



Laboratoire Central de Surveillance de la Qualité de l'Air



Etude n°13 - Géostatistiques / estimation dépassement seuils

Géostatistique et estimation de probabilité de dépassement de seuil

Novembre 2004
Convention : 04000087

Giovanni CARDENAS



GEOSTATISTIQUE ET ESTIMATION DE PROBABILITES DE DEPASSEMENT DE SEUIL

**Laboratoire Central de Surveillance
de la Qualité de l'Air
Etude n°13**

Convention n° 04000087

Financée par la Direction de la Prévention des
Pollutions et des Risques (DPPR)

NOVEMBRE 2004

Giovanni CARDENAS

Ce document comporte 62 pages (hors couverture)

	Rédaction	Vérification	Approbation
NOM	Giovanni CARDENAS	Olivier SAINT-JEAN Laurence ROUIL	Martine RAMEL
Qualité	Ingénieur Etudes et Recherches Direction des Risques Chroniques	Responsable unité 2IEN Ingénieur Etudes et Recherches Direction des Risques Chroniques	Coordination LCSQA Direction des Risques Chroniques
Visa			

TABLE DES MATIERES

1. CONTENU DE LA FICHE.....	5
2. RESUME	7
3. INTRODUCTION	8
3.1. CARTOGRAPHIE DES DÉPASSEMENTS DE SEUILS	8
3.2. CARTOGRAPHIE DES PERCENTILES	8
4. PROBABILITÉ DE DÉPASSEMENT DE SEUIL POUR LA MOYENNE ANNUELLE	9
4.1. GÉNÉRALITÉ SUR LES MÉTHODES DE KRIGEAGE.....	9
4.2. GÉOSTATISTIQUE NON- LINEAIRE	9
4.3. TYPES DE MESURE DISPONIBLES.....	10
4.3.1. Généralités	10
4.3.2. Cas d'étude.....	11
4.4. ANAMORPHOSE GAUSSIENNE	12
4.4.1. Définition	12
4.4.2. Polynômes d'Hermite	13
4.5. MÉTHODE DE L'ESPÉRANCE CONDITIONNELLE	14
4.6. EXEMPLE PRATIQUE DE MISE EN ŒUVRE DE LA METHODE DE L'ESPERANCE CONDITIONNELLE	15
4.6.1. Analyse des données	15
4.6.2. Fonction d'anamorphose pour un support ponctuel	17
4.6.3. Probabilité de dépassement d'un seuil par l'espérance conditionnelle	23
4.7. MÉTHODE DE CONDITIONNEMENT CONFORME	29
4.7.1. La question du support d'estimation	29
4.7.2. Le modèle gaussien discret.....	31
4.7.3. Mise en œuvre de la méthode du Conditionnement Uniforme	33
4.7.3.1 Définition	33
4.7.3.2 Krigeage en dérive externe du dioxyde d'azote.....	34
4.7.3.3 Calcul des coefficients de Changement de Support.....	36
4.7.3.4 Conditionnement Uniforme.....	43
4.8. SIMULATIONS CONDITIONNELLES	47
4.8.1. Définition.....	47
4.8.2. Simulations dans le modèle multigaussien	47
4.8.3. La methode des bandes tournantes.....	48
4.9. EXEMPLE PRATIQUE DES SIMULATIONS CONDITIONNELLES	49
5. CALCUL DE PERCENTILES.....	57
6. CONCLUSION	59
7. RÉFÉRENCES	59
7.1. RAPPORTS LCSQA :	59
7.2. AUTRES RÉFÉRENCES.....	60
8. REMARQUE.....	62

LISTE DE FIGURES

Figure 1 : Présentation du logiciel Isatis de Geovariances	11
Figure 2 : Fonction de répartition d'une variable gaussienne réduite U (moyenne =0, écart-type=1)	12
Figure 3 : Principales statistiques de la moyenne annuelle du dioxyde d'azote	17
Figure 4 : Fenêtre " Gaussian Anamorphosis Modeling" du logiciel Isatis, utilisé pour le calcul de l'anamorphose gaussienne ponctuelle	18
Figure 5 : Fenêtre « Interactive Fitting Parameters » (les valeurs numériques correspondent à l'anamorphose ponctuelle)	20
Figure 6 : Transformation gaussienne de la concentration de dioxyde d'azote et modèle d'anamorphose ajusté par polynômes d'Hermite pour un support ponctuel	21
Figure 7 : Exemple d'interpolation linéaire pour trouver la transformée du seuil de dépassement ($30 \mu\text{g}/\text{m}^3$)	22
Figure 8 : Principales statistiques de la variable gaussienne	24
Figure 9 : Résultats du krigeage simple ponctuel de la transformée gaussienne du dioxyde d'azote	25
Figure 10 : Menu permettant le calcul de la valeur gaussienne du seuil (anamorphose ponctuelle)	27
Figure 11 : Calcul de la probabilité de dépassement d'un seuil ponctuelle	27
Figure 12 : Menu permettant le calcul direct de la probabilité de dépassement ponctuelle	28
Figure 13 : Carte de probabilité de dépassement d'un seuil de $30 \mu\text{g}/\text{m}^3$ pour le dioxyde d'azote par la méthode de l'espérance conditionnelle (commune de Bourg-en-Bresse, année 2001)	29
Figure 14 : Schéma montrant différents types de support	30
Figure 15 : Histogramme des différents types de supports	31
Figure 16 : Coefficient de changement de support en fonction de la variance de bloc	32
Figure 17 : Carte de la maille de densité de population	35
Figure 18 : Nuages de corrélation entre les moyennes saisonnières du dioxyde d'azote et la densité de population	35
Figure 19 : Krigeage Ordinaire des Panneaux	36
Figure 20 : Fenêtre « Bloc Support Correction » du logiciel Isatis, utilisée pour le calcul du coefficient de changement de support Point- Bloc	38
Figure 21 : Fenêtre « Gaussian Anamorphosis Modeling du logiciel Isatis, utilisé pour le calcul de l'anamorphose gaussienne des blocs	40
Figure 22 : Fenêtre « Gaussian Anamorphosis Modeling » du logiciel Isatis, utilisée pour le calcul de l'anamorphose gaussienne des panneaux	41
Figure 23 : Fenêtre « Bloc Support Correction » du logiciel Isatis, utilisée pour le calcul du coefficient de changement de support Point- Panneau	41
Figure 24 : Anamorphose de blocs de $1*1 \text{ m}^2$ et des panneaux de $50*50 \text{ m}^2$	42
Figure 25 : Calcul manuelle de la probabilité de dépassement de seuil par Conditionnement Uniforme	43
Figure 26 : Menu permettant le calcul de la probabilité de dépassement de seuil par Conditionnement Uniforme	44
Figure 27 : Probabilité de dépassement de seuil par Conditionnement Uniforme	45
Figure 28 : Comparaison des deux cartes de probabilité de dépassement obtenues par « Conditionnement Uniforme » et par « l'Espérance Conditionnelle »	46
Figure 29 : Le principe des bandes tournantes en 2D	49
Figure 30 : Menu permettant le calcul des simulations conditionnelles par bandes tournantes	50
Figure 31 : Cinq simulations conditionnelles par bandes tournantes de la concentration de dioxyde d'azote avec la densité de population en dérive externe	53
Figure 32 : Menu permettant le calcul de la probabilité de dépassement de seuil par simulations conditionnelles	54
Figure 33 : Comparaison des deux cartes de probabilité de dépassement obtenues par Simulations Conditionnelles (à gauche) et par « Conditionnement Uniforme » (à droite)	54
Figure 34 : Statistiques et nuage de corrélation de la probabilité de dépasser un seuil de $30 \mu\text{g}/\text{m}^3$ calculée par conditionnement uniforme et simulations conditionnelles	56

LISTE DE TABLEAUX

<i>Tableau 1 : Statistiques des huit campagnes de mesure du dioxyde d'azote.....</i>	<i>12</i>
<i>Tableau 2 : Statistiques de l'écart entre teneurs recalculées et teneurs originales</i>	<i>23</i>
<i>Tableau 3 : Statistiques du krigeage simple et des deux sélections des données (écart- type plus petit que 0.8 et que 0.7).....</i>	<i>26</i>
<i>Tableau 4: Statistiques du Krigeage en dérive externe des panneaux</i>	<i>36</i>
<i>Tableau 5 : Statistiques de la probabilité de dépasser 30 µg/m³</i>	<i>45</i>
<i>Tableau 6 : Statistiques de la moyenne et l'écart-type des 200 simulations conditionnelles par bandes tournantes.....</i>	<i>53</i>

1. CONTENU DE LA FICHE

Thème III : TRAVAUX NUMERIQUES

14 : GEOSTATISTIQUE ET ESTIMATION DE PROBABILITES DE DEPASSEMENT DE SEUIL

PROGRAMME NOUVEAU

CONTEXTES ET OBJECTIFS

L'application, des techniques de la géostatistique linéaire (krigeage ordinaire, cokrigeage, krigeage avec dérive externe) aux données de qualité de l'air a permis d'améliorer la précision des cartes de concentration obtenues par interpolation. Une façon d'exploiter ces cartes serait de les confronter aux seuils fixés par les directives européennes afin de cerner les zones de dépassement. Toutefois, si la carte des valeurs estimées et la carte de variance de l'erreur d'estimation qui lui est associée permettent d'apprécier correctement les niveaux de concentration dans l'air ambiant, elles sont insuffisantes, du fait du lissage opéré par le krigeage, pour répondre à la question du dépassement de seuil.

Le cadre probabiliste de la géostatistique offre néanmoins d'autres techniques pour aborder ce problème. En effet les méthodes géostatistiques dites « non linéaires » ont été développées non pas dans le but d'estimer une valeur vraie inconnue mais afin de calculer la probabilité avec laquelle cette valeur inconnue est supérieure ou inférieure à un seuil donné. Elles ne fournissent plus une carte d'estimation et sa carte de variance associée mais une carte de probabilité du dépassement de seuil.

Ce dernier type de carte se révèle particulièrement utile dans la mesure où il sert à délimiter des zones potentiellement exposées à un risque de dépassement et à quantifier par exemple le nombre d'habitants affectés.

A ce jour, les méthodes de la géostatistique linéaire (espérance conditionnelle, krigeage disjonctif, simulations) n'ont fait l'objet que d'études ponctuelles (ex : AIRNORMAND) et ne sont pas encore répandues dans les AASQA.

En vue de définir une méthodologie d'estimation des probabilités de dépassement de seuil, nous nous proposons donc d'explorer ces différents algorithmes, et d'en évaluer la pertinence dans le contexte de la qualité de l'air.

PROGRAMME PROPOSE POUR 2004

Il est important, en géostatistique, de déterminer le support de travail. Une première partie consistera donc à définir pour chaque polluant réglementé le support spatial et temporel des données qu'il convient de considérer. On tiendra compte à cette fin de la réglementation en vigueur et de la façon dont s'expriment les seuils.

Dans une seconde partie, afin de pouvoir fournir une méthodologie, il est proposé d'appliquer les techniques de la géostatistique non linéaire à différents jeux de données. Ceux-ci pourront correspondre à deux cas critiques, par exemple à des campagnes caractérisées par une grande densité de mesures et à des campagnes limitées à un nombre restreint de points.

Les conclusions de cette analyse et les recommandations pratiques qui en ressortiront seront fournies aux AASQA sous la forme d'un guide remis fin 2004.

COLLABORATIONS

AASQA , Centre de Géostatistique de l'Ecole des Mines de Paris

DUREE DES TRAVAUX

Il s'agit d'une nouvelle étude d'une durée de un an.

2. RESUME

Les réglementations de la qualité de l'air (directives européennes et décrets nationaux) précisent différentes valeurs limites de la concentration des principaux polluants à ne pas dépasser, pour définir ces seuils, plusieurs supports temporels sont employés : concentration horaires, moyennes journalières ou annuelles.

Ces seuils sont à respecter « en tout point du domaine » ; le cadre probabiliste sur lequel sont fondés les méthodes géostatistiques offre la possibilité de répondre à cette normativité.

Néanmoins, le calcul direct du dépassement de seuil sur les cartes d'estimation (par krigeage) est imprécis, surtout dans les zones où l'écart-type d'estimation est élevé, c'est pourquoi d'autres méthodes dites non- linéaires doivent être employées pour estimer la probabilité de dépassement des seuils. Parmi ces méthodes, l'espérance conditionnelle et les simulations conditionnelles par bandes tournantes peuvent être utilisées.

Ces méthodes sont basées sur un modèle « multigaussien » où toutes les combinaisons linéaires des variables gaussiennes sont également gaussiennes. En conséquence, l'implémentation pratique de ces méthodes entraîne le plus souvent une transformation des concentrations brutes en valeurs gaussiennes. Une variante de l'espérance conditionnelle qui permet de prendre également en compte les phénomènes non- stationnaires est le « conditionnement uniforme » dont une description est proposée dans ce document.

Lorsque le support temporel des seuils est exprimé en moyennes saisonnières ou annuelles, les données des tubes passifs peuvent être utilisées pour le calcul de la probabilité de dépassement de seuil.

En revanche lorsque le seuil est exprimé en une unité de temps plus petite comme l'heure ou le jour, les données valables sont les données continues des stations fixes. Dans ce cas il est nécessaire de cartographier les percentiles afin de les comparer au seuil. Des principes méthodologiques relatifs à cette procédure, plus délicate à mettre en œuvre, sont également présentés.

3. INTRODUCTION

3.1. CARTOGRAPHIE DES DEPASSEMENTS DE SEUILS

Le but de cette étude est de développer, dans le contexte de la pollution de l'air, une procédure de calcul de la probabilité avec laquelle la concentration d'un polluant peut dépasser un seuil déterminé, dans un domaine spatial connu et dans une unité de temps donnée, basée sur l'utilisation de méthodes géostatistiques.

L'objectif recherché est d'obtenir une cartographie de cette probabilité. Ce type de cartes pouvant, par exemple, être utilisé comme outil d'aide à la décision et concernant la vérification des objectifs réglementaires de qualité de l'air.

Les réglementations de la qualité de l'air précisent différents seuils de dépassement en fonction du support temporel. Par exemple, pour le dioxyde d'azote la valeur limite pour la prévention de la santé humaine en 2010 sera de $40 \mu\text{g}/\text{m}^3$ en moyenne annuelle et de $200 \mu\text{g}/\text{m}^3$ en valeur horaire à ne pas dépasser plus de dix-huit fois par an (directive européen N°. 1999/30/CE du 22 avril 1999).

Le critère de qualité sur la moyenne annuelle, peut être calculé à partir des techniques de la géostatistique non linéaire, sujet principal de cette étude. Les méthodes présentées sont l'espérance conditionnelle, une variante de cette dernière appelée « conditionnement uniforme » et les simulations conditionnelles par bandes tournantes.

L'application pratique de ces méthodes est illustrée grâce à un jeu des données fournies gracieusement par « l'Association l'Air de l'Ain et des Pays de Savoie ». Cette association a réalisé huit campagnes de mesures par tubes passifs (quatre pendant l'été 2001 et quatre pendant l'hiver 2001-2002), afin de cartographier la pollution par le dioxyde d'azote dans l'agglomération de Bourg-en-Bresse.

Ces méthodes étant complètement implémentées dans le logiciel ISATIS, logiciel spécialisé en géostatistique et dont la plupart des AASQA sont équipées, nous avons privilégié la description de la mise en oeuvre pratique des méthodes, plutôt que les développements mathématiques de la géostatistique non- linéaire qu'elle sous-tendent.

Pour les lecteurs désireux d'approfondir la théorie des méthodes, une liste de références est indiquée à la fin d cette ouvrage.

3.2. CARTOGRAPHIE DES PERCENTILES

Le critère de qualité sur les valeurs horaires ou journalières est plus complexe à cartographier car, considérant l'exemple du dioxyde d'azote, pour pouvoir sélectionner les points où la concentration horaire de $200 \mu\text{g}/\text{m}^3$ a été dépassée plus de 18 fois dans l'année, il faudrait connaître la concentration réelle horaire du dioxyde d'azote en tout point du domaine et pour toute la période en étude. Or on ne dispose des données horaires que dans les sites où des stations fixes sont installées.

Pour résoudre ce problème, dans la partie finale de cette étude, quelques pistes à explorer sont présentées, notamment l'utilisation des simulations géostatistiques, moyennant quelques hypothèses simplificatrices.

4. PROBABILITE DE DEPASSEMENT DE SEUIL POUR LA MOYENNE ANNUELLE

4.1. GENERALITE SUR LES METHODES DE KRIGEAGE

Parmi les méthodes d'interpolation existantes, le krigeage est l'une de plus utilisées au sein des AASQA. Cette méthode, qui fait partie des méthodes de la géostatistique linéaire, a l'avantage d'utiliser la corrélation spatiale des mesures et de permettre la prise en compte de variables auxiliaires. On obtient donc des interpolations plus précises.

L'utilisation des méthodes de la géostatistique linéaire pour la cartographie de la qualité de l'air a fait l'objet de plusieurs rapports du LCSQA, on cite particulièrement les études suivantes :

- « Evaluation des incertitudes associées aux méthodes géostatistiques (Ineris 2003) », dénommé par la suite « l'étude d'Incertitudes » et
- « Guide d'utilisation des méthodes de la géostatistique linéaire (Ineris 2003) ».

Nous rappelons que l'estimation par krigeage est définie comme une combinaison linéaire des données disponibles, qui donne une variance de l'erreur d'estimation minimale.

Les équations du krigeage fournissent les pondérateurs (λ_i) optimaux qui permettent d'obtenir en tout point une valeur estimée dont la variance de l'erreur d'estimation est minimale.

$$Z_o^* = \sum (\lambda_i * Z_i) \quad \text{Var}(Z_o^* - Z_o) = \min$$

On rappelle que la variance d'une distribution est le carré des écarts à la moyenne (mesure de la dispersion autour de la valeur moyenne) ; dans le cas de l'erreur d'estimation, la moyenne est égale à zéro. Par ailleurs, l'écart-type d'estimation est la racine carrée de la variance d'estimation.

Il est important de noter que le krigeage est un interpolateur exact car en tout site de mesure il reproduit la valeur mesurée. En ces points l'erreur d'estimation et sa variance sont donc nulles.

Le krigeage fournit non seulement une carte d'estimation de la concentration du polluant mais aussi une carte des erreurs d'estimation. Celle-ci informe des points où l'estimation est plus ou moins fiable. Les erreurs d'estimation les plus importantes sont souvent rencontrées aux sites les moins bien échantillonnés.

La maille d'estimation représente ainsi une interpolation des concentrations disponibles et non pas une restitution des concentrations réelles.

4.2. GEOSTATISTIQUE NON- LINEAIRE

Le calcul direct du dépassement de seuil sur les cartes d'estimation est imprécis surtout dans les zones où l'écart-type d'estimation est élevé.

Prenons l'exemple d'une estimation par krigeage de la moyenne annuelle du dioxyde d'azote effectuée dans une ville donnée.

Supposons que nous ayons en un point de la maille d'estimation une valeur estimée de $41 \mu\text{g}/\text{m}^3$, avec un écart-type de l'erreur d'estimation de $1 \mu\text{g}/\text{m}^3$ et que en un autre point la valeur estimée soit la même, mais avec un écart-type de l'erreur d'estimation trois fois plus grand ($3 \mu\text{g}/\text{m}^3$).

Si l'on compare la carte de krigeage directement avec la valeur annuelle limite ($40 \mu\text{g}/\text{m}^3$) on en conclut que cette valeur est dépassée dans les deux points pris en considération.

Cependant, si l'on prend en compte les écarts-types d'estimation, on peut dire que dans le premier point il y a toutes les chances pour que la valeur vraie dépasse en effet $40 \mu\text{g}/\text{m}^3$. En revanche pour le point dont l'écart-type est plus élevé, la valeur vraie peut valoir $41 \mu\text{g}/\text{m}^3$ mais elle peut également être égale à $40 \mu\text{g}/\text{m}^3$ ou $39 \mu\text{g}/\text{m}^3$. Le fait que la valeur estimée dépasse $40 \mu\text{g}/\text{m}^3$ ne signifie donc pas que la valeur vraie ne puisse être inférieure à $40 \mu\text{g}/\text{m}^3$.

Ce n'est donc pas la concentration dans une maille qu'il faut estimer mais une variable égale à 1 si la concentration dépasse 40 et à 0 sinon, cette variable est appelée indicatrice et est notée: $I_{Z>40}$. Le résultat de l'estimation de ce type de variable (0-1) a le sens d'une probabilité : en l'occurrence la probabilité que la concentration dans un point dépasse $40 \mu\text{g}/\text{m}^3$.

La seule considération de combinaisons linéaires de données disponibles ne suffit pas à trouver cette probabilité, c'est pourquoi les méthodes employées pour le calcul de dépassement de seuils sont appelés méthodes non linéaires.

Parmi les méthodes de la géostatistique non linéaire nous avons choisi de présenter :

- l'espérance conditionnelle,
- le conditionnement uniforme et,
- les simulations conditionnelles par bandes tournantes.

Ces méthodes peuvent être utilisées pour divers types de polluants dans l'air comme l'ozone, l'azote, le benzène, etc, la seule exigence étant d'avoir un nombre suffisant de mesures du polluant dans le domaine spatial (l'agglomération, le département, la région, etc). Cette contrainte est commune à toute étude géostatistique (Guide d'utilisation des méthodes de la géostatistique linéaire, Ineris 2003).

4.3. TYPES DE MESURE DISPONIBLES

4.3.1. GENERALITES

Les AASQA disposent de différents types de mesures de la qualité de l'air, d'une part, les mesures horaires déjà évoquées, provenant des stations fixes (capteurs) ou des camions mobiles, et d'autre part, les mesures hebdomadaires ou bi hebdomadaires obtenues à partir des échantillonneurs passifs.

Les mesures des stations fixes ont la caractéristique d'être continues sur des longues périodes de temps (saison, année) malheureusement le nombre de stations fixes est faible et leur localisation est très ciblée, car leur principal objectif est la surveillance des dépassements locaux des seuils d'information et/ou d'alerte.

Les camions mobiles sont utilisés pour effectuer des études ciblées comme la surveillance de la qualité de l'air dans des zones sensibles, près des industries ou des aéroports par exemple. A partir de ces études, on peut déterminer la nécessité d'une surveillance plus continue et donc de l'installation ou non d'une station fixe.

Les campagnes par échantillonneurs passifs permettent d'évaluer la répartition spatiale de la pollution sur une période donnée. A partir d'un nombre souvent conséquent de points de mesure, une interpolation spatiale est réalisée pour obtenir une cartographie de la concentration de fond d'un polluant dans une zone qui peut être une agglomération ou toute une région. La durée d'une campagne est le plus souvent d'une à deux semaines.

L'avantage de ce type de campagnes est donc la possibilité d'investiguer une surface du territoire plus étendue par rapport aux stations fixes. Le désavantage est le coût lié au nombre des échantillonneurs à installer, ce qui limite les périodes de mesure à quelques semaines par saison.

Pour le calcul de la probabilité de dépassement de seuil en moyenne annuelle, les types de mesures utiles sont donc les données provenant des échantillonneurs passifs, à condition que les deux saisons (été et hiver) soient suffisamment échantillonnées.

Dans le cas où des mesures de capteurs sont disponibles, on peut les intégrer comme des échantillons supplémentaires en calculant la moyenne annuelle des données horaires.

Afin que les lecteurs tirent profit de ce document nous avons privilégié l'application pratique des méthodes en réduisant le côté mathématique. En effet, nous avons voulu attirer l'attention sur la résolution des problèmes pratiques qui peuvent se poser au praticien plutôt que sur les développements mathématiques de la géostatistique non linéaire.

Pour illustrer les méthodes nous avons utilisé le logiciel ISATIS version 5.0 (Geovariances), logiciel spécialisé en géostatistique et dont la plupart des AASQA sont équipées.

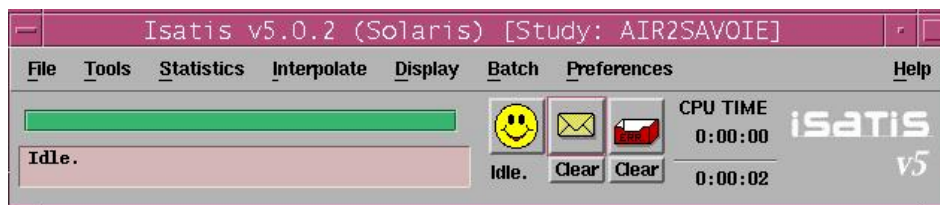


Figure 1 : Présentation du logiciel Isatis de Geovariances

4.3.2. CAS D'ETUDE

Les données chargées dans ce logiciel ont été fournies gracieusement par « l'Association l'Air de l'Ain et des Pays de Savoie ».

Cette association a réalisé huit campagnes de mesures par tubes passifs, quatre pendant l'été 2001 et quatre pendant l'hiver 2001-2002, afin de cartographier la pollution par l'azote dans l'agglomération de Bourg-en-Bresse (Tableau 1).

VARIABLE	Date où commence la période de mesures	Nb	Moyenne	Variance
ETE 1	19-06-01	38	17.6	24.8
ETE 2	05-07-01	42	13.6	16.2
ETE 3	19-07-01	46	17.8	23.8
ETE 4	02 au 20-08-01	44	11.9	12.0
HIVER 1	20-12-01	41	30.4	29.7
HIVER 2	08-01-02	45	44.2	47.0
HIVER 3	21-01-02	46	29.1	36.9
HIVER 4	04 au 18-02-02	44	27.0	30.4

Tableau 1 : Statistiques des huit campagnes de mesure du dioxyde d’azote

A partir de ces données, nous avons calculé la moyenne annuelle. Ce calcul a été effectué en ne prenant en compte que les sites où les mesures d’au moins deux quinzaines par saison sont disponibles. Les exemples du calcul de la probabilité de dépassement, exposés par la suite, ont été effectués sur cette moyenne annuelle du dioxyde d’azote.

4.4. ANAMORPHOSE GAUSSIENNE

4.4.1. DEFINITION

Les méthodes de la géostatistique non linéaire exposées dans ce rapport (Espérance Conditionnelle, Conditionnement Uniforme, Simulations) supposent que la distribution de fréquences de la variable étudiée peut être modélisée par un modèle gaussien.

Il convient de rappeler que la fonction de répartition d’une variable gaussienne centrée réduite U (moyenne nulle et variance égale à 1) est donnée par la formule de la Figure 2.

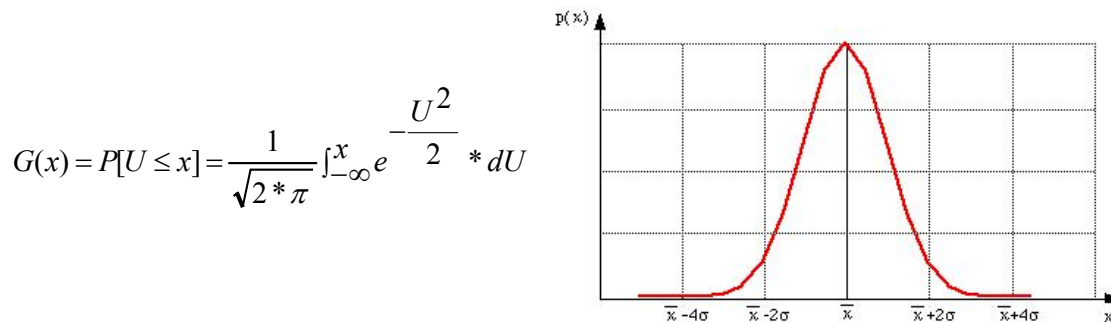


Figure 2 : Fonction de répartition d’une variable gaussienne réduite U (moyenne =0, écart-type=1)

Cette fonction n'a pas d'expression analytique simple. En pratique on se sert soit de tables, soit de formules d'approximation, calculées numériquement.

En qualité de l'air il est très rare de rencontrer des distributions parfaitement gaussiennes, le plus souvent l'histogramme est asymétrique avec quelques valeurs élevées. Ces valeurs représentent la concentration dans certains endroits du domaine où l'exposition au polluant est plus forte, par exemple le dioxyde d'azote en centre ville ou l'ozone en zones rurales.

Par conséquent, on doit procéder à une transformation de la variable réelle en variable gaussienne. Le processus consiste à déformer l'histogramme du polluant en un histogramme gaussien réduit ; sur les histogrammes cumulés (fonctions de répartition F(z) et G(y)) elle consiste à associer à chaque valeur de « z », la valeur « y » (gaussienne) correspondant à la même fréquence cumulée.

4.4.2. POLYNOMES D'HERMITE

La fonction reliant « y » et « z » est appelée fonction d'anamorphose ou tout simplement anamorphose (Φ). Cette fonction enregistre toutes les irrégularités que peuvent présenter la distribution expérimentale du polluant ; pour une variable gaussienne, l'anamorphose est linéaire.

Les fonctions utilisées en géostatistique pour modéliser l'anamorphose gaussienne sont les « polynômes d'Hermite ». Ces polynômes sont définis à partir de la densité de probabilité gaussienne réduite. Les trois premiers polynômes sont donnés par les fonctions suivantes :

$$H_0(Y) = 1; \quad H_1(Y) = -Y; \quad H_2(Y) = \frac{(Y^2 - 1)}{\sqrt{2}}$$

Ils sont ensuite calculables par à une relation de récurrence :

$$H_{n+1}(Y) = -\frac{1}{\sqrt{n+1}} * Y * H_n(Y) - \sqrt{\frac{n}{n+1}} * H_{n-1}(Y) \quad \text{où} \quad n > 0$$

Le principal avantage de ce type de polynômes est que n'importe quelle fonction peut être développée en polynômes d'Hermite, par exemple, pour l'anamorphose gaussienne on obtient :

$$Z = \Phi [Y] = f_0 + f_1 * H_1[Y] + f_2 * H_2[Y] + \dots$$

$$Z = \sum_{n=0}^{+\infty} \{f_n * H_n[Y]\}$$

Y : valeur gaussienne centrée réduite
 f_n : énième coefficient
 $H_i[Y]$: énième polynôme d'Hermite

Le nombre de coefficients calculés « f_n » indique le degré $n=0,1,2,\dots$ du polynôme d'Hermite ajusté. Ces coefficients restent constants et ne dépendent pas du degré du polynôme. On peut ainsi stocker une centaine de coefficients, et choisir ensuite le degré 10, 20 ou 30 de l'approximation polynomiale.

En pratique, on calcule un nombre déterminé de polynômes à partir des valeurs gaussiennes (Y) du polluant, (on peut se contenter d'en calculer au maximum quelques dizaines), ensuite on calcule les coefficients de chaque polynôme à l'aide de la formule :

$$f_n = \int z^n * H_n(y) * g(y) * dy$$

Y : valeur gaussienne centrée réduite
 Z : Valeur réelle
 f_n : énième coefficient
 $H_n[Y]$: énième polynôme d'Hermite
 $g(y)$: fonction de densité d'une loi gaussienne centrée réduite

Cette intégrale est très facile à développer (on remplace la formule des polynômes d'Hermite dans l'intégrale), mais en pratique les logiciels font le calcul et donnent directement les coefficients des polynômes.

On peut valider ce calcul par les deux expressions ci-après :

$$f_0 = E[Z] \quad \text{et} \quad Var[Z] = \sum_1^{\infty} (f_n)^2$$

On vérifie donc que le premier coefficient est égal à la moyenne expérimentale des données et que la somme du carré de tous les coefficients (sauf le premier) est égale à la variance expérimentale de la variable réelle.

Une fois que la fonction d'anamorphose a été ajustée par les polynômes d'Hermite, on peut passer des valeurs gaussiennes aux valeurs réelles du polluant.

La Figure 6 illustre la transformation gaussienne de notre cas d'étude (anamorphose de la moyenne de la concentration de l'année 2001, du dioxyde d'azote de l'agglomération de Bourg-en-Bresse).

4.5. METHODE DE L'ESPERANCE CONDITIONNELLE

Le calcul de la probabilité de dépassement de seuil par la méthode de l'espérance conditionnelle est l'une des plus simples à mettre en œuvre. En effet, la probabilité est calculée directement sur la transformée gaussienne de la concentration de la pollution.

L'espérance conditionnelle suppose gaussienne toute combinaison linéaire des variables gaussiennes pour tout ensemble de points. C'est ce que l'on appelle l'hypothèse multi-gaussienne. Cette hypothèse est très contraignante et n'est pas vérifiable dans la pratique.

Cette méthode entraîne de plus, des hypothèses de stationnarité stricte. Par exemple, on ne peut pas effectuer le calcul de dépassement de seuil en prenant en compte une dérive (ou variogramme linéaire). Dans ce cas il est préférable d'utiliser la méthode du conditionnement uniforme qui sera abordée par la suite.

Partant de la variable gaussienne, et supposant donc qu'une anamorphose gaussienne est déjà effectuée, nous donnons ci-dessous les expressions qui permettent de calculer la probabilité de dépassement de seuil :

$$P[Y \geq \text{seuil} / Y = y_i, i = 1, \dots, N] = E[I_{Y \geq \text{seuil}}]$$

$$E[I_{Y \geq \text{seuil}}] = 1 - G\left(\frac{\text{seuil} - Y}{\sigma_{KS}}\right) \quad E[I_{Y \leq \text{seuil}}] = G\left(\frac{\text{seuil} - Y}{\sigma_{KS}}\right)$$

Dans ces expressions G est la fonction de répartition d'une variable gaussienne centrée réduite (U). Cette probabilité est donnée dans des tableaux ou calculée par des formules d'approximation.

En résumé, il suffit d'effectuer un krigeage simple sur la transformée gaussienne de la concentration du polluant, puis de calculer la relation $[(\text{seuil}-Z^{KS})/\sigma^{KS}]$ et enfin de lire directement la probabilité pour les valeurs normales correspondantes $[G(U)]$ sur une table appropriée (les logiciels la donnent directement).

La valeur numérique du seuil à appliquer doit être sa transformée gaussienne.

Dans la section suivante nous allons illustrer l'utilisation pratique de ces formules, à l'aide de données de la concentration de l'année 2001 du dioxyde d'azote de l'agglomération de Bourg-en-Bresse.

4.6. EXEMPLE PRATIQUE DE MISE EN ŒUVRE DE LA METHODE DE L'ESPERANCE CONDITIONNELLE

La méthode pour accéder à la probabilité de dépassement d'un seuil de concentration en moyenne annuelle ou saisonnière, par l'espérance conditionnelle peut se résumer par la démarche présentée ci-dessous.

4.6.1. ANALYSE DES DONNEES

C'est la phase classique par laquelle toute étude géostatistique doit débiter. On doit notamment :

- calculer des histogrammes,
- supprimer les données doublons,
- identifier la localisation géographique des échantillons,
- détecter les données aberrantes et les valeurs extrêmes,
- si les données sont regroupées effectuer une déclusterisation,
- calculer le variogramme expérimental,
- rechercher les possibles anisotropies,
- explorer la possible corrélation avec des variables auxiliaires,
- décider de l'hypothèse de stationnarité de la variable,
- et dans le cas où la variable est stationnaire modéliser la covariance expérimentale respective.

La Figure 3 met en évidence les principales statistiques de la moyenne annuelle de la concentration du dioxyde d'azote dans l'agglomération de Bourg-en-Bresse, pour l'année 2001. Dans ce cas d'étude, la concentration du polluant présente d'une part une non stationnarité due à la présence de fortes valeurs au centre ville, et d'autre part le variogramme omni- directionnel indique un processus fortement aléatoire quasi-pépitique.

La concentration est loin de présenter une distribution normale, l'histogramme et le nuage de corrélation avec une variable gaussienne le démontrent. La localisation géographique des données indique un échantillonnage plus dense en centre ville qu'en périphérie, le but de la campagne étant une cartographie de la pollution urbaine du dioxyde d'azote.

A ce stade de l'étude, on peut conclure que la structure du phénomène n'est pas clairement accessible par les seules valeurs réelles, peut-être une transformation des données permettrait-elle de tempérer l'influence des valeurs fortes du centre ville ? Ou peut-être une étude non stationnaire doit être envisagée ? Malgré ces constatations et de façon à pouvoir appliquer la méthode nous avons un peu « forcé » l'ajustement d'un modèle stationnaire.

Ce modèle peut être composé d'une ou plusieurs modèles autorisés (fonction de covariance définie positive).

Un modèle particulier est l'effet de pépite, il peut être défini comme « une discontinuité à l'origine » causée par :

- la présence d'une structure qui atteint son palier à une très petite distance (microstructure),
- par des erreur de mesure,
- ou par ces deux phénomènes en même temps.

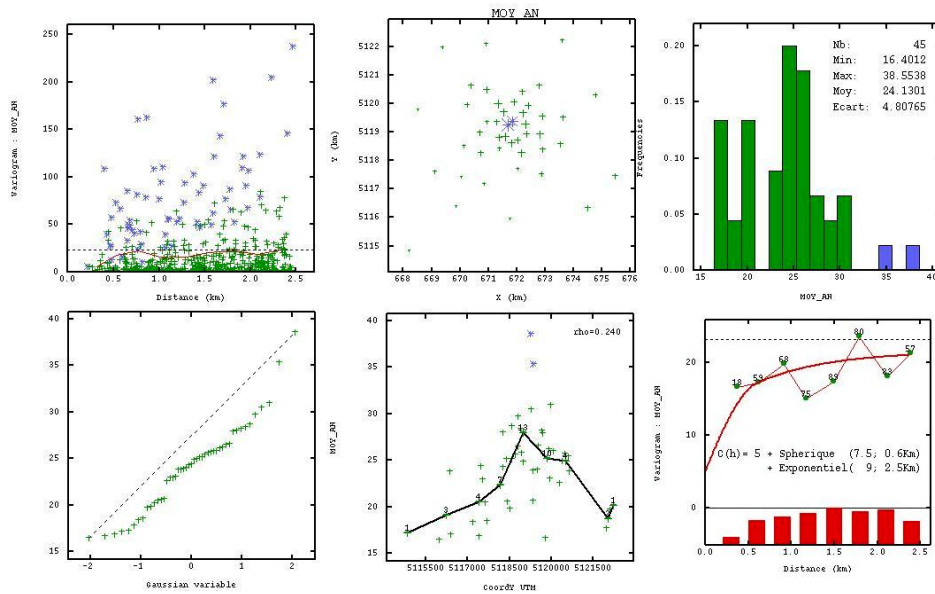
Dans le cas des données de Bourg-en-Bresse, l'absence des échantillons à petite distance (entre 0 et 500 m) ne permet pas de connaître avec certitude la valeur de l'effet de pépite, une valeur de $5 (\mu\text{g}/\text{m}^3)^2$ a donc été ajustée arbitrairement.

Remarque : Ce choix n'est pas sans conséquence, dans la page 33 de « l'étude d'Incertitudes », l'analyse de sensibilité menée sur l'effet de pépite montre qu'au fur et à mesure qu'il augmente, les estimations deviennent plus lisses en se rapprochant de la moyenne arithmétique des données. (Dans le cas d'un effet de pépite pur, la corrélation spatiale des données est nulle et les estimations correspondent à la moyenne arithmétique des points du voisinage).

Un modèle de covariance est défini par la structure choisie (exponentielle, sphérique, gaussienne, etc) et par les deux paramètres portée et palier.

La portée représente la distance après laquelle la corrélation spatiale des données devient nulle, elle indique dans quelle mesure l'influence d'un échantillon s'atténue avec la distance. Le palier représente la valeur du variogramme à une distance égale à la portée (limite supérieure du variogramme, là où il n'y a plus de corrélation entre les échantillons séparés par cette distance).

Une structure sphérique de portée 600 m (palier 7.5) et une structure gaussienne de portée 2.5 Km (palier 9) ont été choisies pour modéliser le variogramme ; ce modèle n'atteint pas la variance des données.



En Haut: A gauche : Nuée variographique
 Au milieu : Carte de localisation des tubes passifs
 A droite : Histogramme des concentrations mesurées

En Bas : A gauche : Nuage de corrélation entre les concentrations mesurées et une variable gaussienne centrée réduite.
 Au milieu : Nuage de corrélation entre les concentrations et la latitude (coordonnée Y)
 A droite : Variogrammes expérimentale des concentrations mesurées et modèle ajusté

Les astérisques et barres bleus correspondent au deux concentrations les plus élevées.

Figure 3 : Principales statistiques de la moyenne annuelle du dioxyde d'azote

4.6.2. FONCTION D'ANAMORPHOSE POUR UN SUPPORT PONCTUEL

Si la distribution de la concentration du polluant n'est pas gaussienne, on effectue une anamorphose et on la modélise à l'aide des polynômes d'Hermite de façon à avoir une fonction permettant de passer des valeurs réelles aux valeurs gaussiennes ou inversement.

Remarque : La moyenne annuelle du dioxyde d'azote pour l'année 2001 pour l'agglomération de Bourg-en-Bresse est de $24 \mu\text{g}/\text{m}^3$, valeur très inférieure au seuil de dépassement (valeur fixée à $58 \mu\text{g}/\text{m}^3$). On a donc considéré un autre seuil pour le calcul de probabilité de dépassement, celui de $30 \mu\text{g}/\text{m}^3$ (c'est la valeur à ne pas dépasser en moyenne annuelle pour les oxydes d'azote (monoxydes + dioxydes) en ce qui concerne la protection de la végétation).

Pour ce premier calcul, on a supposé que les seuils de dépassement indiqués dans les normes ont comme référence un support ponctuel. On a donc calculé la probabilité avec laquelle la concentration ponctuelle du dioxyde d'azote dépasse une concentration ponctuelle de $30 \mu\text{g}/\text{m}^3$. Si un support plus grand (bloc) doit être pris en compte il faut opérer un changement de support, sujet expliqué dans la section 4.7. Les calculs ont été effectués sur le centre des mailles d'une grille (maille de 50m de côté).

Les calculs sont effectués à l'aide du menu interactif du logiciel Isatis : « Gaussian Anamorphosis Modelling » (Figure 4). L'objectif de ce menu est de calculer les valeurs gaussiennes « discrètes » du polluant, puis de calculer les polynômes d'Hermite nécessaires pour l'ajustement de la courbe. Ce menu permet aussi le calcul du coefficient de changement de support nécessaire lors de calculs sur des blocs (section 4.7).

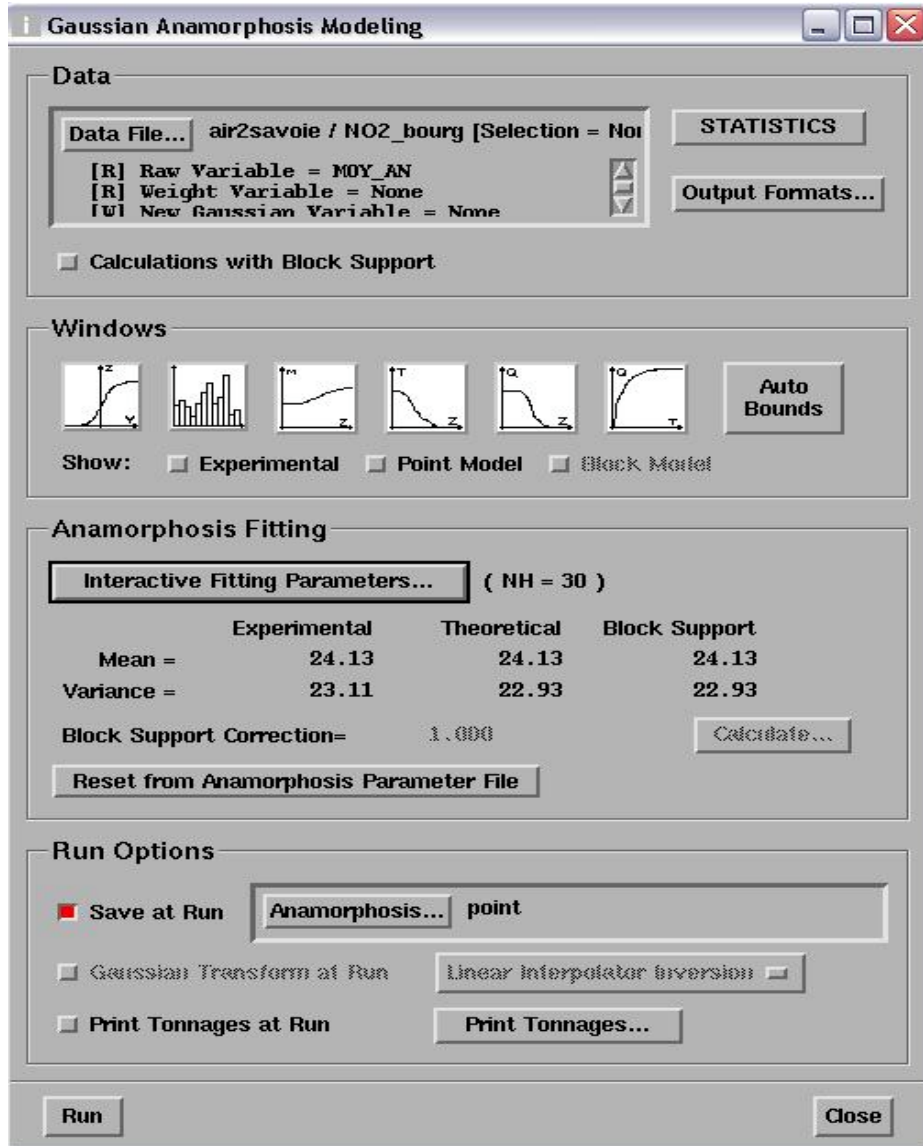


Figure 4 : Fenêtre " Gaussian Anamorphosis Modeling" du logiciel Isatis, utilisé pour le calcul de l'anamorphose gaussienne ponctuelle

La démarche de renseignement de ce menu est détaillé ci-dessous :

- Renseigner le nom de la variable représentant les concentrations réelles (MOY_AN) ; il est conseillé de garder en « tête » les valeurs extrêmes (minimales et maximales) ainsi que la moyenne et la variance du polluant, le bouton « STATISTICS » permettant d'accéder à ces valeurs.

Remarque : Cette fenêtre propose de créer la variable gaussienne, pour cela il faut renseigner le nom à assigner à cette variable (MOY_AN_gauss) ; ensuite cocher l'option « Gaussian Transform at Run » (partie inférieure de la fenêtre); et choisir une de trois méthodes d'inversion proposées : «fréquences » (Frequency Inversion), « empirique » (Empirical Inversion) ou « interpolation linéaire » (Linear Interpolator Inversion).

- La fenêtre « Interactive Fitting Parameters » (Figure 5) permet de définir principalement deux paramètres, à savoir l'intervalle de calcul de l'anamorphose et le nombre de polynômes d'Hermite à ajuster.
- Lorsque l'on veut passer des valeurs gaussiennes aux valeurs réelles, il est important de prêter attention aux intervalles. Par exemple dans la méthode de simulations, étudiée dans les sections suivantes, lorsque l'on tire aléatoirement plusieurs réalisations d'une variable il est difficile d'estimer avec certitude l'intervalle des valeurs dans lequel les résultats vont être compris. C'est pourquoi il est conseillé de prévoir à l'avance un intervalle suffisamment grand de façon à ne pas influencer artificiellement les statistiques des résultats.
- Pour définir le nombre de polynômes à ajuster il n'y a pas de règle générale, le conseil à retenir est qu'il faut veiller à ce que l'ajustement graphique obtenu soit acceptable, le bouton « anamorphosis » (premier icône à gauche de la section « WINDOWS ») permet d'afficher les ajustements et donc de piloter le paramétrage.
- L'intervalle par défaut dans l'onglet « Authorized Interval on Raw Variable », est défini par les valeurs extrêmes des données expérimentales. Dans le cas de l'espérance conditionnelle, on peut conserver cet intervalle. Dans l'exemple présenté, nous avons choisi un intervalle allant de 10 à 45 $\mu\text{g}/\text{m}^3$ pour la moyenne annuelle du dioxyde d'azote. De cette façon, on peut utiliser la même anamorphose ponctuelle pour les autres méthodes présentées dans ce rapport.

$$Z_{\min}^a = 10; \quad Z_{\max}^a = 45$$

- Pour contrôler l'ajustement du modèle dans les extrêmes (limites supérieures et inférieures) le logiciel propose d'ajouter deux points supplémentaires (appelés « points de contrôle »), dont les coordonnées ne peuvent être plus petites en valeur absolue que les valeurs extrêmes des données expérimentales. Cette option peut être utile dans le cas où le modèle s'écarte des données expérimentales à cause, par exemple, d'un nombre insuffisant de points. On a conservé les valeurs proposées par défaut (pour la variable réelle : valeurs les plus proches aux minimum et maximum des données expérimentales).

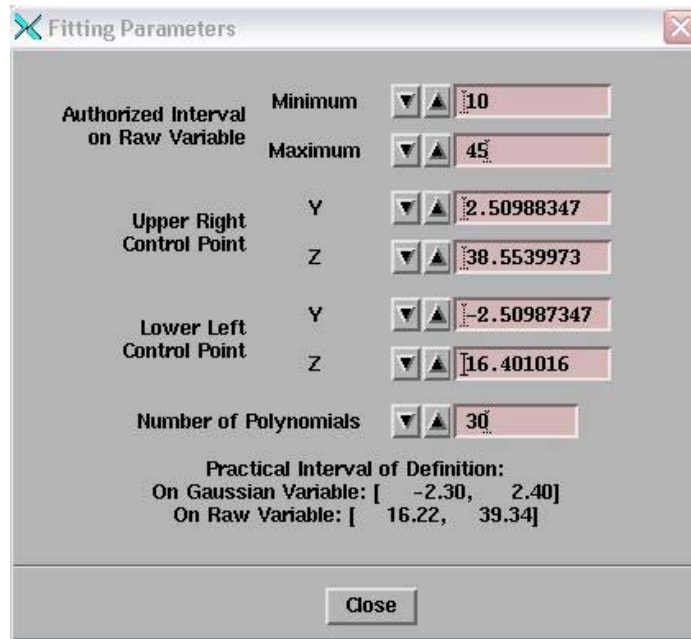


Figure 5 : Fenêtre « Interactive Fitting Parameters » (les valeurs numériques correspondent à l'anamorphose ponctuelle)

- Pour le nombre de polynômes, on a également conservé la valeur proposée par défaut (30), les courbes de couleur bleu dans la Figure 6 montrent l'ajustement obtenu par l'anamorphose ponctuelle.
- La Figure 6 illustre la fonction de répartition de la variable réelle en fonction des valeurs gaussiennes de la moyenne annuelle du dioxyde d'azote et le modèle ajusté par 30 polynômes d'Hermite.
- L'intervalle absolu est égal à l'intervalle autorisé. Dans notre cas, les valeurs gaussiennes de l'intervalle absolu ponctuel sont :

$$Y_{\min}^a = -4.58; \quad Y_{\max}^a = 4.62$$

- Les résultats de transformations ne sortiront pas des intervalles absolus, que ce soit dans un sens ou dans l'autre (Figure 6).
- Pour des valeurs plus grandes que les valeurs extrêmes, le modèle devient périodique. Pour cette raison, on limite l'utilisation du modèle à l'intervalle compris entre les deux premiers points d'inflexion de la courbe polynomiale ajustée (premiers points où la pente de la tangente à la courbe est nulle). Cet intervalle est appelé "intervalle pratique" (Figure 6).

- Si on sollicite le modèle en dehors de l'intervalle pratique (cas où l'intervalle absolu est plus grand que l'intervalle pratique) une interpolation linéaire est réalisée entre l'intervalle pratique et l'intervalle absolu (Figure 6).

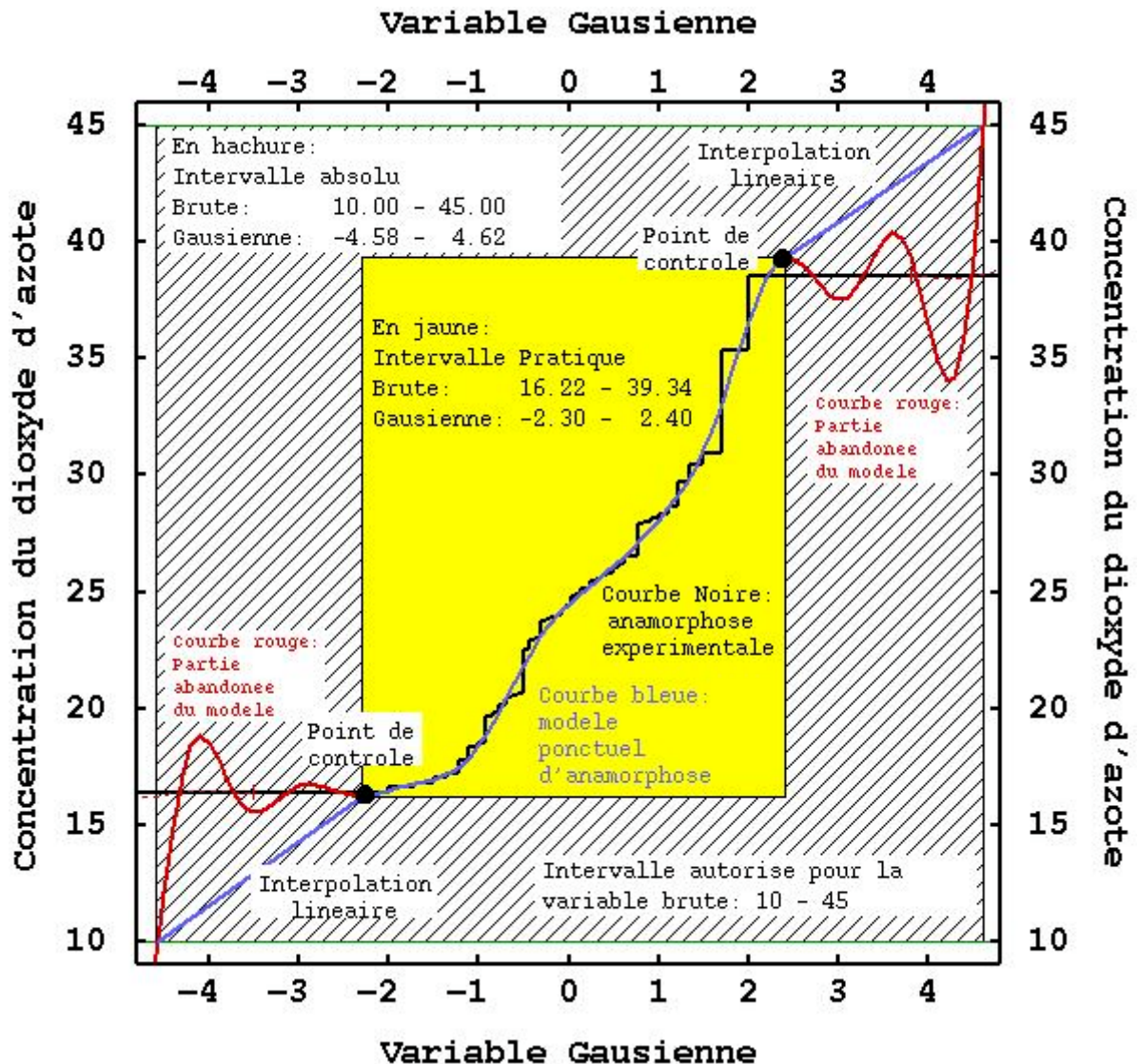


Figure 6 : Transformation gaussienne de la concentration de dioxyde d'azote et modèle d'anamorphose ajusté par polynômes d'Hermite pour un support ponctuel

- Notons que si l'intervalle autorisé est plus petit en valeur absolue que les points d'inflexion du modèle, alors il est égal à l'intervalle absolu et aucune interpolation linéaire n'est nécessaire. Nous avons sauvegardé l'anamorphose ponctuelle de la Figure 6 pour effectuer les transformations gaussiennes.

Remarque : Cette opération se fait dans la partie inférieure de la fenêtre de la Figure 4, il faut cocher l'option « Save at Run » et donner un nom à la anamorphose (par exemple « point »), puis il faut cliquer sur le bouton « Run ».

- Les polynômes d’Hermite permettent le passage de la valeur gaussienne (Y) à la valeur réelle (Z) directement (la formule de calcul est fonction des valeurs gaussiennes) mais il n’y a pas d’expression qui donne directement la valeur gaussienne à partir des valeurs réelles. Pour effectuer cette transformation Isatis propose donc trois méthodes d’inversion, à savoir :

➤ « **Interpolation linéaire** » (Linear Interpolator Inversion) : c’est la seule méthode qu’utilise le modèle d’anamorphose. D’abord on crée à l’aide de la fonction d’anamorphose $Z = \Phi(Y)$ une table avec toutes les valeurs Y (gaussiennes) et ses correspondants Z (réelles). Ensuite pour calculer la valeur Y_i correspondant à la valeur réelle Z_i , on effectue une interpolation linéaire entre les valeurs les plus proches de Z_i du tableau. La Figure 7 montre l’exemple de la transformation du seuil de dépassement en valeur gaussienne par interpolation linéaire de l’anamorphose ponctuelle.

Dans cette méthode, deux valeurs réelles identiques obtiennent la même valeur gaussienne. Il est important de prendre en compte les intervalles autorisés et pratiques : si la teneur se trouve en dehors des intervalles absolus (plus petite que $10 \mu\text{g}/\text{m}^3$ ou plus grand que $45 \mu\text{g}/\text{m}^3$), la valeur gaussienne assignée correspondra à l’intervalle absolu le plus proche (-4.58 ou 4.62) ; si la teneur se trouve entre l’intervalle absolu et l’intervalle pratique (entre 10 et $16.22 \mu\text{g}/\text{m}^3$ ou entre 39.34 et $45 \mu\text{g}/\text{m}^3$) la valeur gaussienne assignée correspondra à une valeur interpolée linéairement comme il a été déjà expliqué. (cf. Figure 6 et Figure 7).

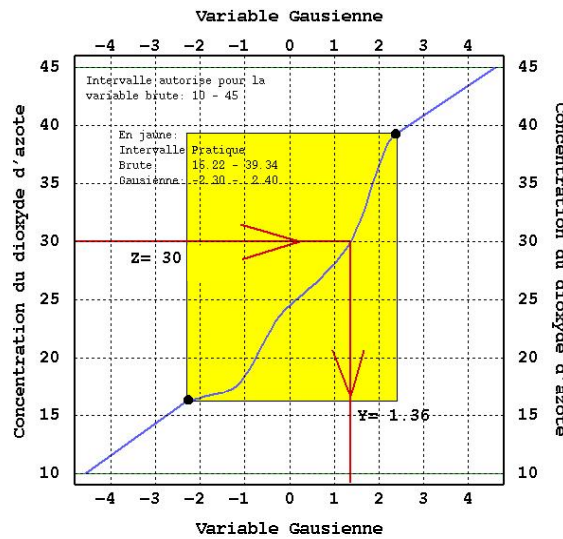


Figure 7 : Exemple d’interpolation linéaire pour trouver la transformée du seuil de dépassement ($30 \mu\text{g}/\text{m}^3$)

➤ « **Empirique** » (Empirical Inversion) : sur les histogrammes cumulés (fonctions de répartition F(z) et G(y)) cette méthode consiste à associer à chaque valeur de « z », la valeur « y » (gaussienne) correspondant à la même fréquence cumulée, deux teneurs identiques obtiennent ainsi une même valeur gaussienne.

- «*Fréquences*» (Frequency Inversion) : d’abord on trie les teneurs par ordre croissant, puis la fréquence cumulée est calculée en additionnant la fréquence (1/N) de chaque teneur. Ensuite on procède comme d’ordinaire (on associe à chaque valeur de « z », la valeur « y » (gaussienne) correspondant à la même fréquence cumulée), deux teneurs identiques obtiennent ainsi une valeur gaussienne différente.
- Un critère pour choisir le type d’inversion est de recalculer les teneurs à partir des valeurs gaussiennes et de les comparer avec les teneurs originales. Isatis effectue cette comparaison et donne la moyenne, l’écart- type et la variance de la différence entre les deux teneurs. Le Tableau 2 présente ces statistiques pour les trois méthodes d’inversion.
- Dans notre cas, la méthode de l’interpolation linéaire à l’aide de l’anamorphose redonne la même teneur, toutes les statistiques sont donc nulles, c'est pourquoi on a choisi cette méthode.

<i>Méthode</i>	<i>Moyenne</i>	<i>Variance</i>	<i>Ecart- type</i>
Fréquentielle	0.022731	0.863550	0.929274
Empirique	0.004485	0.652372	0.807696
Interpolation Linéaire	0.0	0.0	0.0

Tableau 2 : Statistiques de l’écart entre teneurs recalculées et teneurs originales

4.6.3. PROBABILITE DE DEPASSEMENT D’UN SEUIL PAR L’ESPERANCE CONDITIONNELLE

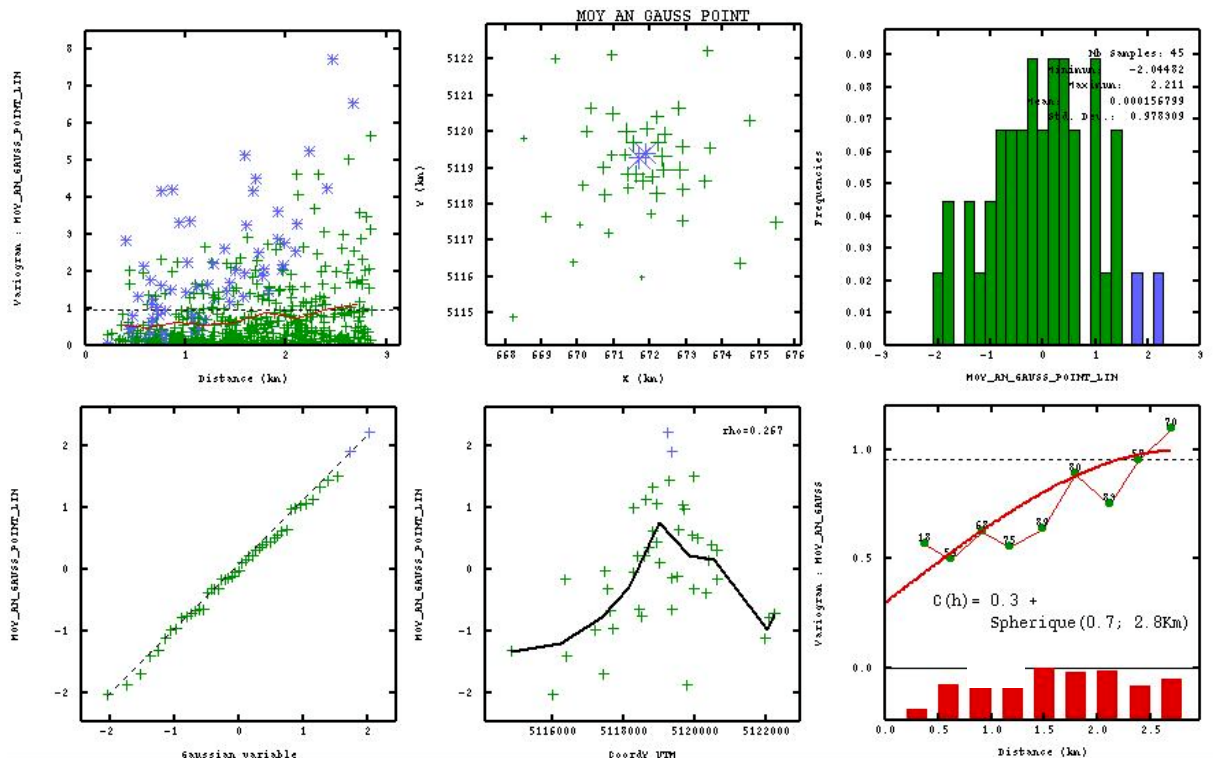
Pour effectuer le calcul de la probabilité de dépassement il faut d’abord effectuer un krigeage simple sur la variable gaussienne. Celle-ci a été obtenue par la méthode de l’interpolation linéaire, d’après les résultats du Tableau 2 de la section 4.6.2.

La Figure 8 illustre les principales statistiques de la transformée gaussienne de la moyenne du dioxyde d’azote de Bourg-en-Bresse. Cette figure est à comparer directement avec la Figure 3 qui montre les statistiques des concentrations originales.

La nouvelle variable a une moyenne presque nulle et un écart-type proche de 1 (propriétés d’une variable gaussienne réduite) mais l’histogramme n’est pas parfaitement gaussien, le nuage de corrélation entre la variable transformée et une variable gaussienne montre quelques valeurs de la transformée qui s’écartent légèrement d’un comportement gaussien parfait (valeurs les plus petites et quelques valeurs proches de 1).

Toutefois, la transformation est acceptable et a permis de réduire l'écart entre les valeurs extrêmes et le reste des données. Maintenant le variogramme expérimental est moins affecté par ces valeurs, le phénomène est moins pépétique (moins aléatoire) et par conséquent un peu plus structuré.

L'effet de pépité représente maintenant environ 30% de la variance totale et malgré l'évidente non stationnaire liée aux fortes concentrations du centre ville on peut supposer une stationnarité à petite échelle modélisée par une structure sphérique de portée 2.8 Km et de palier 0.7.



- En Haut:
 - A gauche : Nuée variographique
 - Au milieu : Carte de localisation des tubes passifs
 - A droite : Histogramme des transformées gaussiennes de la concentration
 - En Bas :
 - A gauche : Nuage de corrélation entre les transformées gaussiennes et une variable gaussienne centrée réduite.
 - Au milieu : Nuage de corrélation entre les transformées gaussiennes et la latitude (coordonnée Y)
 - A droite : Variogrammes expérimentale des transformées gaussiennes et modèle ajusté
- Les astérisques et barres bleus correspondent au deux concentrations les plus élevées.

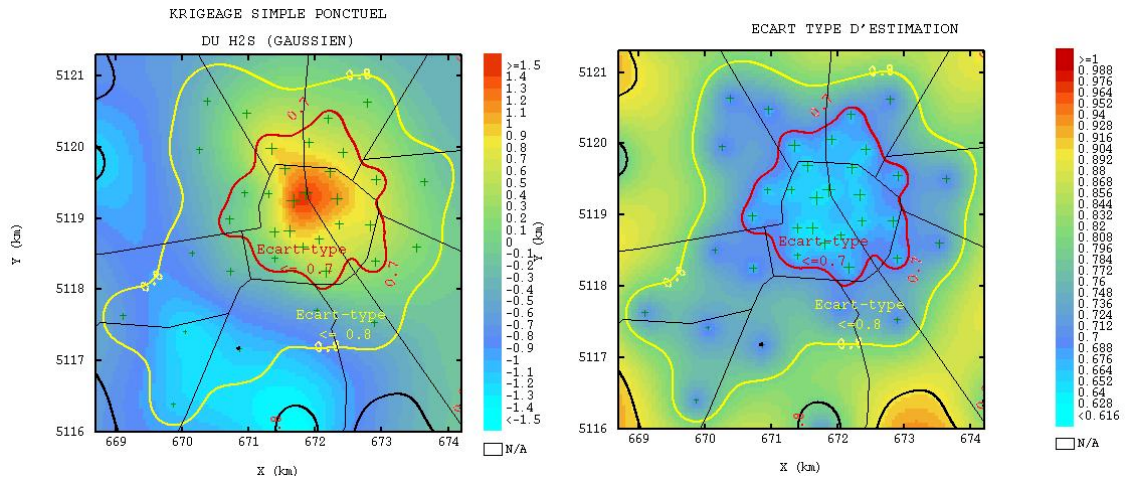
Figure 8 : Principales statistiques de la variable gaussienne

La Figure 9 représente les cartes résultantes du krigeage simple (à moyenne connue = 0).

Deux sélections de données ont été effectuées afin de définir les zones où la qualité de l'estimation est la meilleure. Ces zones sont celles où l'écart-type d'estimation est le plus faible. En utilisant ce critère on a repéré les écarts-types inférieures à 0.7 et à 0.8. Ces valeurs ont été choisies arbitrairement, le but n'étant que d'apercevoir l'allure des zones sélectionnées et de comparer les statistiques des sélections.

La sélection des zones où l'écart-type est inférieur à 0.7 (surtout zones urbaines) représente les endroits où l'estimation est plus fiable car l'échantillonnage est plus dense et le poids assigné à la moyenne (zéro) des données est moins important. Au delà (zones périphériques suburbaines et bords du domaine) le modèle est très sollicité en extrapolation à cause du manque de données conditionnelles. Le poids assigné à la moyenne globale est beaucoup plus important.

Remarque : « Lorsqu'en krigeage simple, le poids de la moyenne est faible (voisin de 0), l'estimation dépend principalement des valeurs locales, c'est-à-dire des points qui appartiennent à la configuration de krigeage, et non pas de la teneur moyenne dans le domaine. Le degré de stationnarité requis est donc moins important, de plus les estimateurs du krigeage ordinaire et du krigeage simple sont proches, ainsi que les variances d'estimation »: extrait de la page 16 de l'annexe B de « l'étude d'Incertitudes ».



Les cellules à l'intérieur de la courbe de niveau rouge représentent un écart-type d'estimation inférieure à 0.7, celles à l'intérieur de la courbe jaune représentent un écart inférieure à 0.8.

Les lignes noires représentent le réseau routier de Bourg-en-Bresse

Figure 9 : Résultats du krigeage simple ponctuel de la transformée gaussienne du dioxyde d'azote

Le Tableau 3 montre les statistiques de l'ensemble des résultats (et des deux sélections des cellules effectuées). La première sélection ($\sigma_{KS} \geq 0.8$) réunit 27.5 % des points de la grille d'estimation. Dans cette sélection la moyenne est plus proche de zéro, qui est la moyenne des données expérimentales.

En théorie, d'après l'hypothèse de stationnarité, la moyenne et la variance devraient être constantes partout dans le domaine. Or dans les deux sélections ces statistiques changent. Par exemple la variance augmente de 0.24 à 0.44 pour l'ensemble du domaine si l'on considère la première sélection et elle rechute à 0.14 pour la seconde sélection des points.

<i>Toutes les données</i>	Nb de cellules	Minimum	Maximum	Moyenne	Ecart- type	Variance	CV
Krigeage simple ponctuel du H2S gaussien	23 864	-1.49	1.46	-0.37	0.49	0.24	-1.3
Ecart-type d'estimation	23 864	0.63	1	0.84	0.09	0.01	0.1
<i>Ecart- type <=0.8</i>	27.5% des points						
Krigeage simple ponctuel	6 571	-1.3	1.46	-0.1	0.66	0.44	-6.64
<i>Ecart- type <=0.7</i>	6.8% des points						
Krigeage simple ponctuel	1 632	-0.22	1.46	0.66	0.37	0.14	0.56

Tableau 3 : Statistiques du krigeage simple et des deux sélections des données (écart- type plus petit que 0.8 et que 0.7)

On observe finalement que les valeurs extrêmes des trois jeux des estimations (l'ensemble des cellules et les deux sélections) sont plus petites que les valeurs extrêmes des données expérimentales, ceci est dû à l'effet de lissage du krigeage.

Les valeurs estimées sont une moyenne pondérée des données du voisinage d'estimation, le résultat est donc un resserrement de la distribution des concentrations.

Le calcul de probabilité peut s'effectuer à l'aide du menu « calculator » ou en utilisant un menu spécifique qui est apparu dans la dernière version du logiciel ISATIS (version 5.0). Les deux méthodes (qui donne des résultats identiques) sont détaillées.

1) Le calcul peut-être réalisé manuellement. Il faut alors transformer le seuil en valeur gaussienne à l'aide de l'anamorphose gaussienne ponctuelle créée dans la section 4.6.2, pour cela on utilise le menu de la Figure 10. Puis il faut créer dans la maille d'estimation une variable contenant la valeur gaussienne de ce seuil de dépassement. Enfin calculer la probabilité à l'aide de la formule présentée dans la section 4.5 (ces deux derniers calculs se font à l'aide du menu de la Figure 11).

La Figure 10 montre le menu utilisé pour la transformation du seuil en valeur gaussienne. Pour une concentration de 30 µg/m3 on a obtenu une valeur gaussienne ponctuelle égale à 1.36 est obtenue (Figure 7).

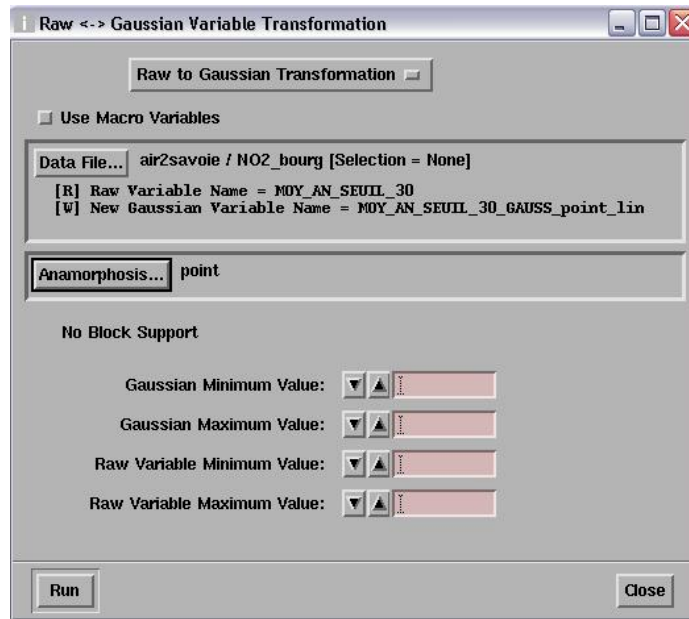


Figure 10 : Menu permettant le calcul de la valeur gaussienne du seuil (anamorphose ponctuelle)

2) La Figure 12 présente le nouveau menu permettant le calcul direct de la probabilité de dépassement. A part le fait d'éviter de taper la formule de la probabilité gaussienne l'avantage de l'utilisation de cette fenêtre est de calculer simultanément la probabilité de dépassement de plusieurs seuils.

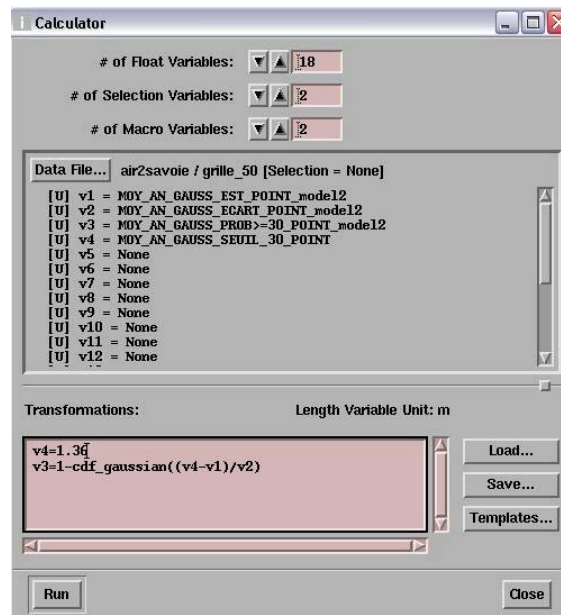


Figure 11 : Calcul de la probabilité de dépassement d'un seuil ponctuelle

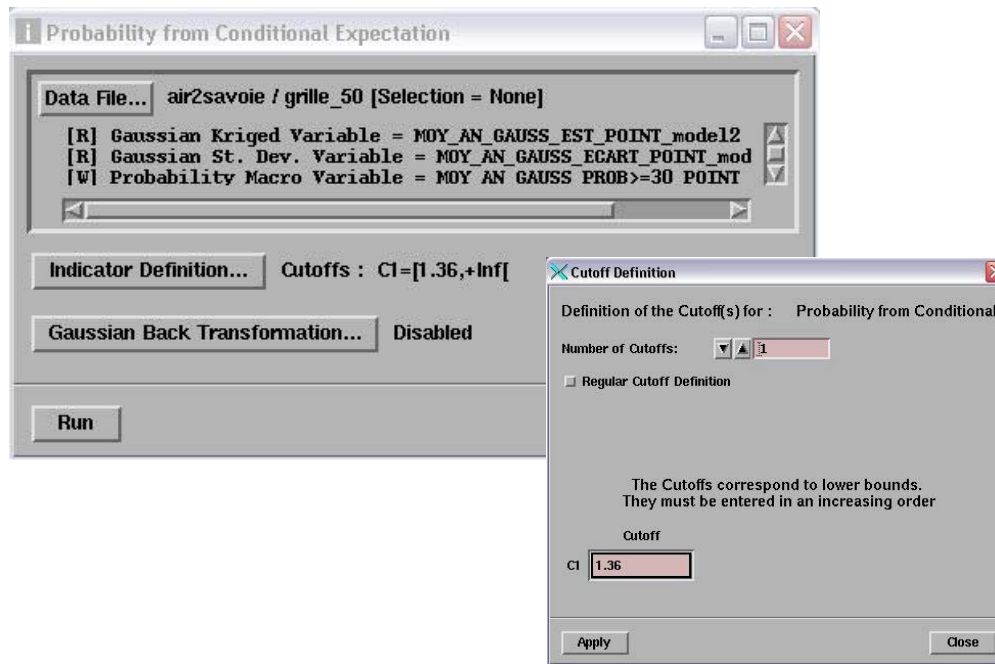


Figure 12 : Menu permettant le calcul direct de la probabilité de dépassement ponctuelle

La carte de la probabilité de dépasser un seuil de $30 \mu\text{g}/\text{m}^3$ est affichée dans la Figure 13, les courbes de niveau représentent les lignes d'isoprobabilité à 10%, 20%, 30%, 40% et 50%. En particulier la courbe d'isoprobabilité 50% entoure les deux points extrêmes repérés en bleu dans la Figure 3, la valeur de ces deux points est de 35 et $38.5 \mu\text{g}/\text{m}^3$, il est donc normal que dans cette zone la probabilité de dépasser $30 \mu\text{g}/\text{m}^3$ soit la plus forte.

Dans la figure on a aussi ajouté les courbes de niveau représentant un écart type d'estimation (ponctuelle par krigeage simple) égal à 0.7 et de 0.8 (comme dans la Figure 9). Dans ces régions les calculs sont supposés plus fiables.

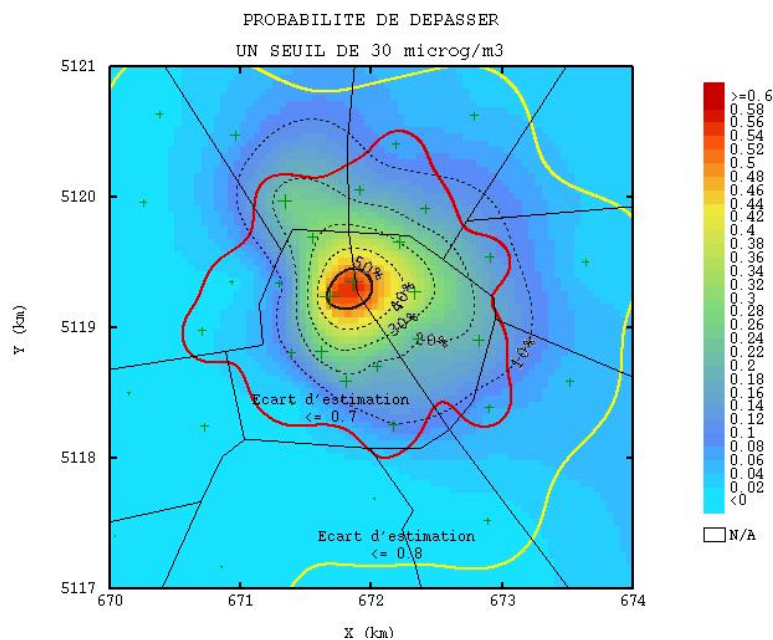


Figure 13 : Carte de probabilité de dépassement d'un seuil de 30 µg/m³ pour le dioxyde d'azote par la méthode de l'espérance conditionnelle (commune de Bourg-en-Bresse, année 2001)

4.7. METHODE DE CONDITIONNEMENT CONFORME

4.7.1. LA QUESTION DU SUPPORT D'ESTIMATION

La géostatistique est une discipline qui s'est développée à l'origine pour répondre à des problèmes posés par l'évaluation des gisements miniers. En particulier les méthodes de la géostatistique non linéaire ont été développées pour pouvoir prédire les résultats d'une exploitation sélective. Les gisement miniers sont exploités par des unités de production dénommées « Panneaux », (dont la dimension peut être par exemple 100*100*5 m³), ils sont divisés à la fois en unités basiques de production, appelées « blocs », ces blocs (dont la dimension peut être de 10*10*5 m³) représentent la quantité de matériel qu'une machine peut transporter. La sélectivité consiste à déterminer quels blocs doivent être envoyés à la usine de production pour exploitation ou à la décharge. Pendant l'exploitation, plusieurs échantillons (dont les dimensions sont de quelques cm³) sont pris dans chaque bloc. La sélection est effectuée uniquement pour les blocs dont les mesures ont en moyenne une concentration supérieure à un seuil (cut-off). Ce seuil est déterminé de façon à obtenir le profit maximal.

Un schéma typique représentant les unités d'exploitation est représenté dans la Figure 14.

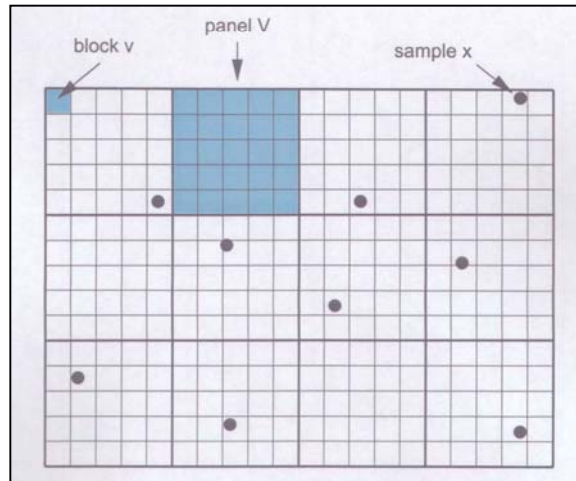


Figure 14 : Schéma montrant différents types de support

Au moment de la prévision on peut s'intéresser soit à l'estimation des réserves (minerai et métal) récupérables pour chacun des blocs (estimation locale), soit à l'estimation des réserves pour un panneau ou pour tout le gisement (estimation globale). Or, sur les échantillons ponctuels les teneurs prennent facilement des valeurs beaucoup plus élevées ou beaucoup plus faibles que sur les blocs, ceux-ci présentant plutôt des teneurs "moyennes" (Figure 15). Naturellement, en moyenne sur l'ensemble du domaine, la teneur du gisement est la même, que l'on divise celui-ci en petits échantillons ou en blocs.

Le problème du changement de support consiste donc à effectuer des estimations sur des blocs à partir de mesures représentant un support ponctuel.

Il est intéressant d'observer que la réglementation de la qualité de l'air (Directives Européennes ou Décrets Nationaux) ne fait pas de distinction sur le support spatial qui doit être utilisé dans le calcul de dépassements de seuil ou pour l'élaboration de cartographies en général. Il est juste indiqué que les niveaux de concentration doivent être évalués sur l'ensemble du territoire.

Dans le cas de mesures par tubes passifs, le support de mesure est quasi-ponctuel, car la zone de représentativité d'un tube passif est très petite par rapport à l'échelle des cartes interpolées (cartes d'une agglomération, par exemple).

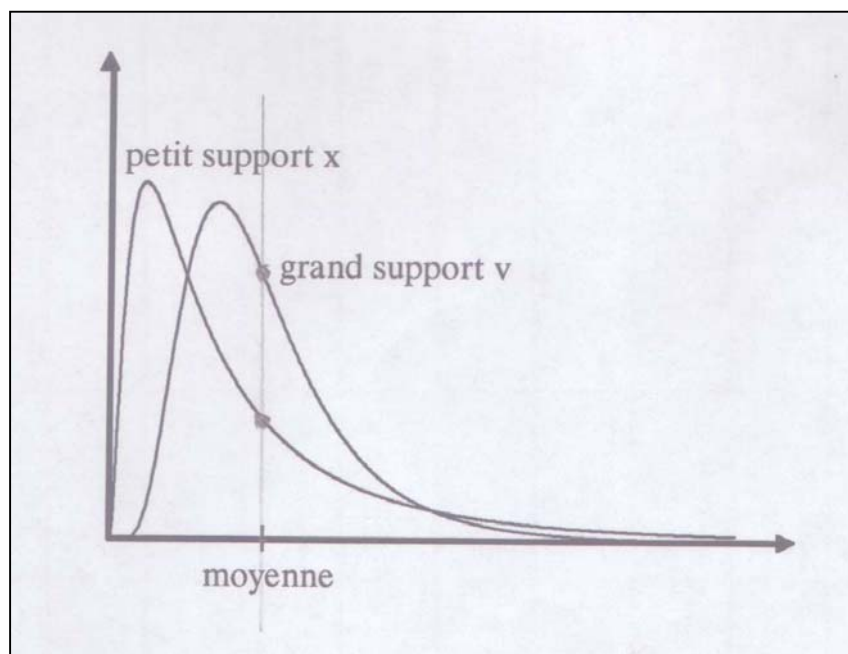


Figure 15 : Histogramme des différents types de supports

Dans « l'étude d'Incertitudes » (LCSQA, 2003), des comparaisons des statistiques sur des estimations menées sur un support ponctuel et sur un support de bloc avaient été présentées. Dans ce rapport on confirme que l'intervalle des valeurs estimées est plus large pour les estimations ponctuelles ainsi que les variances de l'erreur d'estimation, *il y a donc théoriquement plus d'incertitudes dans le calcul d'une concentration ponctuelle par rapport à une concentration moyenne sur un bloc.*

Une variante de la méthode de l'espérance conditionnelle est le « Conditionnement Uniforme ». Ce développement fait partie des méthodes qui ont été créées à l'origine pour l'industrie minière : le but était de calculer la proportion des blocs à l'intérieur des panneaux que dépassent un seuil de référence (Figure 14).

L'avantage de cette méthode est qu'elle permet de s'affranchir des hypothèses de stationnarité stricte et d'inclure une composante non stationnaire dans le calcul de la probabilité. L'adaptation de cette méthode dans le contexte de la qualité de l'air est présentée par la suite.

4.7.2. LE MODELE GAUSSIEN DISCRET

Pour effectuer le calcul de probabilité de dépassements sur des blocs il faut connaître l'anamorphose gaussienne des blocs (Φ_v, v :bloc). Le fait de connaître cette transformation équivaut à connaître la distribution des blocs ; pour ce calcul, le modèle gaussien discret est utilisé.

Ce modèle repose sur l'hypothèse que le couple $(Y_x, Y_v, x$: ponctuelle, v : bloc), suit une loi bi-gaussienne (combinaison linéaire de deux variables est aussi gaussienne), en conséquence l'anamorphose des blocs peut être calculée à partir de l'anamorphose ponctuelle (Φ_x) et d'un coefficient de changement de support:

$$Z_v = \Phi[Y_v] = f_0 + \{f_1 * r^1 * H_1[Y_x]\} + \{f_2 * r^2 * H_2[Y_x]\} + \dots = \sum_{n=0}^{+\infty} \{f_n * r^n * H_n[Y_x]\}$$

Cette expression signifie que l'anamorphose des blocs est obtenue en multipliant les polynômes d'Hermite et les coefficients des polynômes d'Hermite ponctuels par le coefficient de changement de support élevé à la puissance n, n étant égal au degré du polynôme, le premier polynôme (f_0) reste inchangé, comme dans le cas de l'anamorphose ponctuelle, on a les expressions suivantes :

$$f_0 = E[\Phi[Y_x]] = E[\Phi[Y_v]]; \quad Var[Z_v] = Var[\Phi[Y_v]] = \sum_1^{\infty} (f_n * r^n)^2$$

A partir de la dernière expression le coefficient de changement de support peut être calculé :

$$Var [Z_v] = \sum_1^{\infty} (f_n * r^n)^2 \qquad Var [Z_x] = \sum_1^{\infty} (f_n)^2$$

$$Var [Z_v] = Var [Z_x] - \bar{\gamma}(v, v) \qquad \sum_1^{\infty} (f_n)^2 - \bar{\gamma}(v, v) = \sum_1^{\infty} (f_n * r^n)^2$$

Ces formules impliquent la connaissance du modèle du variogramme ajusté à la variable réelle. Pour calculer le variogramme moyen des blocs $[\bar{\gamma}(v,v)]$, la méthode utilisée consiste à diviser le bloc en petits blocs, puis à calculer tous les variogrammes entre ces petits blocs et à prendre la moyenne de ces variogrammes.

Le coefficient de changement de support est compris entre 0 et 1, si le support est ponctuelle le coefficient est égal à 1.

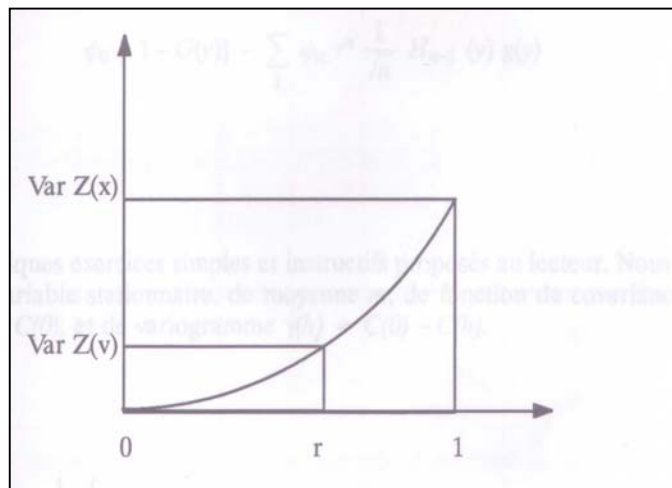


Figure 16 : Coefficient de changement de support en fonction de la variance de bloc

Dans la section suivante nous allons illustrer l'utilisation pratique de ces formules, à l'aide de données de la concentration de l'année 2001 du dioxyde d'azote de l'agglomération de Bourg-en-Bresse.

4.7.3. MISE EN ŒUVRE DE LA METHODE DU CONDITIONNEMENT UNIFORME

4.7.3.1 DEFINITION

Cette méthode est une variante de la méthode de l'espérance conditionnelle, cette fois le conditionnement est effectué sur les concentrations supposées connues sur des zones étendues, appelées par la suite panneaux. L'objectif est de calculer la proportion des blocs à l'intérieur des panneaux que dépassent un seuil de référence (Figure 14).

L'expression que permet le calcul de cette proportion (ou probabilité) est la suivante :

$$P[Z(v) \geq \text{seuil} \mid Z(V)] = E[I_{Z(v) \geq \text{seuil}}] = E[I_{Y(v) \geq \text{seuil}_{\text{gaussien}}}]$$

$$E[I_{Y(v) \geq \text{seuil}_{\text{gaussien}}}] = 1 - G\left(\frac{\text{seuil}_{\text{gaussien}} - [R * Y(V)]}{\sqrt{(1 - R^2)}}\right)$$

Dans cette expression « v », symbolise le support des blocs et « V » celui des panneaux, « Y(V) » représente la transformée gaussien de la concentration du polluant dans les panneaux, et « R » symbolise le coefficient de changement de support entre blocs et panneaux.

Si la concentration dans les panneaux n'est pas connue, elle peut être remplacée par son estimation. On effectue une estimation de bloc de la concentration du polluant dans les panneaux, ensuite on transforme l'estimation en variable gaussienne à l'aide d'une anamorphose gaussienne, de façon à obtenir « Y(V) ».

Cette estimation peut prendre en compte une variable auxiliaire comme dérive externe ou comme variable secondaire dans un cokrigeage. De cette façon, le calcul de la probabilité de dépassement d'un seuil peut être affiné.

Le coefficient de changement de support bloc - panneau (« R ») peut être calculé à partir des deux coefficients point – bloc (r) et point – panneau (r'), par l'expression suivante :

$$R = \frac{r'}{r}$$

On passe donc du bloc au panneau de la même façon que l'on était passé du point au bloc.

Dans le contexte de la qualité de l'air, les panneaux peuvent être concrètement représentés par une grille d'estimation : pour Bourg-en-Bresse, on conserve la même maille de 50 m de côté utilisée dans le cas de l'espérance conditionnelle (Figure 13).

Dans le contexte minier la taille des blocs est choisie par rapport aux unités basiques de production, représentant la quantité de matériel qu'une machine peut transporter. En qualité de l'air on ne peut pas appliquer un critère similaire, le but étant d'effectuer une cartographie et non pas une sélection des blocs à exploiter. Il n'y a pas donc de critère particulier à respecter, par exemple pour Bourg-en-Bresse une taille des blocs de 1 m² a été choisie ; de cette façon on calcule la proportion des mailles de 1 m² (à l'intérieur des mailles de 2500 m²) dont la concentration du dioxyde d'azote dépasse un seuil de 30 µg/m³.

4.7.3.2 KRIGEAGE EN DERIVE EXTERNE DU DIOXYDE D'AZOTE

Un krigeage en dérive externe de la moyenne du dioxyde d'azote a été effectué. Le modèle utilisé est celui qui a été ajusté dans la section 4.6.1 (effet de pépite de 5, une structure sphérique de petite portée 0.6 Km et de palier 7.5 et une structure gaussienne de grande portée 2.5 Km et de palier 9).

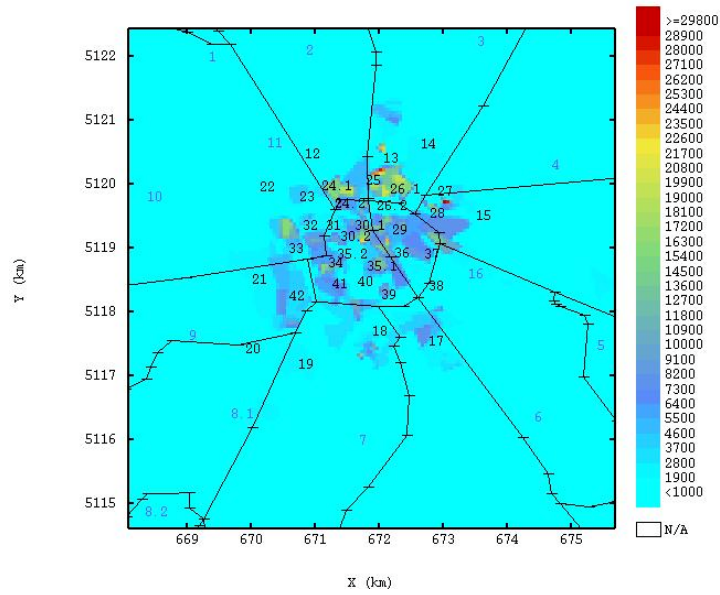
La densité de population a été prise comme dérive externe afin de modéliser la non stationnarité due à la présence de fortes concentrations au centre ville. Cette variable a été calculée à l'aide d'un SIG sur la maille de 50 mètres de côté, les estimations seront réalisées dans cette maille.

Pour calculer les valeurs de densité de population sur la maille, l'information existante au niveau des îlots (chaque îlot représente un ou plusieurs quartiers d'une commune) est considérée. Le nombre d'habitants et la surface en kilomètres carrés des îlots composant l'agglomération de Bourg-en-Bresse, sont donc connus.

La Figure 17 montre la maille de la densité de population. Nous observons que les valeurs les plus fortes sont localisées en centre ville et que les zones périphériques ont une valeur de population quasi nulle.

Aux points où sont situés les tubes on a attribué la valeur de densité de population de la maille la plus proche (migration). La Figure 18 montre le nuage de corrélation de la population « migrée » et de la concentration annuelle du dioxyde d'azote en Bourg-en-Bresse.

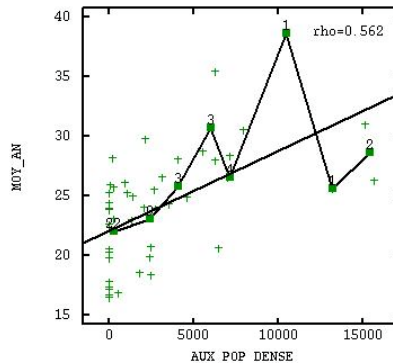
Le coefficient de corrélation est de 0.56. La plupart des concentrations des tubes se trouve dans les deux premières classes de densité de population prises en compte, c'est-à-dire jusqu'à une valeur de 5 000 hab/Km² environ. Aux trois dernières classes de densité ne sont associés que quatre tubes passifs. On observe que les zones de plus forte densité de population sont soumises à des concentrations plutôt moyennes pour toutes les saisons.



Les numéros correspondent aux échantillonneurs passifs

Les lignes représentent le réseau routier

Figure 17 : Carte de la maille de densité de population



Les lignes brisées représentent la moyenne par classe de densité de population, les numéros indiquent le nombre des tubes pris en compte dans le calcul de chaque moyenne

La ligne droite représente la régression linéaire de la concentration en fonction de la densité de population

Figure 18 : Nuages de corrélation entre les moyennes saisonnières du dioxyde d'azote et la densité de population

Après avoir effectué le krigeage en dérive externe, les mailles où l'écart -type d'estimation est inférieur ou égal à $3.5 \mu\text{g}/\text{m}^3$ sont sélectionnées (Figure 19). Cet écart- type équivaut à une variance d'estimation de 12.25 (Tableau 4) . Ce seuil, choisi arbitrairement, permet de repérer les zones où les estimations sont plus précises, celles-ci correspondent aux zones urbaines.

Tous les données	Nb de cellules	Minimum	Maximum	Moyenne	Ecart- type	Variance	CV
Krigeage ordinaire des panneaux en dérive externe	23864	17.62	38.73	21.81	2	4.01	0.09
Variance d'estimation	23864	3.26	27.92	13.92	2.68	7.21	0.19
Variance d'estimation <= 12.25	14.7 % des panneaux						
Krigeage ordinaire des panneaux en dérive externe	3 517	19.65	37.92	24.89	2.74	7.48	0.11
Variance d'estimation	3 517	3.26	12.25	9.24	2.06	4.26	0.22

Tableau 4: Statistiques du Krigeage en dérive externe des panneaux

La zone sélectionnée représente 15% des panneaux, la moyenne sur cette sélection (24.9 µg/m³) est légèrement supérieure à la moyenne sur l'ensemble du domaine (21.8 µg/m³) mais est plus proche de la moyenne expérimentale des données (23.1 µg/m³).

Les valeurs extrêmes des concentrations estimées sont plus petites que celles des données expérimentales. Ce résultat est dû à l'effet de lissage de l'interpolation et à l'effet de support, car les estimations représentent une surface plus grande (il y a donc plus de chance de retrouver de valeurs intermédiaires).

La variance ponctuelle qui était de 23.1 (µg/m³)² est passée à un peu plus de 4.0 (µg/m³)² pour toutes les mailles de 2500 m² et à 7.5 pour la sélection de 15% des mailles.

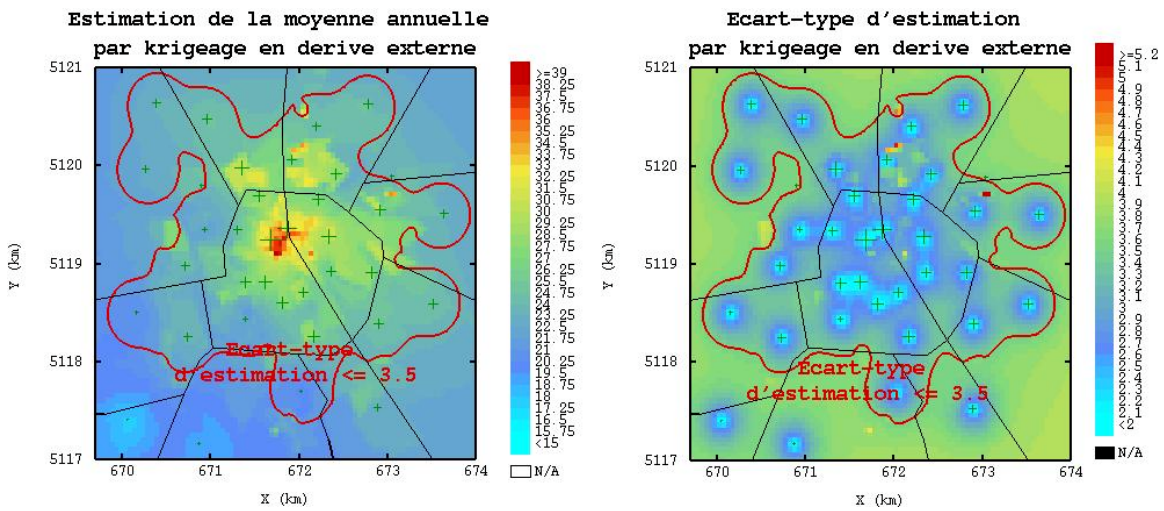


Figure 19 : Krigeage Ordinaire des Panneaux

4.7.3.3 CALCUL DES COEFFICIENTS DE CHANGEMENT DE SUPPORT

Pour calculer les coefficients de changement de support nous utilisons la même fenêtre que celle utilisée lors du calcul de l'anamorphose gaussienne de la section 4.6.2 (Figure 4).

A. Coefficient de Changement de Support Point-Blocs

- Nous sélectionnons l'option «Calculation with Bloc Support », le bouton « Calculate » est alors activé, il donne accès à un menu permettant de calculer le coefficient de changement de support (Figure 20).
- Par défaut l'option « Bloc correction » est cochée, l'autre option « Kriged Panel Correction » est utilisée pour le calcul du coefficient de changement de support points-panneaux.
- Il faut rentrer le variogramme ponctuel des données réelles (ajusté dans la section 4.6.1, Figure 3) et les dimensions du bloc ; *dans notre cas la taille des blocs est de 1 m²*.
- Pour le calcul du variogramme moyen de bloc on divise le bloc en un nombre déterminé de petits blocs et on calcule le variogramme à l'intérieur des petits blocs résultants. Le variogramme moyen sera la moyenne des variogrammes calculés. Les paramètres NX et NY aident à déterminer le nombre de subdivisions à effectuer. Nous avons choisi cinq subdivisions pour chaque axe, ce qui donne une discrétisation du bloc en 25 petits blocs.
- Les options « Bloc Partition » et « Bloc discretization » nous permettent de choisir la façon dont les variogrammes entre les petits blocs vont être calculés, la première possibilité calcule le variogramme entre le centre de tous les blocs ; dans cette méthode il est possible d'avoir des distances nulles (cas où on calcule le variogramme dans le même bloc), et il n'y a qu'une seule possibilité d'effectuer le calcul, on obtient donc une seule valeur numérique ; cette méthode est conseillée quand le variogramme a été calculé sur un support proche du support des blocs.
- Pour la seconde possibilité le variogramme est calculé différemment, cette fois le premier point de chaque couple est le centre d'un petit bloc et le second point est un point localisé aléatoirement à l'intérieur d'un autre bloc, il n'y a pas de possibilité que le second point coïncide avec le premier point et donc le variogramme n'est pas calculé pour une distance nulle.
- Comme les points sont localisés au hasard on peut obtenir différents résultats, il est donc conseillé de lancer plusieurs calculs (bouton « calculate » de la Figure 20) et de vérifier que les résultats ne sont pas trop différents ; si tel est le cas il faut faire varier les paramètres NX et NY et ressayer.

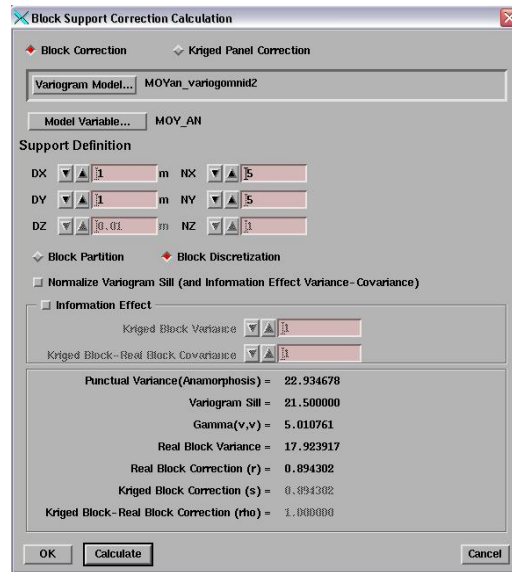


Figure 20 : Fenêtre « Bloc Support Correction » du logiciel Isatis, utilisée pour le calcul du coefficient de changement de support Point- Bloc

- Dans notre cas nous avons conservé un variogramme moyen de bloc de 5.01, la variance de bloc est donc égale à 17.9.
- Un variogramme moyen de bloc plus grand que la variance expérimentale des données réelles provoque une erreur au moment du calcul du coefficient de changement de support (Figure 16).
- Pour éviter cet inconvénient il est proposé d’effectuer une correction du variogramme moyen de bloc en le multipliant par un coefficient représentant le rapport existant entre la variance expérimentale et la somme des paliers du modèle. Cette correction est effectuée automatiquement lorsque l’on sélectionne l’option « Normalize Variogram Sill ».
- La variance utilisée dans ce calcul est la variance ponctuelle théorique (calculée à l’aide des coefficients de l’anamorphose ponctuelle : 22,93). Le coefficient de correction est calculé de la façon suivante :

$$\text{si } \Sigma(\text{paliers}) \neq \text{Var}[Z_p], \text{ alors : } \frac{\text{Var}[Z_p]}{\Sigma(\text{paliers})} = \frac{22,93}{21,5} = 1,066$$

- Dans notre cas nous n’avons pas effectué de correction, le coefficient de changement de support (point -bloc de 1 m²) est de 0.894.

- L'ajustement de l'anamorphose de bloc donne des valeurs plus « lisses » que l'anamorphose ponctuelle. Par exemple, une valeur gaussienne égale à 2.0 dans l'anamorphose ponctuelle équivaut à environ 38 $\mu\text{g}/\text{m}^3$ (Figure 7). Dans l'anamorphose de bloc cette valeur correspond à environ 34 $\mu\text{g}/\text{m}^3$ (Figure 24).
- Les valeurs gaussiennes des intervalles absolu et pratique pour l'anamorphose de bloc sont plus grandes (en valeurs absolue) que celles de l'anamorphose ponctuelle. On rappelle que si des transformations des valeurs entre ces deux intervalles ont lieu, les transformations respectives seront calculées par interpolation linéaire.

$$Y_{\min}^a - 5.7 =; Y_{\max}^a = 6.0$$

$$Y_{\min}^p - 3.5 =; Y_{\max}^p = 3.7$$

- Pour sauvegarder le modèle d'anamorphose il faut cocher l'option «Save At Run». Notre but étant de calculer l'anamorphose, nous n'avons pas créé une nouvelle variable gaussienne (Figure 21).
- Une impression des résultats est alors effectuée, on y voit notamment les coefficients de l'anamorphose, le coefficient de changement de support, les intervalles autorisés, absolus et pratiques, si une variable gaussienne a été créée, on affiche ses principales statistiques, la méthode d'inversion choisie et les statistiques de l'écart entre concentrations recalculées et originales.

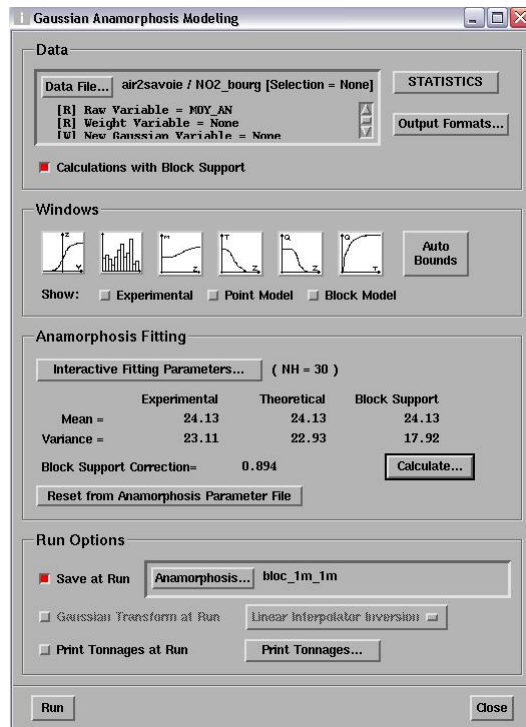


Figure 21 : Fenêtre « Gaussian Anamorphosis Modeling du logiciel Isatis, utilisé pour le calcul de l’anamorphose gaussienne des blocs

B. Coefficient de Changement de Support Point -Panneau

Pour calculer le coefficient point – panneau, nous utilisons toujours le même menu, mais cette fois nous sélectionnons l’option « Kriged Panel Correction » (Figure 22 et Figure 23) où nous devons tout simplement renseigner la variance des estimations des panneaux, la moyenne de cette variance étant de 9.24 (Tableau 4), nous avons décidé de l’arrondir à une valeur de 9.0 (choix arbitraire), le coefficient de changement de support est alors de 0.64.

Les valeurs gaussiennes des intervalles absolu et pratique pour l’anamorphose des panneaux sont plus grandes (en valeurs absolu) que celles de l’anamorphose des blocs:

$$Y_{\min}^a - 8.85 =; Y_{\max}^a = 9.09$$

$$Y_{\min}^p - 6.30 =; Y_{\max}^p = 6.80$$

Pour sauvegarder le modèle d’anamorphose il faut cocher l’option « Save At Run». Notre but étant de calculer l’anamorphose, nous n’avons pas crée une nouvelle variable gaussienne.

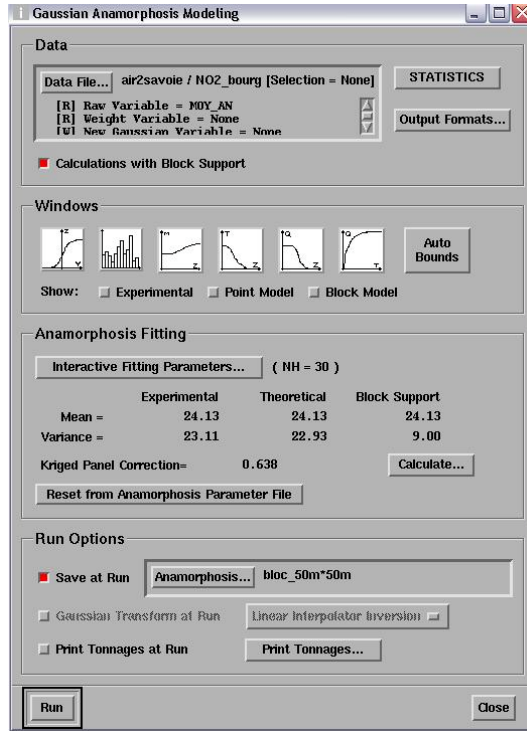


Figure 22 : Fenêtre « Gaussian Anamorphosis Modeling » du logiciel Isatis, utilisée pour le calcul de l'anamorphose gaussienne des panneaux

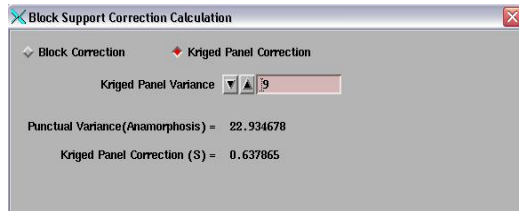
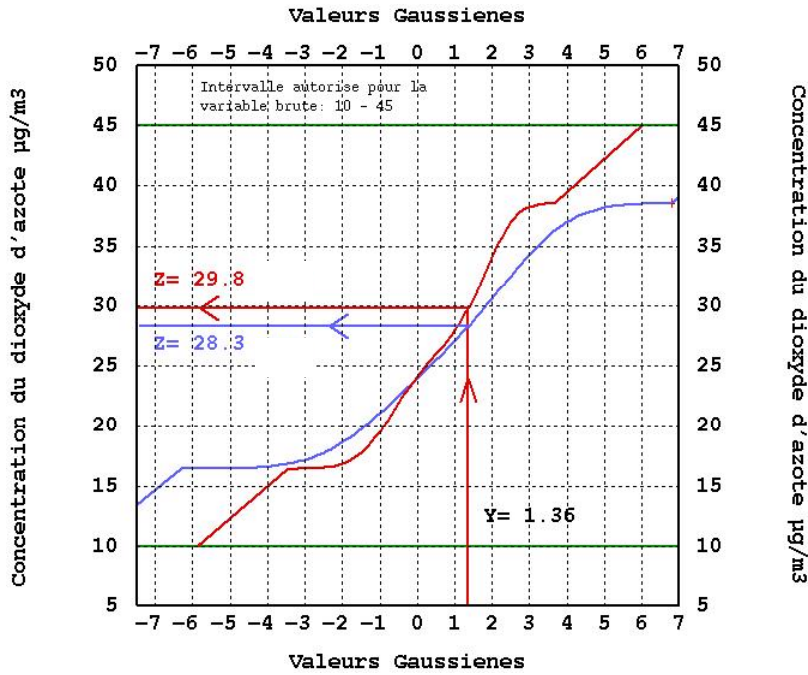


Figure 23 : Fenêtre « Bloc Support Correction » du logiciel Isatis, utilisée pour le calcul du coefficient de changement de support Point- Panneau

Afin de bien apprécier l'effet de support, nous avons superposé les modèles des polynômes d'Hermite des anamorphoses point -bloc et point –panneau (Figure 24).

Dans la Figure 24, l'effet de support est ainsi représenté. Nous observons que plus la taille du support augmente, plus l'intervalle de variation des concentration possibles est faible.

Dans le cas d'un support ponctuel le seuil pour le dioxyde d'azote de $30 \mu\text{g}/\text{m}^3$ correspondait bien à une valeur gaussienne de 1.36, mais si nous considérons un support de 1 m^2 , cette valeur gaussienne correspond maintenant à une concentration de $29.8 \mu\text{g}/\text{m}^3$ ou à une concentration de $29.3 \mu\text{g}/\text{m}^3$ si le support considéré augmente à 250 m^2 .



Courbe bleu : anamorphose des panneaux, courbe rouge : anamorphose des blocs

Figure 24 : Anamorphose de blocs de $1 \times 1 \text{ m}^2$ et des panneaux de $50 \times 50 \text{ m}^2$

Remarque : Il existe une autre manière de calculer le coefficient de changement de support des panneaux, elle consiste à utiliser l'option « Bloc Correction » de la même façon qu'on a calculé le coefficient de changement de support des blocs de 1 m^2 (Figure 20), on rappelle que ce menu utilise le variogramme ponctuel. Cette option a été préalablement testée, en obtenant les résultats suivants :

- un variogramme moyen des panneaux de $5.76 (\mu\text{g}/\text{m}^3)^2$,
- une variance des panneaux de $17.17 (\mu\text{g}/\text{m}^3)^2$,
- et un coefficient de changement de support de 0.87.

Cette variance théorique des panneaux (17.2) est largement supérieure à la variance des estimations des panneaux obtenues par krigeage en dérive externe (environ 9.0). La différence entre ces deux variances (en théorie elles devraient être égales puisqu'il s'agit de la même variable) peut être imputée au fait qu'en plus du variogramme ponctuel, les estimations prennent en compte une variable auxiliaire (la population comme dérive externe). C'est d'ailleurs pour cette raison que l'option « Kriged Panel Correction » a été retenue.

Le coefficient de changement de support (0.87) calculé par l'option « Bloc Correction » est aussi plus élevé que la valeur finalement retenue (0.64).

4.7.3.4 CONDITIONNEMENT UNIFORME

Le coefficient de changement de support blocs- panneaux est alors de 0.72 (=0.64/0.89). Nous pouvons maintenant appliquer l'expression de la section 4.7.3, pour calculer la probabilité de dépassement de seuil.

Cette expression peut être calculée manuellement à l'aide du menu « Calculator » (Figure 25) ou en utilisant le menu « Uniform Conditioning » montré dans la Figure 26.

Pour le calcul manuel, il faut connaître la transformée gaussienne des estimations des panneaux et l'équivalent gaussien de la valeur du seuil ($30 \mu\text{g}/\text{m}^3$). Pour y accéder nous utilisons le menu « Raw to Gaussian Transformation » de la Figure 25.

Pour le calcul du seuil gaussien des blocs nous utilisons l'anamorphose des blocs (« bloc_1m*1m »), pour le calcul des estimations gaussiennes des panneaux nous utilisons l'anamorphose des panneaux (« bloc_50m_50m »).

Le seuil gaussien de blocs calculé est égal à 1.40, nous rappelons que le seuil gaussien ponctuel était de 1.36 (Figure 7).

La variable gaussienne représentant les estimations des panneaux a été nommée : « MOY_AN_DEXT_EST_GAUSS ». Dans la fenêtre « Calculator » de la Figure 25 nous montrons la formule pour calculer la probabilité de dépassement.

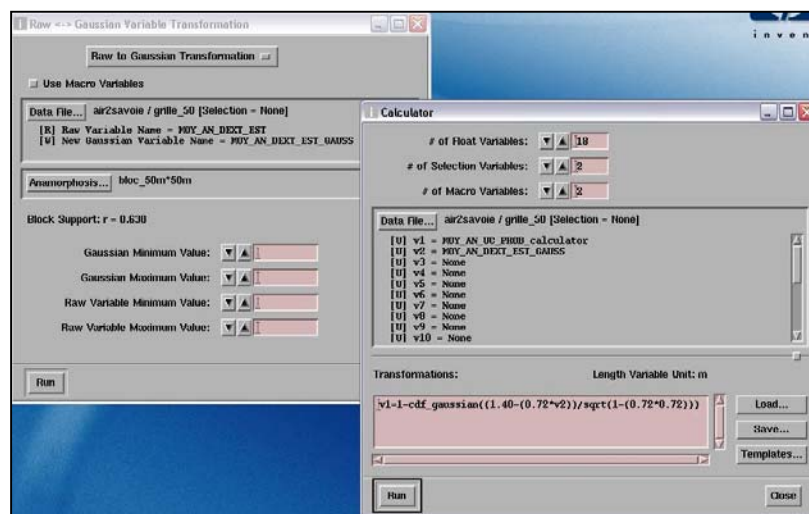


Figure 25 : Calcul manuelle de la probabilité de dépassement de seuil par Conditionnement Uniforme

L'autre façon d'accéder à cette probabilité est d'utiliser directement la fenêtre « Uniform Conditioning » (Figure 26). Ce menu demande à renseigner :

- la variable représentant les estimations des panneaux,
- une variable « erreur » dans le cas où l'option « Tonnage Corrections » est sélectionnée,
- les deux anamorphoses point- bloc et point- panneau,
- la concentration du seuil en valeur brute

- et le nom à donner aux variables résultantes.

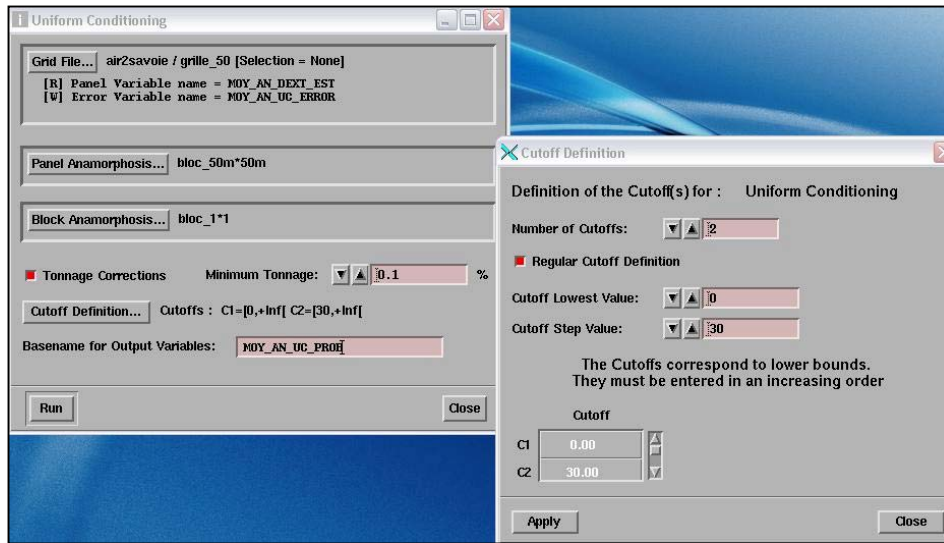


Figure 26 : Menu permettant le calcul de la probabilité de dépassement de seuil par Conditionnement Uniforme

Remarque : La fenêtre « Uniform Conditioning » a l'avantage de permettre le calcul de la probabilité de dépassement de plusieurs seuils à la fois et de calculer deux autres fonctions non-linéaires de la concentrations des blocs, à savoir :

La « Teneur Moyenne » : c'est la moyenne de la concentration des blocs de 1 m² que dépassant le seuil à l'intérieur des panneaux, l'expression de calcul est la suivante :

$$Teneur Moyenne = m(z) = E[Z(v) / Z(v) \geq seuil]$$

La « quantité de Métal » : c'est la concentration des blocs de 1 m² que dépassent le seuil à l'intérieur des panneaux, moyennée sur le nombre total des blocs que forment le panneaux, l'expression de calcul est la suivante :

$$Métal = Q(z) = E[Z(v) * I_{Z(v) \geq seuil}]$$

L'étude de la pertinence du calcul de ces deux quantités dans le contexte de la qualité de l'air n'a pas été approfondie dans ce rapport.

L'option «Tonnage Corrections» permet de corriger automatiquement les probabilités lorsque les résultats sont en dehors d'un intervalle compris entre 0 et 1. L'option « Minimum Tonnage » établie le seuil à appliquer, par exemple une valeur choisie de 0.01% indique que si la probabilité calculée est inférieure ou égale à 0.01%, la valeur est approchée à 0, de même si la probabilité calculée est supérieure ou égale à 99.99%, la probabilité est approché à 100% ; à chaque fois qu'une probabilité est corrigée, la variable « erreur » est renseignée avec la valeur 1 (dans notre cas, le logiciel a effectué la correction sur 284 mailles où la probabilité calculée était égale à 0.01%).

La Figure 27 montre la carte de la probabilité avec laquelle la concentration des blocs de 1 m² dépasse un seuil de 30 µg/m³ à l'intérieur des panneaux (mailles de 2500 m²) ; dans cette figure sont indiquées d'une part la sélection des estimations les plus fiables (écart-type de krigeage ordinaire des panneaux en dérive externe <= 3.5, Figure 19), d'autre part les courbes de 50% de probabilité : dans ces quartiers, le risque de dépasser le seuil est plus grand que 50%.

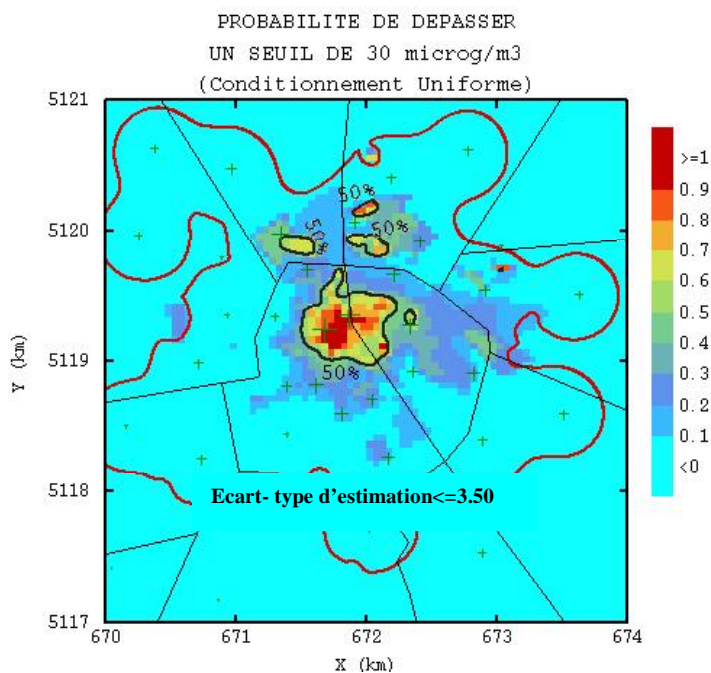


Figure 27 : Probabilité de dépassement de seuil par Conditionnement Uniforme

Probabilité de dépasser 30 µg/m ³ par conditionnement uniforme	Nb de cellules	Minimum	Maximum	Moyenne	Ecart- type	Variance	CV
Tout le domaine	23864	0%	100%	2%	8%	100% ²	4.24
Ecart- type d'estimation <= 3.50	3552			10%	18%	300% ²	1.67

Tableau 5 : Statistiques de la probabilité de dépasser 30 µg/m³ (conditionnement uniforme)

La Figure 28 permet de comparer visuellement les deux méthodes de calcul de la probabilité de dépassement employées, à gauche, la méthode du « Conditionnement Uniforme » (estimations sur des blocs de 1 m²) , à droite, « l'Espérance Conditionnelle » (estimations ponctuelles), pour cette dernière méthode la probabilité maximale de dépassement est de 60%, en revanche pour le « Conditionnement Uniforme » en arrive à calculer des probabilités de 100%.

La carte de probabilité calculée par la méthode du « Conditionnement Uniforme » est plus détaillée. Cette carte prend en compte le tissu urbain introduit indirectement par la densité de population lors du krigeage en dérive externe ; cette méthode permet donc de s'affranchir de l'hypothèse de stationnarité stricte nécessaire dans le cas de l'espérance conditionnelle.

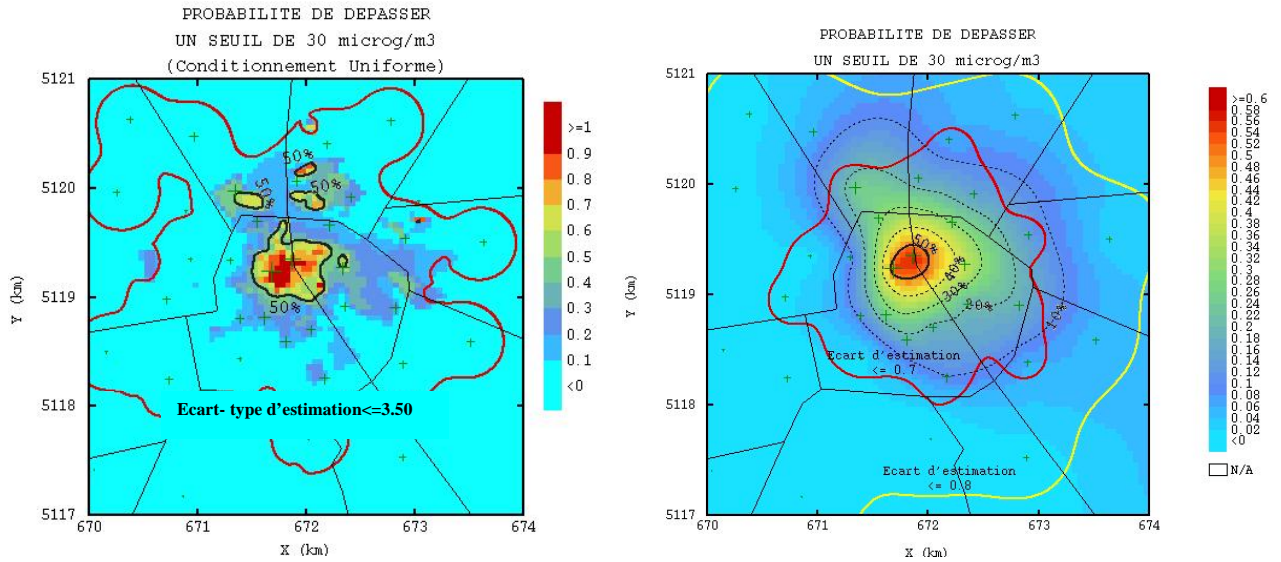


Figure 28 : Comparaison des deux cartes de probabilité de dépassement obtenues par « Conditionnement Uniforme » et par « l'Espérance Conditionnelle »

4.8. SIMULATIONS CONDITIONNELLES

4.8.1. DEFINITION

Les techniques de krigeage permettent d'estimer localement les valeurs prises par une variable régionalisée à partir d'un échantillonnage de cette dernière.

En qualité de l'air il peut s'agir de la concentration d'un polluant mesurée par tubes passifs.

Cependant, la propriété de lissage induite par le krigeage empêche d'apprécier la variabilité des valeurs inconnues (non mesurées) : ainsi, l'histogramme des valeurs estimées est moins dispersé que celui des valeurs réelles ; le variogramme des valeurs estimées n'est pas non plus représentatif de la réalité.

Cette propriété de lissage du krigeage a déjà été évoquée dans cette étude.

Le but des simulations est de reproduire la variabilité spatiale de la variable régionalisée respectant ses propriétés statistiques (histogramme, variogramme, etc.), chaque simulation est alors considérée comme une version (réalisation) possible de la réalité. De même une simulation conditionnelle est une simulation restituant aux points de mesures les valeurs qui y sont connues.

Les simulations conditionnelles sont utiles pour estimer des grandeurs qui ne sont pas directement accessibles par krigeage, *par exemple en qualité de l'air la probabilité de dépassement de seuil peut être évaluée par la fréquence avec laquelle les concentrations simulées dépassent le seuil de référence.*

4.8.2. SIMULATIONS DANS LE MODELE MULTIGAUSSIEN

La construction de simulations est particulièrement simple dans le cadre des fonctions aléatoires de loi spatiale multigaussienne, c'est-à-dire telles que toute combinaison linéaire de valeurs suit une distribution gaussienne.

Aussi, comme dans le cas de l'espérance conditionnelle, simule-t-on le plus souvent non pas directement la variable brute mais sa transformée gaussienne, avec sa propre structure, avant de repasser dans l'espace de travail initial.

De nombreux algorithmes existent pour obtenir des simulations non conditionnelles des fonctions aléatoires de loi spatiale multigaussienne: décomposition matricielle, moyennes mobiles discrètes, méthodes spectrales, bandes tournantes, méthode séquentielle, etc.

Pour les données de Bourg-en-Bresse, la méthode des bandes tournantes a été choisie pour effectuer les simulations non conditionnelles (cf. §4.9).

Le conditionnement à un ensemble de données se résout à l'aide d'un krigeage; en effet, on a en tout point x :

$$Y(x) = Y(x)^K + [Y(x) - Y(x)^K] \quad \text{où :} \quad Y(x)^K = \text{Krigeage de la gaussienne}$$

Dans le cas d'une fonction aléatoire multigaussienne, le résidu en tout point $[Y(x) - Y(x)^K]$ est indépendant des valeurs aux points de données.

En conséquence ce résidu ne dépend pas d'une combinaison linéaire des données, par exemple l'estimation par krigeage : n'importe quelle estimation et le résidu sont donc deux fonctions aléatoires indépendantes non nécessairement stationnaires. Par exemple en un point donné le résidu est zéro ; de ce fait on peut utiliser un krigeage simple, un krigeage ordinaire, un krigeage ordinaire avec dérive externe, ou un cokrigeage (voir page 200 de Lantuejoul, 2002).

L'idée est alors de substituer à ce résidu indépendant mais inconnu un résidu simulé ayant exactement la même structure.

Pour ce faire, on fabrique une simulation non conditionnelle de la variable, soit $Y(x)_s$, sur le domaine considéré, puis on calcule en tout point x , le résidu de son krigeage à partir des valeurs prises par Y_s aux points de données :

$$[Y(x)_s - Y(x)^K_s]$$

La recombinaison :

$$Y(x)_{sc} = Y(x)^K + [Y(x)_s - Y(x)^K_s]$$

donne alors une autre simulation de la fonction aléatoire, mais qui est maintenant conditionnelle : en un point de donnée on retrouve bien la valeur continue :

$$Y(x_i)_{sc} = Y(x_i)^K + [Y(x_i)_s - Y(x_i)^K_s] = Y(x_i) \quad \text{car :} \quad Y(x_i)^K = Y(x_i)$$

En résumé, la méthode de conditionnement d'une fonction aléatoire multigaussienne consiste à effectuer une simulation non conditionnelle de la variable gaussienne (par bandes tournantes par exemple), puis à effectuer deux krigeages: le premier sur la variable gaussienne, le second sur les valeurs de la simulation non conditionnelle aux points des données.

Pour la simulation de la concentration du dioxyde d'azote de Bourg-en-Bresse, l'estimation utilisée a été un krigeage ordinaire avec la densité de population en dérive externe, de même, la transformation gaussienne utilisée a été l'anamorphose gaussienne ponctuelle employée dans la méthode de l'espérance conditionnelle.

4.8.3. LA METHODE DES BANDES TOURNANTES

La méthode est basée sur la décomposition d'un processus stationnaire en processus sinusoïdal et de sa covariance en somme de cosinus¹.

En appliquant la théorie mathématique appropriée et en introduisant le système de coordonnées polaires on peut remplacer la simulation d'une fonction aléatoire gaussienne avec covariance $C(h)$ par la réalisation des simulations unidimensionnelles avec covariance, C_θ , où θ représentant une série de directions dans l'espace.

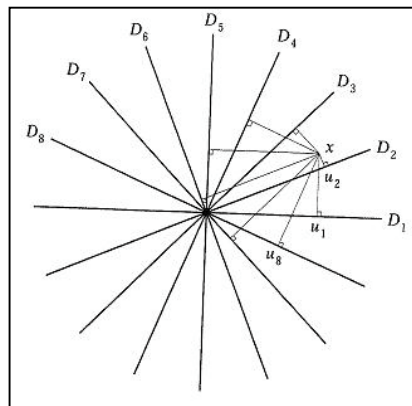
L'algorithme peut se résumer ainsi :

- Tirer uniformément un nombre déterminé de directions θ_i .

¹ cette décomposition peut se faire grâce au théorème de Bochner, lequel établit qu'une covariance spatiale correspond à la transformation de Fourier d'une mesure positive, appelée le spectre.

- Effectuer une simulation le long de chaque droite de direction θ_i , chaque simulation doit respecter chaque covariance C_θ .
- Effectuer un épannage de chacune de ces simulations dans l'espace pour obtenir une simulation de covariance $C(h)$.

La façon dont sont générées les directions θ_i , déterminées les covariances C_θ et dont est construite chaque simulation unidimensionnelle, qui respecte chaque covariance C_θ , est discutée dans Chiles 1999 et Lantuejoul 2002.



La valeur de la simulation dans un point x du plan est égale à la somme des simulations unidimensionnelles effectuées dans les projections u_i sur les droites D_i .

Figure 29 : Le principe des bandes tournantes en 2D

4.9. EXEMPLE PRATIQUE DES SIMULATIONS CONDITIONNELLES

Le but est d'effectuer un nombre déterminé de réalisations de la moyenne annuelle de la concentration de dioxyde d'azote de Bourg-en-Bresse (année 2001). Puis de calculer en tout point du domaine d'estimation le nombre de réalisations que dépassent le seuil de référence ($30 \mu\text{g}/\text{m}^3$). Cette fréquence sera associée directement à la probabilité de dépassement du seuil.

Sur Isatis, la méthode de simulation choisie, parmi celles proposées, est la méthode des bandes tournantes (menu Interpolate/Conditional Simulations/Turning Bands, Figure 30).

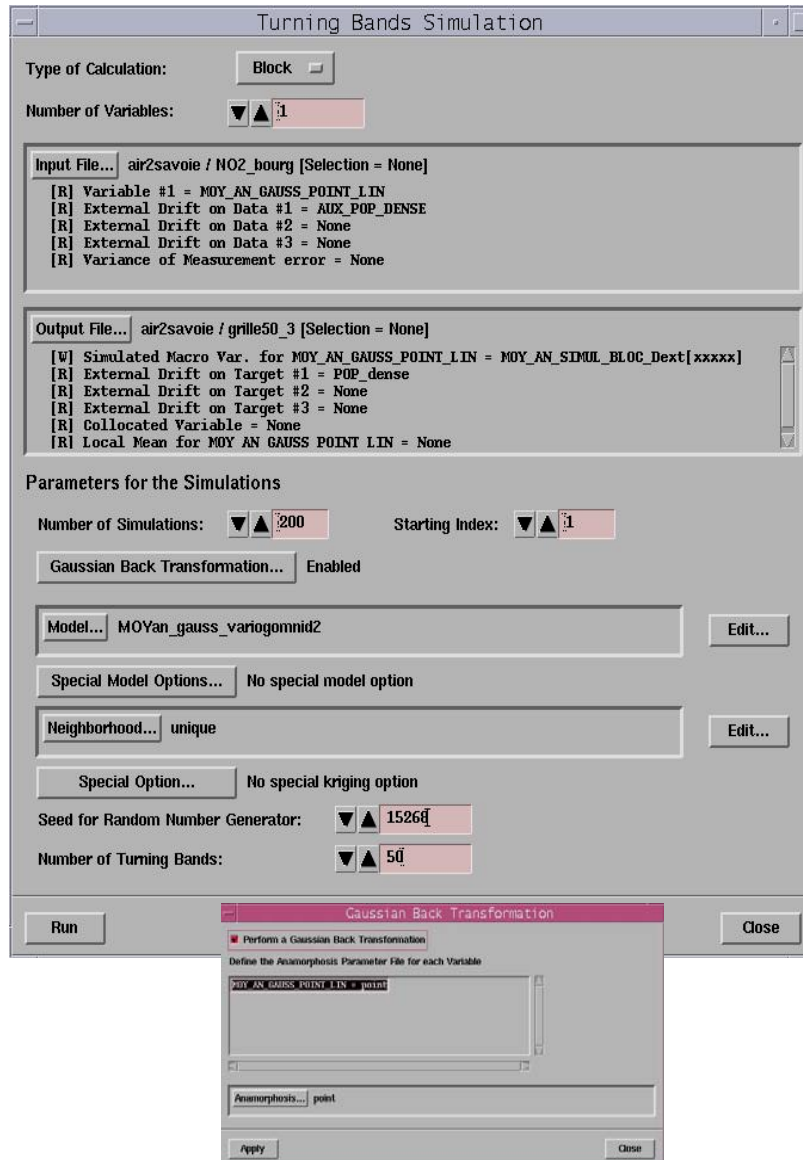


Figure 30 : Menu permettant le calcul des simulations conditionnelles par bandes tournantes

Les informations qui doivent être renseignées dans cette fenêtre sont les suivantes :

- « Type of calculation »: cette option permet le choix de la taille de support sur lequel les estimations (pour le conditionnement) seront réalisées : à savoir, point ou bloc ; *des calculs sur un support de blocs ont été effectués.*
Lorsque l’option bloc est choisie, la variable à renseigner par la suite doit rester une variable ponctuelle (avec sa respective fonction d’anamorphose ponctuelle et son respective modèle de covariance ponctuel). En effet, pour effectuer le krigeage des blocs, ISATIS calcule la covariance moyenne des blocs par la méthode expliquée dans la section 4.7.3.3 (page 36).
- « Number of variables » : il s’agit du nombre de variables à simuler, dans ce cas on a simulé *la transformée gaussienne de la concentration de dioxyde d’azote.*

Cependant le module est capable d'effectuer des cosimulations, dans ce cas la covariance utilisée doit être cohérente avec le nombre de variables : variogramme multivariable.

- « Input file » : on renseigne le nom de la variable à simuler ainsi que le nom de la dérive externe ; *des simulations sur la transformée gaussienne de la concentration de dioxyde d'azote avec la densité de population comme dérive externe ont été calculées.*
- « Output file » : il faut d'abord choisir la maille d'estimation dans laquelle les simulations seront stockées, puis déterminer le nom de la variable qui contiendra les simulations, et enfin il faut sélectionner la maille contenant la dérive externe.
- « Number of simulations » : c'est le nombre de réalisations de la concentration annuelle de dioxyde d'azote qui seront calculées.

Il n'existe pas de critère absolu permettant de définir avec certitude le nombre de simulations qui doivent être effectuées (dans Lantuejoul 2002, page 197 quelques critères théoriques sont données). Il faut savoir que plus on a des simulations, plus les statistiques calculées seront robustes, cependant un grand nombre de simulations peut demander beaucoup de ressources informatiques et de temps de calcul (surtout dans le cas de simulations sur des blocs). Dans le cas de Bourg-en-Bresse, 200 simulations de la concentration de dioxyde d'azote ont été réalisées.

- « Starting Index » : la variable que contient les simulations est appelée sur Isatis « macro variable », car elle contient autant des variables que de simulations. Ces variables sont numérotées dans un ordre croissant au fur et à mesure que les simulations sont stockées. A l'aide de cette option on peut définir l'indice par lequel la numérotation des simulations doit commencer, dans ce cas on a laissé la valeur proposée par défaut : 1.
- « Gaussian Back Transformation... » : Lorsque on a renseigné une variable gaussienne dans « Input file », et que l'on est intéressé à stocker les simulations non pas en valeur gaussiennes mais en valeurs réelles, on doit sélectionner, dans cette onglet, le nom de l'anamorphose gaussienne à utiliser.

Dans la partie inférieure de la Figure 30 est signalé le nom de l'anamorphose utilisée, à savoir : l'anamorphose des points (« point ») ajustée lors de la mise en œuvre de l'espérance conditionnelle. cf §4.6.2).

- « Model » : il s'agit du modèle de variogramme ajusté sur la transformé gaussienne de la concentration annuelle du dioxyde d'azote. Le modèle est celui présenté dans la Figure 8 avec en plus la densité de population comme dérive externe.
- « Neighborhood »: on détermine ici, le voisinage d'estimation à utiliser lors du conditionnement par krigeage. Si l'option « bloc » a été choisie dans « Type of calculation » on pourra choisir, dans ce menu, le nombre de blocs à utiliser pour le processus de discrétisation. Un nombre des blocs trop grand entraînera une augmentation considérable du temps de calcul, dans notre cas nous avons choisi cinq subdivisions pour chaque axe (NX et NY), ce qui donne une discrétisation du bloc en 25 petits blocs.
- « Seed for Random Number Generator » : Pour générer des nombres aléatoires, Isatis utilise un générateur dit « congruentiel », cette méthode génère une suite de nombres par une formule déterministe de manière à obtenir une suite qui semble aléatoire (indépendance et distribution uniforme dans l'intervalle de variation), l'expression est la suivante :

$$X_i = (a * X_{i-1}) \bmod m$$

X_{i-1} = seed (graine)
 $a=105$
 $m= 20000159$
 (X module m) : reste de la division entière de x par m.

Le graine (seed) permet d'initialiser le générateur, un numéro est proposé automatiquement par Isatis à chaque fois que le menu (Interpolate/Conditional Simulations/Turning Bands, Figure 30) est lancé. Toutefois si l'on doit faire des traitements statistiques portant sur différents jeux de simulations, il est alors nécessaire de changer le graine de façon à obtenir une suite différente à chaque exécution. Il est conseillé de conserver le numéro proposé par défaut pour Isatis, dans ce cas 15268, ou de choisir un nombre assez large (5 ou 6 digits).

- « Number of Turning Bands » : Le nombre de directions θ_i tirées uniformément doit être suffisant pour balayer tout l'espace; dans la pratique on doit en utiliser quelques centaines.

Toutefois pour des raisons de temps de calcul nous n'avons choisi qu'une cinquantaine des bandes tournantes, les résultats ont été un peu pénalisés par ce choix.

D'autres options, qui n'ont pas été utilisées, sont aussi disponibles, par exemple la possibilité d'effectuer un krigeage simple, d'utiliser une variance de l'erreur de mesure ou d'effectuer un cokrigeage colocalisée.

- Une fois que les options nécessaires pour effectuer les simulations sont renseignées, on procède à l'exécution en cliquant sur le bouton « RUN », le temps de calcul est bien sûr fonction des ressources informatiques disponibles : vitesse de calcul du processeur, quantité de mémoire vive disponible, etc., mais aussi du nombre de simulations à effectuer et du nombre de bandes tournantes sélectionné.

A titre d'information, la version windows du logiciel Isatis utilisée était installée sur un ordinateur portable Hewlett Packard avec 192 Mo de RAM et dont le processeur est un Intel Pentium de 2.8 GHz. », avec cette configuration le temps utilisé pour effectuer les 200 simulations de bloc de la moyenne annuelle de dioxyde d'azote par 50 bandes tournantes, a été d'à peu près 24 heures (1j).

La Figure 31 montre cinq des 200 simulations conditionnelles calculées, ces cartes représentent cinq versions possibles de la distribution spatiale de la moyenne de l'année 2001 de la concentration de dioxyde d'azote de l'agglomération de Bourg-en-Bresse.

Dans les zones urbaines les concentrations se trouvent en général dans un intervalle situé entre 25 et 35 $\mu\text{g}/\text{m}^3$; si l'on regarde attentivement les cartes, on s'aperçoit de l'existence de quelques « bandes » ou lignes, ce phénomène est propre à la méthode de simulation utilisée, il faudrait en effet augmenter le nombre des bandes tournantes pour les éliminer.

En théorie la moyenne des simulations doit être égale au krigeage et l'écart- type doit être égal à l'écart- type de krigeage. Les statistiques du Tableau 6 sont effectivement très proches des statistiques de l'estimation par krigeage en dérive externe effectué dans la section 4.7.3.2 (Tableau 4).

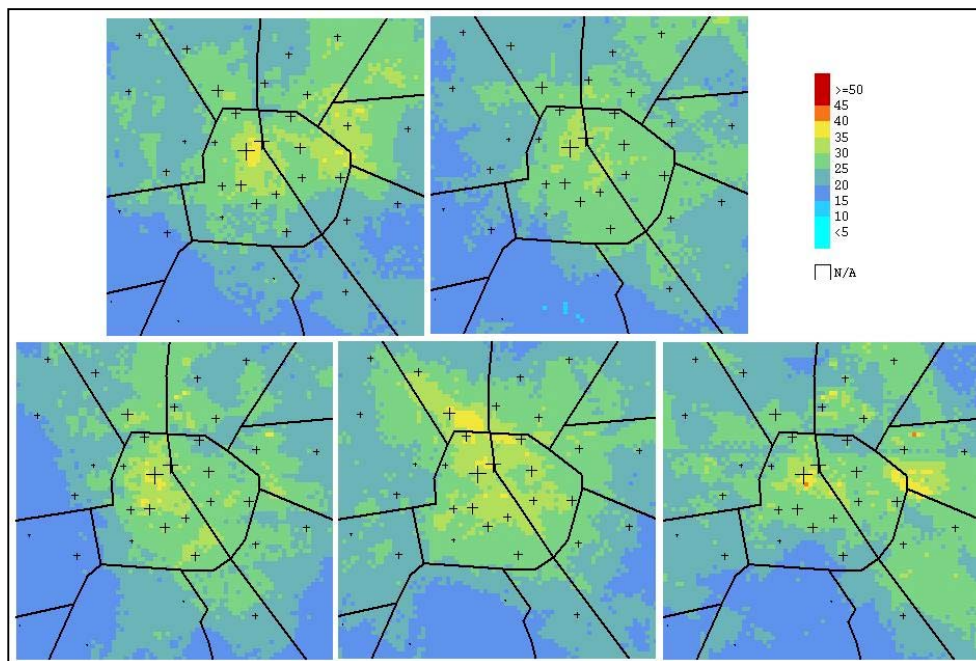


Figure 31 : Cinq simulations conditionnelles par bandes tournantes de la concentration de dioxyde d'azote avec la densité de population en dérive externe

Statistique des simulations	Nb de cellules	Minimum	Maximum	Moyenne	Ecart- type	Variance	CV
Moyenne	5 929	18.03	35.82	23.82	3.03	9.20	0.13
Ecart- type		1.37	5.30	2.29	0.32	0.10	0.14

Tableau 6 : Statistiques de la moyenne et l'écart-type des 200 simulations conditionnelles par bandes tournantes

La Figure 32 montre le menu qui permet de calculer la probabilité de dépassement (Tools/Simulation Post-processing) ; dans ce menu il faut d'abord choisir la « macro » variable contenant les simulations, puis il faut cliquer sur le bouton « Edit... » de l'option « Iso-Cutoff Maps » et renseigner la valeur du seuil de référence.

Dans cette option on peut calculer en même temps la probabilité de dépasser plusieurs seuils, ainsi que d'autres fonctions comme la probabilité de ne pas dépasser le seuil ou la moyenne des concentrations au-dessous ou au-dessus du seuil de référence. Dans le cas de l'agglomération de Bourg-en-Bresse, un seul seuil a été appliqué : $30 \mu\text{g}/\text{m}^3$.

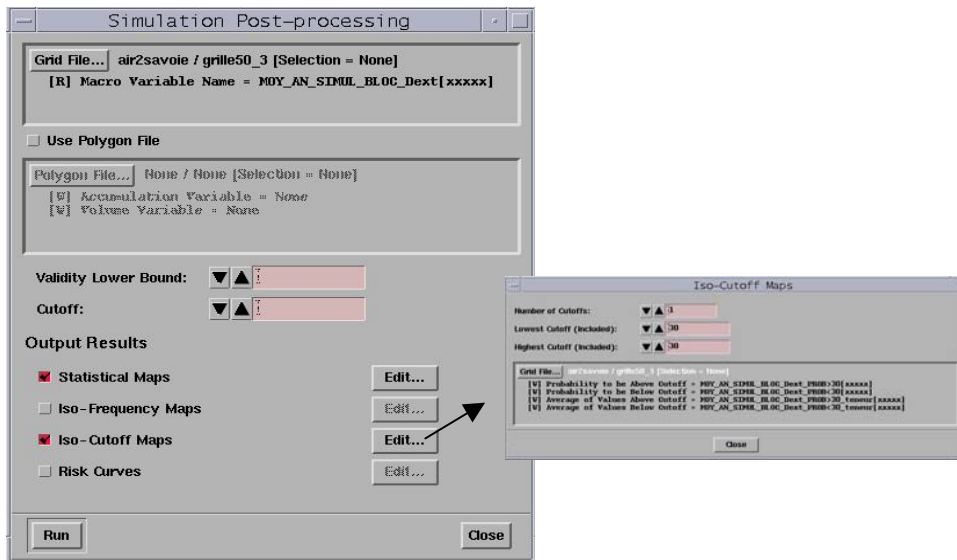


Figure 32 : Menu permettant le calcul de la probabilité de dépassement de seuil par simulations conditionnelles

Les deux cartes de probabilité de dépassement calculées par les deux dernières méthodes présentées (Conditionnement Uniforme et Simulations Conditionnelles) sont présentées dans la Figure 33.

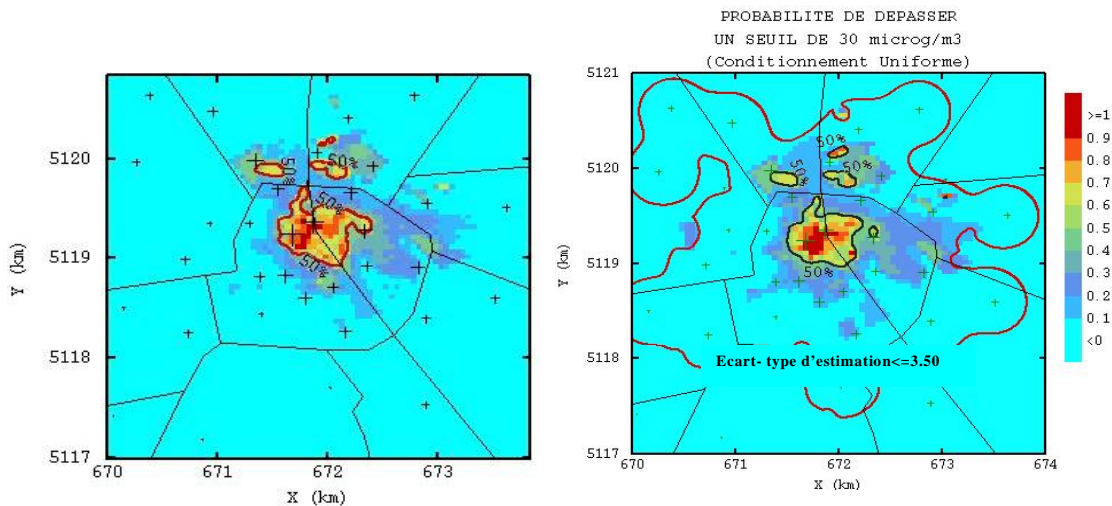
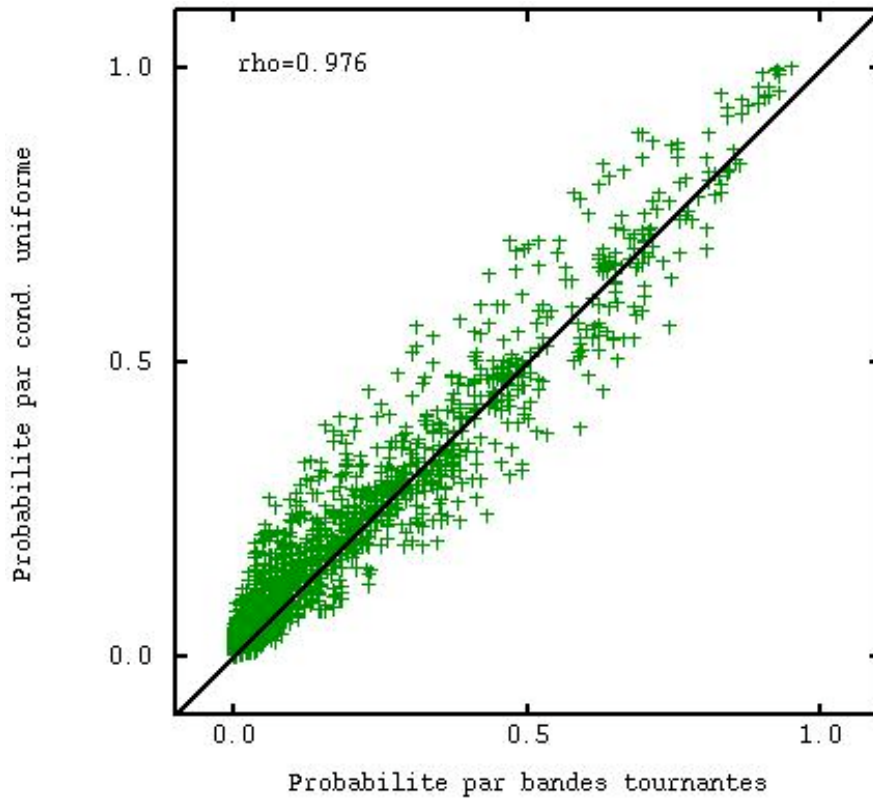


Figure 33 : Comparaison des deux cartes de probabilité de dépassement obtenues par Simulations Conditionnelles (à gauche) et par « Conditionnement Uniforme » (à droite)

L'allure des deux cartes de probabilité est très proche, en théorie les résultats devraient être les mêmes car les deux méthodes s'appuient sur la même hypothèse multigaussienne. Les quelques écarts constatés peuvent être imputés au faible nombre des bandes tournantes ou au calcul des anamorphoses des blocs lors de la mise en oeuvre du « Conditionnement Uniforme ».

Malgré cela, les résultats sont satisfaisants ; en effet une assez bonne corrélation dans le nuage de la Figure 34 (coefficient de 0.976) est constatée.



Les probabilités calculées par « Conditionnement Uniforme » sont légèrement supérieures à celles calculées par simulations

<i>Statistique des simulations</i>	Nb de cellules	Minimum	Maximum	Moyenne	Ecart- type	Variance	CV
Probabilité par simulations conditionnelles	5 929	0%	95%	6%	14%	200% ²	2.45
Probabilité par conditionnement uniforme		0%	100%	6.7%	14.4%	208% ²	2.15

Figure 34 : Statistiques et nuage de corrélation de la probabilité de dépasser un seuil de $30 \mu\text{g}/\text{m}^3$ calculée par conditionnement uniforme et simulations conditionnelles

5. CALCUL DE PERCENTILES

Comme il a été indiqué en introduction, dans le contexte de la qualité de l'air, il existe un critère de qualité basé sur la connaissance de valeurs de percentiles des concentrations horaires ou journalières du polluant. Par exemple, pour le dioxyde d'azote, la directive européenne N°. 1999/30/CE du 22 avril 1999, stipule que pour l'année 2010 les concentrations horaires ne devraient pas dépasser une valeur limite de $200 \mu\text{g}/\text{m}^3$ plus de dix-huit fois par an.

Pour répondre à la norme, le mieux serait de cartographier, sur le domaine d'étude, le percentile 99.8 des concentrations horaires (sur une année de 8760 heures) de dioxyde d'azote, puis de le comparer à la valeur seuil de $200 \mu\text{g}/\text{m}^3$.

La cartographie des percentiles pose 2 problèmes spécifiques :

- ❖ Les percentiles ne peuvent être calculés directement que pour les stations fixes ayant des données continues pendant une année. ***Contrairement à l'estimation par krigeage de la concentration d'un polluant, les percentiles ne peuvent pas être interpolés directement car il s'agit d'une quantité non linéaire qui ne peut pas être calculée à partir d'une combinaison linéaire de percentiles disponibles dans un voisinage.***

Il serait donc nécessaire, des disposer des concentrations horaires pendant une année dans tout le domaine, ce qui est impossible. Il est donc nécessaire de recourir à la simulation pour reproduire des concentrations possibles, sur lesquelles les percentiles peuvent être calculés directement.

- ❖ Le deuxième problème est l'aspect temporel, car il faudrait effectuer une modélisation spatio-temporelle (par exemple ajuster un variogramme en trois dimensions : les coordonnées géographiques et le temps) pour pouvoir reproduire non seulement la variabilité spatiale mais aussi la variabilité temporelle du phénomène de pollution, de plus une telle modélisation devrait prendre en compte des informations auxiliaires comme les conditions météo (vitesse et direction du vent, température), les cadastres des émissions, etc ; ces informations devraient être disponibles pour toute l'année et sur tout le domaine d'étude.

Si l'on fait abstraction de l'aspect temporel on peut appliquer les simulations conditionnelles (Deraisme, 2003) afin de reproduire les concentrations horaires, pour ensuite calculer les percentiles et les comparer à la valeur limite (par exemple $200 \mu\text{g}/\text{m}^3$ pour les concentrations horaires de dioxyde d'azote).

Cette méthode dont l'algorithme est présenté par la suite fait actuellement l'objet de plusieurs tests à AIRNORMAND :

- ❖ On effectue un nombre déterminé de simulations de la concentration du polluant dans l'unité du temps demandée par les normes.

Dans le cas du dioxyde d'azote, on peut envisager d'effectuer 200 simulations de la concentration horaire pour toutes les heures de l'année, cela implique la réalisation de 1 752 000 simulations (=200 simulations par 8760 données horaires).

Dans la pratique ces simulations ne sont possibles que lorsque la covariance spatiale peut être calculée dans de bonnes conditions : c'est-à-dire dans les régions où un réseau dense de stations fixes est disponible (comme dans les agglomérations).

Un autre inconvénient est la modélisation du variogramme des mesures horaires ; pour ne pas effectuer l'ajustement d'un grand nombre de variogramme (8 760), des hypothèses simplificatrices doivent être faites. Ces hypothèses entraînent une source d'imprécision à la méthode.

Une possibilité peut être de n'utiliser qu'un seul variogramme moyen calculé par exemple sur la moyenne annuelle, ou d'effectuer des ajustements des variogrammes de la concentration moyenne sur une semaine (52 ajustements) ou sur un mois (12 ajustements).

Il est possible aussi de réduire les calculs, en n'effectuant des simulations que pour des périodes de temps où l'on sait que les concentrations sont les plus élevées, par exemple pendant l'hiver pour le dioxyde d'azote ou l'été pour l'ozone. Si une réduction du temps est effectuée il faut recalculer le percentile correspondant: par exemple 18 heures dans six mois représentent le percentile 99.6.

Les simulations peuvent être réalisées par la méthode de bandes tournantes, dans cette méthode on peut effectuer des simulations conditionnelles en prenant une variable auxiliaire en cokrigeage ou en dérive externe (comme la densité de population) et de cette façon améliorer la précision des résultats.

- ❖ Une fois les simulations effectuées, l'étape suivante est de calculer le percentile pour chaque simulation.

Pour le calcul du percentile on doit créer une variable où l'on stocke une seule simulation à la fois, pour chaque heure (on sélectionne donc une carte du dioxyde d'azote pour chaque heure, la variable stockera ainsi 8760 cartes). Puis on calcule le percentile sur cette variable (menu présenté dans la Figure 32, option Isofrequency Map : Frequency 99.8) et on répète le processus jusqu'à la dernière simulation effectuée, on obtient à la fin 200 percentiles.

- A la fin de la procédure la variable contiendra un nombre déterminé des valeurs possibles du percentile. Sur cette variable l'on peut déterminer des probabilités de dépassement d'un seuil de $200 \mu\text{g}/\text{m}^3$ (menu Tools/Simulation Post-processing, option « Iso-Cutoff Maps, Figure 32).

6. CONCLUSION

Cette étude montre, dans le contexte de la qualité de l'air, les possibilités offertes par la géostatistique pour aider les AASQA dans le calcul des probabilités de seuils et de percentiles.

Parmi les méthodes de la géostatistique non- linéaire, « l'espérance conditionnelle » et « le conditionnement uniforme » offrent une rapide mise en oeuvre. Ces méthodes permettent de plus la prise en compte de variables auxiliaires grâce au cokrigeage ou à la dérive externe (il est possible donc de prendre en compte une composante non stationnaire).

Les simulations conditionnelles par bandes tournantes permettent aussi la prise en compte des variables auxiliaires et la modélisation de non- stationnarités, mais leur mise en œuvre est plus lourde et nécessite d'une plus grand temps de calcul.

Les simulations deviennent irremplaçables dès que l'on veut effectuer des calculs plus complexes mettant en jeu soit plusieurs points dans l'espace soit des relations entre des variables (comme le niveau de population et la population).

La simulation de concentrations journalières(ou horaires) et le calcul des percentiles sur chaque simulation, peut servir, en faisant abstraction de l'aspect temporel, à caractériser les zones comportant des risques plus ou moins élevés. Plusieurs aspects doivent encore être étudiés comme l'automatisation du calcul du variogramme journalier (ou horaire) et le nombre optimal des stations fixes nécessaires pour obtenir des résultats acceptables.

Cependant un calcul plus poussé des données temporelles devrait prendre en compte aussi la variabilité temporelle du phénomène de pollution (par exemple ajuster un variogramme en trois dimensions : les coordonnées géographiques et le temps).

Enfin les méthodes exposées dans cette étude sont basées sur une distribution multigaussienne des concentrations.

Cette hypothèse est très difficile à vérifier, on peut se contenter dans la pratique de vérifier si la distribution des données transformées est bigaussienne, en calculant le « H-scatter plot »: (nuage de corrélation décalé d'une distance h), ce nuage doit présenter une allure elliptique et en calculant la racine carré du variogramme divisée par le madogramme, cette relation doit être égale à 1.77

Dans le cas où l'hypothèse multigaussienne n'est pas vérifiée, il faut implémenter d'autres modèles de la fonction aléatoire (modèles booléen, des feuilles mortes, de substitution, etc), le lecteur trouvera une présentation détaillée de ces méthodes non gaussiennes dans Lantuejoul 2002 et Chiles1999.

7. REFERENCES

7.1. RAPPORTS LCSQA :

Barrois Q., 2004. Utilisation du module « Geostatistical Analyst » d'ARCVIEW dans le cadre de la qualité de l'air, rapport de stage INERIS.

Cardenas G., Malherbe L., Evaluation des incertitudes associées aux méthodes géostatistiques, Convention N° 115/03, INERIS, Décembre 2003. Téléchargeable sur l'adresse :

http://www.lcsqa.org/rapport/rap/prog2003/ineris/Etude15_rapport_incertitude_avril2004.pdf

Cardenas G., Malherbe L., Représentation de la qualité de l'air dans les zones peu/pas couvertes par les stations de mesure fixes :partie II, application à la problématique d'une association, Convention N° 31/2001, Etude 20, INERIS, Décembre 2002. Téléchargeable sur l'adresse :

http://www.lcsqa.org/rapport/rap/prog2002/ineris/Etude_19%20_cartographie2002.pdf

Fouquet C.(de), 2003. Méthodologie de cartographie de la concentration annuelle de NO2 sur l'agglomération de Mulhouse, Rapport d'avancement N-6/03/G.

Fouquet C.(de), 2003. Etude sur la réalisation de cartographies de la qualité de l'air dans les zones peu/pas couvertes par les réseaux de stations fixes à l'aide de méthodes géostatistiques (complément d'étude et synthèse).

Fouquet C. (de), Gallois D., Malherbe L., Cardenas G., 2004. Mapping annual nitogene dioxide concentrations above Mulhouse urban area. GeoEnv 2004, Neuchâtel (Suisse), 13-15 octobre.

Le Loc'h G., 2003. Etude exploratoire du dioxyde d'azote sur l'agglomération de Montpellier, Rapport N-8/03/G.

Rouil Laurence, Malherbe Laure, Méthodes de représentation de la qualité de l'air, Guide d'utilisation des méthodes de la géostatistique linéaire, Convention 115/2003, Décembre 2003. Téléchargeable sur l'adresse :

http://www.lcsqa.org/rapport/rap/prog2003/ineris/Etude14_1_guide_nelle_version.pdf

http://www.lcsqa.org/rapport/rap/prog2003/ineris/Etude14_1_Annexe1_m%E9thodes.pdf

http://www.lcsqa.org/rapport/rap/prog2003/ineris/Etude14_1_Annexe2ACP.pdf

http://www.lcsqa.org/rapport/rap/prog2003/ineris/Etude14_1_Annexe3VEM.pdf

Rouil Laurence, Malherbe Laure, Assistance à l'utilisation des outils de modélisation, Convention 115/03, Décembre 2003. Téléchargeable sur l'adresse :

http://www.lcsqa.org/rapport/rap/prog2003/ineris/Etude14.2_rapport_assistance_2003_corrige.PDF

Rouil Laurence, Malherbe Laure, Méthodes de représentation de la qualité de l'air dans les zones peu/pas couvertes, Partie I : stratégies et préconisations, Rapport LCSQA, Convention 31/2001, Décembre 2002. Téléchargeable sur l'adresse :

http://www.lcsqa.org/rapport/rap/prog2002/ineris/Etude_19%20cartographie2002.pdf

7.2. AUTRES REFERENCES

Bobbia, Michel, Juillet 2001, Investigations en vue d'une approche géostatistique de la qualité de l'air – Analyse du risque : Les simulations Conditionnelles, Air normand.

Deraisme, J, Février 2003, Complément d'étude des probabilité de dépassement de seuil et d'exposition de la population à la pollution. Geovariances.

Chiles, J.P., Delfiner P., Geostatistics Modeling Spatial Uncertainty, 1999, Ed. John Wiley & Sons INC.

European Environment Agency (EEA), 1998. Guidance report on preliminary assessment under EC air quality directives. Technical report.

Gallois D., Fouquet C. (de), LeLoc'h, G., Malherbe L., Cardenas G. Mapping annual nitogene dioxide concentrations in urban areas. Geostats 2004, Banff (Canada), 27 septembre-1^{er} octobre.

Isatis, divers manuels et notices d'utilisation, Geovariances.

Lantuéjoul, C, Geostatistical Simulation, Model and Algorithms, Springer, 2002.

Rivoirard, J, 1994, Introduction to Disjunctive Kriging and Non-linear Geostatistics. Oxford University Press, Oxford.

Saporta, G 1990, Probabilités, analyse des données et statistique. Editions Technip, Paris

Wackernagel, H., 1995, Multivariate Geostatistics- An Introduction with Applications. Springer, Berlin.

8. REMARQUE

Le présent rapport a été établi sur la base des informations fournies à l'INERIS, des données (scientifiques ou techniques) disponibles et objectives et de la réglementation en vigueur.

La responsabilité de l'INERIS ne pourra être engagée si les informations qui lui ont été communiquées sont incomplètes ou erronées.

Les avis, recommandations, préconisations ou équivalents qui seraient portés par l'INERIS dans le cadre des prestations qui lui sont confiées, peuvent aider à la prise de décision. Etant donné la mission qui incombe à l'INERIS de par son décret de création, l'INERIS n'intervient pas dans la prise de décision proprement dite. La responsabilité de l'INERIS ne peut donc se substituer à celle du décideur.

Le destinataire utilisera les résultats inclus dans le présent rapport intégralement ou sinon de manière objective. Son utilisation sous forme d'extraits ou de notes de synthèse sera faite sous la seule et entière responsabilité du destinataire. Il en est de même pour toute modification qui y serait apportée.

L'INERIS dégage toute responsabilité pour chaque utilisation du rapport en dehors de la destination de la prestation.

Fin du rapport envoyé au client. Ce qui suit est destiné au dossier maître