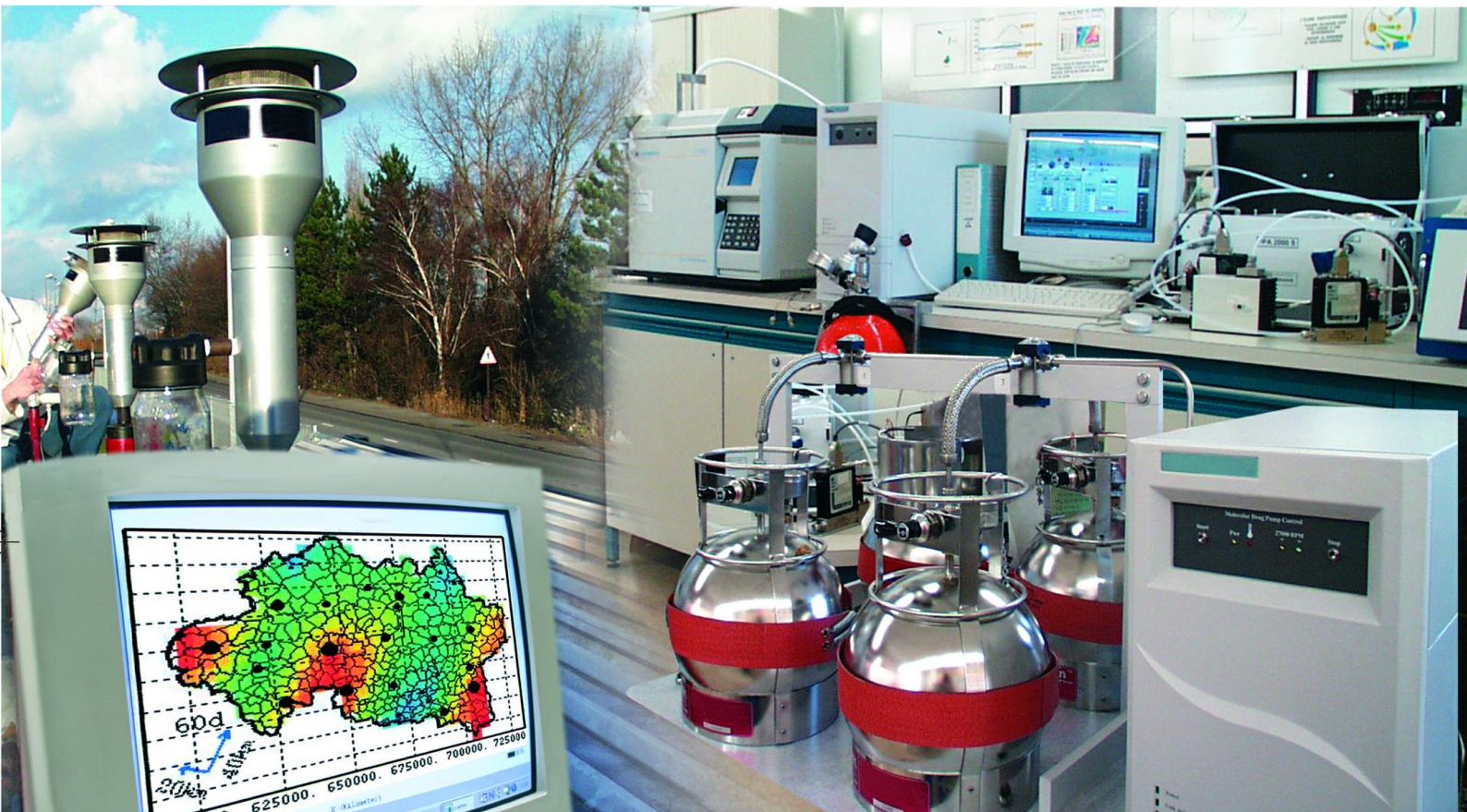




Laboratoire Central de Surveillance de la Qualité de l'Air



Modélisation et Traitements Numériques

Travaux relatifs à la plate-forme nationale de modélisation
PREV'AIR : Réalisation de cartes analysées d'ozone (3/3)

Décembre 2008

Programme 2008

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, sous la coordination technique de l'ADEME et en concertation avec les Associations Agréées de Surveillance de la Qualité de l'Air (AASQA). Ces travaux en matière de pollution atmosphérique supportés financièrement par le 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 en apportant un appui scientifique et technique aux AASQA.

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 (3/3)

Laboratoire Central de Surveillance
de la Qualité de l'Air

Thème : Modélisation et Traitements Numériques

Programme financé par le
Ministère de l'Écologie, de l'Énergie, du Développement durable et de la Mer
(MEEDDM)

2008

L. MALHERBE, A. UNG (INERIS)
Ch. de FOUQUET (MINES ParisTech/ARMINES)

Ce document comporte 51 pages (hors couverture et annexes)

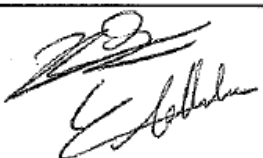
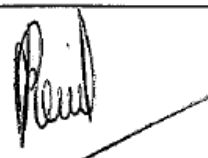

	Rédaction	Vérification	Approbation
NOM	L. MALHERBE A. UNG	L. ROUÏL	M. RAMEL
Qualité	Ingénieurs Direction des Risques Chroniques	Responsable du pôle DECI Direction des Risques Chroniques	Responsable LCSQA/INERIS Direction des Risques Chroniques
Visa			

TABLE DES MATIÈRES

1. RESUME.....	8
2. INTRODUCTION.....	10
2.1 Contexte.....	10
2.2 Contenu de l'étude	10
3. DONNEES D'ENTREE	11
3.1 Variables	11
3.2 Période d'étude	11
3.3 Sélection des stations de mesure.....	11
4. KRIGEAGE DES INNOVATIONS.....	13
4.1 Rappel des hypothèses	13
4.2 Evaluation statistique.....	13
4.3 Vérification des hypothèses.....	14
5. ETUDE ET MODELISATION DES RELATIONS ENTRE LE MODELE CHIMERE ET LES MESURES	16
5.1 Etude temporelle par station.....	16
5.1.1 Correlation CHIMERE-mesures.....	16
5.1.2 Variogrammes expérimentaux	16
5.1.3 Modélisation	25
5.1.3.1 Modèle de référence	25
5.1.3.2 Généralisation du modèle	25
5.1.3.3 Validation du modèle	27
5.1.3.4 Analyse des coefficients.....	32
5.1.3.5 Modèle moyen par zone	34
5.2 Etude spatiale : comparaison par heure du jour.....	36
5.2.1 Remarques préliminaires	36
5.2.2 Variogrammes expérimentaux	36
5.2.3 Modèle spatial	38
6. APPLICATION DU COKRIGEAGE	39
6.1 Evaluation statistique.....	39
6.2 Cartographies.....	45
7. BILAN ; VOIES D'AMELIORATION.....	48
8. REFERENCES.....	50

9. LISTE DES ANNEXES	51
ANNEXE 1	53
ANNEXE 2	56
ANNEXE 3	61
ANNEXE 4	65
ANNEXE 5	97
ANNEXE 6	102

1. RESUME

Cette étude, conduite en collaboration avec le Centre de Géosciences (équipe de géostatistique) de l'École des Mines de Paris, s'inscrit dans les travaux du LCSQA relatifs au système PREV'Air. Elle a pour objet d'améliorer la réalisation des cartes analysées, qui sont obtenues en corrigeant les simulations du modèle CHIMERE à l'aide des données d'observation. On s'intéresse ici aux simulations et aux mesures horaires (stations fixes) des concentrations d'ozone et de PM10 sur la France.

Ce rapport présente les résultats préliminaires obtenus pour l'ozone sur deux périodes d'étude : quatre semaines de l'été 2005 et quatre semaines de l'été 2006. Le travail sur l'ozone et les PM10 se poursuivra en 2009.

La méthode de correction actuellement utilisée 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 à CHIMERE mais structuré spatialement. 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.

Une consultation de l'ensemble des AASQA a permis de remettre à jour la liste des stations de mesure jugées pertinentes pour l'analyse (i.e. dont la représentativité spatiale est du même ordre de grandeur que la maille CHIMERE). Cette liste inclut 219 sites pour l'ozone.

A partir de cette nouvelle sélection de stations, le krigeage des innovations a été appliqué aux simulations d'ozone sur les deux périodes d'étude. Les statistiques d'erreur confirment l'efficacité de la méthode. Après correction, les concentrations modélisées sont en meilleur accord avec la mesure. Ce résultat vaut aussi bien pour les sites d'observation utilisés dans le krigeage (stations d'analyse) que pour les stations de mesure exclues de cette sélection (stations de vérification). En validation, la RMSE (racine carrée de l'erreur moyenne quadratique) diminue ainsi de 40 à 40% et la corrélation augmente de 20 à 25%.

Une analyse de la variabilité temporelle (station par station) et spatiale (heure par heure) des observations et de CHIMERE montre cependant que l'hypothèse d'un modèle à résidu n'est pas toujours vérifiée. Dans le temps comme dans l'espace, les innovations sont souvent corrélées à CHIMERE. Nous avons donc cherché à développer un nouveau modèle géostatistique capable de représenter plus exactement les relations spatiales ou temporelles entre CHIMERE et les observations. Comme il s'agit d'un problème complexe, l'aspect temporel et l'aspect spatial ont été abordés séparément. Dans les deux cas, et exception faite de situations particulières (zones de montagne notamment), un modèle de corrélation intrinsèque ou ce qui en est la forme plus générale, le modèle linéaire de corégionalisation, apparaît généralement bien adapté : les variogrammes simples et croisés de CHIMERE, des observations et des innovations peuvent s'écrire comme une combinaison linéaire des mêmes structures. Ce modèle a l'avantage de décrire de façon simple les rapports de variabilité entre CHIMERE et les observations aux différentes échelles temporelles ou spatiales. Dans les zones où la simulation

déterministe se révèle moins satisfaisante, il permet de préciser la nature des écarts entre CHIMERE et la mesure.

Dans le temps, les variogrammes de chaque station ont pu être correctement modélisés par une structure de longue portée (deux à trois jours), caractéristique de la durée de dispersion des panaches, une structure périodique de période 24 heures, caractéristique du cycle diurne, et un effet de pépité (pour le variogramme des observations), caractéristique de l'erreur de mesure.

Dans l'espace, les variogrammes horaires ont été modélisés par une structure exponentielle de portée comprise entre 90 et 150 km selon l'heure du jour et par un effet de pépité (variogramme des observations). Cette modélisation pourra être améliorée : prise en compte d'une structure de petite portée sur le variogramme des observations, représentation des anisotropies, etc. De plus, la corrélation spatiale entre CHIMERE et les mesures se révèle assez faible. Ce point requiert des investigations complémentaires : on examinera notamment si cette corrélation s'accroît lorsqu'on travaille par zone géographique et non plus de manière globale sur la France.

A titre d'essai, le modèle spatial global ici défini a été mis en œuvre dans un cokrigeage ; il a été testé par validation croisée sur les stations d'analyse et par validation sur les stations de vérification. Les statistiques d'erreur indiquent que par rapport aux simulations brutes de CHIMERE, les concentrations estimées par cokrigeage sont en meilleure adéquation avec la mesure. Ces statistiques restent toutefois comparables à celles du krigeage des innovations. D'autre part, les cartes résultant du cokrigeage sont plus lisses que les cartes issues du krigeage des innovations et le poids des observations y est généralement prépondérant. Les travaux complémentaires prévus pour 2009 permettront de mieux évaluer l'intérêt respectif des méthodes.

2. INTRODUCTION

2.1 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. 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 plusieurs études 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.

Les approfondissements requis au fil de l'étude nous ont conduits à limiter notre travail à l'**ozone**. La cartographie des PM₁₀ sera traitée en 2009, ce qui permettra d'intégrer les dernières mises à jour de la BDQA en ce qui concerne les particules (mises à jour faisant suite aux corrections des mesures de PM₁₀ non volatiles).

2.2 CONTENU DE L'ETUDE

Cette étude repose sur l'exploitation des données horaires de deux périodes estivales de 28 jours chacune : du 18 juillet au 14 août 2005 et du 18 juillet au 14 août 2006. Les chapitres de ce rapport en reprennent les principales étapes:

- Mise à jour de la liste des stations de mesure jugées pertinentes pour l'analyse.
- Mise en œuvre, à titre de contrôle, du krigeage des innovations. Cette technique d'analyse est fondée sur un modèle géostatistique appelé *modèle à résidu* dont on vérifie les hypothèses sous-jacentes.
- Etude des différences de variabilité temporelle et spatiale entre CHIMERE et les observations. A l'issue de cette analyse, un nouveau modèle est proposé pour décrire les relations entre ces deux variables : il s'agit du *modèle de corrélation intrinsèque* et de sa généralisation en *modèle linéaire de corégionalisation*.
- Correction spatiale des simulations horaires de CHIMERE : le modèle défini précédemment est considéré dans sa dimension spatiale et utilisé dans un cokrigeage entre CHIMERE et les mesures. Les résultats sont comparés à

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

Rapports disponibles à l'adresse www.lcsqa.org

ceux du krigeage des innovations. Les avantages de la méthode et les difficultés rencontrées sont alors discutés.

- Conclusion sur les améliorations à apporter et sur les suites de ce travail.

Tous les traitements statistiques et géostatistiques ont été effectués avec le logiciel R (version 2.5.0) et ses packages associés (chron, akima, splancs, geoR, gstat).

3. DONNEES D'ENTREE

3.1 VARIABLES

Les données suivantes ont été exploitées :

- simulations horaires du champ d'ozone sur la France issues du modèle CHIMERE (version V200709B, résolution 0.1° x 0.15°) ;
- données horaires de concentration d'ozone extraites de la BDQA ;

Pour l'interprétation des résultats ont été également prises en compte :

- les données d'altitude aux points de mesure extraites de la BDQA ;
- la densité d'occupation du sol (CORINE Land Cover) par maille CHIMERE ;
- la densité de population estimée sur chaque maille CHIMERE en fonction de l'occupation du sol.

3.2 PERIODE D'ETUDE

La période d'étude doit être suffisamment longue pour assurer la fiabilité des résultats et suffisamment courte pour qu'on puisse multiplier les tests dans un temps de calcul raisonnable. Issues d'un compromis entre ces deux nécessités, deux périodes de calcul allant du 18 juillet au 14 août 2005 et du 18 juillet au 14 août 2006 ont été définies. Dans la suite, elles seront plus simplement désignées par « été 2005 » et « été 2006 ».

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 estimée comparable aux dimensions de la maille de CHIMERE (soit environ 10 x 15 km).

Elles se répartissent comme suit :

- 10 stations rurales nationales
- 48 stations rurales régionales
- 75 stations périurbaines
- 78 stations urbaines

- 6 stations industrielles
- 2 stations d'observation spécifique.

Dans le rapport, nous les nommerons « stations d'analyse ». Les 228 stations exclues de cet ensemble seront appelées « stations de vérification ».

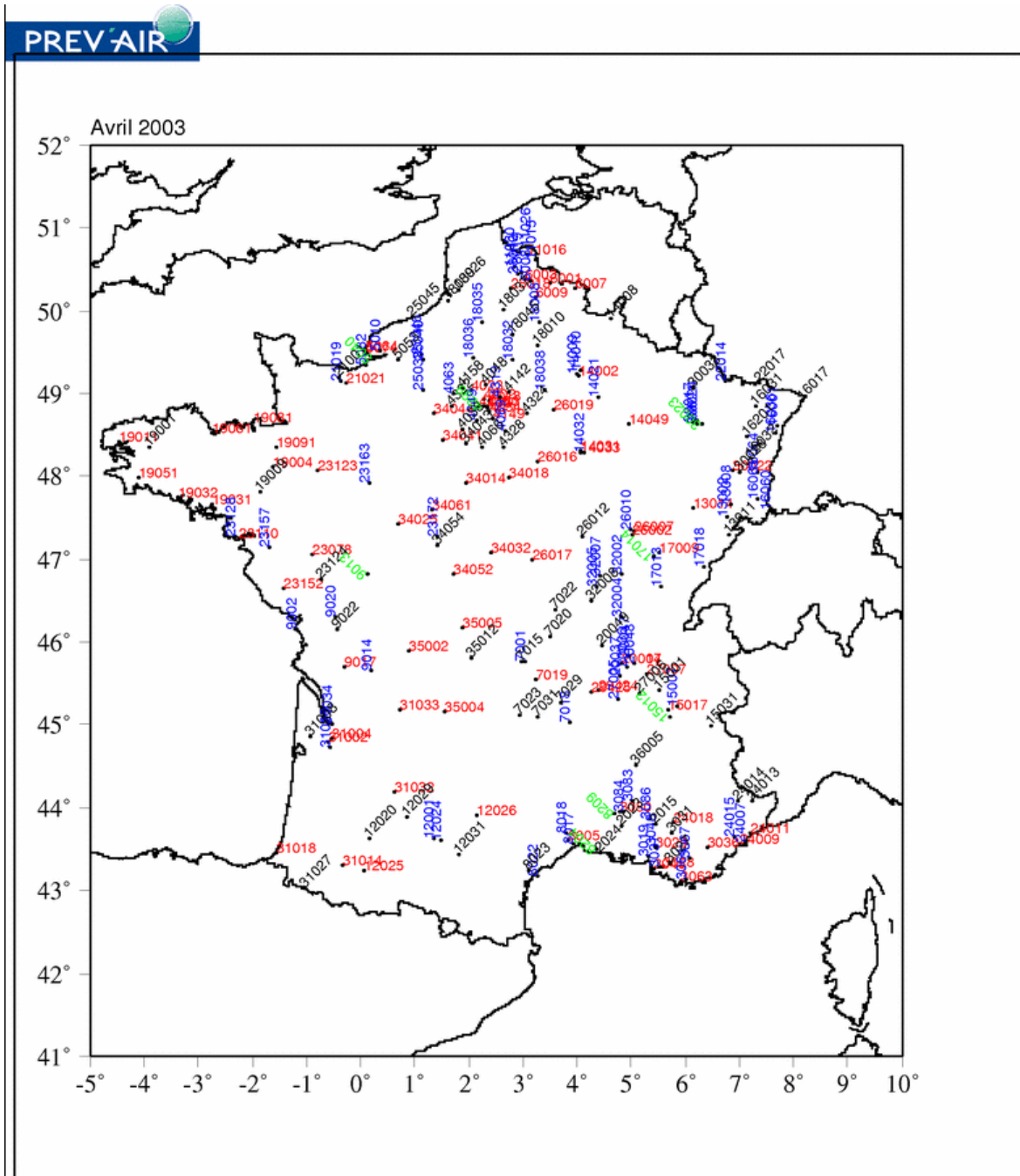


Figure 1 – Stations de mesure sélectionnées pour l'analyse. Numéros = codes des stations. Noir : stations rurales ; bleu : stations périurbaines ; rouge : stations urbaines ; vert : stations d'observation spécifique.

4. KRIGEAGE DES INNOVATIONS

4.1 RAPPEL DES HYPOTHESES

Le krigeage des innovations est fondé sur l'hypothèse que la concentration en un point x et à un instant t , $Z(x,t)$, est la somme du modèle CHIMERE, $mod(x,t)$, et d'un résidu aléatoire non corrélé, $R(x,t)$:

$$Z(x,t) = \underbrace{mod(x,t)}_{\perp} + R(x,t) \quad (1)$$

Aux points de mesure x_i , les résidus expérimentaux $R(x_i,t)$ sont donnés par l'écart entre les observations et le modèle CHIMERE.

En un point de grille x_g , à un instant t_0 , la concentration modélisée par CHIMERE est corrigée par krigeage de la façon suivante :

$$\hat{Z}(x_g, t_0) = \underbrace{mod(x_g, t_0)}_{\text{Modèle CHIMERE interpolé à la station } i} + \underbrace{\hat{R}(x_g, t_0)}_{\text{innovation}} = \underbrace{mod(x_g, t_0)}_{\text{Modèle CHIMERE interpolé à la station } i} + \sum_i \underbrace{\lambda_i(t_0)}_{\text{Poids de krigeage}} \cdot \underbrace{[Z(x_i, t_0) - mod(x_i, t_0)]}_{\text{Concentration mesurée à la station } i}$$

4.2 EVALUATION STATISTIQUE

Le krigeage des innovations, actuellement opérationnel dans PREV'Air, est testé pour contrôle sur chaque période d'étude.

Deux types d'évaluation sont réalisés :

- **Validation croisée** : en chaque station d'analyse, les concentrations horaires estimées à partir de CHIMERE et des 228 stations restantes (hors la station considérée) sont comparées aux vraies valeurs. Station par station et sur l'ensemble des sites, on calcule des indicateurs statistiques de performance que l'on confronte aux indicateurs associés à CHIMERE seul (i.e. avant correction par krigeage).
- **Validation sur les sites de vérification** : en chaque station de vérification, les concentrations horaires estimées à partir de CHIMERE et des 219 stations d'analyse sont comparées aux vraies valeurs. Comme pour la validation croisée, on en déduit des indicateurs statistiques que l'on compare à ceux de CHIMERE brut.

D'après ces statistiques, et conformément aux résultats des études antérieures, le krigeage des innovations améliore sensiblement l'adéquation entre CHIMERE et les données de mesure. En particulier, il réduit sensiblement l'erreur quadratique, accroît

la corrélation et augmente significativement le pourcentage de cas où l'erreur est inférieure à 50% (Annexe 1). Des statistiques plus détaillées seront fournies dans la comparaison avec le krigeage (cf. §6.1 et annexe 5).

4.3 VERIFICATION DES HYPOTHESES

Rappel sur les variogrammes :

Le variogramme simple d'une variable régionalisée Z_1 représente la variabilité moyenne de Z_1 en fonction de la distance h qui sépare deux points. Expérimentalement, il se calcule de la manière suivante :

$$\hat{\gamma}_1(h) = \frac{1}{2N(h)} \sum_{N(h)} \left[z_1(s_i) - z_1(s_j) \right]^2 \quad (\text{cas omnidirectionnel})$$

$N(h)$: nombre de couples de points de mesure distants de h

$z_1(s_i)$: valeur expérimentale de Z_1 au point de mesure s_i .

Le variogramme croisé est une extension du variogramme au cas de deux variables Z_1 et Z_2 . Expérimentalement, il se calcule selon l'expression :

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

$z_1(s_i), z_2(s_i)$: valeurs expérimentales de Z_1 et Z_2 au point de mesure s_i .

Le variogramme croisé entre Z_1 et Z_2 permet de décrire la corrélation spatiale entre ces deux variables.

N.B. : les variogrammes peuvent être calculés dans l'espace géographique, comme c'est ici le cas ; h est alors une distance spatiale. Ils peuvent être aussi calculés dans le temps, comme on le verra au chapitre 5 ; h est alors une durée.

Si l'hypothèse du krigeage des innovations est satisfaite, autrement dit, si la concentration d'ozone est la somme du modèle CHIMERE et d'un résidu non corrélé à ce dernier (relation 1, §4.1), alors à un instant t_0 , les variogrammes simples et croisés doivent vérifier les relations suivantes (2) :

$$\begin{cases} \gamma_Z(h, t_0) = \gamma_{\text{CHIM}}(h, t_0) + \gamma_{\text{Innov}}(h, t_0) \\ \gamma_{Z.\text{CHIM}}(h, t_0) = \gamma_{\text{CHIM}}(h, t_0) \\ \gamma_{\text{Innov}.\text{CHIM}}(h, t_0) = 0 \end{cases}$$

avec :

γ_Z : variogramme des observations

γ_{CHIM} : variogramme de CHIMERE

γ_{INNOV} : variogramme des innovations

$\gamma_{Z.\text{CHIM}}$: variogramme croisé observations-CHIMERE

$\gamma_{\text{INNOV}.\text{CHIM}}$: variogramme croisé innovations-CHIMERE

Les variogrammes expérimentaux montrent que la variabilité spatiale des observations est toujours supérieure à celle de CHIMERE, ce qui s'accorde avec un modèle du type *concentration = CHIMERE + résidu*. Toutefois, si les relations entre

les variogrammes (2) sont vérifiées pour certaines heures (Figure 2, ligne du haut, exemple du 25/07/2005 15h), divers exemples montrent que ce n'est pas systématique (Figure 2, ligne du bas, exemple du 06/08/2006 15h).

Aussi proposons-nous d'examiner plus en détail les relations entre le modèle CHIMERE et les observations afin d'établir, s'il est possible, un modèle de correction plus adapté.

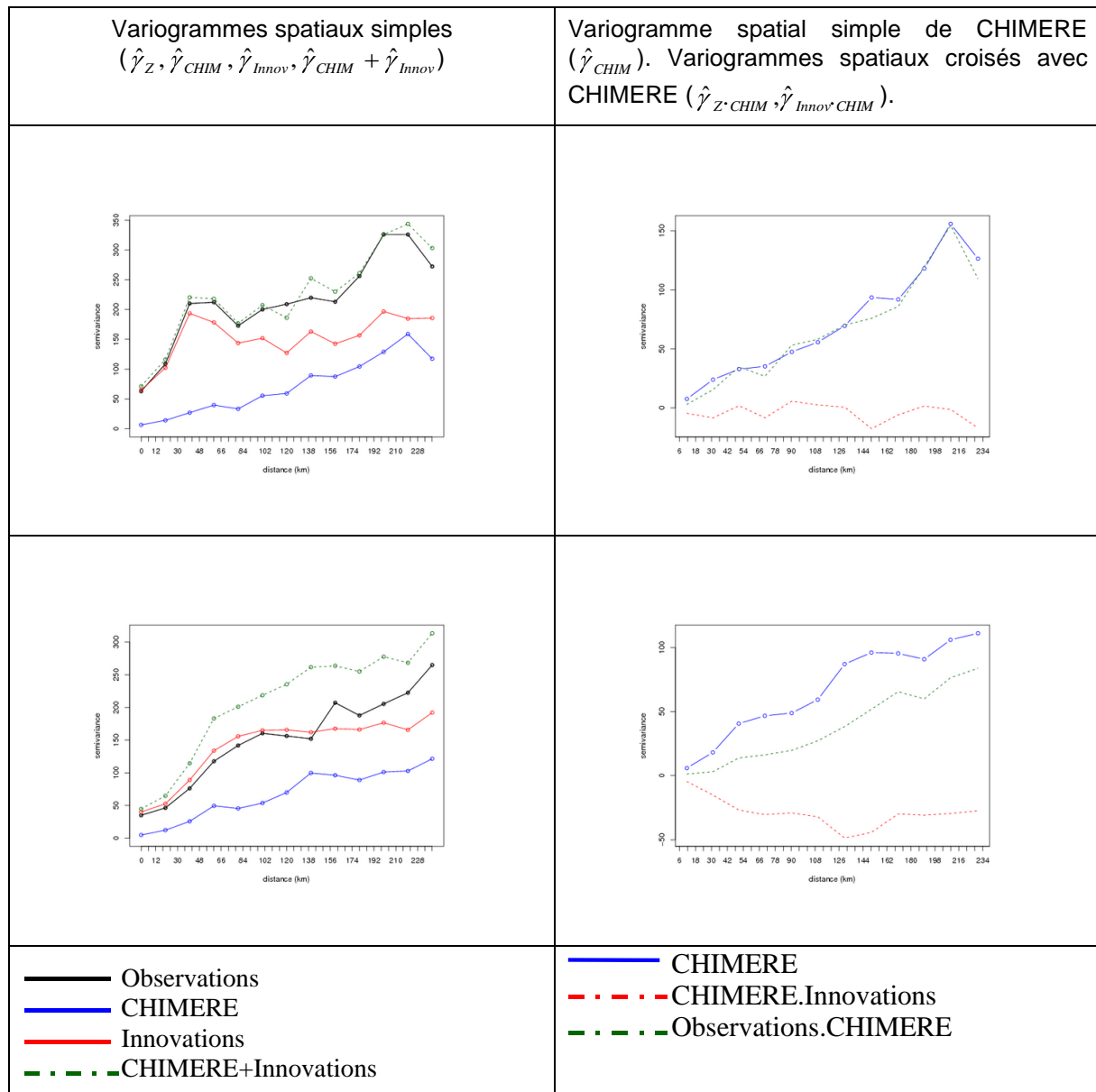


Figure 2 – Exemples de variogrammes simples et croisés expérimentaux. Ligne du haut : la variabilité des observations et des innovations est très supérieure à celle de CHIMERE, cependant, l'hypothèse d'un krigeage des innovations est vérifiée. Ligne du bas : les variogrammes ne remplissent pas les conditions requises, les innovations sont spatialement corrélées à CHIMERE.

5. ETUDE ET MODELISATION DES RELATIONS ENTRE LE MODELE CHIMERE ET LES MESURES

5.1 ETUDE TEMPORELLE PAR STATION

Le modèle CHIMERE a été interpolé aux points d'observation par la fonction *interp* de R (interpolation linéaire de type triangulaire).

La comparaison entre CHIMERE et les mesures s'effectue par le calcul et l'examen, station par station, des séries temporelles, du nuage et du coefficient de corrélation ainsi que des variogrammes temporels.

5.1.1 CORRELATION CHIMERE-MESURES

Les séries des données modélisées et mesurées ont été considérées de manière synchrone puis décalées l'une par rapport à l'autre de plus ou moins 1 à 3 heures.

En chaque station, la corrélation entre les séries a été calculée toutes heures confondues et par plage horaire (1-4h ; 5-8h ; 9-12h ; 13-18h ; 19-21h ; 22-0h). Les statistiques sur les coefficients de corrélation indiquent qu'en général, la corrélation est supérieure entre les mesures à l'heure h et CHIMERE à l'heure h-1. Cette augmentation de corrélation n'est pas très importante (de l'ordre de 0,02 en moyenne) mais elle se révèle quasi systématique. On la retrouve pour les deux étés, les différentes plages horaires et les diverses typologies de station (cf. annexe 2).

Ce résultat pourrait s'expliquer par les différences de support temporel entre la mesure et le modèle : une donnée de mesure d'ozone attribuée à l'heure h est la moyenne des 4 mesures quart-horaires (mesures instantanées tous les quarts d'heure) réalisées entre h-1 et h. En revanche, une donnée de simulation affectée à l'heure h est une valeur instantanée. A priori, une mesure à l'heure h peut donc être aussi bien corrélée à la donnée CHIMERE de l'heure h-1 qu'à celle de l'heure h.

5.1.2 VARIOGRAMMES EXPERIMENTAUX

Sont ici examinés les variogrammes simples et croisés de CHIMERE, des observations et des innovations.

Afin de se placer dans la situation de corrélation maximale (cf. §5.1.1), on utilise les données de CHIMERE à l'heure h-1 et les données d'observation à l'heure h.

Les variogrammes ont été calculés dans la dimension temporelle par pas d'une heure, jusqu'à une distance maximale de 150 heures (soit un peu plus de 6 jours).

Pour les observations comme pour CHIMERE, deux échelles de variation temporelle se superposent. On distingue :

- un phénomène de grande portée dont la variabilité est représentée par la base des oscillations. Cette portée est d'environ 2 à 3 jours, ce qui correspond à la durée moyenne de dispersion d'un panache ;
- un phénomène cyclique de période égale à 24h. Cette période coïncide avec celle du cycle journalier des concentrations.

Si dans la dimension temporelle, les relations entre CHIMERE et les observations s'accordent avec l'hypothèse d'un modèle à résidu (cf. § 4.3), alors, en un point x_0 , les variogrammes simples et croisés vérifient les relations :

- pour le variogramme des observations :

$$\gamma_Z(x_0, \tau) = \gamma_{\text{CHIM}}(h, \tau) + \gamma_{\text{Innov}}(h, \tau)$$

- pour le variogramme croisé observations-CHIMERE :

$$\gamma_{Z.\text{CHIM}}(x_0, \tau) = \gamma_{\text{CHIM}}(h, \tau)$$

- pour variogramme croisé innovations-CHIMERE :

$$\gamma_{\text{Innov.CHIM}}(x_0, \tau) = 0$$

Or ce n'est généralement pas le cas. L'étude des variogrammes met en évidence plusieurs cas de figure, dont nous présentons ci-après un échantillon (cas de figure numérotés de 1 à 4).

Légende :

Dans les graphiques suivants,

- l'axe des ordonnées est intitulé *semivariance* ; ce terme désigne la valeur du variogramme.
- l'axe des abscisses est intitulé *distance* ; il s'agit d'une distance temporelle exprimée en nombre d'heures.

1) La composante de grande portée de CHIMERE (base du variogramme) coïncide avec celle des observations mais non la composante périodique qui est d'amplitude moindre (cas le plus fréquent, cf. Figure 3) ou supérieure. Les innovations sont temporellement corrélées à CHIMERE (le variogramme croisé entre CHIMERE et les innovations est non nul et négatif).

Interprétation : en moyenne sur la période, CHIMERE reproduit correctement l'amplitude des variations de concentration sur plusieurs jours mais il sous- ou surestime les fluctuations diurnes.

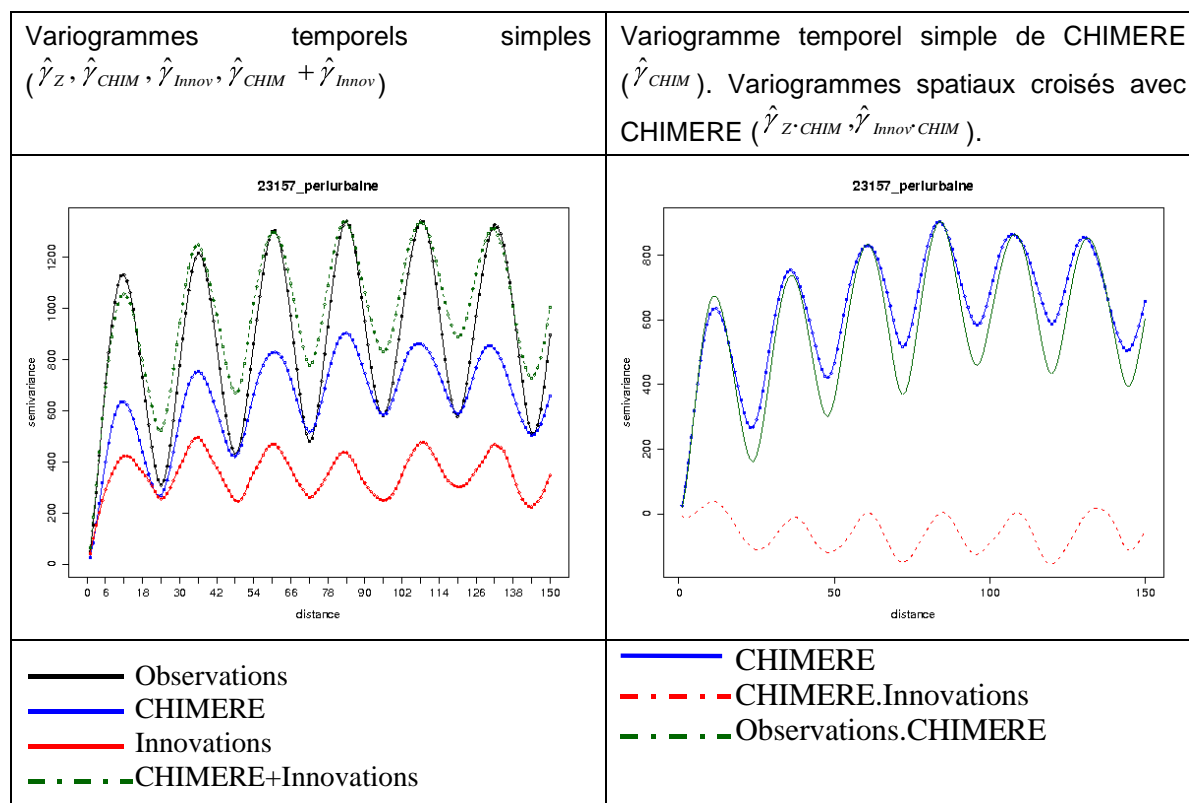


Figure 3 – Cas de figure n°1. Exemple de la station 23157 (AIR Pays de Loire, Epinettes, périurbaine), été 2006.

Les variogrammes de la Figure 3 sont cohérents avec les séries temporelles qui montrent une sous-estimation par CHIMERE de l'amplitude du signal journalier (Figure 4).

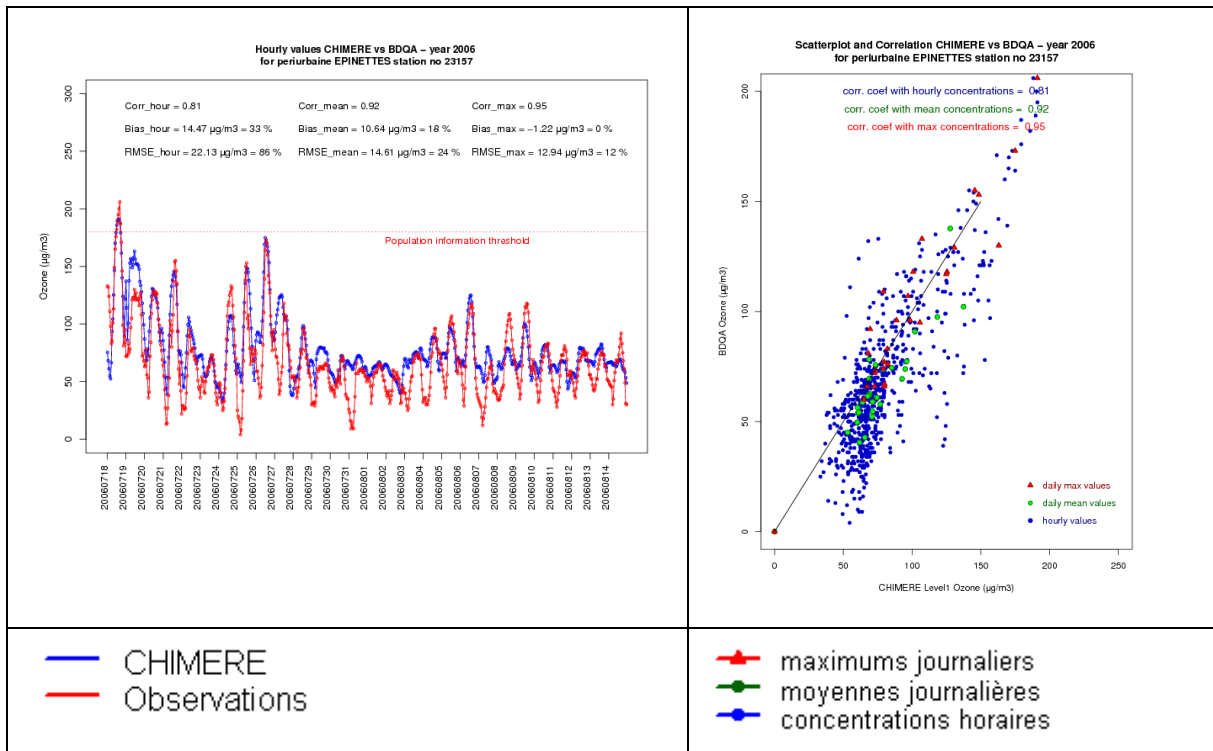


Figure 4 – Cas de figure n°1. Exemple de la station 23157, été 2006. A gauche : série des valeurs horaires. A droite : nuage de corrélation correspondant. La corrélation est sensiblement plus faible sur l'ensemble des valeurs horaires (points bleus, $\rho=0,61$) que sur les maximums horaires journaliers (points rouges, $\rho=0,95$).

2) La composante périodique de CHIMERE est approximativement de même amplitude que celle des observations mais le phénomène de grande portée a une variabilité plus faible ou plus grande (la base du variogramme est respectivement moins ou plus élevée).

Interprétation : CHIMERE atténue ou accentue les variations de concentration sur plusieurs jours. Cette dernière configuration est illustrée par l'exemple ci-après (Figure 5, Figure 6). Les innovations sont temporellement corrélées à CHIMERE.

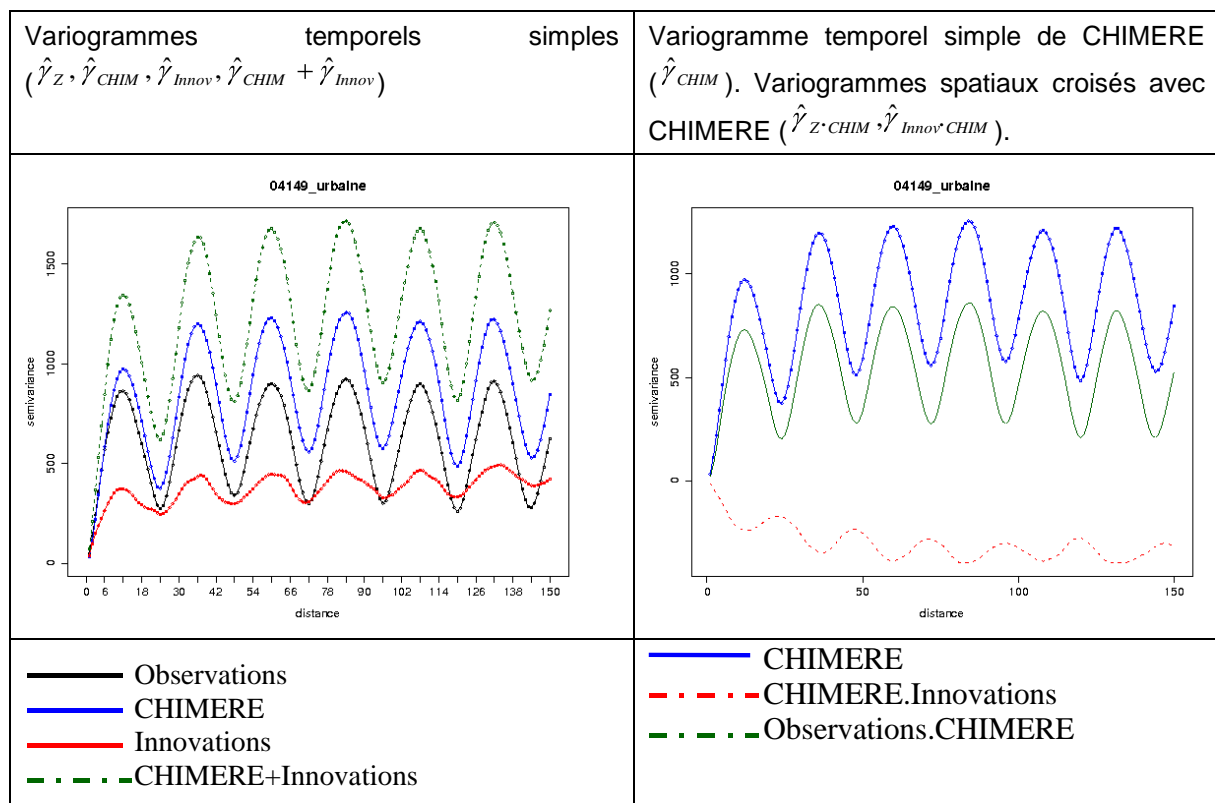


Figure 5 – Cas de figure n°2. Exemple de la station 04149 (AIRPARIF, Montgeron, urbaine), été 2005.

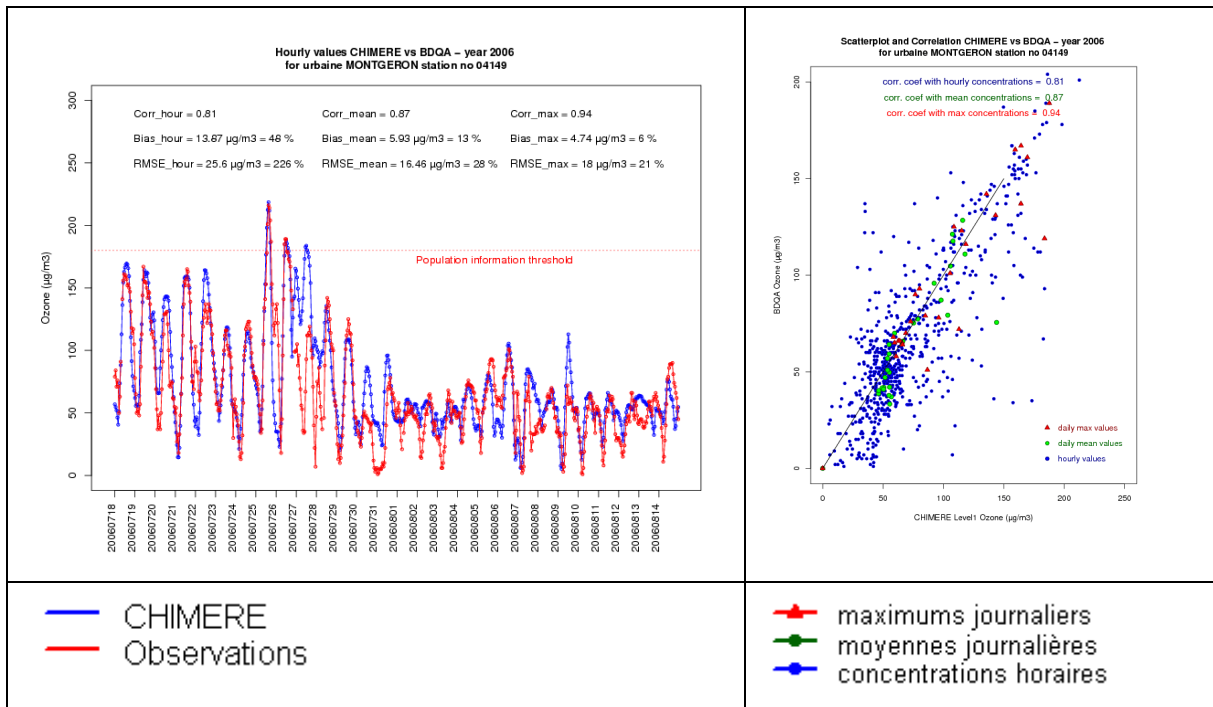


Figure 6 – Cas de figure n°2. Exemple de la station 04149, été 2006. A gauche : série des valeurs horaires. A droite : nuage de corrélation correspondant.

3) Le variogramme de CHIMERE et le variogramme des observations coïncident approximativement mais CHIMERE est corrélé temporellement avec les innovations (le variogramme croisé entre CHIMERE et les innovations est non nul et négatif).

Interprétation : en moyenne, CHIMERE simule correctement la variabilité des concentrations mais il ne varie pas toujours en parfaite corrélation avec la mesure (Figure 7, Figure 8) (en termes géostatistiques, on parle de simulation non conditionnelle).

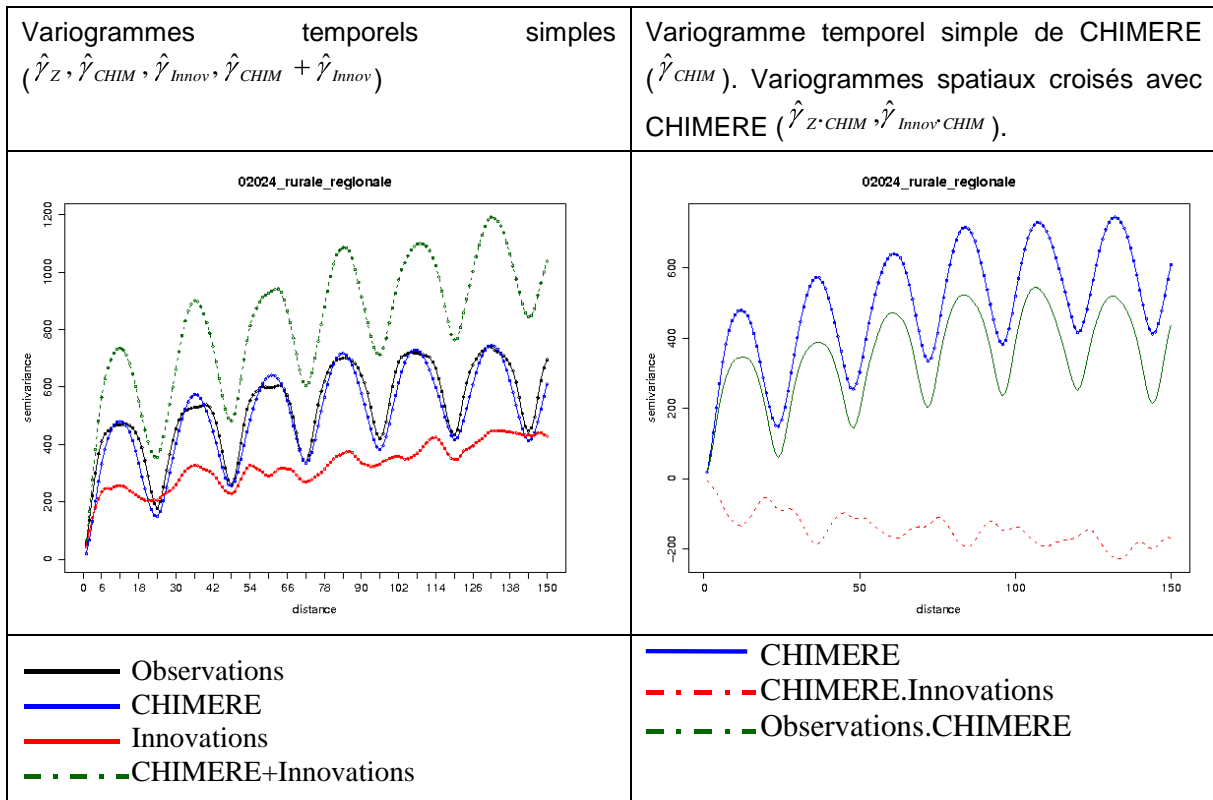


Figure 7 – Cas de figure n°3. Exemple de la station 02024 (AIRFOBEP, Saintes-Maries-de-la-Mer, rurale régionale), été 2006.

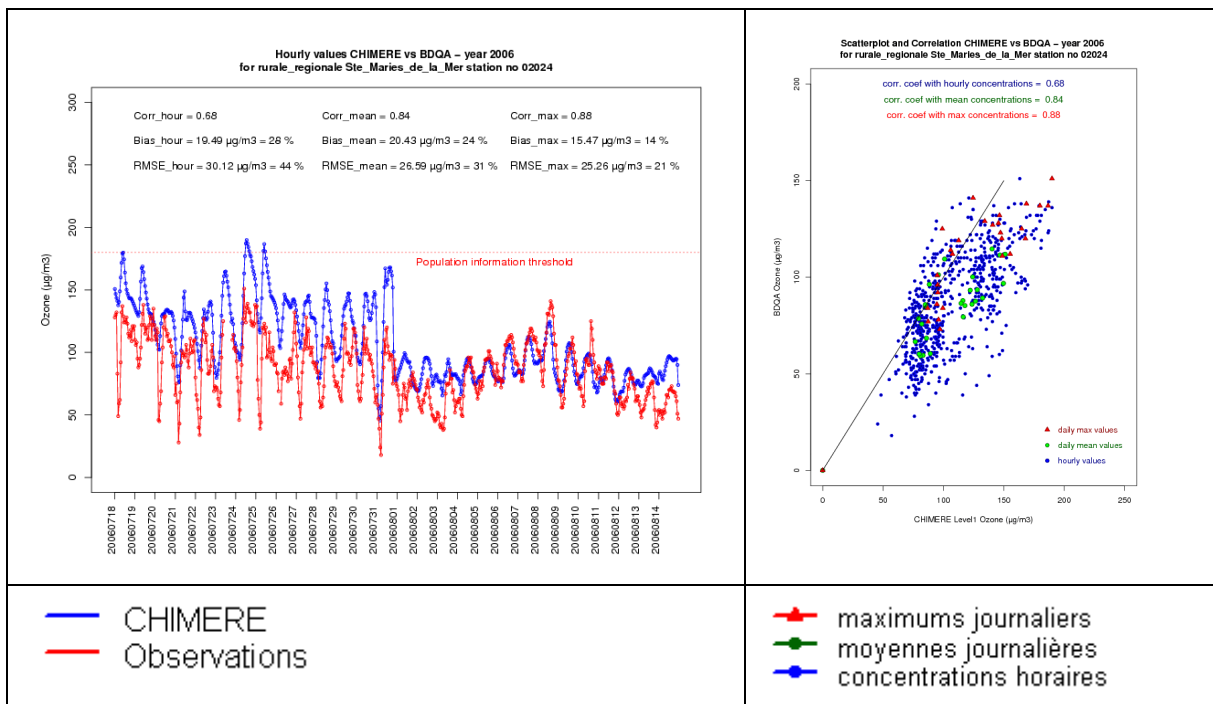


Figure 8 – Cas de figure n°3. Exemple de la station 02024, été 2006. A gauche : série des valeurs horaires. A droite : nuage de corrélation correspondant.

4) Le variogramme de CHIMERE diffère du variogramme des observations par ses deux composantes (Figure 9, Figure 10).

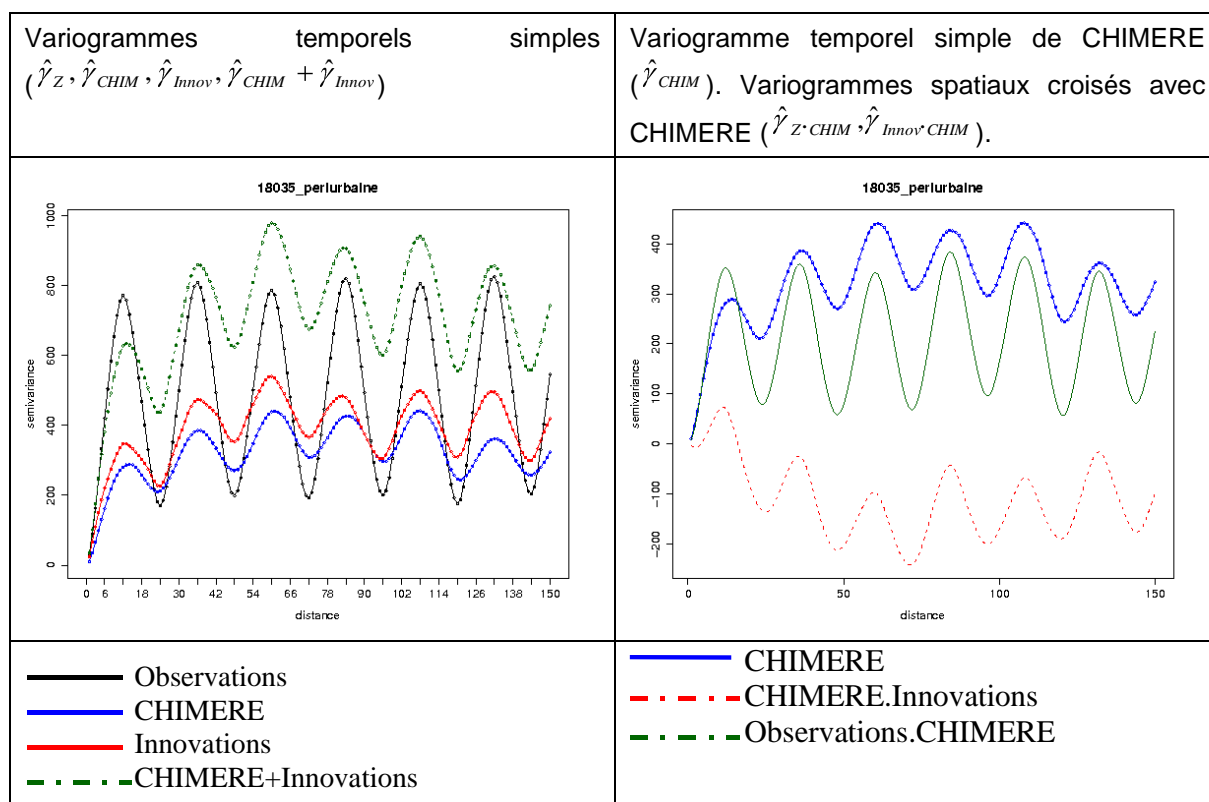


Figure 9 – Cas de figure n°4. Exemple de la station 18036 (ATMO Picardie, Beaumont_Beauvais, périurbaine), été 2006

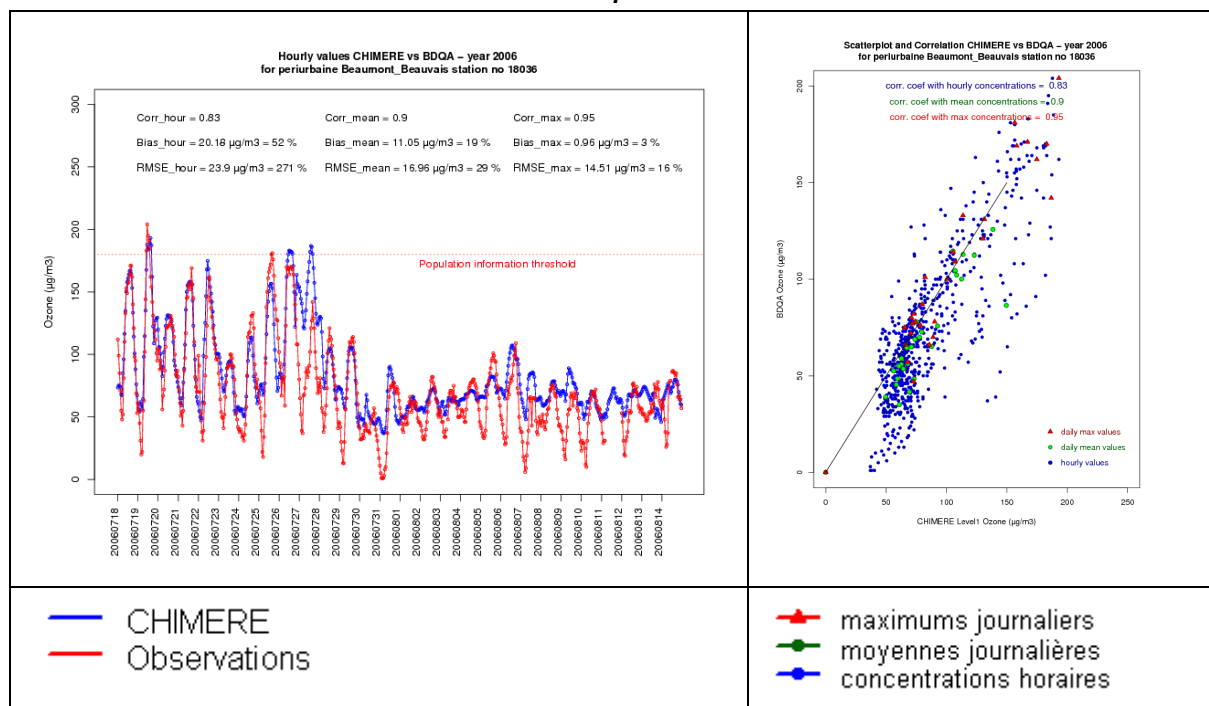


Figure 10 – Cas de figure n°4. Exemple de la station 18036, été 2006. A gauche : série des valeurs horaires. A droite : nuage de corrélation correspondant.

5) Cas particuliers

Un petit nombre d'autres stations se distinguent par leurs variogrammes temporels. Il s'agit notamment des sites de montagne.

Pour la majorité – mais non la totalité - des sites localisés à plus de 600 m d'altitude, la composante périodique journalière du variogramme des observations est sensiblement atténuée, voire nulle, ou relativement erratique. La composante de grande portée diffère selon les stations. Pour plusieurs d'entre elles, elle est proche de la linéarité, ce qui peut traduire une non stationnarité à cette échelle temporelle (Figure 11).

Suivant les cas, le variogramme de CHIMERE s'écarte plus ou moins du variogramme des observations.

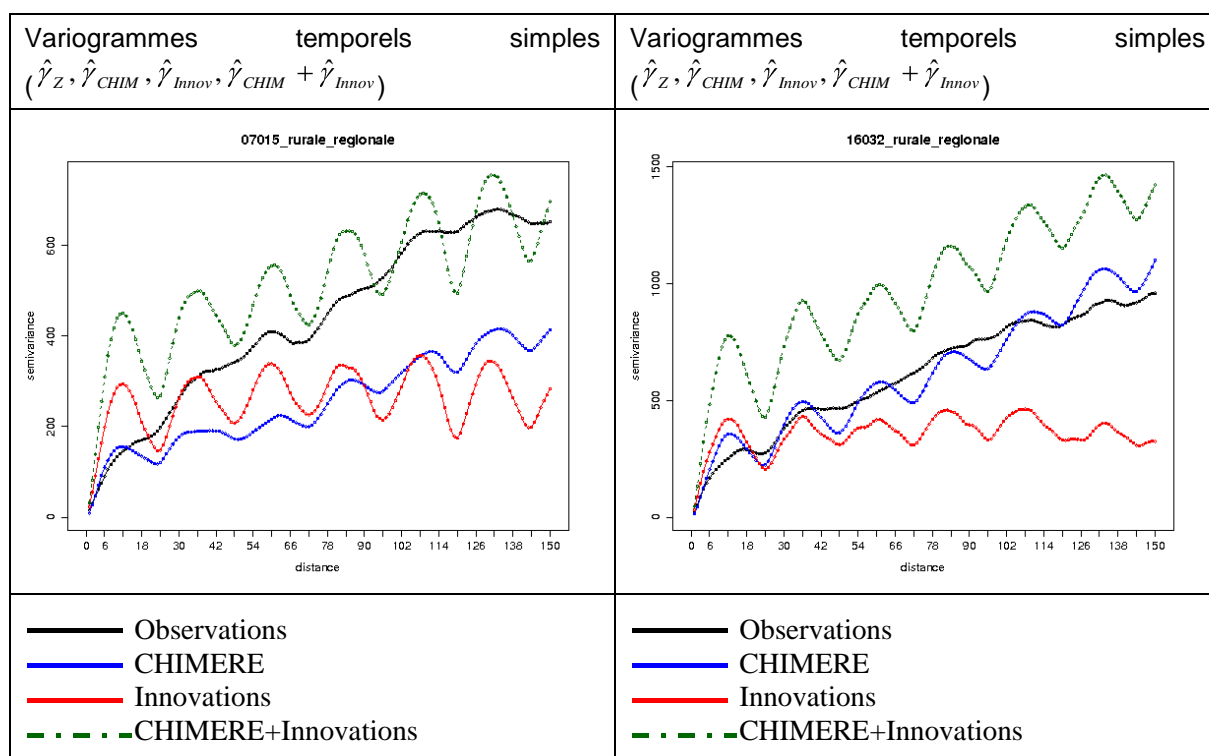


Figure 11 – Exemple de stations de montagne : la station 07015 (Puy-de-Dôme, ATMO Auvergne, rurale régionale) et la station 16032 (Hautes-Vosges, ASPA, rurale régionale), été 2006

Ces résultats mettent en évidence la complexité des sites de montagne, pour lesquels la variabilité des concentrations mesurées ne peut se décrire de manière générale, et la difficulté qu'a le modèle pour simuler les concentrations dans ces zones, en raison de la résolution considérée.

5.1.3 MODELISATION

5.1.3.1 MODELE DE REFERENCE

Etudiant les moyennes *journalières* de différentes espèces chimiques à l'échelle de l'Europe, Chilès et Séguret (2008) ont confronté la mesure (de support spatial « ponctuel ») à la simulation CHIMERE (concentrations simulées sur des mailles d'environ 50 km x 50 km).

Ils montrent que les relations entre la concentration d'ozone vraie mais inconnue, la mesure, CHIMERE et l'innovation peuvent être efficacement représentées par un modèle géostatistique de *corrélation intrinsèque*. Ce dernier suppose que les variogrammes simples et croisés des différentes variables sont proportionnels. Ce modèle est détaillé dans l'annexe 3.

5.1.3.2 GENERALISATION DU MODELE

Pour un grand nombre de stations, les variogrammes temporels de CHIMERE et des observations présentent la même allure (portées voisines pour la base du variogramme, même période de 24h pour la composante cyclique). Comme on l'a vu cependant (§5.1), la composante de grande portée et/ou la composante cyclique de CHIMERE se caractérisent, selon les sites, par une variabilité supérieure ou inférieure à celle des observations.

En conséquence, une adaptation logique du modèle de Chilès et Séguret est un **modèle linéaire de corégionalisation**. Dans ce modèle, tous les variogrammes s'expriment comme des combinaisons linéaires des mêmes structures élémentaires. Sont ici considérées les structures suivantes :

- effet de pépite σ_ε^2 (affecté au variogramme de Z uniquement, il représente l'erreur de mesure) ;
- structure exponentielle γ_Y (composante de grande portée)
- structure sinusoïdale γ_P (composante périodique)

En une station située au point x_i , les variogrammes s'écrivent donc :

$$\begin{cases} \gamma_Z(x_i, \tau) = \gamma_Y(x_i, \tau) + \gamma_P(x_i, \tau) + \sigma_\varepsilon^2(x_i) \\ \gamma_{CHIM}(x_i, \tau) = \omega_1(x_i)^2 \cdot \gamma_Y(x_i, \tau) + \omega_2(x_i)^2 \cdot \gamma_P(x_i, \tau) \\ \gamma_{Z.CHIM}(x_i, \tau) = \omega_1(x_i)\rho_1(x_i) \cdot \gamma_Y(x_i, \tau) + \omega_2(x_i)\rho_2(x_i) \cdot \gamma_P(x_i, \tau) \end{cases}$$

γ_Z est le variogramme des observations.

γ_{CHIM} est le variogramme de CHIMERE.

$\gamma_{Z.CHIM}$ est le variogramme croisé entre les observations et CHIMERE.

$\rho_1(x_i)$ représente le coefficient de corrélation entre les composantes exponentielles de Y (concentration réelle) et de CHIMERE.

$\rho_2(x_i)$ représente le coefficient de corrélation entre les composantes périodiques de Y et de CHIMERE.

$\omega_1(x_i)$ et $\omega_2(x_i)$ représentent le rapport des amplitudes entre CHIMERE et la mesure pour chacune des composantes.

A l'issue de nombreux essais, une procédure a été élaborée afin de définir automatiquement un modèle linéaire de corégionalisation par période et par station⁴ :

- *Evaluation de la grande portée de la composante exponentielle :*

On extrait la base du variogramme expérimental des observations en sélectionnant le premier point ainsi que les points multiples de 24 heures (qui correspondent au minimum de chaque sinusoïde) ; la modélisation de cette base par une structure exponentielle fournit une estimation de la grande portée.

- *Modélisation de la composante exponentielle :*

La base des variogrammes simples et du variogramme croisé des observations et de CHIMERE est modélisée par une structure exponentielle. La portée a la valeur précédemment établie ; seuls les paliers sont ajustés. Pour le variogramme des observations, l'effet de pépité est positif ou nul ; pour le variogramme de CHIMERE et le variogramme croisé, il est fixé à zéro. Deux modes d'ajustement sont testés. La méthode retenue est celle qui conduit au moindre écart par rapport aux variogrammes expérimentaux et qui assure la positivité de la matrice des paliers.

- *Calcul de la composante périodique:*

Les fonctions exponentielles précédemment ajustées sont discrétisées selon un pas horaire et soustraites aux variogrammes expérimentaux. Les variogrammes résiduels représentent la composante périodique.

- *Modélisation de la composante périodique :*

Une structure sinusoïdale est ajustée sur la composante périodique du variogramme des mesures, de CHIMERE et du croisé. La période du signal est fixée à 24 h ; son amplitude est estimée par l'amplitude expérimentale moyenne sur les deux premières sinusoïdes. Le cas échéant, cet ajustement est modifié de façon que la matrice des paliers soit positive.

Des modèles ainsi ajustés, on peut extraire aisément⁵ les valeurs de $\omega_1(x_i)$, $\rho_1(x_i)$, $\omega_2(x_i)$, $\rho_2(x_i)$ en chaque station $i \in \{1 \dots 219\}$ et en déduire le variogramme des innovations.

⁴ NB : la fonction *fit.lmc* du module *gstat* de R (Pesbema, 2004) permet l'ajustement d'un modèle linéaire de corégionalisation mais elle n'a pas donné entière satisfaction dans la modélisation simultanée des composantes exponentielle et périodique des variogrammes. En particulier, elle a tendance à surestimer l'amplitude de la composante périodique. C'est pourquoi une méthode d'ajustement a été spécifiquement développée.

5.1.3.3 VALIDATION DU MODELE

Si le modèle linéaire de corégionalisation au point x_i décrit correctement les relations entre CHIMERE et les mesures, alors le variogramme des innovations issu de ce modèle doit concorder avec le variogramme des innovations calculé expérimentalement.

Cette cohérence a été vérifiée pour les 2 étés 2005 et 2006 et les 219 stations d'analyse.

Dans la plupart des cas, le modèle linéaire de corégionalisation se révèle bien adapté (exemples de 2 stations : Figure 12, Figure 13).

Dans les cas restants, le variogramme des innovations modélisé s'écarte plus ou moins sensiblement du variogramme expérimental. Cela concerne principalement les stations implantées dans des zones géographiques complexes (montagne, bord de mer, cf. Figure 14) où la résolution du modèle CHIMERE ne permet pas la simulation des phénomènes locaux. Pour une minorité de stations, l'inadéquation du modèle linéaire de corégionalisation ne trouve pas d'explication évidente (exemple de la Figure 15). On notera également la singularité de la station 24015 (station de Grasse, cycle de période 12h très marqué en 2006, Figure 16).

⁵ en calcule pour chaque composante le rapport entre les paliers de γ_z , γ_{CHIM} et $\gamma_{z.CHIM}$.

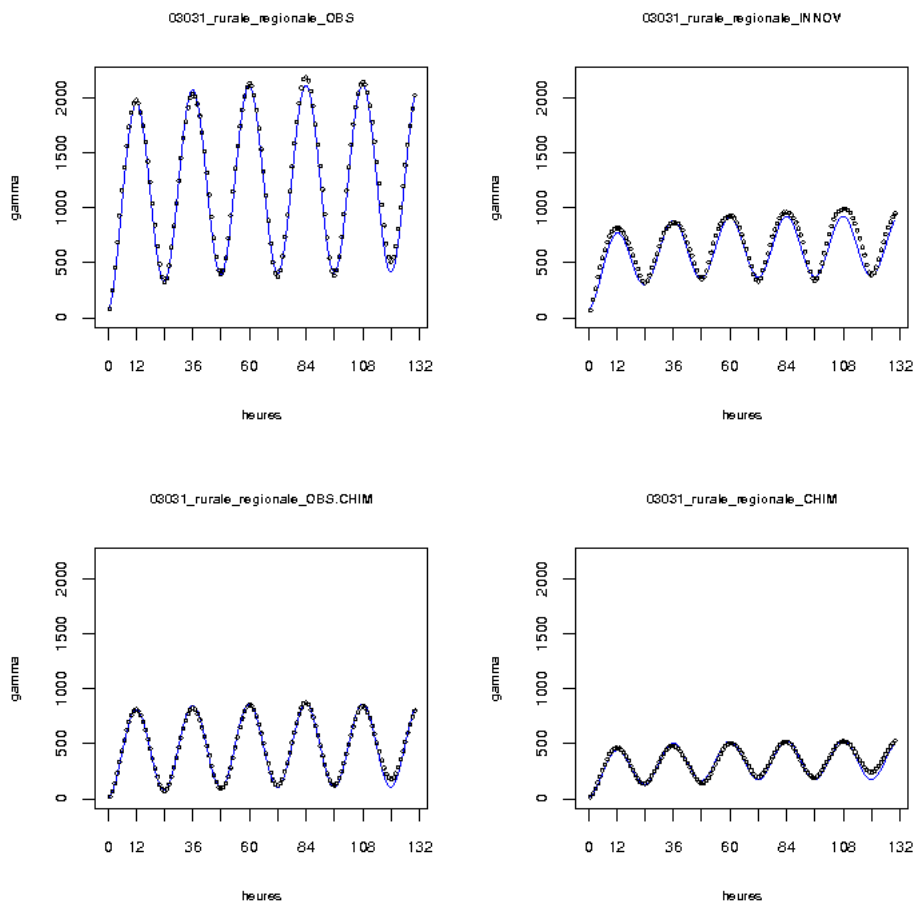


Figure 12 – Ajustement automatique d’un modèle linéaire de corégionalisation et calcul du variogramme des innovations (en haut à droite). Exemple de la station rurale 03031 en 2005. Points noirs : variogrammes expérimentaux ; trait continu : variogrammes modélisés. De gauche à droite et de haut en bas : observations ; innovations ; croisé observations-CHIMERE ; CHIMERE.

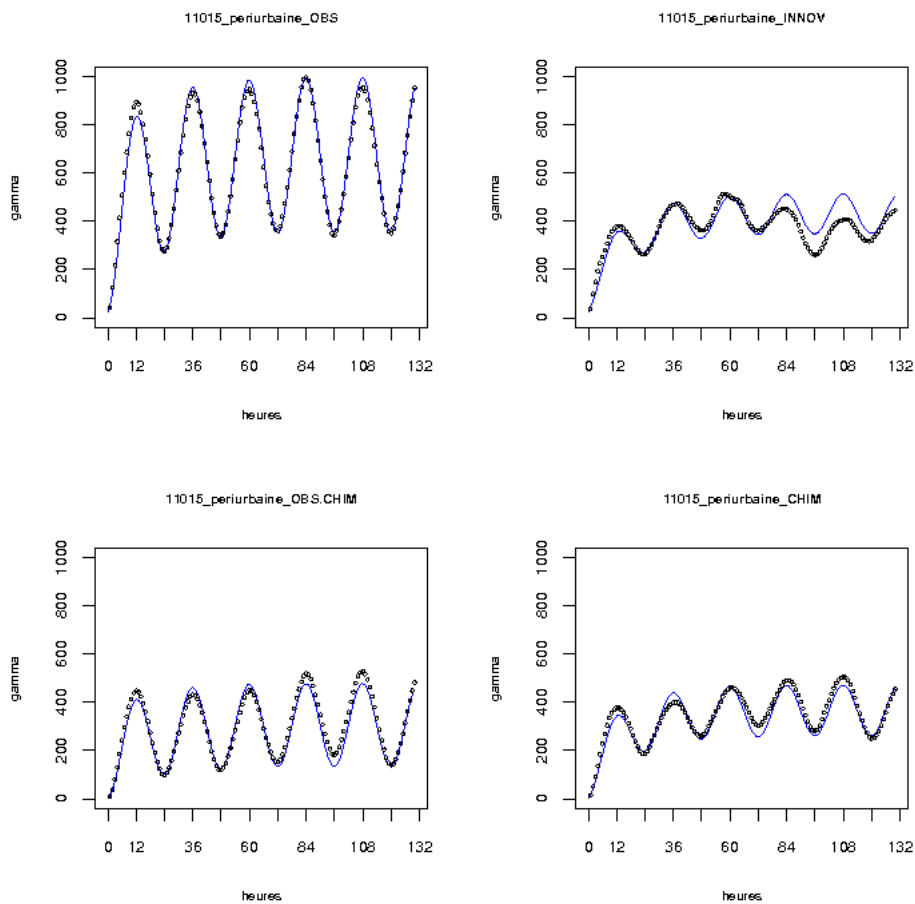


Figure 13 – Ajustement automatique d’un modèle linéaire de corégionalisation et calcul du variogramme des innovations. Exemple de la station périurbaine 11015 en 2005. Points noirs : variogrammes expérimentaux ; trait continu : variogrammes modélisés. De gauche à droite et de haut en bas : observations ; innovations ; croisé observations-CHIMERE ; CHIMERE.

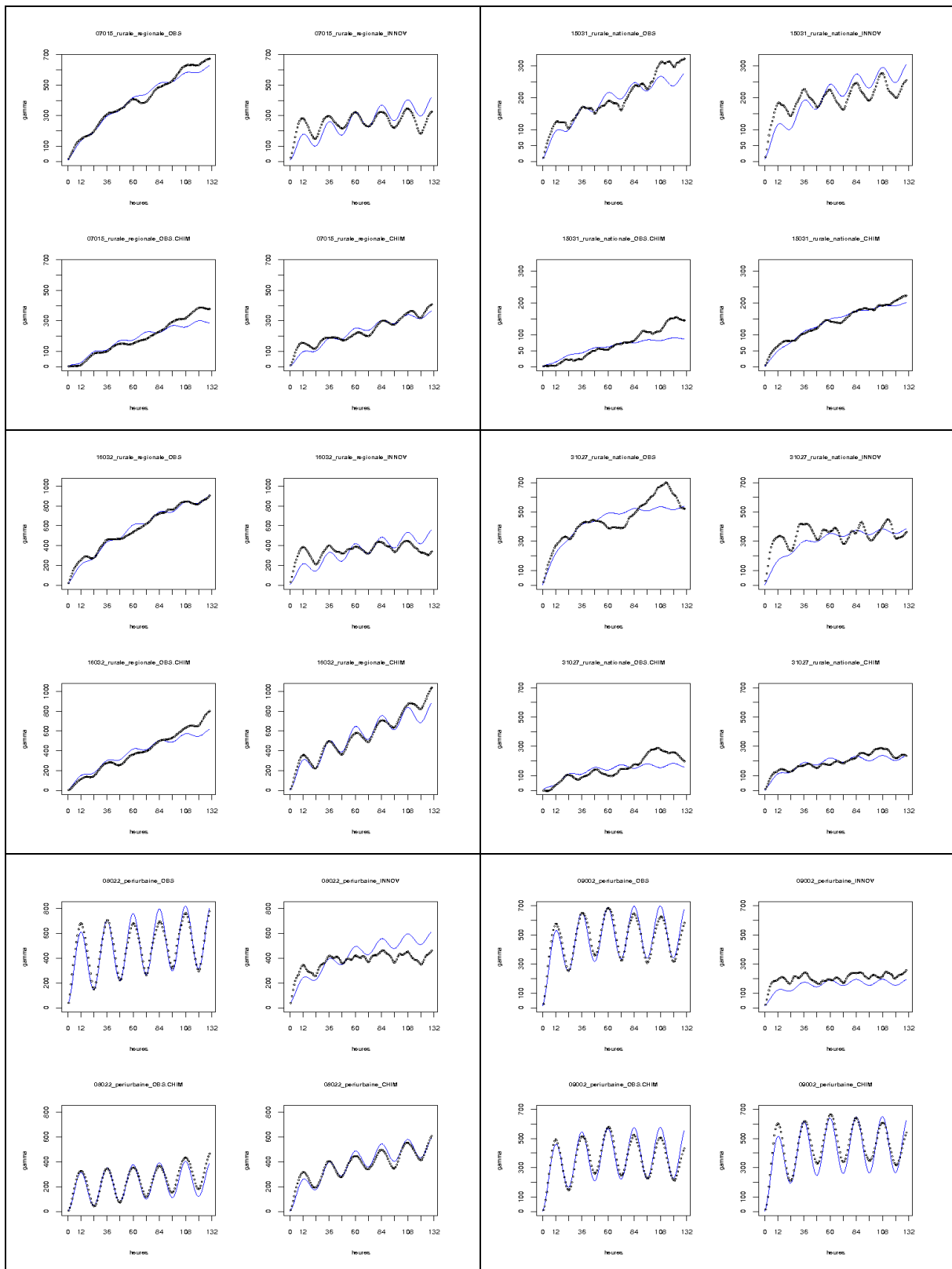


Figure 14 – Ajustement automatique d'un modèle linéaire de corégionalisation et calcul du variogramme pour les innovations. Exemple des stations rurales de montagne 07015 (1460 m), 15031 (1750 m), 16032 (1107 m) et 31027 (1400 m) et des stations de bord de mer 08022 et 09002, en 2006. Points noirs : variogrammes expérimentaux ; trait continu : variogrammes modélisés. De gauche à droite et de haut en bas : observations ; innovations ; croisé observations-CHIMERE ; CHIMERE.

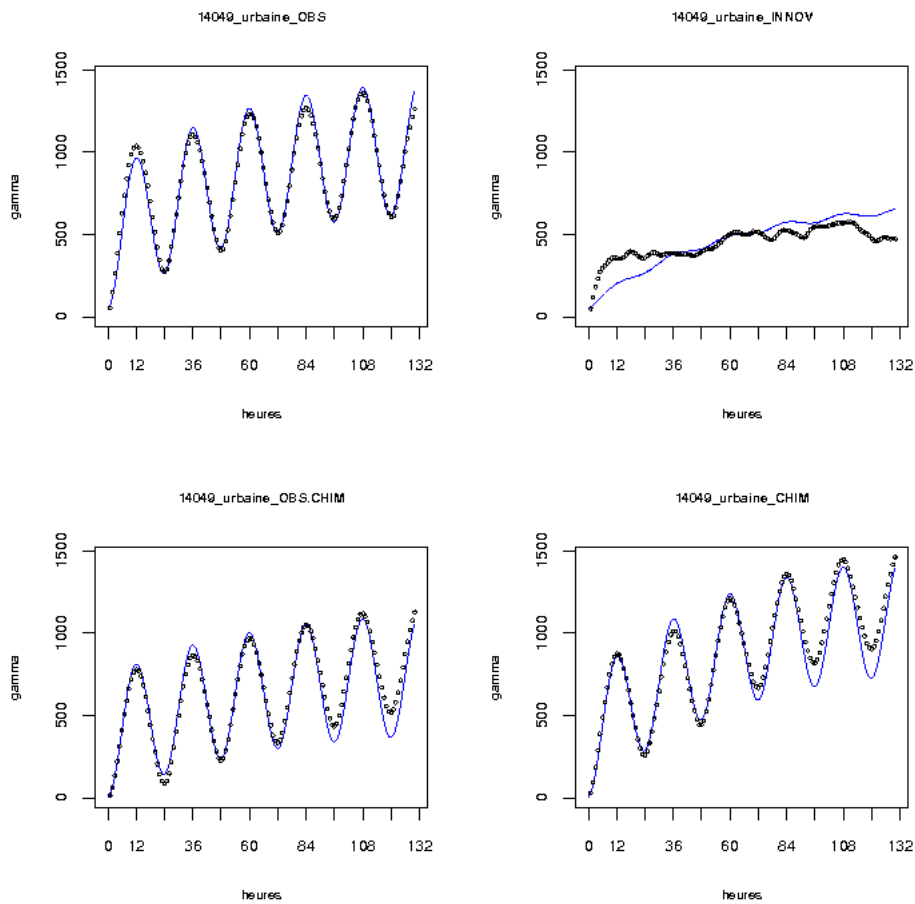


Figure 15 – Ajustement automatique d'un modèle linéaire de corégionalisation et calcul du variogramme pour les innovations. Exemple de la station périurbaine 14049 en 2006. Points noirs : variogrammes expérimentaux ; trait continu : variogrammes modélisés. De gauche à droite et de haut en bas : observations ; innovations ; croisé observations-CHIMERE ; CHIMERE.

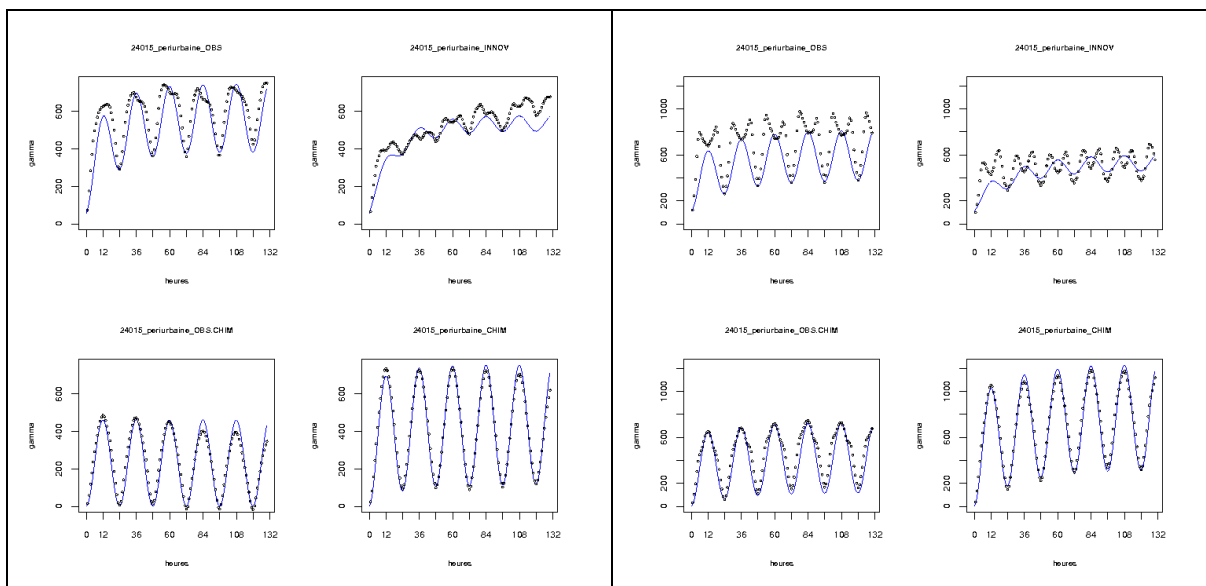


Figure 16 – Ajustement automatique d'un modèle linéaire de corégionalisation et calcul du variogramme pour les innovations. Exemple de la station périurbaine 24015 en 2005 (à gauche) et 2006 (à droite). Points noirs : variogrammes expérimentaux ; trait continu : variogrammes modélisés. De gauche à droite et de haut en bas : observations ; innovations ; croisé observations-CHIMERE ; CHIMERE.

5.1.3.4 ANALYSE DES COEFFICIENTS

Les coefficients du modèle linéaire de corégionalisation (ω_1 , ω_2 , ρ_1 et ρ_2) offrent l'intérêt de nous renseigner sur la variabilité de CHIMERE relativement à celle des observations (Tableau 1). En examinant les statistiques et la distribution spatiale de ces coefficients, nous pouvons ainsi évaluer si ce rapport de variabilité dépend de la zone géographique et du type de site.

L'annexe 4 présente l'ensemble des figures réalisées à cette fin : représentations cartographiques, statistiques, histogrammes, ACP, variogrammes.

Tableau 1 – Signification des coefficients du modèle linéaire de corégionalisation

$\omega_1 \approx 1$	CHIMERE simule correctement la variabilité de la composante de grande portée.
$\omega_2 \approx 1$	CHIMERE simule correctement la variabilité de la composante périodique.
$\omega_1 < 1$ ($\omega_1 > 1$)	CHIMERE sous-estime (surestime) la variabilité de la composante de grande portée.

$\omega_2 < 1$ ($\omega_2 > 1$)	CHIMERE sous-estime (surestime) la variabilité de la composante périodique.
$\rho_1 \approx 1$	La composante de grande portée de CHIMERE est bien corrélée à celle des observations (à l'échelle de plusieurs jours, CHIMERE et les mesures varient de manière corrélée).
$\rho_1 \ll 1$	Les variations de CHIMERE et des mesures à l'échelle de plusieurs jours sont peu corrélées.
$\rho_2 \approx 1$	La composante périodique de CHIMERE est bien corrélée à celle des observations (dans une journée, CHIMERE et les mesures varient de manière corrélée).
$\rho_2 \ll 1$	Les variations de CHIMERE et des mesures à l'échelle d'une journée sont peu corrélées.
Cas particulier : $\omega_1 \approx 1$ et $\omega_2 \approx 1$ $\rho_1 \ll 1$; $\rho_2 \ll 1$	CHIMERE simule correctement la variabilité temporelle des concentrations mais ne reproduit pas correctement leur évolution chronologique (« simulation non conditionnelle »).

Composante de grande portée (composante exponentielle)

Les composantes exponentielles de CHIMERE et des observations sont modérément corrélées: sur l'ensemble des stations, la moyenne et la médiane de ρ_1 sont voisines de 0,5.

En moyenne sur les stations, la variabilité de la composante exponentielle de CHIMERE est inférieure de 28% pour l'été 2005 et de 9% pour l'été 2006 à celle des observations (voir les statistiques de ω_1^2).

Interprétation

Pour la plupart des sites, l'adéquation du modèle linéaire de corégionalisation signifie qu'en moyenne, CHIMERE évalue correctement la portée temporelle des phénomènes. Cependant, les deux remarques précédentes nous indiquent que

- *d'un point de vue chronologique, les concentrations horaires simulées par CHIMERE n'évoluent pas toujours en parfaite corrélation avec la mesure ;*
- *CHIMERE tend à légèrement sous-estimer l'amplitude des variations sur le long terme.*

Les coefficients ω_1 et ρ_1 ne montrent pas de lien particulier avec la typologie des stations ni avec l'occupation du sol. Toutefois, les valeurs de ρ_1 et ω_1 s'organisent selon une certaine structure spatiale, comme on peut le voir sur les variogrammes et les représentations cartographiques.

Composante périodique

Les composantes périodiques de CHIMERE et des observations sont très corrélées entre elles: sur l'ensemble des stations, la moyenne et la médiane de ρ_1 sont respectivement supérieures à 0,9 et 0,95.

En moyenne sur les stations, la variabilité de la composante périodique de CHIMERE est inférieure de 46% pour l'été 2005 et de 30% pour l'été 2006 à celle des observations (voir les statistiques de ω_2^2).

Interprétation

Les fluctuations diurnes simulées par CHIMERE s'accordent temporellement avec la mesure mais leur amplitude est atténuée.

5.1.3.5 MODELE MOYEN PAR ZONE

Six zones relativement homogènes en ce qui concerne ω_1 et ρ_1 sont délimitées empiriquement sur la France pour l'été 2005. Un modèle linéaire de corégionalisation est ajusté sur les variogrammes expérimentaux moyens par zone (stations de montagne exclues) selon la procédure décrite au paragraphe 5.1.3.2. La bonne qualité de l'ajustement confirme l'efficacité de ce type de modèle.

Comme les coefficients varient d'une année à l'autre, le découpage établi pour l'été 2005 ne s'adapte pas tout à fait aux coefficients de l'été 2006. Le résultat de l'ajustement reste cependant satisfaisant (Figure 17).

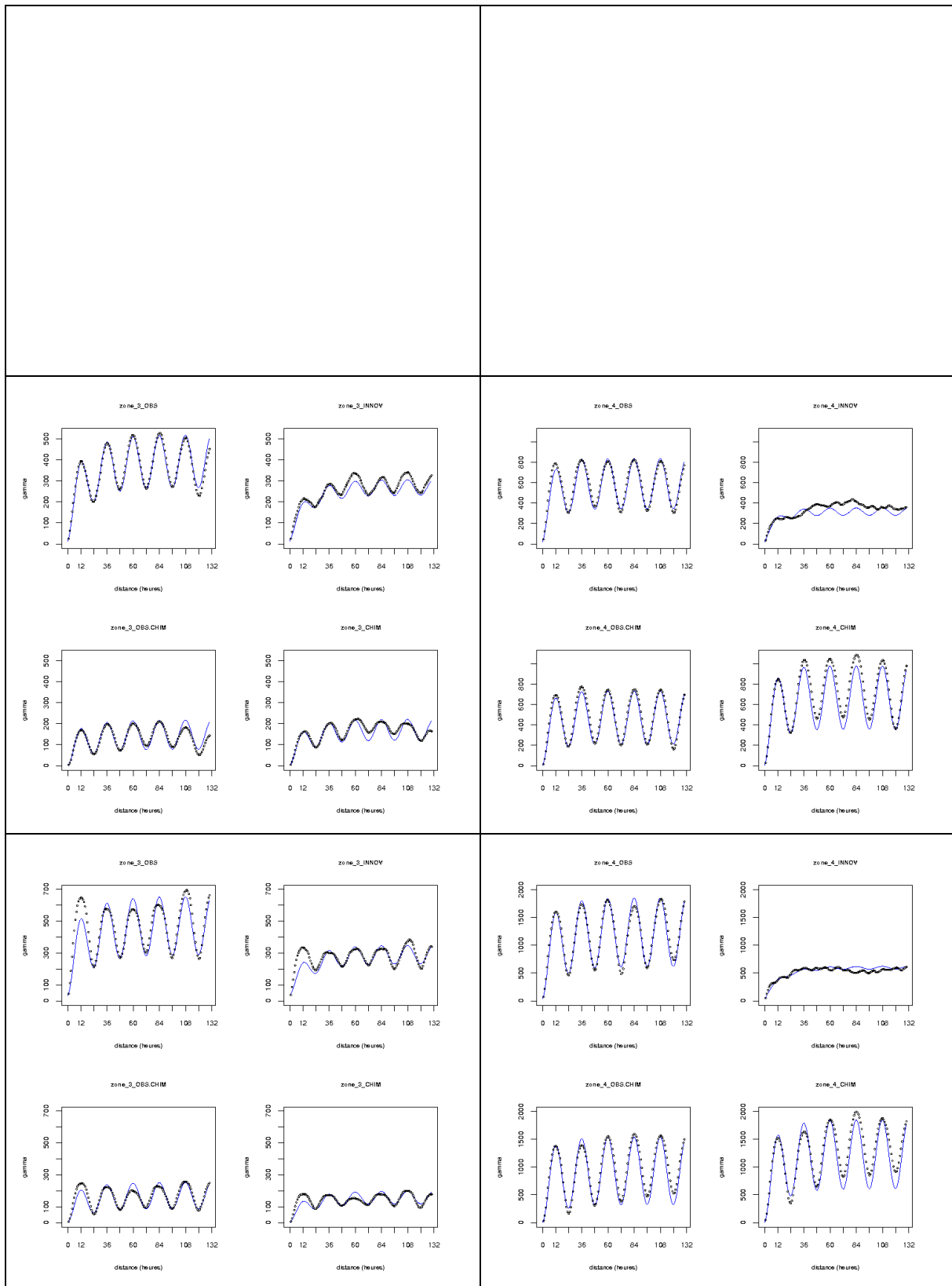


Figure 17 – Ajustement d'un modèle linéaire de corégionalisation sur les variogrammes expérimentaux moyens par zone. Exemple de deux zones : modèle ajusté pour l'été 2005 (2^e ligne) et l'été 2006 (3^e ligne)

Une perspective est de spatialiser les coefficients $\omega_1(x)$, $\rho_1(x)$, $\omega_2(x)$, $\rho_2(x)$ obtenus par station afin de modéliser les relations entre CHIMERE et la mesure en tout point du domaine et de développer, s'il est possible, un modèle spatio-temporel.

Il s'agit d'un problème complexe, c'est pourquoi, dans une première étape, l'aspect spatial de la question est abordé séparément (§ 5.2). La cohérence des modélisations temporelle et spatiale et la possibilité de les raccorder entre elles pourront être évaluées en 2009.

5.2 ETUDE SPATIALE : COMPARAISON PAR HEURE DU JOUR

5.2.1 REMARQUES PRELIMINAIRES

- 1) Du fait de leur singularité, les stations implantées à plus de 600 m d'altitude et la station 24015 ne sont pas prises en compte dans le calcul des variogrammes spatiaux mais elles sont utilisées en entrée du krigeage.
- 2) A cause de la densité inégale de stations, le variogramme de CHIMERE calculé sur les sites de mesure ne coïncide pas nécessairement avec le variogramme de CHIMERE calculé sur tous les points de grilles. Parfois, les deux variogrammes concordent mais généralement, le variogramme calculé sur la grille est plus linéaire.

Dans la suite, tous les variogrammes spatiaux, qu'ils soient simples ou croisés, seront calculés sur les mêmes points, *i.e.* sur les stations d'analyse. L'effet de la répartition préférentielle des sites sera évalué en 2009 par une étude de sensibilité.

5.2.2 VARIOGRAMMES EXPERIMENTAUX

Pour avoir une idée d'ensemble de la variabilité spatiale de CHIMERE et des mesures, nous avons tracé les 24 variogrammes horaires moyens sur chaque été⁶ (comme dans l'étude temporelle, les données de CHIMERE et de mesure correspondent respectivement aux heures h-1 et h).

Ces variogrammes, dont nous donnons ici quelques exemples (Figure 18), montrent qu'en moyenne :

- La continuité spatiale du phénomène augmente au cours de la journée (augmentation de la portée).
- La variabilité spatiale des concentrations modélisées par CHIMERE (en bleu sur les graphiques) est inférieure voire très inférieure à celle des observations (en noir) ; pour certaines heures, le variogramme des mesures révèle une structure de petite portée (30 à 40 km) que ne possède pas le variogramme de CHIMERE.

⁶ en pondérant les points de chaque variogramme par le nombre de couples ayant permis de les calculer.

- La variabilité spatiale des innovations (en rouge) est comparable voire supérieure à celle des observations ;
- La corrélation spatiale entre CHIMERE et les observations (variogramme croisé en vert) reste assez faible quelle que soit l'heure. En particulier, elle devient quasi nulle entre 20h et 21h puis se remet à augmenter aux environs de 6h.

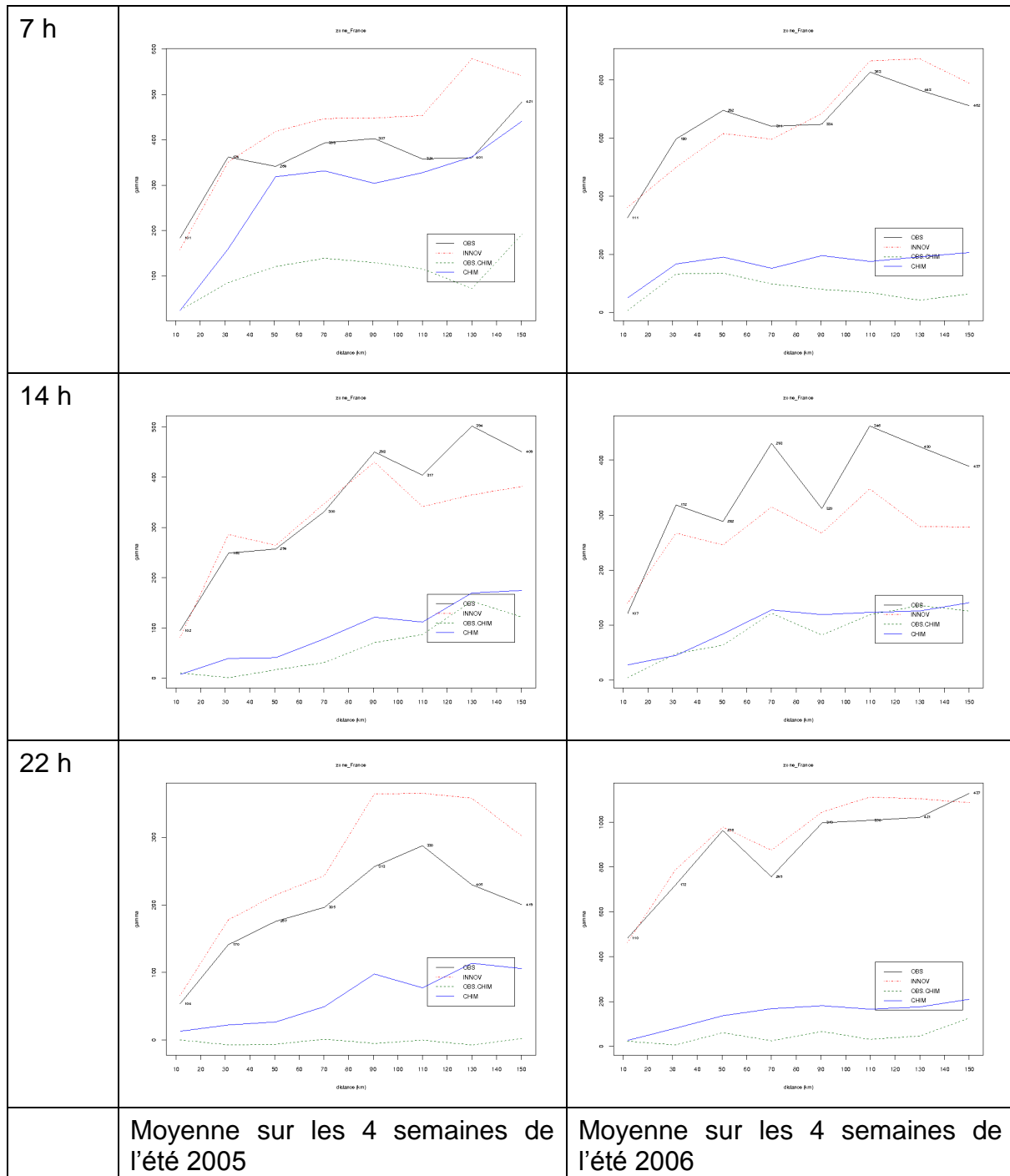


Figure 18 – Variogrammes spatiaux moyens par heure du jour : variogrammes simples des observations (en noir), de CHIMERE (en bleu), des innovations (en rouge) ; variogramme croisé des observations et de CHIMERE (en vert)

Ces constats restent généralement valables lorsqu'on examine les variogrammes heure par heure (voir l'exemple d'une journée en annexe 5). De cela découlent deux commentaires.

- **Dans un krigeage des innovations, la variabilité spatiale des concentrations analysées reflète principalement la variabilité spatiale des innovations.**
- Dans un cokrigeage entre CHIMERE et les mesures, les observations pèsent généralement plus que CHIMERE dans le calcul des concentrations analysées.

A la suite de premiers tests, nous proposons d'examiner en 2009 jusqu'à quel point la corrélation entre CHIMERE et les mesures peut être améliorée lorsqu'on calcule des variogrammes par zone (et non plus sur la France entière) et par direction de l'espace (étude anisotrope).

5.2.3 MODELE SPATIAL

Un modèle spatial est défini afin d'effectuer un cokrigeage entre CHIMERE et les observations.

Pour ce premier essai de modélisation, on se place dans le cadre du modèle de corrélation intrinsèque (5.1.3.1). A une heure t_0 , les variogrammes simples et croisé des observations et de CHIMERE s'écrivent :

$$\begin{cases} \gamma_Z(h, t_0) = \gamma_Y(h, t_0) + \sigma_\varepsilon^2(t_0) \\ \gamma_{CHIM}(h, t_0) = \omega_1(t_0)^2 \cdot \gamma_Y(h, t_0) \\ \gamma_{Z.CHIM}(h, t_0) = \omega_1(t_0)\rho_1(t_0) \cdot \gamma_Y(h, t_0) \end{cases}$$

γ_Z : variogramme des observations

γ_Y : composante de grande portée

σ_ε^2 : composante pépitique

γ_{CHIM} : variogramme de CHIMERE

$\gamma_{Z.CHIM}$: variogramme croisé observations-CHIMERE

Afin de permettre la comparaison avec le krigeage des innovations, le modèle CHIMERE et les observations sont considérés sans décalage temporel (conditions actuelles d'application du krigeage des innovations).

Le type de structure (choix d'une structure exponentielle) et sa portée (choix d'une valeur par heure du jour) sont fixés préalablement d'après l'étude des variogrammes expérimentaux moyens par été.

Les paliers des variogrammes sont ajustés dans R pour chaque heure de la période d'étude.

L'annexe 5 présente à titre d'illustration le résultat de cette modélisation pour une journée de l'été 2005. Les choix de modélisation (nombre de structures, portées

spatiales) pourront être affinés en 2009. En particulier, on pourra prendre en compte l'existence d'une structure de petite portée.

6. APPLICATION DU COKRIGEAGE

Les concentrations horaires d'ozone pendant les étés 2005 et 2006 ont été estimées par cokrigeage ordinaire entre CHIMERE et les observations, avec la mesure comme variable principale et CHIMERE comme variable auxiliaire. Le modèle spatial est le modèle de corrélation intrinsèque précédemment défini, dont les paliers sont réajustés heure par heure (5.2.3).

Trois groupes de stations sont distingués :

- les stations d'analyse utilisées dans le calcul des variogrammes et en entrée du krigeage ;
- les stations d'analyse exclues du calcul des variogrammes mais utilisées en entrée du krigeage ;
- les stations de vérification.

Afin de pouvoir comparer rigoureusement les méthodes, on a réappliqué le krigeage des innovations en distinguant ces mêmes groupes de stations.

Pour la validation croisée et la vérification (cf. §6.1), un cokrigeage (ou un krigeage des innovations) ponctuel a été réalisé, ce qui permet de confronter l'estimation et la mesure en chaque point d'observation. Pour la cartographie, il convient d'établir un cokrigeage ou un krigeage de blocs de façon que l'estimation finale ait le même support spatial que CHIMERE.

6.1 EVALUATION STATISTIQUE

La qualité de l'estimation est évaluée :

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

Pour chaque méthode (simulation déterministe brute, krigeage des innovations, cokrigeage) et chaque été (2005 et 2006), pour la validation croisée d'une part et pour la vérification d'autre part, différentes statistiques d'erreur sont calculées :

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

En moyenne sur les périodes et les stations, le cokrigeage réduit significativement l'écart entre le modèle et les mesures (Figure 19, Figure 20, Figure 21, Figure 22), y compris sur les sites urbains (diminution de l'erreur, augmentation de la corrélation).

D'un point de vue statistique, son efficacité est comparable à celle du krigeage des innovations.

Validation croisée

Cokrigeage (été 2005 et été 2006) : Sur l'ensemble des stations d'analyse, la moyenne des erreurs en valeur absolue diminue de 45%, la RMSE diminue de 40%, le coefficient de corrélation augmente de 25% (2005) et 17% (2006).

Krigeage des innovations (été 2005 et été 2006) : Sur l'ensemble des stations d'analyse, la moyenne des erreurs en valeur absolue diminue de 47% et 46%, la RMSE diminue de 42% et 41%, le coefficient de corrélation augmente de 26% et 17%.

Vérification

Cokrigeage (été 2005 et été 2006) : Sur l'ensemble des stations de vérification, la moyenne des erreurs en valeur absolue diminue de 52 et 51%, la RMSE diminue de 49 et 47%, le coefficient de corrélation augmente de 37 et 23%.

Krigeage des innovations (été 2005 et été 2006) : Sur l'ensemble des stations de vérification, la moyenne des erreurs en valeur absolue diminue de 53 et 51%, la RMSE diminue de 49 et 46%, le coefficient de corrélation augmente de 37 et 23% comme pour le cokrigeage.

Des statistiques plus détaillées par typologie sont fournies dans les tableaux de l'annexe 6.

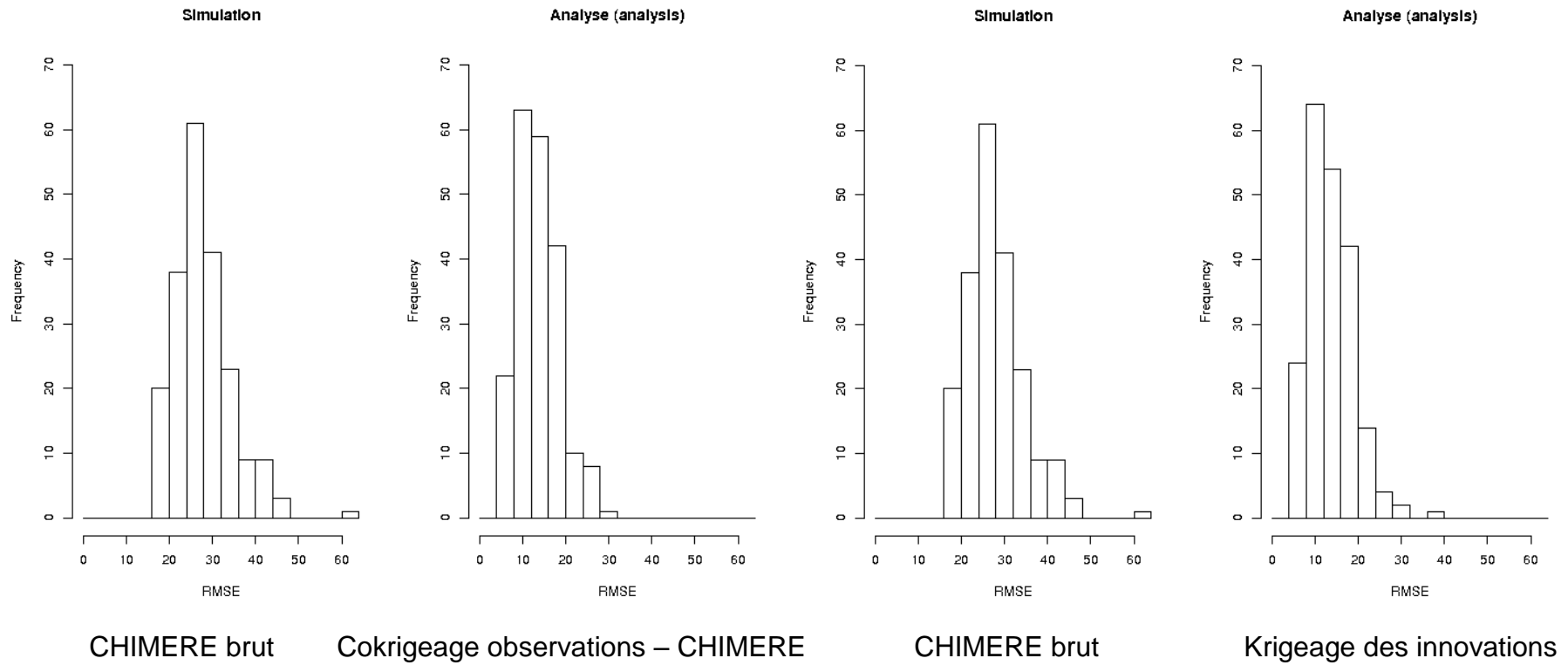


Figure 19 – Été 2005. Distribution de la RMSE calculée station par station sur l'ensemble des sites de vérification.

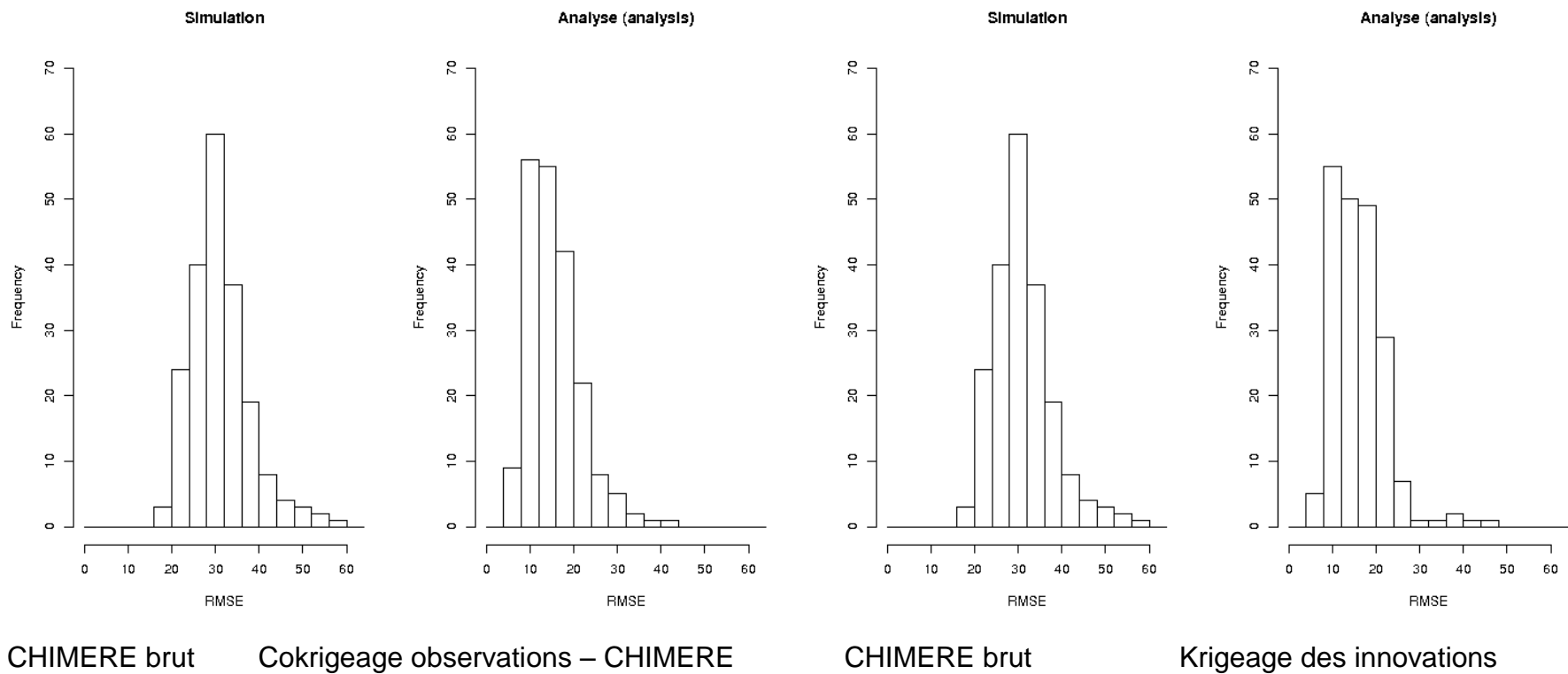


Figure 20 – Été 2006. Distribution de la RMSE calculée station par station sur l'ensemble des sites de vérification.

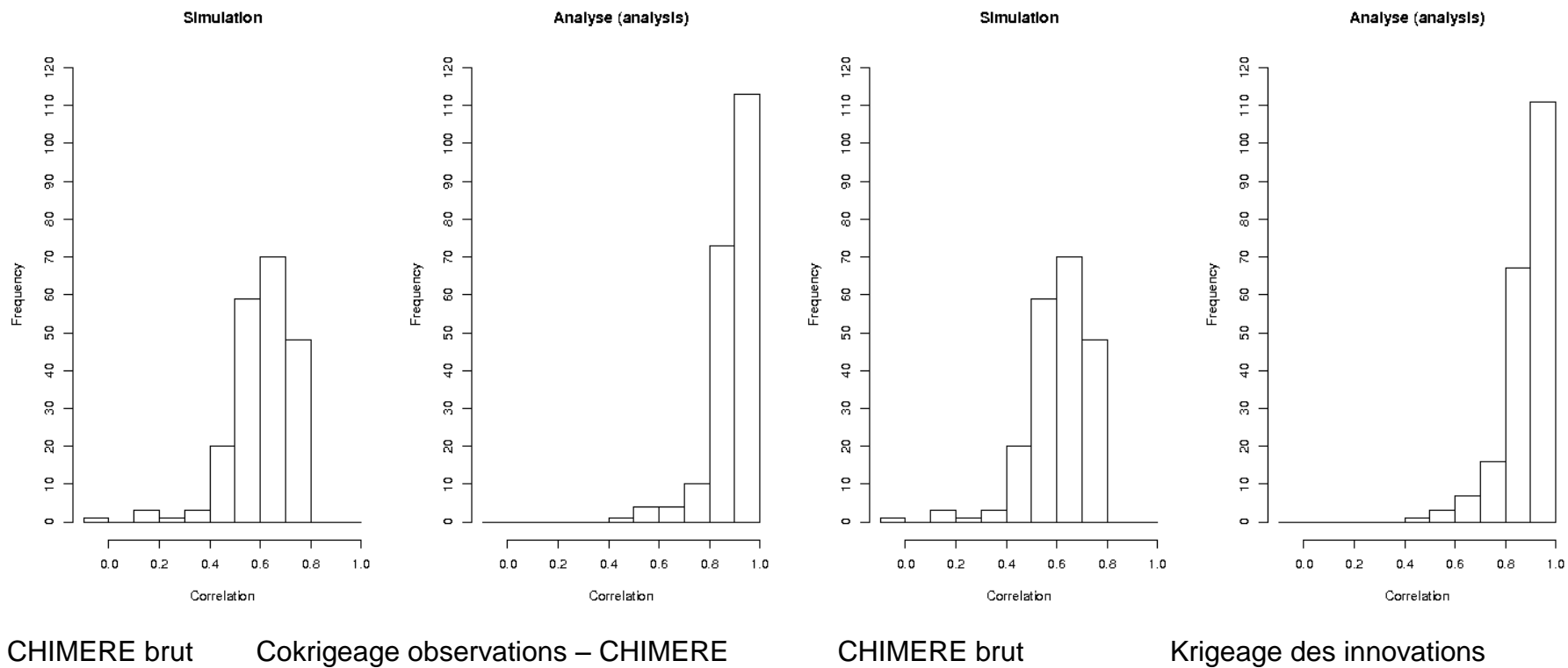


Figure 21 – Été 2005. Distribution du coefficient de corrélation calculé station par station sur l'ensemble des sites de vérification.

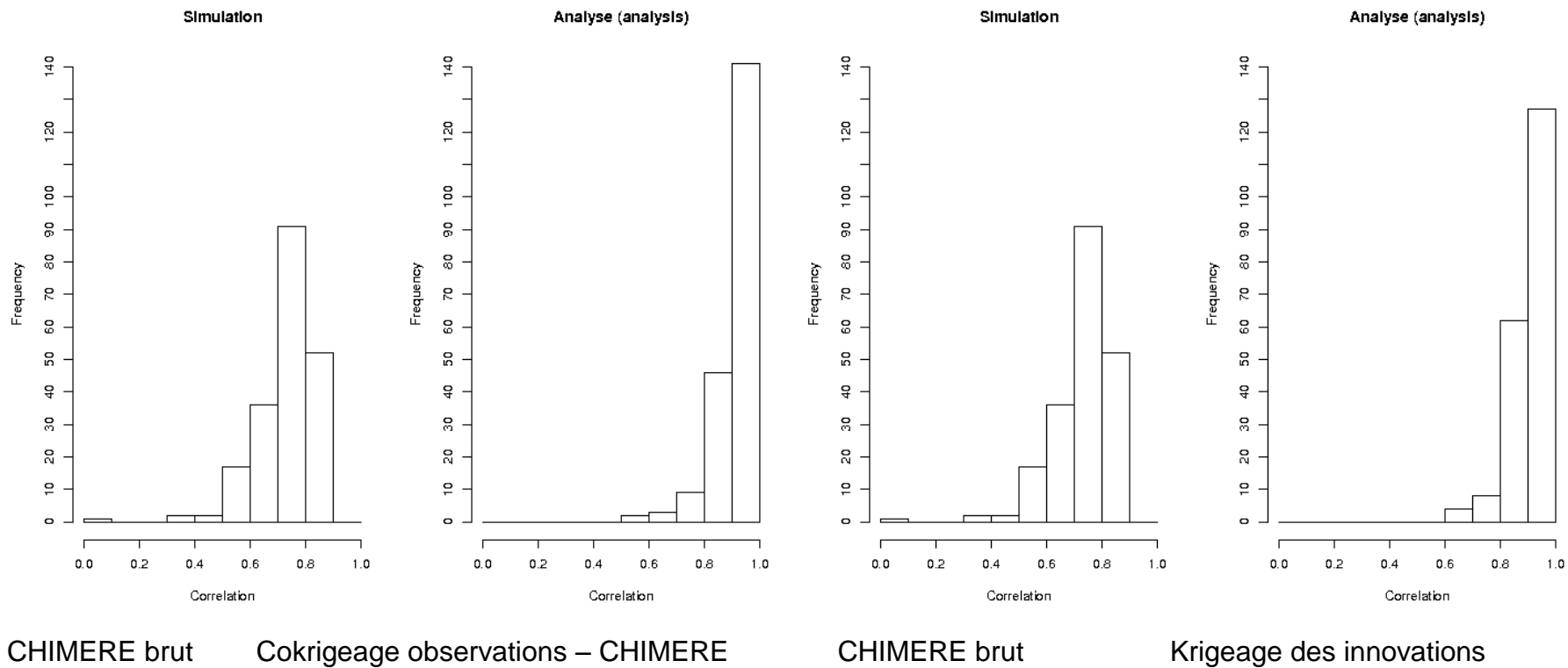


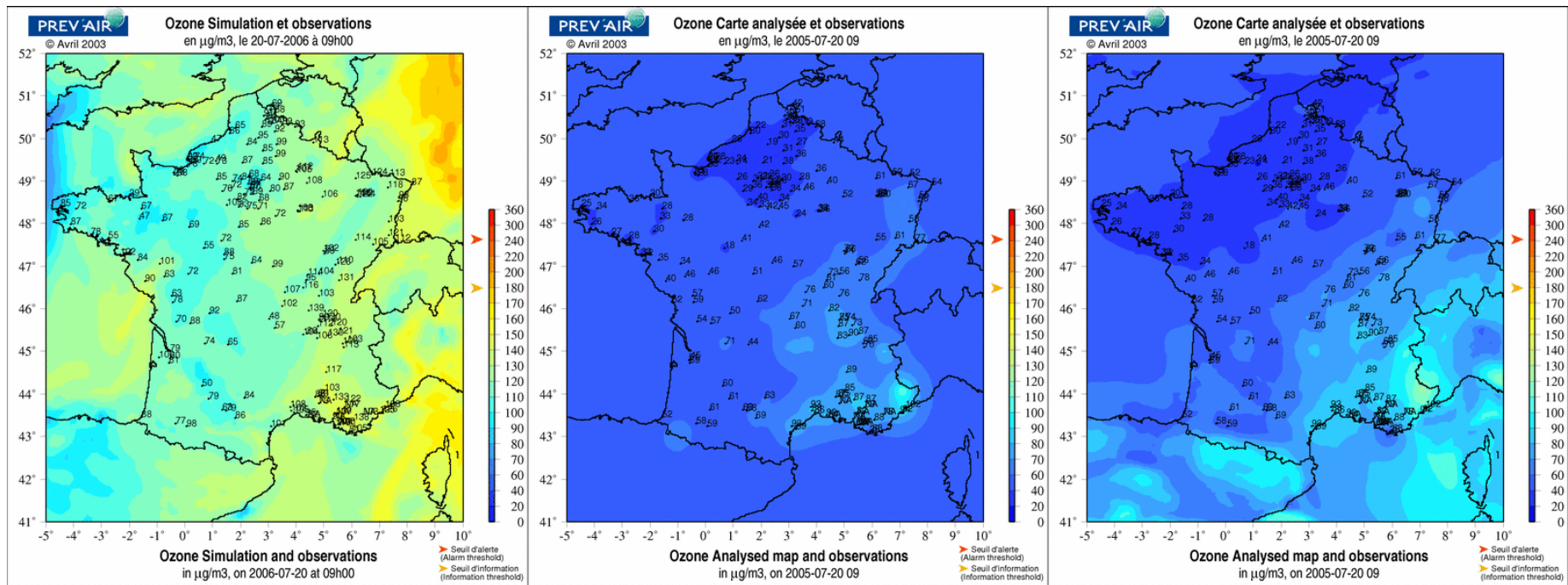
Figure 22 – Eté 2006. Distribution du coefficient de corrélation calculé station par station sur l'ensemble des sites de vérification.

6.2 CARTOGRAPHIES

Les cartographies établies par cokrigeage sont plus lisses que les cartographies issues du krigeage des innovations (Figure 23). Des différences de voisinage d'estimation pourraient en être notamment la cause. Cette question nécessite d'être examinée en 2009.

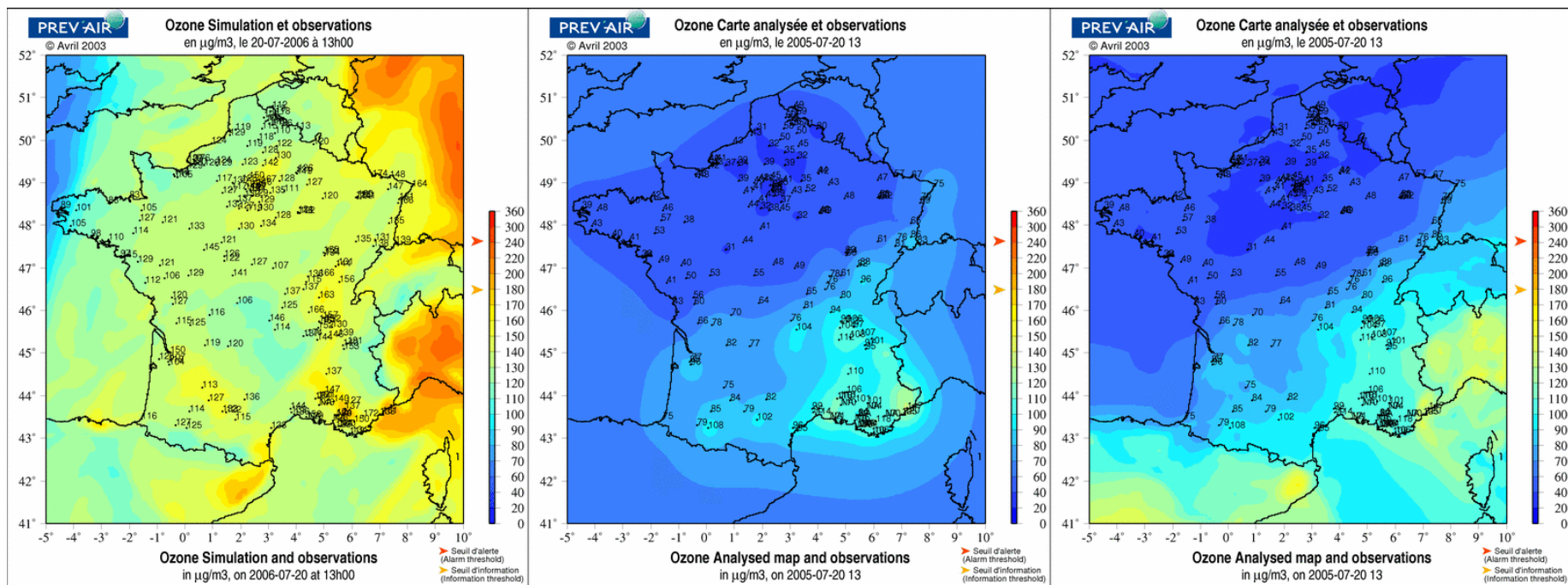
Lorsque la corrélation spatiale (à l'échelle de la France) entre CHIMERE et les mesures est faible, les observations ont plus de poids que CHIMERE : le cokrigeage ressemble logiquement à un krigeage des observations.

La modélisation spatiale bivariable n'a pas été suffisamment approfondie pour qu'on puisse entièrement apprécier les avantages et limites du cokrigeage. Un nouveau bilan sera établi en 2009 à la suite des améliorations proposées (cf. §5.2).



Concentrations d'ozone 1) simulées par CHIMERE ; 2) après cokrigage ; 3) après krigeage des innovations.

Figure 23 – Cartographie des concentrations d'ozone. 20 juillet 2005, 9h (faible corrélation spatiale entre les observations et CHIMERE, cf. annexe 4).



Concentrations d'ozone 1) simulées par CHIMERE ; 2) après cokrigeage ; 3) après krigeage des innovations.

Figure 24 – Cartographie des concentrations d'ozone. 20 juillet 2005, 13h (il existe une corrélation spatiale entre les observations et CHIMERE, cf. annexe 4).

7. BILAN ; VOIES D'AMELIORATION

Cette étude qui se poursuit en 2009 a pour objet de mettre à jour, et s'il est possible d'améliorer, la méthode d'analyse utilisée dans PREV'Air.

Le travail réalisé en 2008 porte sur la cartographie horaire des concentrations d'ozone. Constatant que l'hypothèse du modèle à résidu, sur laquelle repose le krigeage des innovations, n'était pas toujours vérifiée, nous avons cherché à caractériser les relations de variabilité entre le modèle CHIMERE et les mesures des stations.

219 stations dont la représentativité spatiale peut être comparée à la maille CHIMERE ont été sélectionnées à cette fin d'après les informations communiquées par les AASQA.

Dans une première partie, la **variabilité temporelle** de CHIMERE, des observations et des mesures a été analysée station par station. On montre qu'elle est généralement bien décrite par un modèle géostatistique appelé « modèle linéaire de corégionalisation ». Dans cette étude, ce modèle comprend deux composantes : une composante de grande portée (2 à 3 jours), que l'on peut relier au temps de dispersion des panaches, et une composante cyclique (période : 24h) caractéristique des variations diurnes. Un tel modèle a été ajusté automatiquement en chacune des 219 stations d'analyse. La plupart des stations pour lesquelles il est inadapté sont situées en terrain complexe (ex : montagne), dans des zones où CHIMERE et la mesure réagissent différemment. Pour un petit nombre de stations, l'inadéquation du modèle linéaire de corégionalisation reste inexplicée.

Il est intéressant d'étudier les paramètres ou coefficients des modèles de corégionalisation ajustés par station : leurs valeurs nous renseignent, pour chaque échelle de temps (grande portée, cycle journalier) et en fonction du lieu, sur la façon dont CHIMERE est corrélé à la mesure et varie relativement à celle-ci. Cette étude montre ainsi :

- que les composantes périodiques de CHIMERE et des observations sont très corrélées entre elles (corrélation moyenne supérieure à 0,9) et que CHIMERE atténue le plus souvent l'amplitude du signal journalier (de 30 à 50% en moyenne) ;
- que les composantes de grande portée sont modérément corrélées entre elles (corrélation moyenne voisine de 0,5) et que CHIMERE, en moyenne, atténue légèrement l'amplitude des variations.

Dans une seconde partie, la **variabilité spatiale** de CHIMERE a été analysée heure par heure. Pour la modéliser, on s'est placé à nouveau dans le cadre du modèle linéaire de corégionalisation. En première approche, on a réduit ce modèle à une structure exponentielle unique dont la portée a été ajustée selon l'heure du jour (un effet de pépite a été pris en compte pour les observations). Ce modèle a été évalué et appliqué dans un cokrigeage ordinaire entre CHIMERE et les observations. D'après les statistiques de test (validation croisée, validation sur les

stations complémentaires des 219 sites), les concentrations estimées par cokrigeage s'accordent mieux avec la mesure que les résultats bruts de CHIMERE. En revanche, ces statistiques n'indiquent pas de gain notable par rapport au krigeage des innovations.

Si les statistiques d'erreur associées respectivement au cokrigeage ordinaire et au krigeage des innovations sont comparables, en revanche, les cartographies diffèrent sensiblement. Les cartes établies par cokrigeage sont plus lisses.

En 2009, nous proposons d'approfondir les points suivants :

- La corrélation spatiale entre CHIMERE et les observations peut être faible voire quasi nulle. Dans ce cas le cokrigeage est proche du krigeage ordinaire des observations : le modèle CHIMERE, que l'on prétend corriger, ne contribue pratiquement pas à la cartographie, ce qui est problématique.
 - Cette corrélation spatiale s'améliore-t-elle si l'on calcule des variogrammes par zone géographique au lieu de variogrammes globaux sur la France ? si l'on calcule des variogrammes anisotropes ? si l'on autorise des décalages temporels de 0 à ± 2 heures entre CHIMERE et les observations ? Dans cette éventualité, l'objectif de la cartographie doit être revu : à une heure h donnée, veut-on corriger la simulation CHIMERE de l'heure h ? Ou cherche-t-on à élaborer une cartographie la plus précise possible, à partir des observations de l'heure h et d'une simulation (ou d'une moyenne de simulations) qui leur est bien corrélée ?
- Le modèle spatial de corégionalisation pourrait prendre en compte la structure de petite portée observable sur le variogramme des mesures.
- Les cartes de variance d'estimation (variance de cokrigeage ou de krigeage) seront exploitées. En s'aidant des résultats de la validation croisée et de la vérification, on évaluera si la variance d'estimation représente correctement la variance d'erreur expérimentale et si elle peut être ou non utilisée comme mesure de l'incertitude. Si tel est le cas, on pourra s'appuyer sur les cartes de variance d'estimation pour identifier les parties moins fiables des cartes analysées.
- Enfin, une réflexion sur la manière de raccorder les modélisations temporelle et spatiale pourrait être amorcée.

8. 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.

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

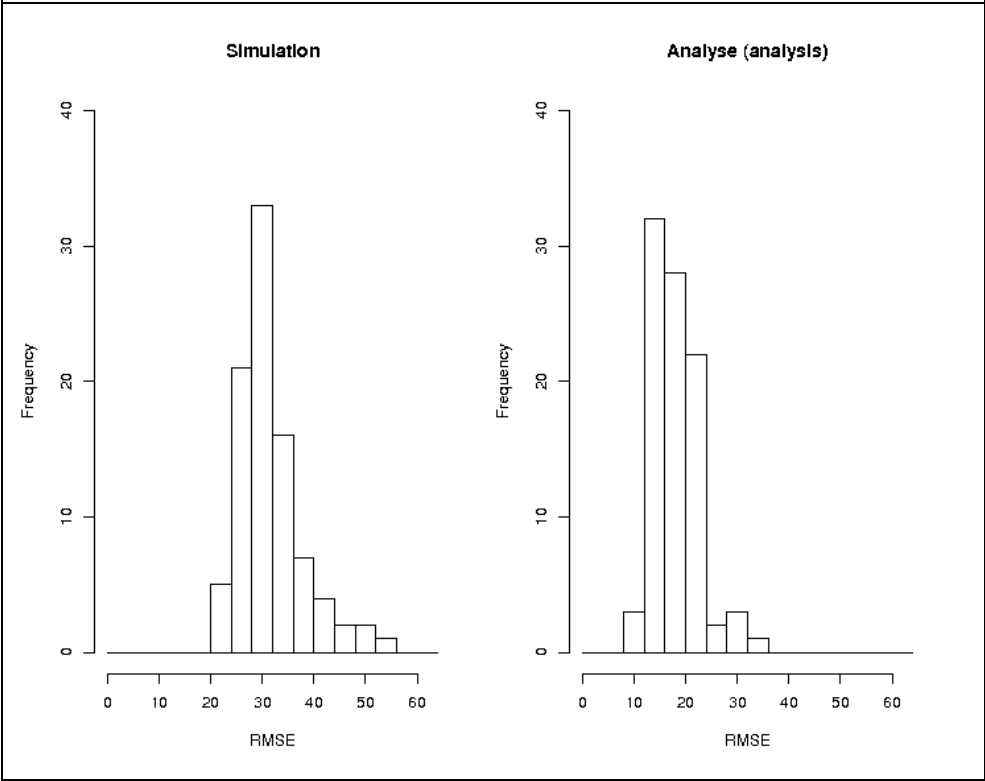
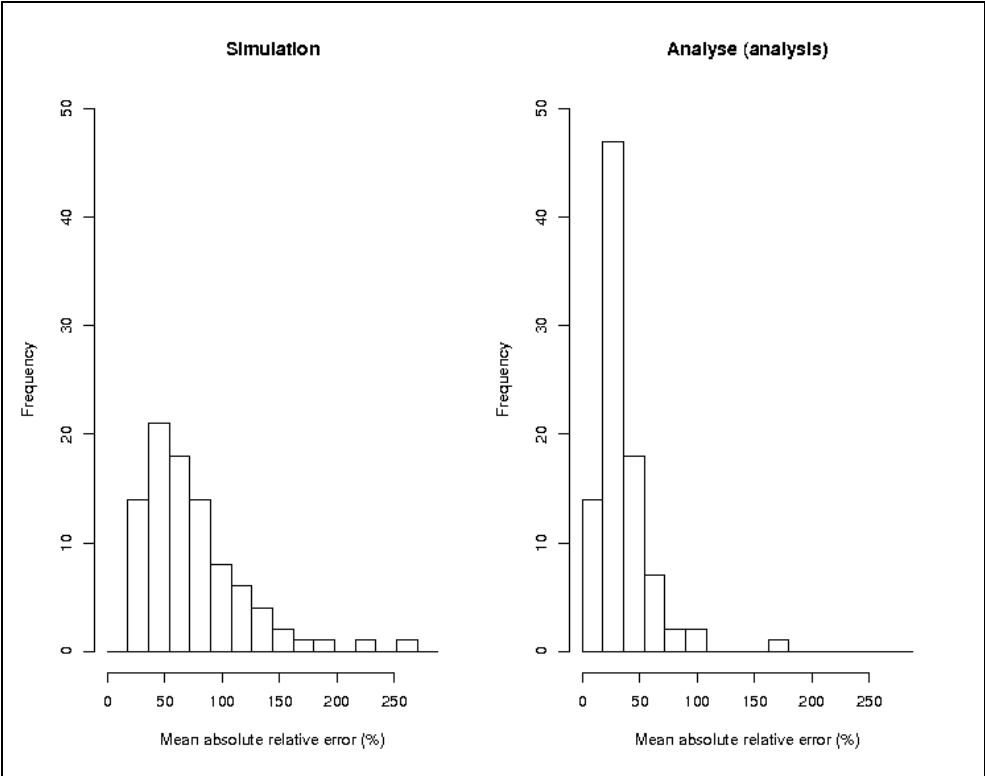
Rossiter D. G., 2007. Technical note. Co-kriging with the gstat package of the R environment for statistical computing. Version 2.1, January 2007, 81 pp. <http://www.itc.nl/personal/rossiter>.

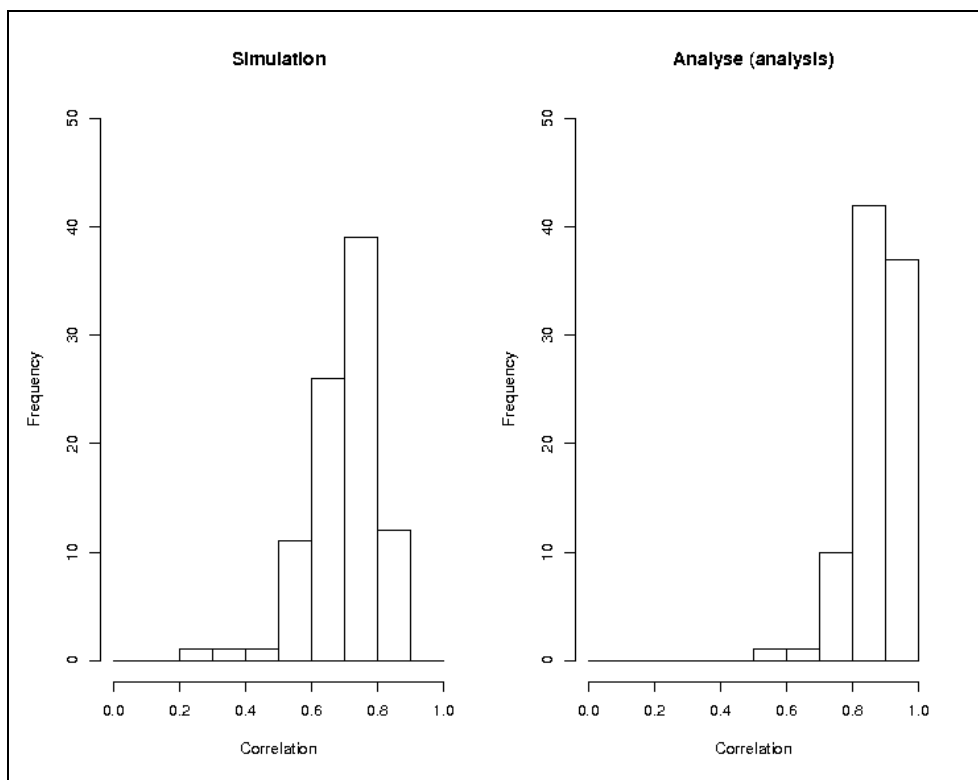
9. LISTE DES ANNEXES

Repère	Désignation	Nombre de pages
Annexe 1	Vérification initiale du krigeage des innovations	3
Annexe 2	Statistiques du coefficient de corrélation par station entre CHIMERE et les observations	4
Annexe 3	Présentation du modèle de corrélation intrinsèque	3
Annexe 4	Modélisation temporelle : cartographies des coefficients du modèle linéaire de corégionalisation	30
Annexe 5	Variogrammes spatiaux heure par heure sur la France. Exemple de la journée du 20 juillet 2005.	4
Annexe 6	Quantification du gain lié à l'analyse. Krigeage des innovations et cokrigeage ordinaire entre les observations et CHIMERE.	4

ANNEXE 1

Vérification initiale du krigeage des innovations





*Figure 25 – Comparaison entre CHIMERE et les observations avant (fig. de gauche) et après (fig. de droite) krigeage des innovations.
 Histogrammes de différentes statistiques de comparaison calculées station par station : de haut en bas, erreur relative en valeur absolue, racine carrée de l'erreur moyenne quadratique, coefficient de corrélation.
 Validation sur les sites de vérification, toutes typologies confondues. Eté 2006.*

ANNEXE 2

Statistiques du coefficient de corrélation par station entre le modèle CHIMERE et les mesures

Les coefficients de corrélation ont été calculés station par station pour différentes plages horaires (00-23h ; 1-4h ; 5-8h ; 9-12h ; 13-18h ; 19-21h ; 22-0h)

- sur la période allant du 18 juillet au 14 août 2005
- sur la période allant du 18 juillet au 14 août 2006

Comme on le voit dans les tableaux ci-après (sont donnés pour exemple les résultats pour toutes les heures confondues et pour les heures de l'après-midi), **les corrélations sont légèrement supérieures** en moyenne entre les **mesures à l'heure h+1 et CHIMERE (interpolé aux stations) à l'heure h**. On retrouve ce résultat pour les différentes plages horaires et les différentes typologies de station.

Autrement dit, CHIMERE semble légèrement en avance par rapport aux mesures.

2005 : 00-23h

Typologie		Mesures à l'heure h-1 et données CHIMERE à l'heure	Mesures à l'heure h et données CHIMERE à l'heure h	Mesures à l'heure h+1 et données CHIMERE à l'heure h
R	Nombre de stations	57	57	57
R	P50	0.603	0.642	0.671
R	Moy	0.582	0.627	0.639
PU	Nombre de stations	71	71	71
PU	P50	0.605	0.667	0.694
PU	Moy	0.592	0.649	0.663
U	Nombre de stations	77	77	77
U	P50	0.593	0.660	0.681
U	Moy	0.586	0.649	0.669

2005 : 13-18h

Typologie		Mesures à l'heure h-1 et données CHIMERE à l'heure h	Mesures à l'heure h et données CHIMERE à l'heure h	Mesures à l'heure h+1 et données CHIMERE à l'heure h
R	Nombre de stations	57	57	57
R	P50	0.603	0.642	0.671
R	Moy	0.582	0.627	0.639
PU	Nombre de stations	71	71	71
PU	P50	0.605	0.667	0.694
PU	Moy	0.769	0.800	0.797
U	Nombre de stations	77	77	77
U	P50	0.593	0.660	0.681
U	Moy	0.586	0.649	0.669

2006 : 00-23h

Typologie		Mesures à l'heure h-1 et données CHIMERE à l'heure h	Mesures à l'heure h et données CHIMERE à l'heure h	Mesures à l'heure h+1 et données CHIMERE à l'heure h
R	Nombre de stations	58	58	58
R	P50	0.765	0.798	0.799
R	Moy	0.736	0.764	0.769
PU	Nombre de stations	75	75	75
PU	P50	0.742	0.783	0.796
PU	Moy	0.714	0.756	0.764
U	Nombre de stations	78	78	78
U	P50	0.593	0.660	0.681
U	Moy	0.710	0.753	0.763

2006 : 13-18h

Typologie		Mesures à l'heure h-1 et données CHIMERE à l'heure h	Mesures à l'heure h et données CHIMERE à l'heure h	Mesures à l'heure h+1 et données CHIMERE à l'heure h
R	Nombre de stations	58	58	58
R	P50	0.765	0.798	0.799
R	Moy	0.736	0.764	0.769
PU	Nombre de stations	75	75	75
PU	P50	0.714	0.756	0.764
PU	Moy	0.734	0.780	0.796
U	Nombre de stations	78	78	78
U	P50	0.710	0.765	0.781
U	Moy	0.695	0.733	0.739

ANNEXE 3

Présentation du modèle de corrélation intrinsèque

Les notations sont les suivantes :

- $Z(x,t)$ représente la mesure à la date t , en la station implantée au « point » x ;
- $Y(x,t)$ représente la variable d'intérêt (état réel) ;
- $S(x,t)$ représente la simulation CHIMERE.

La mesure s'écrit $Z_{x,t} = Y_{x,t} + \varepsilon_{x,t}$

où l'erreur de mesure $\varepsilon_{x,t}$ est supposée d'espérance nulle et de variance σ_ε^2 constante, et est indépendante d'une mesure à l'autre. Cette erreur de mesure se traduit par une composante pépitique d'amplitude σ_ε^2 sur le variogramme expérimental spatial ou temporel des mesures Z .

L'hypothèse fondamentale est la suivante : la simulation S et la variable Y sont en corrélation intrinsèque, c'est-à-dire que leurs variogrammes simples et croisés sont proportionnels. Ceci peut s'écrire :

$$S_{x,t} = \omega\rho Y_{x,t} + \omega\sqrt{1-\rho^2} R_{x,t} \quad (3)$$

R est une variable aléatoire indépendante de Y mais dont le variogramme est proportionnel à celui de S ou à celui de Y .

Dans ce modèle, ρ est le coefficient de corrélation « caché », car inaccessible en pratique, entre la variable Y et la simulation CHIMERE S .

ω^2 représente le rapport de variabilité entre la simulation CHIMERE S et la variable Y .

$\gamma_{h,\tau}$ désignant le variogramme de la variable Y , ceux de la simulation S et de la mesure Z s'écrivent respectivement :

$$\gamma_Z(h,\tau) = \sigma_\varepsilon^2 + \gamma_{h,\tau} \quad (4)$$

$$\gamma_S(h,\tau) = \omega^2 \gamma_{h,\tau} \quad (5)$$

et le variogramme croisé entre les données et CHIMERE s'écrit :

$$\begin{aligned} \gamma_{ZS}(h,\tau) &= \frac{1}{2} E \left[Z_{x+h,t+\tau} - Z_{x,t} \cdot S_{x+h,t+\tau} - S_{x,t} \right] \\ &= \omega\rho \gamma_{h,\tau} \end{aligned} \quad (6)$$

En introduisant la différence ou « innovation » $D=Z-S$ entre la mesure et la simulation, on a

$$D_{x,t} = 1 - \omega\rho Y_{x,t} - \omega\sqrt{1-\rho^2} R_{x,t} + \varepsilon_{x,t} \quad (7)$$

Chacun des termes étant non corrélé aux deux autres, le variogramme de l'innovation s'écrit

$$\gamma_D(h,\tau) = \sigma_\varepsilon^2 + (1 - \rho\omega)^2 + \omega^2 (1 - \rho^2) \gamma_{h,\tau} \quad (8)$$

Ce modèle

- est très souple, car suivant les valeurs de ω et de ρ , le variogramme de la simulation, celui de la mesure et celui de l'innovation s'ordonnent différemment, à l'effet de pépite près (cf. Chilès et Séguret, 2008).
- n'est compatible avec un « modèle à résidu », c'est-à-dire avec un modèle du type $Y=\omega S+Q$, où le résidu Q est non spatialement corrélé à S , que dans le cas où $\rho=1$ ou $\omega=0$ (cas inintéressant). Expérimentalement le modèle à résidu apparaît effectivement inadapté.

En pratique, les paramètres du modèle se calculent simplement :

- l'ajustement du variogramme de la mesure Z donne la composante pépitique σ_ε^2 ainsi que le variogramme $\gamma_{h,\tau}$ de Y ;
- l'ajustement du variogramme de la simulation fournit le facteur positif ω ;
- l'ajustement du variogramme croisé entre Z et S fournit le coefficient de corrélation ρ ;
- et l'ajustement du variogramme de l'innovation permet de contrôler la validité de l'ajustement.

ANNEXE 4

Analyse statistique des coefficients du modèle linéaire de corégionalisation

Cartes

Statistiques élémentaires

Histogrammes

ACP

Variogrammes

Etude de la grande portée (composante exponentielle) et des coefficients ω et ρ (composantes exponentielle et périodique) du modèle linéaire de corégionalisation ajusté par station

Signification des termes utilisés dans les figures :

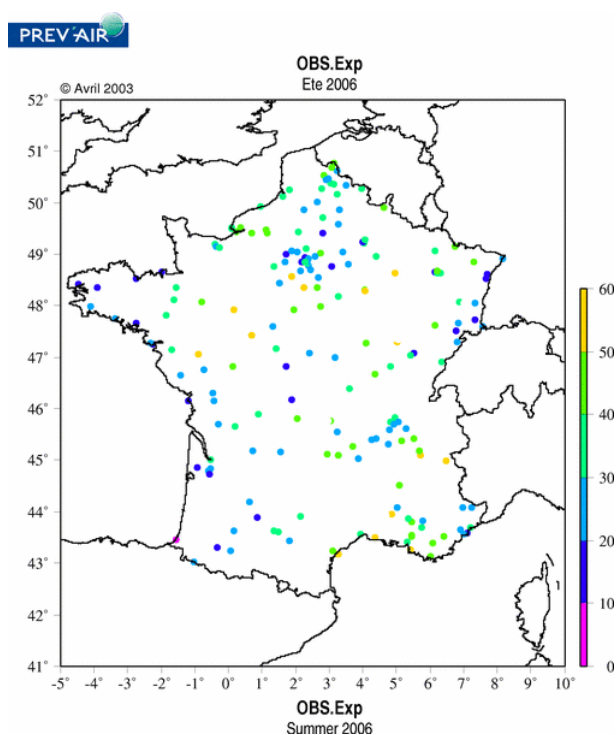
OBS.Exp : portée du modèle exponentielle, exprimée en heures ($h/3$)

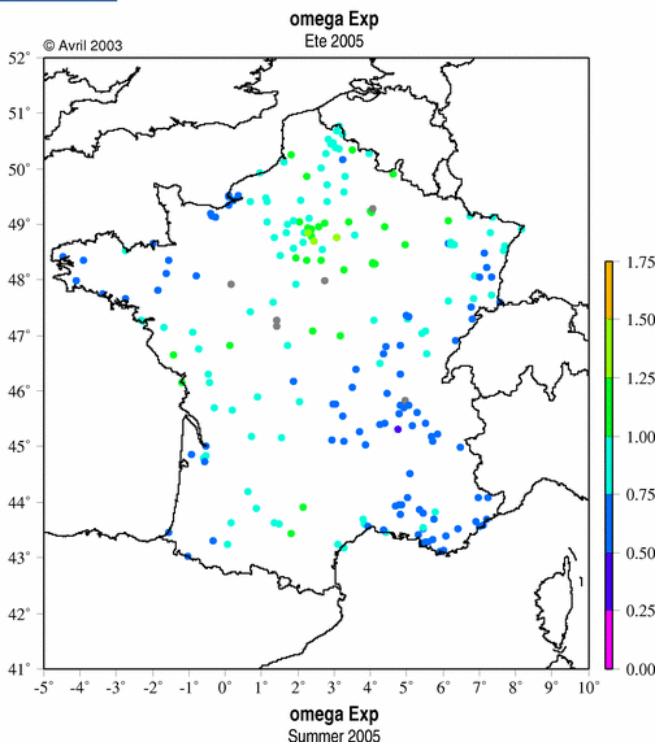
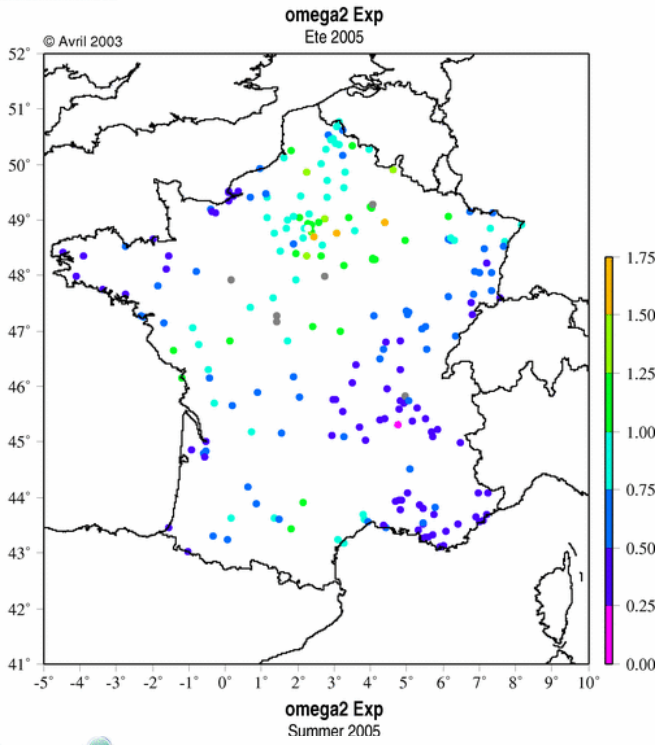
Omega2 : ω^2 (_exp ou _per : composante exponentielle ou périodique)

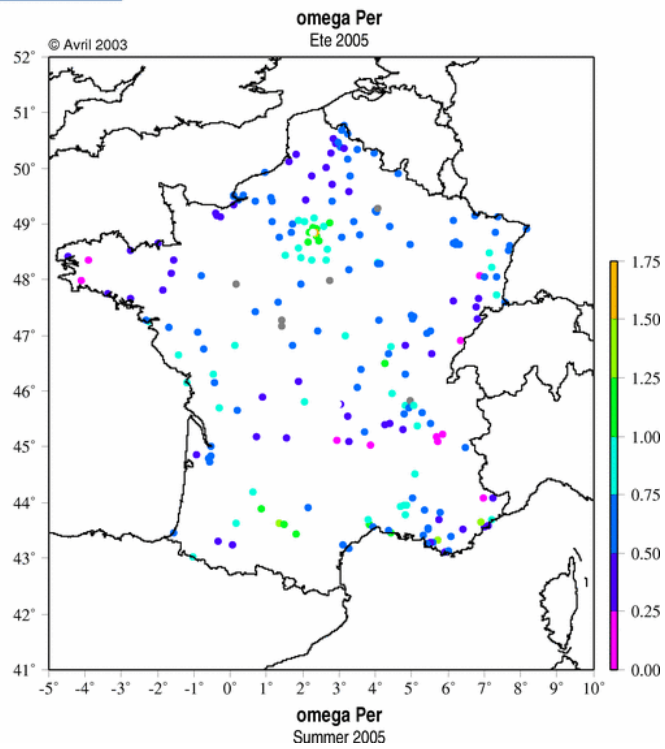
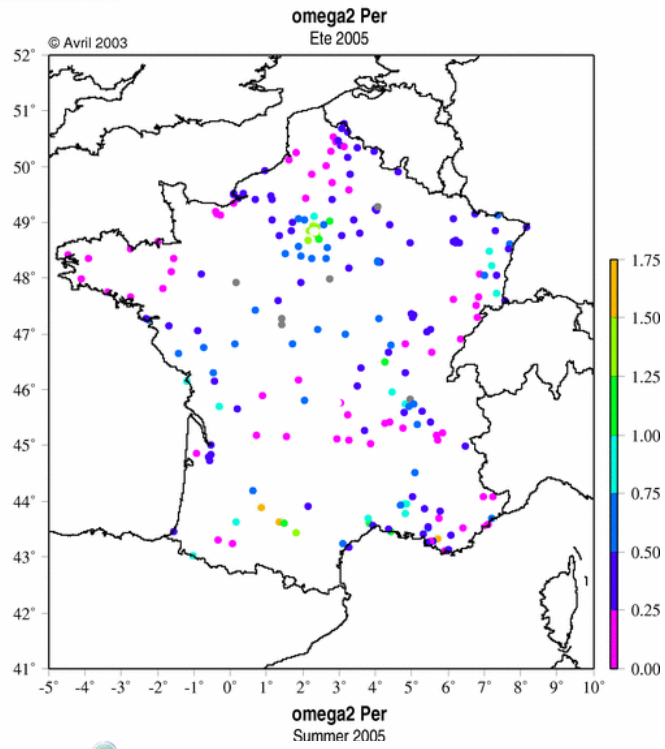
Omega : ω (_exp ou _per : composante exponentielle ou périodique)

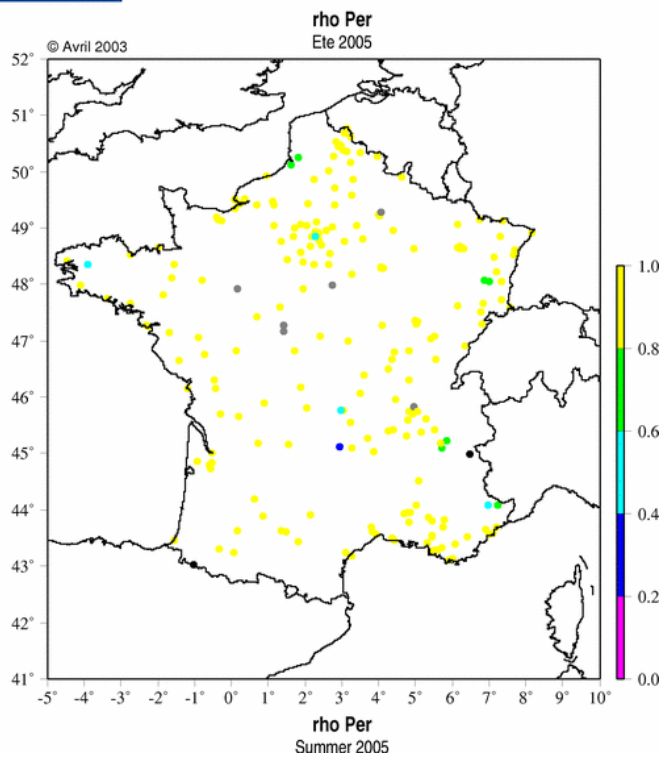
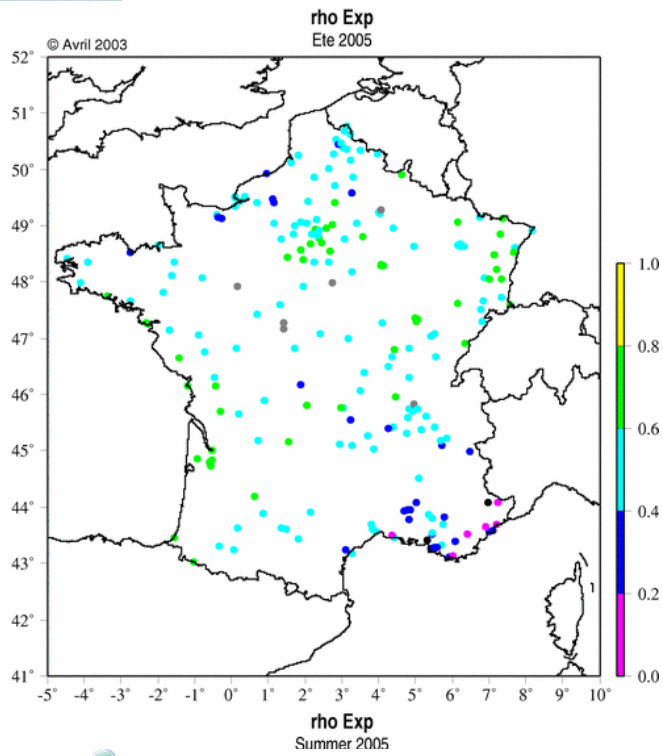
rho : ρ (_exp ou _per : composante exponentielle ou périodique)

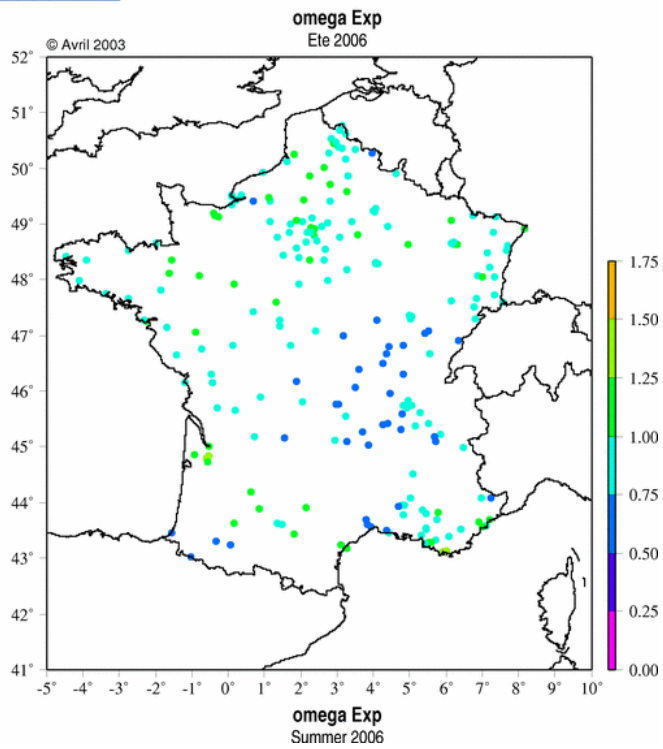
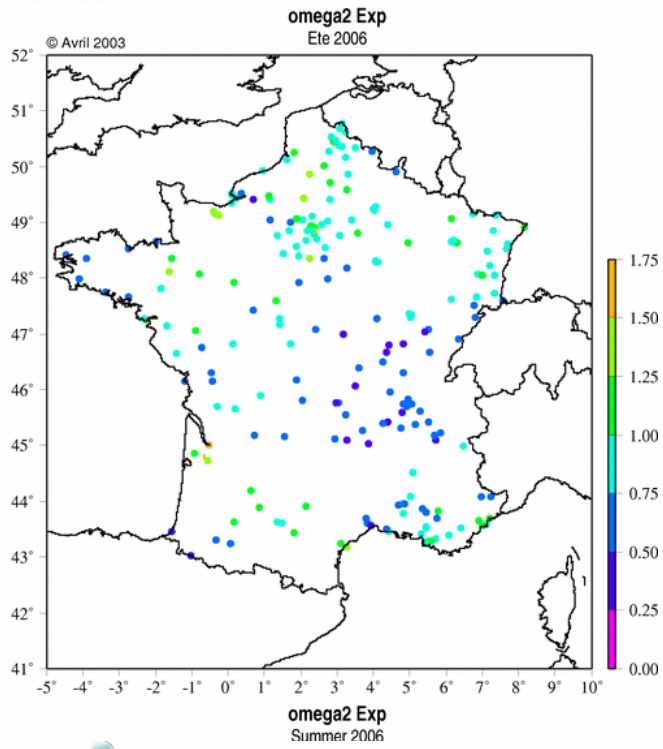
Cartographies

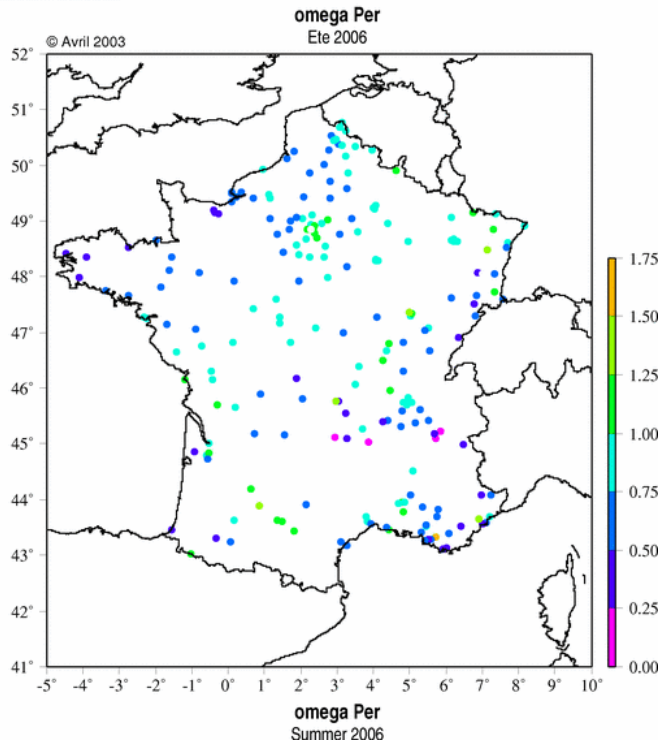
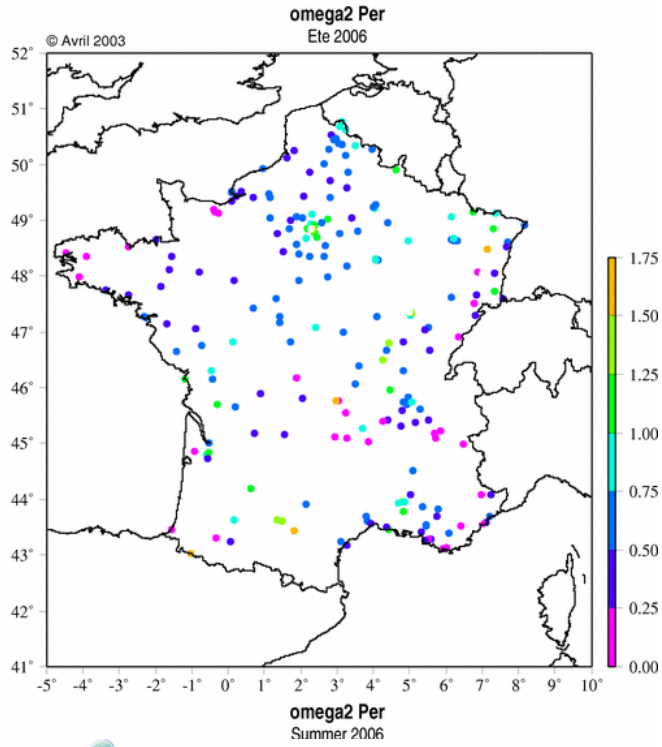


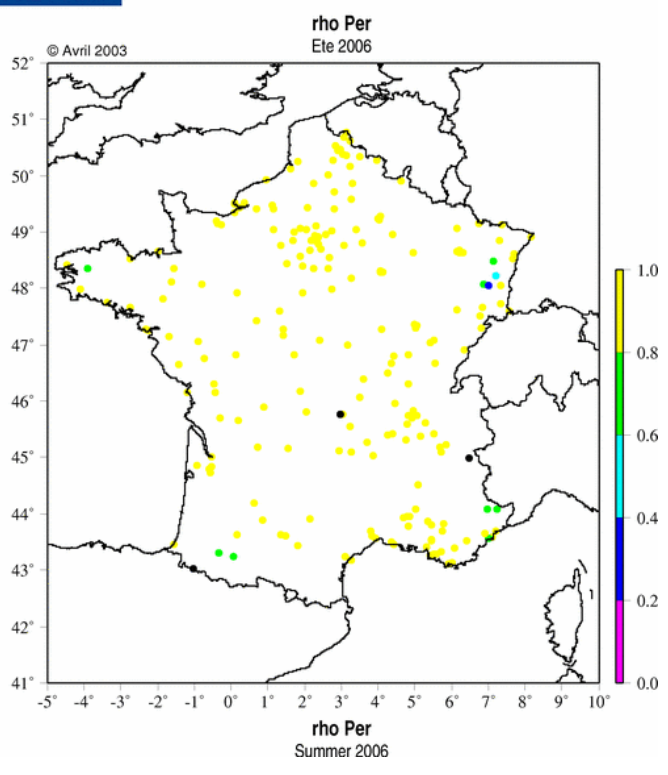
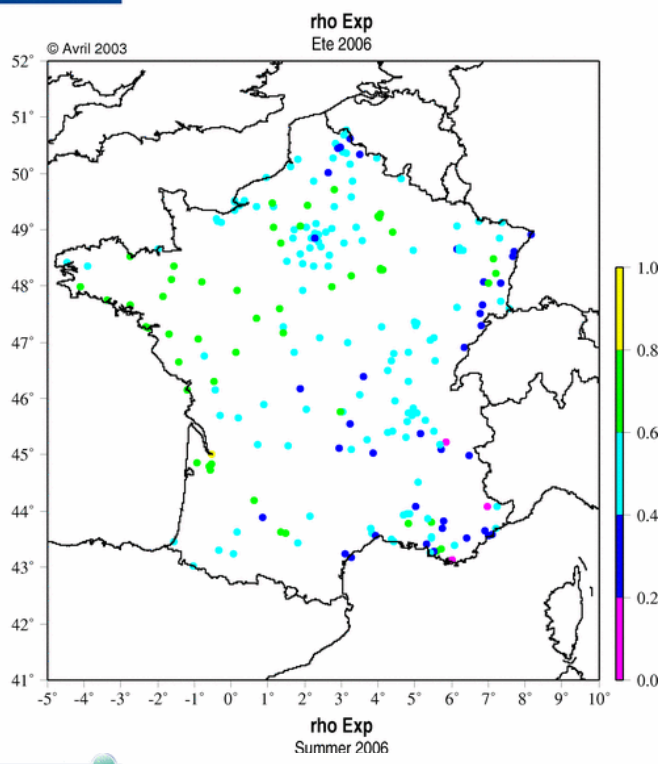












Statistiques

Eté 2005

	Minimum	P5	P50	P95	Maximum	Mean	Standard deviation
Portée (h/3 en heures)	2.8	8.4	19.2	33.6	54.8	20.2	8.2
ω^2 exp	0.24	0.36	0.64	1.24	1.97	0.72	0.30
ω exp	0.49	0.60	0.80	1.12	1.40	0.83	0.17
ρ exp	- 0.04	0.23	0.54	0.70	0.78	0.52	0.15
ω^2 per	0.01	0.08	0.39	1.43	7.31	0.54	0.68
ω per	0.09	0.28	0.63	1.20	2.70	0.66	0.31
ρ per	-1.00	0.76	0.96	1.00	1.00	0.91	0.20

Eté 2006

	Minimum	P5	P50	P95	Maximum	Mean	Standard deviation
Portée	5.0	15.5	34.1	80.9	157.2	38.7	21.3
ω^2 exp	0.26	0.52	0.94	1.37	2.08	0.91	0.29
ω exp	0.51	0.72	0.97	1.17	1.44	0.94	0.15
ρ exp	0.01	0.32	0.54	0.71	0.79	0.53	0.12
ω^2 per	0.01	0.12	0.57	1.54	8.03	0.70	0.72
ω per	0.10	0.35	0.75	1.24	2.83	0.78	0.31
ρ per	-1.00	0.79	0.97	1.00	1.00	0.93	0.21

Portées : différences assez sensibles entre l'été 2005 et l'été 2006 ; portées plus grandes en 2006.

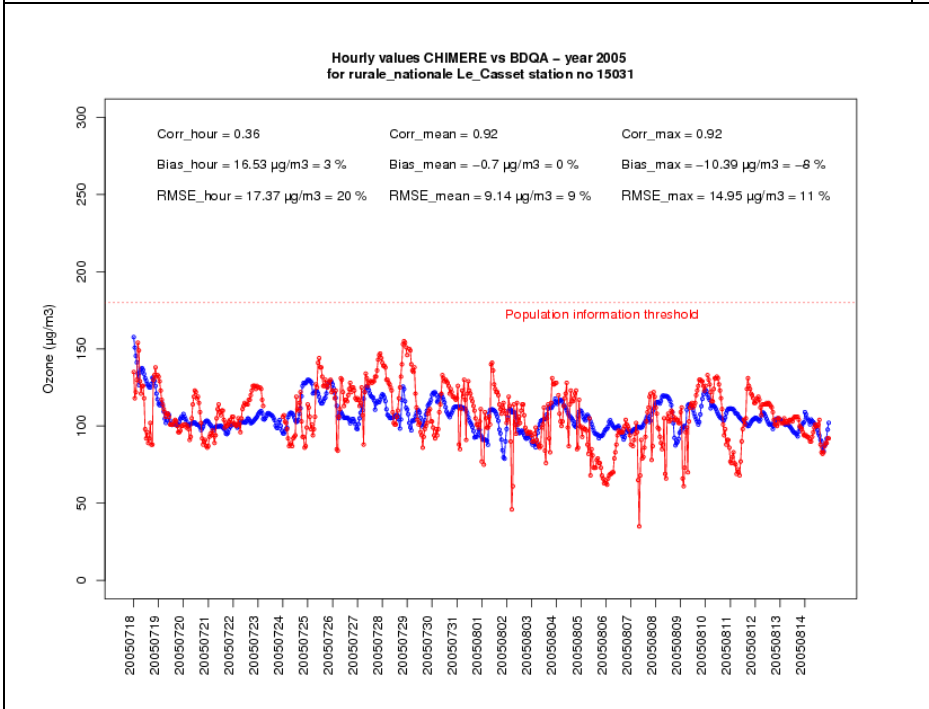
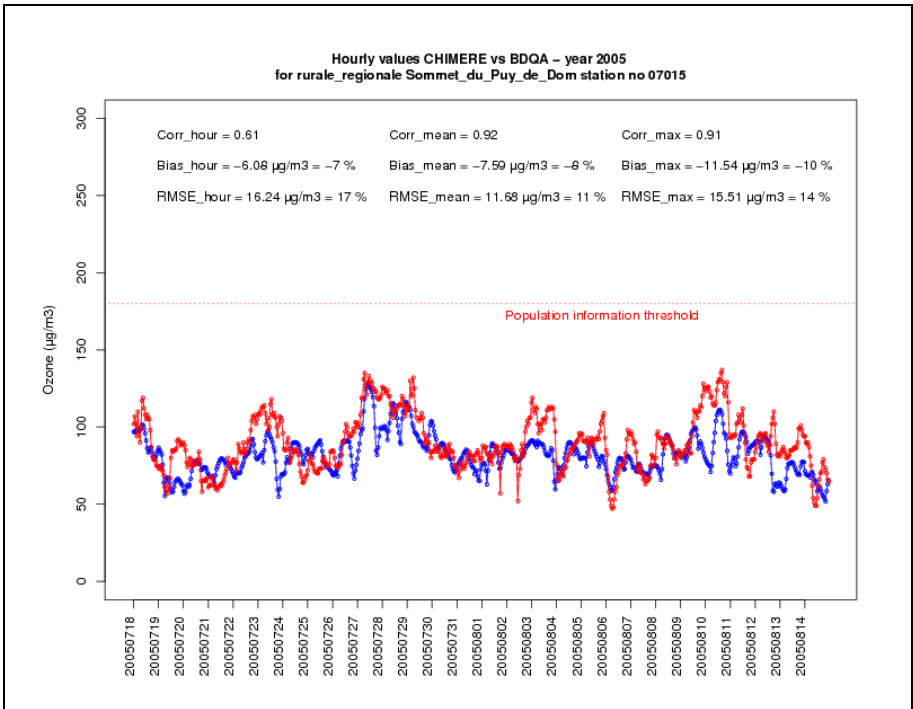
Pour les coefficients, statistiques assez similaires.

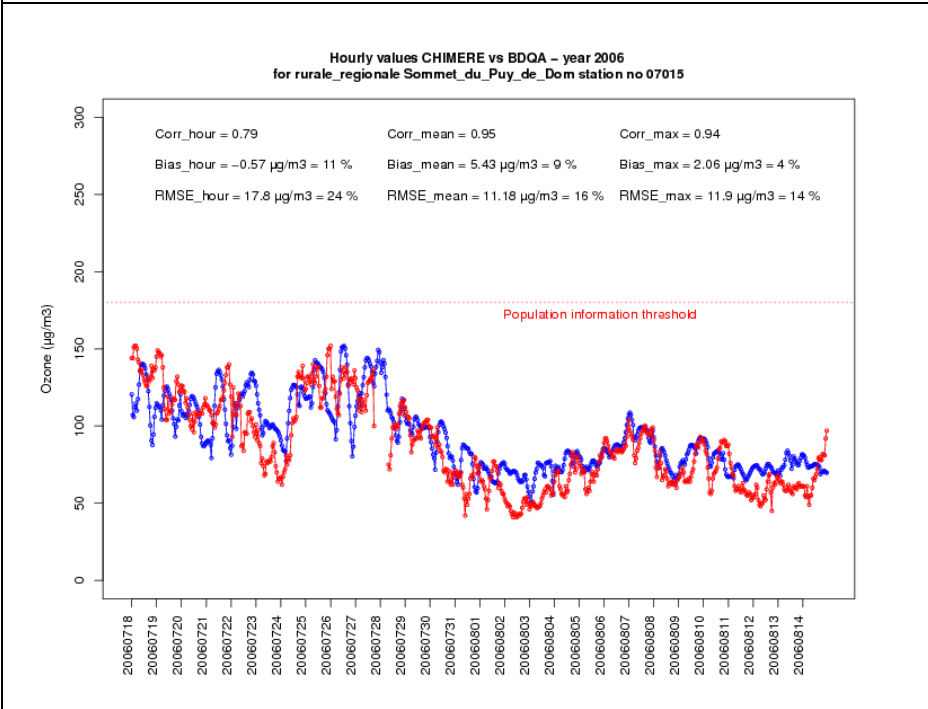
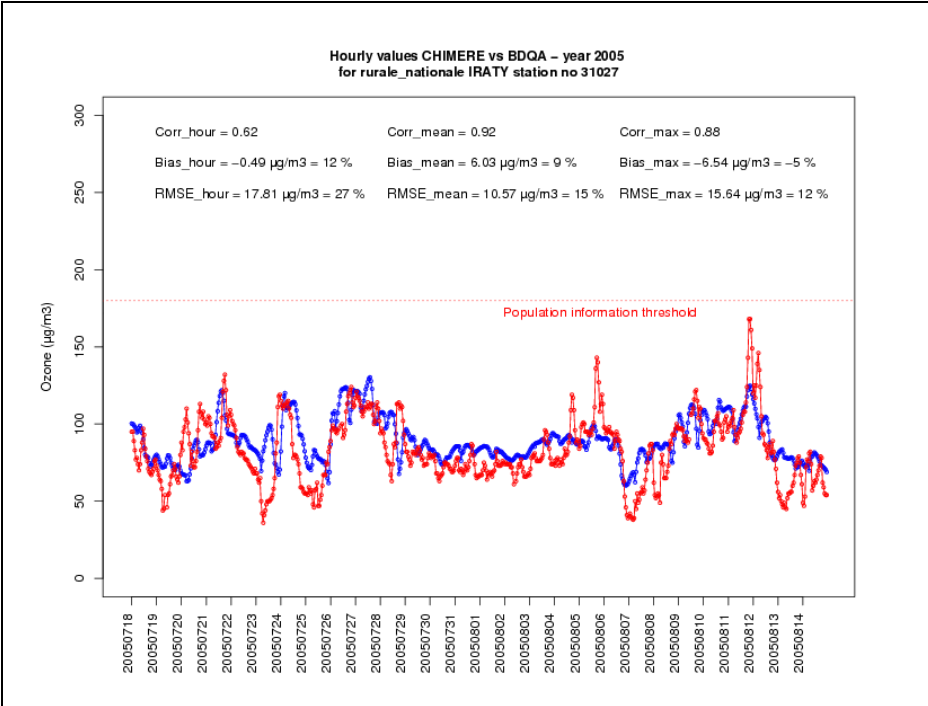
Rho_périodique négatif pour certaines stations de montagne, voir ci-après

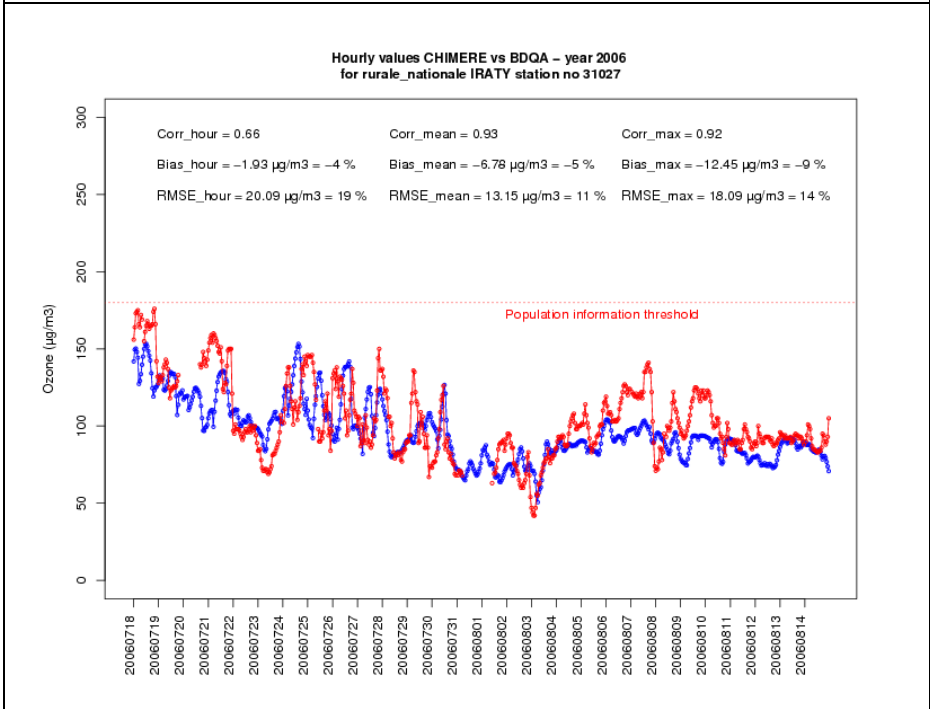
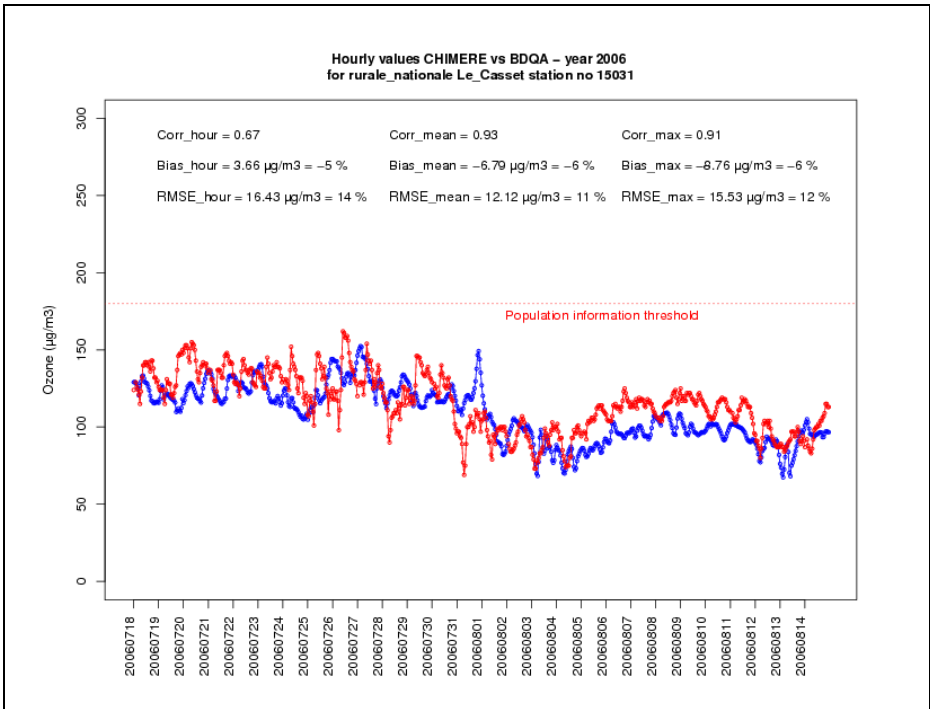
2005 (cf. exemples de variogrammes) : 15031 : r=-1 et 31027 : r=-0.87

2006 (cf. exemples de variogrammes) : 31027 : r= -1 ; 07015 : r= -0,82 et 15031 : r=-0,38

De haut en bas : 07015, 15031, 31027 en 2005 puis 07015, 15031, 31027 en 2006



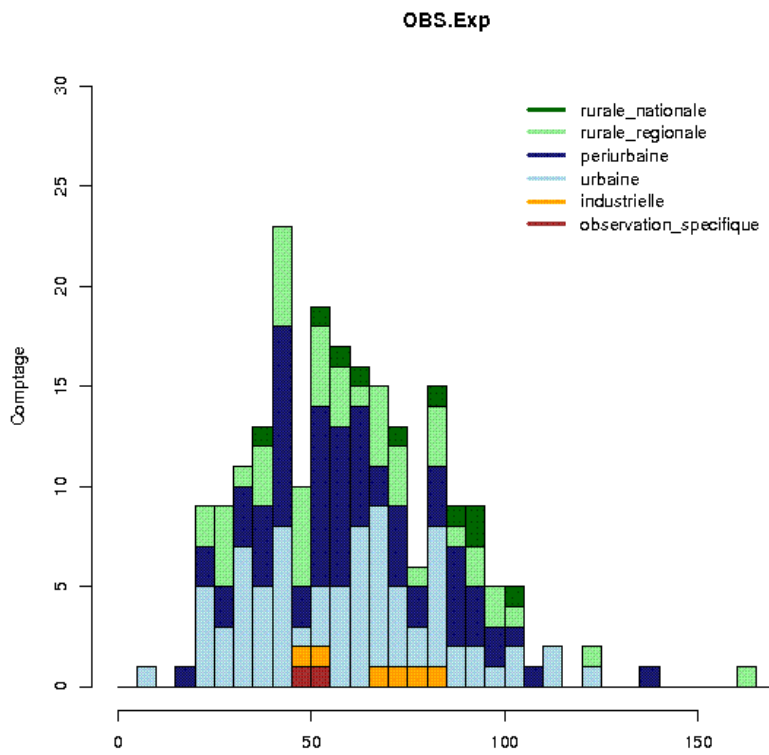


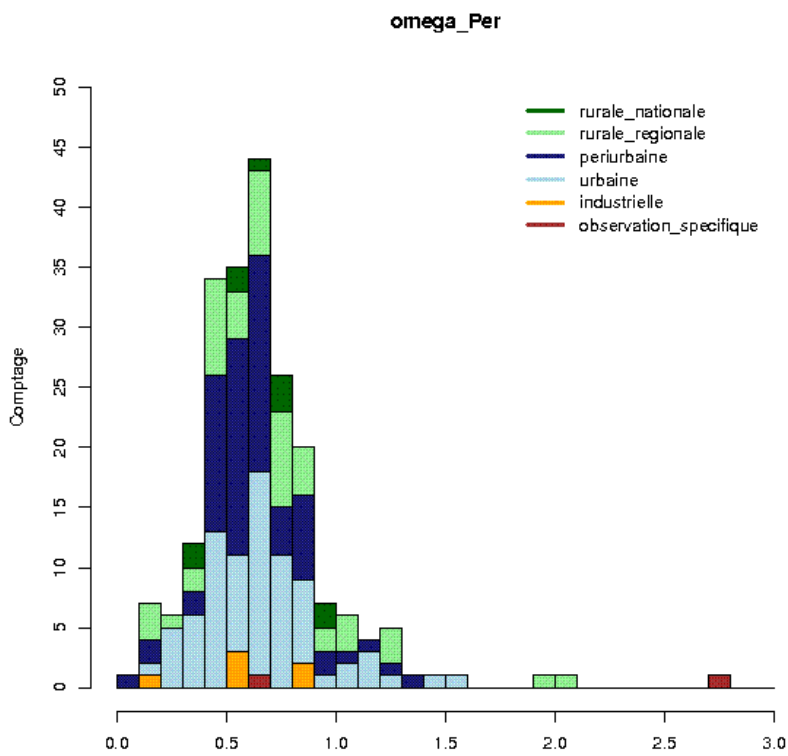
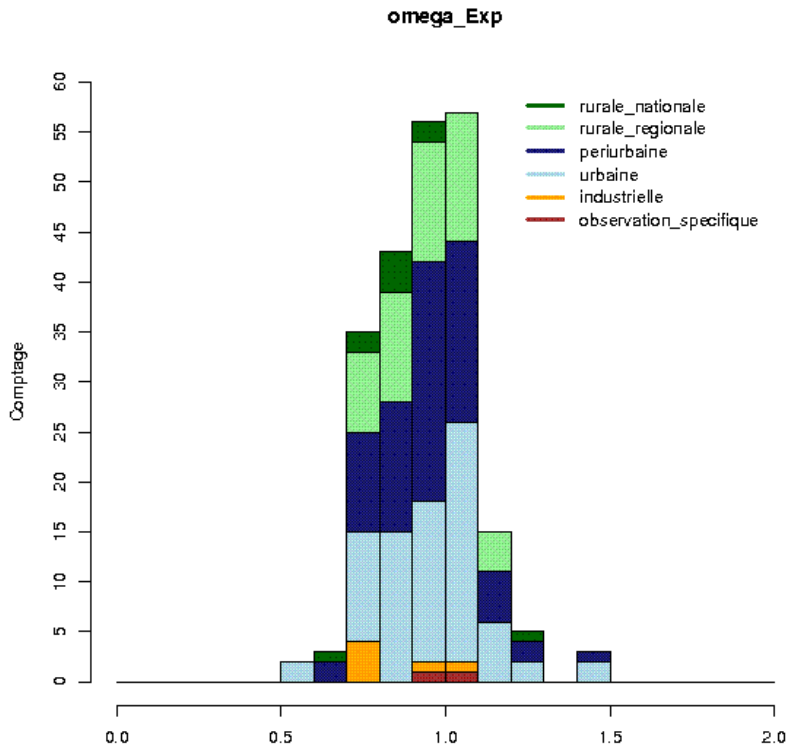


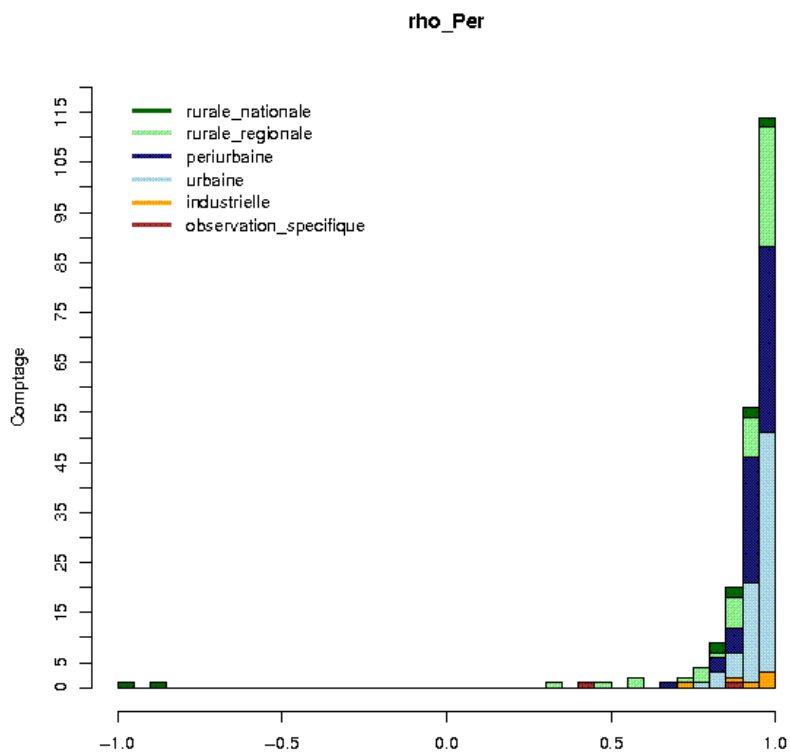
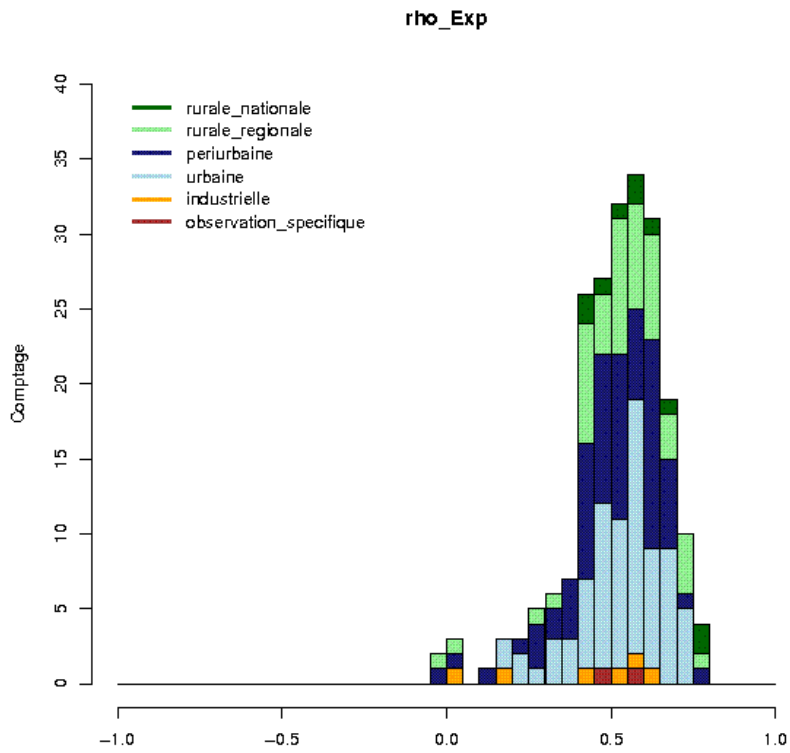
Histogrammes

En fonction de l'occupation des sols

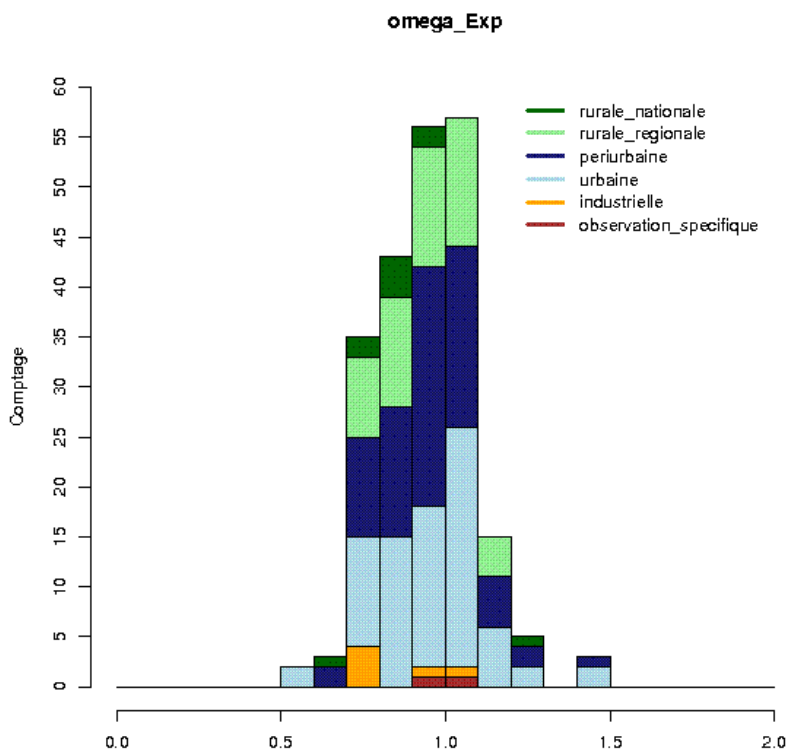
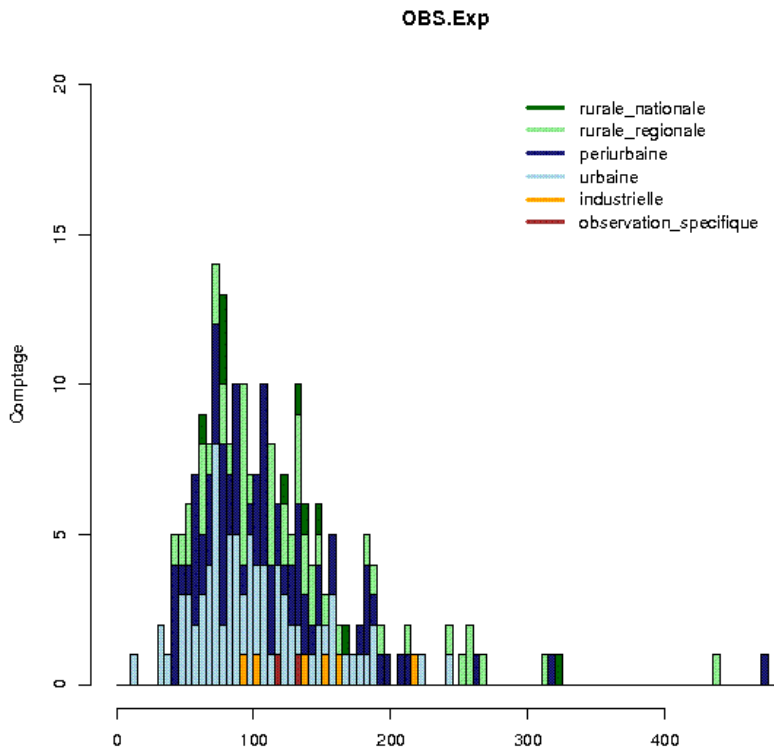
Eté 2005 :

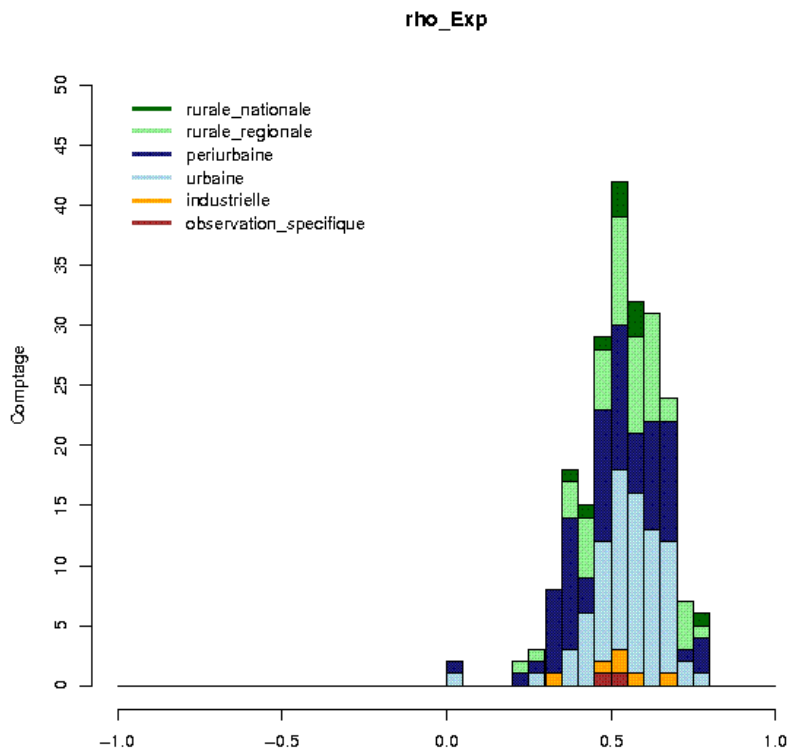
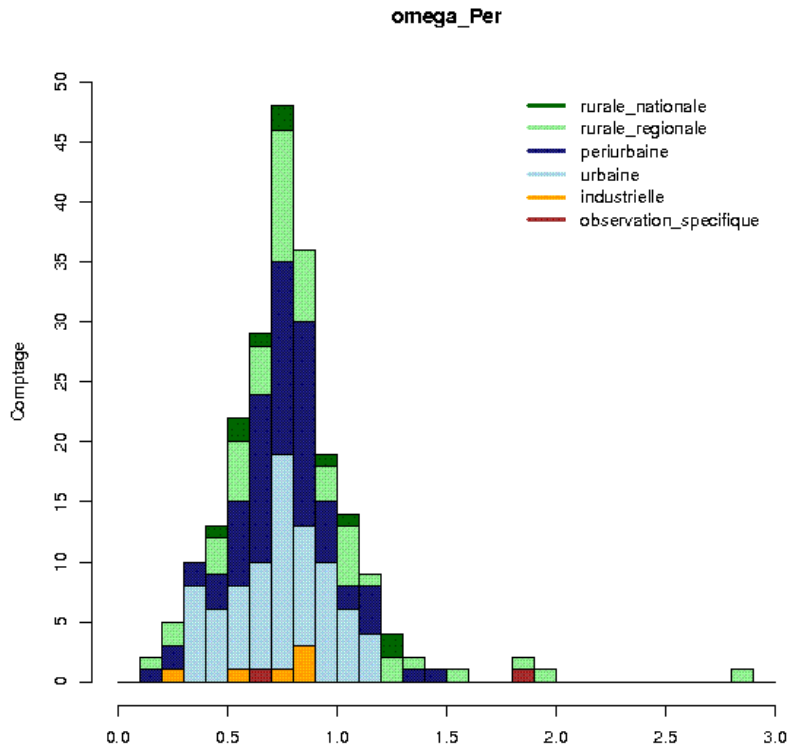




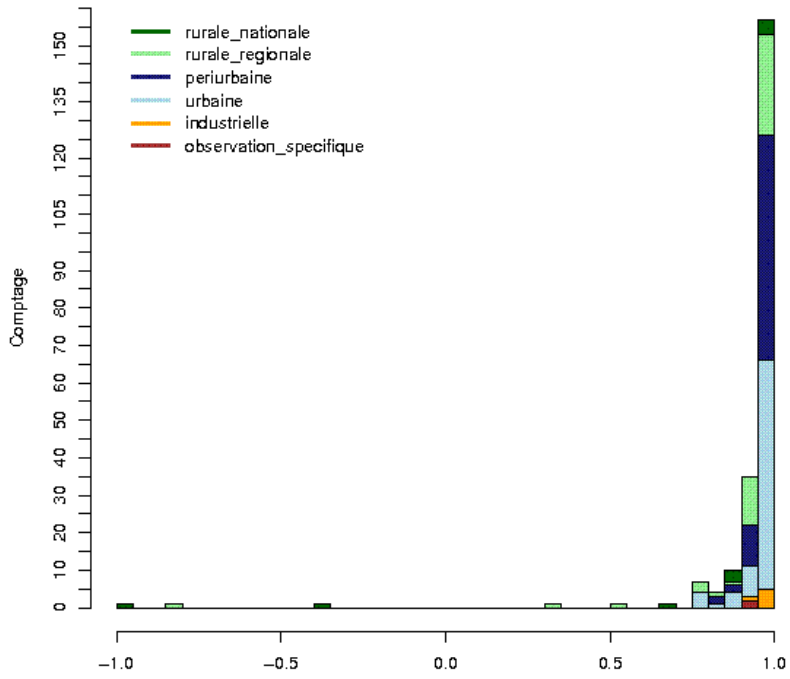


Eté 2006 :



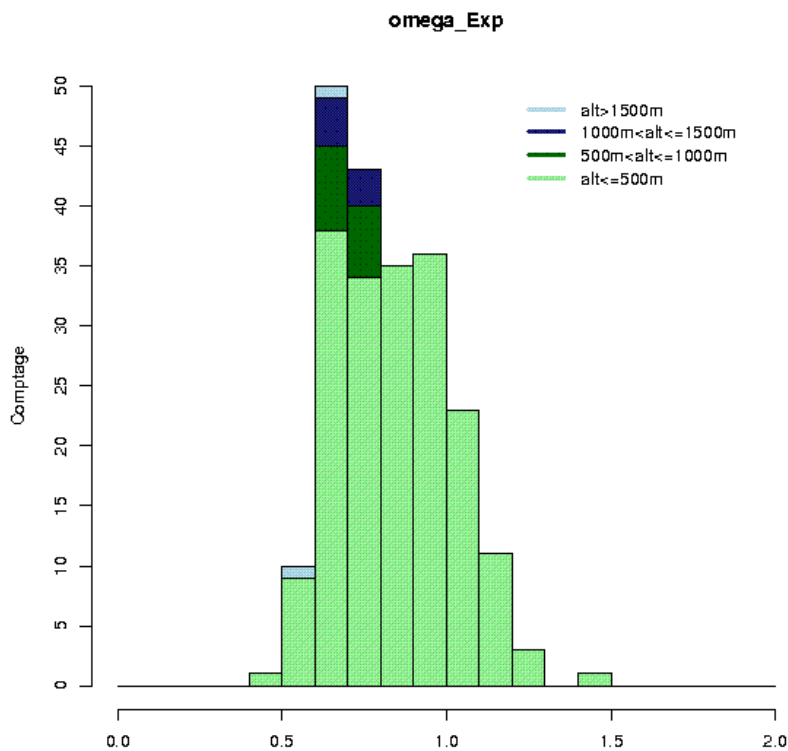
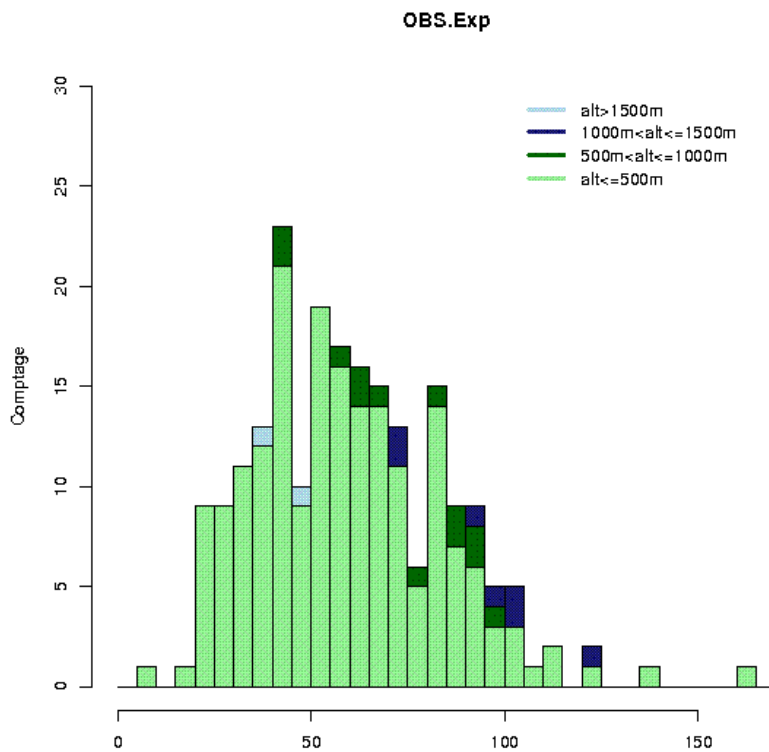


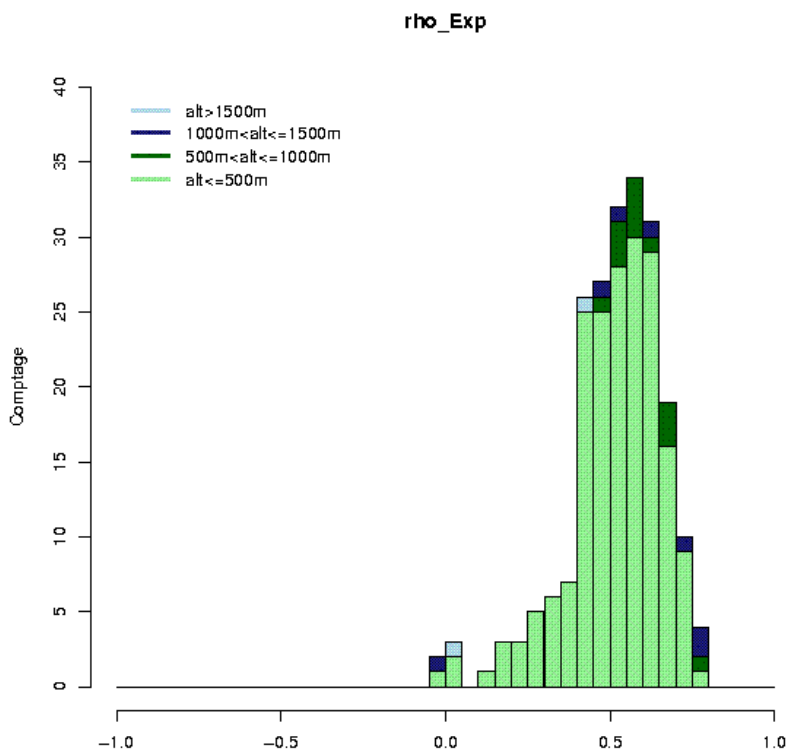
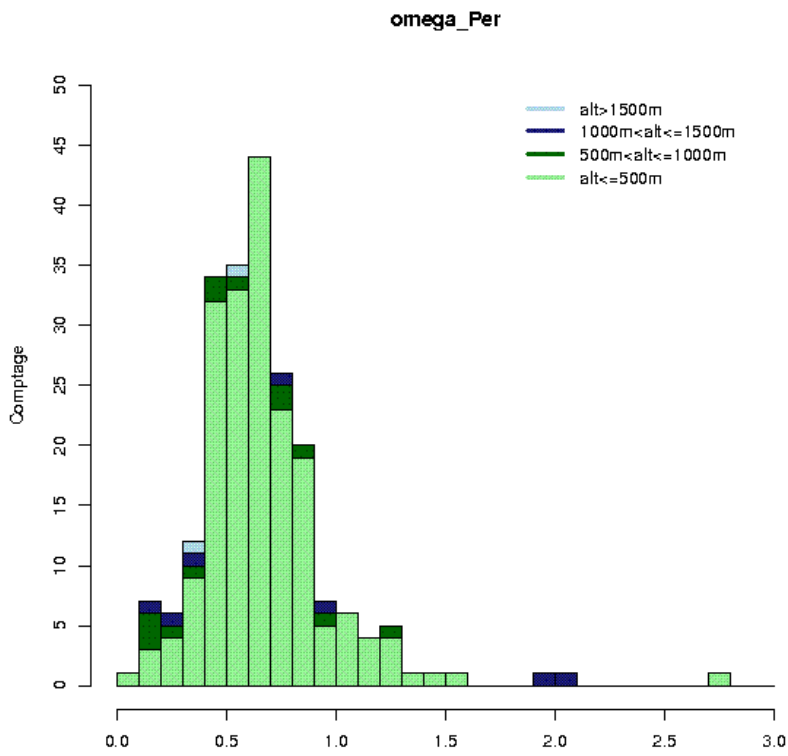
rho_Per



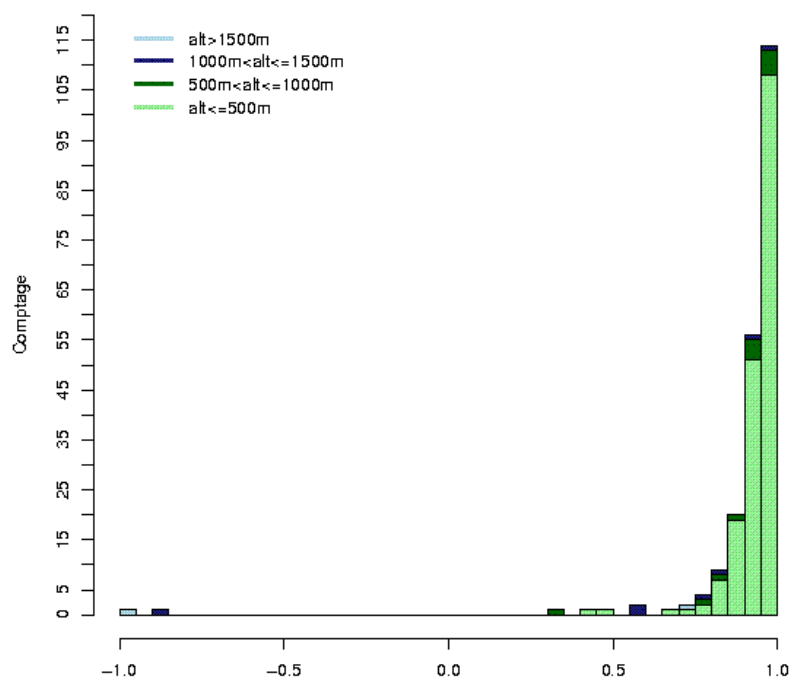
En fonction de l'altitude

Été 2005 :



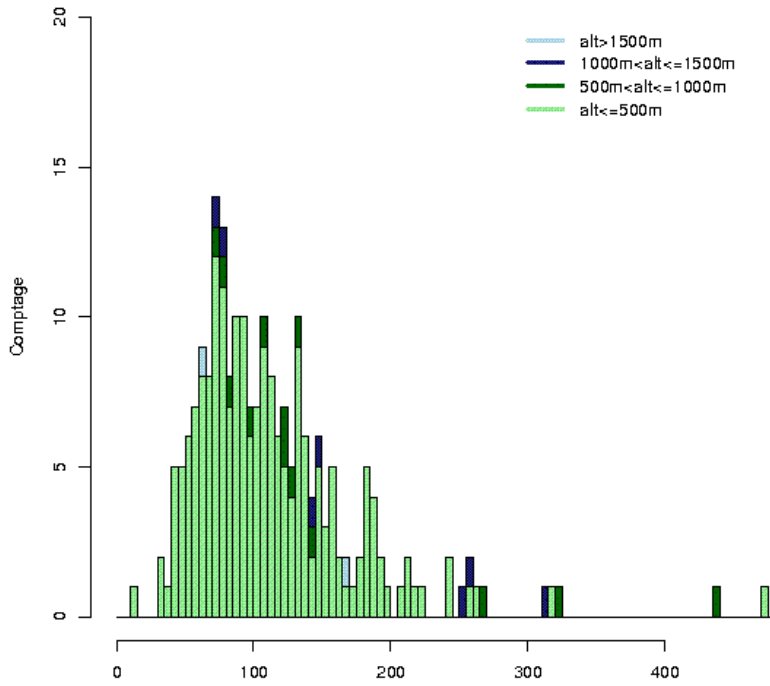


rho_Per

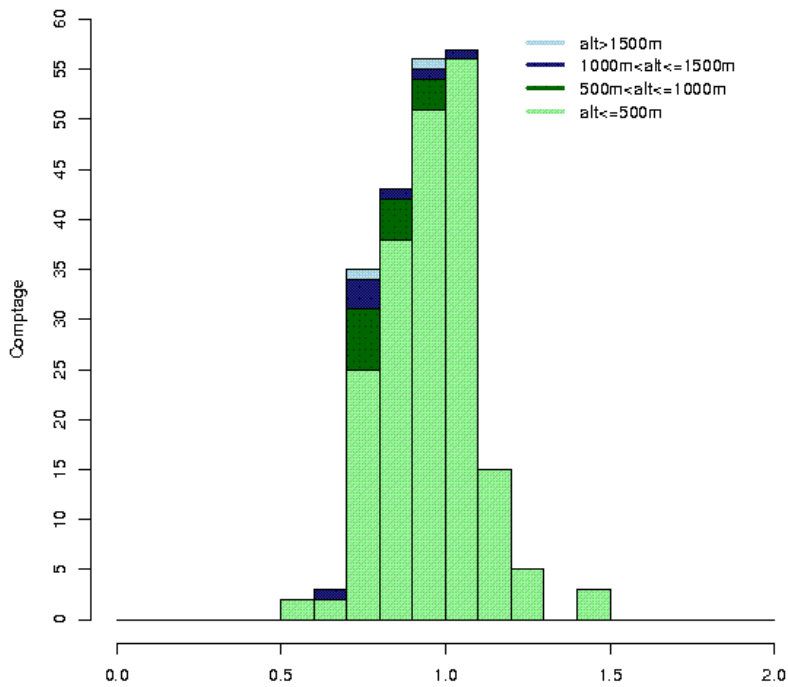


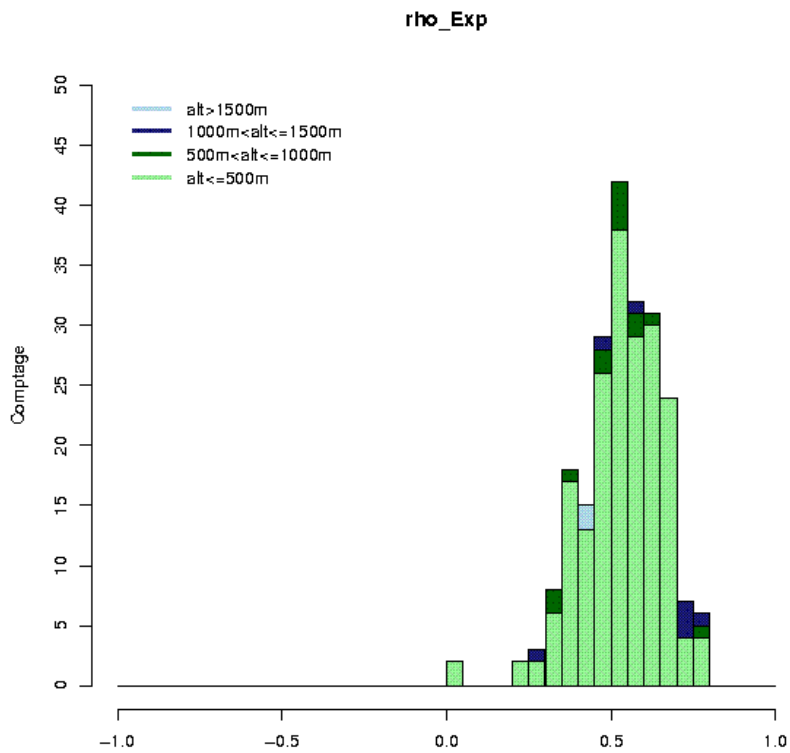
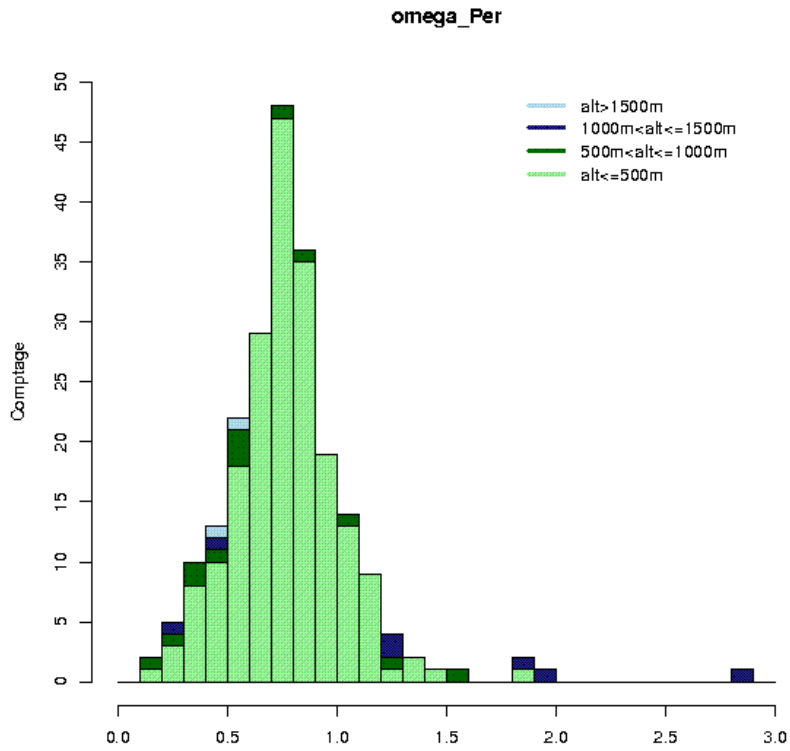
Été 2006 :

OBS.Exp

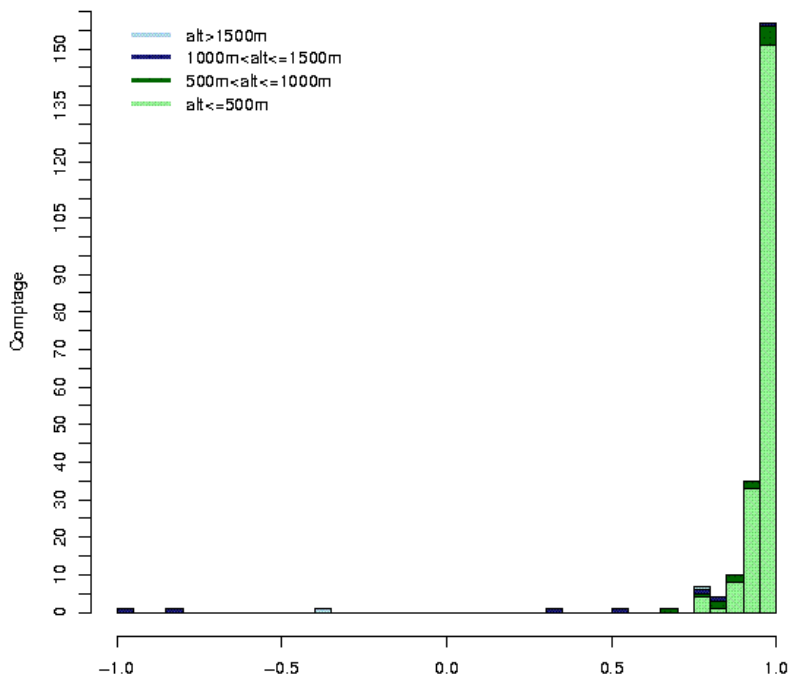


omega_Exp

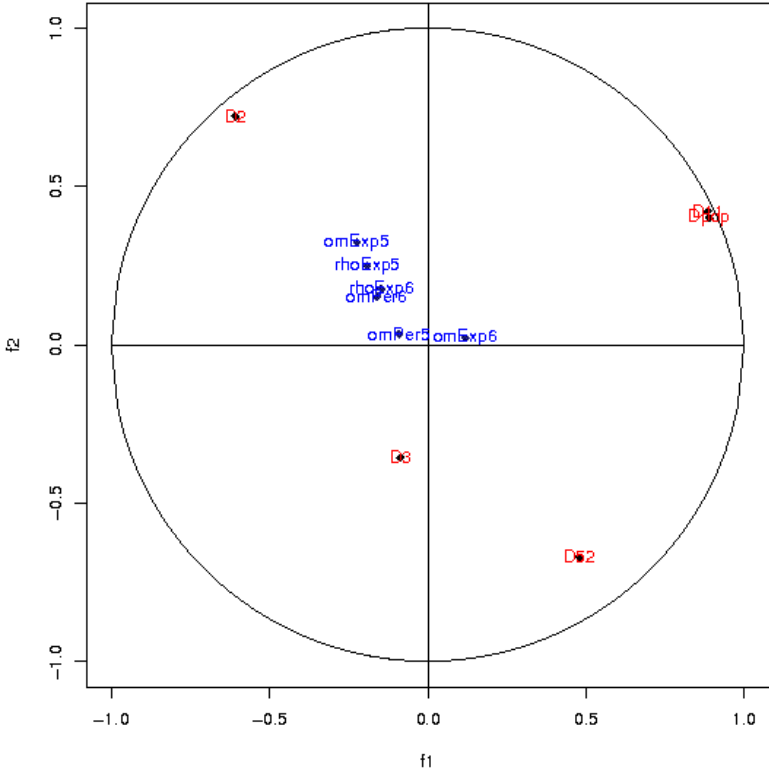
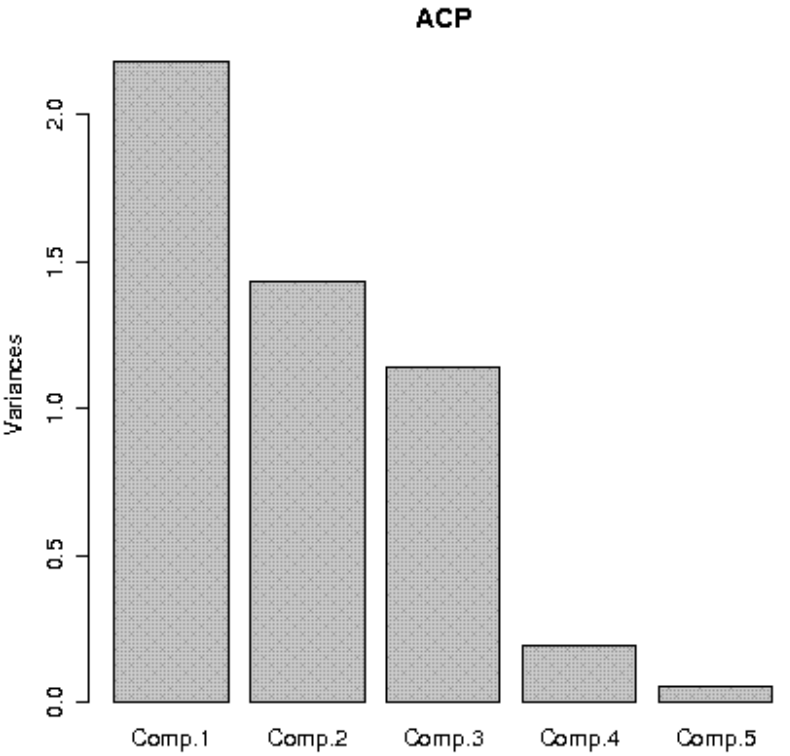


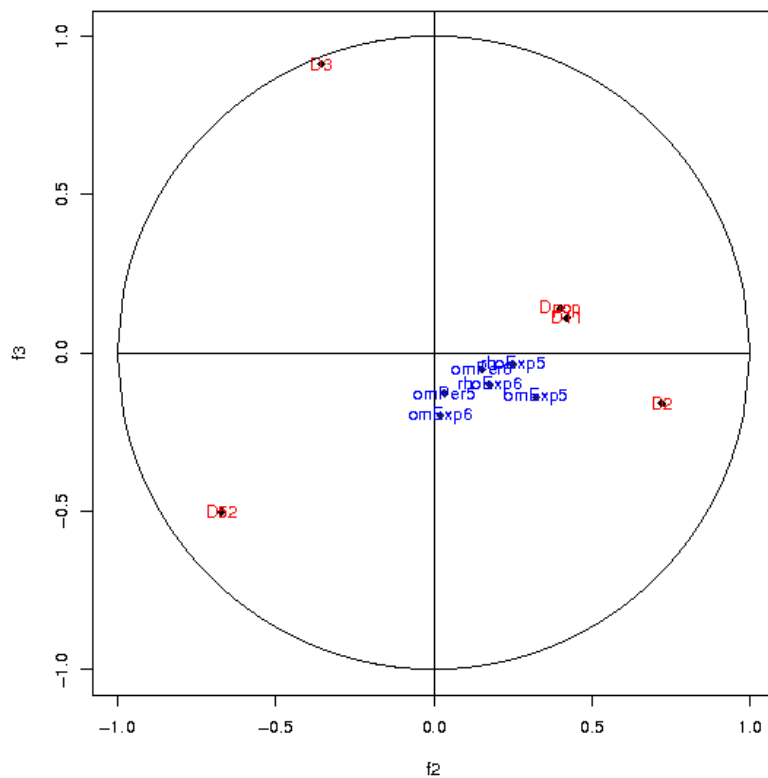
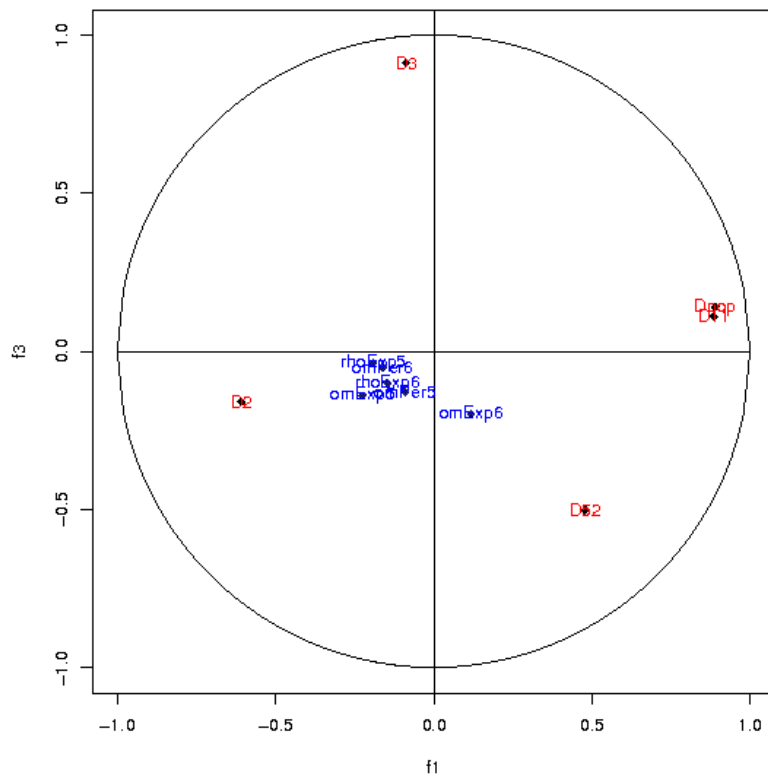


rho_Per



ACP sur les transformées logarithmiques ($\log(X+1)$) des densités d'occupation du sol et de population interpolées aux stations. Projection a posteriori des coefficients sur les plans factoriels.

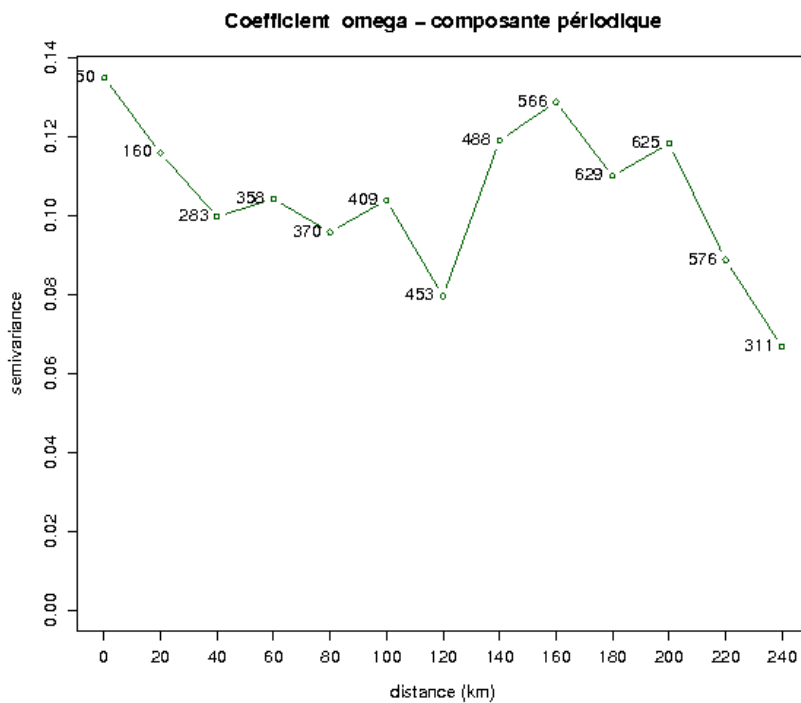
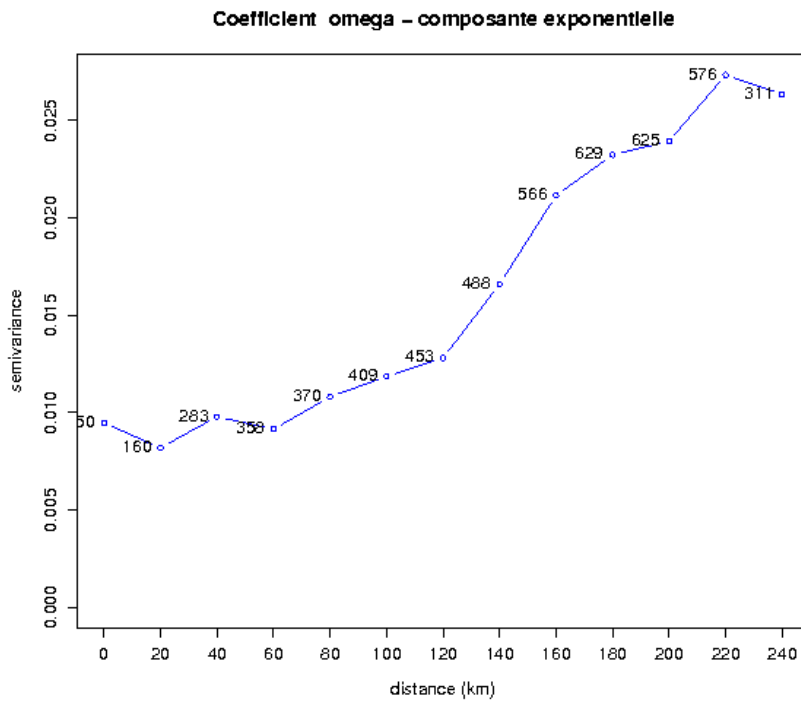


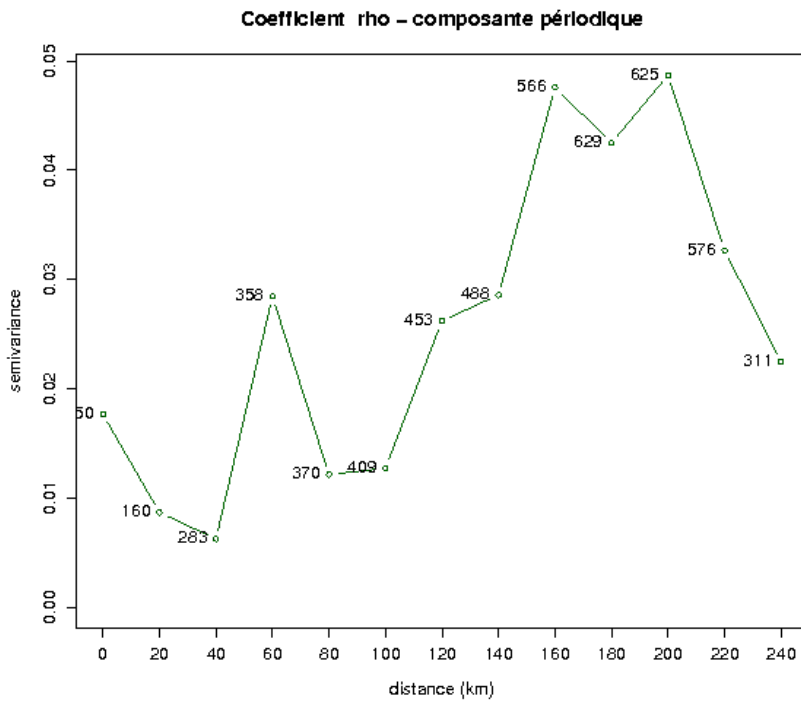
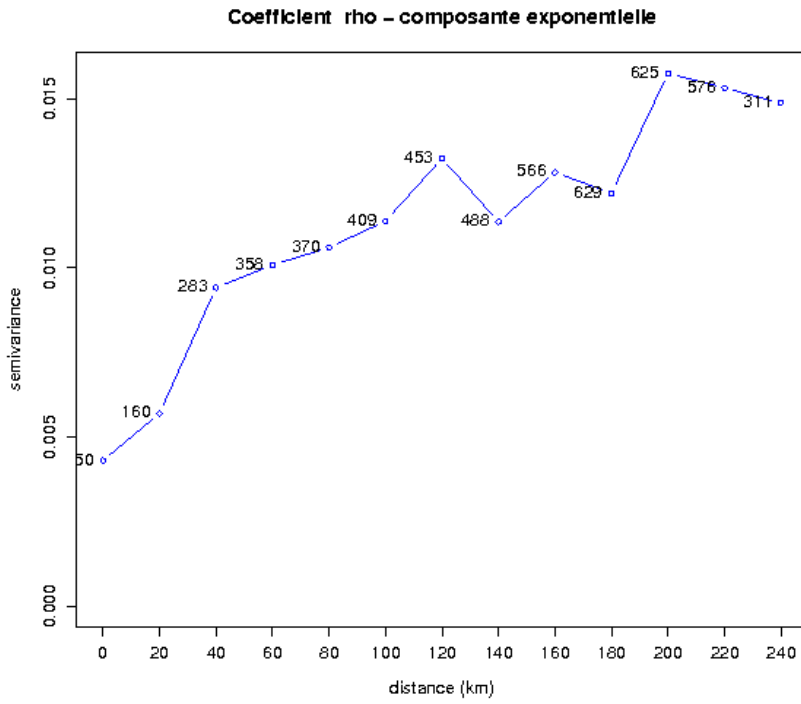


D11 : densité de bâti ; D2 : densité de surface agricole ; D3 : densité de surface forestière ; D52 : densité de surface marine

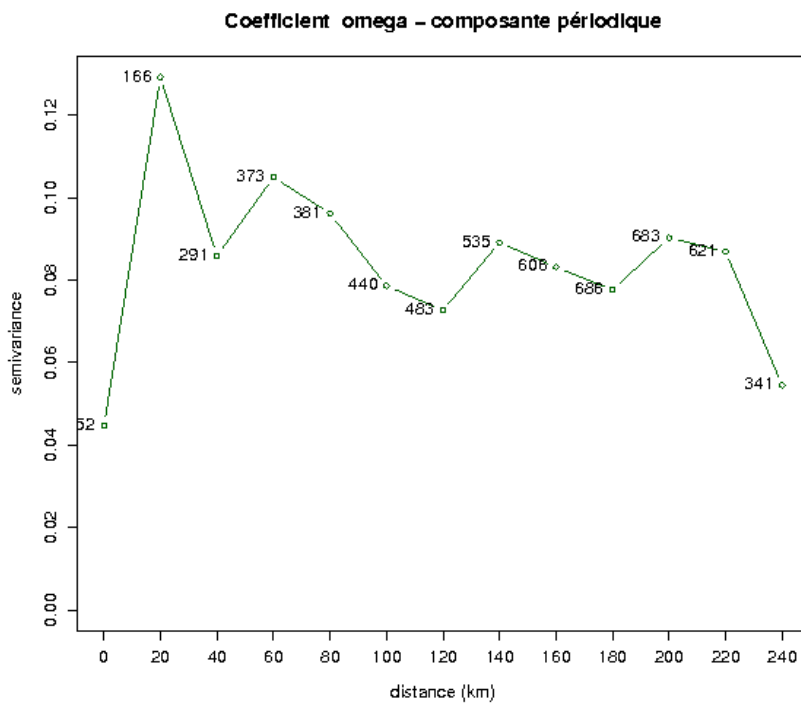
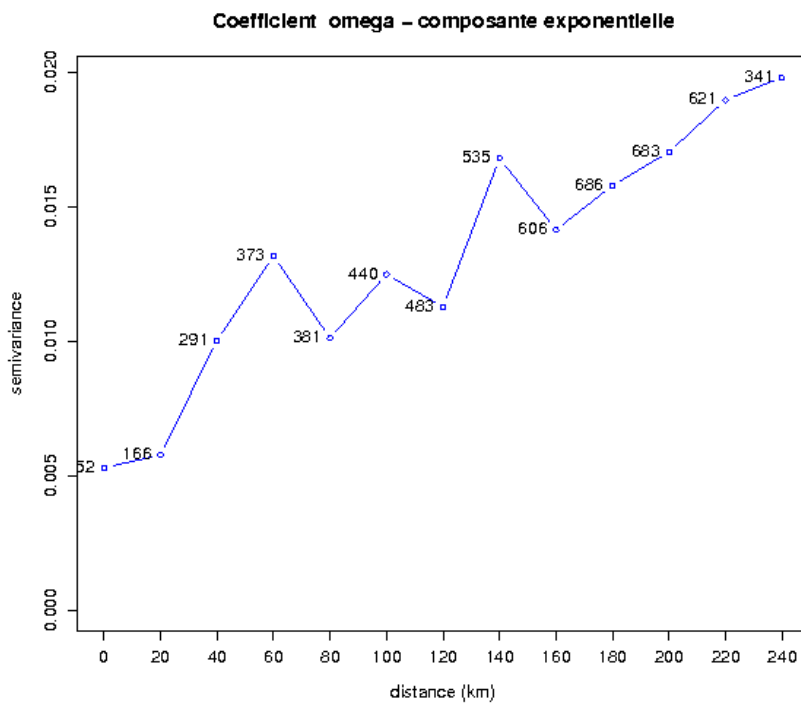
Variogrammes spatiaux

Eté 2005 :

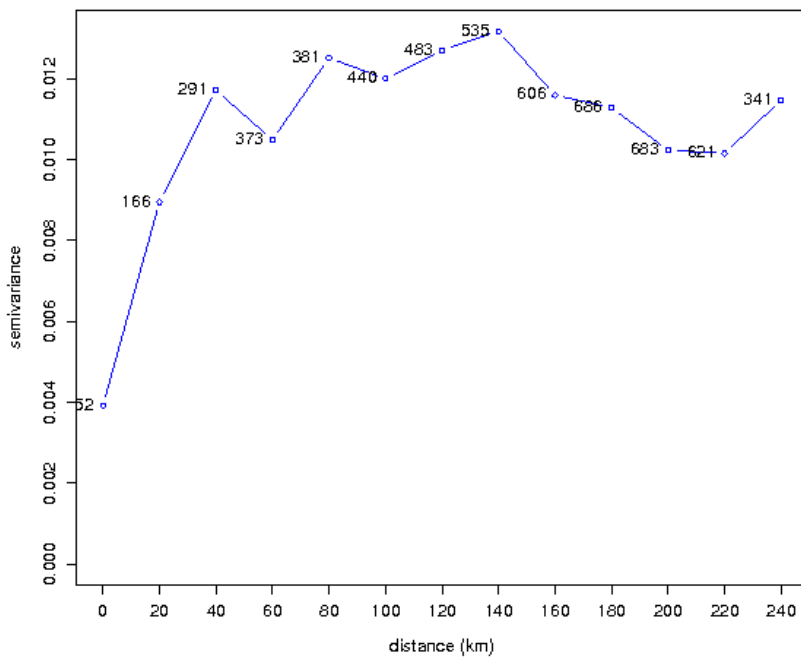




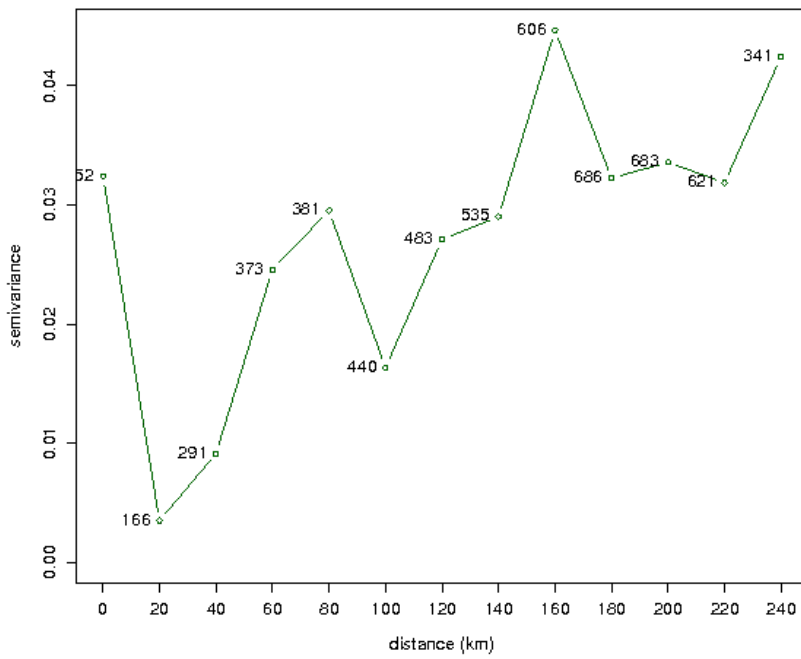
Eté 2006 :



Coefficient rho - composante exponentielle



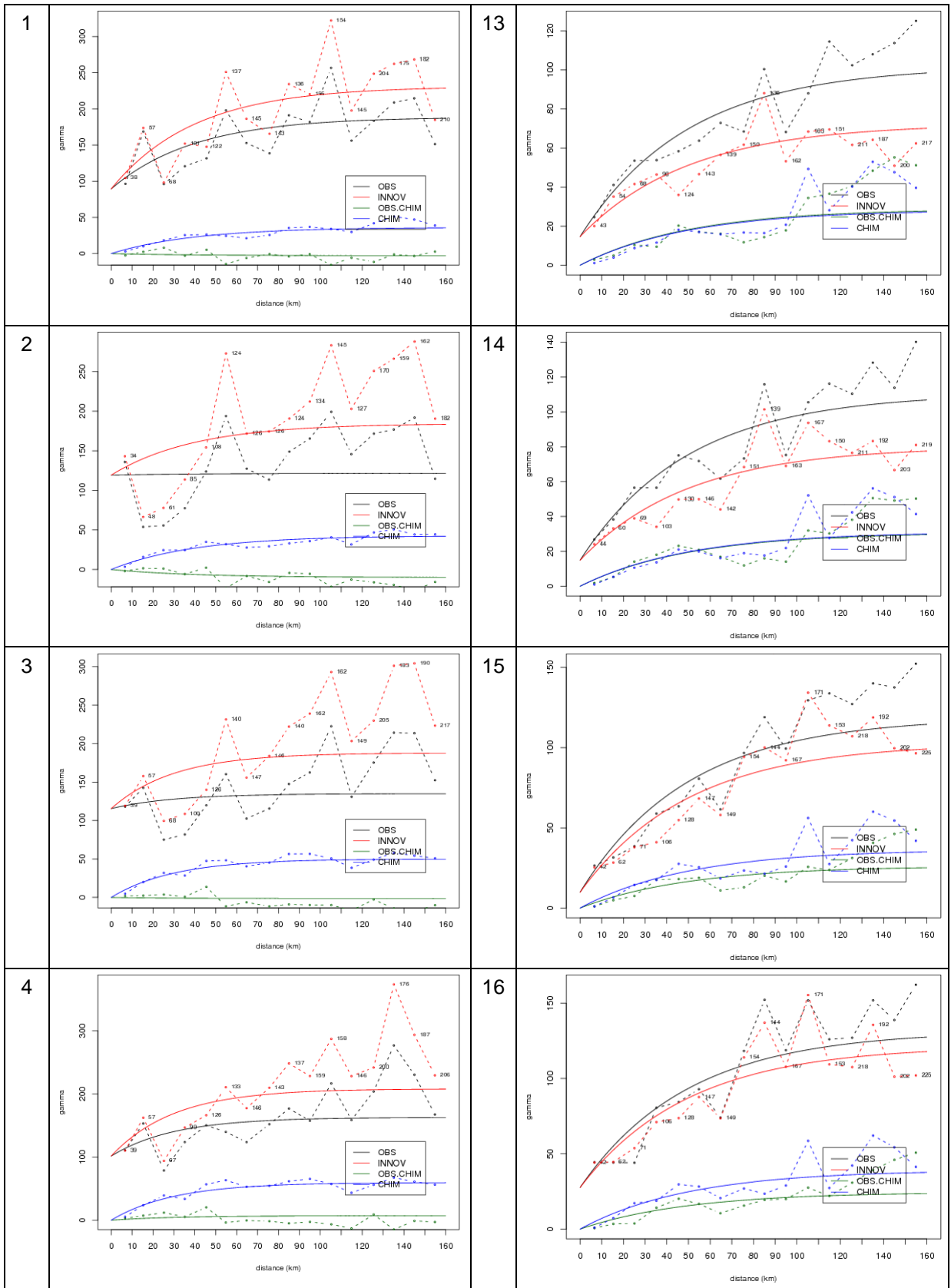
Coefficient rho - composante périodique

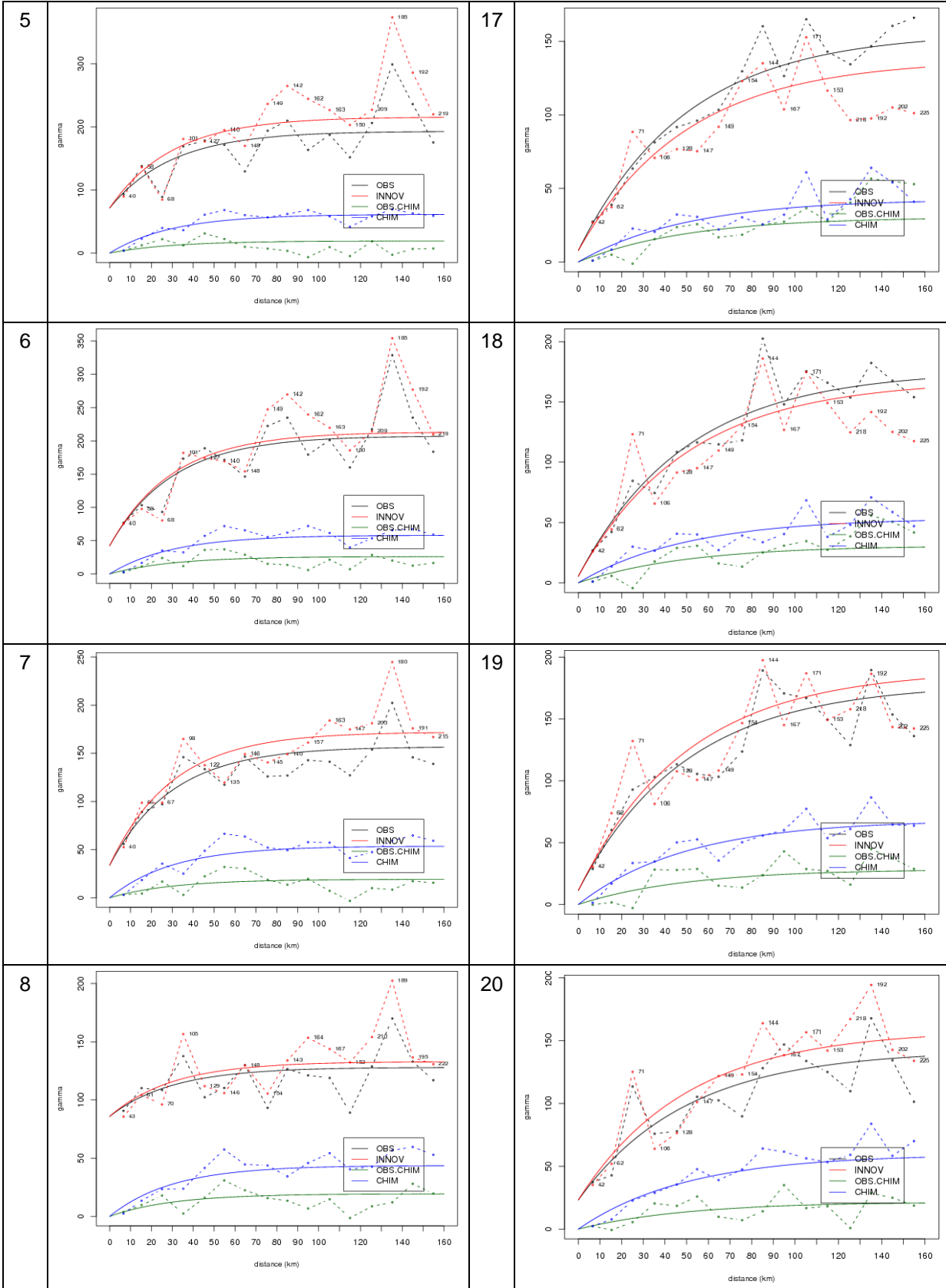


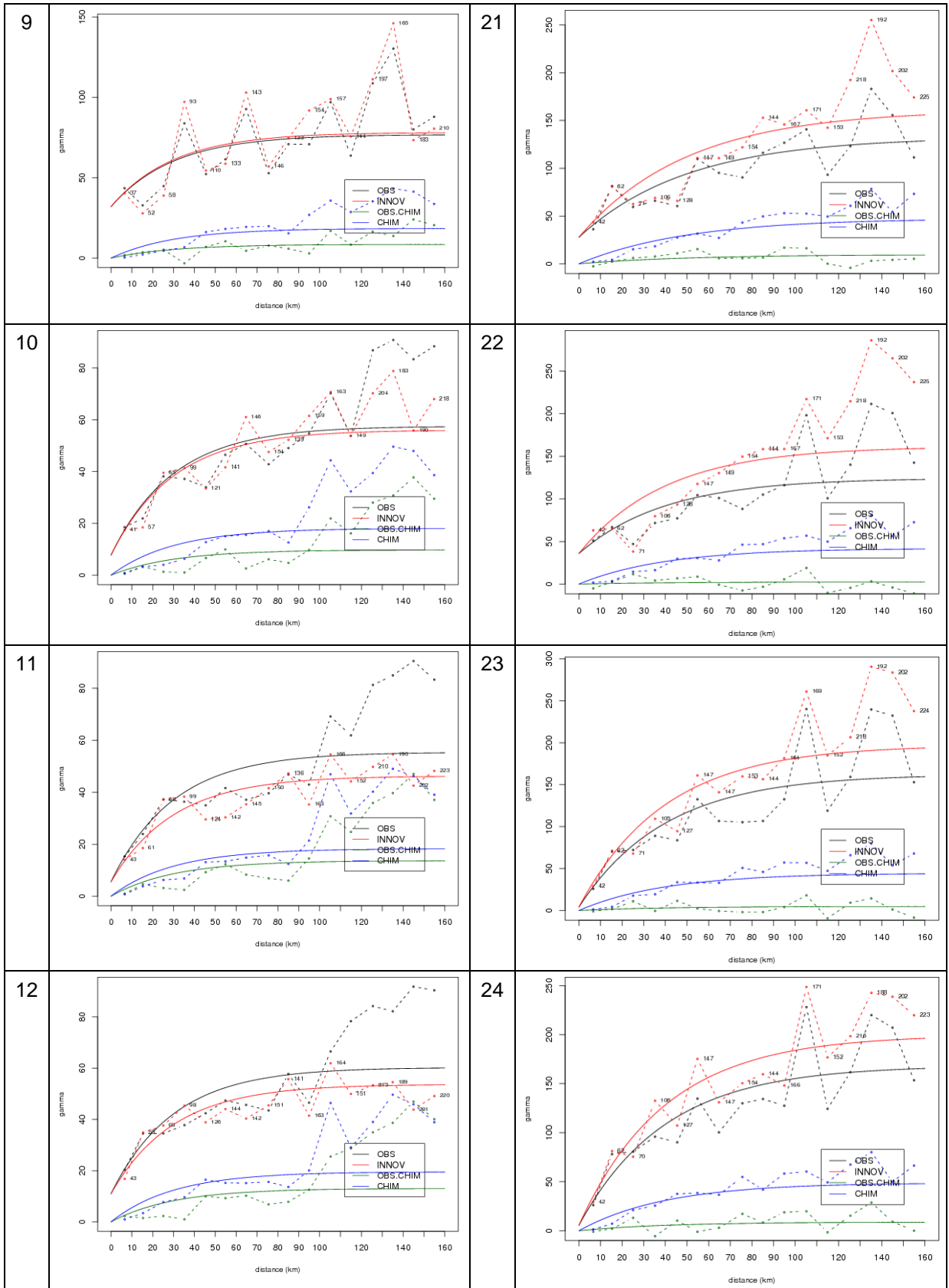
ANNEXE 5

Variogrammes spatiaux heure par heure sur la France. Exemple de la journée du
20 juillet 2005

Variogrammes expérimentaux et essai de modélisation







ANNEXE 6

Quantification du gain lié à l'analyse

Rapports y_i / x_i entre les statistiques d'erreurs ou coefficients de corrélation calculés après analyse (y_i) et les statistiques d'erreurs ou coefficients de corrélation calculés avant analyse

Comparaison entre le krigeage des innovations et le cokrigeage

Légende :

moyerr : biais moyen

stderr : écart-type du biais

rmse : racine carrée de l'erreur moyenne quadratique

moyerrabs : moyenne des erreurs en valeur absolue

moyerrrel : moyenne des erreurs relatives

stderrrel : écart-type des erreurs relatives

moyerrrelabs : moyenne des erreurs relatives en valeur absolue

cor : coefficient de corrélation

Tableau 2 – Validation croisée. Tableau des quotients y_i/x_i où y_i est une statistique d'erreur (ou un coefficient de corrélation) calculée après krigeage sur l'ensemble des heures et des stations d'une typologie donnée et x_i , cette même statistique avant krigeage. L'analyse améliore les statistiques d'erreur et le coefficient de corrélation si ce quotient est respectivement inférieur et supérieur à 1.

année et méthode	typo	moy err	stderr	rmse	moyerrabs	moyerrel	stderrel	moyerrelabs	cor
2005 CK	industrielle	-0.12	0.59	0.52	0.50	0.28	0.45	0.41	1.42
2005 KI	industrielle	-0.09	0.58	0.51	0.49	0.23	0.36	0.37	1.42
2006 CK	industrielle	-0.24	0.58	0.52	0.54	0.24	0.42	0.42	1.25
2006 KI	industrielle	-0.18	0.57	0.51	0.51	0.20	0.46	0.39	1.25
2005 CK	observation_specifique	0.07	0.51	0.37	0.32	0.24	0.39	0.34	1.60
2005 KI	observation_specifique	0.01	0.48	0.34	0.30	0.14	0.28	0.26	1.62
2006 CK	observation_specifique	0.27	0.46	0.40	0.38	0.38	0.48	0.42	1.32
2006 KI	observation_specifique	0.19	0.49	0.41	0.38	0.26	0.38	0.34	1.31
2005 CK	periurbaine	0.10	0.64	0.52	0.48	0.38	0.55	0.45	1.30
2005 KI	periurbaine	0.10	0.64	0.52	0.48	0.33	0.52	0.42	1.29
2006 CK	periurbaine	0.08	0.62	0.54	0.50	0.40	0.61	0.49	1.20
2006 KI	periurbaine	0.08	0.64	0.55	0.50	0.35	0.59	0.47	1.19
2005 CK	rurale_nationale	-0.96	1.13	1.10	1.00	-0.13	0.57	0.68	0.87
2005 KI	rurale_nationale	-0.80	1.03	0.99	0.89	-0.09	0.54	0.63	0.98
2006 CK	rurale_nationale	-1.75	1.08	1.15	1.06	-0.28	0.54	0.75	0.95
2006 KI	rurale_nationale	-1.25	0.90	0.94	0.89	-0.19	0.42	0.64	1.06
2005 CK	rurale_regionale	-0.45	0.83	0.78	0.72	0.29	0.54	0.54	1.13
2005 KI	rurale_regionale	-0.42	0.78	0.74	0.68	0.23	0.49	0.48	1.16
2006 CK	rurale_regionale	-0.37	0.80	0.75	0.69	0.26	0.56	0.55	1.11
2006 KI	rurale_regionale	-0.29	0.77	0.72	0.65	0.20	0.49	0.49	1.12
2005 CK	urbaine	0.16	0.61	0.51	0.47	0.36	0.50	0.45	1.32
2005 KI	urbaine	0.14	0.61	0.51	0.47	0.31	0.49	0.42	1.32
2006 CK	urbaine	0.17	0.58	0.50	0.46	0.37	0.60	0.47	1.21

année et méthode	typo	moy err	stderr	rmse	moyerrabs	moyerrel	stderrel	moyerrelabs	cor
2006 KI	urbaine	0.11	0.62	0.52	0.47	0.29	0.64	0.45	1.20
2005 CK	toutes	0.00	0.71	0.60	0.55	0.35	0.53	0.47	1.25
2005 KI	toutes	0.00	0.69	0.58	0.53	0.30	0.51	0.43	1.26
2006 CK	toutes	-0.01	0.68	0.60	0.55	0.35	0.59	0.50	1.17
2006 KI	toutes	-0.01	0.68	0.59	0.54	0.29	0.59	0.47	1.17

Tableau 3 – Vérification. Tableau des quotients y_i/x_i où y_i est une statistique d'erreur (ou un coefficient de corrélation) calculée après krigeage sur l'ensemble des heures et des stations d'une typologie donnée et x_i , cette même statistique avant krigeage. L'analyse améliore les statistiques d'erreur et le coefficient de corrélation si ce quotient est respectivement inférieur et supérieur à 1.

année et méthode	typo	moy err	stderr	rmse	moyerrabs	moyerrel	stderrel	moyerrelabs	cor
2005 CK	industrielle	0.07	0.60	0.78	0.62	0.41	0.51	0.50	1.16
2005 KI	industrielle	0.04	0.54	0.72	0.57	0.33	0.47	0.43	1.19
2006 CK	industrielle	0.07	0.62	0.73	0.63	0.42	0.49	0.51	1.22
2006 KI	industrielle	0.21	0.58	0.71	0.62	0.44	0.55	0.51	1.23
2005 CK	observation_specifique	0.08	0.57	0.72	0.56	0.25	0.38	0.43	1.27
2005 KI	observation_specifique	0.13	0.55	0.70	0.55	0.26	0.37	0.41	1.29
2006 CK	observation_specifique	-0.34	0.62	0.69	0.62	0.07	0.40	0.47	1.22
2006 KI	observation_specifique	-0.22	0.58	0.65	0.58	0.09	0.37	0.43	1.24
2005 CK	periurbaine	0.19	0.49	0.65	0.52	0.39	0.52	0.46	1.37
2005 KI	periurbaine	0.15	0.48	0.65	0.51	0.33	0.48	0.42	1.37
2006 CK	periurbaine	0.17	0.52	0.65	0.56	0.38	0.55	0.49	1.24
2006 KI	periurbaine	0.15	0.52	0.67	0.57	0.32	0.50	0.45	1.23
2005 CK	rurale_regionale	0.12	0.52	0.70	0.53	0.28	0.41	0.42	1.68
2005 KI	rurale_regionale	0.38	0.59	0.74	0.61	0.47	0.58	0.54	1.62
2006 CK	rurale_regionale	-0.51	0.77	0.84	0.76	-0.12	0.55	0.59	1.10

année et méthode	typo	moy err	stderr	rmse	moyerrabs	moyerrel	stderrel	moyerrel abs	cor
2006 KI	rurale_regionale	0.14	0.63	0.75	0.65	0.24	0.61	0.55	1.16
2005 CK	trafic	0.33	0.58	0.79	0.62	0.45	0.55	0.51	1.29
2005 KI	trafic	0.22	0.54	0.77	0.58	0.35	0.54	0.45	1.32
2006 CK	trafic	0.34	0.56	0.80	0.62	0.52	0.66	0.58	1.14
2006 KI	trafic	0.26	0.50	0.74	0.57	0.38	0.54	0.47	1.18
2005 CK	urbaine	0.17	0.45	0.59	0.48	0.35	0.49	0.42	1.39
2005 KI	urbaine	0.16	0.45	0.60	0.49	0.30	0.46	0.39	1.39
2006 CK	urbaine	0.18	0.46	0.58	0.49	0.37	0.54	0.45	1.24
2006 KI	urbaine	0.19	0.47	0.60	0.51	0.33	0.54	0.43	1.23
2005 CK	toutes	0.17	0.48	0.63	0.51	0.37	0.50	0.44	1.37
2005 KI	toutes	0.16	0.47	0.63	0.51	0.31	0.47	0.40	1.37
2006 CK	toutes	0.16	0.49	0.62	0.53	0.37	0.54	0.47	1.23
2006 KI	toutes	0.17	0.49	0.63	0.54	0.33	0.53	0.44	1.23