64	0,34	0,30
Fives / Armentières	0,79	1,08	6,49	37,92	0,00	64,56	0,21	0,20
Armentières / Fives	0,79	0,67	3,71	29,11	0,00	40,03	0,22	0,20
Fives / Halluin	0,79	1,01	13,97	39,45	0,00	84,67	0,23	0,23
Halluin / Fives	0,79	0,67	-1,02	25,25	0,00	55,85	0,30	0,29
MOYENNE U / P	0,82	0,84	5,53	31,66	0,00	57,27	0,24	0,22
Armentières / Halluin	0,88	0,81	8,31	28,63	0,00	31,36	0,20	0,16
Halluin / Armentières	0,88	0,95	-2,11	25,03	0,00	36,63	0,24	0,22
MOYENNE P / P	0,88	0,88	3,10	26,83	0,00	33,99	0,22	0,19

L'évolution des paramètres des régressions linéaires (Fig. 13) est semblable quelle que soit la typologie des stations. La pente diminue lorsque le pas de temps augmente tandis que, par compensation, l'ordonnée à l'origine augmente. Néanmoins, ces variations sont assez modérées. Il semble donc possible de pouvoir estimer convenablement une valeur hebdomadaire ou quotidienne à partir des régressions établies pour des valeurs bihebdomadaires (avec cependant une diminution probable de la qualité de l'estimation, que nous évaluerons par la suite).

Bien que les corrélations aient été optimisées pour chaque pas de temps, l'exactitude des estimations est meilleure pour les pas de temps longs que pour les pas de temps courts (Fig. 14). Cela semble cohérent avec la nature de certains phénomènes qui influent sur les concentrations en NO₂ (telles que les conditions météorologiques). En effet, l'influence peut être la même en tous points pour des pas de temps longs (influence saisonnière) mais différente pour des pas de temps courts (phénomène pluvieux localisé, par exemple).

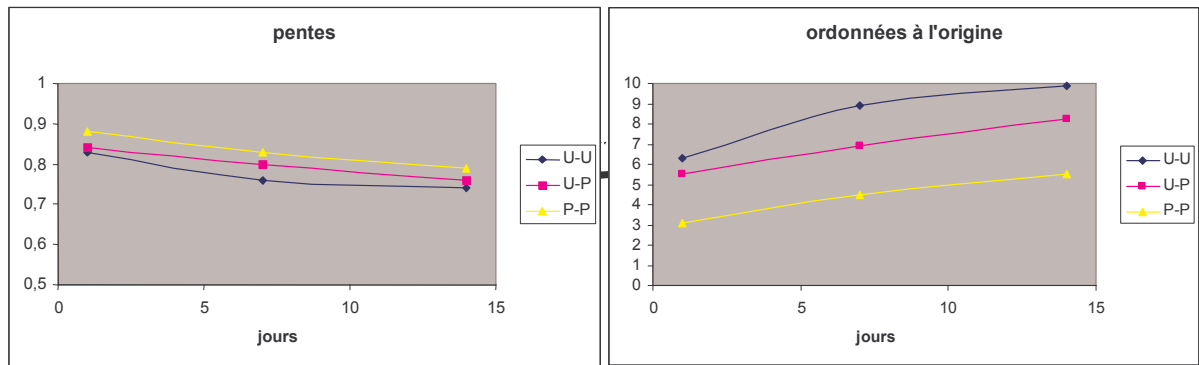


Figure 13 – Evolution des paramètres des régressions linéaires (pente et ordonnée à l'origine) en fonction du pas de temps et de la typologie des stations (U : urbaine ; P : périurbaine).

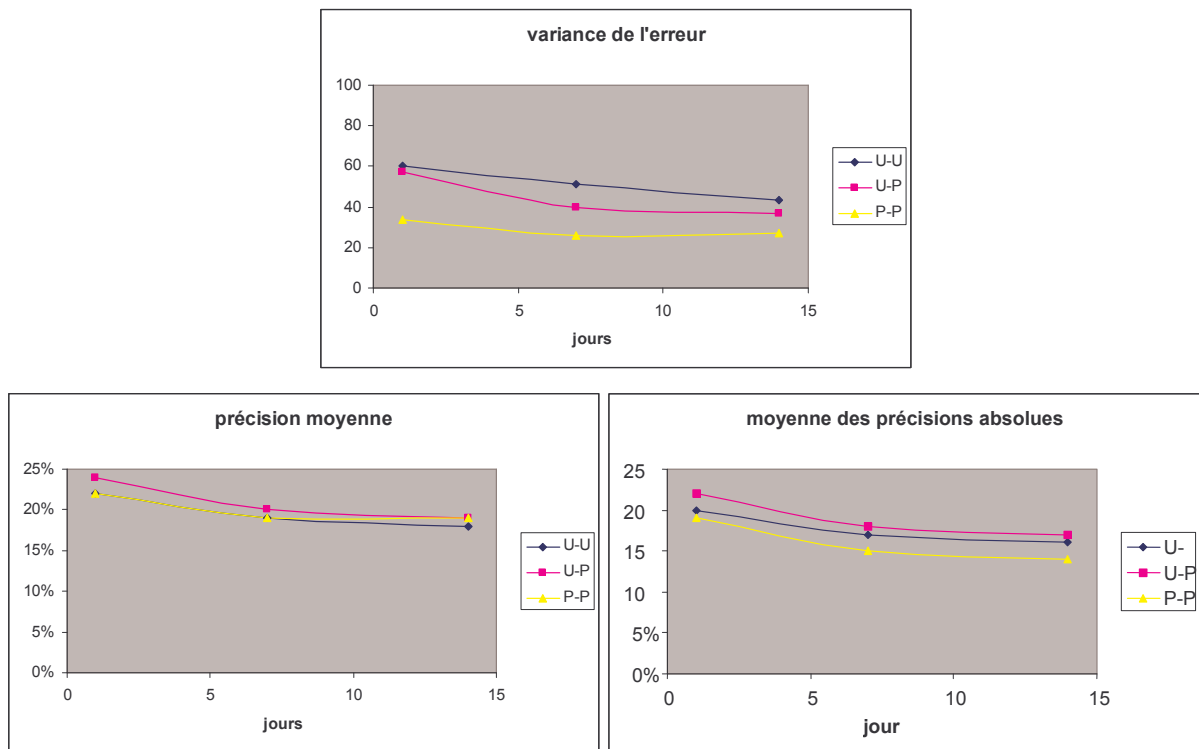


Figure 14 – Evolution des paramètres mesurant la qualité des estimations (variance de l'erreur, précision moyenne et moyenne des précisions absolues) en fonction du pas de temps et de la typologie des stations (U : urbaine ; P : périurbaine).

Validation absolue des corrélations bihebdomadaires et journalières : application aux données de l'année 2002.

Afin de tester la validité et la qualité des corrélations trouvées, nous les avons appliquées aux données mesurées au cours de l'année 2002, non utilisées pour établir les corrélations. Il s'agit donc d'un test de « validation absolue ».

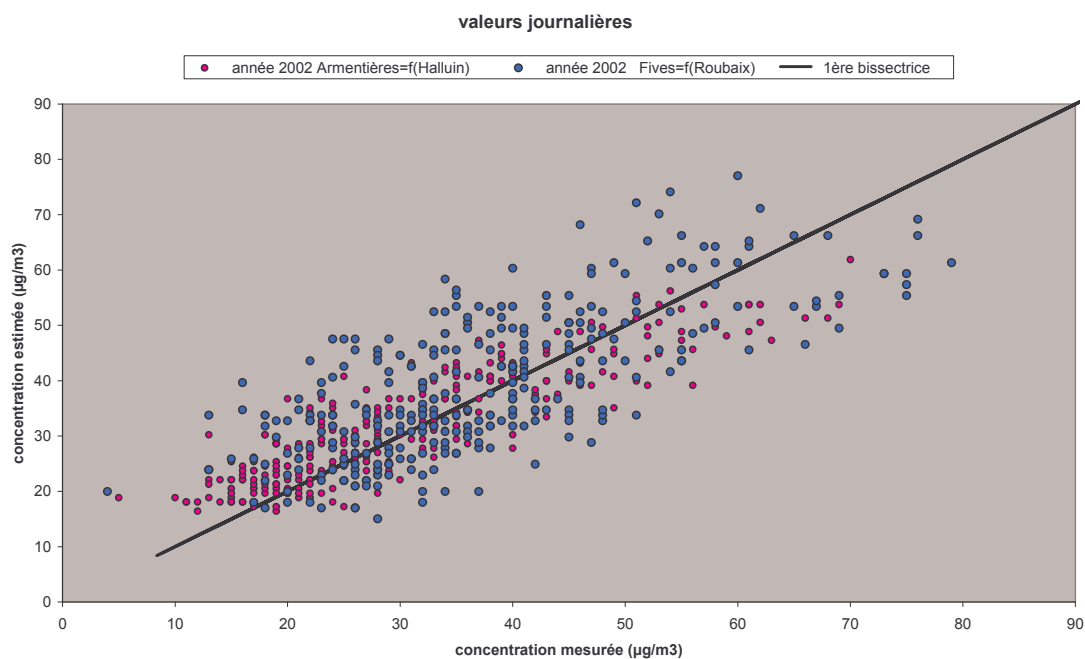


Figure 15 – Estimations des valeurs journalières (calculées à partir des corrélations établies pour les années 1998 à 2003 à l'exclusion de 2002) **pour l'année 2002.**

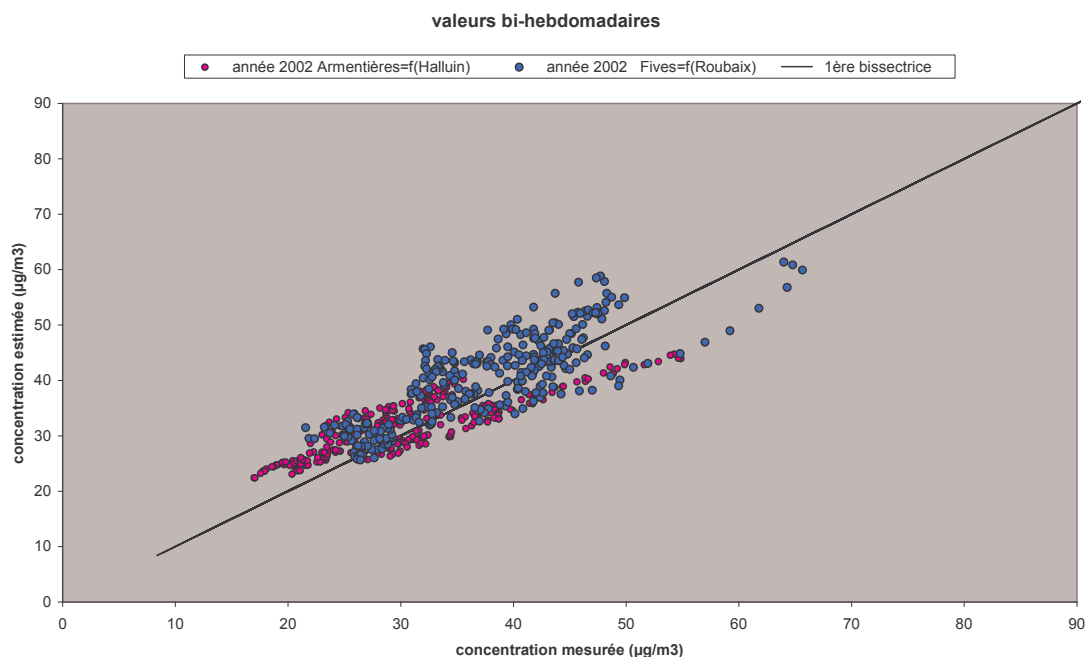


Figure 16 – Estimations des valeurs bi-hebdomadaires (calculées à partir des corrélations établies pour les années 1998 à 2003 à l'exclusion de 2002) **pour l'année 2002.**

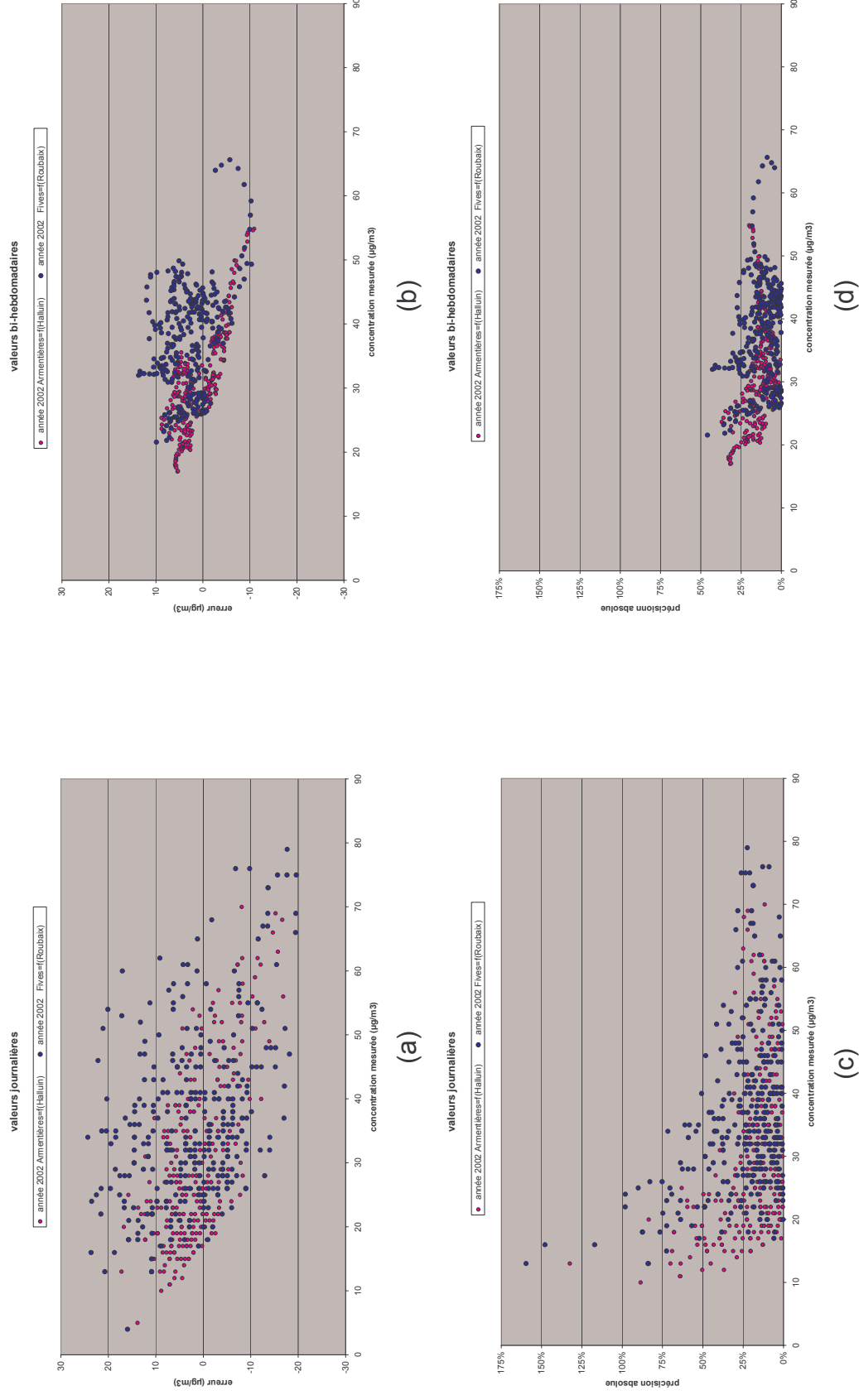


Figure 17 – Erreur (a, b) et précision absolue (c, d) pour l'estimation des données de l'année 2002, à partir des corrélations établies pour d'autres années.

La validation est globalement satisfaisante, les valeurs estimées étant proches des valeurs mesurées (Fig 15 et 16).

Concernant les erreurs d'estimation (Tab. 12), leur moyenne n'est plus nulle, ce qui témoigne du fait que la régression linéaire n'a pas été optimisée pour l'année 2002. La précision moyenne (Tab. 12) reste du même ordre de grandeur que pour les autres années, et reste meilleure pour des valeurs bi-hebdomadaires.

On peut donc conclure que les corrélations, établies pour des années données, peuvent servir à estimer les valeurs d'années différentes.

Tableau 12 – Erreur et précision obtenues lors de l'estimation des données de l'année 2002, à partir des corrélations calculées pour d'autres années.

	Pas de temps	Erreur moyenne	Précision moyenne	% de valeurs avec une erreur relative > 50%
<i>Stations périurbaine :</i>				
Armentières = f (Halluin)	24h	1,28	20%	8%
Armentières = f (Halluin)	14j	1,07	14,5%	0%
<i>Stations urbaines :</i>				
Lille Fives = f (Roubaix château)	24h	2,12	26%	11,9%
Lille Fives = f (Roubaix château)	14j	2,84	12,5%	0%

Estimation de valeurs journalières, à partir de corrélations établies pour des valeurs bi-hebdomadaires.

Nous souhaitons tester l'hypothèse selon laquelle les corrélations, établies sur une base bi-hebdomadaire entre un site passif et une station, pourraient servir à estimer les valeurs journalières au site passif à partir des valeurs journalières de la station.

Comme nous ne disposons pas de mesures journalières aux sites passifs, nous simulons la situation en considérant deux stations, l'une jouant le rôle de la station de référence, l'autre celui du site passif.

Nous avons précédemment établi les corrélations entre stations, sur une base bi-hebdomadaire, correspondant au temps nécessaire pour effectuer les prélèvements de NO₂ par échantillonnage passif (Tab. 9)

En utilisant les paramètres (pente et ordonnée à l'origine) de ces régressions bi-hebdomadaires, nous avons estimé les valeurs journalières d'une station à partir des valeurs journalières d'une autre station, et calculé l'erreur commise. Les données utilisées sont celles des années 1998 à 2003 (en excluant l'année 2002). Les résultats sont consignés au tableau 13.

Tableau 13 (U : urbaine ; P : périurbaine)

Tableau récapitulatif pour le calcul des moyennes journalières à partir des régressions bihebdomadaires

Stations sti/stj sti = f(stj)	Nombre valeurs	a a sti(x)	b b sti(x)	moy sti moy sti	moy_err _stj-(a.val_sti+b)	var_err	précision moyenne racine(var_err)/moy_sti
Marcq / Roubaix	1494	0,61	15,35	35,52	0,88	68,62	0,23
Roubaix / Marcq	1494	0,68	9,44	34,44	-0,91	64,43	0,23
Marcq / Fives	1494	0,49	16,24	34,98	0,39	68,55	0,24
Fives / Marcq	1494	0,97	5,20	38,88	0,42	82,83	0,23
Roubaix / Fives	1634	0,68	7,05	34,42	-1,12	51,51	0,21
Fives / Roubaix	1634	1,00	6,00	38,85	1,43	60,06	0,20
Moyenne U-U	1541	0,74	9,88	36,18	0,18	66,00	0,22
Marcq / Armentières	1411	0,83	11,13	35,66	-0,21	52,46	0,20
Armentières / Marcq	1411	0,68	5,10	29,25	-0,04	41,76	0,22
Marcq / Halluin	1385	0,73	17,19	35,93	1,19	54,78	0,21
Halluin / Marcq	1385	0,69	1,18	27,36	-1,26	47,94	0,25
Roubaix / Armentières	1478	1,01	3,26	34,23	-1,38	61,64	0,23
Armentières / Roubaix	1478	0,70	6,33	29,33	0,97	47,33	0,23
Roubaix / Halluin	1436	0,67	15,46	34,09	-1,00	67,17	0,24
Halluin / Roubaix	1436	0,61	5,73	26,49	0,09	54,70	0,28
Fives / Armentières	1504	1,10	5,66	37,99	-0,45	63,18	0,21
Armentières / Fives	1504	0,54	8,67	29,05	0,16	47,70	0,24
Fives / Halluin	1458	0,99	13,86	39,39	0,64	76,97	0,22
Halluin / Fives	1458	0,52	5,35	26,42	-0,58	60,62	0,29
Moyenne U-P	1445	0,76	8,24	32,10	-0,16	56,35	0,24
Armentières / Halluin	1393	0,70	11,30	29,64	0,70	37,35	0,21
Halluin / Armentières	1393	0,88	-0,21	27,09	-1,25	43,26	0,24
Moyenne P-P	1393	0,79	5,54	28,36	-0,28	40,31	0,22

Comme en témoignent les valeurs des erreurs moyennes d'estimation, proches de zéro, et des précisions moyennes, toutes inférieures à 30%, l'estimation des valeurs journalières est satisfaisante. Les graphes des figures 18 et 19 montrent que les couples de points (mesure ; estimation) sont proches de la 1^{ère} bissectrice.

Cependant, les erreurs d'estimation peuvent parfois être importantes, notamment pour des valeurs élevées (Fig. 20 et 21). Mais, exprimées en erreurs relatives, ce sont plutôt les valeurs faibles qui sont à l'origine d'erreurs relatives importantes.

L'erreur relative absolue peut dépasser 50%, dans environ 5,3% des cas (stations urbaines, Fig. 22) jusqu'à 7,3% des cas (stations périurbaines, Fig. 23). Elle dépasse même parfois 75% (exactitude admissible pour l'évaluation objective, Directive 199/30/CE) dans 1,7% à 2,3% des cas, respectivement pour les stations urbaines ou périurbaines (Fig. 22 et 23).

Par rapport aux recommandations de la Directive 1999/30/CE, il s'avère donc possible d'estimer des valeurs journalières, à partir de corrélations établies sur une base bi-hebdomadaire :

- dans environ 90% des cas si l'on considère l'exactitude admissible pour la modélisation (50%) ;
- ou dans environ 95% des cas si l'on considère l'exactitude admissible pour l'évaluation objective (75%).

Dans 5 à 10% des cas, cependant, la précision des estimations n'est pas suffisante pour répondre aux objectifs de qualité des données requis dans la Directive 1999/30/CE. Ces journées correspondent vraisemblablement à des situations météorologiques ou d'émissions particulières.

Il serait souhaitable, dans une étude ultérieure, de mieux caractériser ces situations particulières afin de mieux préciser l'applicabilité de la méthode.

L'une des voies pour y parvenir serait de classer les concentrations journalières en dioxyde d'azote en fonction des paramètres météorologiques, afin de repérer si les journées pour lesquelles l'exactitude des estimations est mauvaise se situent dans des classes particulières.

On pourrait ainsi s'appuyer sur la méthodologie de l'étude n°3 du rapport d'activités LCSQA 2002 de l'EMD, menée par Jean-Luc HOUDRET :

« influence de paramètres météorologiques sur la stratégie de mesure à l'aide de moyens mobiles ».

Le principe de cette étude est basé sur la norme :

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

Cette norme consiste à faire des ventilations des mesures de pollution en strates, élaborées selon différents critères météorologiques et temporels.

La première partie de cette étude portait sur les données journalières de plusieurs polluants atmosphériques mesurés par l'AREMALM (même zone d'étude que le présent rapport) de 1996 à 1999 dans plusieurs stations de mesure. Les paramètres météorologiques utilisés étaient les observations journalières de la vitesse et la direction du vent, la température, la pression, la hauteur d'eau, la visibilité et l'état du ciel.

L'application de cette méthodologie au cas du NO₂ permettrait de mettre en évidence les strates principales pour ce polluant. On pourrait alors s'y référer pour classer les concentrations journalières en NO₂ et identifier les situations où les corrélations fonctionnent de façon satisfaisante.

Estimation des valeurs journalières de la station "Lille Fives Lakanal" à partir des valeurs journalières de la station "Roubaix château", en utilisant la corrélation établie pour des valeurs bi-hebdomadaires.

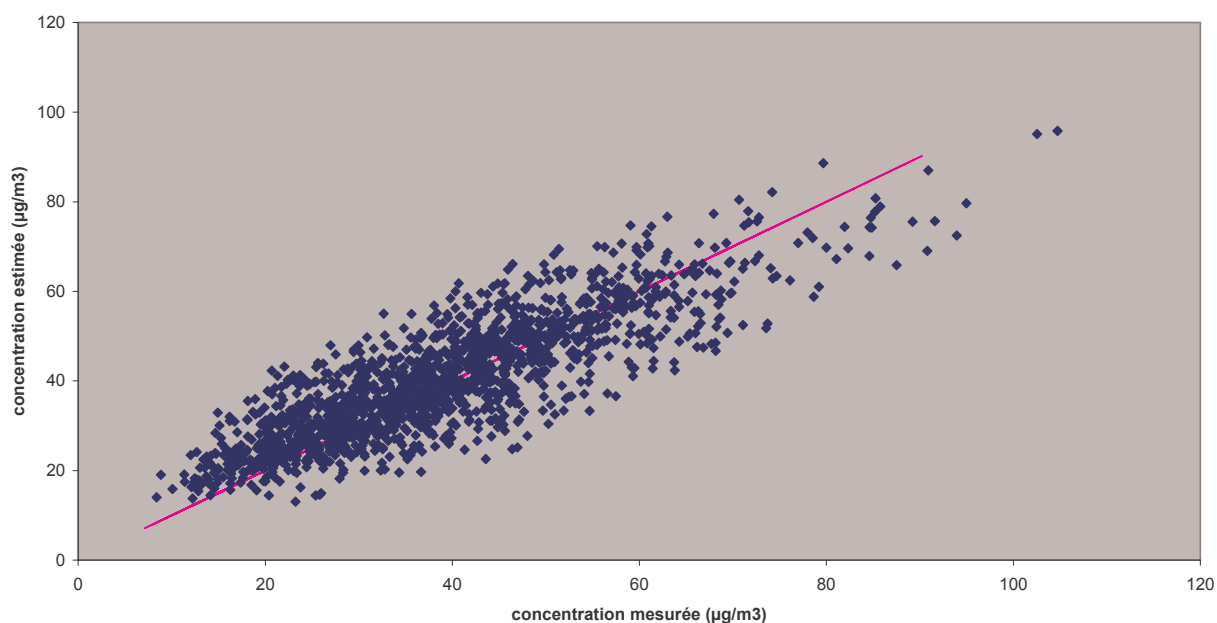


Figure 18 – Estimation des valeurs journalières d'une station urbaine (Lille Fives Lakanal) à partir d'une corrélation bihebdomadaire (avec la station urbaine Roubaix château).

Estimation des valeurs journalières de la station "Armentières Beauvais" à partir des valeurs journalières de la station "Halluin cailloux", en utilisant la corrélation établie pour des valeurs bi-hebdomadaires.

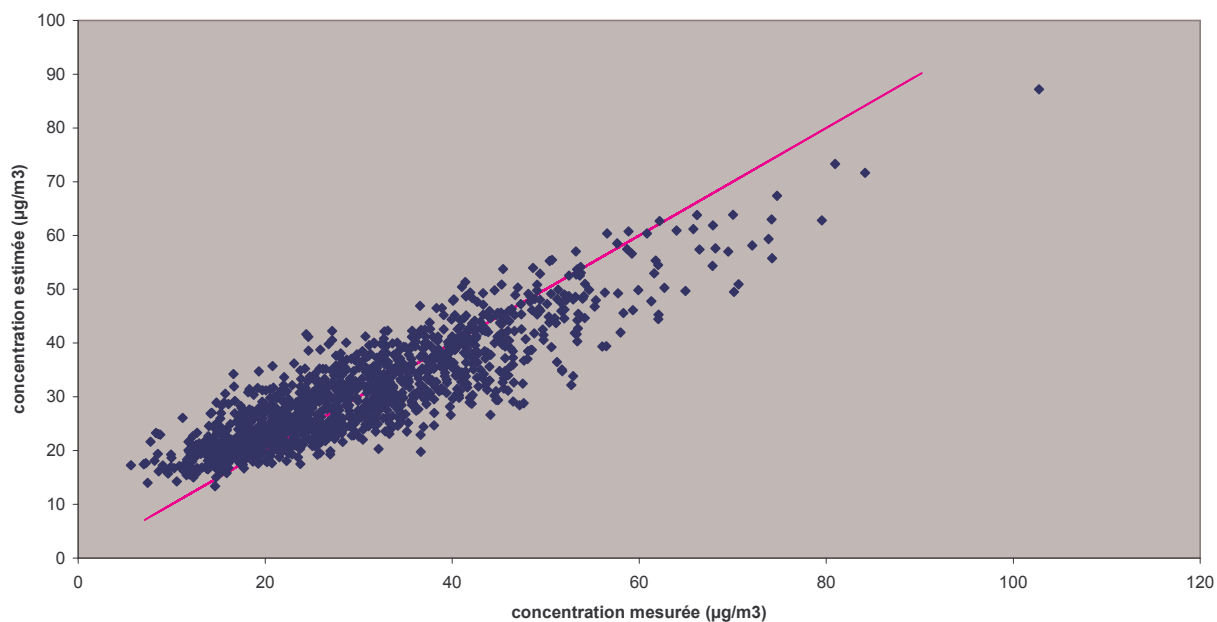


Figure 19 – Estimation des valeurs journalières d'une station périurbaine (Armentières Beauvais) à partir d'une corrélation bihebdomadaire (avec la station périurbaine Halluin Cailloux).

Erreur d'estimation des valeurs journalières de la station "Lille Fives Lakanal" à partir des valeurs journalières de la station "Roubaix château", en utilisant la corrélation établie pour des valeurs bi-hebdomadaires.

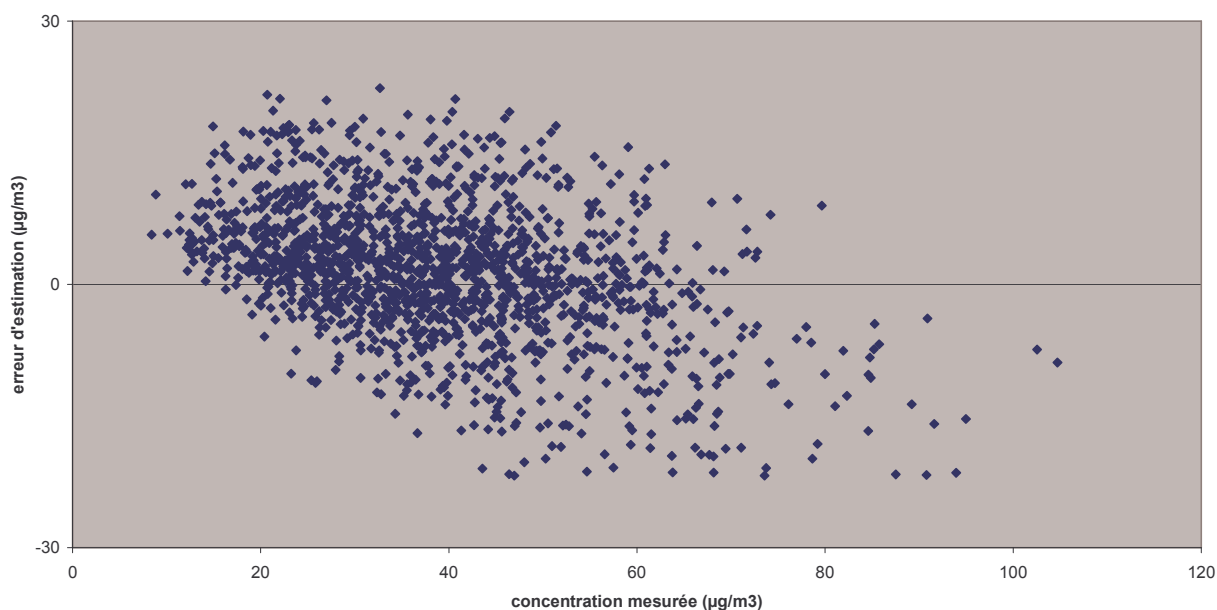


Figure 20 – Erreur d'estimation des valeurs journalières d'une station urbaine (Lille Fives) à partir d'une corrélation bihebdomadaire (avec la station urbaine Roubaix château).

Erreur d'estimation des valeurs journalières de la station "Armentières Beauvais" à partir des valeurs journalières de la station "Halluin cailloux", en utilisant la corrélation établie pour des valeurs bi-hebdomadaires.

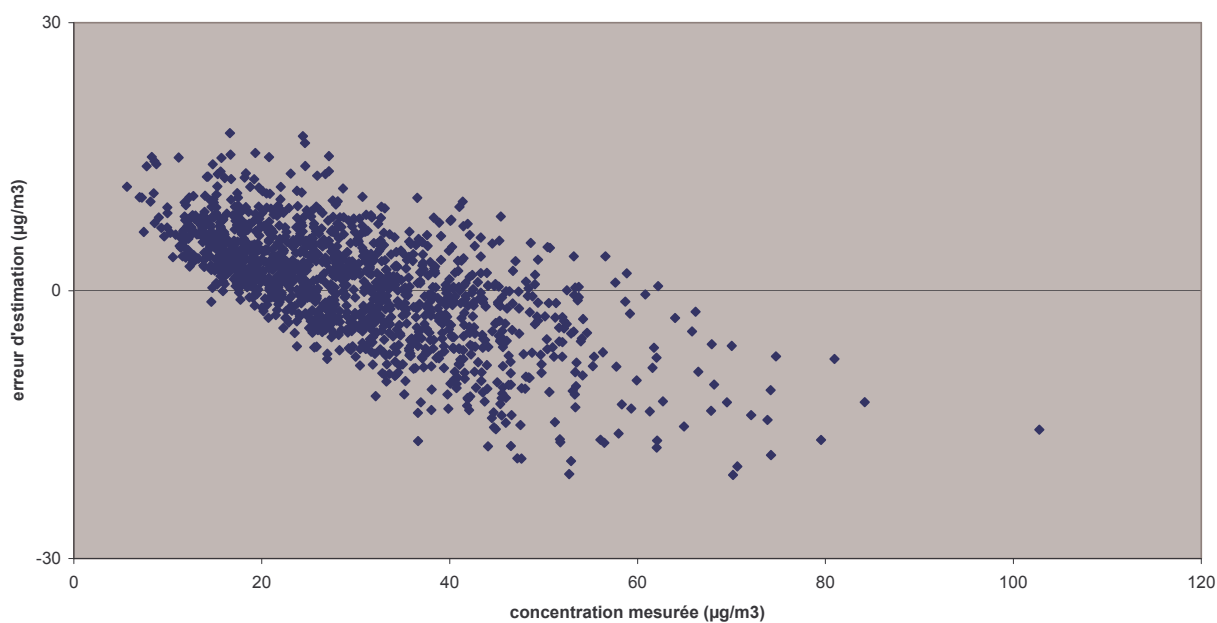


Figure 21 – Erreur d'estimation des valeurs journalières d'une station périurbaine (Armentières Beauvais) à partir d'une corrélation bihebdomadaire (avec la station périurbaine Halluin cailloux).

Précision absolue d'estimation des valeurs journalières de la station "Lille Fives Lakanal" à partir des valeurs journalières de la station "Roubaix château", en utilisant la corrélation établie pour des valeurs bi-hebdomadaires.

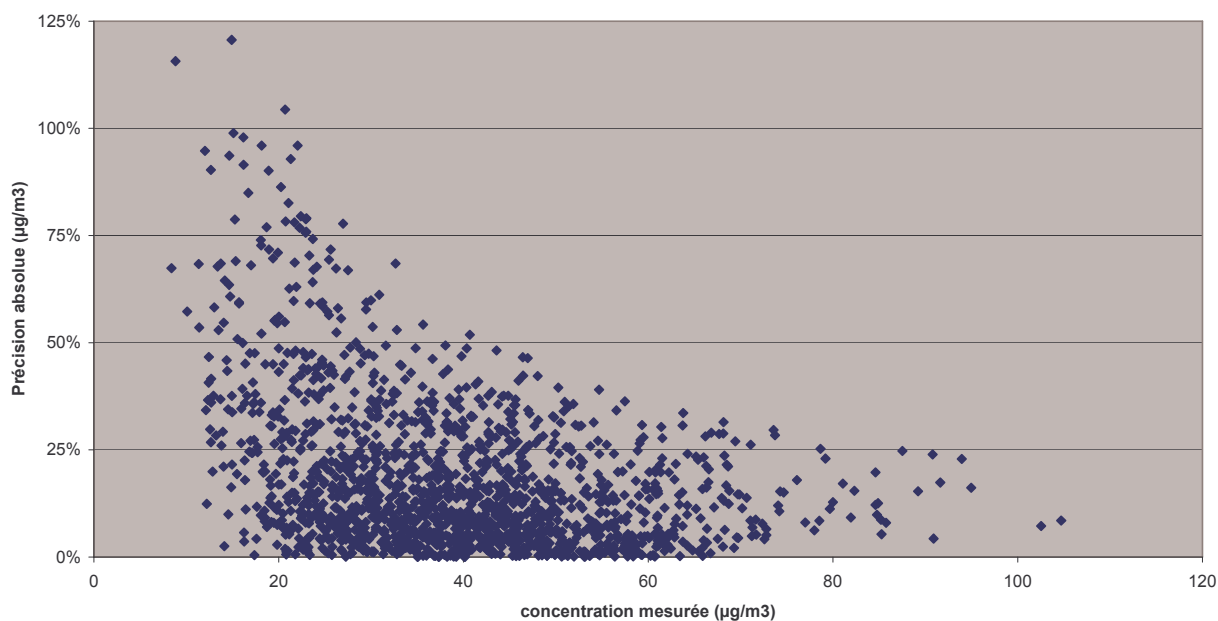


Figure 22 - Erreur relative absolue (précision absolue) des valeurs journalières d'une station urbaine (Lille Fives Lakanal) estimées à partir d'une corrélation bihebdomadaire (avec la station urbaine Roubaix château).

Précision absolue d'estimation des valeurs journalières de la station "Armentières Beauvais" à partir des valeurs journalières de la station "Halluin cailloux", en utilisant la corrélation établie pour des valeurs bi-hebdomadaires.

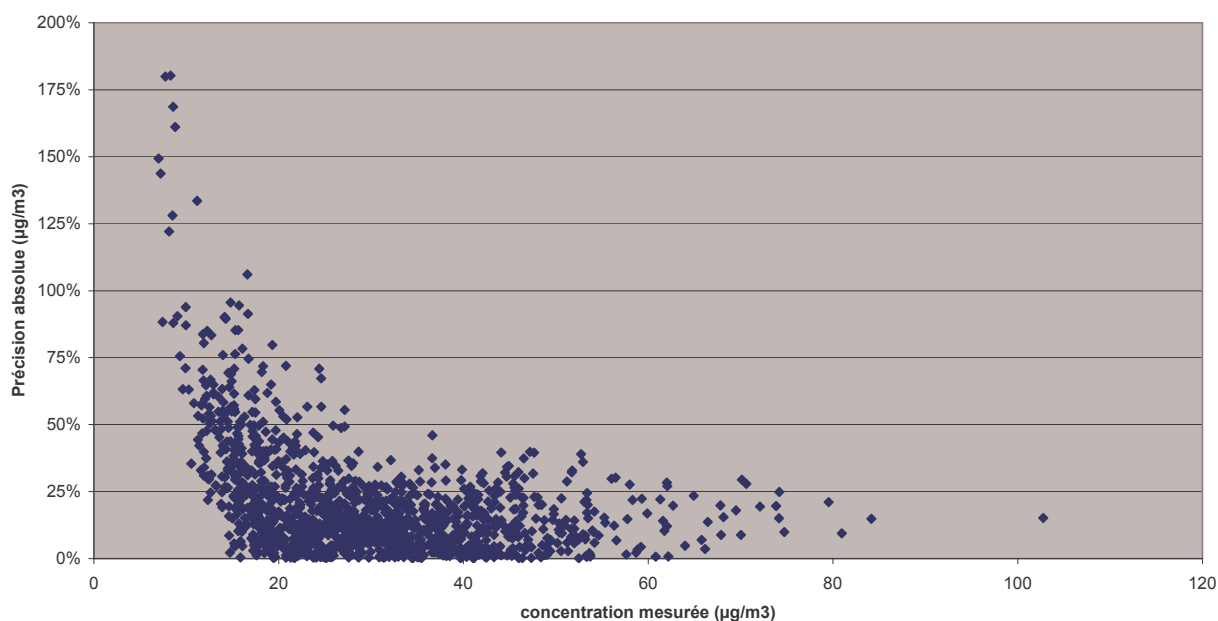


Figure 23 - Erreur relative absolue (précision absolue) des valeurs journalières d'une station périurbaine (Armentières) estimées à partir d'une corrélation bihebdomadaire (avec la station périurbaine Halluin cailloux).

CONCLUSION

Ce travail a permis de tester la faisabilité de réalisation de cartes journalières de répartition spatiale des concentrations en dioxyde d'azote, à partir des mesures fournies par quelques stations automatiques, en nombre limité, et en tirant profit de l'information fournie par des campagnes pré-existantes de mesure du dioxyde d'azote par échantillonnage passif.

Nous avons montré que, dans le cas du jeu de données étudié, correspondant à des mesures effectuées par l'AREMA Lille Métropole pendant les années 1998 à 2003, plus de la moitié des sites, utilisés régulièrement pour l'échantillonnage passif, présentaient des valeurs corrélées à celles de stations automatiques.

Les stations automatiques prises en compte sont uniquement de typologie « urbaine » ou « périurbaine ».

Ce résultat est en accord avec une étude précédente, qui avait montré que les zones de représentativité spatiale de ces types de stations pouvaient être très étendues [5] et englober, par conséquent, un grand nombre d'emplacements utilisés pour l'échantillonnage passif du NO₂.

Le nombre de sites passifs corrélés s'est révélé suffisant pour qu'une carte estimative des teneurs en NO₂ soit réalisée par interpolation des données, en utilisant le krigeage ordinaire comme technique d'interpolation.

Cette carte estimative rend bien compte des niveaux de fond, mais ne permet pas forcément de rendre compte de la pollution à proximité d'émetteurs industriels ou d'axes de grande circulation.

En effet, les sites d'échantillonnage passif placés à proximité d'une installation industrielle ou d'un axe de circulation sont majoritairement influencés par des sources locales. Ils ne peuvent donc pas être corrélés, de façon satisfaisante, avec des stations de typologie « urbaine » ou « périurbaine ».

De même, il s'est avéré difficile de trouver des sites passifs corrélés aux stations urbaines ou périurbaines dans les centres-villes, fortement influencés par le trafic automobile.

Le recours à des variables auxiliaires, telles que les sources d'émission d'oxydes d'azote, lors de l'étape d'interpolation des données par krigeage, devrait certainement permettre d'améliorer la représentativité et précision des cartes.

Il est envisagé, dans une étude ultérieure, d'optimiser et d'automatiser l'étape d'interpolation géostatistique des données, afin de pouvoir établir des cartes estimatives et fiables, quotidiennement.

Nous avons montré, également, que le nombre de stations automatiques disponibles n'était pas un facteur limitant pour l'application de la méthode. En effet, deux stations automatiques semblent suffire, à condition qu'elles soient de typologie différente : l'une urbaine, l'autre périurbaine.

Néanmoins, il serait utile, au cours d'une étude ultérieure, d'étudier des données relatives à des territoires présentant plus de relief, pour lesquels le nombre de stations automatiques nécessaires pourrait être différent.

Concernant l'estimation de concentrations journalières, à partir des mesures journalières des stations automatiques et des corrélations bi-hebdomadaires calculées, nous avons montré que, dans environ 90 à 95% des cas (selon que l'on considère l'exactitude nécessaire pour la modélisation ou pour l'estimation objective), de telles estimations pouvaient être effectuées, avec une précision suffisante au regard des préconisations de la directive européenne 1999/30/CE.

Néanmoins, dans 5 à 10% des cas environ, la précision s'est révélée insuffisante. Ces cas semblent relever de situations de pollution peu fréquentes, vraisemblablement liées à des situations météorologiques particulières, qu'il conviendrait d'étudier ultérieurement.

L'établissement automatique de cartes journalières estimatives des teneurs en NO_2 , par cette méthode, nécessiterait en effet de connaître un peu mieux les différentes situations de pollution rencontrées sur la zone d'étude, en liaison avec les conditions météorologiques.

On pourrait s'appuyer, par exemple, sur l'approche développée et appliquée par Jean-Luc Houdret à de nombreuses régions françaises [4], basée sur la norme internationale ISO 9359 : « Qualité de l'air – Echantillonnage aléatoire stratifié pour l'évaluation de la qualité de l'air ambiant ».

Notre but serait de pouvoir identifier les différentes situations météorologiques pour lesquelles les corrélations s'appliquent ou pas.

La réalisation de cartes journalières est, d'autre part, conditionnée par l'automatisation de l'étape d'interpolation des données.

Il serait nécessaire, dans une étude ultérieure, d'étudier les variations possibles des paramètres de krigeage, en fonction des différentes situations de pollution rencontrées, et d'intégrer des variables auxiliaires afin de mieux rendre compte de certaines pollutions locales.

Cette étude a par ailleurs nécessité un traitement des données de qualité de l'air sous EXCEL (calcul des paramètres et exactitude des régressions). Le classeur EXCEL et les macros développées pour cette étude sont disponibles.

REMERCIEMENTS

Nous remercions sincèrement l'AREMA Lille Métropole pour la fourniture du jeu de données.

BIBLIOGRAPHIE

[1] « Echantillonneurs passifs pour le dioxyde d'azote », ADEME réf. 4414, septembre 2002.

[2] Malherbe L. et Rouil L., « Méthodes de représentation de la qualité de l'air – Guide d'utilisation des méthodes de la géostatistique linéaire », étude LCSQA INERIS DRC-03-45597-LR1-Lma-LCSQA-n°140_v1, 2003.

[3] Giovanni C. et Malherbe L., « Evaluation des incertitudes associées aux méthodes géostatistiques », étude LCSQA INERIS DRC-03-45599-2IEN-MECO-Gca-Lma-LCSQA-n°141_v1, 2003.

[4] Houdret J.L., « Influence de paramètres météorologiques sur la stratégie de mesure à l'aide de moyens mobiles », étude LCSQA EMD n°6, 2003.

[5] Perdrix E. et al., « Représentativité spatiale des stations urbaines », étude LCSQA EMD n°7, 2001.

[6] Sagnier I., « Etude de la répartition de la pollution urbaine par le dioxyde d'azote sur l'agglomération lilloise par la méthode des échantillonneurs passifs », rapport d'étude AREMA Lille Métropole, juin 1999, 49 p.

[7] Sagnier I., Dusart A., Deguine J., « Etude de la répartition de la pollution urbaine sur l'arrondissement de Lille par la méthode des échantillonneurs passifs. Eté 2003 – Hiver 2004 (rapport intermédiaire campagne été) », rapport d'étude AREMA Lille Métropole, août 2003, 92 p.

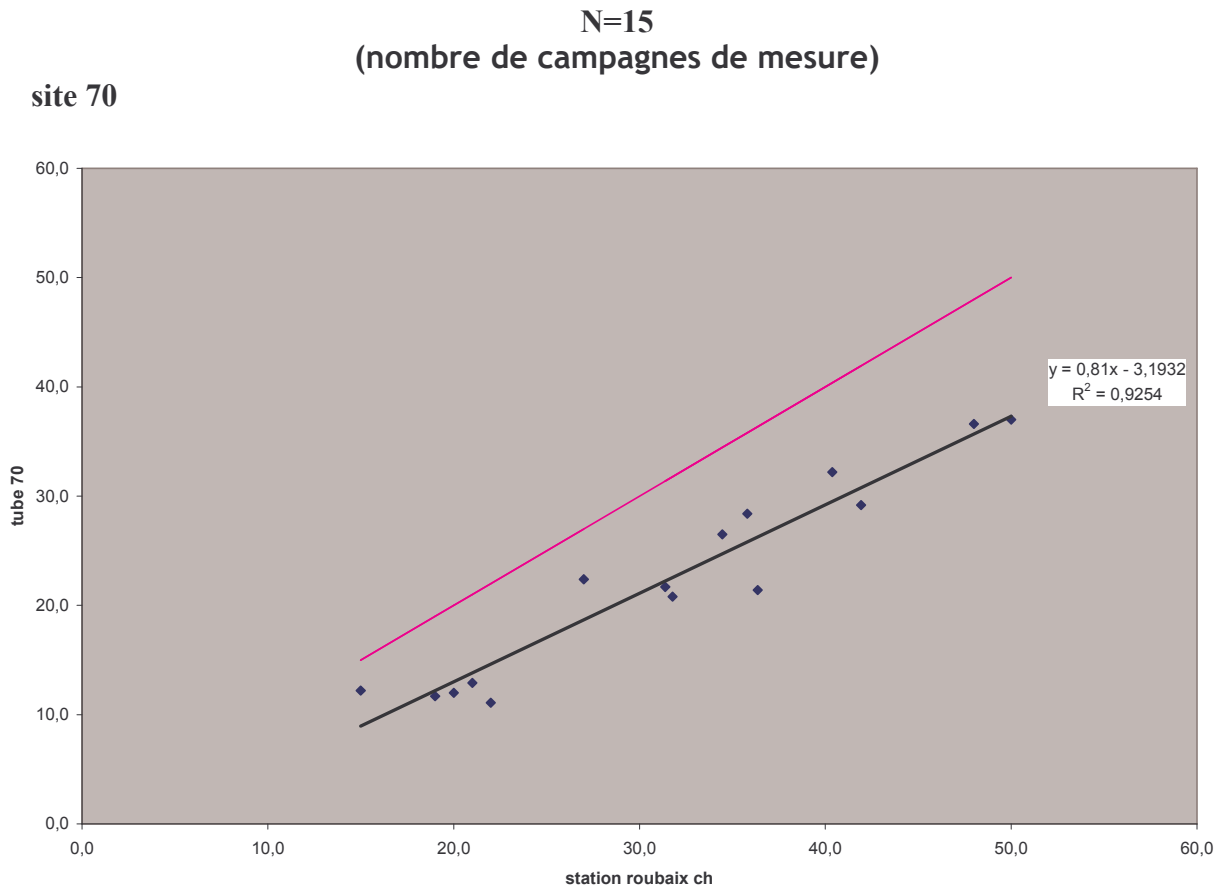
[8] Dusart A., Deguine J. et co-auteurs « Etude de la répartition de la pollution urbaine sur l'arrondissement de Lille par la méthode des échantillonneurs passifs. Eté 2003 – Hiver 2004 (rapport final campagnes été 2003/hiver 2004) », rapport d'étude AREMA Lille Métropole, 2004, 119 p.

ANNEXE 1

Corrélations entre des sites d'échantillonnage passif du NO₂ et la station urbaine "Roubaix château"

Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

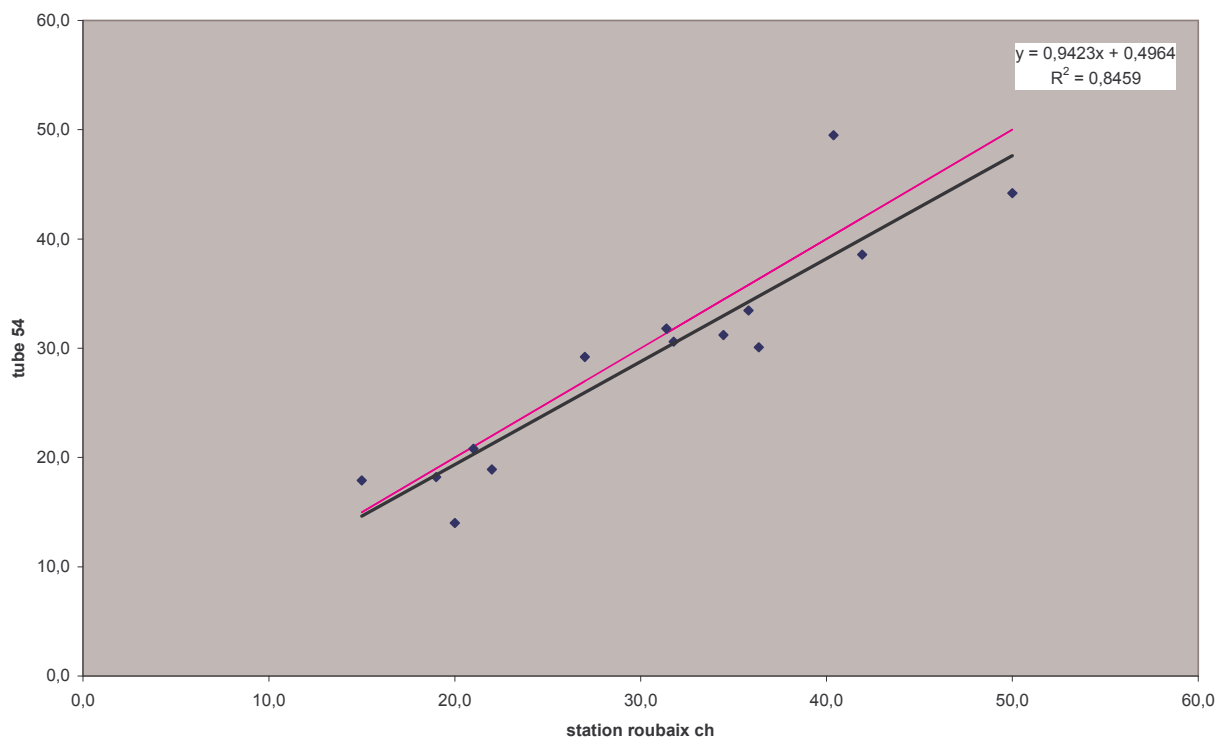
Remarque: Pour toutes les figures de l'annexe1, la droite en rose est la 1^{ère} bissectrice, celle en noir correspond à la régression linéaire trouvée.



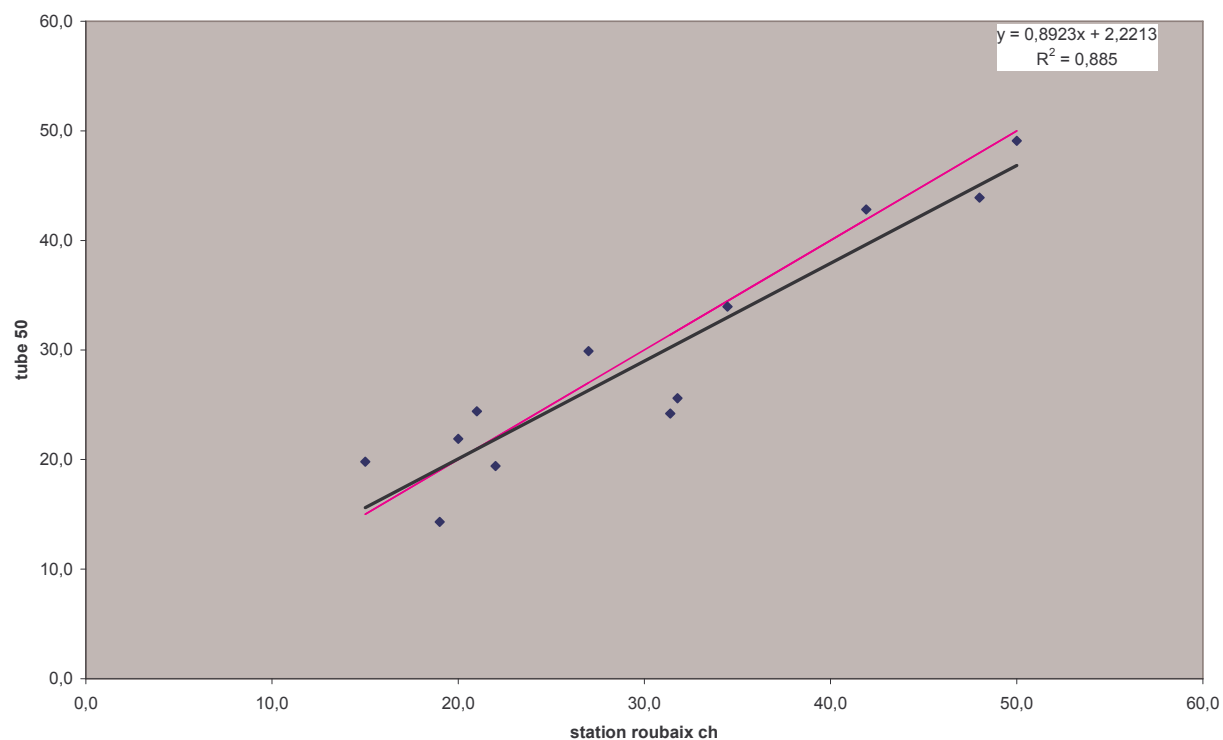
Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

N=14
(nombre de campagnes de mesure)

site 54

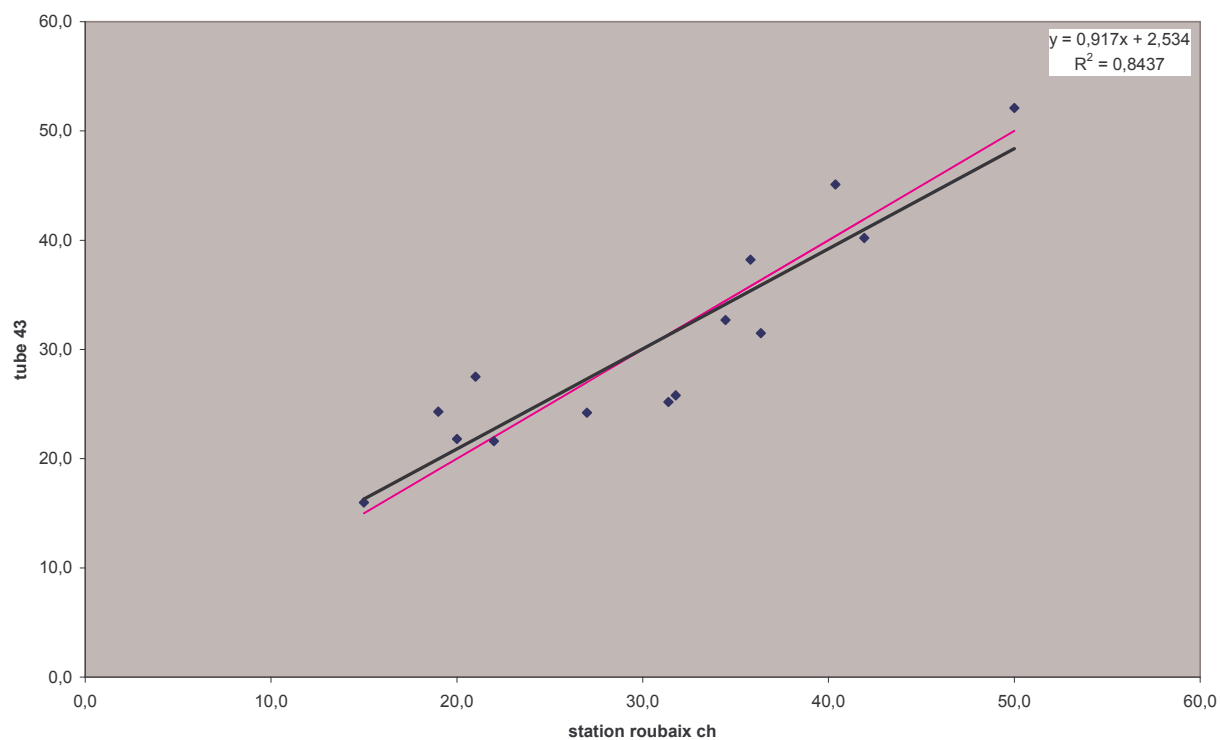


site 50

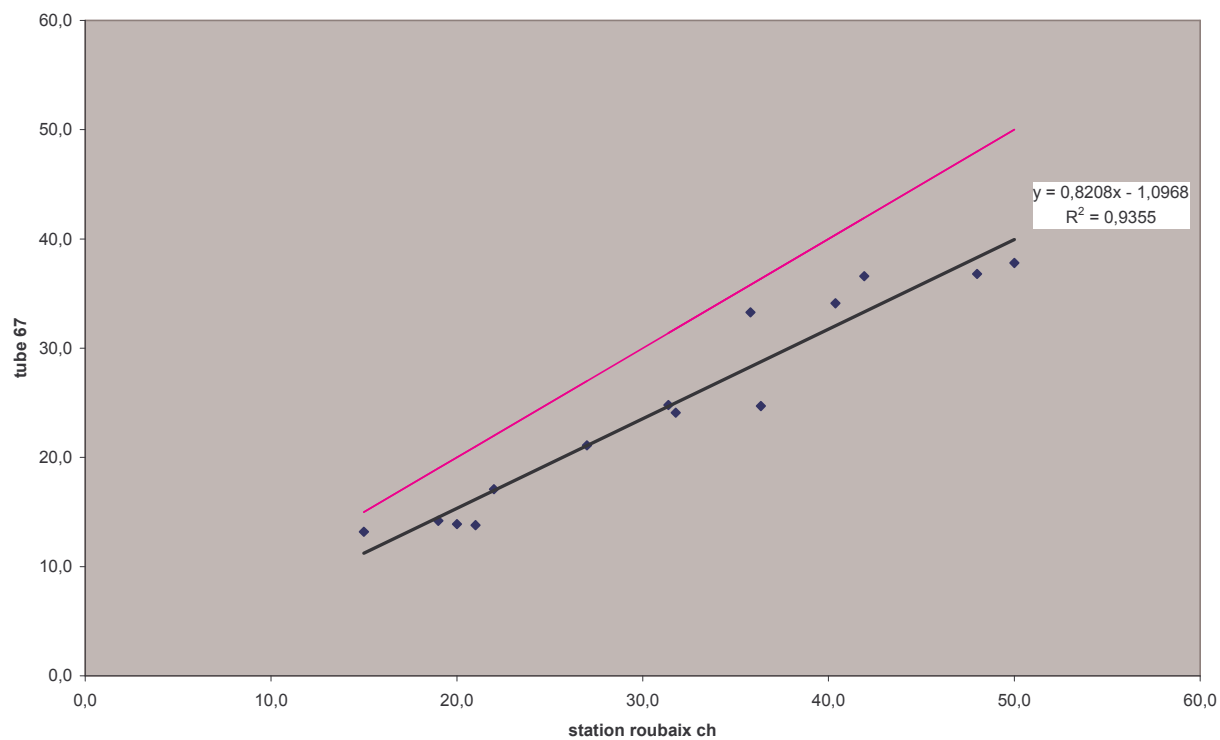


Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

site 43

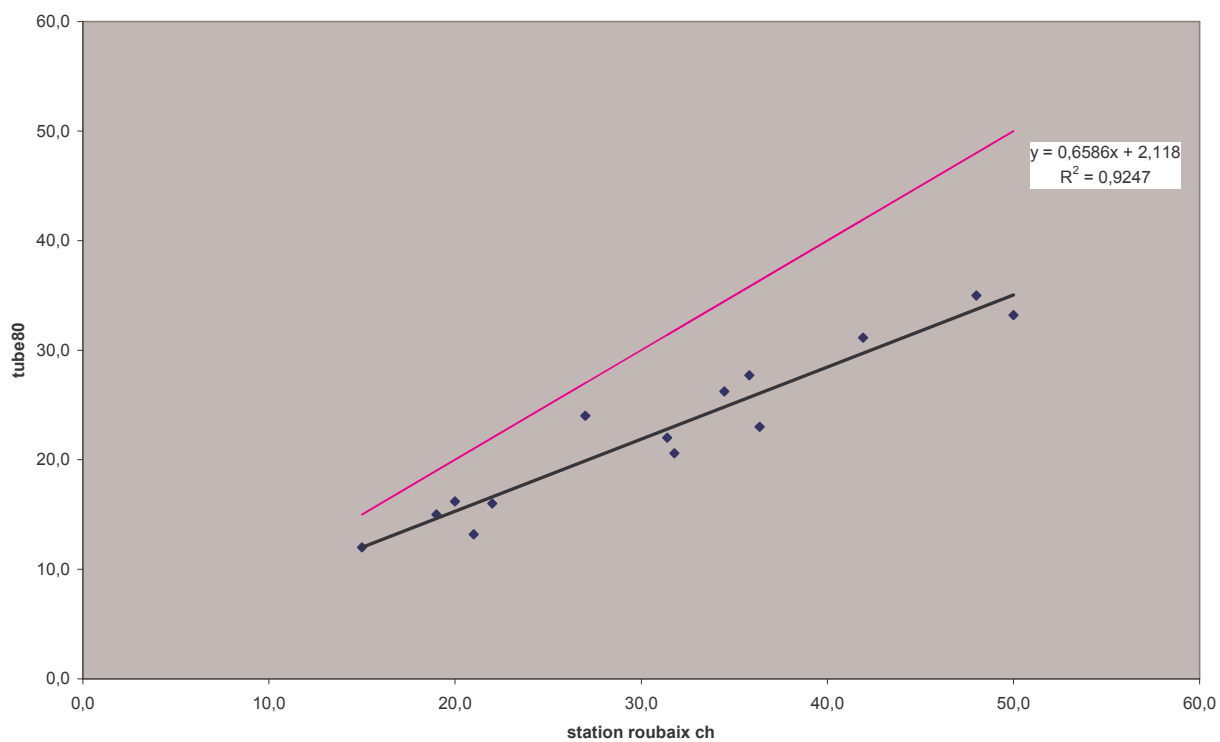


site 67

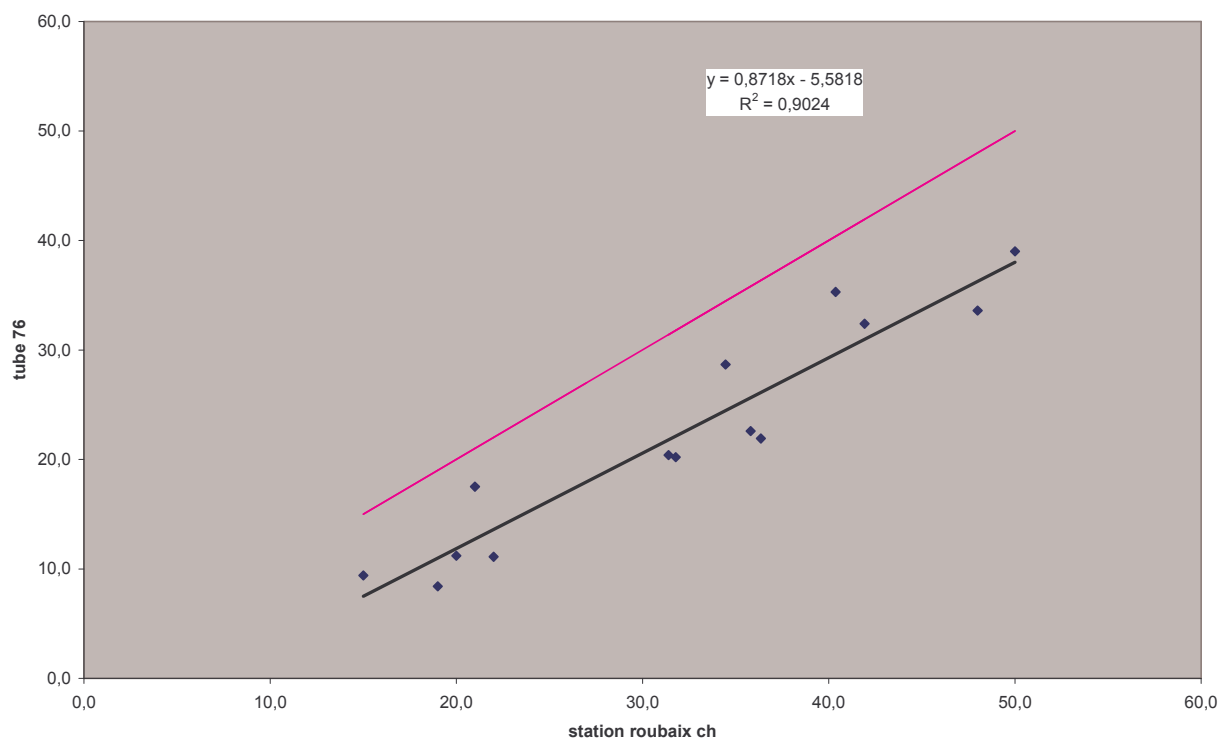


Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

Site 80



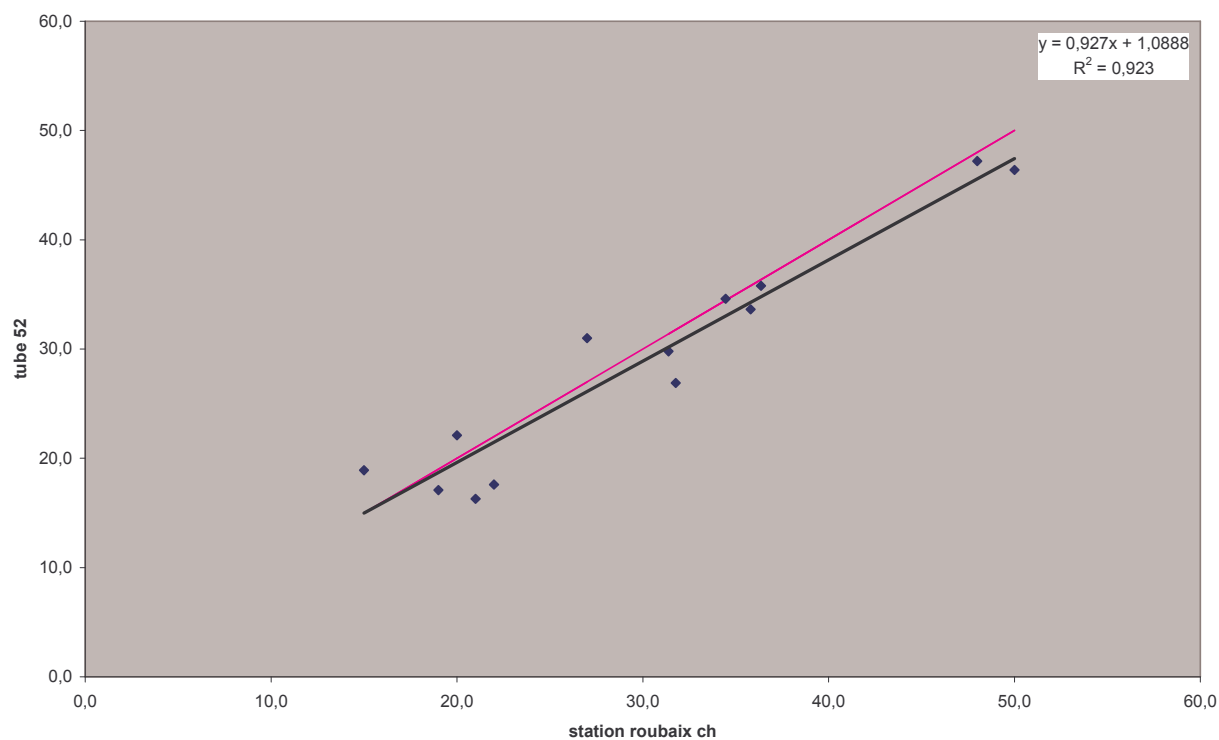
Site 76



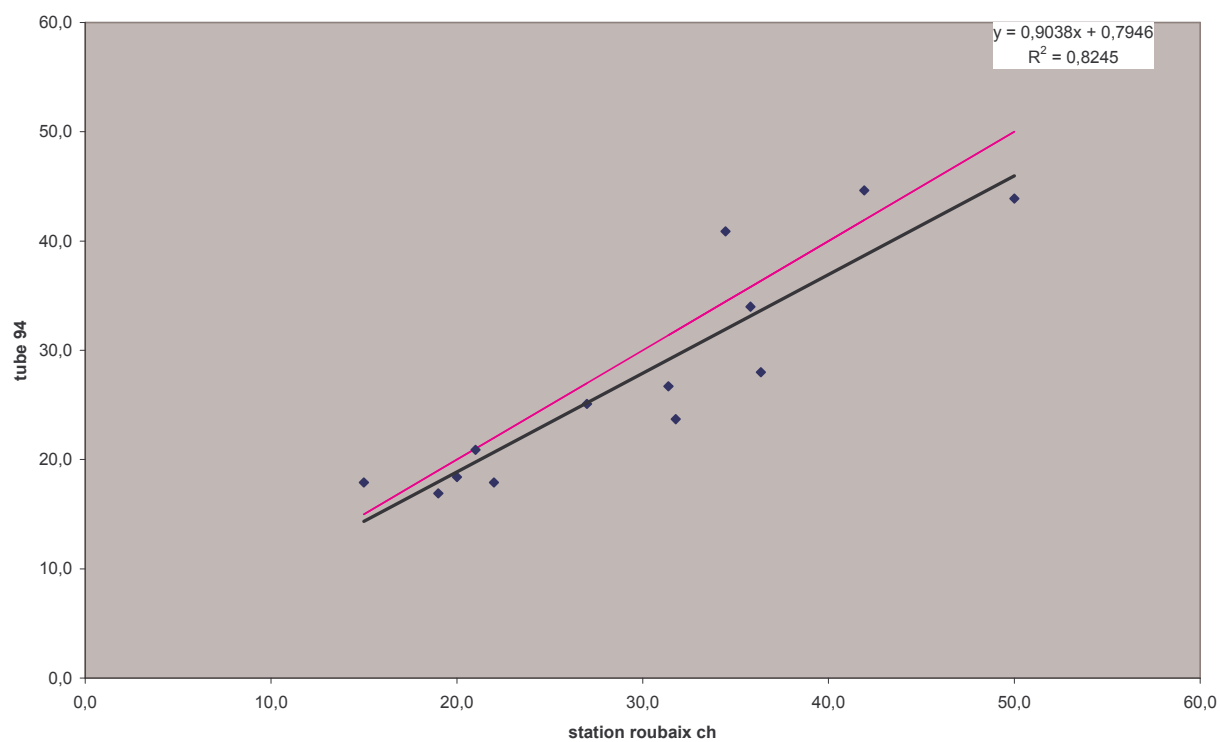
Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

N=13
(nombre de campagnes de mesure)

site 52

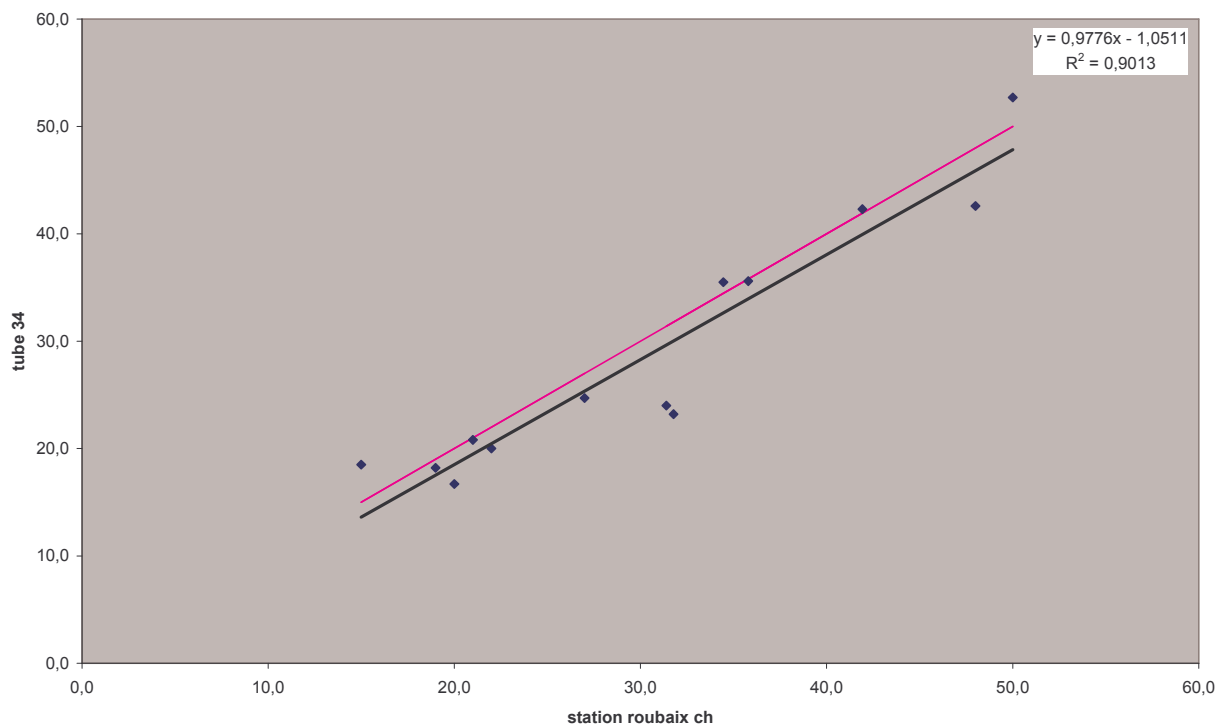


site 94

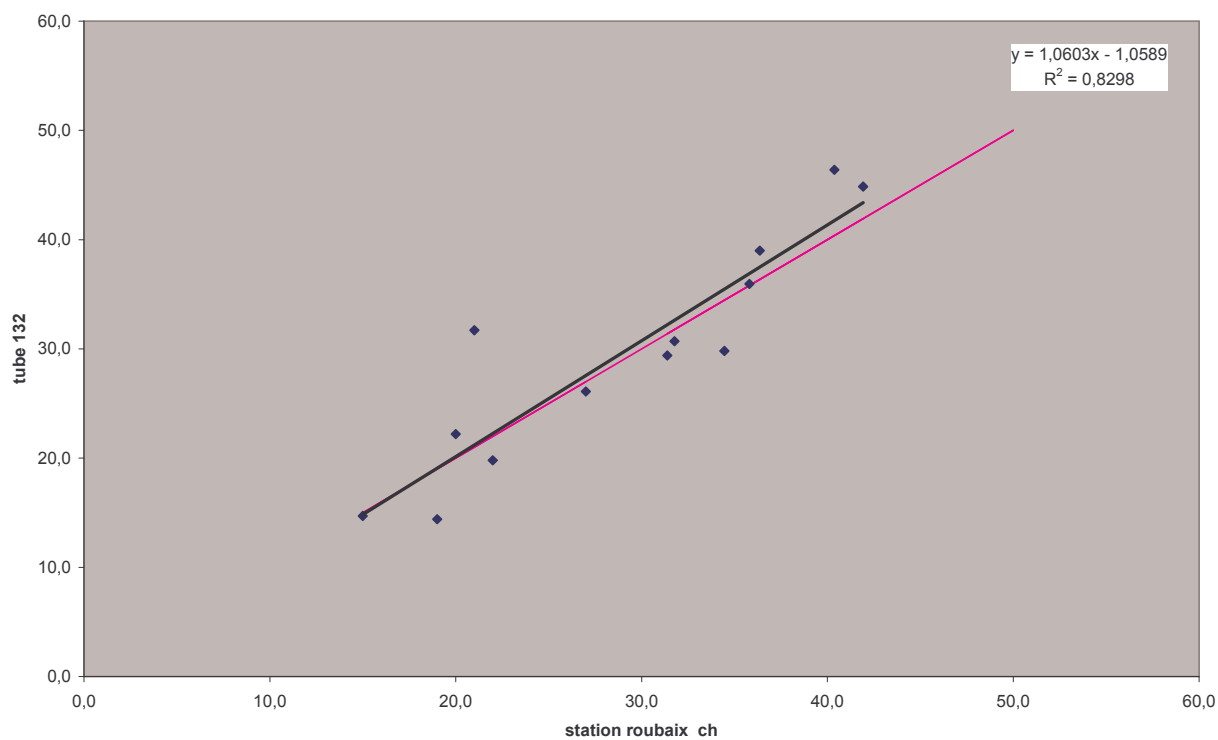


Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

site 34

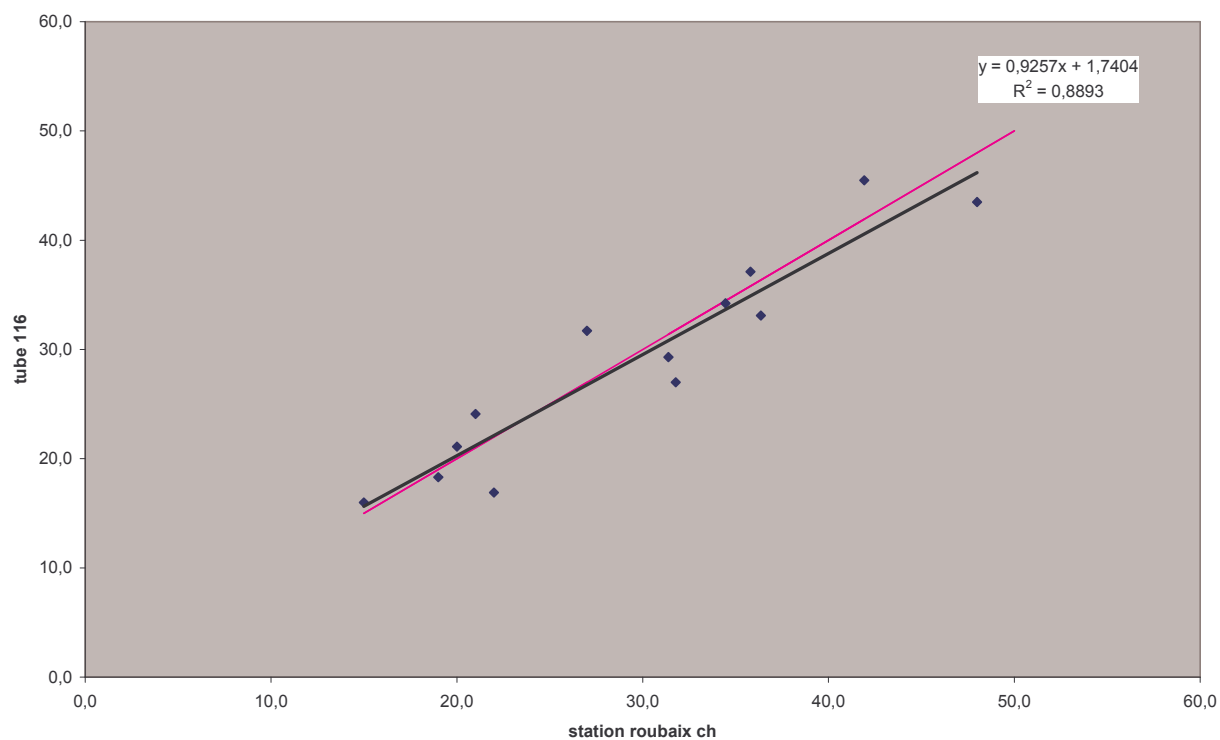


site 132

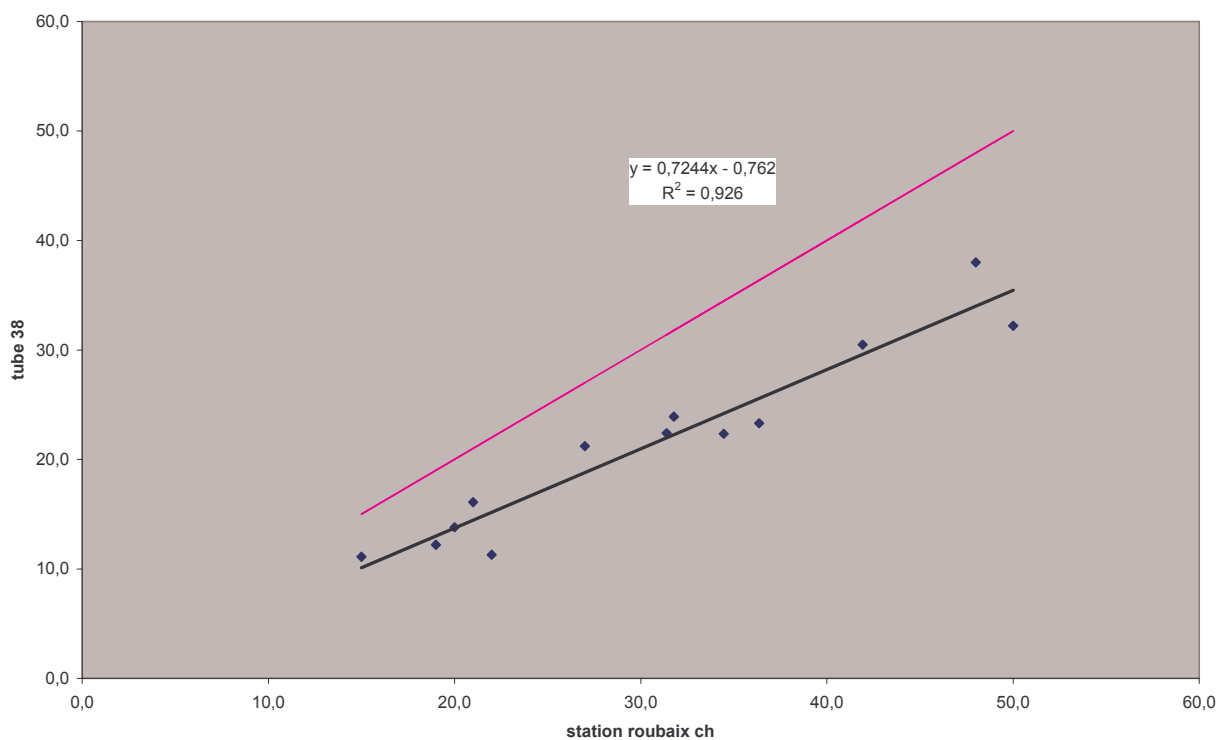


Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

Site 116

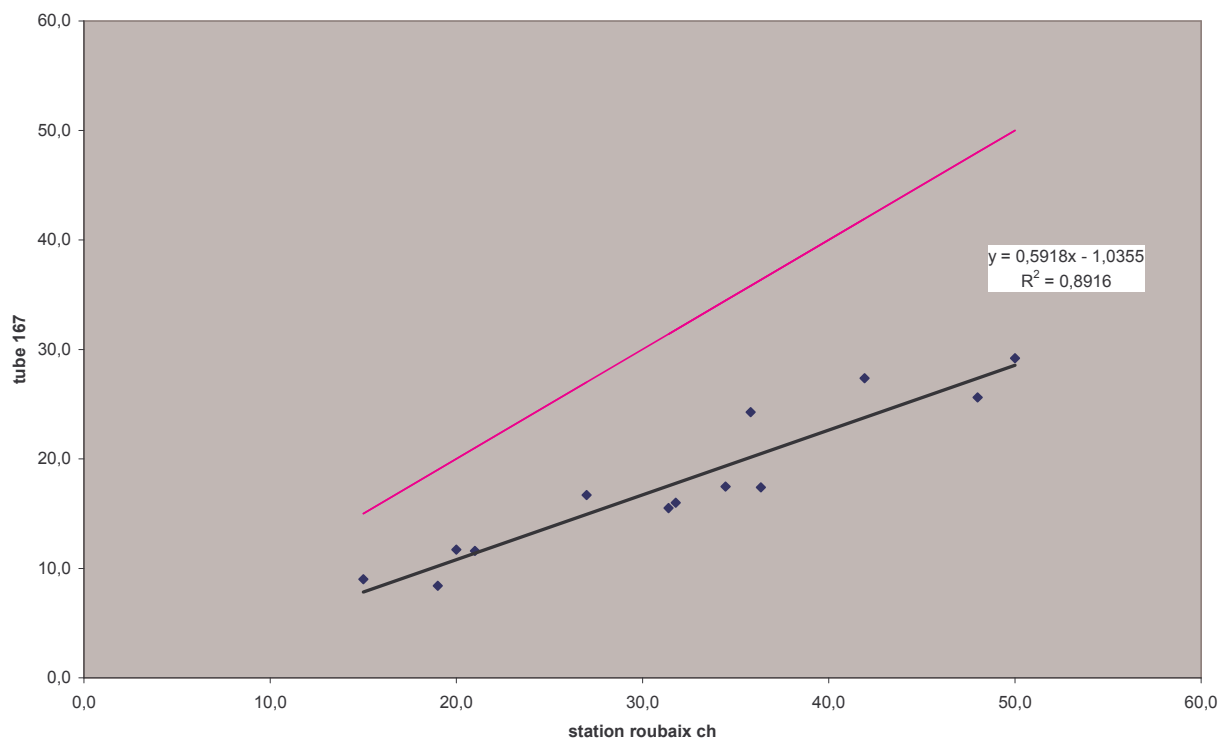


Site 38

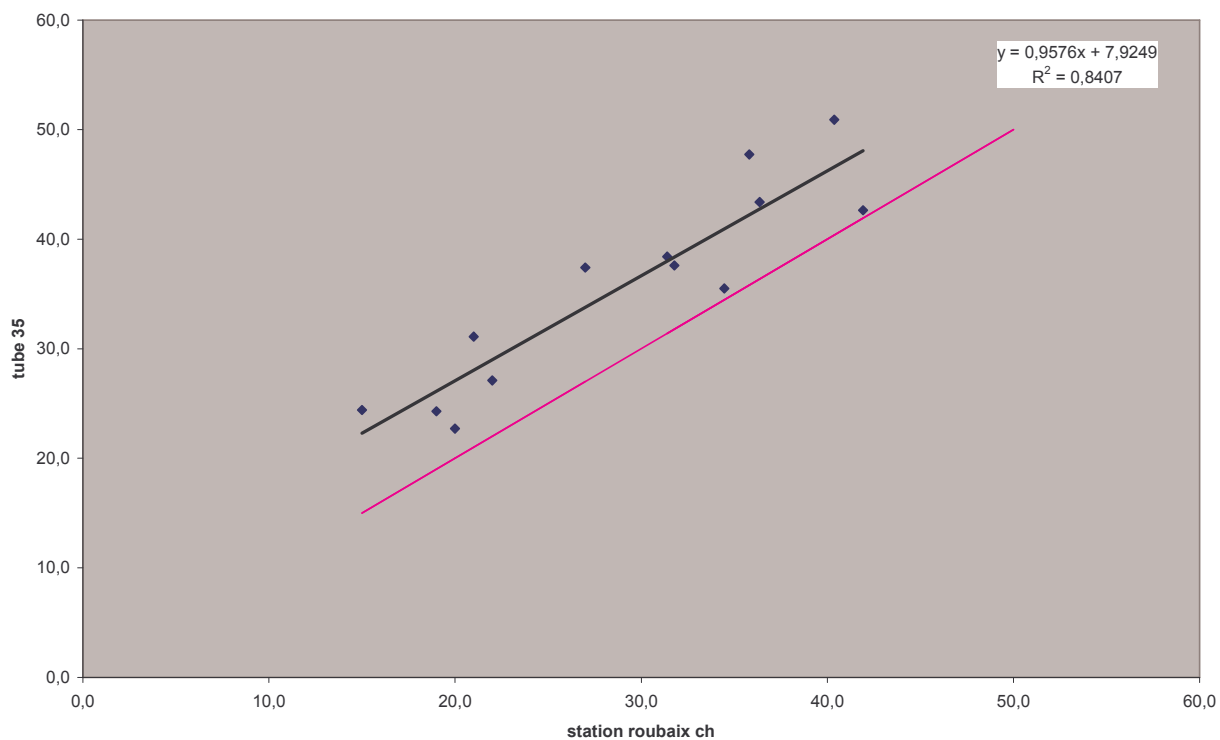


Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

Site 167



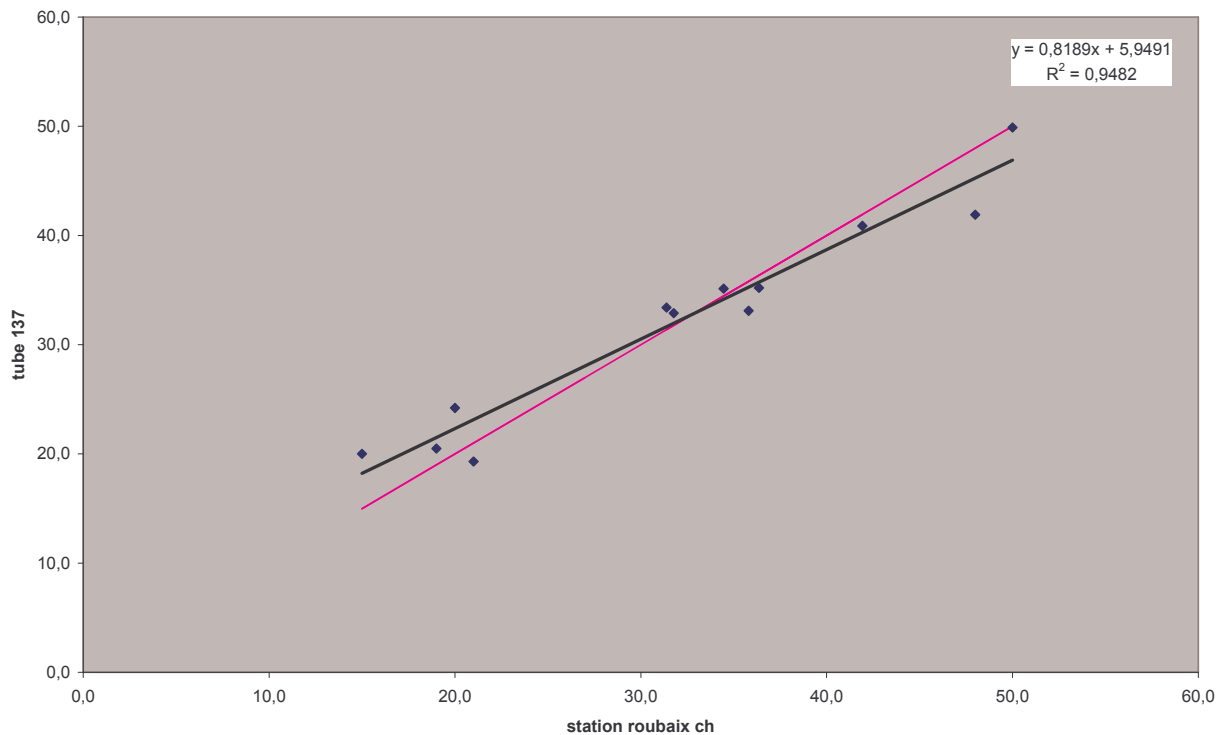
Site 35



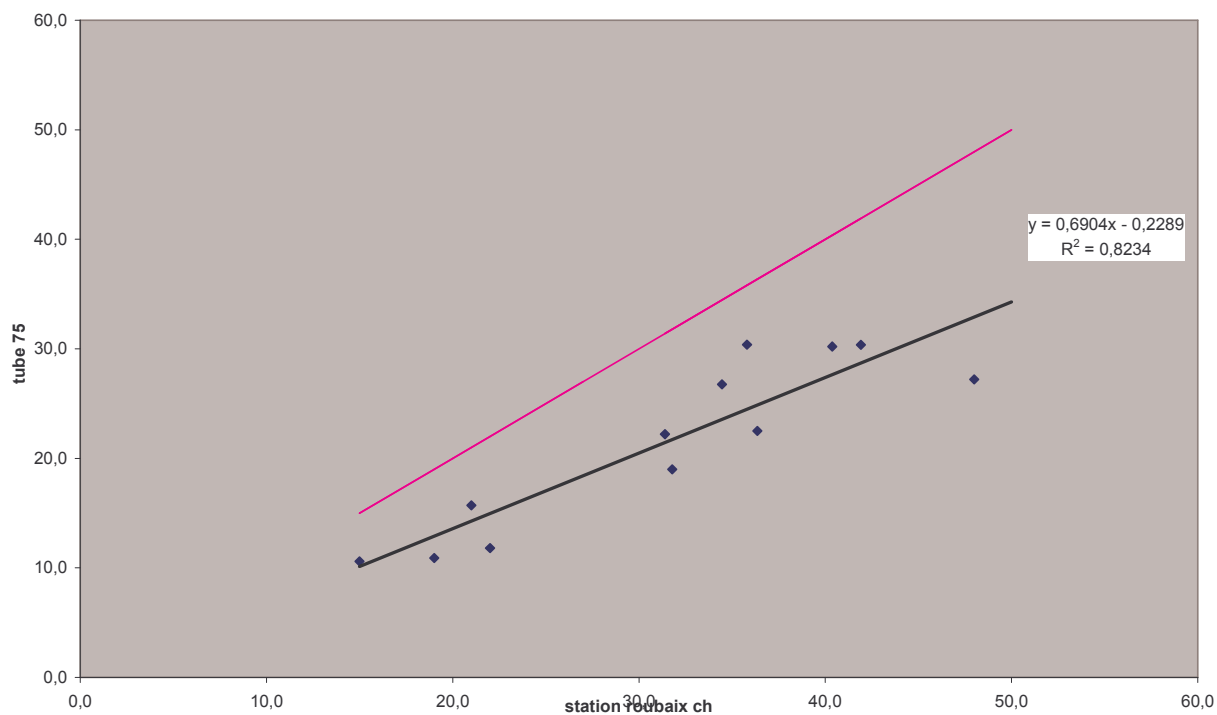
Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

N=12
(nombre de campagnes de mesure)

site 137

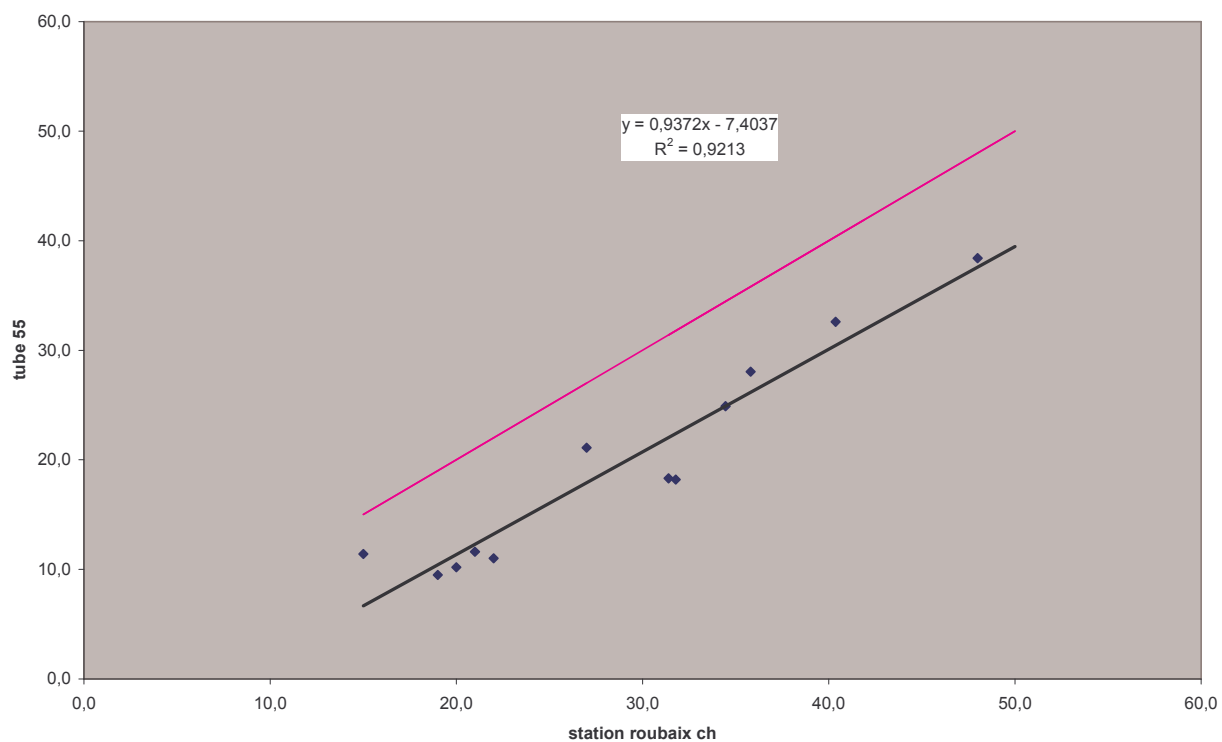


site 75

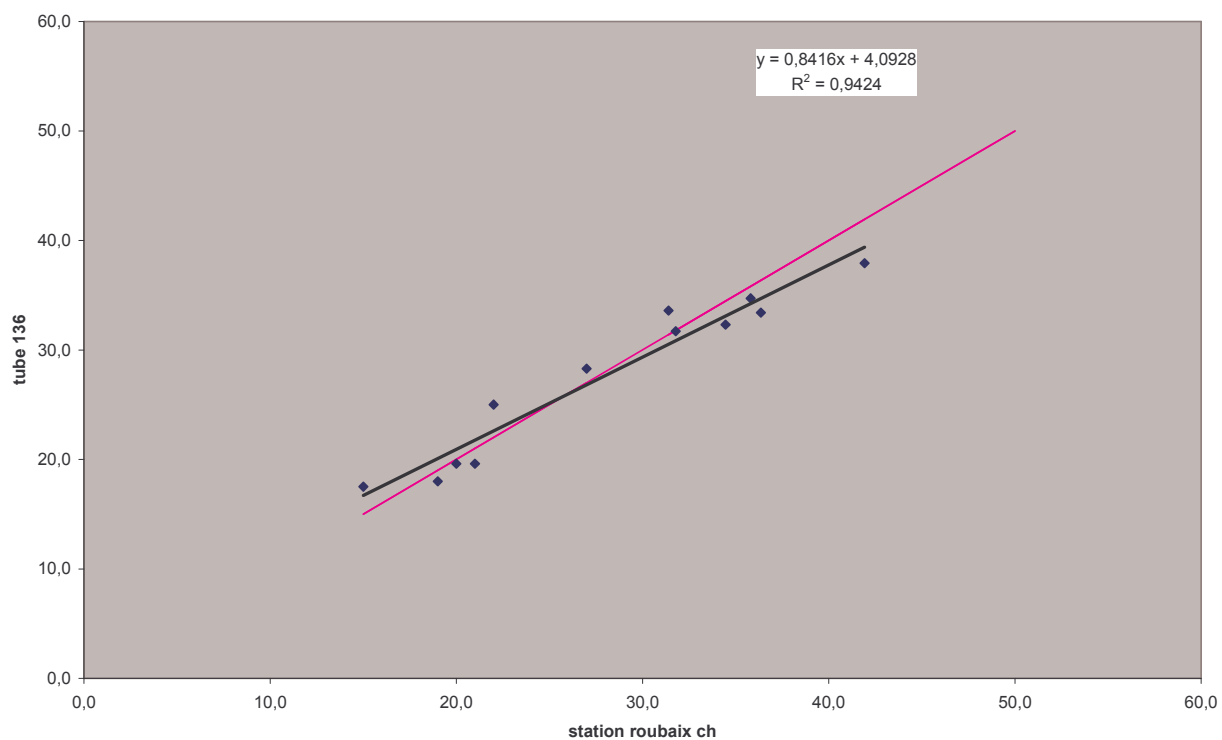


Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

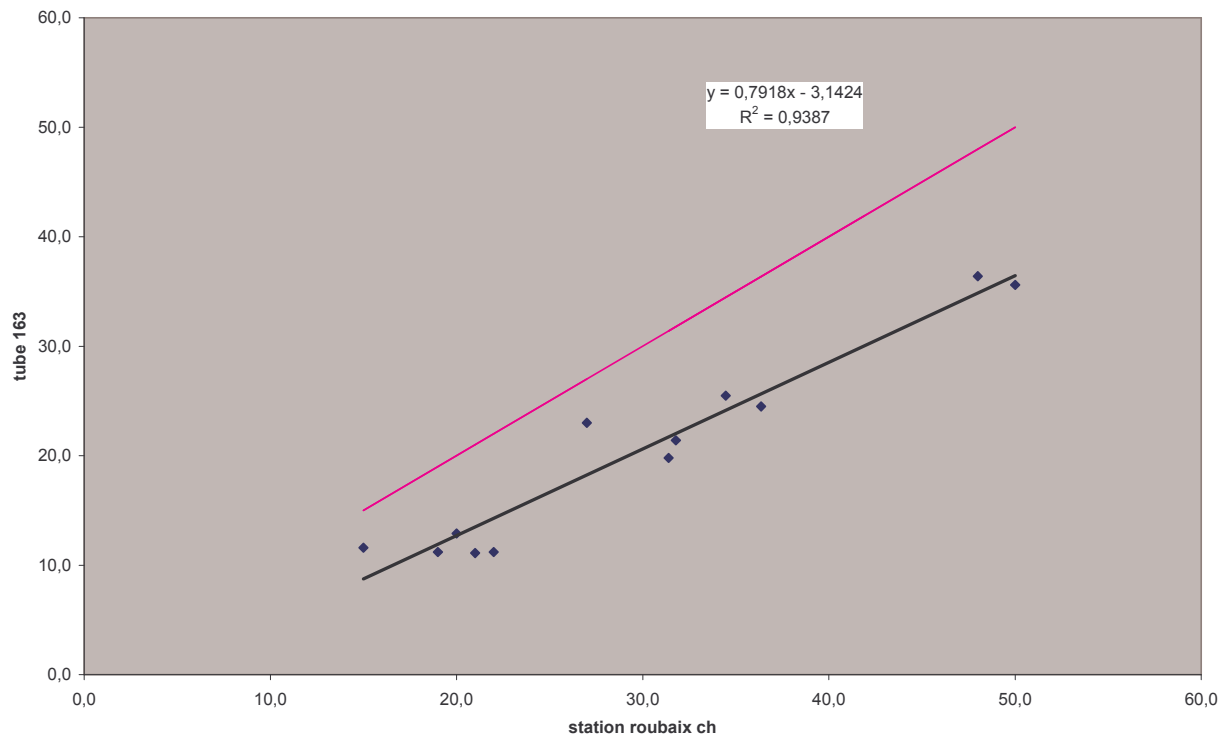
site 55



site 136



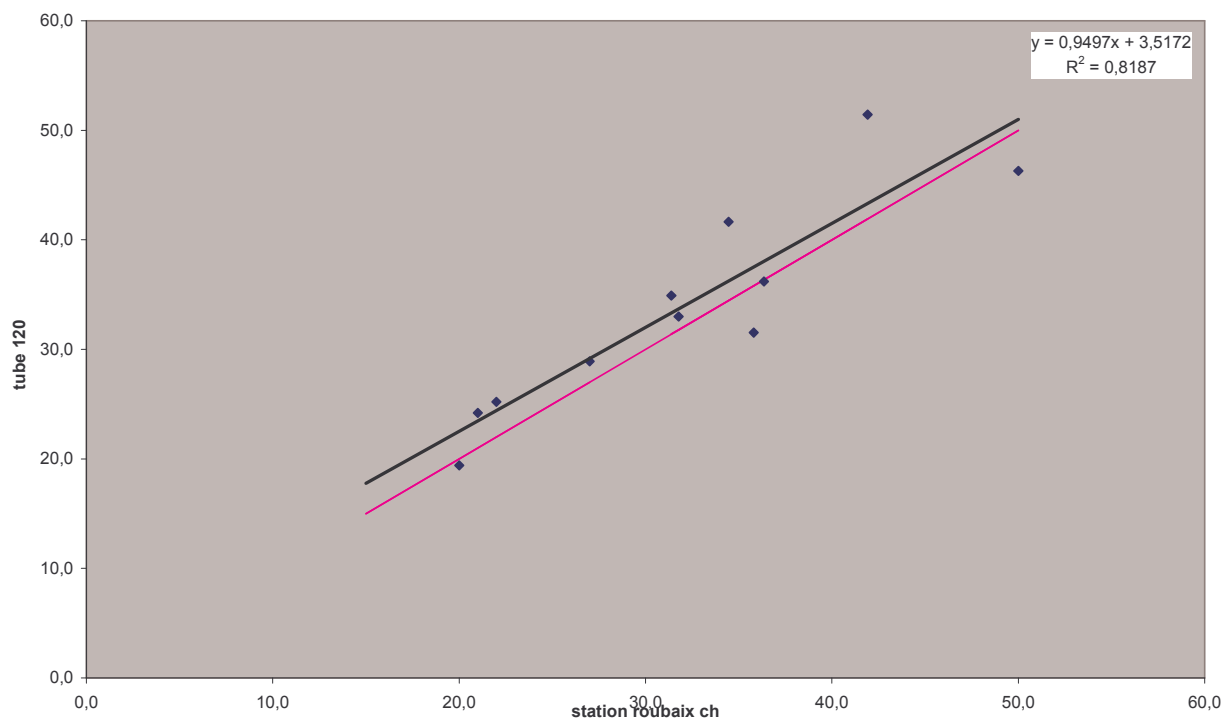
Site 163



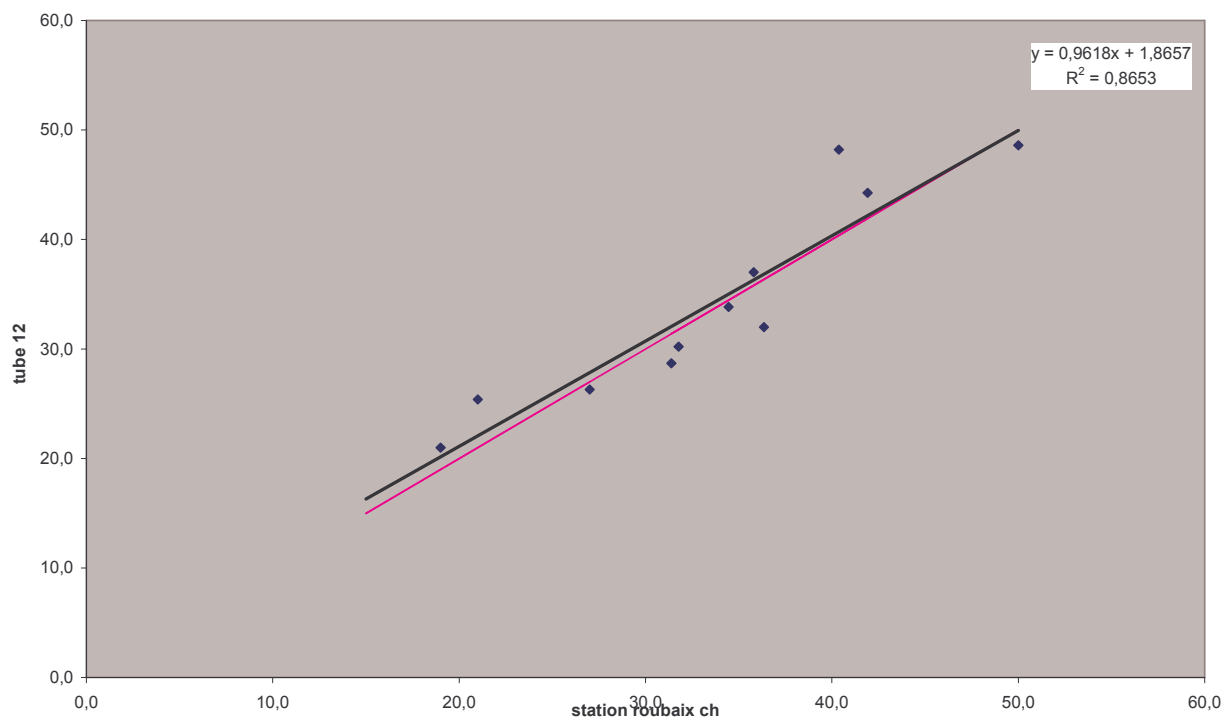
Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

N=11
(nombre de campagnes de mesure)

site 120

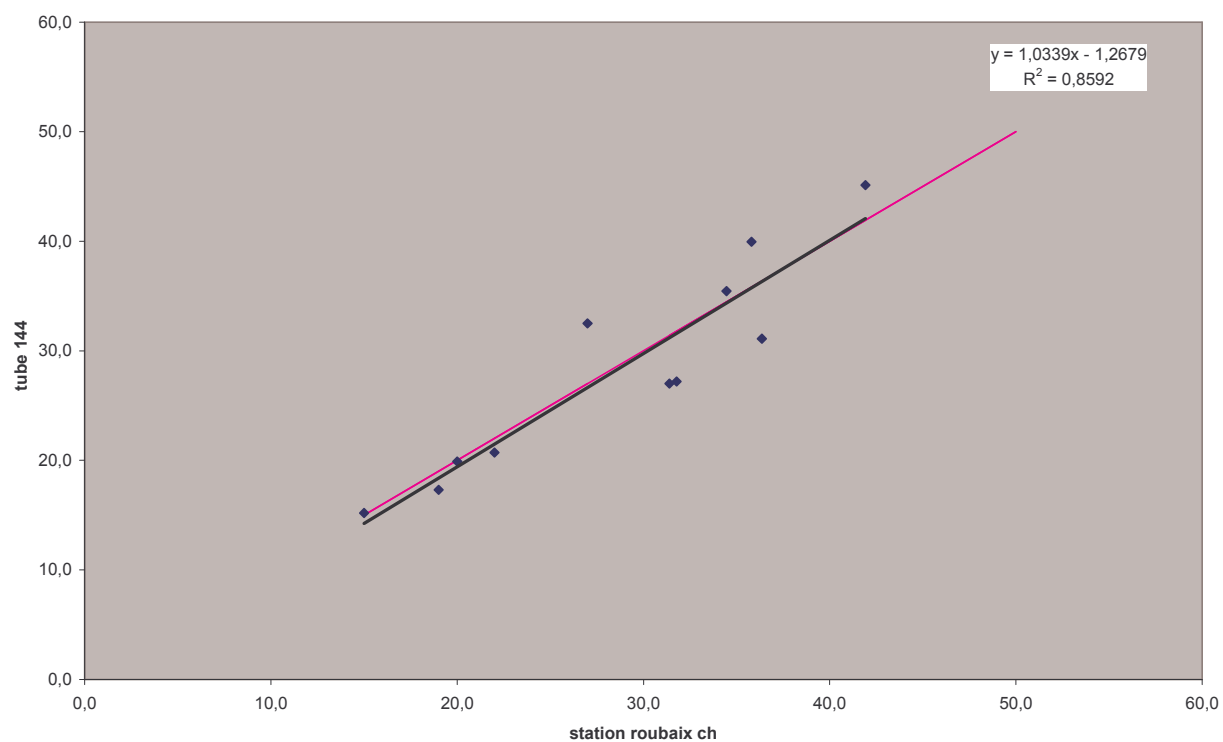


site 12

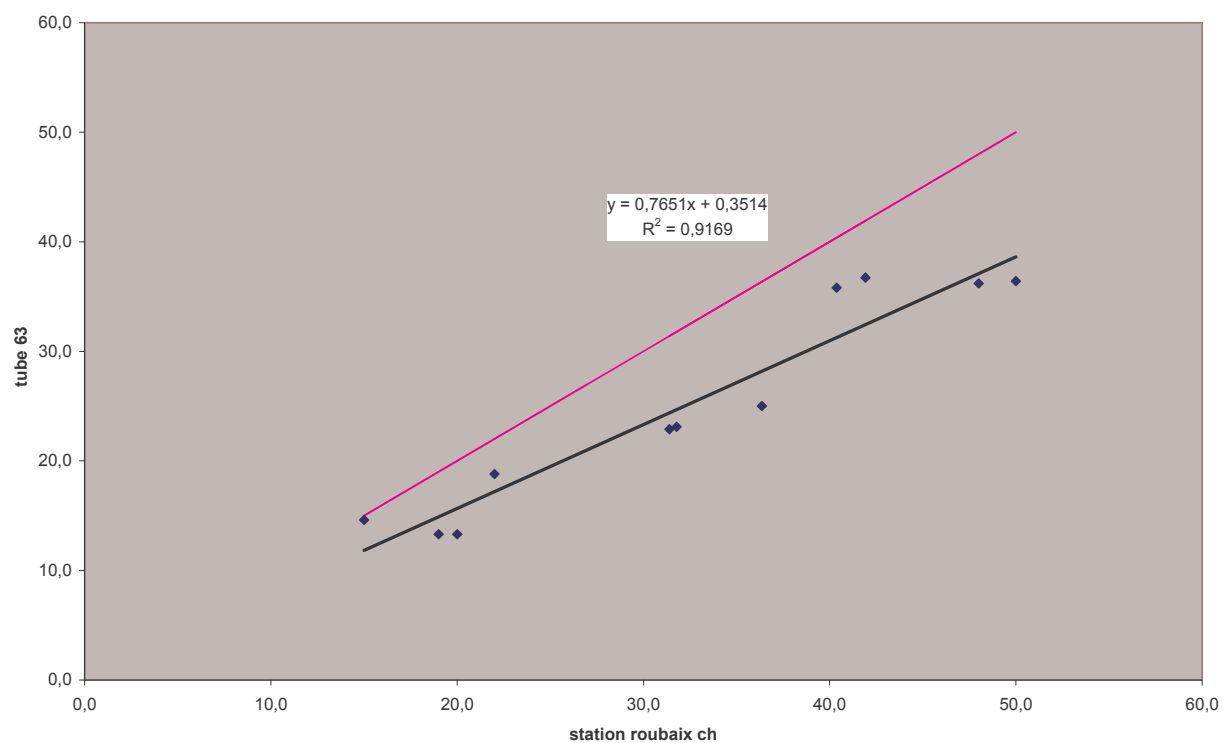


Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

site 144

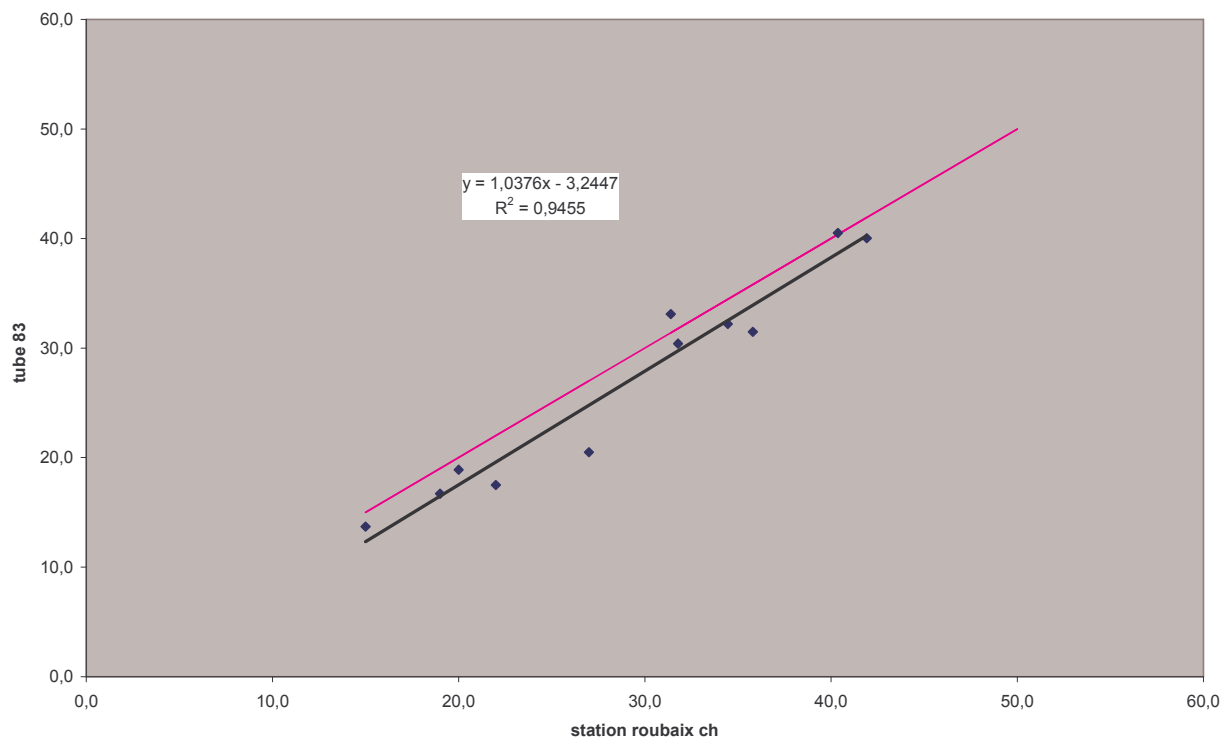


site 63

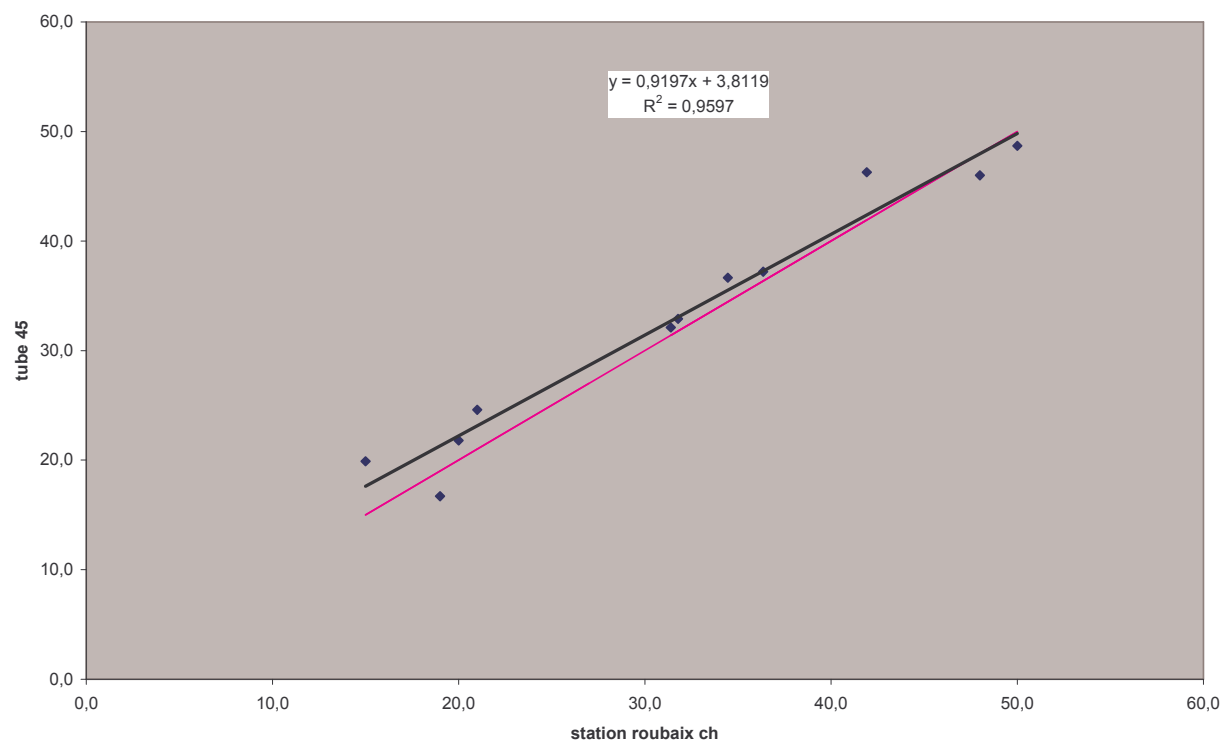


Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

Site 83

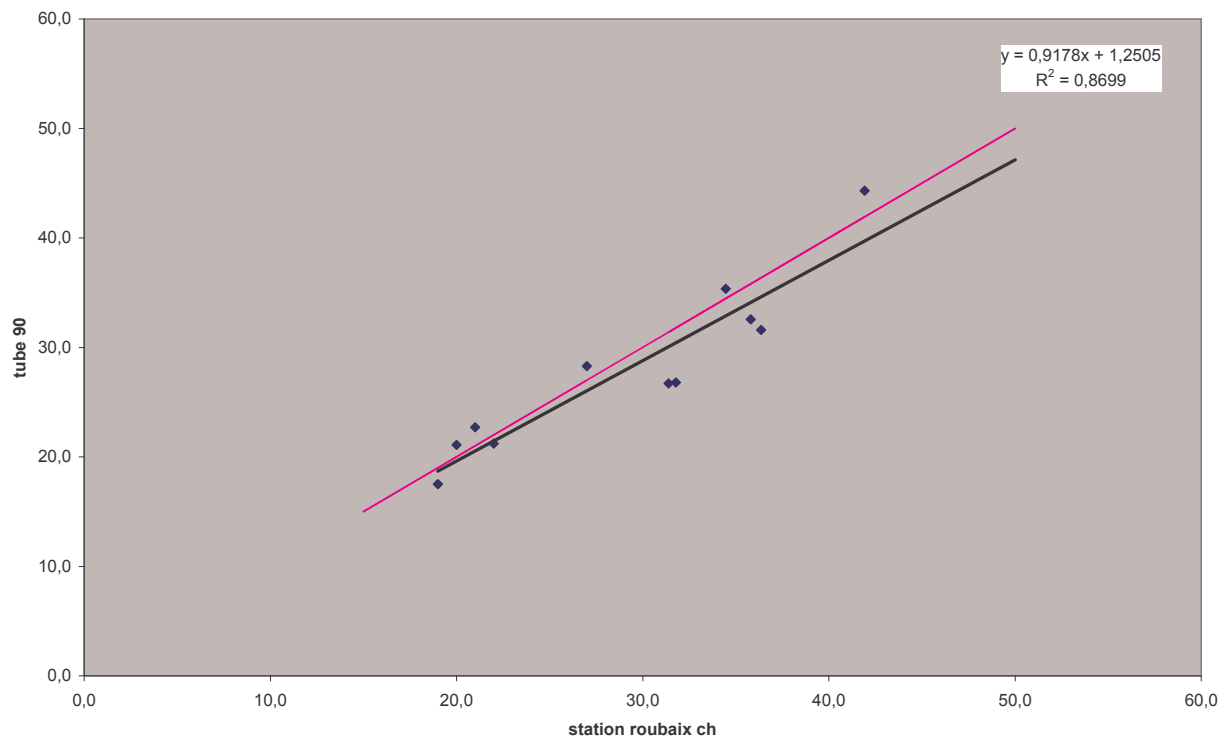


Site 45



Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

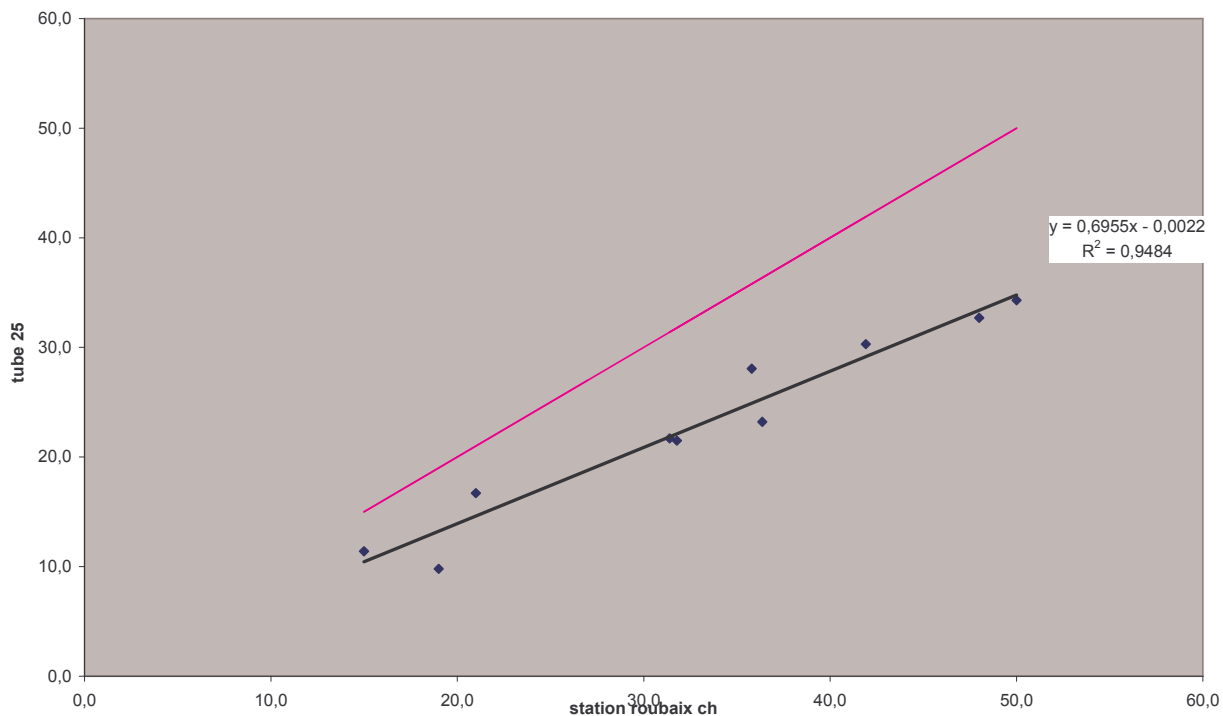
Site 90



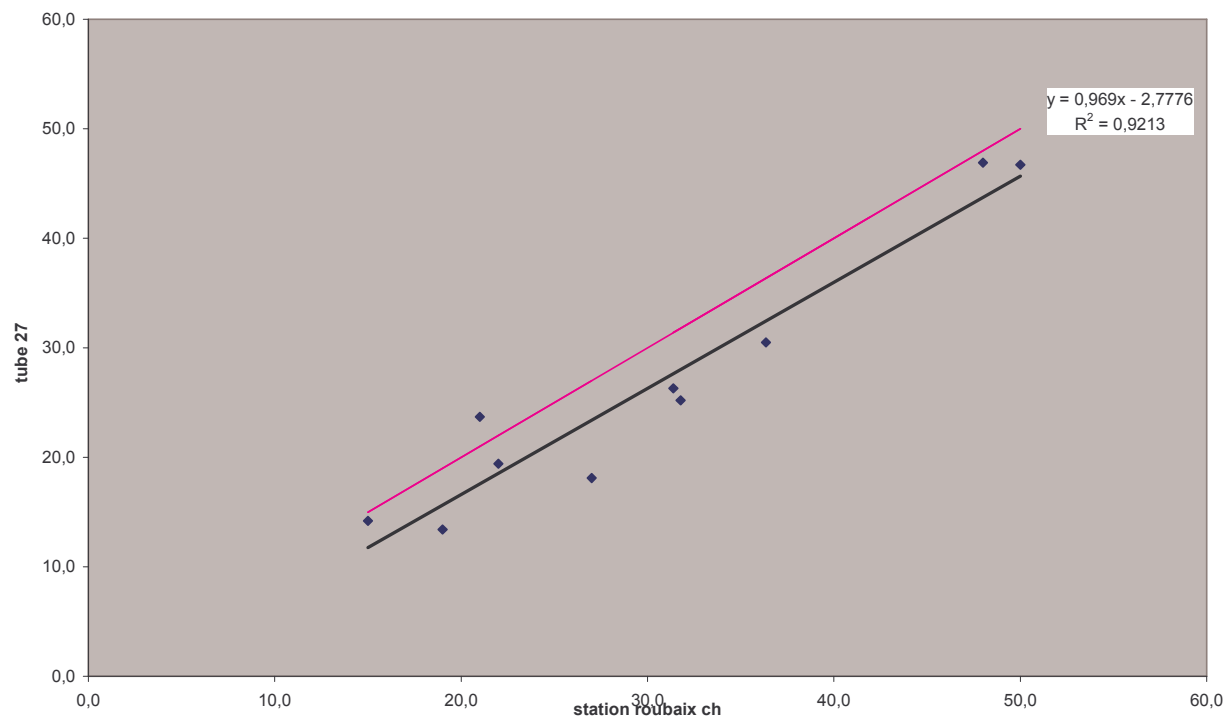
Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

N=10
(nombre de campagnes de mesure)

site 25

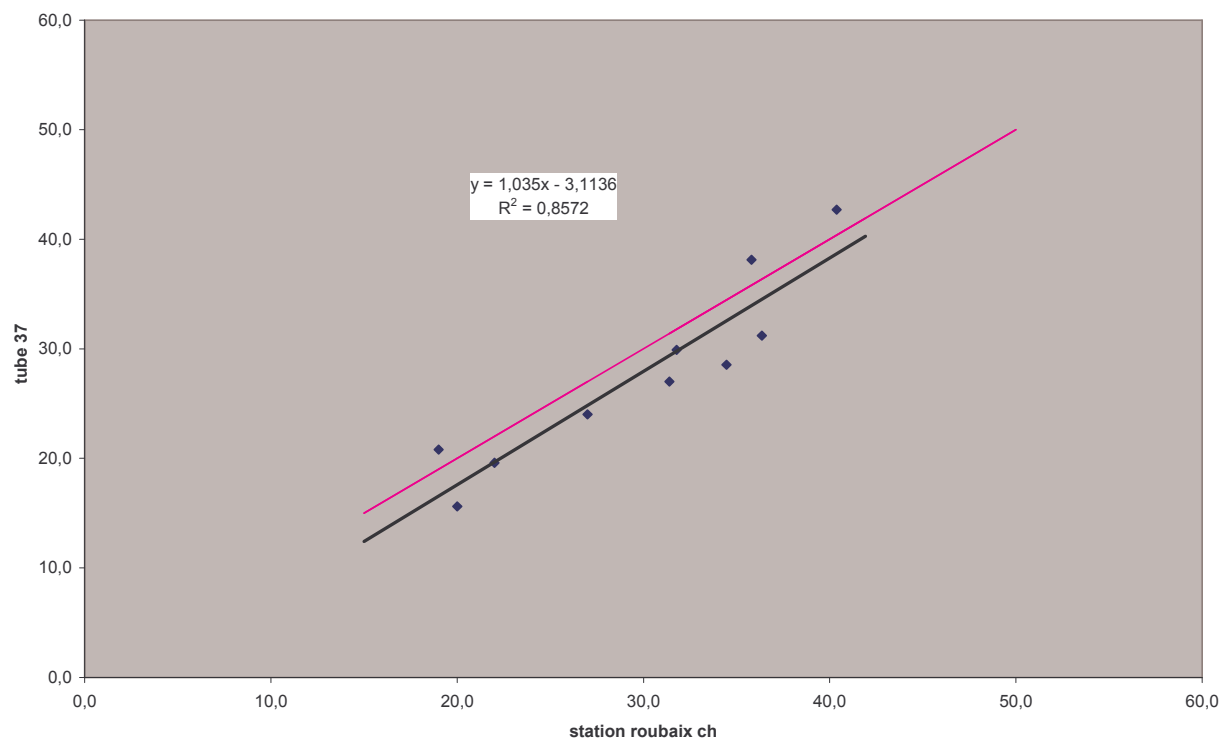


site 27



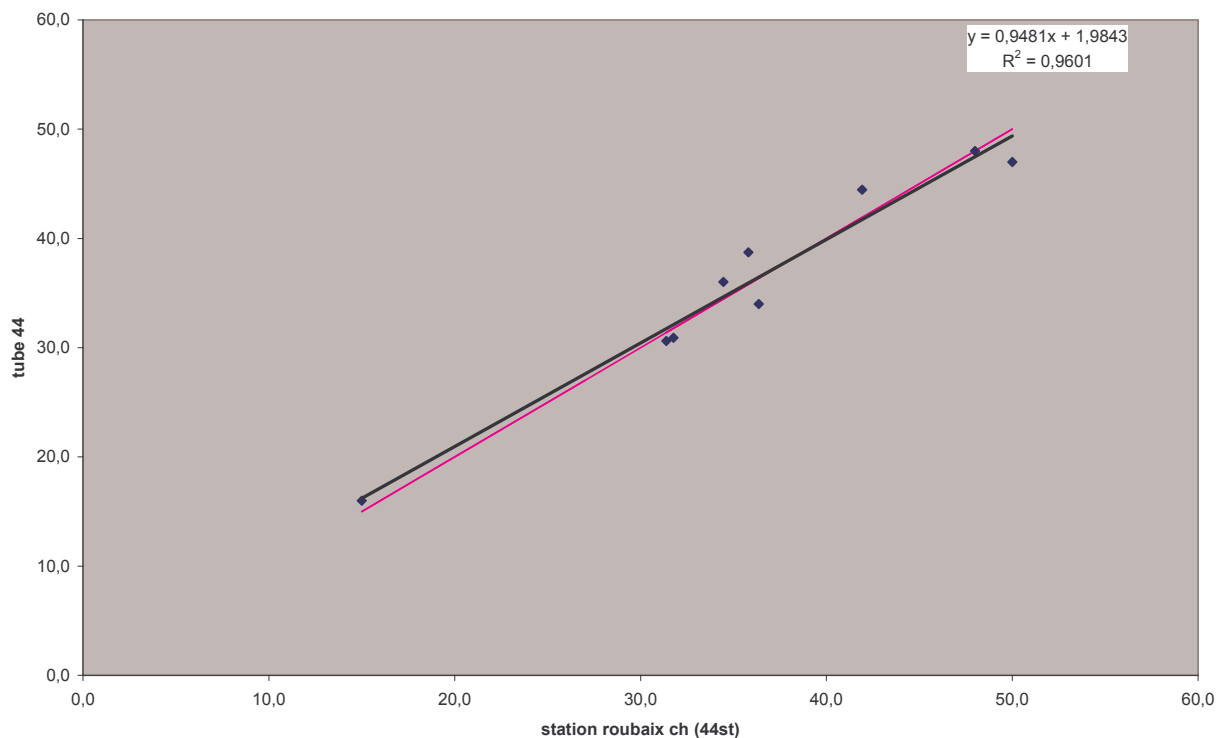
Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

site 37

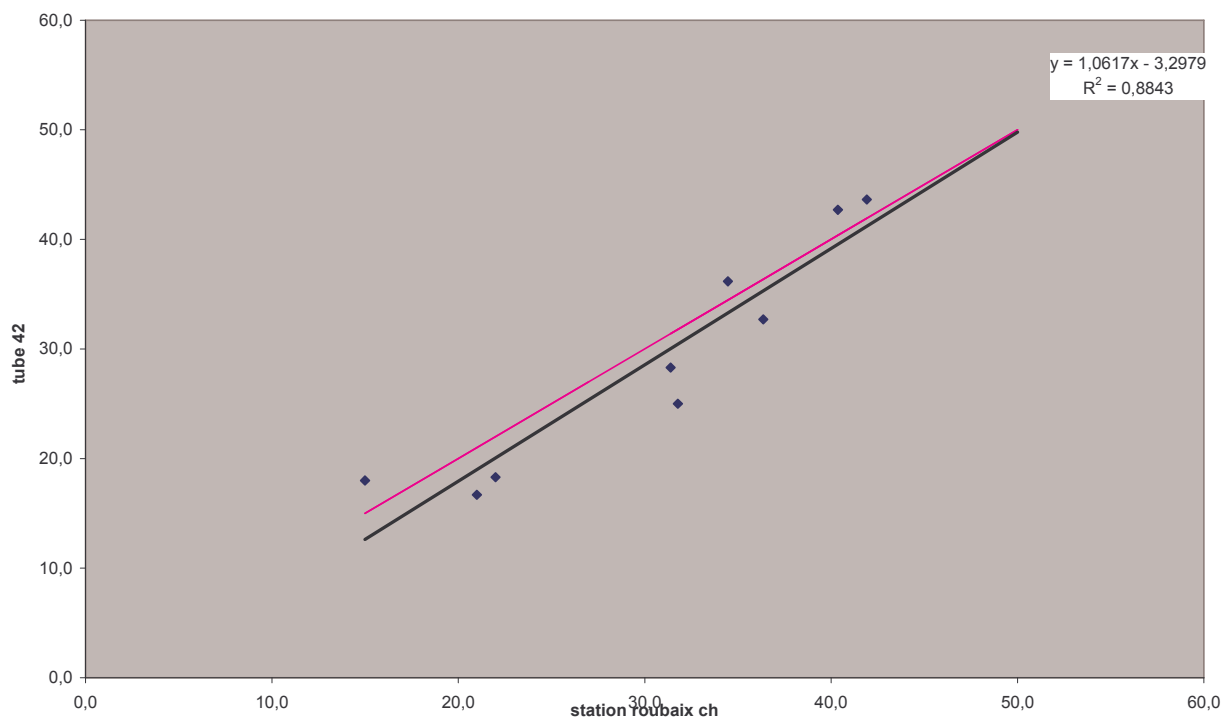


Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

N=9
(nombre de campagnes de mesure)
site 44 (site de la station)

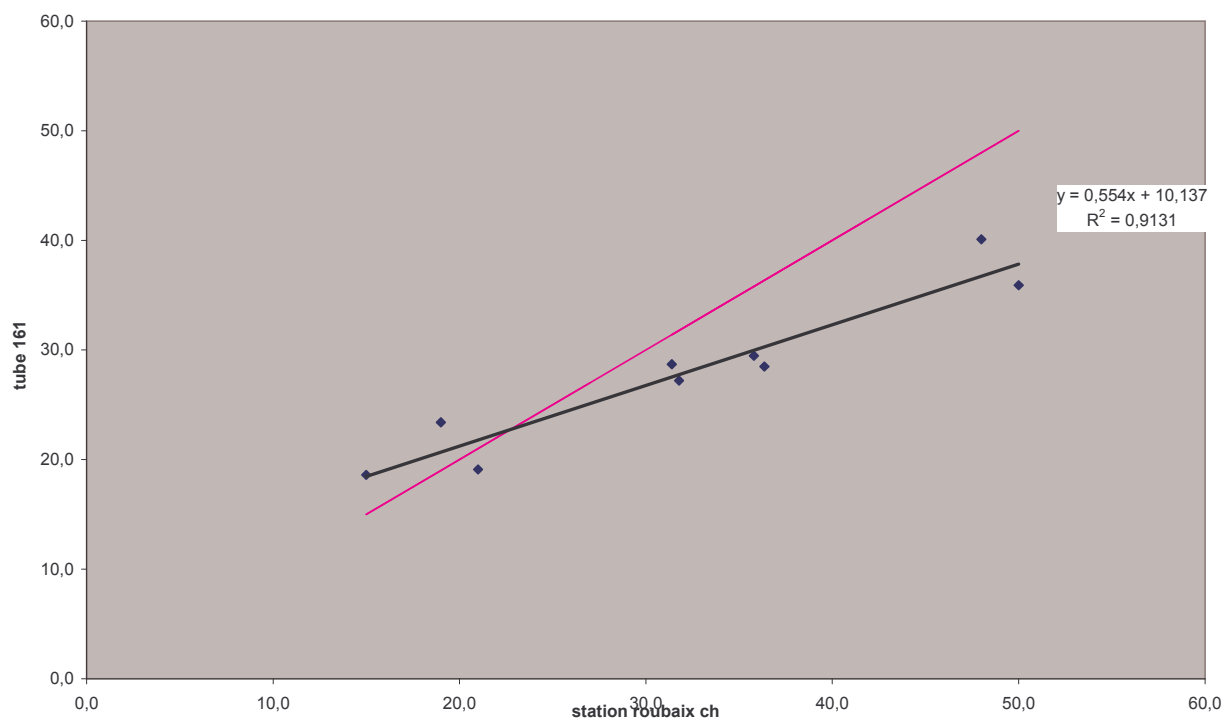


site 42

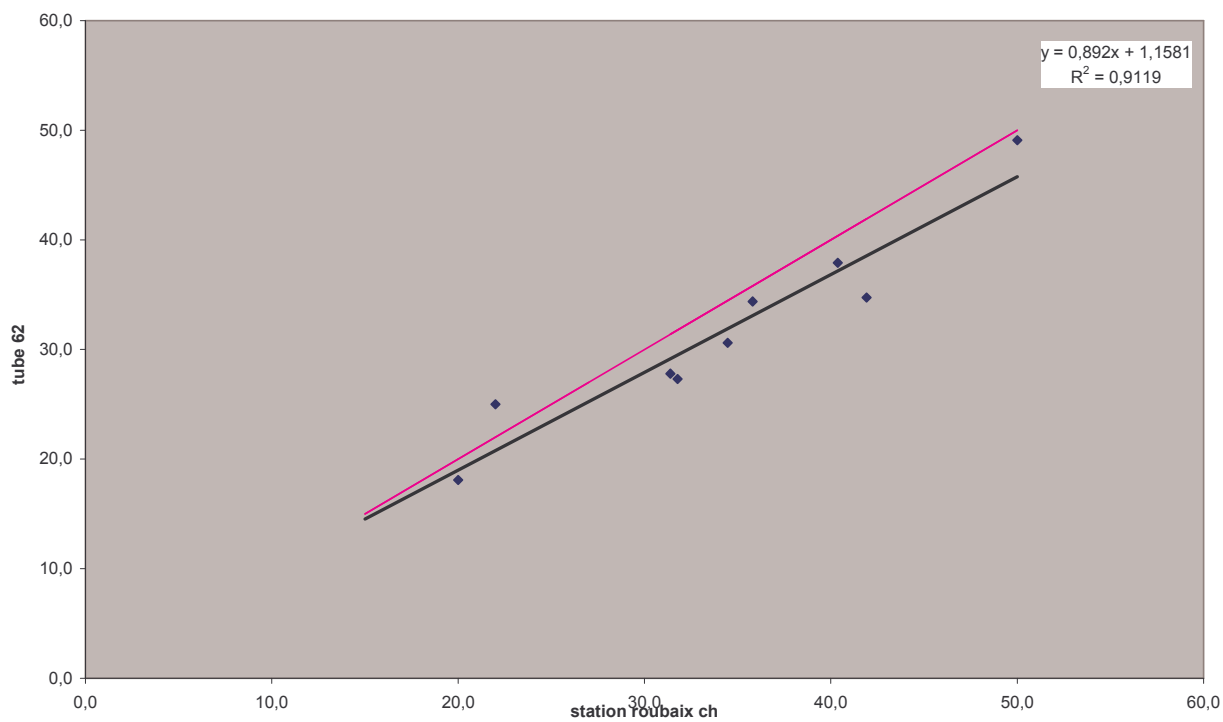


Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

site 161

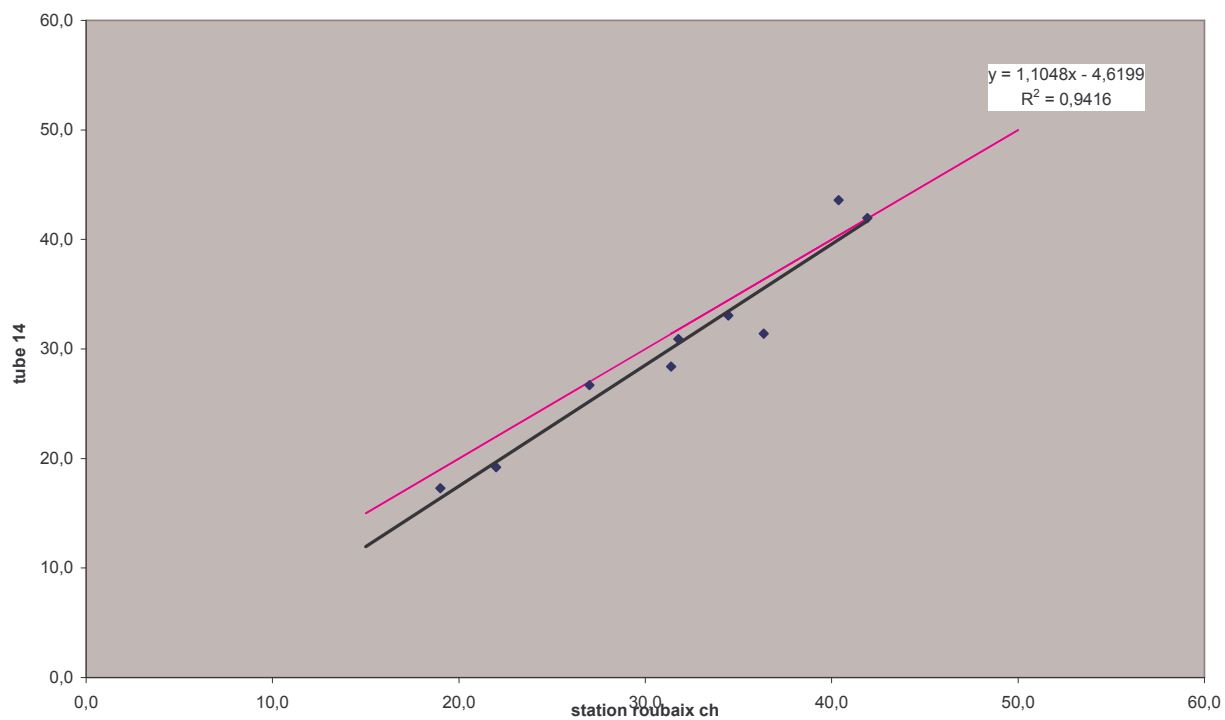


site 62



Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

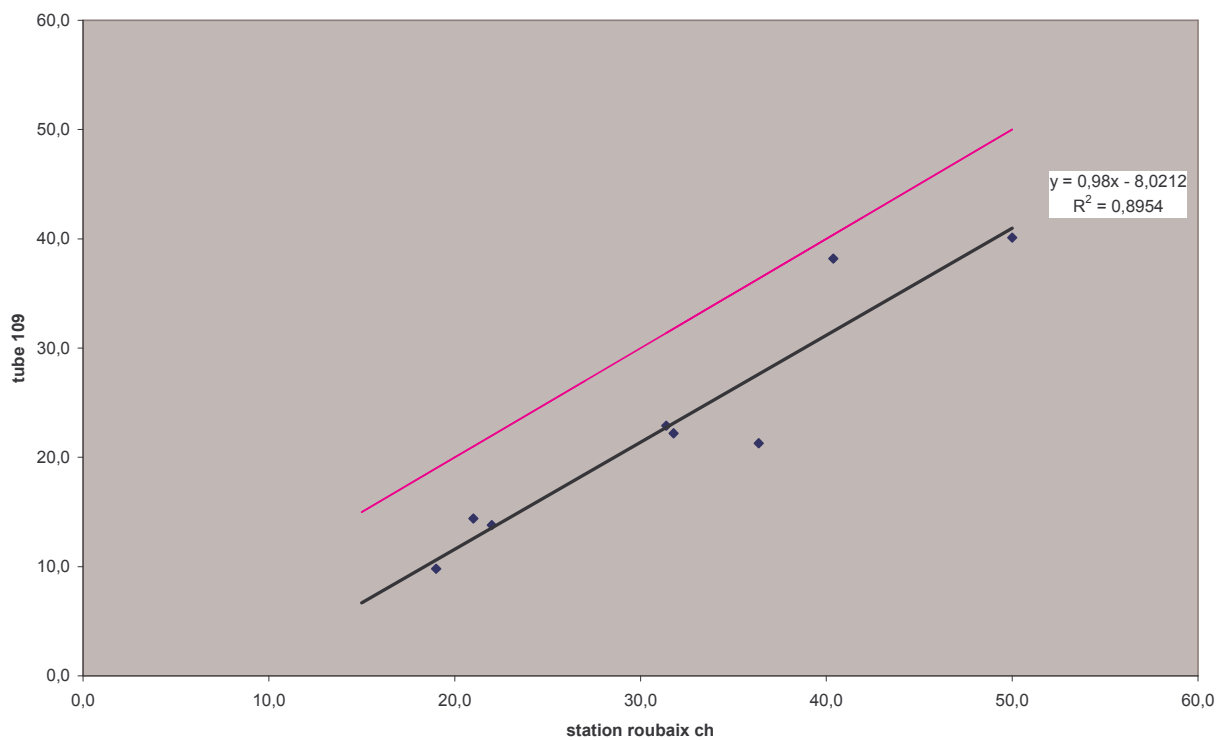
Site 14



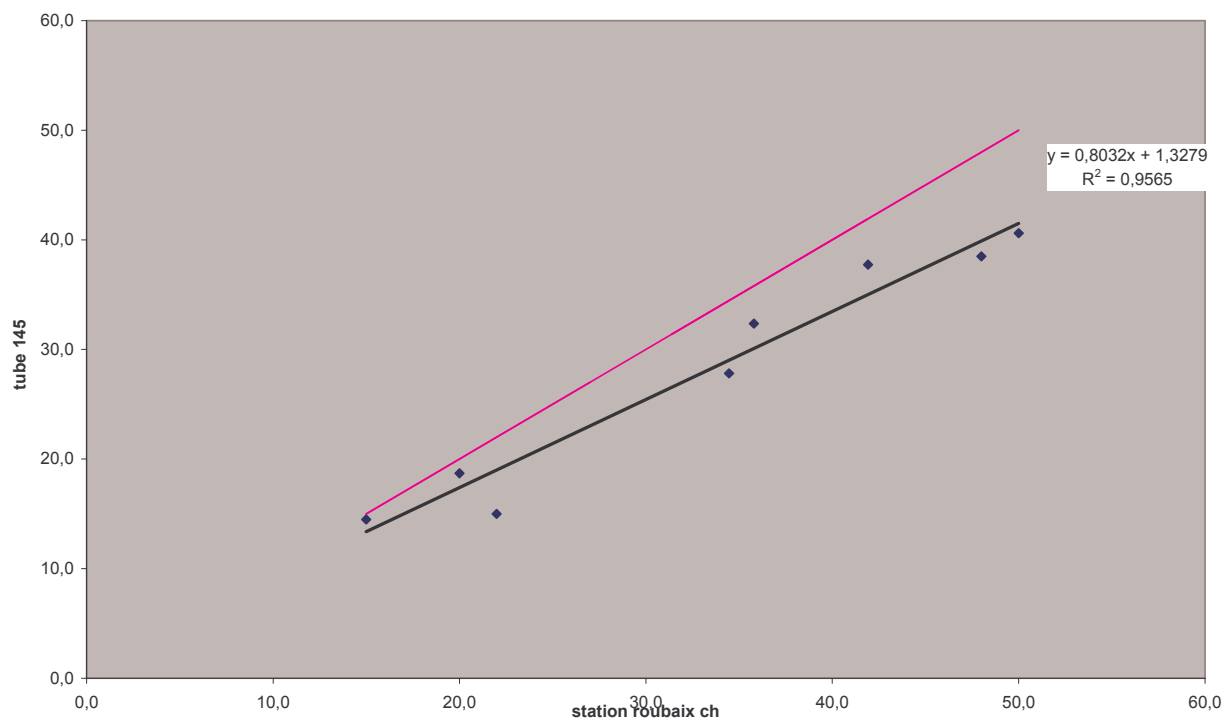
Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

N=8
(nombre de campagnes de mesure)

site 109

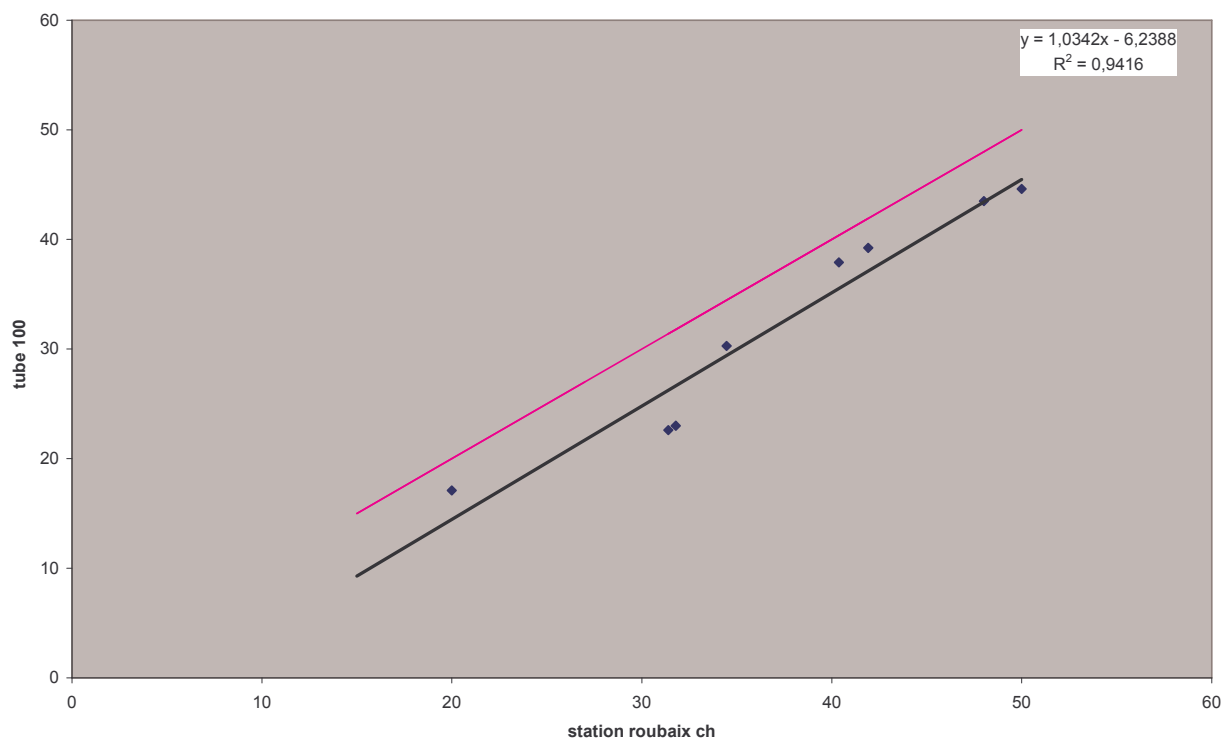


site 145



Annexe 1 : Corrélations avec la station urbaine « Roubaix château »

site 100



ANNEXE 2

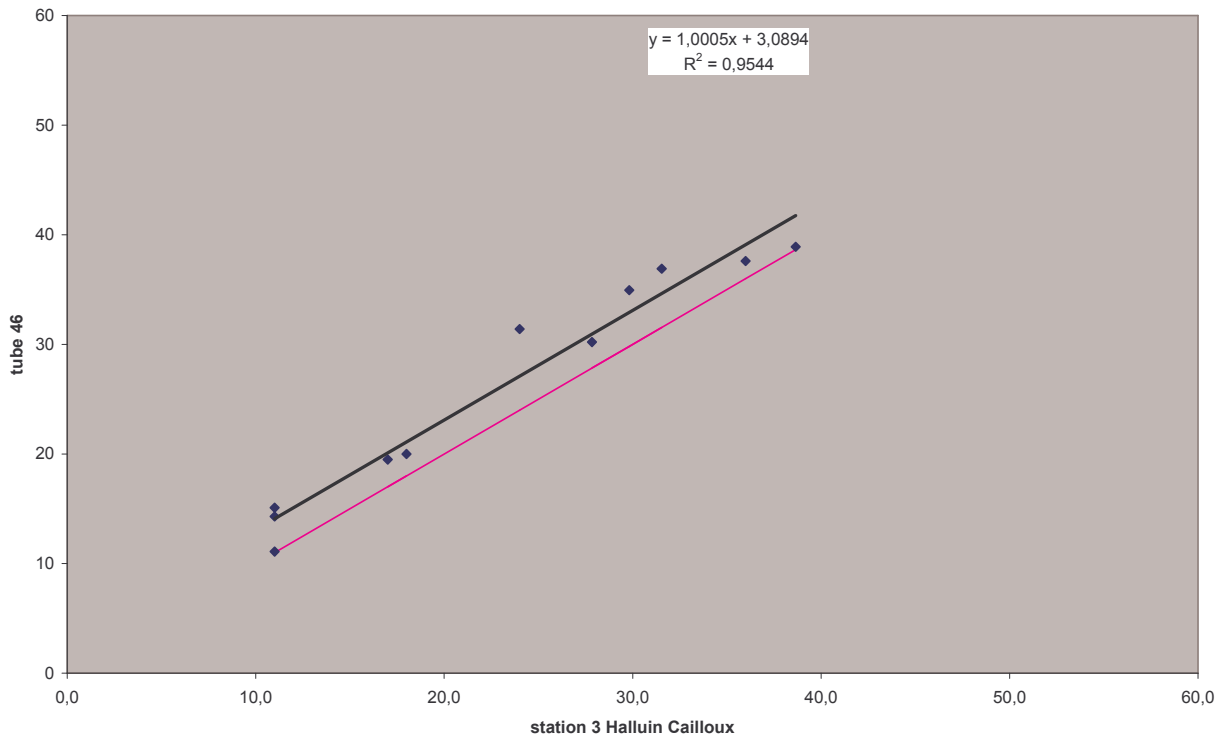
Corrélations entre des sites d'échantillonnage passif du NO₂ et la station péri-urbaine "Halluin cailloux"

Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

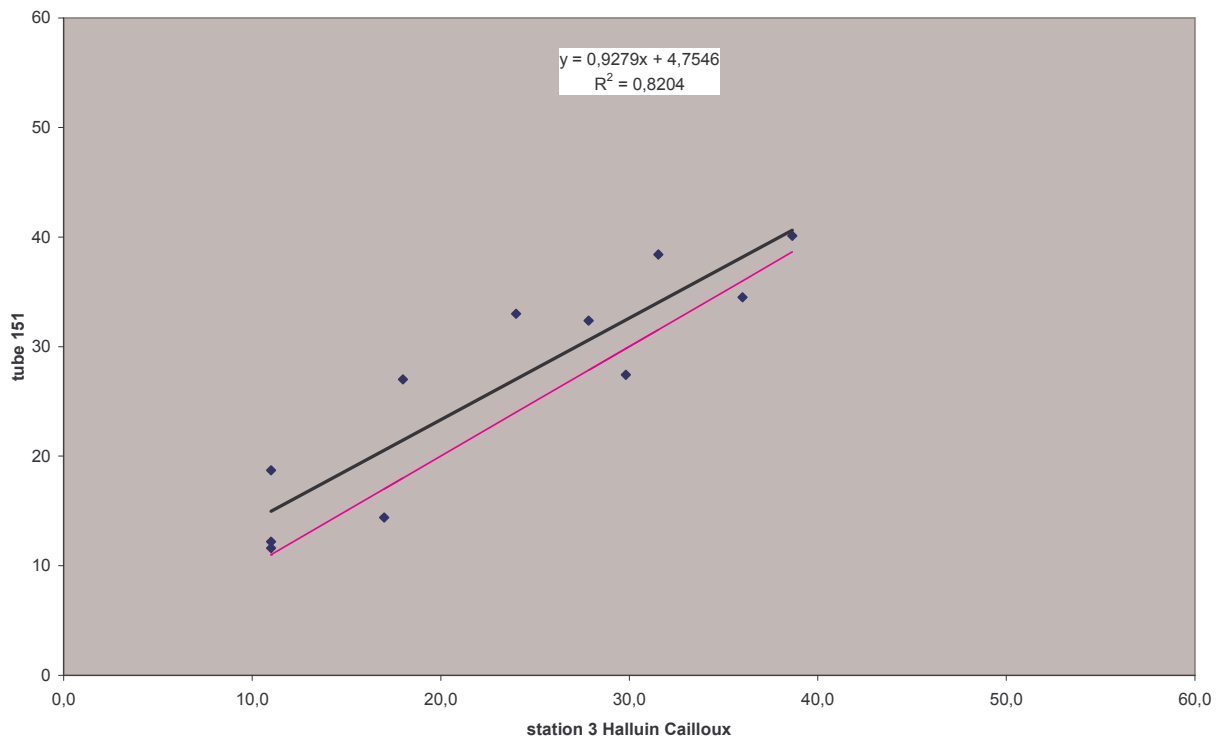
Remarque: Pour toutes les figures de l'annexe 2, la droite en rose est la 1^{ère} bissectrice, celle en noir correspond à la régression linéaire trouvée.

N=11
(nombre de campagnes de mesure)

site 46

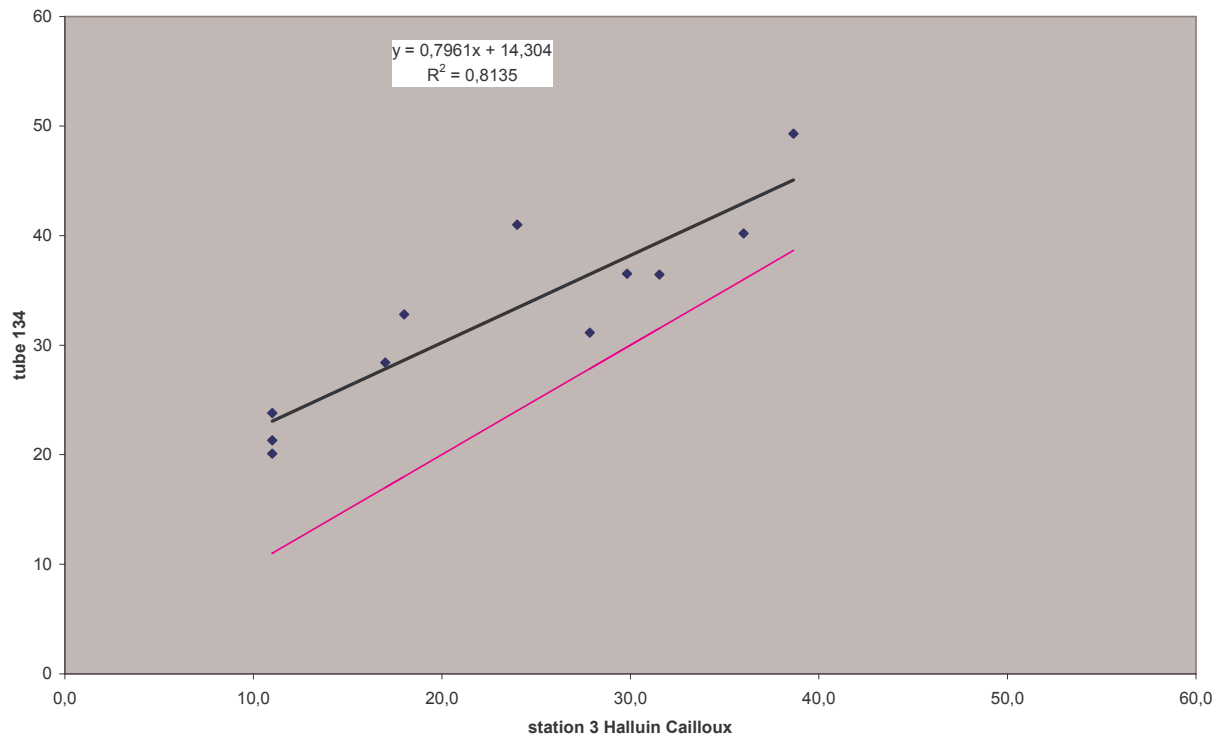


site 151

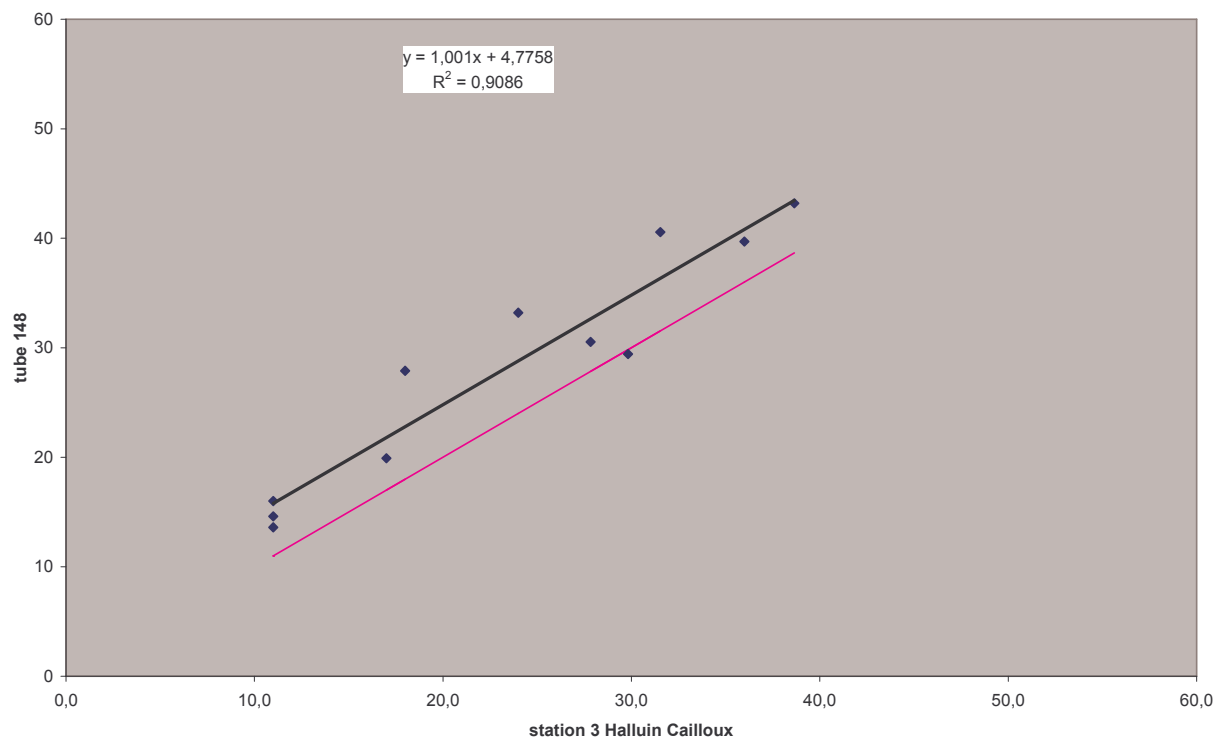


Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

site 134

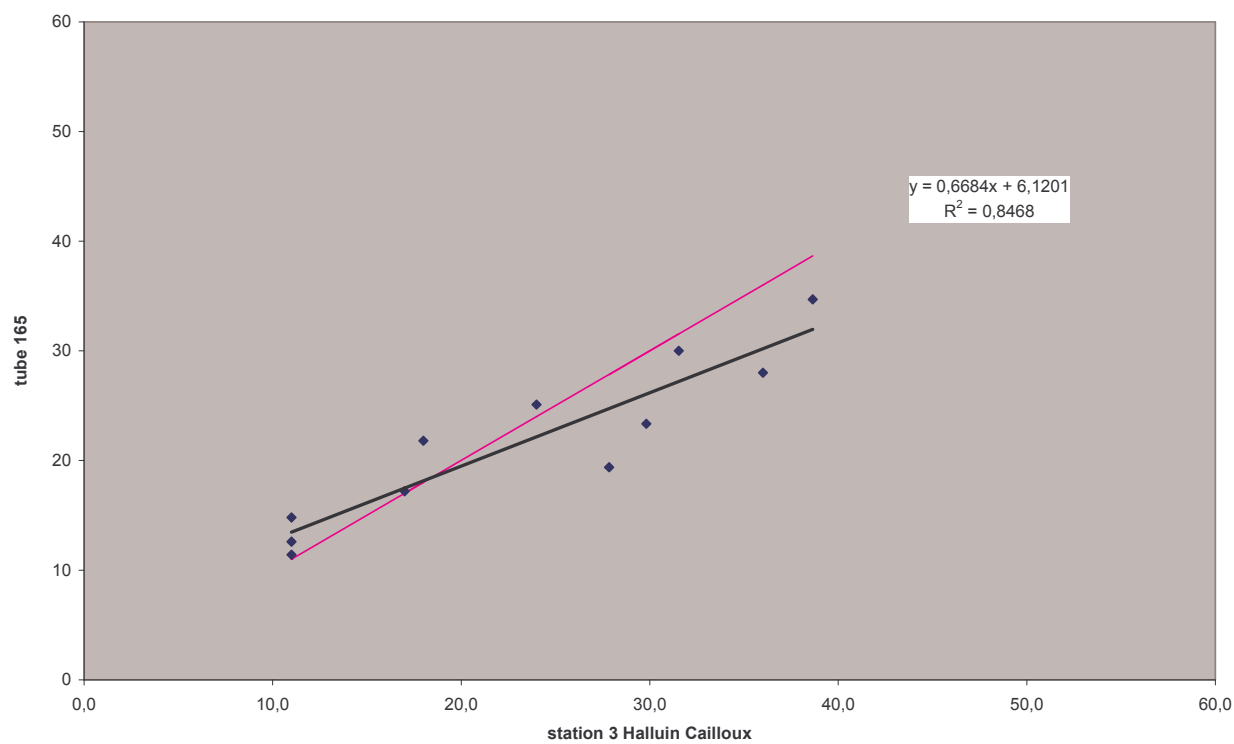


site 148



Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

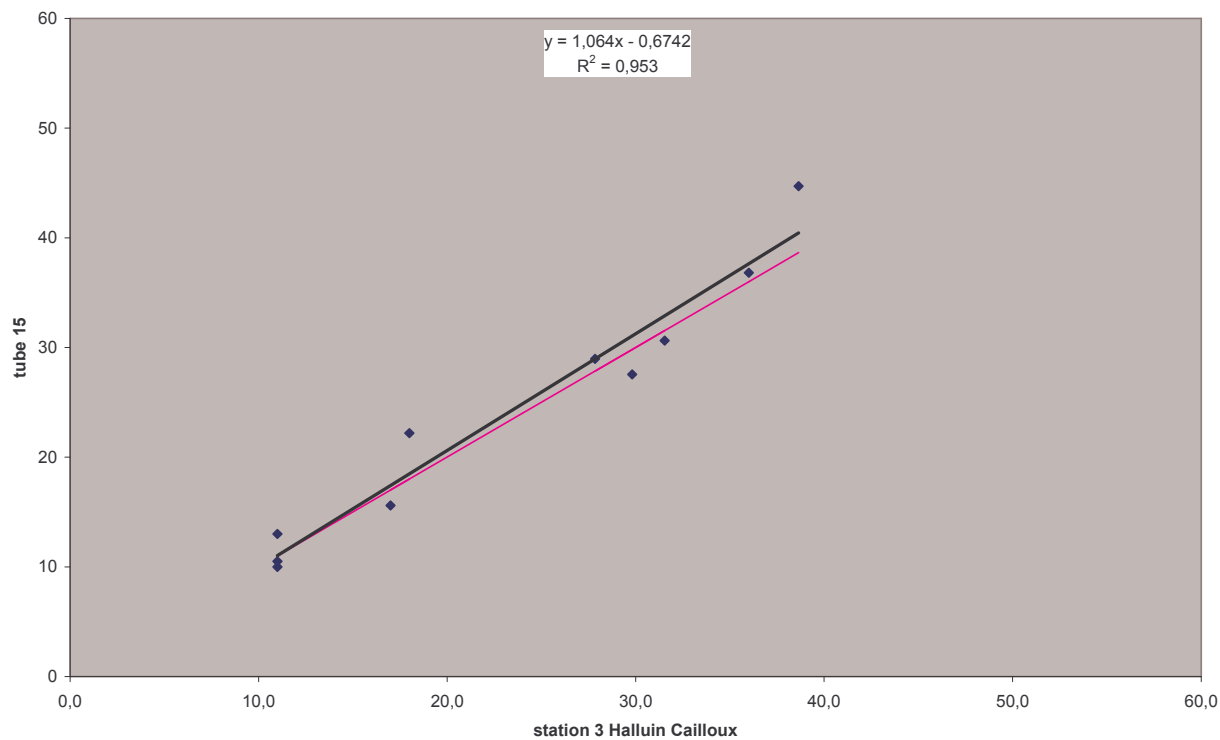
Site 165



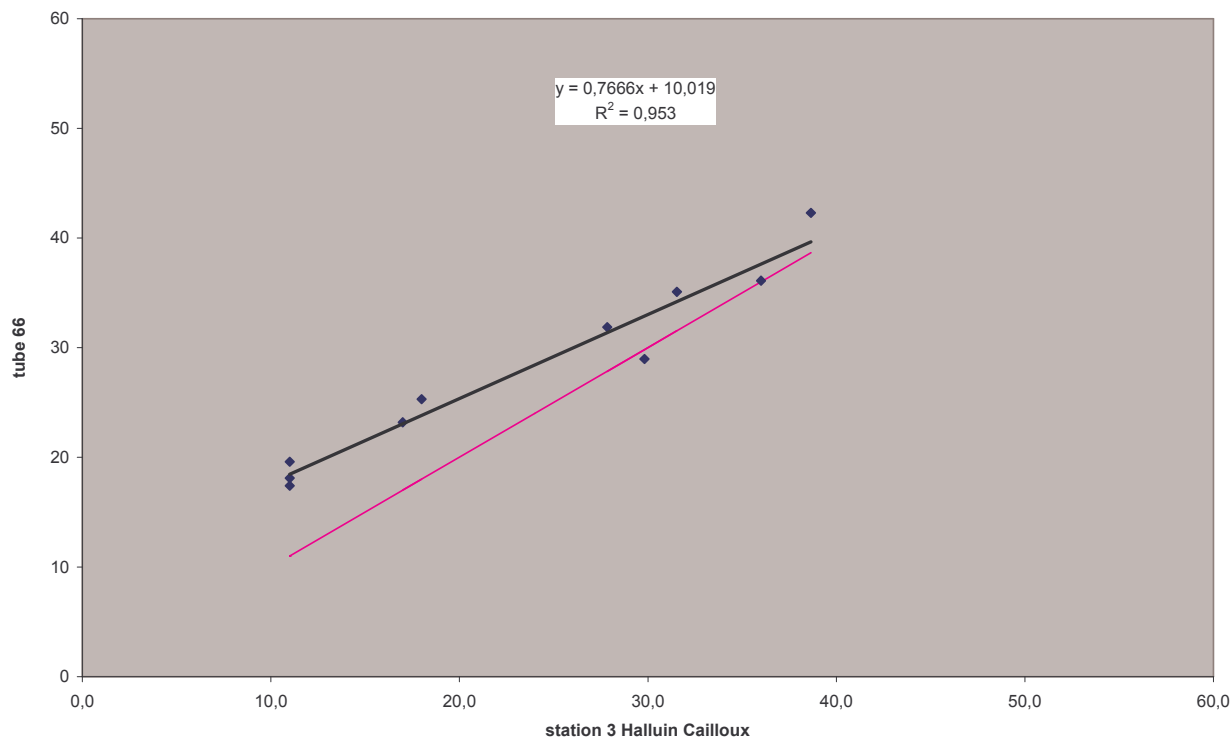
Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

N=10
(nombre de campagnes de mesure)

site 15

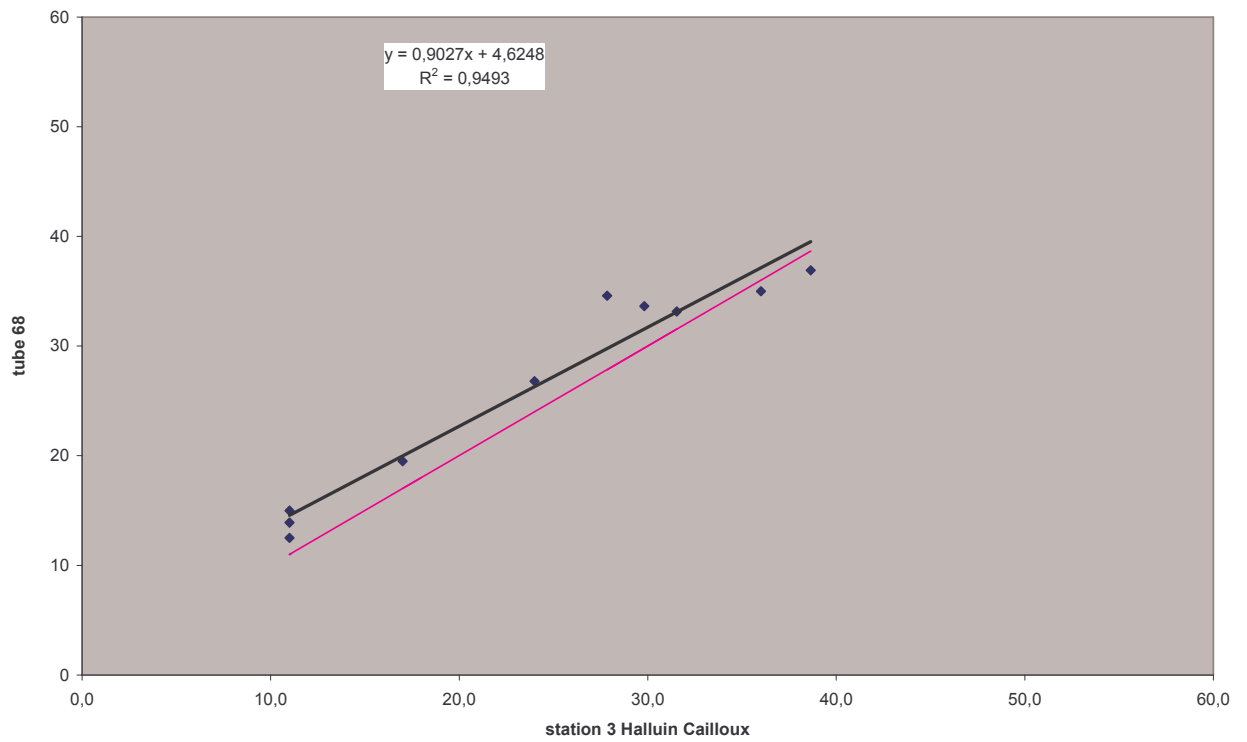


site 66

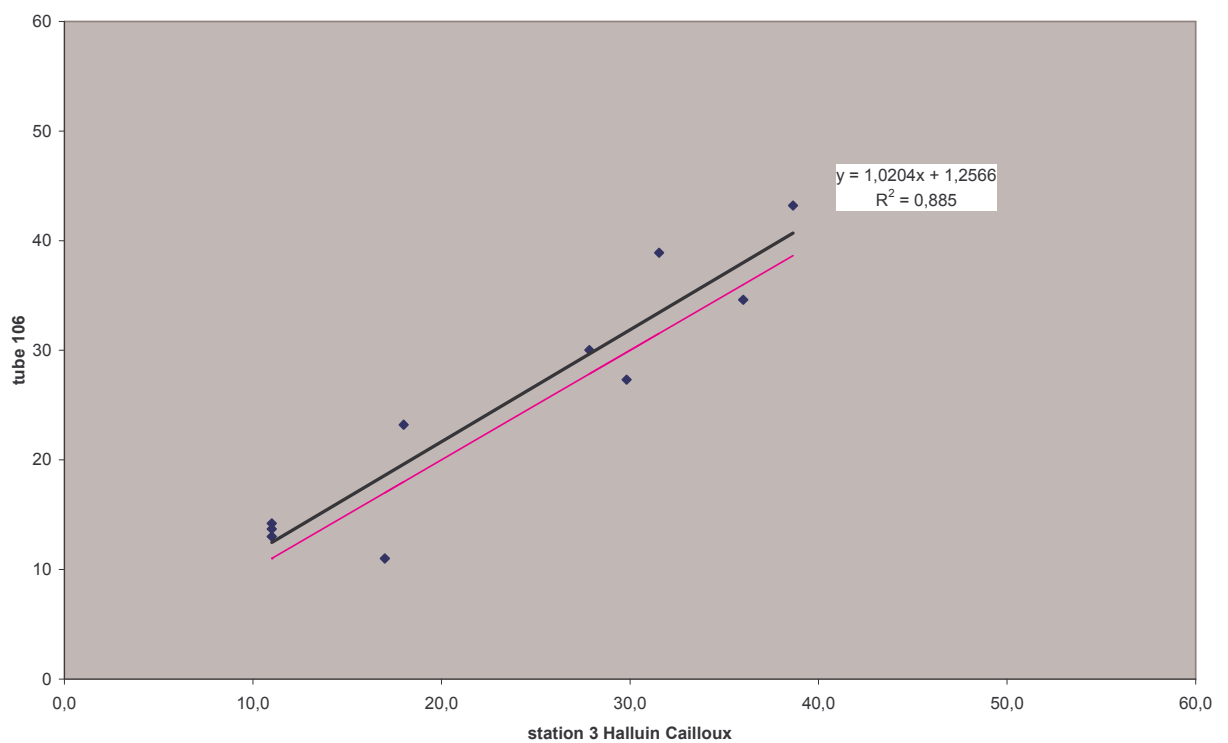


Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

site 68

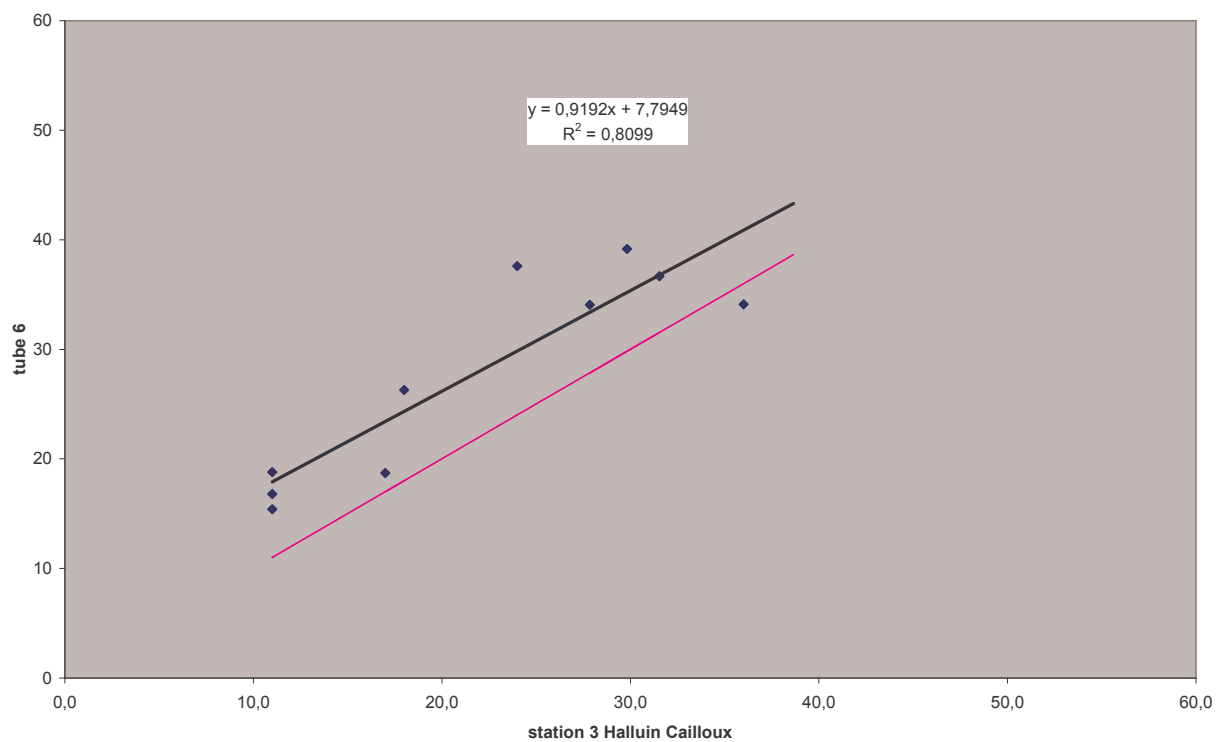


site 106

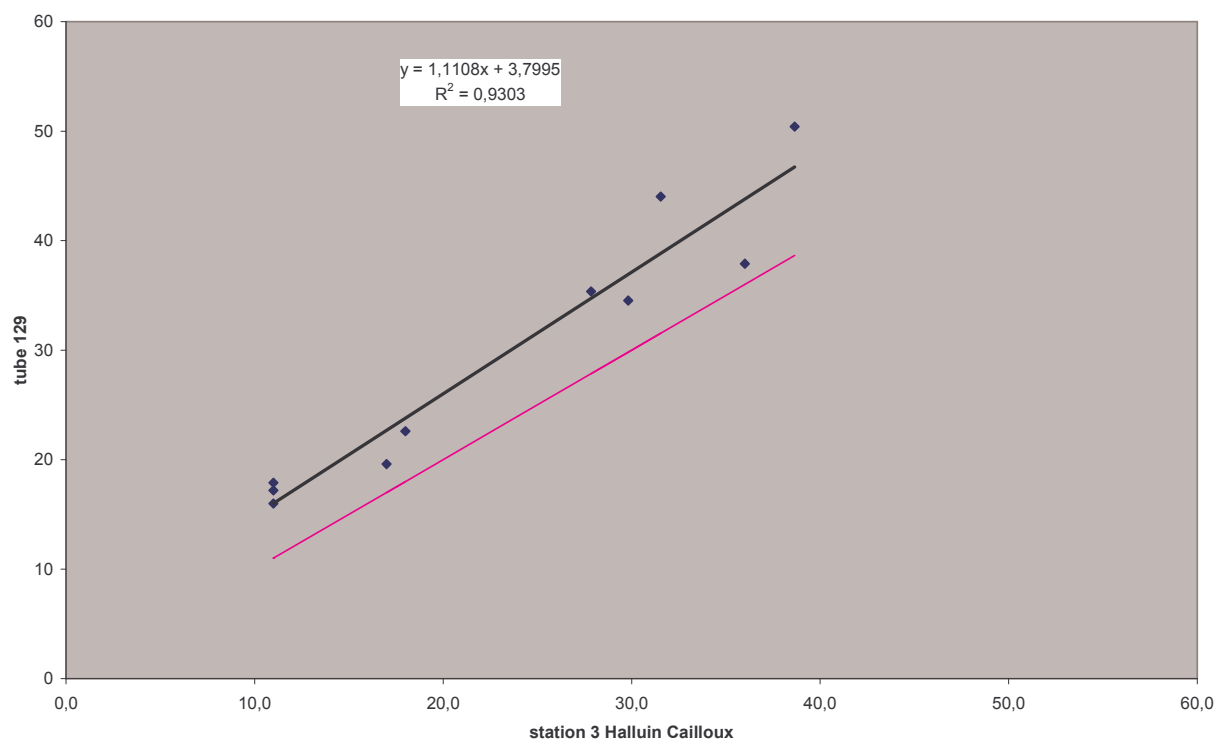


Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

Site 6

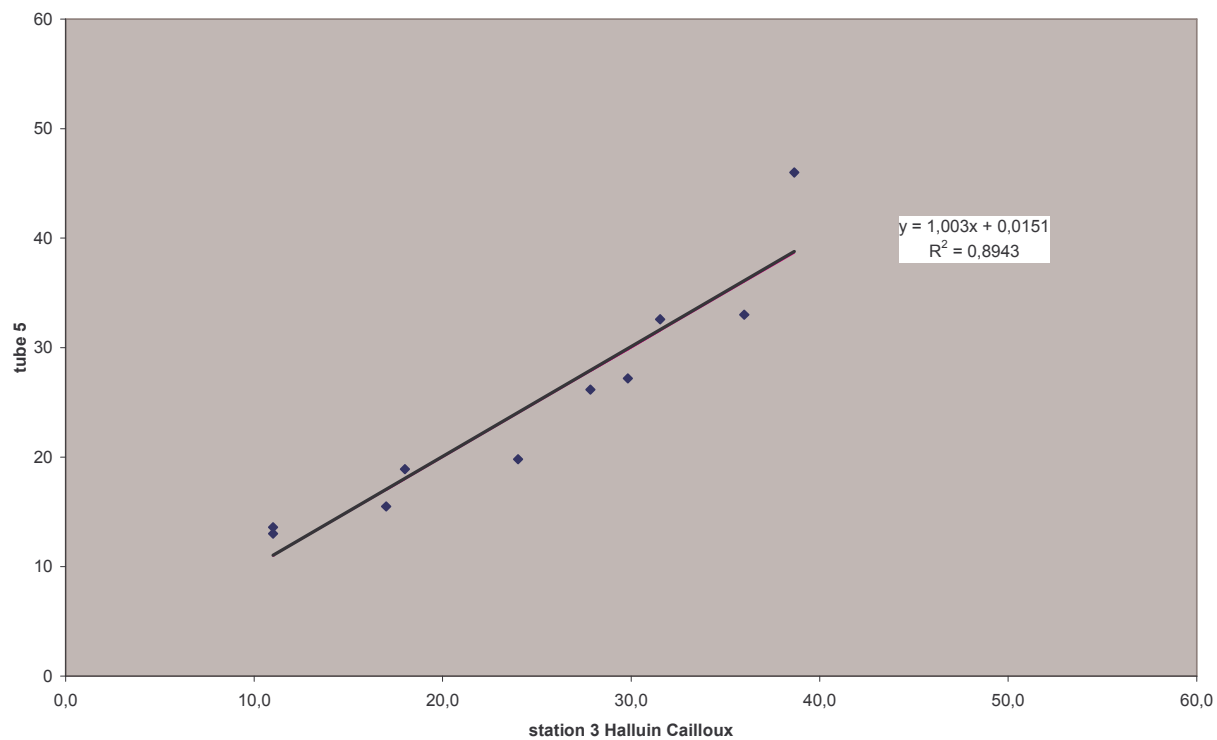


Site 129



Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

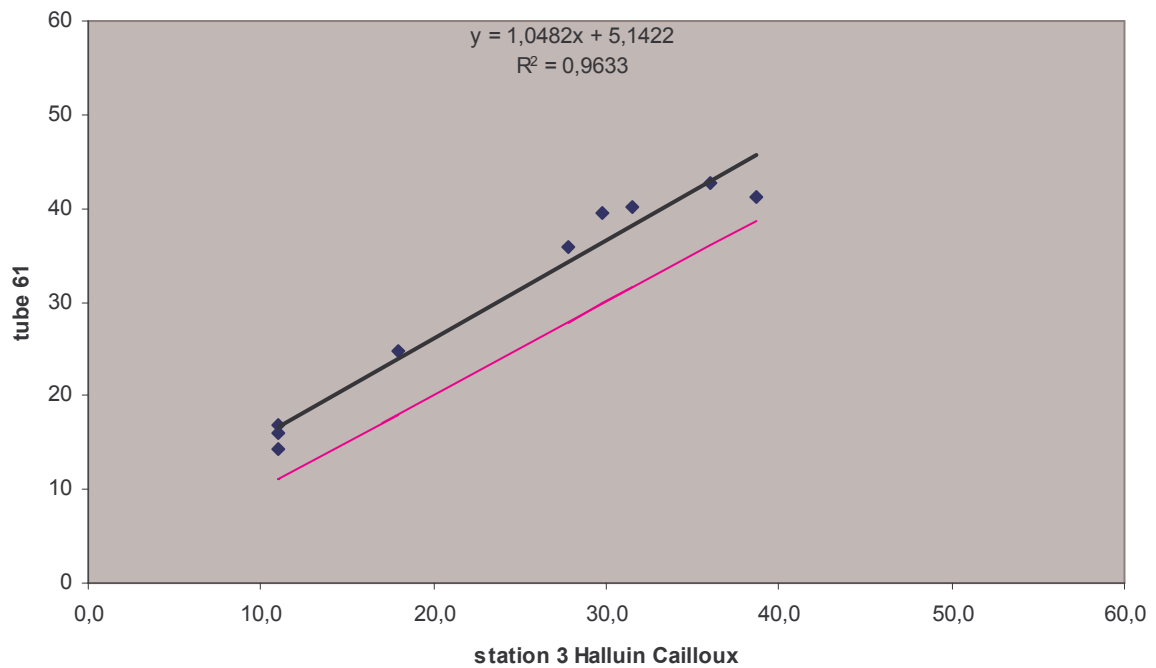
Site 5



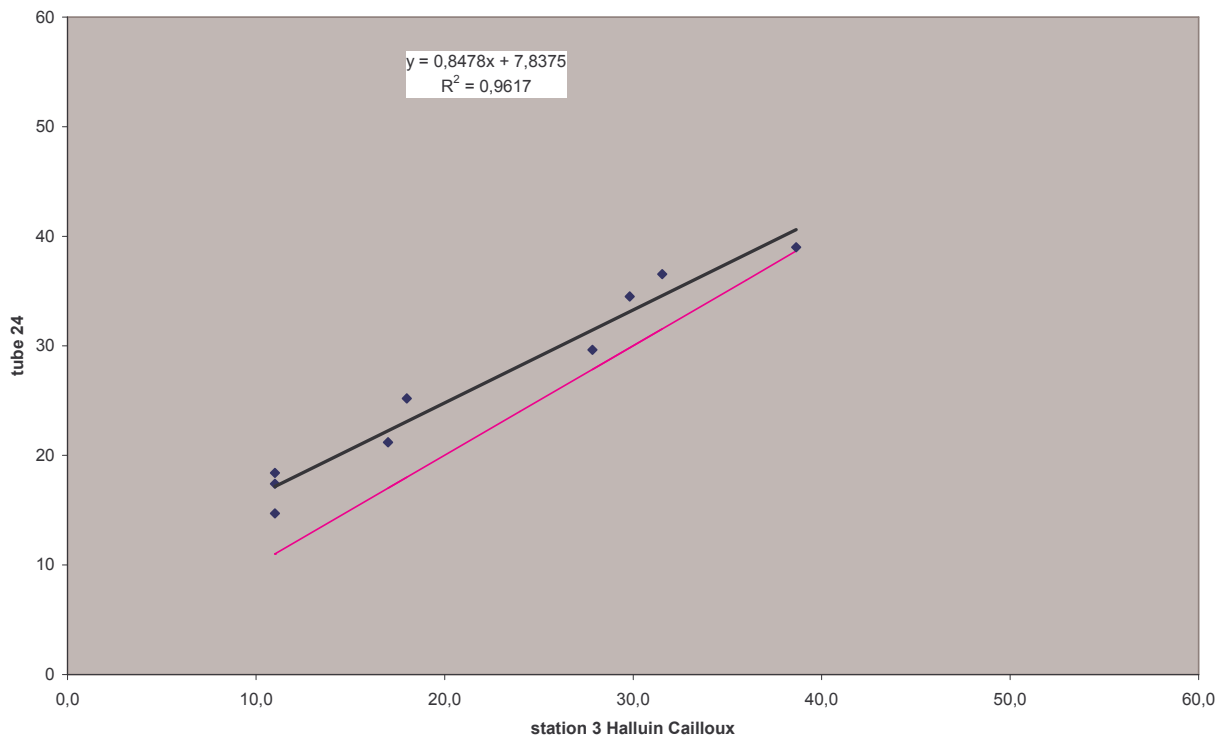
Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

N=9
(nombre de campagnes de mesure)

site 61

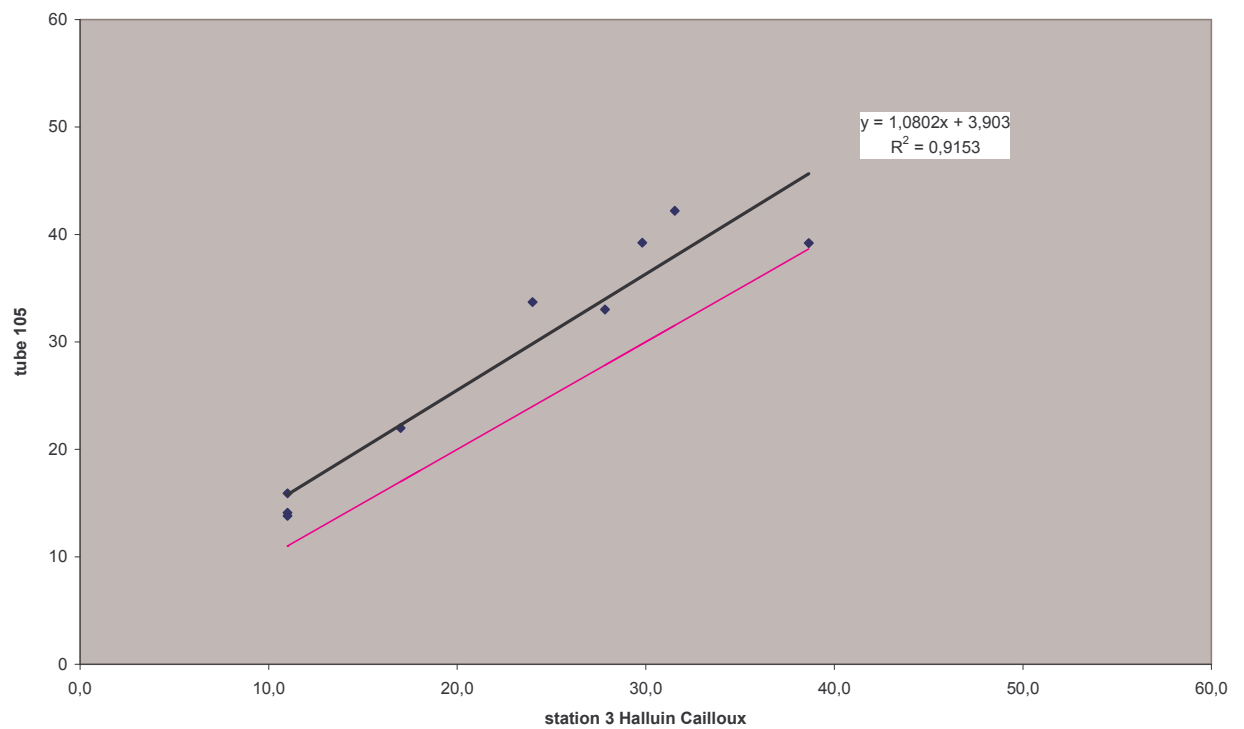


site 24

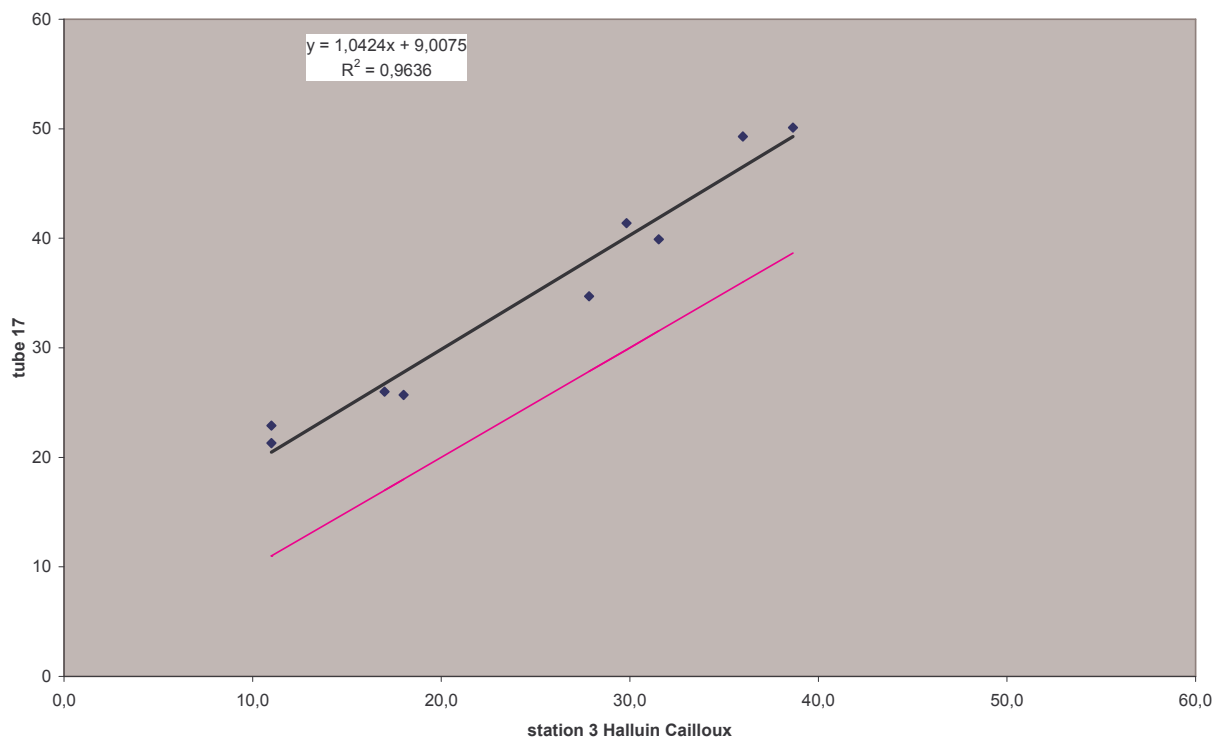


Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

site 105

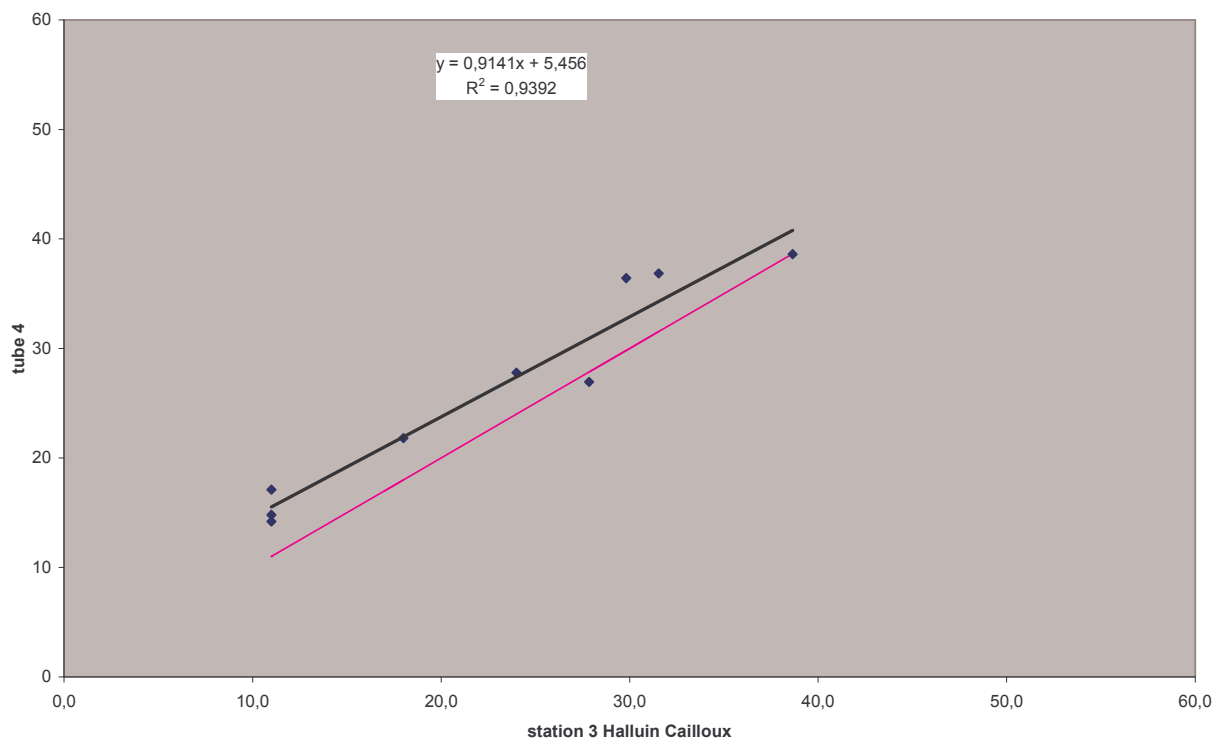


site 17

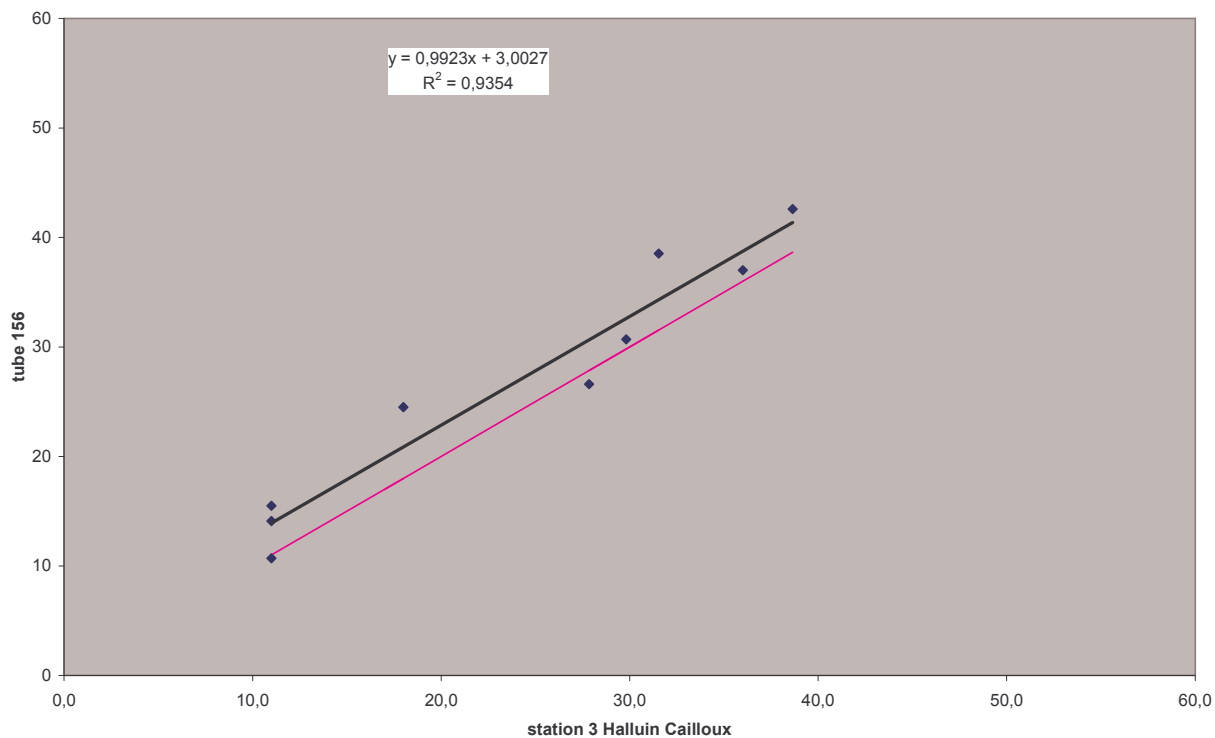


Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

Site 4

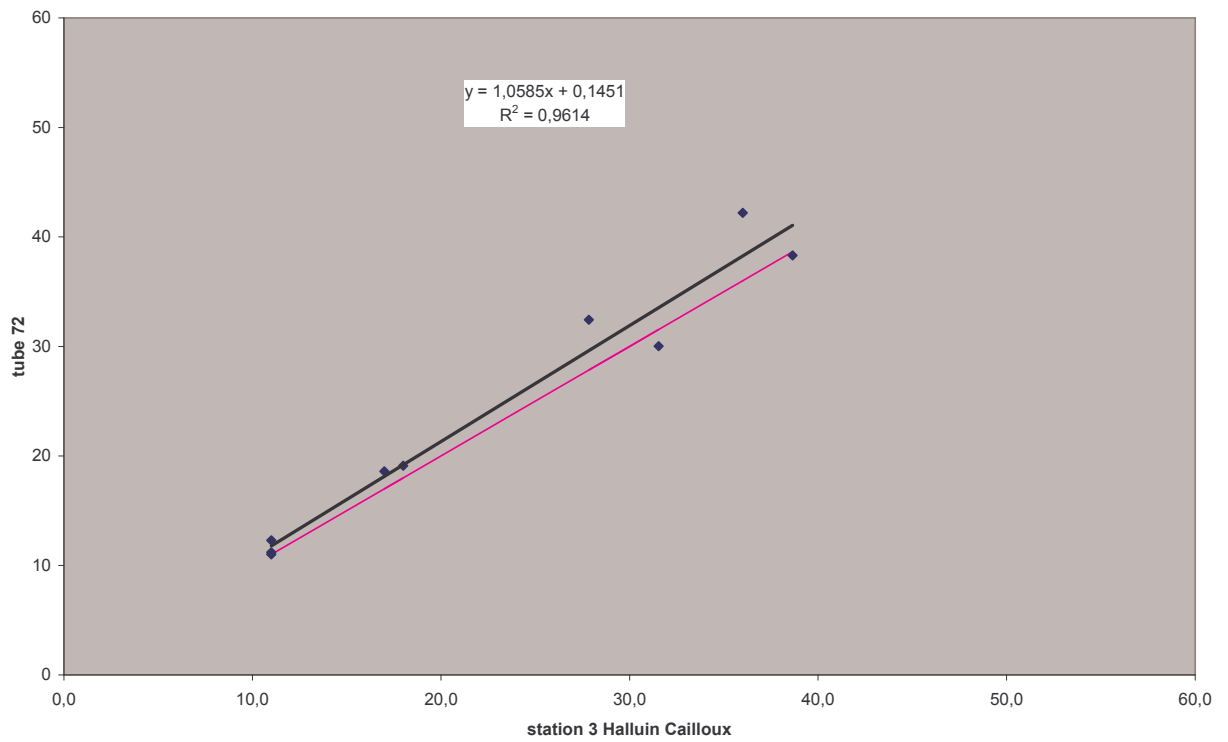


Site 156

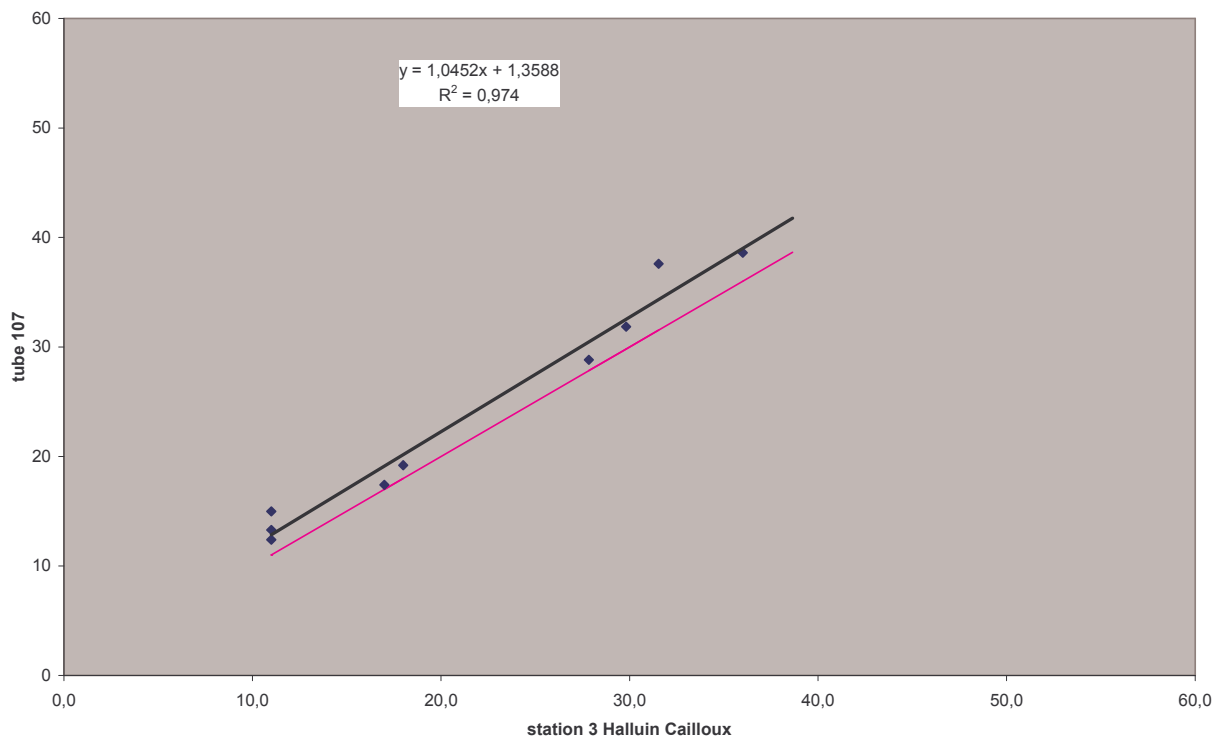


Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

Site 72

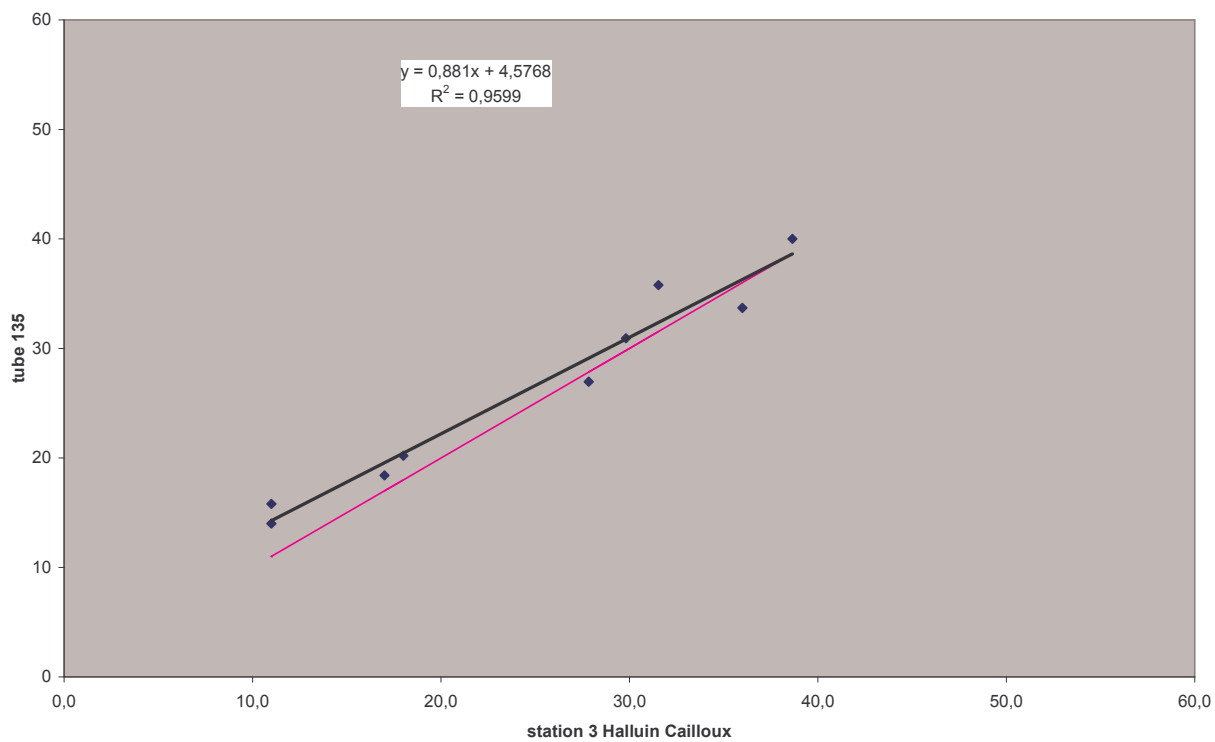


Site 107

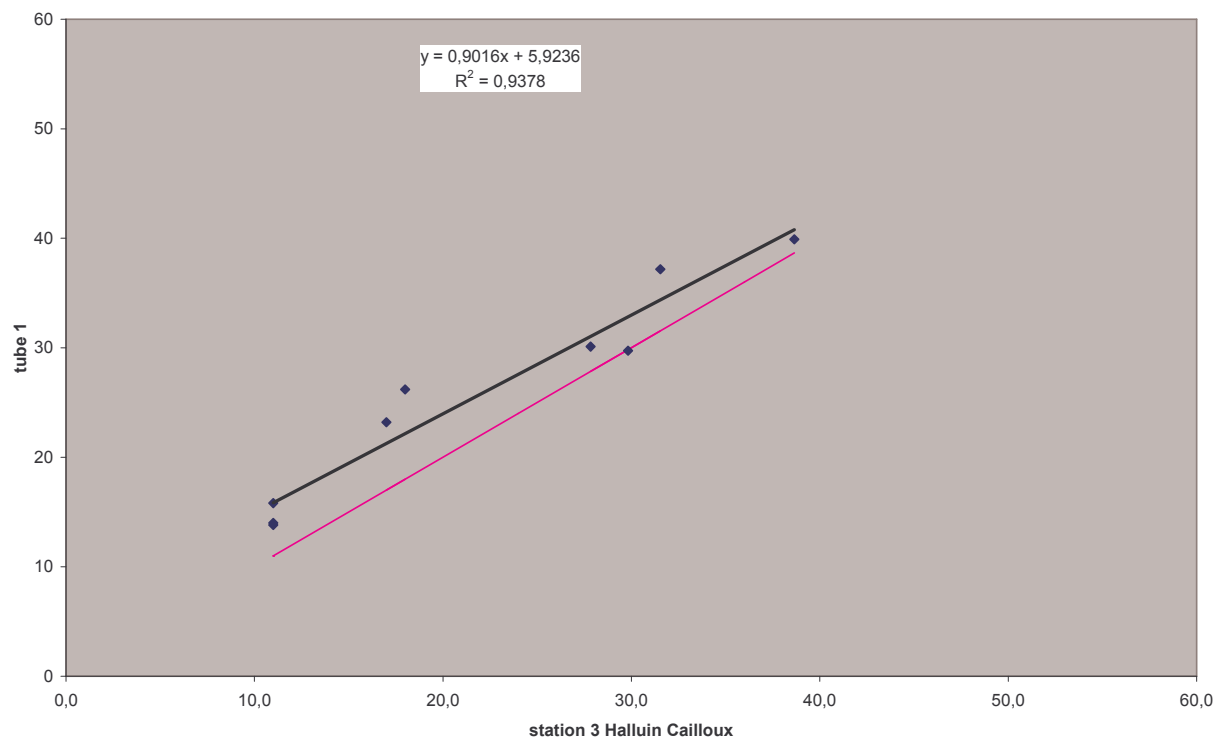


Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

Site 135

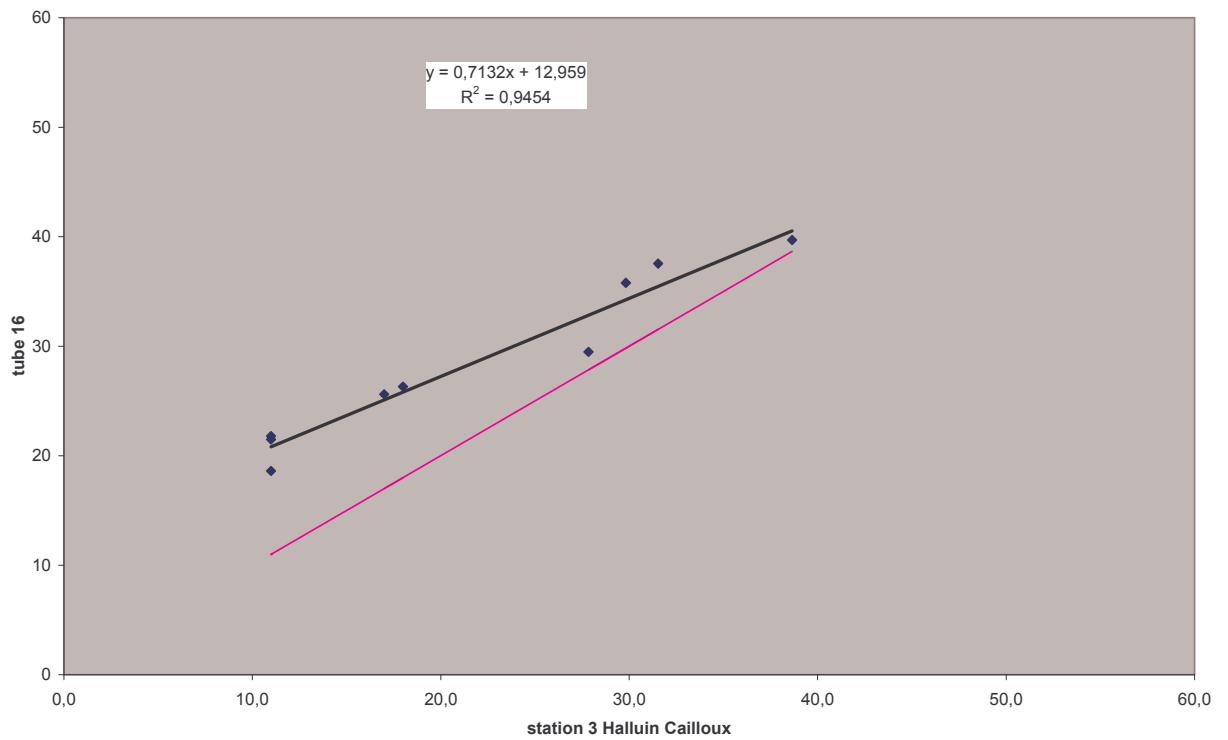


Site 1

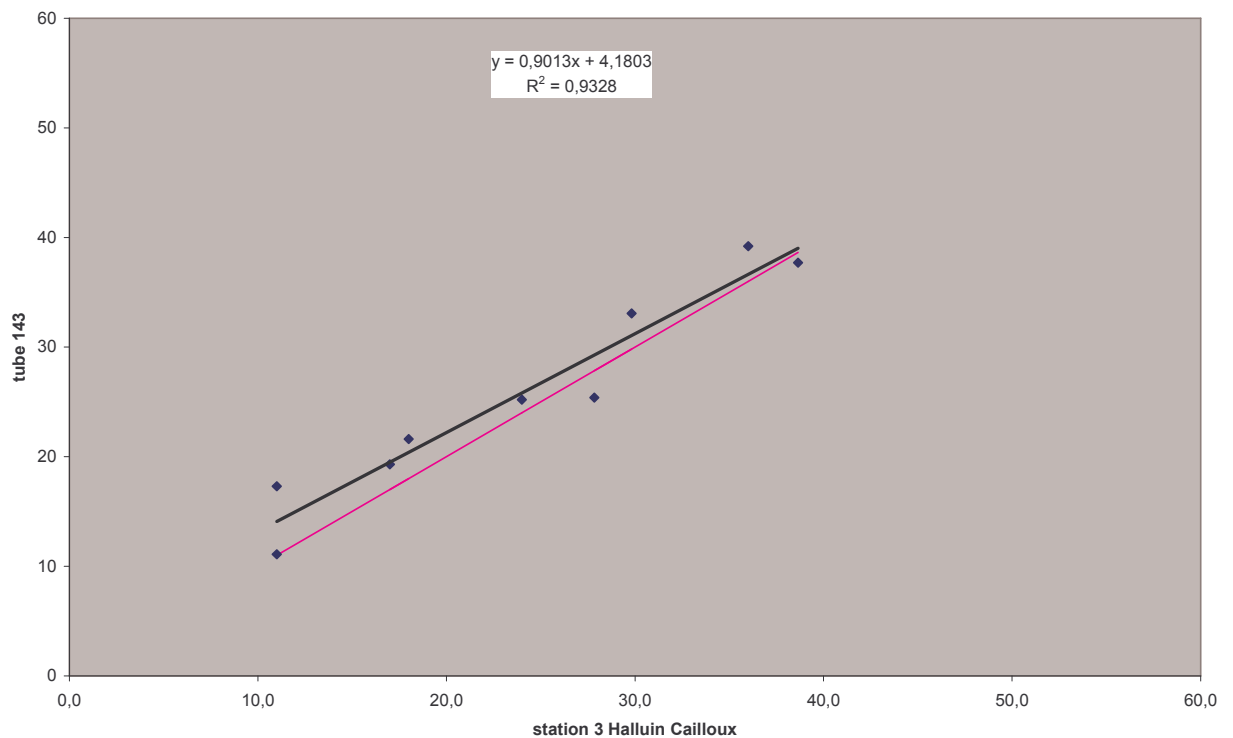


Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

Site 16

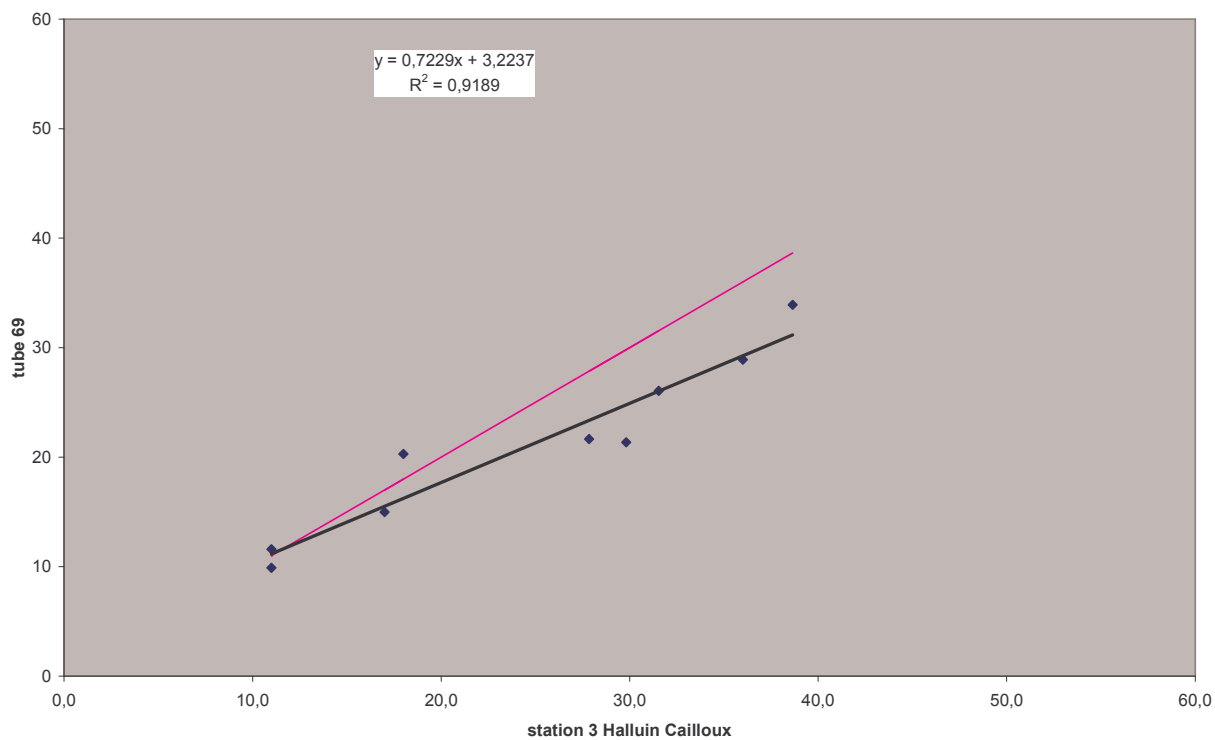


Site 143

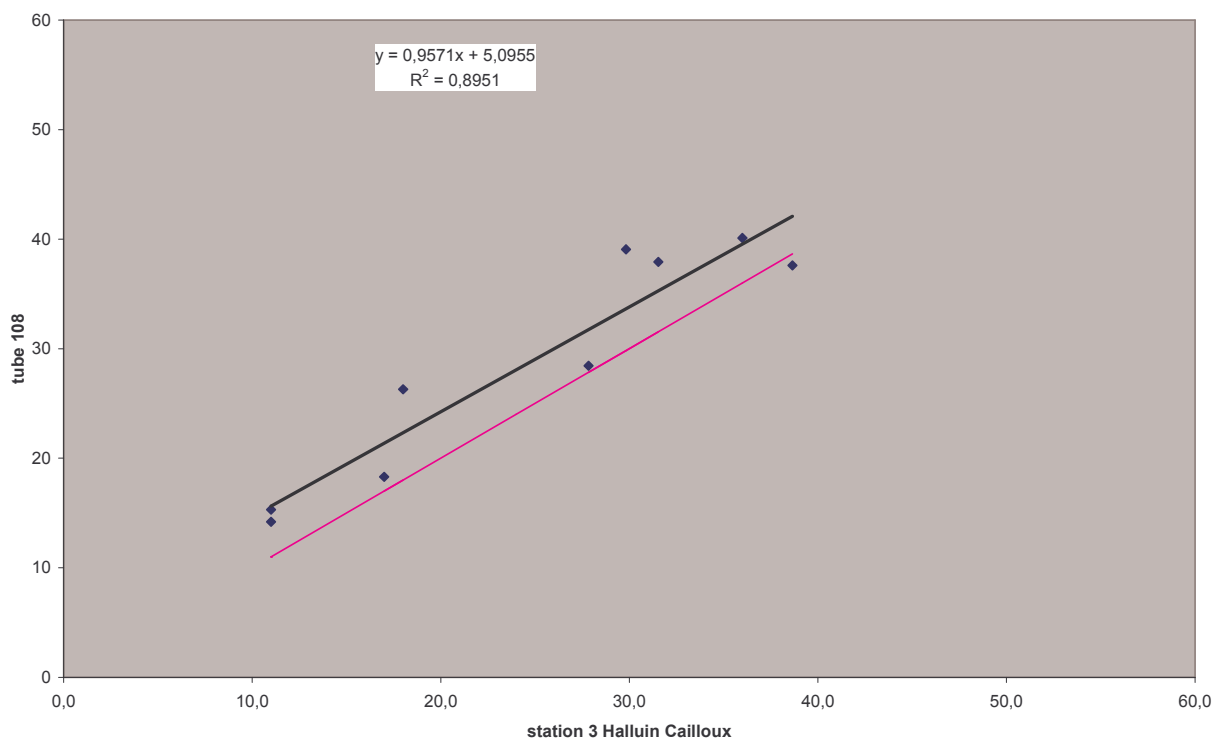


Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

Site 69



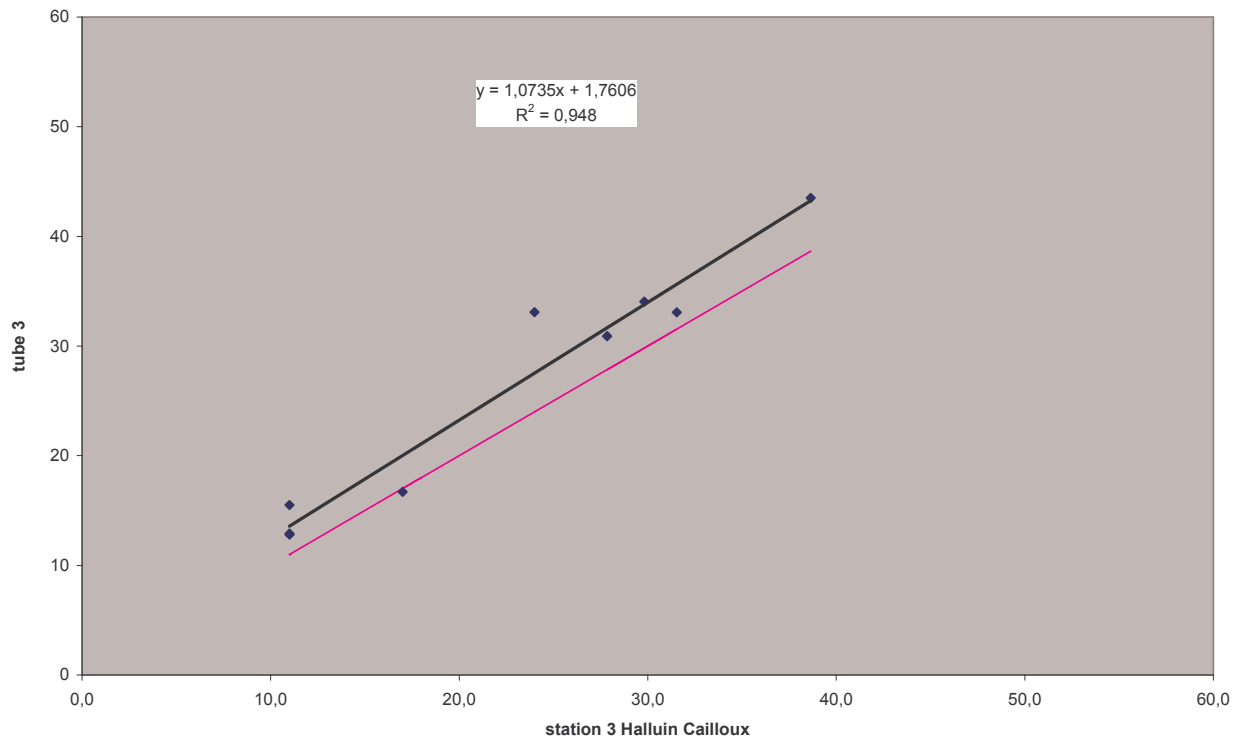
Site 108



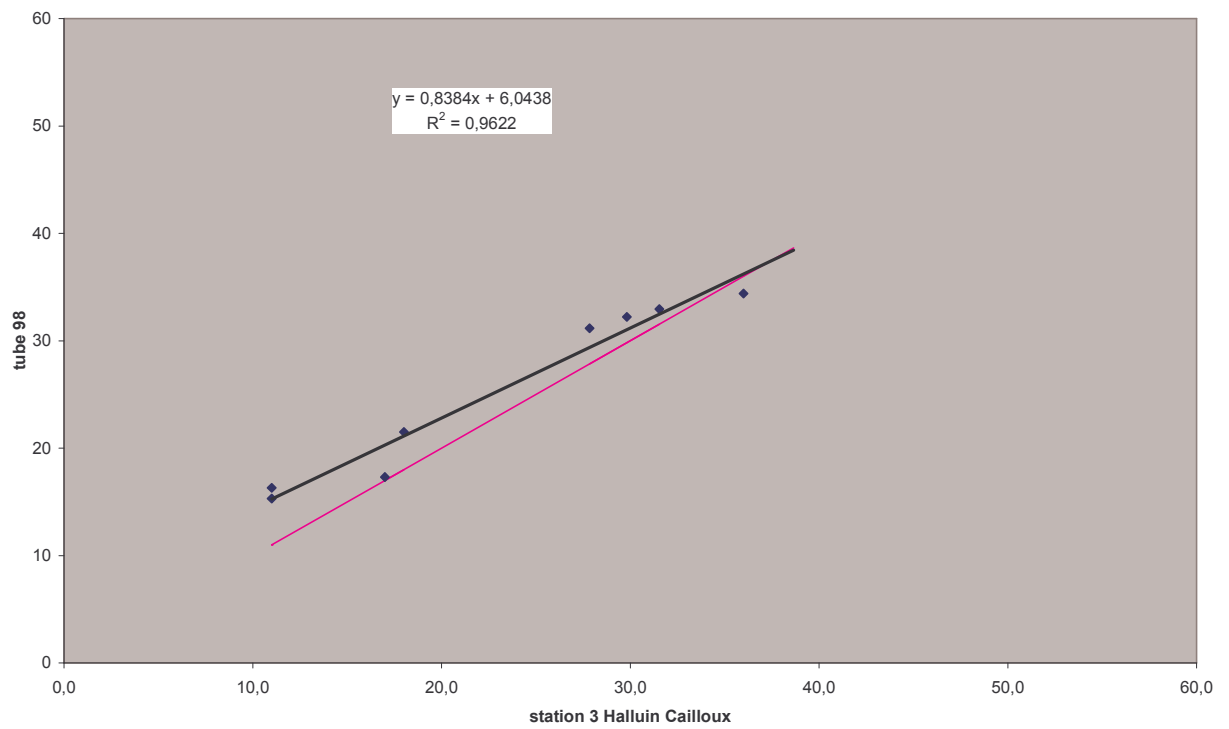
Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

N=8
(nombre de campagnes de mesure)

site 3 (site de la station)

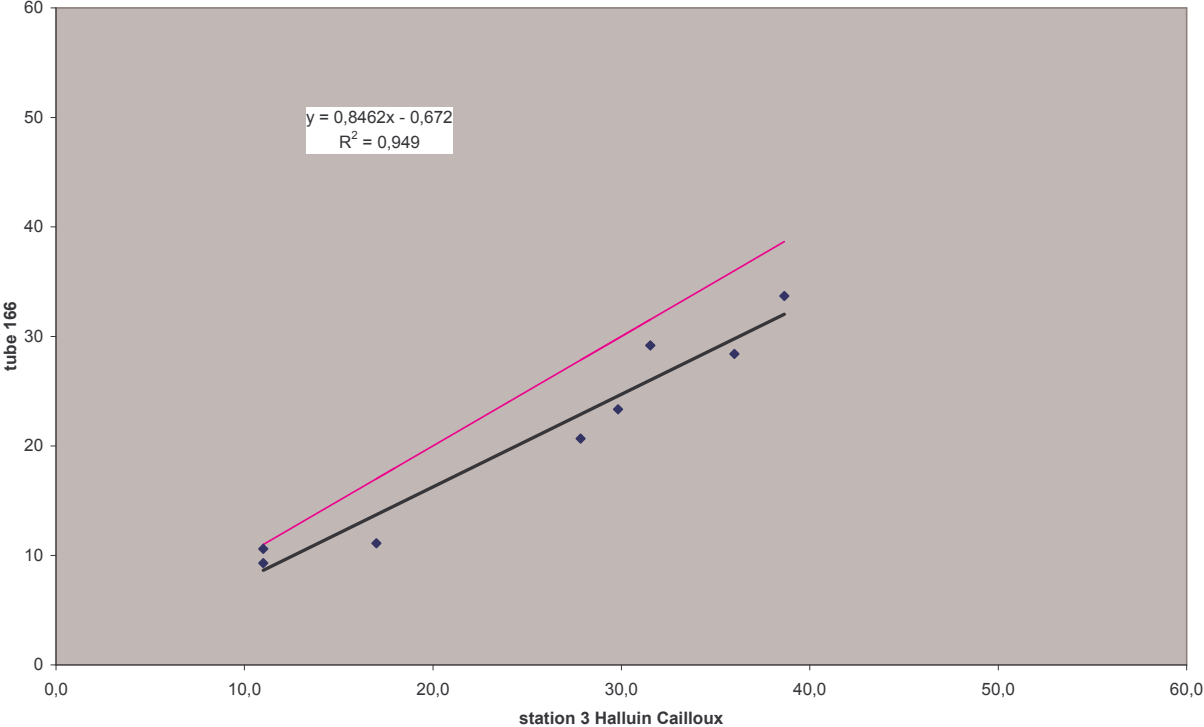


site 98



Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

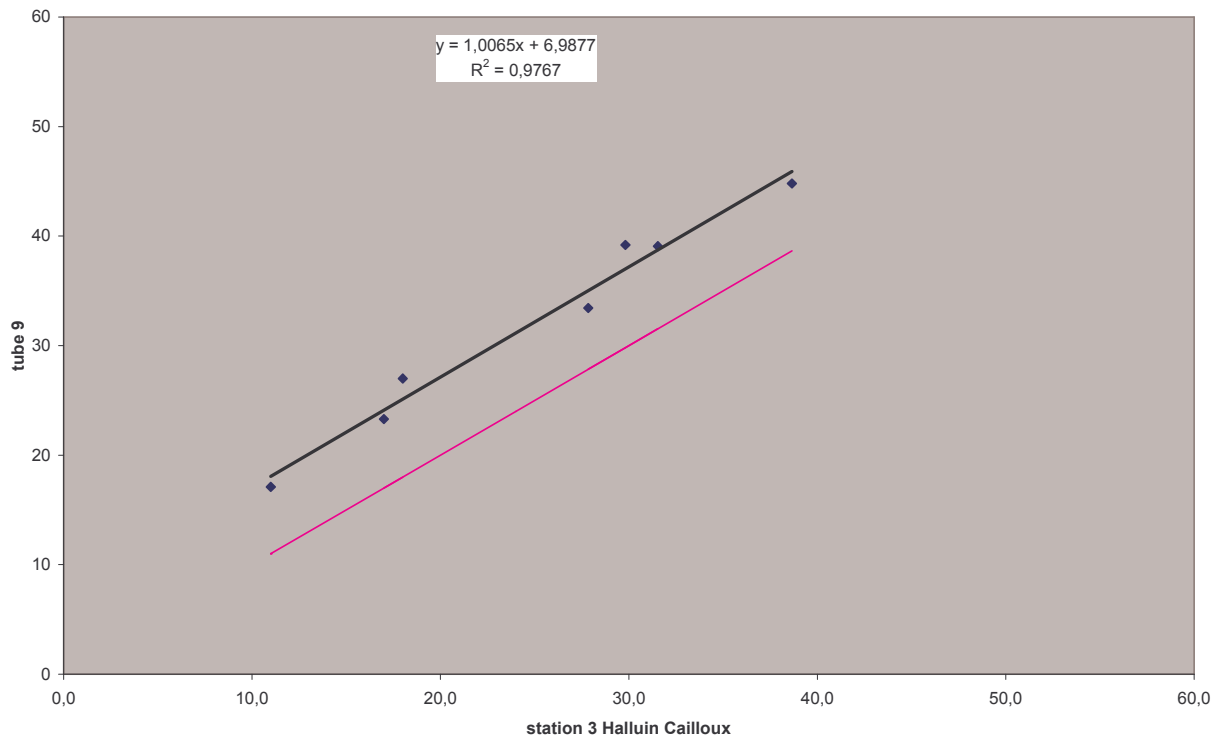
site 166



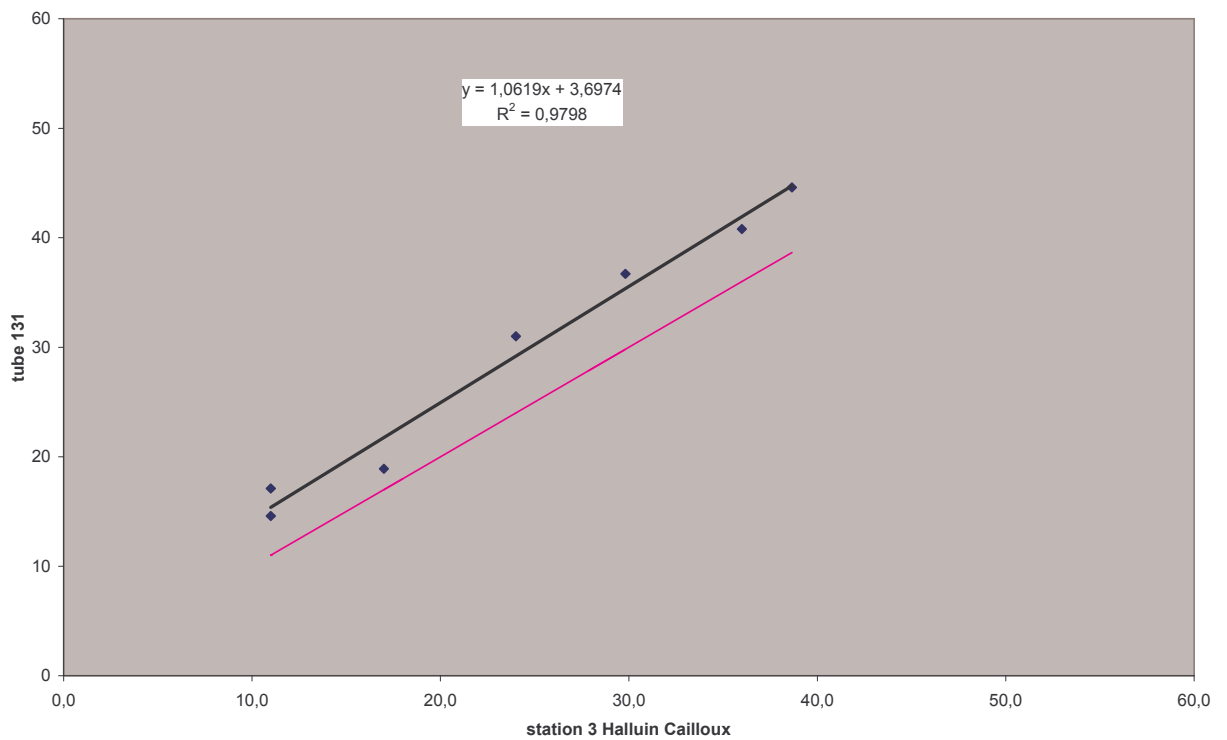
Annexe 2 : Corrélations avec la station péri-urbaine « Halluin cailloux »

N=7
(nombre de campagnes de mesure)

site 9



site 131



ANNEXE 3

Recherche des corrélations entre stations automatiques

Etude des régressions linéaires entre stations automatiques

Démarche suivie :

- Calcul des régressions linéaires entre stations automatiques, sur une base temporelle de 14 jours, à partir de données établies sur plusieurs années.
- Vérification de la validité des régressions pour une année différente.
- Estimation des valeurs moyennées sur 7 jours ou 24h à partir des régressions obtenues sur une base de 14 jours, et calcul de la précision des estimations.

Calculs effectués automatiquement au moyen du package Excel:

Préparation des données

Une feuille par station comprenant :

- Moyenne glissante sur 24h des mesures de la station
- Moyenne glissante sur 7 jours des mesures de la station
- Moyenne glissante sur 14 jours des mesures de la station

Calculs des régressions

Une feuille par couple de station et par type de moyenne (24h, 7j, 14j) comprenant les paramètres de corrélation :

- le coefficient de détermination (r^2), égal au rapport de la variance expliquée par le modèle sur la variance totale des données;
- la pente de la droite de régression (a) et son ordonnée à l'origine (b) ;
- la moyenne des valeurs de la station de référence (en abscisse) ;
- la moyenne de l'erreur d'estimation des valeurs de la station en ordonnée à partir des valeurs de la station de référence ;
- la variance de l'erreur d'estimation, calculée selon l'équation (1) :

$$\sigma_{\text{erreur}}^2 = \frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2 \quad (\text{équation 1})$$

- avec e_j : erreur commise lors de l'estimation de la $j^{\text{ème}}$ valeur ;
 \bar{e} : moyenne des erreurs ;
 n : nombre total de valeurs.

Annexe 3 : Recherche des corrélations entre stations automatiques

- la précision⁶ moyenne, calculée selon l'équation (2) :

$$\text{precis_moy} = \frac{\sqrt{\sigma_{\text{erreur}}^2}}{\bar{y}} \quad (\text{équation 2})$$

- avec \bar{y} : moyenne des valeurs mesurées de la station en ordonnée.

- la moyenne de la précision absolue, calculée selon l'équation (3) :

$$\text{moy_precis} = \frac{1}{n} \sum_{i=1}^n \frac{|e_i|}{y_i} \quad (\text{équation 3})$$

- avec y_j : i^{ème} valeur mesurée de la station en ordonnée.

Présentations des résultats

Pour chaque type de moyenne (24h, 7j, 14j) présentation d'un tableau de synthèse alimenté dynamiquement. Ce tableau comporte pour chaque couple de station :

- r^2
- a
- b
- la valeur moyenne de la station de référence
- la moyenne de l'erreur d'estimation
- la variance de l'erreur d'estimation
- la précision moyenne
- la précision absolue

⁶ Les termes « précision » et « exactitude » sont utilisés indifféremment dans ce rapport, tout comme dans la version française de l'Annexe VIII (objectifs de qualité des données) de la directive 1999/30/CE. Cependant, selon le VIM (Vocabulaire International des termes fondamentaux et généraux de Métrologie, AFNOR NFX 07-001, déc. 1994), le terme adéquat est « exactitude ».

Illustrations :

Préparations des données

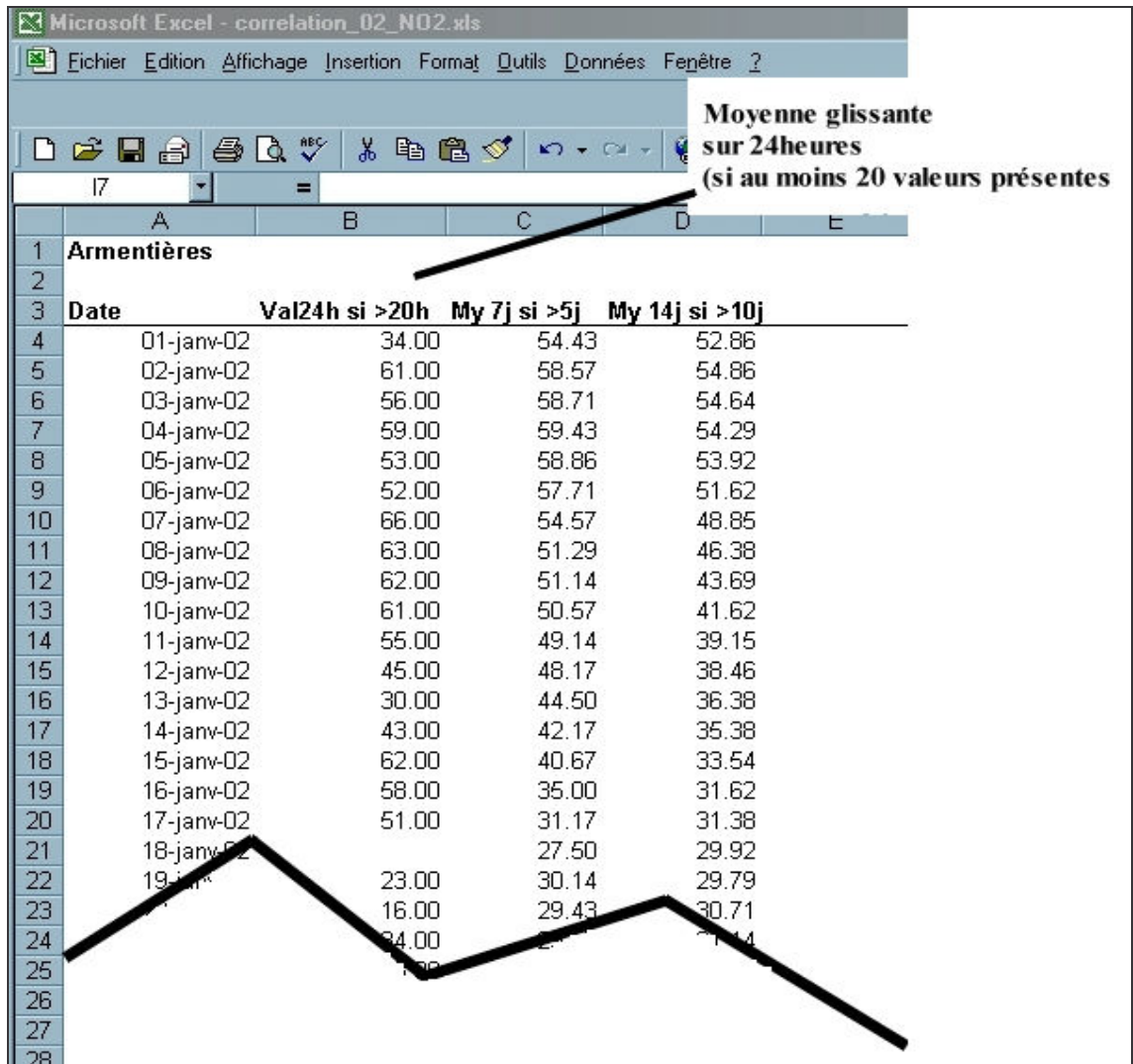


figure n°3.1 : calcul des moyennes glissantes (24h,7j,14j) pour chaque stations

Annexe 3 : Recherche des corrélations entre stations automatiques

Calculs des régressions

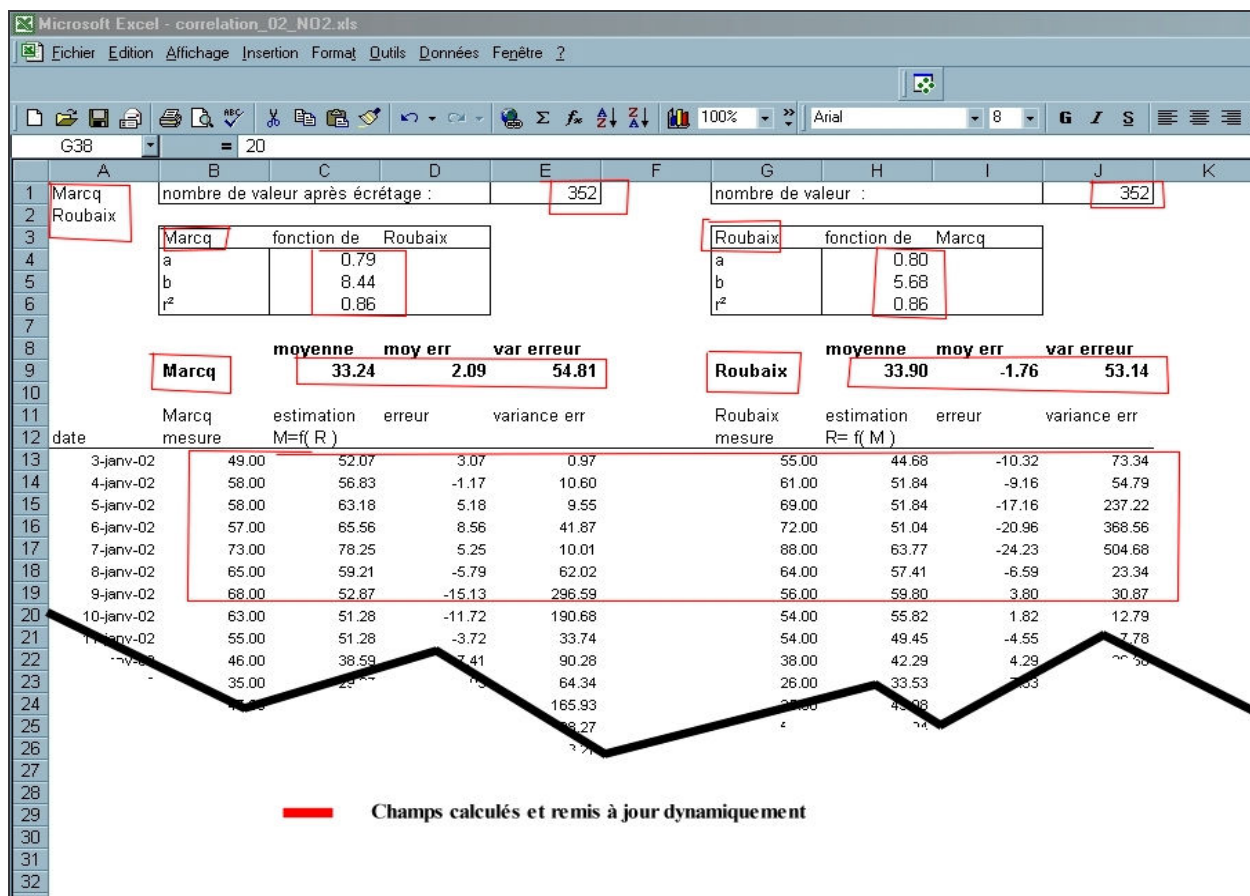


figure n°3.2 : Feuille des calculs pour un couple de stations

Annexe 3 : Recherche des corrélations entre stations automatiques

Présentations des résultats

Tableau récapitulatif pour les moyennes horaires									
Stations sti/stj sti = f(stj)	r ²	nb Valeurs avant/apres	a a sti(x)	b b sti(x)	moy sti moy sti	moy_err val stj-(a.val sti+b)	var_err	recision racine(var_err)/moy sti	moyenne
Marcq / Roubaix	0.86	352.00	0.79	8.44	33.24	2.09	54.81	0.22	
Roubaix / Marcq	0.86	352.00	0.80	5.68	33.90	-1.76	53.14	0.22	
Marcq / Fives	0.76	321.00	0.70	7.06	32.98	-0.62	85.66	0.28	
Fives / Marcq	0.76	317.00	0.93	7.13	35.99	1.82	78.59	0.25	
Marcq / Armentières	0.90	288.00	0.98	6.19	34.89	1.53	40.82	0.18	
Armentières / Marcq	0.90	287.00	0.77	2.47	30.84	-1.57	32.47	0.18	
Marcq / Halluin	0.94	365.00	0.90	11.67	33.32	3.90	30.21	0.16	
Halluin / Marcq	0.94	365.00	0.81	-2.27	28.41	-3.81	16.06	0.14	
Roubaix / Fives	0.86	303.00	0.77	3.32	32.45	-1.68	76.97	0.27	
Fives / Roubaix	0.86	297.00	0.98	6.18	35.57	2.55	83.98	0.26	
Roubaix / Armentières	0.86	279.00	0.96	4.78	35.61	-1.49	49.43	0.20	
Armentières / Roubaix	0.86	277.00	0.75	4.89	30.64	0.81	37.62	0.20	
Roubaix / Halluin	0.86	351.00	0.76	13.65	33.68	1.53	56.18	0.22	
Halluin / Roubaix	0.86	351.00	0.73	1.87	28.33	-1.91	36.49	0.21	
Fives / Armentières	0.86	263.00	1.08	6.49	37.92	1.17	83.56	0.24	
Armentières / Fives	0.86	260.00	0.67	3.71	30.18	-1.08	61.24	0.26	
Fives / Halluin	0.86	325.00	1.01	13.97	35.79	6.47	76.72	0.24	
Halluin / Fives	0.86	318.00	0.67	-1.02	28.03	-5.22	63.98	0.29	
Armentières / Halluin	0.88	289.00	0.81	8.31	30.74	1.35	40.08	0.21	
Halluin / Armentières	0.88	289.00	0.95	-2.11	29.29	-2.26	36.94	0.21	

Coupure des données pour : var(erreur)=500

figure n°3.3 : Tableau des résultats alimenté de manière automatique



941, rue Charles Bourseul - BP 838 - 59508 DOUAI Cedex
Tél. 03 27 71 22 22 - Fax 03 27 71 25 25
mél : mines@ensm-douai.fr - <http://www.ensm-douai.fr>

Imprimé à l'Ecole des Mines de Douai - 59500 DOUAI