



**HAL**  
open science

# Spatialisation des données de modélisation par Krigeage

Marie Hennequi

► **To cite this version:**

Marie Hennequi. Spatialisation des données de modélisation par Krigeage. Méthodologie [stat.ME]. 2010. dumas-00520260

**HAL Id: dumas-00520260**

**<https://dumas.ccsd.cnrs.fr/dumas-00520260v1>**

Submitted on 22 Sep 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Spatialisation des données de modélisation par Krigeage

---

Marie HENNEQUI

Master Statistique et Applications  
Stage de 1<sup>ère</sup> année du 14 juin au 15 août

# Remerciements

Je tiens tout d'abord à remercier Alain TARGET et Joseph KLEIPETER, respectivement directeur et directeur adjoint, pour m'avoir offert l'opportunité de réaliser mon stage de Master au sein de l'ASPA.

Un grand merci à Matthieu BERNARD qui m'a guidé au cours de ce stage et m'a permis de mener à bien ce projet.

Merci à Gilles PERRON pour ses conseils avisés sur la géostatistique et à Raphaèle DEPROST pour le temps qu'elle m'a consacré .

Merci à mes collègues stagiaires Thibaut DE RUFFRAY, Matthieu BOSANSKY et Guillaume BOSCHETTO pour leur bonne humeur.

Enfin, merci à tout le personnel de l'ASPA pour leur accueil chaleureux et pour m'avoir permis de découvrir différents métiers présents à l'ASPA.

# Table des matières

<b>1</b>	<b>Contexte et objectifs du stage</b>	<b>5</b>
1.1	Présentation de l'ASPA	5
1.2	Objectifs du stage	6
<b>2</b>	<b>Matériel et méthodes</b>	<b>8</b>
2.1	Une méthode géostatistique : Le Krigeage	8
2.1.1	Analyse variographique	8
2.1.2	Le krigeage	11
2.1.3	Validation croisée	12
2.2	Présentation des données	13
2.2.1	Stations de mesure	13
2.2.2	Les différents polluants étudiés	14
2.3	Mise en oeuvre du krigeage	17
<b>3</b>	<b>Présentation des résultats</b>	<b>21</b>
3.1	Résultats pour le reporting	21
3.1.1	NO <sub>2</sub>	21
3.1.2	PM10	26
3.1.3	Ozone	35
3.1.4	Conclusion	38
3.2	Résultats pour la prévision	38
3.2.1	Maximum journalier de NO <sub>2</sub>	39
3.2.2	Moyenne journalière en PM10	42
3.2.3	Maximum journalier d'ozone	43
	<b>Annexes</b>	<b>46</b>
<b>A</b>	<b>Le Krigeage avec R</b>	<b>48</b>
A.1	Analyse variographique avec R	48
A.1.1	Le variogramme expérimental	48
A.1.2	Modélisation du variogramme	50
A.2	Le krigeage avec R	52
A.3	La validation croisée avec R	54
<b>B</b>	<b>Code R</b>	<b>55</b>
B.1	Script pour le reporting	55
B.2	Script pour le nombre de jours de dépassement en PM10	60
B.3	Script pour la prévision	64
B.4	Fichier Base	69

# Chapitre 1

## Contexte et objectifs du stage

### 1.1 Présentation de l'ASPA <sup>1</sup>

Regroupant des représentants de l'Etat, des industriels, des collectivités locales et territoriales ainsi que des mouvements associatifs (environnement, consommateurs, milieu de la santé) et de personnalités qualifiées, l'Association pour la Surveillance et l'étude de la Pollution atmosphérique en Alsace (ASPA) ne poursuit qu'un seul objectif : l'amélioration de la qualité de l'air. D'un effectif d'une trentaine de personnes (techniciens, ingénieurs, administratifs...), l'équipe est dirigée par Alain Target.

Créée pour mesurer la pollution industrielle dans l'agglomération strasbourgeoise et assurer la mise en oeuvre des procédures d'alerte, l'ASPA n'a cessé de voir s'accroître son expertise, son champ d'action et ses moyens, sous les effets conjugués de la demande et du développement du cadre législatif et réglementaire. Une étape décisive a été, en 1988, l'extension des mesures de la qualité de l'air à tout le territoire alsacien avec l'adhésion de la région Alsace, de la préfecture du Haut-Rhin, de collectivités locales et d'industriels haut-rhinois. En 1996, la Loi sur l'air a donné une impulsion et une légitimité nouvelle à l'ASPA et à toutes les associations françaises, qui se sont finalement regroupées sous la bannière de la fédération atmo. Dans cette évolution, l'ASPA a toujours joué un rôle de pionnier. En effet, toujours à la pointe de la conquête de l'air, elle a dû, pour faire face à des attentes croissantes, étoffer son expertise et ses moyens.

#### Les missions de l'ASPA

Le rôle de l'ASPA ne se résume pas à la simple production et diffusion de données élémentaires de qualité de l'air. Le découpage ci-après en quatre familles fait apparaître des missions répondant à des objectifs bien identifiés des récents textes réglementaires (mesure de la qualité de l'air, alerte, information permanente, surveillance en référence aux normes, prévision des pointes) et fait ressortir l'expertise acquise progressivement par les réseaux.

---

<sup>1</sup>source : <http://www.atmo-alsace.net/>

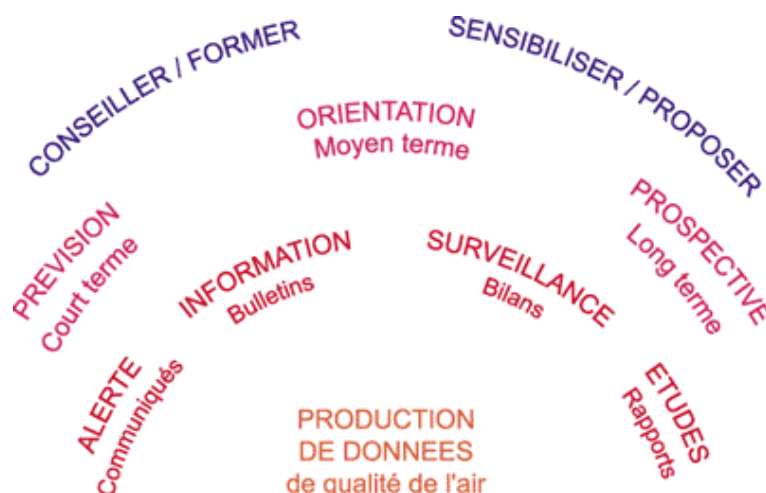


FIG. 1.1: Les missions de l'ASPA

- Mission centrale : **La production des données élémentaires de qualité de l'air** comprend d'une part la gestion d'un réseau de stations permanentes de mesure en continu de la qualité de l'air (cf. paragraphe 2.2.1) et l'organisation de campagnes de mesures temporaires et d'autre part la mise en oeuvre de modèles de prévision, de spatialisation ou de diagnostic fournissant des données estimées dans le temps et/ou dans l'espace.
- Missions de bases : La diffusion de données commentées est à réaliser pour les situations **d'alerte**, pour **l'information** permanente, sous forme de bilans récapitulatifs en référence aux normes (appelée **surveillance** au sens restreint du terme) et plus largement sous forme **d'études** diverses sur les phénomènes de pollution, les impacts des activités, etc.
- Missions d'expertises : Elles concernent la prévision des situations de pollution (à terme mission de base) allant au delà de la simple utilisation de modèle, **l'orientation** à moyen terme des planifications et projets urbains, industriels et routiers touchant à la qualité de l'air, la **prospective** visant à anticiper sur les évolutions probables de celle-ci en projetant dans l'avenir les évolutions des paramètres qui la conditionnent.
- Missions pédagogiques et stratégiques : Valorisant les connaissances, compétences et expériences acquises dans le domaine de la qualité de l'air, ces missions contribuent au **conseil/formation** et à la **sensibilisation/proposition** auprès des autorités, des acteurs publiques ou privés, du grand public, des milieux scolaires et universitaires dans l'optique d'une meilleure compréhension des enjeux liés à la qualité de l'air et des réponses qu'il convient d'apporter voire de susciter. La valorisation du savoir faire technique pour la mise en place et la mise en oeuvre d'un outil de surveillance de la qualité de l'air entre également dans cette catégorie.

## 1.2 Objectifs du stage

Le stage comporte un double objectif. La spatialisation des sorties de modélisation est appliquée à deux échelles temporelles différentes au sein de l'ASPA : une journalière pour les données de prévision et une annuelle dans le cadre du reporting européen.

Jusqu'à présent ces deux traitements sont complètement découplés et ne font pas intervenir les mêmes méthodes de spatialisation. La spatialisation des données de prévision est basée sur une technique de pondération de l'inverse distance, automatisée en interne sous Fortran. La spatialisation des données nécessaires au reporting fait intervenir la technique géostatistique dite du Krigeage par l'intermédiaire du logiciel spécialisé dans le traitement géostatistique «Isatis». Cette procédure n'est pas automatisée puisque cette méthode demande une forte expertise.

Le but de ce stage est donc d'unifier les méthodes utilisées en conservant uniquement la méthode

dite de Krigeage. Cela est possible grâce à l'utilisation du langage de programmation R dans lequel cette méthode est implémentée ; ce qui rend son automatisation réalisable.

Par la suite, la cohérence des résultats du Krigeage sous R sera estimée en fonction de plusieurs critères dont la cohérence avec les cartes produites par l'ancien système qui bénéficiait d'une forte expertise humaine.

Si les résultats sont jugés satisfaisants sur les données annuelles, cela permettra d'automatiser la tâche du reporting et de libérer des unités d'oeuvre de la personne en charge de la géostatistique.

Suivant la cohérence des sorties annuelles, la méthode sera appliquée en journalier, tout d'abord pour tester la faisabilité technique. La phase de validation nécessite une phase de test sur plusieurs semaines de manière à appréhender les différents comportements et bugs éventuels de la spatialisation. Des résultats sur quelques dates seront néanmoins présentés à la fin de ce rapport.

## Chapitre 2

# Matériel et méthodes

Afin de réaliser les cartes de reporting et de prévision, une méthode géostatistique a été retenue : le krigeage. Ce chapitre présente brièvement cette technique de modélisation spatiale puis expose les données qui sont à disposition pour permettre de réaliser les cartes.

### 2.1 Une méthode géostatistique : Le Krigeage

Le krigeage est une technique géostatistique de modélisation spatiale permettant, à partir de données dispersées, d'obtenir une représentation homogène des informations étudiées. Dans l'analyse de la pollution atmosphérique, on ne dispose que d'un certain nombre de stations de mesures qui vont fournir les données. Le krigeage nous permettra donc, à l'aide des mesures de concentrations obtenues en stations, d'estimer les concentrations hors station. Il sera alors possible de créer une carte étendant les relevés à tout l'espace. D'autres techniques géostatistiques permettent de faire ce travail mais le krigeage a l'avantage de prendre en compte les distances entre les données (i.e. les stations de mesure), les distances entre les données et la cible (i.e. le point pour lequel on veut estimer la mesure) et la structure spatiale (grâce à l'analyse variographique).

Les méthodes statistiques classiques telles que la régression linéaire se basent sur une hypothèse fondamentale qui est l'indépendance des variables. Or, lorsqu'une variable est spatialement autocorrélée, cette hypothèse n'est plus vérifiée. Ainsi, le krigeage va se baser sur cette nouvelle hypothèse : l'autocorrélation spatiale des données. Concrètement, cela signifie que deux données rapprochées dans l'espace tendent à posséder des caractéristiques similaires.

Dans la théorie de l'interpolation spatiale, la région de l'espace géographique concernée par l'étude est appelée «champ», notée  $D$ . La variable régionalisée que l'on étudie est vue comme une réalisation d'une fonction aléatoire  $\{Z(s), s \in D\}$  et toute valeur régionalisée  $z(s_i)$  est considérée comme une réalisation d'une variable aléatoire  $Z(s_i)$ .

#### 2.1.1 Analyse variographique

Pour appliquer ce type d'interpolation par krigeage, il est nécessaire d'explorer la structure spatiale des données afin de vérifier si celles-ci sont bien autocorrélées. L'analyse variographique va nous permettre de mener cette étude à bien. L'outil principal permettant cette analyse est le semi-variogramme qui décrit l'évolution de la semi-variance en fonction de la distance entre les mesures et permet ainsi de d'étudier le lien spatial entre les données. Il est défini de la manière suivante :

$$\gamma(h) = \frac{1}{2} \text{Var}(Z(s+h) - Z(s)) \quad \forall s \in D$$

où :

- $Z(\cdot)$  est la variable régionalisée étudiée
- $s$  est le vecteur des coordonnées



- $h$  est le vecteur distance
- $D$  est le domaine géographique considéré.

Le graphique 2.1 expose un exemple de variogramme ainsi que les propriétés que l'on peut rencontrer.

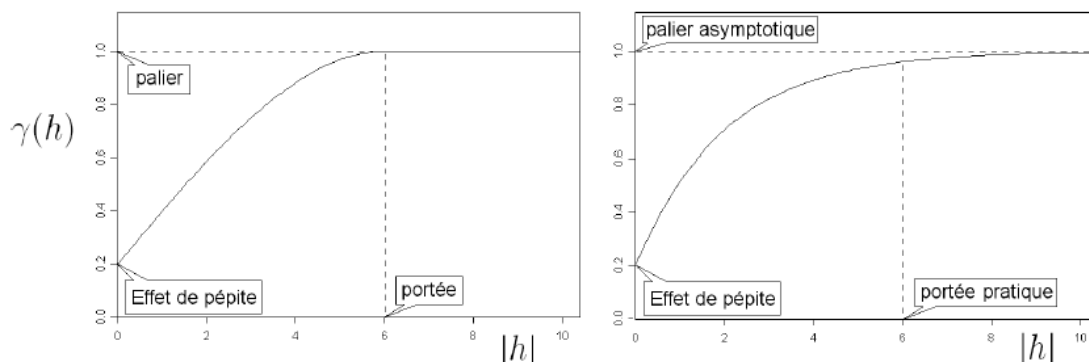


FIG. 2.1: Propriétés d'un variogramme

Ces propriétés sont les suivantes :

- **Effet de pépité :**

L'effet de pépité correspond à la limite du variogramme en zéro. Elle représente donc la variation entre deux mesures très proches et peut donc provenir de trois effets :

- une variabilité naturelle du paramètre mesuré (dans le cas où il est mesuré à deux instants différents par exemple).
- une variabilité de l'instrument de mesure : la pépité mesure donc en partie l'erreur statistique de l'instrument de mesure.
- un réel effet pépité : une variation brutale du paramètre mesuré.

- **Portée et Palier :**

Lorsque  $h$  augmente, le variogramme  $\gamma(h)$  peut ou non atteindre un plateau. L'atteinte d'un plateau indique qu'à partir d'une certaine distance, il n'y a plus de dépendance spatiale entre les données. Cette distance est nommée **portée** et le terme **palier** dénote la variance à laquelle le plateau se présente. Un palier peut n'être atteint qu'asymptotiquement. Dans ce cas, la portée réelle est infinie mais une portée pratique est définie par la distance à laquelle le variogramme atteint 95% de la valeur de son palier.

En pratique, l'analyse variographique va se dérouler en deux étapes :

### Estimation du variogramme

A l'aide des mesures obtenues sur les différentes stations, on estime le variogramme par la méthode des moments (méthode qui consiste à estimer l'espérance par la moyenne empirique). Cet estimateur est défini de la façon suivante :

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{N(h)} [z(s_i) - z(s_j)]^2$$

où :

- $h$  correspond à une distance
- $z(s_i)$  correspond à la valeur de la variable régionalisée au point  $s_i$
- $N(h) = \{(i, j) \text{ tel que } s_i - s_j = h\}$
- $|N(h)|$  est le nombre de paires distinctes de l'ensemble  $N(h)$

Si les données sont réparties sur une grille régulière couvrant le champ  $D$ ,  $\gamma(\cdot)$  est estimable pour un petit nombre de distances, chacune associée à plusieurs couples de points. Toutefois, les données sont souvent réparties de façon irrégulière sur  $D$ . Dans ce rapport, nous serons dans ce cas, les stations de mesures étant irrégulièrement réparties sur le territoire alsacien.

Une notion de **pas**, noté  $p$ , est alors introduite. Le variogramme sera calculé seulement pour certaines distances  $\{r_1 = p, r_2 = 2p, \dots, r_k = kp, \dots, r_n = P\}$  où  $P$  est la **portée maximale** qui correspond à la distance maximale de prise en compte des mesures. L'ensemble  $N(r_k)$  est ainsi redéfini par  $N(r_k) = \{(i, j) \text{ tel que } r_{k-1} < |s_i - s_j| \leq r_k\}$  avec  $r_0=0$ .

Le calcul du variogramme expérimental (variogramme estimé) requiert donc la spécification de ces deux paramètres. Une «mauvaise» spécification de ces derniers peut conduire à des variogrammes expérimentaux impossibles à interpréter et à modéliser. Un pas trop petit fera que peu de couples de stations seront pris en compte pour le calcul du variogramme alors qu'un pas trop grand génèrera un variogramme trop lisse.

Pour les besoins du stage, nous utiliserons le logiciel R pour estimer le variogramme expérimental (cf. Annexe A pour le manuel d'utilisation). Les paramètres **pas** et **portée maximale** sont à fournir en degré. Pour la région Alsace, on appliquera la correspondance suivante : 1 degré vaut environ 100km.

### Modélisation du variogramme

Une fois le variogramme estimé, il convient de le modéliser par une fonction de référence (modèle exponentiel, sphérique...). Ce modèle aura alors toutes les propriétés requises pour effectuer un krigeage sur nos données.

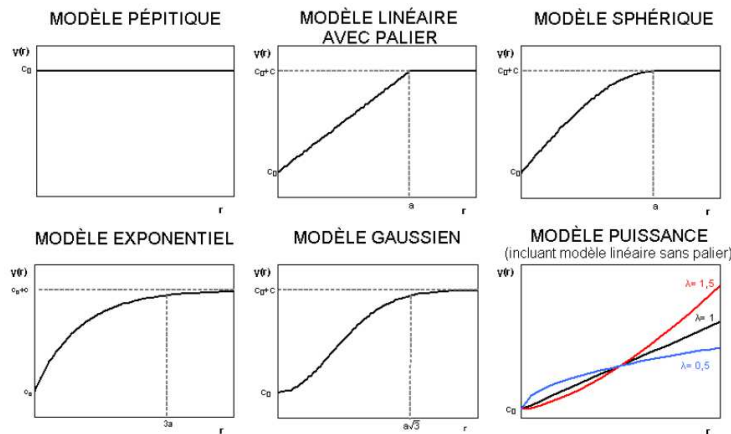


FIG. 2.2: Représentation des différents modèles de variogramme

Le modèle le plus couramment utilisé (et celui utilisé pour ce stage) est le modèle Sphérique défini de la manière suivante :

$$\gamma(r) = \begin{cases} c_0 + c \left( \frac{3}{2} \frac{r}{a} - \frac{1}{2} \frac{r^3}{a^3} \right) & \text{pour } 0 \leq r \leq a \\ c_0 + c & \text{pour } r > a \end{cases}$$

où :

- $c_0$  correspond à l'effet pépité
- $c_0 + c$  correspond au palier
- $a$  correspond à la portée

## 2.1.2 Le krigeage

Le modèle de base du krigeage a la même forme que le modèle de régression classique mais les erreurs sont maintenant supposées dépendantes spatialement. Il s'énonce de la façon suivante :

$$Z(s) = \mu(s) + \delta(s) \quad \forall s \in D$$

où :

- $\mu(\cdot)$  constitue la structure déterministe pour l'espérance de  $Z$ .
- $\delta(\cdot)$  est une fonction aléatoire stationnaire.

Pour formuler complètement le modèle, il faut spécifier la forme de la tendance  $\mu(\cdot)$ . C'est en fait cette tendance qui précise le type de krigeage effectué. Les quatre types classiques de krigeage sont les suivants :

- Le krigeage simple :  $\mu(s) = m$  est une constante connue.
- Le krigeage ordinaire :  $\mu(s) = \mu$  est une constante inconnue.
- Le krigeage universel :  $\mu(s) = \sum_{j=0}^p f_j(s)\beta_j$  est une combinaison linéaire de fonctions de la position  $s$ .
- Le krigeage avec dérive externe :  $\mu(s) = \sum_{j=0}^p f_j(w)\beta_j$  où  $w = (w_1(s), \dots, w_q(s))$  est le vecteur des valeurs prises par les  $q$  variables régionalisées auxiliaires au point  $s$ . Ce type de krigeage consiste à interpoler la variable régionalisée  $Z(\cdot)$  en ajoutant des variables explicatives au modèle (appelées variables auxiliaires).

L'idée de base du krigeage est de prévoir la valeur d'une variable régionalisée  $z$  en un site non échantillonné  $s_0$  par une combinaison linéaire des données ponctuelles adjacentes. On veut une estimation non biaisée et de variance minimale. Ces conditions se résument en quatre contraintes qui vont permettre de construire le système de krigeage :

### - Contrainte de linéarité

L'estimateur est une combinaison linéaire pondérée des données. Ainsi, dans le cas où l'on souhaite estimer la valeur de la variable régionalisée au point  $s_0$ , on posera :

$$\hat{Z}(s_0) = a + \sum_{i \in V(s_0)} \lambda_i Z(s_i)$$

où les  $s_i$  ( $i=1 \dots n$ ) sont les sites échantillonnés utilisés pour l'estimation, et où la constante  $a$  et les poids  $\lambda_i$  sont les inconnues du problème.

### - Contrainte d'autorisation

L'erreur d'estimation doit être une combinaison linéaire autorisée, c'est-à-dire que son espérance et sa variance doivent exister.

### - Contrainte de non-biais

L'espérance de l'erreur d'estimation doit être nulle. Intuitivement, cela signifie que si l'on répète l'estimation sur un grand nombre de configurations de krigeage identiques, les erreurs commises se neutralisent.

– *Contrainte d'optimalité*

La dernière étape consiste à rechercher des pondérateurs qui minimisent la variance de l'erreur d'estimation, sous les contraintes précédentes. En terme plus concrets, si l'on répète l'estimation sur de nombreuses configurations de krigeage identiques, la variance des erreurs est la plus faible possible.

Quelque soit le type de krigeage utilisé, le système de krigeage est toujours obtenu au terme de ces quatre étapes et dépend du modèle de variogramme utilisé.

Nous ne développerons pas dans ce rapport la construction du krigeage ni les formules obtenues mais il est possible de se référer au livre d'Arnaud et Emery ([1], 2000) pour plus de détails.

**Remarque :** Le type de krigeage utilisé pour la suite de ce rapport est le krigeage dit ordinaire qui suppose que pour tout point du domaine  $s_i$ , la variable aléatoire  $Z(s_i)$  a une espérance constante inconnue.

Le krigeage est une méthode d'interpolation exacte c'est-à-dire qu'il restitue les valeurs régionalisées mesurées aux sites d'observations. Il ne sera donc pas possible d'analyser directement les résidus du modèles (car ils seront tous nuls). C'est pour cette raison que pour valider un modèle ou comparer deux modèles entre eux, on utilise la validation croisée présentée ci-dessous.

### 2.1.3 Validation croisée

Lors de ce stage, nous avons utilisé la validation croisée dite «leave-one-out».

Supposons que l'on ait un échantillon d'observation  $\{Z(s_i), i = 1 \dots N\}$ .

Le processus de la validation croisée «leave-one-out» est le suivant :

- On retire l'observation  $i$ , notée  $Z(s_i)$  de l'échantillon et on lance le modèle avec le nouvel échantillon.
- Puis à l'aide de ce nouveau modèle, on calcule la valeur prédite de l'observation retirée  $i$  que l'on note  $\hat{Z}_{[i]}(s_i)$ .

On réitère ainsi ce processus sur toutes les observations dont on dispose. On obtient alors une série de valeurs prédites que l'on peut comparer aux valeurs observées en calculant les résidus du modèle.

Il est ainsi possible d'étudier les erreurs commises sur l'ensemble des données en regardant leur moyenne et leur écart-type. Cependant, l'étude des erreurs n'est pas toujours satisfaisante. En effet, l'amplitude des erreurs commises en un site de données ne dépend pas seulement de la qualité du modèle variographique, mais aussi de l'éloignement par rapport aux autres sites. Ainsi, un site isolé tend à produire une erreur d'estimation élevée. Cet effet peut être corrigé en examinant les erreurs standardisées, c'est-à-dire les erreurs divisées par les écarts-types de krigeage. Celles-ci sont alors ramenées à la même échelle et ne sont plus sensibles à la configuration géométrique des échantillons.

Une fois la validation croisée effectuée, le calcul des différents critères permet de comparer deux modèles entre eux. Voici une liste des critères les plus utilisés dans la validation croisée :

– **Biais moyen** =  $\frac{1}{N} \sum_{i=1}^N (\hat{Z}_{[i]}(s_i) - Z(s_i))$

– **MSE** =  $\frac{1}{N} \sum_{i=1}^N (\hat{Z}_{[i]}(s_i) - Z(s_i))^2$

– **Critère d'adéquation** =  $\frac{1}{N} \sum_{i=1}^N \left( \frac{\hat{Z}_{[i]}(s_i) - Z(s_i)}{\sigma_{[i]}(s_i)} \right)^2$

où  $\sigma_{[i]}^2(s_i)$  correspond à la variance de krigeage, c'est-à-dire  $\sigma_{[i]}^2(s_i) = Var(\hat{Z}_{[i]}(s_i) - Z(s_i))$

$$- \text{PRESS} = \sum_{i=1}^N \left( \widehat{Z}_{[i]}(s_i) - Z(s_i) \right)^2.$$

Pour ces quatre critères, le «meilleur» modèle est celui qui a le plus petit critère étudié. Un modèle sera d'autant plus performant si ces critères sont proches de 0.

## 2.2 Présentation des données

### 2.2.1 Stations de mesure

L'ASPA gère aujourd'hui, des bords du Rhin jusqu'aux sommets vosgiens, un réseau régional de surveillance de la qualité de l'air avec une trentaine de stations (urbaines, industrielles et rurales). Au total, environ 80 analyseurs mesurent en continue 24/24 heures les niveaux d'oxydes d'azote, de particules, d'ozone et d'autres polluants (tels que le plomb, le dioxyde soufre, le monoxyde de carbone...). Une station effectue un relevé de mesure toutes les heures; les données à disposition seront donc des données horaires.

Les concentrations relevées en stations constituent les données de départ pour effectuer le krigeage et réaliser les cartes attendues. En fonction du polluant étudié, différentes stations entreront en compte pour le krigeage. En effet, comme le montre la carte 2.3, il existe plusieurs types de stations qui ne mesurent pas toutes les mêmes polluants. Les stations «trafic» et «industrielles» ne seront jamais prises en compte dans le krigeage car ces dernières donnent uniquement une information de proximité sur respectivement le trafic routier et l'industrie concernée. Elles ne représentent donc pas des phénomènes régionaux mais des phénomènes à petite échelle. Le but étant de représenter les comportements des différents polluants à une échelle régionale, ces deux types de stations ont donc été évincées.

Les stations utilisées pour chaque polluant seront présentées dans le chapitre 3 lors de la présentation des cartes obtenues.

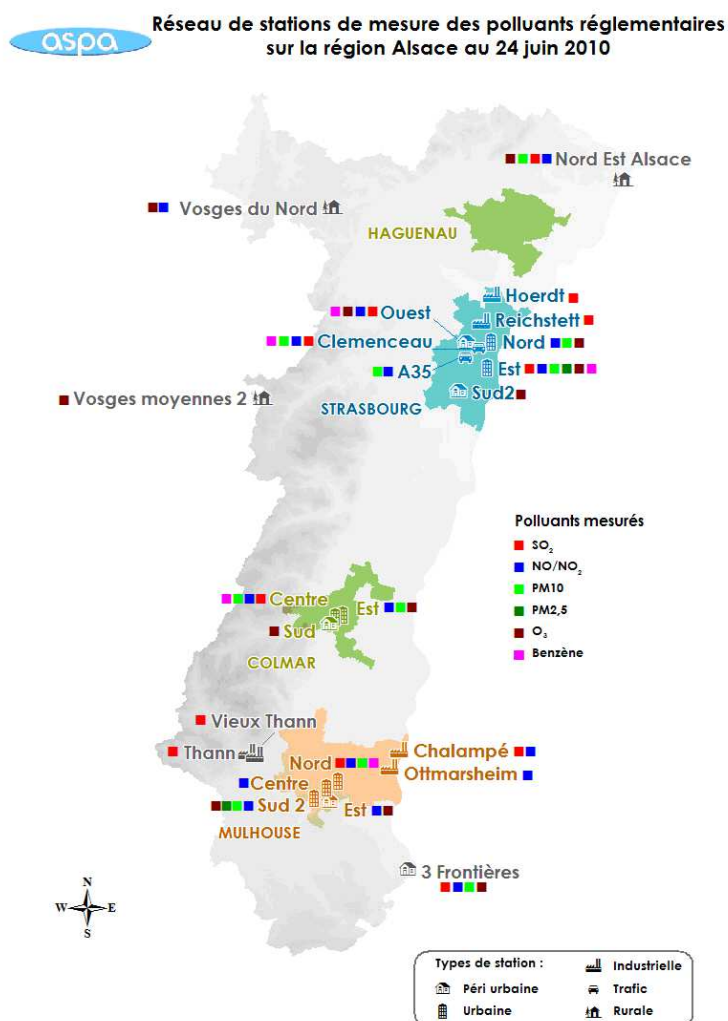


FIG. 2.3: Réseau des Stations de mesures de l’ASPA

### 2.2.2 Les différents polluants étudiés

La directive européenne de 1992 (modifiée en 2003) fixe les seuils d’information et d’alerte de la population ainsi que les valeurs limites pour la protection de la santé humaine, de la végétation et de la forêt qu’il convient, à terme, de ne pas dépasser. Cette directive rend obligatoire la mesure dans l’air de teneurs des différents polluants afin d’étudier l’impact de ces derniers sur la santé et l’environnement.

Durant mon stage, deux missions m’ont été confiées. La première consistait à produire des cartes nécessaires pour le reporting 2009, c’est-à-dire reconstruire de la donnée à une échéance annuelle (moyennes annuelles, nombre de jours de dépassement du seuil d’alerte, etc.). La deuxième avait pour but d’établir les cartes de prévision, c’est-à-dire travailler sur des données journalières (maximum journalier, moyenne journalière). Voici un récapitulatif des polluants et des indices demandés :

#### Cartes pour le dioxyde d’azote (NO<sub>2</sub>)

Les rejets de NO<sub>2</sub> proviennent essentiellement de la combustion de combustibles de tous types (gazole, essence, charbons, fiouls, GN...). Ils se forment par combinaison de l’azote (atmosphérique et contenu

dans les combustibles) et de l'oxygène de l'air à hautes températures. Tous les secteurs utilisateurs de combustibles sont concernés, en particulier les transports routiers.

*Les effets sur la santé* : Le NO<sub>2</sub> est un gaz irritant pour les bronches. Chez les asthmatiques, il augmente la fréquence et la gravité des crises. Chez l'enfant, il favorise les infections pulmonaires.

*Les effets sur l'environnement* : Le NO<sub>2</sub> participe aux phénomènes des pluies acides, à la formation de l'ozone troposphérique, dont ils sont les précurseurs, et à l'atteinte de la couche d'ozone stratosphérique comme à l'effet de serre.

- Objectif de Qualité : 40 µg/m<sup>3</sup> en moyenne annuelle
- Seuil d'Information et de recommandation : 200 µg/m<sup>3</sup> en moyenne horaire
- Seuil d'Alerte : 400 µg/m<sup>3</sup> en moyenne horaire abaissé à 200 µg/m<sup>3</sup> en cas de persistance

Les cartes demandées pour le NO<sub>2</sub> sont les suivantes :

#### Reporting

- **Moyenne annuelle** de NO<sub>2</sub>.
- **Percentile horaire 99,8** de NO<sub>2</sub>. Ce percentile se base sur les données horaires de concentration de NO<sub>2</sub>. Il correspond à la concentration pour laquelle 99,8% des valeurs horaires sur une année sont en-dessous de cette concentration.

#### Prévision

- **Maximum horaire journalier** en NO<sub>2</sub>.

### Cartes pour les PM10 (Particules)

Les particules ou poussières en suspension liées à l'activité humaine proviennent majoritairement de la combustion des matières fossiles, du transport automobile (gaz d'échappement, usure, frottements...) et d'activités industrielles très diverses (sidérurgie, incinération...). Leur taille et leur composition sont très variables. Les PM10 représentent la catégorie de particules dont le diamètre est inférieur 10 micromètres (les PM2,5, ou très fines particules, ont un diamètre inférieur à 2,5 micromètres mais ne seront pas abordées dans ce rapport).

*Les effets sur la santé* : Selon leur taille, les particules pénètrent plus ou moins profondément dans l'arbre pulmonaire. Les particules les plus fines peuvent, à des concentrations relativement basses, irriter les voies respiratoires inférieures et altérer la fonction respiratoire dans son ensemble. Certaines particules ont des propriétés mutagènes et cancérigènes.

*Les effets sur l'environnement* : Les effets de salissure des bâtiments et des monuments sont les atteintes à l'environnement les plus évidentes.

- Objectif de Qualité : 30 µg/m<sup>3</sup> en moyenne annuelle

Les cartes demandées pour les PM10 sont les suivantes :

#### Reporting

- **Moyenne annuelle** de PM10.
- **Percentile journalier 90,4** de PM10. Ce percentile est défini de la manière suivante : Pour chaque jour, on calcule la moyenne (sur 24 heures) de la concentration en PM10. Le percentile 90,4 correspond alors à la concentration de PM10 pour laquelle 90,4% des moyennes obtenues sur une année sont situées en-dessous de cette valeur.
- **Nombre de jours de dépassement** sur une année. Pour chaque jour, on calcule la moyenne horaire des concentrations de PM10 et on regarde le nombre de jours pour lesquels la moyenne est supérieure à 50µg/m<sup>3</sup>.

## Prévision

- Moyenne journalière en PM10.

### Cartes pour l'ozone (O<sub>3</sub>)

Dans la stratosphère (entre 10 et 60 km d'altitude), l'ozone constitue un filtre naturel qui protège la vie sur terre de l'action néfaste des ultraviolets "durs". Le "trou dans la couche d'ozone" est une disparition partielle de ce filtre, liée à l'effet "destructeur d'ozone" de certains polluants émis dans la troposphère et qui migrent lentement dans la stratosphère.

Dans la troposphère (entre le sol et 10 km), les taux d'O<sub>3</sub> devraient être naturellement faibles. Cet ozone est un polluant dit "secondaire". Il résulte généralement de la transformation chimique dans l'atmosphère de certains polluants dits "primaires" (en particulier NO (oxyde d'azote), NO<sub>2</sub> (dioxyde d'azote) et COV (Composés Organiques Volatils)), sous l'effet des rayonnements solaires. Les mécanismes réactionnels sont complexes et les plus fortes concentrations d'O<sub>3</sub> apparaissent l'été, en périphérie des zones émettrices des polluants primaires, puis peuvent être transportées sur de grandes distances. Le cycle de l'ozone est résumé par la figure 2.4.

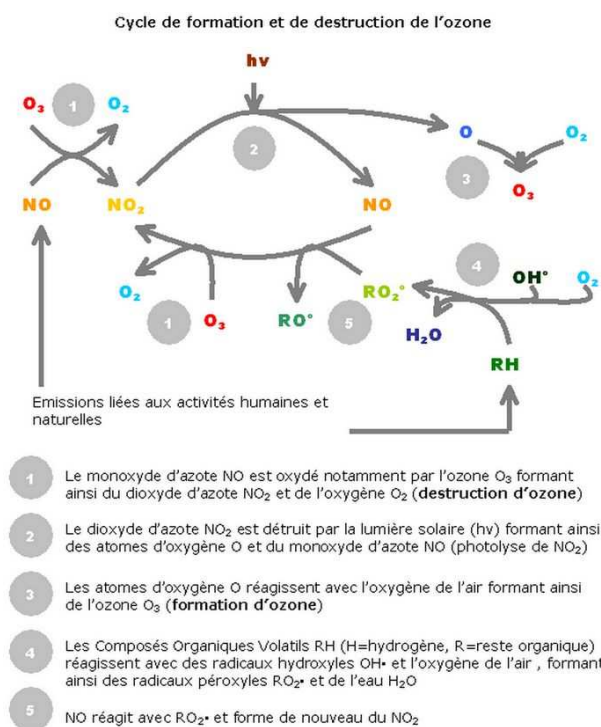


FIG. 2.4: Cycle de formation et destruction de l'ozone

*Les effets sur la santé :* L'O<sub>3</sub> est un gaz agressif qui pénètre facilement jusqu'aux voies respiratoires les plus fines. Il provoque toux, altération pulmonaire ainsi que des irritations oculaires. Ses effets sont très variables selon les individus.

*Les effets sur l'environnement :* L'O<sub>3</sub> a un effet néfaste sur la végétation (sur le rendement des cultures par exemple) et sur certains matériaux (caoutchouc...). Il contribue également à l'effet de serre.

- Objectif de Qualité : 120 µg/m<sup>3</sup> en maximum journalier de la moyenne sur 8h
- Seuil d'Information et de recommandation : 180 µg/m<sup>3</sup> en moyenne horaire
- Seuil d'Alerte : 240 µg/m<sup>3</sup> en moyenne horaire



Les dispositions de la réglementation concernant la protection de la forêt et de la végétation sont peu connues du grand public. La végétation est sensible aux effets chroniques de l'ozone (pollution de fond) plutôt qu'aux effets ponctuels dus aux pics de pollution. Les niveaux critiques à long terme pour l'ozone sont exprimés à partir d'une exposition cumulative au-dessus d'une concentration de 40 parties par milliard (ppb) (ou  $80 \mu\text{g}/\text{m}^3$ ). L'indice utilisé pour évaluer le degré d'exposition des plantes à cette pollution est l'AOT 40. Il est exprimé en  $\mu\text{g}/\text{m}^3$  par heure et correspond à la somme des différences entre les concentrations horaires supérieures à 40 ppb durant une période donnée en utilisant uniquement les valeurs sur 1 heure mesurées quotidiennement entre 8 heures et 20 heures. Les valeurs limites de l'AOT 40, calculée à partir de valeurs moyennes horaires mesurées de mai à juillet pour la végétation (AOTV) et d'avril à septembre pour la forêt (AOTF) sont les suivantes :

- Végétation :  $6000 \mu\text{g}/\text{m}^3/\text{h}$  (3000 ppb/h)
- Forêt :  $20000 \mu\text{g}/\text{m}^3/\text{h}$  (10000 ppb/h)

Les cartes demandées pour l'ozone sont les suivantes :

#### Reporting

- **AOTF** (forêt), défini ci-dessus.
- **AOTV** (végétation), défini ci-dessus.

#### Prévision

- **Maximum horaire journalier** en ozone.

## 2.3 Mise en oeuvre du krigeage

Pour les prévisions et le reporting, l'ASPA utilise des données de modélisation, à échelle régionale, issues de modèle déterministes. La plateforme de modélisation est composée du modèle météorologique MM5 couplé au modèle de chimie/transport CHIMERE (cf. figure 2.5). Ce dernier permet de prévoir les champs de concentration des polluants précités en prenant en compte les phénomènes de dispersion pure et de chimie de l'atmosphère en fonction des champs météorologiques et des émissions. Ce modèle présente un biais négatif, c'est-à-dire qu'il sous-estime souvent les concentrations réelles des polluants mesurés. Afin d'obtenir des cartes plus justes, le modèle CHIMERE est corrigé à l'aide des mesures effectuées en stations. C'est lors de cette étape que le krigeage intervient.

## Chaîne de modélisation de la qualité de l'air

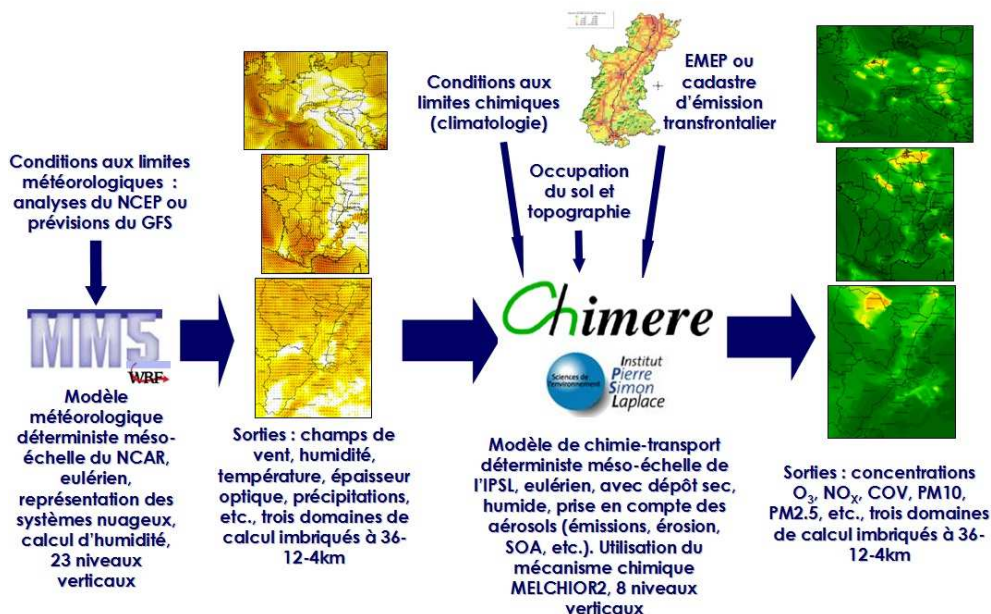


FIG. 2.5: Plateforme de modélisation de l'ASPA

Le krigeage ne va pas s'effectuer sur les données brutes mais sur les différences entre les mesures et le modèle CHIMERE. Ce type de krigeage est encore appelé «Krigeage des innovations». A des fins de représentation spatiale des sorties de modélisation une projection de type lambert conformal a été choisie pour définir la grille de calcul. Pour des raisons de facilité de calcul sous R, il a été choisi d'interpoler les sorties de modélisation sur une grille régulière en latitude/longitude. R disposant d'un package de conversion cartographique, il sera tout à fait possible d'utiliser un grille d'interpolation dans un quelconque système de projection.

C'est sur cette grille régulière que l'on va effectuer le krigeage des innovations. Puis on ajoute le modèle CHIMERE au krigeage alors effectué pour obtenir une carte de prévision ou de reporting corrigée par les données relevées en stations.

Les différentes étapes pour réaliser les cartes sont résumées sur le schéma ci-après.

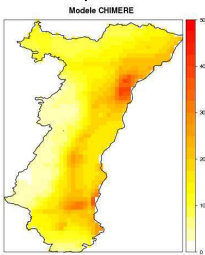
# Mise en œuvre du krigage

Données du modèle CHIMERE

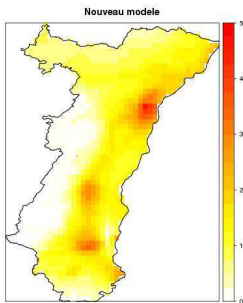
Interpolation sur une grille régulière

25	30	32
26	25	75
25	68	70

Représentation du modèle



Addition de la carte du modèle CHIMERE et des différences krigées



— Point aux stations

Mesures des différentes stations

	30	
		73

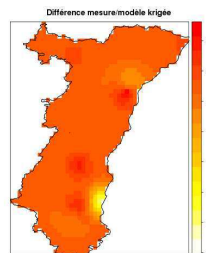
Différence mesure/modèle aux stations

	5	
		3

Krigage des différences mesure/modèle

5	7	2
4	6	4
10	7	3

Représentation des différences krigées



— Point krigé

Pour ce qui est des moyennes annuelles, l'étape géostatistique se base sur une carte moyennée des sorties du modèle ainsi que les moyennes annuelles des mesures. Le pré-traitement des données permet de travailler directement sur des données à une échéance annuelle.

L'étude du nombre de jours de dépassement en PM10 a été testée dans le cadre d'un projet de détermination de zones potentiellement exposées à plus de 35 jours de dépassement du seuil de  $50 \mu\text{g}/\text{m}^3$  sur une année. En première approche, le nombre de jours de dépassement avait été pré-traité en sortie de modélisation et une étude géostatistique a été faite directement sur le nombre de jours. Cependant, les résultats n'étaient pas satisfaisants. Le problème vient du fait que l'on krige une variable correspondant à un nombre de jours et cette dernière n'est pas additive et ne vérifie pas l'hypothèse d'autocorrélation indispensable pour appliquer le krigeage. Une autre approche consistant à post-traiter les données a donc été choisie pour contourner ce problème. La démarche est la suivante :

- Pour chaque jour de l'année, on effectue un krigeage sur nos données de concentrations journalières (maximum ou moyenne). On procède comme ci-dessus, c'est-à-dire avec un krigeage des innovations puis addition au modèle CHIMERE.
- On cherche les points de la grille où il y a dépassement. On note 1 dans la maille s'il y a dépassement et 0 sinon.
- En effectuant cette opération sur tous les jours, on obtient une grille de dépassement journalière. En sommant toutes ces grilles, on obtient une grille de dépassement annuel ; à chaque maille correspond le nombre de jours de dépassement à cet endroit.
- Enfin, on représente cette grille sur une carte afin de pouvoir l'analyser.

## Chapitre 3

# Présentation des résultats

L'objectif de ce chapitre est de présenter les résultats de krigeage obtenus à l'aide du logiciel R et d'en vérifier la validité ou tout au moins la cohérence. Cette vérification se base sur plusieurs critères :

- la cohérence physique avec les processus physico-chimiques
- la comparaison avec l'ancien modèle
- la validation croisée du krigeage effectué

### 3.1 Résultats pour le reporting

Pour chaque polluant, les points de mesures pris en compte dans l'étude sont représentés sur une carte de manière à visualiser la distribution spatiale des mesures. Dans l'objectif d'analyser le modèle obtenu, quatre cartes seront présentées :

- La carte obtenue par le logiciel R (après krigeage des différences et ajout du modèle CHIMERE), appelée «Nouveau Modèle».
- La carte représentant le modèle obtenu à l'aide du logiciel Isatis, appelée «Modèle Isatis». Aucun calcul n'a été effectué par le logiciel R, il a seulement été utilisé pour représenter les données obtenues par l'autre logiciel.
- La carte de différence «Nouveau Modèle» - «Modèle Isatis». Cette carte permettra de comparer nos résultats au modèle déjà existant.
- La carte du modèle CHIMERE.

Enfin, la carte des résidus et l'histogramme des erreurs standardisées obtenus lors de la validation croisée seront exposés afin de valider le modèle de krigeage utilisé.

Les données utilisées pour obtenir les différents résultats de krigeage correspondant aux données de l'année 2009.

Les scripts du logiciel R utilisés pour obtenir les cartes du reporting sont tous inspirés de celui présenté en annexe B.1.

#### 3.1.1 NO<sub>2</sub>

**Moyenne annuelle** A l'aide des différentes stations de mesure (cf. fig 3.1) et du modèle CHIMERE, nous allons représenter la moyenne annuelle 2009 pour le NO<sub>2</sub>.

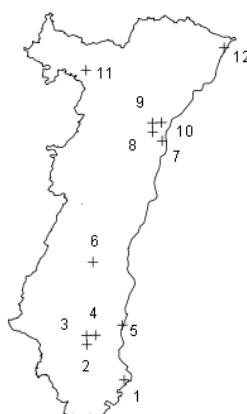


FIG. 3.1: Carte des stations de mesure pour le NO<sub>2</sub>

On commence par tracer le variogramme expérimental des différences Mesure/Modèle avant d'ajuster ce dernier avec un modèle sphérique. La figure 3.2 représente le variogramme ainsi obtenu. Nous avons pris comme paramètres du variogramme expérimental, une portée de 150km et un pas de 20km. Le choix de la portée et du pas se base plutôt sur la disposition du réseau que sur des processus physique. En effet, le NO<sub>2</sub> étant un polluant à petite échelle, une portée et un pas plus faibles seraient sans doute préférables pour modéliser la concentration de NO<sub>2</sub>. Cependant, les stations étant éloignées et d'un nombre restreint, il convient d'augmenter le pas et la portée pour avoir des résultats cohérents.

**Remarque :** On rappelle que notre système de conversion revient à dire que 1°=100km.

Variogramme avec portée= 1.5 et pas= 0.2 ajuste avec sphérique

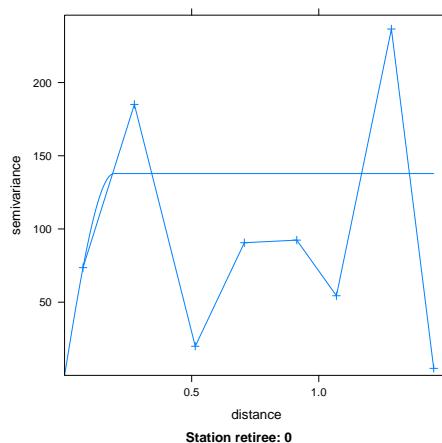


FIG. 3.2: Variogramme de la différence Mesure/Modèle de la moyenne annuelle de NO<sub>2</sub>

Grâce au variogramme modélisé, on effectue le krigeage sur nos différences puis on y ajoute le modèle

CHIMERE pour obtenir le «Nouveau modèle» (cf. figure 3.3). Les différentes cartes obtenues permettent de constater que le modèle réalisé avec le logiciel R est très proche de celui obtenu à l'aide du logiciel Isatis. On remarque en effet que la carte des différences «Nouveau modèle - modèle Isatis» ne prend ses valeurs qu'entre -3 et  $3 \mu\text{g}/\text{m}^3$ , ce qui montre la faible différence entre le nouveau modèle et l'ancien. Sur les cartes des deux modèles est ajoutée la valeur mesurée en chaque station afin de voir si notre modèle est en adéquation avec la réalité.

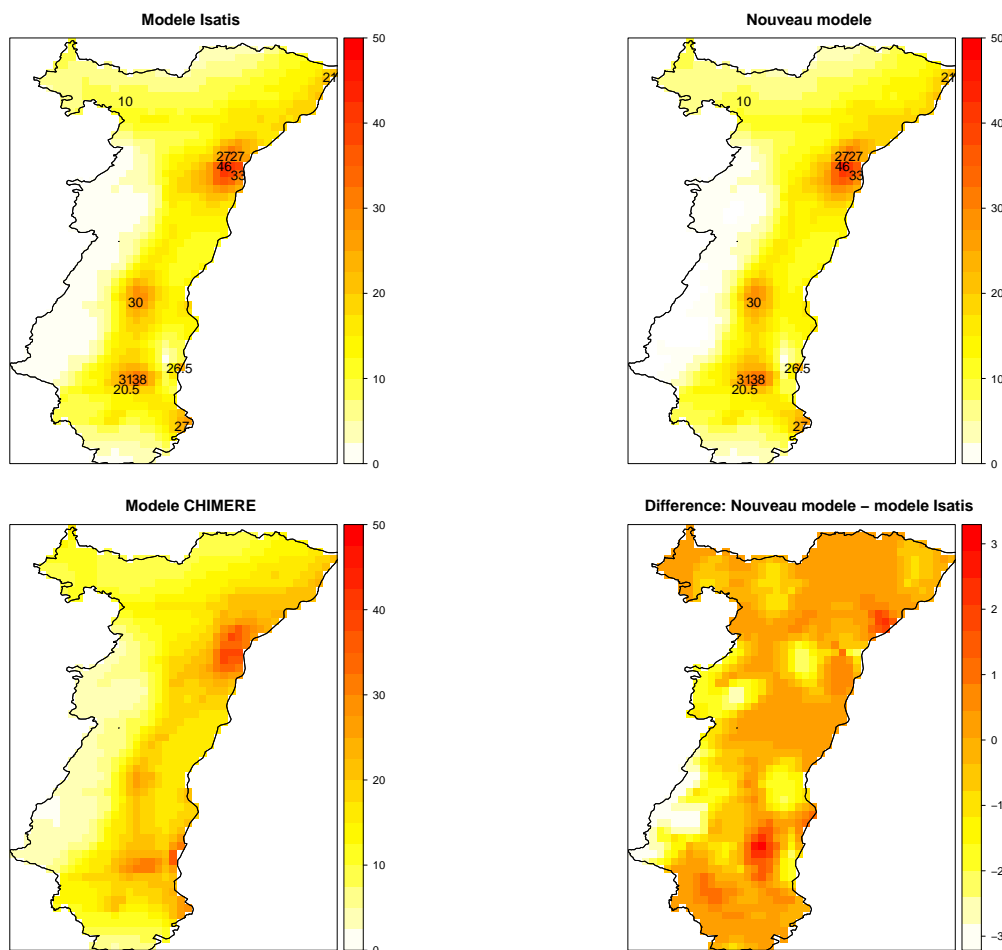


FIG. 3.3: Représentation des différents modèles de la moyenne annuelle de  $\text{NO}_2$

Le modèle obtenu est donc très proche de celui obtenu par Isatis. Il convient cependant de valider le krigeage effectué pour accepter définitivement le modèle. En effectuant la validation croisée sur le krigeage innovations (cf. figure 3.4), on peut noter que les erreurs standardisées sont majoritairement centrées en zéro. De plus, les résidus de la validation croisée sont relativement faibles mise à part sur le couple de station Calampe/Ottmarsheim pour laquelle le résidu est de  $-29,4 \mu\text{g}/\text{m}^3$ . La valeur mesurée en ce point étant de  $26,5 \mu\text{g}/\text{m}^3$ , on a une sur-évaluation du modèle de l'ordre de 100%. Cette sur-évaluation peut être expliquée par le gradient Ouest-Est sur les stations de Mulhouse (Sud, Centre et Nord). Ce gradient n'est pas représentatif d'une évolution à échelle régionale mais simplement de la zone dans laquelle elles sont implantées. Mais qui, faute de point de mesure complémentaire et de variable auxiliaire, a tendance à fortement influencer les estimations sur le secteur de Chalampe.

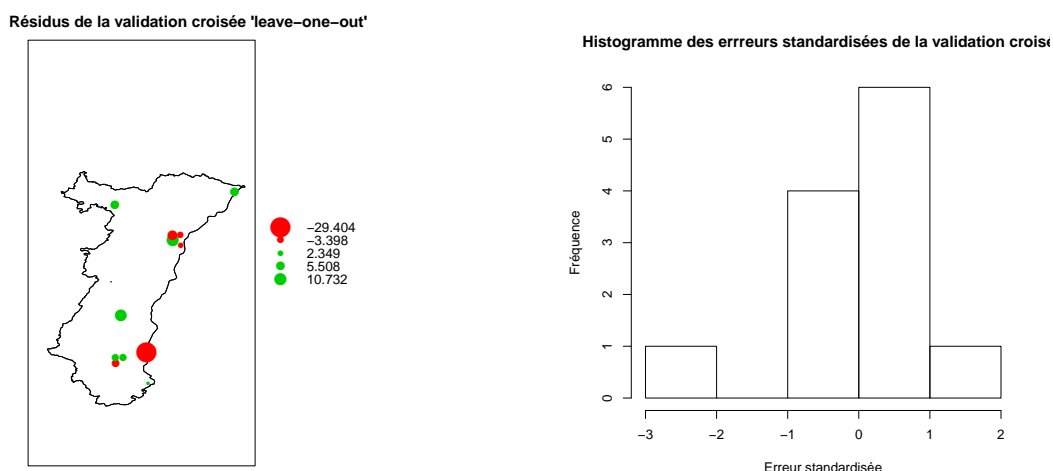


FIG. 3.4: Validation croisée du krigeage des innovations de la moyenne annuelle de NO<sub>2</sub>

**Percentile horaire 99,8** Pour représenter le percentile horaire, l'opération consiste à tracer le variogramme expérimental des différences Mesure/Modèle à l'aide des mesures effectuées en stations (c.f. fig 3.1) avant d'ajuster ce dernier avec un modèle sphérique. La figure 3.5 représente le variogramme ainsi obtenu.

Nous avons pris comme paramètres du variogramme expérimental, une portée de 150km et un pas de 20km. Le choix de ces paramètres se justifie de la même manière que dans le paragraphe précédent.

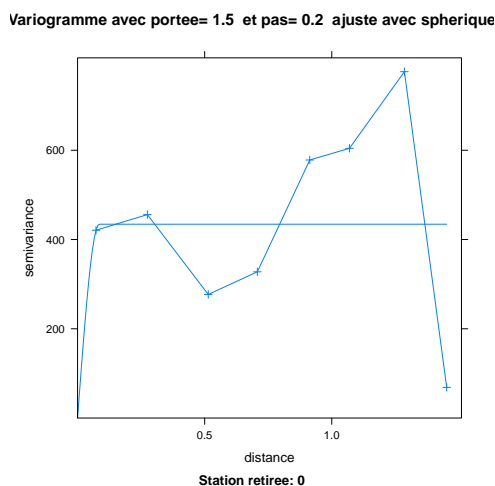


FIG. 3.5: Variogramme de la différence Mesure/Modèle du percentile horaire 99,8 de NO<sub>2</sub>

De la même manière que précédemment, on obtient les quatre cartes ci-après. On remarque à nouveau, grâce à la carte des différences, que le «nouveau modèle» est très proche de celui obtenu à l'aide du logiciel Isatis.



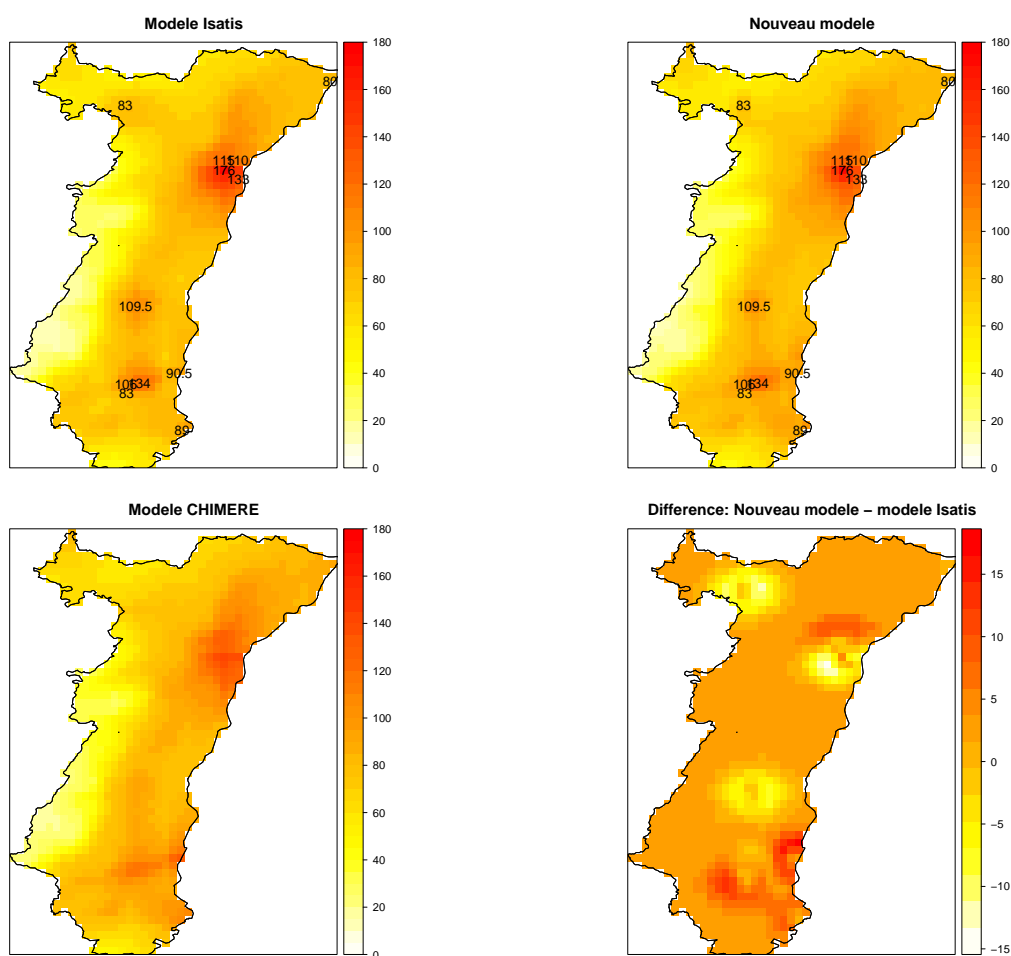


FIG. 3.6: Représentation des différents modèles du percentile horaire 99,8 de  $\text{NO}_2$

Afin de valider le krigeage, une validation croisée est effectuée sur le krigeage des innovations (cf. figure 3.7). Grâce à l'histogramme des erreurs standardisées, on remarque que ces dernières sont bien centrées autour de zéro. Comme pour la moyenne annuelle, on note un fort résidu ( $-47,3 \mu\text{g}/\text{m}^3$ ) pour le couple de station Chalampe/Ottmarsheim (5). La validation croisée met également en évidence un résidu élevé ( $40,94 \mu\text{g}/\text{m}^3$ ) pour la station 8 (Strasbourg Est) qui traduit une sous-estimation du modèle pour cette station. La valeur relevée à cette station est de  $176 \mu\text{g}/\text{m}^3$  alors que les trois autres stations de Strasbourg présentent un relevé de l'ordre de  $110 \mu\text{g}/\text{m}^3$ . Ce fort gradient entre les stations de Strasbourg n'étant pas reproductible à échelle régionale, il est normal d'obtenir une sous-estimation de la station Strasbourg Est lors de la validation croisée.

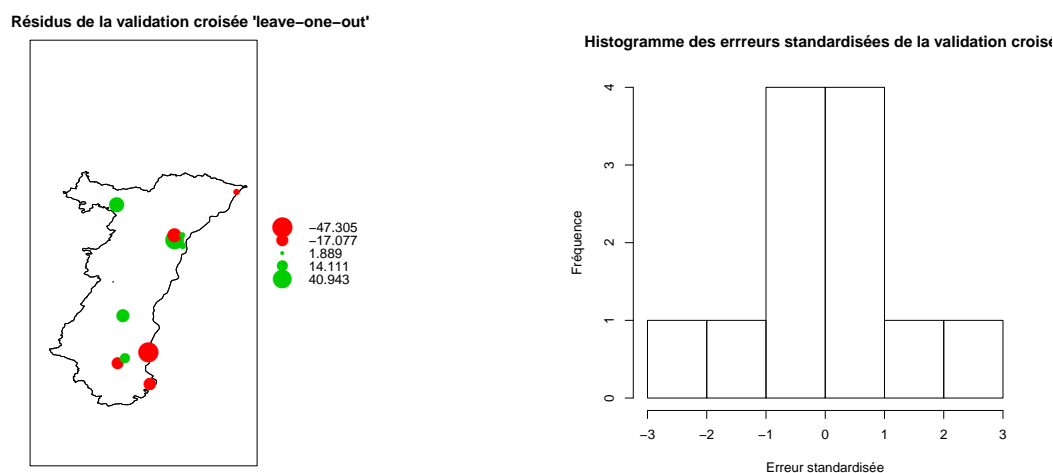


FIG. 3.7: Validation croisée du krigeage des innovations du percentile horaire 99,8 de  $\text{NO}_2$

### Conclusion pour le $\text{NO}_2$

En sortie de modélisation (cartes du modèle CHIMERE) l'impact des émissions des agglomérations et des axes routiers (A35) est diffus à cause de la taille des mailles de calculs et de la diffusion numérique du modèle eulérien alors utilisé. L'application du krigeage sur nos données permet ici de limiter cet effet de diffusion.

Sur les cartes de  $\text{NO}_2$  obtenues à l'aide du logiciel R, on retrouve bien des fortes valeurs sur les grosses agglomérations (Strasbourg, Colmar et Mulhouse). Le krigeage des innovations permet d'avoir un gradient zone urbaine/péri-urbaine plus fort. De même l'autoroute A35 ressort plus du fond et est moins «étalée» que pour le modèle CHIMERE, ce qui correspond d'avantage à la réalité. Enfin, les valeurs dans les Vosges sont aussi abaissées de manière cohérente avec les campagnes de mesures effectuées par l'ASPA (cf. Rapport ASPA-10060102-ID) et les processus physico-chimiques (temps de résilience,...).

La limite de ce modèle intervient pour les stations Mulhouse Est et Strasbourg Est qui présentent des résidus de validation croisée élevés. Le modèle ne permet pas de différencier une évolution à petite échelle d'une évolution à échelle régionale et peut donc produire des erreurs d'estimation comme observées ci-dessus.

### 3.1.2 PM10

Les stations utilisées pour le krigeage des PM10 sont les suivantes :

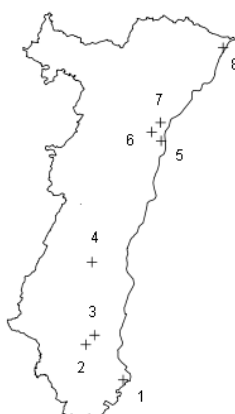


FIG. 3.8: Carte des stations de mesure pour les PM10

**Moyenne annuelle** A l'aide des différentes stations de mesure (cf. figure 3.8) et du modèle CHIMERE, nous allons représenter la moyenne annuelle 2009 pour les PM10.

On commence par tracer le variogramme expérimental des différences Mesure/Modèle puis on ajuste ce dernier avec un modèle sphérique. La figure 3.9 représente le variogramme ainsi obtenu. Nous avons pris comme paramètres de variogramme expérimental une portée de 120km et un pas de 25km. Ce choix se justifie, comme pour le  $\text{NO}_2$ , par la répartition du réseau de stations qui implique de prendre un pas assez grand.

Variogramme avec portée= 1.2 et pas= 0.25 ajuste avec sphérique

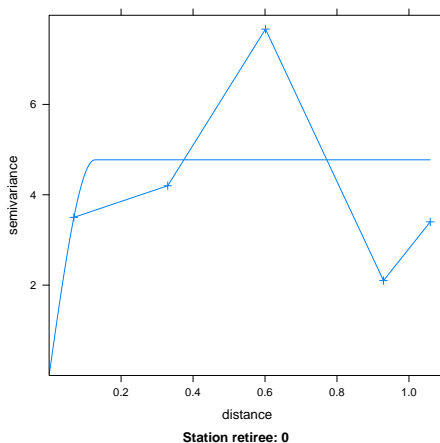


FIG. 3.9: Variogramme de la différence Mesure/Modèle de la moyenne annuelle de PM10

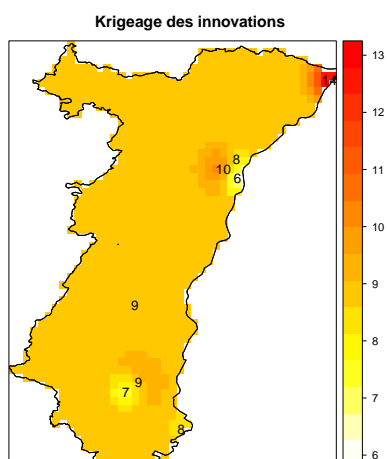


FIG. 3.10: Krigage des innovations de la moyenne annuelle de PM10

La figure 3.10 représente le krigage effectué sur les différences Mesure/Modèle. On peut observer, grâce à cette carte, que le krigage change peu la répartition spatiale du modèle CHIMERE. Il corrige seulement un biais du modèle de l'ordre de  $9 \mu\text{g}/\text{m}^3$ . Seules les valeurs en proximité directe des points de mesures peuvent être plus ou moins corrigées mais toujours avec des valeurs de correction dans un intervalle de 7 à  $14 \mu\text{g}/\text{m}^3$  (proche du biais moyen).

Ce biais quasi constant peut être expliqué par une valeur de pas trop importante aux vues des phénomènes physico-chimiques qui régissent les imissions de PM10. La marge de manoeuvre sur les valeurs du pas de krigage est limitée par la disposition du réseau de mesure. Une solution envisageable serait la mise en place d'une dérive externe.

Cependant, au vu de la validation croisée (cf. figure 3.12), rien ne permet de dire que les résultats ne sont pas acceptables.

De plus, les cartes 3.11 montre que le modèle que l'on obtient est très proche de celui obtenu à l'aide du logiciel Isatis.

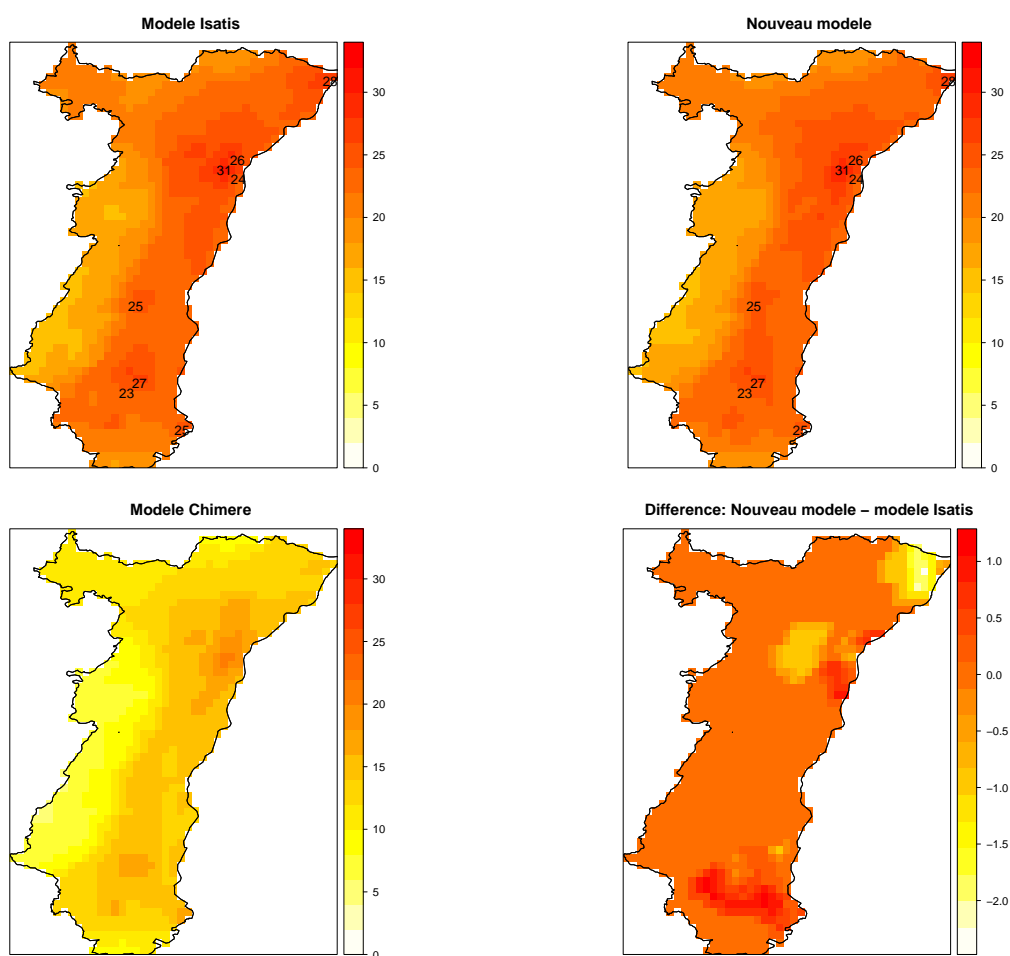


FIG. 3.11: Représentation des différents modèles de la moyenne annuelle de PM10

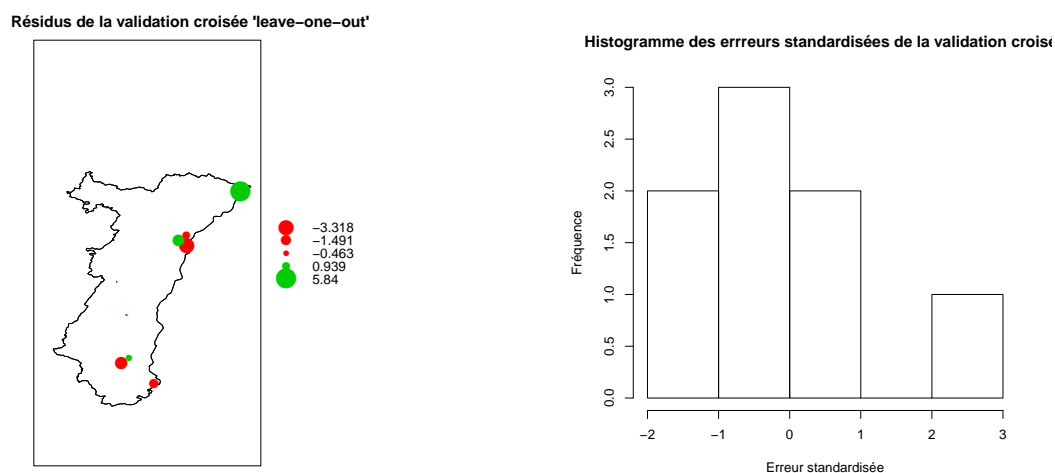


FIG. 3.12: Validation croisée du krigeage des innovations de la moyenne annuelle de PM10

**Percentile journalier 90,4** Pour représenter le percentile journalier, on commence par tracer le variogramme expérimental des différences Mesure/Modèle à l'aide des mesures effectuées en station (c.f. fig 3.8) avant d'ajuster ce dernier avec un modèle sphérique. La figure 3.13 représente le variogramme ainsi obtenu.

Nous avons pris comme paramètres du variogramme expérimental, une portée de 150km et un pas de 25km.

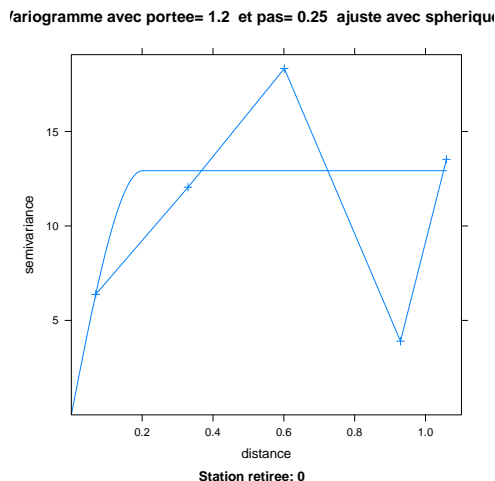


FIG. 3.13: Variogramme de la différence Mesure/Modèle du percentile journalier 90,4 des PM10

Les quatre cartes ci-après sont obtenues en appliquant le même processus que précédemment. On remarque que le modèle obtenu par le logiciel R est très proche du modèle obtenu par Isatis. La carte des différences montre en effet que celles-ci se situent entre 1 et  $-1 \mu\text{g}/\text{m}^3$ .

Il est à noter que, pour des raisons de visibilité des cartes, l'échelle des deux modèles corrigés est différente de celle du modèle CHIMERE. La répartition spatiale des données est la même pour les trois cartes mais le modèle CHIMERE prend des valeurs plus faibles que nos modèles corrigés.

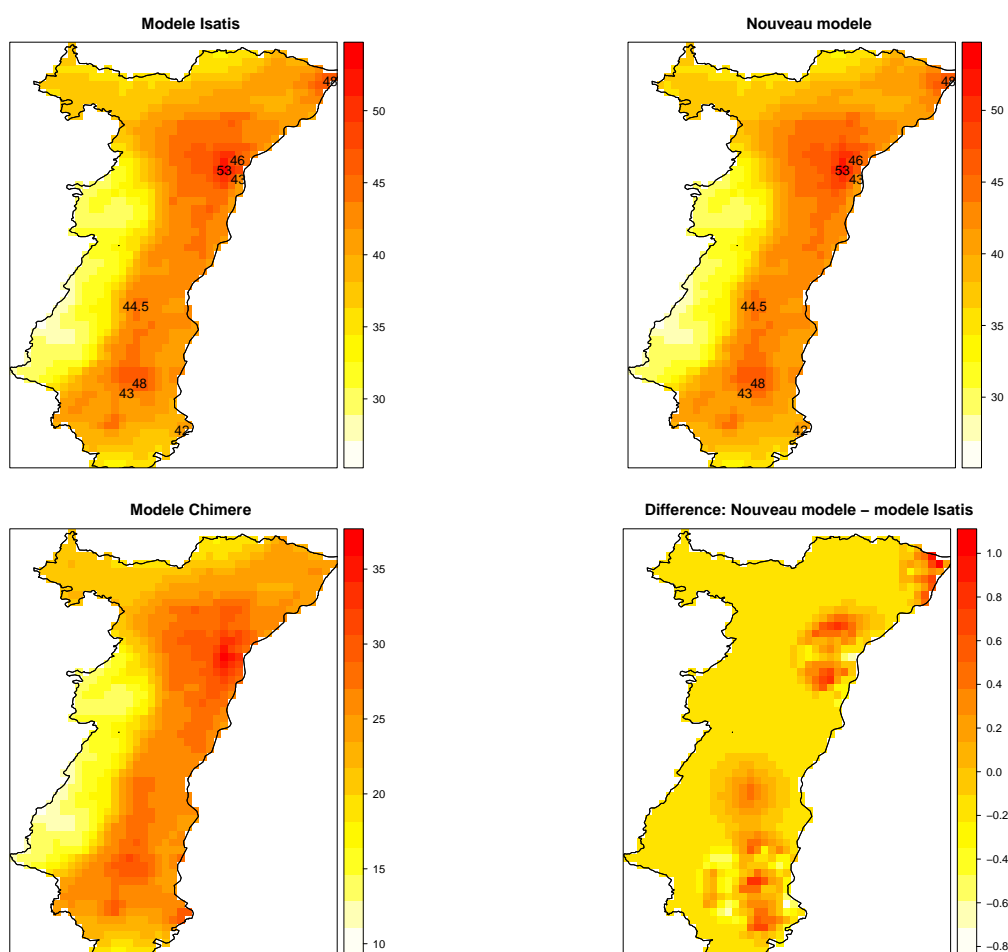


FIG. 3.14: Représentation des différents modèles du percentile journalier 90,4 des PM10

Le modèle ainsi créé est donc très proche de celui obtenu par Isatis. On remarque cependant le même phénomène que pour la moyenne annuelle de PM10 sur le krigeage des innovations. En effectuant la validation croisée sur le krigeage effectué, on peut noter que les erreurs standardisées sont plutôt centrées en zéro. Les résidus élevés des stations 1 (3 frontières) et 8 (Nord-Est Alsace) s'expliquent par le fait que ces stations sont isolées et tendent donc à produire des erreurs d'estimation élevées (cf. paragraphe 2.1.3).

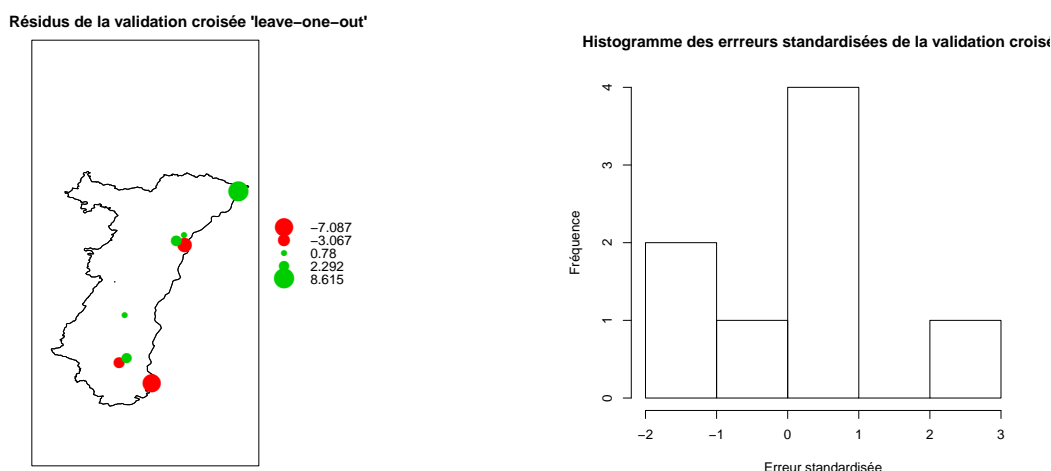


FIG. 3.15: Validation croisée du krigeage des innovations du percentile journalier 90,4 des PM10

**Nombre de jours de dépassement en PM10** Nous allons représenter le nombre de jours où la concentration des PM10 dépasse le seuil de  $50 \mu\text{g}/\text{m}^3$  sur l'année 2009.

Le script R utilisé pour obtenir les cartes ci-après est différent de celui utilisé pour les autres résultats du reporting. Il est présenté en annexe B.2

En représentant le nombre de jours de dépassement uniquement sur le mois de janvier avec les stations mesurant les PM10 (c.f. fig 3.16), on a pu observer que l'absence de stations dans les Vosges ne permet pas d'obtenir des cartes de concentrations en accord avec les valeurs estimées à l'aide de différentes campagnes de mesures. Le krigeage des innovations fait ressortir certains dépassements du seuil de  $50 \mu\text{g}/\text{m}^3$  en désaccord avec les concentrations mesurées à l'aide du labo-mobile qui indiquent des valeurs de l'ordre de  $12 \mu\text{g}/\text{m}^3$  en moyenne journalière. C'est pour cette raison que nous avons décidé de rajouter deux stations «virtuelles»; une dans les Vosges moyennes et une dans les Vosges du Sud.

Pour la station des Vosges du Sud nous nous sommes basés sur la campagne Grand Ballon/Ballon D'Alsace menée en 2009/2010 (Rapport ASPA-10060102-ID). Cette campagne a permis de recueillir des relevés de concentration en PM10 et d'ainsi observer comment ce polluant se comporte à cet endroit. Après l'étude des données de cette campagne, nous avons décidé de prendre comme mesure de PM10 la moyenne de la campagne c'est-à-dire  $13 \mu\text{g}/\text{m}^3$ . La moyenne de cette campagne se situe très nettement en-dessous du seuil de dépassement, ce qui nous conforte dans l'idée que la première carte obtenue ne concordait pas avec la réalité. Pour la station des Vosges moyennes, nous avons utilisé des données récoltées sur une période de deux mois sur cette station et nous avons décidé de prendre une valeur de  $12 \mu\text{g}/\text{m}^3$ .

Le carte ci-dessous résume les stations prises en compte pour le krigeage.



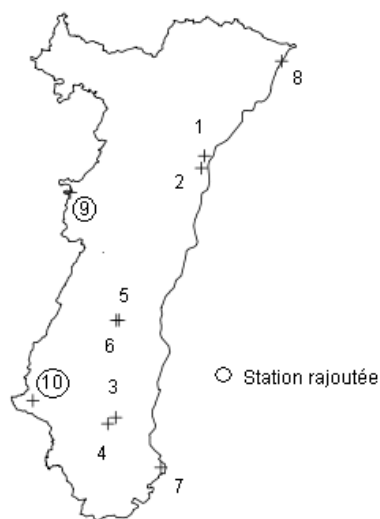


FIG. 3.16: Stations utilisées pour représenter le nombre de jours de dépassement de PM10

A l'aide du logiciel R, on applique le modèle de krigeage des différences Mesure/Modèle pour tous les jours de l'année. Pour chaque jour, le même modèle est donc utilisé et il existe des jours pour lesquels le modèle ne s'ajuste pas. Le nombre de ces jours étant relativement faible et ne présentant pas de dépassement de PM10, ils ont été retirés de l'historique de données. Ces jours sont les suivants :

- 27/06/2009
- 08/08/2009
- 12/08/2009
- 25/10/2009

De plus, pour les dates du 29 et 30 août 2009 le fichier CHIMERE est endommagé. Le processus de modélisation régionale ayant confirmé qu'il y a eu des problèmes de calculs pour ces dates, nous les avons également retirées de nos données.

La figure 3.17 représente à gauche le modèle testé par l'ASPA et à droite le nouveau modèle. On peut noter que l'ancien modèle présente des dépassements de PM10 dans les Vosges ce qui n'est pas concordant avec la réalité à l'inverse du nouveau modèle qui reste à zéro dans les Vosges.

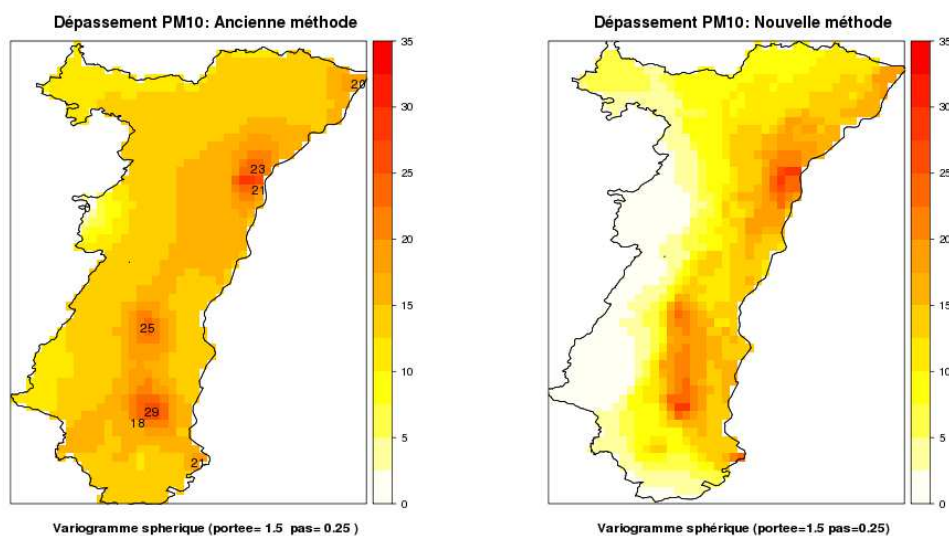


FIG. 3.17: Nombre de jours de dépassement sur l'année 2009 en PM10

Rappelons que la représentation du nombre de jours de dépassement en PM10 est nécessitée par un projet de détermination de zones potentiellement exposées à plus de 35 jours de dépassement du seuil sur une année. Dans le cadre de ce projet, les valeurs de krigeage doivent être assorties d'un intervalle de confiance défini de façon à ce qu'il y ait 40% de chances que la vraie valeur soit dans cet intervalle. Il est construit de la façon suivante :

$$[K_i - 0.5244\sigma_{K_i}; K_i + 0.5244\sigma_{K_i}]$$

où

- $K_i$  est la valeur krigée du point  $i$  de la grille.
- $\sigma_{K_i}$  est l'écart-type du krigeage en ce point.

**Remarque :** Les coefficients de cet intervalle de confiance sont basés sur une hypothèse (très optimiste) de distribution gaussienne des résidus.

Cet intervalle de confiance permet de prendre en compte l'incertitude du krigeage. On représentera donc la carte correspondant au modèle précédent auquel on ajoute  $0.5244\sigma$ . Cette carte permet alors d'identifier les zones en dehors desquelles le risque de dépassement est considéré comme nul.

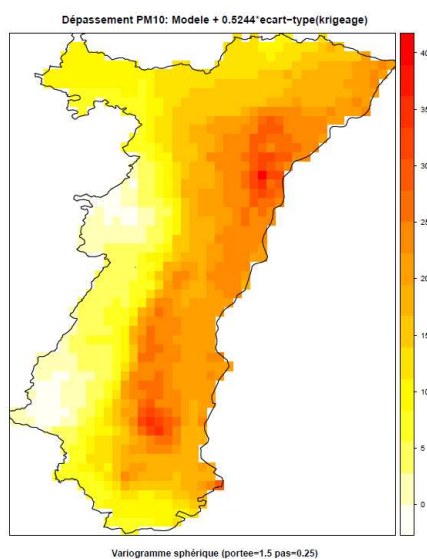


FIG. 3.18: Carte des zones potentiellement exposées à plus de 35 jours de dépassement du seuil de  $50 \mu\text{g}/\text{m}^3$  en PM10

### Conclusion pour les PM10

Pour la moyenne annuelle et le percentile 90,4 des PM10, nous avons réussi à trouver un modèle de krigeage répondant aux attentes, tant au niveau de la cohérence physique que de la similitude avec le modèle obtenu par le logiciel Isatis.

En ce qui concerne le nombre de jours de dépassement, la nouvelle méthode expérimentée semble plus concluante que la méthode précédemment testée par l'ASPA.

### 3.1.3 Ozone

Les stations utilisées pour le krigeage des AOT de l'ozone sont les suivantes :



FIG. 3.19: Carte des stations de mesure pour les AOT d'ozone

**AOTF** Pour représenter l'AOTF (forêt) d' $O_3$ , on commence par tracer le variogramme expérimental des différences Mesure/Modèle à l'aide des mesures effectuées en stations (cf. figure 3.19) avant d'ajuster ce dernier avec un modèle sphérique. La figure 3.20 représente le variogramme ainsi obtenu. Les paramètres du variogramme expérimental sont : une portée de 150km et un pas de 24km.

Variogramme avec portee= 1.5 et pas= 0.24 ajuste avec spherique

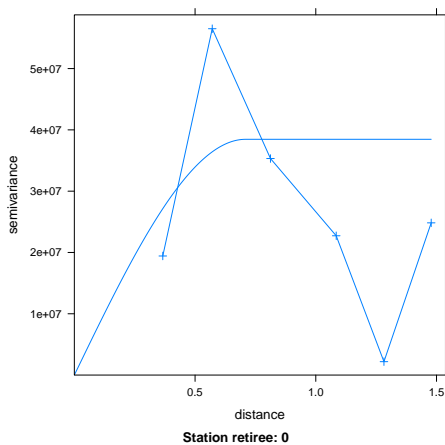


FIG. 3.20: Variogramme de la différence Mesure/Modèle de l'AOTF

Toujours sur le même processus que pour le  $NO_2$  et les particules, on obtient les quatre cartes ci-après. Les différences entre le nouveau modèle et l'ancien sont de manière générale inférieures à 15%. La répartition géographique est sensiblement identique en sortie des deux modèles. Le nouveau modèle a tendance à donner des valeurs plus faibles sur la partie Nord des Vosges et à la partie Nord-Est de l'Alsace. Sur ce point, la carte krigée grâce à l'ancien modèle semble plus correspondre aux attentes en terme d'expertise humaine.

On notera aussi que les cartes krigées réduisent l'impact des axes routiers et des zones urbaines en augmentant toutes les valeurs en plaine de manière homogène.

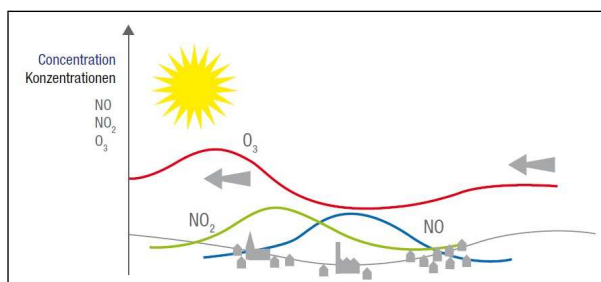


FIG. 3.21: Evolution des niveaux d'ozone et de  $\text{NO}_x$  aux abords d'une agglomération

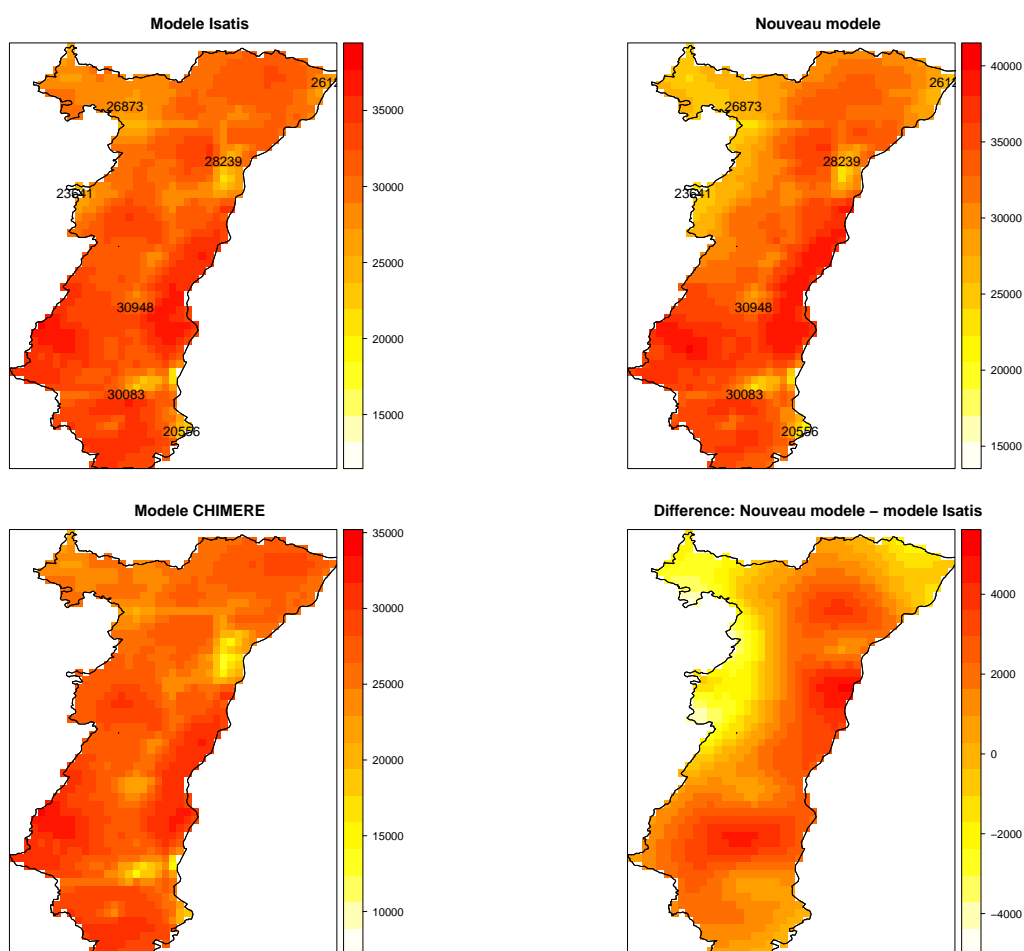


FIG. 3.22: Représentation des différents modèles de l'AOTF

Sur les trois cartes obtenues à l'aide des différents modèles, on remarque que les concentrations les plus faibles en ozone se trouvent au niveau des grandes agglomérations et de l'autoroute A35, alors

que les Vosges présentent des concentrations plus élevées. Ce phénomène s'explique par le processus de formation et de destruction de l'ozone vu dans le chapitre précédent (cf. 2.4). Ce dernier est représenté schématiquement sur la figure 3.21. L'ozone est détruit par les  $\text{NO}_x$  (émis par les agglomérations et axes routiers) par un processus rapide et est également formé par ces derniers en zones rurales par un processus plus long.

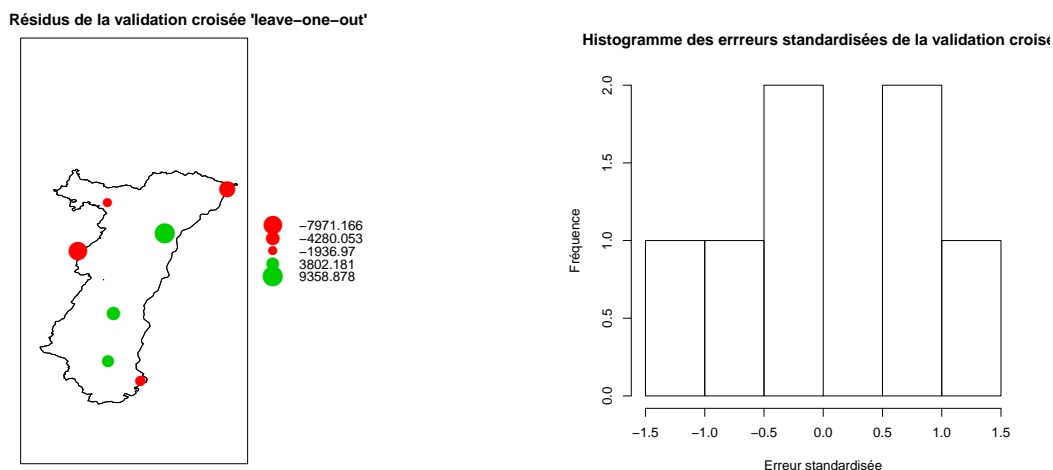


FIG. 3.23: Validation croisée du krigeage des innovations de l'AOTF

### Conclusion pour l'ozone

Contrairement aux deux précédents polluants, l'ozone semble plus difficilement modélisable par le krigeage. On a cependant réussi à trouver un modèle représentant correctement la spatialisation de ce polluant (au regard des connaissances de ce polluant) mais qui sous-estime les concentrations par rapport au modèle obtenu par Isatis.

#### 3.1.4 Conclusion

Pour chaque polluant ( $\text{NO}_2$ ,  $\text{PM}_{10}$  et ozone), il a été possible de trouver un modèle de krigeage représentant les différentes quantités demandées par le reporting avec le logiciel R. Les cartes ainsi obtenues se rapprochent fortement de celles réalisées à l'aide du logiciel Isatis. De plus, les validations croisées effectuées sur les krigeages des innovations renvoient des résidus faibles pour la majorité des stations, ce qui conforte les choix effectués.

## 3.2 Résultats pour la prévision

Deux fois par jours (à 10h et 16h), des cartes de la moyenne journalière de  $\text{PM}_{10}$  et des maximums journaliers d'ozone ( $\text{O}_3$ ) et de dioxyde de carbone ( $\text{NO}_2$ ) sont mises en ligne sur le site de l'ASPA (<http://www.atmo-alsace.net/site/Accueil-2.html>).

Ces cartes issues de la plateforme de modélisation sont corrigées par les mesures. Cette correction consiste en une correction statistique, qui prend en compte les dernières mesures disponibles au niveau des stations. Le ratio  $\text{ModèleCorrigé}/\text{ModèleBrut}$  est calculé au niveau de ces stations et spatialisé de manière à produire les cartes corrigées. La spatialisation consiste actuellement en une interpolation bilinéaire. Le but de cette partie est d'explorer la possibilité de remplacer cette spatialisation par un krigeage des innovations.

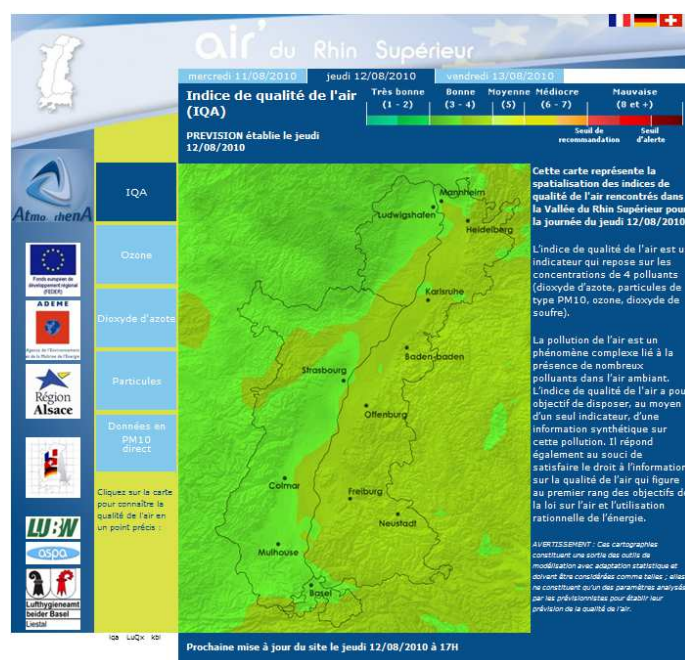


FIG. 3.24: Carte de prévision du site de l’ASPA

Les scripts R utilisés pour obtenir les différentes cartes de prévision sont tous inspiré de celui présenté en annexe B.3.

### 3.2.1 Maximum journalier de NO<sub>2</sub>

Pour représenter le maximum journalier de NO<sub>2</sub>, on utilise la méthode de krigeage présentée dans le paragraphe 2.3 pour chaque jour de l’année. Le but étant de trouver un modèle que l’on puisse automatiser chaque jour, le même modèle devra être utilisé tous les jours. Nous présentons ci-dessous les résultats du modèle appliqué sur un jour en particulier, le 24 janvier 2009.

On remarque que le modèle après krigeage relève le modèle CHIMERE et devient alors plus proche de la réalité (cf. 3.26).

En appliquant ce modèle sur d’autres jours de l’année, on tombe sur des cartes satisfaisantes mais aussi sur des cartes moins cohérentes. Il y a en effet des jours où notre modèle est singulier, c’est-à-dire que le variogramme est «mal» modélisé avec le modèle que l’on applique. Il a donc été décidé de garder un modèle convenant au plus grand nombre de jours mais il est malgré tout inévitable de rencontrer des jours pour lesquels le modèle ne conviendra pas.

Les paramètres retenus pour le variogramme expérimental sont donc : une portée de 150 km et un pas de 25 km.

Variogramme avec portee= 1.5 et pas= 0.25 ajuste avec spheriqu

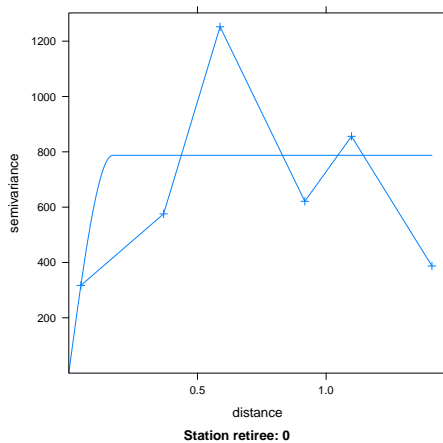


FIG. 3.25: Variogramme de la différence Mesure/Modèle du maximum journalier de NO<sub>2</sub> du 24-01-2009

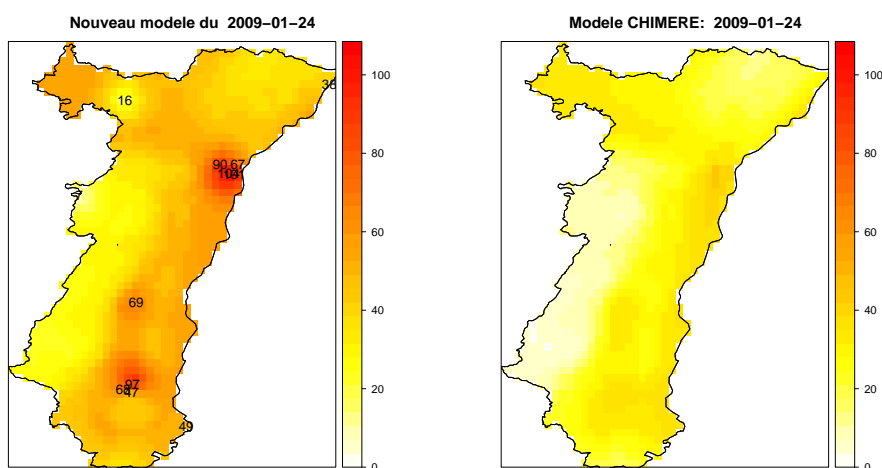


FIG. 3.26: Représentation des différents modèles du maximum journalier de NO<sub>2</sub> du 24-01-2009



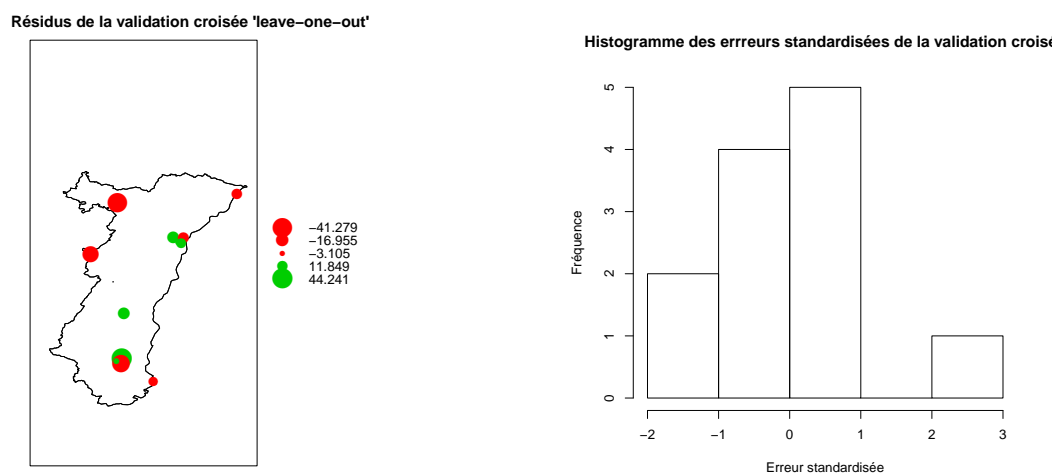


FIG. 3.27: Validation croisée du krigeage des innovations du maximum journalier de  $\text{NO}_2$  du 24-01-2009

Comme énoncé précédemment, les limites d'automatisation de la procédure avec R, dans le but d'obtenir les cartes de prévision, sont que le modèle peut être singulier pour certains jours. Nous présentons ci-dessous, la carte de maximum journalier de  $\text{NO}_2$  du 10 avril 2009, date pour laquelle le modèle est dit singulier.

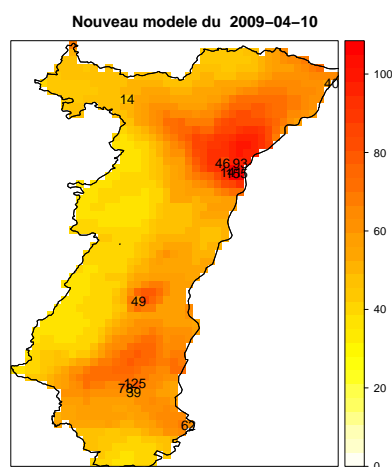


FIG. 3.28: Représentation du maximum journalier de  $\text{NO}_2$  du 10-04-2009 (Modèle singulier)

Cette carte n'est pas satisfaisante car les valeurs estimées en stations et les mesures relevées en stations sont très différentes. On peut néanmoins remarquer que ce jour de l'année présente de grandes différences de mesure au sein d'une même agglomération. La station Strasbourg Est mesure une concentration de  $141 \mu\text{g}/\text{m}^3$  au maximum de la journée alors que celle de Strasbourg Nord ne mesure qu'une concentration de  $46 \mu\text{g}/\text{m}^3$ . Ce fort écart de mesure peut expliquer l'erreur de modélisation car c'est un phénomène à petite échelle qui n'est pas pris en compte dans le krigeage.

Le nombre de jours ayant un modèle singulier étant relativement faible, une solution serait, pour ces jours posant problème, d'effectuer une moyenne des stations d'une même agglomération (Strasbourg, Colmar et Mulhouse), en se basant sur les zones des indices de qualité de l'air diffusées par l'ASPA, afin d'obtenir

une seule mesure par zone et effacer les disparités mises en évidence précédemment. Cela se rapprocherait de la méthode actuellement utilisée par l'ASPA.

### 3.2.2 Moyenne journalière en PM10

Pour représenter la moyenne journalière en PM10, le modèle retenu est celui qui consiste à tracer le variogramme expérimental avec les paramètres suivants : une portée de 150 km et un pas de 25 km. Les résultats présentés ci-dessous correspondent à la représentation de la moyenne journalière en PM10 pour le 10 avril 2009.

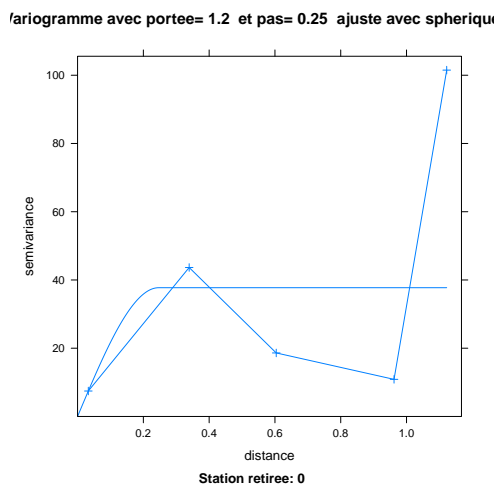


FIG. 3.29: Variogramme de la différence Mesure/Modèle du maximum journalier de PM10 du 10-04-2009

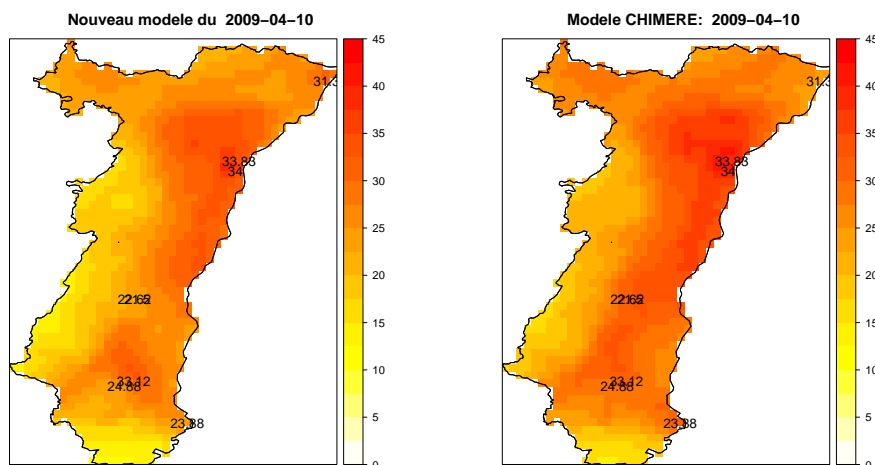


FIG. 3.30: Représentation des différents modèles du maximum journalier de PM10 du 10-04-2009

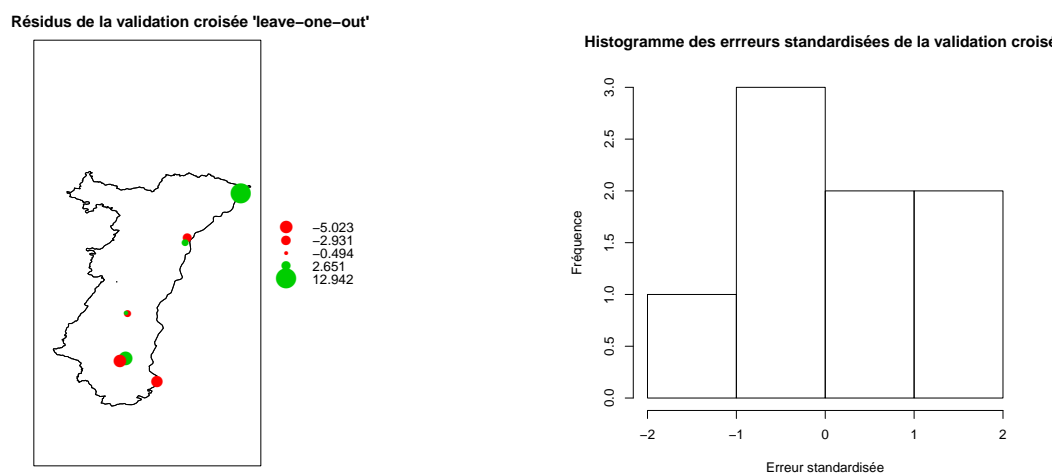


FIG. 3.31: Validation croisée du krigeage des innovations du maximum journalier de PM10 du 10-04-2009

Pour les mêmes raisons énoncées dans le paragraphe précédent, il existe certains jours pour lesquels le variogramme expérimental se modélise «mal». Nous présentons ci-dessous, la carte de la moyenne journalière de PM10 du 15 janvier 2009, date pour laquelle le modèle est dit singulier.

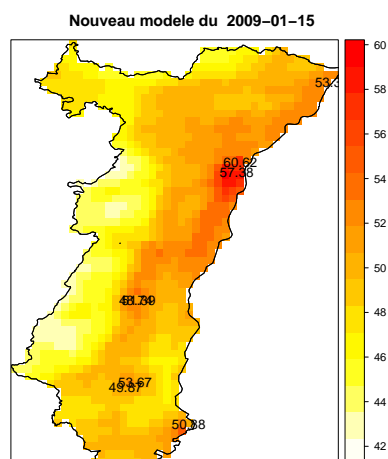


FIG. 3.32: Représentation de la moyenne journalière de PM10 du 15-01-2009 (Modèle singulier)

On peut noter que malgré le fait que le modèle soit singulier, la carte obtenue semble plutôt cohérente. Elle relève la carte du modèle CHIMERE (carte qui ne sera pas présentée ici) et semble cohérente par rapport aux mesures effectuées en stations.

### 3.2.3 Maximum journalier d'ozone

Il a été plus difficile de trouver un modèle cohérent pour l'ozone car ce dernier est difficile à modéliser à l'aide du krigeage. Il a cependant été décidé de garder le modèle qui consiste à tracer le variogramme expérimental avec les paramètres suivants : une portée de 150 km et un pas de 25 km. Nous présentons dans cette partie, le résultat de la modélisation pour le maximum journalier en ozone

pour la date du 11 avril 2009. Le variogramme expérimental des différences Mesure/Modèle est présenté sur la figure 3.33.

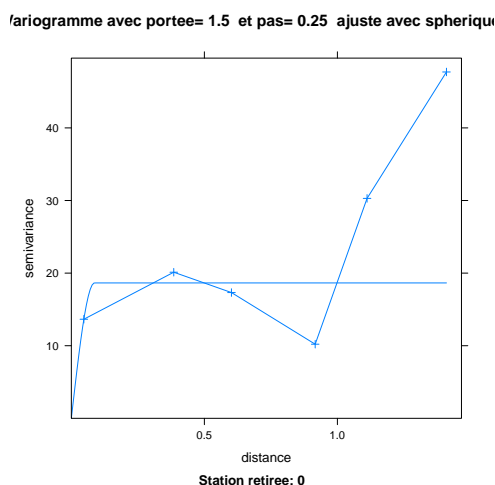


FIG. 3.33: Variogramme de la différence Mesure/Modèle du maximum journalier de  $O_3$  du 11-04-2009

Les cartes présentées ci-après représentent le maximum journalier du 24 avril 2009 pour le modèle CHIMERE et pour le modèle obtenu après krigeage. Il faut cependant analyser ces deux cartes avec précaution car pour une meilleure visualisation, les échelles des deux cartes ne sont pas les mêmes. Le modèle CHIMERE sous-estime très largement la concentration en ozone ; les valeurs de l'échelle varient entre 48 et 62  $\mu g/m^3$  alors que les valeurs mesurées en stations sont de l'ordre de 130  $\mu g/m^3$ . Le fait d'intégrer les mesures en stations pour obtenir le modèle permet donc de corriger la carte du modèle CHIMERE.

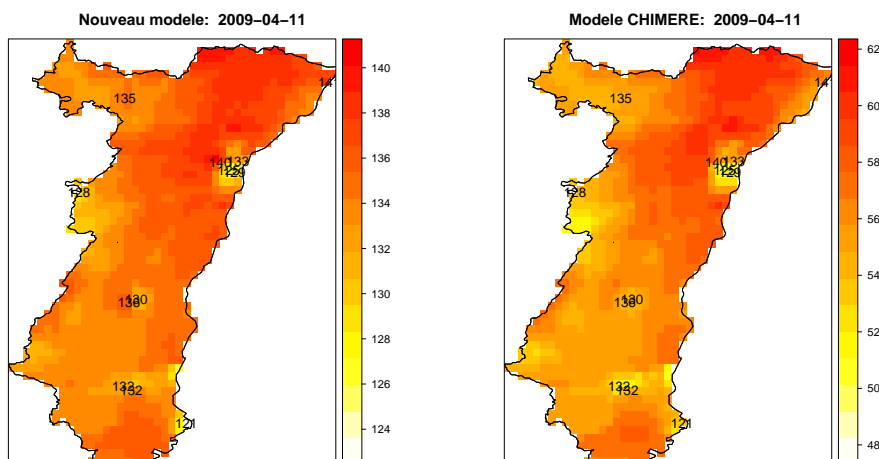


FIG. 3.34: Représentation des différents modèles du maximum journalier de  $O_3$  du 11-04-2009

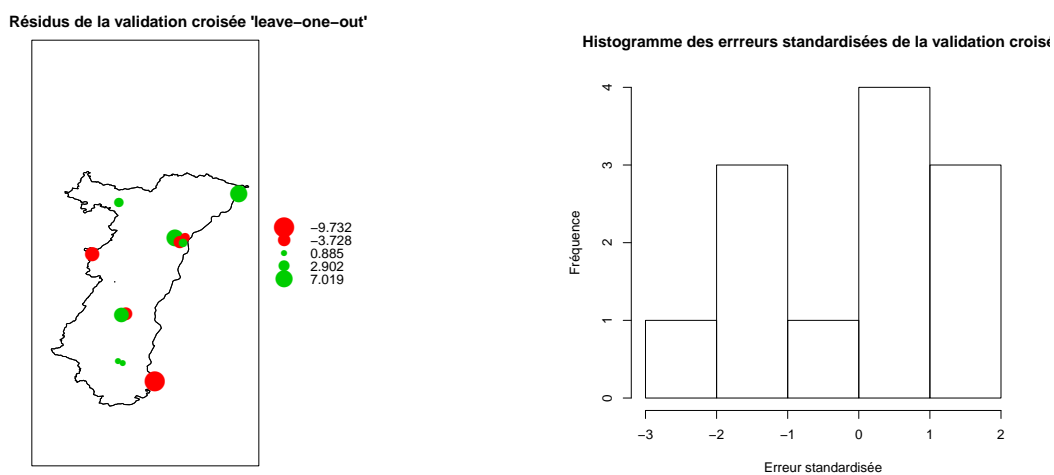


FIG. 3.35: Validation croisée du krigeage des innovations du maximum journalier de  $O_3$  du 11-04-2009

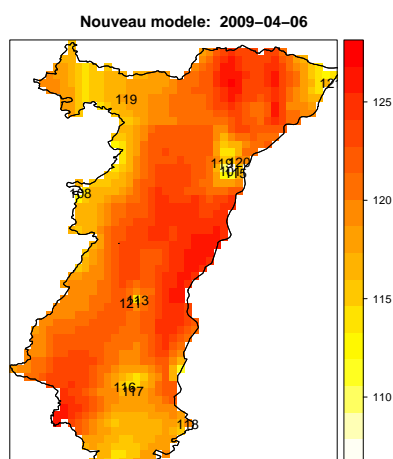


FIG. 3.36: Représentation du maximum journalier d'ozone du 06-04-2009 (Modèle singulier)

# Conclusion

Les deux missions proposées lors de ce stage ont été remplies. Le krigeage a réussi à être mis en oeuvre tant pour les cartes du reporting que pour les cartes de prévision.

Concernant le reporting, le logiciel R permet d'obtenir des cartes très ressemblantes à celles obtenues avec le logiciel Isatis actuellement utilisé par l'ASPA. Les faibles différences entre les deux logiciels semblent donc négligeables par rapport au gain de temps apporté par le logiciel R.

Pour la prévision, les conclusions du travail effectué se feront sur un plus long terme. D'un point de vue technique, la méthode de spatialisation de la prévision par le krigeage est très facilement automatisable. Il suffira d'intégrer la gestion des modèles singuliers grâce à la manipulation expliquée dans le paragraphe 3.2.1. Avant de remplacer la méthode actuelle de spatialisation par celle du krigeage, la mise en place en doublon est nécessaire de manière à appréhender les cas particuliers qui peuvent apparaître en prévision et qui doivent être gérés de manière automatique.

Il est à noter que l'aspect le plus limitant en ce qui concerne la spatialisation n'est pas du à un problème de méthodologie mais surtout à la disposition du réseau de mesure. La grande majorité des mesures couvrent la plaine et sont par conséquent quasiment toutes sur un même axe Nord/Sud. Dans cette configuration plus on s'éloigne de cet axe plus les estimations sont incertaines. De même les stations de type urbaines et périurbaines sont prédominantes ce qui tend à fausser les estimations en zone rurale.

# Annexes

## Annexe A

# Le Krigeage avec R

Voici un résumé des principales fonctions de R permettant d'effectuer une analyse variographique, un krigeage et une validation croisée. Le package de R utilisé est : **gstat**.

### A.1 Analyse variographique avec R

Dans tout ce que nous faisons avec R, on considère que nous avons des données géographiques. Cependant, lorsque l'on fournit un tableau au logiciel, il faut lui préciser quelles colonnes correspondent aux coordonnées de nos points. Pour cela, on utilise la fonction : **coordinates(...)**.

*Exemple :*

```
> Mesure
  moy    lat    lon i j zone
1 33.83333 48.6105 7.78163 53 65 1
2 34.00000 48.5704 7.76645 53 63 1
3 33.12500 47.7576 7.34725 45 42 2

> coordinates(Mesure)<-c("lon","lat")
# ou coordinates(Mesure)<- ~lon+lat

> Mesure
      coordinates      moy i j zone
1 (7.78163, 48.6105) 33.83333 53 65 1
2 (7.76645, 48.5704) 34.00000 53 63 1
3 (7.34725, 47.7576) 33.12500 45 42 2
```

#### A.1.1 Le variogramme expérimental

Cette étape consiste à calculer et tracer le variogramme estimé à l'aide de nos observations.

##### Calcul du variogramme expérimental

La fonction de R permettant de calculer le variogramme expérimental est la suivante :

```
> variogram(formule, data...)
```

Les arguments de cette fonction sont les suivants :



- **formule** : `variable_réponse ~ régresseurs`.
  - `variable_réponse` = variable que l'on veut étudier.
  - `régresseurs` = variables explicatives, comme par exemple les émissions, les reliefs... (s'il n'y en a pas, on met simplement `... ~ 1`).
- **data** : Objet de type «dataframe» où se trouve nos variables utilisées.
- **cutoff** : Distance maximale entre deux points au-delà de laquelle le variogramme n'est plus calculé.
- **width** :
- **cloud** : TRUE/FALSE. Si True, cette option calcule la nuée variographique.
- **map** : TRUE/FALSE. Si True, cette option calcule la carte variographique.

### Valeurs renvoyées

Si `cloud=TRUE`, cette fonction renvoie un objet du type «`variogramCloud`». Sinon, elle renvoie un objet du type «`gstat.variogram`» avec les données suivantes :

- **np** : Nombre de paires de points utilisées pour chaque point du variogramme.
- **dist** : Distance moyenne entre les points d'une paire pour chaque point du variogramme.
- **gamma** : Donne les valeurs calculées du variogramme.

### Représentation graphique du variogramme expérimental

Une fois le variogramme expérimental calculé, il est important de le représenter graphiquement afin de pouvoir l'analyser et de déterminer le meilleur modèle variographique. Pour cela, on utilise la fonction suivante :

```
> plot(x,...)
```

où `x` correspond à un objet du type «`gstat.variogram`» ou «`variogramCloud`», c'est-à-dire calculé avec la fonction vue ci-dessus.

Les arguments de cette fonction sont les suivants :

- **type** : Permet de tracer différents types de variogramme
  - “p” : seuls les points calculés sont tracés (option par défaut)
  - “l” : les points calculés sont reliés
  - “o” : les points calculés sont tracés et reliés
  - ...
- **pch** : Permet de changer le type de points tracés (peut prendre des valeurs de 1 à 25)
  - 1 = o (option par défaut)
  - 3 = +
  - ...
- **identify** : TRUE/FALSE : Si on a calculé la nuée variographique (c'est-à-dire si on a ajouté l'option `cloud=TRUE` dans la fonction `variogram()`), cette option permet d'identifier pour chaque point du nuage, quelle paire de points correspond à ce point.

*Exemple :*

```
> vgm<-variogram(moy~1,Mesure,cutoff=0.4)
> vgm
  np      dist      gamma dir.hor dir.ver  id
1  1 0.01165545 0.6328125      0      0 var1
2  2 0.04315513 17.0225694      0      0 var1
3  1 0.27441185 42.7812500      0      0 var1
4  3 0.31039364 41.0651042      0      0 var1
5  2 0.34100534 4.1132812      0      0 var1
```

```
> plot(vgm,pch=3,type="o")
```

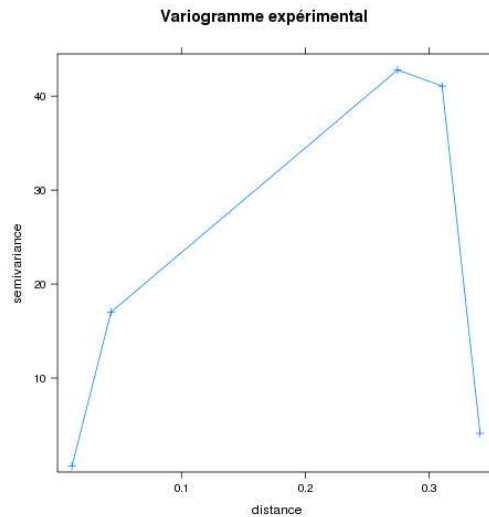


FIG. A.1: Variogramme expérimental

### A.1.2 Modélisation du variogramme

Une fois le variogramme expérimental calculé, l'étape suivante consiste à modéliser ce variogramme à l'aide de modèles connus afin de pouvoir lui appliquer le krigeage. Pour cela, on a besoin de deux fonctions du logiciel R : `vgm(...)` et `fit.variogram(...)`.

#### Générer un variogramme modélisé

```
> vgm(psill, model,...)
```

Cette fonction permet de générer un modèle de variogramme. Elle renvoie un objet de type «variogramModel». Les arguments de cette fonction sont les suivants :

- **psill** : Palier (ou palier asymptotique) du modèle
- **model** : Type de modèle ("Exp", "Sph", ...)
- **range** : Portée du modèle
- **nugget** : Effet pépite du modèle
- **anis** : Paramètres d'anisotropie (si on veut faire dépendre le variogramme d'une direction)

En tapant la commande `vgm()`, R nous renvoie tous les modèles possibles.

```
> vgm()
```

short	long
1 Nug	Nug (nugget)
2 Exp	Exp (exponential)
3 Sph	Sph (spherical)
4 Gau	Gau (gaussian)
5 Exc	Exclass (Exponential class)
6 Mat	Mat (Matern)
7 Ste Mat	(Matern, M. Stein's parameterization)

```
8   Cir          Cir (circular)
9   Lin          Lin (linear)
...

```

Les modèles les plus fréquemment utilisés sont les suivants :

### Ajuster le variogramme expérimental par un modèle de variogramme

```
> fit.variogram(object, model...)
```

Cette fonction permet d'ajuster le variogramme expérimental par un modèle de variogramme. Les arguments de cette fonction sont les suivants :

- **object** : Variogramme expérimental de type «gstat.variogram» (issu de la fonction **variogram(...)**)
- **model** : Modèle de variogramme de type «variogramModel» (issu de la fonction **vgm(...)**)
- **fit.sills** : TRUE/FALSE → détermine si le palier du modèle doit être ajusté.
- **fit.ranges** : TRUE/FALSE → détermine si la portée du modèle doit être ajusté.

Cette fonction renvoie un objet du type «variogramModel» et «data.frame». Elle renvoie le type de modèle (model), le palier (psill) et la portée (range) du variogramme modélisé. Les données **psill** et **range** de la fonction **vgm()** sont données à titre indicatif pour R. A l'aide de ces données, il va automatiquement ajuster le variogramme. Les valeurs psill et range renvoyées peuvent donc différer de celles que l'on a rentré dans la fonction **vgm()**. Si le variogramme n'est pas «ajustable» c'est-à-dire qu'il n'arrive pas à le modéliser, R renvoie un message d'erreur : *warning : «singular model»*.

Lorsque l'on a ajusté le variogramme expérimental, on trace ce dernier avec le variogramme ajusté à l'aide la commande **plot(vario\_exp,vario\_fit)** (c.f. exemple ci-dessous).

*Exemple :*

```
# Modèle sphérique: palier=33, portée=0.17
> vgm.fit<-fit.variogram(vgm,vgm(psill=33,model="Sph",range=0.17))
> vgm.fit
  model  psill  range
1   Sph 32.40504 0.1732973

# Modèle exponentiel: palier=39, portée=0.14
> vgm.fit2<-fit.variogram(vgm,vgm(psill=39,model="Exp",range=0.14))
> vgm.fit2
  model  psill  range
1   Exp 39.33449 0.1351448

# Graphique du variogramme ajusté avec le modèle sphérique
> plot(vgm,vgm.fit)
# Graphique du variogramme ajusté avec le modèle exponentiel
> plot(vgm,vgm.fit2)

```

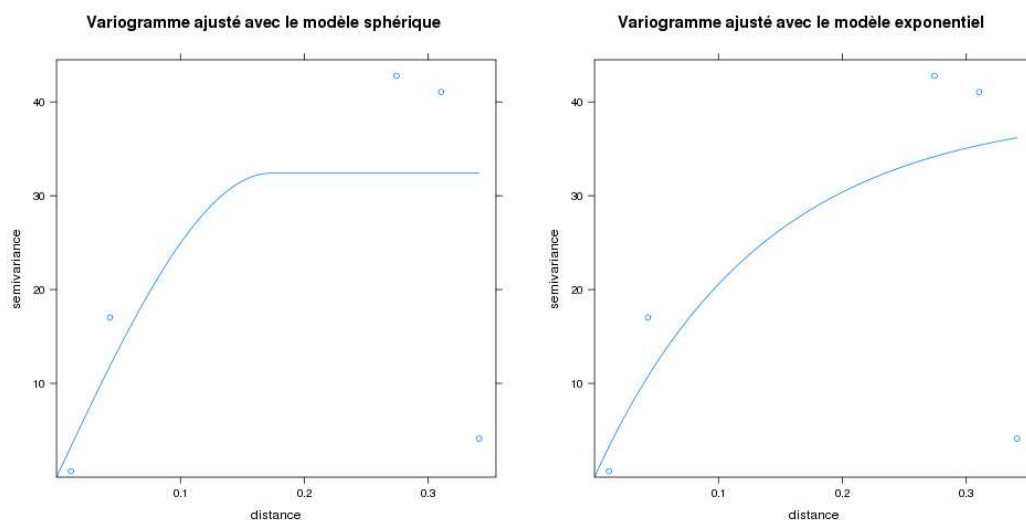


FIG. A.2: Variogrammes ajustés

## A.2 Le krigeage avec R

Afin de kriger la variable régionalisée avec R, il faut tout d'abord créer une grille sur laquelle on pourra effectuer le krigeage.

Voici un exemple de création d'une grille :

```
# Coordonnées lat/long de l'Alsace
> est<-8.22
> ouest<-6.83
> nord<-49.07
> sud<-47.42

> grx<- seq(ouest,est,by=0.05)
> gry<- seq(sud,nord,by=0.05)
> x<-rep(grx, length(gry))
> y<-rep(gry, length(grx))
> y<-sort(y, decreasing=F)
> Grid<-data.frame(x=x, y=y)
> coordinates(Grid)=~x+y
> gridded(Grid)<-TRUE

> class(Grid)
[1] "SpatialPixels"
attr(,"package")
[1] "sp"
```

Une fois la grille créée, on peut effectuer le krigeage sur nos données à l'aide de la fonction suivante :

```
> krige(formula, data,...)
```

Les arguments de cette fonction sont les suivants :

- **formula** :  $\text{variable\_réponse} \sim \text{régresseurs}$ .
  - $\text{variable\_réponse}$  = variable que l'on veut étudier.
  - $\text{régresseurs} = 1$  si on veut faire un krigeage simple ou ordinaire.
  - $\text{régresseurs} = \text{régresseur1} + \text{régresseurs2} + \dots$  si on veut faire un krigeage avec dérive externe
- **data** : Objet de type «dataframe» où se trouvent nos variables utilisées.
- **model** : Variogramme modélisé du type «variogramModel» (issu de la fonction **fit.variogram**).
- **beta** : Si on effectue un krigeage simple, **beta** correspond à l'espérance de la variable régionalisée étudiée.
- **nmax** : Nombre d'observations à utiliser pour kriger chaque point du domaine. Par défaut, toutes les observations sont utilisées
- **maxdist** : Seules les observations étant à une distance inférieure à **maxdist** du point à prédire sont utilisés pour cette prédiction. Si on donne en plus un **nmax**, les deux critères sont appliqués.
- **nmin** : Si le nombre d'observations étant à une distance inférieure à **maxdist** du point à prédire est inférieur à **nmin**, des observations manquantes sont générées (voir l'aide de R pour plus d'informations).

*Remarque* : Il faut aussi donner à cette fonction une grille (de type «SpatialPixels») pour lui indiquer où effectuer le krigeage. (voir exemple ci-dessus).

### Valeurs renvoyées

- Cette fonction renvoie un objet de type «SpatialPixelsDataFrame» avec les données suivantes :
- les coordonnées des points qui ont été krigés
  - les valeurs prédites en ces points ( $\text{var1.pred}$ )

### Obtenir une représentation graphique du krigeage

Pour visualiser le krigeage effectué sur nos données, on utilise la fonction suivante :

```
> spplot(obj, ...)
```

Les arguments de cette fonction sont :

- **obj** : Objet que l'on veut représenter de type «Spatial».
- **xlab** : Nommer l'axe x.
- **ylab** : Nommer l'axe y.
- **aspect** : Ratio des deux axes. Par défaut, une unité de l'axe x est égal à une unité de l'axe y. Mais ce ratio peut être modifié pour avoir une meilleure représentation (si on a des coordonnées lat/long par exemple).

La fonction **spplot** renvoie un graphique de type «trellis». Pour le représenter dans un fichier pdf ou png, il faut utiliser la commande **plot(spplot(...))**.

*Exemple* :

```
> krigeage<-krige(moy~1,Mesure,Grid,vgm.fit)

# On représente la carte krigée seulement pour l'Alsace
> Alsace <- readShapePoly("../alsace_wgs84_region.shp")
> toto <- which(!is.na(overlay(krigDiff_sph, Alsace)))
> plot(spplot(krigeage[toto,"var1.pred"],main="Krigeage",
+ panel=function(...){panel.gridplot(...);sp.polygons(Alsace)},aspect=1.3))
```

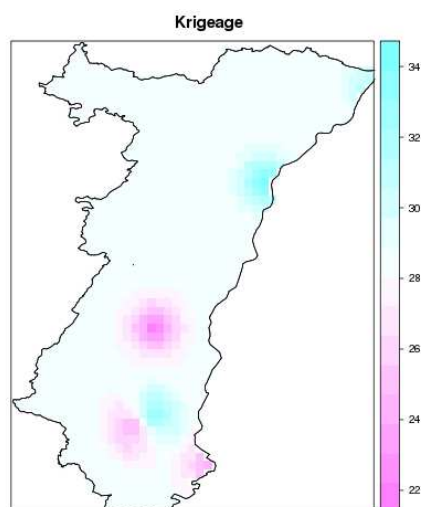


FIG. A.3: Représentation graphique du krigeage

### A.3 La validation croisée avec R

Le package **gstat** que l'on utilise depuis le début permet de faire une validation croisée automatique pour la méthode du krigeage. La fonction utilisée est la suivante :

```
> krige.cv(formula,data ...)
```

Les arguments de cette fonction sont :

- **formula** :variable\_réponse ~ régresseurs (comme pour la fonction 'krige').
- **data** : objet de type «dataframe» où se trouvent nos variables utilisées..
- **model** : Variogramme modélisé du type «variogramModel» (issu de la fonction **fit.variogram**).
- **nfold** : Pour une validation «leave-one-out», ce paramètre vaut 1. (voir l'aide R pour les autres validations croisées possibles).

Cette fonction renvoie un tableau avec les variables suivantes :

- **coordinates** : Coordonnées du point
- **var1.pred** : Valeur estimée par la validation croisée en ce point
- **var1.var** : Variance du krigeage en ce point
- **observed** : Valeur observée en ce point
- **residual** : Résidu de la validation croisée : **observed - var1.pred**
- **zscore** : Résidu standardisé :  $\text{residual}/\sqrt{\text{var1.pred}}$

## Annexe B

# Code R

### B.1 Script pour le reporting

Ce script présente les fonctions utiles pour effectuer les cartes de reporting puis donne un exemple d'application de ces fonctions pour la représentation de la moyenne annuelle de NO<sub>2</sub>.

```
library(ncdf)
library(maptools)
library(sp)
library(DBI)
library(RMySQL)
library(gstat)
library(rgdal)
library(xtable)
library(akima)
library(graphics)

# Fichier permettant de tracer les contours de l'Alsace
Alsace <- readShapePoly("/home/atmo/workspace/Stage_Marie/donnees_entree/shapefiles/
alsace_wgs84_region.shp")

#### Fonction qui transforme des données CHIMERE en geodata ####

## Paramètres de la fonction:
## - La matrice ou vecteur des latitudes (lat)
## - La matrice ou vecteur des longitudes (lon)
## - La matrice ou vecteur des valeurs du modèle en chaque point (polluant)

## Valeur renvoyée: Le data.frame des données transformées en geodata

chimereAsGeodata<-function(lat,lon,polluant){

Lat<-as.vector(lat)
Lon<-as.vector(lon)
Polluant<-as.vector(polluant)
geodata<-data.frame(lat=Lat,lon=Lon,data=Polluant)
coordinates(geodata)<-~lon+lat
```

```
return(geodata)
}
```

```
#### Fonction qui représente le modèle CHIMERE ####
```

```
## Paramètres de la fonction:
```

```
## - La matrice des latitudes (lat)
```

```
## - La matrice des longitudes (lon)
```

```
## - La matrice des valeurs du modèle en chaque point (z_mod)
```

```
## Valeur renvoyée: L'interpolation sur une grille régulière en lat/lon des données CHIMERE
```

```
ModChim<-function(lat,lon,z_mod){
```

```
# Interpolation des données CHIMERE sur une grille régulière en lat/long
```

```
intepChim <- interp(as.vector(lon), as.vector(lat), as.vector(z_mod), xo = grx, yo = gry)
```

```
grille_intepChim <- expand.grid(intepChim$x,intepChim$y)
```

```
intep_geoChim <- chimereAsGeodata(grille_intepChim$Var2,grille_intepChim$Var1,intepChim$z)
```

```
gridded(intep_geoChim) <- TRUE
```

```
# Représentation du modèle CHIMERE
```

```
toto2<-which(!is.na(overlay(intep_geoChim, Alsace)))
```

```
plot(spplot((intep_geoChim[toto2,"data"]),
```

```
panel=function(...){
```

```
panel.gridplot(...)
```

```
sp.polygons(Alsace)
```

```
},
```

```
aspect=1.3,
```

```
col.regions=heat.colors(50)[50:1],
```

```
at=c(seq(0,50,2.5)),
```

```
main="Modele CHIMERE"))
```

```
return(intep_geoChim)
```

```
}
```

```
#### Fonction qui calcule le variogramme expérimental et qui le modélise puis représente les deux variogrammes ####
```

```
## Paramètres de la fonction:
```

```
## - Fichier de mesure de type geodata (Mesure)
```

```
## - La portée maximale du variogramme expérimental exprimée en degré (portee)
```

```
## - Le pas du variogramme expérimental exprimé en degré (pas)
```

```
## - La station à retirer pour calculer le variogramme expérimental (StationRetiree)
```

```
## - numéro de la station s'il y en a une
```

```
## - 0 sinon
```

```
## Valeur renvoyée: Le variogramme modélisé
```

```
Vario<-function(Mesure,portee,pas,StationRetiree){
```

```
if(StationRetiree==0){
```



```

Donnees<-Mesure
}

else{

Donnees<-RetraitStation(Mesure,StationRetiree)
}

# Calcul du variogramme expérimental et ajustement de ce dernier
vgmDiff<-variogram(diff~1,Donnees,cutoff=portee,width=pas)
vgmDiff.fit<-fit.variogram(vgmDiff,vgm(psill=1,model="Sph",range=0.4))
# Représentation des variogrammes expérimental et modélisé
plot(plot(vgmDiff,vgmDiff.fit,pch=3,type="o",main=paste("Variogramme avec portee=",portee,
" et pas=",pas," ajuste avec spherique"),sub=paste("Station retiree:",StationRetiree)))

return(vgmDiff.fit)
}

#### Fonction qui effectue et représente le krigeage des innovations (différence Mesure/Modele) ####

## Paramètres de la fonction:
## - Le fichier de mesure de type geodata (Mesure)
## - Le variogramme modélisé (variofit): issu de la fonction 'Vario'
## - La grille sur laquelle effectuer le krigeage (Grid)

## Valeur renvoyée: Le krigeage des différences

KrigDiff<-function(Mesure,variofit,Grid){

# Krigeage des différences mes/mod (avec toutes les stations)
KrigDiff<-krige(diff~1,Mesure,Grid,variofit)

# Représentation du krigeage
toto1 <- which(!is.na(overlay(KrigDiff , Alsace)))
plot(spplot((KrigDiff[toto1,"var1.pred"]),
panel=function(...)
{
panel.gridplot(...)
sp.text(coordinates(Mesure),as.vector(Mesure$mes))
sp.polygons(Alsace)
},
aspect=1.3,
col.regions=heat.colors(50)[50:1],
main=paste("Krigeage des Innovations")
))

return(KrigDiff)
}

#### Fonction qui représente le Modèle CHIMERE + Différence Krigée ####

```

```

## Paramètres de la fonction:
## - Le fichier de mesure de type geodata (Mesure)
## - La Différence Mesure/Modele krigée (DiffKrig): issue de la fonction 'KrigDiff'
## - Le modèle CHIMERE (modChim): issu de la fonction 'ModChim'

## Valeur renvoyée: Le nouveau modèle obtenu

ModR<-function(Mesure,KrigDiff,modChim){

Krig<-KrigDiff

# On ajoute le modèle CHIMERE à la différence Mesure/Modele krigée
Krig@data[,1] <- KrigDiff@data[,1]+modChim@data

# Représentation de la moyenne annuelle de NO2
toto1 <- which(!is.na(overlay(KrigDiff , Alsace)))
plot(spplot((Krig[toto1,"var1.pred"]),
panel=function(...)
{
panel.gridplot(...)
sp.text(coordinates(mesureFrame),as.vector(mesureFrame$mes))
sp.polygons(Alsace)
},
aspect=1.3,
col.regions=heat.colors(50)[50:1],
at=c(seq(0,50,2.5)),
main=paste("Nouveau modele")
))

return(Krig)
}

##### Fonction qui effectue la validation croisée du krigeage des innovations #####

## Paramètres de la fonction:
## - Le fichier de mesure de type geodata (Mesure)
## - Le variogramme modélisé (variofit): issu de la fonction 'Vario'

## Valeur renvoyée: Le tableau de la validation croisée et affiche différents critères de validation

Valid<-function(Mesure,variofit){

valid<-krige.cv(diff~1,Mesure,variofit)

MSE<-mean((valid$residual)^2)
PRESS<-sum((valid$residual)^2)
Ad<-mean((valid$zscore)^2)
B<-mean(valid$residual)

criteres<-data.frame(MSE,PRESS,Ad,B)
print(criteres)
return(valid)
}

```

```
}

```

```
##### Représentation de la moyenne annuelle de NO2 #####

```

```
# Création de la grille régulière sur l'Alsace

```

```
est<-8.3
ouest<-6.7
nord<-50
sud<-47
grx<- seq(ouest,est,by=0.03)
gry<- seq(sud,nord,by=0.03)
x<-rep(grx, length(gry))
y<-rep(gry, length(grx))
y<-sort(y, decreasing=F)
Grid<-data.frame(x=x, y=y)
coordinates(Grid)=~x+y
gridded(Grid)<-TRUE

```

```
#####

```

```
##### Importation des données (mesures et modèle) #####

```

```
# Pour le reporting, on travaille sur des données pré-traitées. Ces dernières se trouvent dans
# un fichier texte que l'on importe. On les place ensuite dans une matrice 78*93.

```

```
# On lit les coordonnées et on les met dans une matrice: x pour latitude et y pour longitude
coord <- read.table("/home/atmo/workspace/Stage_Marie/donnees_entree/coord.txt", header=TRUE,
  colClasses=c('numeric', 'numeric', 'numeric'),dec = ",", sep="\t")
lon <- coord[,2]
lat <- coord[,1]
x <- matrix(lat, 78,93)
y <- matrix(lon, 78,93)

```

```
# On lit les données du modèle et on les met dans une matrice z_mod
Modele <- read.table("/home/atmo/workspace/Stage_Marie/donnees_entree/MoyAnnuNO2_mod.txt",
  colClasses=c("numeric"), header=FALSE, sep=";")
z_mod <- matrix(Modele[,1], nrow=78, ncol=93, byrow='FALSE')

```

```
# On lit les données de mesure et on les met dans une matrice z_mes
Mesure <- read.table("/home/atmo/workspace/Stage_Marie/donnees_entree/MoyAnnuNO2_mes.txt",
  colClasses=c("numeric"), header=FALSE, sep=";")
z_mes <- matrix(Mesure[,1], nrow=78, ncol=93, byrow='FALSE')

```

```
# On cherche les points qui ont une mesure (càd != -999)
ind_mes <- which(z_mes!=-999, arr.ind=TRUE)
mesnotnull <- z_mes[ind_mes]
diffmesnotnull <- z_mes[ind_mes]- z_mod[ind_mes]
latnotnull <- x[ind_mes]
lonnotnull <- y[ind_mes]

```

```

# On récapitule les mesures et différences mes/mod dans un dataframe
mesureFrame <- data.frame(mes=mesnotnull,diff=diffmesnotnull, lat=latnotnull, lon=lonnotnull)
coordinates(mesureFrame) <- c('lon', 'lat')
#####

#####
# Creation du pdf
nom_pdf <- paste("/home/atmo/workspace/Stage_Marie/Krigeage/Krigeage2009/Moyenne2009N02.pdf")
pdf(nom_pdf, width=10, heigh=12, paper="a4")
opar <- par()

# Variogramme
variofit<-Vario(mesureFrame,1.5,0.2,0)
# Krigeage des innovations
diffKrig<-KrigDiff(mesureFrame,variofit,Grid)
# Modèle CHIMERE
modeleChimere<-ModChim(x,y,z_mod)
# Nouveau Modèle: Krigeage des innovations + modèle CHIMERE
modele<-ModR(mesureFrame,diffKrig,modeleChimere)
# Validation croisée
valid<-Valid(mesureFrame,variofit)

dev.off()
#####

```

## B.2 Script pour le nombre de jours de dépassement en PM10

Ce script permet de représenter le nombre de jours de dépassement en PM10 à l'aide la méthode présentée dans le paragraphe 2.3.

```

library(ncdf)
library(maptools)
library(sp)
library(DBI)
library(RMySQL)
library(gstat)
library(rgdal)
library(xtable)
library(akima)
library(graphics)

# On utilise le script de base
source("/home/atmo/workspace/Stage_Marie/Rapport/BaseRapport.R")

# Fichier permettant de tracer les contours de l'Alsace
Alsace <- readShapePoly("/home/atmo/workspace/Stage_Marie/donnees_entree/shapefiles/
alsace_wgs84_region.shp")

# Création d'une grille régulière sur l'Alsace

```

```

est<-8.5
ouest<-6.5
nord<-50.5
sud<-46.5

grx<- seq(ouest,est,by=0.03)
gry<- seq(sud,nord,by=0.03)
x<-rep(grx, length(gry))
y<-rep(gry, length(grx))
y<-sort(y, decreasing=F)
Grid<-data.frame(x=x, y=y)
coordinates(Grid)=~x+y
gridded(Grid)<-TRUE

# Fonction qui interpole le modèle CHIMERE sur la grille définie ci-dessus

## Paramètres de la fonction:
## - Les données du modèle CHIMERE (DonneesChimere)
## - La date pour laquelle on veut interpoler le modèle (Date)

## Valeur renvoyée: Les données interpolées sur la grille régulière

InterpChimere<-function(DonneesChimere,Date){

# On met les données Chimère sur notre grille
Chimere_intep <- interp(as.vector(DonneesChimere$lon), as.vector(DonneesChimere$lat),
as.vector(DonneesChimere$polluant),xo = grx, yo = gry)
grille_intep <- expand.grid(Chimere_intep$x,Chimere_intep$y)
# Transformation en données geodata
Chimere_intep_geo <- chimereAsGeodata(grille_intep$Var2,grille_intep$Var1,Chimere_intep$z)
gridded(Chimere_intep_geo) <- TRUE

return(Chimere_intep_geo)
}

##### Recherche des points de dépassement journaliers après krigeage #####

annee<-'2009'
polluant<-'PM10'

# Initialisation de la date au 1er janvier
Date <- as.Date(paste(annee,'-01-01',sep=''))

# Initialisation des listes de dépassement
DepJour <- list()
DepJourChim<-list()

# On fait un boucle sur toute l'année pour kriger chaque jour
# Attention: comme on supprime les jours qui nous dérange, on ne fait pas tourner sur 365 jours
# mais sur 365-6 (qui correspond au nombre de jours que l'on retire)

```

```
for(i in c(1:359)){

# Dates qui posent un problème
if(Date=="2009-06-27" || Date=="2009-08-08" || Date=="2009-08-12" || Date=="2009-10-25")

{
Date<-Date+1
}

# Fichier Chimere non existant pour le 29 et 30 aout
if(Date=="2009-08-29")
{
Date<-Date+2
}

print(Date)

##### Import des donnees #####

# Import des mesures
Mesure<-extractionMesure(annee,polluant,Date)

# Rajout des stations vituelles dans les Vosges
ind<-length(Mesure$moy)+1
# Vosges moyennes
Mesure$lat[ind]<-48.4906
Mesure$lon[ind]<-7.12664
Mesure$moy[ind]<-12
Mesure$i[ind]<-41
Mesure$j[ind]<-62
Mesure$zone[ind]<-8

# Vosges du sud
Mesure$lat[ind+1]<-47.8129
Mesure$lon[ind+1]<-6.9426
Mesure$moy[ind+1]<-13
Mesure$i[ind+1]<-37
Mesure$j[ind+1]<-44
Mesure$zone[ind+1]<-8

# Transformation en geodata et suppression des lignes où il manque une mesure
Mesure_geo<-mesureAsGeodata(Mesure)

# Import des données du modèle CHIMERE et transformation en geodata
# Comme on a des données horaires, on prend la valeur moyenne de la journée
Chimere<-importChimere(annee,Date,polluant)
Chimere$polluant <- rowMeans(Chimere$polluant, na.rm = FALSE, dims = 2)

# Différence Mesure/Modèle
```

```

Diff<-Mesure_geo
for (j in c(1:nrow(Mesure_geo))){
Diff$moy[j] <- Mesure_geo$moy[j] - Chimere$polluant[Mesure_geo$i[j],Mesure_geo$j[j]]
}

##### Variogramme des différences mesure/modèle #####
Vgm<-variogram(moy~1,data=Diff,cutoff=1.5,width=0.25)
Vgm.fit<-fit.variogram(Vgm,vgm(psill=1,model="Sph",range=0.4))

##### Krigeage des différences mesure/modèle #####
Krig<-krige(moy~1,Diff,Grid,Vgm.fit)

##### Interpolation sur la grille de krigeage du modèle Chimère #####
IntepChim<-InterpChimere(Chimere,Mesure_geo,Date)

##### Représentation des différences + modèle Chimère #####
Krig@data[,1] <- Krig@data[,1]+IntepChim@data

# Dépassement du krigeage : on crée une liste par jour. S'il y a dépassement au point de
# la maille->1 sinon 0
# S'il n'y a pas de valeur, on met 0
DepJour[[i]] <- ifelse(Krig@data[,1]>=50,1,0)
DepJour[[i]][is.na(DepJour[[i]])] <- 0

# Dépassement du modele CHIMERE : on crée une liste par jour. S'il y a dépassement au
# point de la maille->1 sinon 0
# S'il n'y a pas de valeur, on met 0
DepJourChim[[i]]<- ifelse(as.vector(Chimere$polluant)>=50,1,0)
DepJourChim[[i]][is.na(DepJourChim[[i]])] <- 0

# On passe au jour suivant
Date <- Date + 1

}

#####
# Creation du pdf
nom_pdf <- paste("/home/atmo/workspace/Stage_Marie/Krigeage/KrigeagePM10/DepassementAnnuelPM10.pdf")
pdf(nom_pdf, width=10, heigh=12, paper="a4")
opar <- par()

##### Représentation du nombre de jours de dépassement après krigeage des innovations #####
DepAnnu <- do.call('cbind',DepJour)
DepAnnu <- apply(DepAnnu, 1, sum)

intep <- interp(Krig@coords[,1], Krig@coords[,2], DepAnnu , xo = grx, yo = gry)

```

```

grille_interp <- expand.grid(intep$x,intep$y)
intep_geo <- chimereAsGeodata(grille_interp$Var2,grille_interp$Var1,intep$z)
gridded(intep_geo) <- TRUE
toto2<-which(!is.na(overlay(intep_geo, Alsace)))

plot(spplot((intep_geo[toto2,"data"]),
panel=function(...){panel.gridplot(...);sp.polygons(Alsace)},
aspect=1.3,
col.regions=heat.colors(50)[50:1],
at=c(seq(0,40,2.5)),
main="Dépassement PM10: Nouvelle méthode",
sub="Variogramme sphérique (portee=1.5 pas=0.25)"
))

##### Représentation du dépassement du modele CHIMERE #####
DepAnnuChim<-do.call('cbind',DepJourChim)
DepAnnuChim <- apply(DepAnnuChim, 1, sum)

intepChim <- interp(as.vector(Chimere$lon),as.vector(Chimere$lat) , as.vector(DepAnnuChim) ,
  xo = grx, yo = gry)
grille_interpChim <- expand.grid(intepChim$x,intepChim$y)
intep_geoChim <- chimereAsGeodata(grille_interpChim$Var2,grille_interpChim$Var1,intepChim$z)
gridded(intep_geoChim) <- TRUE
toto2<-which(!is.na(overlay(intep_geoChim, Alsace)))
plot(spplot((intep_geoChim[toto2,"data"]),
panel=function(...){panel.gridplot(...);sp.polygons(Alsace)},
aspect=1.3,
col.regions=heat.colors(50)[50:1],
at=c(seq(0,40,2.5)),
main="Dépassement PM10: Modèle CHIMERE"
))

dev.off()

```

### B.3 Script pour la prévision

Ce script expose les différentes fonctions nécessaires pour effectuer les cartes de prévision du maximum journalier de NO<sub>2</sub> et donne un exemple d'application en représentant le 24-01-2009.

Ce script fait intervenir le script 'Base' exposé dans le paragraphe B.4.

```

library(ncdf)
library(maptools)
library(sp)
library(DBI)
library(RMySQL)
library(gstat)
library(rgdal)
library(xtable)

```



```
library(akima)
library(graphics)

source("/home/atmo/workspace/Stage_Marie/Krigeage/base.R")

Alsace <- readShapePoly("/home/atmo/workspace/Stage_Marie/donnees_entree/shapefiles/
alsace_wgs84_region.shp")

# Création de la grille

est<-8.3
ouest<-6.7
nord<-50
sud<-47

grx<- seq(ouest,est,by=0.03)
gry<- seq(sud,nord,by=0.03)
x<-rep(grx, length(gry))
y<-rep(gry, length(grx))
y<-sort(y, decreasing=F)
Grid<-data.frame(x=x, y=y)
coordinates(Grid)=~x+y
gridded(Grid)<-TRUE

#### Fonction qui représente le modèle CHIMERE ####

## Paramètres de la fonction:
## - Les données du modèle CHIMERE (DonneesChimere)
## - La date (Date)

## Valeur renvoyée: L'interpolation sur une grille régulière en lat/lon des données CHIMERE

ModChim<-function(DonneesChimere,Date){

intepChim <- interp(as.vector(DonneesChimere$lon), as.vector(DonneesChimere$lat),
as.vector(DonneesChimere$polluant), xo = grx, yo = gry)
grille_interpChim <- expand.grid(intepChim$x,intepChim$y)
intep_geoChim <- chimereAsGeodata(grille_interpChim$Var2,grille_interpChim$Var1,intepChim$z)
gridded(intep_geoChim) <- TRUE
toto2<-which(!is.na(overlay(intep_geoChim, Alsace)))
plot(spplot((intep_geoChim[toto2,"data"]),
panel=function(...){panel.gridplot(...);sp.polygons(Alsace)},
aspect=1.3,
col.regions=heat.colors(50)[50:1],
at=c(seq(0,110,3.5)),
main=paste("Modele CHIMERE: ",Date)))
return(intep_geoChim)
}
```

```
##### Fonction qui calcule le variogramme expérimental et qui le modélise et qui représente
les deux variogrammes #####

## Paramètres de la fonction:
## - Fichier de mesure de type geodata (Mesure)
## - La portée maximale du variogramme expérimental exprimée en degré (portee)
## - Le pas du variogramme expérimental exprimé en degré (pas)

## Valeur renvoyée: Le variogramme modélisé

Vario<-function(Mesure,portee,pas){

# Calcul du variogramme expérimental et ajustement de ce dernier
vgmDiff<-variogram(diff~1,Mesure,cutoff=portee,width=pas)
vgmDiff.fit<-fit.variogram(vgmDiff,vgm(psill=1,model="Sph",range=0.4))
# Représentation des variogrammes expérimental et modélisé
plot(plot(vgmDiff,vgmDiff.fit,pch=3,type="o",main=paste("Variogramme avec portee=",portee,
" et pas=",pas," ajuste avec spherique")))

return(vgmDiff.fit)
}

##### Fonction qui effectue et représente le krigeage des innovations (différence Mesure/Modele) ##

## Paramètres de la fonction:
## - Le fichier de mesure de type geodata (Mesure)
## - Le variogramme modélisé (variofit): issu de la fonction 'Vario'
## - La grille sur laquelle effectuer le krigeage (Grid)
## - La date (Date)

## Valeur renvoyée: Le krigeage des différences

KrigDiff<-function(Mesure,variofit,Grid,Date){

# Krigeage des différences mes/mod (avec toutes les stations)
KrigDiff<-krige(diff~1,Mesure,Grid,variofit)

# Représentation du krigeage
toto1 <- which(!is.na(overlay(KrigDiff , Alsace)))
plot(spplot((KrigDiff[toto1,"var1.pred"]),
panel=function(...)
{
panel.gridplot(...)
sp.polygons(Alsace)
},
aspect=1.3,
col.regions=heat.colors(50)[50:1],
main=paste("Krigeage des Innovations")
))

return(KrigDiff)
}
```

```
}

```

```
##### Fonction qui représente le Modèle CHIMERE + Différence Krigée #####

```

```
## Paramètres de la fonction:

```

```
## - Le fichier de mesure de type geodata (Mesure)

```

```
## - La Différence Mesure/Modele krigée (DiffKrig): issue de la fonction 'KrigDiff'

```

```
## - Le modèle CHIMERE (modChim): issu de la fonction 'ModChim'

```

```
## - La date (Date)

```

```
## Valeur renvoyée: Le nouveau modèle obtenu

```

```
ModR<-function(Mesure,KrigDiff,modChim,Date){

```

```
Krig<-KrigDiff

```

```
# On ajoute le modèle CHIMERE à la différence Mesure/Modele krigée

```

```
Krig@data[,1] <- KrigDiff@data[,1]+modChim@data

```

```
# Representation du nouveau modele (representation du modele chimere + diff mes/mod)

```

```
toto1 <- which(!is.na(overlay(KrigDiff , Alsace)))

```

```
plot(spplot((Krig[toto1,"var1.pred"]),

```

```
panel=function(...)
```

```
{
```

```
panel.gridplot(...)
```

```
sp.text(coordinates(Mesure),as.vector(Mesure$moy))
```

```
sp.polygons(Alsace)
```

```
},
```

```
aspect=1.3,
```

```
col.regions=heat.colors(50)[50:1],
```

```
at=c(seq(0,110,3.5)),
```

```
main=paste("Nouveau modele du ",Date)
```

```
))
```

```
return(Krig)
```

```
}

```

```
##### Fonction qui effectue la validation croisée du krigeage des innovations #####

```

```
## Paramètres de la fonction:

```

```
## - Le fichier de mesure de type geodata (Mesure)

```

```
## - Le variogramme modélisé (variofit): issu de la fonction 'Vario'

```

```
## Valeur renvoyée: Le tableau de la validation croisée et affiche différents critères

```

```
## de validation

```

```
Valid<-function(Mesure,variofit){

```

```
valid<-krige.cv(diff~1,Mesure,variofit)

```

```

MSE<-mean((valid$residual)^2)
PRESS<-sum((valid$residual)^2)
Ad<-mean((valid$zscore)^2)
B<-mean(valid$residual)

criteres<-data.frame(MSE,PRESS,Ad,B)
print(criteres)
return(valid)

}

#### Fonction principale qui réunit les fonctions ci-dessus pour représenter le maximum journalier
de NO2 en un jour donné. Elle crée un pdf dans lequel sont représenté les sorties graphiques ####

## Paramètres de la fonction:
## - L'année (annee)
## - Le polluant (polluant)
## - La date (Date)

main<-function(annee,polluant,Date){

# Creation du pdf
nom_pdf <- paste("/home/atmo/workspace/Stage_Marie/Krigeage/KrigeageNO2/
KrigMaxJourNO2(",Date,").pdf")
pdf(nom_pdf, width=10, heigh=12, paper="a4")
opar <- par()

##### Import des donnees #####

# Import des mesures et transformation en mesures géo
Mesure<-extractionMesure(annee,polluant,Date)
Mesure_geo<-mesureAsGeodata(Mesure)

# Import des données du modèle Chimère et transformation en mesures géo
# Comme on a des données horaires, on prend la valeur max de la journée
Chim<-importChimere(annee,Date,polluant)
Chimere<-list()
Chimere$lat<-Chim$lat
Chimere$lon<-Chim$lon
Chimere$polluant<-matrix(nrow=78,ncol=93)
for(i in c(1:78)){
for(j in c(1:93)){
Chimere$polluant[i,j]<-max(Chim$polluant[i,j,])
}
}

# Différence mesure/modèle
mesureFrame<-Mesure_geo

```

```

for (j in c(1:nrow(Mesure_geo))){
mesureFrame$diff[j] <- Mesure_geo$moy[j] - Chimere$polluant[Mesure_geo$i[j],Mesure_geo$j[j]]
}

##### Variogramme et Krigeage #####
variofit<-Vario(mesureFrame,portee=1.5,pas=0.25)
DiffKrig<-KrigDiff(mesureFrame,variofit,Grid,Date)
modChim<-ModChim(Chimere,Date)
modele<-ModR(mesureFrame,DiffKrig,modChim,Date)

dev.off()

}

##### Représentation du maximum journalier de N02 pour la date du 24-01-2009 #####

annee<-'2009'
polluant<-'N02'
Date<-as.Date("2009-01-24")
print(Date)

main(annee,polluant,Date)

```

## B.4 Fichier Base

Ce script réunit les principales fonctions que l'on utilise pour importer des données du modèle CHIMERE et les mesures de la base Polair.

```

library(ncdf)
library(maptools)
library(sp)
library(DBI)
library(RMySQL)
library(gstat)
library(rgdal)
library(xtable)
library(akima)
library(graphics)

# Fichier source contenant les fonctions pour l'utilisation de base de données MySQL
source("/home/atmo/workspace/Stage_Marie/Krigeage/tools_db_common.R")

# Chemin fichiers
datesFiles <- list('PM10' = list(
'2009' = "/home/atmo/workspace/Stage_Marie/donnees_entree/dates_depassement_pm10_2009.csv",
'2010' = "/home/atmo/workspace/Stage_Marie/donnees_entree/dates_depassement_pm10_2010.csv"),
'03' = list('2009' = "/home/atmo/workspace/Stage_Marie/donnees_entree/dates_depassement_o3_2009.csv"))

```

```

'NO2' = list('2009' = "/home/atmo/workspace/Stage_Marie/donnees_entree/
dates_depassement_no2_2009.csv")
)
chimereRepo <- list('2007' = "/mnt/Modelisation/2007/CHIMERE_PAMINA/",
'2008' = "/mnt/Modelisation/2008/CHIMERE_04km/",
'2009' = "/mnt/Modelisation/2009/CHIMERE_04km/",
'2010' = "/mnt/Modelisation/CHIMERE_04km/")

chimereBaseName <- list('2007' = "CHIMERE_V200501H_PAMINA_",
'2008' = "out.",
'2009' = "out.",
'2010' = "out.")

chimereSuffix <- list('2007' = ".nc",
'2008' = "_previ.nc",
'2009' = "_previ.nc",
'2010' = "_previ.nc")

chimereLatLonName <- list('2007' = c('latitcrs', 'longicrs'),
'2008' = c('lat', 'lon'),
'2009' = c('lat', 'lon'),
'2010' = c('lat', 'lon'))

configFile <- "/home/atmo/workspace/Stage_Marie/Krigeage/stations_krigeage.csv"
configFrame <- read.table(configFile, sep=";", header=TRUE)

# Paramètres de connection a Polair
DSN <- list( 'polair' = list( 'user' = 'polair',
'password' = 'petunia',
'dbname' = 'polair',
'host' = '172.16.26.10')
)

# Fonction permettant de lire un fichier de dates
### Paramètres de la fonction: l'année et le polluant

LectureFichierDate <- fonction(annee, polluant, datesF=datesFiles){

datesFile <- datesF[[polluant]][[annee]]
datesDep <- read.table(datesFile, sep=";", colClasses=c("character"), col.names=c("datesDep"))
datesDep$datesDep <- as.Date(datesDep$datesDep, format="%d/%m/%Y")
return(datesDep)
}

##### Extraction des données du modèle CHIMERE #####

# Création du nom du fichier CHIMERE

```

```

## Paramètres de la fonction: l'année et la date

nomFichierChimere <- fonction(annee, dateDbt){

dateDbt_str <- format(dateDbt, "%Y%m%d")
dateFin_str <- format(dateDbt + 4, "%Y%m%d")
chimereFileName <- paste(chimereRepo[[annee]], chimereBaseName[[annee]], dateDbt_str, "_",
dateFin_str, chimereSuffix[[annee]], sep="")

return(chimereFileName)
}

# Ouverture du fichier CHIMERE
## Paramètres de la fonction: le nom du fichier CHIMERE (issu de "nomFichierChimere")

ouvertureFichierChimere <- fonction(chimereFileName){

tryCatch( netcdf <- open.ncdf(chimereFileName),
error = function(e) (print(paste("ERREUR de connection", e))),
finally=print(paste("fichier netCDF ouvert"))
)

return(netcdf)
}

# Import des données Chimère
## Paramètres de la fonction: l'année, la date et le polluant

importChimere<-fonction(annee,dateDep,polluant){

chimereFileName <- nomFichierChimere(annee, dateDep)
netcdf <- ouvertureFichierChimere(chimereFileName)
lat <- get.var.ncdf(netcdf, chimereLatLonName[[annee]][1], start=c(1,1),count=c(-1,-1))
lon <- get.var.ncdf(netcdf, chimereLatLonName[[annee]][2], start=c(1,1),count=c(-1,-1))
Polluant <- get.var.ncdf(netcdf, polluant, start=c(1, 1, 1, 1),count=c(-1,-1,1,24))
close.ncdf(netcdf)

return(list(lat=lat,lon=lon,polluant=Polluant))
}

# Transformation des données CHIMERE en geodata
## Paramètres de la fonction: une liste de latitude, une de longitude et une des concentrations
## du polluant

chimereAsGeodata<-fonction(lat,lon,polluant){
Lat<-as.vector(lat)
Lon<-as.vector(lon)
Polluant<-as.vector(polluant)
geodata<-data.frame(lat=Lat,lon=Lon,data=Polluant)
coordinates(geodata)<-~lon+lat
return(geodata)
}

```

```
##### Extraction des mesures du fichier Polair #####
```

```
getMesurePolair <- fonction(date_dbt, date_fin, vh_osm, basePolair=Base_Polair){

query = paste("SELECT VH_DATE_HOR AS date, VH_VALV AS Valid, VH_ETAV AS CodeValid,
VH_VALB AS Brute, VH_ETAB AS CodeBrute ",
"FROM POLAIR.VAL_HOR WHERE ",
"VH_DATE_HOR >= '",date_dbt,"%' AND VH_DATE_HOR < '",date_fin,"%' AND `VH_OSM` = '",
vh_osm,'" AND VH_ETAV <> 'C' ",
"AND VH_ETAV <> 'D' and VH_ETAV <> 'P' and VH_ETAV <> 'I' and VH_ETAV <> 'M' and VH_ETAV <>
'Z';",sep="")

# Exportation des données validées et Brutes avec leur code de validation
result <- dbGetQuery(basePolair, query)
if (length(result)!=0) {
# test du code de validation des donnees valide
valide <- which(result$CodeValid == "A" | result$CodeValid == "0" | result$CodeValid == "R")
# Test de la présence des mesures ('N' <- mesure pas encore rapatriées)
noval <- which(result$CodeValid == "N")

# On retiens les données validées et si pas de validation brutes
Mesure <- result$Brute
Mesure[valide] <- result$Valid[valide]
Mesure[noval] <- NA
Mesure.frame <- data.frame(result$date, Mesure); names(Mesure.frame) <- c("date",vh_osm)
Mesure.frame <- na.omit(Mesure.frame)
}else{
return(NULL)
}

return(Mesure.frame)
}

# Extraction des numéros de station et de polluant et création du numéro polair associé
## Paramètres de la fonction: le polluant et le data.frame de configuration (issue de la lecture
## fichier conf)

extract_codes_polair <- fonction(polluantNom, configF=configFrame) {

config_donnee <- subset(configF, get(polluantNom)!=0, select=c("Code_Station", polluantNom,
'Latitude', 'Longitude', "i_PAMINA", "j_PAMINA", "i_PREVEST", "j_PREVEST", "Zone"))
Codes_Polair <- as.numeric(config_donnee[,1]) + as.numeric(config_donnee[,2])
return(list(Codes_Polair=Codes_Polair, lat= config_donnee$Latitude, lon=config_donnee$Longitude,
i_PAMINA=config_donnee$i_PAMINA, j_PAMINA=config_donnee$j_PAMINA,
i_PREVEST=config_donnee$i_PREVEST, j_PREVEST=config_donnee$j_PREVEST,
Zone=config_donnee$Zone))
}
```



```
# Extraction de la mesure journalière du fichier polair
## Paramètres de la fonction: l'année, le polluant et la date

extractionMesure <- fonction(annee, polluantNom, date){

# La fonction "getMesurePolair" renvoie des données horaires par station.
# La fonction "extractionMesure" transforme ces données en journaliere:
# moyenne journalière pour PM10 et maximum journalier pour les autres

codesPolair <- extract_codes_polair(polluantNom)
date_fin <- as.Date(date) + 1
mesure <- NULL
i <- 1
for (codeStat in codesPolair$Codes_Polair) {
temp <- getMesurePolair(date, date_fin, codeStat)

if (is.data.frame(temp) & length(temp)==2){
j <- which(codesPolair$Codes_Polair==codeStat)

if (polluantNom=="PM10"){
mesure$moy[i] <- mean(temp[,2], na.rm=TRUE)
mesure$lat[i] <- codesPolair$lat[j]
mesure$lon[i] <- codesPolair$lon[j]
mesure$i[i] <- codesPolair$i_PREVEST[j]
mesure$j[i] <- codesPolair$j_PREVEST[j]
mesure$zone[i] <- codesPolair$Zone[j]

}else{
if(length(temp[,2])!=0){
mesure$moy[i] <- max(temp[,2], na.rm=TRUE)
mesure$lat[i] <- codesPolair$lat[j]
mesure$lon[i] <- codesPolair$lon[j]
mesure$i[i] <- codesPolair$i_PREVEST[j]
mesure$j[i] <- codesPolair$j_PREVEST[j]
mesure$zone[i] <- codesPolair$Zone[j]

}
}

i <- i+1
}
}
return(mesure)
}

# Transformation des mesures en geodata
## Paramètre de la fonction: une liste (issue de la fonction "extractionMesure")

mesureAsGeodata<-function(MesurePolluant){
```

```
MesurePolluant<-as.data.frame(MesurePolluant)
MesurePolluant<-na.omit(MesurePolluant)
coordinates(MesurePolluant)<-c("lon","lat")
return(MesurePolluant)
}

# Retrait d'une station du fichier de mesure
## Paramètres de la fonction: Fichier de données geodata (issue de la fonction "mesureAsGeodata")
## et le numéro de la station dans ce fichier

RetraitStation<-function(Donnees,NumeroStation){

lon<-coordinates(Donnees)[,1]
lat<-coordinates(Donnees)[,2]
mes<-Donnees$mes
diff<-Donnees$diff

lon1<-NULL
lat1<-NULL
mes1<-NULL
diff1<-NULL
station1<-NULL

for(j in c(1:nrow(Donnees))){

if(j!= NumeroStation){
lon1[j]<-lon[j]
lat1[j]<-lat[j]
mes1[j]<-mes[j]
diff1[j]<-diff[j]
station1[j]<- NumeroStation
}
}

tab<-data.frame("lon"=lon1,"lat"=lat1,"mes"=mes1,"diff"=diff1,"station_retirée"=station1)

tab<-na.omit(tab)
coordinates(tab)<-c("lon","lat")

return(tab)

}

Base_Polair <- db_connect(DSN$polair, dbDriver("MySQL"))
```

# Bibliographie

- [1] ARNAUD M., EMERY X. (2000), *Estimation et interpolation spatiale*. Hermes, Paris, France.
- [2] BIVAND R., PEBESMA E., GOMES-RUBIO V. (2008), *Applied Spatial Data Analysis with R*. Springer, Baltimore, USA. 1ère édition.
- [3] BAILLARGEON S. (2005), Le krigeage : revue de la théorie et application à l'interpolation spatiale de données de précipitations. [En ligne], [http://newton.mat.ulaval.ca/theses/S-Baillargeon\\_05.pdf](http://newton.mat.ulaval.ca/theses/S-Baillargeon_05.pdf)
- [4] R Development Core Team (2009). *R : A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- [5] Pebesma, E.J. (2004). *Multivariable geostatistics in S : the gstat package*, Computers & Geosciences, 30 : 683-691. <http://cran.r-project.org/web/packages/gstat/>.
- [6] Pebesma, E.J. (2004). Package 'gstat'. [En ligne], <http://www.gstat.org/bin/gstat.pdf>.
- [7] Pebesma, E.J. (1999). Gstat user's manual. [En ligne], <http://www.gstat.org/gstat.pdf>.