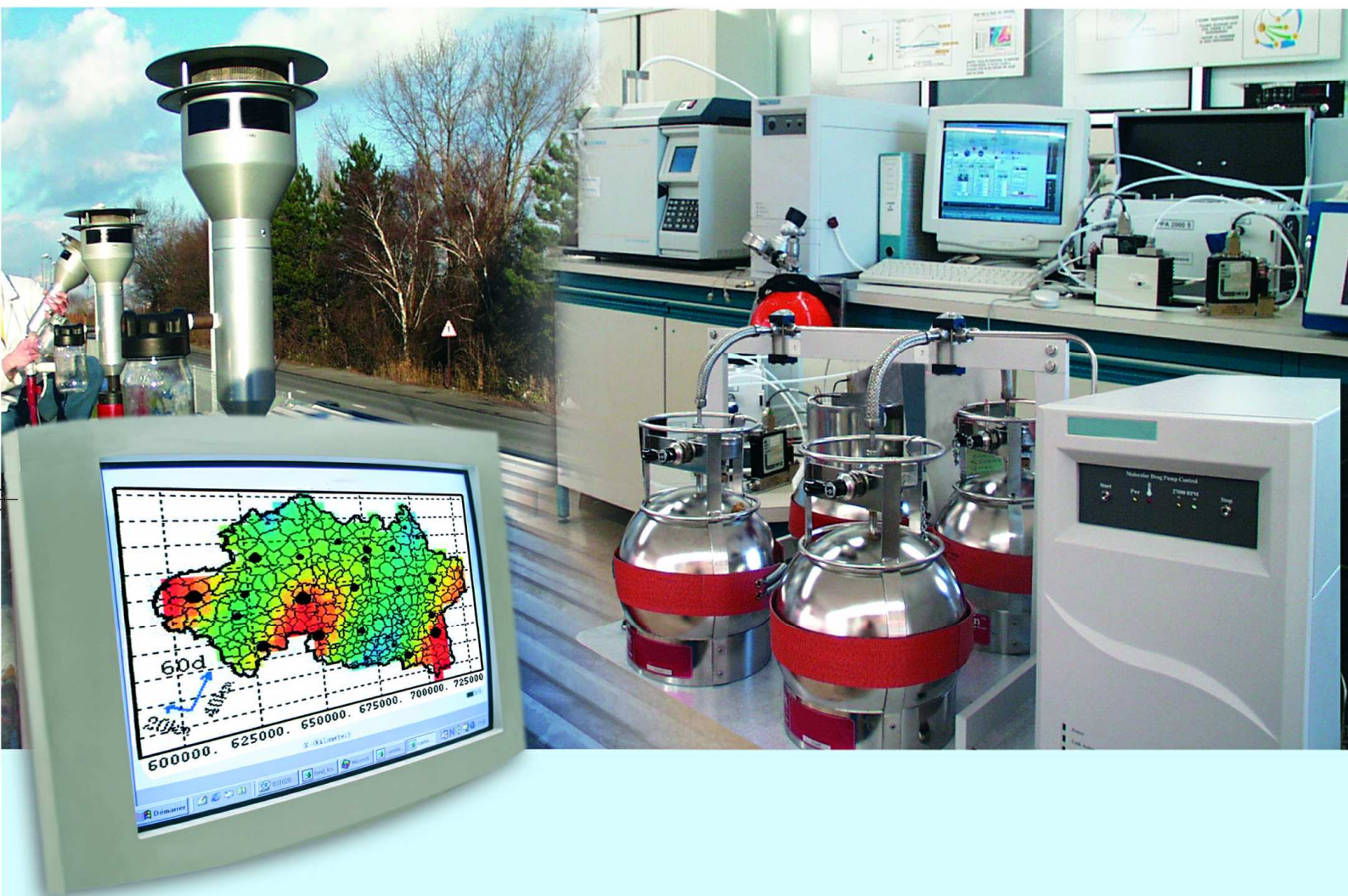




Laboratoire Central de Surveillance de la Qualité de l'Air



Coordination

Réduction optimisée du réseau de mesures fixes d'ozone

Septembre 2012

Programme 2012

Marjorie Délias





PREAMBULE

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

Le Laboratoire Central de Surveillance de la Qualité de l'Air est constitué de laboratoires de l'Ecole des Mines de Douai, de l'INERIS et du LNE. Il mène depuis 1991 des études et des recherches finalisées à la demande du Ministère chargé de l'environnement, et en concertation avec les Associations Agréées de Surveillance de la Qualité de l'Air (AASQA). Ces travaux en matière de pollution atmosphérique ont été financés par la Direction Générale de l'Energie et du Climat (bureau de la qualité de l'air) du Ministère de l'Ecologie, du Développement durable, des Transports et du Logement. Ils sont réalisés avec le souci constant d'améliorer le dispositif de surveillance de la qualité de l'air en France en apportant un appui scientifique et technique au MEDDE et aux AASQA.

L'objectif principal du LCSQA est de participer à l'amélioration de la qualité des mesures effectuées dans l'air ambiant, depuis le prélèvement des échantillons jusqu'au traitement des données issues des mesures. Cette action est menée dans le cadre des réglementations nationales et européennes mais aussi dans un cadre plus prospectif destiné à fournir aux AASQA de nouveaux outils permettant d'anticiper les évolutions futures.



Réduction optimisée du réseau de mesures fixes d'ozone

Laboratoire Central de Surveillance
de la Qualité de l'Air

Coordination

Programme financé par la
Direction Générale de l'Énergie et du Climat (DGEC)

2012

MARJORIE DELIAS

Ce document comporte 28 pages (hors couverture et annexes)

	Rédaction	Vérification	Approbation
NOM	DELIAS Marjorie	BESSAGNET Bertrand	ROUIL Laurence
Qualité	Ingénieur de l'unité Modélisation Atmosphérique et Cartographie Environnementale (MOCA) Direction des Risques Chroniques	Responsable de l'unité Modélisation Atmosphérique et Cartographie Environnementale (MOCA) Direction des Risques Chroniques	Responsable du pôle Modélisation Environnementale et Décision (DECI) Direction des Risques Chroniques
Visa			

TABLE DES MATIÈRES

RESUME	6
1. INTRODUCTION	7
2. METHODOLOGIE	7
2.1 Le réseau à optimiser	8
2.2 La fonction de coût	10
2.3 Les simulations CHIMERE	11
2.4 La procédure de krigeage	11
2.5 L'optimisation par recuit simulé	12
2.6 Réalisation de cartes analysées	15
3. RESULTATS	15
3.1 Analyse géostatistique préliminaire sur le réseau de fond utilisé	15
3.2 Optimisation par recuit simulé	18
3.3 Comparaison de cartes analysées	21
4. COMPARAISON DES RESEAUX OPTIMISES.....	25
5. CONCLUSION ET PERSPECTIVES.....	26
6. REFERENCES	28
7. LISTE DES ANNEXES	29

RESUME

Cette étude s'inscrit dans le cadre des travaux de coordination du LCSQA relatifs au dispositif métrologique français. Elle a pour objet l'optimisation du réseau de mesures fixes français grâce aux outils de géostatistique et à l'optimisation statistique.

Ce rapport présente les résultats obtenus dans le cadre de l'étude sur l'optimisation du réseau de mesures de fond d'ozone. L'objectif de celle-ci a été de rechercher un réseau optimisé, à partir du dispositif actuel, opérationnel au 1^{er} mai 2012, et de taille correspondant au nombre minimal de points de mesure fixes d'ozone, imposé par la Directive 2008-50-CE concernant « la qualité de l'air ambiant et un air pur pour l'Europe » et calculé d'après le zonage défini au 1^{er} janvier 2010.

Cet exercice d'optimisation, tel qu'il est réalisé ici, consiste en la minimisation d'une fonction de coût, définie selon les besoins de l'exercice, et en la recherche de la valeur pour laquelle le minimum est atteint. La première partie de ce rapport détaille et justifie la méthodologie employée, du choix de la fonction de coût à l'algorithme statistique utilisé pour la minimiser.

La deuxième partie présente les résultats obtenus selon les paramètres choisis dans le processus d'optimisation et compare les cartes analysées des maxima journaliers estivaux d'ozone des moyennes glissantes sur 8h, basées d'une part sur le réseau opérationnel au 1^{er} mai 2012 et d'autre part sur les réseaux optimisés obtenus.

Enfin, la dernière partie de ce rapport compare les réseaux optimisés entre eux. Cette étude fait ainsi apparaître un groupe de stations nécessaires à l'élaboration de cartographies d'ozone et un deuxième groupe de stations dont les données d'ozone apportent peu à la qualité de ces cartographies.

Ces résultats constituent le début du travail sur l'optimisation du réseau français et démontrent l'efficacité de la méthodologie développée. Elle sera appliquée prochainement aux réseaux NO₂ et PM₁₀ et pourrait être adaptée à l'échelle régionale comme outil de recherche d'implantation de stations.

1. INTRODUCTION

Le dispositif français de mesures fixes s'est construit progressivement, sans plan d'ensemble ni stratégie globale, la surveillance étant confiée à des associations agréées locales qui doivent concilier les contraintes de leur territoire (politiques, sociales et financières) avec les engagements nationaux. Aujourd'hui le coût financier du développement et de la maintenance du réseau et la montée en puissance de la modélisation justifient une certaine rationalisation du réseau.

La surveillance de la qualité de l'air nécessite un réseau de mesures dont les objectifs peuvent être multiples :

- Contrôle du respect des normes en concentrations en vigueur dans le pays
- Activation des mesures d'urgence lors d'épisodes de pollution
- Vérification des modèles numériques et initialisation des simulations pour la réalisation de prévision à court terme
- Analyse des tendances sur le moyen/long terme
- Élaboration des cartographies de la qualité de l'air
- Analyse des responsabilités des sources de pollution...

Dans cette étude, le choix a été fait de mener une optimisation du réseau avec un objectif cartographique afin d'identifier les points de mesures qui pourraient être soustraits au réseau actuel sans porter préjudice à la qualité des cartes générées à partir d'une procédure de krigeage. Ce rapport présente les résultats du travail de réduction optimisée du réseau de mesures de l'ozone exclusivement.

2. METHODOLOGIE

Le problème d'optimisation d'un réseau consiste à déterminer le minimum d'une fonction de coût et la valeur pour laquelle il est atteint.

Il convient tout d'abord de définir quelques notions. Notons :

- $\mathcal{D}$ le domaine géographique sur lequel porte l'étude
- $\mathcal{A}$ l'ensemble formé par le réseau actuel (instrumenté). La taille de cet ensemble est notée n .
- $\mathcal{G}$ un sous-réseau testé (inclus dans $\mathcal{A}$). La taille de ce sous-ensemble est noté p .

Le problème d'optimisation se traduit alors par la formulation suivante :

$$\omega^* = \underset{\omega}{\operatorname{argmin}} \varphi(\omega) \quad [1]$$

Où :

- ω est une des configurations de réseau possibles
- φ désigne la fonction de coût (ou fonction objectif).

ω représente un vecteur à n composantes. La composante i de ce vecteur est 1 ou 0 suivant que le site s_i est sélectionné ou non dans le sous-réseau instrumenté G . La somme des composantes vaut par conséquent p .

Les choix méthodologiques développés ci-dessous sont résumés sur la Figure 1.

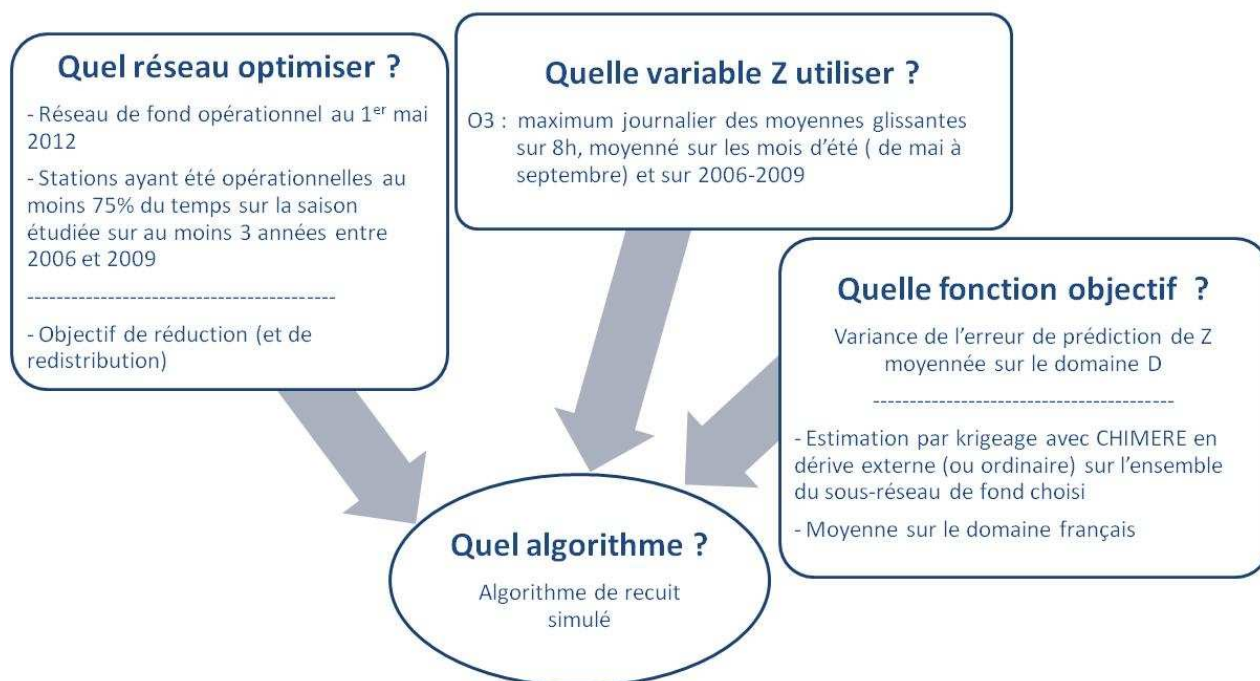


Figure 1 : Choix méthodologiques effectués dans cette étude

2.1 LE RESEAU A OPTIMISER

Les sites du réseau à optimiser sont des points de mesure O₃ de la Base de Données de la Qualité de l'Air (BDQA). Seules les stations de fond urbain, périurbain et rural sont considérées dans cette étude.

Le réseau optimisé devrait prendre en compte les exigences européennes en matière de surveillance minimale au sein même d'une zone administrative de surveillance (ZAS). Cependant, compte tenu de la complexité de la détermination du nombre minimal de points de mesures fixes au sein d'une ZAS, il serait sans doute difficile d'inclure dans la recherche d'optimisation des contraintes numériques et typologiques par zone de surveillance administrative. Cette contrainte supplémentaire imposée au réseau optimisé devra donc être vérifiée a posteriori par « validation d'expert ».

Afin d'appréhender l'effet maximal d'une réduction des points de mesure d'ozone sur la qualité des cartographies obtenues par analyse, la taille du réseau optimisé est par ailleurs déterminée en fonction des exigences européennes basées sur le zonage français 2010 : au minimum 153 points fixes d'ozone doivent couvrir la France hors DOM-TOM. La taille des sous-réseaux G testés sera donc $p=153$.

Ainsi le réseau choisi est le réseau de fond, en fonctionnement au 1^{er} mai 2012, pour lequel sont disponibles au moins 75% de données sur la période considérée (période qui sera décrite dans le paragraphe suivant) pendant au moins 3 années sur 4 entre 2006 et 2009. Ce réseau se répartit comme indiqué dans le Tableau 1 :

Tableau 1 : Taille du réseau français mesurant l'ozone pris en compte dans l'étude selon la typologie des stations et leur taux de fonctionnement sur les étés 2006-2009

Taille du réseau O3 de fond rural pris en compte	49
Taille du réseau O3 de fond urbain+périurbain pris en compte	266
Taille du réseau O3 total de fond pris en compte	315

La carte suivante présente la répartition spatiale des stations utilisées dans cette étude :

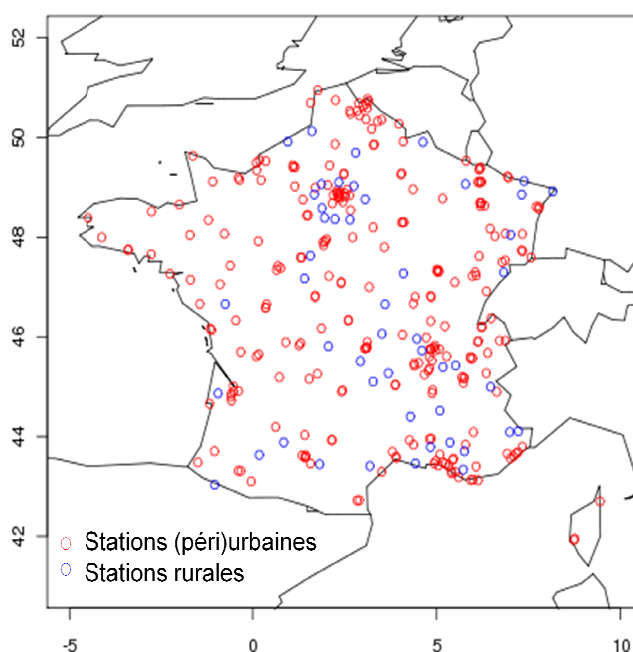


Figure 2: Répartition des stations françaises utilisées dans cette étude

2.2 LA FONCTION DE COUT

En 2007, Brus et Heuvelink (1) ont proposé une méthode d'échantillonnage optimal pour le krigeage universel de variables environnementales, méthode consistant à générer un plan d'échantillonnage optimal minimisant la variance de prédiction du krigeage universel. Cette méthode serait particulièrement bien adaptée aux objectifs de cartographie des concentrations des polluants atmosphériques par krigeage. Sur cette base, l'INERIS, en collaboration avec l'École des Mines de Paris, a publié des *Recommandations sur le schéma d'échantillonnage pour les campagnes de mesures de la qualité de l'air* (2) en 2010, résultats repris et complétés par Romary *et al.* (3)

Aussi, dans cette étude, le choix s'est porté sur la même fonction de coût, à savoir la moyenne sur le domaine étudié de la variance de prédiction du krigeage :

$$\varphi(\omega) = \frac{1}{|D|} \int_D \text{Var}[\hat{Z}(x) - Z(x)] dx \quad [2]$$

Où :

- Z est la variable aléatoire pour le polluant étudié (décrite plus loin)
- $\hat{Z}$ représente l'estimation par krigeage de la variable Z
- D est le domaine étudié de taille $|D|$.

Puisque la variance de l'erreur d'estimation ne dépend pas des données d'observations de la variable étudiée, cette intégrale est calculée sur un maillage régulier de D .

Le krigeage choisi dans (1; 2) est un krigeage avec les sorties du modèle CHIMERE en dérive externe (les simulations sont décrites au paragraphe suivant). Dans cette étude, un krigeage ordinaire est également testé.

L'introduction d'une pondération afin d'orienter l'optimisation du réseau est testée dans Romary *et al.* (3). Dans cette étude, un test visant à orienter l'optimisation vers un réseau plus dense dans les zones où les performances du modèle sont moindres est présenté en Annexe 4.

Le choix de la variable Z

Un krigeage horaire des concentrations caractérise mieux le cycle diurne de l'ozone. Pour cette raison, la procédure d'analyse des cartes dans PREV'AIR utilise un krigeage horaire de l'ozone et des PM_{10} . Toutefois, la recherche du minimum de la fonction de coût nécessite le test d'un grand nombre de sous-réseaux. Réaliser un krigeage horaire sur un jour-type pour chaque sous-réseau testé serait alors très lourd en termes de temps de calcul.

Aussi pour l'ozone, l'étude est réalisée en considérant le maximum journalier des moyennes glissantes sur 8h (critère utilisé dans la définition de la valeur cible pour l'ozone dans la directive européenne 2008/50/CE). Les pics d'ozone apparaissant principalement en été, Z représente, pour l'ozone, la moyenne de cette valeur sur les mois de mai à septembre 2006-2009.

2.3 LES SIMULATIONS CHIMERE

Dans cette étude, le modèle de chimie-transport CHIMERE (4) est utilisé avec les configurations suivantes :

- données météorologiques : simulations WRF¹ version 3.2.1 (2-way nesting, résolutions de 45 x 45 km et 15 x 15 km, 32 niveaux verticaux, pas de temps de 3h) ;
- couverture terrestre : carte globale d'occupation des surfaces GLOBCOVER² (résolution 300m)
- émissions : inventaire EMEP, spatialisé par l'INERIS

La simulation a été réalisée sur deux domaines imbriqués qui couvrent :

- pour le premier, l'Europe ([15°W, 35°N]x[35°E,70°N]) avec une grille 71x101 à une résolution spatiale de 50 km
- pour le deuxième, la France (métropole, [5°W,41°N] x[10°E,52°N]) avec une grille 101x111 à une résolution spatiale de 10 km : (domaine FRA10 utilisé dans PREV'AIR)

9 niveaux verticaux répartis de la surface à 500hPa sont utilisés.

Dans cette note, la simulation couvre la période du 1^{er} janvier au 31 décembre 2009. Les conditions aux limites sont les climatologies mensuelles GOCART³ pour les particules et LMDzINCA⁴ pour les gaz.

2.4 LA PROCEDURE DE KRIGEAGE

Comme indiqué précédemment, les types de krigeage réalisés dans cette étude sur la variable Z sont un krigeage avec CHIMERE en dérive externe et un krigeage ordinaire.

¹ Configuration WRF choisie : modèle de surface Noah ; modèle de couche limite YSU ; schéma convectif Kain-Fritsch ; schéma radiatif RRTMG/RRTMG ; schéma de microphysique MP4 ; forçage grande échelle ERA-INTERIM ;

² Projet de l'ESA : produire une carte globale d'occupation des surfaces pour l'année 2005-2006, en utilisant les données à 300 m de résolution du capteur MERIS embarqué à bord de la plateforme satellitaire ENVISAT.

³ Goddard Chemistry Aerosol Radiation and Transport : modele global développé par Chin et al. (*Atmospheric sulfur cycle simulated in the global model GOCART: model description and global properties* – J.Geophys. Res. – 2000)

⁴ Modèle chimie-aérosols INCA (développé par le Laboratoire des Sciences du Climat et de l'Environnement) couplé au modèle de circulation général, LMDz, du Laboratoire de Météorologie Dynamique, tous deux laboratoires de l'Institut Pierre-Simon Laplace.

➤ **Krigeage avec CHIMERE en dérive externe**

$$Z(x) = Y'(x)\beta + S(x) \quad [3]$$

où x est la coordonnée spatiale, Y le vecteur colonne des valeurs modélisées, β un vecteur de paramètres et $S(x)$ le champ aléatoire appelé résidu.

La variance de l'erreur de krigeage est calculée à l'aide de la matrice de covariance C (obtenue grâce au modèle variographique) selon le développement suivant :

$$\begin{aligned} \text{Var}(Z(x_0) - \widehat{Z}(x_0)) \\ = C(0) - c_0' C^{-1} c_0 + (Y_0 - Y' C^{-1} c_0)' (Y' C^{-1} Y)^{-1} (Y_0 - Y' C^{-1} c_0) \end{aligned} \quad [4]$$

où c_0 est le vecteur des covariances des résidus entre les points de prédiction et les observations.⁵

Le krigeage est réalisé sur la grille CHIMERE 10 km décrite au paragraphe précédent, le pas de calcul du variogramme expérimental est de 25 km, la distance maximale de 300 km. La fonction de coût est obtenue en moyennant la variance sur les centres de maille à l'intérieur de la France Métropolitaine.

➤ **Krigeage ordinaire**

Dans le cas du krigeage ordinaire, les étapes de calcul sont les mêmes, seules changent les équations [3] et [4] qui deviennent :

$$Z(x) = S(x) \quad [5]$$

$$\text{Var}(Z(x_0) - \widehat{Z}(x_0)) = C(0) - c_0' C^{-1} c_0 \quad [6]$$

2.5 L'OPTIMISATION PAR RECUIT SIMULE

La fonction de coût choisie peut avoir plusieurs minima. Tester toutes les configurations de réseau possibles aurait été extrêmement coûteux en temps de calcul (réduire le réseau de 315 points de mesure à 153 aurait nécessité le test de 2.10^{93} configurations). L'algorithme le plus fréquemment utilisé pour résoudre ce problème d'optimisation est alors une méthode stochastique : l'algorithme de recuit simulé (5), moins coûteux en temps de calcul et permettant d'échapper aux minima locaux.

⁵ L'ensemble des calculs statistiques sont réalisés sous R et en particulier, en ce qui concerne la recherche du modèle variographique lors de l'étude préliminaire, avec la librairie « *gstat* ».

Le principe de l'algorithme de recuit simulé est basé sur un processus de recuit utilisé en métallurgie (alternance de cycles de refroidissement lent et de réchauffage - recuit) qui ont pour effet de minimiser l'énergie du matériau) et est résumé sur la Figure 3 :

La solution initiale ω_0 peut être prise au hasard dans l'espace des solutions possibles. À cette solution correspond une énergie initiale E_0 , valeur de la fonction de coût φ en ω_0 . À chaque itération de l'algorithme une modification élémentaire de la solution est effectuée. Cette modification entraîne une variation ΔE de l'énergie du système (toujours calculée à partir du critère que l'on cherche à optimiser). Si cette variation est négative (c'est-à-dire qu'elle fait baisser l'énergie du système), elle est appliquée à la solution courante. Sinon, elle est acceptée avec une certaine probabilité. L'avantage d'un tel algorithme est d'éviter en théorie de rester bloqué dans le bassin d'attraction d'un minimum local.

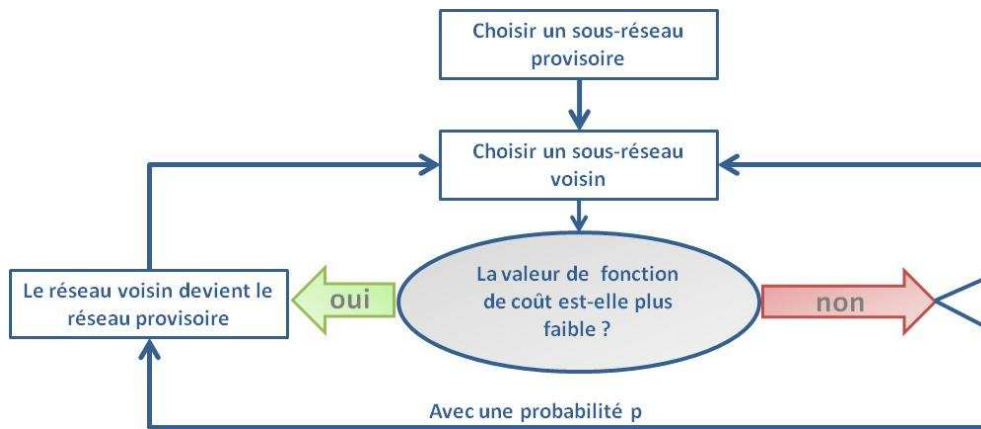


Figure 3 : Principe de fonctionnement de l'algorithme de recuit simulé

La probabilité évoquée ci-dessus est classiquement (1) celle issue de la statistique de Boltzmann :

$$P(\omega^j, \omega^{j+1}, \tau) = \min(1, e^{-\frac{\varphi(\omega^{j+1}) - \varphi(\omega^j)}{\tau}}) \quad [7]$$

avec τ la température de recuit. τ est initialement grand et le processus itératif engendre de grandes variations de φ . Lorsque τ diminue selon une certaine loi de décroissance, les itérations cherchent de plus fines variations. A chaque température, le système doit rester suffisamment de temps pour que le système atteigne un état stable. Ainsi, l'utilisation d'un tel algorithme nécessite le choix de plusieurs paramètres :

➤ **La température initiale**

La température initiale généralement choisie (1; 2) l'est de telle façon que la probabilité moyenne d'accepter une combinaison empirant la fonction coût est de 0,8. Celle-ci peut être calculée en lançant une première fois l'algorithme dans lequel toutes les augmentations de la fonction de coût sont acceptées. A partir des résultats, la moyenne des augmentations de la fonction de coût $\overline{\Delta\varphi^+}$ est calculée et la température initiale est obtenue par la formule :

$$\tau_0 = \frac{\overline{\Delta\varphi^+}}{\ln(0,8)} \quad [8]$$

➤ **La loi de décroissance de la température de recuit**

La loi de décroissance de la température de recuit est le plus souvent géométrique :

$$\tau_{k+1} = \alpha\tau_k \quad [9]$$

où α est le facteur de décroissance compris entre 0 et 1 et généralement proche de 1 ($\alpha=0,99$ pour Wu *et al.* (6) $\alpha=0,9999$ pour Wu *et al.* (7), $\alpha=0,95$ pour Brus *et al.* (1)).

Dans cette étude, d'après les recommandations de Pibouleau *et al.* (8), le facteur de décroissance a été pris égal à 0,9 dans une loi de décroissance qui s'applique par paliers d'acceptations : la température de recuit décroît seulement après un nombre d'acceptations correspondant à la taille du vecteur ω .

Le terme « acceptation » désigne l'action, au sein de l'algorithme, d'autoriser l'utilisation du réseau voisin comme nouveau réseau de référence, parce qu'il engendre une valeur de la fonction de coût inférieure à celle engendrée par le réseau précédent, mais aussi, avec la probabilité de Boltzmann, même s'il est moins bon que le réseau précédent.

➤ **Le mécanisme de génération du réseau voisin**

Le mécanisme de génération du réseau voisin choisi lors de la réduction du réseau de mesures de fond (urbain, périurbain et rural sans distinction) est le suivant (Figure 4) :

A la $n+1^{\text{ème}}$ itération, on choisit aléatoirement 50 points parmi ceux non instrumentés à la $n^{\text{ème}}$ itération. Chaque point s_i choisi est instrumenté à la place d'un point s_j instrumenté à la $n^{\text{ème}}$ itération et sélectionné aléatoirement dans le voisinage de s_i . Le voisinage est défini par un rayon de voisinage R_{vois} initialement fixé à 300km et qui décroît au cours du processus itératif. Si aucun point instrumenté n'est trouvé dans ce rayon de voisinage, le point s_i est exclu des points potentiels. Le vecteur configuration ω_n est alors remplacé par le vecteur ω_{n+1} où la $i^{\text{ème}}$ composante passe de 1 à 0 et la $j^{\text{ème}}$ de 0 à 1.

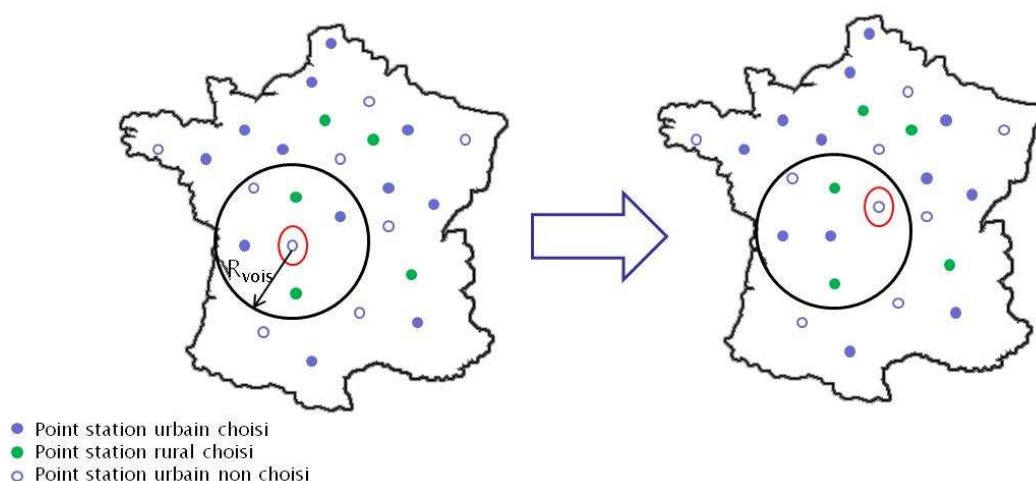


Figure 4 : Génération d'un sous-réseau voisin dans le cas d'une réduction du réseau de fond

➤ **Le critère d'arrêt de l'algorithme**

Le critère d'arrêt de l'algorithme peut être un nombre d'itérations fixé, une température finale fixée, ou encore un critère de non-progression (la qualité de la mesure n'augmente plus pendant un certain nombre d'itérations).

2.6 REALISATION DE CARTES ANALYSEES

Afin de visualiser l'impact d'une diminution du nombre de stations sur la qualité des cartes analysées, un krigeage, avec CHIMERE en dérive externe ou ordinaire, a été réalisé sur les maxima journaliers d'ozone des moyennes glissantes sur 8h sur le réseau initial et sur les réseaux obtenus après optimisation. Pour cette étape et afin de disposer de quelques stations pour l'étape de validation absolue, un groupe de 20 points n'appartenant pas au réseau optimisé est retiré du set initial (315 stations). Après l'étape de calcul du variogramme et avant l'étape de krigeage, sont rajoutées les stations AIRBASE, hors France mais incluses dans le domaine FRA10, de typologie de fond. Le voisinage de krigeage est pris égal à 80 stations.

3. RESULTATS

3.1 ANALYSE GEOSTATISTIQUE PRELIMINAIRE SUR LE RESEAU DE FOND UTILISE

Afin de calculer la matrice de covariance, une analyse géostatistique est pratiquée sur le réseau de fond d'ozone total (315 points de mesure) auquel 20 points sont retirés pour la validation. Dans le cas du krigeage avec CHIMERE en dérive externe, un modèle, composé d'une structure pépitique (pépité=13,3) additionnée d'une structure exponentielle (portée 32km ; palier 8,9), a été ajusté

aux données expérimentales. Dans le cas du krigeage ordinaire, les structures modélisées sont les suivantes : pépite(12,7) + exponentiel (portée 500km ; palier 68,3).

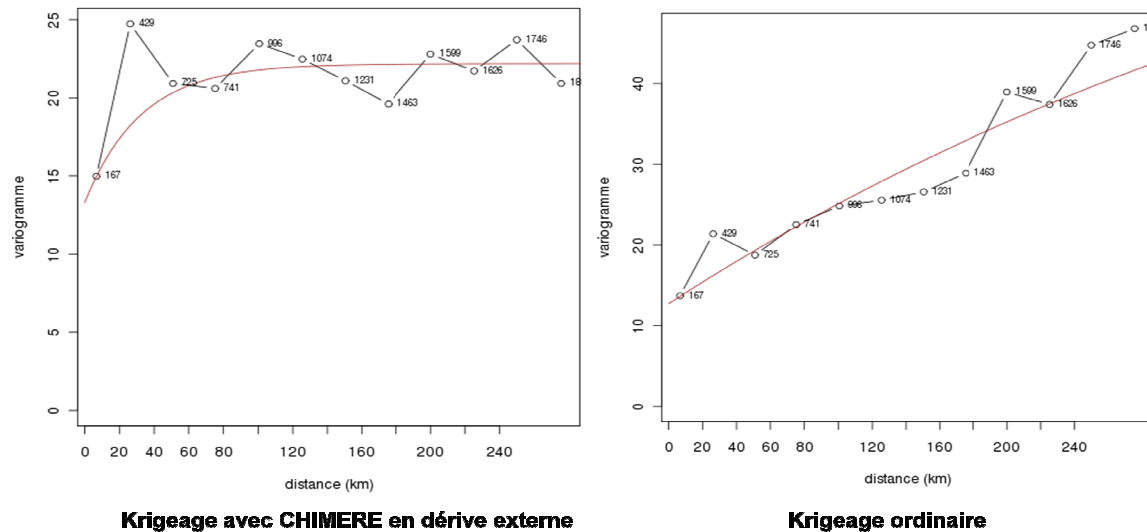


Figure 5 : Variogrammes expérimentaux et modélisés

➤ **Test de validation croisée, en voisinage unique**

Afin d'apprécier la qualité du modèle dans les deux cas, un test de validation croisée, en voisinage unique, est réalisé et les valeurs estimées sont comparées aux valeurs mesurées, notamment en calculant l'erreur d'estimation (différence entre valeur estimée et valeur vraie). Comme le montre la Figure 6 suivante ainsi que les Tableau 3 et Tableau 4 en Annexe 1, les valeurs estimées sont bien corrélées aux valeurs obtenues par les mesures, avec un coefficient de corrélation supérieur à 0,86.

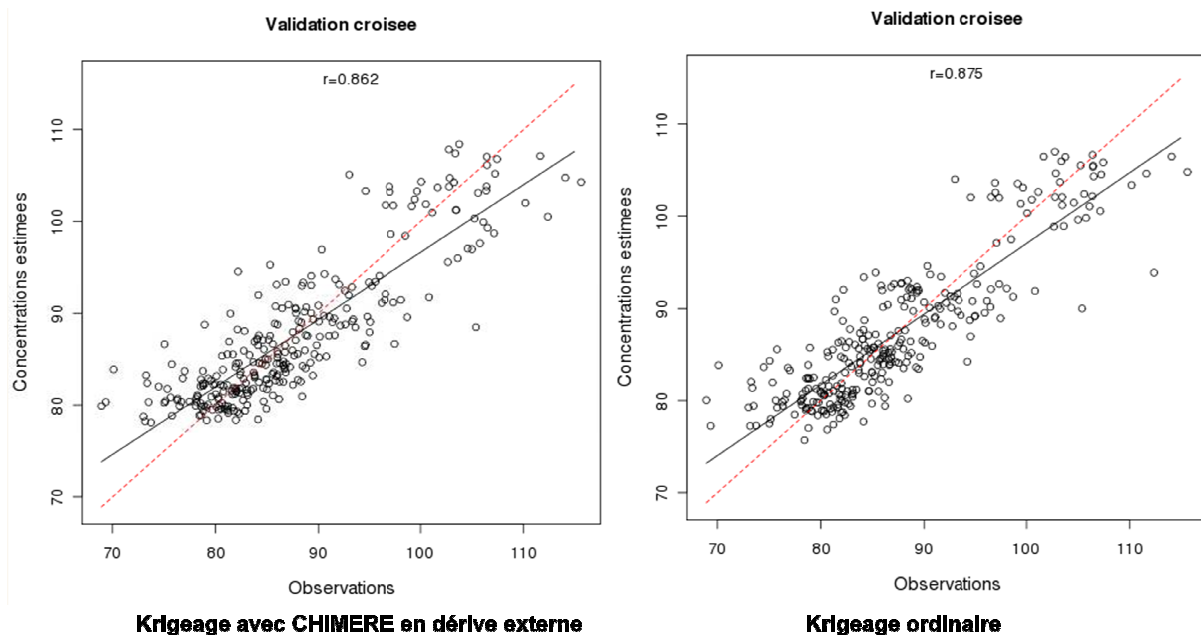


Figure 6 : Test de validation croisée : valeurs estimées en fonction des valeurs vraies

➤ **Test de validation absolue mené sur les données complémentaires**

Des estimations ont été réalisées aux 20 points laissés de côté lors de l'étude variographique et sont ensuite comparées aux valeurs vraies. Comme le montre la Figure 7, ainsi que le Tableau 5 en Annexe 1, les résultats du test de validation absolue (avec des coefficients de corrélation supérieurs à 0.865) confirment, au même titre que ceux de la validation croisée, que les modèles variographiques établis pour chacun des krigeages sont satisfaisants.

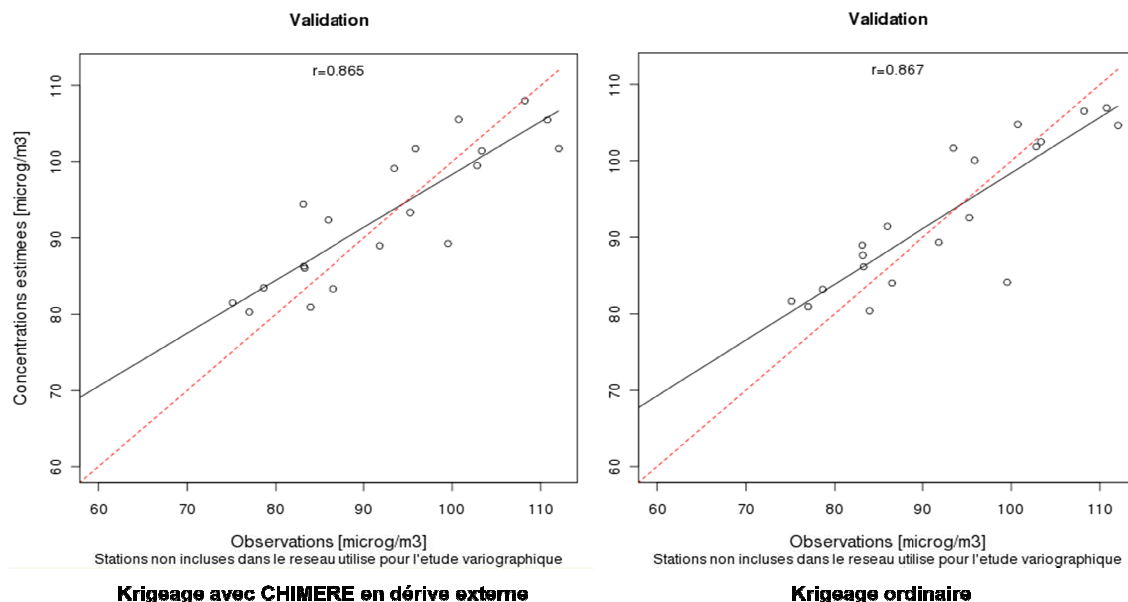


Figure 7 : Test de validation absolue : valeurs estimées en fonction des valeurs vraies

Ces tests permettent donc de confirmer le choix des structures et des paramètres de variogramme qui sont utilisés dans le calcul de la fonction de covariance dont l'équation, calculée à partir de celle du variogramme modélisé γ , est rappelée ci-dessous :

$$\left\{ \begin{array}{l} \gamma(h) = C(0) - C(h) \\ \lim_{h \rightarrow \infty} C(h) = 0 \\ \gamma(h) = C + \sigma(1 - \exp(-\frac{h}{a})) \end{array} \right. \Rightarrow C(h) = \begin{cases} C + \sigma \text{ pour } h = 0 \\ C + \sigma \cdot \exp(-\frac{h}{a}) \text{ pour } h > 0 \end{cases} \quad [10]$$

avec h la distance entre deux points et les valeurs suivantes :

	Krigeage avec CHIMERE en dérive externe	Krigeage ordinaire
Pépite = C	13,3	12,7
Portée = a	32 km	500 km
Deltapalier = σ	8,9	68,3

3.2 OPTIMISATION PAR RECUIT SIMULE

L'optimisation du réseau de fond a été réalisée en utilisant les paramètres du variogramme identifiés ci-dessus pour les deux types de krigeage. Les Figure 8 et Figure 9 présentent les résultats de l'optimisation en utilisant un krigeage en dérive externe et un krigeage ordinaire respectivement. En haut de chaque figure, sont d'abord représentés, de gauche à droite, l'évolution de la valeur de la fonction de coût au cours du processus itératif (φ), la moyenne empirique du taux d'acceptation de transition (ma) et un graphique comparant les valeurs de la fonction de coût avant et après optimisation ainsi que les valeurs prises par cette fonction dans le cas d'une réduction aléatoire de la taille du réseau. En bas, est représenté le réseau optimisé : pour des raisons de facilité de lecture, les réseaux urbains et périurbains sont présentés à gauche et le réseau rural à droite.

Les graphiques correspondant à l'évolution de la fonction de coût et de la moyenne empirique montrent que l'algorithme s'est comporté de manière satisfaisante : les premières transitions sont acceptées avec un taux proche de 1, ce qui permet une bonne exploration de l'espace des solutions. Le taux décroît ensuite en même temps que la température. Parallèlement, la valeur de la fonction de coût décroît de manière satisfaisante puis se stabilise, ce qui indique la convergence de l'algorithme.

En haut à droite, les graphiques démontrent l'intérêt de la démarche d'optimisation à l'aide de l'algorithme de recuit simulé développé dans cette étude puisque les valeurs prises par la fonction de coût dans le cas d'une réduction aléatoire (diagramme « à moustaches » avec 500 itérations) sont plus élevées que celle prise avec le réseau optimisé.

Enfin des simulations répétées avec le même jeu de paramètres d'entrée et une « graine » (nombre utilisé dans le générateur pseudo-aléatoire du programme et intervenant dans la génération du voisin) différente, donne des résultats similaires, ce qui prouve la stabilité des résultats.

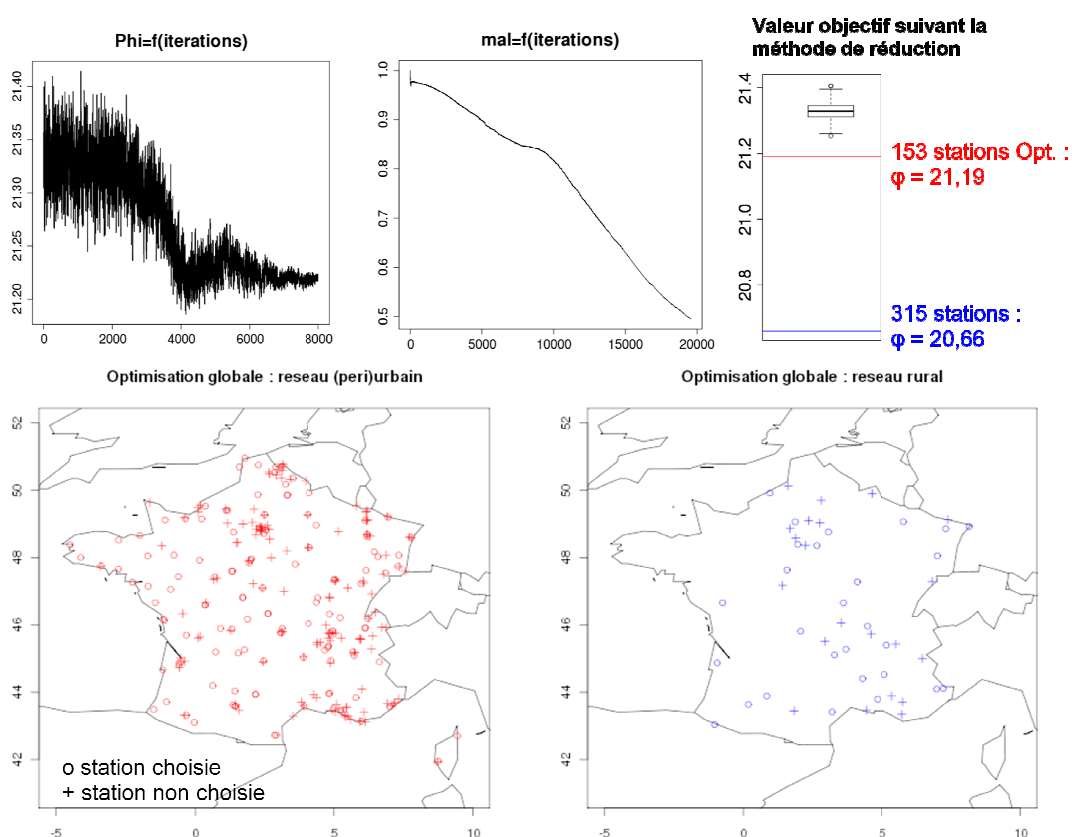


Figure 8 : Krigeage avec CHIMERE en dérive externe : Résultats de la procédure d'optimisation du réseau de fond

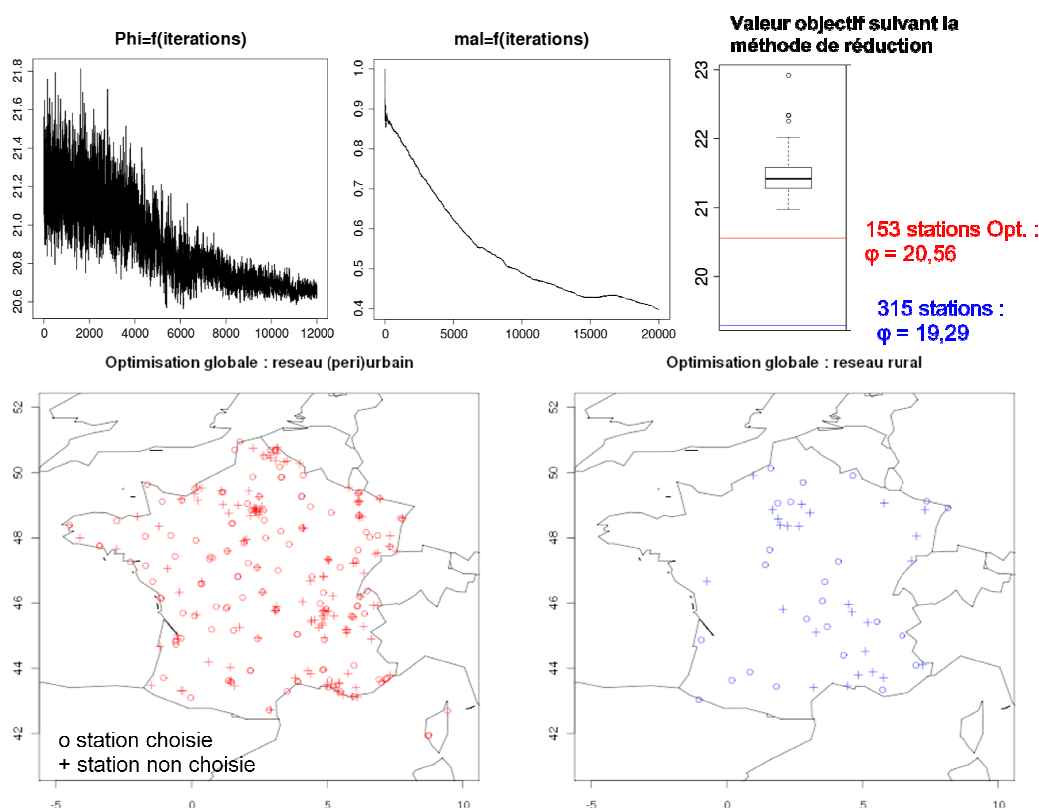


Figure 9 : Krigeage ordinaire : Résultats de la procédure d'optimisation du réseau de fond

L'examen des cartes de répartition des stations dans le réseau optimisé montre que dans les deux cas, les zones actuellement denses en stations sont les plus fortement touchées par le retrait de points de mesure, comme c'est le cas pour les grosses agglomérations telles que Paris, Marseille et Lyon. Ces résultats avaient par ailleurs été observés par Wu *et al.* (6) qui utilisaient une fonction de coût différente, à savoir la moyenne de l'erreur quadratique moyenne de prédiction aux points non évalués.

Cependant la proportion de points ruraux dans le réseau optimisé se maintient par rapport à la proportion dans le réseau initial, comme indiqué dans le Tableau 2.

Tableau 2: Nombre de stations des différentes typologies avant et après optimisation par réduction

	Réseau initial	Réseau optimisé – Krigeage avec CHIMERE en dérive externe	Réseau optimisé – Krigeage ordinaire
Stations (péri)urbaines	266	122	129
Stations rurales	49	31	24
Pourcentage de stations rurales dans le réseau total	15,6%	20%	16%

3.3 COMPARAISON DE CARTES ANALYSEES

Afin de visualiser l'impact d'une diminution du nombre de stations sur la qualité des cartographies, des cartes analysées des maxima journaliers d'ozone des moyennes glissantes sur 8h, moyennés sur la période estivale sont réalisées avant et après optimisation.

➤ *Krigeage avec CHIMERE en dérive externe*

La Figure 10 présente les variogrammes expérimentaux et modélisés avant et après optimisation. Les 2 modèles utilisés ont une structure pépitique, additionnée d'une structure exponentielle dont les paramètres sont indiqués sous les graphiques. Comme précédemment, une validation absolue est réalisée sur 20 points, non pris en compte dans l'analyse variographique.

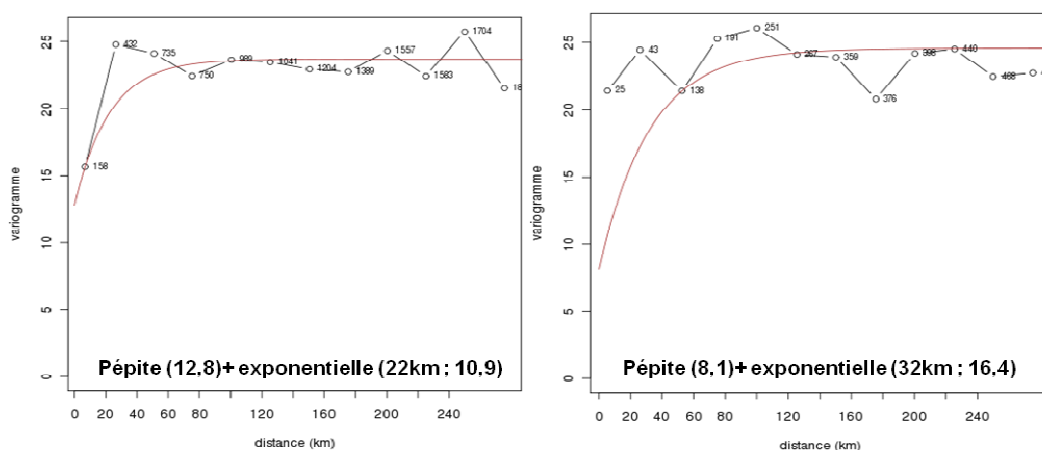


Figure 10 : Krigeage avec CHIMERE en dérive externe : variogrammes expérimentaux et modélisés avant et après optimisation et paramètres des variogrammes modélisés

Les estimations des valeurs des maxima journaliers des moyennes glissantes sur 8h ont ensuite été réalisées par krigeage avec CHIMERE en dérive externe sur la base de ces modèles et du réseau avant et après optimisation. La Figure 11 suivante présente les cartes analysées à partir du réseau initial (en haut à gauche) et à partir du réseau optimisé (en haut à droite). En bas, la cartographie souligne les différences entre les deux cartes analysées.

On remarque que les structures des maxima journaliers des moyennes glissantes sur 8h d'ozone sont globalement conservées en cartographie lorsque l'on utilise seulement 153 stations françaises au lieu de 315. Les différences observées sur le territoire français sont globalement faibles, en moyenne $-0,005 \mu\text{g}/\text{m}^3$ et au maximum en valeur absolue de $6,8 \mu\text{g}/\text{m}^3$ qui correspond à la région Rhône-Alpes. Ces différences dans le Sud-Est correspondent également à des zones de fortes concentrations estivales d'ozone. Une étude plus locale pourrait permettre de comprendre et réduire ces différences.

Dans le cadre du test de validation absolue, les estimations aux points laissés de côté lors de l'analyse variographique sont calculées par krigeage et comparées aux valeurs vraies : la comparaison entre estimations et valeurs vraies est présentée en Figure 12. On remarque que les valeurs sont globalement mieux estimées après optimisation (en bleu) qu'avant optimisation (en noir). Le coefficient de corrélation est d'ailleurs meilleur ($r=0,872$ contre $0,866$) et les erreurs d'estimations sont plus faibles avec une moyenne de $1,1 \mu\text{g}/\text{m}^3$ (contre $1,4 \mu\text{g}/\text{m}^3$) et une médiane de $-0,2 \mu\text{g}/\text{m}^3$ (contre $0,6 \mu\text{g}/\text{m}^3$). (Les statistiques détaillées sont présentées en Annexe 2 Tableau 6).

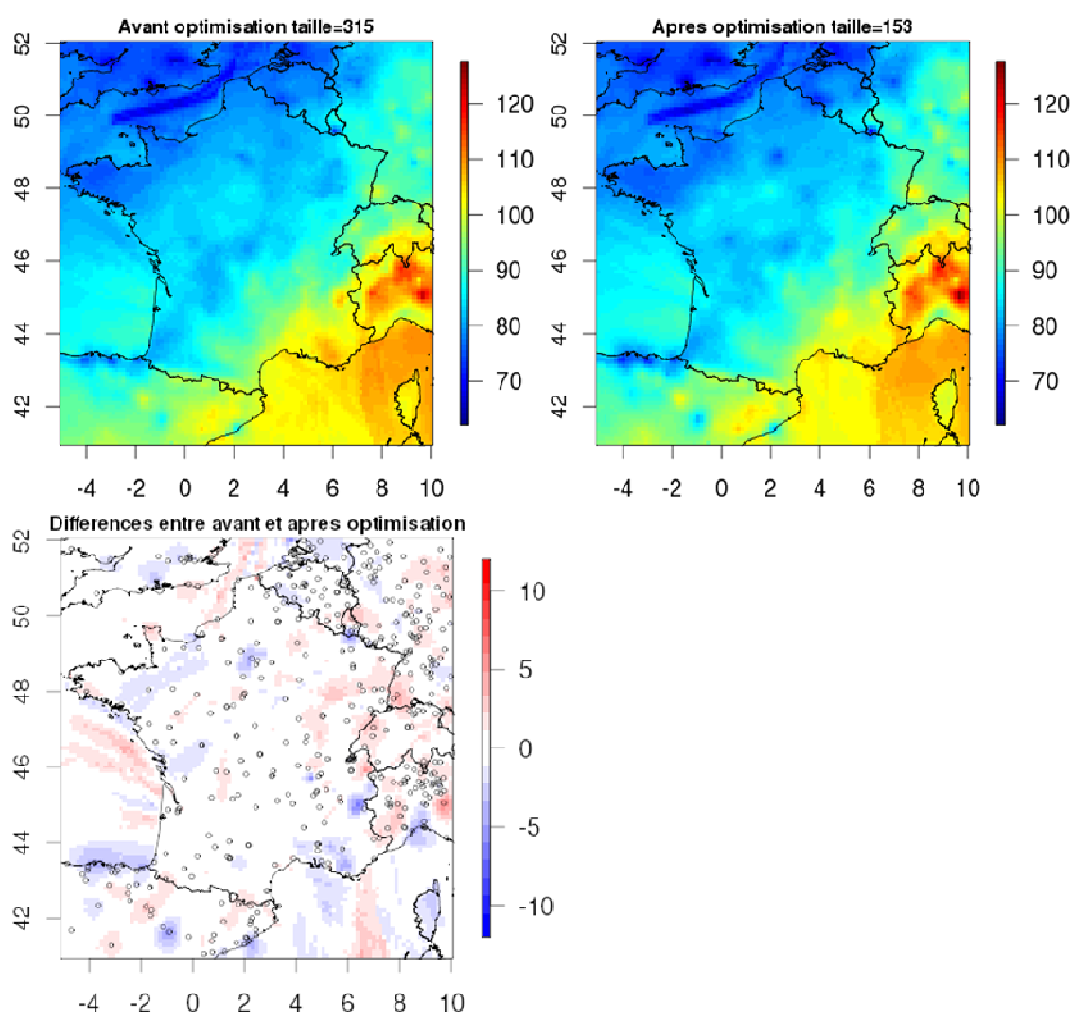


Figure 11 : Krigeage avec CHIMERE en dérive externe : cartes analysées en $\mu\text{g}/\text{m}^3$ des maxima journaliers pour l'ozone des moyennes glissantes sur 8h, moyennée sur les périodes estivales 2006-2009

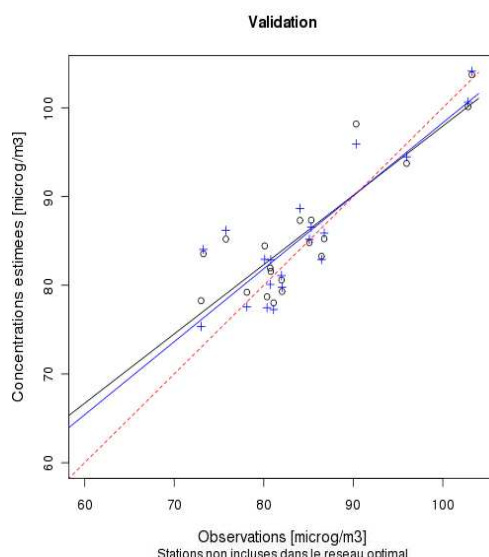


Figure 12 : Test de validation absolue par krigeage en dérive externe : valeurs estimées en fonction des valeurs vraies. En noir, les résultats après analyse avec 295 stations françaises ; en bleu avec les 153 stations du réseau optimisé

➤ Krigeage ordinaire

De la même manière, une analyse variographique avant et après optimisation est réalisée par krigeage ordinaire. La Figure 13 présente les variogrammes expérimentaux et modélisés ainsi que les paramètres des modèles avant et après optimisation.

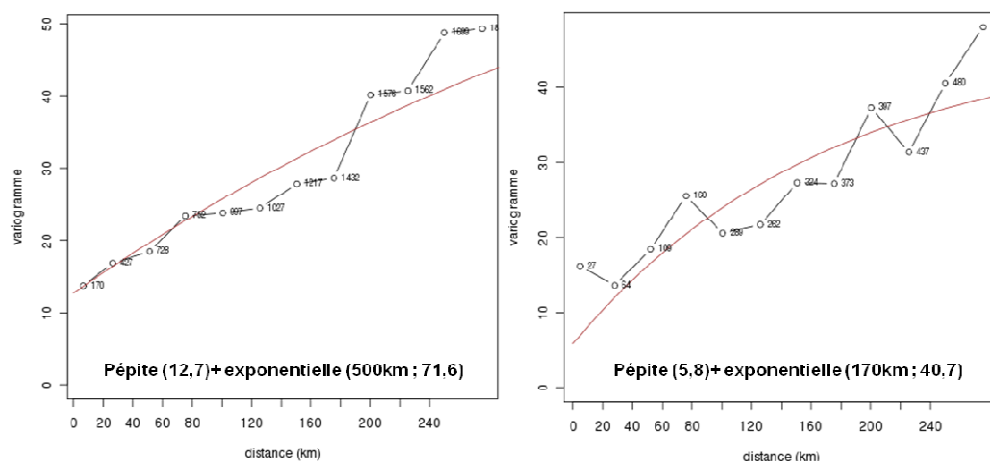


Figure 13 : Krigeage ordinaire : variogrammes expérimentaux et modélisés avant et après optimisation et paramètres des variogrammes modélisés

Les estimations des valeurs des maximas journaliers ont ensuite été réalisées par krigeage ordinaire sur la base de ces modèles et du réseau avant et après optimisation et la Figure 14 présente les cartes analysées.

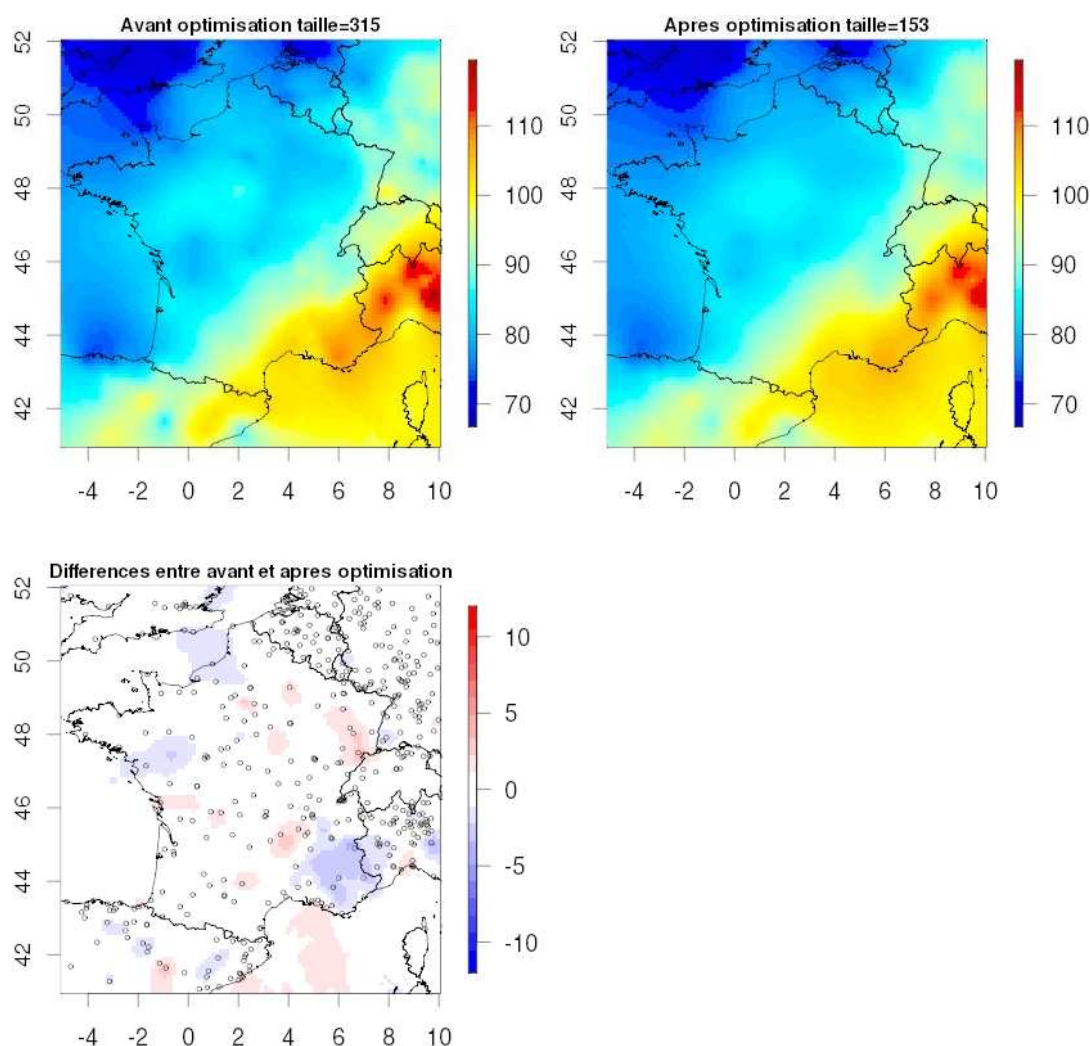


Figure 14 : Krigeage ordinaire : cartes analysées en $\mu\text{g}/\text{m}^3$ des maxima journaliers pour l'ozone des moyennes glissantes sur 8h, moyennée sur les périodes estivales 2006-2009

Comme précédemment, on remarque que les cartographies présentées en Figure 14 sont similaires. Les différences s'observent localement sur le territoire français, en Rhône-Alpes en particulier.

Dans le cadre du test de validation absolue, les estimations aux points écartés lors de l'analyse variographique sont calculées par krigeage et comparées aux valeurs vraies : la comparaison entre estimations et valeurs vraies est présentée en Figure 15. L'avantage (ou non) de l'optimisation n'est pas aussi clair sur ce graphique que sur la Figure 12. Le coefficient de corrélation après optimisation est un peu moins bon ($r=0,749$) qu'avant optimisation ($r=0,777$) mais les moyennes des erreurs d'estimations sont équivalentes ($0,4 \mu\text{g}/\text{m}^3$ après optimisation contre $0,3 \mu\text{g}/\text{m}^3$ avant). (Les statistiques détaillées sont présentées en Annexe 2 Tableau 7).

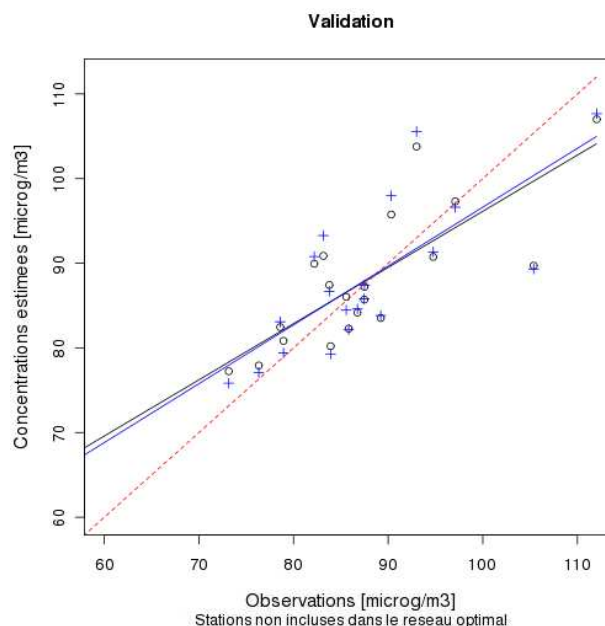


Figure 15 : Test de validation absolue par krigeage ordinaire : valeurs estimées en fonction des valeurs vraies. En noir, les résultats après analyse avec 295 stations françaises ; en bleu avec les 153 stations du réseau optimisé

4. COMPARAISON DES RESEAUX OPTIMISES

L'optimisation avec krigeage ordinaire ne fait à aucun moment intervenir les simulations CHIMERE au contraire de l'optimisation avec krigeage avec dérive externe. Certes, les réseaux optimisés obtenus sont différents mais on peut distinguer un groupe de points de mesure d'ozone choisis dans les deux cas, ainsi qu'un set de stations qui ne sont choisies dans aucun des cas.

Afin de compléter la comparaison, une troisième procédure d'optimisation est menée après bruitage grossier des sorties CHIMERE (ajout aux données de concentrations en $\mu\text{g}/\text{m}^3$ d'un bruit gaussien centré en 0 et d'écart-type égal à 3). Même en ajoutant au set d'optimisations un cas extrême où les performances du modèle sur l'ozone sont dégradées, il est facile d'identifier un groupe de 46 points de mesure ozone qui sont choisis systématiquement et à l'inverse un groupe de 62 points qui ne sont jamais choisis dans les réseaux optimisés. Le détail de ces points de mesure est donné en Annexe 3 et la carte de la Figure 16 présente la répartition géographique de ces stations qui sont réparties de façon uniforme sur le territoire et plus densément au niveau des agglomérations.

C'est également à proximité des grandes agglomérations que la majorité des stations qui sont systématiquement laissées de côté sont situées : 61% de ce groupe de stations sont des stations urbaines ; ce qui confirme le fait que la densité du réseau d'ozone dans les agglomérations est supérieure à ce qui est nécessaire pour l'élaboration de cartographies. En ce qui concerne les points de mesure ruraux, seulement 5 d'entre eux ne sont jamais sélectionnés et ils représentent 24% des points systématiquement choisis dans les réseaux optimisés.

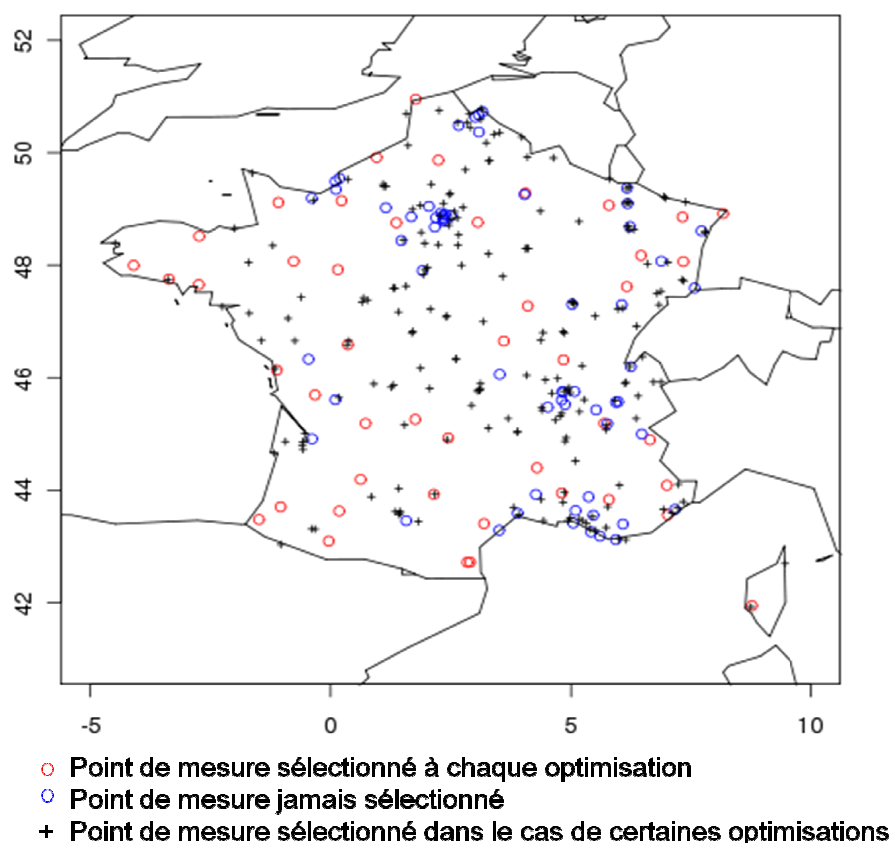


Figure 16 : Carte présentant les points de mesure d'ozone sélectionnés dans les 3 procédures d'optimisation (en rouge) et ceux jamais sélectionnés (en bleu). Les croix représentent les autres points de mesure qui peuvent être sélectionnés dans un ou deux des 3 réseaux.

5. CONCLUSION ET PERSPECTIVES

Le réseau de mesures français d'ozone est dense. Cette étude a permis de montrer que réduire le nombre de stations jusqu'au minimum européen, imposé par le zonage français défini en 2010, permet d'obtenir des cartographies quasiment similaires aux cartographies actuelles que ce soit avec un krigeage avec CHIMERE en dérive externe ou avec un krigeage ordinaire. Les tests de validation confirment que l'estimation des valeurs de concentrations ne semble pas souffrir de la réduction de moitié du nombre de mesures fixes en France. Une étude plus locale au niveau de la région Rhône-Alpes pourrait permettre de répartir différemment les stations et ainsi diminuer les différences observées.

Ces résultats constituent le début d'un travail de faisabilité et d'opportunité sur l'optimisation du réseau français, ils démontrent l'efficacité de la méthodologie développée et permettent d'ores et déjà d'identifier des points de mesure d'ozone apparaissant comme non indispensables à une bonne estimation de la qualité de l'air et d'autres qui, au contraire, fournissent une information nécessaire à l'élaboration de cartographies (voir Annexe 3).

Ces résultats laissent également entrevoir diverses utilisations de cette méthode. Celle-ci sera appliquée prochainement aux réseaux NO₂ et PM₁₀ mais on peut également imaginer adapter la méthodologie à l'échelle régionale comme outil de recherche d'implantation de stations.

6. REFERENCES

1. **Brus, D.J. et Heuvelink, G.B.M.** Optimization of sample patterns for universal kriging of environmental variables. *Geoderma*. 2007, 138, pp. 86-95.
2. **Romary, T. et Fouquet (de), C.** *Recommandations sur le schéma d'échantillonnage pour les campagnes de mesures de la qualité de l'air : une approche par optimisation*. LCSQA / Mines ParisTech. avril 2010.
3. **Romary, T., de Fouquet, C. et Malherbe, L.** Sampling design for air quality measurement surveys : an optimization approach. *Atmospheric environment*. 2011, 45, pp. 3613-3620.
4. **Vautard R., M. Beekmann, J. Roux, D. Gombert.** Validation of a deterministic forecasting system for the ozone concentrations over the Paris area. *Atmospheric Environment*. 2001, 35, pp. 2449-2461.
5. **Kirkpatrick, S., Gelatt Jr, C.D. et Vecchi, M.P.** Optimization by simulated annealing. *Science*. 1983, Vol. 220, 4598, pp. 671-680.
6. **Wu, L., Bocquet, M. et Chevallier, M.** Optimal reduction of the ozone monitoring network over France. *Atmospheric Environment*. 2010, 44, pp. 3071-3083.
7. **Wu, L. et Bocquet, M.** Optimal redistribution of the background ozone monitoring stations over France. *Atmospheric Environment*. 2011, 45, pp. 772-783.
8. **Pibouleau, L., et al.** Expérimentations numériques sur les variantes et paramètres de la méthode du recuit simulé. *Chemical Engineering Journal*. 15 janvier 2005, Vol. 105, 3, pp. p 117-130.

7. LISTE DES ANNEXES

Repère	Désignation	Nombre de pages
Annexe 1	Analyse géostatistique préliminaire - Tableaux	3
Annexe 2	Réalisation de cartes analysées : validation absolue	2
Annexe 3	Identification des points de mesure d'ozone choisis et non choisis quelle que soit l'optimisation testée	4
Annexe 4	Introduction d'une pondération par la RMSE entre observations et modèle	7

ANNEXE 1

Analyse géostatistique préliminaire – Tableaux

➤ **Test de validation croisée, en voisinage unique**

Tableau 3 : Résultats de la validation croisée sur toutes les données pour le krigeage avec CHIMERE en dérive externe

	Validation croisée par krigeage avec CHIMERE en dérive externe selon le modèle : $\text{pépité}(13,3) + \exp(32\text{km} ; 8,9)$		
	Valeurs calculées à partir des mesures	Valeurs estimées	Erreur d'estimation
Nombre de données	295	295	295
Minimum	68.92	78.10	-16.89
25 ^{ème} percentile	81.23	81.77	-2.55
Médiane	85.64	84.98	-0.28
75 ^{ème} percentile	91.62	90.62	2.94
Maximum	115.57	108.41	13.78
Étendue (=Max-Min)	46.65	30.31	30.67
Moyenne	87.39	87.39	0.004
Écart-type	8.9	7.58	4.51
Variance	79.3	57.38	20.30
Corrélation mesures/estimations		0.862	

Tableau 4 : Résultats de la validation croisée sur toutes les données pour le krigeage ordinaire

	Validation croisée par krigeage ordinaire selon le modèle : $\text{pépité}(12,7) + \exp(500\text{km} ; 68,3)$		
	Valeurs calculées à partir des mesures	Valeurs Estimées	Erreur d'estimation
Nombre de données	295	295	295
Minimum	68.92	75.70	-18.47
25 ^{ème} percentile	81.23	81.03	-2.53
Médiane	85.64	85.39	-0.35
75 ^{ème} percentile	91.62	91.29	2.77
Maximum	115.57	106.99	13.72
Étendue (=Max-Min)	46.65	31.29	32.19
Moyenne	87.39	87.37	-0.02
Écart-type	8.9	7.8	4.3
Variance	79.3	60.9	18.6
Corrélation mesures/estimations		0.875	

➤ **Test de validation absolue menée sur les données complémentaires**

Tableau 5 : Résultats de la validation absolue pour les deux types de krigeage

		Validation absolue par krigeage avec CHIMERE en dérive externe selon le modèle : pépité(13,3) + exp(32km ; 8,9)	Validation absolue par krigeage ordinaire selon le modèle : pépité(12,7) + exp(500km ; 68,3)
	Valeurs calculées à partir des mesures	Valeurs estimées	Valeurs estimées
Nombre de données	20	20	20
Minimum	75.12	80.29	80.39
25 ^{ème} percentile	83.23	85.39	84.08
Médiane	92.64	92.85	90.38
75 ^{ème} percentile	101.30	101.50	102.02
Maximum	112.00	108.00	106.94
Moyenne	92.52	93.13	92.97
Écart-type	11.4	9.15	9.58
Variance	130	84	92
Corrélation mesures/estimations		0.865	0.867

ANNEXE 2

Réalisation de cartes analysées **Validation absolue**

➤ **Krigeage avec CHIMERE en dérive externe**

Tableau 6 : Résultats de la validation absolue par krigeage en dérive externe avant et après optimisation

	Valeurs calculées à partir des mesures	Avant optimisation pépité (12.8) + exp(22km ; 10.9)		Après optimisation pépité (8.1) + exp(32km ; 16.4)	
		Valeurs Estimées	Erreur d'estimation	Valeurs Estimées	Erreur d'estimation
Nombre de données	20	20	20	20	20
Minimum	73.00	78.02	-3.19	75.36	-3.83
25 ^{ème} percentile	80.27	80.27	-1.80	80.01	-1.65
Médiane	82.01	83.99	0.64	83.48	-0.24
75 ^{ème} percentile	86.53	87.31	3.53	87.08	2.47
Maximum	103.20	103.73	10.30	104.12	10.79
Étendue (=Max-Min)	30.20	25.71	13.49	28.76	14.62
Moyenne	84.36	85.73	1.37	85.45	1.09
Écart-type	8.35	7.52	4.17	7.98	4.13
Variance	69.8	56.6	17.4	62.1	17.1
Corrélation mesures/estimations		0.866		0.872	

➤ **Krigeage ordinaire**

Tableau 7 : Résultats de la validation absolue par krigeage ordinaire avant et après optimisation

	Valeurs calculées à partir des mesures	Avant optimisation pépité (12.7) + exp(500km ; 71.6)		Après optimisation pépité (5.8) + exp(170km ; 40.7)	
		Valeurs Estimées	Erreur d'estimation	Valeurs Estimées	Erreur d'estimation
Nombre de données	20	20	20	20	20
Minimum	73.16	77.24	-15.66	75.84	-16.05
25 ^{ème} percentile	82.91	82.42	-3.59	82.85	-3.49
Médiane	86.29	86.63	0.34	86.24	-0.27
75 ^{ème} percentile	30.99	90.79	3.92	91.82	3.30
Maximum	112.00	107.01	10.79	107.67	12.56
Étendue	38.84	29.77	26.45	31.83	28.61
Moyenne	87.75	88.01	0.26	88.11	0.36
Écart-type	9.4	8.0	5.9	8.7	6.4
Variance	87.9	64.1	35.3	75.8	41.5
Corrélation mesures/estimations		0.777		0.749	

ANNEXE 3

Identification des points de mesure d'ozone choisis et non choisis quelle que soit l'optimisation testée

➤ **Points de mesure fixes d'ozone retenus dans le réseau optimisé**
quelle que soit la procédure d'optimisation choisie

Code station	Typologie	Zone	Nom station
03080	urbaine	PACA-AVIGNON	AVIGNON MAIRIE
04324	rurale	ILE-DE-FRANCE-ZR	Zone Rurale Est
07017	urbaine	AUVERGNE-ZUR	Aurillac Mairie
07022	rurale	AUVERGNE-ZR	Paray le Fresil
08023	rurale	LANGUEDOC-ROUSSILLON-ZUR	Biterrois-Narbonnais
08712	périurbaine	LANGUEDOC-ROUSSILLON-ZUR	St Estève
08714	urbaine	LANGUEDOC-ROUSSILLON-ZUR	Les Carmes
09008	périurbaine	POITOU-CHARENTES-ZUR	Aytré
09015	périurbaine	POITOU-CHARENTES-ZUR	Les Couronneries
09017	urbaine	POITOU-CHARENTES-ZR	COGNAC
10025	périurbaine	NORD-PAS-DE-CALAIS-ZUR	Sangatte
12020	rurale	MIDI-PYRENEES-ZR	PEYRUSSE
12026	urbaine	MIDI-PYRENEES-ZUR	ALBI DELMAS
12042	urbaine	MIDI-PYRENEES-ZR	LOURDES LAPACCA
13014	urbaine	FRANCHE-COMTE-ZR	Vesoul Pres Caillet
14010	périurbaine	CHAMPAGNE-ARDENNE-ZUR	BETHENY
15017	urbaine	RHONE-ALPES-GRENOBLE	Fontaine Les Balmes
16017	rurale	ALSACE-ZR	Nord-est Alsace
16031	rurale	ALSACE-ZR	Vosges du Nord
16064	périurbaine	ALSACE-ZUR	Colmar Sud
18035	périurbaine	PICARDIE-ZUR	Salouel
19021	urbaine	BRETAGNE-ZUR	Lorient CTM
19031	urbaine	BRETAGNE-ZR	Vannes Roscanvec
19051	urbaine	BRETAGNE-ZR	Quimper Ferry
19061	urbaine	BRETAGNE-ZR	Saint Brieuc Balzac
21021	urbaine	BASSE-NORMANDIE-ZR	Lisieux
21040	urbaine	BASSE-NORMANDIE-ZR	SAINT LO EGLISE
23123	urbaine	PAYS DE LA LOIRE-ZUR	MAZAGRAN
23163	périurbaine	PAYS DE LA LOIRE-ZUR	Fillé
24009	urbaine	PACA-NICE	CANNES BROUSSAILLES
24014	rurale	PACA-ZR	CIANS
24018	urbaine	PACA-ZR	MANOSQUE
24024	urbaine	PACA-ZR	BRIANCON
25045	rurale	HAUTE-NORMANDIE-ZR	Phare d Ailly
26012	rurale	BOURGOGNE-ZR	Station MORVAN
30019	urbaine	LORRAINE-ZUR	Épinal
30033	rurale	LORRAINE-ZR	Jonville en Woevre
31016	urbaine	AQUITAINE-ZUR	SAINT-CROUTS
31032	urbaine	AQUITAINE-ZUR	AGEN
31033	urbaine	AQUITAINE-ZUR	PERIGUEUX
31036	urbaine	AQUITAINE-ZR	DAX
32006	périurbaine	BOURGOGNE-ZR	Macon Paul Bert
34044	urbaine	CENTRE-ZR	Dreux Nord
35006	urbaine	LIMOUSIN-ZR	TULLE - Hugo
36010	rurale	RHONE-ALPES-ZR	Ardèche Rurale Sud
41007	périurbaine	CORSE-ZUR	AJACCIO SPOSATA

➤ **Points de mesure fixes d'ozone exclus des réseaux optimisés obtenus quelle que soit la procédure d'optimisation choisie**

Code station	Typologie	Zone	Nom station
01017	urbaine	LORRAINE-METZ	Metz-Sablon
01019	urbaine	LORRAINE-METZ	Thionville-Piscine
02004	urbaine	PACA-ZI	Martigues P. Central
02016	urbaine	PACA-ZUR	Salon de Provence
03045	urbaine	PACA-MARSEILLE	MARSEILLE MARGUERITE
03046	périurbaine	PACA-TOULON	LA CIOTAT
03048	périurbaine	PACA-MARSEILLE	AIX PLATANES
03063	urbaine	PACA-TOULON	TOULON ARSENAL
03067	périurbaine	PACA-ZR	BRIGNOLES
03086	rurale	PACA-ZR	APT
04002	urbaine	ILE-DE-FRANCE-PARIS	GENNEVILLIERS
04004	urbaine	ILE-DE-FRANCE-PARIS	PARIS 18eme
04008	urbaine	ILE-DE-FRANCE-PARIS	CACHAN
04018	urbaine	ILE-DE-FRANCE-PARIS	AUBERVILLIERS
04023	urbaine	ILE-DE-FRANCE-PARIS	CERGY-PONTOISE
04034	urbaine	ILE-DE-FRANCE-PARIS	VITRY-SUR-SEINE
04037	urbaine	ILE-DE-FRANCE-PARIS	PARIS 13eme
04049	périurbaine	ILE-DE-FRANCE-PARIS	LES ULIS
04100	urbaine	ILE-DE-FRANCE-PARIS	VILLEMOMBLE
04145	urbaine	ILE-DE-FRANCE-PARIS	GARCHES
04160	urbaine	ILE-DE-FRANCE-PARIS	PARIS 6eme
04322	rurale	ILE-DE-FRANCE-ZR	Zone Rurale Ouest
05074	urbaine	HAUTE-NORMANDIE-ZI	Le Havre Ec. Herriot
05079	périurbaine	HAUTE-NORMANDIE-ZI	Montivilliers
05082	périurbaine	BASSE-NORMANDIE-ZR	Touques
06003	urbaine	NORD-PAS-DE-CALAIS-DOUAI	Douai Theuriet
07020	rurale	AUVERGNE-ZR	Busset
08016	urbaine	LANGUEDOC-ROUSSILLON-MONTPELLIER	Prés Arènes
08022	périurbaine	LANGUEDOC-ROUSSILLON-ZR	Agathois-piscénois
08617	périurbaine	LANGUEDOC-ROUSSILLON-ZUR	La Calmette
09016	périurbaine	POITOU-CHARENTES-ZUR	La Couronne
09019	urbaine	POITOU-CHARENTES-ZUR	École Jules Ferry
11016	urbaine	NORD-PAS-DE-CALAIS-LILLE	Marcq CTM
11027	urbaine	NORD-PAS-DE-CALAIS-LILLE	Tourcoing Centre
11029	urbaine	NORD-PAS-DE-CALAIS-LILLE	Lomme Hugo
12041	périurbaine	MIDI-PYRENEES-TOULOUSE	SICOVAL
14004	urbaine	CHAMPAGNE-ARDENNE-ZUR	MAIRIE REIMS
15001	rurale	RHONE-ALPES-GRENOBLE	Charavines
15031	rurale	PACA-ZR	Le Casset2
15038	urbaine	RHONE-ALPES-GRENOBLE	St Martin dHeres
16001	périurbaine	ALSACE-STRASBOURG	STG Ouest
16060	périurbaine	ALSACE-ZR	C.C.3 Frontières
17006	périurbaine	FRANCHE-COMTE-ZUR	Besançon Chailluz
20004	urbaine	RHONE-ALPES-LYON	ST JUST

20017	urbaine	RHONE-ALPES-LYON	GERLAND
20037	périurbaine	RHONE-ALPES-LYON	TERNAY
20048	périurbaine	RHONE-ALPES-LYON	SAINT EXUPERY
21001	urbaine	BASSE-NORMANDIE-ZUR	Chemin Vert CAEN
24020	urbaine	PACA-NICE	CAGNES LADOUMEGUE
25039	urbaine	HAUTE-NORMANDIE-ZUR	Évreux Centre
26002	urbaine	BOURGOGNE-ZUR	Station TARNIER
27004	urbaine	RHONE-ALPES-LYON	Vienne Centre
28020	périurbaine	NORD-PAS-DE-CALAIS-DOUAI	Noeux S. Sports
29426	urbaine	RHONE-ALPES-ST ETIENNE	SAINT-CHAMOND
30022	urbaine	LORRAINE-ZR	Gérardmer
30024	périurbaine	LORRAINE-NANCY	Tomblaine
31031	périurbaine	AQUITAINE-BORDEAUX	SAINT-SULPICE
33101	urbaine	RHONE-ALPES-ZUR	PASTEUR
33103	périurbaine	RHONE-ALPES-ZUR	BARBY
33211	urbaine	RHONE-ALPES-ZUR	ANNEMASSE
34011	urbaine	CENTRE-ORLEANS	Préfecture
34042	urbaine	CENTRE-ZUR	Lucé

ANNEXE 4

Introduction d'une pondération par la RMSE entre observations et modèle

Il peut être intéressant d'orienter l'optimisation du réseau vers un réseau qui permettrait de mieux appréhender la pollution atmosphérique en combinant modèle et mesures. Pour ce faire, on introduit une pondération dans la fonction de coût, comme cela avait été réalisé dans Romary *et al* (3). Dans cette étude, l'influence des performances du modèle CHIMERE sur la distribution du réseau optimisé est testée en modifiant la fonction de coût comme suit :

$$\varphi(\omega) = \frac{1}{|D|} \int_D \text{Var}[\hat{Z}(x) - Z(x)] w(x) dx \quad [11]$$

où w est la pondération introduite, ici l'inverse de la RMSE entre CHIMERE et les observations.

CALCUL DES SCORES DE RMSE

La pondération introduite dans la fonction de coût est basée sur le calcul de la RMSE entre la valeur de Z observée aux stations de mesure et la valeur de Z estimée par le modèle CHIMERE. Ce calcul est réalisé sur les 315 stations de fond françaises utilisées dans le cadre de l'optimisation ainsi que sur les stations AIRBASE.

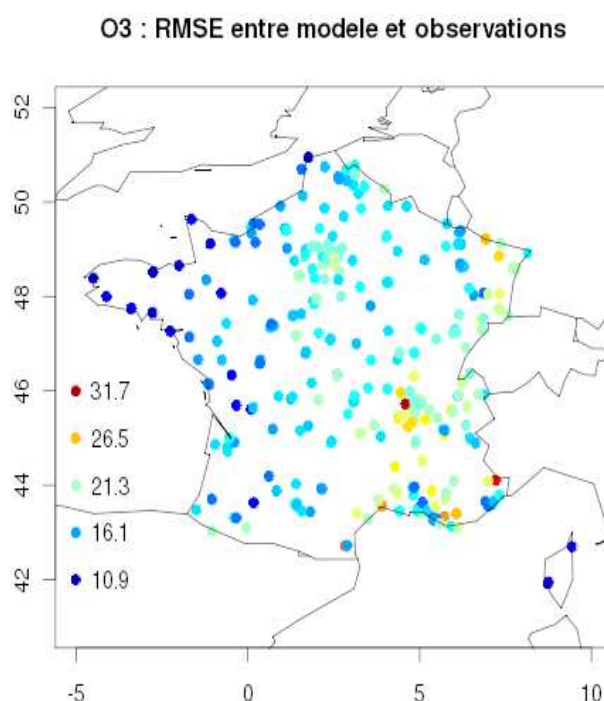


Figure 17 : Calcul des scores de RMSE pour la moyenne sur l'été des maxima journaliers des moyennes glissantes sur 8h d'ozone

Sur la Figure 17, on observe une certaine spatialisation des scores de RMSE avec des zones où celle-ci est plus faible : Bretagne et côte ouest ; des zones où, au contraire, elle est plus élevée : sud-est.

ANALYSE GEOSTATISTIQUE DES SCORES DE RMSE

La variance de l'erreur d'estimation ne dépendant pas des données d'observations de la variable étudiée, l'intégrale de l'équation [11] est calculée sur un maillage régulier de $\mathcal{D}$, à savoir la grille CHIMERE à 10km sur la France (domaine FRA10). Il convient donc de connaître la valeur de la pondération en tout point de cette grille. Pour ce faire, une étude variographique sur la RMSE est réalisée. Celle-ci est justifiée par la présence de zones de valeurs de RMSE similaires sur la Figure 17.

L'analyse variographique est d'abord réalisée sur les 315 stations françaises. Un modèle, composé d'une structure pépétique (pépité=1,60) additionnée d'une structure exponentielle (portée 20km ; palier 6,1), a été ajusté aux données expérimentales comme indiqué sur la Figure 18.

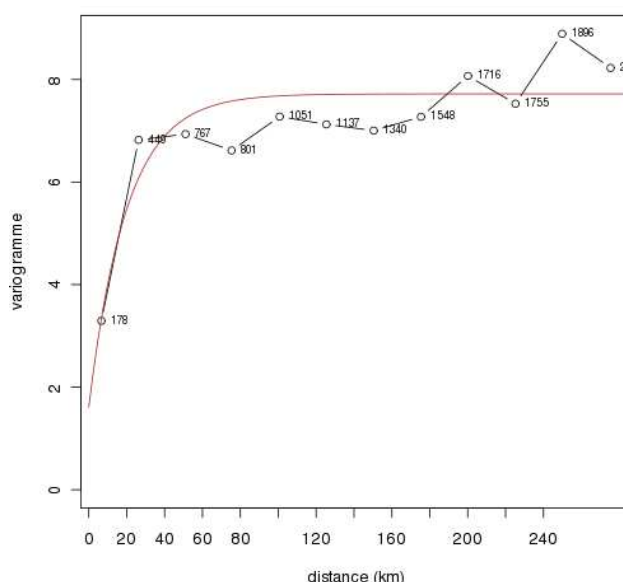


Figure 18 : Variogrammes expérimental et modélisé des scores de RMSE pour l'ozone (krigeage ordinaire)

Après le calcul du variogramme, les stations AIRBASE hors France mais incluses dans le domaine FRA10, de typologie de fond sont ajoutées au set de stations pour réaliser le krigeage.

➤ Test de validation croisée, en voisinage unique

Afin d'apprécier la qualité du modèle, un test de validation croisée, en voisinage unique, est réalisé et les valeurs estimées sont comparées aux valeurs « vraies », notamment en calculant l'erreur d'estimation (différence entre valeur estimée moins valeur vraie). Comme le montre la Figure 19, les valeurs estimées sont corrélées aux valeurs « vraies », avec un coefficient de corrélation de 0,66.

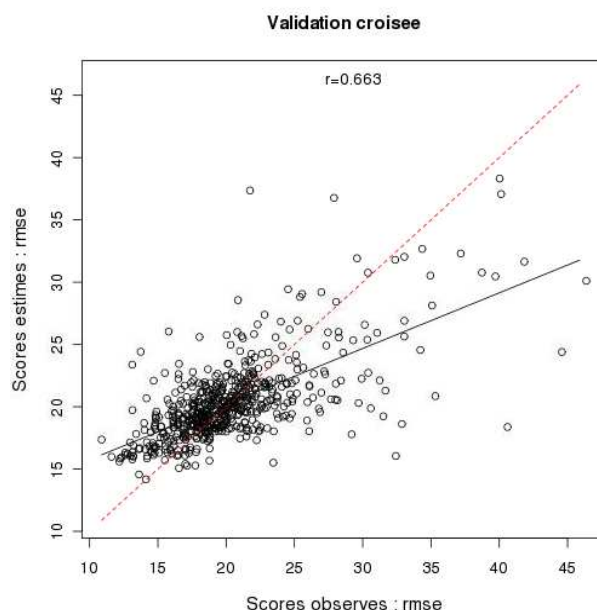


Figure 19 : Test de validation croisée : valeurs estimées en fonction des valeurs vraies

Un krigeage ordinaire est alors réalisé de manière à obtenir des valeurs en tout point de la grille FRA10. À titre d'information, la Figure 20 présente une cartographie de la RMSE : on retrouve bien les zones françaises de valeurs de RMSE similaires, déjà observées sur la Figure 17.

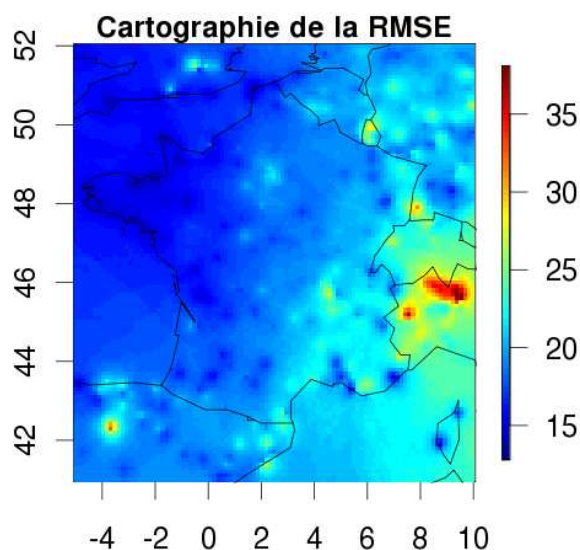


Figure 20 : Cartographie par krigeage ordinaire de la RMSE entre observations et modèle CHIMERE pour le maximum journalier des moyennes glissantes sur 8h des concentrations d'ozone, moyenné sur l'été entre 2006 et 2009.

OPTIMISATION PAR RECUIT SIMULE

La procédure d'optimisation est rigoureusement identique à celle utilisée précédemment dans le cadre d'une réduction avec un krigeage en dérive externe, à l'exception de l'introduction, dans le calcul de la fonction de coût, de l'inverse de la valeur de RMSE au point de grille considéré (voir Équation [11]). Sur la Figure 21 est présentée la répartition des points de mesures périurbains et urbains (à gauche) et ruraux (à droite) dans le dispositif optimisé trouvé. Le réseau initial était composé de 15,6% de stations rurales ; le nouveau réseau conserve cette proportion avec 23 stations rurales pour 130 stations périurbaines et urbaines.

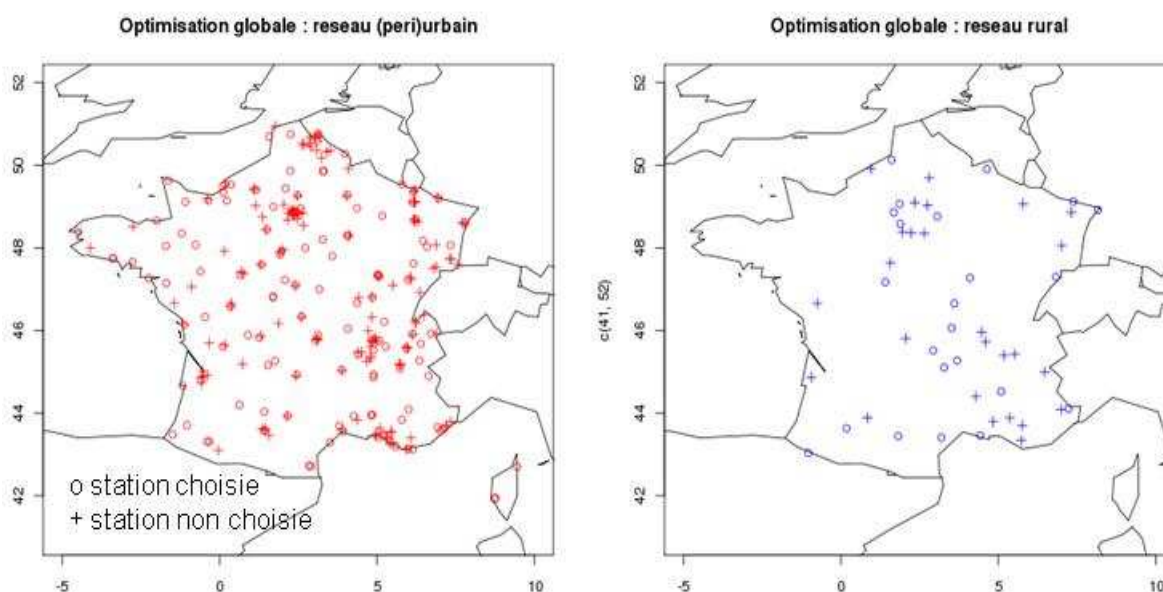


Figure 21 : Optimisation avec pondération : Résultats de la procédure d'optimisation du réseau de fond

Si l'on compare la répartition du réseau optimisé trouvée grâce à cette optimisation pondérée à celle trouvée précédemment (Figure 22), on observe qu'un set de stations est commun aux deux types d'optimisation (87 stations communes). Les autres points de mesure fixes sont répartis différemment sur le territoire français. L'examen détaillé des zones où la RMSE prenait des valeurs plus fortes et plus faibles (la Bretagne et la côte ouest ; et le sud-est respectivement) montre que l'algorithme a favorisé l'implantation des mesures fixes là où la RMSE était plus élevée avec 37 mesures contre 31 sans pondération tout en allégeant le dispositif à l'ouest (34 mesures fixes pour le réseau obtenu avec pondération contre 41 sinon).

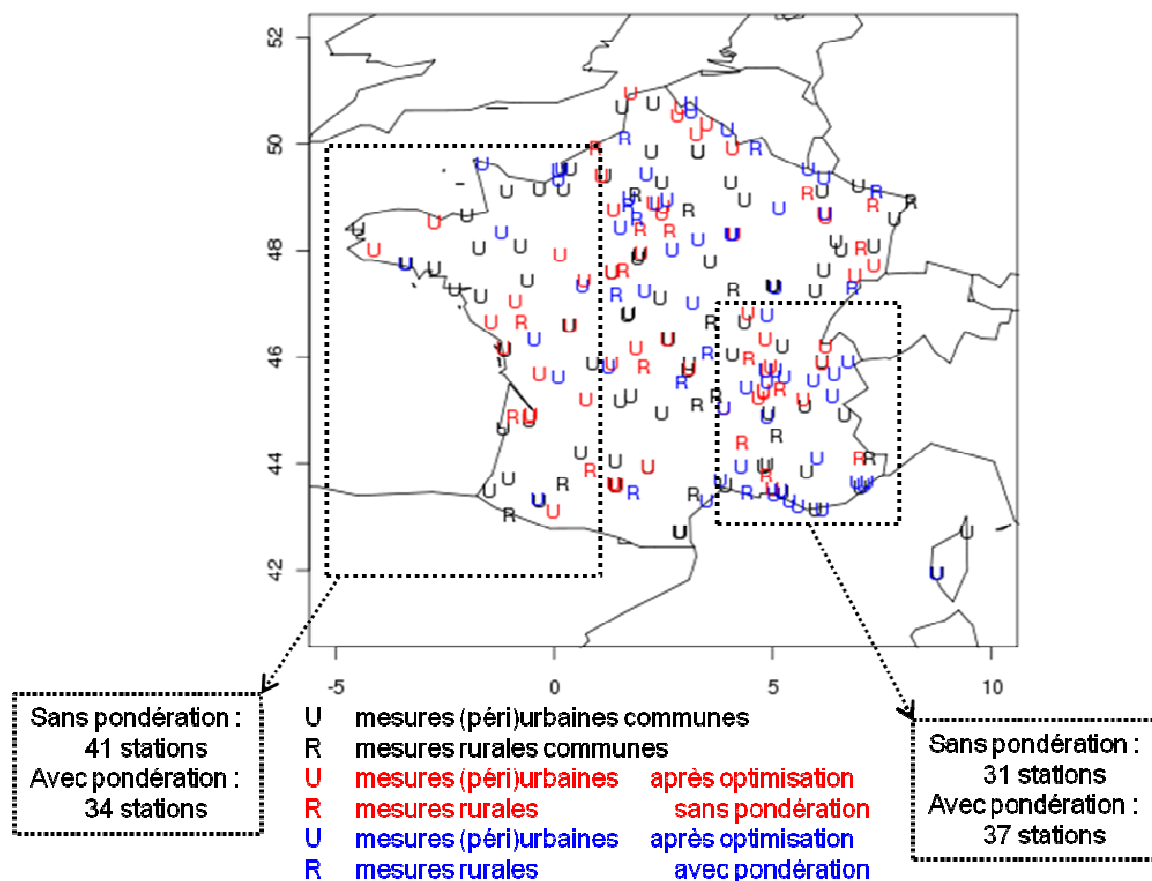


Figure 22 : Comparaison des réseaux optimisés après optimisation sans et avec pondération par l'inverse de la RMSE.

COMPARAISON DES CARTES ANALYSEES

La Figure 23 présente les cartes analysées des maxima journaliers d'ozone des moyennes glissantes sur 8h, moyennée sur la période estivale, réalisées à partir du réseau initial (en haut à gauche), du réseau après optimisation sans pondération (en haut au centre) et du réseau après optimisation après pondération avec l'inverse de la RMSE entre modèle et observation (les estimations ont été réalisées avec un krigeage en dérive externe). En bas, les cartographies soulignent les différences entre les cartes analysées avant et après optimisation sans pondération (à gauche) ; entre les cartes analysées avant et après optimisation sans pondération (au centre) ; et entre le cartes analysées après optimisation sans et avec pondération (à droite).

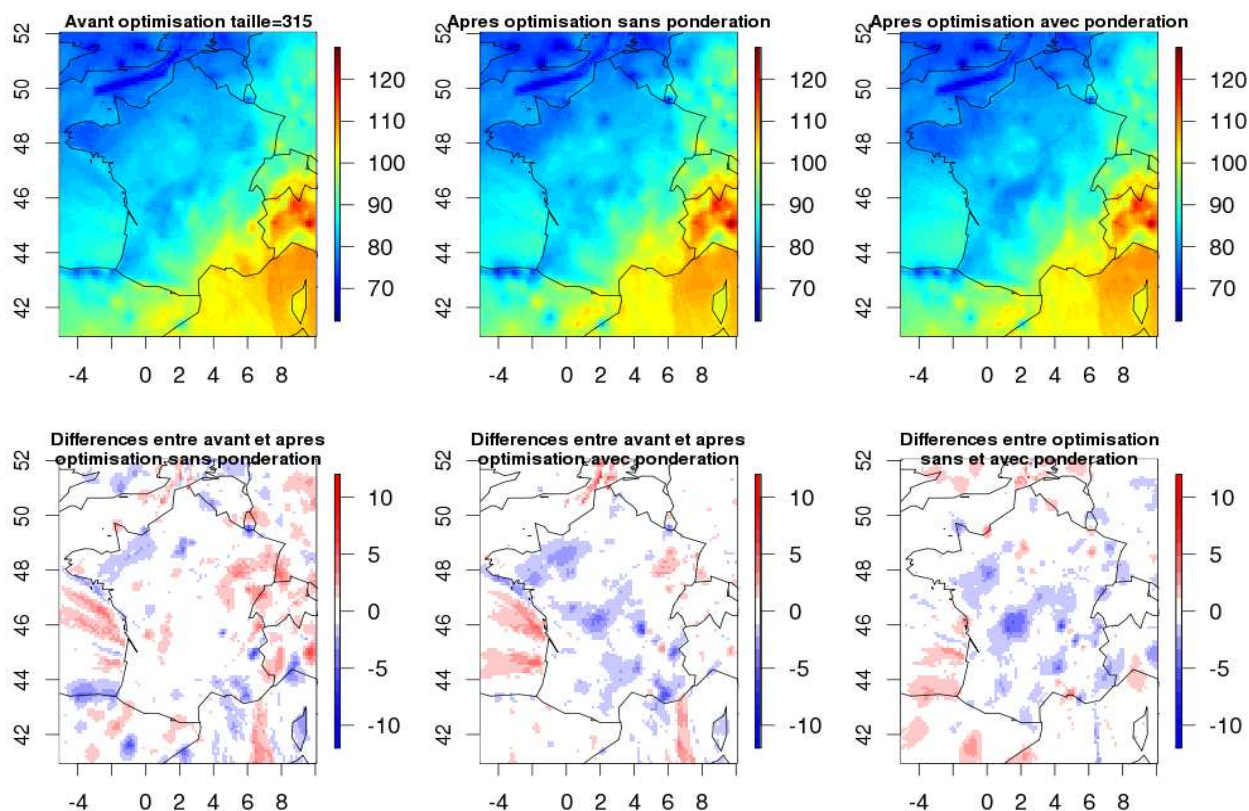


Figure 23 : Cartes analysées en $\mu\text{g}/\text{m}^3$ des maxima journaliers pour l'ozone des moyennes glissantes sur 8h, moyennées sur les périodes estivales 2006-2009 et comparaisons entre les différentes cartes.

L'examen de ces cartes analysées montre que les deux types d'optimisation génèrent des structures similaires à celles observées dans la cartographie initiale. Néanmoins, les cartes de différences entre les estimations obtenues avant et après optimisation montrent des différences plus marquées lorsque la pondération par l'inverse de la RMSE entre modèle et observations est introduite. En effet sur la France, l'introduction de cette pondération conduit à des grandes zones de maxima plus faibles qu'avec le réseau initial (zones en bleu sur la carte des différences). Une optimisation sans pondération génère quant à elle des différences plus locales.

CONCLUSION

L'introduction d'une pondération par l'inverse de la RMSE entre modèle et observations dans la fonction de coût génère des différences dans la répartition du dispositif optimisé ; densifiant comme attendu les zones où les performances du modèle sont moindres, aux dépens des zones où les performances sont meilleures.

Cependant, la carte analysée par krigeage en dérive externe avec le réseau optimisé obtenu après optimisation avec pondération montre des différences avec celle obtenue avec le réseau initial, différences plus marquées qu'après une optimisation sans pondération.