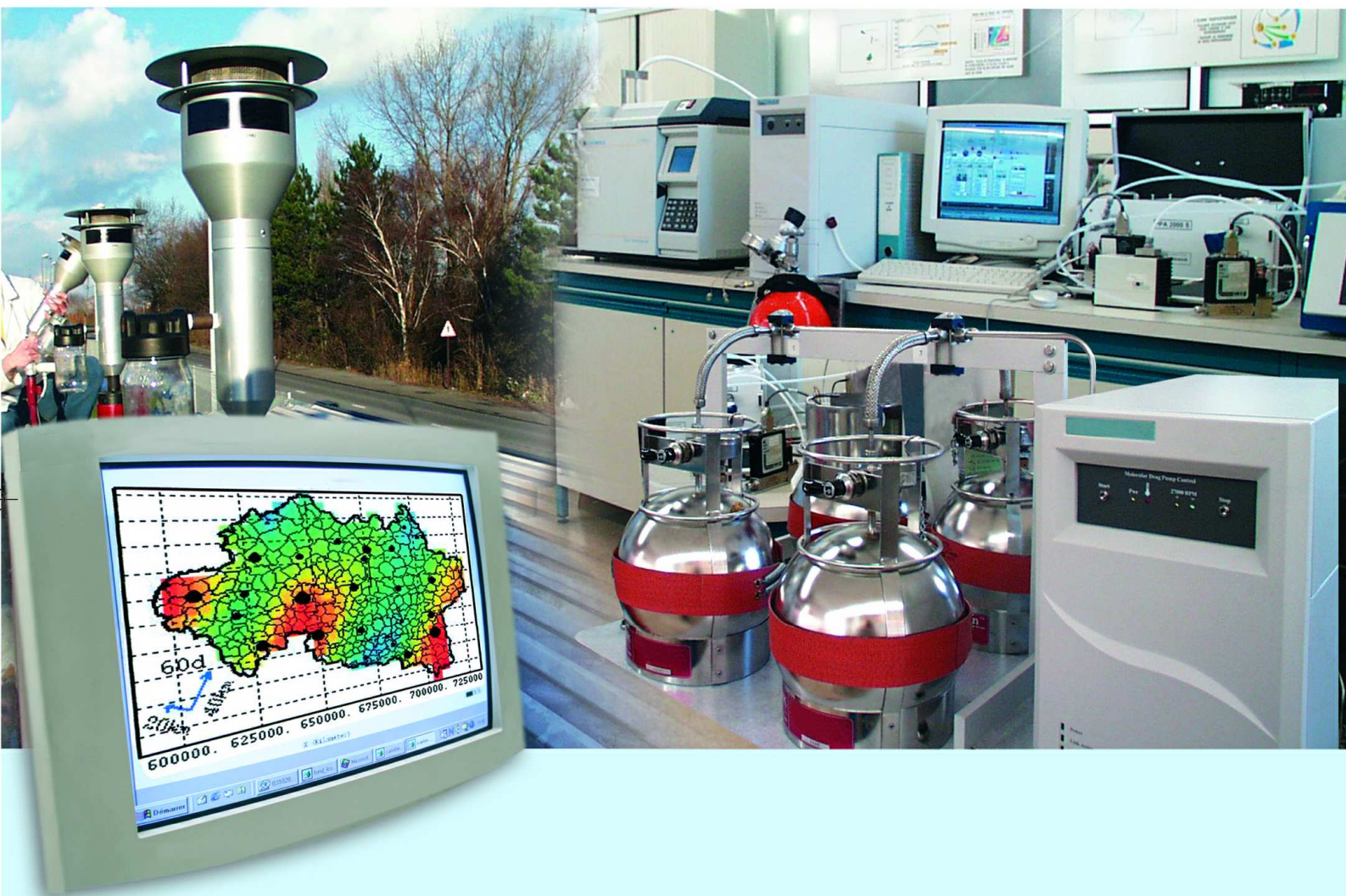




Laboratoire Central de Surveillance de la Qualité de l'Air



Modélisation et Traitements Numériques

Travaux relatifs à la plate-forme nationale de modélisation
PREV'AIR : Réalisation de cartes analysées d'ozone (2/2)

Décembre 2009

Programme 2009

L. MALHERBE, A. UNG





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. Ces travaux en matière de pollution atmosphérique supportés financièrement par la Direction Générale de l'Energie et du Climat du Ministère de l'Ecologie, de l'Energie, du Développement Durable et de la Mer sont réalisés avec le souci constant d'améliorer le dispositif de surveillance de la qualité de l'air en France, coordonné au plan technique par l'ADEME, en apportant un appui scientifique et technique aux AASQA.

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

Travaux relatifs à la plate-forme nationale de modélisation PREV'AIR : Réalisation de cartes analysées d'ozone (2/2)

Laboratoire Central de Surveillance
de la Qualité de l'Air

Thème : Modélisation - Traitements numériques

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

2009

L. MALHERBE, A. UNG

Ce document comporte 33 pages (hors couverture et annexes)

	Rédaction	Vérification	Approbation
NOM	L. MALHERBE A. UNG	B. BESSAGNET	M. RAMEL
Qualité	Ingénieurs de l'Unité MOCA Direction des Risques Chroniques	Responsable de l'Unité MOCA Direction des Risques Chroniques	Responsable LCSQA/INERIS Direction des Risques Chroniques
Visa			

TABLE DES MATIÈRES

RESUME	9
1. INTRODUCTION.....	11
1.1 Contexte.....	11
1.2 Contenu de l'étude	11
2. RAPPELS SUR LES METHODES D'ESTIMATION MULTI VARIABLE UTILISEES EN CARTOGRAPHIE.....	12
2.1.1 Bilan théorique des méthodes	12
2.1.2 Prise en compte de variables auxiliaires	13
3. APPLICATION DES METHODES A LA CARTOGRAPHIE HORAIRE D'OZONE ET EVALUATION	17
3.1 Période d'étude	17
3.2 Variables de concentration	17
3.3 Sélection des stations de mesure	17
3.4 Chaîne de calcul.....	19
3.5 Contrôle des hypothèses.....	22
3.6 Evaluation statistique.....	24
3.7 Cartographies.....	28
4. CONCLUSION	31
5. REFERENCES.....	32
6. LISTE DES ANNEXES	33

RESUME

Cette étude, conduite en collaboration avec le Centre de Géosciences (équipe de géostatistique) de Mines ParisTech, s'inscrit dans le cadre des travaux du LCSQA relatifs au système PREV'AIR. Elle a pour objet d'améliorer la réalisation des cartes analysées, qui sont produites en corrigeant les simulations du modèle CHIMERE à l'aide des données d'observation.

Ce rapport présente les résultats obtenus pour la cartographie horaire d'ozone sur la France. Un travail sur la cartographie journalière des PM_{10} a été effectué pour les besoins de l'étude LCSQA 2009 « Evaluation des zones géographiques touchées par les dépassements de valeurs limites ». Les compléments et adaptations nécessaires à une mise en œuvre opérationnelle dans PREV'AIR seront réalisés en 2010.

Rappelons que la méthode de correction utilisée à ce jour dans PREV'Air est la technique géostatistique du *krigeage des innovations*. Elle suppose que la concentration d'ozone est la somme du modèle CHIMERE et d'un résidu aléatoire non corrélé spatialement à CHIMERE mais structuré. Des valeurs expérimentales de ce résidu, les innovations, peuvent être calculées aux stations en faisant la différence entre les concentrations mesurées et modélisées. La méthode d'analyse consiste à ajouter à CHIMERE un champ correctif obtenu par krigeage ordinaire des innovations.

En 2008, la variabilité relative des mesures et des simulations a été examinée dans le temps et l'espace. A la lumière de cette analyse, il est apparu que les hypothèses sous-jacentes au krigeage des innovations n'étaient pas toujours vérifiées. Le modèle linéaire de corégionalisation a été proposé pour décrire de façon plus satisfaisante les relations entre les observations et CHIMERE. Ce modèle, utilisable dans un cokrigeage, permet de différencier les rapports de variabilité par échelle de temps ou d'espace.

En 2009, le krigeage des innovations et le cokrigeage ont été testés et comparés dans un cadre purement spatial, sur deux périodes estivales d'environ un mois. A cette occasion, le krigeage avec CHIMERE en dérive externe a été également appliqué. Il ne distingue pas les échelles de variabilité comme le cokrigeage mais, plus général que le krigeage des innovations, il permet d'ajuster localement une relation linéaire entre les observations et CHIMERE.

Dans le domaine convexe défini par les stations de mesure, les trois méthodes produisent des résultats comparables. Ainsi la précision de l'estimation dépend-elle en premier lieu de la qualité des données disponibles et de la bonne répartition des stations utilisées en entrée du krigeage. Les statistiques d'erreurs donnent toutefois l'avantage au krigeage avec dérive externe qui sera mis en fonctionnement dans PREV'AIR en 2010. Pour une pleine efficacité, le cokrigeage, dont on a pu constater aussi la pertinence pour la cartographie des PM_{10} , appelle des ajustements complémentaires.

1. INTRODUCTION

1.1 CONTEXTE

Depuis 2003 pour l'ozone et 2005 pour les PM₁₀, les cartes des concentrations simulées pour la veille par le modèle CHIMERE sont corrigées chaque matin à l'aide des observations extraites de la base BASTER : on parle de *cartes analysées*. La méthode employée, appelée *krigeage des innovations*, consiste à ajouter au modèle un champ correctif estimé par interpolation géostatistique. La mise en place de cette méthode, son évaluation et la définition des paramètres de krigeage ont fait l'objet de travaux du LCSQA (2003¹, 2004², 2005³).

Après plusieurs années de fonctionnement, qui ont vu évoluer le modèle CHIMERE ainsi que le réseau des stations de mesure, il s'est révélé utile d'évaluer une nouvelle fois l'approche et d'examiner si la correction pouvait être améliorée.

En 2008, une analyse approfondie de la variabilité temporelle et spatiale des concentrations d'ozone modélisées et mesurées a été réalisée à cette fin⁴. L'étude conduite en 2009, ciblée sur la cartographie horaire d'ozone, en constitue le prolongement. Comme en 2008, elle a bénéficié de la collaboration du Centre de Géosciences/géostatistique de Mines ParisTech.

Le problème de la cartographie des PM₁₀ a été traité dans le contexte spécifique de l'évaluation des zones géographiques touchées par les dépassements de valeurs limites⁵. Notons que la nature des relations de variabilité entre CHIMERE et la mesure montre des similitudes avec l'ozone. L'adaptation de ces résultats à la cartographie opérationnelle des PM₁₀ dans PREV'AIR sera effectuée en 2010.

1.2 CONTENU DE L'ETUDE

Trois techniques d'analyse relevant de la géostatistique multi variable ont été évaluées et comparées. A l'occasion de ces tests, les hypothèses sous-jacentes à chaque méthode ont été examinées. Cette mise au point théorique est synthétisée dans la première partie du rapport et détaillée dans l'annexe 2.

La seconde partie présente et commente les résultats des évaluations réalisées sur deux périodes estivales d'environ un mois.

¹ Honoré et Malherbe, 2003. Application de modèles grande échelle à la problématique régionale : cas de l'ozone. Rapport LCSQA.

² Honoré et Malherbe, 2004. Application de modèles grande échelle à la problématique régionale : cartographie de l'ozone et du dioxyde d'azote. Rapport LCSQA.

³ INERIS, 2005. Travaux relatifs à PREV'AIR. Rapport LCSQA

⁴ Malherbe L., Ung A., 2008. Travaux relatifs à la plate-forme nationale de modélisation PREV'AIR : Réalisation de cartes analysées d'ozone (3/3).

⁵ Malherbe L., Cárdenas G., 2009. Evaluation des zones géographiques touchées par les dépassements de valeurs limites. Note méthodologique. Version 1. Rapport LCSQA.

Rapports disponibles à l'adresse www.lcsqa.org

Les prolongements possibles de ce travail sont discutés en conclusion.

Tous les traitements statistiques et géostatistiques ont été effectués avec le logiciel R (version 2.9.0) et ses packages associés (chron, akima, splancs, geoR, gstat).

2. RAPPELS SUR LES METHODES D'ESTIMATION MULTI VARIABLE UTILISEES EN CARTOGRAPHIE

Ce chapitre présente trois méthodes d'estimation multi variable qui sont applicables à la réalisation de cartes analysées :

- Krigeage du résidu
- Krigeage avec dérive externe
- Cokrigeage

De plus amples détails sur les hypothèses, le principe, l'intérêt et les limites de ces techniques sont fournis en annexe 2.

Dans toute la suite :

- Z est la variable de concentration à estimer sur le domaine d'étude à un instant t_0 donné (ex. : concentration horaire d'ozone, concentration journalière de PM_{10}). Elle est mesurée sur un ensemble de points d'observation notés x_α .
- S est la simulation issue de CHIMERE pour cet instant t_0 . Elle est définie sur une grille de points notés x_β .
- Les concentrations sont estimées aux points d'une grille notés x_j . Ceux-ci coïncident ou non avec les x_β . Dans PREV'AIR, les x_j et les x_β coïncident.
- V_j est le voisinage d'estimation associé à un point x_j .

2.1.1 BILAN THEORIQUE DES METHODES

Les tableaux ci-après (Tableau 1, Tableau 2) récapitulent les hypothèses et les formules d'estimation propres à chaque méthode.

Les relations supposées entre CHIMERE et la mesure diffèrent selon la méthode :

- ***Krigeage du résidu/krigeage des innovations*** : relation linéaire globale de la forme $a.x+b$ (a et b constants sur le domaine)
- ***Krigeage avec dérive externe*** : relation linéaire ajustable localement (par voisinage)
- ***Cokrigeage*** : relation globale fixée par un modèle bi variable (dans le cas le plus fréquent où les paramètres du modèle ajusté sont constants sur le domaine) mais **différenciable par échelle de variabilité spatiale**.

Le choix d'une méthode sera ainsi guidé par la plus ou moins bonne adéquation entre ces modèles et les données disponibles.

2.1.2 PRISE EN COMPTE DE VARIABLES AUXILIAIRES

Les cartes analysées peuvent être affinées à l'aide de variables auxiliaires descriptives de l'environnement et des sources (données d'émissions, de population, d'occupation du sol, ...).

Si elles se révèlent corrélées aux concentrations, ces variables peuvent être introduites dans le krigeage en tant que dérivées. Pour chaque méthode d'analyse, le tableau de l'annexe 4 indique la façon de les prendre en compte.

Tableau 1 – Géostatistique multi variable appliquée à la réalisation de cartes analysées : hypothèses et formulations

Méthode	Hypothèses et relations sous-jacentes		Modélisation du variogramme
Krigeage du résidu	Relation linéaire globale (identique sur tout le domaine) entre Z et S Absence de corrélation spatiale entre S et le résidu W	$Z(x) = a.S(x) + b + T(x)$ En pratique : $Z(x) = a^*.S(x) + b^* + W(x)$ où a^* et b^* sont le plus souvent estimés par moindres carrés.	Calcul des résidus $W(x_\alpha)$ aux points expérimentaux. Modèle de variogramme ajusté sur le variogramme expérimental de W.
Cas particulier du krigeage du résidu : krigeage des innovations	Relation linéaire globale entre Z et S, avec $a^*=1$ et $b^*=0$ Absence de corrélation spatiale entre S et le résidu (innovation) I	$Z(x) = S(x) + I(x)$ où $I(x)$ est l'innovation.	Calcul des innovations $I(x_\alpha)$ aux points expérimentaux. Modèle de variogramme ajusté sur le variogramme expérimental de I.
Krigeage des observations avec CHIMERE en dérive externe	Relation linéaire locale (dans un voisinage de x) entre Z et S Absence de corrélation spatiale entre S et le résidu R	$Z(x) = a_x.S(x) + b_x + R(x)$	Modélisation non stationnaire indirecte (par validation croisée). En première approximation : <i>modèle</i> de variogramme ajusté sur le variogramme expérimental de W (cf. krigeage du résidu)
Cokrigeage ordinaire (à moyenne inconnue) entre les observations (variable principale) et CHIMERE (variable secondaire)	Cas du modèle linéaire de corégionalisation :		
	Les variogrammes simples (γ_Z et γ_S) et le variogramme croisé γ_{ZS} de Z et S sont des combinaisons linéaires des mêmes structures élémentaires γ^j , par exemple :	Z et S s'écrivent comme la somme de composantes d'échelle indépendantes (Z_k^i), par exemple : $Z(x) = \varepsilon(x) + Z_1^{pp}(x) + Z_1^{gp}(x)$ $S(x) = Z_2^{pp}(x) + Z_2^{gp}(x)$	Modèle linéaire de corégionalisation ajusté sur les variogrammes simples γ_Z , γ_S , et le variogramme croisé γ_{ZS}

Méthode	Hypothèses et relations sous-jacentes	Modélisation du variogramme
	γ^0 : effet de pépite, γ^{pp} : structure exponentielle de petite portée, γ^{gp} : structure exponentielle de grande portée. avec: $\gamma^{\varepsilon}(h) \propto \gamma^0(h)$ (erreur de mesure) $\gamma^{Z_1^{pp}}(h) \propto \gamma^{pp}(h)$; $\gamma^{Z_2^{pp}}(h) \propto \gamma^{pp}(h)$ $\gamma^{Z_1^{gp}}(h) \propto \gamma^{gp}(h)$; $\gamma^{Z_2^{gp}}(h) \propto \gamma^{gp}(h)$	

Tableau 2 – Géostatistique multi variable appliquée à la réalisation de cartes analysées : estimation

Méthode	Formule d'estimation	Poids de krigeage	Données d'entrée
Krigeage du résidu	$Z^{KR}(x_j) = a^* \cdot S(x_j) + b^* + W^{KO}(x_j)$ $= S(x_j) + \sum_{x_\alpha \in V_j} \lambda_\alpha^{KO} \cdot W(x_\alpha)$ avec $W(x_\alpha) = Z(x_\alpha) - [a^* \cdot S(x_\alpha) + b^*]$	$\sum_{V_j} \lambda_\alpha^{KO} = 1$	Données d'observation aux points x_α Données de simulation aux points d'observation x_α et au point d'estimation x_j .
Cas particulier du krigeage du résidu : krigeage des innovations	$Z^{KI}(x_j) = S(x_j) + I^{KO}(x_j)$ $= S(x_j) + \sum_{x_\alpha \in V_j} \lambda_\alpha^{KO} \cdot I(x_\alpha)$ avec $I(x_\alpha) = Z(x_\alpha) - S(x_\alpha)$	$\sum_{V_j} \lambda_\alpha^{KO} = 1$	Données d'observation aux points x_α Données de simulation aux points d'observation x_α et au point d'estimation x_j .
Krigeage des observations avec CHIMERE en dérive externe	Estimation directe à partir des données expérimentales : $Z^{KDE}(x_j) = \sum_{x_\alpha \in V_j} \lambda_\alpha^{KDE} \cdot Z(x_\alpha)$	$\begin{cases} \sum_{V_j} \lambda_\alpha^{KDE} = 1 \\ S(x_j) = \sum_{V_j} \lambda_\alpha^{KDE} \cdot S(x_\alpha) \end{cases}$	Données d'observation aux points x_α Données de simulation aux points d'observation x_α et au point d'estimation x_j .

Méthode	Formule d'estimation	Poids de krigeage	Données d'entrée
Cokrigeage ordinaire (à moyenne inconnue) entre les observations (variable principale) et CHIMERE (variable secondaire)	Estimation directe à partir des données expérimentales : $Z^{CK}(x_j) = \sum_{x_\alpha \in V_j^Z} \lambda_\alpha^{CK} \cdot Z(x_\alpha) + \sum_{x_\beta \in V_j^S} \lambda_\beta^{CK} \cdot S(x_\beta)$ V_j^Z : voisinage d'estimation pour la variable Z V_j^S : voisinage d'estimation pour la variable S	$\begin{cases} \sum_{V_j^Z} \lambda_\alpha^{CK} = 1 \\ \sum_{V_j^S} \lambda_\beta^{CK} = 0 \end{cases}$	Données d'observation aux points x_α Données de simulation aux points de la grille CHIMERE (x_β)

Remarque : le voisinage d'estimation V_j^Z (V_j^S) autour d'un point d'estimation x_j représente l'ensemble des points de données de Z (de S) utilisés en entrée du krigeage. Le choix du voisinage est un point délicat. L'extension du voisinage doit tenir compte de la portée spatiale du phénomène et s'accorder avec le domaine de validité du variogramme. Dans le cas de variables denses, la définition du voisinage est également contrainte par les ressources informatiques (nécessité de limiter le voisinage à cause des temps de calcul).

3. APPLICATION DES METHODES A LA CARTOGRAPHIE HORAIRE D'OZONE ET EVALUATION

3.1 PERIODE D'ETUDE

Deux périodes estivales ont été sélectionnées :

- une période de quatre semaines déjà considérée dans l'étude LCSQA 2008, elle s'étend du 18 juillet au 14 août 2006.
- une période plus récente (août 2009).

Ce choix résulte d'un compromis entre la nécessité de travailler sur des durées suffisamment longues et celle de pouvoir multiplier les tests dans un temps de calcul raisonnable.

Pour simplifier, nous appellerons ces périodes *été 2006* et *été 2009*.

3.2 VARIABLES DE CONCENTRATION

Les données suivantes ont été exploitées :

- simulations horaires du champ d'ozone sur la France issues du modèle CHIMERE (résolution $0.1^\circ \times 0.15^\circ$, version V200709B pour les étés 2005-2006, version V2008B pour l'été 2009) ;
- données horaires de concentration d'ozone extraites de la BDQA (étés 2005-2006) ou de BASTER (été 2009).

3.3 SELECTION DES STATIONS DE MESURE

Après consultation des AASQA, 219 stations de mesure (Figure 1) ont été sélectionnées parmi les 447 sites de mesure d'ozone recensés pour 2005-2006 (sites de proximité inclus). Il s'agit de stations dont la représentativité spatiale pour l'ozone est jugée comparable aux dimensions de la maille de CHIMERE (soit une dizaine de kilomètres environ). Dans le rapport, nous les nommerons **stations d'analyse**.

Pour l'été 2009, cette liste a été mise à jour afin de prendre en compte les modifications du réseau de mesure. Elle comprend 234 sites (Figure 1).

Les typologies de ces stations et le nombre de stations par typologie sont présentés dans le tableau ci-après (Tableau 3) :

Tableau 3 – Répartition par typologie des stations sélectionnées pour l'analyse

Typologie	Etés 2005-2006	Eté 2009
rurale nationale	10	12
rurale régionale	48	53
périurbaine	75	80
urbaine	78	82
industrielle	6	6
observation spécifique	2	1

Les stations périurbaines et urbaines qui ne font pas partie de cet ensemble seront appelées **stations de vérification**. Leurs mesures serviront de données indépendantes pour l'évaluation des méthodes.

Remarque : du fait de leur spécificité mise en évidence par le travail de 2008, certaines stations d'analyse ne seront pas utilisées dans la modélisation géostatistique : il s'agit des stations de montagne et de quelques stations situées en bord de mer. En revanche, elles seront prises en compte en entrée du krigeage.

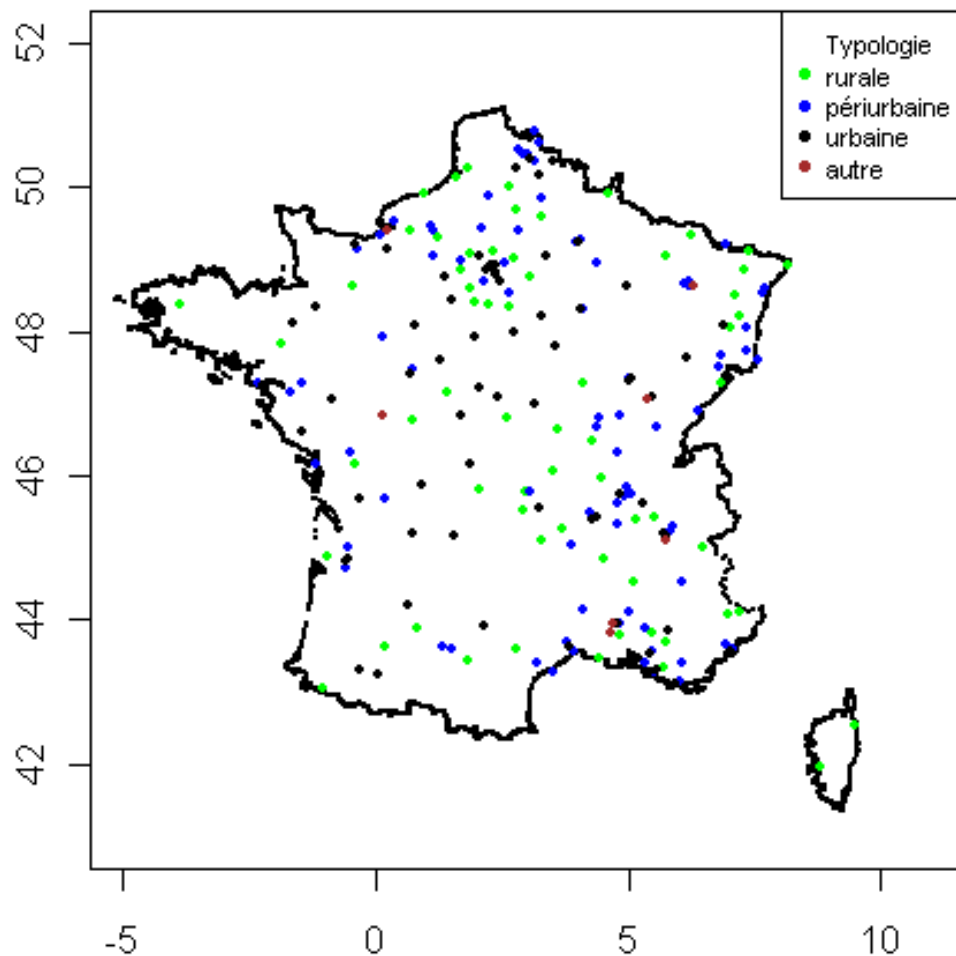


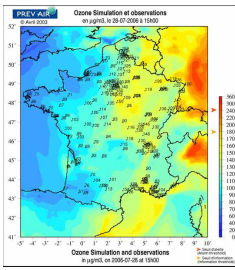
Figure 1 – Stations de mesure sélectionnées pour l'analyse en 2009.

3.4 CHAÎNE DE CALCUL

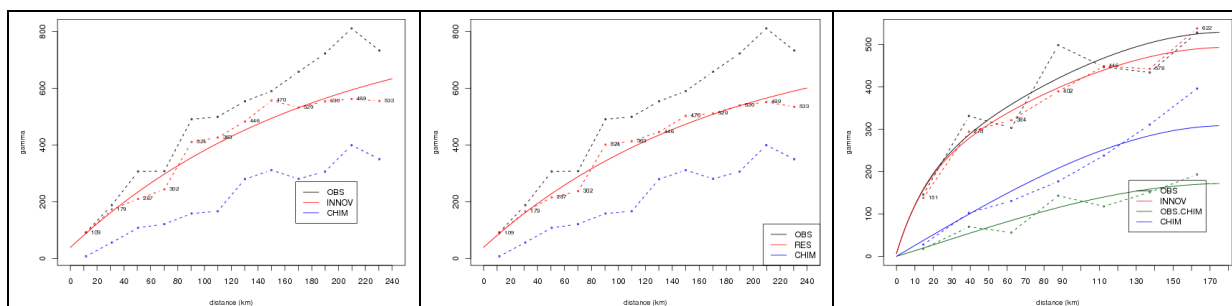
Tous les calculs géostatistiques ont été réalisés avec le logiciel R (module *gstat*, Rossiter, 2007 ; Pebesma, 2004) selon le mode opératoire ci-après (Tableau 4). Les fonctions utilisées sont indiquées en italique.

Les simulations de CHIMERE de l'heure *h* sont corrigées par les données de mesure correspondant à cette même heure. Du fait des contraintes liées à un usage automatique et opérationnel des méthodes, on n'a pas cherché à prendre en compte les éventuels décalages temporels entre les variables (avance ou retard de CHIMERE).

Tableau 4 – Réalisation de cartes analysées. Déroulement des calculs.

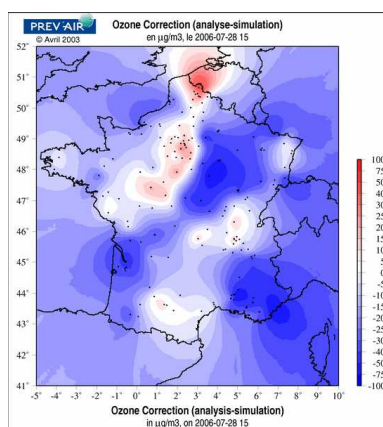
Récupération des données de CHIMERE pour la date considérée			
			
Récupération des données de la BDQA ou de BASTER pour la date considérée			
Ouverture de R			
Chargement des données et interpolation linéaire de CHIMERE aux stations de mesure			
Krigeage des innovations	Krigeage avec dérive externe⁶	Cokrigeage	
Calcul des innovations expérimentales	Estimation par moindres carrés d'une dérive globale de la forme $a \cdot \text{CHIMERE} + b$		
Calcul du variogramme expérimental des innovations (<i>variogram</i>) et ajustement d'un modèle (<i>fit.variogram</i>) (effet de pépité + 1 structure exponentielle de portée et de palier ajustés automatiquement)	Calcul du variogramme expérimental des résidus (<i>variogram</i>) et ajustement d'un modèle (<i>fit.variogram</i>) (effet de pépité + 1 structure exponentielle de portée et de palier ajustés automatiquement)	Calcul des variogrammes simples des observations et de CHIMERE et de leur variogramme croisé (<i>variogram</i>) et ajustement d'un modèle linéaire de corégionalisation (<i>fit.lmc</i>) (effet de pépité pour le variogramme des observations + 2 structures exponentielle et sphérique de portées fixées et de paliers ajustés automatiquement)	

⁶ En première approche, le variogramme défini pour le krigeage avec dérive externe a été modélisé à partir d'une dérive globale estimée par moindres carrés. Pour une modélisation indirecte par validation croisée, des développements informatiques sont nécessaires.

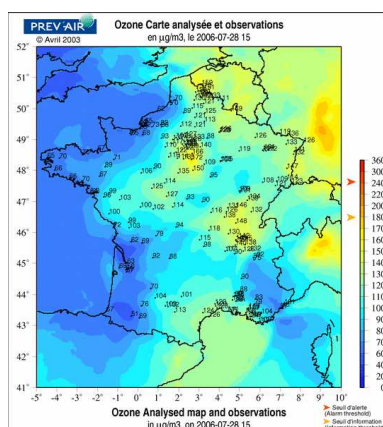


Validation croisée

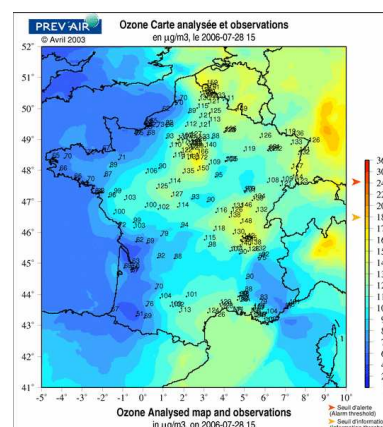
Estimation ponctuelle des innovations sur la grille en voisinage glissant⁷
(predict.gstat)



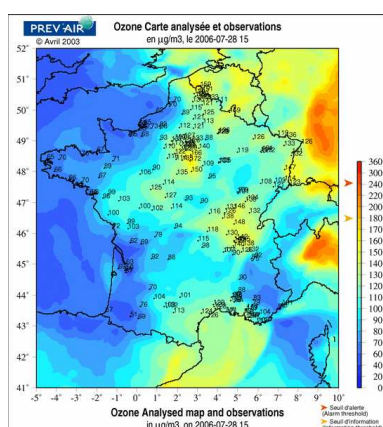
Estimation ponctuelle des concentrations sur la grille en voisinage glissant⁷
(predict.gstat)



Estimation ponctuelle des concentrations sur la grille en voisinage glissant⁸
(predict.gstat)



Ajout du modèle CHIMERE pour obtenir l'estimation finale :



Evaluation sur les stations de vérification

⁷ Pour que les données d'entrée du krigeage soient suffisamment nombreuses quelle que soit la position du point d'estimation, il s'est révélé préférable de définir le voisinage comme un nombre de stations (choisi ici égal à 80) plutôt que comme un rayon autour du point.

⁸ Le voisinage est défini pour les données de mesure d'une part, pour CHIMERE d'autre part. Pour les observations, il s'élève comme précédemment à 80 points. Pour CHIMERE, il vaut 289 (17x17) points.

3.5 CONTROLE DES HYPOTHESES

Les hypothèses propres à chaque méthode (cf. §) ont été contrôlées en examinant heure par heure les variogrammes simples et croisés des observations, de CHIMERE et des variables résiduelles. De cette analyse ressortent les conclusions suivantes :

- Krigeage des innovations : les hypothèses théoriques ne sont pas systématiquement vérifiées. Assez souvent, l'innovation se révèle corrélée spatialement à CHIMERE (leur variogramme croisé n'est pas nul) et le variogramme des observations ne s'écrit pas comme la somme du variogramme de CHIMERE et de celui de l'innovation (Figure 2a et b).
- Krigeage avec dérive externe : les hypothèses sont plus difficilement contrôlables du fait que le résidu vrai est inaccessible. Le résidu estimé à partir d'une dérive globale est peu ou non corrélé spatialement à CHIMERE (voir leur variogramme croisé, Figure 2c et d).
- Cokrigeage : dans l'ensemble, le modèle linéaire de corégionalisation décrit correctement les relations de variabilité entre les observations et CHIMERE. On peut s'en assurer en déduisant du modèle ajusté le variogramme des innovations (cf. rapport LCSQA 2008) : celui-ci concorde généralement avec le variogramme expérimental des innovations (Figure 2e). Il arrive toutefois que pour certaines heures, les types de structures ou valeurs de portées prédéfinis pour la modélisation automatique ne correspondent pas aux données expérimentales: l'ajustement du modèle est alors moins précis (Figure 2f).

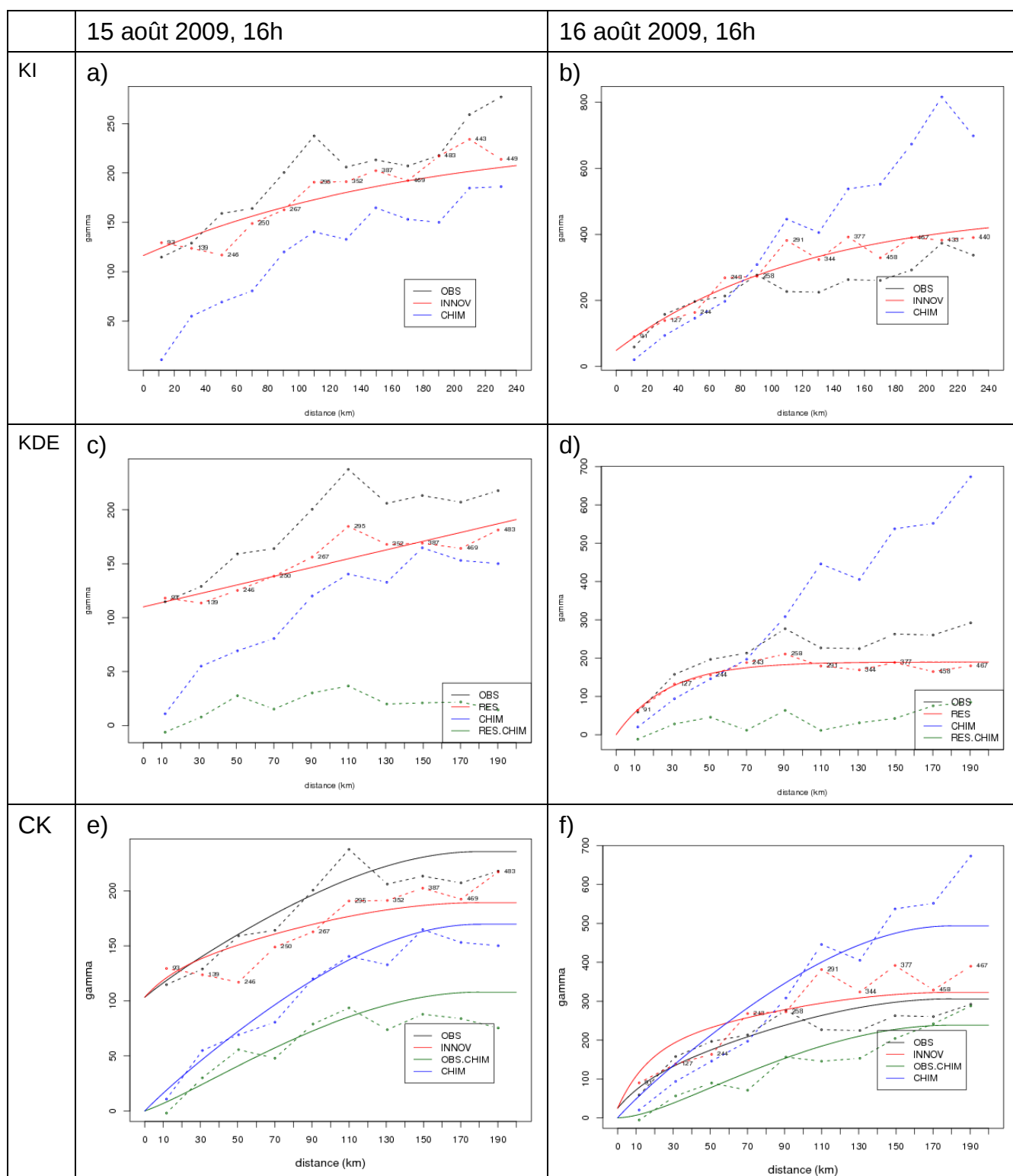


Figure 2 – Exemples de variogrammes. En pointillés : variogrammes expérimentaux ; en trait plein : variogrammes ajustés. KI : krigeage des innovations ; KDE : krigeage avec dérive externe ; CK : cokrigeage.

3.6 EVALUATION STATISTIQUE

La qualité de l'estimation a été évaluée :

- par validation croisée sur les stations d'analyse ;
- par validation sur les stations de vérification.

Nous nous limitons à la présentation des résultats de validation.

Pour chaque méthode (simulation déterministe brute, krigeage des innovations, krigeage avec dérive externe et cokrigeage) et chaque période (2006 et 2009), différentes statistiques d'erreurs ont été calculées :

- station par station,
- par typologie,
- et sur l'ensemble des stations.

En moyenne sur la période et les stations, les trois méthodes réduisent significativement l'écart entre le modèle et les mesures (Figure 3, Figure 4, Tableau 5, Tableau 6, et annexe 5), y compris pour les sites urbains : l'erreur est réduite, la corrélation augmente.

Tableau 5 – Été 2006 (18 juillet-14 août). Validation sur les stations de vérification périurbaines et urbaines. Statistiques d'erreurs et corrélation toutes heures confondues.

Méthode	Typologie	N	biais ^a [µg/m ³]	RMSE ^b [µg/m ³]	Moy_err. rel_abs ^c [%]	Coefficient de corrél.
mod	PU+U	117741	17.0	31.3	79.6	0.734
KI	PU+U	117741	2.3	15.7	31.9	0.910
KDE	PU+U	117741	2.0	14.8	31.2	0.920
CK	PU+U	117741	3.0	15.4	33.5	0.915

Tableau 6 – Été 2009 (août). Validation sur les stations de vérification périurbaines et urbaines. Statistiques d'erreurs et corrélation toutes heures confondues.

Méthode	Typologie	N	biais ^a [µg/m ³]	RMSE ^b [µg/m ³]	Moy_err. rel_abs ^c [%]	Coefficient de corrél.
mod	PU+U	114784	21.4	32.6	111.2	0.712
KI	PU+U	114784	2.5	15.4	43.5	0.888
KDE	PU+U	114784	1.7	14.6	42.5	0.899
CK	PU+U	114784	3.2	15.0	47.4	0.895

*mod : CHIMERE ; KI : krigeage des innovations ; KDE : krigeage avec dérive externe ; CK : cokrigeage

a : moyenne des erreurs ; b : racine carrée de l'erreur moyenne quadratique ; c : moyenne des erreurs relatives en valeur absolue (cf. formules en annexe 5)

Des statistiques plus détaillées par typologie sont fournies dans les tableaux de l'annexe 5.

Si les statistiques d'erreurs ne diffèrent pas sensiblement selon les méthodes, en particulier lorsqu'elles se restreignent aux maximums journaliers, le krigeage avec dérive externe, qui recalcule localement la relation entre les observations et CHIMERE, a néanmoins l'avantage. Selon la période et la typologie de station, le cokrigeage ou le krigeage des innovations arrivent en seconde position (annexe 5).

Malgré l'aptitude du modèle linéaire de corégionalisation à décrire les relations entre les observations et CHIMERE, le cokrigeage n'a pas l'efficacité attendue. Plusieurs raisons pourraient en rendre compte :

- pour certains jours ou certaines heures, le modèle de corégionalisation tel qu'il a été prédéfini (structures, portées) ne s'ajuste pas bien sur les variogrammes expérimentaux ;
- un modèle unique est défini sur le domaine ; un modèle non stationnaire dont les paramètres varieraient dans l'espace pourrait être plus adapté ;
- le voisinage d'estimation utilisé pour CHIMERE est peut-être trop limité (à cause des temps de calcul, un voisinage d'environ 80 km de rayon a été adopté pour CHIMERE.)

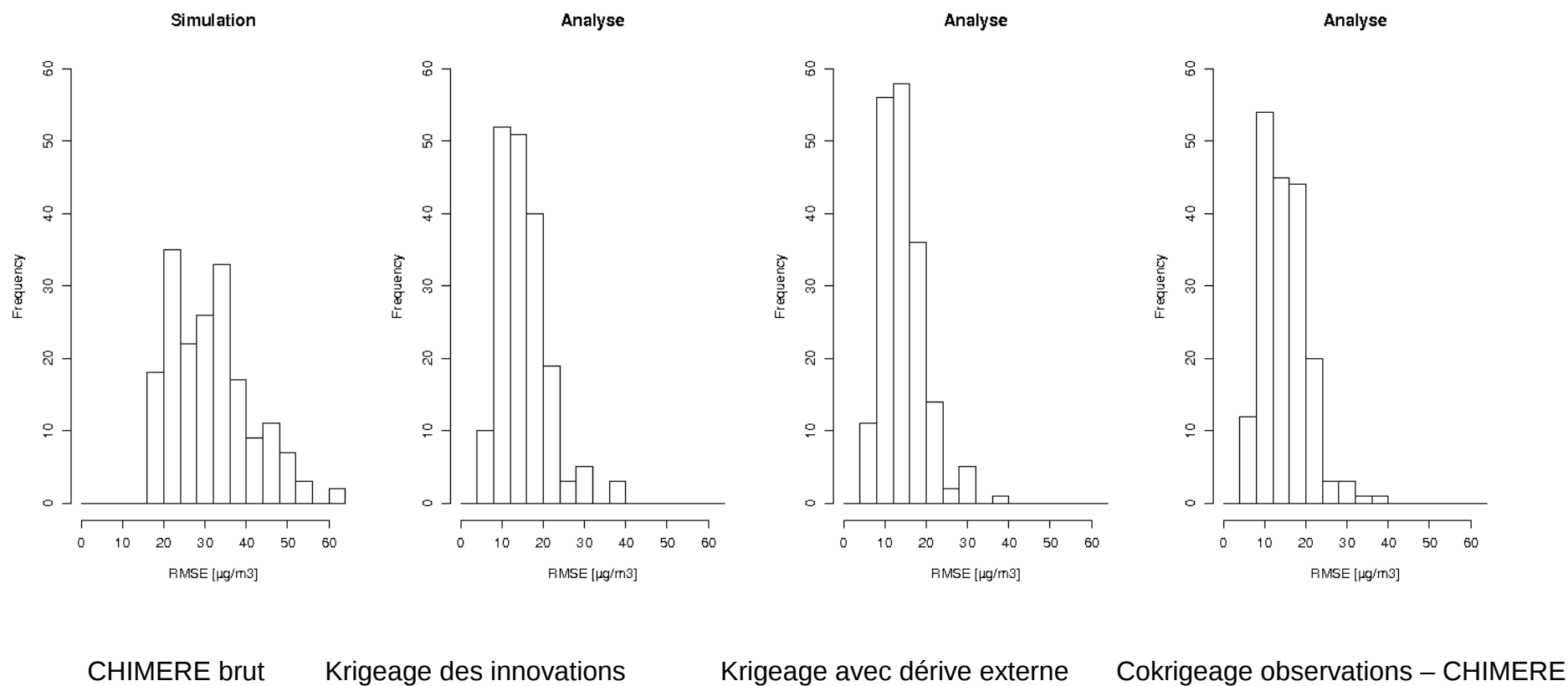


Figure 3 – Été 2009. Distribution de la **RMSE** calculée sur toute la période en chaque site de vérification périurbain ou urbain.

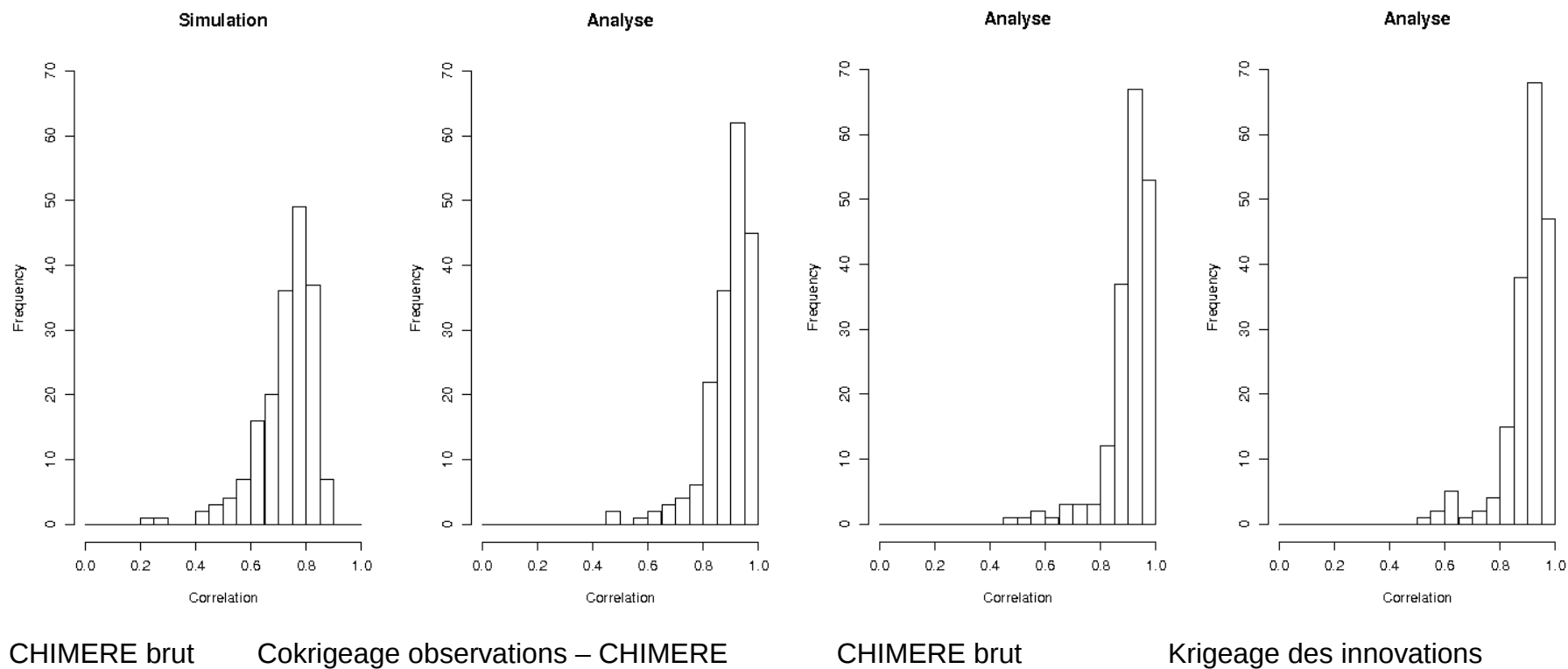


Figure 4 – Été 2009. Été 2009. Distribution de la **corrélation** calculée sur toute la période en chaque site de vérification périurbain ou urbain.

3.7 CARTOGRAPHIES

Des exemples de cartes analysées sont présentés pour deux dates (Figure 6 ; Figure 7). Dans l'enveloppe convexe des stations, les cartographies obtenues par les trois méthodes d'analyse se ressemblent. Les zones de correction positive ou négative se répartissent de façon similaire.

En revanche, hors de cette enveloppe, *i.e.* en extrapolation, les méthodes montrent des différences assez marquées. Dans ces parties du domaine, l'estimation est nettement moins précise (la variance de krigeage y est plus élevée, voir par exemple la Figure 5).

Pour évaluer la fiabilité des méthodes en extrapolation, des comparaisons seront faites avec les observations des pays frontaliers. Par la suite, des stations de ces pays pourront être utilisées comme stations d'analyse afin d'améliorer l'estimation en bordure du domaine.

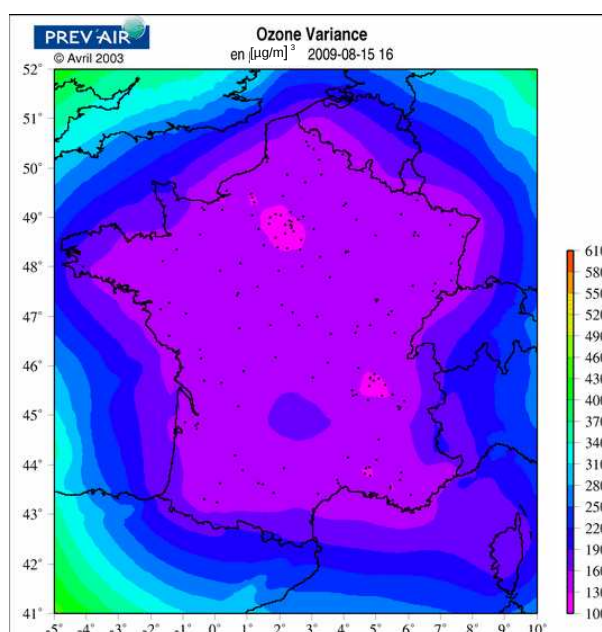


Figure 5 – Carte de la variance de l'erreur d'estimation associée au krigeage avec dérive externe. 15 août 2009, 16h.

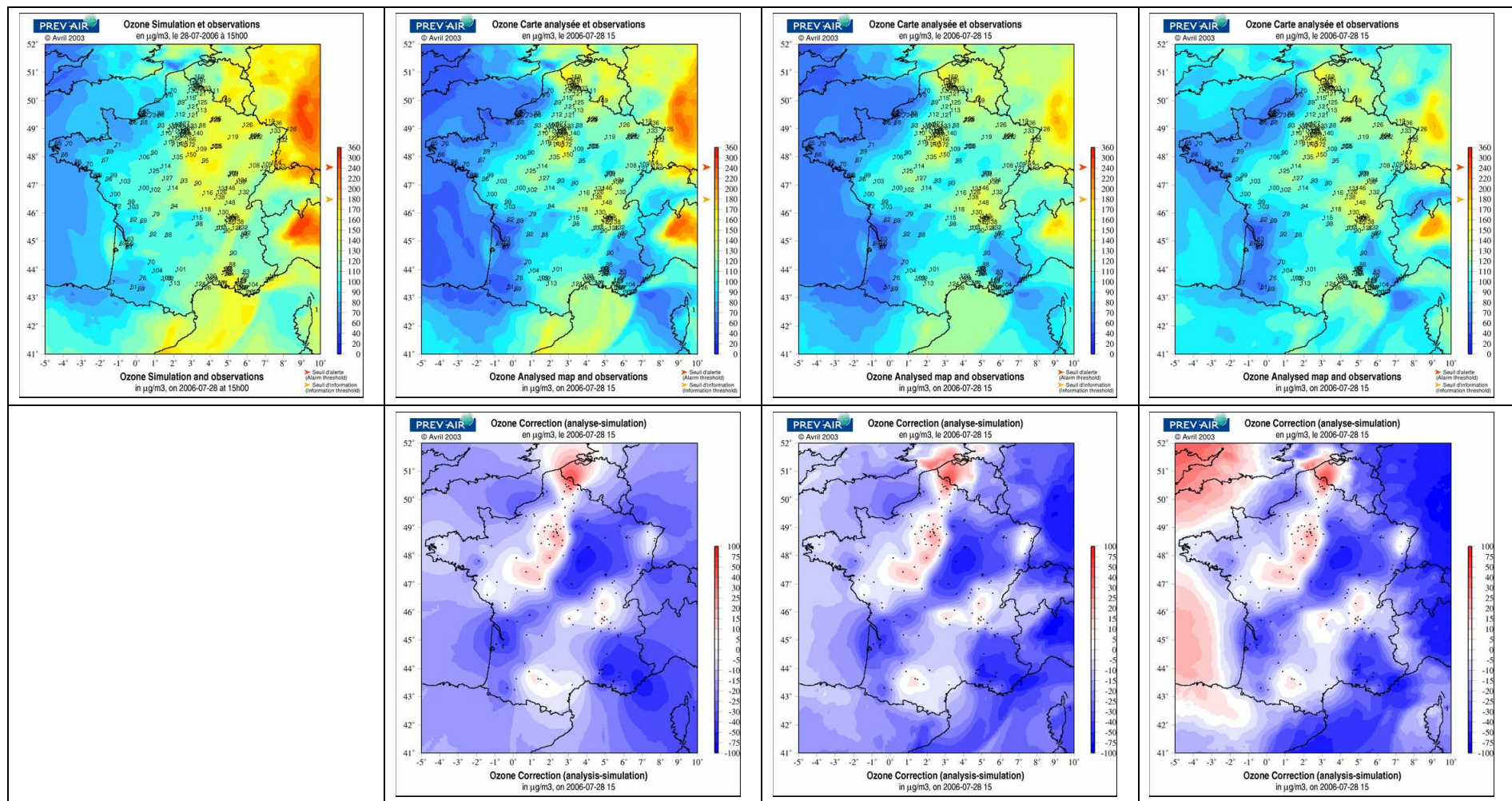


Figure 6 – 1^{ère} ligne : Cartographie des concentrations d'ozone 1) simulées par CHIMERE ; 2) après krigeage des innovations ; 3) après krigeage avec dérive externe ; 4) après cokrigeage. 28 juillet 2006, 15h.

2^{nde} ligne : différences entre les cartes analysées et la simulation initiale

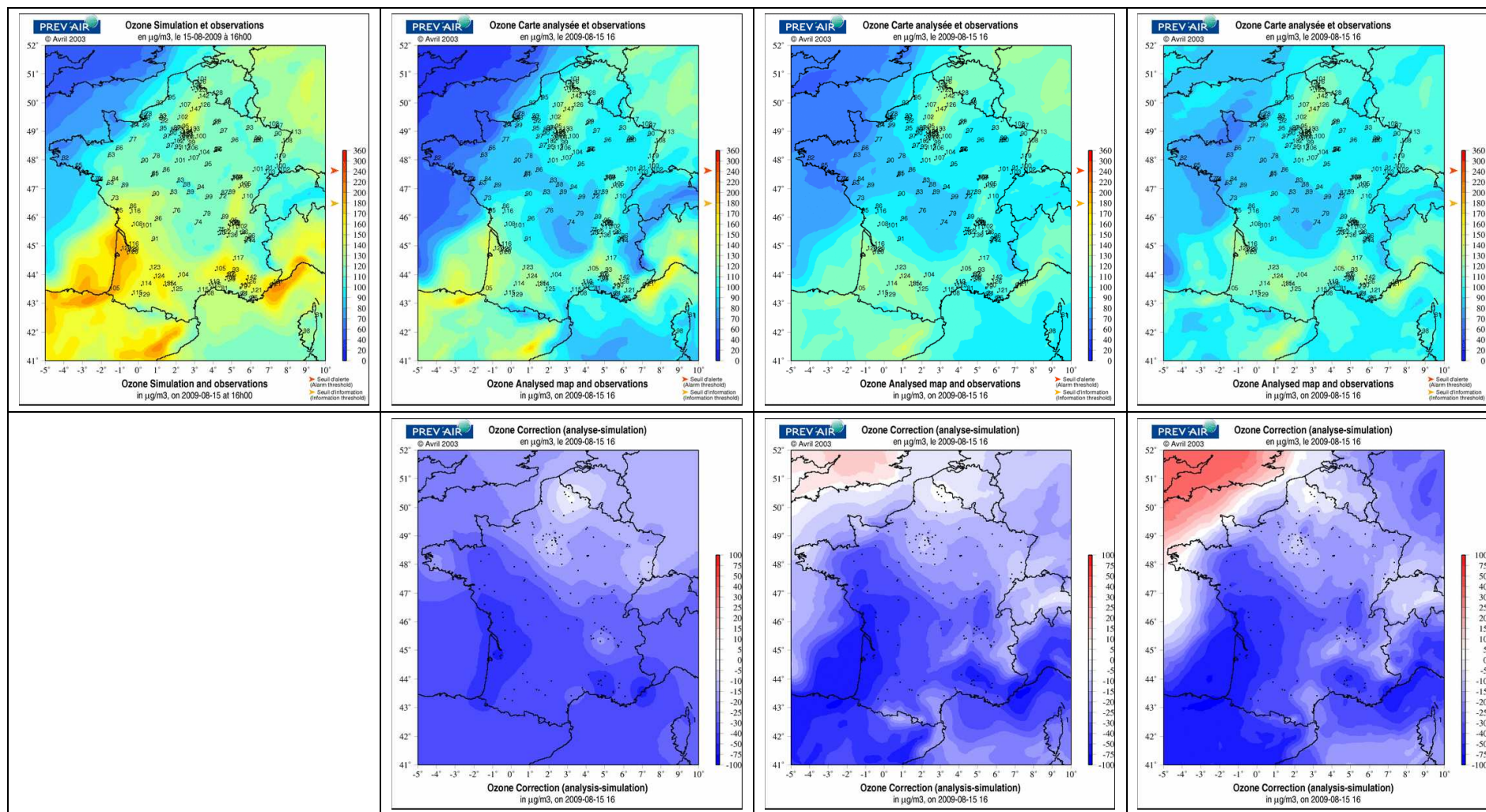


Figure 7 – 1^{ère} ligne : Cartographie des concentrations d'ozone 1) simulées par CHIMERE ; 2) après krigeage des innovations ; 3) après krigeage avec dérive externe ; 4) après cokrigeage. 15 août 2009, 16h.

2^{nde} ligne : différences entre les cartes analysées et la simulation initiale

4. CONCLUSION

Cette étude pluriannuelle a pour objet d'améliorer et mettre à jour la méthode d'analyse utilisée dans PREV'AIR. Elle prolonge le travail exploratoire conduit en 2008, dans lequel la variabilité temporelle et spatiale des observations et de CHIMERE a été plus particulièrement examinée. Cette précédente analyse montre que généralement, dans le temps comme dans l'espace, le modèle géostatistique appelé *modèle linéaire de corégionalisation* permet de rendre compte des relations de variabilité entre les mesures et les simulations. Ce type de modèle peut être utilisé dans un cokrigage, technique géostatistique d'estimation multi variable. En 2008, des tests préliminaires de cokrigage entre les observations et le modèle CHIMERE ont été ainsi effectués afin de corriger spatialement ce dernier ; ils n'ont cependant débouché sur aucune conclusion ferme, certains paramètres de calcul nécessitant des réglages complémentaires.

En 2009, de nouveaux essais de cokrigage ont été réalisés sur deux périodes d'environ un mois, extraites des étés 2006 et 2009. Le krigeage des innovations et le krigeage avec dérive externe ont été également appliqués pour comparaison. A l'occasion de ces calculs, un bilan théorique des hypothèses propres à chaque méthode a été établi. Le krigeage avec dérive externe et le cokrigage reposent sur des modèles plus généraux que le krigeage des innovations : le premier permet un ajustement linéaire local de CHIMERE ; le second différencie la relation observations-CHIMERE par échelle de variabilité spatiale.

Les concentrations horaires estimées par chaque méthode sur des stations de vérification indépendantes (stations périurbaines et urbaines) ont été comparées aux valeurs mesurées. Les trois méthodes produisent des résultats proches, en particulier pour le pic journalier. Les statistiques d'erreurs accordent cependant un petit avantage au krigeage avec dérive externe. Cette méthode pourra être mise en œuvre dans PREV'AIR en 2010 en remplacement du krigeage des innovations.

En ce qui concerne les PM₁₀, le travail a été effectué pour l'évaluation des zones de dépassement de valeurs limites (Malherbe et al., 2009). Comme pour l'ozone, le modèle linéaire de corégionalisation décrit de manière satisfaisante les relations de variabilité entre les observations et CHIMERE. Il a été utilisé dans un cokrigage pour cartographier les concentrations journalières de PM₁₀ sur la France. Dans ce contexte spécifique, la technique du cokrigage offre la possibilité d'estimer directement les concentrations sur une grille kilométrique,

- à partir des données de mesure issues des stations de fond rurales, périurbaines ou urbaines ;
- à partir des données de simulation disponibles sur une grille plus lâche (0,10 x 0,15°) ;
- en introduisant une ou plusieurs variables auxiliaires en dérive.

Ces travaux seront mis à profit pour améliorer les cartes analysées de PM₁₀ dans PREV'AIR.

5. REFERENCES

Chilès J.-P., Séguret S., 2008. Geostatistical analysis of validation data of an air pollution simulator. In GEOSTATS 2008 – VIII International Geostatistics Congress, 1-5 December, Santiago, Chile, Editors: Julián M. Ortiz and Xavier Emery, Mining Engineering Department, University of Chile.

Malherbe L., Cárdenas G., 2009. Evaluation des zones géographiques touchées par les dépassements de valeurs limites. Rapport LCSQA n°DRC-10-103355-00551A, www.lcsqa.org.

Malherbe L., Bessagnet B., Létinois L., Rouïl L., 2009. Rencontre Statistique et pollution : problèmes et méthodes. Exposé disponible à l'adresse suivante : <http://www.sfds.asso.fr/93-Activites>.

Pebesma E. J., 2004. Multi variable geostatistics in S: the gstat package. Computers & Geosciences, 30, 683–691.

Rossiter D. G., 2007. Technical note. Co-kriging with the gstat package of the R environment for statistical computing. Version 2.1, January 2007, 81 pp. <http://www.itc.nl/personal/rossiter>.

6. LISTE DES ANNEXES

Repère	Désignation	Nombre de pages
Annexe 1	Fiche descriptive de l'étude	4
Annexe 2	Méthodes d'estimation en géostatistique multi variable	7
Annexe 3	Compléments théoriques	2
Annexe 4	Prise en compte de variables auxiliaires	4
Annexe 5	Résultats de validation	6

ANNEXE 1

Fiche descriptive de l'étude

THEME 6 : Modélisation – traitements numériques

ETUDE N° 6/3 : TRAVAUX RELATIFS A LA PLATE FORME NATIONALE DE MODELISATION PREV'AIR

Responsable de l'étude : INERIS

Objectif

L'objectif de cette étude est de réaliser les actions de support technique aux utilisateurs du système PREV'AIR et les développements nécessaires en vue de répondre à leurs besoins futurs.

Contexte et travaux antérieurs

Le système PREV'AIR de prévision et de cartographie de la qualité de l'air a été mis en place au cours de l'été 2003 par l'INERIS, en collaboration avec le CNRS et l'ADEME. Une convention de développement et d'exploitation du système a également été signée en juin 2004 entre le MEEDDAT, l'INERIS, l'ADEME, le CNRS et Météo France.

Outre la diffusion par internet d'observations en temps réel, de prévisions et cartographies d'ozone, de particules (PM₁₀ et PM_{2.5}) et d'oxydes d'azote, les données numériques de concentrations calculées par CHIMERE sont téléchargeables par les organismes qui en font la demande. A ce jour presque toutes les AASQA disposent d'un compte sur PREV'AIR et accèdent chaque jour aux prévisions et analyses calculées à l'INERIS.

Travaux proposés pour 2009

Assistance utilisateurs

Le système PREV'AIR fournit notamment une prévision de l'ozone, des oxydes d'azote et des particules sur l'Europe et la France, à partir des résultats issus du modèle déterministe CHIMERE appliqué à une résolution spatiale de 50 et 10 km respectivement. Certaines AASQA disposent d'un modèle déterministe régional de prévision de qualité de l'air auquel PREV'AIR fournit des conditions aux limites. D'autres AASQA utilisent les sorties brutes de PREV'AIR qu'elles traitent en fonction de leurs besoins.

L'INERIS fournit une assistance aux utilisateurs de PREV'AIR en terme d'appui technique pour l'utilisation et l'interprétation des données issues des modèles ; l'ensemble de ces actions d'assistance pourront faire l'objet, le cas échéant, d'une session de présentation des modèles, leurs apports et leurs limites, auprès des AASQA.

L'INERIS met également à disposition l'ensemble des outils de post-traitement qu'il développe. Il répond à toute demande ponctuelle d'extraction de données. Enfin l'INERIS propose et développe des représentations d'indicateurs cohérentes avec les besoins exprimés par les AASQA (séries temporelles, cartographies nationales....).

Bilan des performances des modèles

Les modèles CHIMERE et MOCAGE sont les deux modèles dont les résultats sont disponibles sur PREV'AIR à ce jour. Après la période estivale, l'INERIS s'est engagé à fournir un rapport détaillé sur le bilan de performances de ces deux modèles région par région. Le bilan est établi par comparaison aux observations selon des critères statistiques classiques. Cette information est disponible à la fin du premier semestre de chaque année pour les résultats relatifs à l'été précédent.

Modélisation dans les DOM

Des prévisions de concentrations en polluants atmosphériques sur les départements et territoires d'outremer peuvent être élaborées en utilisant les versions globales des modèles CHIMERE et MOCAGE qui ont une résolution de 1°. Ces codes permettent en particulier de détecter les épisodes de poussières qui sont à l'origine de fréquents dépassements des seuils réglementaires de PM dans ces zones géographiques. En 2009, il est proposé d'analyser les performances du modèle CHIMERE-Dust sur ces zones en regard des données de mesure des AASQA. De plus nous examinerons la situation pour les autres polluants et évaluerons la possibilité d'affiner les prévisions pour les autres polluants en utilisant les données du futur Inventaire National Spatialisé qui couvre les départements et territoires d'outremer.

Réalisation de cartes analysées d'ozone et de PM₁₀ : évaluation, amélioration, test de nouvelles procédures

Depuis 2003 pour l'ozone et 2005 pour les PM₁₀, les cartes de la veille simulées par le modèle CHIMERE sur la France sont corrigées chaque matin à l'aide des observations extraites de la base BASTER. En 2008, après cinq années de fonctionnement qui ont vu évoluer aussi bien le modèle CHIMERE que le réseau des stations de mesure, il a paru utile de réévaluer le choix des stations prises en compte dans l'analyse et la méthode de correction mise en œuvre (krigeage des innovations).

Dans un premier temps, nous avons cherché à mieux appréhender les limites de la méthode de correction spatiale, ce qui nous a conduit à analyser en profondeur les différences entre CHIMERE et les mesures. Dans un second temps, nous avons exploré une nouvelle approche qui tient mieux compte des relations effectives entre CHIMERE et les mesures. La méthode mise en place pour l'ozone a soulevé quelques difficultés techniques qu'il faudra achever de résoudre en 2009 avant application aux PM₁₀. Des tests de validation complémentaires sont également prévus pour l'ozone et les PM₁₀ à la suite desquels la méthode sera introduite dans PREV'Air. Enfin, la pertinence d'une approche spatio-temporelle plus développée et le gain qui pourrait en résulter seront discutés.

Renseignements synthétiques

Titre de l'étude	Travaux relatifs à la Plate-forme nationale de modélisation PREV'AIR		
Personne responsable de l'étude	Frédéric Meleux		
Travaux	pérennes		
Durée des travaux pluriannuels			
Collaboration AASQA	toutes		
Heures d'ingénieur	EMD :	INERIS : 650	LNE :
Heures de technicien	EMD :	INERIS :	LNE :
Document de sortie attendu	Bilan de performance des modèles région par région Document sur la généralisation du système aux DOM-TOM Rapport sur la cartographie du NO2 Rapport sur l'analyse Mise à jour du site/ hot line		
Lien avec le tableau de suivi CPT	Comité utilisateurs PREV'AIR		
Matériel acquis pour l'étude			

ANNEXE 2

Méthodes d'estimation en géostatistique multi variable

Soient Z une variable aléatoire observée sur un ensemble de points et y (ou Y , selon qu'elle est traitée comme déterministe ou aléatoire) une variable connue de façon dense sur le domaine d'étude.

L'objectif est d'estimer Z sur l'ensemble de la zone à partir des valeurs expérimentales de Z et des données de y .

La géostatistique propose plusieurs méthodes (krigeage du résidu, krigeage avec dérive externe, cokrigeage) dont nous détaillons ci-après les hypothèses, la démarche, les avantages et les limites.

Dans toute la suite, les équations sont présentées pour un instant t_0 donné : $t_0=j/m/a$ h:m pour une cartographie horaire ; $t_0=j/m/a$ pour une cartographie moyenne journalière.

Les points de mesure de la variable Z sont notés x_α .

Les points de la grille d'estimation (centres de mailles) sont notés x_j .

1. KRIGEAGE DU RESIDU

1.1 MODELE

Hypothèse : la moyenne m de la variable Z est expliquée par une variable déterministe y connue exhaustivement sur le domaine d'étude :

$$\begin{aligned} Z(x) &= m(x) + T(x) \\ &= a y(x) + b + T(x) \end{aligned}$$

T , résidu « vrai », est d'espérance nulle : $E[T(x)]=0$. D'où : $E[Z(x)]=ay(x)+b$

En pratique, la variable y est supposée connue aux points expérimentaux x_α (stations de mesure) et aux points de grille x_j . Le modèle précédent est remplacé par :

$$Z(x) = a^*y(x) + b^* + W(x)$$

où a^* et b^* sont déterminés à partir des données expérimentales de Z et y aux points x_α , par exemple par moindres carrés.

Sur l'ensemble des x_α , W , résidu « expérimental », vérifie $E[W(x)]=0$.

W est sans corrélation spatiale⁹ avec la variable dense y .

Ce modèle est appelé **modèle à résidu**.

⁹ Pour la notion de corrélation spatiale, se reporter à l'annexe 3.

1.2 ESTIMATION

Le modèle $Z(x) = a^*y(x) + b^* + W(x)$ étant établi, la variable Z est estimée sur la grille par :

$$Z^{KR}(x_j) = a^* \cdot y(x_j) + b^* + W^{KO}(x_j)$$

où $W^{KO}(x_j)$ est l'estimation par krigeage ordinaire de W sur les points de grille x_j .

Les étapes de cette estimation sont les suivantes :

- Calcul des résidus expérimentaux $W(x_\alpha) = Z(x_\alpha) - (a^*y(x_\alpha) + b^*)$;
- Calcul et modélisation du variogramme des résidus ;
- Krigeage ordinaire (à moyenne inconnue) de W aux points x_j :

$$W^{KO}(x_j) = \sum_{x_\alpha \in V_j} \lambda_\alpha^{KO} \cdot W(x_\alpha)$$

λ_α^{KO} : poids de krigeage

V_j : voisinage d'estimation déterminé en fonction de la portée et du domaine de validité du variogramme

Les poids de krigeage sont solutions d'un système d'équations permettant de satisfaire aux contraintes de **non biais** et d'**optimalité** (variance d'erreur minimale). Ils vérifient en particulier la relation $\sum_{V_j} \lambda_\alpha^{KO} = 1$ (1).

- Calcul de la variance de l'erreur d'estimation $[\sigma_w^{KO}(x_j)]^2$.

1.3 INTERET ET LIMITES

La méthode est d'une mise en œuvre particulièrement aisée ; en contrepartie :

- Si les coefficients a^* et b^* de la dérive, estimés le plus souvent par moindres carrés, sont sans biais, ils ne sont généralement pas optimaux. L'optimalité nécessite en effet l'indépendance des résidus. Or, si W est spatialement corrélé, ce qui fait précisément l'intérêt du krigeage, cette condition n'est pas vérifiée.

Quel que soit le mode d'estimation de a^* et b^* , le variogramme des résidus estimés est biaisé par rapport au variogramme des résidus vrais.

- La variance de l'erreur d'estimation de Z ne se réduit pas au seul terme $[\sigma_w^{KO}(x_j)]^2$. En toute rigueur, elle devrait aussi prendre en compte l'incertitude d'estimation de la dérive $m^*(x)=a^*y(x)+b^*$. Le krigeage du résidu ne calcule pas cette seconde composante.

1.4 APPLICATION A LA CARTOGRAPHIE

Le krigeage des innovations utilisé jusqu'à présent dans PREV'AIR constitue un cas particulier du krigeage du résidu. La relation $Z(x) = a^*y(x) + b^* + W(x)$ y est remplacée par $Z(x) = CHIMERE(x) + I(x)$, où I représente l'innovation.

Ce modèle suppose implicitement que :

- $a^*=1$ et $b^*=0$;
- I est non corrélée spatialement à CHIMERE.

En un point de grille x_j , à un instant t_0 , la concentration est estimée par la formule :

$$Z^{KI}(x_j) = CHIMERE(x_j) + I^{KO}(x_j)$$

où I^{KO} est l'estimation par krigeage ordinaire de I :

$$I^{KO}(x_j) = \sum_{x_\alpha \in V_j} \lambda_\alpha^{KO} \cdot [Z(x_\alpha) - CHIMERE(x_\alpha)]$$

λ_α^{KO} : poids de krigeage

V_j : voisinage d'estimation

$Z(x_\alpha) - CHIMERE(x_\alpha)$: innovation expérimentale

2. KRIGEAGE AVEC DERIVE EXTERNE

2.1 MODELE

Comme pour le krigeage du résidu, on suppose que la variable déterministe y , connue sur tout le champ (en pratique : aux points expérimentaux x_α et aux points de grille x_j), constitue une dérive de Z , à une constante et un facteur près :

$$Z(x) = a y(x) + b + R(x)$$

Toutefois, à la différence du modèle considéré en 1.1, **les coefficients a et b ne sont pas fixés** et peuvent varier dans l'espace ($a=a(x)$; $b=b(x)$). Seule compte la linéarité de la moyenne m par rapport à la variable explicative y

R est sans corrélation spatiale¹⁰ avec la variable dense y .

2.2 ESTIMATION

La variable Z est estimée directement sur la grille par combinaison linéaire des données expérimentales :

$$Z^{KDE}(x_j) = \sum_{x_\alpha \in V_j} \lambda_\alpha^{KDE} \cdot Z(x_\alpha)$$

¹⁰ Pour la notion de corrélation spatiale, se reporter à l'annexe 3.

Les poids de krigeage λ_{α}^{KDE} doivent garantir l'absence de biais **quels que soient les coefficients inconnus a et b**. Aux conditions du krigeage ordinaire, **s'ajoute en conséquence la contrainte suivante** :

$$y(x_j) = \sum_{V_j} \lambda_{\alpha}^{KDE} \cdot y(x_{\alpha}) \quad (2)$$

Sous les contraintes (1) et (2), les poids λ_{α}^{KDE} sont calculés de façon à minimiser la variance d'erreur d'estimation.

Les étapes de cette estimation sont les suivantes :

- Modélisation du variogramme de R ;
- Krigeage de Z aux points x_j : $Z^{KDE}(x_j) = \sum_{x_{\alpha} \in V_j} \lambda_{\alpha}^{KDE} \cdot Z(x_{\alpha})$

λ_{α}^{KDE} : poids de krigeage

V_j : voisinage d'estimation

- Calcul de la variance de l'erreur d'estimation $[\sigma_Z^{KDE}(x_j)]^2$.

2.3 INTERET ET LIMITES

Comparé au krigeage du résidu, le krigeage avec dérive externe offre plusieurs avantages:

- En voisinage glissant, **la relation entre Z et y s'ajuste localement** ;
- Les coefficients $a(x)$ et $b(x)$ inconnus sont filtrés par la contrainte de non biais ;
- L'incertitude sur $m(x)$ est prise en compte dans la variance de krigeage.

Le point délicat de la méthode réside dans la modélisation du variogramme des résidus. Les coefficients a et b n'étant pas fixés, les résidus $R(x)$ sont inaccessibles, y compris aux points expérimentaux. Le variogramme expérimental de R est donc lui aussi inaccessible.

Dans la pratique, différentes méthodes sont utilisées pour définir un modèle de variogramme :

- En première approximation, le modèle est défini comme pour un krigeage du résidu : 1) estimation d'une dérive globale de la forme $a^* \cdot y(x) + b^*$ (généralement par moindres carrés) ; 2) ajustement d'un modèle de variogramme sur le variogramme expérimental des résidus $Z(x_{\alpha}) - (a^* \cdot y(x_{\alpha}) + b^*)$.
- Plus rigoureusement, un modèle non stationnaire est ajusté de manière indirecte (validation croisée).

2.4 APPLICATION A LA CARTOGRAPHIE

Considérant la relation $Z(x) = a.CHIMERE(x) + b + R(x)$

en un point de grille x_j , à un instant t_0 , la concentration est estimée par la formule :

$$Z^{KDE}(x_j) = \sum_{x_\alpha \in V_j} \lambda_\alpha^{KDE} \cdot Z(x_\alpha)$$

avec la condition : $CHIMERE(x_j) = \sum_{x_\alpha \in V_j} \lambda_\alpha^{KDE} \cdot CHIMERE(x_\alpha)$

3. COKRIGEAGE

3.1 MODELE

La variable dense, précédemment notée y , est considérée comme une variable aléatoire Y . On se place dans la situation où **Z et Y sont spatialement corrélées**¹¹. Y est supposée connue sur tout ou partie des points de mesure x_α (pour la construction du modèle variographique) et sur un ensemble de points $\{x_\beta\}$ qui coïncide ou non avec la grille d'estimation.

Deux modèles sont fréquemment utilisés pour décrire les relations entre deux variables : le modèle de corrélation intrinsèque (Chilès et Séguret, 2008) et le modèle linéaire de corégionalisation¹¹.

Dans l'hypothèse du modèle linéaire de corégionalisation, qui généralise le modèle de corrélation intrinsèque, les relations entre Z et Y s'écrivent de la façon suivante :

$$\begin{cases} Z(x) = Z_1^1(x) + Z_1^2(x) + \dots + m_Z \\ Y(x) = Z_2^1(x) + Z_2^2(x) + \dots + m_Y \end{cases}$$

où $Z_1^i(x)$ et $Z_2^i(x)$ $\{i=1,2,\dots\}$ sont des composantes de variogrammes proportionnels à la même structure $\gamma(h)$.

3.2 ESTIMATION

La variable Z (variable principale) est estimée directement sur la grille par combinaison linéaire des données expérimentales :

$$Z^{CK}(x_j) = \sum_{x_\alpha \in V_j^Z} \lambda_\alpha^{CK} \cdot Z(x_\alpha) + \sum_{x_\beta \in V_j^Y} \lambda_\beta^{CK} \cdot Y(x_\beta)$$

¹¹ Se reporter à l'annexe 3.

où les poids de krigeage vérifient : $\sum_{V_j^1} \lambda_{\alpha}^{CK} = 1$ et $\sum_{V_j^2} \lambda_{\beta}^{CK} = 0$

V_1 et V_2 sont les voisinages d'estimation autour du point x_j définis respectivement pour la variable Z et la variable Y.

Les étapes de cette estimation sont les suivantes :

- Modélisation simultanée des variogrammes simples de Z et Y et de leur variogramme croisé (par exemple par ajustement d'un modèle linéaire de corégionalisation).

- Cokrigeage de Z aux points x_j : $Z^{CK}(x_j) = \sum_{x_{\alpha} \in V_j^Z} \lambda_{\alpha}^{CK} \cdot Z(x_{\alpha}) + \sum_{x_{\beta} \in V_j^Y} \lambda_{\beta}^{CK} \cdot Y(x_{\beta})$

λ_{α}^{KDE} : poids de krigeage

V_j^Z, V_j^Y : voisinages d'estimation pour les variables Z et Y

- Calcul de la variance de l'erreur d'estimation $[\sigma_Z^{CK}(x_j)]^2$.

3.3 INTERET ET LIMITES

Le cokrigeage est plus général que les méthodes précédentes. Le modèle linéaire de corégionalisation, lorsqu'il est valide, permet de **différencier la corrélation entre Z et Y en fonction de l'échelle de variabilité** (ex : petite échelle / grande échelle), ce que ne font pas les méthodes précédentes.

En revanche, si le krigeage avec dérive externe présente l'avantage d'ajuster localement la relation entre Z et Y, le modèle défini pour le cokrigeage fixe celle-ci sur l'ensemble du domaine. Une modulation locale des paramètres de ce modèle est envisageable mais plus difficile à réaliser.

Une autre difficulté de la méthode réside dans le choix des voisinages d'estimation relatifs à chacune des variables et en particulier à la variable dense Y.

On parle de cokrigeage colocalisé lorsque le voisinage utilisé pour Y est constitué des points de mesure de Z (les x_{α}) et du point d'estimation x_j . Sauf dans l'hypothèse d'un modèle à résidu, il représente une simplification non optimale du cokrigeage (perte possible d'information).

3.4 APPLICATION A LA CARTOGRAPHIE

Soit un modèle bivariable entre Z et CHIMERE.

En un point de grille x_j , à un instant t_0 , la concentration est estimée par la formule :

$$Z^{CK}(x_j) = \sum_{x_{\alpha} \in V_j^1} \lambda_{\alpha}^{CK} \cdot Z(x_{\alpha}) + \sum_{x_{\beta} \in V_j^2} \lambda_{\beta}^{CK} \cdot CHIMERE(x_{\beta})$$

ANNEXE 3

Compléments théoriques

Corrélation spatiale entre deux variables

L'existence ou l'absence de corrélation spatiale entre deux variables peut être contrôlée par le calcul du **variogramme croisé**.

Soient deux variables Z_1 et Z_2 . Le variogramme croisé est défini par :

$$\gamma_{12}(h) = \frac{1}{2} E [Z_1(x+h) - Z_1(x)] [Z_2(x+h) - Z_2(x)]$$

Il vérifie :

$$\gamma_{12}(h) = \gamma_{12}(-h)$$

$$\gamma_{12}(0) = 0$$

$\gamma_{12}(h)$ peut être négatif si Z_1 et Z_2 varient en sens inverse.

Un variogramme croisé constamment nul traduit l'absence de corrélation spatiale.

Expérimentalement, le variogramme croisé se calcule de la façon suivante:

$$\hat{\gamma}_{12}(h) = \frac{1}{2N(h)} \sum_{i,j} [z_1(s_i) - z_1(s_j)] [z_2(s_i) - z_2(s_j)]$$

où $N(h)$ est le nombre de couples de points (s_i, s_j) distants de h ($\pm$ tolérance).

Modèle de corrélation intrinsèque

Dans ce modèle, tous les variogrammes simples et croisés sont proportionnels à la même structure. Pour deux variables Z_1 et Z_2 , cela donne :

$$\begin{cases} \gamma_{Z_1}(h) = b_1 \cdot \gamma(h) \\ \gamma_{Z_2}(h) = b_2 \cdot \gamma(h) \\ \gamma_{Z_1 Z_2}(h) = b_{12} \cdot \gamma(h) \end{cases}$$

Modèle linéaire de corégionalisation

Il généralise le modèle précédent dans le cas de plusieurs composantes. Tous les variogrammes simples et croisés s'écrivent comme une combinaison linéaire des mêmes structures.

$$\begin{cases} \gamma_1(h) = b_1^1 \gamma^1(h) + b_1^2 \gamma^2(h) \\ \gamma_2(h) = b_2^1 \gamma^1(h) + b_2^2 \gamma^2(h) \\ \gamma_{12}(h) = b_{12}^1 \gamma^1(h) + b_{12}^2 \gamma^2(h) \end{cases}$$

ANNEXE 4

Prise en compte de variables auxiliaires

Dans le tableau suivant, nous prendrons l'exemple de deux variables auxiliaires notées ϕ_1 et ϕ_2 .

Tableau 7 – Géostatistique multi variable appliquée à la réalisation de cartes analysées : prise en compte de variables auxiliaires

Méthode	Hypothèses et relations sous-jacentes	Modélisation du variogramme	Données d'entrée
Krigeage des innovations	On suppose que l'écart modèle-mesure (l'innovation) s'explique à l'aide des variables auxiliaires: $Z(x) = S(x) + a_{1x} \cdot \phi_1(x) + a_{2x} \cdot \phi_2(x) + b_x + J(x)$	Modélisation non stationnaire indirecte du variogramme de J. En première approximation : 1) estimation d'une dérive globale, 2) calcul du résidu aux points expérimentaux : $J'(x_\alpha) = Z(x_\alpha) - [S(x_\alpha) + a_1^* \cdot \phi_1(x_\alpha) + a_2^* \cdot \phi_2(x_\alpha) + b^* + W'(x_\alpha)]$ où a_1^* , a_2^* et b^* sont estimés par moindres carrés par exemple 3) ajustement d'un modèle sur le variogramme expérimental du résidu.	Données d'observation aux points x_α Données de simulation, de ϕ_1 et de ϕ_2 aux points d'observation x_α et au point d'estimation x_j .
	Formule d'estimation : $Z^{KI}(x_j) = S(x_j) + \sum_{x_\alpha \in V_j} \lambda'_\alpha \cdot I(x_\alpha)$ avec les conditions supplémentaires : $\phi_1(x_j) = \sum_{x_\alpha \in V_j} \lambda'_\alpha \cdot \phi_1(x_\alpha)$ et $\phi_2(x_j) = \sum_{x_\alpha \in V_j} \lambda'_\alpha \cdot \phi_2(x_\alpha)$		
Krigeage des observations avec	$Z(x) = a_{0x} \cdot S(x) + a_{1x} \cdot \phi_1(x) + a_{2x} \cdot \phi_2(x) +$	Modélisation non stationnaire indirecte	Données d'observation aux

Méthode	Hypothèses et relations sous-jacentes	Modélisation du variogramme	Données d'entrée
CHIMERE en dérive externe	$b_x + R(x)$	du variogramme de R En première approximation : 1) estimation d'une dérive globale, 2) calcul du résidu aux points expérimentaux : $W(x_\alpha) = Z(x_\alpha) - [a_0^* \cdot S(x_\alpha) + a_1^* \cdot \phi_1(x_\alpha) + a_2^* \cdot \phi_2(x_\alpha) + b^* + W'(x_\alpha)]$ où a_0^*, a_1^*, a_2^* et b^* sont estimés par moindres carrés par exemple 3) ajustement d'un modèle sur le variogramme expérimental du résidu.	points x_α Données de simulation, de ϕ_1 et de ϕ_2 aux points d'observation x_α et au point d'estimation x_j.
	Estimation directe à partir des données expérimentales : $Z^{KDE}(x_j) = \sum_{x_\alpha \in V_j} \lambda_\alpha^{KDE} \cdot Z(x_\alpha)$ avec les conditions supplémentaires $\phi_1(x_j) = \sum_{x_\alpha \in V_j} \lambda_\alpha^{KDE} \cdot \phi_1(x_\alpha) \text{ et } \phi_2(x_j) = \sum_{x_\alpha \in V_j} \lambda_\alpha^{KDE} \cdot \phi_2(x_\alpha)$		
Cokrigage à moyenne inconnue entre les observations (variable principale) et CHIMERE (variable secondaire)	$Z(x) = a_{1x} \cdot \phi_1(x) + a_{2x} \cdot \phi_2(x) + b_x + U(x)$ U et S s'écrivent comme la somme de composantes d'échelle indépendantes (Z_k^i), par exemple : $U(x) = \varepsilon(x) + Z_1^{pp}(x) + Z_1^{gp}(x)$	En première approximation : estimation d'une dérive globale, par exemple par moindres carrés : $Z(x) = a_1^* \cdot \phi_1(x) + a_2^* \cdot \phi_2(x) + b^* + V(x),$ calcul des résidus expérimentaux $V(x_\alpha)$ et du variogramme expérimental de V	Données d'observation aux points x_α Données de simulation aux points de la grille CHIMERE (x_β) Données de ϕ_1 et de ϕ_2 aux points d'observation x_α et au

Méthode	Hypothèses et relations sous-jacentes	Modélisation du variogramme	Données d'entrée
	$S(x) = Z_2^{pp}(x) + Z_2^{gp}(x)$	Modèle linéaire de corégionalisation ajusté sur les variogrammes simples et croisé $\gamma_v, \gamma_s, \gamma_{vs}$	point d'estimation x_j
	Estimation directe à partir des données expérimentales : $Z^{CK}(x_j) = \sum_{x_\alpha \in V_j^Z} \lambda_\alpha^{CK} \cdot Z(x_\alpha) + \sum_{x_\beta \in V_j^Y} \lambda_\beta^{CK} \cdot S(x_\beta)$ avec les conditions supplémentaires : $\phi_1(x_j) = \sum_{x_\alpha \in V_j} \lambda_\alpha^{CK} \cdot \phi_1(x_\alpha)$ et $\phi_2(x_j) = \sum_{x_\alpha \in V_j} \lambda_\alpha^{CK} \cdot \phi_2(x_\alpha)$		

ANNEXE 5

**Validation sur les stations de vérification.
Statistiques d'erreurs avant et après analyse.**

Légende :

mod : CHIMERE seul

KI : krigeage des innovations

KDE : krigeage des observations avec CHIMERE en dérive externe

CK : cokrigeage entre les observations et CHIMERE

moy_err : erreur moyenne (*biais*) $\frac{1}{N} \sum_i (Z_i^* - Z_i)$

std_err : écart-type de l'erreur $\sqrt{\frac{1}{N-1} \sum_i \left[(Z_i^* - Z_i) - \left(\overline{Z_i^* - Z_i} \right) \right]^2}$

rmse : racine carrée de l'erreur quadratique moyenne $\sqrt{\frac{1}{N} \sum_i (Z_i^* - Z_i)^2}$

moy_err_abs : moyenne des erreurs en valeur absolue $\frac{1}{N} \sum_i |Z_i^* - Z_i|$

moy_errel_abs : moyenne des erreurs relatives en valeur absolue $\frac{1}{N} \sum_i \frac{|Z_i^* - Z_i|}{Z_i}$

sd_errel_abs : écart-type des erreurs relatives en valeur absolue

$$\sqrt{\frac{1}{N-1} \sum_i \left[\frac{|Z_i^* - Z_i|}{Z_i} - \left(\overline{\frac{|Z_i^* - Z_i|}{Z_i}} \right) \right]^2}$$

cor : coefficient de corrélation

Tableau 8 – Été 2006 (18 juillet-14 août). Statistiques des erreurs toutes heures confondues.

Méthode	Typologie	N	moy_err	moy_errabs	std_err	rmse	moy_errela_bs	std_errel_abs	cor
mod	périurbaine (PU)	30259	16.7	23.3	26.3	31.1	79.4	278.4	0.727
KI	périurbaine (PU)	30259	1.9	11.2	16.2	16.3	33.1	127.6	0.902
KDE	périurbaine (PU)	30259	1.5	10.6	15.3	15.4	32.7	131.4	0.912
CK	périurbaine (PU)	30259	2.6	11.3	16.3	16.5	35.6	142.1	0.899
mod	urbaine (U)	87482	17.2	23.6	26.3	31.4	79.7	300.9	0.736
KI	urbaine (U)	87482	2.4	10.6	15.4	15.5	31.4	134.6	0.913
KDE	urbaine (U)	87482	2.1	10.0	14.4	14.6	30.6	136.5	0.923
CK	urbaine (U)	87482	3.2	10.2	14.6	15.0	32.7	146.0	0.921
mod	PU + U	117741	17.0	23.5	26.3	31.3	79.6	295.3	0.734
KI	PU + U	117741	2.3	10.7	15.6	15.7	31.9	132.8	0.910
KDE	PU + U	117741	2.0	10.2	14.6	14.8	31.2	135.2	0.920
CK	PU + U	117741	3.0	10.5	15.1	15.4	33.5	145.0	0.915

Tableau 9 – Été 2006 (18 juillet-14 août). Statistiques des erreurs sur le pic journalier.

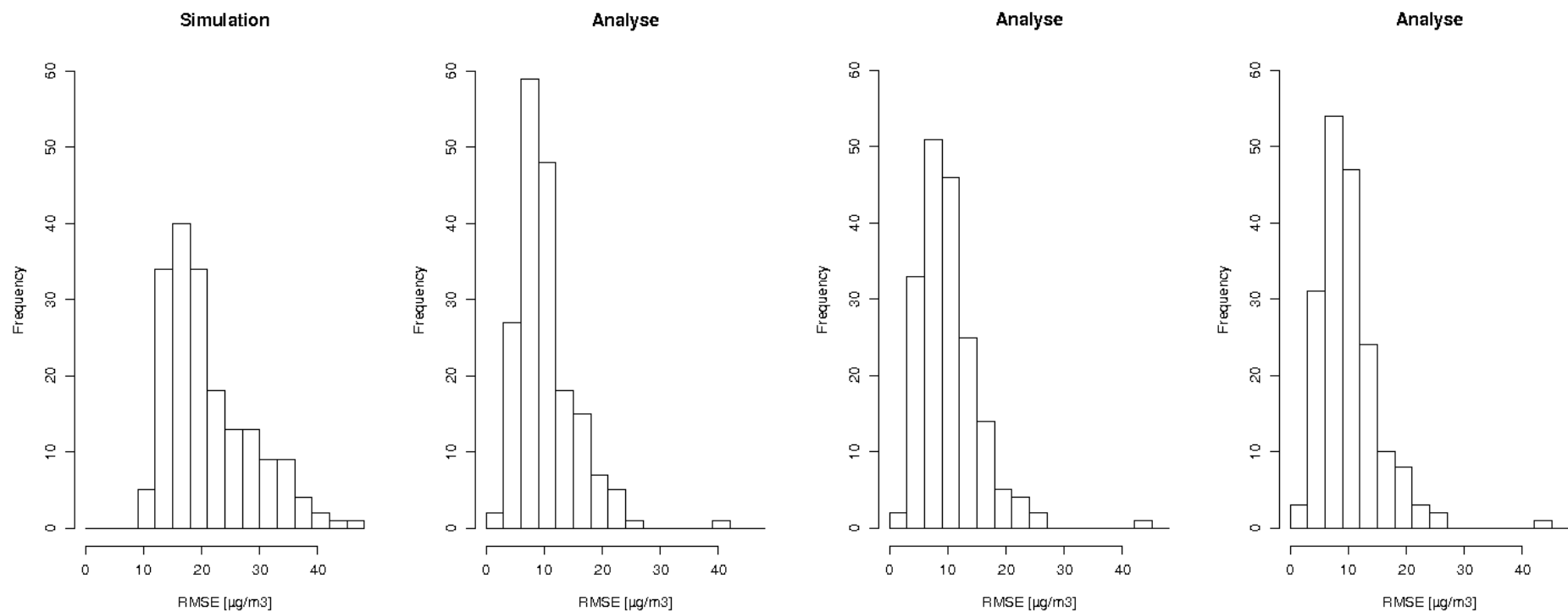
Méthode	Typologie	N	moy_err	moy_errabs	std_err	rmse	moy_errela_bs	std_errel_abs	cor
mod	périurbaine (PU)	1264	4.2	14.3	19.4	19.8	14.0	14.2	0.872
KI	périurbaine (PU)	1264	-1.3	7.6	10.8	10.9	7.2	7.3	0.959
KDE	périurbaine (PU)	1264	-2.1	7.6	10.6	10.8	7.2	7.2	0.961
CK	périurbaine (PU)	1264	-1.9	7.8	11.0	11.2	7.4	8.1	0.957
mod	urbaine (U)	3643	5.0	15.0	20.2	20.8	14.5	13.7	0.872
KI	urbaine (U)	3643	-0.2	7.9	12.3	12.3	7.2	7.4	0.951
KDE	urbaine (U)	3643	-0.4	7.6	11.7	11.7	7.0	7.1	0.955
CK	urbaine (U)	3643	-0.6	7.6	11.9	12.0	7.0	8.0	0.953
mod	PU + U	4907	4.8	14.8	20.0	20.5	14.4	13.8	0.872
KI	PU + U	4907	-0.5	7.8	11.9	11.9	7.2	7.4	0.953
KDE	PU + U	4907	-0.9	7.6	11.5	11.5	7.0	7.1	0.956
CK	PU + U	4907	-1.0	7.6	11.7	11.8	7.1	8.0	0.954

Tableau 10 – Été 2009 (18 juillet-14 août). Statistiques des erreurs toutes heures confondues.

Méthode	Typologie	N	moy_err	moy_errabs	std_err	rmse	moy_errela_bs	std_errel_abs	cor
mod	périurbaine (PU)	28838	22.3	25.2	24.2	32.9	114.8	418.0	0.713
KI	périurbaine (PU)	28838	1.8	11.2	15.5	15.6	44.8	166.2	0.882
KDE	périurbaine (PU)	28838	0.9	11.0	15.2	15.2	44.4	172.1	0.888
CK	périurbaine (PU)	28838	2.5	11.4	15.5	15.7	49.3	188.1	0.883
mod	urbaine (U)	85946	21.2	25.1	24.7	32.6	110.0	408.1	0.712
KI	urbaine (U)	85946	2.7	10.9	15.1	15.4	43.0	167.1	0.890
KDE	urbaine (U)	85946	2.0	10.3	14.2	14.3	41.9	165.6	0.902
CK	urbaine (U)	85946	3.5	10.7	14.4	14.8	46.8	188.5	0.900
mod	PU + U	114784	21.4	25.1	24.6	32.6	111.2	410.6	0.712
KI	PU + U	114784	2.5	10.9	15.2	15.4	43.5	166.9	0.888
KDE	PU + U	114784	1.7	10.5	14.5	14.6	42.5	167.3	0.899
CK	PU + U	114784	3.2	10.9	14.7	15.0	47.4	188.4	0.895

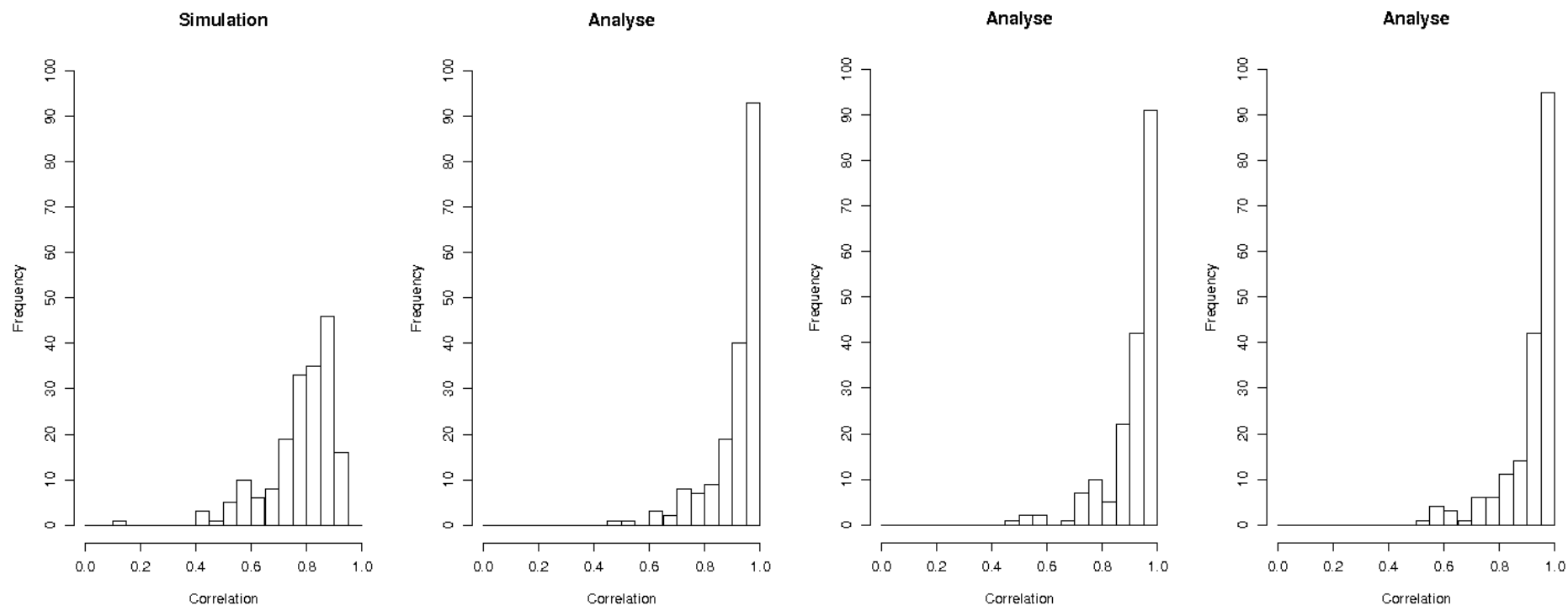
Tableau 11 – Été 2009 (18 juillet-14 août). Statistiques des erreurs sur le pic journalier.

Méthode	Typologie	N	moy_err	moy_errabs	std_err	rmse	moy_errela_bs	std_errel_abs	cor
mod	périurbaine (PU)	1191	11.6	15.8	17.4	20.9	16.5	16.7	0.811
KI	périurbaine (PU)	1191	-2.6	7.6	10.1	10.4	7.4	6.8	0.926
KDE	périurbaine (PU)	1191	-3.0	7.6	9.9	10.4	7.3	6.7	0.929
CK	périurbaine (PU)	1191	-3.1	7.7	10.1	10.6	7.4	6.9	0.926
mod	urbaine (U)	3561	12.6	16.8	18.1	22.1	18.0	16.8	0.815
KI	urbaine (U)	3561	-0.1	7.7	10.6	10.6	7.7	7.2	0.927
KDE	urbaine (U)	3561	-0.5	7.6	10.6	10.6	7.6	7.3	0.928
CK	urbaine (U)	3561	-0.7	7.7	10.6	10.6	7.7	7.4	0.927
mod	PU + U	4752	12.3	16.6	18.0	21.8	17.6	16.8	0.814
KI	PU + U	4752	-0.7	7.7	10.5	10.6	7.6	7.1	0.926
KDE	PU + U	4752	-1.1	7.6	10.5	10.5	7.5	7.1	0.927
CK	PU + U	4752	-1.3	7.7	10.6	10.6	7.7	7.3	0.926



CHIMERE brut Krigage des innovations Krigage avec dérive externe Cokrigage observations – CHIMERE

Figure 8 – Été 2009. Distribution de la **RMSE** calculée sur le **pic journalier** en chaque site de vérification périurbain ou urbain.



CHIMERE brut Krigeage des innovations Krigeage avec dérive externe Cokrigeage observations – CHIMERE

Figure 9 – Eté 2009. Distribution de la **corrélation** calculée sur le **pic journalier** en chaque site de vérification périurbain ou urbain.