

Chapitre 4

Construction du cas d'étude synthétique

Un cas d'étude synthétique, représentant une situation "réelle" et assimilé à un aquifère alluvial a été créé dans le but de permettre l'évaluation qualitative et quantitative des résultats obtenus par l'application de la méthode stochastique alternative MSA proposée pour la détermination des zones de protection. Ce cas d'étude synthétique a été construit à partir d'informations géologiques et hydrogéologiques recueillies sur de nombreux sites d'études appartenant à la plaine alluviale de la Meuse, et ce, afin qu'il soit aussi proche que possible des conditions réelles de terrain. La génération du champ de conductivité hydraulique hétérogène de cet aquifère hypothétique a nécessité de définir les paramètres géostatistiques décrivant la structure de corrélation spatiale de la conductivité hydraulique. Ces paramètres n'ont pas été tirés de la littérature comme c'est le cas chez de nombreux auteurs, mais ils ont été déduits de l'analyse exploratoire des données du site de captage de Vivegnis (au Nord-Est de Liège), site de la plaine alluviale de la Meuse pour lequel le nombre de piézomètres est le plus important. A partir du champ de conductivité hydraulique ainsi généré, trois jeux de données constituant l'ensemble des données conditionnantes du cas synthétique ont été créés artificiellement : un jeu de conductivités hydrauliques, un jeu de hauteurs piézométriques et un jeu de résistivités électriques.

Dans ce travail, tous les traitements statistiques et géostatistiques des données ont été réalisés à l'aide du logiciel ISATIS [ISATIS, 2000], code développé au Centre de Géostatistique de l'Ecole des Mines de Paris.

4.1 Analyse exploratoire des données de conductivité hydraulique du site de captage de Vivegnis

Le site de captage de Vivegnis, appartenant à la Société Wallonne des Eaux SWDE, est situé dans la plaine alluviale de la Meuse, en bordure du Canal Albert, à environ dix kilomètres au Nord-Est de Liège (figure 4.1).

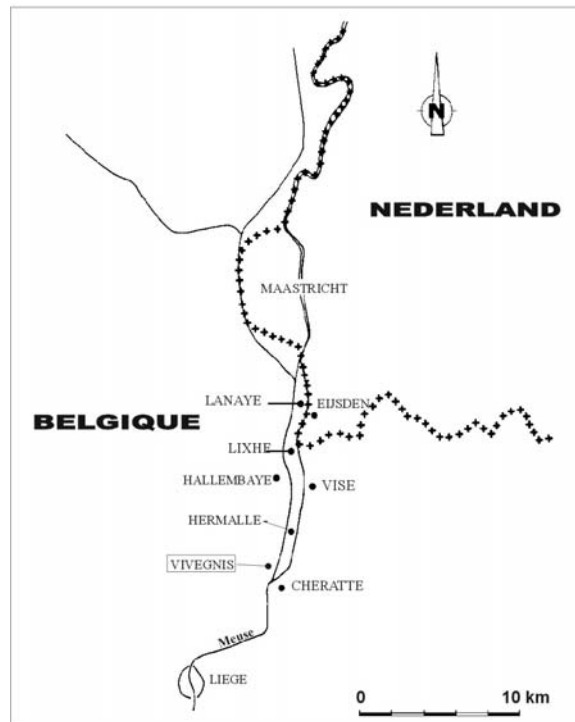


Figure 4.1 : Plan de situation

Le site comporte six puits de production, P1 à P6, et dix piézomètres de surveillance, Pz1 à Pz10 positionnés sur la figure 4.2. Les tableaux I et II fournis à l'annexe C reprennent les caractéristiques géologiques et hydrogéologiques de chacun de ces puits [DEROUANE, 1994 et GUYOT, 2000]. Les dépôts meubles constituant la plaine alluviale, épais de 5 à 15 mètres, sont composés de haut en bas de limons fluviaux (2 à 5 mètres) et de graviers roulés (3 à 10 mètres) enrobés dans une matrice sableuse, limoneuse ou argileuse plus ou moins abondante suivant les endroits. On observe, à la base de ces dépôts, la présence de graviers plus propres, plus grossiers et donc plus perméables. Des essais de pompage réalisés sur le

site [DEROUANE, 1994] ont permis de mesurer les conductivités hydrauliques^(*) en phase de rabattement et de remontée.

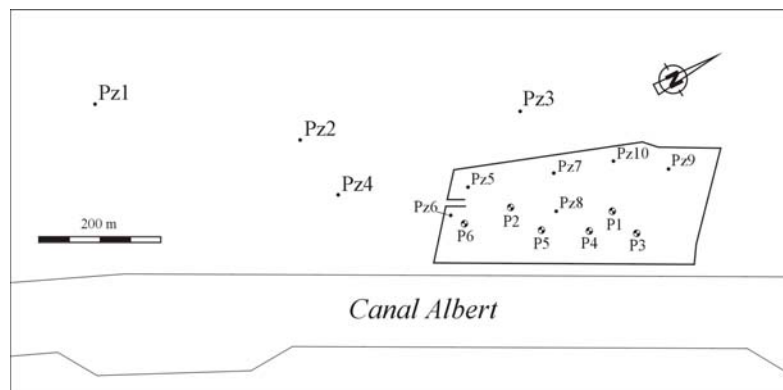


Figure 4.2 : Plan du site de captage de Vivegnis

4.1.1 Paramètres statistiques et histogrammes expérimentaux

4.1.1.1 Statistiques sur les données naturelles

Le tableau suivant présente les résultats statistiques calculés sur les deux jeux de 16 données de conductivité hydraulique (en rabattement et en remontée) issus du site de Vivegnis : minimum, maximum, moyenne arithmétique m_A , écart-type σ , variance σ^2 , et coefficient de variation (défini comme le rapport de l'écart-type à la moyenne σ/m_A).

	Minimum (m/s)	Maximum (m/s)	Moyenne m_A (m/s)	Ecart-type σ (m/s)	Variance σ^2 (m/s) ²	Coeff. Variat. σ/m_A
$K_{\text{rabattement}}^{\text{Vivegnis}}$	0,00018	0,037	0,011	0,010	0,00010	0,92
$K_{\text{remontée}}^{\text{Vivegnis}}$	0,00018	0,034	0,010	0,009	0,00008	0,90

Tableau 4.1 : Statistiques sur les données naturelles de la conductivité hydraulique du site de Vivegnis

Le coefficient de variation montre que l'écart-type est de l'ordre de grandeur de la moyenne, signe d'une variabilité importante. Les statistiques sur les deux jeux étant relativement semblables, la poursuite de l'analyse n'a été effectuée qu'avec un seul de ces

(*) Les résultats des essais de pompage ont été interprétés par les équations de Theis en régime de non-équilibre. Ces équations de non-équilibre d'une nappe captive sont applicables à une nappe libre si son épaisseur est uniforme, ses dimensions horizontales très grandes par rapport à sa puissance et sa surface piézométrique de faible gradient hydraulique [CASTANY, 1963], ce qui est vérifié sur ce site. Les équations de Theis fournissent les valeurs de transmissivité T de l'aquifère. La conductivité hydraulique K est déduite de la transmissivité par l'équation $T=K.h$ avec h , la hauteur d'eau au puits lors du pompage.

deux jeux, soit la conductivité hydraulique mesurée en remontée. La figure 4.3 présente l'histogramme de ces données de conductivité hydraulique.

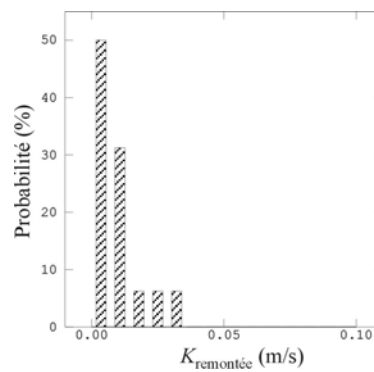


Figure 4.3 : Histogramme de la conductivité hydraulique en remontée sur le site de Vivegnis

4.1.1.2 Probabilités et loi normale

4.1.1.2.1 Définition de la loi monovariante normale

Loi normale centrée réduite : Une variable aléatoire X est une variable normale centrée réduite, notée $N(0,1)$, si elle admet la fonction de densité de probabilité $g(s)$ suivante :

$$g(s) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{s^2}{2}\right) \quad (4.1a)$$

$$G(s) = \int_{-\infty}^s g(s) \cdot ds \quad (4.1b)$$

Par définition, X est une variable d'espérance mathématique nulle et de variance unitaire. La figure suivante montre la densité de probabilité $g(s)$ et la fonction de distribution des probabilités ou fdp $G(s)$ d'une variable normale centrée réduite.

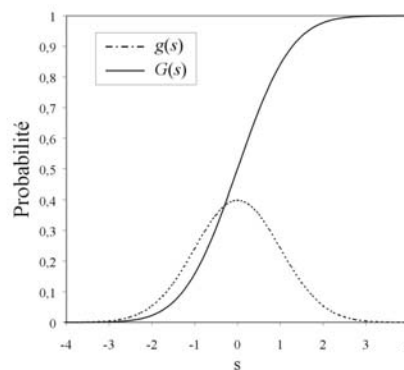


Figure 4.4 : Fonctions de densité de probabilité $g(s)$ et fonction de distribution $G(s)$

Loi normale : Y est une variable normale (ou gaussienne), notée $N(m, \sigma)$, si $X = (Y - m)/\sigma$ est une variable normale centrée réduite. Y est alors entièrement caractérisée par son espérance mathématique m et sa variance σ^2 . La fonction de densité de probabilité $g(s)$ de la variable aléatoire Y s'écrit dès lors:

$$g(s) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(s-m)^2}{2\sigma^2}\right) \quad (4.2)$$

Loi lognormale : X est une variable lognormale si $Y = \ln X$ est une variable normale.

4.1.1.2.2 Loi spatiale normale

Par définition, la fonction aléatoire $Z(\underline{x})$ est dite normale si sa loi spatiale est normale, c'est-à-dire si $\forall n \in \mathbb{N}^{(*)}$, $\forall (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)$, $(Z(\underline{x}_1), Z(\underline{x}_2), \dots, Z(\underline{x}_n))$ est un vecteur normal. Un vecteur est dit normal si toute combinaison linéaire de ses composantes est une variable normale.

La loi normale est très souvent utilisée pour la caractérisation statistique des phénomènes physiques. De plus, cette loi est symétrique par rapport à la moyenne et une fonction aléatoire stationnaire de loi normale est entièrement définie par son espérance mathématique et sa covariance. Cette dernière propriété est particulièrement intéressante pour la géostatistique linéaire : on est en général incapable, en partant d'un jeu de données expérimentales de taille limitée, d'estimer la loi spatiale au-delà de l'ordre 2 (cf. chapitre 1). Dans le cas de la loi normale, la connaissance de la loi spatiale à l'ordre 2 suffit à la définir entièrement.

4.1.1.2.3 Transformation anamorphose

Les propriétés naturelles ne présentent pas toujours une loi de comportement gaussienne. Pour preuve l'histogramme de la conductivité hydraulique présenté à la figure 4.3. Or de nombreuses méthodes géostatistiques, et en particulier toutes les méthodes de génération des simulations stochastiques présentées au chapitre 1, sont basées sur l'hypothèse fondamentale que les variables sont gaussiennes. Il est dès lors important de normaliser la distribution par une anamorphose. Divers types d'anamorphoses existent, elles constituent des artifices utiles pour contourner l'obstacle de la variable non gaussienne. Ces anamorphoses supposent que la variable de départ est bien définie par son histogramme.

(*) \mathbb{N} représentant l'ensemble de entiers naturels.

L'anamorphose logarithmique \ln (la plus classique) associée à la variable X est une transformation telle que la fonction de répartition de la variable transformée $Y = \ln X$ est normale $N(m, \sigma)$.

L'anamorphose gaussienne ϕ associée à la variable X est une transformation telle que la fonction de répartition de la variable transformée $Y = \phi(X)$ est normale centrée réduite $N(0,1)$. L'anamorphose gaussienne possède une hypothèse sous-jacente très forte : elle suppose la stationnarité de l'histogramme de la variable étudiée. Un inconvénient majeur de cette transformation est que ϕ n'est pas nécessairement bijective. Son inversion en fin d'étude géostatistique pour donner des résultats sur la variable naturelle $X = \phi^{-1}(Y)$ peut donc poser problème.

4.1.1.3 Statistiques sur les données transformées

De nombreux auteurs, tel que FREEZE [1975] et HOEKSEMA & KITANIDIS [1985], ont mis en évidence le caractère log-normal de la distribution des conductivités hydrauliques. Pour le site de Vivegnis, le tableau 4.2 récapitule les statistiques sur le logarithme népérien de la conductivité hydraulique (la conductivité hydraulique étant exprimée en m/s) : minimum, maximum, moyenne arithmétique, écart-type, variance et coefficient de variation.

	<i>Minimum</i>	<i>Maximum</i>	<i>Moyenne</i> m_A	<i>Ecart-type</i> σ	<i>Variance</i> σ^2	<i>Coeff. Variat.</i> $ \sigma/m_A $
$\ln K^{\text{Vivegnis}}$	-8,62	-3,38	-5,1	1,31	1,72	0,26

Tableau 4.2 : Statistiques sur le logarithme népérien de la conductivité hydraulique du site de Vivegnis

La transformation logarithmique a pour effet principal, du fait même de sa forme "aplatie", d'atténuer les écarts entre les valeurs extrêmes. Le coefficient de variation est donc beaucoup plus faible que pour les données naturelles (cf. tableau 4.1).

La figure 4.5 présente l'histogramme du logarithme népérien de la conductivité hydraulique et le compare à la densité de probabilité normale définie par sa moyenne et son écart-type, respectivement $-5,1$ et $1,31$. Cet histogramme est unimodal, c'est-à-dire qu'il présente un maximum de fréquence unique et, dans l'hypothèse de la loi normale, confondu avec la moyenne. Le logarithme de la conductivité hydraulique n'est pas une distribution parfaitement normale, mais l'allure générale de l'histogramme incite à penser que cette approximation est acceptable.

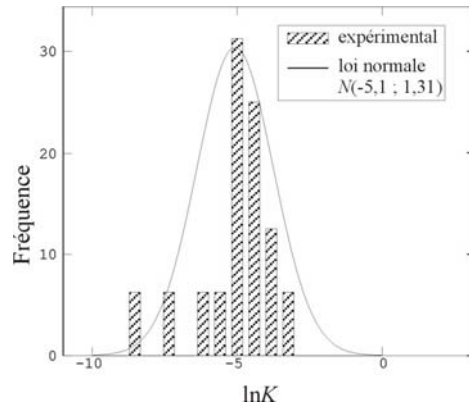


Figure 4.5 : Histogramme du logarithme népérien de la conductivité hydraulique du site de Vivegnis

Il ne faut pas perdre de vue que les données manipulées sont des conductivités hydrauliques de terrain, mesurées avec une marge d'erreur. L'anamorphose gaussienne est un raffinement que nous avons décidé de ne pas utiliser, en admettant que l'incertitude sur la mesure de la conductivité hydraulique est au moins du même ordre de grandeur que l'approximation sur le caractère gaussien de la fonction de répartition du logarithme népérien de la conductivité hydraulique. De plus, contrairement à l'anamorphose gaussienne, la transformation logarithmique est facilement inversible et ne nécessite pas l'hypothèse de stationnarité de l'histogramme. Dans la suite de l'étude, nous travaillerons donc sur le logarithme népérien de la conductivité hydraulique, en faisant l'hypothèse que cette variable $Y = \ln K$ suit approximativement une loi normale.

4.1.2 Variogramme expérimental et modèles de variogramme

4.1.2.1 Variogramme expérimental

Le variogramme expérimental est la clé de voûte de l'analyse géostatistique. Il permet de mettre en lumière les corrélations spatiales existant entre les variables. Sa définition statistique a été donnée par l'équation (1.15). Il s'agit ici de l'inférer à partir de la seule information fournie par les données. Si $(Z(x_i))_{i=1,p}$ désignent les p valeurs expérimentales de la propriété Z , mesurées aux points $(x_i)_{i=1,p}$, le variogramme expérimental stationnaire est défini comme la moyenne des écarts quadratiques entre valeurs :

$$\gamma(h) = \frac{1}{2N_c(h)} \cdot \sum_{i=1}^{N_c(h)} (Z(x_i + h) - Z(x_i))^2 \quad (4.3)$$

où $N_c(h)$ désigne le nombre de couples de points expérimentaux distants de h .

La figure 4.6 présente le variogramme du logarithme népérien de la conductivité hydraulique du site de Vivegnis. La variance expérimentale du jeu de données, $\sigma_{\ln K}^2$ (cf. tableau 4.2), est symbolisée en trait horizontal pointillé.

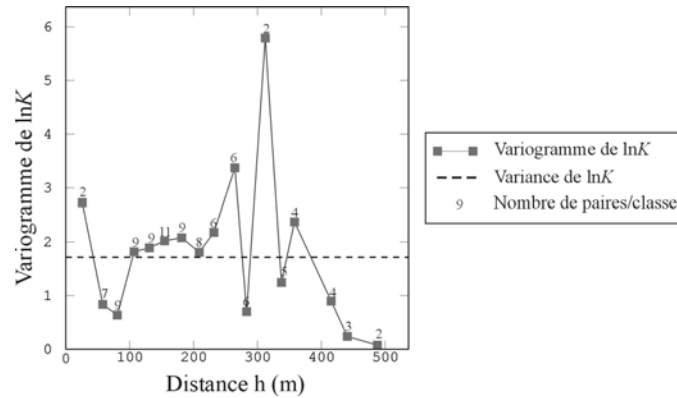


Figure 4.6 : Variogramme du logarithme népérien de la conductivité hydraulique du site de Vivegnis

Le manque de stabilité du variogramme pour le point initial et pour les distances croissantes (points fluctuants de manière erratique autour de la variance) peut s'expliquer en partie par la diminution du nombre de couples de points de calcul. Le variogramme étant défini comme une valeur moyenne (cf. équation 4.3), si le nombre de couples de points diminue, les fluctuations augmentent et la moyenne n'est plus significative.

4.1.2.2 Modèles de variogramme

L'intérêt de la modélisation d'un variogramme expérimental est essentiellement pratique : il s'agit de passer d'une fonction définie par des points à une fonction continue dans l'espace et possédant une expression mathématique. La méthode la plus simple pourrait être de relier simplement les points expérimentaux par une interpolation de type *spline*. Mais il n'est pas forcément judicieux de conserver toutes les oscillations du variogramme pour les grandes distances. Comme évoqué précédemment, celles-ci peuvent être dues à un nombre de paires de points trop petit pour que la moyenne des écarts quadratiques $[\ln K(x+h) - \ln K(x)]^2$ soit représentative. De plus, n'importe quelle fonction $\gamma(h)$ ne peut être choisie comme variogramme : MATHERON [1970] a démontré que la fonction γ doit être conditionnellement définie positive et que pour $h \rightarrow \infty$, $\gamma(h)$ doit nécessairement croître moins que h^2 , c'est-à-dire que $\lim_{|h| \rightarrow \infty} \frac{\gamma(h)}{h^2} \rightarrow 0$. La modélisation choisie est un compromis entre le calage au plus près du variogramme expérimental et la stabilisation de son comportement dans les grandes distances. Au modélisateur de faire la part des choses, en gardant à l'esprit le réalisme de l'information physique contenue dans le variogramme. Il s'agit de rendre compte de la corrélation spatiale

d'un phénomène physique bien précis : dans ce cas-ci, la conductivité hydraulique mesurée par essai de pompage dans un aquifère alluvial.

Plusieurs types de fonctions analytiques existent et peuvent être utilisés au choix du modélisateur. Elles sont définies par deux paramètres : le palier S et la portée a , distance à laquelle le variogramme atteint le palier (s'il existe). Dans le cas où le variogramme atteint ce dernier de façon asymptotique, la portée a , appelée alors "portée pratique", est définie comme la distance à laquelle la fonction analytique atteint 95% de la valeur du palier. De nombreux modèles de variogramme existent, mais seuls ceux utilisés dans ce travail seront définis.

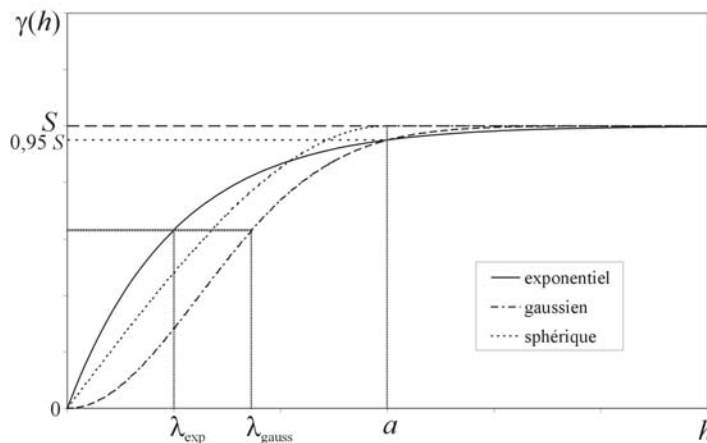


Figure 4.7 : Modèles de variogramme

Variogramme exponentiel :
$$\gamma(h) = S \left[1 - \exp\left(-\frac{3h}{a}\right) \right] \quad (4.4a)$$

pour lequel la longueur de corrélation λ , définie par l'intégrale (1.27), est égale au tiers de la portée pratique : $\lambda_{\text{exp}} = a/3$.

Variogramme gaussien :
$$\gamma(h) = S \left[1 - \exp\left(-\left(\frac{\sqrt{3}h}{a}\right)^2\right) \right] \quad (4.4b)$$

pour lequel la longueur de corrélation $\lambda_{\text{gauss}} = a/\sqrt{3}$.

Variogramme sphérique :
$$\gamma(h) = \begin{cases} S \left[\frac{3}{2} \left(\frac{h}{a}\right) - \frac{1}{2} \left(\frac{h}{a}\right)^3 \right] & h < a \\ S & h \geq a \end{cases} \quad (4.4c)$$

Il est défini comme le complémentaire du volume d'intersection de deux sphères de même diamètre a et distantes de h .

Le variogramme expérimental du logarithme népérien de la conductivité hydraulique du site de Vivegnis a été modélisé par ces trois fonctions analytiques. Pour optimiser le calage de ces modèles théoriques sur le variogramme expérimental, on a "joué" sur les paramètres S et a . La figure 4.8 présente les modèles de variogramme qui s'ajustent visuellement au mieux au variogramme expérimental, soit

- un modèle exponentiel de portée pratique $a = 150$ mètres et de palier $S = 2$ (proche de la variance théorique $\sigma_{\ln K}^2 = 1,72$)
- un modèle exponentiel de portée pratique $a = 300$ mètres et de palier $S = 2,5$
- un modèle gaussien de portée pratique $a = 150$ mètres et de palier $S = 2,15$
- un modèle sphérique de portée $a = 150$ mètres et de palier $S = 2$

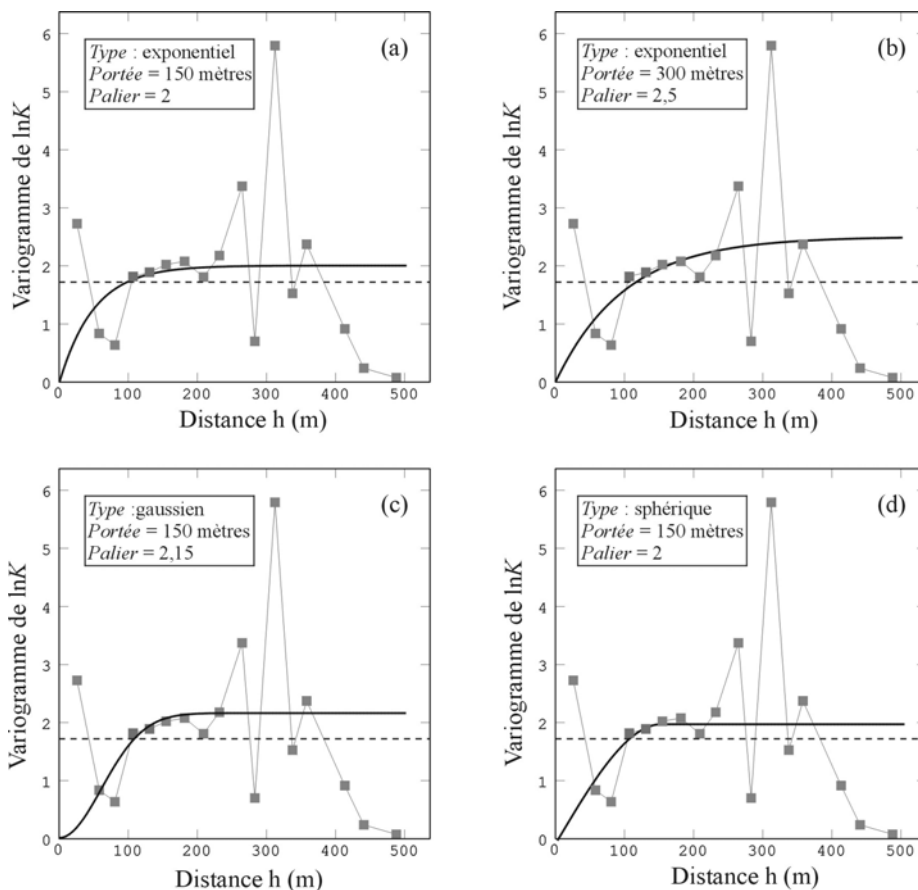


Figure 4.8 : Comparaison des modèles de variogramme s'ajustant au mieux au variogramme expérimental du logarithme népérien de la conductivité hydraulique du site de Vivegnis

Ces modèles tentent d'ajuster au mieux la croissance du variogramme pour les courtes distances (excepté le premier point), mais ne tiennent pas compte des oscillations du variogramme expérimental pour les grandes distances.

Il arrive en pratique que les variogrammes montrent une ordonnée non nulle à l'origine, que l'on appelle **effet de pépité**. Cet effet de pépité peut provenir de fortes variations locales des valeurs de la propriété ou être attribué soit à des erreurs de mesure, soit au fait que les données n'ont pas été récoltées à un intervalle de mesure suffisamment petit pour pouvoir observer la structure sous-jacente continue du phénomène. Un comportement avec effet de pépité est donc très courant dans l'analyse spatiale des données. Néanmoins, pour la présente étude, aucun effet de pépité n'a dû être introduit pour modéliser le variogramme expérimental.

4.1.2.3 Validation croisée

Une fois les modèles de variogramme choisis, la procédure de validation croisée permet de comparer l'impact de ces différents modèles sur les résultats de l'estimation. Le principe de cette technique de validation des modèles de variogramme est le suivant : une mesure $z(x_j)$ est temporairement soustraite du jeu de données $(z(x_i))_{i=1,p}$. La valeur de z en ce point x_j est alors réestimée par krigeage en utilisant l'information sur les points de mesure restants $(z(x_i))_{i \neq j}$ et le modèle de variogramme $\gamma_Z(h)$. Notons $z^*(x_j)$ cette estimation. Suite à la résolution du système de krigeage (1.36), la variance théorique de l'erreur d'estimation en ce point x_j peut être calculée par l'équation (1.38), notons-la $\hat{\sigma}_j^2$. Cette opération, menée pour tous les points $(x_i)_{i=1,p}$ pris un à un, permet de comparer l'erreur d'estimation expérimentale (différence entre la valeur ré-estimée par krigeage $z^*(x_j)$ et la valeur réelle mesurée $z(x_j)$) et l'erreur d'estimation due au modèle de variogramme à travers le calcul des expressions suivantes :

- *Moyenne de l'erreur*
$$m_\varepsilon = \frac{1}{p} \sum_{i=1}^p (z^*(x_i) - z(x_i)) \quad (4.5a)$$

- *Variance de l'erreur*
$$\sigma_\varepsilon^2 = \frac{1}{p} \sum_{i=1}^p (z^*(x_i) - z(x_i))^2 \quad (4.5b)$$

- *Moyenne de l'erreur standardisée*
$$\hat{m}_\varepsilon = \frac{1}{p} \sum_{i=1}^p \left(\frac{z^*(x_i) - z(x_i)}{\hat{\sigma}_i} \right) \quad (4.5c)$$

- *Variance de l'erreur standardisée*
$$\hat{\sigma}_\varepsilon^2 = \frac{1}{p} \sum_{i=1}^p \left(\frac{z^*(x_i) - z(x_i)}{\hat{\sigma}_i} \right)^2 \quad (4.5d)$$

La moyenne de l'erreur, qui doit tendre vers zéro, permet de vérifier si le krigeage est effectivement non-biaisé. La variance de l'erreur standardisée, qui doit se rapprocher de 1, correspond au rapport entre les variances expérimentale et théorique de krigeage. Elle permet

de vérifier que les erreurs de krigeage sont cohérentes avec la variance calculée. Les points x_i pour lesquels l'erreur standardisée $(z^*(x_i) - z(x_i)) / \hat{\sigma}_i$ est comprise dans l'intervalle $[-2.5 ; 2.5]$ (par analogie au cas de la loi normale centrée réduite, où cet intervalle contient 95% des valeurs) sont dits robustes. Le tableau 4.3 reprend les valeurs des expressions (4.5a) à (4.5d) obtenues par validation croisée sur les logarithmes népériens de la conductivité hydraulique du site de Vivegnis, pour les quatre modèles de variogrammes choisis.

	m_ε	σ_ε^2	\hat{m}_ε	$\hat{\sigma}_\varepsilon^2$	Données robustes
1 ^{er} modèle exponentiel, $a = 150\text{m}$, $S = 2$	0,021	2,41	0,008	1,47	15
2 ^{ème} modèle exponentiel, $a = 300\text{m}$, $S = 2,5$	0,06	2,57	0,023	1,86	14
3 ^{ème} modèle gaussien, $a = 150\text{m}$, $S = 2,15$	0,044	3,16	0,039	6,75	11
4 ^{ème} modèle sphérique, $a = 150\text{m}$, $S = 2$	0,031	2,36	0,013	1,9	14

Tableau 4.3 : Comparaison des résultats statistiques de la validation croisée pour les quatre modèles de variogrammes choisis

Ces résultats statistiques varient fort d'un modèle à l'autre. Pour le premier modèle, la condition de non-biais semble vérifiée puisque la moyenne des erreurs est faible (0,021). La variance de l'erreur standardisée est de 1,47, signe que l'erreur de krigeage expérimentale est légèrement supérieure à l'erreur théorique. Pour les trois autres modèles, la moyenne de l'erreur est moins proche de zéro et annonce donc un biais plus important. La variance de l'erreur standardisée, plus élevée, montre une erreur de krigeage plus importante, principalement dans le cas du modèle gaussien pour lequel cette quantité vaut 6,75. De plus, sur 16 points krigés, 15 sont robustes pour le premier modèle contre 14 pour le second (variogramme exponentiel, portée double) et le quatrième (variogramme sphérique), et 11 pour le troisième (variogramme gaussien). Cette procédure de validation croisée indique une bonne adéquation entre le premier modèle de variogramme et les valeurs expérimentales du logarithme népérien de la conductivité hydraulique sur le site de Vivegnis.

Afin de se trouver dans les conditions les plus proches possible des sites de la plaine alluviale de la Meuse, le champ de conductivité hydraulique du cas synthétique sera construit sur base des paramètres de ce modèle exponentiel, $a = 150$ mètres et $S = 2$, noté :

$$\gamma_{\ln K}(h) = 2[1 - \exp(-h/50)]$$

4.2 Construction du cas d'étude synthétique

Un cas d'étude synthétique représentant la situation "réelle" a été construit numériquement dans le but de permettre l'évaluation qualitative et quantitative des résultats obtenus par l'application de la MSA pour la détermination des zones de protection autour des puits de captage. Il est en effet intéressant, lorsque l'on réalise une analyse de l'incertitude, de pouvoir se référer à une situation connue.

Pour modéliser l'écoulement souterrain et délimiter les isochrones advectives correspondantes, le code d'écoulement MODFLOW et le logiciel de *particle tracking* MODPATH, tous deux développés par le *U.S. Geological Survey*, ont été utilisés dans ce travail. Pour faciliter leur utilisation, le pré- et post-processeur GMS [GMS, 1996] a été employé.

Le logiciel MODFLOW [MCDONALD & HARBAUGH, 1988] utilise la méthode des différences finies pour simuler l'écoulement souterrain. Il calcule la distribution piézométrique et le champ des vitesses à partir de l'équation d'écoulement (4.6) (dans notre cas en régime permanent) et de la loi de Darcy (4.7):

$$\operatorname{div}[\underline{\underline{K}} \underline{\underline{grad}}h] - q = 0 \quad (4.6)$$

$$\underline{\underline{v}} = -\frac{\underline{\underline{K}}}{n_e} \underline{\underline{grad}}h \quad (4.7)$$

où $\underline{\underline{K}}$ est le tenseur de conductivité hydraulique [LT^{-1}], h la hauteur piézométrique [L], q le terme général de sollicitation extérieure [T^{-1}], $\underline{\underline{v}}$ la vitesse d'écoulement [LT^{-1}] et n_e la porosité efficace.

L'approximation par différences finies des équations différentielles consiste à remplacer les dérivées spatiales par des différences finies. Chaque cellule de la grille de calcul est affectée d'un nœud central auquel on attribue les valeurs des paramètres hydrogéologiques. Ces dernières sont des valeurs moyennes, équivalentes et représentatives de l'ensemble de la cellule. Le principe des différences finies consiste à calculer la valeur de la variable dépendante h au centre de chaque cellule du domaine. Une équation d'écoulement sera donc écrite pour chacune de ces cellules : si N cellules sont définies, un système de N équations à N inconnues devra dès lors être résolu [DE MARSILY, 1986]. Pour les modèles multicouches, le logiciel MODFLOW résout l'équation d'écoulement à deux dimensions pour chaque couche individuellement. Lorsque chaque couche du modèle correspond à une unité

hydrostratigraphique différente ou que plusieurs couches du modèle représentent une seule unité hydrostratigraphique, les communications entre nappes sont réalisées à l'aide de flux verticaux calculés sur base de l'épaisseur des couches et de leurs valeurs de conductivité hydraulique verticale.

Le logiciel de *particle tracking* MODPATH [POLLOCK, 1989, 1994] permet, quant à lui, de calculer la trajectoire tri-dimensionnelle advective de particules d'eau "imaginaires" dans un champ d'écoulement obtenu par simulation du modèle MODFLOW, ainsi que les temps de transfert associés.

4.2.1 Cadre géologique général

La MSA étant destinée à être appliquée à des aquifères alluviaux "réels", on a désiré se placer directement dans des conditions aussi proches que possible de celles rencontrées sur le terrain. Le cas d'étude synthétique, assimilé à un aquifère alluvial type, a dès lors été construit à partir des informations géologiques et hydrogéologiques recueillies sur de nombreux sites d'études (figure 4.9) appartenant à la plaine alluviale de la Meuse [HADDOUCHI, 1987; MONJOIE & *al.*, 1987; DASSARGUES, 1991; DEROUANE, 1994; PETERS, 1996; RENTIER, 1996]. Ces informations (dimensions de la plaine alluviale, épaisseur, valeurs de conductivité hydraulique et valeurs de résistivité électrique des formations alluviales, type et gradient de la nappe, etc.) sont synthétisées dans le tableau III repris à l'annexe C.

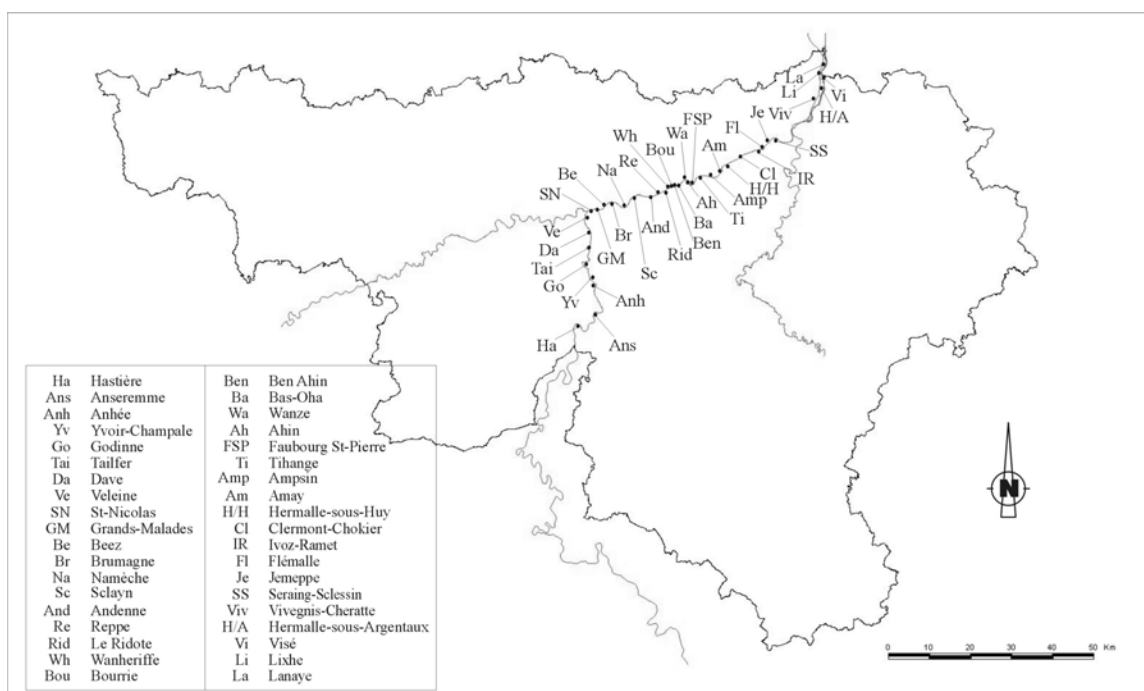


Figure 4.9 : Localisation des sites appartenant à la plaine alluviale de la Meuse d'où proviennent les informations géologiques et hydrogéologiques

Les dépôts de la plaine alluviale de la Meuse évoquent une rivière à bras multiples et à chenaux anastomosés. Ce régime est caractéristique des régions périglaciaires. Il s'établit dans des rivières surchargées en sédiments hétérogènes. Les matériaux excédant la compétence locale se déposent en îlots de graviers entre lesquels l'eau s'écoule à vitesse accrue dans les chenaux. Verticalement, on dénombre en général deux séquences granoclassées dans les dépôts meubles :

- une couche supérieure composée de limons fluviaux et sables fins constituant une phase d'alluvionnement tardive et surmontant indifféremment les îlots, les bras et les chenaux, dont l'épaisseur varie entre 2 et 4 mètres (mais peut être inférieure à 0,5 mètres au droit des îlots et supérieure à 5 mètres au niveau d'anciens bras) ;
- une couche inférieure de graviers roulés enrobés dans une matrice sableuse, limoneuse ou argileuse, d'abondance variable, dont l'épaisseur varie en moyenne entre 4 et 8 mètres (mais peut descendre sous les 2 mètres près du versant et au droit des anciens chenaux fluviaux et dépasser les 12 mètres au niveau des îlots). Sous cette couche graveleuse se trouve le bedrock imperméable ou non.

Sur base de ces informations, une coupe schématisée à la figure 4.10 a été établie.

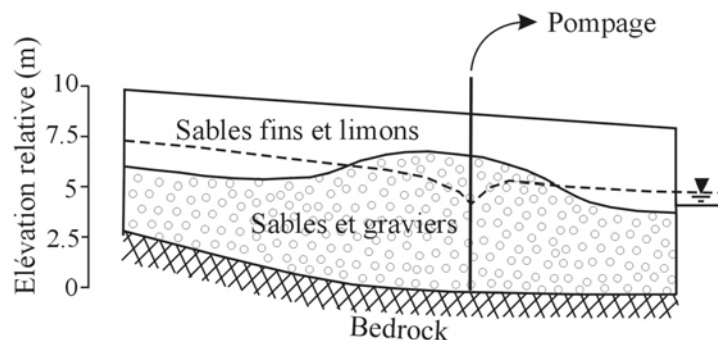


Figure 4.10 : Coupe schématisée pour un aquifère type de la plaine alluviale de la Meuse

D'une manière générale pour l'ensemble des sites de la plaine alluviale de la Meuse, la nappe alluviale est tantôt captive, tantôt libre, dépendant de son niveau par rapport à la base de la couche sablo-limoneuse et du taux d'argile rendant cette couche imperméable. Elle est alimentée par :

- *l'impluvium direct* : l'infiltration efficace est conditionnée par le coefficient de conductivité hydraulique du recouvrement sablo-limoneux semi-perméable ;
- *les apports des versants* : écoulement hypodermique dans le colluvium et dans le manteau d'altération du substratum, se déversant dans la nappe alluviale en pied de versant ;
- *la Meuse* : des interactions entre la nappe alluviale et le cours d'eau existent. La Meuse est navigable et équipée de plusieurs barrages-écluses qui régularisent le débit du cours d'eau et qui peuvent à l'amont favoriser l'alimentation de la nappe des graviers par le cours d'eau.

Cependant, quand les berges du cours d'eau sont aménagées, des murs plus ou moins étanches peuvent empêcher partiellement les échanges entre la nappe alluviale et la Meuse ;

- *le bedrock* : la nappe alluviale est alimentée par le bedrock quand ce dernier est aquifère.

4.2.2 Modèle conceptuel

Deux couches distinctes ont été modélisées correspondant pour la première à la couche de sables fins et de limons de surface et pour la deuxième à la couche graveleuse, le bedrock étant supposé imperméable. La première couche étant moins perméable, la conductivité hydraulique a été supposée uniforme sur l'entièreté de la couche avec une valeur de 10^{-5} m/s. Le champ de conductivité hydraulique de la deuxième couche a été divisé en trois parties de dimensions identiques (figure 4.11) : une partie centrale constituant le site d'étude proprement dit, caractérisée par un champ de conductivité hydraulique hétérogène, que nous appellerons, dans la suite de ce travail, **champ de conductivité hydraulique de référence**, représentant le champ "réel", et deux extensions latérales situées de part et d'autre, et qui ont pour but de reporter les conditions aux frontières afin qu'elles n'influencent pas les résultats d'écoulement et de transport. La conductivité hydraulique appliquée sur ces deux extensions latérales est uniforme et la valeur sera fixée égale à celle de la conductivité hydraulique équivalente du champ de référence. La plaine alluviale étant généralement limitée en largeur (cours d'eau d'un côté et limite de la plaine alluviale de l'autre), aucun report de conditions aux frontières n'a été réalisé dans la direction y .

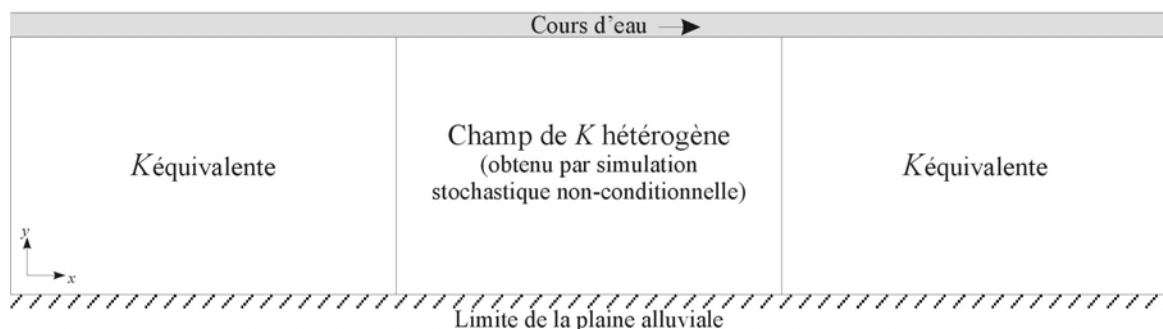


Figure 4.11 : Modèle conceptuel

Une recharge de 250 mm/an, correspondant à l'infiltration efficace moyenne calculée sur le territoire belge, a été appliquée au modèle.

Pour permettre la simulation du transport advectif, une porosité efficace uniforme de 5% a été choisie pour l'entièreté de la couche graveleuse, ce qui correspond à la porosité efficace moyenne mesurée dans la plaine alluviale de la Meuse (cf. tableau III à l'annexe C).

4.2.2.1 Dimensions de la grille de calcul

L'une des deux hypothèses fortes du développement de la géostatistique est l'hypothèse d'ergodicité (cf. paragraphe 1.1.3.1) qui suppose l'équivalence entre les moyennes statistiques et les moyennes spatiales. Pour respecter cette hypothèse, il est nécessaire de modéliser un domaine suffisamment grand de manière à prendre en compte la plupart des structures spatiales possibles du milieu. Tout le problème réside alors dans l'estimation de ce "domaine suffisamment grand". S'il n'existe pas de règle absolue en la matière, les expériences tirées de la littérature [GELHAR, 1993] indiquent qu'il est plausible de supposer l'ergodicité si la dimension totale du milieu est supérieure d'au moins un ordre de grandeur à l'échelle intégrale. Le modèle de variogramme exponentiel choisi possède une portée pratique de 150 mètres qui, comme on l'a vu plus haut (équation 4.4a), équivaut pour une fonction exponentielle à une longueur de corrélation $\lambda_{\ln K}$ de 50 mètres. En se basant sur les dimensions habituelles de la plaine alluviale de la Meuse (voir tableau III à l'annexe C), la grille de calcul choisie pour le champ de conductivité hydraulique hétérogène, de type rectangulaire, a les dimensions suivantes :

- axe des x : longueur de la plaine alluviale, $L = 600$ mètres (soit $12 \lambda_{\ln K}$)
- axe des y : largeur de la plaine alluviale, $l = 400$ mètres (soit $8 \lambda_{\ln K}$)

La grille de calcul totale, constituée des trois grilles réduites de taille identique, possède dès lors une longueur de 1800 mètres et une largeur de 400 mètres.

4.2.2.2 Taille des mailles

Dans le but de reproduire au mieux la structure du champ de conductivité hydraulique (c'est-à-dire de prendre en compte les effets de la variabilité spatiale), une grille très raffinée est requise pour la partie centrale. En effet, la taille des cellules de la grille Δx doit correspondre à une fraction de la longueur de corrélation du logarithme népérien de la conductivité hydraulique $\lambda_{\ln K}$. ABABOU & al. [1989] ont suggéré la relation empirique suivante : $\lambda_{\ln K} / \Delta x = 1 + \sigma_{\ln K}^2$. BELLIN & al. [1992] ont démontré que le rapport $\lambda_{\ln K} / \Delta x = 4$ fournit une précision satisfaisante. En choisissant des mailles carrées de 5 mètres de côté, le rapport $\lambda_{\ln K} / \Delta x$ est de 10, soit supérieur au rapport minimum recommandé. Avec ce choix de mailles de 25 m^2 , la grille centrale comporte 9600 éléments. La conductivité hydraulique des deux extensions latérales étant uniforme, la grille ne nécessite pas un raffinement aussi important. Des cellules rectangulaires de 30 mètres sur 5 mètres ont été choisies, élevant le nombre total de cellules à 12800 par couche. Puisque la grille est identique pour les deux couches, on obtient un modèle totalisant 25600 cellules.

4.2.2.3 Conditions aux limites

Etant donné les limites physiques choisies pour le domaine étudié, deux types de conditions frontières ont été imposés au modèle :

- 1) condition de potentiel imposé, dite condition de Dirichlet, qui revient à spécifier le potentiel sur les limites où celui-ci est indépendant des flux échangés. En terme de hauteur piézométrique h , ce type de condition s'exprime par $h(x,y,z,t)=f(x,y,z,t)$ où f est une fonction connue.
- 2) condition de flux imposé, dite condition de Neumann, qui revient à spécifier un flux perpendiculairement aux limites du modèle, indépendamment des hauteurs piézométriques. Ce type de condition s'exprime par $\partial h(x,y,z,t)/\partial n = f'(x,y,z,t)$ où f' est la valeur spécifiée.

Les conditions aux limites imposées sur toutes les frontières du modèle sont identiques pour les deux couches :

- la nappe étant supposée en équilibre avec le cours d'eau, un potentiel variant linéairement de l'amont vers l'aval a été imposé le long de ce dernier ;
- une condition de potentiel imposé a également été appliquée sur les deux frontières latérales, déterminant ainsi le gradient de la nappe ;
- le long de la limite de la plaine alluviale, on a considéré une alimentation en provenance du versant : un flux à travers cette frontière a été imposé correspondant à un débit de $0,0064 \text{ m}^3/\text{s}$, calculé en supposant une recharge de $250 \text{ mm}/\text{an}$ sur un versant de 1800 mètres de long et d'environ 450 mètres de large.

4.2.3 Champ de conductivité hydraulique

Pour terminer la construction du cas d'étude synthétique, il reste à définir le champ de conductivité hydraulique de la couche graveleuse.

4.2.3.1 Champ de conductivité hydraulique de référence

Le champ de conductivité hydraulique de référence, que l'on veut le plus réaliste possible, a été obtenu à l'aide du logiciel ISATIS en générant une simulation stochastique non-conditionnelle unique par la méthode des bandes tournantes (décrite au chapitre 1), à partir du modèle de variogramme choisi sur base des données de conductivité hydraulique du site captage de Vivegnis (correspondant à une distribution gaussienne des $\ln K$ de moyenne $m_A = -5,1$ et de variance $\sigma^2 = 1,72$), soit un variogramme isotrope, exponentiel, de portée

$a = 150$ mètres (donc de longueur de corrélation $\lambda = 50$ mètres) et de palier $S = 2$, noté $\gamma_{\ln K}(h) = 2[1 - \exp(-h/50)]$. Un nombre de 1000 bandes a été utilisé pour générer la simulation stochastique non-conditionnelle. D'après le manuel d'ISATIS, une loi empirique couramment utilisée consiste à dire que le nombre de bandes doit être supérieur au nombre de lignes et de colonnes du modèle, soit respectivement 120 et 80, et être inférieur au nombre de cellules de la grille, soit 9600.

A ce stade de l'étude, les propriétés statistiques (minimum, maximum, moyenne arithmétique, variance et coefficient de variation) du logarithme népérien de la conductivité hydraulique de la simulation stochastique non-conditionnelle ont été calculées et sont présentées au tableau 4.4.

	Minimum	Maximum	Moyenne m_A	Ecart-type σ	Variance σ^2	Coeff. Variat. $ \sigma/m_A $
$\ln K^{SNC}$	-9,51	-1,17	-5,46	1,38	1,9	0,25

Tableau 4.4 : Statistiques calculées sur les valeurs du logarithme népérien de la conductivité hydraulique de la simulation stochastique non-conditionnelle

La variance calculée sur une simulation stochastique est une variance dite "de dispersion", estimée sur un champ **fini**. Elle peut être différente de la variance *a priori* qui correspond à un champ **infini**. MATHERON [1970] a montré que pour une fonction aléatoire stationnaire, la variance de dispersion est inférieure à la variance *a priori* et converge vers celle-ci quand le champ tend vers l'infini. Ce constat est vérifié ici : en effet, la variance de dispersion calculée sur les résultats de la simulation stochastique non-conditionnelle vaut 1,9 alors que la variance *a priori* est égale à 2 (valeur du palier du variogramme choisi). La variance calculée sur la grille est donc proche de la valeur théorique à moins de 5%. La réalisation sur la grille de calcul constitue un échantillon suffisamment grand pour que la variance de dispersion soit proche de la variance *a priori*. Avec une grille de calcul d'une extension de plus de 10 longueurs de corrélation, l'ergodicité est donc bien atteinte pour ce modèle.

La figure 4.12 montre l'histogramme des valeurs du logarithme népérien de la conductivité hydraulique de la simulation stochastique non-conditionnelle et le compare à la fonction de répartition gaussienne définie par sa moyenne et son écart-type, respectivement $-5,46$ et $1,38$. Cette fonction de répartition est très proche de l'histogramme.

La figure 4.13 présente le variogramme du logarithme népérien de la conductivité hydraulique de la simulation stochastique non-conditionnelle calculé à partir de tous les points de la grille, dans les directions des axes x et y .

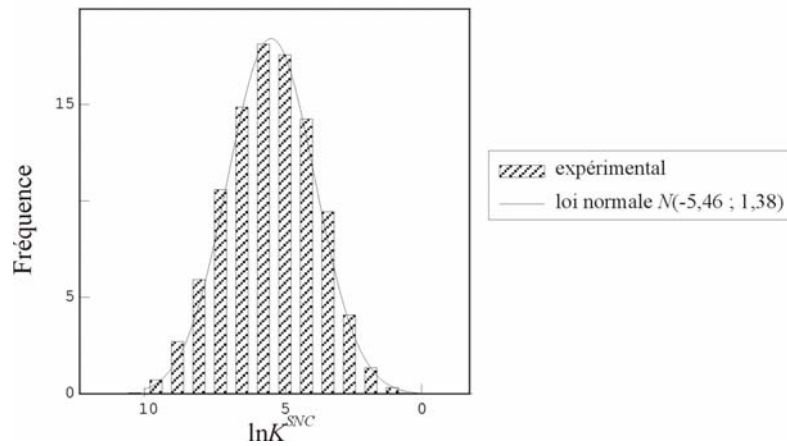


Figure 4.12 : Histogramme des valeurs du logarithme népérien de la conductivité hydraulique de la simulation stochastique non-conditionnelle

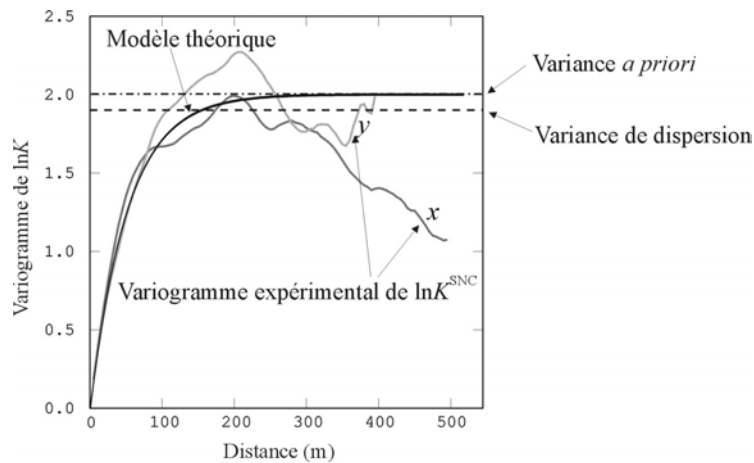


Figure 4.13 : Variogramme du logarithme népérien de la conductivité hydraulique de la simulation stochastique non-conditionnelle

La croissance du variogramme est conforme au modèle exponentiel imposé. Le manque de stabilité pour les distances plus importantes peut être expliqué par une diminution du nombre de couples de calcul du variogramme.

Les vérifications statistiques réalisées sur les valeurs du logarithme népérien de la conductivité hydraulique semblent indiquer une bonne cohérence statistique entre la simulation stochastique non-conditionnelle et la loi spatiale.

Pour obtenir le champ de conductivité hydraulique de référence, il reste à transformer le champ des logarithmes en un champ de la variable naturelle. Cependant, il est impossible d'obtenir simultanément des estimateurs non-biaisés du logarithme $Y = \ln K$ et de sa variable naturelle K . En effet, la relation entre les moments statistiques de K et de Y n'est pas $E(K) = \exp[E(Y)]$, mais $E(K) = \exp\left[E(Y) + \frac{1}{2} \text{var}(Y)\right]$. On verra plus loin que l'estimation optimale de la conductivité hydraulique équivalente pour un modèle bi-dimensionnel, dans

des conditions d'écoulement uniforme, est obtenue par sa moyenne géométrique. Lors du krigeage des conductivités hydrauliques équivalentes sur des blocs, il est préférable d'obtenir une estimation non-biaisée dans l'espace des logarithmes Y plutôt que dans l'espace normal. Le champ de conductivité hydraulique de référence (figure 4.14) a donc été obtenu par simple transformation $K = \exp(\ln K)$ sur l'entièreté des cellules de la grille (constituant un estimateur biaisé).

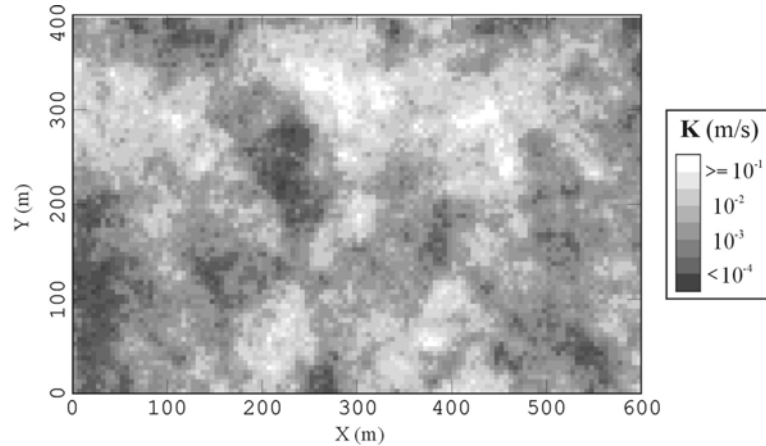


Figure 4.14 : Champ de conductivité hydraulique de référence

Les statistiques calculées sur les valeurs de conductivité hydraulique du champ de référence sont présentées au tableau 4.5.

	<i>Minimum</i> (m/s)	<i>Maximum</i> (m/s)	<i>Moyenne</i> m_A (m/s)	<i>Ecart-type</i> σ (m/s)	<i>Variance</i> σ^2 (m/s) ²	<i>Coeff. Variat.</i> σ_K / m_A
$K^{Réf}$	0,000074	0,31	0,01	0,018	0,00033	1,8

Tableau 4.5 : Statistiques calculées sur les valeurs naturelles de conductivité hydraulique du champ de référence

4.2.3.2 Conductivité hydraulique équivalente

Pour rappel, les moyennes arithmétique, géométrique et harmonique des données $(Z_i)_{i=1,n}$ d'une propriété Z s'écrivent

- moyenne arithmétique : $m_{A_z} = \frac{1}{n} \sum_{i=1}^n Z_i$ (4.8a)

- moyenne géométrique : $m_{G_z} = \left(\prod_{i=1}^n Z_i \right)^{1/n}$ (4.8b)

- moyenne harmonique : $m_{H_z} = \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{Z_i} \right)^{-1}$ (4.8c)

De nombreux auteurs, tel que MATHERON [1967], ont pu vérifier que la conductivité hydraulique équivalente est toujours comprise entre la moyenne harmonique et la moyenne arithmétique des conductivités hydrauliques locales :

$$m_{H_K} \leq K_{\text{équivalente}} \leq m_{A_K}, \quad \text{appelée inégalité fondamentale [RENARD, 1997]} \quad (4.9)$$

De plus, la fonction de répartition de la conductivité hydraulique étant log-normale, la moyenne arithmétique sur les logarithmes de la conductivité hydraulique correspond à la moyenne géométrique de la conductivité hydraulique,

$$m_{A_{\ln K}} = m_{G_K} \quad (4.10)$$

La moyenne géométrique est comprise entre les moyennes arithmétique et harmonique. En effet, ces dernières peuvent s'exprimer en fonction de la moyenne géométrique [DAGAN, 1989]

$$m_{H_K} = m_{G_K} \exp\left(\frac{-\sigma_{\ln K}^2}{2}\right) \quad \text{et} \quad m_{A_K} = m_{G_K} \exp\left(\frac{\sigma_{\ln K}^2}{2}\right) \quad (4.11)$$

Cette moyenne géométrique des conductivités hydrauliques est classiquement utilisée pour représenter la conductivité hydraulique équivalente, pour différentes raisons :

- c'est une valeur intermédiaire comprise entre les deux bornes limites m_{H_K} et m_{A_K} ;
- GELHAR [1993] a montré, pour plusieurs cas de figures, le lien entre la conductivité hydraulique équivalente et la moyenne géométrique :

$$\text{cas 2D isotrope : } K_{\text{équivalente}} = m_{G_K} \quad (4.12a)$$

$$\text{cas 3D isotrope : } K_{\text{équivalente}} = m_{G_K} \left(1 + \frac{\sigma_{\ln K}^2}{6}\right) \quad \text{pour } \sigma_{\ln K}^2 < 2 \quad (4.12b)$$

La moyenne géométrique des conductivités hydrauliques comme valeur équivalente constitue donc un bon compromis.

Les moyennes arithmétique, géométrique et harmonique de l'ensemble des valeurs de conductivité hydraulique du champ de référence sont reprises au tableau 4.6. La moyenne géométrique est comprise entre les valeurs des moyennes harmonique et arithmétique, ce qui constitue bien un résultat classique.

	Moyenne arithmétique m_A (m/s)	Moyenne géométrique m_G (m/s)	Moyenne harmonique m_H (m/s)
$K^{Réf}$	10^{-2}	$4,2 \cdot 10^{-3}$	$1,7 \cdot 10^{-3}$

Tableau 4.6 : Valeurs des différentes moyennes de conductivité hydraulique du champ de référence

La conductivité équivalente appliquée uniformément sur les deux extensions latérales du modèle synthétique a donc été choisie égale à la moyenne géométrique du champ de conductivité hydraulique de référence, soit $4,2 \cdot 10^{-3}$ m/s.

4.2.4 Isochrone de référence

Une zone de protection est définie autour d'un puits de captage pour un débit fixé. Dans ce cas d'étude synthétique, un puits fictif pompant un débit de $60 \text{ m}^3/\text{h}$ a été introduit dans la deuxième couche au sein d'une zone de conductivité hydraulique élevée (comme c'est généralement le cas dans la pratique). L'équation d'écoulement (4.6) a été résolue numériquement à l'aide du code aux différences finies MODFLOW pour le champ de conductivité hydraulique créé. Suite à la simulation de l'écoulement souterrain et en considérant une porosité efficace uniforme dans la couche graveleuse égale à 0,05, plusieurs isochrones advectives relatives à différents temps t ont été délimitées autour du puits par *particle tracking* à l'aide du logiciel MODPATH.

La technique de *particle tracking* consiste, pour rappel, à suivre la trajectoire de particules imaginaires placées dans un champ d'écoulement. Dans ce travail, la procédure de délimitation des isochrones est réalisée de la manière suivante. Une particule est associée à chaque élément de la grille de calcul et le logiciel MODPATH suit le trajet de ces particules dans la direction de l'écoulement (*forward tracking*) depuis leur position initiale sur la grille. Le mouvement de ces particules est arrêté au temps t spécifié. Durant ce temps t , certaines particules sont "capturées" par le puits de pompage. L'ensemble des éléments de la grille de calcul, correspondant aux positions initiales des particules ayant atteint le puits en un temps $T \leq t$, constitue par définition la zone de protection au temps t dont l'enveloppe délimite l'isochrone au temps $t^{(*)}$. Les isochrones à 1 jour, 5 jours, 10 jours et 20 jours ont ainsi été délimitées et seront appelées, dans la suite de ce travail, **isochrones de référence** $\tau_{Réf}(\underline{x}, t)$. Elles serviront, comme leur nom l'indique, de références auxquelles les résultats de la MSA seront comparés. La figure 4.15 représente l'une d'entre elles.

(*) L'avantage du "*forward tracking*" est que la position finale et les temps de transferts sont calculés pour toutes les cellules de la grille, permettant de délimiter précisément les zones de protection. Par la méthode du "*backward tracking*", pour laquelle un grand nombre de particules sont initialement positionnées au point de pompage et leur mouvement est suivi dans le sens inverse de l'écoulement, la zone de protection est seulement définie "là où les particules sont passées", résultant alors en une délimitation moins précise de la zone.

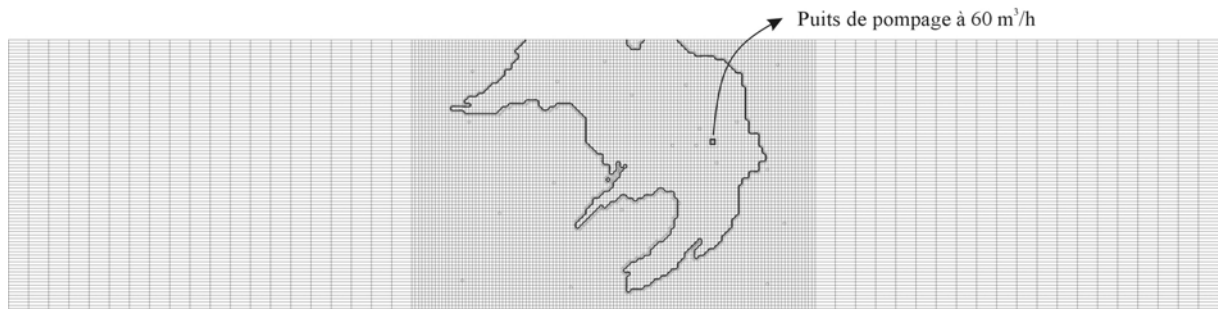


Figure 4.15 : Isochrone de référence à 20 jours

4.3 Création des jeux de données conditionnantes

Les trois types de données initiales du cas d'étude synthétique qui seront considérés comme les données conditionnantes dans la MSA ont été créés artificiellement à partir du champ de conductivité hydraulique de référence. Ces données consistent en un jeu de conductivités hydrauliques (données *hard*), un jeu de hauteurs piézométriques (données *soft*) et un jeu de résistivités électriques (données *soft*).

4.3.1 Données de conductivité hydraulique

Pour construire le jeu de données de conductivité hydraulique du cas d'étude synthétique tout en se plaçant dans des conditions d'aquifères réels, 25 points "d'observation" ont été répartis de manière irrégulière sur la grille de calcul (figure 4.16) ; le point n°1 représentant le puits de captage. Les valeurs de conductivité hydraulique prélevées directement sur le champ de référence, au droit des points de mesure, constituent l'ensemble des valeurs de ce jeu de données, supposées correspondre, par exemple, à des résultats d'essais de pompage.

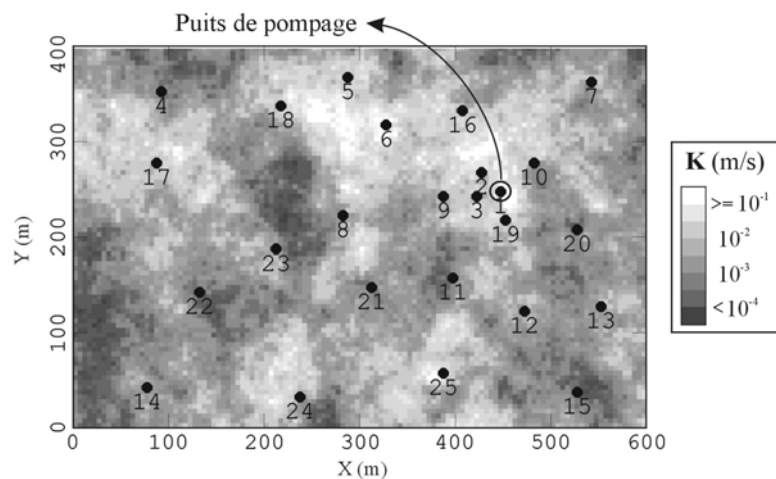


Figure 4.16 : Position des points de mesure sur le champ de conductivité hydraulique de référence (le jeu de 15 données étant constitué des points 1 à 15, le jeu de 25 données, des points 1 à 25)

A partir de ces 25 valeurs de conductivité hydraulique, deux jeux ont été créés – un jeu de 15 données (points 1 à 15) et un jeu de 25 données (points 1 à 25) – dans le but d'évaluer l'apport de données *hard* supplémentaires sur les résultats.

4.3.2 Données de hauteur piézométrique

Le modèle d'écoulement a été simulé en régime permanent dans les conditions de pompage considérées (débit de $60 \text{ m}^3/\text{h}$ au puits n°1) et a ainsi fourni la distribution piézométrique de référence de l'ensemble de la grille de calcul (figure 4.17).

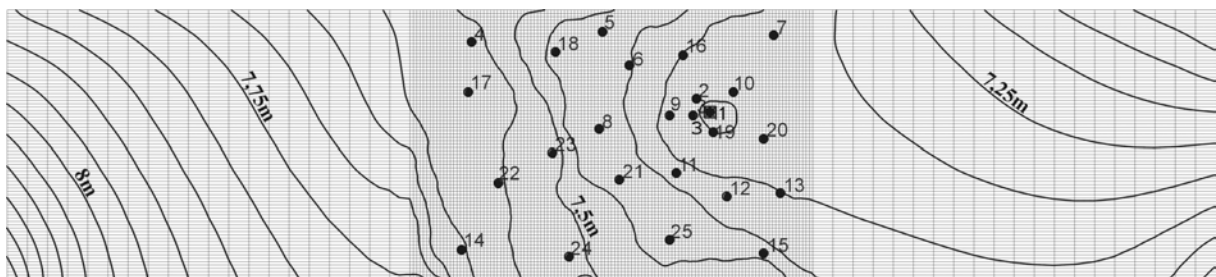


Figure 4.17 : Distribution piézométrique du modèle synthétique en condition de pompage ($60 \text{ m}^3/\text{h}$ au puits n°1), isochrones espacées d'un intervalle de 0,05m

Un jeu de 15 données et un jeu de 25 données d'observations piézométriques ont dès lors été créés à partir des niveaux d'eau observés en chaque point de mesure (ces données de hauteur piézométrique ont été prélevées aux mêmes points de mesure que les données de conductivité hydraulique). Ces deux jeux constitueront, dans la MSA, l'ensemble des données *soft* conditionnantes du processus de calibration du modèle d'écoulement.

4.3.3 Données de résistivité électrique

4.3.3.1 Corrélation conductivité hydraulique-résistivité électrique

HADDOUCHI [1987], MONJOIE & al. [1987] et DASSARGUES [1991] ont évoqué l'existence, dans la plaine alluviale de la Meuse, d'une corrélation entre la conductivité hydraulique K mesurée aux puits par essai de pompage et la résistivité électrique ρ des graviers saturés mesurée par sondage électrique à proximité immédiate des forages. Un total de 63 couples de données provenant de divers sites de la plaine alluviale de la Meuse pour lesquelles il existe simultanément une valeur de conductivité hydraulique et une valeur de résistivité électrique ont été mis en relation. Le diagramme de dispersion ($\ln K$, $\ln \rho$) pour l'échantillon de taille

$n = 63$ est représenté à la figure 4.18. Une première idée de la relation pouvant exister entre les deux variables est obtenue par l'examen de ce diagramme de dispersion.

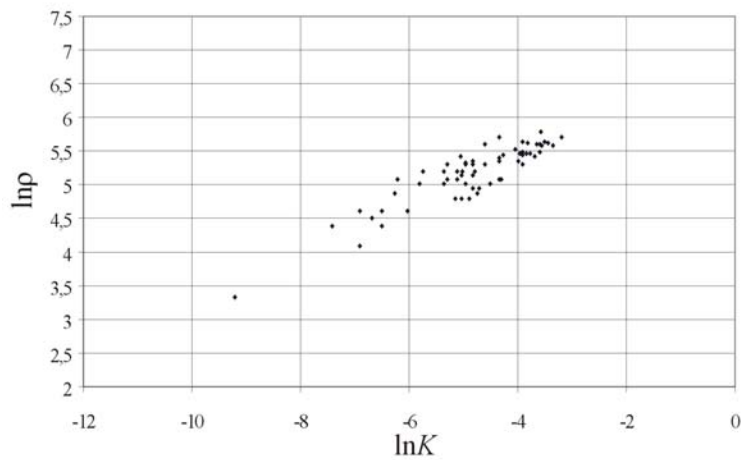


Figure 4.18 : Diagramme de dispersion ($\ln K, \ln \rho$) pour les 63 couples mesurés sur le terrain

Soit un échantillon aléatoire de taille n pour un couple $(X, Y) : (X_1, Y_1) \dots (X_i, Y_i) \dots (X_n, Y_n)$.

Les moyennes et les variances expérimentales (respectivement \bar{X} , \bar{Y} et S_x^2 , S_y^2) représentent respectivement les coordonnées du centre de gravité du nuage et la dispersion relative à ce centre de gravité et s'écrivent :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \qquad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \qquad (4.13)$$

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \qquad S_y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 \qquad (4.14)$$

La covariance expérimentale exprimant la dispersion du nuage est définie par :

$$S_{xy} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \qquad (4.15)$$

Le coefficient de corrélation expérimental a pour objectif de quantifier le degré de liaison (linéaire) entre deux variables. Il s'écrit :

$$\rho = \frac{S_{xy}}{S_x \cdot S_y} \qquad (4.16)$$

Le tableau 4.7 présente les valeurs expérimentales de la moyenne et de la variance pour chaque variable de l'échantillon étudié, ainsi que la covariance et le coefficient de corrélation expérimentaux.

	Moyenne \bar{X}	Variance S_x^2
$\ln K$	-4,83	1,29
$\ln \rho$	5,17	0,19
	Covariance S_{xy}	Coeff. corrélation ρ
$\ln K, \ln \rho$	0,45	0,9

Tableau 4.7 : Statistiques sur $\ln K$ et $\ln \rho$

Le coefficient de corrélation expérimental, égal à 0,9, traduit une forte corrélation entre les deux variables. D'après la forme du nuage de dispersion et la valeur du coefficient de corrélation, une relation linéaire entre $\ln K$ et $\ln \rho$ apparaît clairement. Cette relation peut être décrite par l'utilisation d'une régression. Le modèle de régression linéaire pour lequel la variable Y dépend d'une seule variable explicative X est appelé régression simple et s'écrit : $Y = b_0 + b_1 X$. Cette équation décrit une droite dans le plan (x,y) , appelée droite de régression. Une observation est notée $Y_i = b_0 + b_1 X_i + \varepsilon_i$, où ε_i représente la différence entre la valeur observée Y_i et sa valeur estimée Y , et est appelé *résidu*. L'ajustement de la droite de régression aux valeurs observées se fait par la méthode des moindres carrés. Cette méthode consiste à choisir comme estimateurs (\hat{b}_0, \hat{b}_1) , les valeurs qui rendent minimum la somme des carrés des résidus. La droite de régression estimée est reprise sur la figure 4.19 et s'écrit : $\ln \rho = 6,836 + 0,345 \ln K$.

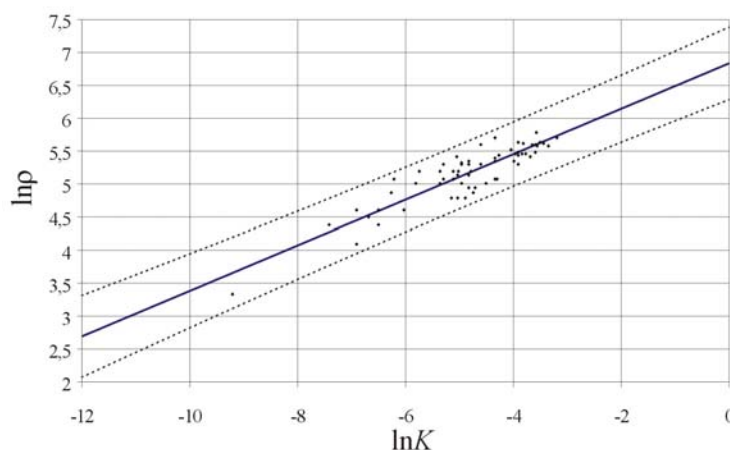


Figure 4.19 : Diagramme de dispersion $(\ln K, \ln \rho)$ des 63 données de terrain : droite de régression (en trait plein) et zone de prédiction à 95% (en pointillés)

La qualité de l'explication fournie par le modèle de régression est exprimée par le coefficient ρ^2 . Dans ce cas d'étude, $\rho^2 = 0,81$, ce qui signifie que la régression explique 81% de la variation de $\ln \rho$.

Le résidu ε est considéré comme une variable de loi normale $N(0,\sigma)$. En pratique, la variance résiduelle σ^2 n'est pas connue, mais peut être estimée à partir de la somme des carrés des résidus expérimentaux :

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \varepsilon_i^2}{n-2}, \text{ la variance des résidus expérimentaux} \quad (4.17)$$

où n est le nombre d'observations.

La moyenne des résidus $\bar{\varepsilon}$ calculée pour l'échantillon étudié est égale à $0,00014 \approx 0$. La figure 4.20, reprenant l'histogramme des résidus, permet de mettre en évidence le caractère normal de cette variable en la comparant à la loi normale $N(0;0,19)$.

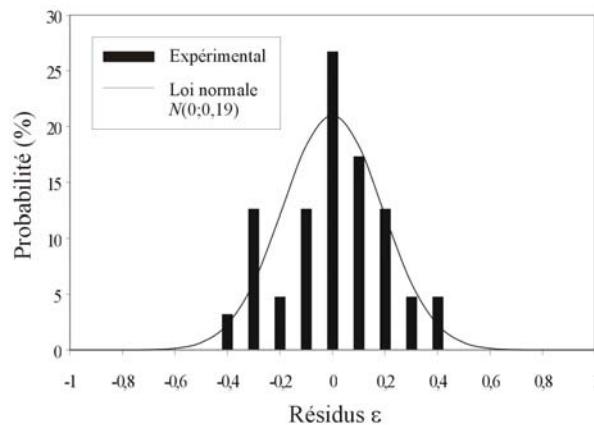


Figure 4.20 : Histogramme des résidus comparé à la loi normale $N(0;0,19)$

La zone de prédiction (ZP) d'une valeur individuelle quelconque Y en fonction de toute valeur possible X est calculée à partir de l'équation donnée ci-dessous [WEISBERG, 1980] :

$$ZP(Y(X), \forall X) : \left[\hat{b}_0 + \hat{b}_1 X \pm \sqrt{2F_{2,n-2}^{1-\alpha}} \cdot \sqrt{\hat{\sigma}^2 \cdot \left(1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{SC_X} \right)} \right] \quad (4.18)$$

avec SC_X , la somme des carrés des écarts en X : $SC_X = \sum_{i=1}^n (X_i - \bar{X})^2$

$F_{2,n-2}^{1-\alpha}$, le quantile d'ordre $1-\alpha$ de la distribution de Fisher-Snedecor à 2 et $n-2$ degrés de libertés.

La zone de prédiction à 95%, représentée par deux droites en pointillé sur la figure 4.19, a été calculée à partir des valeurs suivantes :

n	p	\hat{b}_0	\hat{b}_1	$\hat{\sigma}^2$	α	$F_{2;61}^{0,95}$
63	2	6,836	0,345	0,037	0,05	3,15

Tableau 4.8 : Valeurs des paramètres à introduire dans l'équation de la zone de prédiction

4.3.3.2 Création du jeu de résistivité électrique

La relation entre le logarithme népérien de la conductivité hydraulique et celui de la résistivité électrique a été vérifiée et quantifiée. Il est dès lors possible de créer un jeu de résistivités électriques à partir du champ de conductivités hydrauliques de référence. Plusieurs profils fictifs de tomographie électrique d'environ 150 mètres de long ont été positionnés sur la grille de calcul (figure 4.21).

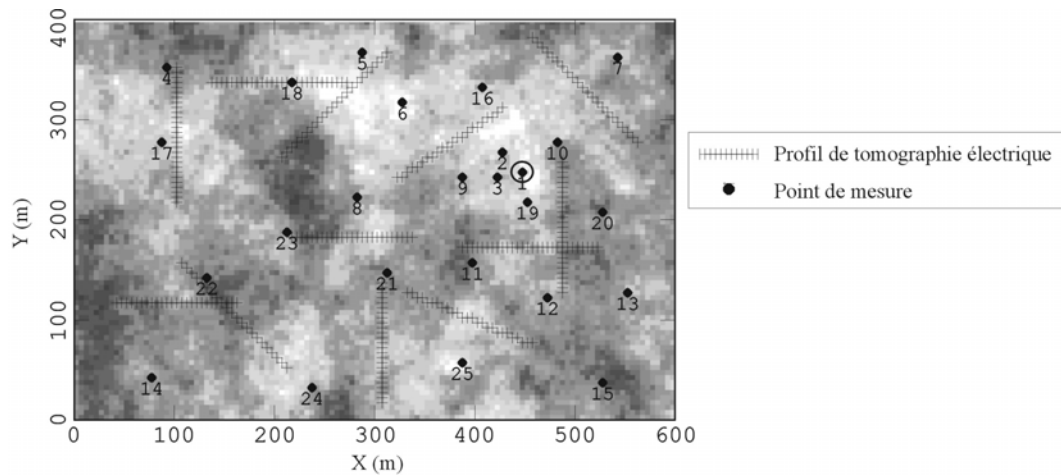


Figure 4.21 : Position des profils fictifs de tomographie électrique sur la grille de calcul

Les valeurs de conductivité hydraulique des p cellules situées au droit de ces profils ont été transformées en valeurs de résistivité électrique de la manière suivante :

$$\begin{aligned} \ln \hat{\rho}_i &= \hat{b}_0 + \hat{b}_1 \ln K_i + \hat{\sigma} \cdot N_i(0,1) \\ &= 6,836 + 0,345 \ln K_i + 0,19 \cdot N_i(0,1) \quad , \quad \forall i = 1, \dots, p \end{aligned} \quad (4.19)$$

où $N_i(0,1)$ représente un tirage au sort parmi une distribution normale centrée réduite.

Au total, 300 données de résistivité électrique (correspondant à 12 profils tomographiques) ont été créées, définissant deux jeux de 150 et de 300 données. La figure 4.22 représente le

diagramme de dispersion ($\ln K$, $\ln \rho$) pour l'échantillon de taille $n = 300$. Par construction, ce nuage est bien situé à l'intérieur de la zone de prédiction à 95% définie par l'équation (4.18).

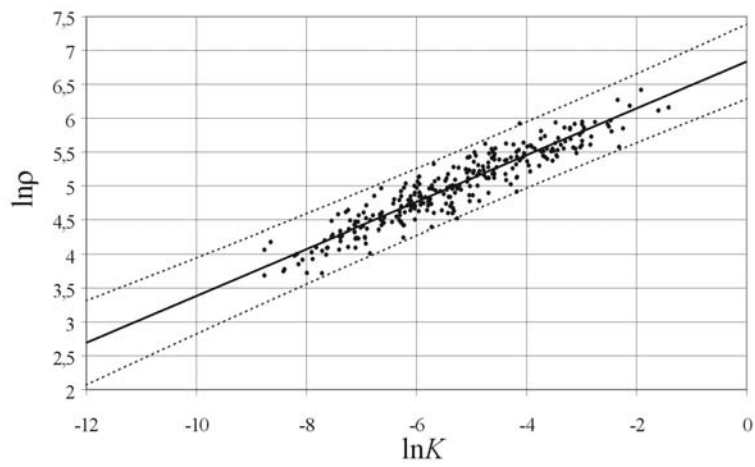


Figure 4.22 : Diagramme de dispersion ($\ln K$, $\ln \rho$) des 300 données du cas synthétique