

Construction optimale de réseaux de mesure: application à la surveillance des polluants aériens

Notes de cours de
l'École Nationale Supérieure des Techniques Avancées
ParisTech
Révision 1.13

Marc Bocquet
CEREA, École des Ponts ParisTech
Université Paris-Est et INRIA



Janvier 2006 - 19 janvier 2009

Table des matières

Remerciements	v
1 Généralités	1
1.1 Pourquoi un réseau de mesure	1
1.2 Optimalité des réseaux de mesure	2
2 Capacité du réseau à estimer : méthodes d'interpolation statistiques	7
2.1 Généralités	7
2.2 Un objectif récurrent : la capacité à interpoler	8
2.2.1 Techniques d'interpolation empiriques	8
2.2.2 Bref aperçu de la technique de krigeage	8
3 Mesure de l'utilité d'un réseau : trois approches différentes	13
3.1 Les méthodes exploratoires : distance, remplissage	13
3.2 Les méthodes historiquement liées à la régression	16
3.2.1 Classification des critères	17
3.2.2 Le théorème d'équivalence de Kiefer-Wolfowitz	18
3.2.3 Équivalence asymptotique des D- et G-optimalités	19
3.2.4 Calcul des fonctions de sensibilité	20
3.3 Les méthodes fondées sur le maximum d'entropie	21
3.3.1 Gain d'information	21
3.3.2 Lien avec l'entropie	22
3.3.3 Calcul de l'information mutuelle de deux réseaux à partir d'un modèle gaussien	23
4 Exemple d'application : surveillance des émanations d'un site industriel	27
4.1 Plan d'expérience	27
4.1.1 Modèle physique, interpolation et performance du réseau	27
4.1.2 Choix secondaires	29
4.2 Résultats	30
4.2.1 Influence du critère sur la conception du réseau	30
4.2.2 Influence de la physique : conditions météorologiques	30

Remerciements

Merci à Yelva Roustan pour les figures des réseaux de mesure de la qualité de l'air. Merci à Rachid Abida et Nikki Vercauteren pour de nombreuses discussions, et pour certaines figures. Merci également à Bruno Sportisse et Werner Müller, de l'Institut für angewandte Statistik, Autriche, pour leur soutien et leur conseils. Merci à l'Institut de Radioprotection et de Sécurité Nucléaire, dont le soutien financier a permis le lancement de l'axe de recherche *network design* au CEREa, et en particulier à Olivier Isnard et Anne Mathieu, du service des situations d'urgence et d'organisation de la crise. Enfin merci aux élèves de l'École Nationale des Techniques Avancées et aux élèves de l'École Nationale des Ponts et Chaussées, pour avoir subi plusieurs *réglages* sur ce cours, pour lequel il n'existe pas de manuel de référence.

Le CEREa est un laboratoire commun Ecole des Ponts ParisTech / EDF R&D. Le groupe assimilation de données / modélisation inverse, auquel se rattache l'activité *network design* est partie intégrante de l'équipe-projet CLIME commune à l'INRIA et l'ENPC.

Site web du CEREa :

<http://cerea.enpc.fr>.

Page groupe assimilation de données du CEREa :

http://cerea.enpc.fr/HomePages/bocquet/PageWebAssim/assim_fr.html.

Site web de l'auteur :

<http://cerea.enpc.fr/HomePages/bocquet>.

Chapitre 1

Généralités

La mesure de la qualité de l'air sur une région, c'est-à-dire la mesure de concentration en polluants (par exemple à l'échelle de l'Île de France), s'effectue principalement aux moyens de stations terrestres de mesure. Les polluants récurrents que l'on souhaite analyser au moyen de ces stations sont l'ozone (O_3), les oxydes d'azote (NO et NO_2), le dioxyde de soufre (SO_2), le monoxyde de carbone (CO), les particules fines (les PM_{10} , plus petites que $10\ \mu m$ et les $PM_{2.5}$, plus petites que $2.5\ \mu m$) et les composés organiques volatils (dérivés des hydrocarbures). Une spatialisation de l'analyse de l'état chimique de l'air au niveau du sol est obtenue par un réseau de stations dispersées sur la zone à instrumenter. Sur la figure 1.1, sont représentées les stations automatiques de mesure des polluants utilisées par l'agence AIRPARIF sur l'Île de France [5]. Près de 16000 mesures de concentrations sont ainsi collectées quotidiennement [2]. La fréquence de ces mesures est horaire.

AIRPARIF fait partie des associations agréées pour la surveillance de la qualité de l'air (AASQUA) (il y en a une quarantaine), qui couvrent le territoire français (en particulier les agglomérations de plus de 100000 habitants). Le réseau des stations des AASQUA [4] est représenté sur la figure 1.2. Les mesures des stations du réseau des AASQUA sont collectées dans la base de données BDQA (Banque de données sur la qualité de l'air), gérée par l'ADEME.

À l'échelle européenne existe le réseau EMEP, représenté sur la figure 1.3. EMEP est le programme de coopération européenne pour la surveillance et l'évaluation du transport à longue distance des polluants aériens. L'ensemble des stations régionales des agences régionales des pays européens ainsi que le réseau EMEP forment un réseau permettant d'abonder la base de données AIRBASE [6]. Le réseau correspondant est visualisé sur la figure 1.4.

1.1 Pourquoi un réseau de mesure

Il est clair que la surveillance de la qualité de l'air nécessite la constitution de réseaux de mesure. On peut invoquer plusieurs objectifs possibles pour ces réseaux :

1. Le contrôle de la qualité de l'air, c'est-à-dire le respect de certaines normes en concentrations qui ont été décrétées.
2. L'activation des mesures d'urgence rendue nécessaire par le dépassement de seuils critiques lors d'épisodes de pollution aigus.

3. La surveillance à terme de l'évolution des concentrations en polluants. En particulier il s'agit d'estimer l'impact de mesures de réduction des émissions.
4. Constitution d'une base de données, afin d'alimenter en données la recherche amont : l'étude de la sensibilité des concentrations de polluants à certains facteurs.

D'autres objectifs sont plus spécifiques :

1. Obtenir des données afin d'établir des prévisions à court terme (de un à cinq jours), grâce à l'assimilation de données.
2. Déterminer les concentrations représentatives dans les zones denses en population (impact sanitaire), ainsi que les pics de concentrations.
3. Déterminer les concentrations de fond de certaines espèces traces (métaux lourds, composés organiques volatils, polluants organiques persistants (POP)).
4. Déterminer le transport de polluants d'une région à l'autre (pollueur, pollué).

1.2 Optimalité des réseaux de mesure

L'optimalité de ces réseaux est un problème assez souvent occulté. Le réseau a pu se construire progressivement, sans plan d'ensemble ni stratégie globale. Les motivations sociologiques et politiques ont pu dominer les décisions d'implantation. On peut néanmoins s'attendre à ce qu'avec la rationalisation des réseaux et le coût financier élevé des stations et du traitement des données, la question prenne de l'ampleur.

Cependant qu'entend-on par réseau optimal ? La question est déterminante. En effet, un réseau est optimal pour un objectif assigné à un programme de surveillance permis par le réseau. Suivant l'objectif, un même réseau sera satisfaisant ou ne le sera pas.

Les critères d'optimalité de l'on peut en tirer résultent de la quantification de ces objectifs.

Leur formalisation (attribution d'un score numérique, un coût, à chaque réseau potentiel) est souvent ardue. Par exemple, il est naturel que la surveillance des concentrations près des zones denses en population et près des zones de forte émission soit prioritaire. (Il est clair que la répartition des stations sur la figure 1.1 suit plus ou moins, et par la force des choses, cette hiérarchisation.) Pour obtenir un critère numérique, il est donc nécessaire de quantifier ces hiérarchisations des priorités, ce qui n'a rien d'immédiat. Dans l'exemple évoqué, des complications surviennent si l'on prend conscience que la densité de population diffère entre jour et nuit, comme les émissions.

Les critères les plus fréquents sont ceux qui permettent de remplir les objectifs (tels que ceux décrits plus haut) assignés au réseau de surveillance. Ces critères se focalisent sur les statistiques nécessairement mises en œuvre pour l'interprétation des données fournies par le réseau. En effet pour remplir un ou plusieurs des objectifs sus-mentionnés, on devrait disposer d'une carte continue en espace et en temps des champs de concentrations. Ce qui n'est pas ce que fournit le réseau, parcimonieux par nature. De plus les mesures sont entachées d'erreur, ce qui complique l'estimation. Par conséquent, afin de remplir quelques uns des grands objectifs mentionnés ci-dessus, on doit s'assurer que plusieurs conditions de nature statistiques soient remplies. On peut citer quelques exemples parmi d'autres :

1. Les valeurs moyennes obtenues des données du réseau devraient pouvoir être estimées à une précision déterminée.

2. Les données recueillies doivent permettre l'interpolation des concentrations en tout point du domaine et à une précision suffisante.
3. Les maxima locaux de concentrations doivent pouvoir être soit observés directement soit diagnostiqués à partir des données du réseau, autant que possible.
4. La mise en berne d'une ou plusieurs stations doit avoir une influence limitée sur les performances du réseau.

Ce type de critères pour un réseau a pour objectif la *prévision spatiale*, c'est-à-dire, à un temps donné, reconstruire le champ de concentration au moyen des mesures collectées par le réseau.

Lorsque l'on dispose d'un modèle fin de dispersion des polluants surveillés, de nouveaux enjeux peuvent apparaître. On peut notamment chercher à déterminer un ou plusieurs des paramètres du modèle, ou bien des champs de forçage. Il s'agit donc d'inverser des paramètres à l'aide des données fournies par le réseau. On souhaite donc que le réseau fournissent des données telles que l'erreur moyenne commise lors de l'inversion soit réduite.

En conséquence la première étape de la construction d'un réseau de mesure consiste à :

1. se fixer les objectifs généraux du programme de surveillance attribué au réseau.
2. à définir les marges de tolérances pour les erreurs statistiques commises sur les observables lorsqu'on utilise les données du réseau.

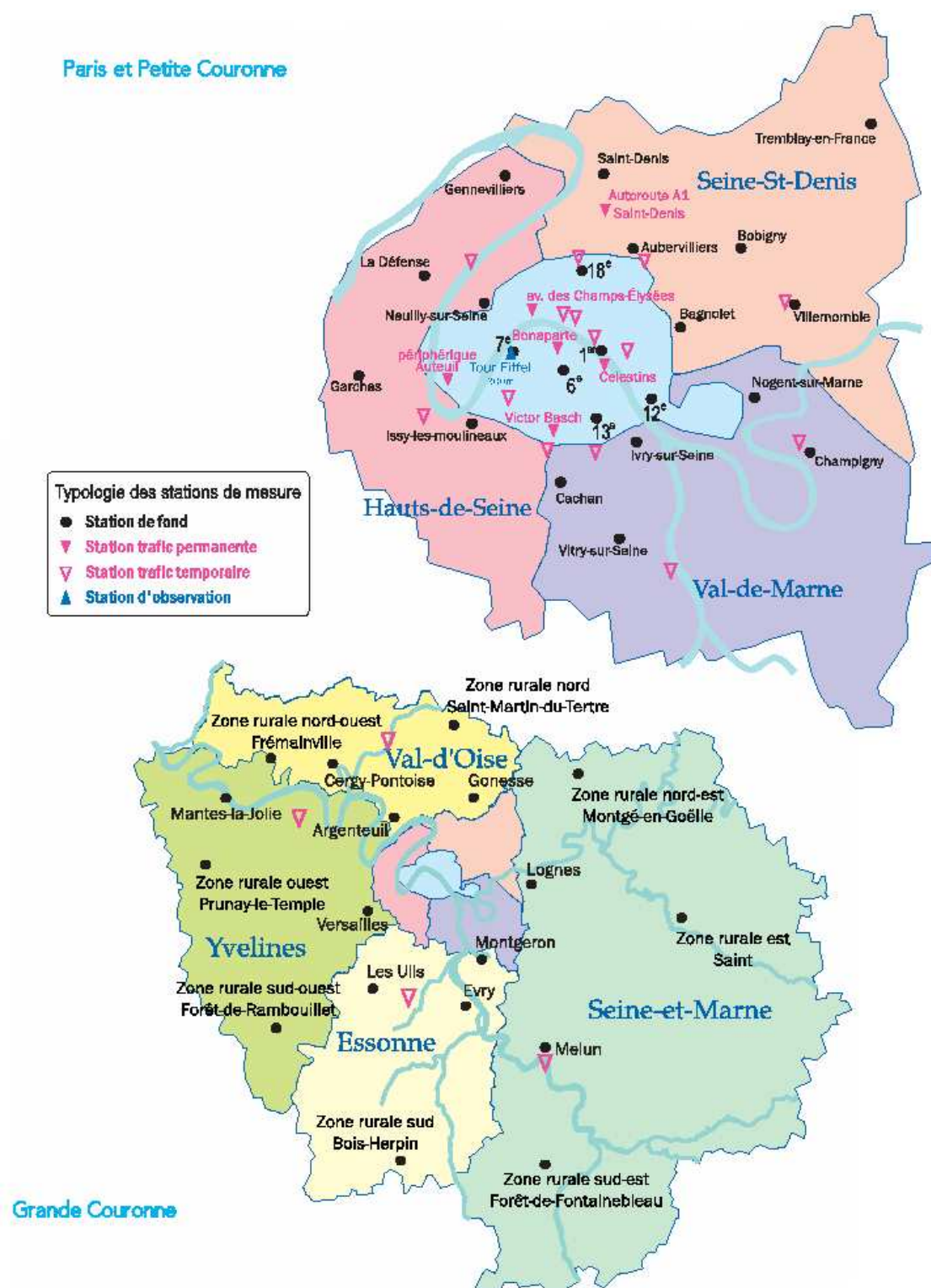


FIG. 1.1 – Disposition des stations de mesure utilisées par AIRPARIF, tiré du rapport d'activité 2007 d'AIRPARIF.

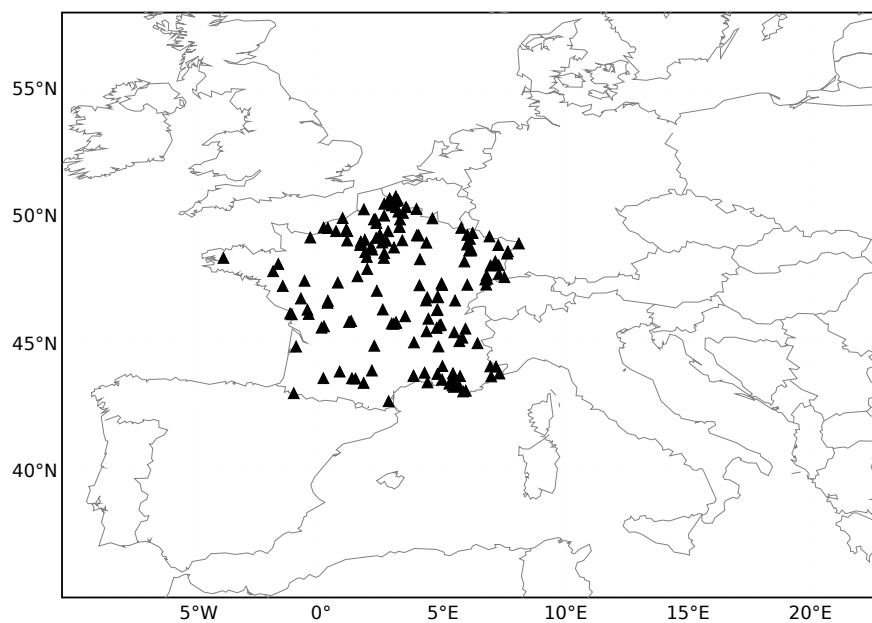


FIG. 1.2 – Disposition des stations de mesure utilisées par les AASQUA.

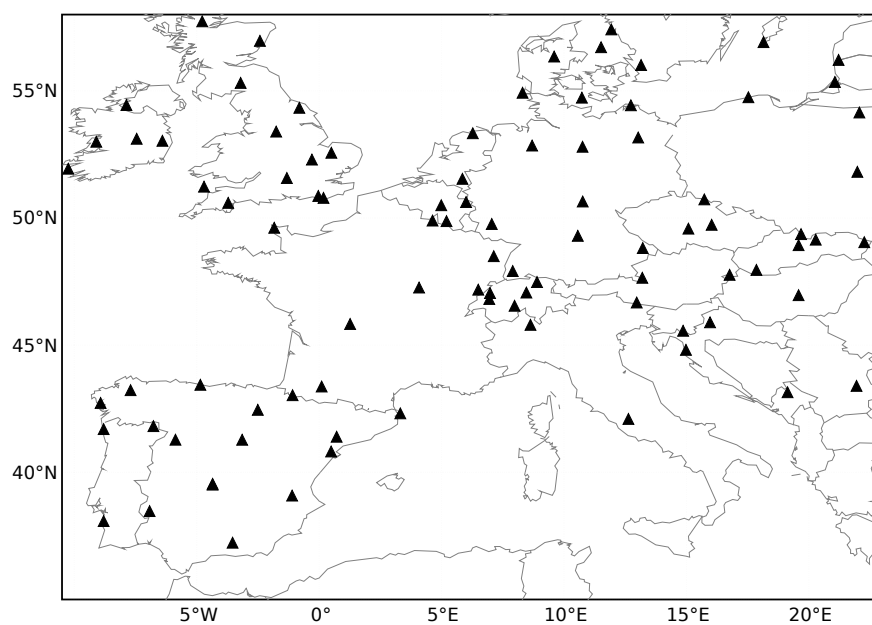


FIG. 1.3 – Disposition des stations de mesure utilisées par EMEP.

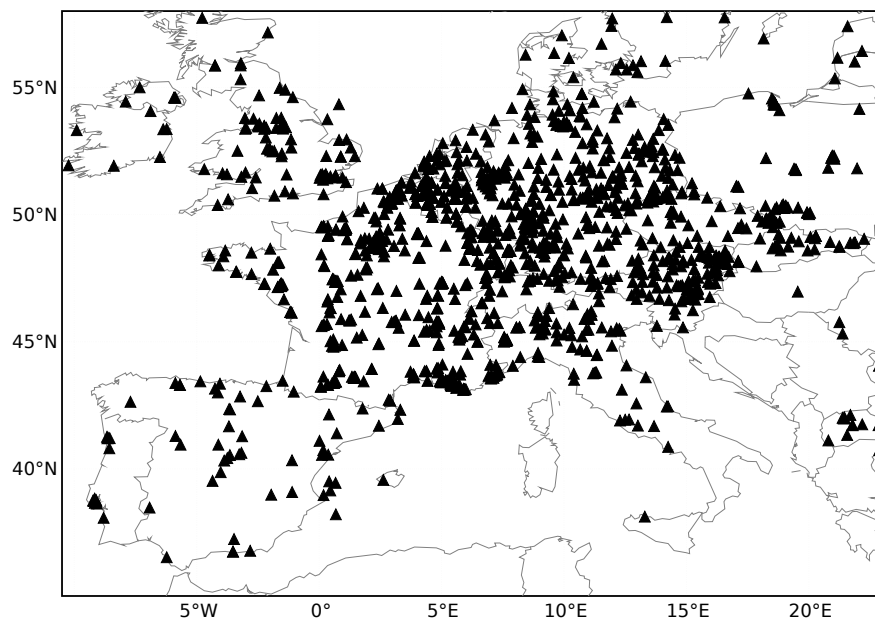


FIG. 1.4 – Disposition des stations de mesure utilisées par AIRBASE.

Chapitre 2

Capacité du réseau à estimer : méthodes d'interpolation statistiques

Dans cette section, on montre comment on peut définir une fonction coût particulière pour un réseau de mesure. Les fonctions coût correspondent à autant de critères de rationalisation du réseau. De nombreux choix sont possibles et légitimes. On ne donnera donc ici que quelques exemples. L'objectif étant ici de davantage détailler un choix parmi d'autres.

2.1 Généralités

On peut distinguer génériquement trois étapes successives dans la conception d'une fonction d'utilité pour un réseau :

1. Définir précisément ce que l'on veut obtenir du réseau : moyenne, extrêmes, seuils d'alerte, etc. Ces observables peuvent être distinctes des mesures directes sur le réseau.
2. Définir la manière (algorithme) d'obtenir ces quantités à partir des mesures sur le réseau.
3. Proposer un critère quantitatif d'évaluation de la performance du réseau au regard de son objectif défini à l'étape qui précède.

Il y a donc plusieurs choix successifs distincts à faire, lorsque l'on souhaite parler d'optimalité pour un réseau.

Considérons donc un réseau de n emplacements potentiels de stations de mesure. Chaque station effectue une mesure de concentration d'un polluant par intervalle de temps. Compte tenu ou non d'un existant pour le réseau, on distingue plusieurs options :

- On peut chercher à réduire un réseau déjà existant (soucis efficacité/coût).
- On peut chercher à augmenter un réseau, notamment un réseau qui aura été construit au fur et à mesure (besoins croissants, apports financiers successifs).
- On peut vouloir créer un réseau à partir de rien.

La troisième option (réseau *ab nihilo*) conduit à des difficultés, du fait de la complexité mathématique dans la résolution du problème. Elle est en pratique plus rare puisqu'elle

nécessite un investissement massif. Les deux premières options sont plus aisées à mettre en œuvre et sont rencontrées plus fréquemment. Mais ces deux premières approches conduisent à des formulations distinctes. En effet, pour rajouter un site à un réseau on recherchera une station dont les variances des mesures a priori est importante (par exemple la station de plus grande variances). Cela caractérise en effet une plus grande incertitude. Au contraire, pour ôter un site d'un réseau existant, on cherchera à retirer un site de faible variance (par exemple la station de plus petite variance).

2.2 Un objectif récurrent : la capacité à interpoler

On considère maintenant un cas intermédiaire où l'on cherche à sélectionner un certain nombre de stations parmi une sélection de sites prédéterminés. On suppose l'existence d'un tel ensemble de sites potentiels, car il est, au moins du point de vue combinatoire, plus difficile d'envisager de pouvoir positionner un site en tout point géographique (continuum de possibilité).

On suppose que le coût d'installation d'une station est égal quelque soit l'emplacement retenu. On dispose d'un budget permettant d'installer p stations, avec $p \leq n$.

Il s'agit donc de choisir au mieux p des n emplacements potentiels. Les emplacements retenus seront dits *instrumentés*, alors que les $n - p$ autres seront *non-instrumentés*. Connaissant les mesures effectuées aux p sites instrumentés, on peut chercher estimer les concentrations aux $n - p$ sites non-instrumentés. Un critère objectif pour la constitution du réseau consiste à demander à ce que l'erreur commise par l'estimation sur les sites non-instrumentés soit la plus faible possible.

Pour atteindre cet objectif, il faut être capable d'estimer des valeurs de concentration en des sites extérieurs au réseau instrumenté, et d'en estimer l'erreur attachée de façon a priori. Il existe de nombreuses méthodes, plus ou moins empiriques, répondant à cet objectif.

2.2.1 Techniques d'interpolation empiriques

Parmi les plus simples, on compte :

- Le point le plus proche : se décline sous la forme des polygones de Thiessen (ou de Voronoi, ou de Dirichlet). Par exemple ces polygones sont définis par $V_i = \{\mathbf{s} = (x, y) : \|\mathbf{s} - \mathbf{s}_i\| \leq \|\mathbf{s} - \mathbf{s}_j\|, \forall i \neq j\}$.
- La triangularisation (par exemple les triangles de Delaunay).
- Des combinaisons dont les pondérations sont liées à la distance aux points du réseau.
- La tendance (*trend surface*)

Une méthode alternative issue de la *géostatistique* est le *krigeage*, méthode qui a pour vocation de répondre à cette question [3] : donner une estimation d'un champ statistique en un point à partir de mesures spatialisées.

2.2.2 Bref aperçu de la technique de krigeage

On note \mathbf{Z} le vecteur des observations sur les p stations d'un réseau potentiel \mathcal{G} :

$$\mathbf{Z}^T = (Z(s_1), Z(s_2), \dots, Z(s_p)), \quad (2.1)$$

où \mathcal{G} est constitué des sites $\{s_1, s_2, \dots, s_p\}$. On note \mathbf{Y} le vecteur des concentrations réelles sur les $n - p$ stations complémentaires du réseau \mathcal{G} (ce réseau est noté \mathcal{U}) :

$$\mathbf{Y}^T = (Y(s_{p+1}), Y(s_{p+2}), \dots, Y(s_n)), \quad (2.2)$$

où $\mathcal{U} = \{s_{p+1}, s_{p+2}, \dots, s_n\}$. Enfin on note $\hat{\mathbf{Y}}$ le vecteur des estimations pour les concentrations sur les sites de \mathcal{U} .

Comme l'on cherche à déterminer une estimation sur plusieurs sites simultanément, la technique de krigeage porte le nom de *cokrigeage*.

On cherche à obtenir une estimation sur \mathcal{U} à partir des mesures sur \mathcal{G} qui soit optimale dans un sens à définir. Pour cela, on va faire appel à l'analyse dite **BLUE**, pour Best Linear Unbiased Estimator. On suppose qu'il est possible de déduire \mathbf{Y} à partir de \mathbf{Z} à partir du modèle :

$$\mathbf{Y} = \mathbf{L}\mathbf{Z} + \boldsymbol{\epsilon}. \quad (2.3)$$

où \mathbf{L} est une matrice $(n - p) \times p$ à déterminer. Cet Ansatz simple, mais néanmoins non-trivial, est un modèle de régression linéaire. $\boldsymbol{\epsilon}$ mesure l'erreur modèle commise. \mathbf{Y} et \mathbf{Z} sont vus comme des vecteurs aléatoires corrélés. Un estimateur est alors donné par

$$\hat{\mathbf{Y}} \equiv \mathbf{L}\mathbf{Z}. \quad (2.4)$$

On suppose que le modèle de régression est sans biais, c'est-à-dire $E[\boldsymbol{\epsilon}] = 0$. Il en résulte que $E[\hat{\mathbf{Y}}] = \mathbf{L}E[\mathbf{Z}]$. Et de là :

$$\mathbf{Y} - E[\mathbf{Y}] = \mathbf{L}(\mathbf{Z} - E[\mathbf{Z}]) + \boldsymbol{\epsilon}. \quad (2.5)$$

On note alors les écarts à la moyenne $\Delta\mathbf{Y} = \mathbf{Y} - E[\mathbf{Y}]$ et $\Delta\mathbf{Z} = \mathbf{Z} - E[\mathbf{Z}]$.

On veut choisir \mathbf{L} de sorte que l'erreur commise par le modèle de régression soit la plus faible possible, en moyenne. Le critère d'optimalité de l'analyse BLUE consiste à demander que la variance de l'estimateur soit minimale. On veut donc minimiser $E[\text{Tr}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T)]$. Or

$$\begin{aligned} E[\text{Tr}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T)] &= \text{Tr}\left(E\left[(\Delta\mathbf{Y} - \mathbf{L}\Delta\mathbf{Z})(\Delta\mathbf{Y} - \mathbf{L}\Delta\mathbf{Z})^T\right]\right) \\ &= \text{Tr}\left(E\left[\Delta\mathbf{Y}\Delta\mathbf{Y}^T\right] + \mathbf{L}E\left[\Delta\mathbf{Z}\Delta\mathbf{Z}^T\right]\mathbf{L} \right. \\ &\quad \left. - \mathbf{L}E\left[\Delta\mathbf{Z}\Delta\mathbf{Y}^T\right] - E\left[\Delta\mathbf{Y}\Delta\mathbf{Z}^T\right]\mathbf{L}^T\right). \end{aligned} \quad (2.6)$$

Pour poursuivre le raisonnement, il apparaît donc qu'il faut se donner les matrices de covariances :

- des mesures effectuées sur les sites instrumentés $\Sigma_{\mathbf{Z}\mathbf{Z}} = E[\Delta\mathbf{Z}\Delta\mathbf{Z}^T]$,
- des mesures issues des sites instrumentés avec les mesures issues des sites non-instrumentés : $\Sigma_{\mathbf{Y}\mathbf{Z}} = E[\Delta\mathbf{Y}\Delta\mathbf{Z}^T]$. On a aussi $\Sigma_{\mathbf{Z}\mathbf{Y}} = E[\Delta\mathbf{Z}\Delta\mathbf{Y}^T] = \Sigma_{\mathbf{Y}\mathbf{Z}}^T$.

C'est une information a priori, un modèle de nature statistique, qu'il est nécessaire de se donner, sur lequel on reviendra ensuite. Au minimum de la fonctionnelle, la variation de $\text{Tr}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T)$ est nulle au premier ordre :

$$0 = \text{Tr}\left(\delta\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T + \boldsymbol{\epsilon}\delta\boldsymbol{\epsilon}^T\right) = \text{Tr}\left(\delta\mathbf{L}\Sigma_{\mathbf{Z}\mathbf{Z}}\mathbf{L}^T + \mathbf{L}\Sigma_{\mathbf{Z}\mathbf{Z}}\delta\mathbf{L}^T - \delta\mathbf{L}\Sigma_{\mathbf{Z}\mathbf{Y}} - \Sigma_{\mathbf{Y}\mathbf{Z}}\delta\mathbf{L}^T\right) \quad (2.7)$$

soit encore $0 = 2\text{Tr}\left((\mathbf{L}\Sigma_{\mathbf{Z}\mathbf{Z}} - \Sigma_{\mathbf{Y}\mathbf{Z}})\delta\mathbf{L}^T\right)$. Il en résulte que

$$\mathbf{L} = \Sigma_{\mathbf{Y}\mathbf{Z}}\Sigma_{\mathbf{Z}\mathbf{Z}}^{-1}. \quad (2.8)$$

La détermination de \mathbf{L} achève la construction de l'estimateur BLUE

$$\hat{\mathbf{Y}} = \Sigma_{\mathbf{Y}\mathbf{Z}}\Sigma_{\mathbf{Z}\mathbf{Z}}^{-1}\mathbf{Z}. \quad (2.9)$$

La valeur du minimum de l'erreur moyenne vaut alors

$$\mathcal{E} = \min_{\mathbf{L}} \mathbb{E} [\text{Tr}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T)] = \text{Tr}(\Sigma_{\mathbf{Y}\mathbf{Y}} - \Sigma_{\mathbf{Y}\mathbf{Z}}\Sigma_{\mathbf{Z}\mathbf{Z}}^{-1}\Sigma_{\mathbf{Z}\mathbf{Y}}). \quad (2.10)$$

Donc si le réseau instrumenté \mathcal{G} est choisi, et si l'on décide d'utiliser la technique de krigeage issue de BLUE, alors il est possible de déterminer l'erreur moyenne associée à ce choix.

Un réseau optimal est donc une solution de la minimisation du critère \mathcal{E} :

$$\mathcal{G}^* = \underset{\mathcal{G}}{\text{argmin}} \mathcal{E}(\mathcal{G}). \quad (2.11)$$

\mathcal{G} varie sur toutes les configurations possibles de p stations parmi les n stations candidates.

En général, le calcul a priori des matrices de covariances $\Sigma_{\mathbf{X}\mathbf{X}}$, $\Sigma_{\mathbf{X}\mathbf{Y}}$ se fonde sur des statistiques passées. Les géostatisticiens utilisent peu directement les matrices de covariances (ou *corrélogrammes*), mais considèrent un objet qui lui est indirectement lié, la *semi-variance*. Elle est de la forme

$$\gamma(\mathbf{h}) = \frac{1}{2} \mathbb{E} \left[(Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h}))^2 \right] \quad (2.12)$$

Cette définition insiste sur la séparation spatiale entre deux sites. En développant le carré, on comprend que la semi-variance peut s'exprimer en fonction des éléments des matrices de covariance. Elle est préférée au corrélogramme car sa définition est plus générale (un semi-variogramme peut être bien défini quand le corrélogramme associé peut ne pas exister).

Il existe un certain nombre de modèles statistiques, dit modèle de *semivariogramme* qui fournissent une expression pour $\gamma(h)$. Par exemple le modèle exponentiel est de la forme

$$\gamma(h) = c_0 + c \left(1 - \exp\left(-\frac{h}{a}\right) \right), \quad (2.13)$$

où c_0 est le paramètre dit *effet de pépîte*, $c + c_0$ est le paramètre de *palier*, et a est la *portée*.

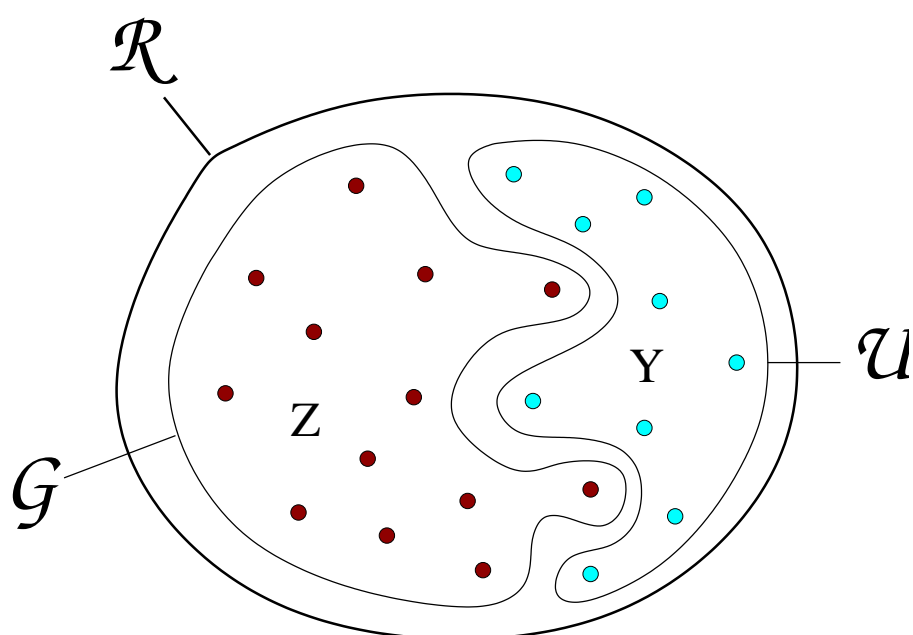


FIG. 2.1 – Représentation symbolique du réseau total \mathcal{R} , du réseau instrumenté \mathcal{G} et du réseau non-instrumenté \mathcal{U} .

Chapitre 3

Mesure de l'utilité d'un réseau : trois approches différentes

On présente ici trois approches couramment utilisées dans la construction optimale de réseaux, en particulier en science environnementale. Plus ou moins distinctement, ces trois démarches répondent toutes aux questions : utilité du réseau ? critère d'évaluation ?

3.1 Les méthodes exploratoires : distance, remplissage

La construction optimale d'un réseau est aujourd'hui considérablement aidée par la croissance vertigineuse de la puissance informatique. Néanmoins cela n'a pas toujours été le cas. De plus il subsiste de très nombreuses applications, notamment en qualité de l'air, où l'utilisation du modèle a un coût numérique important et l'utilisation de méthodes d'optimisation reste difficile. C'est pourquoi de nombreuses méthodes dites *exploratoires* ont été développées en ayant à l'esprit l'exigence de la simplicité. Ces méthodes sont donc généralement indépendantes de tout modèle explicite. On peut citer :

- La méthode de projection : il s'agit de plaquer un réseau prédéfini où les stations sont disposées de façon figée sur le domaine que l'on veut instrumenter (par exemple à l'aide d'une transformation conforme).
- Les méthodes stochastiques : on crée un réseau aléatoire dans le domaine par tirage. Ceci garantit une absence de biais dans la construction. En revanche, la méthode a contre elle qu'il est presque toujours possible de construire un réseau déterministe légèrement plus performant en retouchant localement quelques stations. C'est pourquoi elle n'a que peu de faveurs. Elle peut cependant servir à initialiser un algorithme de construction.
- Les méthodes de remplissage : elles permettent comme les deux méthodes précédentes d'engendrer facilement des réseaux ayant une très bonne couverture, mais la démarche est davantage raisonnée qu'auparavant. On rappelle que \mathcal{G} est le réseau des sites instrumentés alors que \mathcal{U} est le réseau des sites non-instrumentés. On considère ici plusieurs déclinaisons du concept :

1. La *U-optimalité* : pour tout réseau ξ_n comportant n stations, il s'agit de tester

son uniformité sur le territoire à instrumenter en définissant l'indicateur :

$$\mathcal{D}(\xi_n) = \left(\frac{1}{|\mathcal{U}|} \sum_{x \in \mathcal{U}} (d(x, \xi_n))^\alpha \right)^{1/\alpha}, \quad (3.1)$$

où $d(x, \xi_n)$ est la distance du point x non-instrumenté au réseau, c'est-à-dire $d(x, \xi_n) = \min_{y \in \xi_n} \|x - y\|$. $|\mathcal{U}|$ est le nombre de sites non-instrumentés. On recherche alors un réseau optimal ξ_n^* tel que

$$\xi_n^* = \operatorname{argmax}_{\xi_n} \mathcal{D}(\xi_n). \quad (3.2)$$

C'est un cas particulier d'une distance généralisée [16]

$$\mathcal{C}_{\alpha, \beta} = \left(\frac{1}{n} \sum_{x \in \mathcal{U}} (d_\beta(x, \xi_n))^\alpha \right)^{1/\alpha}, \quad (3.3)$$

avec

$$d_\beta(x, \xi_n) = \left(\frac{1}{n} \sum_{u \in \xi_n} \|x - u\|^\beta \right)^{1/\beta}. \quad (3.4)$$

Dans le cas présent $\mathcal{D}(\xi_n)$ correspond alors à la limite asymptotique β tend vers $-\infty$.

2. La ***S-optimalité*** : il s'agit pour tout réseau ξ_n comportant n stations de tester sa capacité à se répartir les uns par rapport aux autres de façon optimale :

$$\mathcal{D}(\xi_n) = \frac{n}{\sum_{x \in \xi_n} (d(x, \xi_n - x))^{-1}}, \quad (3.5)$$

où $\mathcal{D}(\xi_n)$ représente la moyenne harmonique des distances d'un point du réseau aux autres points du réseau. On recherche alors un réseau optimal ξ_n^* tel que

$$\xi_n^* = \operatorname{argmax}_{\xi_n} \mathcal{D}(\xi_n). \quad (3.6)$$

3. Distance ***minimax*** : il ne s'agit pas de tester l'uniformité des points du réseau sur l'ensemble du territoire à étudier mais de tester l'étalement des points du réseau les uns par rapport aux autres.

$$\xi_n^* = \operatorname{argmax}_{\xi_n} \left\{ \min_{x_i, x_j \in \xi_n} \|x_i - x_j\| \right\}. \quad (3.7)$$

Ce critère correspond au cas limite $\lim_{\alpha \rightarrow +\infty, \beta \rightarrow -\infty} \mathcal{C}_{\alpha, \beta}$. Sur la figure 3.1 est donné un exemple d'application du critère minimax.

4. Algorithme du "Cafetier" : Cette technique s'inspire la philosophie du maxmin, mais elle devient constructive. Selon cette technique, on ajoute chaque point du réseau l'un après l'autre. Lorsque l'on cherche le $(n + 1)$ -ème point du réseau, qui s'ajoute par récurrence au réseau à n points précédemment construit, on utilise :

$$x^{(n+1)} = \operatorname{argmax}_{x \in \Omega} \left\{ \min_{x_i \in \xi} \|x_i - x\| \right\}. \quad (3.8)$$

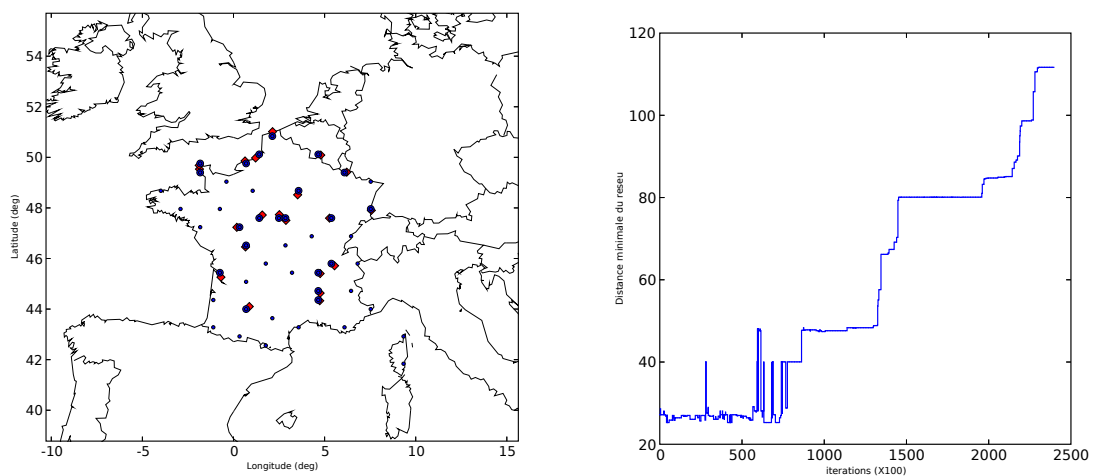


FIG. 3.1 – On cherche à optimiser un réseau de 46 stations. 21 stations sont fixes et sont implantées (fictivement) près des sites nucléaires civils français (centrales nucléaires, usines de retraitement). Les losanges représentent les sites nucléaires, les grands disques les stations fixes. Reste donc 25 stations dont la distribution est à optimiser. Le critère étant purement géométrique, aucun modèle physique n'intervient. À gauche est représenté un réseau proche de l'optimalité. À droite est tracé l'évolution du critère au fur et à mesure que l'algorithme (basé sur des méthodes stochastiques) progresse. Ce résultat est lié à une étude menée avec l'IRSN (projet Descartes).

Une deuxième raison pour lesquels les réseaux exploratoires sont prisés est qu'ils ne sont pas nécessairement plus mauvais que des réseaux construits sur des critères sophistiqués (confère [16]) sur la base de modèles statistiques ou de modèles numériques complexes déterministes.

3.2 Les méthodes historiquement liées à la régression

A priori la théorie de la *régression linéaire* est assez éloignée de la construction optimale de réseau. C'est pourtant de ce domaine qu'ont émergé les premiers travaux sur l'*optimal design theory*. Cette approche est développée dans les monographies [11, 8].

Le problème de la régression linéaire est le suivant. On considère une expérience dont les entrées sont les paramètres x_1, x_2, \dots, x_n , et dont les sorties sont les observations y_1, y_2, \dots, y_n . On dispose de p fonctions f_j sur lesquelles on souhaite décomposer ce jeu de mesures :

$$y_i = \sum_{j=1}^p f_j(x_i)\beta_j + \epsilon_i. \quad (3.9)$$

Les coefficients β_j ($j = 1, \dots, p$) sont à déterminer. Ils décident de la pondération de chacune des p fonctions de base f_j , connues a priori. ϵ_i représente l'erreur commise sur la sortie y_i . On suppose ces erreurs sans biais, décorrelées et homocédastiques d'écart-type σ . On définit la matrice $\mathbf{H} = [f_1, f_2, \dots, f_p]$ de taille $n \times p$:

$$\mathbf{H} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_p(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ f_1(x_n) & f_2(x_n) & \cdots & f_p(x_n) \end{pmatrix} \quad (3.10)$$

que l'on peut récrire à l'aide de n vecteurs colonnes h_i

$$h_i(x_i) = (f_1(x_i), f_2(x_i), \dots, f_p(x_i))^T, \quad (3.11)$$

qui ne dépendent chacun que d'une variable x_i . On a ainsi

$$\mathbf{H}^T = [h_1(x_1), h_2(x_2), \dots, h_n(x_n)]. \quad (3.12)$$

Puisqu'on peut distinguer les $h_i(x_i)$ par leur variable, on peut abandonner l'indice i avec $h(x_i)$. L'idée sous-jacente est d'introduire une dimension spatiale au problème. Les x_i sont alors interprétés comme la (les) coordonnée(s) d'un point x_i . La solution *normale* du problème de régression

$$\mathbf{y} = \mathbf{H}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (3.13)$$

est donnée par

$$\boldsymbol{\beta}^* = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}. \quad (3.14)$$

Elle s'obtient par la minimisation du problème variationnel :

$$\mathcal{J}(\boldsymbol{\beta}) = \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{H}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{H}\boldsymbol{\beta}). \quad (3.15)$$

Construction optimale de réseaux de mesure

$\mathbf{H}^T \mathbf{H} / \sigma^2$ s'identifie donc au Hessien du problème variationnel associé. Son inverse caractérise la variance du résultat, c'est-à-dire l'incertitude qui lui est associée. On cherche à réduire cette variance, en choisissant au mieux les points x_1, x_2, \dots, x_n . Comme les expressions mathématiques ne sont pas faciles à manipuler vues comme des fonctions de ces points, on introduit une mesure $\xi(x)$ positive, définie sur le domaine Ω et normalisée : $\int_{\Omega} d\xi(x) = 1$, de sorte que

$$\mathbf{H}^T \mathbf{H} = \sum_{i=1}^n h(x_i) h^T(x_i) = n \int_{\Omega} d\xi(x) h(x) h^T(x). \quad (3.16)$$

Ω est l'espace des points sur lequel est défini la mesure, donc le réseau. En terme de distributions de Dirac $\xi(x) = \frac{1}{n} \sum_{i=1}^n \delta(x - x_i)$. On note

$$M(\xi) = \int_{\Omega} d\xi(x) h(x) h^T(x). \quad (3.17)$$

Cette matrice de taille $p \times p$ est appelée **matrice d'information**. En tant que Hessien (à un facteur constant près : $\mathbf{H}^T \mathbf{H} = nM(\xi)$) du problème variationnel de régression associé, cette matrice caractérise la confiance que l'on porte dans le résultat de la régression. Dans un sens à préciser, plus la norme de cette matrice est grande, plus grande sera la qualité du réseau sous-jacent, qui permettra la plus grande réduction d'erreur.

3.2.1 Classification des critères

Il existe une classification de critères que l'on peut retenir pour optimiser ce réseau de points et qui se fonde sur la matrice d'information $M(\xi)$, dont l'inverse est, rappelons-le, directement proportionnel à la variance de la régression linéaire.

D-optimalité

La **D-optimalité** : Le critère d'optimalité a pour fonction coût $\Psi\{M(\xi)\} = -\ln |M(\xi)|$. Le déterminant $|M(\xi)|$ mesure le volume décrit par les directions principales de $M(\xi)$, c'est-à-dire le volume d'un ellipsoïde de confiance, à niveau de confiance donné. Le réseau optimal est donc solution de

$$\xi^* = \underset{\xi}{\operatorname{argmin}} \Psi\{M(\xi)\}. \quad (3.18)$$

En termes de valeurs propres de $M(\xi)$, le critère équivaut à minimiser $\prod_i \frac{1}{\lambda_i}$.

A-optimalité

La **A-optimalité** : $\Psi\{M(\xi)\} = \operatorname{Tr}(M^{-1}(\xi))$. Ce critère correspond à la variance moyenne sur les paramètres. En terme de valeurs propres de M , il s'agit de minimiser $\sum_i \frac{1}{\lambda_i}$.

E-optimalité

La **E-optimalité** : $\Psi\{M(\xi)\} = \max_{\|\mathbf{u}\|=1} \mathbf{u}^T M^{-1}(\xi) \mathbf{u}$. En terme de valeurs propres de M , il s'agit de minimiser le maximum des $\frac{1}{\lambda_i}$.

G-optimalité

La **G-optimalité** : $\Psi\{M(\xi)\} = \max_{x \in \Omega} d(x, \xi)$, où $d(x, \xi) = h(x)^T M(\xi)^{-1} h(x)$. Il est simple d'en déduire que $\sigma^2 d(x, \xi)$ est la variance de l'erreur associée à la réponse estimée $h(x)^T \beta$ à la position x , de sorte que le critère de G-optimalité est celui qui minimise la variance maximale sur toutes les positions x .

Comme $\beta^* = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}$, on a $\mathbf{y}^* = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}$, d'où l'on extrait la matrice d'influence :

$$\mathbf{A} = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T = \frac{1}{n} \begin{bmatrix} h^T(x_1) \\ h^T(x_2) \\ \vdots \\ h^T(x_n) \end{bmatrix} M(\xi)^{-1} [h(x_1) \quad h(x_2) \quad \cdots \quad h(x_n)] \quad (3.19)$$

Et donc en terme de variables de position continues :

$$A = \int_{\Omega} d\xi(x) \otimes d\xi(x') h^T(x) M(\xi)^{-1} h(x'). \quad (3.20)$$

La matrice \mathbf{A} , de taille $n \times n$, se décompose donc en des contributions dépendant des paires de point du réseau $A(x, x')$. Les scalaires $d(x, \xi)$ s'interprètent alors comme les contributions diagonales de la matrice d'influence.

La moyenne de $d(x, \xi)$ sur tout le réseau est

$$\begin{aligned} \int_{\Omega} d\xi d(x, \xi) &= \int_{\Omega} d\xi \text{Tr} \{ M(\xi)^{-1} h(x) h(x)^T \} \\ &= \text{Tr} \{ M(\xi)^{-1} M(\xi) \} = \text{Tr}(\mathbf{I}_p) = p. \end{aligned} \quad (3.21)$$

Ce qui implique qu'il existe au moins un point x tel que $d(x, \xi) \leq p$ et qu'il existe au moins un point x tel que $d(x, \xi) \geq p$. Il en résulte aussi qu'une configuration optimale est telle que $d(x, \xi) = p$ uniformément.

3.2.2 Le théorème d'équivalence de Kiefer-Wolfowitz

Il est rarement pratique d'optimiser la fonctionnelle $\Psi\{M(\xi)\}$. On préfère travailler sur la fonction de sensibilité de $\phi(x, \xi)$, qui est une sorte de dérivée de $\Psi\{M(\xi)\}$. Pour cela, on compare $\Psi\{M(\xi)\}$ évaluée en la mesure ξ et $\Psi\{M(\xi')\}$ évaluée en la mesure $\xi' = (1 - \alpha)\xi + \alpha\delta_x$. α est un paramètre tel que $0 < \alpha < 1$. δ_x est une mesure de Dirac au point physique x . Autrement dit, on rajoute le point x au réseau.

Il en résulte pour la matrice d'information évaluée sur cette nouvelle mesure :

$$M(\xi') = (1 - \alpha)M(\xi) + \alpha M(\delta_x). \quad (3.22)$$

La fonction de sensibilité ϕ est alors définie par

$$\phi(x, \xi) = \lim_{\alpha \rightarrow 0} \frac{1}{\alpha} [\Psi\{(1 - \alpha)M(\xi) + \alpha M(\delta_x)\} - \Psi\{M(\xi)\}]. \quad (3.23)$$

Le théorème *général* d'équivalence énonce alors que les trois conditions d'optimalité suivantes sont équivalentes [12] :

Construction optimale de réseaux de mesure

1. ξ^* minimise $\Psi \{M(\xi)\}$.
2. pour tout x , $\phi(x, \xi^*) \geq 0$.
3. $\phi(x, \xi^*) \geq 0$ atteint son minimum aux points appartenant au réseau optimum ξ^* , c'est-à-dire aux points où la mesure ξ^* est strictement positive.

3.2.3 Équivalence asymptotique des D- et G-optimalités

On considère un réseau ξ_n , constitué de points x_1, x_2, \dots, x_n , et de poids $1/n$ associé à chacun. Ce réseau permet la mesure de $\mathbf{y} = \mathbf{H}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, avec les mêmes notations que précédemment. En particulier $\mathbf{H}^T = [h(x_1), h(x_2), \dots, h(x_n)]$.

Le critère retenu pour juger du mérite du réseau est $\Psi(M(\xi_n)) = -\ln |M(\xi_n)|$ (c'est-à-dire la D-optimalité). On considère alors une modification du réseau comme suit :

$$\xi_{n+1} = \frac{n}{n+1}\xi_n + \frac{1}{n+1}\delta_x \quad (3.24)$$

On a donc rajouté un point x au réseau, auquel on confère le même poids que tout les autres. On a alors

$$\mathbf{H}_{n+1}^T \mathbf{H}_{n+1} = \mathbf{H}_n^T \mathbf{H}_n + h(x)h(x)^T. \quad (3.25)$$

Donc

$$\begin{aligned} |\mathbf{H}_{n+1}^T \mathbf{H}_{n+1}| &= |\mathbf{H}_n^T \mathbf{H}_n + h(x)h(x)^T| \\ &= |\mathbf{H}_n^T \mathbf{H}_n| | \mathbf{I} + (\mathbf{H}_n^T \mathbf{H}_n)^{-1} h(x)h(x)^T | \\ &= |\mathbf{H}_n^T \mathbf{H}_n| (1 + h(x)^T (\mathbf{H}_n^T \mathbf{H}_n)^{-1} h(x)) \end{aligned} \quad (3.26)$$

En passant de la deuxième à la troisième ligne, on a utilisé le fait que pour deux matrices \mathbf{A} et \mathbf{B} de taille respective $p \times n$ et $n \times p$, on a

$$\begin{aligned} |\mathbf{I}_p + \mathbf{A}\mathbf{B}| &= \exp(\ln |\mathbf{I}_p + \mathbf{A}\mathbf{B}|) = \exp(\text{Tr} \ln (\mathbf{I}_p + \mathbf{A}\mathbf{B})) \\ &= \exp \left(\text{Tr} \left\{ \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (\mathbf{A}\mathbf{B})^k \right\} \right) = \exp \left(\text{Tr} \left\{ \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (\mathbf{B}\mathbf{A})^k \right\} \right) \\ &= \exp(\text{Tr} \ln (\mathbf{I}_n + \mathbf{B}\mathbf{A})) = |\mathbf{I}_n + \mathbf{B}\mathbf{A}|. \end{aligned} \quad (3.27)$$

Comme $\mathbf{H}_n^T \mathbf{H}_n = nM(\xi_n)$, on a $|\mathbf{H}_n^T \mathbf{H}_n| = n^p |M(\xi_n)|$, et

$$\Psi(M(\xi_{n+1})) - p \ln(n+1) = \Psi(M(\xi_n)) - p \ln n - \ln \left(1 + \frac{1}{n} h^T(x) M(\xi_n)^{-1} h(x) \right). \quad (3.28)$$

Et de là

$$\begin{aligned} \lim_{n \rightarrow \infty} n [\Psi \{M(\xi_{n+1})\} - \Psi \{M(\xi_n)\}] &= \lim_{n \rightarrow \infty} np \ln \left\{ \frac{n+1}{n} \right\} - \lim_{n \rightarrow \infty} n \ln \left\{ 1 + \frac{d(x, \xi)}{n} \right\} \\ &= p - d(x, \xi_n). \end{aligned} \quad (3.29)$$

Pour un grand réseau ($n \gg 1$), cette limite doit être comprise comme la fonction de sensibilité ($\alpha = 1/n$). Par conséquent, pour ce réseau, et lorsque n est grand :

$$\phi(x, \xi) = p - d(x, \xi). \quad (3.30)$$

Pour que le réseau ξ soit optimal, cette fonction doit être positive pour tout x d'après le théorème d'équivalence. Il faut donc que $d(x, \xi) \leq p$ pour tout x . Or c'est la condition de G-optimalité. Dans ce cas il y a donc équivalence entre la D-optimalité et la G-optimalité. Plus généralement, le théorème d'équivalence (Kiefer-Wolfowitz), qui précède le théorème général d'équivalence, affirme qu'il y a une équivalence entre ces deux optimalités pour des mesures ξ continues [13].

3.2.4 Calcul des fonctions de sensibilité

Incidentement, grâce à l'exemple précédent, nous avons calculé la fonction de sensibilité associée au critère G , qui s'exprime en terme des fonctions de réponse. Un calcul plus direct permet d'obtenir

$$\begin{aligned}
& \Psi \{(1 - \alpha)M(\xi) + \alpha M(\delta_x)\} - \Psi \{M(\xi)\} \\
&= -\ln |(1 - \alpha)M(\xi) + \alpha M(\delta_x)| + \ln |M(\xi)| \\
&= -\ln |(1 - \alpha)M(\xi)| - \ln \left| I + \frac{\alpha}{1 - \alpha} M^{-1}(\xi) h(x) h^\dagger(x) \right| + \ln |M(\xi)| \\
&\simeq -\ln(1 - \alpha)^p - \text{Tr} \left(\frac{\alpha}{1 - \alpha} M^{-1}(\xi) h(x) h^\dagger(x) \right) \\
&\simeq p\alpha - \alpha h^\dagger(x) M^{-1}(\xi) h(x) \\
&\simeq \alpha (p - d(x, \xi)) ,
\end{aligned} \tag{3.31}$$

ce qui se traduit en

$$\phi_G(x, \xi) = p - d(x, \xi) . \tag{3.32}$$

On peut immédiatement utiliser conjointement ce résultat et le théorème général d'équivalence pour montrer l'équivalence des optimalités D et G.

Ce type de calcul peut également être mené pour d'autres critères. Prenons par exemple le cas du critère A :

$$\begin{aligned}
& \Psi \{(1 - \alpha)M(\xi) + \alpha M(\delta_x)\} - \Psi \{M(\xi)\} \\
&= \text{Tr} [(1 - \alpha)M(\xi) + \alpha M(\delta_x)]^{-1} - \text{Tr} (M^{-1}(\xi)) \\
&= \text{Tr} \left([(1 - \alpha)M(\xi)]^{-1} \left(I + \frac{\alpha}{1 - \alpha} M^{-1}(\xi) M(\delta_x) \right)^{-1} \right) - \text{Tr} (M^{-1}(\xi)) \\
&\simeq -\text{Tr} \left([(1 - \alpha)M(\xi)]^{-1} \frac{\alpha}{1 - \alpha} M^{-1}(\xi) h(x) h^\dagger(x) \right) - \text{Tr} (M^{-1}(\xi)) \\
&\quad + (1 - \alpha)^{-1} \text{Tr} (M^{-1}(\xi)) \\
&\simeq -\alpha h^\dagger(x) M^{-2}(\xi) h(x) + \alpha \text{Tr} (M^{-1}(\xi)) \\
&\simeq \alpha \left(-h^\dagger(x) M^{-2}(\xi) h(x) + \text{Tr} (M^{-1}(\xi)) \right) ,
\end{aligned} \tag{3.33}$$

ce qui se traduit en

$$\phi_A(x, \xi) = \text{Tr} (M^{-1}(\xi)) - h^\dagger(x) M^{-2}(\xi) h(x) . \tag{3.34}$$

3.3 Les méthodes fondées sur le maximum d'entropie

Plutôt que de chercher à réduire l'erreur d'un réseau de mesure, on va ici chercher à maximiser l'information que permet d'obtenir ce réseau de mesure. La théorie de l'information (initié par Shannon, suivant Boltzmann et Gibbs) attribue un contenu en information rigoureux à une densité de probabilité. En particulier, on sait calculer la différence de contenu en information entre deux densités, disons $p(x)$ et $q(x)$. Cette différence s'exprime par la formule de Kullback-Leibler :

$$K[p, q] = \int dx p(x) \ln \frac{p(x)}{q(x)}. \quad (3.35)$$

K n'est pas une distance au sens mathématique du terme ne serait-ce que parce que $K[p, q] \neq K[q, p]$ en général.

3.3.1 Gain d'information

Cette formule peut être appliquée au réseau instrumenté \mathcal{G} , dont les observations suivent la densité de probabilité $p(\mathbf{Z})$. Les mesures sur les sites non-instrumentés suivent la densité $p(\mathbf{Y})$. Le gain en information est donné par

$$\int d\mathbf{Y} p(\mathbf{Y}|\mathbf{Z}) \ln \frac{p(\mathbf{Y}|\mathbf{Z})}{p(\mathbf{Y})}, \quad (3.36)$$

où $p(\mathbf{Y})$ est la connaissance a priori sur les mesures du réseau non-instrumenté, et $p(\mathbf{Y}|\mathbf{Z})$ est la connaissance a posteriori obtenu sur les sites non-instrumentés \mathcal{U} à partir des mesures sur \mathcal{G} . Ce gain (toujours positif ou nul) est calculé pour un jeu de données \mathbf{Z} . Or on recherche la moyenne de l'information apportée par le réseau. C'est pourquoi la moyenne de l'information sur \mathbf{Z} doit être prise. On cherche alors à maximiser l'information :

$$\int d\mathbf{Z} p(\mathbf{Z}) \int d\mathbf{Y} p(\mathbf{Y}|\mathbf{Z}) \ln \frac{p(\mathbf{Y}|\mathbf{Z})}{p(\mathbf{Y})}. \quad (3.37)$$

Cette expression est en réalité symétrique pour les réseaux \mathcal{G} et \mathcal{U} . Il est en effet facile de constater qu'elle peut être réécrite

$$\begin{aligned} I[\mathbf{Y}, \mathbf{Z}] &= \int d\mathbf{Y} d\mathbf{Z} p(\mathbf{Z}, \mathbf{Y}) \ln \frac{p(\mathbf{Z}, \mathbf{Y})}{p(\mathbf{Y})p(\mathbf{Z})} \\ &= \int d\mathbf{Y} p(\mathbf{Y}) \int d\mathbf{Z} p(\mathbf{Z}|\mathbf{Y}) \ln \frac{p(\mathbf{Z}|\mathbf{Y})}{p(\mathbf{Z})} \end{aligned} \quad (3.38)$$

C'est l'*information mutuelle* des réseaux [10]. Le point de vue est donc symétrique : c'est aussi l'information moyenne que pourrait apporter le réseau \mathcal{U} sur le réseau \mathcal{G} .

Application à la construction *ab nihilo* d'un réseau On considère un ensemble de sites potentiels \mathcal{R} . On souhaite construire un réseau optimal \mathcal{G} inclus dans \mathcal{R} ($\mathcal{R} = \mathcal{U} + \mathcal{G}$.) On souhaite obtenir le réseau \mathcal{G} qui prédise au mieux les mesures sur le reste des sites (non-instrumentés) \mathcal{U} . Il s'agit donc de maximiser $I[\mathbf{Y}, \mathbf{Z}]$. Cette construction a un défaut, puisqu'elle ignore totalement l'information brute apportée par le réseau \mathcal{G} indépendamment de \mathcal{U} .

3.3.2 Lien avec l'entropie

L'entropie, qui caractérise le désordre, le contraire de l'information (mathématiquement son opposé), est définie comme

$$H_{\mathbf{X}}[X] = - \int d\mathbf{X} p(\mathbf{X}) \ln p(\mathbf{X}) . \quad (3.39)$$

Il est aisé de voir que l'information mutuelle Eq.(3.37) est égale à

$$I[\mathbf{Y}, \mathbf{Z}] = H_{\mathbf{Y}}[\mathbf{Y}] - E_{\mathbf{Z}} [H_{\mathbf{Y}}[\mathbf{Y}|\mathbf{Z}]] . \quad (3.40)$$

Application à la construction *ab nihilo* d'un réseau, bis On considère un ensemble de sites potentiels \mathcal{R} . On souhaite construire un réseau optimal \mathcal{G} inclus dans \mathcal{R} ($\mathcal{R} = \mathcal{U} + \mathcal{G}$.) On souhaite obtenir le réseau \mathcal{G} qui apporte le plus d'information possible, non plus seulement sur les sites de \mathcal{U} comme dans le cas d'application précédent mais aussi sur \mathcal{G} , donc en fait sur \mathcal{R} . L'ensemble des mesures prises sur les sites potentiels de \mathcal{R} est noté par le vecteur aléatoire \mathbf{X} . On souhaite minimiser la perte d'information obtenue en passant de \mathcal{R} à \mathcal{U} (donc toujours maximiser l'information obtenue). L'information obtenue entre la situation où on ne dispose que d'une information a priori sur \mathbf{X} , $p(\mathbf{X})$, et une situation où de l'information a été déduite sur \mathbf{X} , à partir de mesures \mathbf{Z} sur le réseau \mathcal{G} et notée $p(\mathbf{X}|\mathbf{Z})$ est :

$$\int d\mathbf{Y} p(\mathbf{X}|\mathbf{Z}) \ln \frac{p(\mathbf{X}|\mathbf{Z})}{p(\mathbf{X})} . \quad (3.41)$$

Pour évaluer l'information gagnée en moyenne, on doit moyenner sur les configurations \mathbf{Z} , pondérées par leur densité $p(\mathbf{Z})$, soit

$$\begin{aligned} \int d\mathbf{Z} p(\mathbf{Z}) \int d\mathbf{Y} p(\mathbf{X}|\mathbf{Z}) \ln \frac{p(\mathbf{X}|\mathbf{Z})}{p(\mathbf{X})} &= \int d\mathbf{Z} d\mathbf{Y} p(\mathbf{Z}) p(\mathbf{Y}|\mathbf{Z}) \ln \frac{p(\mathbf{Y}|\mathbf{Z})}{p(\mathbf{Y}|\mathbf{Z})p(\mathbf{Z})} \\ &= - \int d\mathbf{Z} p(\mathbf{Z}) \ln p(\mathbf{Z}) = H_{\mathbf{Z}}(\mathbf{Z}) . \end{aligned} \quad (3.42)$$

On en conclut qu'il est équivalent de maximiser l'entropie associé au réseau recherché. Cette fois-ci le critère tient compte de l'information lié au réseau instrumenté. On notera que ce raisonnement s'applique directement à la réduction d'un réseau \mathcal{R} à \mathcal{U} . Toutefois, cela ne peut être numériquement un bon critère, car il repose sur la manipulation d'un grand réseau \mathcal{U} , plutôt que sur le calcul de propriétés liées à un nombre de sites à retirer limité.

Application à l'extension d'un réseau Supposons que le réseau des sites instrumentés \mathcal{G} soit figé. On souhaite l'agrandir en lui ajoutant plusieurs stations \mathcal{U}' prises parmi un ensemble de stations \mathcal{U} . On veut alors maximiser l'information obtenue grâce aux mesures des stations de \mathcal{U}' , compte tenu des informations obtenues du réseau pré-existant \mathcal{G} . On souhaite donc maximiser $I[\mathbf{Y}', \mathbf{Z}]$. Puisque \mathcal{G} est figé, $H_{\mathbf{Z}}[\mathbf{Z}]$ est constant. Puisque

$$I[\mathbf{Y}', \mathbf{Z}] = H_{\mathbf{Z}}[\mathbf{Z}] - E_{\mathbf{Y}'} [H_{\mathbf{Z}}[\mathbf{Z}|\mathbf{Y}']] . \quad (3.43)$$

Il est donc équivalent de minimiser l'entropie conditionnelle $E_{\mathbf{Y}'} [H_{\mathbf{Z}}[\mathbf{Z}|\mathbf{Y}']]$. Cette construction ne souffre pas de défaut de principe.

Avec cette définition, on remarque que

$$H_{\mathbf{Y},\mathbf{Z}}[\mathbf{Y}, \mathbf{Z}] = H_{\mathbf{Z}}[\mathbf{Z}] + E_{\mathbf{Z}} [H_{\mathbf{Y}}[\mathbf{Y}|\mathbf{Z}]] \quad (3.44)$$

On peut prouver cette formule directement par

$$\begin{aligned} H_{\mathbf{Y},\mathbf{Z}}[\mathbf{Y}, \mathbf{Z}] &= - \int d\mathbf{Z} d\mathbf{Y} p(\mathbf{Z}, \mathbf{Y}) \ln p(\mathbf{Z}, \mathbf{Y}) \\ &= - \int d\mathbf{Z} d\mathbf{Y} p(\mathbf{Y}|\mathbf{Z})p(\mathbf{Z}) \ln p(\mathbf{Y}|\mathbf{Z})p(\mathbf{Z}) \\ &= - \int d\mathbf{Z} p(\mathbf{Z}) \ln p(\mathbf{Z}) - \int d\mathbf{Y} p(\mathbf{Y}) \int d\mathbf{Z} p(\mathbf{Y}|\mathbf{Z}) \ln p(\mathbf{Y}|\mathbf{Z}) \\ &= H_{\mathbf{Z}}[\mathbf{Z}] + E_{\mathbf{Z}} [H_{\mathbf{Y}}[\mathbf{Y}|\mathbf{Z}]] . \end{aligned} \quad (3.45)$$

On en déduit donc que

$$E_{\mathbf{Z}} [H_{\mathbf{Y}}[\mathbf{Y}|\mathbf{Z}]] = H_{\mathbf{Y},\mathbf{Z}}[\mathbf{Y}, \mathbf{Z}] - H_{\mathbf{Z}}[\mathbf{Z}] . \quad (3.46)$$

qui peut être exploitée pour des calculs analytiques (notamment pour des variables analytiques gaussiennes).

Application à la réduction d'un réseau On cherche à réduire un réseau \mathcal{R} , en éliminant les sites \mathcal{U} et en retenant les sites \mathcal{G} . On a vu qu'une solution optimale du point de vue de l'information, consiste à maximiser l'entropie de \mathcal{G} : $H_{\mathbf{Z}}[\mathbf{Z}]$. Mais cela est incommode numériquement. Or on a vu que :

$$H_{\mathbf{Z}}[\mathbf{Z}] = H_{\mathbf{X}}[\mathbf{X}] - E_{\mathbf{Z}} [H_{\mathbf{Y}}[\mathbf{Y}|\mathbf{Z}]] . \quad (3.47)$$

Comme l'entropie du réseau initial est fixée, il est équivalent de minimiser l'entropie conditionnelle $E_{\mathbf{Z}} [H_{\mathbf{Y}}[\mathbf{Y}|\mathbf{Z}]]$. On verra notamment dans le cas gaussien que cette formule présente l'avantage numérique de requérir des calculs dans l'espace des sites de \mathcal{U} et non de \mathcal{G} comme précédemment.

Application à la construction *ab nihilo* d'un réseau On considère un ensemble de sites potentiels \mathcal{R} . On souhaite construire un réseau optimal \mathcal{G} à partir de ces sites tels que $\mathcal{R} = \mathcal{U} + \mathcal{G}$. Quelque soit la répartition des sites potentiels entre les réseaux \mathcal{G} et \mathcal{U} , l'entropie globale $H_{\mathbf{Y},\mathbf{Z}}[\mathbf{Y}, \mathbf{Z}]$ associée au réseau \mathcal{R} est constante. Donc maximiser $-H_{\mathbf{Y},\mathbf{Z}}[\mathbf{Y}|\mathbf{Z}]$ est équivalent à maximiser $H_{\mathbf{Z}}[\mathbf{Z}]$. Il est donc équivalent de maximiser l'entropie associée aux mesures du réseau que l'on veut instrumenter.

3.3.3 Calcul de l'information mutuelle de deux réseaux à partir d'un modèle gaussien

On considère deux réseaux \mathcal{G} et \mathcal{U} , permettant les observations (vecteurs) \mathbf{Z} et \mathbf{Y} respectivement. Ces observations suivent un modèle statistique gaussien :

$$p(\mathbf{Y}, \mathbf{Z}) \sim N \left(0, \begin{pmatrix} \Sigma_{\mathbf{Y}\mathbf{Y}} & \Sigma_{\mathbf{Y}\mathbf{Z}} \\ \Sigma_{\mathbf{Z}\mathbf{Y}} & \Sigma_{\mathbf{Z}\mathbf{Z}} \end{pmatrix} \right) \quad (3.48)$$

L'information mutuelle partagée par les deux réseaux (reprise de Eq.(3.38)) vaut

$$I[\mathbf{Y}, \mathbf{Z}] = \int d\mathbf{Y} d\mathbf{Z} p(\mathbf{Z}, \mathbf{Y}) \ln \frac{p(\mathbf{Z}, \mathbf{Y})}{p(\mathbf{Y})p(\mathbf{Z})}. \quad (3.49)$$

On déduit de Eq.(3.48) les densités marginales :

$$p(\mathbf{Y}) \sim N(0, \Sigma_{\mathbf{Y}\mathbf{Y}}) \quad \text{et} \quad p(\mathbf{Z}) \sim N(0, \Sigma_{\mathbf{Z}\mathbf{Z}}). \quad (3.50)$$

L'information mutuelle se décompose en

$$\begin{aligned} I[\mathbf{Y}, \mathbf{Z}] &= \int d\mathbf{Y} d\mathbf{Z} p(\mathbf{Z}, \mathbf{Y}) \ln \frac{p(\mathbf{Z}, \mathbf{Y})}{p(\mathbf{Y})p(\mathbf{Z})} \\ &= \int d\mathbf{Y} d\mathbf{Z} p(\mathbf{Z}, \mathbf{Y}) \ln p(\mathbf{Z}, \mathbf{Y}) \\ &\quad - \int d\mathbf{Y} p(\mathbf{Y}) \ln p(\mathbf{Y}) - \int d\mathbf{Z} p(\mathbf{Z}) \ln p(\mathbf{Z}) \\ &= H_{\mathbf{Y}}[\mathbf{Y}] + H_{\mathbf{Z}}[\mathbf{Z}] - H_{\mathbf{Y}, \mathbf{Z}}[\mathbf{Y}, \mathbf{Z}]. \end{aligned} \quad (3.51)$$

On rappelle que l'entropie s'écrit $H_{\mathbf{X}}[\mathbf{X}] = - \int d\mathbf{X} p(\mathbf{X}) \ln p(\mathbf{X})$. Pour calculer l'information mutuelle, il faut donc calculer l'entropie des trois densités gaussiennes $p(\mathbf{Y}, \mathbf{Z})$, $p(\mathbf{X})$ et $p(\mathbf{Y})$.

Génériquement, considérons la distribution gaussienne normalisée ($\sim N(0, \mathbf{P})$)

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{P}|^{n/2}} \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{P}^{-1} \mathbf{x}\right), \quad (3.52)$$

Alors un calcul élémentaire d'intégrale gaussienne donne :

$$\begin{aligned} \int d\mathbf{x} p(\mathbf{x}) \ln p(\mathbf{x}) &= \int d\mathbf{x} p(\mathbf{x}) \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{P}^{-1} \mathbf{x} - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\mathbf{P}| \right\} \\ &= -\frac{1}{2} \text{Tr}(\mathbf{I}) - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\mathbf{P}| \\ &= -\frac{n}{2} (1 + \ln(2\pi)) - \frac{1}{2} \ln |\mathbf{P}| \end{aligned} \quad (3.53)$$

On en déduit donc pour le calcul de l'information mutuelle que

$$I[\mathbf{Y}, \mathbf{Z}] = -\frac{1}{2} \ln |\Sigma_{\mathbf{Y}\mathbf{Y}}| - \frac{1}{2} \ln |\Sigma_{\mathbf{Z}\mathbf{Z}}| + \frac{1}{2} \ln \left| \begin{array}{cc} \Sigma_{\mathbf{Y}\mathbf{Y}} & \Sigma_{\mathbf{Y}\mathbf{Z}} \\ \Sigma_{\mathbf{Z}\mathbf{Y}} & \Sigma_{\mathbf{Z}\mathbf{Z}} \end{array} \right| \quad (3.54)$$

Mais on a (utilisant la formule du déterminant d'une matrice bloc)

$$\left| \begin{array}{cc} \Sigma_{\mathbf{Y}\mathbf{Y}} & \Sigma_{\mathbf{Y}\mathbf{Z}} \\ \Sigma_{\mathbf{Z}\mathbf{Y}} & \Sigma_{\mathbf{Z}\mathbf{Z}} \end{array} \right| = |\Sigma_{\mathbf{Z}\mathbf{Z}}| |\Sigma_{\mathbf{Y}\mathbf{Y}} - \Sigma_{\mathbf{Y}\mathbf{Z}} \Sigma_{\mathbf{Z}\mathbf{Z}}^{-1} \Sigma_{\mathbf{Z}\mathbf{Y}}| \quad (3.55)$$

et donc

$$I[\mathbf{Y}, \mathbf{Z}] = \frac{1}{2} \ln |\mathbf{I} - \Sigma_{\mathbf{Y}\mathbf{Z}} \Sigma_{\mathbf{Z}\mathbf{Z}}^{-1} \Sigma_{\mathbf{Z}\mathbf{Y}} \Sigma_{\mathbf{Y}\mathbf{Y}}^{-1}|. \quad (3.56)$$

Ce calcul est utile pour calculer à moindres frais l'entropie $H_{\mathbf{Y}, \mathbf{Z}}[\mathbf{Y}|\mathbf{Z}]$. En effet

$$\mathbb{E}_{\mathbf{Z}} [H_{\mathbf{Y}}[\mathbf{Y}|\mathbf{Z}]] = H_{\mathbf{Y}}[\mathbf{Y}] - I[\mathbf{Y}, \mathbf{Z}]. \quad (3.57)$$

Construction optimale de réseaux de mesure

On en déduit donc que

$$E_{\mathbf{Z}} [H_{\mathbf{Y}}[\mathbf{Y}|\mathbf{Z}]] = -\frac{n-p}{2}(1 + \ln(2\pi)) + \frac{1}{2} \ln |\Sigma_{\mathbf{Y}\mathbf{Y}} - \Sigma_{\mathbf{Y}\mathbf{Z}}\Sigma_{\mathbf{Z}\mathbf{Z}}^{-1}\Sigma_{\mathbf{Z}\mathbf{Y}}|. \quad (3.58)$$

Ceci est à rapprocher du résultat Eq.(2.10). Par conséquent le critère fondé sur la maximisation de l'entropie est proche de la D-optimalité. Il est équivalent dans le cas gaussien. Le critère fondé sur la variance minimum du type Eq.(2.11), est par opposition proche de l'A-optimalité.

Chapitre 4

Exemple d'application : surveillance des émanations d'un site industriel

Dans cette section, on cherche à illustrer la construction optimale d'un réseau de mesure sur un exemple simple mais non trivial. On s'intéresse à la petite échelle (encore appelé champ proche). Il s'agit de la surveillance dans un rayon de quelques centaines de mètres à quelques kilomètres autour d'un site industriel (usine chimique, centrale nucléaire). Il est demandé de concevoir un réseau de mesure permettant la détection d'un polluant en situation accidentelle. De plus le réseau doit permettre de reconstituer la carte du nuage de polluant, à partir d'une méthode d'interpolation adéquate. On dispose de la rose des vents locaux (distribution statistique de la direction et de l'intensité du vent sur le site). Dans la suite, on supposera toutefois pour simplifier que l'intensité du vent est fixe. On se placera dans deux cas de figures : vent modéré (3 m s^{-1}) et vent fort (10 m s^{-1}). De plus, on supposera que la distribution angulaire est soit uniforme, soit piquée sur une direction (vent dominant).

4.1 Plan d'expérience

Pour réaliser l'expérience, on a besoin d'un modèle numérique qui soit une représentation de la réalité des phénomènes physiques. On espère que l'erreur modèle est faible, mais la simplicité du modèle est aussi un critère déterminant, car le modèle n'est qu'une des pièces du plan d'expérience.

4.1.1 Modèle physique, interpolation et performance du réseau

Le modèle physique de dispersion

La dispersion du polluant sera représentée par un modèle gaussien. On suppose que le rejet atteint rapidement un régime permanent. Le système de coordonnées spatiales est choisi pour chaque accident de telle manière que Ox désigne la direction du vent, après rotation appropriée du système de coordonnées. On note u l'intensité moyenne du vent. En un point (x, y, z) , on note $\tau = x/u$, le temps de transfert. Alors le modèle gaussien de

τ	A_h	k_h	A_z	k_z
$\tau < 240$ s	0.405	0.859	0.42	0.814
240 s $< \tau < 3280$ s	0.135	1.13	1	0.685
3280 s $< \tau < 97000$ s	0.135	1.13	20	0.5
97000 s $< \tau < 508000$ s	0.463	1	20	0.5
508000 s $< \tau < 1.3 \cdot 10^6$ s	6.5	0.824	20	0.5
$\tau > 1.3 \cdot 10^6$	$2 \cdot 10^5$	0.5	20	0.5

TAB. 4.1 – Tableau des paramètres de turbulence servant au calcul des coefficients de Doury. Conditions de diffusion normales (gradient de température inférieur à $-0.5^\circ/100$ m).

Doury estime le champ de concentration par

$$c(x, y, z) = \frac{1}{2\pi\sigma_h\sigma_z} \exp\left(-\frac{y^2}{2\sigma_h^2}\right) \times \left\{ \exp\left(-\frac{(z-h)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+h)^2}{2\sigma_z^2}\right) \right\}. \quad (4.1)$$

où les coefficients de Doury (1976) sont donnés par

$$\sigma_h = (A_h\tau)^{k_h} \quad \sigma_z = (A_z\tau)^{k_z}. \quad (4.2)$$

Les paramètres A_h , A_z , k_h et k_z sont lus dans le tableau 4.1.1. h est la hauteur du rejet.

Le premier facteur exponentiel décrit la diffusion transversale à l'axe du panache gaussien. Les deux exponentielles présentes dans le dernier terme décrivent la diffusion verticale du panache et son rebond au sol (la deuxième exponentielle est obtenue par la méthode des images).

Les figures 4.1 représentent des résultats d'expériences réalisées dans la soufflerie de l'École Centrale de Lyon sur une maquette du site de la centrale nucléaire du Bugey. Une source de traceur (un hydrocarbure qui autorise une mesure précise) est relâché à partir d'une des enceintes. Des isocontours du champ de concentration au sol sont représentés. Le champ de concentration présente une forme similaire au champ engendré par un panache gaussien. Le modèle gaussien, aussi rudimentaire soit-il, reste donc un outil simple et efficace pour la dispersion en champ proche. Ceci justifie son utilisation dans ce chapitre.

Fonctionnalité du réseau : interpolation du champ de concentration

Comme méthode d'interpolation, on retient le krigeage simple. On se contente d'appliquer la technique décrite dans la section 2.2.2.

Critère d'optimalité du réseau

À partir des mesures \mathbf{Z} sur les sites instrumentés, la méthode d'interpolation prédit les concentrations \mathbf{Y}^{prev} sur les stations non-instrumentées de \mathcal{U} . Ces concentrations sont comparées aux concentrations observées (par le modèle numérique, représentation de la réalité) \mathbf{Y}^{obs} . On retient ici trois critères pour mesurer cette différence. Le critère de norme l_∞ et de norme l_2

$$\mathcal{L}_\infty = \sup_{s \in \mathcal{U}} (|Y_s^{\text{prev}} - Y_s^{\text{obs}}|) \quad \mathcal{L}_2 = \left(\frac{1}{|\mathcal{U}|} \sum_{s \in \mathcal{U}} |Y_s^{\text{prev}} - Y_s^{\text{obs}}|^2 \right)^{1/2}, \quad (4.3)$$

Construction optimale de réseaux de mesure

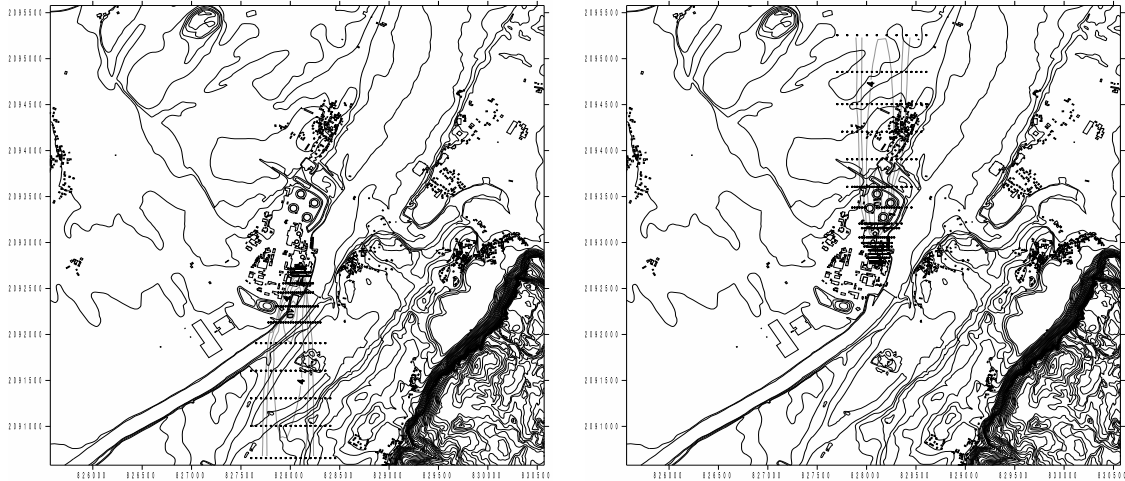


FIG. 4.1 – Isocontours du champ de concentrations au sol du traceur rejeté par une des enceintes miniature de la maquette de la centrale du Bugey placée dans la veine de la soufflerie de l'École Centrale de Lyon. Les deux images correspondent à une situation vent du Nord et une situation vent du Sud.

où $|\mathcal{U}|$ est le cardinal de \mathcal{U} , l'ensemble des sites potentiels non-instrumentés, et le critère de biais fractionnel

$$\mathcal{L}_{\text{FB}} = \frac{2}{|\mathcal{U}|} \sum_{s \in \mathcal{U}} \frac{|Y_s^{\text{prev}} - Y_s^{\text{obs}}|}{|Y_s^{\text{prev}} + Y_s^{\text{obs}}|}. \quad (4.4)$$

4.1.2 Choix secondaires

Nombre de points du réseau et nombre de configurations

On cherche à construire un réseau de 60 stations. On considère un réseau potentiel de 120 stations réparties en cercles concentriques autour du site. Cette manière de procéder réduit considérablement le champ d'investigation, puisque 120 sites sont concernés au lieu d'un continuum de possibilités. Toutefois, l'optimisation reste une opération combinatoirement très difficile puisque qu'il y a $\binom{120}{60}$ réseaux distincts possibles. À l'aide de la formule de Stirling, il est facile de montrer que lorsque n est suffisamment grand

$$\binom{n}{n/2} \sim \sqrt{\frac{2}{\pi n}} 2^n. \quad (4.5)$$

Pour $n = 120$, on obtient environ $9.7 \cdot 10^{34}$ configurations possibles! Seules des méthodes stochastiques permettent d'explorer (relativement) efficacement ce vaste espace de configurations. On considérera aussi des cas où 20 stations sont à choisir sur un total de 40 sites potentiels (soit un total d'environ $1.4 \cdot 10^{11}$ configurations possibles).

Algorithme d'optimisation

La méthode d'optimisation que l'on retient est le *recuit simulé* (*simulated annealing* en anglais [14]). Il s'agit d'une méthode stochastique développée en physique statistique

dans les années 80 et qui dérive de l'algorithme de Métropolis (c'est un algorithme de type Monte-Carlo). On part d'un réseau donné possédant le nombre de stations voulu (ici 60). La capacité du réseau à interpoler le champ de concentration sur les 120 sites potentiels à partir des mesures effectuées sur les 60 sites instrumentés est ensuite notée. La note obtenue vaut E_a . Elle mesure l'écart entre les concentrations prédites et les concentrations observées. Donc plus la note est faible, meilleur est le réseau. Puis on choisit aléatoirement une station du réseau courant que l'on va exclure et on choisit au hasard une station parmi les sites potentiels non instrumentés. On obtient donc un nouveau réseau qui possède toujours 60 stations. On peut alors noter ce nouveau réseau : E_b . Alors

- si $E_b < E_a$, on accepte le nouveau réseau.
- si $E_b > E_a$, on accepte le nouveau réseau avec la probabilité $e^{-(E_b-E_a)/T}$, sinon on le rejette et on revient au réseau précédent.

$e^{-(E_b-E_a)/T}$ est un facteur de Boltzmann, où le paramètre T est interprété comme une mesure de la température du système. On itère alors le processus.

Afin d'amener le système d'optimisation vers un réseau satisfaisant (de score le plus faible possible), on diminue progressivement la température T , au fur et à mesure des itérations. Il existe de nombreux schémas de refroidissement, prescrivant la décroissance de T . Des schémas de refroidissement populaires sont

- une décroissance exponentielle : $T_k = T_0 (T_f/T_0)^k$,
- le schéma de Lundy : $T_{k+1} = T_k/(1 + aT_k)$

En général, il n'y a pas de garantie pour qu'un tel algorithme converge, sauf si le système est refroidi très progressivement, par exemple de façon logarithmique ($T_k \geq a/\ln(b+k)$). À cause du temps de calcul, c'est rarement acceptable en pratique. C'est pourquoi dans les conditions d'utilisation usuelles du recuit simulé, il n'y a pas de garantie de convergence. Toutefois, l'absence de convergence n'exclut en rien que

- le réseau optimal ait été atteint au moins une fois, lors de l'exploration stochastique,
- ou qu'un réseau proche de l'optimalité (c'est-à-dire suffisant en pratique) ait été atteint.

Un exemple de recuit simulé est donné sur la figure 4.2.

4.2 Résultats

4.2.1 Influence du critère sur la conception du réseau

Comme on peut le constater sur la figure 4.3, le choix du critère a une influence significative sur le réseau optimal. Cela est surtout visible sur la configuration de vent dominant. Selon que l'on opte pour un critère construit sur la norme l_∞ , ou sur le biais fractionnel les stations du réseau sont davantage positionnées en amont ou en aval du rejet typique ! Le choix du critère peut donc, selon la physique, devenir crucial.

4.2.2 Influence de la physique : conditions météorologiques

La physique qui sous-entend le problème (dispersion de polluants ou de radionucléides) est aussi cruciale dans la détermination du réseau. On l'a remarqué sur la figure 4.3 pour deux distributions des vents très distinctes. Mais une telle sensibilité existe également lorsque l'on fait varier d'autres paramètres physiques, comme l'intensité du vent (à distribution angulaire égale).

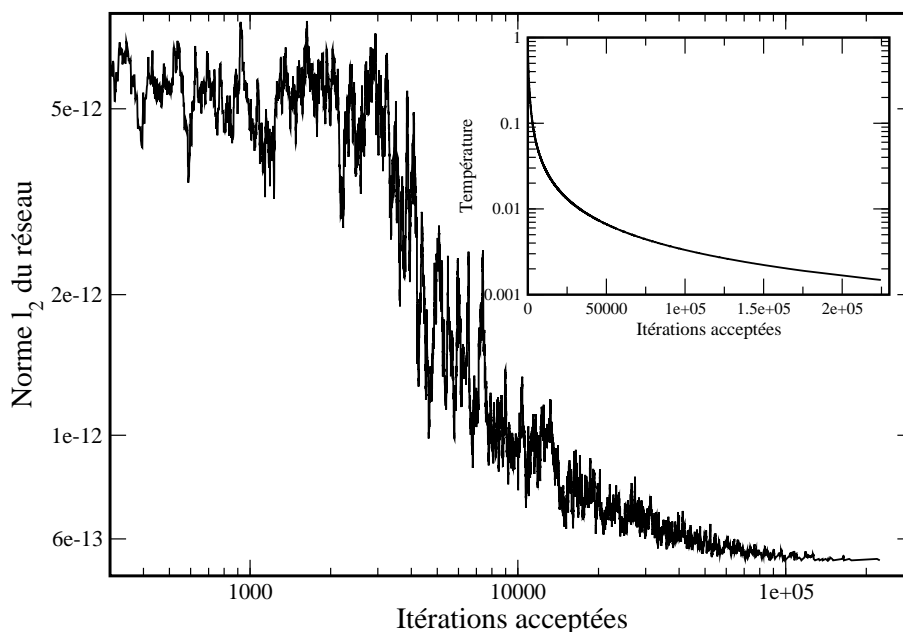


FIG. 4.2 – Évolution de la note du réseau lors du recuit simulé. En insert est représenté la décroissance de la température qui suit un schéma de Lundy.

On optimise un réseau de 60 stations pour 120 sites potentiels. Dans un premier cas le vent souffle à 3 m s^{-1} . Le réseau optimal est alors dispersé afin de bien appréhender la dispersion transversale du panache (figure de gauche de 4.4). Dans un second cas le vent souffle à 10 m s^{-1} . Le réseau optimal se densifie au plus près du site, de façon à ce que le panache, beaucoup plus fin, ne puisse échapper à la surveillance du réseau (figure de droite de 4.4).

Erreur commise par le réseau optimal (vent fort)

Sur la figure 4.5 est étudiée la réponse du réseau et du système qui en exploite les données (krigeage simple) sur un rejet spécifique sous un vent d'intensité forte. Dans ce premier cas, le réseau est aussi dense que possible autour du site. Le vent est fort et le panache est fin. Il s'agit donc d'éviter que celui-ci ne passe entre les capteurs.

Erreur commise par le réseau optimal (vent faible)

Sur la figure 4.6 est étudiée la réponse du réseau et du système qui en exploite les données (krigeage simple) sur un rejet spécifique sous un vent d'intensité faible. Dans ce deuxième cas, le panache est suffisamment large pour que l'algorithme tende vers un réseau plus relâché. Le panache sera toujours détecté et il s'agit de l'évaluer, au mieux.

La zone intérieure au site industriel n'est pas instrumentée. Les erreurs commises dans cette région par le krigeage simple ne sont pas pertinentes.

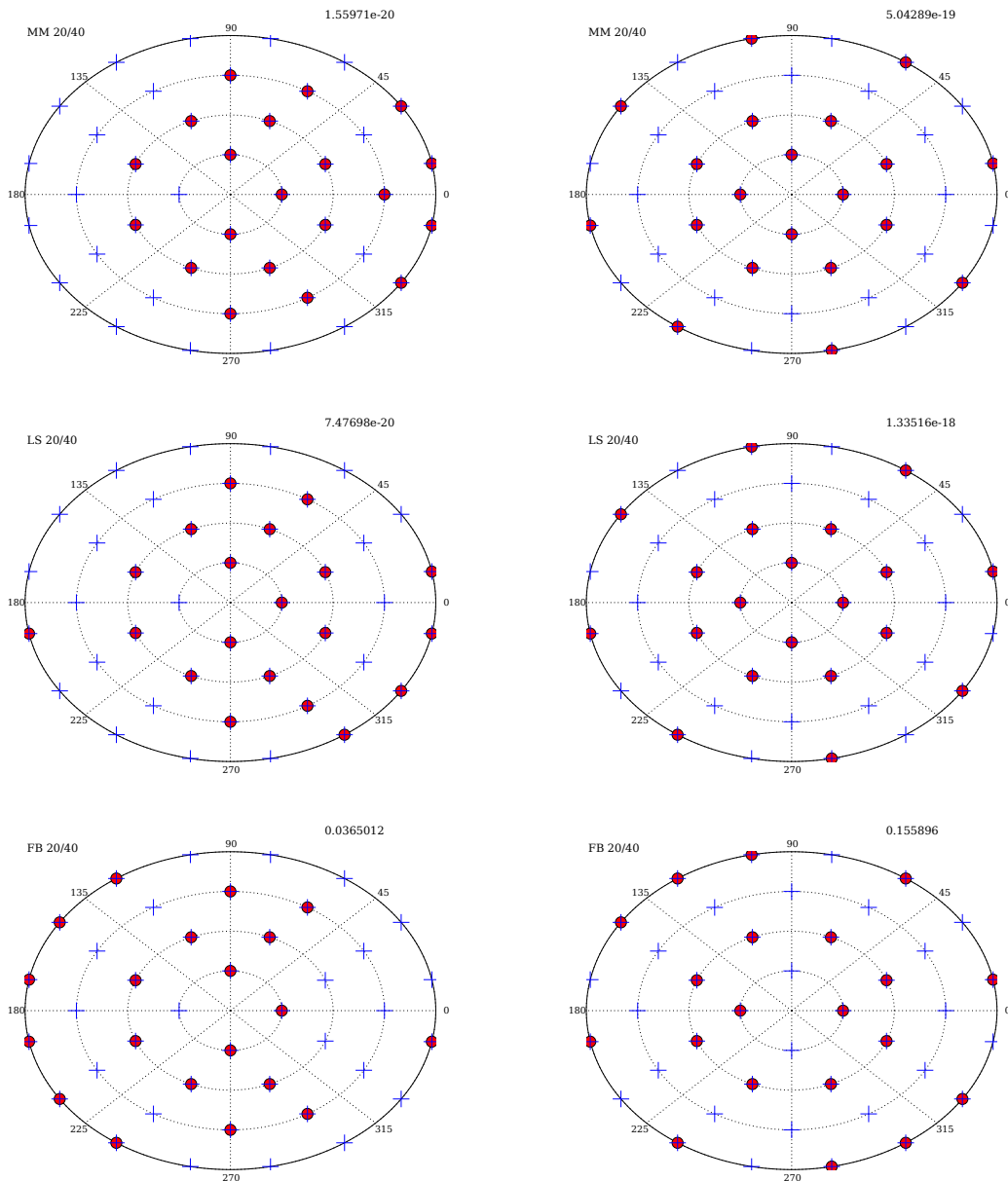


FIG. 4.3 – Réseaux optimaux de 20 stations (représentées par des ●) calculés sur une base de 40 sites potentiels (représentés par des +) disposés en cercles concentriques. La colonne de gauche correspond à une rose des vents dominée par un vent directionnel (vers la droite, direction 0°). La colonne de droite correspond à une distribution angulaire uniforme. Le critère utilisé pour les réseaux de la première ligne est la norme l_∞ . Le critère utilisé pour les réseaux de la seconde ligne est la norme l_2 . Le critère utilisé pour les réseaux de la troisième ligne est le biais fractionnel.

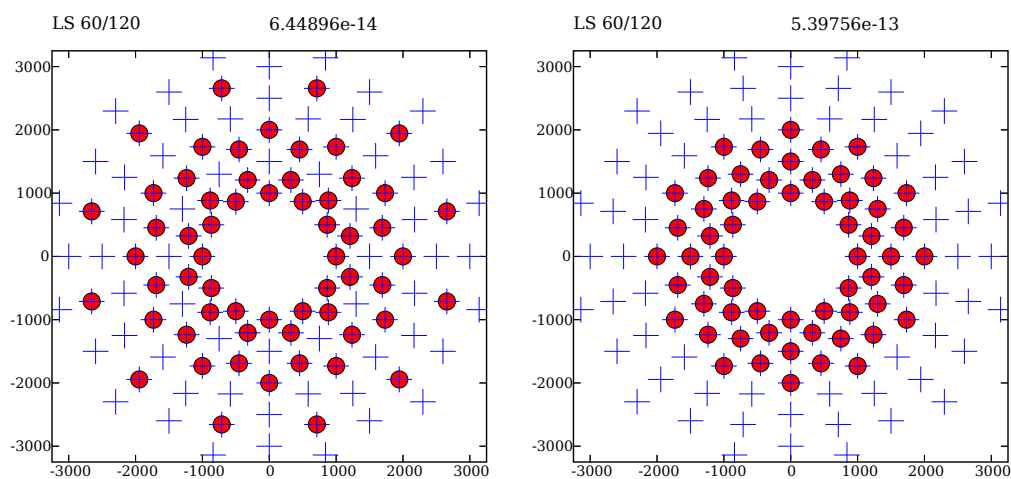


FIG. 4.4 – Réseaux optimaux de 60 stations (représentées par des \bullet) calculés sur une base de 120 sites potentiels (représentés par des $+$) disposés en cercles concentriques. Le critère retenu est celui des moindres carrés (norme l_2). La figure de gauche correspond à un vent de distribution angulaire uniforme et de force 3 m s^{-1} . La figure de droite correspond à un vent de distribution angulaire uniforme et de force 10 m s^{-1} .

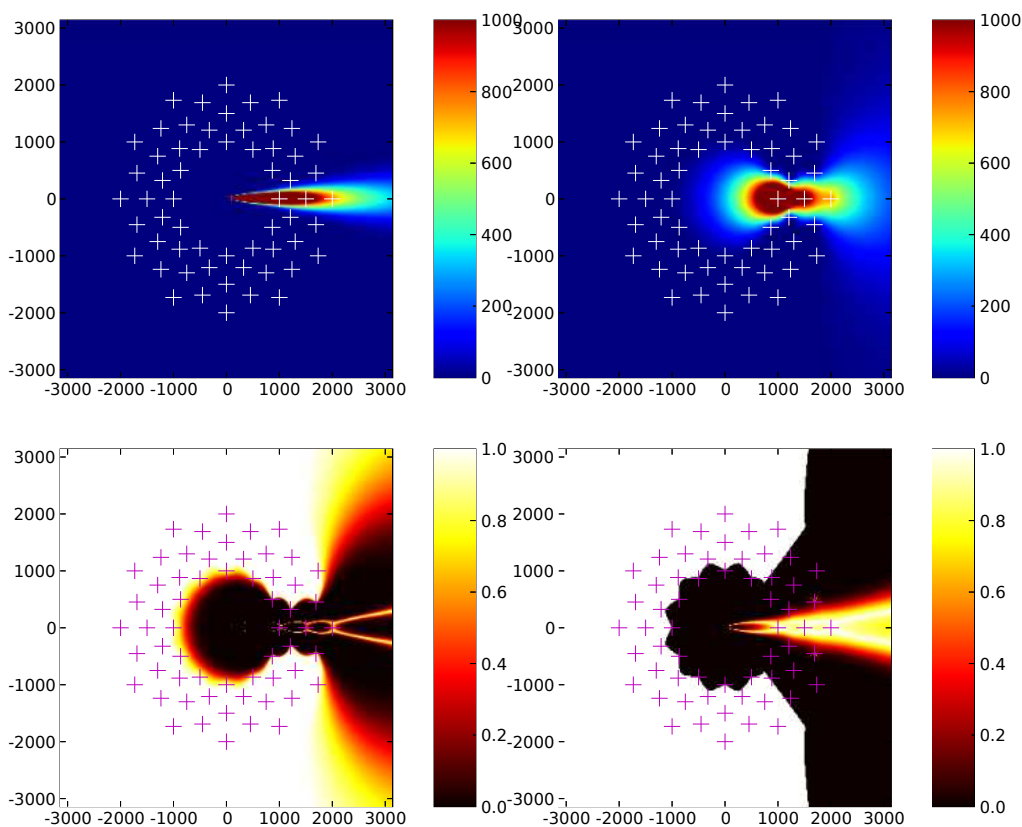


FIG. 4.5 – On considère le réseau optimal de 60 stations (représentées par des +), construit sur une base des vents de vitesse 10 m s^{-1} (voir réseau de droite de la figure 4.4). On étudie la réponse du système d'interpolation se fondant sur le réseau optimal à un rejet accidentel. En haut à gauche est représenté le rejet accidentel. En haut à droite est représenté le résultat du krigeage simple du champ de concentrations en utilisant les données du réseau optimal. Les figures du bas représentent les erreurs commises en fonction du point, pour un critère de type moindre carré (pour lequel le réseau est optimisé) et un critère de type biais fractionnel (pour lequel le réseau n'est pas optimisé).

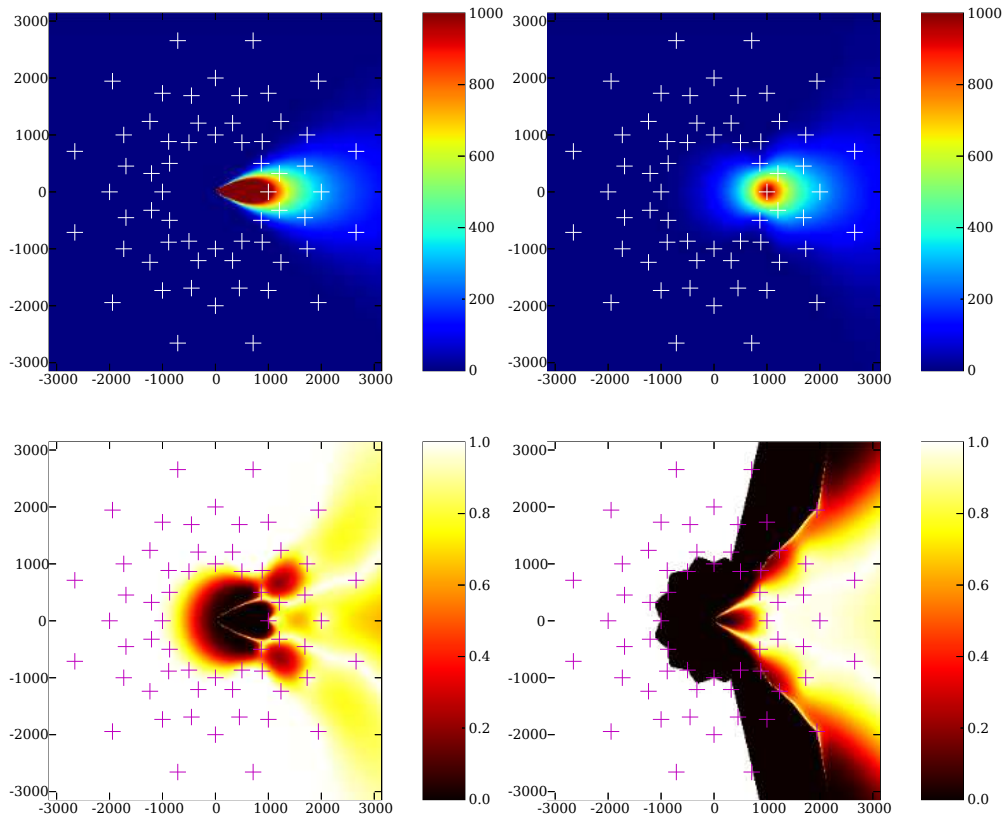


FIG. 4.6 – On considère le réseau optimal de 60 stations (représentées par des +), construit sur une base des vents de vitesse 3 m s^{-1} (voir réseau de gauche de la figure 4.4). On étudie la réponse du système d'interpolation se fondant sur le réseau optimal à un rejet accidentel. En haut à gauche est représenté le rejet accidentel. En haut à droite est représenté le résultat du krigeage simple du champ de concentrations en utilisant les données du réseau optimal. Les figures du bas représentent les erreurs commises en fonction du point, pour un critère de type moindre carré (pour lequel le réseau est optimisé) et un critère de type biais fractionnel (pour lequel le réseau n'est pas optimisé). Les distance sont en mètres.

Bibliographie

- [1] Munn, R. E. *The Design of Air Quality Monitoring Networks*. MacMillan Publishers Ltd, 1981.
- [2] Programme de Surveillance de la Qualité de l'air, AIRPARIF, juillet 2005. Consultable sur <http://www.airparif.asso.fr>.
- [3] Wackernagel, H., *Multivariate geostatistics, an introduction with applications*, third edition. Springer, 2003.
- [4] Site de la fédération ATMO : <http://atmo-france.org/>
- [5] Site d'air parif (qui regorge d'informations synthétiques et claires sur la pollution atmosphérique urbaine et péri-urbaine) : <http://www.airparif.asso.fr/>
- [6] Site regroupant les données AIRBASE : <http://air-climate.eionet.europa.eu/databases/airbase/> et aussi http://www.copernicus.org/EGU/acp/\-acpd/6/12769/acpd-6-12769_p.pdf
- [7] Webster, R. and Oliver, M. A., *Geostatistics for Environmental Scientists*”, Statistics in Practice, Series Editor Vic Barnett, Wiley, 2001.
- [8] Müller, W. G., *Collecting Spatial Data : Optimum Design of Experiments for Random Fields*, Springer-Verlag, 2001.
- [9] Bocquet, M., *Introduction aux principes et méthodes de l'assimilation de données en géophysique*, Notes de cours de l'ENSTA, 2005.
- [10] Cover, T. M. and Thomas, J. A., *Elements of Information Theory*, Wiley-Interscience, 1991.
- [11] Fedorov, V., V. and Hackel, P., *Model-Oriented Design of Experiments*, lecture notes in statistics, Springer-Verlag, 1997.
- [12] Kiefer, J. General equivalence theory for optimum designs (approximate theory), *Annals of Statistics*, 2, p. 849-879, 1974.
- [13] Kiefer, J. and Wolfowitz, J., The equivalence of two extremal problems, *Canad. J. Math.*, 14, p.363-366, 1960.
- [14] Kirkpatrick, S., Gelatt, C. D. and Vecchi, M., P. Optimization by Simulated Annealing, *Science*, Vol 220, Number 4598, pages 671-680, 1983. <http://citeseer.ist.psu.edu/kirkpatrick83optimization.html>
- [15] Quélo, D. *Assimilation de données pour la qualité de l'air, en particulier Annexe A*. Thèse de doctorat de l'École Nationale des Ponts et Chaussées, 2004. http://www.enpc.fr/cerea/fich/these_Denis_QUELO.pdf

- [16] Nychka, D. and Saltzman, N. Design of air quality networks. Dans Case Studies in Environmental Statistics, eds. D. Nychka, W. Piegorsch and L.H. Cox, Lecture Notes in Statistics number 132, Springer Verlag, New York, pp.51-76, 1998.