Institut national de la recherche scientifique – Centre eau Terre et environnement
(Université du Québec, Canada)
Université d'Avignon et des Pays du Vaucluse
(Avignon, France)

Application de la modélisation neuronale à l'évaluation du risque de contamination des eaux souterraines par les pesticides

Présentée par
Cécile Doukouré

THESE EN COTUTELLE

pour obtenir le grade de

Philosophiae Doctor (Ph.D.) en Sciences de l'Eau de l'Institut National de la Recherche Scientifique – Centre Eau Terre et Environnement

et de

Docteur de l'Université d'Avignon et des Pays du Vaucluse
Spécialité : Hydrogéologie

Soutenue le 17 décembre 2007 devant le jury composé de :

Président du jury	**Professeur Jacky Mania** Polytech'Lille
Examinateur externe	**Philippe Lagacherie** INRA Montpellier
Examinateur interne	**Professeur Taha Ouarda** INRS-ETE
Directeurs de thèse	**Professeur Olivier Banton** Université d'Avignon et des Pays du Vaucluse
	Professeur Pierre Lafrance INRS-ETE

REMERCIEMENTS

Ce travail de thèse, financé par les Fonds québécois de la recherche sur la nature et les technologies (FQNRT), a été réalisé en cotutelle entre le centre Eau, Terre et environnement de l'institut national de la recherche scientifique (INRS-ETE) à Québec et le laboratoire d'hydrogéologie de l'Université d'Avignon et des Pays de Vaucluse (LHA).

Je tiens en premier lieu à remercier mes directeurs de recherche, les professeurs Olivier Banton et Pierre Lafrance de m'avoir permis de réaliser ce travail. Ils ont su m'encadrer en étant présents lorsque j'en avais besoin tout en me laissant explorer à ma guise ce travail de recherche, et c'est précisément ce qui me convenait. Un grand merci.

Je tiens également à remercier les membres de mon jury, messieurs Philippe Lagacherie, Jacky Mania et Taha Ouarda d'avoir accepté de juger mon travail.

Je remercie l'Agence de l'eau Rhône-Méditerranée et Corse qui a financé une partie des analyses de pesticide.

J'ai été amenée à travailler dans le cadre des sites d'études avec Frédéric Lalbat et Rémi de la Vaissière que je remercie pour leur collaboration et pour toutes les données précieuses qu'ils m'ont fournies. Du côté québécois, je remercie Pauline Fournier qui m'a beaucoup aidée avec les analyses de pesticides ainsi que René Lefebvre et Alexandre Bonton pour leurs données sur le site de Portneuf.

J'ai eu la chance durant ma thèse de travailler en parallèle sur deux projets de recherche. Ce furent des expériences très enrichissantes et qui m'ont énormément appris. Je tiens donc à remercier sincèrement Jean-Christophe Comte avec qui j'ai travaillé sur un projet de géophysique et qui a un don particulier pour rendre les choses passionnantes. Je remercie également Michel Nolin, Gérard Laflamme et Isabelle Beaudin avec qui j'ai travaillé sur un projet en pédologie.

Ayant partagé mon temps de travail entre le Québec et la France, j'ai été amenée à rencontrer beaucoup de personnes ; des collègues doctorants, des professeurs, des compagnons de bureau ou

de couloirs, et de véritables amis. Vous faites tous partie de cette période de ma vie et je vous en remercie.

Mes plus loyaux remerciements s'adressent à mes parents. Je ne saurais trouver les mots pour exprimer toute ma reconnaissance qui dépasse largement ces quelques années et ces quelques lignes. Une pensée également pour mes frères et Fatim.

Enfin, mes derniers mots vont à JB… merci pour ta patience, merci pour ta gentillesse, merci pour ton soutien.

RÉSUMÉ

Application de la modélisation neuronale à l'évaluation du risque de contamination des eaux souterraines par les pesticides

En milieu rural, la majorité de la population est alimentée par les ressources d'eau souterraine. Dans un souci de santé publique il est nécessaire de prévoir les risques de contamination de ces ressources. Les outils disponibles ne sont cependant pas toujours adéquats du fait de la difficulté à les mettre en place à une large échelle ou à cause du niveau de précision obtenu. Ce travail s'inscrit dans une démarche de prévision du risque de contamination des eaux souterraines par les pesticides.

Dans un premier temps, trois sites présentant des caractéristiques climatiques, hydrogéologiques et culturales différentes ont fait l'objet de campagnes d'échantillonnage afin d'étudier les corrélations pouvant exister entre la détection de pesticides et certains autres contaminants d'origine agricole.

Par la suite, l'application de la modélisation par réseaux de neurones artificiels a été testée. L'utilisation dans les méthodes neuronales des autres contaminants agricoles précédemment étudiés et d'autres variables chimiques caractérisant la géologie locale ou les temps de séjour de l'eau a permis d'évaluer le potentiel d'occurrence des pesticides avec des performances de l'ordre de 80 %. Cette approche, testée et validée sur différents jeux de données, se présente donc comme un compromis entre des modèles de simulation performants mais difficiles à mettre en place à de larges échelles et des indices de vulnérabilité d'application aisée mais de précision plus faible.

Mots clés : Réseaux de neurones, classification, pesticides, contamination de l'eau souterraine.

Merci à Matthieu Fournier

TABLE DES MATIERES

LISTE DES FIGURES

LISTE DES TABLEAUX

INTRODUCTION GÉNÉRALE

Les dernières décennies ont vu naître une prise de conscience des conséquences de l'emploi massif des produits phytosanitaires tant sur la santé humaine que sur l'environnement. Des concentrations mesurables en pesticide ont fréquemment été observées dans les eaux de surface, dans les eaux souterraines, dans les sédiments et dans l'atmosphère (Barbash *et al.* 1999). Alors que les transferts vers l'eau de surface sont relativement bien étudiés, les craintes concernant l'eau souterraine sont de plus en plus présentes. En effet, la fragilité de cette ressource, la perte des usages qui découle de son altération et l'extrême difficulté liée à sa réhabilitation font des eaux souterraines un point de protection fondamental. De nombreuses études ont mis en évidence une augmentation de la contamination des aquifères, conséquence de l'augmentation de la production agricole depuis les années 60 et de la généralisation en pratique courante de l'utilisation des pesticides. Cette préoccupation grandissante se traduit par l'établissement de normes essentiellement basées sur la potabilité des eaux de consommation. Ainsi, la directive européenne (98/83/EEC) limite les concentrations à 0.1 µg/l pour un pesticide donné et à 0.5 µg/l la concentration totale en pesticides dans l'eau destinée à la consommation. Au Québec, avec la Loi sur les pesticides (chapitre P-9.3), adoptée par l'Assemblée nationale du Québec en 1987, la réglementation quant à l'accès, l'entreposage et la vente des pesticides est contrôlée rigoureusement.

Il semble de plus en plus nécessaire de protéger les ressources en eau souterraine. Il est alors essentiel, pour y parvenir, de pouvoir prédire le transfert des solutés vers les eaux souterraines. Cependant, la l'évaluation de la contamination potentielle des aquifères est extrêmement difficile, car les mécanismes de transfert des polluants vers les aquifères sont très complexes et résultent de l'interaction d'un grand nombre de processus tel que l'adsorption sur les sédiments, la volatilisation et la dégradation. De plus, la grandeur de ces processus est elle-même influencée par une multitude de facteurs. Ces facteurs concernent les propriétés physico-chimiques du composé, les caractéristiques intrinsèques au site telles que les propriétés du sol, mais également des facteurs externes comme les pratiques culturales et les conditions climatiques.

De nombreuses méthodes ont été développées afin d'estimer le transport des pesticides dans le compartiment souterrain. Ces méthodes, des plus simples aux plus complexes, présentent une

grande variabilité d'approches en fonction de l'échelle à laquelle on souhaite travailler, de l'objectif visé et de la précision attendue. Les méthodes les plus simples permettent de visualiser des zones plus vulnérables qui, cependant, ne nous renseignent pas sur les concentrations potentielles de pesticides dans les eaux souterraines. D'un autre côté, plusieurs modèles numériques de simulation ont fait leur preuve dans la prédiction des transferts de pesticides à travers le sol. Cependant, ce type de modèle est souvent appliqué localement. Or des besoins se font ressentir à l'échelle régionale. En effet, c'est à cette échelle que les conséquences de l'emploi massif des produits phytosanitaires peuvent être appréhendées à long terme et que le suivi environnemental est délicat. Il devient donc indispensable de disposer de méthodes permettant d'évaluer les risques de contaminations à large échelle et c'est dans ce cadre que le travail de thèse s'insère.

Le mémoire s'articule en quatre chapitres :

- Le premier chapitre est essentiellement consacré à l'étude bibliographique du transport des pesticides vers les eaux souterraines et à la présentation des objectifs de la thèse. Nous nous intéresserons dans une première partie à la complexité du transport des pesticides qui rend compte de la difficulté de la mise en place systématique des outils de prédiction des contaminations ou des risques potentiels de contamination. La seconde partie permettra de replacer le travail au sein de cette problématique et d'en présenter les objectifs principaux.
- Le deuxième chapitre concerne l'étude de la relation entre les pesticides et différents contaminants agricoles. Dans une première partie, les sites d'étude et la stratégie d'échantillonnage seront présentés. Nous verrons ensuite les résultats quantitatifs de l'identification des pesticides dosés et présents sur les trois sites d'étude avant d'aborder statistiquement leur corrélation avec la présence des différents autres contaminants.
- Le troisième chapitre aborde l'application de la modélisation neuronale sur les données obtenues expérimentalement sur les trois sites d'étude. Les généralités sur les réseaux de neurones seront tout d'abord présentées avant d'aborder la méthodologie utilisée, les résultats et l'analyse des performances de classification en termes de détection de pesticides.

- Le quatrième chapitre traite de l'application des réseaux de neurones sur des données issues d'un suivi temporel. Ce jeu de données va nous permettre d'évaluer deux autres approches. Dans une première partie, nous verrons si la prise en compte l'historique de contamination d'un ouvrage permet d'améliorer les performances de classification. La seconde partie tentera d'évaluer si les variations de contamination dans un ouvrage peuvent être détectées.

Chapitre 1 Cadre de l'étude

Ce premier chapitre est essentiellement consacré à l'étude bibliographique des pesticides dans les eaux souterraines et à la présentation des objectifs de la thèse. Nous nous intéresserons dans une première partie à la complexité du transport des pesticides qui rend compte de la difficulté de la mise en place systématique des outils de prédiction des contaminations ou des risques potentiels de contamination. La seconde partie permettra de replacer le travail au sein de cette problématique et d'en présenter les objectifs principaux.

1.1 Les pesticides dans l'eau souterraine

Les produits phytosanitaires jouent un rôle majeur dans l'agriculture en protégeant les cultures et en assurant les productions végétales. Cependant, leur large utilisation depuis plusieurs décennies a mené à leur fréquente apparition dans les divers compartiments de l'environnement sous l'action de nombreux processus.

1.1.1 Définitions et caractéristiques

D'après le dictionnaire encyclopédique des sciences de l'eau (Ramade 1998), les pesticides sont des substances chimiques minérales ou organiques de synthèse utilisées à vaste échelle contre les ravageurs des cultures, les animaux nuisibles et les agents vecteurs d'affections parasitaires ou microbiologiques de l'homme et des animaux domestiques. Le terme produit phytosanitaire est plus spécifique des pesticides utilisés pour la protection des végétaux, cependant les deux termes seront employés ici sans distinction. Les pesticides de formulation commerciale sont composés de deux types de substance :

- Les matières actives qui confèrent au produit son effet toxique. Ce sont elles qui sont analysées et qui sont réglementées. Plusieurs matières actives peuvent être mélangées dans un même produit commercialisé.
- Les additifs qui ont pour but de permettre, voire renforcer, l'efficacité du produit.

La production de pesticides a connu une croissance considérable depuis la fin de la seconde guerre mondiale. Une tendance au ralentissement s'est observée depuis les années 80, liée en partie à la découverte de substances plus actives nécessitant des tonnages plus faibles et aussi à cause des graves problèmes écologiques et toxicologiques, soulevés par certaines de ces substances. En 2003, on dénombre plus de 800 matières actives entrant dans la composition de plus de 6000 produits commercialisés à travers le monde (IFEN 2003).

Les pesticides sont généralement classés en familles selon les organismes-cibles dont principalement les herbicides, les insecticides et les fongicides. Il existe cependant une multitude de groupes chimiques de pesticides dont les propriétés diffèrent grandement. Plusieurs

classifications ont été établies afin de les regrouper d'une façon pertinente en fonction des objectifs visés. Dans le contexte de la contamination des eaux souterraines, d'après (Weber 1994), l'essentiel est de considérer les principales propriétés physico-chimiques des pesticides, soit la solubilité dans l'eau, le degré de volatilisation, la rétention sur le sol et la persistance.

Au Québec, la proportion des puits individuels dans lesquels on a décelé la présence de pesticides est de 50 % pour les zones étudiées à proximité de la production de pomme de terre, de 40 % pour les zones étudiées à proximité des vergers et de 20 % pour les zones étudiées à proximité des cultures de maïs. Les principaux herbicides rencontrés dans l'eau souterraine sont l'atrazine et ses sous-produits de dégradation, puis le diuron, le métolachlore, la métribuzine et la simazine. Les insecticides les plus fréquemment rencontrés dans l'eau souterraine sont le carbaryl, le diazinon et l'imidaclopride (Giroux *et al.* 1997;Giroux 2003).

En France, le réseau national de surveillance de l'Institut Français de l'Environnement (IFEN) permet de faire un bilan des contaminations tous les ans. En 2004, des concentrations de pesticides ont été quantifiées sur 61 % des 910 points interprétables du réseau de connaissance y compris sur des nappes supposées captives (IFEN 2006). Sur ces 910 points de mesures, 27 % sont de qualité médiocre ou mauvaise et nécessiteraient un traitement spécifique d'élimination des pesticides s'ils étaient destinés à la consommation. Les principales molécules rencontrées sur l'ensemble du réseau de surveillance sont l'atrazine, la terbuthylazine et leurs sous-produits de dégradation, puis la simazine, l'oxadixyl et le diuron.

1.1.2 Difficulté de prédiction de la contamination des eaux souterraine par les pesticides

La prédiction du transport des pesticides vers les eaux souterraines est délicate car plusieurs processus interviennent. Ces processus vont en effet déterminer le devenir des pesticides dans l'environnement, mais leur grandeur est influencée par de nombreux facteurs qui regroupent d'une part les propriétés physico-chimiques du composé, les propriétés intrinsèques du sol, mais également des facteurs extrinsèques tels que les conditions climatiques et les pratiques culturales.

1.1.2.1 L'intervention de plusieurs processus

Lors de l'application de pesticides agricoles, seule une partie de la quantité épandue atteint réellement la cible visée (herbes, insectes ravageurs, champignons, etc.). Cette quantité peut représenter moins de 0.03 % de la dose appliquée (Pimentel 1995). Le restant est dissipé dans les différents compartiments de l'environnement que sont l'atmosphère, les eaux de surface, le sol et les eaux souterraines. La répartition dans chacun de ces compartiments est fonction de l'interaction d'un certain nombre de processus tels que la volatilisation pour le compartiment atmosphérique, le ruissellement pour les eaux de surface, l'adsorption et le lessivage pour le sol (Figure 1-1).

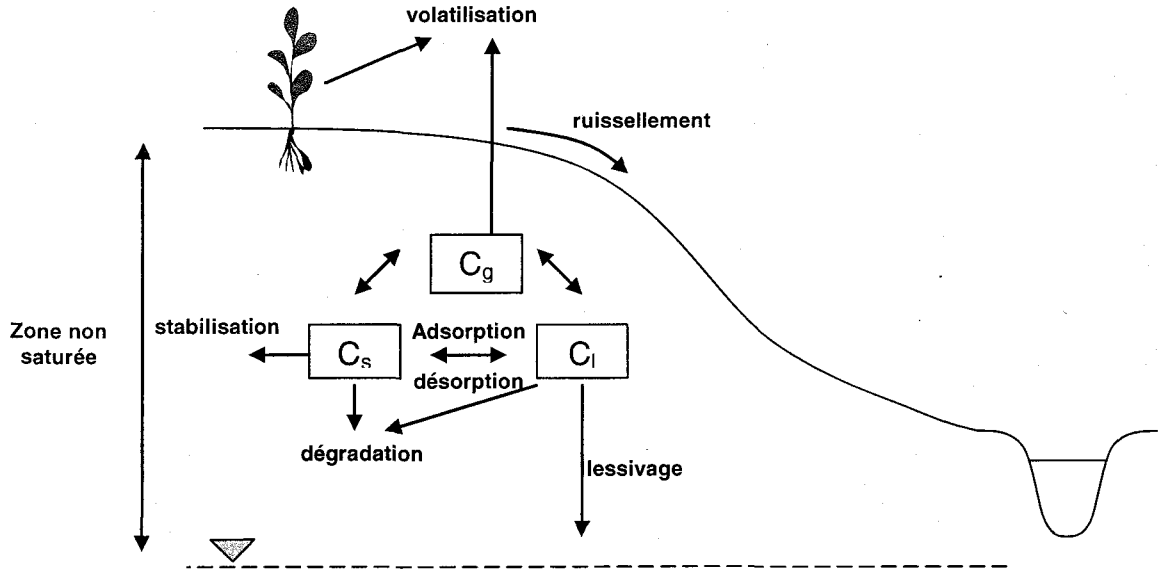


Figure 1-1. Processus intervenant dans le devenir des pesticides après l'application sur le sol. C_g : Concentration des pesticides en phase gazeuse ; C_s : Concentration des pesticides sur la phase solide ; C_l : Concentration des pesticides en phase liquide

La volatilisation

La volatilisation est le processus physico-chimique par lequel un composé est transféré en phase gazeuse. Elle peut avoir lieu à partir du sol ou des plantes et peut résulter soit de l'évaporation à partir d'une phase liquide, soit de la sublimation à partir d'une phase solide. Les pertes par volatilisation sont loin d'être négligeables, pouvant représenter jusqu'à 90 % de la dose appliquée (Carter 2000). De telles valeurs sont observées pour certains composés comme le lindane en

moins d'une semaine alors que pour d'autres composés tels que l'atrazine, les pertes cumulées après 24 jours représentent environ 2 % de la dose appliquée (Bedos *et al.* 2002). Ceci illustre bien l'importance des différentes propriétés des molécules. Cependant, même pour un composé donné, les pertes par volatilisation peuvent fortement varier en fonction des conditions environnementales.

Pour certains auteurs (Jury *et al.* 1983) cités dans (Bedos *et al.* 2002), le degré de volatilisation peut-être appréhendé en se basant essentiellement sur la constante de Henry, K_H. La constante de Henry représente le rapport, à l'équilibre, entre la fraction molaire du produit dans l'air et sa fraction molaire dans l'eau. Elle se présente ainsi comme l'équivalent d'un coefficient de partage air-eau qui permet d'évaluer la tendance d'un produit à se volatiliser. Plus la valeur de K_H (sans dimension) est élevée, plus le produit aura tendance à se volatiliser. Les auteurs déterminent trois classes de composés:

- les pesticides de catégorie I ($Kh >> 2.65.10^{-5}$) très volatils ;
- les pesticides de catégorie II (ayant des valeurs intermédiaires) moyennement volatils ;
- les pesticides de catégorie III ($Kh << 2.65.10^{-5}$) peu volatils.

D'après Bedos *et al.* (2002), la constante de Henry n'est pas suffisante pour appréhender les phénomènes de volatilisation. Bien que la pression de vapeur et la solubilité dans l'eau aient un rôle essentiel, l'adsorption sur le sol est également à considérer. En effet, si un composé s'adsorbe de manière significative au sol, les possibilités de volatilisation seront diminuées.

Outre les propriétés physico-chimiques des composés, le contexte environnemental a aussi une influence sur la grandeur du processus de volatilisation. Tout d'abord, les conditions atmosphériques agissent de diverses façons. Par exemple, la pression de vapeur est thermo-dépendante. Elle a donc tendance à augmenter avec la température ambiante. La vitesse du vent et l'humidité de l'air facilitent aussi la volatilisation (Grass *et al.* 1994). Les propriétés du sol interviennent également. La teneur en matière organique par exemple présente un effet indirect en modulant le processus d'adsorption du composé à la surface du sol.

Le ruissellement

Le ruissellement apparaît quand l'intensité des précipitations excède les capacités d'infiltration du sol (ruissellement hortonien) ou lorsque le sol est saturé (ruissellement par saturation). Les pesticides se trouvant dans la partie supérieure du sol peuvent alors être entraînés dans l'eau de ruissellement, responsable des principaux cas de contamination des eaux de surface. Ils se présentent sous forme dissoute ou adsorbée à la surface des sédiments (Leonard 1990). D'après plusieurs études, ce sont les pesticides présents dans les 2 à 10 premiers millimètres du sol qui peuvent être transférés dans l'eau de ruissellement (Triegel et Guo 1994).

Les pertes par ruissellement représentent environ 0.5 % de la quantité appliquée pour une grande majorité des pesticides (Wauchope 1978). Cependant en fonction du type d'application et du type d'événement pluvial qui survient, les concentrations peuvent atteindre jusqu'à 11 % de la quantité appliquée (Wauchope 1978).

Plusieurs facteurs vont avoir une influence sur la quantité de pertes par ruissellement. C'est le cas de l'intensité de l'événement pluvial, mais surtout du temps écoulé depuis l'application, car la fraction de pesticides disponible dans la partie du sol sujette au ruissellement diminue très rapidement par volatilisation, décomposition et lessivage. D'autres paramètres sont également à prendre en considération dans la capacité de ruissellement tels que la nature du sol, la topographie et le couvert végétal (de Bruyn 2004).

Lessivage dans le sol

Le lessivage est la résultante des processus de transport des solutés à travers la zone non saturée du sol et peut être à l'origine de la pollution des eaux souterraines. Ce transport est régi par trois processus physiques principaux : l'advection, la dispersion et la diffusion moléculaire. L'impact des processus chimiques (adsorption) sur le contaminant sera décrit à la section suivante.

L'advection correspond à l'écoulement de l'eau souterraine causant la migration du soluté. Dans le cas de la zone saturée, le mouvement résulte de la présence d'un gradient hydraulique (dh/dx) dans le milieu poreux. Il est proportionnel à la conductivité hydraulique K et inversement proportionnel à la porosité cinématique n_c:

$$\bar{v} = -\frac{K}{n_c}\frac{dh}{dx}$$

où $\bar{v}$ est la vitesse moyenne de l'eau. Cependant, les vitesses d'écoulement varient en fonction des pores créant, pour le soluté, un étalement du profil de concentration. C'est la dispersion mécanique. Un autre processus crée également un étalement du soluté dû à la présence d'un gradient de concentration. Il s'agit de la diffusion moléculaire qui uniformise la concentration au travers du système.

Des études ont par ailleurs constaté que des solutés ont été transportés plus profondément que ce qui était attendu d'après la vitesse d'écoulement en milieu poreux homogène. Les flux préférentiels (macroporosité structurale du sol) sont souvent tenus comme responsables d'une telle différence entre les vitesses anticipées et les observations de terrain (Jury et Fluhler 1992). La forme sous laquelle les solutés sont transportés par ces flux préférentiels est incertaine, mais le transport d'une forme adsorbée sur les colloïdes est l'un des mécanismes possibles (Gooddy *et al.* 2001).

L'adsorption et la complexation

L'adsorption est le mécanisme réversible par lequel un composé chimique en solution se fixe à la surface d'une particule de sol. Le degré pour lequel un pesticide préfère une phase ou une autre va affecter tous les autres aspects de son comportement dans le sol et déterminer son potentiel de contamination pour les eaux souterraines. Généralement, l'adsorption est caractérisée par un coefficient de partage sol-eau, Kd, mesuré expérimentalement, représenté par la pente d'une isotherme d'adsorption linéaire (Wauchope *et al.* 2002).

$$Kd = \frac{x/m_s}{C_e}$$

Avec x/m_s la concentration de pesticides sur la phase solide ($g.g^{-1}$) et C_e la concentration à l'équilibre en phase aqueuse ($g.l^{-1}$). Lorsque cette constante prend des valeurs importantes de l'ordre de 100, les pesticides sont fortement adsorbés par le sol et restent quasiment immobiles (Wauchope *et al.* 2002).

L'adsorption crée un retard dans le transport du soluté par rapport à la vitesse moyenne de l'eau. Le plus souvent ce phénomène est décrit par le facteur retard (R) :

$$R = 1 + Kd\frac{\rho}{\theta}$$

où ρ est la densité apparente sèche du milieu solide et θ le contenu en eau. La vitesse du contaminant v_c est alors décrite par la relation suivante : $v_c = \frac{\bar{v}}{R}$

Des études ont montré une bonne corrélation entre le Kd et la teneur en matière organique du sol, d'où l'hypothèse que la matière organique du sol constitue le principal sorbant pour de nombreuses molécules hydrophobes. Il en découle l'utilisation fréquente du paramètre Koc qui correspond au Kd rapporté à la fraction en carbone organique du sol :

$$K_{OC} = \frac{Kd}{f_{OC}}$$

Avec f_{OC} la fraction de carbone organique du sol.

Cependant, Wauchope *et al.* (2002) précisent que l'utilisation de ce coefficient d'adsorption présente des limites qu'il est nécessaire de connaître :

- La matière organique du sol n'est pas le seul sorbant. D'autres éléments du sol tels que les surfaces minérales argileuses (Laird *et al.* 1994) peuvent également adsorber des composés légèrement plus solubles ou chargés, particulièrement dans le cas des sols où la teneur en matière organique est faible (Means *et al.* 1982).
- Les processus d'adsorption présentent deux types de cinétiques. Une première phase d'adsorption rapide et réversible apparaît en quelques minutes au niveau de l'interface sol/eau. Ensuite vient une phase lente peu réversible qui nécessite quelques semaines à plusieurs années pour atteindre un pseudo-équilibre (Kan *et al.* 1994).
- Des phénomènes de dégradation peuvent apparaître pendant les expérimentations de sorption. Les pertes par dégradation peuvent alors être considérées comme des pertes dues à une adsorption irréversible.

D'autre part, les composantes de la matière organique du sol n'ont pas toutes la même affinité pour les composés organiques de synthèse. Or la composition et la structure de la matière organique varient en fonction de son origine et de son histoire pédologique ce qui influence son affinité pour les composés (Grathwohl 1990). Garbarini et Lion (1986) ont montré une variation dans la sorption de composés organiques pour différentes fractions de matière organique. Il y a donc, dans certains cas des limites dans l'application des valeurs de Koc trouvées dans la littérature si l'on ne considère pas la nature de la matière organique impliquée. Ces auteurs ont montré que le rapport C/O était mieux corrélé au Kd que le carbone organique seul.

La dégradation

Dans le sol et dans l'eau, les pesticides sont soumis à des dégradations par voie biotique et abiotique. Le taux de dégradation d'un pesticide est un facteur clé conditionnant le potentiel de contamination de l'eau souterraine par les pesticides (Di *et al.* 1998). Le taux de dégradation est généralement caractérisé par la demi-vie du composé, DT50 exprimée en jours. D'après (Kerle *et al.* 1996), les pesticides peuvent être regroupés en trois catégories :

- les pesticides non persistants avec un DT50 < 30 jours ;
- les pesticides moyennement persistants avec un DT50 entre 30 et 100 jours ;
- les pesticides persistants avec un DT50 > 100 jours.

Dégradation abiotique

Les transformations chimiques peuvent avoir lieu pour les composés en solution ou adsorbés à la surface des particules de sol. Les facteurs qui influencent le plus ce type de dégradation chimique sont la nature du pesticide, la température, le pH et la teneur en eau du sol. Les principales voies de dégradation abiotique sont :

- L'hydrolyse : c'est le processus par lequel un composé réagit avec l'atome d'oxygène de l'eau. Il en résulte un clivage de la molécule en plusieurs sous-produits. Elle est influencée par le pH de l'eau.

- Les réactions d'oxydoréduction : en général, on observe plutôt une oxydation dans les eaux de surface et une réduction dans les sédiments et les zones anaérobies (Triegel et Guo 1994).
- La photolyse : ce phénomène concerne essentiellement les composés exposés à la surface du sol. Les UV sont les rayons qui dégradent le plus les pesticides. L'intensité de la lumière, le temps d'exposition et les propriétés du pesticide vont affecter le taux de photodégradation (Kerle *et al.* 1996).

Dégradation biotique

C'est la voie principale de dégradation des pesticides dans l'environnement. L'activité microbienne peut conduire à la métabolisation complète des molécules. Les micro-organismes utilisent les pesticides comme source de carbone (minéralisation). Tous les facteurs, tels que la température, l'humidité du sol, le pH, qui affectent l'activité microbienne, auront donc un effet sur les processus de biodégradation. En général, le taux de dégradation est plus élevé dans l'horizon de surface du sol où la teneur en matière organique est plus importante et diminue avec la profondeur dans le sol, là où les conditions telles l'humidité, la température et l'aération sont moins favorables à l'activité microbienne. Cette dégradation microbienne est catalysée par des enzymes avec des températures préférentielles entre 10 et 45 °C. En deçà ou au-delà, la bioactivité est nettement réduite.

Bien que certains sous-produits de dégradation apparaissent comme moins toxiques que les composés parents (Heydens *et al.* 1996), d'autre possèdent une toxicité similaire, voire plus élevée (Tessier et Clark 1995). C'est pourquoi la prise en compte des sous-produits de dégradation dans les systèmes hydrologiques est essentielle à la compréhension des conséquences sur l'environnement et la santé humaine de l'utilisation des pesticides (Kolpin *et al.* 1998).

1.1.2.2 L'intervention de plusieurs facteurs

De nombreux facteurs, tant naturels qu'anthropiques, vont affecter la probabilité de retrouver des pesticides dans les eaux souterraines. Certains de ces facteurs ont été brièvement abordés dans la première partie, car ils sont essentiels à la compréhension des processus majeurs. Cependant,

d'autres variables tout aussi essentielles sont également à prendre en considération pour comprendre l'occurrence des produits phytosanitaires dans les eaux souterraines. Nous ferons donc dans cette section un rappel des principaux paramètres ayant une influence sur le devenir des pesticides dans les sols et les eaux souterraines.

Les propriétés physico-chimiques des composés

Plusieurs études ont tenté d'évaluer l'importance des propriétés physico-chimiques des molécules sur leur transport vers les eaux souterraines (Worrall F. *et al.* 2002) ; (Tariq *et al.* 2004). Il a été montré que celles-ci sont tout aussi importantes que les caractéristiques du site afin d'appréhender l'occurrence des contaminations. Aussi, il en ressort que si l'on considère tous les autres facteurs égaux (caractéristiques environnementales, propriétés physiques du site), la probabilité de trouver un pesticide plutôt qu'un autre dans l'eau souterraine est principalement liée à son degré de partition entre les phases solide et aqueuse (Koc) et à sa résistance aux dégradations chimiques et biologiques dans le sol (DT50) (Barbash *et al.* 2001).

Les propriétés intrinsèques du site

Les propriétés intrinsèques du site sont indissociables de la compréhension du devenir des pesticides dans l'eau souterraine. Elles concernent tant les caractéristiques du sol que les caractéristiques hydrogéologiques du site. Dans la zone non saturée, le transport par advection des pesticides est directement lié à la conductivité hydraulique du sol et à la porosité. La structure du sol détermine les mouvements de l'eau et, dans une certaine mesure, le taux d'interaction entre celui-ci et les pesticides. Le degré d'agrégation des particules de sol, l'activité biologique qui y est présente, la teneur en matière organique et en argile vont ainsi influencer de nombreux processus et avoir un impact sur le comportement des pesticides (Carter 2000).

Les caractéristiques hydrogéologiques qui vont influencer le devenir des pesticides sont en premier lieu celles qui contrôlent le mouvement de l'eau. Une fois que les pesticides ont atteint la zone saturée, leur mouvement dépend du gradient hydraulique et de la conductivité hydraulique. Le degré de protection de l'aquifère va également déterminer la probabilité de déceler des pesticides dans l'eau souterraine.

Les facteurs externes

Les facteurs externes qui vont avoir une influence sur le devenir des pesticides dans l'environnement concernent essentiellement les pratiques culturales et les facteurs climatiques. Les pratiques culturales vont tout d'abord déterminer la quantité de pesticide disponible pour un lessivage potentiel en profondeur. Ainsi, l'occupation du sol est un facteur essentiel. Alors qu'il peut sembler évident de retrouver les pesticides dans les zones où ils sont utilisés, Barbash *et al.* (2001) ont tenu à le démontrer statistiquement. Il en ressort que la détection de pesticide est significativement plus fréquente dans les zones de forte utilisation agricole. (McKenna *et al.* 1990) ont étudié l'influence des pratiques culturales sur la qualité de l'eau souterraine. Pour cela, ils regroupent ces facteurs en deux catégories : d'une part, ceux qui concernent les pratiques au sens strict (composé, mode d'application, pratiques de labour et rotations) et d'autre part les pratiques de gestion de l'eau qui concernent essentiellement l'irrigation et le drainage.

Les facteurs climatiques tels que la température, les précipitations, le vent et l'humidité relative influencent le comportement des pesticides appliqués à la surface du sol. D'autre part, l'intensité et la distribution des précipitations vont moduler la quantité d'eau disponible pour l'entraînement des composés dissous au travers de la zone racinaire.

1.1.3 Les outils de quantification de la contamination

Il est aujourd'hui essentiel de pouvoir évaluer la contamination des eaux souterraines par les pesticides. Cette évaluation devrait passer par la considération de tous les processus et des facteurs qui ont une influence. Certes, les mesures directes restent le meilleur moyen de déterminer l'impact de l'utilisation des produits sur les aquifères, mais ces mesures restent très coûteuses et les campagnes à large échelle sont donc limitées. Un certain nombre d'outils permettant d'appréhender ces contaminations sont aujourd'hui disponibles. Ces outils présentent chacun des avantages et des inconvénients en mettant l'accent ou en négligeant certains des aspects du transfert des contaminants. D'après Zhang *et al.* (1996), il existe trois types d'outils nous permettant d'appréhender la contamination de l'eau souterraine par les pesticides :

- les indices qui combinent des caractéristiques physiques et fournissent un résultat sous forme de score ;

- les méthodes statistiques qui tentent de trouver des corrélations entre les différents facteurs et les zones où les contaminations sont observées ;
- la modélisation mathématique basée sur la représentation des processus afin de simuler le comportement des substances.

1.1.3.1 Indices et indicateurs

Plusieurs approches et outils ont été développés afin d'évaluer l'impact relatif des pesticides sur l'environnement (van der Werf 1996). Ces approches varient dans leur complexité et dans les paramètres pris en compte. On distingue les indices de vulnérabilité non spécifiques d'un contaminant donné des autres méthodes spécifiques des pesticides qui tiennent compte de leurs propriétés physico-chimiques. Dans la littérature, ces dernières sont souvent désignées par le terme indicateur, mais la distinction entre indice et indicateurs reste floue.

Les indices de vulnérabilité de l'eau souterraine sont fondés sur l'hypothèse qu'un nombre défini de paramètres majeurs contrôle largement le devenir des contaminants dans la zone non saturée du sol, que ces paramètres sont connus et qu'ils peuvent être évalués. La plus répandue de ces méthodes est l'indice DRASTIC (Aller *et al.* 1987) qui utilise un système établi sur sept caractéristiques pédologiques et hydrogéologiques de la région concernée. D'autres indices de ce genre ont par la suite été développés, tels que GOD (Foster 1987) et EPIK (Doerfliger *et al.* 1999). Typiquement, ces méthodes relient des variables concernant la recharge de l'aquifère, le niveau d'eau et les propriétés du sol. Elles sont basées sur des concepts simples issus de l'opinion d'expert et non pas sur la représentation des processus. L'avantage de ces indices est de fournir des critères de décision simples, le plus souvent destinés aux gestionnaires. Cependant, ce type de méthode présente un certain nombre d'inconvénients. Premièrement, le manque de base physique en fait des outils subjectifs dont les résultats sont souvent incertains et parfois contradictoires (Gogu *et al.* 2003). La prise en compte ou l'exclusion de certaines variables est arbitraire de même que le poids des paramètres dans le calcul de l'indice final (Worrall F. *et al.* 2002). Ensuite, ces indices ne sont pas basés sur des observations ou sur des mesures de la contamination de l'eau souterraine et sont rarement validés ou confrontés à des observations de terrain (Merchant 1994) cité dans Worrall *et al.* 2002). D'autre part, les propriétés physico-chimiques du contaminant ne sont généralement pas considérées, ce qui implique que

l'occurrence d'un produit varie uniquement en fonction des caractéristiques physiques du site ou des conditions climatiques. D'après Banton et Villeneuve (1989), les caractéristiques chimiques des contaminants qui ne sont pas considérés dans ce type d'indice sont déterminantes dans l'évaluation du potentiel de contamination des eaux souterraines par les produits agricoles.

D'autres méthodes ne considèrent que les propriétés physico-chimiques du contaminant. Gustafson (1989) a développé un indice basé sur des observations afin de prédire si un composé peut ou non être lessivé dans l'eau souterraine. De telles méthodes considèrent que les propriétés du composé conditionnent à elles seules le devenir du composé dans le sol et ne prennent pas en compte les variations naturelles de ces paramètres tels que le Kd qui peut varier fortement en fonction du site.

Les indicateurs permettent d'accéder à un niveau de précision supérieur en combinant les données du site et les propriétés du composé. Ils diffèrent en fonction de la procédure de calcul et de l'importance accordée aux facteurs d'application et aux conditions environnementales. Reus *et al.* (2002) ont comparé et évalué huit indicateurs dont sept prennent en considération le compartiment souterrain. Les propriétés des pesticides que tous ces indicateurs prennent en compte sont la persistance dans le sol (DT50) et la mobilité (Koc). Trois de ces indicateurs utilisent la constante d'Henry afin de prendre en compte la volatilisation. Les indicateurs tels que EPRIP (Padovani *et al.* 2004) et SyPEP (Pussemier 1999) utilisent une approche par ratio, c'est-à-dire un rapport entre l'exposition (le plus souvent la concentration dans un compartiment donné) et la toxicité. Les concentrations sont généralement déterminées par des algorithmes complexes issus de modèles numériques. Le score produit peut être basé directement sur la concentration alors que certaines méthodes convertissent les concentrations en indice compris entre 0 et 5. PERI et p-EMAb (Lewis *et al.* 2003) sont des indicateurs qui utilisent l'indice GUS comme base pour déterminer le potentiel de contamination de l'eau souterraine tandis que IPEST (van der Werf et Zimmer 1998) combine l'indice GUS avec un système basé sur la logique floue.

Reus *et al.* (2002) précisent qu'un indicateur doit fournir une information fiable, ce qui passe nécessairement par une validation. Cependant, la validation de ce type d'outil est difficile, car les méthodes d'indicateur produisent le plus souvent un système de score sans unité.

1.1.3.2 Modélisation

Les modèles numériques de simulation du transfert de pesticides dans l'eau souterraine sont des outils très utiles à la compréhension des processus et à l'analyse des problèmes pouvant affecter les eaux souterraines, en prédisant des sorties et en permettant d'appliquer différents scénarios. Ils ont l'avantage de présenter des résultats quantitatifs, en terme de concentration par exemple qui peuvent être comparés avec les critères de la qualité de l'eau (Banton et Villeneuve 1989 ; Lindström 2005).

Il existe une vaste gamme de modèles numériques, allant des plus simples aux plus complexes. D'après Loague et Abrams (2001), le choix d'un modèle se base en premier lieu sur deux critères : l'échelle à laquelle on souhaite travailler et la finalité de l'outil. Les finalités peuvent être diverses et nombreuses. En effet, les modèles peuvent servir à prédire et à comprendre le comportement des contaminants dans le système, à faire des tests sur la mobilité et la persistance des nouvelles molécules en développement, ou servir de base à la gestion des cultures et des pratiques (Wagenet et Rao 1990). Il existe ainsi différents types de modèles présentant une grande variabilité d'approches en fonction du degré de complexité, de l'environnement pour lequel ils ont été développés, et du développeur lui-même.

En règle générale, on distingue les modèles mécanistes des modèles empiriques. Les modèles empiriques tels que GLEAMS (Leonard 1990) utilisent souvent des représentations simplifiées des processus fondamentaux, ce qui en fait des outils attrayants, car ils ne requièrent pas de grandes quantités de données d'entrée. Ce type d'approche, s'adressant le plus souvent à des gestionnaires, ne nécessite que des approximations des variables du système. Cependant, la capacité de prédiction de ces modèles est limitée et reste essentiellement qualitative en terme de mouvement d'eau et de solutés (Wagenet et Rao 1990). Ces modèles peuvent cependant constituer des outils très intéressants lorsqu'ils sont correctement utilisés, par exemple pour grouper les pesticides en classes comportementales. Les modèles mécanistes tels que LEACHM (Wagenet et Hudson, 1987), AgriFlux (Banton *et al.* 1997), PRZM (Carsel *et al.* 1985) ou PEARL (Tiktat *et al.* 2000) utilisent des représentations physiques et mathématiques afin de reproduire le plus fidèlement possible les processus qui interviennent. À l'inverse des modèles

empiriques, la difficulté de l'utilisation de modèles plus complexes est la quantité de données requises. Leur capacité de prédiction a souvent été validée avec succès à l'échelle de la parcelle.

Le deuxième degré de distinction se fait dans la prise en compte des incertitudes. Il en découle deux types de modèles : les modèles déterministes et les modèles stochastiques. Un modèle est déterministe s'il ne fait pas appel au calcul de probabilité. Il est stochastique si les variables sont décrites par une distribution de probabilité. Ce dernier type de modèle s'observe de plus en plus avec à la prise de conscience d'un certain nombre d'incertitudes impliquées dans le transfert des pesticides dans le sol. Ces incertitudes peuvent être liées aux données d'entrée, mais également au modèle lui-même.

Les incertitudes sur les données d'entrée des modèles peuvent comprendre notamment les caractéristiques du site, les propriétés du sol et les conditions climatiques. Elles sont dues aux variabilités spatiales et temporelles des paramètres, mais également aux procédures d'échantillonnage ou d'analyse en laboratoire (Dubus *et al.* 2003). Les incertitudes du modèle représentent son incapacité à représenter correctement les processus même si tous les paramètres et données sont estimés avec une bonne fiabilité (Loague et Corwin 1998). Il est important de connaître la dimension des incertitudes sur les sorties d'un modèle, particulièrement lorsque le modèle est utilisé à des fins de prédiction. La modélisation stochastique des processus de devenir des pesticides est plus adaptée à l'évaluation de la quantité de contaminants atteignant la nappe d'eau souterraine, car cette approche prend en compte les variabilités spatiales des caractéristiques physiques et hydrodynamiques du sol (Lafrance et Banton 1995).

Dans le contexte du transfert de pesticides dans les eaux souterraines à l'échelle régionale, les modèles calculent soit le transport vertical dans la zone non saturée, soit le transport dans la zone saturée (Herbst *et al.* 2005). On assiste de plus en plus à des techniques de couplage de ces deux types de modèles afin de représenter réellement des processus complets de transfert dans l'eau souterraine, avec une prise en compte des flux latéraux qui interviennent dans la zone saturée. Loague *et al.* (1998a ; 1998b) par exemple, ont utilisé le modèle PRZM-2 (Mullins *et al.* 1993) pour caractériser le lessivage à l'échelle régionale du DBCP et ont intégré les résultats comme entrée dans un modèle 3D de transport dans la zone saturée. Les couplages engendrent cependant

un certain nombre d'inconvénients tels que la fixation d'une profondeur du sol à laquelle sont couplés les deux modèles ou le fait de négliger les remontées capillaires.

Étant donné que la plupart des modèles de transfert de pesticides dans la zone non saturée sont unidimensionnels, le couplage avec les modèles en 3D de la zone saturée requiert un moyen de spatialiser les sorties du modèle unidimensionnel. Refsgaard et Butts (1999) (cités dans (Stenemo *et al.* 2005) identifient quatre approches pour l'application d'un modèle unidimensionnel à large échelle :

- l'agrégation qui consiste à appliquer le modèle à fine échelle et à en agréger les résultats ;
- les équations sont considérées valides à large échelle ;
- les équations du modèle sont étendues à large échelle sur une base théorique qui explique la prise en compte de la variabilité spatiale des paramètres du modèle ;
- de nouvelles équations sont développées.

Heuvelink et Pebesma (1999) recommandent d'utiliser l'agrégation spatiale qui permet d'une part d'éliminer les biais dus à l'application des équations à une échelle pour laquelle elles n'ont pas été développées, et d'autre part de montrer comment les incertitudes dans les entrées du modèle se propagent à la sortie.

1.1.3.3 Méthodes statistiques

Les méthodes statistiques sont le plus souvent utilisées pour déterminer la relation ou la dépendance entre des contaminations observées et les conditions environnementales. Une fois ces relations établies, elles permettent d'évaluer une probabilité de contamination. Teso *et al.* (1996) par exemple ont utilisé une régression logistique pour prédire la vulnérabilité d'un aquifère à partir des caractéristiques du sol. Le désavantage de ces méthodes statistiques est qu'elles sont difficiles à développer, et une fois établies, elles ne peuvent le plus souvent être appliquées que dans des régions qui possèdent des caractéristiques environnementales similaires à la région pour laquelle elles ont été développées (Lindström 2005).

De nombreuses autres méthodes statistiques sont utilisées en parallèle des autres outils, soit pour évaluer les incertitudes, soit pour spatialiser les données par krigeage par exemple (Lindström 2005). Depuis une dizaine d'années, on assiste également au développement de l'utilisation de réseaux de neurones artificiels pour quantifier les contaminations. Les réseaux de neurones sont des modèles non linéaires multi-entrées multi-sorties agissant comme une boîte noire, et permettant de représenter les interactions complexes entre les paramètres d'entrée et les paramètres de sortie (Bockstaller et Girardin 2003). Leur application sera développée au chapitre 3.

1.2 Justification et objectifs de la présente étude

Il est à présent reconnu que les eaux souterraines sont sujettes aux contaminations par les produits phytosanitaires aussi bien en Europe qu'en Amérique du Nord. Ce phénomène est d'autant plus préoccupant que cette ressource constitue l'approvisionnement en eau potable de la majorité des populations rurales et qu'elle est difficilement réhabilitable. Alors que les analyses pour déceler la présence de certains contaminants tels les nitrates sont fréquentes dans les puits et forages privés, les produits phytosanitaires ne font pas partie des paramètres standards du suivi de la qualité de l'eau et leur analyse reste rare ou fragmentaire. L'échantillonnage systématique des ouvrages domestiques à large échelle n'est pas économiquement réalisable. Ceci est dû au coût élevé des analyses, mais également à la grande diversité des produits phytosanitaires utilisés.

Comme l'a montré la revue de littérature, la prédiction du potentiel de contamination est difficile pour les raisons suivantes :

- plusieurs processus complexes interviennent dans le devenir des pesticides ;
- ces processus interagissent entre eux ;
- la grandeur de ces processus est influencée par des facteurs externes variables dans le temps et dans l'espace.

La prédiction du risque de contamination devrait donc passer par la prise en compte de tous ces processus et facteurs, ce qui serait extrêmement lourd à mettre en pratique. Cela implique une excellente connaissance des sites, tant sur le plan physique qu'au niveau des pratiques culturales qui y sont appliquées. Les outils plus complexes tels les modèles mécanistes permettent effectivement d'estimer des concentrations. Cependant, ces outils sont souvent développés pour une échelle locale. Les méthodes appliquées à large échelle sont le plus souvent des indices de vulnérabilité, qui délimitent de grandes zones, mais qui ne traitent pas du risque d'occurrence pour un point donné. De nouvelles méthodes alternatives susceptibles de prédire les contaminations par les pesticides des eaux captées par les ouvrages municipaux et domestiques sont donc nécessaires.

Les réseaux de neurones artificiels (RNA) sont des outils de modélisation non linéaires qui ne nécessitent pas de formulation physique du problème. Ils sont utilisés dans les sciences de l'eau

depuis une dizaine (Maier et Dandy 2000 ; Maier et Dandy 1996). Plus spécifiquement pour les pesticides, plusieurs travaux ont été effectués avec des résultats intéressants en termes de performance du réseau (Mishra *et al.* 2004;Ray et Klindworth 2000;Sahoo *et al.* 2006;Sahoo *et al.* 2005;Yang *et al.* 2003). Cependant, les données d'entrée de ces modèles restent le plus souvent les mêmes que celles des modèles classiques, ce qui permet donc d'éviter la formulation physique de la relation, mais ce qui ne simplifie pas le travail de collecte des données.

Parallèlement, plusieurs études ont montré des corrélations positives entre la détection de pesticides dans l'eau souterraine et les concentrations de certains paramètres chimiques (Istok et Rautman 1996) ; (Burow *et al.* 1998) ; (Kolpin *et al.* 1998). Par exemple, Kolpin *et al.* (1998) ont observé que des caractéristiques similaires d'aquifères étaient corrélées avec la présence simultanée de pesticides et de concentrations excessives en nitrates. Burow *et al.* (1998) ont montré que 83% des puits dans lesquels les concentrations en nitrates étaient supérieures à la norme de potabilité présentaient au moins un pesticide en concentration décelable. Cependant, d'autres études ont démontré que si une corrélation pouvait être observée entre la présence de pesticides et des concentrations excessives en nitrate, l'inverse ne l'était pas forcément, c'est-à-dire que l'absence de pesticides n'est pas corrélée avec l'absence de nitrates. Ainsi, une présence de nitrates en quantité importante peut être observée aussi fréquemment dans des puits qui ne présentent pas de contamination décelable par les pesticides. Les nitrates à eux seuls ne permettent donc pas de prédire des contaminations par les pesticides.

Burow *et al.* (1998) ont établi une corrélation entre la présence de pesticides et les concentrations en sulfates. Les sulfates sont souvent appliqués comme acidifiants du sol. Une augmentation des teneurs en sulfates peut correspondre à une augmentation des effets de l'agriculture sur la qualité de l'eau souterraine. De même dans plusieurs études, les concentrations en chlorures qui entrent dans la composition des fertilisants sont corrélées à celles des nitrates (Navarro *et al.* 2004).

Ces rapprochements, dont découlent les objectifs de la thèse, sont le résultat de processus complexes et difficilement quantifiables qui s'apprêtent bien aux RNA. L'objectif général de la thèse est en conséquence d'étudier l'utilité de certains paramètres chimiques de l'eau (essentiellement des contaminants et autres variables chimiques reliés aux activités agricoles) dans la prédiction du potentiel d'occurrence de pesticides dans l'eau souterraine.

Le premier objectif sera donc de confirmer ou d'infirmer une relation possible entre la détection de pesticides et les concentrations de certains paramètres chimiques de l'eau. Cette étude sera menée à partir d'investigations *in situ* de la qualité de l'eau. Trois sites d'études présentant des caractéristiques différentes ont ainsi été échantillonnés en vue d'étudier la relation entre les concentrations en pesticides et certains paramètres chimiques de l'eau souterraine.

Le deuxième objectif sera d'évaluer si ces paramètres chimiques peuvent être utilisés afin de prédire ou d'améliorer la prédiction de l'occurrence des pesticides dans les puits ou forages domestiques par modélisation neuronale.

Chapitre 2 Étude de la relation entre l'occurrence des pesticides et les autres contaminants d'origine agricole

Ce chapitre a pour objectif l'étude de la relation entre la présence de pesticides et celle d'autres éléments d'origine agricole. Dans une première partie, les sites d'étude et la stratégie d'échantillonnage seront présentés. Nous verrons ensuite les résultats quantitatifs de l'identification des pesticides dosés et présents sur les trois sites d'étude avant d'aborder statistiquement leur corrélation avec la présence des différents autres éléments.

2.1 Présentation des sites d'étude

2.1.1 Bassin de Valence (France)

2.1.1.1 Situation géographique

Le bassin de Valence est situé dans le sud-est de la France au niveau de la moyenne vallée du Rhône et couvre environ 1600 km² (Figure 2-1). Le secteur d'étude est limité à l'Ouest par le Rhône, à l'Est par les reliefs du Vercors et au Nord par la vallée fluvio-glaciaire de Bièvre-Valloire. Le secteur présente deux territoires physiographiques aux reliefs différents séparés par l'Isère. Au Nord de l'Isère, la zone appelée Drôme des collines est constituée de collines molassiques et de plateaux en altitude. Au Sud de l'Isère, la Plaine de Valence présente un relief beaucoup moins marqué malgré la présence de quelques buttes molassiques au Sud.

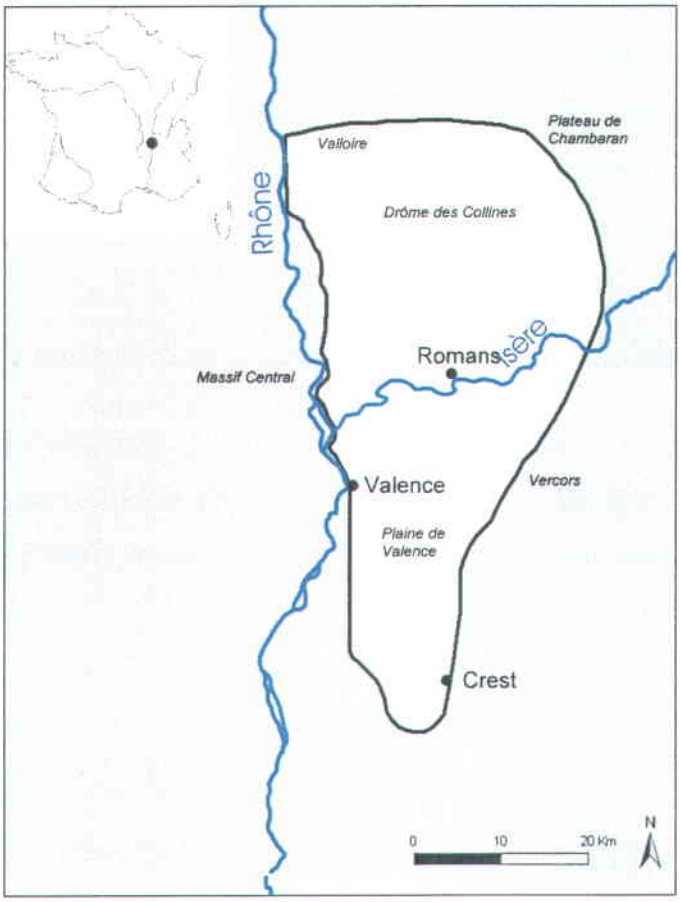


Figure 2-1. Localisation du bassin de Valence

2.1.1.2 Géologie (de la Vaissière. R. 2006)

Le bassin de Valence présente des dépôts d'âge miocène constitués principalement de sables molassiques, plus ou moins grésifiés et/ou argileux, et de cailloutis à leurs sommets, où circule une puissante nappe aquifère. Une bonne partie de ces dépôts est masquée par des recouvrements d'âges plus récents qui sont :

- d'une part, des argiles plus ou moins sableuses surmontées de cailloutis d'âge pliocène souvent masquées par les alluvions ; toutefois, quelques affleurements d'argiles subsistent notamment à l'Est de Chabeuil et sous forme de cailloutis au niveau des buttes de Montoison et de Montmeyran, mais aussi au niveau de l'interfluve Valloire-Galaure ;
- d'autre part, les alluvions et cailloutis d'âge quaternaire qui affleurent très largement (vallée fluvio-glaciaire de la Valloire, cailloutis d'Alixan et alluvions de l'ancienne Isère au niveau de la plaine de Valence).

Ce bassin molassique s'est, en fait, surimposé sur un ancien fossé d'effondrement d'âge oligocène dont les affleurements, composés principalement de calcaires, de marnes et de marnes sableuses, sont assez restreints, uniquement en bordure de la plaine de Valence :

- à l'Ouest au niveau d'Etoile-sur-Rhône ;
- au Sud de Grane à Crest ;
- de manière discontinue en bordure du Vercors, de Crest jusqu'à Pont-en-Royans.

Sous ces terrains, nous retrouvons aussi en affleurements les sables kaolinitiques rouge et blanc d'âge éocène, exploités dans des carrières en bordure et du Massif Central (sables de Douevas). Quatre coupes géologiques est-ouest du bassin (de la Vaissière 2006) permettent de mieux se rendre compte de la structure du bassin.

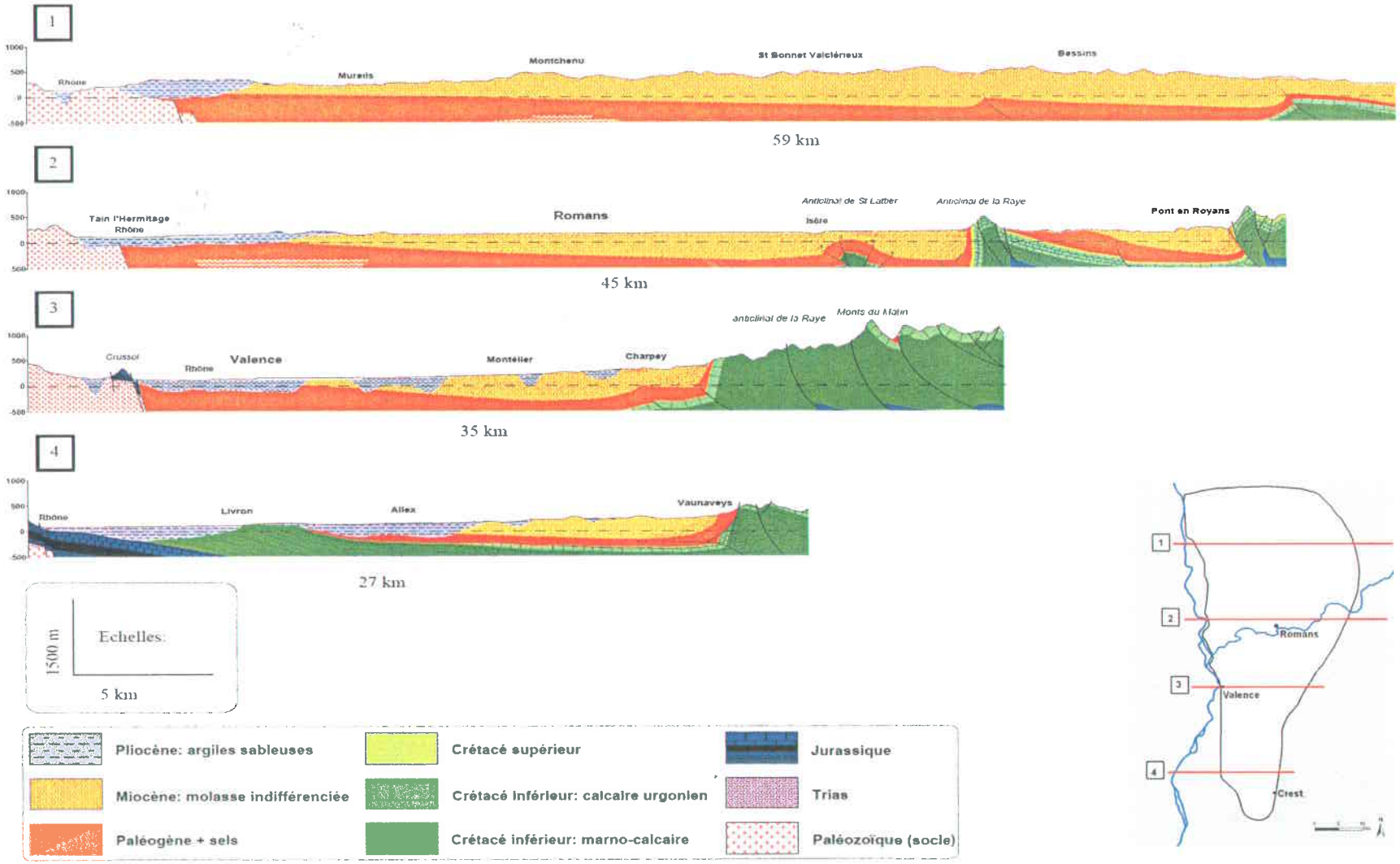


Figure 2-2. Coupes géologiques est-ouest du bassin molassique de Valence (de la Vaissière2006)

2.1.1.3 Principaux aquifères

La région d'étude présente deux structures aquifères principales :

- les aquifères alluvionnaires qui sont constitués des alluvions de l'Isère et de ses anciens cours, du Rhône et de la Drôme. Leur épaisseur varie de quelques mètres pour les alluvions les plus fines à 70 m pour les hautes terrasses des alluvions de l'Isère. Leur perméabilité est généralement bonne, ce qui les rend très intéressants pour l'exploitation.
- la molasse miocène, avec une épaisseur moyenne d'environ 400 m, qui représente la plus importante ressource régionale en eau.

En plus de ces deux structures aquifères, la région présente également des structures au potentiel aquifère plus faible, il s'agit des cailloutis d'Alixan et de ressources karstiques.

2.1.1.4 Activité culturale et types de sols

Le bassin de Valence est une région essentiellement agricole avec une Surface Agricole Utilisée de 56 %. Les cultures dominantes sont les céréales (45 %) avec principalement maïs, sorgho et blé puis les fourrages (21 %) et les vergers (17 %). Les élevages sont nombreux dans le bassin avec principalement l'aviculture qui représente 70 % des élevages.

Au nord de l'Isère, dans la Drôme des collines, les sols présentent une grande variabilité. La majorité des sols sont issus des collines à pente faible molassiques et marneuses. Ce sont des sols développés directement sur les marnes ou la molasse. Ils présentent une faible évolution et sont le plus souvent sablo-argileux. Au Sud, la plaine de Valence est dominée par deux types d'unités pédologiques. Il y a d'une part les sols des alluvions anciennes du Rhône et de ses affluents. Ce sont des sols caillouteux en surface. Et d'autre part les mêmes sols avec un recouvrement limoneux éolien récent.

2.1.1.5 Qualité de l'eau

Une dizaine d'ouvrages de la plaine de Valence font l'objet d'un suivi de qualité. Les résultats montrent des détections fréquentes d'un ou plusieurs pesticides dont la somme des concentrations

dépasse parfois la norme de 0.5 µg/l. En ce qui concerne les nitrates, des campagnes d'échantillonnage réalisées en 2003 (de la Vaissière 2006) sur le secteur d'étude montrent que 14 % des prélèvements dépassent la norme de potabilité. D'un point de vue géochimique, les aquifères superficiels présentent des eaux essentiellement bicarbonatées calciques tandis que l'aquifère profond est bicarbonaté calcique à calcique magnésien.

2.1.2 Bassins de Carpentras et Valréas (France)

2.1.2.1 Situation géographique

Le site d'étude est composé du regroupement de deux bassins dont les caractéristiques hydrogéologiques sont en continuité. Les bassins de Valréas et Carpentras se situent dans le sud-est de la France, majoritairement dans le département du Vaucluse, à proximité de la vallée du Rhône (Figure 2-3). Le secteur d'étude est limité au Nord, à l'Est et à l'Ouest par des massifs montagneux (Montagne de la Lance, massif du Tricastin, massif de Lafare-Suzette, plateau de Vaucluse) et au Sud par la Valée de la Durance. La limite entre les deux bassins se situe entre le massif d'Uchaux et celui de Lafare-Suzette. L'ensemble des deux bassins couvre une surface d'environ 1000 km².

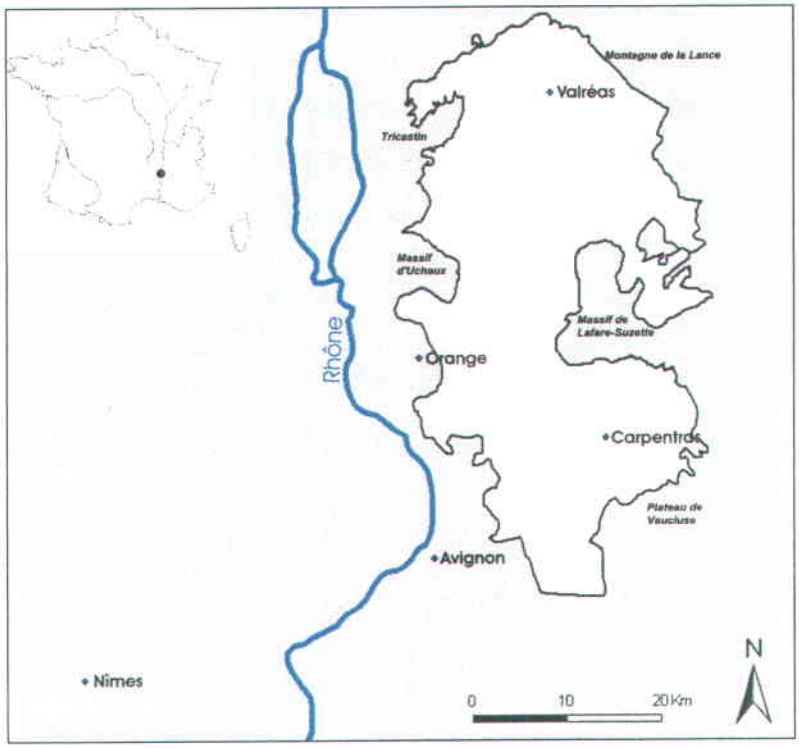


Figure 2-3. Localisation des bassins de Valréas et Carpentras

Le réseau hydrographique des bassins est relativement dense. Au nord, sur le bassin de Valréas, ce sont essentiellement le Lez et l'Aigues qui drainent le bassin. Du côté de Carpentras, ce sont la Sorgue, l'Ouvèse et l'Auzon qui parcourent le bassin. Tous ces cours d'eau se jettent dans le Rhône.

2.1.2.2 Principaux aquifères

Les bassins de Valréas et de Carpentras comprennent plusieurs structures aquifères qui sont détaillées dans (Huneau 2000) et (Lalbat 2006). Le substratum correspond à des calcaires du Crétacé inférieur sur lequel reposent des dépôts de l'Oligocène, du Miocène puis du quaternaire. Des coupes géologiques (Figure 2-4) permettent de se faire une meilleure idée de la structure géologique des bassins. Les dépôts du Miocène et les alluvions du quaternaire représentent les principales formations aquifères :

- L'aquifère miocène des bassins de Valréas et Carpentras d'âge helvétien est constitué d'un empilement de strates alternativement sablo-gréseuses (safres) et argilo-marneuses avec de rapides variations latérales de faciès. L'épaisseur de la formation aquifère atteint 300 à 400 m et localement 600 m, constituant ainsi le principal réservoir de la région. L'alimentation de la nappe se fait au niveau des affleurements du Miocène.
- L'aquifère alluvial est constitué d'éléments détritiques grossiers emballés dans une matrice argilo-limoneuse. Les nappes alluviales, dans l'ensemble libres et continues, sont quelques fois captives. Leur perméabilité et leur transmissivité sont généralement fortes.

2.1.2.3 Activité culturale et occupation du sol

Les bassins de Valréas et de Carpentras sont très agricoles. La surface agricole utile (SAU) représente environ 50 % de la surface. Les principales productions sont la viticulture, les cultures de céréale et le maraîchage.

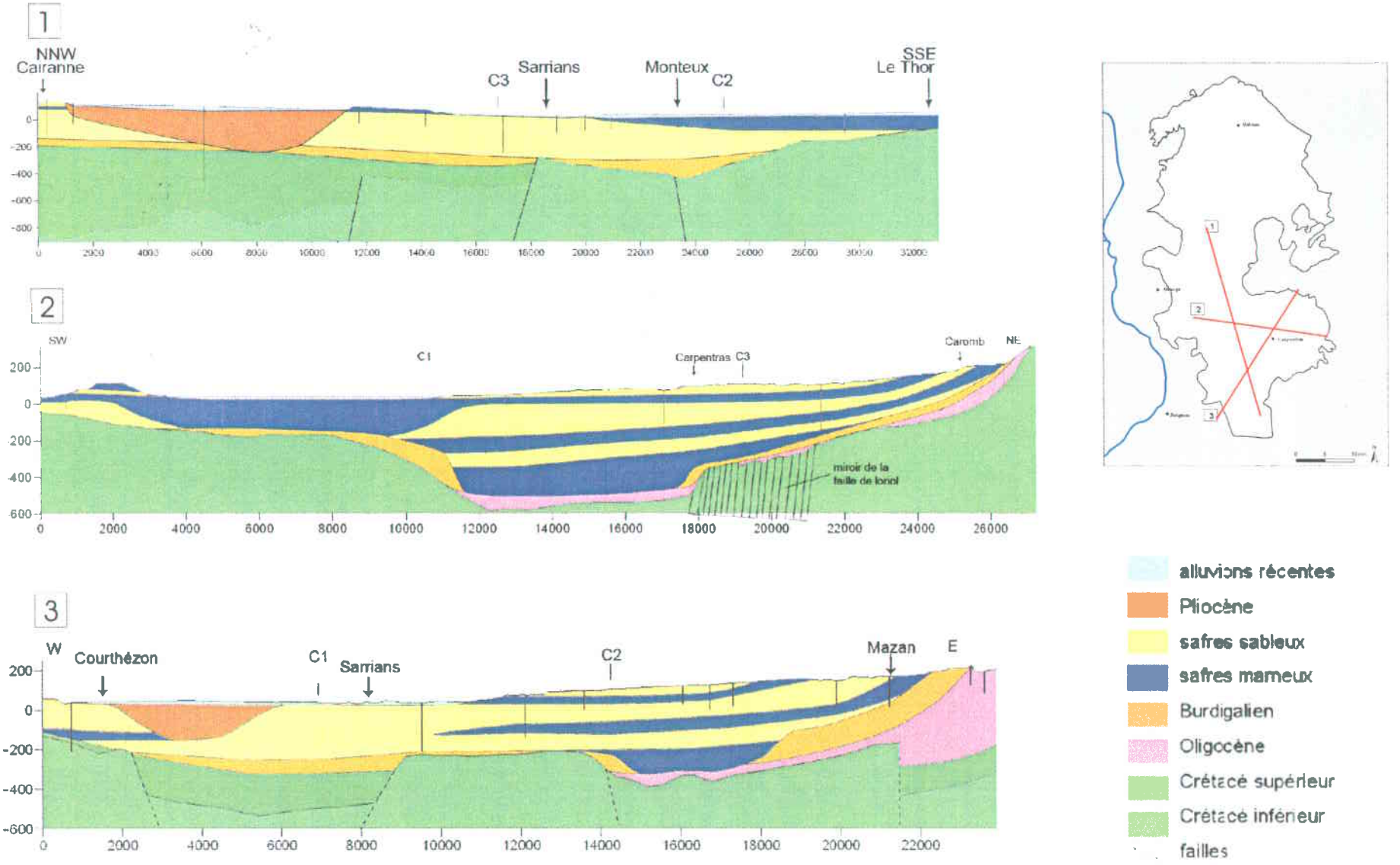


Figure 2-4. Coupes géologiques du bassin de Carpentras (Lalbat, 2006)

2.1.2.4 Qualité de l'eau

Quelques ouvrages des bassins de Valréas et Carpentras ont fait l'objet d'un suivi de qualité. Globalement, les teneurs en pesticides restent faibles ou non détectées, mais certains ouvrages présentent néanmoins des concentrations supérieures à la norme européenne. En ce qui concerne les nitrates, plusieurs études géochimiques ont permis de faire le point sur les teneurs couvrant une bonne partie des bassins. Il en ressort que les eaux souterraines semblent présenter des contaminations essentiellement au niveau des bordures là où a lieu l'alimentation de l'aquifère.

2.1.3 Comté de Portneuf (Québec)

2.1.3.1 Situation géographique

Le comté de Portneuf (Québec) se situe sur la rive nord du fleuve Saint-Laurent près de la ville de Québec (**Erreur ! Source du renvoi introuvable.**). Le comté occupe deux territoires physiographiques distincts, soient les Basses-Terres du Saint-Laurent et les Laurentides. Les Basses-Terres du Saint-Laurent, sur lesquelles se situe le secteur d'étude, prennent la forme d'une immense plaine fertile au relief peu marqué constituée d'une succession de terrasses qui bordent le fleuve. Le site est délimité en grande partie par le relief marqué des Laurentides et par le fleuve Saint-Laurent dans la partie Sud.

2.1.3.2 Géologie

La géologie du secteur est caractérisée par une couverture de sédiments marins et continentaux d'âge quaternaire recouvrant en discordance le socle rocheux. Ce dernier est constitué au Nord de roches ignées métamorphiques précambriennes à faible potentiel aquifère. Au Sud, ce sont des roches sédimentaires calcaires de l'Ordovicien, dont la majeure partie est recouverte par les argiles de l'ancienne mer de Champlain (Fagnan *et al.* 1999).

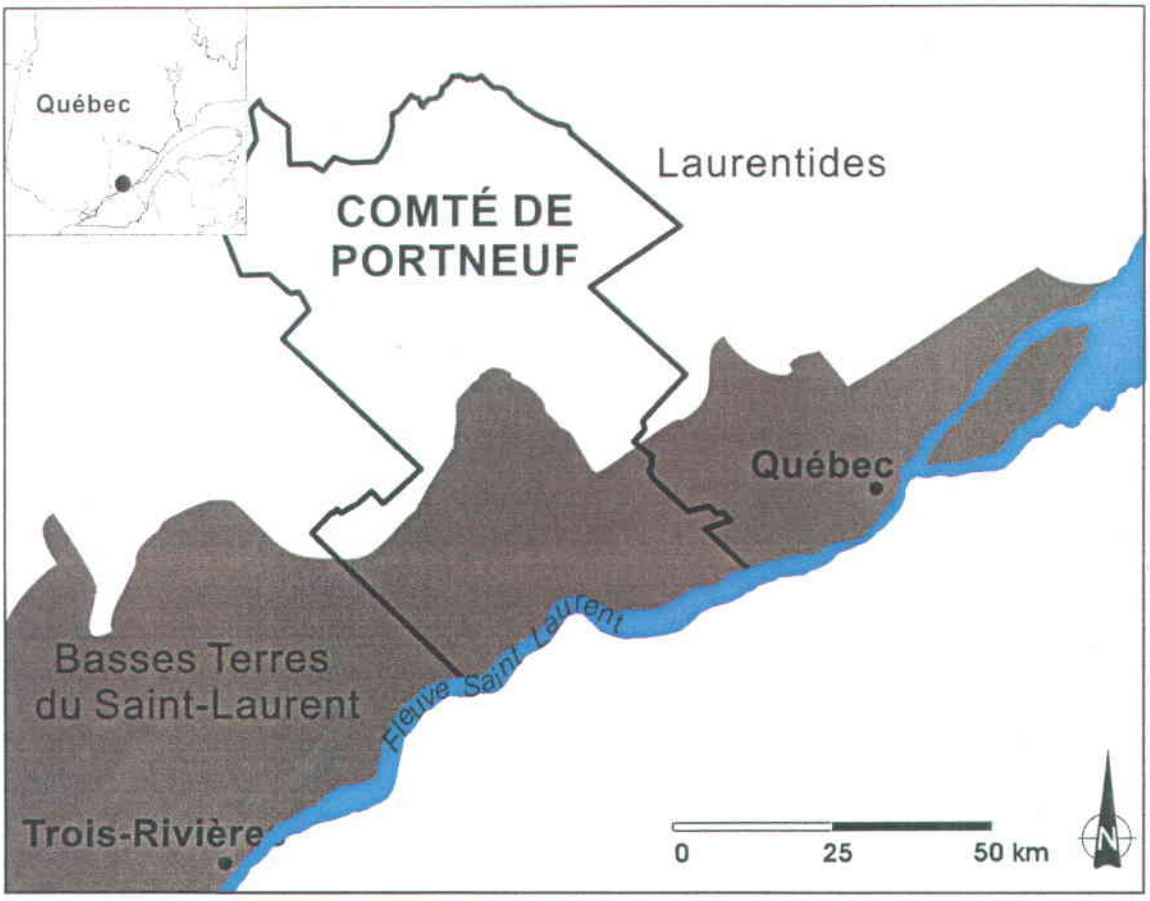


Figure 2-5. Localisation du Comté de Portneuf

La couverture de dépôt quaternaire est relativement épaisse (40 m). Les plus étendues des formations superficielles ont été mises en place lors de la déglaciation qui a suivi le dernier épisode glaciaire (Wisconsinien supérieur). Les principaux événements qui ont marqué la déglaciation sont l'incursion d'un vaste bras de mer (Mer de Champlain) et son retrait subséquent. L'épisode de la mer de Champlain a permis la formation de grands deltas sablo-graveleux et la sédimentation de séquences silteuses et argileuses parfois très épaisses au-dessus des dépôts glaciaires et du socle rocheux.

2.1.3.3 Principaux aquifères

D'après les travaux de Fagnan *et al.* (1999), cinq contextes hydrogéologiques ont été distingués (Figure 2-6) :

- Les sables et graviers deltaïques présentent une grande perméabilité qui en font les principaux aquifères de la région. Ils reposent soit directement sur le roc, soit sur les

argiles marines. Leur épaisseur varie de 5 m en bordure à une cinquantaine de mètres à l'intérieur.

- La Moraine de Sainte-Narcisse présente une grande variabilité granulométrique avec présence de zones sablo-graveleuses qui suggèrent un bon potentiel aquifère. Elle permettrait d'augmenter la recharge des formations rocheuses et granulaires enfouies.
- Les roches sédimentaires et ignées possèdent un faible potentiel aquifère.
- Les silts et argiles de la mer de Champlain : ils sont peu perméables et constituent une formation captive relativement continue dans la région. Ils peuvent affleurer ou être recouverts par les sables et graviers deltaïques. Cette unité forme généralement la base des aquifères deltaïques.

Deux coupes hydrostratigraphiques transversales des deltas des rivières Sainte-Anne et Jacques-Cartier (voir la Figure 2-6 pour la localisation des coupes) permettent de mieux comprendre le contexte hydrogéologique des aquifères de sables et graviers de surface (Figure 2-7). Le niveau de la surface libre (surface de la nappe phréatique) est présenté sur chacune des coupes. Le niveau de la nappe permet d'apprécier de façon générale les épaisseurs saturées des formations de sables et graviers de surface, la pente de la nappe ainsi que les profondeurs de la nappe dans ces secteurs.

La coupe de la rivière Sainte-Anne montre l'aquifère de sables et graviers de surface surmontant des sédiments fins (silts et argiles marins) de la Mer de Champlain qui constituent la base de la formation aquifère. La rivière Sainte-Anne s'écoule sur les dépôts fins de sorte que l'aquifère de sables et graviers n'est pas en contact hydraulique direct avec la rivière. Des sources suintent le long du contact géologique entre la formation aquifère et l'unité de sédiments fins. En s'éloignant de la rivière, l'épaisseur de l'aquifère diminue graduellement pour ne faire que quelques mètres (1 à 2 m) d'épaisseur en périphérie des affleurements rocheux. Ces secteurs constituent les limites de l'aquifère de sables et graviers de surface.

La coupe de la rivière Jacques-Cartier montre un contexte hydrogéologique quelque peu différent. À cet endroit, la rivière s'écoule sur la formation aquifère. En effet, les informations de forages indiquent que les sédiments fins (silts et argiles) de la Mer de Champlain ne sont pas

présents à cet endroit. Le socle rocheux qui est affleurant aux extrémités nord-est et sud-ouest de la coupe constitue la limite de la formation aquifère (Fagnan *et al.* 1999).

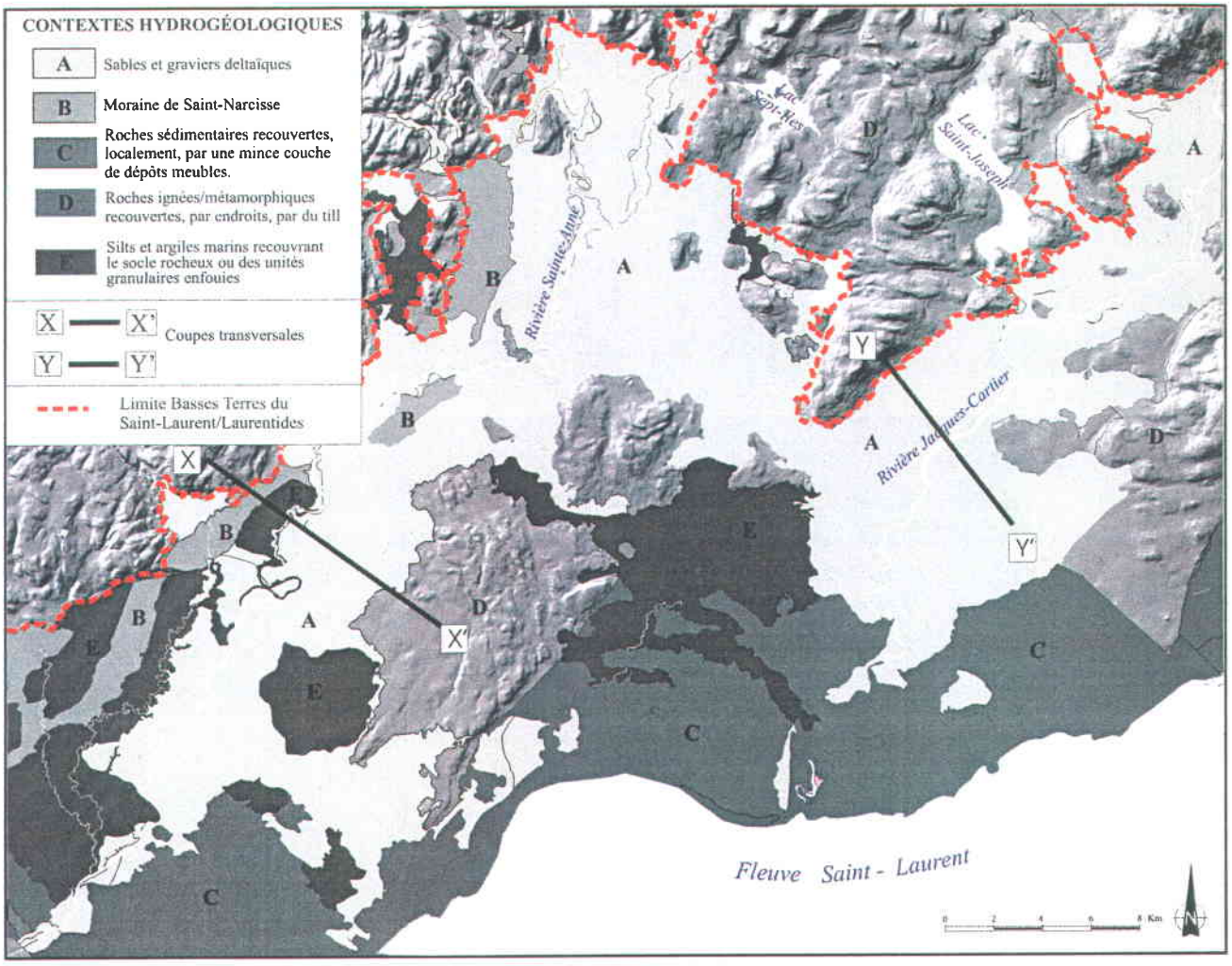


Figure 2-6. Contextes hydrogéologiques du comté de Portneuf (Fagnan et al. 1999, modifié)

2.1.3.4 Activité culturale et occupation du sol

Le comté de Portneuf constitue le territoire le plus agricole de la région de la ville de Québec. La majorité de l'agriculture est représentée par la culture de pomme de terre. On retrouve la culture de pomme de terre associée aux sols sableux en bordure du territoire. Le centre et le nord du territoire sont essentiellement dominés par la forêt avec des sols plus argileux. Étant donné le relief accidenté des Laurentides, l'activité agricole n'y est pas développée.

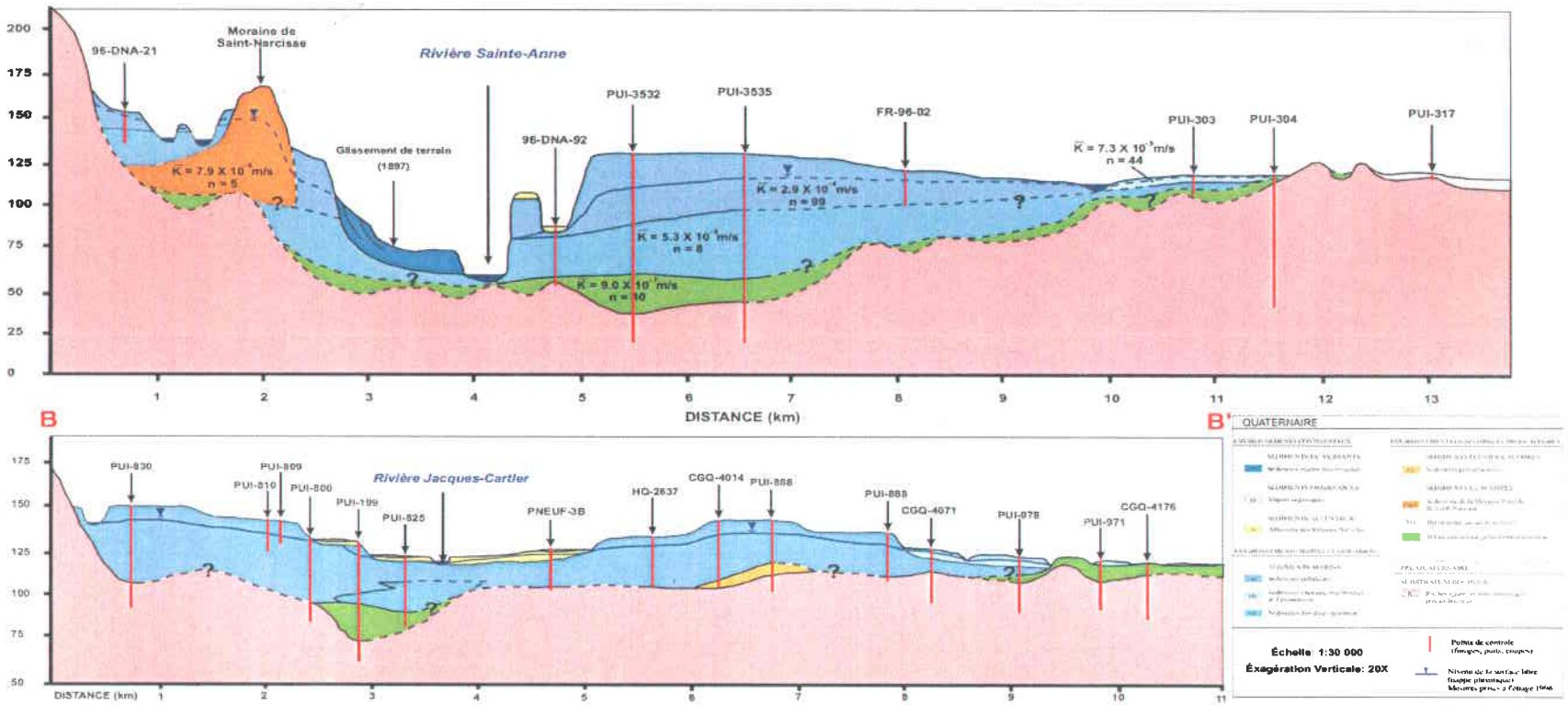


Figure 2-7. Coupes hydrostratigraphiques du Comté de Portneuf (Fagnan et al.)

2.1.3.5 Qualité de l'eau

Plusieurs analyses de pesticides et nitrates ont été effectuées dans le comté de Portneuf entre 1999 et 2001 dans les zones de culture intensive de la pomme de terre (Giroux 2003). Dans cette étude, près de 80 % des échantillons présentent des détections en pesticides. L'insecticide imidaclopride et l'herbicide métribuzine sont les composés les plus fréquemment détectés. En ce qui concerne les nitrates, 61 % des puits échantillonnés présentent des concentrations supérieures ou égales à la norme québécoise de 10 mg/l N-NO3 (45 mg/l NO3). D'un point de vue géochimique, l'eau des aquifères deltaïques est surtout bicarbonaté calcique et sulfaté calcique. Elle est en général peu minéralisé et son pH est bas (< 6.5).

2.2 Prélèvements et méthodologie analytique

2.2.1 Choix des pesticides

Le choix des pesticides à analyser a été réalisé afin de cibler les molécules potentiellement présentes parmi la grande variété de produits appliqués. Le choix des molécules s'est basé sur les critères suivants :

- les recommandations de traitements phytosanitaires des cultures dominantes ;
- les applications réelles à l'aide de données fournies par les producteurs locaux ;
- les cas de contaminations connues au cours de précédentes études ;
- les possibilités d'analyses en laboratoire.

Les pesticides ainsi analysés pour chaque site d'étude sont présentés dans le Tableau 2-1.

Tableau 2-1. Liste des pesticides analysés pour les trois sites

Site d'étude	Culture dominante	Pesticides analysés	Type	LQM[1] (µg/l)
Bassin de Valence	Céréales	Atrazine + DEA[2]	Herbicide	0.02
Bassin de Carpentras	Vigne	Simazine	Herbicide	0.02
		Diuron	Herbicide	0.02
		Terbuthylazine + DET[3]	Herbicide	0.02
		Oxadixyl	Fongicide	0.02
Comté de Portneuf	Pomme de terre	Atrazine + DEA	Herbicide	0.008 ; 0.02
		Métolachlore	Herbicide	0.03
		Métribuzine	Herbicide	0.02
		Linuron	Herbicide	0.02

[1] Limite de quantification de la méthode; [2] Désethyl- atrazine (sous-produit de dégradation de l'atrazine); [3] Désethyl-terbuthylazine (sous-produit de dégradation de la terbuthylazine)

Les pesticides analysés dans cette étude sont brièvement présentés ici, leurs caractéristiques physico-chimiques étant présentées dans le *Tableau 2-2* :

- L'atrazine est un herbicide de la famille des triazines. Elle est utilisée pour détruire les mauvaises herbes en pré-émergence et en post-émergence, particulièrement dans le maïs, mais également dans d'autres cultures telle la vigne. Dans les régions où elle est fortement utilisée, l'atrazine est l'un des pesticides les plus fréquemment rencontrés. Dans l'eau potable, la concentration maximale acceptable (CMA) au Canada est de

5 µg/l (Santé Canada 2002). En France, où l'atrazine est interdite d'utilisation depuis octobre 2003, la concentration est de 0.1 µg/l,. Le déséthyl-atrazine est le sous-produit de dégradation de l'atrazine le plus rencontré dans les eaux souterraines.

- Le métolachlore est un herbicide de la famille des chloro-acétanilides utilisé pour lutter contre les mauvaises herbes dans le maïs et le soja essentiellement. Il s'adsorbe facilement sur la matière organique du sol et se trouve donc rarement lessivé dans les sols où la teneur en matière organique est élevée. Dans l'eau potable, la concentration maximale acceptable (CMA) au Canada est de 50 µg/l et de 0.1 µg/l en France.
- La métribuzine est un herbicide de la famille des triazines employé en pré-levée et en post-levée pour lutter contre les mauvaises herbes qui parasitent diverses cultures agricoles. Dans l'eau potable, la concentration maximale acceptable (CMA) au Canada est de 80 µg/l et de 0.1 µg/l en France.
- Le linuron est un herbicide de pré-levée épandu sur le sol pour réfréner la germination des mauvaises herbes annuelles et vivaces : il est absorbé par le système racinaire. Aucune norme ne s'applique pour ce composé au Canada En France la concentration maximale réglementaire dans l'eau potable est de 0.1 µg/l.
- La simazine est une triazine utilisée comme agent de stérilisation du sol et comme herbicide de pré-levée contre les mauvaises herbes dicotylédones et graminées qui infestent un large éventail de cultures. La pénétration de la simazine à travers les couches du sol dépend du pH : elle est plus soluble aux faibles valeurs de pH, car elle se lie davantage à la matière organique et aux argiles du sol à des valeurs de pH plus élevées (Anderson 1986). Dans l'eau potable, la concentration maximale acceptable (CMA) au Canada est de 10 µg/l et de 0.1 µg/l en France. Son utilisation en zone agricole est interdite en France depuis le 1er octobre 2003.
- Le diuron est un herbicide de la famille des urées substituées. Stable à l'oxydation et à la dégradation, il persiste dans les sols pendant une saison complète ou davantage. Dans l'eau potable, la concentration maximale acceptable (CMA) au Canada est de 150 µg/l et de 0.1 µg/l en France.
- La terbuthylazine est un herbicide de la famille des triazines. Aucune recommandation ne s'applique pour ce composé au Canada. En France la norme est de 0.1 µg/l dans

l'eau potable et son utilisation est interdite depuis le 30 juin 2004 sur la vigne et depuis le 1er octobre 2003 sur les autres cultures. Le déséthyl-terbuthylazine est son principal sous-produit de dégradation.

- L'oxadixyl est un fongicide utilisé essentiellement dans les cultures de la vigne. Aucune recommandation ne s'applique pour ce composé au Canada. En France, la concentration maximale réglementaire est de 0.1 µg/l dans l'eau potable.

Tableau 2-2. Caractéristiques physico-chimiques des pesticides analysés (source FOOTPRINT, 2006)

Nom	Type de pesticide	Famille chimique	Formule brute	Masse molaire (g/mol)	Solubilité dans l'eau 20\ (mg/l)	DT50 Sol (j)	Koc (ml/g)
Atrazine	Herbicide	Triazine	$C_8H_{14}ClN_5$	215.7	35	75	100
Desethylatrazine	Métabolite	Triazine	$C_6H_{10}ClN_5$	187	3200	45	18
Diuron	Herbicide	Urée	$C_9H_{10}Cl_2N_2O$	233.1	35.6	75.5	1067
Linuron	Herbicide	Urée	$C_9H_{10}Cl_2N_2O_2$	249.1	63.8	48	620
Métolachlore	Herbicide	Chloroacetamide	$C_{15}H_{22}ClNO_2$	283.8	530	20	200
Métribuzine	Herbicide	Triazine	$C_8H_{14}N_4OS$	214.3	1165	11.5	37.9
Oxadixyl	Fongicide	Phenylamide	$C_{14}H_{18}N_2O_4$	278.3	3400	75	12
Simazine	Herbicide	Triazine	$C_7H_{12}ClN_5$	201.7	5	90	130
Terbuthylazine	Herbicide	Triazine	$C_9H_{16}ClN_5$	229.7	8.5	45	220

2.2.2 Prélèvements d'eau souterraine

Les points d'eau choisis pour les prélèvements sont en général des captages privés. Il peut s'agir de puits de surface, de forages profonds ou de pointes filtrantes. Le choix des ouvrages à échantillonner s'est basé sur plusieurs critères en fonction du site d'étude et de l'accessibilité des ouvrages.

Sur le site du bassin de Valence, 95 points d'eau souterraine ont été échantillonnés dans le cadre de cette étude et de celle de l'aquifère (de la Vaissière 2006). Ils sont repartis sur l'ensemble du site (Figure 2-8) dont 83 dans des forages profonds captant l'aquifère molassique et 12 dans des puits de surface captant l'eau dans des niveaux quaternaires ou molassiques. Ces prélèvements ont été effectués en juillet 2005. Sur les bassins de Valréas et Carpentras, 100 échantillons ont été prélevés essentiellement dans l'aquifère molassique profond. La répartition des prélèvements a été effectuée de façon à suivre des lignes d'écoulement (Figure 2-9). Ce choix avait pour objectif de pouvoir éventuellement lier les concentrations en pesticides avec la structure géologique du site. Les prélèvements ont été effectués en juillet 2005, dont 60 sur le bassin de Carpentras et 40

sur le bassin de Valréas. Dans le Comté de Portneuf, 53 échantillons ont été prélevés, essentiellement dans des puits de surface correspondant à l'aquifère des sables et graviers de surface. (Figure 2-10). Certains prélèvements supplémentaires ont été effectués dans quelques ouvrages atteignant le roc. Les échantillons ont été prélevés en octobre 2005.

Le prélèvement de l'eau des puits a été réalisé selon le même protocole pour les trois sites d'étude. Les prélèvements proviennent d'ouvrages sans filtre au robinet ou directement à la sortie du forage. L'eau du puit a été renouvelée afin de s'assurer de la représentativité de l'échantillon. De façon à tester cette représentativité, le contrôle des paramètres température, pH et conductivité électrique de l'eau a été effectué in situ au cours du renouvellement. À leur stabilisation, l'échantillon a été prélevé et considéré comme représentatif de l'eau souterraine. En règle générale, le temps de renouvellement était de 15 à 30 minutes en fonction du type d'ouvrage.

Pour le Comté de Portneuf, les échantillons ont été prélevés dans des bouteilles en polyéthylène haute densité (HDPE). Des études montrent en effet que l'adsorption des pesticides étudiés au Québec sur les bouteilles HDPE est négligeable (Topp et Smith 1992). Ces échantillons ont été placés dans une glacière à 4°C pour leur transport au laboratoire où ils ont été immédiatement congelés à – 18°C. Ceci assure une conservation appropriée des composés jusqu'au moment de l'analyse. Pour les bassins de Valence et Carpentras, les échantillons destinés aux analyses de pesticides ont été prélevés dans des bouteilles en verre ambré et extraits dans les 48 h suivant le prélèvement. Cette méthodologie a été effectuée afin d'être en règle avec les exigences françaises en matière de prélèvement et d'analyse.

La liste des flaconnages et des éventuels agents de conservation pour chacun des sites d'étude est présentée dans le Tableau 2-3.

Tableau 2-3. Flaconnage et conservation des échantillons

Site	Composés analysés	Contenant	Conservation
Porneuf	Pesticides	HDPE 1 l	Congélation -18°C
	Anions	HDPE 30 ml	4°C
	COD	Verre ambré 50 ml	4°C
Bassin de Valence et Bassin de Carpentras	Pesticides	Verre ambré 1 l	4°C pendant 48 h
	Anions	HDPE 30 ml	4°C
	COD	Verre ambré 30 ml	4°C - Chlorure mercurique

*HDPE: Polyéthylène haute densité
** COD : Carbone organique dissous

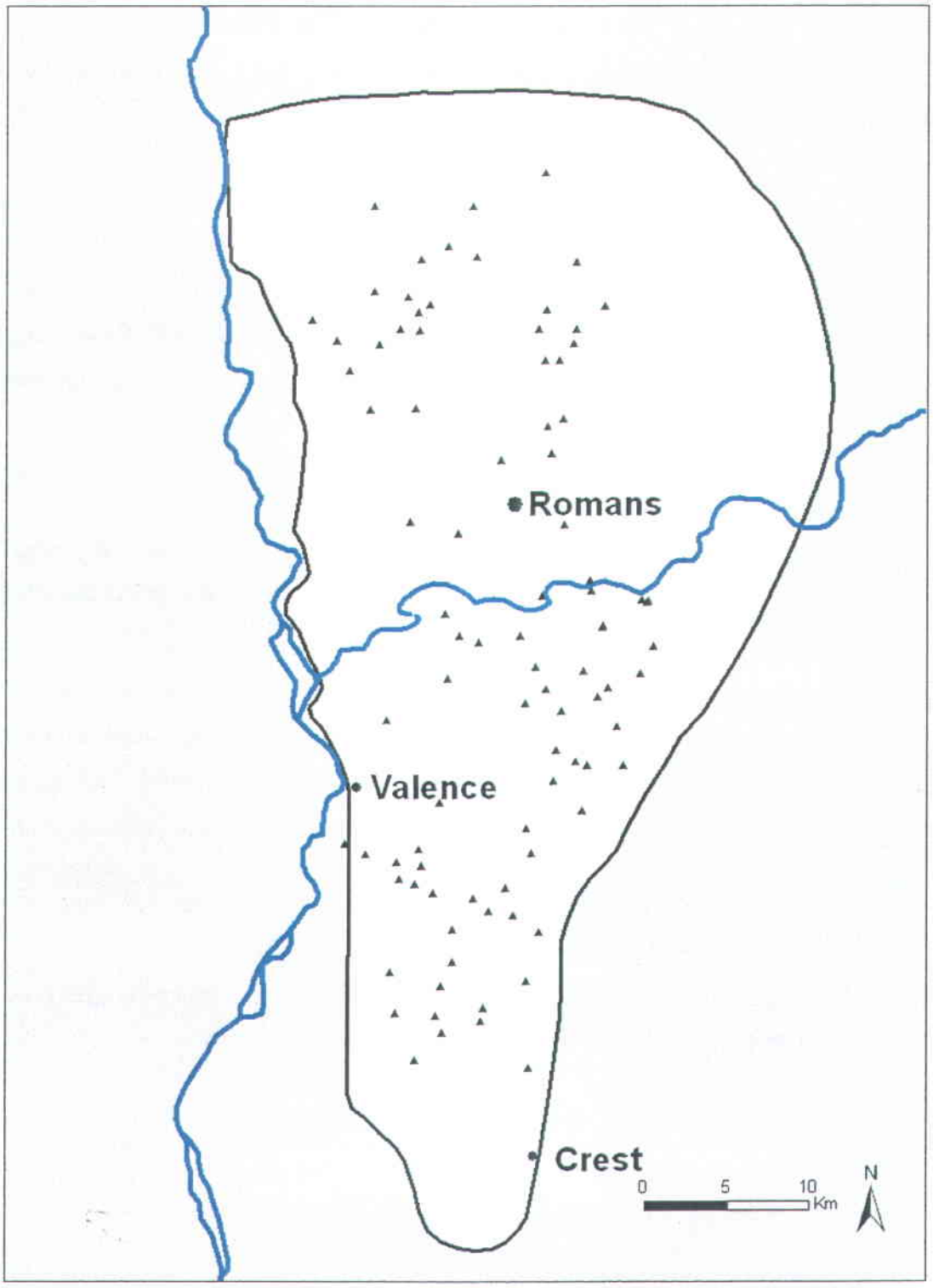


Figure 2-8. Localisation des points de prélèvement sur le bassin de Valence

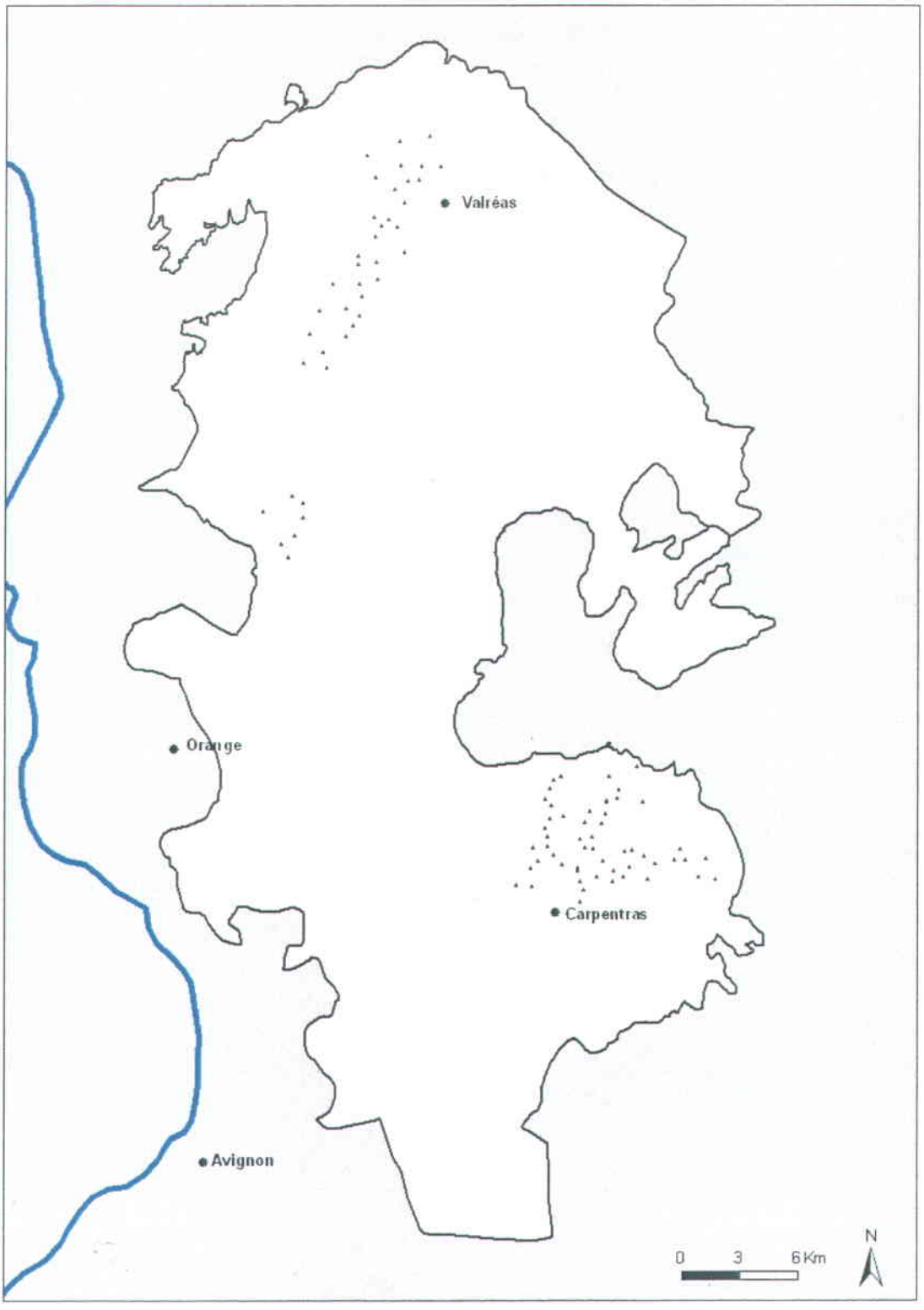


Figure 2-9. Localisation des points de prélèvement sur les bassins de Valréas et de Carpentras

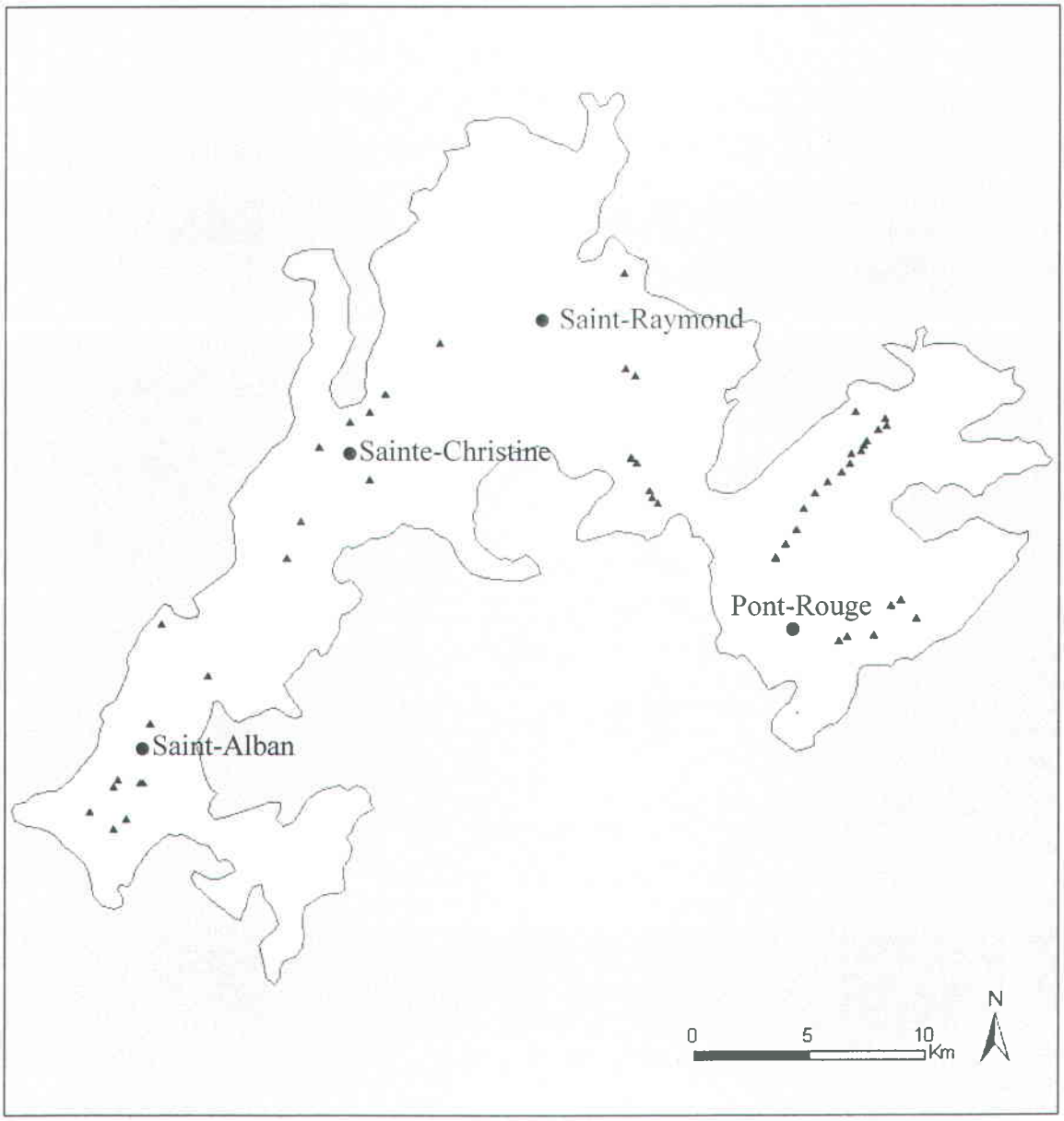


Figure 2-10. Localisation des points de prélèvement sur le bassin de Portneuf

2.2.3 Analyses

Toutes les extractions de pesticides ont été effectuées selon le même protocole. Tous les composés ont été extraits en phase solide en utilisant des cartouches octadécyl C18. Cette méthode a été démontrée comme étant une méthode efficace d'extraction des pesticides dans l'eau (Hennion et Pichon 1994) et permet de réduire la quantité de solvants par rapport à des techniques d'extraction liquide-liquide (Deger *et al.* 2000). Les volumes extraits sont de 1 l pour tous les sites concentrés dans 1 ml pour les bassins de Valence et Carpentras et dans 0.5 ml pour le comté de Portneuf (Tableau 2-4).

Les composés ont été quantifiés par chromatographie en phase gazeuse/spectrométrie de masse (GC/MS) pour le Comté de Portneuf et par chromatographie liquide haute performance (HPLC) pour les bassins de Valence et Carpentras. Les limites de quantification de la méthode (LQM) utilisées sont données pour chaque composé dans les figures de résultats (Figure 2-11, Figure 2-12 et Figure 2-13). Elles prennent en compte tous les facteurs de dilution et de concentration, mais ne tiennent pas compte du pourcentage de récupération de la solution étalon d'extraction.

Tableau 2-4. Méthodes analytiques et facteurs de concentration pour le dosage des pesticides

Site	Volume extrait (ml)	Volume concentré (ml)	Facteur de concentration	Méthode analytique
Bassin de Valence	1000	1	1000	HPLC
Bassin de Carpentras	1000	1	1000	HPLC
Comté de Portneuf	1000	0.5	2000	GC/MS

Un contrôle de qualité est effectué pour chaque douzaine d'échantillons. Ces contrôles permettent d'estimer la justesse et la répétitivité de la méthode. Ils comprennent :

- un blanc de méthode qui permet de s'assure de l'absence ou de la présence de contamination lors des manipulations ou de l'analyse ;
- un dupliqua qui permet d'estimer la précision de la méthode ;
- un échantillon de contrôle avec une concentration connue ;
- un échantillon fortifié qui permet d'établir si l'extraction est efficace.

2.3 Occurrence des pesticides

Cette section va permettre de faire le point sur les résultats quantitatifs des campagnes de pesticides sur les trois sites d'étude. Pour chaque site, nous présenterons donc quelles molécules ont été détectées et quel est l'ordre de grandeur des concentrations.

2.3.1 Bassin de Valence

Sur les 95 points d'eau souterraine prélevés dans le bassin de Valence, 56 échantillons (soit 59 %) présentent au moins un pesticide en concentration détectable (limite 0.02 µg/l). Bien qu'étant interdit à l'utilisation depuis 2003, les composés les plus détectés sont l'atrazine et son sous-produit de dégradation le déséthyl-atrazine (DEA) avec respectivement 40 et 43 cas de détection (Tableau 2-5). Pour la majorité des échantillons, les concentrations en atrazine et DEA sont inférieures à la norme française de 0.1 µg/l (Figure 1-1). Cependant, dans 11 et 10 cas respectifs pour ces deux composés, cette norme est dépassée avec des valeurs atteignant 1.11 µg/l pour l'atrazine et 0.43 µg/l pour le DEA.

Tableau 2-5. Nombre d'échantillons présentant des détections par composé analysé sur le bassin de Valence

Composés analysés	Nombre de détections	Nombre de détections > 0.1µ/l
Déséthylatrazine	43	10
Atrazine	40	11
Simazine	5	1
Diuron	5	2
Terbuthylazine	2	0
Déséthylterbuthylazine	2	0
Oxadixyl	1	0

La simazine et le diuron présentent chacun cinq cas de détection avec respectivement un et deux cas de concentrations supérieures à la norme. Les valeurs atteignent 0.38 µg/l pour le diuron. La terbuthylazine, son sous-produit de dégradation le déséthyl-terbuthylazine (DET) et l'oxadixyl sont peu détectés (Tableau 2-1) et ne présentent pas de concentrations supérieures à la norme.

La somme des pesticides dont la norme établit en France la concentration maximale à 0.5 µg/l varie de la limite de détection à 1.56 µg/l. Dans la limite des composés mesurés, cette norme est dépassée dans le cas de cinq échantillons.

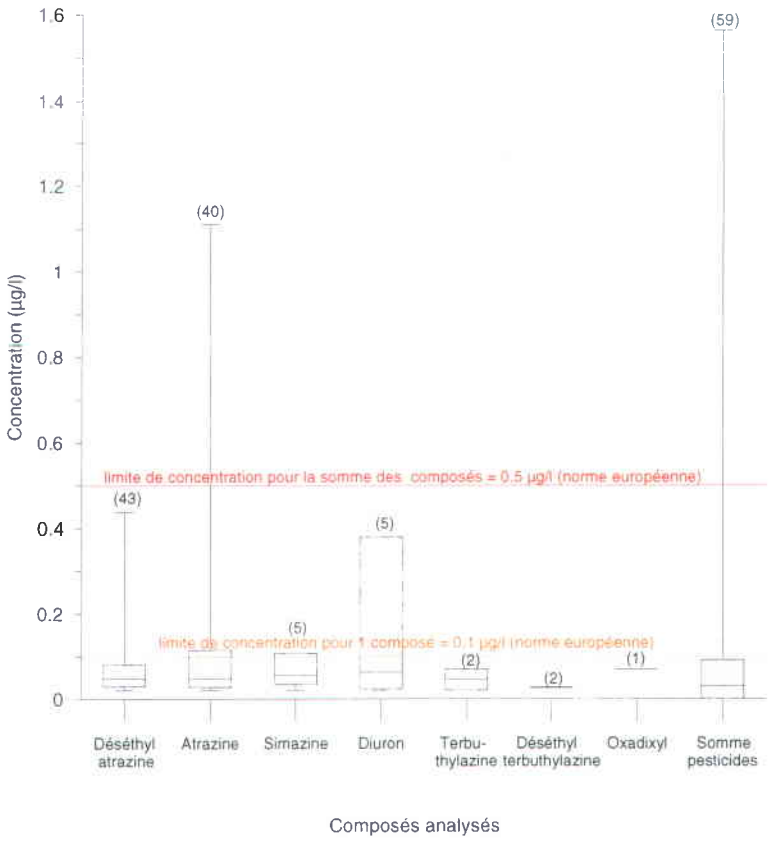


Figure 2-11. Médianes, quartiles et extrêmes des concentrations détectées par composé et pour la somme des composés pour le bassin de Valence. Les valeurs entre parenthèses indiquent le nombre de détections

2.3.2 Bassin de Carpentras-Valréas

Sur les bassins de Carpentras-Valréas, trois composés ont été détectés : la simazine, la terbuthylazine et le DET. Sur l'ensemble des 100 prélèvements répartis sur les deux bassins, 18 échantillons présentent au moins un de ces pesticides en concentration détectable.

Les composés les plus détectés sont la terbuthylazine mais surtout son sous produit de dégradation le déséthyl-terbuthylazine (DET). Ils présentent respectivement 8 et 16 cas de contamination (Tableau 2-6). Les concentrations en terbuthylazine (comprises entre 0.02 µg/l et

0.03 µg/l) sont proches de la limite de détection. Pour le DET, les concentrations varient entre la limite de détection et 0.3 µg/l (Figure 2-12). Dans 8 cas sur 16, ces concentrations sont supérieures à la norme de 0.01 µg/l. Le troisième composé détecté est la simazine avec trois cas de contamination. Les concentrations varient entre 0.038 et 0.056 µg/l.

La somme des concentrations de tous les pesticides analysés n'excède jamais la norme de 0.5 µg/l. Les valeurs varient entre la limite de détection et 0.3 µg/l.

Tableau 2-6. Nombre d'échantillons présentant des détections par composé analysé sur les bassins de Carpentras et de Valréas

Composés analysés	Nombre de détections	Nombre de détections > 0.1µ/l
Simazine	3	0
Terbuthylazine	8	0
Déséthylterbuthylazine	16	8

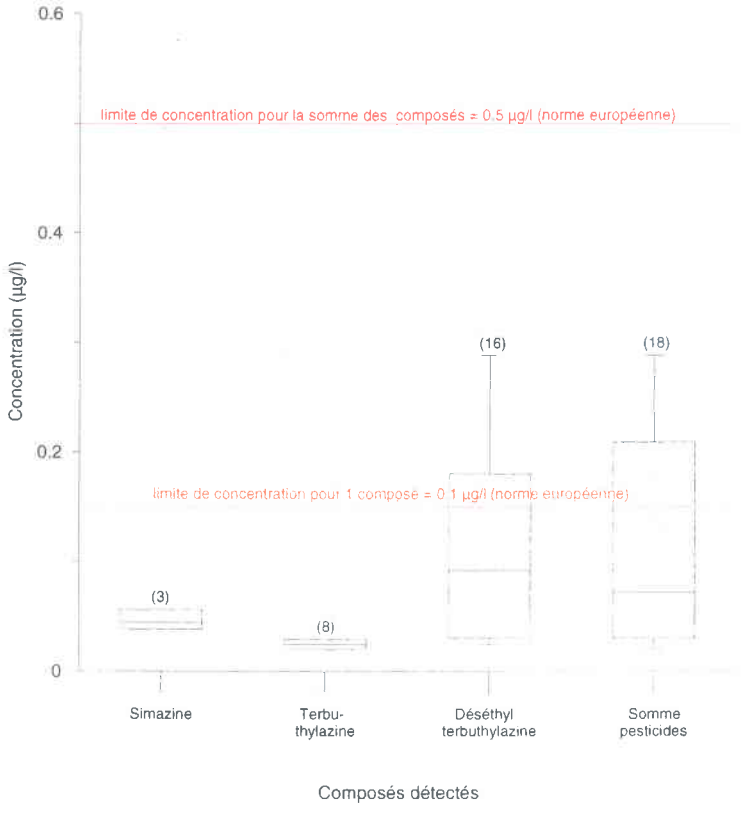


Figure 2-12. Médiane, quartiles et extrêmes des concentrations détectées par composé et pour la somme des composés pour les bassins de Carpentras et de Valréas. Les valeurs entre parenthèses indiquent le nombre de détections

2.3.3 Comté de Portneuf

Le contrôle de qualité effectué lors des analyses nous permet d'utiliser 50 échantillons parmi les 53 prélevés. Pour trois échantillons, il y a eu une mauvaise récupération du contrôle d'extraction. D'autre part, les données de métolachlore ne seront pas traitées, suite à la détection dans les blancs de méthode de concentrations du même ordre de grandeur que celles détectées dans les échantillons (environ 0.02 µg/l). Ceci signifie entre autres une contamination mineure lors des manipulations et également de très faibles concentrations en métolachlore dans ces échantillons. Parmi les 50 échantillons, 38 présentent ainsi des pesticides en concentration détectable et les quatre composés analysés ont été détectés (Tableau 2-7).

Tableau 2-7. Nombre d'échantillons présentant des détections par composé analysé sur le Comté de Portneuf. Les chiffres entre parenthèses correspondent aux dépassements de la norme européenne

Composés analysés	Nombre de détections	Nombre de détections > norme (>0.1µg/l)
Atrazine	21	0 (0)
DEA	25	0 (3)
Métribuzine	25	0 (10)
Linuron	28	nd

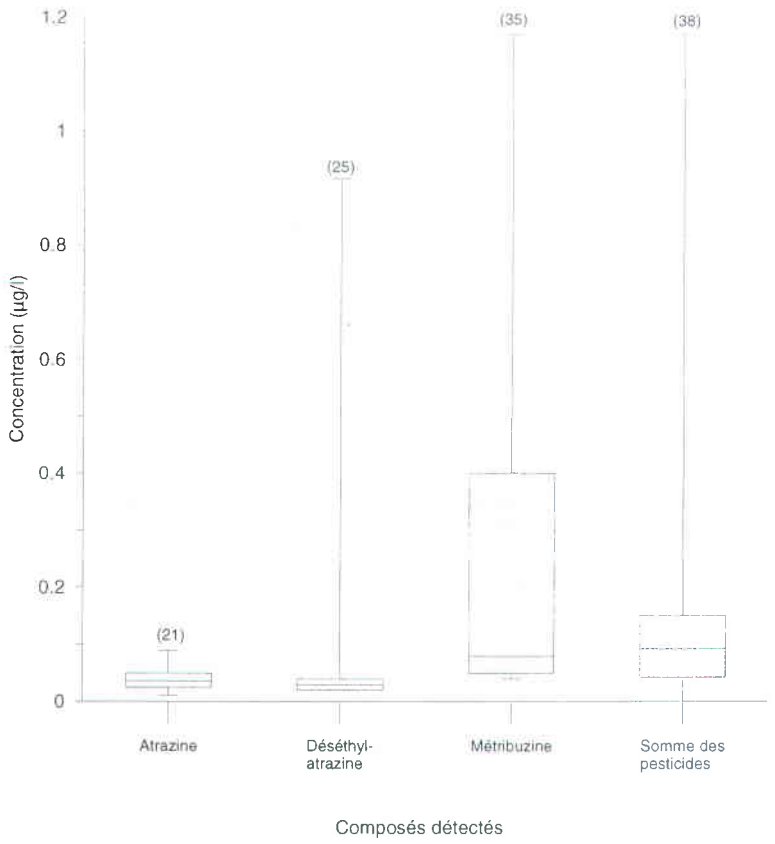


Figure 2-13. Médiane, quartiles et extrêmes des concentrations détectées par composé et pour la somme des composés pour le Comté de Portneuf. Les valeurs entre parenthèses indiquent le nombre de détections

Le composé le plus détecté est le linuron avec 28 cas de détection. Cependant et par suite d'un problème de dégradation du linuron dans des solutions standard lors des analyses, les concentrations obtenues sont surestimées et ne seront donc pas utilisées quantitativement ci-après. Seules la détection ou la non-détection peuvent être utilisées. La métribuzine a été détectée dans 25 échantillons avec des concentrations allant de 0.04 µg/l à 1.17 µg/l (Figure 2-13). La recommandation (CMA) canadienne étant à 80 µg/l n'est donc pas dépassée. L'atrazine et son sous-produit le DEA sont respectivement détectés dans 21 et 24 échantillons avec des concentrations allant de la limite de détection à 0.09 µg/l pour l'atrazine et 0.15 µg/l pour le DEA.

2.4 Relation entre pesticides et autres éléments d'origine agricole

L'occurrence des pesticides et des autres éléments d'origine agricole (notamment les nitrates, les chlorures et les sulfates) dans l'eau souterraine est mise en relation avec certaines caractéristiques physiques et chimiques du milieu. En effet, certains des processus qui contrôlent l'occurrence de ces différents éléments peuvent également contrôler l'occurrence des pesticides, et la co-occurrence de ces composés peut indiquer une vulnérabilité globale de l'aquifère à la contamination d'origine agricole. Par exemple, plusieurs études ont montré une corrélation entre la détection de pesticides et la présence de nitrates en concentrations excessives (Kolpin *et al.* 1998) ; Hamilton et Helsel 1995). Istok et Rautman (1996) ont utilisé cette corrélation afin de mieux prédire l'occurrence des herbicides dans l'eau souterraine. L'objectif de cette partie sera donc d'évaluer s'il y a effectivement une relation entre les concentrations en pesticides et celles des autres contaminants agricoles afin de voir si ces derniers, dont l'analyse est plus facile à mettre en œuvre, peuvent aider à la prédiction des risques de contamination par les pesticides. On appellera contaminant agricole un composé dont la concentration dans l'eau souterraine semble être supérieure à ce qu'elle serait en l'absence d'activité agricole même s'il ne provient pas directement de substance appliquée en surface (Bohlke 2002).

2.4.1 Origine des nitrates, chlorures et sulfates dans les trois bassins

2.4.1.1 Cas du bassin de Valence

La géochimie du bassin de Valence a été étudiée dans le cadre de la thèse de Rémi de la Vaissière (2005). Les nitrates étant un bon indicateur de pollution anthropique des eaux souterraines, leurs teneurs ont été comparées avec celles en chlorures et en sulfates afin d'étudier l'origine de ces derniers.

Deux forages présentent de fortes concentrations en chlorures sans présence de nitrates. Ces forages se situant tous deux à proximité de sites d'extraction de saumure, leurs concentrations élevées en chlorures sont attribuées à des pollutions ponctuelles (de la Vaissière 2006). Sans ces deux forages, des corrélations significatives (R Spearman = 0.75 ; $p = 1.10^{-17}$) entre les

concentrations en nitrates et en chlorures (Figure 2-14) laissent envisager une origine principalement agricole des fortes teneurs en chlorures.

Les sulfates sont également significativement corrélés aux nitrates (Figure 2-15 R Spearman = 0.80 ; $p = 8.10^{-22}$) même si plusieurs ouvrages comportent de fortes teneurs en sulfates par rapport aux nitrates. L'origine des sulfates n'est donc pas clairement identifiable. Bien que la corrélation significative avec les nitrates laisse envisager une source principalement anthropique, la présence de gypses est connue sur le bassin (de la Vaissière 2006) et peut être à l'origine de certaines valeurs élevées de sulfates.

2.4.1.2 Cas des Bassins de Carpentras-Valréas

Sur les basins de Carpentras-Valréas, les corrélations entre les nitrates et les chlorures et sulfates sont significatives, mais plus faibles que sur le bassin de Valence (R Spearman = 0.58 ; $p = 2.10^{-10}$, R Spearman = 0.38 ; $p = 1.10^{-4}$ respectivement pour les chlorures et les sulfates). Ces valeurs plus faibles laissent penser que les chlorures et les sulfates ont d'autres origines que les contaminations anthropiques.

Le bassin de Carpentras présente parfois de très fortes valeurs en sulfates. Deux sources sont envisagées. Tout d'abord les évaporites du Trias et de l'Oligocène puisque, en effet, l'influence de la dissolution du gypse sur l'hydrochimie des eaux du bassin a été mise en évidence dans Lalbat (2006). Cependant, la corrélation entre certaines fortes concentrations en sulfates et en nitrates justifie également d'une contamination anthropique agricole (Figure 2-16). La part de chacune de ces sources est donc difficile à estimer, mais il semblerait que la provenance des deux justifie de telles concentrations en sulfates (Lalbat 2006).

En ce qui concerne les chlorures, une origine anthropique peut également être envisagée du fait de la corrélation avec les nitrates (Figure 2-17). Cependant, la présence de certaines fortes valeurs de chlorures associées à Na^+ suppose une influence évaporitique bien qu'aucune trace de halite n'ait été signalée à l'affleurement (Lalbat 2006).

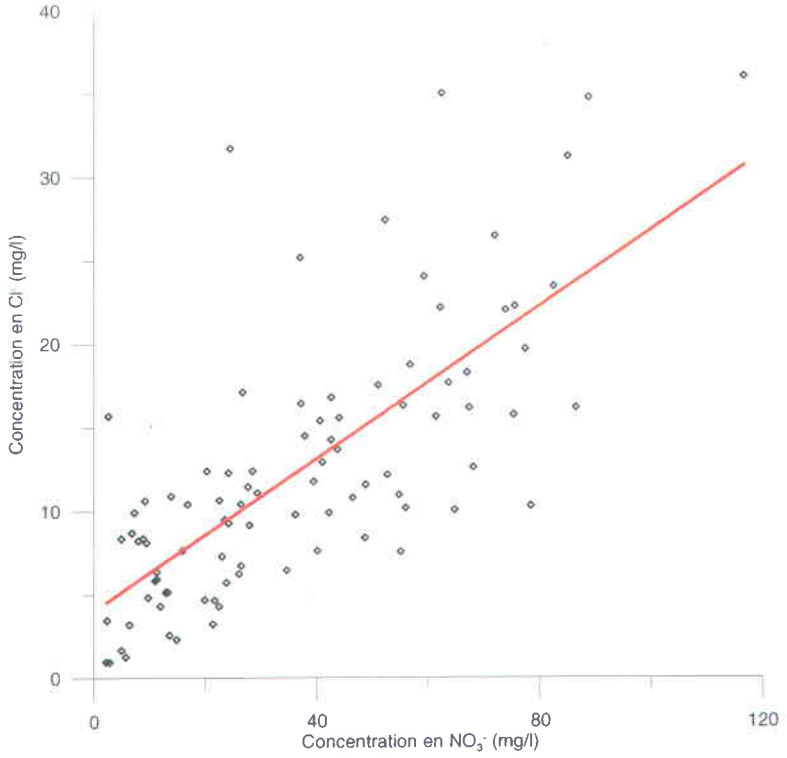


Figure 2-14. Concentrations en Cl- en fonction de celles en NO_3^- dans le bassin de Valence

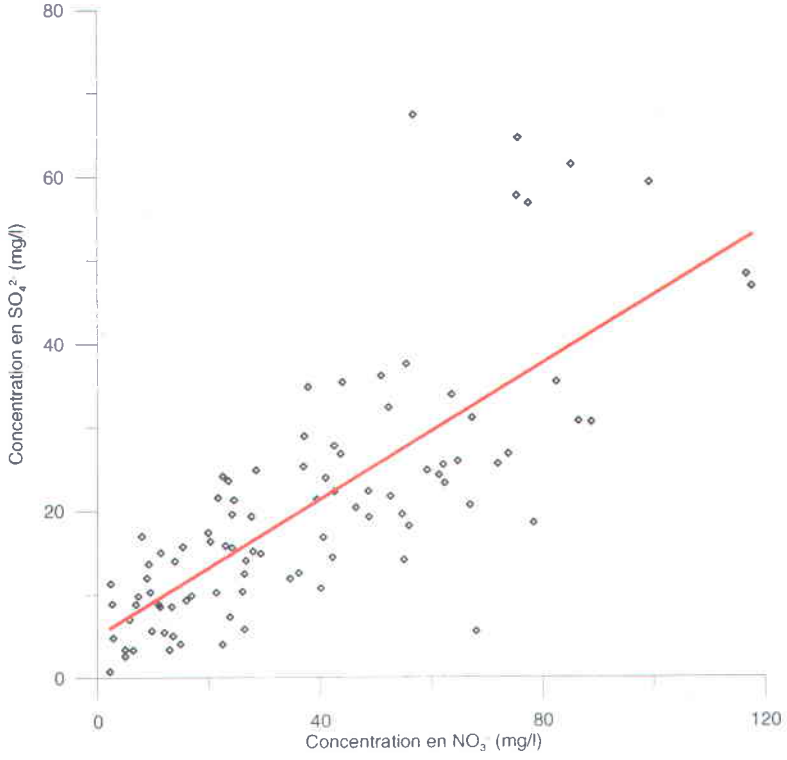


Figure 2-15. Concentrations en SO_4^{2-} en fonction de celles en NO_3^- dans le bassin de Valence

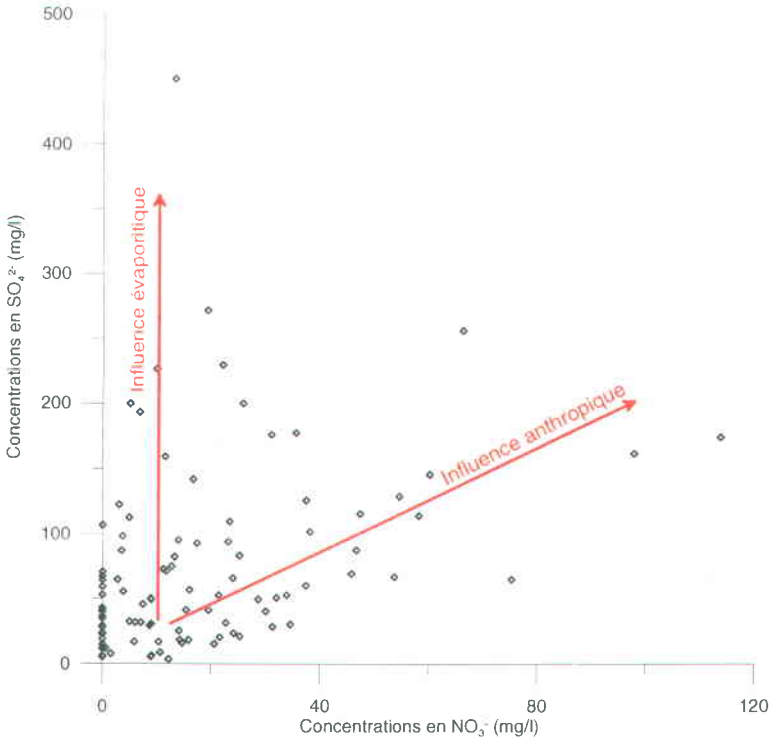


Figure 2-16. Concentrations en SO_4^{2-} en fonction de celles en NO_3^- dans les bassins de Carpentras-Valréas

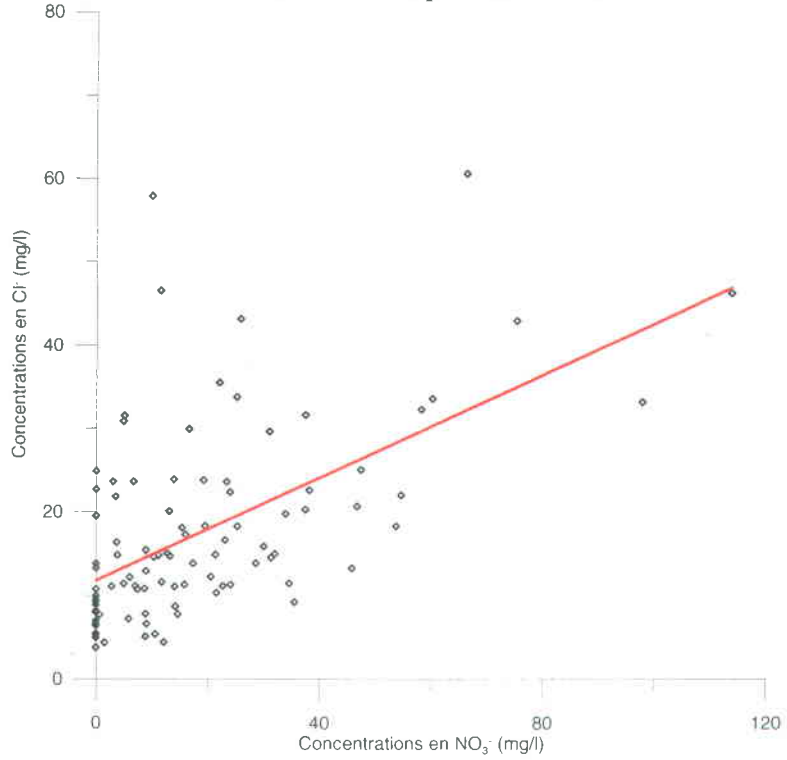


Figure 2-17. Concentrations en Cl^- en fonction de celles en NO_3^- dans les bassins de Carpentras-Valréas

2.4.1.3 *Cas du Comté de Portneuf*

Dans le comté de Portneuf, les sulfates sont corrélés positivement aux nitrates (Figure 2-18 R Spearman = 0.62 ; p = 2.10^{-6}) ce qui laisse envisager une source principalement due aux apports anthropiques. Le sulfate de cuivre est fréquemment utilisé pour la lutte contre le mildiou en culture de pomme de terre prédominante sur le bassin.

Pour les chlorures, on observe deux tendances (Figure 2-19 R Spearman = 0.45 ; p = 8.10^{-4}). Il y a en effet une corrélation entre les concentrations en chlorures et celles en nitrates. Cependant pour certains points, on observe de fortes teneurs en chlorures avec peu ou pas de nitrates, ce qui laisse envisager une autre source. Celle-ci peut être naturelle, mais également issue d'une pollution qui n'est pas agricole. En effet, au Québec, de nombreux aquifères superficiels présentent des contaminations aux chlorures dues aux sels de déglaçage des voiries.

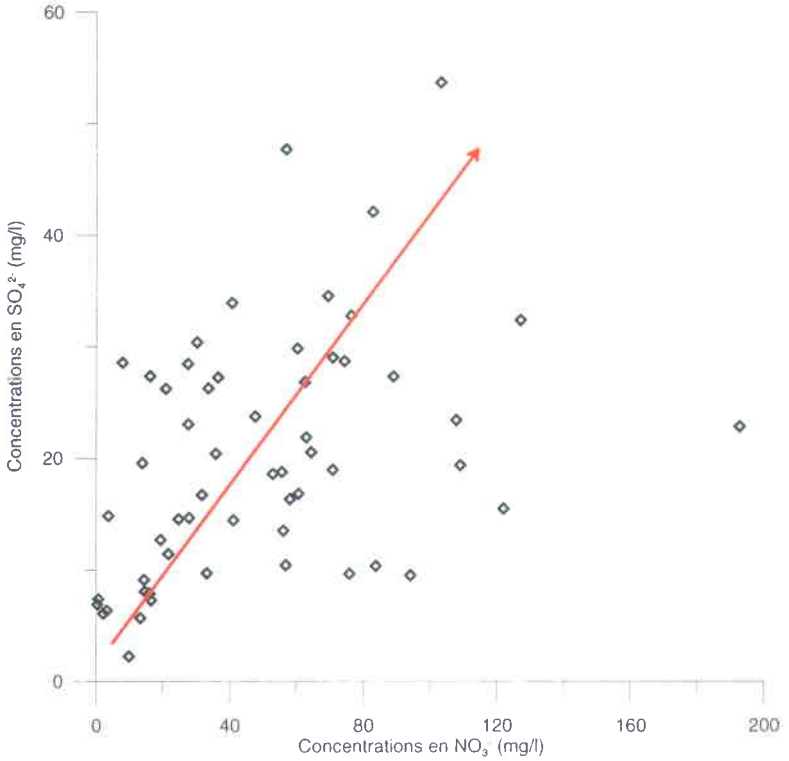


Figure 2-18. Concentrations en SO_4^{2-} en fonction de celles en NO_3^- dans le Comté de Portneuf

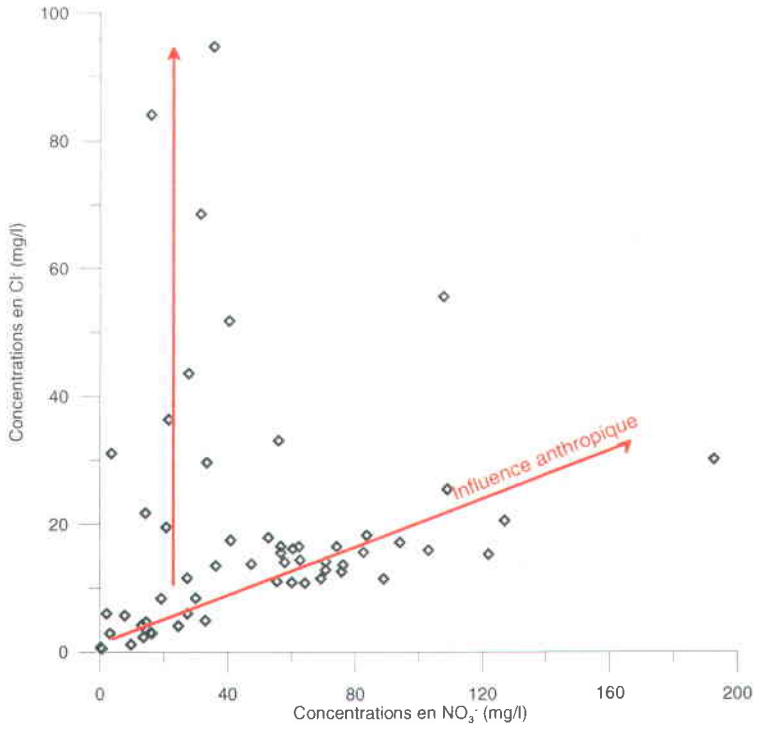


Figure 2-19. Concentrations en Cl^- en fonction de celles en NO_3^- dans le Comté de Portneuf

2.4.2 Relation entre les pesticides et les nitrates

Pour les trois sites d'étude l'ordre de grandeur des concentrations en nitrates est représenté sur la Figure 2-20. Les médianes sont de l'ordre de 45 mg/l pour le bassin de Valence, de 15 mg/l pour les bassins de Carpentras-Valréas et de 40 mg/l pour le Comté de Portneuf. Les trois sites présentent cependant des échantillons avec des concentrations excessives en nitrates (>100 mg/l). La norme européenne pour l'eau de consommation étant fixée à 50 mg/l et à 10 mg/l N-NO3 (soit 45 mg NO3 /l) au Québec.

La présence de corrélation entre les concentrations en nitrates et les concentrations des pesticides présentant des détections pour chaque site a été étudiée (Tableau 2-8). Pour chacun des sites, bien que relativement faibles, les coefficients de corrélation sont significatifs (seuil 0.05) entre les nitrates et la somme des concentrations en pesticides. Excepté pour le Comté de Portneuf, où les nitrates ne semblent corrélés qu'avec la métribuzine, ils sont corrélés avec tous les pesticides présentant suffisamment (>5) d'observations sur les deux autres bassins. Dans le cas de Portneuf, les concentrations en atrazine et en DEA sont proches de la limite de détection et varient peu, ce qui justifie l'absence de corrélation significative.

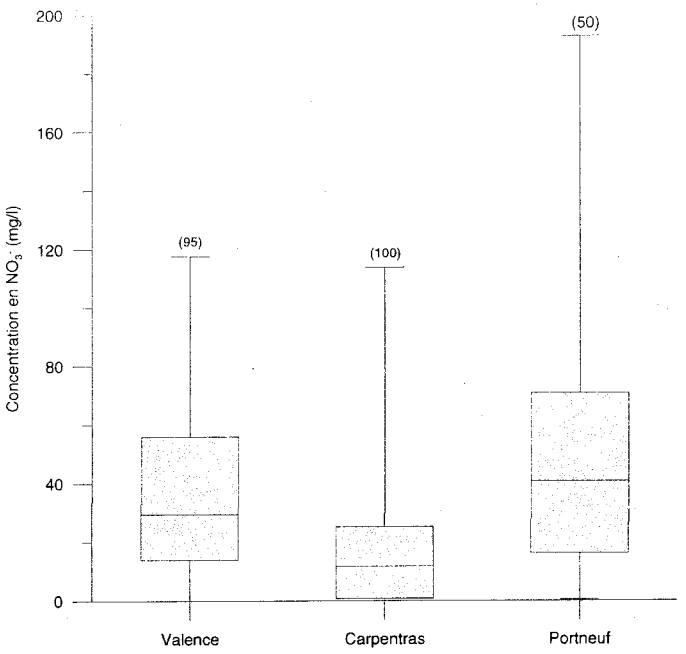


Figure 2-20. Médiane, quartiles et extrêmes des concentrations en nitrates mesurées sur les trois sites d'étude. Les valeurs entre parenthèses indiquent le nombre de détections

L'observation graphique des concentrations de pesticides en fonction des concentrations en nitrates semble montrer que ces dernières sont variables pour les échantillons sans détection de pesticides, mais sont élevées pour les échantillons présentant de fortes teneurs en pesticides. Ceci pourrait expliquer les faibles valeurs des coefficients de corrélation puisque les nitrates sont présents également lorsqu'il n'y a pas de détection de pesticide.

Tableau 2-8. Coefficient de corrélation de Spearman entre les concentrations en nitrates et les concentrations mesurées des pesticides sur chacun des trois sites d'étude. L'astérisque signifie que la corrélation est significative au seuil 0.05.

	Valence	Carpentras	Portneuf
Atrazine	0.36*	0	0.20
Déséthylatrazine	0.42*	0	0.22
Terbuthylazine	-0.05	0.30*	-
Déséthylterbuthylazine	0.03	0.31*	-
Simazine	0.21	0.10	-
Diuron	0	0	-
Métribuzine	-	-	0.42*
Somme des pesticides	0.43*	0.31*	0.55*

Afin de corroborer la présence d'une différence entre les échantillons présentant des détections en pesticides et ceux n'en présentant pas, le test de Mann Whitney a été appliqué afin de comparer les échantillons. Pour les trois sites, les concentrations en nitrates sont statistiquement différentes (seuil 0.05) pour les échantillons avec et sans pesticides (Tableau 2-9). Cette observation est en accord avec des études précédentes (Burrow 1996) où de fortes concentrations en nitrates étaient corrélées avec le nombre de composés détectés.

Tableau 2-9. Test de Mann Whitney pour les trois sites. Comparaison des concentrations en nitrates pour les échantillons avec et sans détection de pesticides

	Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type	Significativité
Valence	Détection	59	5.0	117.6	44.9	26.8	Différence significative au seuil 0.05 (p=0.0002)
	Non détection	36	0.0	86.5	26.0	24.3	
Carpentras	Détection	18	5.1	58.2	26.0	15.8	Différence significative au seuil 0.05 (p=0.002)
	Non détection	82	0.0	113.8	16.0	21.8	
Portneuf	Détection	38	0.3	192.6	57.1	39.8	Différence significative au seuil 0.05 (p=0.004)
	Non détection	12	0.0	75.7	25.1	26.2	

2.4.3 Relation entre les pesticides et les chlorures

Pour les trois sites d'études l'ordre de grandeur des concentrations en chlorures est représenté sur la Figure 2-21. Les médianes sont de l'ordre de 10 mg/l pour le bassin de Valence, de 15 mg/l pour les bassins de Carpentras-Valréas et le Comté de Portneuf. Les concentrations maximales mesurées sont de 83 mg/l pour le bassin de Valence, 60 mg/l pour les bassins de Carpentras-Valréas et 55 mg/l pour le Comté de Portneuf. Ces valeurs, bien qu'élevées, restent largement en dessous de la norme européenne de 250 mg/l pour l'eau de consommation. Au Canada, les chlorures ne sont soumis à aucune recommandation, cependant l'objectif organoleptique est de 250 mg/l.

Les corrélations entre les concentrations en chlorures et les concentrations en pesticides sont plus faibles que dans le cas des nitrates, mais restent significatives (Tableau 2-10). On observe ainsi que pour les trois sites, les chlorures sont positivement corrélés avec la somme des pesticides.

Le test de Mann Whitney (Tableau 2-11) effectué afin de comparer les concentrations en chlorures dans les échantillons avec et sans détection de pesticides montre que les concentrations

en chlorures sont statistiquement plus faibles dans les échantillons ne présentant pas de détection de pesticide.

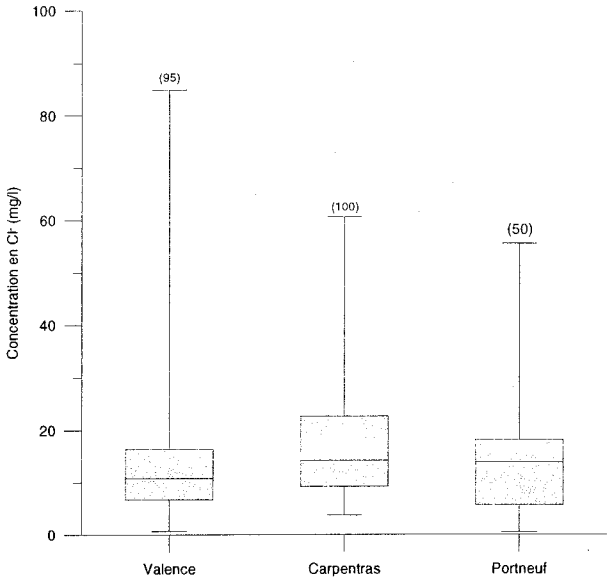


Figure 2-21. Médiane, quartiles et extrêmes des concentrations en chlorures mesurées sur les trois sites d'étude. Les valeurs entre parenthèses indiquent le nombre de détections

Tableau 2-10. Coefficient de corrélation de Spearman entre les concentrations en chlorures et les concentrations mesurées des pesticides sur chacun des trois sites d'étude

	Valence	Carpentras	Portneuf
Atrazine	0.27*	0	0.20
Déséthylatrazine	0.30*	0	0.33*
Terbuthylazine	-0.05	0.21*	-
Déséthylterbuthylazine	0.05	0.27*	-
Simazine	0.21	0.05	-
Diuron	-0.03	0	-
Métribuzine	-	-	0.30*
Somme des pesticides	0.30*	0.29*	0.40*

Tableau 2-11. Test de Mann Whitney pour les trois sites. Comparaisons des concentrations en chlorures pour les échantillons avec et sans détection de pesticides

	Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type	Significativité
Valence	Détection	59	1.7	84.9	15.9	14.0	Différence significative au seuil 0.05 (p=0.012)
	Non détection	36	0.7	63.0	11.4	11.6	
Carpentras	Détection	18	9.3	46.5	22.8	10.2	Différence significative au seuil 0.05 (p=0.004)
	Non détection	82	3.8	60.6	16.1	11.5	
Portneuf	Détection	38	0.7	55.5	17.9	12.8	Différence significative au seuil 0.05 (p=0.004)
	Non détection	12	0.5	31.0	7.9	8.3	

2.4.4 Relation entre les pesticides et les sulfates

L'ordre de grandeur des concentrations en sulfates pour les trois sites est représenté sur la Figure 2-22. Les médianes sont de l'ordre de 20 mg/l pour le bassin de Valence et le Comté de Portneuf et de 50 mg/l pour les bassins de Carpentras-Valréas. On remarque également les grandes variations sur ce dernier site avec des concentrations allant jusqu'à 925 mg/l visiblement influencées par la présence de gypse.

Les corrélations entre les concentrations en sulfates et celle des différents pesticides (Tableau 2-12) sont du même ordre de grandeur que pour les nitrates pour le bassin de Valence et le Comté de Portneuf. Cependant, pour les bassins de Carpentras-Valréas, la corrélation avec la somme des pesticides, bien que significative, reste tout de même assez faible ce qui s'explique par la présence d'une source de sulfates autre que la contamination anthropique.

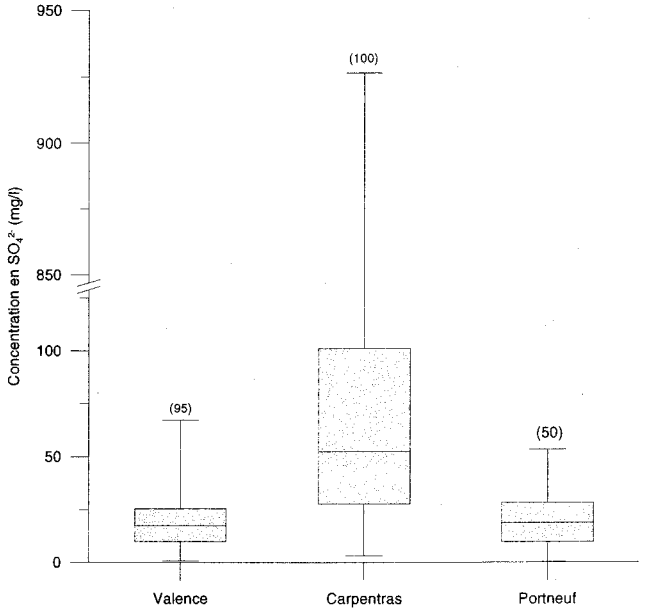


Figure 2-22. Médiane, quartiles et extrêmes des concentrations en sulfates mesurées sur les trois sites d'étude. Les valeurs entre parenthèses indiquent le nombre de détections

Tableau 2-12. Coefficient de corrélation de Spearman entre les concentrations en sulfates et les concentrations mesurées des pesticides sur chacun des trois sites d'étude

	Valence	Carpentras	Portneuf
Atrazine	0.32*	0	0.22
Déséthylatrazine	0.38*	0	0.11
Terbuthylazine	0	0.15	-
Déséthylterbuthylazine	0.05	0.27*	-
Simazine	0.21	-0.02	-
Diuron	-0.06	0.00	-
Métribuzine	-	-	0.52*
Somme des pesticides	0.38*	0.26*	0.56*

Néanmoins, même pour les bassins de Carpentras-Valréas, le test de Mann Whitney (Tableau 2-13) confirme que les concentrations en sulfates des échantillons avec et sans détection de pesticides sont significativement différentes au seuil 5 % ce qui pourrait confirmer l'importance d'une source anthropique.

Tableau 2-13. Test de Mann Whitney pour les trois sites. Comparaisons des concentrations en sulfates pour les échantillons avec et sans détection de pesticides

	Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type	Significativité
Valence	Détection	59	3.3	67.3	24.3	15.8	Différence significative au seuil 0.05 (p=0.0005)
	Non détection	36	0.7	36.1	14.2	9.7	
Carpentras	Détection	18	30.3	200.1	97.6	55.9	Différence significative au seuil 0.05 (p=0.011)
	Non détection	82	3.2	926.5	79.5	119.9	
Portneuf	Détection	38	2.2	53.6	22.2	11.7	Différence significative au seuil 0.05 (p=0.007)
	Non détection	12	0.3	28.5	12.2	8.1	

2.5 Relation entre les pesticides et le carbone organique dissous

Dans le sol, la matière organique naturelle se présente sous deux formes : une fraction solide (retenue ou adsorbée aux particules de sol) et une fraction dissoute. Les pesticides peuvent interagir avec l'une ou l'autre de ces deux fractions (Spark et Swift 2002) ; (Ben-Hur *et al.* 2003). Il a été montré en laboratoire que certains pesticides hydrophobes sont capables d'une complexation avec la matière organique dissoute (Devitt et Wiesner 1998) ; (Fitch et Du 1996) et que celle-ci peut jouer un rôle important dans la distribution et la migration de ces composés dans l'environnement (Lafrance *et al.* 1990;Stites et Kraft 2001).

D'autre part, Dunnivant *et al.* (1992) (cités dans (Nelson *et al.* 1998) ont montré que le carbone organique dissous (COD) était capable de se transporter rapidement à travers une colonne de sol sans flux préférentiel. La formation de complexes non adsorbables peut donc affecter grandement la mobilité des contaminants dans le sol (Lafrance *et al.* 1990). Dans cette optique, bien que le COD ne soit pas un contaminant agricole, sa relation avec les pesticides est étudiée.

L'ordre de grandeur des concentrations en COD est représenté sur la Figure 2-23. Les médianes sont comprises entre 1 et 1.5 mg/l pour les trois sites. Les maximums sont de 3.2 mg/l pour le bassin de Valence, 3.3 mg/l pour les bassins de Carpentras-Valréas et 3.9 mg/l pour le comté de Portneuf.

Les corrélations observées entre les concentrations en COD et celles des différents pesticides sur les trois sites ne sont significatives que dans le cas de la somme des pesticides pour le bassin de Valence (Tableau 2-14). Les autres coefficients de corrélation sont très faibles (< 0.2).

Le test de Mann Whitney (Tableau 2-15) montre que les moyennes en COD sont toujours plus faibles pour les échantillons sans détection de pesticides néanmoins cette différence n'est significative que pour le bassin de Valence.

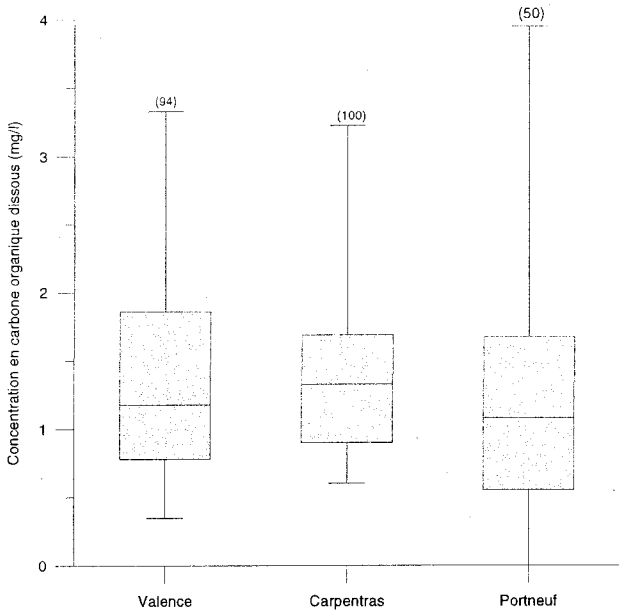


Figure 2-23. Médiane, quartiles et extrêmes des concentrations en carbone organique dissous mesurées sur les trois sites d'étude. Les valeurs entre parenthèses indiquent le nombre de détections

Tableau 2-14. Coefficient de corrélation de Spearman entre les concentrations en carbone organique dissous et les concentrations mesurées des principaux pesticides sur chacun des trois sites d'étude

	Valence	Carpentras	Portneuf
Atrazine	0.117*	nd	0.196
Déséthylatrazine	0.180	nd	0.069
Terbuthylazine	0.180	-0.013	nd
Déséthylterbuthylazine	nd	0.125	nd
Métribuzine	nd	nd	0.050
Somme des pesticides	0.274	0.098	0.126

Tableau 2-15. Test de Mann Whitney pour les trois sites. Comparaisons des concentrations en carbone organique dissous pour les échantillons avec et sans détection de pesticides

	Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type	Significativité
Valence	Détection	58	0.4	3.3	1.5	0.7	Différence significative au seuil 0.05 (p=0.004)
	Non détection	36	0.3	2.6	1.1	0.5	
Carpentras	Détection	18	0.6	2.7	1.5	0.6	Différence non significative au seuil 0.05 (p=0.342)
	Non détection	82	0.6	3.2	1.3	0.5	
Portneuf	Détection	38	0.0	3.2	1.2	0.8	Différence non significative au seuil 0.05 (p=0.328)
	Non détection	12	0.1	3.9	1.1	1.1	

2.6 Pouvoir discriminant des variables nitrates, chlorures et sulfates

Les paragraphes précédents ont montré la présence d'une relation entre les concentrations des différents contaminants agricoles et la détection des pesticides. Les corrélations sont significatives, mais restent cependant assez faibles. Les tests de Mann Whitney ont mis en évidence une différence de concentrations entre les échantillons présentant des détections en pesticides et ceux n'en présentant pas. Afin d'étudier si ces variables présentent réellement un pouvoir discriminant sur les détections de pesticides, la méthode des courbes ROC (Receiver Operating Characteristics) a été appliquée.

2.6.1 Principes des courbes ROC

Les courbes ROC permettent d'évaluer la performance de classifieurs dans des problèmes de classification à deux classes. Elles sont couramment utilisées dans le domaine de l'intelligence artificielle ou en statistiques médicales où elles permettent d'évaluer la pertinence d'une variable dans le calcul de diagnostic. Elles permettent en outre de déterminer quelle valeur de cette variable permet de mieux discriminer la cible. De nombreux articles décrivent l'utilisation des courbes ROC dans le domaine statistique (Bradley 1997) ; (Lasko *et al.* 2005). Dans le cas de tests binaires, on choisit une valeur du classifieur afin de séparer la population en deux, l'une où l'événement est présent, et l'autre où il est absent. La performance de ces tests est en général déterminée par les mesures de spécificité (Sp) et de sensibilité (Se) (Figure 2-23).

Les courbes ROC permettent de visualiser sur un seul graphique les tests de sensibilité et de spécificité pour toute la gamme de valeur de la variable en faisant varier le seuil sur celle-ci.

La performance du test est mesurée par l'aire sous la courbe (AUC). Une valeur de 1 représente une discrimination parfaite tandis qu'une aire de 0.5 (ligne diagonale) correspond à une discrimination aléatoire. On considère habituellement que le modèle est bon dès lors que la valeur de l'AUC est supérieure à 0.7. Un modèle ayant une AUC supérieure à 0.9 est excellent. D'autre part, le point pour lequel la somme de la sensibilité et de la spécificité est la plus grande correspond à la valeur qui permet la meilleure discrimination.

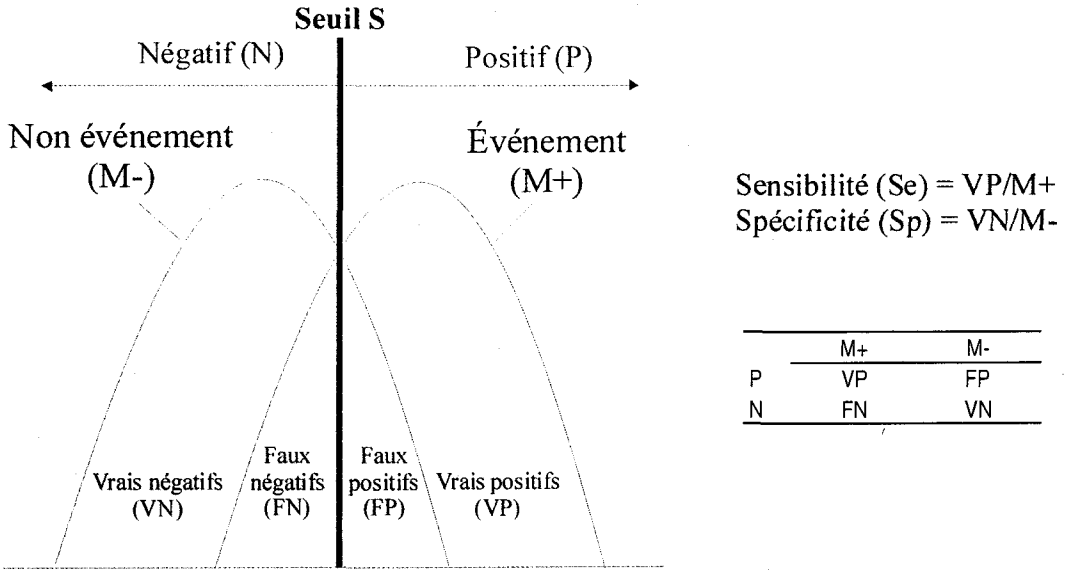


Figure 2-24. Détermination de la sensibilité et de la spécificité

2.6.2 Application des courbes ROC

Pour chacune des variables, les courbes ROC ont été calculées pour les trois sites (Figure 2-25, Figure 2-26, Figure 2-27). Les événements positifs et négatifs correspondent respectivement à la détection et à la non détection de pesticides dans les échantillons.

Pour les nitrates, les valeurs des aires sous la courbe pour les trois sites sont comprises entre 0.72 et 0.78 (Tableau 2-16) ce qui signifie qu'ils présentent une discrimination raisonnable.

Tableau 2-16. Aire sous la courbe et valeur optimale de nitrates obtenues à partir des courbes ROC

	Aire sous la courbe	Valeur la plus discriminative (mg/l)
Valence	0.72	24.5
Carpentras	0.73	33.0
Portneuf	0.78	23.4

Les valeurs de nitrates qui représentent le meilleur seuil de discrimination sont du même ordre de grandeur pour les trois sites, autour de 30 mg/l. Ces valeurs, relativement élevées, illustrent bien le fait que certains points présentent de fortes teneurs en nitrates sans pour autant avoir des pesticides.

Pour les chlorures, la discrimination sur les sites de Carpentras-Valréas et de Portneuf est raisonnable tandis qu'elle est faible sur le bassin de Valence (Tableau 2-17). Nous avons vu au paragraphe 2.4.1 que ce bassin présentait des contaminations ponctuelles en chlorures dues à une ancienne usine, ce qui pourrait expliquer la faible discrimination.

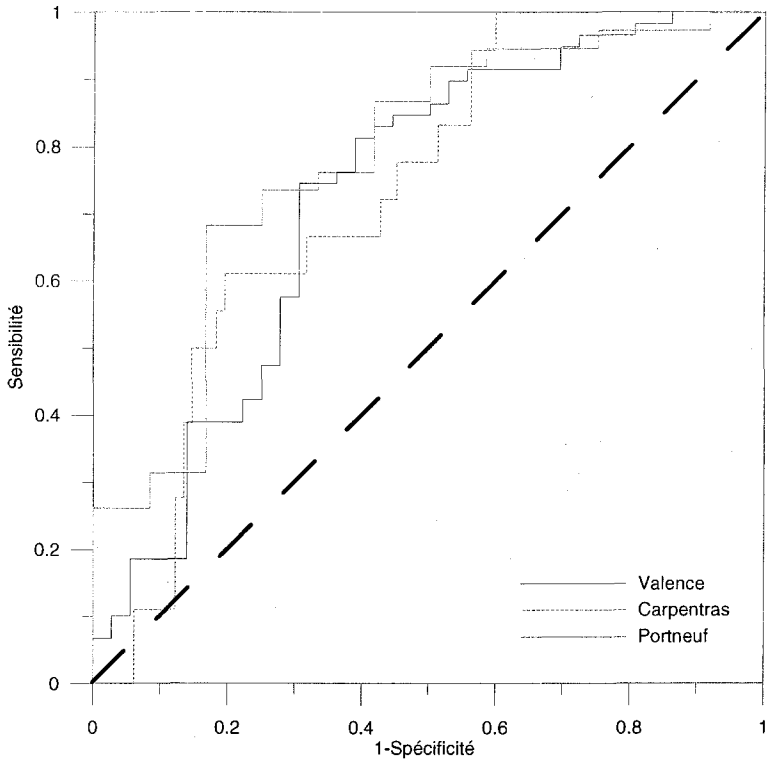


Figure 2-25. Courbe ROC pour les nitrates

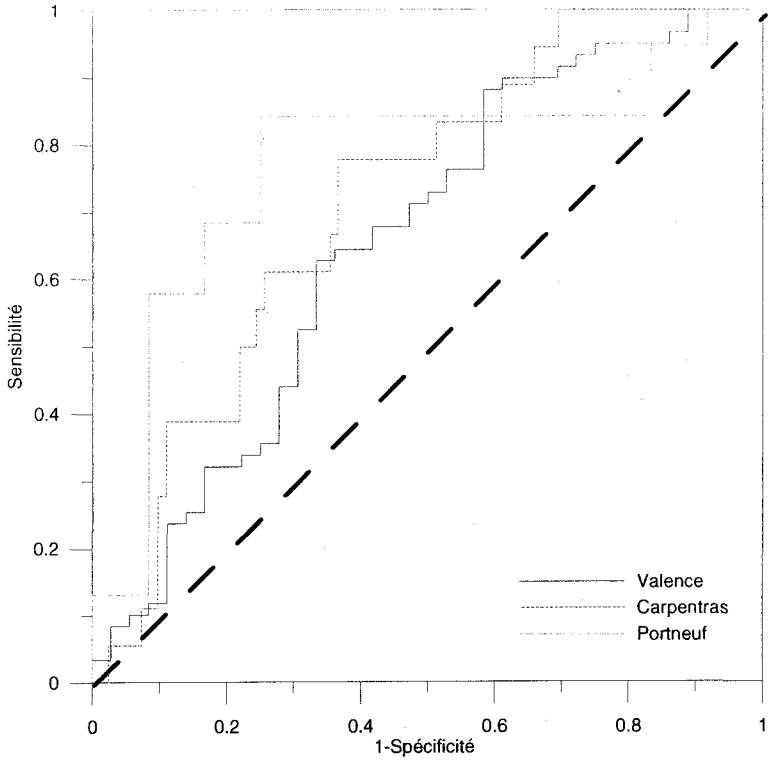


Figure 2-26. Courbe ROC pour les chlorures

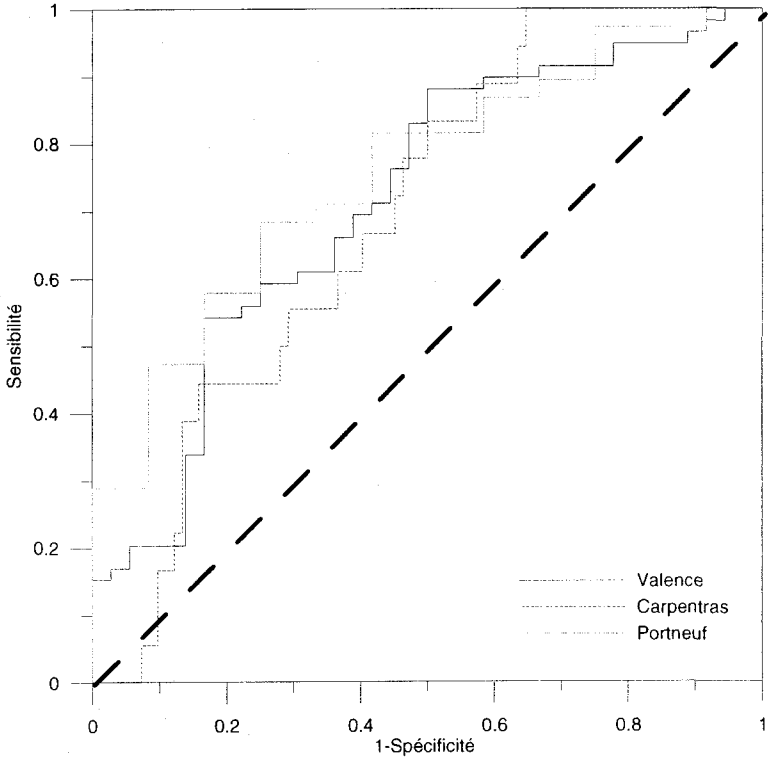


Figure 2-27. Courbe ROC pour les sulfates

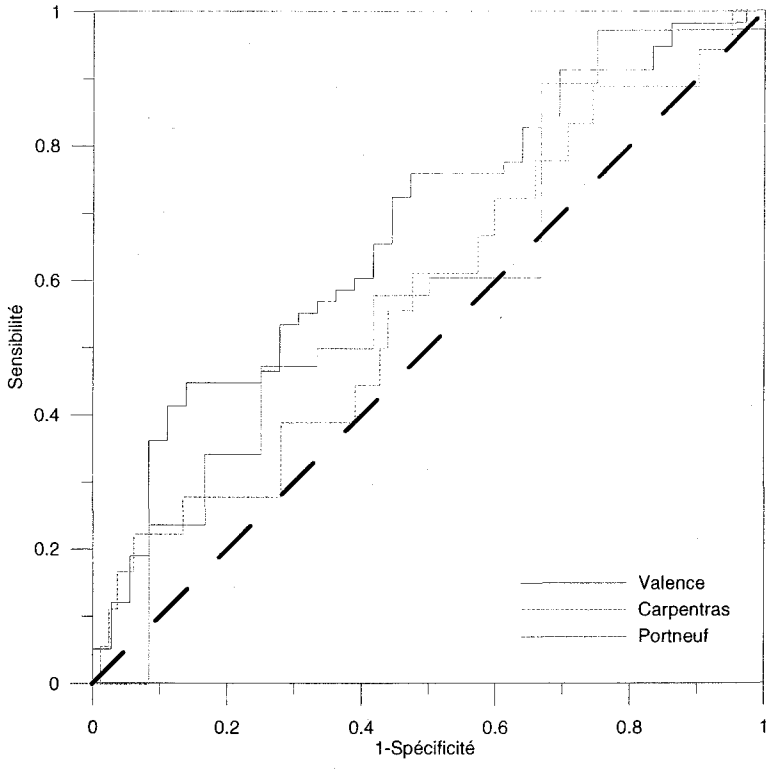


Figure 2-28. Courbe ROC pour le carbone organique dissous.

Tableau 2-17. Aire sous la courbe et valeur optimale de chlorures obtenues à partir des courbes ROC

	Aire sous la courbe	Valeur la plus discriminative (mg/l)
Valence	0.66	6.3
Carpentras	0.71	14.8
Portneuf	0.77	6.0

Pour les sulfates (Tableau 2-18), la capacité de discrimination est raisonnable (AUC comprise entre 0.69 et 0.76). Cependant la valeur seuil des bassins de Carpentras-Valréas est très élevée par rapport aux deux autres. Alors que pour le bassin de Valence et le Comté de Portneuf, les seuils sont inférieurs à 15 mg/l, cette valeur s'élève à 30 mg/l pour Carpentras-Valréas. Nous savons que Carpentras-Valréas présente de très fortes concentrations en sulfates, ce qui peut expliquer cette forte valeur, mais montre l'importance des propriétés géochimiques du site.

Tableau 2-18. Aire sous la courbe et valeur optimale de sulfates obtenues à partir des courbes ROC

	Aire sous la courbe	Valeur la plus discriminative (mg/l)
Valence	0.71	10.2
Carpentras	0.69	30.1
Portneuf	0.76	14.8

Le carbone organique dissous présente un pouvoir discriminant assez faible. En effet, les AUC ne dépassent pas 0.7.

Tableau 2-19. Aire sous la courbe et valeur optimale en carbone organique dissous obtenues à partir des courbes ROC

	Aire sous la courbe	Valeur la plus discriminative (mg/l)
Valence	0.66	1.4
Carpentras	0.57	2.0
Portneuf	0.59	0.6

Les courbes ROC appliquées sur les trois sites montrent que les nitrates, les chlorures et les sulfates présentent des capacités de discrimination en termes de détection de pesticides. Les discriminations ne sont cependant pas suffisantes individuellement (AUC de l'ordre de 0.70) pour permettre d'évaluer un risque de contamination par les pesticides. D'autre part, on remarque que pour les trois sites, les valeurs les plus discriminatives ne sont pas toujours du même ordre de grandeur. Pour les bassins de Carpentras-Valréas par exemple, les valeurs de chlorures et de sulfates sont plus élevées que pour les deux autres bassins. Ceci montre que l'on retrouve des tendances similaires sur les trois sites avec des concentrations plus élevées en nitrates, chlorures et sulfates dans les échantillons présentant des pesticides. Cependant, l'ordre de grandeur de ces concentrations n'est pas le même dans tous les sites du fait des caractéristiques géochimiques qui leur sont propres.

2.7 Conclusion partielle

Ce chapitre avait pour objectif de confirmer ou d'infirmer la présence d'une relation entre la présence de certains contaminants d'origine agricole et celle des pesticides analysés. Pour cela, trois sites ont été échantillonnés avec près de 250 prélèvements sur lesquels les concentrations des différents contaminants, solutés et pesticides ont été analysées. Ces trois sites présentent des caractéristiques différentes, tant d'un point de vue hydrogéologique et climatique que culturales.

Les trois sites présentent des corrélations significatives entre les concentrations en nitrates, chlorures et sulfates et celles des différents pesticides analysés. Cependant, ces corrélations sont relativement faibles. Plusieurs explications peuvent être fournies :

- les nitrates, chlorures et sulfates n'ont pas le même comportement que les pesticides et les processus qui interviennent dans leurs transports ne sont pas les mêmes ;
- les nitrates, chlorures et sulfates peuvent avoir d'autres origines (naturelles ou anthropiques) qu'une pollution d'origine agricole, certaines fortes concentrations ne correspondant donc pas forcément à une vulnérabilité de l'aquifère ;
- un nombre limité de pesticide a été choisi ; bien que ceux-ci correspondent aux plus fortes probabilités de détection, il se peut que certaines molécules présentes n'aient pas été analysées ;
- les concentrations en pesticides sont faibles, parfois très proches de la limite de détection, et les incertitudes sont donc importantes par rapport à l'ordre de grandeur des concentrations obtenues.

Une deuxième approche discriminante permettant de s'affranchir des valeurs de concentrations en pesticides a été abordée. Il s'agissait d'évaluer si les concentrations des contaminants agricoles possédaient un pouvoir discriminant en terme de détection ou non détection de pesticides. Pour les trois sites et les trois contaminants, les concentrations sont statistiquement supérieures dans les échantillons qui présentent des détections de pesticides. Leur pouvoir discriminant évalué à partir de courbes ROC est raisonnable (AUC supérieure à 0.7), mais n'est pas suffisant individuellement pour prédire la détection de pesticides. La combinaison de ces différentes variables peut néanmoins augmenter ce pouvoir discriminant. C'est l'objet du prochain chapitre.

Chapitre 3 Modélisation neuronale

Le troisième chapitre traite de l'application de la modélisation neuronale sur les données obtenues expérimentalement sur les trois sites d'étude. Les généralités sur les réseaux de neurones seront tout d'abord présentées avant d'aborder la méthodologie utilisée, les résultats et l'analyse des performances de classification en termes de détection de pesticide.

3.1 Introduction

Les réseaux de neurones artificiels (RNA) sont, depuis une quinzaine d'années, de plus en plus utilisés dans les sciences environnementales. Leur intérêt se révèle surtout dans les situations où la représentation physique des différents processus est complexe et soumise à des incertitudes ou lorsque les données présentent du bruit. En sciences de l'eau, les RNA sont appliqués dans des domaines très variés tel que la modélisation pluie-débit (Kuo-Lin Hsu *et al.* 1995), la modélisation de la qualité de l'eau (Maier et Dandy 1996), du couvert de glace (Seidou *et al.* 2006) ou du niveau piézométrique (Shukla *et al.* 1996). Maier et Dandy (2000) ont établi une revue des différentes applications et méthodologies appliquées pour les ressources en eau.

Pour le cas plus particulier des pesticides dans l'eau souterraine, une équipe a publié plusieurs travaux sur la prédiction des concentrations en pesticides dans les puits domestiques (Mishra *et al.* 2004) ; (Ray et Klindworth 2000) ; Sahoo *et al.* 2005 ; (Sahoo *et al.* 2006;Sahoo *et al.* 2005). Dans ces études, les entrées sont des indices représentant certaines propriétés physiques du milieu, des pesticides ou des ouvrages de captage. Les résultats en terme de performances du réseau sont bons, mais l'approche sous forme de catégories des variables d'entrée a obligé les auteurs à réaliser de fortes approximations.

Les réseaux de neurones sont souvent perçus comme une alternative à certaines difficultés rencontrées dans les méthodes statistiques classiques (Maier et Dandy 2000). Cependant, plusieurs auteurs ont démontré que les réseaux sont parfois très proches, voire équivalents, à certains modèles statistiques (Cheng et Titterington 1994) ; Fortin *et al.* 1997). Les réseaux de neurones permettent de regrouper ces modèles en un seul outil et de les rendre accessibles aux praticiens (Maier *et al.* 2000) tout en gardant à l'esprit que ces modèles présentent cependant certaines limites. Ce sont par exemple des outils qui peuvent difficilement extrapoler au-delà de la marge de grandeur des données d'apprentissage, et qui fonctionnent d'autre part comme des boîtes noires, ce qui en limite l'extraction de connaissances.

3.2 Les réseaux de neurones artificiels – Concepts généraux

Un réseau de neurones est un ensemble d'unités de calcul, appelées neurones, connectées les unes aux autres et capables de se transmettre des informations au moyen des connexions qui les relient. Bien qu'il existe plusieurs sortes de réseaux de neurones artificiels, seuls les perceptrons multicouches (PMC) utilisés dans l'étude seront abordés ici.

3.2.1 Le neurone de base

Les neurones (Figure 3-1) sont le reflet de l'inspiration biologique qui a été à l'origine de la première vague d'intérêt pour les neurones formels, dans les années 1940 à 1970.

Le premier modèle du neurone biologique est proposé par (McCulloch et Pitts 1943). En s'appuyant sur les propriétés des neurones biologiques connues à cette époque, issues d'observations neurophysiologiques et anatomiques, McCulloch et Pitts proposent un modèle simple de neurone formel.

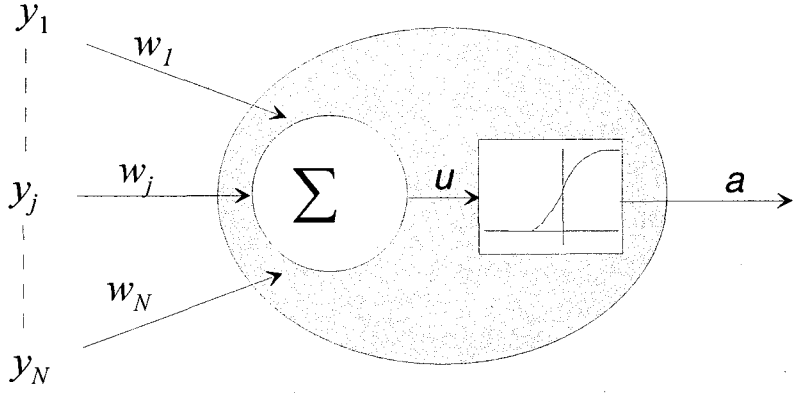


Figure 3-1. Schématisation d'un neurone

Chaque neurone est défini par trois caractéristiques : son signal a_i, ses connexions avec les autres neurones et sa fonction de transfert. Nous utiliserons par la suite les notations suivantes :

- *ai* : le signal du neurone *i*.
- *fi* : la fonction de transfert associée au neurone *i*.
- *ui* : le potentiel du neurone *i*.

- *wij* : le poids de la connexion du neurone *j* vers le neurone *i*.

Le neurone *i*, recevant les informations de n neurones différents, effectue l'opération suivante :

$$a_i = f(u_i) \text{ avec } u_i = \sum_{j=1}^{n} w_{ij} a_{ij} \qquad [3.1]$$

Différentes fonctions de transfert peuvent être utilisées mais celle la plus couramment utilisée est la fonction sigmoïde :

$$f(u) = \frac{1}{1+e^{-cx}} \qquad [3.2]$$

Lorsque c tend vers l'infini, la fonction sigmoïde tend vers une fonction seuil. Pour c=1 on retrouve la fonction logistique (Figure 3-2). La fonction logistique donne une réponse sur l'espace [0,1]. L'intérêt de cette fonction réside dans le fait qu'elle est facilement dérivable.

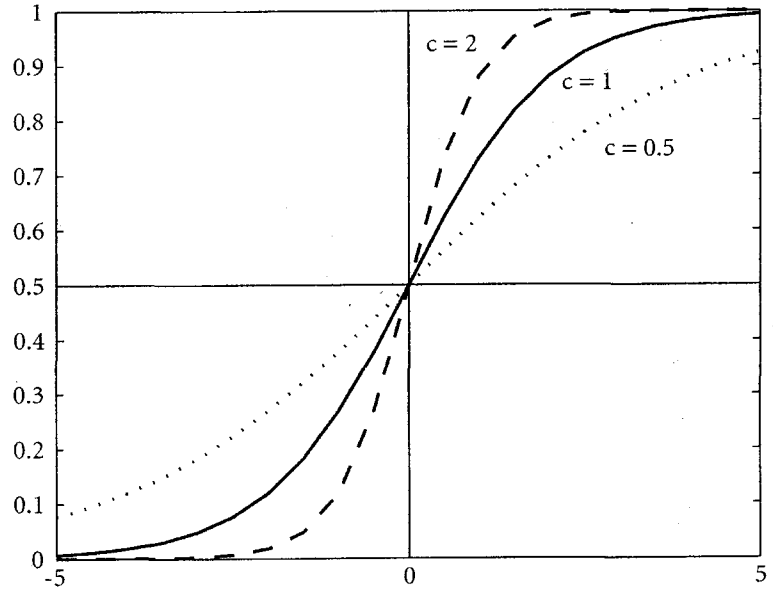


Figure 3-2. Différents paramètres de la fonction sigmoïde (Rennard 2006)

3.2.2 Le perceptron multi-couches (PMC)

3.2.2.1 Structure

Les neurones que l'on vient de définir doivent être assemblés pour former un réseau. Le type de réseau le plus simple s'appelle le perceptron. Il est constitué d'un seul neurone et permet de réaliser des opérations très simples. Il est cependant limité. C'est pourquoi on utilise un type de réseau plus complexe, le perceptron multicouche (Figure 3-3). Comme son nom l'indique, il est constitué de plusieurs couches de neurones connectées entre elles. Une couche est un ensemble de neurones n'ayant pas de connexion entre eux. Un neurone ne peut donc transmettre son état qu'à un neurone situé dans une couche postérieure à la sienne.

La première couche du réseau est la couche d'entrée, on suppose qu'elle contient p entrées. Par abus de langage, on parle parfois de neurones d'entrée, cependant les entrées (représentées par des triangles sur la Figure 3-3) ne sont pas des neurones. Elles ne réalisent aucun traitement de l'information puisqu'elles ne font que transmettre les valeurs des variables. Les états des neurones de cette première couche seront fixés par le problème traité à travers un vecteur $x = (x_1,x_2,\ldots x_p)$. Les états de la première couche étant fixés, le réseau va pouvoir calculer les états des autres neurones en appliquant l'équation [3.1] de proche en proche. Cette partie du calcul est appelée propagation avant, en opposition à la rétropropagation qui sera présentée dans la section suivante.

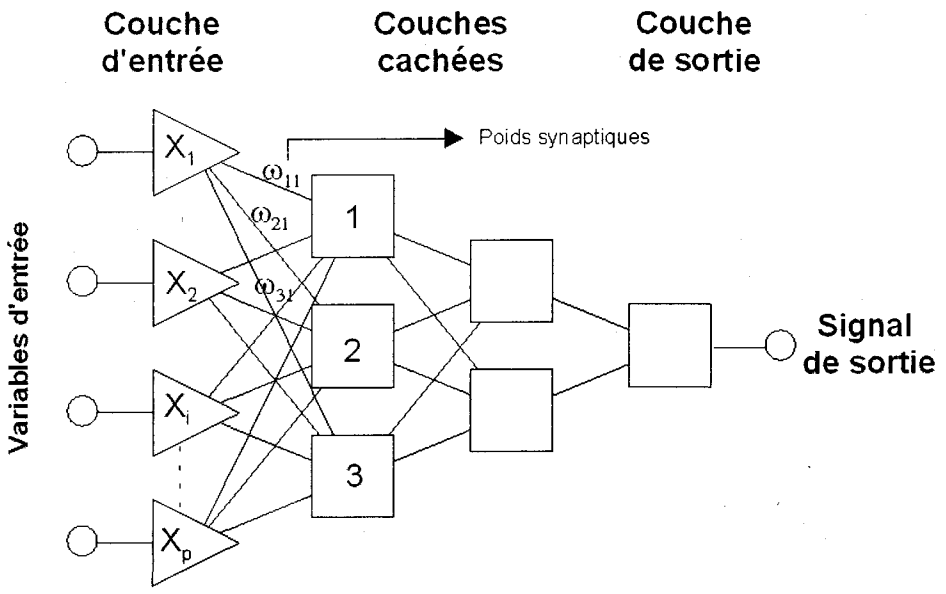


Figure 3-3. Illustration d'un perceptron multi-couches à 2 couches cachées

3.2.2.2 Apprentissage et algorithme de rétropropagation du gradient

La technique de rétropropagation du gradient consiste à définir une notion d'erreur puis à calculer la contribution à cette erreur des différents éléments. Les poids synaptiques qui contribuent à engendrer une erreur importante se verront alors modifiés de manière plus significative que les poids qui ont engendré une erreur négligeable. Les poids du réseau sont tout d'abord initialisés avec des valeurs aléatoires.

Étape 1 : Propagation avant et calcul de l'erreur totale du réseau

Les sorties du réseau sont calculées en appliquant l'équation [3.1] de proche en proche. Dans le cas des problèmes à apprentissage supervisé, les valeurs de sortie seront comparées aux valeurs désirées afin de calculer l'erreur totale du réseau. Si cette erreur est jugée suffisamment faible, l'apprentissage sera considéré comme terminé. Dans le cas contraire, le réseau va modifier les poids afin de minimiser cette erreur. Pour cela, la première étape consiste à calculer le signal d'erreur sur chacun des neurones de sortie.

Étape 2 : Calcul du signal d'erreur sur les neurones de sortie

Le signal d'erreur sur chacun des neurones de sortie est calculé tel que :

$$\delta_i^s = (d_i - a_i) \times f'(u_i^s) \qquad [3.3]$$

Où δ_i^s est le signal d'erreur du neurone i de la couche de sortie s ; d_i est le signal désiré sur le neurone de sortie i ; a_i est le signal de sortie du neurone i et f' est la dérivée de la fonction de transfert. Les signaux d'erreur de la couche de sortie vont permettre de calculer ceux des couches cachées.

Étape 3 : Calcul du signal d'erreur des neurones cachés

Les signaux d'erreur des neurones cachés sont fonction de ceux de la couche suivante (notion de rétropropagation) :

$$\delta_j^{(n-1)} = f'(u_j^{(n-1)}) \times \sum_{i=1}^{N} \delta_i^{(n)} \omega_{ij} \qquad [3.4]$$

Où $\delta_j^{(n-1)}$ est le signal d'erreur du neurone j de la couche $n\text{-}1$ et ω_{ij} le poids synaptique du neurone j vers le neurone i.

Étape 4 : Correction des poids sur toutes les couches

La correction utilise les fonctions suivantes :

$$\begin{aligned} \omega_{ij}^{(t+1)} &= \omega_{ij}^{(t)} + \Delta\omega_{ij} \\ \Delta\omega_{ij} &= \eta \delta_i^{(n)} a_j^{(n-1)} \end{aligned} \qquad [3.5]$$

Où η est le pas d'apprentissage.

Tous les poids du réseau ont ainsi été ajustés afin de minimiser l'erreur. L'algorithme est itératif et les étapes seront alors appliquées autant de fois que nécessaire pour que l'erreur soit jugée acceptable. Le nombre d'itérations est appelé époque.

3.3 Méthodologie

Dans cette étude, les réseaux de neurones sont appliqués pour la classification, c'est-à-dire qu'ils sont utilisés afin d'évaluer leur capacité à classer les échantillons en deux classes :

- la classe 1 qui correspond aux échantillons dont les concentrations en pesticides sont supérieures à la limite de quantification ;
- la classe 0 qui correspond aux échantillons dont les concentrations en pesticides sont inférieures à la limite de quantification.

Les données des trois sites ont été utilisées conjointement, ce qui représente un total de 245 échantillons, dont 117 pour la classe 1 et 128 pour la classe 0.

Les réseaux utilisés sont uniquement les perceptrons multicouches où toutes les fonctions de transfert sont des fonctions logistiques et l'algorithme d'apprentissage utilisé est celui de rétropropagation du gradient. La mise en place d'un réseau nécessite plusieurs étapes préalables :

- la préparation des données du modèle ;
- le choix d'un critère de performance ;
- la sélection des variables d'entrée ;
- la sélection de l'architecture du modèle.

3.3.1 Préparation des données d'entrée

3.3.1.1 Normalisation

Dans la pratique, les valeurs reçues en entrée du modèle peuvent varier sans limites. Ceci peut poser d'importants problèmes de convergence puisque l'ajustement synaptique dépend des valeurs des signaux (équation [3.5]). Le réseau va ainsi accorder une grande importance aux entrées de valeurs élevées au détriment d'entrées tout aussi importantes, mais dont les valeurs sont faibles. Il est donc important de normaliser les entrées afin que le réseau accorde la même importance à toutes les variables.

Les variables d'entrée sont donc toutes centrées et réduites afin d'obtenir une distribution de moyenne 0 et d'écart type 1 :

$$W_n' = \frac{(X_n - \mu_n)}{\sigma_n}$$

Où X_n est le vecteur d'entrée, μ_n sa moyenne et σ_n son écart-type.

Dans la plupart des modèles statistiques, les données doivent présenter une distribution normale afin que les coefficients puissent être estimés correctement. Si les données ne sont pas normales, il faut donc les transformer. Dans le cas des RNA, la littérature semble dire que la distribution des données d'entrée n'a pas besoin d'être connues (Burke et Ignizio 1992). Cependant, certains auteurs expliquent que l'utilisation de la RMSE (root mean square error) nécessiterait la normalité des données d'entrée (Fortin *et al.* 1997).

3.3.1.2 Division des données

Une fois que le procédé d'apprentissage a été effectué, la performance du modèle doit être évaluée sur des données qui n'ont pas été utilisées pour la phase d'apprentissage. Les données sont ainsi divisées en trois séries :

- Une série d'apprentissage est utilisée afin d'estimer la fonction qui lie les entrées du modèle à la sortie. C'est donc à partir de cette série d'exemples que la matrice de poids sera calculée. Le nombre d'observations présentes dans la série d'apprentissage a une influence significative sur les performances du réseau. Plus il est grand, plus l'information disponible est importante, augmentant ainsi le potentiel de performance pouvant être atteint par le réseau (Flood et Kartam 1994).
- Une série de validation est utilisée comme critère d'arrêt de l'apprentissage. Après chaque itération, le groupe de validation est présenté au réseau sans procéder à la phase de rétropropagation. Si l'erreur est trop importante, l'apprentissage reprend. On constate généralement qu'au-delà d'un certain seuil, alors que l'erreur continue de décroître sur les données d'apprentissage, elle a tendance à augmenter sur les données de validation. Ceci tient à l'existence de bruit dans les données. Il est alors recommandé de stopper l'apprentissage (Figure 3-4) avant que celui-ci ne les prenne en compte. Ce phénomène, qui peut fortement dégrader les capacités de généralisation du réseau, est appelé sur-apprentissage.

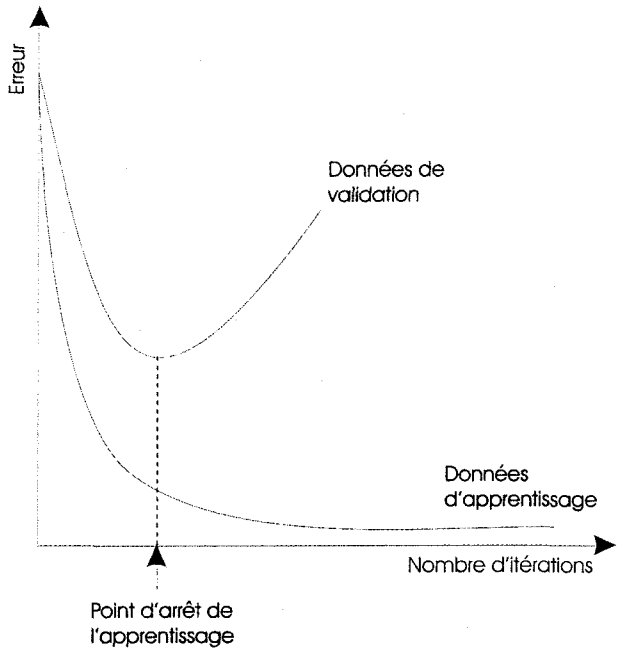


Figure 3-4. Méthode d'arrêt de l'apprentissage

- Une série de test permet d'évaluer la capacité de généralisation du modèle, c'est-à-dire sa capacité à fonctionner sur des données qui ne lui ont jamais été présentées (Cheng et Titterington 1994). Elle ne doit en aucun cas intervenir dans le processus d'apprentissage. Il s'agit donc d'une série indépendante.

Il est recommandé en général de s'assurer de la bonne répartition des données dans les trois séries, car celle-ci peut avoir une influence sur les performances du réseau. Les réseaux n'étant pas capables d'extrapoler en dehors des données d'apprentissage, celles-ci doivent être représentatives de la population. Pour l'étude, environ 70 % des échantillons ont été utilisés pour l'apprentissage, soit 169 échantillons. Le reste a été réparti dans les deux autres séries avec chacune 38 échantillons.

3.3.2 Critères de performance

Plusieurs critères ont été utilisés pour évaluer et comparer les performances des réseaux. Dans les problématiques de classification, c'est essentiellement la performance (P%) qui est importante, c'est-à-dire la capacité du réseau à classer correctement les objets :

$$P(\%) = \frac{\text{Nombre d'éléments bien classés}}{\text{Nombre total d'éléments}} \times 100 \quad [3.6]$$

La sensibilité Se et la spécificité Sp, qui représentent respectivement le nombre d'éléments positifs (classe 1) et négatifs (classe 0) bien classés, permettent d'évaluer la performance des réseaux sur chacune des classes et de s'assurer que ceux-ci n'ont pas tendance à regrouper tous les échantillons dans une même classe.

$$Se = \frac{N_1 \text{ bien classés}}{N_1 total} \quad [3.7]$$

$$Sp = \frac{N_0 \text{ bien classés}}{N_0 total} \quad [3.8]$$

Où N_1 est le nombre d'éléments dans la classe 1 et N_0 le nombre d'élément dans la classe 0.

L'aire sous la courbe ROC (AUC) permet de représenter le pouvoir de classification global du modèle.

La RMSE (*root mean square error)* est l'erreur sur laquelle se base l'apprentissage. Elle est fournie à titre indicatif puisqu'une erreur faible n'indique pas forcément de bonnes capacités en termes de classification.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(T_i - P_i)^2} \quad [3.9]$$

Où T_i est la valeur cible et P_i la valeur calculée.

3.4 Sélection des variables d'entrée

3.4.1 Introduction

La sélection des variables dans les techniques d'apprentissage est une étape incontournable. D'importantes quantités de données sont parfois disponibles, mais toutes ne possèdent pas le même degré d'information. Certaines variables peuvent être fortement bruitées, d'autres peuvent être corrélées entre elles, redondantes ou non pertinentes pour l'objectif visé. L'objectif de la sélection de variables sera donc de déterminer un sous-ensemble de variables pertinent pour le problème en question. De grands bénéfices peuvent être tirés d'une telle sélection (Guyon et Elisseeff 2003) ; (Bowden *et al.* 2005a; Bowden *et al.* 2005b):

- amélioration des performances du réseau ; la présence de variables non pertinentes rend plus difficile l'apprentissage et dégrade la capacité de généralisation du réseau ;
- réduction du nombre d'échantillons ; plus il y a de variables d'entrée, plus la série d'apprentissage doit être importante, donc plus le nombre d'échantillons doit être grand (croissance exponentielle) ;
- réduction du temps de calcul ;
- facilité d'extraction de connaissances ;
- réduction des coûts de futures mesures et du traitement de données.

La sélection des variables d'entrée pour les méthodes d'apprentissage a fait l'objet de nombreuses recherches. Depuis une dizaine d'années, ont pris place des études plus spécifiques à l'application aux réseaux de neurones (Leray et Gallinari 2001) (Bowden *et al.* 2005a). Dans le domaine des sciences de l'eau, les méthodes de sélection de variables les plus utilisées sont :

- Les connaissances *a priori* du système représenté. De nombreux travaux n'utilisent que cette méthode afin de sélectionner les variables d'entrée pertinentes. Il est reconnu dans le domaine que le jugement d'expert est essentiel à la détermination des entrées du modèle (ASCE 2000a) afin de ne pas omettre une variable importante. Cependant, cette sélection reste subjective et dépendante du praticien. Une combinaison avec des méthodes analytiques est généralement préférée (Maier et Dandy 1997).

- Les méthodes indépendantes du modèle. Dans ces méthodes, la sélection des variables est réalisée par diverses méthodes analytiques indépendamment des résultats fournis par le réseau. Ce sont souvent les corrélations qui sont utilisées. L'inconvénient de ces méthodes est que la sélection des variables est le plus souvent basée sur des relations linéaires entre les entrées et les sorties. On risque alors d'omettre certaines variables qui présentent une relation non linéaire avec la sortie. D'autre part, l'effet combinatoire des variables entre elles n'est pas pris en compte.
- Les méthodes dépendantes du modèle. La sélection des variables est réalisée en fonction des résultats du réseau. La méthode la plus appliquée est l'analyse de sensibilité qui permet d'estimer l'impact de chaque variable. La difficulté réside alors dans le choix d'un seuil de sélection.

3.4.2 Méthodologie

Dans la présente étude, plusieurs méthodes de sélection ont été comparées sur les données des sites afin de sélectionner un jeu optimal de variables.

- Une méthode basée sur le jugement : les contaminants.
- Une méthode de sélection indépendante du modèle : stepdisc.
- Deux méthodes dépendantes du modèle : l'analyse de sensibilité et l'algorithme génétique.

3.4.2.1 Les contaminants

Nous avons mis en évidence dans le chapitre 2 que certains contaminants chimiques présentaient un pouvoir discriminant sur la présence de pesticides dans les échantillons. Le premier jeu est donc composé de ces quatre variables, soit les nitrates, les chlorures, les sulfates et le carbone organique dissous.

3.4.2.2 Stepdisc

Dans cette méthode, le pouvoir discriminant est évalué par la statistique du lambda de Wilks (Λ). Géométriquement, il s'agit de trouver le sous-espace de représentation qui permet un écartement maximal entre les centres de gravité des nuages de points conditionnels, c'est-à-dire des nuages

de points associés à chaque valeur de la variable à prédire. Si les nuages sont totalement confondus, Λ prend une valeur de 1 ; plus Λ se rapproche de 0, plus les nuages conditionnels sont distincts (Rakotomalala 2005).

Pour la sélection des variables, l'approche *forward* avec un seuil de 5 % a été utilisée. Cette approche consiste, à partir de l'ensemble vide, à choisir la variable induisant la meilleure amélioration du Λ et la sélectionner si l'amélioration est statistiquement significative. La procédure est itérative en ajoutant une à une les variables jusqu'à ce que l'adjonction d'une variable n'apporte plus d'amélioration.

Au seuil 5 %, quatre variables sur huit ont été sélectionnées. Les résultats sont présentés au Tableau 3-1.

Tableau 3-1. Résultats de la statistique du lambda de Wilks. Les variables en gras correspondent à celles sélectionnées au seuil 5 %

Rang	Variables	Λ de Wilks	F
1	**NO_3^-**	**0.824**	**51.87**
2	**pH**	**0.754**	**22.19**
3	**T**	**0.735**	**6.25**
4	**COD**	**0.721**	**4.45**
5	$SO4^{2-}$	0.715	2.08
6	σ	0.699	5.62
7	Cl^-	0.697	0.42
8	Prof.	0.697	0.06

Les variables sélectionnées par cette méthode sont donc les nitrates, le pH, la température et le carbone organique dissous.

3.4.2.3 Algorithme génétique

Les algorithmes génétiques constituent une méthode de sélection des variables de plus en plus utilisée pour les réseaux de neurones. Ils permettent de sélectionner *a priori* la meilleure combinaison de variables en testant toutes les combinaisons possibles. L'intérêt réside essentiellement dans la prise en compte de la redondance des informations. Bien que cette méthode soit davantage adaptée pour les problèmes présentant une grande quantité de variables

d'entrée, l'information peut tout de même être intéressante dans les problèmes où le nombre de variables disponibles est plus faible.

Trois variables ont été sélectionnées par l'algorithme génétique, il s'agit des nitrates, de la température et du pH.

3.4.2.4 Analyse de sensibilité

L'analyse de sensibilité (Sn) consiste à réévaluer la performance du réseau lorsque l'on retire la variable à tester:

$$Sn = \frac{\text{Performance du réseau avec la variable}}{\text{Performance du réseau sans la variable}} \qquad [3.10]$$

Une sensibilité de 1 indique ainsi que la variable n'a pas d'influence sur le réseau, une valeur inférieure à 1 indique que la variable a tendance à dégrader la performance du réseau et une valeur supérieure à 1 qu'elle augmente la capacité du réseau.

Les résultats n'indiquent pas forcément l'importance de l'information. Par exemple dans le cas où deux variables sont corrélées, celles-ci peuvent présenter une faible sensibilité dans la mesure où l'information sera toujours fournie par l'autre variable. À l'inverse, une variable présentant une information relativement peu importante peut avoir une forte sensibilité dans la mesure où cette information est unique. Le seuil de sélection doit donc être choisi avec précaution et les analyses effectuées plusieurs fois. Il est généralement conseillé de retirer même les variables dont la sensibilité est légèrement supérieure à 1. L'apprentissage sera plus facile si le nombre de variables diminue et la performance est souvent compensée, voire améliorée, par les autres variables. Tenant compte de ces éléments et de la gamme de grandeur des sensibilités, le seuil de sélection a été fixé à 1.05.

La Figure 3-5 présente les sensibilités moyennes pour dix procédures d'apprentissages. En moyenne, aucune des variables ne présente de sensibilité inférieure à 1, ce qui signifie qu'aucune des variables ne semble dégrader la performance du réseau. Cependant, les variables qui présentent une sensibilité supérieure au seuil de 1.05 sont les chlorures, les nitrates, la température et le pH.

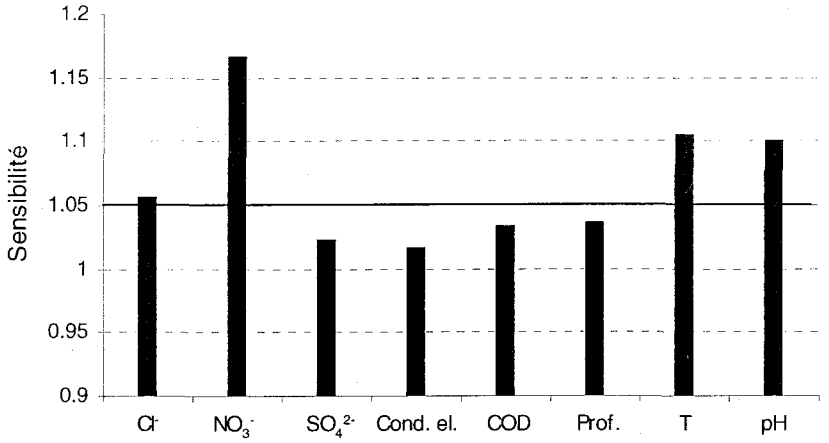


Figure 3-5. Analyse de sensibilité des variables d'entrée du réseau

3.4.3 Résultats et comparaison des différents jeux de variables

Les trois méthodes de sélection de variables ont, toutes, retenu les nitrates, le pH et la température. Il semblerait donc que ces trois variables (qui correspondent à celles sélectionnées par l'algorithme génétique) soient les plus importantes pour le réseau. Dans les analyses de sensibilité, la variable nitrates est celle qui présente les plus fortes valeurs. On peut l'associer à une pollution anthropique d'origine agricole ce qui est pertinent pour le problème étudié. La température et le pH permettent de définir les caractéristiques fondamentales de l'eau. Des contrastes entre ces paramètres peuvent faciliter la mise en évidence de zones d'alimentation, de pollutions ou de géologies différentes. Par exemple le pH a tendance à être plus faible pour des aquifères sableux pouvant ainsi traduire une plus grande vulnérabilité. Ces paramètres ont également une influence sur de nombreuses réactions chimiques telles que les processus de dégradation des pesticides.

La conductivité électrique et les sulfates n'ont été sélectionnés par aucune méthode. Ces deux variables sont corrélées entre elles, mais également aux autres anions. La redondance de l'information qu'elles fournissent peut alors expliquer le fait qu'elles n'aient jamais été retenues.

Les chlorures ont été sélectionnés uniquement par la méthode de sensibilité bien que très proches du seuil de 1.05. Ils peuvent comme les nitrates, apporter une information sur une éventuelle pollution d'origine agricole.

Le carbone organique dissous a été sélectionné par la méthode stepdisc. Nous avons vu dans le chapitre 2 qu'il peut être lié au comportement des pesticides. Associé à d'autres caractéristiques de l'eau, il est donc possible qu'il apporte une information pertinente au réseau.

3.4.3.1 Application et sélection des variables

Afin de sélectionner la meilleure combinaison de variables, les scénarios suivants (Tableau 3-2) correspondant aux différentes méthodes ont été testés.

Tableau 3-2. Cinq scénarios utilisés pour la sélection des variables d'entrée

		Variables sélectionnées						
Méthode	n	Cl^-	NO_3^-	SO_4^{2-}	CE	COD	T	pH
Aucune	7	1	1	1	1	1	1	1
Contaminants	4	1	1	1	0	1	0	0
Stepdisc	4	0	1	0	0	1	1	1
Sensibilité	4	1	1	0	0	0	1	1
Génétique	3	0	1	0	0	0	1	1

Afin de s'assurer de la représentativité des résultats en vue d'une comparaison des réseaux, les mesures suivantes ont été appliquées :

- Étant donné que la répartition des échantillons dans les trois séries peut avoir une influence sur les performances du réseau, les réseaux ont donc été testés pour cinq répartitions différentes et aléatoires des données. En effet, si les caractéristiques des trois séries diffèrent trop, il se peut que le réseau présente de faibles capacités de généralisation. Il en résulterait alors de faibles performances sur les séries de test ou de validation.
- Étant donné que le résultat final d'un réseau dépend des conditions initiales (attribution aléatoire des poids synaptiques), si celles-ci ne sont pas optimales, il se peut que le réseau ne converge pas suffisamment et cela quel que soit le nombre d'époques. Les procédures d'apprentissage et de validation ont donc été répétées dix fois afin de s'exempter des incertitudes dues aux conditions initiales.

Les résultats présentés à la Figure 3-6 correspondent à 50 essais de chacune des combinaisons de variables. Chaque réseau a été testé avec une couche cachée dont le nombre de neurones était identique à celui des variables d'entrée.

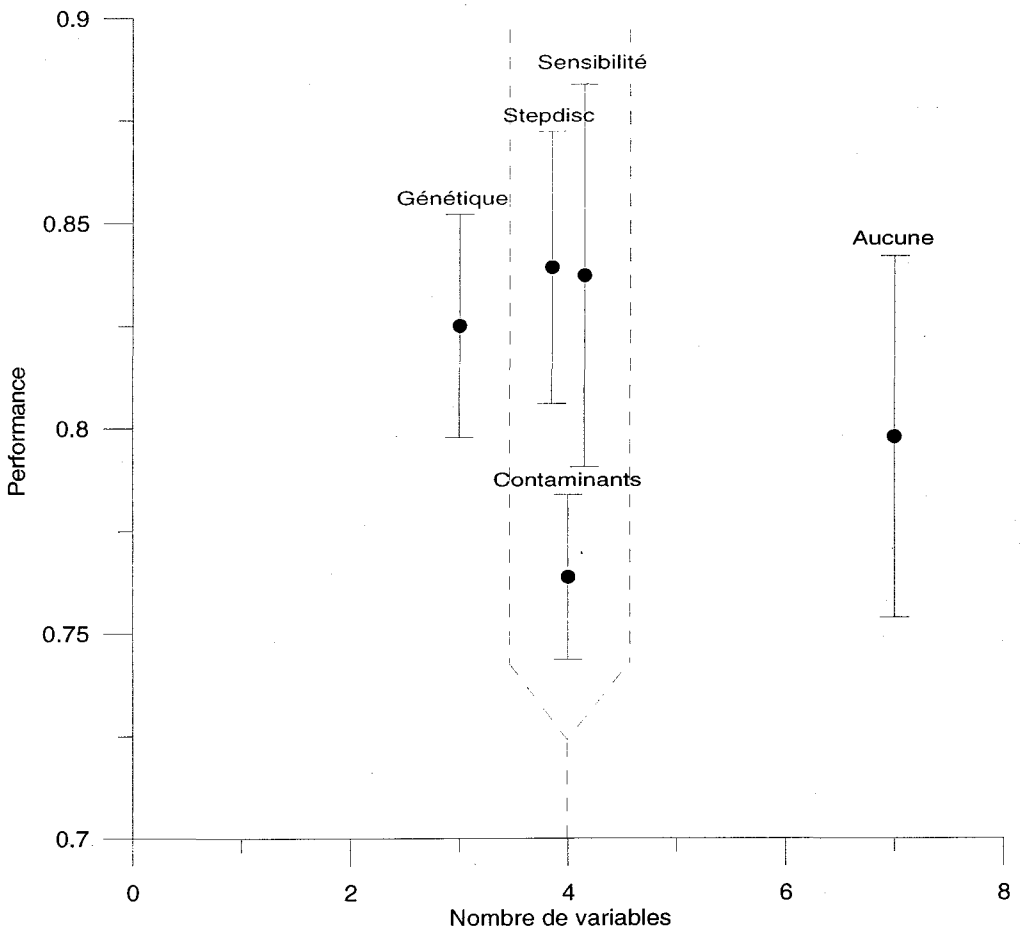


Figure 3-6. Comparaison des performances de différentes méthodes de sélection de variables en fonction du nombre de variables

La Figure 3-6 montre que la sélection des contaminants (chlorures, nitrates, sulfates et carbone organique dissous) est la combinaison de variables qui présente les moins bonnes performances avec un maximum de 75 % de bons classements. Les variables étant corrélées entre elles, cela peut provoquer des problèmes de colinéarité qui sont réputés pour diminuer les capacités des réseaux de neurones. Les méthodes stepdisc et de sensibilité présentent des performances du même ordre de grandeur (près de 85 % en moyenne). Leur légère supériorité à l'algorithme génétique illustre le fait que la présence des chlorures et du carbone organique dissous apporte une information au réseau. Ces deux variables seront donc conservées pour la suite de l'étude.

3.5 Architecture du réseau

Le choix de l'architecture du réseau consiste à fixer le nombre de couches cachées ainsi que le nombre de neurones par couche. En général, les petits réseaux (avec peu de connexions) sont considérés comme ayant de meilleures capacités de généralisation (Castellano *et al.* 1997), mais d'un autre côté la surface d'erreur contient davantage de minima locaux (Bebis et Georgiopoulos 1994). L'avantage des réseaux plus complexes réside dans leur meilleure capacité de convergence. Cependant, ceux-ci nécessitent un plus grand nombre d'échantillons d'apprentissage pour atteindre de bonnes capacités de généralisation. Il n'existe pas d'outil analytique permettant de connaître le nombre idéal de neurones cachés ; on procède le plus souvent par essais et erreurs bien que certains auteurs aient proposé quelques règles. Maren *et al.* (1990) suggèrent ainsi que, dans la majorité des applications, le nombre optimal de neurones cachés est inférieur ou égal à celui des entrées. Hecht-Nielsen (1988) quant à lui suggère de ne pas dépasser la limite suivante, fonction du nombre d'entrées :

$$N^C \leq 2N^E + 1 \qquad [3.11]$$

où N^C est le nombre de neurones cachés et N^E le nombre de variables d'entrée. Cependant afin d'éviter un phénomène de sur-apprentissage, certains auteurs proposent de considérer également le nombre d'exemples présents dans la série d'apprentissage. Ainsi Rogers *et al.* (1994) proposent le critère suivant :

$$N^C \leq \frac{N^{APP}}{N^E + 1} \qquad [3.12]$$

où N^{APP} est le nombre d'exemples d'apprentissage.

Notre étude présentant peu de variables d'entrée par rapport au nombre d'exemples d'apprentissage, l'équation [3.11] a défini la limite supérieure du nombre de neurones cachés soit $Nc < 11$ pour cinq variables d'entrée.

Dix réseaux à 1 couche cachée de 2 à 11 neurones ont été testés afin de définir la meilleure architecture. Pour chaque apprentissage, la division des données dans les trois séries était identique et les performances ont été évaluées sur 20 apprentissages successifs. Les résultats présentés dans le Tableau 3-3 correspondent aux moyennes des RMSE et des performances (P %)

observées sur les 20 apprentissages pour les trois séries de données. L'aire sous la courbe ROC (AUC) correspond aux séries de validation et de test.

Tableau 3-3. Performance des réseaux pour différentes architectures

Réseau	Apprentissage		Validation		Test		AUC
	RMSE	P (%)	RMSE	P (%)	RMSE	P (%)	
5-2-1	0.281	79.91	0.279	81.47	0.293	79.36	0.85
5-3-1	**0.274**	**80.77**	**0.279**	**82.40**	**0.295**	**81.40**	**0.87**
5-4-1	0.278	80.18	0.281	79.74	0.297	78.14	0.86
5-5-1	0.276	80.12	0.281	81.98	0.293	78.94	0.87
5-6-1	0.278	80.06	0.284	80.53	0.293	78.95	0.85
5-7-1	0.282	80.03	0.283	81.48	0.306	77.76	0.85
5-8-1	0.275	80.18	0.283	79.73	0.292	78.54	0.85
5-9-1	**0.275**	**80.56**	**0.278**	**82.52**	**0.297**	**81.12**	**0.86**
5-10-1	0.273	80.33	0.281	80.93	0.303	77.18	0.86
5-11-1	0.278	79.73	0.281	81.32	0.309	76.54	0.86

Les résultats du Tableau 3-3 montrent que les performances globales des réseaux varient peu en fonction du nombre de neurones cachés. Les performances les plus faibles sont observées pour les réseaux à 10 et 11 neurones cachés où la généralisation est légèrement moins bonne ($P\ \%_{test} < 78\ \%$). Les deux réseaux qui présentent les meilleures performances moyennes sont les réseaux à 3 et 9 neurones cachés (Tableau 3-3) avec des performances supérieures à 82 % pour la série de validation et supérieures à 81 % pour la série de test.

Les AUC calculées pour l'ensemble des séries de validation et de test donnent une idée globale du pouvoir discriminant du réseau. Leurs valeurs varient peu, mais le réseau 6-3-1 est légèrement supérieur. Étant donné qu'à performance égale, il est préférable de privilégier le réseau le moins complexe, le réseau 5-3-1 a été sélectionné pour la suite des analyses.

3.6 Classification du réseau

Le Tableau 3-4 présente les résultats pour dix apprentissages effectués en faisant varier la répartition des échantillons dans les différentes séries. La classe 1 correspond aux échantillons dont la concentration en pesticide est supérieure à la limite de quantification et la classe 0 à ceux qui sont inférieurs à cette limite. Les performances moyennes sont de 83.7 % pour la série de validation et 79.9 % pour la série de test. Les performances entre les différentes séries sont du même ordre de grandeur, ce qui montre la capacité de généralisation du réseau, soit sa capacité à fournir des résultats similaires à ceux de l'apprentissage sur des données nouvelles.

Le détail des classifications pour chaque série permet de calculer la sensibilité (Se) et la spécificité (Sp) afin d'évaluer les performances du réseau pour chacune des classes. Dans la majorité des cas, la sensibilité et la spécificité sont du même ordre de grandeur, ce qui permet de mettre en évidence le fait que le réseau n'a pas tendance à surestimer ou sous-estimer l'une des deux classes.

La colonne *Nerr* du Tableau 3-4 indique le nombre total d'échantillons mal classés. On remarque que ce nombre est plutôt constant (environ 45) même dans les cas où la généralisation semble moins bonne (cas 3). Une analyse détaillée a révélé que parmi les échantillons mal classés, 32 l'étaient systématiquement dans toutes les simulations. Il semblerait donc que le réseau ne soit pas capable de classer correctement ces échantillons et que ce soit leur répartition dans les différentes séries qui affecte les performances du réseau.

Tableau 3-4. Résultats de classification du réseau 5-3-1 pour dix apprentissages en faisant varier la distribution des échantillons dans les trois séries. Nerr représente le nombre total d'échantillons mal classés

	Performance P(%)			Classification																	Nerr
	App.	Val.	Test	Apprentissage						Validation						Test					
				Classe 1		Classe 0		Se	Sp	Classe 1		Classe 0		Se	Sp	Classe 1		Classe 0		Se	Sp	
				mesuré	correct	mesuré	correct			mesuré	correct	mesuré	correct			mesuré	correct	mesuré	correct			
1	82.2	81.6	76.3	78	59	91	80	75.6	87.9	18	16	20	15	88.9	75.0	21	17	17	12	81.0	70.6	46
2	79.3	86.8	84.2	81	61	88	73	75.3	83.0	14	13	24	20	92.9	83.3	22	18	16	14	81.8	87.5	46
3	85.2	84.2	65.8	78	70	91	74	89.7	81.3	17	13	21	19	76.5	90.5	22	13	16	12	59.1	75.0	44
4	81.1	76.3	84.2	83	69	86	68	83.1	79.1	15	10	23	19	66.7	82.6	19	16	19	16	84.2	84.2	47
5	80.5	84.2	78.9	81	64	88	72	79.0	81.8	20	16	18	16	80.0	88.9	16	15	22	15	93.8	68.2	47
6	78.7	86.8	78.9	82	65	87	68	79.3	78.2	16	12	22	21	75.0	95.5	19	17	19	13	89.5	68.4	49
7	82.7	81.6	79.5	86	72	82	67	83.7	81.7	14	10	24	21	71.4	87.5	17	14	22	17	82.4	77.3	44
8	79.8	89.5	84.6	81	65	87	69	80.2	79.3	19	17	19	17	89.5	89.5	17	14	22	19	82.4	86.4	44
9	83.3	81.6	82.1	81	69	87	71	85.2	81.6	19	16	19	15	84.2	78.9	17	12	22	20	70.6	90.9	42
10	82.1	84.2	84.6	84	68	84	70	81.0	83.3	19	16	19	16	84.2	84.2	14	13	25	20	92.9	80.0	42
Moy.	81.5	83.7	79.9					81.2	81.7					80.9	85.6					81.7	78.8	45

Tableau 3-5. Résultats de classification d'un réseau avec classe de rejet

	Apprentissage		Validation		Test		Total
	Classe 1	Classe 0	Classe 1	Classe 0	Classe 1	Classe 0	
N	78	91	17	21	22	16	245
Ncorr	65	75	12	19	13	13	197
Nerr	9	9	4	2	7	3	34
Nrejet	4	7	1	0	2	0	14

3.6.1 Analyse des erreurs

3.6.1.1 Classe de rejet

Pour effectuer l'attribution des classes à la sortie du réseau, un seuil optimal est déterminé à partir de la série d'apprentissage. L'échantillon sera affecté à la classe 1 si la sortie du réseau est supérieure à ce seuil et, inversement, il sera affecté à la classe 0 si la sortie est inférieure à ce seuil. Cette méthode est pratique si la finalité du travail est de classer tous les échantillons. Il y a cependant toujours une incertitude concernant les échantillons qui se trouvent à la limite de ce seuil.

Une autre solution consiste à créer une classe intermédiaire de rejet qui peut permettre de mieux interpréter les décisions du réseau et d'isoler les échantillons sur lesquels il y a une forte incertitude. En effet, dans les problèmes de classification à 2 classes, la valeur de sortie du réseau peut être interprétée comme la probabilité d'appartenance à la classe 1 (Statsoft 2007). La classification a ainsi été testée en créant une classe de rejet dans le but d'évaluer si l'origine de ces erreurs était due à une incertitude. Les classes ont été définies telles que :

$$\text{Si } a_s > 0.8 \rightarrow \text{classe 1}$$
$$\text{Si } a_s < 0.2 \rightarrow \text{classe 0}$$
$$\text{Si } 0.2 \leq a_s \leq 0.8 \rightarrow \text{classe de rejet}$$

où a_s est le signal du neurone de sortie.

Le Tableau 3-5 présente un exemple typique de résultats de classification avec classe de rejet. Sur les 245 échantillons, 14 ont été placés dans la classe de rejet, ce qui signifie qu'il y a une incertitude de classement pour ces échantillons. Les concentrations en pesticides des échantillons de classe 1 placés dans la classe de rejet ne sont pas particulièrement faibles par rapport à la moyenne. Leur présence dans la classe de rejet ne peut donc pas être interprétée comme une classe intermédiaire en terme de contamination.

Les 32 échantillons systématiquement mal classés sont toujours présents dans les erreurs. On remarque également qu'à l'exception de deux autres échantillons, ce sont les seuls qui sont mal

classés (Nerr = 34). Les apprentissages réalisés sans ces échantillons fournissent des résultats compris entre 90 % et 100 % de bons classements pour les trois séries.

La présence de ces erreurs systématiques peut avoir plusieurs origines. Tout d'abord, il peut s'agir d'un comportement atypique des ouvrages de captage. Les échantillons de la classe 1 placés dans la classe 0 peuvent être soumis par exemple à des pollutions ponctuelles, ou à une mauvaise isolation de l'ouvrage. À l'inverse, les puits de classe 0 placés en classe 1 peuvent effectivement présenter un risque de contamination par les pesticides sans que ceux-ci n'aient été appliqués à proximité. Mais il peut également s'agir d'un problème dû à l'apprentissage où à la présence des trois sites de caractéristiques contrastées dans le même réseau.

3.6.1.2 Effet des trois sites

La présence des trois sites très différents dans le même réseau pourrait créer un biais dans l'apprentissage. En effet, ces trois sites présentent des caractéristiques contrastées tant au niveau des variables d'entrée qu'au niveau des contaminations en pesticides. Le site de Carpentras ne présente que 18 % de cas de contamination tandis que le site de Portneuf ne présente à l'inverse, que 20 % de cas de non contamination. Lorsque l'on analyse la répartition des 32 erreurs systématiques dans les trois sites (Tableau 3-6) on remarque que sur les 18 échantillons positifs de Carpentras-Valréas, 14 sont systématiquement mal classés ce qui représente près de 78 % des échantillons de classe 1.

Tableau 3-6. Répartition des 32 erreurs systématiques sur les trois sites

	Classe 1			Classe 0		
	N	Nerr	%	N	Nerr	%
Valence	59	4	6.8	36	10	27.8
Carpentras	18	14	77.8	82	2	2.4
Portneuf	40	0	0.0	10	2	20.0
Total	117	18	15.4	128	14	10.9

À l'inverse sur le site de Portneuf, les seuls échantillons systématiquement mal classés sont deux échantillons qui ne présentent pas de pesticides. Le réseau aurait donc tendance à classer tous les échantillons de Carpentras dans la classe 0 et tous ceux de Portneuf dans la classe 1. On peut

alors se demander si le réseau n'aurait pas tendance à reconnaître les caractéristiques chimiques du site et à lui attribuer la classe majoritaire.

Le nombre d'échantillons et la différence d'échantillons dans chaque classe pour les sites de Carpentras et de Portneuf ne permettent pas un apprentissage optimal du réseau sur ces seules données. Le réseau appliqué sur les seules données de Valence où les classes sont mieux réparties donne de bonnes performances (de l'ordre de 80 % pour les trois séries). Parmi les erreurs, on retrouve systématiquement 9 des 14 échantillons présents lorsque les trois sites sont utilisés. Cela pourrait confirmer que la présence de ces erreurs systématiques provient également d'un comportement atypique des ouvrages en question.

3.7 Conclusion partielle

Les réseaux de neurones ont été appliqués sur les données des trois sites d'étude pour la classification. On a ainsi testé la capacité de certaines données chimiques à détecter l'occurrence des pesticides dans l'eau souterraine.

La sélection des variables d'entrée a montré que l'ensemble de variables le plus pertinent pour cette étude est celui constitué des nitrates, des chlorures, du carbone organique dissous, du pH et de la température.

Le réseau présentant les meilleures performances est un perceptron multicouches à une couche cachée de trois neurones. L'algorithme utilisé est celui de rétropropagation du gradient et toutes les fonctions de transfert sont des fonctions logistiques.

Ce réseau présente de bonnes performances, de l'ordre de 80 % de bons classements et les valeurs des AUC sont en moyenne de 0.87. Ces résultats montrent la pertinence de l'utilisation de variables chimiques dans la modélisation neuronale de l'occurrence des pesticides.

L'analyse des erreurs de classement a révélé que 32 échantillons étaient systématiquement mal classés. Plusieurs raisons peuvent être à l'origine de ces erreurs systématiques. Tout d'abord, ces échantillons peuvent présenter un comportement atypique par rapport aux autres. Certains échantillons avec détection de pesticides placés par le réseau dans la classe 0 peuvent par exemple avoir subi des pollutions accidentelles dues à de mauvaises manipulations ou à des ouvrages mal construits. Les échantillons sans détection de pesticides placés par le réseau dans la classe 1 peuvent présenter un réel risque de contamination sans que celle-ci n'ait eu lieu.

Cependant, ces erreurs peuvent également être dues à un possible biais dans l'apprentissage. La présence de trois populations distinctes aux caractéristiques différentes pourrait dérégler l'apprentissage en faisant intervenir ces grandes différences. Cependant, la répartition et le nombre de données ne permettaient pas d'appliquer les réseaux de façon indépendante sur les trois sites.

Chapitre 4 Approche temporelle

Le quatrième chapitre traite de l'application des réseaux de neurones sur des données issues d'un suivi temporel. Ce jeu de donnée va nous permettre d'évaluer deux autres approches. Dans une première partie, nous verrons si la prise en compte de l'historique de contamination d'un ouvrage permet d'améliorer les performances de classification. La seconde partie tentera d'évaluer si les variations de contamination dans un ouvrage peuvent être détectées.

4.1 Introduction

Le chapitre précédent a permis de mettre en évidence la pertinence de certaines variables chimiques dans la modélisation de l'occurrence des pesticides. Cette approche était basée sur un échantillonnage unique dans le temps. Or les concentrations en pesticides sont très variables dans le temps et les échantillons peuvent présenter des niveaux de concentration variables selon les dates de prélèvement. La Figure 4-1 illustre un exemple de variations de concentration sur deux ouvrages suivis par l'Agence de l'Eau Rhône-Méditerranée et Corse (RMC) entre 2001 et 2006. On remarque par exemple que le forage n°1, qui présente régulièrement des concentrations détectables en pesticides, n'en présente pas durant l'été 2004. Ceci peut être dû à des changements de pratiques ou à des conditions climatiques particulières ayant prévalu cette année-là. Le forage n°2 ne présente pas de concentration quantifiable en août 2004 alors que celles-ci sont quantifiées à 0.2 µg/l le mois suivant. Ceci illustre bien que les variations de concentrations en pesticides sont parfois rapides et que, lors d'un échantillonnage unique, la détection peut s'avérer négative alors que les ouvrages peuvent présenter un risque élevé de contamination.

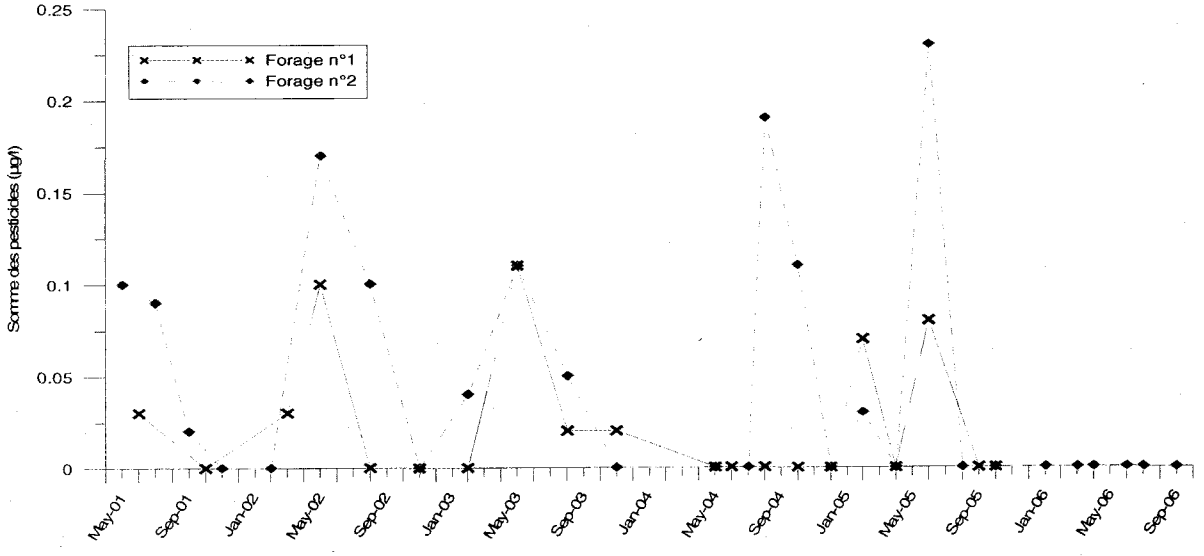


Figure 4-1. Suivi temporel de la contamination dans deux forages

À l'inverse, certains ouvrages peuvent présenter une contamination exceptionnelle de faible durée. Une approche spatiale à une date donnée peut donc ne pas être révélatrice du potentiel réel de contamination des ouvrages échantillonnés. Cela peut expliquer en partie les cas d'erreurs de la méthode identifiés au chapitre précédent.

Nous avons donc décidé de prendre en compte cette variation temporelle et d'évaluer les performances des réseaux de neurones sur des données temporelles.

4.2 Présentation des données

Les données utilisées pour cette approche sont issues du suivi de qualité des eaux souterraines de l'Agence de l'Eau Rhône-Méditerranée et Corse. Parmi l'ensemble des ouvrages suivis, 102 ont été retenus. Ces ouvrages présentent tous des analyses physico-chimiques et de pesticides aux mêmes dates. Chaque ouvrage présente entre 4 et 30 dates de prélèvement réparties entre 1995 et 2006, ce qui représente un total de 2101 analyses. Les ouvrages sélectionnés sont répartis sur l'ensemble du bassin Rhône-Méditerranée et s'étendent sur cinq régions : Bourgogne, Franche-Comté, Rhône-Alpes, Provence-Alpes-Côte-d'Azur et Languedoc-Roussillon (Figure 4-2).

L'étendue de la répartition des données permet d'avoir des conditions géologiques très variées sans pour autant avoir l'inconvénient de l'effet site de la première approche. Les ouvrages forment ainsi un continuum sur une superficie de 121 000 km^2. Le Tableau 4-1 présente la répartition des ouvrages dans les différentes régions ainsi que la fréquence moyenne de détection.

Tableau 4-1. Répartition et fréquence de détection des 102 ouvrages

	Nombre d'ouvrages			Fréquence moyenne
Régions	total	0% détection	100% détection	de détection
Bourgogne	9	0	2	0.65
Franche-Comté	16	3	1	0.28
Languedoc-Roussillon	22	5	1	0.47
Paca	22	10	1	0.32
Rhône-Alpes	33	3	0	0.46

Sur les 102 ouvrages analysés, 5 présentent des détections de pesticides dans 100 % des prélèvements. La moyenne de fréquence de détection tous ouvrages confondus est de 40 %. Sur toutes les analyses, la concentration maximale de la somme des pesticides est de 7.6 µg/l pour un puits du Vaucluse. La moyenne de la somme des pesticides est de 0.2 µg/l et la médiane est de 0.02 µg/l.

Deux approches de classification ont été testées avec ce jeu de données.

- La première approche vise à évaluer le potentiel de contamination pour un ouvrage donné en tenant compte de son historique de contamination.
- La seconde approche vise à évaluer le potentiel des réseaux à classer les échantillons en fonction de la détection ou non de pesticides, c'est-à-dire à déterminer si les variables chimiques peuvent permettre de repérer la détection de pesticides aux différentes dates de prélèvements.

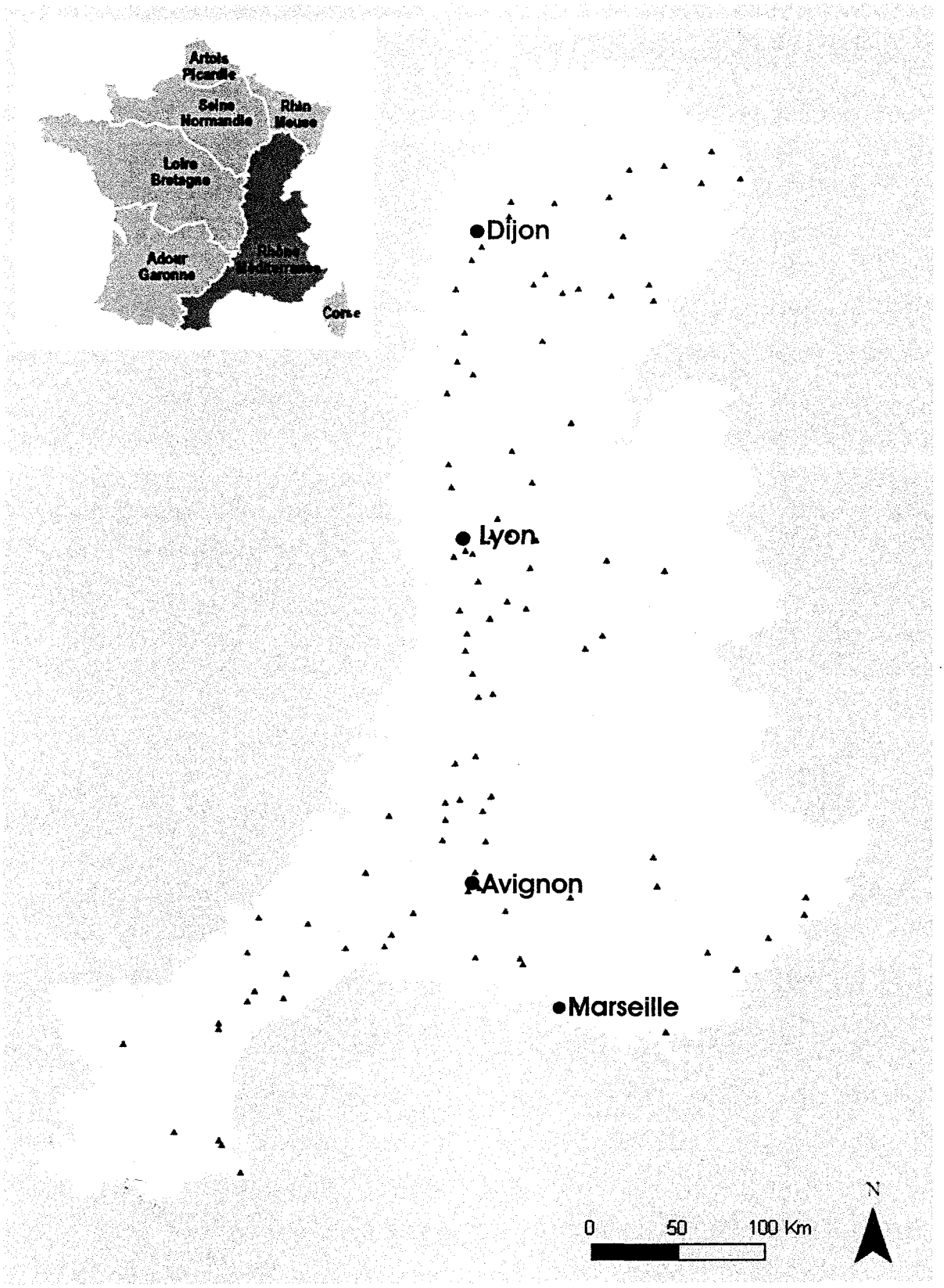


Figure 4-2. Localisation des 102 forages sélectionnés sur le bassin Rhône-Méditerranée

4.3 Etude sur la classification du potentiel de détection des ouvrages

4.3.1 Présentation de l'approche

4.3.1.1 Objectif de l'approche

Nous avons montré au chapitre précédent que les réseaux de neurones pouvaient permettre de classer correctement les échantillons présentant des détections de pesticide avec une performance de l'ordre de 80 %. Cependant, le caractère atemporel de cette méthode ne permet pas de mettre en évidence le risque de contamination global d'un ouvrage.

Dans l'approche qui va suivre, nous avons décidé de prendre en compte l'historique de contamination des ouvrages afin de réellement refléter le risque de contamination. Cette approche moins stricte que la première permettra ainsi de négliger des contaminations exceptionnelles et, à l'inverse, de prendre en compte le facteur risque pour les ouvrages dont les concentrations en pesticides fluctuent rapidement, pouvant ainsi s'avérer négatives à une date donnée. Cette approche va ainsi permettre :

- de valider sur des données nouvelles la capacité des réseaux de neurones à classer les échantillons présentant un risque de contamination par les pesticides à partir de données chimiques.
- d'évaluer si la prise en compte de l'historique de contamination permet d'améliorer les performances du réseau.

4.3.1.2 Classification utilisée

Afin de prendre en compte l'historique de contamination des ouvrages, la classification a été basée sur la fréquence de détection des pesticides, c'est-à-dire pour un ouvrage donné, le nombre de prélèvements où les pesticides ont été détectés rapporté au nombre total de prélèvements. Deux classes ont été créées en fonction de la fréquence de détection :

- Classe HF (haute fréquence) : les ouvrages présentant 25 % ou plus de détection de pesticides.

- Classe BF (basse fréquence) : les ouvrages présentant moins de 25 % de cas de détection.

Le seuil de 25 % a été choisi arbitrairement comme étant relativement significatif en terme de risque de contamination. Sur les 102 ouvrages, 63 appartiennent à la classe HF et 39 à la classe BF. Pour chaque ouvrage, une date unique de prélèvement a été choisie aléatoirement et la classe correspondante à l'ouvrage (HF ou BF) lui a été attribuée.

4.3.2 Sélection des variables

Les données de l'agence de l'eau possédant davantage de paramètres physico-chimiques mesurés, une nouvelle procédure de sélection des variables a été effectuée pour cette approche parmi les 19 variables ainsi disponibles. Certaines d'entre elles pourraient apporter de nouvelles informations et permettre d'améliorer les performances du réseau.

La sélection des variables a été effectuée selon la même méthode que dans le chapitre précédent. Afin de sélectionner le meilleur sous-ensemble de variables, six combinaisons différentes ont ainsi été comparées :

- les contaminants seuls ;
- les variables sélectionnées pour l'approche par site ;
- la méthode Stepdisc ;
- l'analyse de sensibilité ;
- l'algorithme génétique ;
- la présence des 19 variables.

Les différentes méthodes sont présentées dans le chapitre précédent (3.4.2). Sept variables ont été sélectionnées pour la méthode Stepdisc (Tableau 4-3), 10 variables pour l'analyse de sensibilité (Figure 4-1) et 11 variables par l'algorithme génétique.

Cinq variables sont communes aux trois méthodes. Il s'agit des nitrates, du pH, de la dureté, de l'oxygène dissous et de la silice. Le magnésium a été sélectionné par la méthode stepdisc et par

l'algorithme génétique, tandis que chlorures, température et calcium ont été sélectionnés par l'analyse de sensibilité et par l'algorithme génétique (Tableau 4-2).

Les résultats fournis par la ***Figure 4-4*** représentent les performances moyennes et l'erreur réalisée pour 50 apprentissages. Parmi ces apprentissages, 5 distributions aléatoires des données dans les trois séries ont été appliquées (3.4.3.1.).

La méthode stepdisc et la présence des 19 variables sont deux sous-ensembles qui présentent des performances nettement plus faibles que les autres. Ceci révèle bien l'importance de d'une sélection de variables et le fait que la présence de certaines variables non pertinentes ou l'absence de certaines variables clés puisse fortement affecter les performances du réseau. La méthode stepdisc, par exemple, est la seule à avoir sélectionné les bicarbonates. Ces derniers présentent cependant une sensibilité assez faible (Figure 4-3) par rapport aux autres variables. Leur présence dans la sélection pourrait donc être responsable de la performance plus faible de la sélection par la méthode stepdisc. On peut également noter que les chlorures sont absents de cette sélection alors qu'ils ont été sélectionnés par toutes les autres méthodes.

La sélection de l'algorithme génétique est celle qui présente les meilleures performances et le moins de variation. À l'exception du carbone organique dissous, les variables sélectionnées pour la modélisation des sites d'études se retrouvent dans cette sélection. Sept autres variables y ont été ajoutées. Parmi ces variables, on retrouve la dureté, le calcium et le magnésium. Ces trois variables sont très liées entre elles puisque la dureté de l'eau correspond à la présence de sels de calcium et de magnésium. Elle est directement liée à la nature géologique des terrains traversés. Le calcium et le magnésium associés aux sulfates peuvent également être utilisés en agriculture comme fertilisant ou amendement. Il est possible également, mais dans une mesure non quantifiable ici, que l'infiltration des sulfates (d'origine atmosphérique ou terrestre) en profondeur dans le sol s'accompagne du lessivage concomittant de cations bivalents (Ca^{2+}, Mg^{2+}) requis pour assurer l'électroneutralité de la solution de sol transitant vers la zone saturée. Ceci est observable notamment en présence de précipitations acides (désaturation du complexe d'échange du sol en cations basiques, ainsi lessivés, en présence d'acidité). L'oxygène dissous, également présent dans la sélection, a déjà été mis en relation avec les pesticides dans des études précédentes (Burrow *et al.* 1998). Sa concentration était significativement supérieure dans les échantillons qui présentaient au moins une détection de pesticides. Il se peut alors que les

caractéristiques d'un site qui favorisent de fortes concentrations en oxygène dissous (tel que la présence de sédiments à grains grossiers avec une faible teneur en matière organique) favorisent également celles des pesticides par la présence d'infiltration rapide et de temps de séjour plus courts. Les fluorures n'ont été sélectionnés que par l'algorithme génétique. Ils ont tendance à être plus élevés dans les nappes captives et peuvent donc apporter une information sur le degré de protection de l'aquifère.

Il semblerait donc, qu'en plus des variables indiquant un effet de l'agriculture sur la qualité de l'eau, la présence de caractéristiques renseignant sur la géologie ou sur le temps de séjour de l'eau permettrait d'améliorer les performances du réseau. Le sous-ensemble des 11 variables sélectionnées par l'algorithme génétique sera donc conservé pour la suite de l'étude. Ces 11 variables sont les chlorures, les nitrates, les sulfates, la température, le pH, la dureté de l'eau, le calcium, le magnésium, les fluorures, l'oxygène dissous et la silice (forme SiO_2).

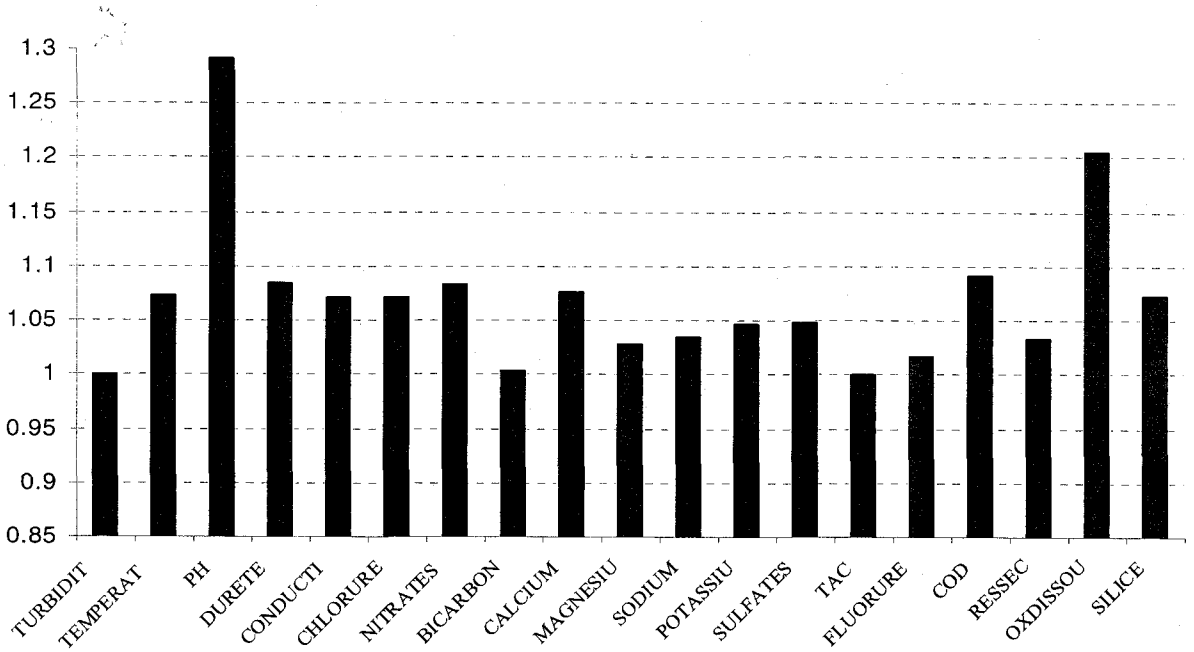


Figure 4-3. Analyse de sensibilité des variables d'entrée du réseau pour la classification sur le potentiel de contamination des ouvrages

Tableau 4-2. Scénarios utilisés pour la sélection des variables d'entrée pour la classification sur le potentiel de contamination des ouvrages

		Variables sélectionnées																		
Méthode	n	Cl^-	NO_3^-	SO_4^{2-}	CE	COD	T	pH	Turb.	Dureté	HCO_3^-	Ca_2^+	Mg_2^+	Na^+	K^+	TAC	F^-	R.Sec	O_2	SiO_2
Aucune	19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Contaminants	4	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sites	5	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Stepdisc	7	0	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	1	1
Sensibilité	10	1	1	0	1	1	1	1	0	1	0	1	0	0	0	0	0	0	1	1
Génétique	12	1	1	1	0	0	1	1	0	1	0	1	1	0	0	0	1	0	1	1

CE = conductivité électrique; Rsec = résidu sec; TAC = Titre alcalimétrique complet

Tableau 4-3. Résultats de la statistique du lambda de Wilks pour la classification sur le potentiel de contamination des ouvrages. Les variables en gras correspondent à celles sélectionnées au seuil 5

rang	Variable	Λ de Wilks	F
1	**NO_3^-**	**0.928**	**7.77**
2	**pH**	**0.835**	**11.04**
3	**SiO2**	**0.802**	**3.98**
4	**O_2**	**0.733**	**9.16**
5	**Mg^{2+}**	**0.683**	**7.05**
6	**Dureté**	**0.672**	**1.53**
7	**$HCO3^-$**	**0.658**	**2.07**
8	CE	0.649	1.18
9	Cl^-	0.642	1.01
10	Ca^{2+}	0.621	3.12
11	Rsec	0.608	1.85
12	Na^{2+}	0.605	0.56
13	$SO4^{2-}$	0.602	0.33
14	TAC	0.581	3.24
15	F^-	0.566	2.22
16	T	0.560	0.89
17	K^+	0.555	0.73
18	COD	0.553	0.44
19	Turbidité	0.552	0.08

CE = conductivité électrique; Rsec = résidu sec;
TAC = Titre alcalimétrique complet

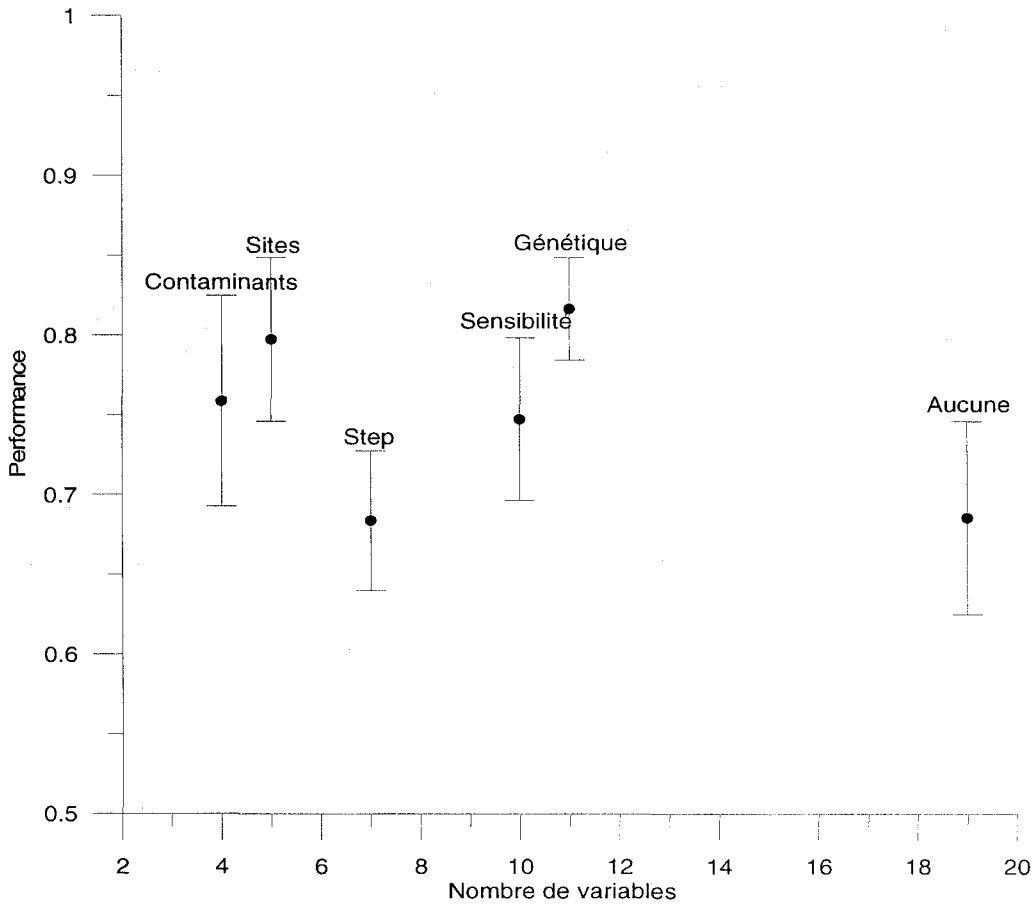


Figure 4-4. Comparaison des performances de différentes méthodes de sélection de variables vs nombre de variables pour la classification sur le potentiel de contamination des ouvrages

4.3.3 Architecture du réseau

Une procédure identique à celle utilisée au chapitre précédent a été appliquée pour ce jeu de données. Dix réseaux à une couche cachée de 2 à 11 neurones ont été testés afin de définir la meilleure architecture. Les résultats présentés dans le Tableau 4-4 correspondent aux moyennes des RMSE et des performances (P %) observés sur les 20 apprentissages pour les trois séries de données. L'aire sous la courbe ROC (AUC) correspond aux séries de validation et de test.

Tableau 4-4. Performances des réseaux avec différentes architectures pour la classification sur le potentiel de contamination des ouvrages

Réseau	Apprentissage		Validation		Test		AUC
	RMSE	P (%)	RMSE	P (%)	RMSE	P (%)	
11-2-1	0.341	83.15	0.289	90.738	0.387	74.83	0.88
11-3-1	0.322	85.92	0.281	89.05	0.395	75.95	0.89
11-4-1	0.316	85.50	0.277	88.57	0.396	76.19	0.89
11-5-1	0.330	85.00	0.283	90.71	0.390	76.90	0.87
11-6-1	0.335	84.50	0.281	87.62	0.386	77.62	0.89
11-7-1	0.318	86.67	0.278	90.71	0.405	77.38	0.90
11-8-1	0.326	85.75	0.279	89.05	0.388	76.90	0.89
11-9-1	**0.332**	**85.42**	**0.279**	**90.95**	**0.383**	**79.05**	0.91
11-10-1	0.330	84.75	0.278	88.81	0.397	75.48	0.89
11-11-1	0.331	85.75	0.278	88.33	0.391	76.67	0.89

D'après ces résultats, les réseaux à 2, 3 et 4 neurones cachés sont ceux qui présentent les moins bonnes performances avec un maximum de 76.2 % de bon classement pour la série test, bien que celles de la série de validation soient de l'ordre de 89 %. Le réseau qui présente les meilleures performances est le réseau 11-9-1 avec des moyennes de 85 %, 90 % et 79 % de bon classement respectivement pour les séries d'apprentissage, de validation et de test. La différence par rapport aux autres réseaux se situe essentiellement dans la série de test. Les autres réseaux ont au maximum 77.7 % de bon classement pour cette série. Ce réseau présente également une AUC de 0.91 ce qui correspond à une très bonne capacité de classement.

Le réseau 11-9-1 semble donc être le réseau le plus approprié pour cette classification et sera donc retenu pour l'étude détaillée de classification.

4.3.4 Capacité de classification du réseau

À partir des variables et du réseau sélectionnés, les performances du réseau ont été évaluées. Les données ont été préalablement standardisées et la répartition dans les différentes séries a été faite de telle sorte que sur les 102 échantillons, 60 soient utilisées pour l'apprentissage, 21 pour la validation et 21 pour la série test. Le Tableau 4-5 présente les résultats de classification pour dix apprentissages réalisés en faisant varier la distribution des données dans les différentes séries. Les performances moyennes sont de 83.7 % pour la série de validation et 82.2 % pour la série de test. Globalement, les performances pour ces deux séries sont toujours du même ordre de grandeur, ce qui montre les capacités de généralisation du réseau. La spécificité et la sensibilité, c'est-à-dire la performance du réseau pour chacune des classes sont du même ordre de grandeur : le réseau accorde donc la même importance aux deux classes.

Le nombre total d'échantillons mal classés est en moyenne de 17. Ce nombre varie de 10 à 20 en fonction de la répartition des données dans les trois séries ce qui montre que le modèle y est tout de même assez sensible. Ceci s'explique par le fait que le modèle n'est pas capable d'extrapoler au-delà de l'ordre de grandeur des données utilisées pour l'apprentissage : il est donc indispensable que la distribution des données dans les séries d'apprentissage et de validation soit similaire.

De plus, contrairement à l'approche par site, où la majorité des échantillons mal classés l'étaient systématiquement, dans cette approche seuls cinq échantillons se retrouvent régulièrement mal classés. Ces cinq échantillons présentent des fréquences de détection comprises entre 0.1 et 0.4 et peu de dates de prélèvement. Du fait de ce faible nombre de dates de prélèvement, le calcul des fréquences de détection n'est peut-être pas suffisamment représentatif.

Dans 90 % des apprentissages, la création d'une classe de rejet avec les seuils de 0.8 et 0.2 ne change pas la classification du réseau. Cela signifie que les erreurs individuelles sont faibles donc que le modèle converge bien. Plusieurs essais ont été effectués en modifiant la date de prélèvement choisie aléatoirement pour chaque ouvrage afin de s'assurer de la représentativité des résultats. Les performances obtenues étaient toujours du même ordre de grandeur.

Tableau 4-5. Résultats de classification du réseau 11-9-1 pour dix apprentissages en faisant varier la distribution des échantillons dans les trois séries.
Nerr représente le nombre total d'échantillons mal classés

	Performance P(%)			Classification																	Nerr
	App.	Val.	Test	Apprentissage						Validation						Test					
				Classe HF		Classe BF		Se	Sp	Classe HF		Classe BF		Se	Sp	Classe HF		Classe BF		Se	Sp	
				mesuré	correct	mesuré	correct			mesuré	correct	mesuré	correct			mesuré	correct	mesuré	correct			
1	80.0	85.7	85.7	31	25	29	23	80.6	79.3	16	13	5	5	81.3	100.0	14	12	7	6	85.7	85.7	18
2	83.3	90.5	71.4	38	32	22	18	84.2	81.8	12	11	9	8	91.7	88.9	11	7	10	8	63.6	80.0	18
3	85.0	76.2	81.0	33	28	27	23	84.8	85.2	15	11	6	5	73.3	83.3	13	11	8	6	84.6	75.0	18
4	75.8	88.9	88.9	37	29	23	18	78.4	78.3	11	10	10	7	90.9	70.0	13	10	8	8	76.9	100.0	20
5	83.3	76.2	85.7	39	32	21	18	82.1	85.7	11	7	10	9	63.6	90.0	11	11	10	7	100.0	70.0	18
6	91.7	81.0	81.0	32	29	28	26	90.6	92.9	15	13	6	4	86.7	66.7	14	11	7	6	78.6	85.7	13
7	86.7	81.0	71.4	35	30	25	22	85.7	88.0	11	9	10	8	81.8	80.0	15	11	6	4	73.3	66.7	18
8	85.0	85.7	81.0	37	30	23	18	81.1	78.3	11	9	10	10	81.8	100.0	13	12	8	7	92.3	87.5	16
9	80.0	81.0	90.5	41	32	19	16	78.0	84.2	10	9	11	8	90.0	72.7	10	9	11	10	90.0	90.9	18
10	91.7	90.5	85.7	35	32	25	23	91.4	92.0	14	13	7	6	92.9	85.7	12	10	9	8	83.3	88.9	10
Moy.	84.2	83.7	82.2					83.7	84.6					83.4	83.7					82.8	83.0	17

4.3.5 Conclusion sur l'approche

Cette approche a permis de montrer sur un nouveau type de données que la classification par les réseaux de neurones à partir de données chimiques permettait de classer les ouvrages en fonction de leur risque potentiel de contamination. La prise en compte de l'historique de contamination permet de s'assurer de la meilleure représentativité des classes utilisées. Pour une même sélection de variables, l'ordre de grandeur des performances de classification est du même ordre de grandeur que dans l'approche atemporelle du chapitre précédent, ce qui nous amène à penser que les variables chimiques peuvent expliquer environ 80 % de la présence des pesticides dans les eaux souterraines captées par ces ouvrages.

La comparaison des différentes méthodes de sélection des variables montre que la présence de certaines variables qui renseignent sur la géologie ou le temps de séjour de l'eau permet d'améliorer les performances du réseau de quelques pour-cents.

4.4 Étude de la variabilité temporelle de la contamination dans les ouvrages

4.4.1 Présentation de l'approche

4.4.1.1 Objectif de l'approche

Nous avons vu dans la partie précédente que l'utilisation de variables chimiques dans les réseaux de neurones pouvait nous permettre de classer les ouvrages en fonction de leur risque de contamination. L'objectif de cette deuxième approche est d'évaluer si une démarche similaire peut nous permettre de détecter les variations de concentrations en pesticides aux différentes dates de prélèvement.

Une telle approche pourrait par exemple permettre de mieux organiser des calendriers d'échantillonnage en ciblant des dates optimales en termes de contamination pour les prélèvements. On tentera donc d'évaluer si le comportement des ouvrages les années passées permet d'appréhender celui des années suivantes et de correctement classer les échantillons.

4.4.1.2 Classification utilisée

Dans cette approche, le problème de classification est le même que celui appliqué aux trois sites d'études dans le chapitre précédent. Les échantillons sont ainsi distribués en deux classes :

- Classe 1 : échantillons dont au moins un des pesticides présente une concentration supérieure au seuil de quantification.
- Classe 0 : échantillons dont les concentrations en pesticides sont toutes inférieures au seuil de quantification.

Pour chacun des 102 ouvrages, toutes les dates de prélèvement ont été conservées. La répartition des données dans les trois séries est présentée dans le Tableau 4-6.

Tableau 4-6. Répartition des données dans les trois séries pour la classification temporelle des ouvrages

	Classe 1	Classe 0	Total
Apprentissage	602	586	1188
Validation	221	191	412
Test	215	196	411
Total	1038	973	2011

Les différentes dates de prélèvement ont été distribuées de telle sorte que les données de validation et de test correspondent aux années 2005 et 2006, et les données d'apprentissage correspondent aux années antérieures. Certains ouvrages, ne présentant pas de date de prélèvement en 2005 et 2006 ou inversement uniquement en 2005 et 2006, ont été supprimés. La répartition des données est présentée dans le Tableau 4-6.

Étant donné que l'approche est différente, la sélection des variables d'entrée et de l'architecture du réseau a été effectuée de façon indépendante pour cette approche.

4.4.2 Sélection des variables d'entrée

La sélection des variables a été effectuée selon la même méthode que dans les autres cas, c'est-à-dire une comparaison de six sous-ensembles. Les résultats détaillés de la sélection par les méthodes stepdisc et sensibilité sont présentés au *Tableau 4-8* et à la Figure 4-5. Les variables sélectionnées pour chaque sous-ensemble sont indiquées dans le Tableau 4-7.

Les nitrates et le magnésium sont les deux seules variables sélectionnées par les trois méthodes. Ce sont également celles qui présentent les plus fortes sensibilités (Figure 4-5). Les différences de performance entre les différentes méthodes sont relativement faibles (de 70 % à 74 %). D'autre part, le fait que la présence de toutes les variables montre des résultats du même ordre de grandeur que les autres méthodes indique que la sélection des variables n'a pas été optimale.

Étant donné que la répartition des données dans l'apprentissage est fixée, les variations de performances pour une même méthode sont uniquement dues au nombre d'époques. Parmi les six sous-ensembles, c'est la sélection par la méthode de sensibilité qui fournit les meilleures performances avec une moyenne de 74%.

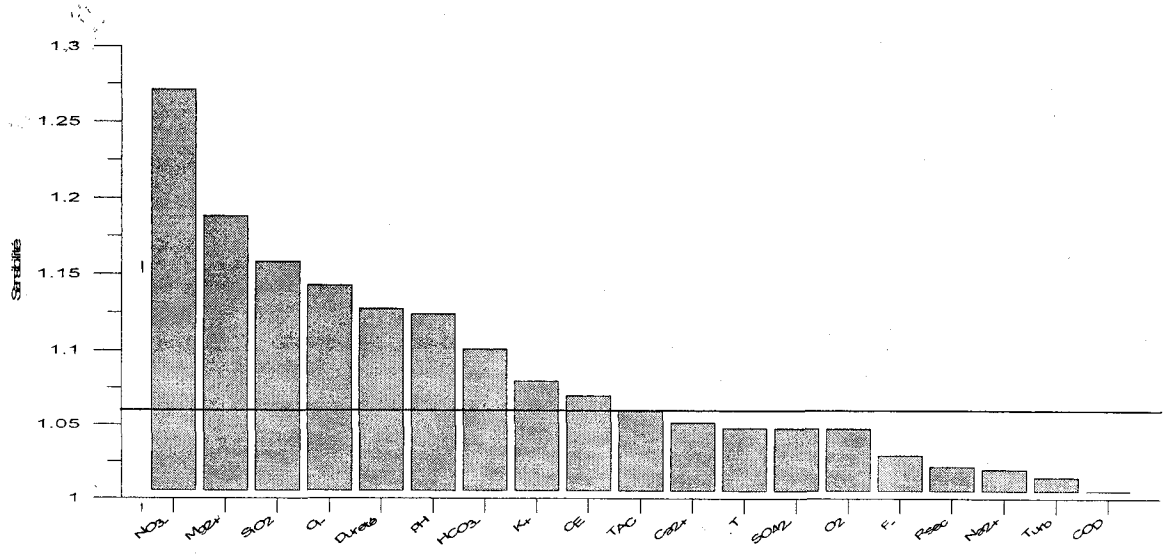


Figure 4-5. Analyse de sensibilité des variables d'entrée du réseau pour la classification temporelle des ouvrages

Tableau 4-7. Scénarios utilisés pour la sélection des variables d'entrée pour la classification temporelle des ouvrages

		Variables sélectionnées																		
Méthode	n	Cl^-	NO_3^-	SO_4^{2-}	CE	COD	T	pH	Turb.	Dureté	HCO_3^-	Ca_2^+	Mg_2^+	Na^+	K^+	TAC	F^-	R.Sec	O_2	SiO_2
Aucune	19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Contaminants	4	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sites	5	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Stepdisc	7	0	1	1	0	0	0	0	0	0	1	0	1	0	0	0	0	1	1	1
Sensibilité	8	1	1	0	0	0	0	1	0	1	1	0	1	0	1	0	0	0	0	1
Génétique	6	0	1	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	1	0

CE = conductivité électrique; Rsec = résidu sec; TAC = Titre alcalimétrique complet

Tableau 4-8 Résultats de la statistique du lambda de Wilks pour la classification temporelle des ouvrages. Les variables en gras correspondent à celles sélectionnées au seuil 5%

rang	Variable	Λ de Wilks	F
1	**Mg^{2+}**	**0.813**	**132.64**
2	**NO_3^-**	**0.768**	**33.54**
3	**$SO4^{2-}$**	**0.752**	**12.01**
4	**$HCO3^-$**	**0.732**	**15.57**
5	**Ca^{2+}**	**0.713**	**15.41**
6	**O2**	**0.704**	**7.37**
7	**Rsec**	**0.699**	**4.18**
8	SiO_2	0.694	3.72
9	Cl^-	0.690	3.34
10	Na^{2+}	0.676	11.33
11	T	0.672	3.58
12	K^+	0.670	1.74
13	CE	0.669	1.10
14	F^-	0.668	0.40
15	Turbidité	0.668	0.42
16	Dureté	0.668	0.02
17	COD	0.668	0.02
18	pH	0.668	0.01
19	TAC	0.668	0.00

CE = conductivité électrique; Rsec = résidu sec;
TAC = Titre alcalimétrique complet

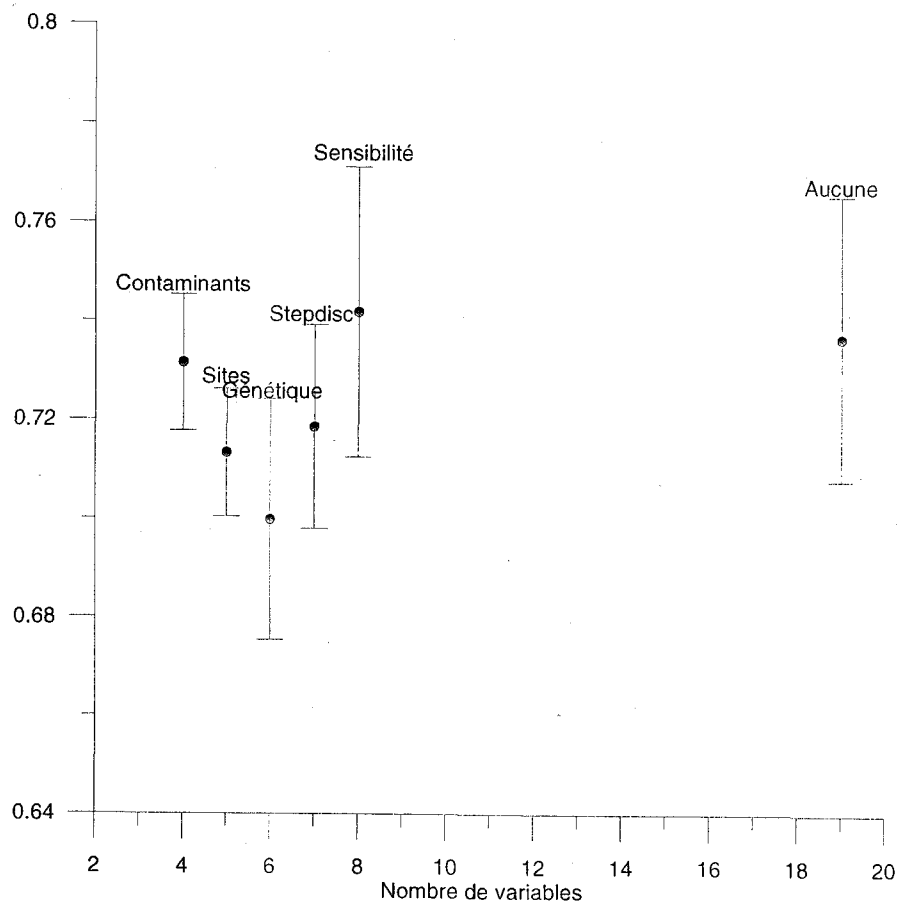


Figure 4-6 Comparaison des performances de différentes méthodes de sélection de variables vs nombre de variables pour la classification temporelle des ouvrages

4.4.3 . Architecture du réseau

L'analyse des différentes architectures pour la meilleure combinaison de variables ne permet pas d'améliorer les performances. Celles-ci restent toujours inférieures à 75 % bien que la configuration 8-7-1 semble donner de meilleurs résultats.

Tableau 4-9. Performance des réseaux avec différentes architectures pour la classification temporelle des ouvrages

Réseau	Apprentissage		Validation		Test		AUC
	RMSE	P (%)	RMSE	P (%)	RMSE	P (%)	
8-2-1	0.427	72.43	0.438	68.68	0.438	69.53	0.68
8-3-1	0.418	73.42	0.431	70.79	0.434	70.34	0.70
8-4-1	0.410	75.43	0.427	71.86	0.431	72.08	0.71
8-5-1	0.409	75.80	0.425	72.46	0.431	72.09	0.71
8-6-1	0.412	75.35	0.429	71.75	0.432	71.33	0.70
8-7-1	0.407	76.04	0.423	73.04	0.430	72.59	0.71
8-8-1	0.412	75.24	0.428	71.77	0.433	71.28	0.70

4.4.4 Capacité de classification du réseau

Les performances du réseau pour cette classification sont faibles à moyennes avec des performances moyennes entre 68 % et 72 % respectivement pour les séries de validation et de test. On remarque également que même sur les données d'apprentissage les performances sont toujours inférieures à 76 %, et cela, quelles que soient les variables d'entrée ou l'architecture du réseau. Les AUC pour cette approche sont de l'ordre de 0.70 c'est-à-dire relativement faibles par rapport aux autres approches où elles étaient supérieures à 0.85. D'autre part, les résultats sont plutôt instables et très variables en fonction du nombre d'itérations, ce qui fait chuter les moyennes.

Plusieurs phénomènes peuvent être mis en cause dans le fait que les performances sont faibles. Il se peut tout d'abord que les variables chimiques ne permettent pas de détecter les variations temporelles en terme de contamination des pesticides : cependant la méthodologie appliquée n'est peut-être pas optimale avec une fois encore la présence de sites différents ou une mauvaise sélection des variables d'entrée.

4.4.4.1 Effet site

L'analyse détaillée des résultats de classification révèle que dans la majorité des cas, une classe unique est attribuée pour un ouvrage donné et cela, quelle que soit la date de prélèvement. Les 25 % d'erreur correspondent donc au pourcentage moyen de variabilité des ouvrages. On se retrouve finalement dans la même configuration que lorsque les trois sites d'étude échantillonnés étaient utilisés ensemble, c'est-à-dire que la variabilité entre les différents ouvrages est supérieure à celle qui est propre à un ouvrage donné. L'apprentissage a donc plutôt tendance à opposer les différents puits, à reconnaître leurs caractéristiques chimiques et à leur attribuer la classe majoritaire présente dans l'apprentissage.

4.4.4.2 Sélection des variables

La sélection des variables quant à elle n'a peut-être pas été optimale pour le problème de classification posé. Les variables sélectionnées favorisent peut-être davantage la reconnaissance des différents ouvrages que leur propre variabilité. Plusieurs variables apportent en effet des informations sur la géologie ou sur les temps de séjour de l'eau. Ces informations sont plutôt caractéristiques de l'ouvrage et risquent alors de masquer les possibles variations temporelles pour un ouvrage donné.

Un suivi temporel sur un seul ouvrage présentant suffisamment de dates de prélèvements pourrait alors permettre de faire une sélection optimale des variables d'entrée et de réellement évaluer la capacité des réseaux à détecter ces variations temporelles. N'ayant au maximum que 32 dates de prélèvement pour un même ouvrage, une telle démarche n'a pas pu être réalisée. Cependant, afin de tester une approche similaire, deux forages d'un même aquifère ont été sélectionnés.

4.4.5 Réduction de la méthode sur deux ouvrages similaires

Les faibles performances pour la classification de la variabilité temporelle des ouvrages peuvent être attribuées à la présence de sites différents et d'une sélection inappropriée ou non adaptée des variables inhérentes. Afin de réévaluer l'approche sur des données ne présentant pas de caractéristiques géochimiques trop différentes, nous avons cherché à sélectionner des ouvrages ayant les caractéristiques suivantes :

- doivent appartenir à la même unité aquifère ;
- doivent présenter des concentrations en pesticides variables selon les dates de prélèvement afin d'équilibrer les classes.

En respectant ces conditions, seuls deux forages ont pu être sélectionnés. Les données cumulées de ces deux ouvrages nous permettent d'obtenir 51 échantillons. La répartition des classes dans les deux ouvrages est présentée dans le Tableau 4-10

Tableau 4-10. Répartition des classes dans les deux ouvrages F1 et F2

	n	Classe 0	Classe 1
F1	20	5	15
F2	31	10	21
	51	15	36

Les analyses de sensibilité montrent que le pH, les nitrates et l'oxygène dissous sont les variables qui ont le plus d'influence sur les performances du réseau.

Les tests de classification ont été effectués avec les variables sites auxquelles a été ajouté l'oxygène dissous qui présente une forte sensibilité et une architecture 6-4-1. Le Tableau 4-11 présente les résultats pour trois apprentissages effectués en faisant varier la distribution des données dans les trois séries.

Tableau 4-11. Résultats de classification pour un réseau 6-4-1 sur les données groupées de deux forages

	Performance P(%)			Classification												Nerr	
	App.	Val.	Test	Apprentissage				Validation				Test				F1	F2
				Classe 1		Classe 0		Classe 1		Classe 0		Classe 1		Classe 0			
				mesuré	correct	mesuré	correct	mesuré	correct	mesuré	correct	mesuré	correct	mesuré	correct		
1	92.3	100.0	83.3	17	16	9	8	10	10	3	3	9	7	3	3	2	2
2	84.6	84.6	91.7	20	16	6	6	9	8	4	3	7	7	5	4	3	4
3	92.3	92.3	75.0	19	19	7	5	8	8	5	4	9	6	3	3	3	3

Pour les trois séries confondues, les performances sont comprises entre 75 % et 100 %. Le nombre d'échantillons mal classés est au maximum de sept ce qui correspond à une performance de 86 %. Le détail des erreurs montre que celles-ci sont bien réparties sur les deux forages et sur

les deux classes. Ceci montre qu'avec une telle approche, le réseau n'a pas tendance à affecter une même classe à tous les prélèvements d'un ouvrage.

De tels résultats laissent envisager que les données chimiques pourraient permettre de déceler les variations temporelles en terme de détection de pesticide. Cependant, nos données ne présentant que 51 échantillons, la représentativité de la méthode est limitée et nécessite d'être validée avec un jeu de données présentant davantage de dates de prélèvement.

4.4.6 Conclusion sur l'approche

Cette approche avait pour objectif d'évaluer la capacité des réseaux de neurones à détecter les variations temporelles en termes de détection de pesticides à partir de données chimiques. Les résultats sur l'ensemble des données présentent des performances moyennes (< 75 %), mais l'effet site peut être mis en cause. La même approche sur deux ouvrages similaires donne de très bonnes performances.

4.5 Conclusion partielle

Dans ce chapitre nous avons utilisé un jeu de données temporelles afin de tester la capacité des réseaux de neurones à classer les échantillons selon leur détection en pesticides à partir de données chimiques. Deux approches ont été testées. Tout d'abord une approche qui permet d'évaluer le risque potentiel de contamination des ouvrages en tenant compte de l'historique de contamination des ouvrages. Une seconde approche avait pour objectif d'évaluer les variations temporelles de contamination des ouvrages.

La première approche présente de bonnes performances avec une moyenne de 83 % de bons classements pour l'ensemble des séries de validation et de test. De tels résultats confirment que l'utilisation de variables chimiques est pertinente dans la modélisation neuronale du risque de contamination par les pesticides et permettrait de rendre compte d'environ 80 % des variabilités. Cette approche permet également de montrer que la présence de certaines variables qui renseignent sur la géologie locale permet d'améliorer les performances du réseau.

La seconde approche présente des résultats faibles à moyens avec des performances inférieures à 75 %, ce qui pourrait laisser croire que les variables chimiques ne permettent pas de rendre compte des variabilités temporelles en termes de détection de pesticides. Cependant, la même approche réalisée sur deux ouvrages similaires donne de très bonnes performances et nous amène à penser que ce serait davantage la diversité géographique et géologique des ouvrages qui seraient la cause des faibles performances. Une telle approche nécessite donc d'être réévaluée à une échelle plus fine où les caractéristiques géochimiques sont similaires.

Chapitre 5 Discussion et conclusion générales

Cette thèse a pour objectifs de répondre essentiellement à deux points. Le premier objectif est de confirmer ou d'infirmer la présence d'une relation entre la détection de pesticides et certains paramètres chimiques de l'eau. Le deuxième objectif est d'évaluer si ces paramètres chimiques peuvent être utilisés afin d'évaluer le potentiel de contamination par les pesticides des eaux souterraines par modélisation neuronale.

5.1 Utilité des variables chimiques dans l'évaluation du risque de contamination des eaux souterraines par les pesticides

La composition chimique de l'eau souterraine résulte de la combinaison de la composition de l'eau qui entre dans le réservoir et des réactions avec les minéraux des roches. L'occupation du sol, par le biais des intrants, a donc une influence sur la composition et la qualité chimique de l'eau souterraine. Ainsi, en milieu agricole, certains composés se retrouvent-ils dans des concentrations supérieures à ce qu'ils seraient sans l'activité agricole. Les nitrates, les chlorures et les sulfates sont les composés les plus souvent associés à des impacts d'origine agricole (Pionke 1985). Ces composés sont en effet souvent présents dans les engrais ou fertilisants et migrent jusqu'à la nappe, modifiant ainsi sa composition chimique naturelle. Les eaux souterraines vulnérables peuvent ainsi présenter une signature chimique particulière (Pacheco et van der Weijden 1996). Cependant, les différents composés ne réagissent pas tous de la même façon, et les modes de transfert varient en fonction de leur solubilité et de leurs interactions avec le milieu environnant.

Notre étude portant sur les pesticides, c'est leur co-occurrence avec ces différents autres composés chimiques qui nous intéresse. La présence de corrélation entre les pesticides et différents autres contaminants agricoles peut donc être d'une aide précieuse. Ces contaminants sont en effet facilement mesurables, et sont donc des données accessibles et plus systématiquement recherchées dans les analyses courantes de suivi de la qualité de l'eau. De telles corrélations, si elles existent, pourraient ainsi permettre de mieux prévoir le risque de contamination de l'eau souterraine par les pesticides.

Dans la littérature, plusieurs études traitant de l'impact de l'agriculture sur les eaux souterraines ont mis en évidence, de façon éparse et parfois contradictoire, une possible relation entre les pesticides et certains contaminants d'origine agricole. Certains auteurs ont cependant infirmé ces relations (Dawson 2001). Ces comparaisons ont porté sur les concentrations en pesticides, parfois sur le nombre de composés détectés. Chacune de ces études a cependant été réalisée pour un site donné et le nombre d'échantillons ne permettait pas forcément de généraliser les résultats. Il était donc utile de réétudier ces relations dans une étude plus vaste comportant d'une part suffisamment de données et d'autre part différents sites en vue d'une comparaison.

Notre étude a donc bénéficié de trois sites d'échantillonnage présentant des caractéristiques climatiques, culturales et hydrogéologiques différentes afin d'évaluer si des tendances se retrouvaient sur chacun des sites. Les études de corrélation révèlent que pour chacun des trois sites, les nitrates, les chlorures et les sulfates sont significativement corrélés avec la somme des concentrations des pesticides dosés. Ces corrélations sont cependant relativement faibles ($0.29 < R\ Spearman < 0.56$). Ceci est entre autres dû au fait que les chlorures et les sulfates ont également des origines naturelles, leur présence n'indique donc pas forcément une contamination d'origine agricole. D'autre part, les variations de concentrations en pesticides sont faibles, d'où l'intérêt d'une classification binaire. Les tests de Mann Whitney ont révélé que les concentrations en nitrates, sulfates et chlorures étaient significativement différentes dans les échantillons avec et sans détection de pesticides. Cette étude confirme donc que pour trois sites présentant des caractéristiques agropédoclimatiques différentes, les échantillons présentant des détections de pesticides sont associés à de fortes concentrations en nitrates, en chlorures et en sulfates.

Indépendamment, le pouvoir discriminant de chaque variable est moyen ($AUC \sim 0.70$) du fait de la présence de ces contaminants dans des concentrations variables pour les échantillons ne présentant pas de détection de pesticides. Cependant, l'intérêt de l'étude se situe dans la combinaison de ces variables chimiques. La co-occurrence de ces composés peut ainsi traduire une vulnérabilité globale de l'aquifère qui, associée à d'autres caractéristiques chimiques, peut permettre de mieux évaluer le risque de contamination par les pesticides.

L'application de la sélection des variables par les réseaux de neurones a mis en évidence que d'autres variables chimiques permettent d'améliorer l'évaluation du risque de contamination des eaux souterraines par les pesticides. Ainsi, les contaminants agricoles potentiels (chlorures, nitrates et sulfates) associés au pH, à la température, à la dureté de l'eau, au calcium, au magnésium et à l'oxygène dissous sont les variables qui, associées les unes aux autres, semblent permettre la meilleure discrimination possible des échantillons avec et sans détection de pesticides. Ces variables caractérisent un équilibre hydrochimique et les variations au sein de cet équilibre semblent permettre d'évaluer un risque de contamination par les pesticides.

5.2 Application des RNA pour l'évaluation du risque de contamination des eaux souterraines par les pesticides

Le deuxième objectif était d'évaluer si les variables chimiques pouvaient permettre d'évaluer le risque de contamination des eaux souterraines par les pesticides. L'étude a été basée sur l'utilisation des réseaux de neurones pour la classification.

5.2.1 Choix d'une approche par classification

Les réseaux de neurones ont été appliqués dans cette étude pour la classification. C'est-à-dire que nous n'avons pas cherché à évaluer des concentrations en pesticides, mais le risque de détection. Ce choix de méthodologie a été motivé pour plusieurs raisons :

- Les concentrations en pesticides sont faibles, parfois proches de la limite de détection, les incertitudes sont donc importantes par rapport à l'ordre de grandeur des concentrations mesurées.
- Les variations de concentrations dans les ouvrages sont rapides, et présentent parfois de fortes amplitudes. Un suivi temporel fréquent serait donc nécessaire pour mieux se rendre compte de l'ordre de grandeur possible des concentrations en pesticide pour un ouvrage donné.
- Une grande variété de pesticides est utilisée. Si on utilise la somme des concentrations mesurées, il faudrait alors s'assurer de l'exclusivité des molécules choisies. La présence dans l'eau d'un pesticide non quantifié pourrait alors changer de façon considérable l'ordre de grandeur des concentrations. L'alternative serait alors de se restreindre à l'étude d'un composé donné.

Plusieurs essais ont été réalisés avec les réseaux de neurones afin de prédire les concentrations mesurées. Cependant, les performances du réseau étaient faibles. L'analyse détaillée des résultats a révélé que les erreurs du réseau portaient essentiellement sur les concentrations supérieures à la limite de détection. Il semblerait donc que les variables chimiques ne permettent pas de rendre compte de l'ordre de grandeur des contaminations, mais plutôt d'un risque global.

Une approche par classification semblait ainsi être le meilleur compromis pour s'affranchir des difficultés mentionnées ci-haut. D'autre part, ces approches ne limitent pas l'étude à une loi du

« tout ou rien ». En effet, les sorties du réseau peuvent être interprétées comme des probabilités d'appartenance à l'une des classes. Les réseaux de neurones fournissent donc une information très riche, qui est loin d'être une simple réponse binaire. Cette propriété (que les réseaux de neurones partagent avec d'autres classifieurs) n'a pas été réellement mise à profit dans cette étude, mais offre d'intéressantes perspectives pour une meilleure compréhension des classifications.

5.2.2 Choix des réseaux de neurones

Il peut être avantageux de mettre en œuvre les réseaux de neurones pour toute application tentant de trouver, par apprentissage, une relation non linéaire entre des données numériques (Dreyfus *et al.* 2004). Cependant, plusieurs conditions sont suggérées par ces mêmes auteurs :

- Une première condition nécessaire mais non suffisante concerne la taille des données. Puisque les réseaux de neurones sont des techniques issues des statistiques, il faut disposer d'échantillons représentatifs et de taille suffisamment grande.
- Le deuxième point est de s'assurer de l'intérêt d'un modèle non linéaire pour l'application considérée. La mise en œuvre de modèle linéaire est plus simple et moins coûteuse en temps de calcul. En l'absence de connaissance *a priori* sur l'intérêt d'un modèle non linéaire, les auteurs suggèrent d'utiliser d'abord des méthodes linéaires. S'il s'avère que la précision du modèle est insuffisante, alors la mise en œuvre de modèles non linéaires tels les réseaux de neurones peut être envisagée.
- Si les données sont disponibles et si l'on s'est assuré de l'utilité d'un modèle non linéaire, on doit alors s'interroger sur l'intérêt d'utiliser un réseau de neurones de préférence à une autre famille de modèle non linéaire. Les auteurs suggèrent que si le nombre de variables d'entrée est supérieur à trois, il peut être avantageux d'utiliser des réseaux de neurones à fonction d'activation sigmoïde.

Dans le cas de notre étude, nous disposions d'un nombre d'échantillons suffisamment grand (245), même si par la suite, l'analyse des résultats a révélé qu'un nombre supérieur d'échantillons aurait été utile. En effet, les données devant être divisées en trois séries (apprentissage, validation et test), il est important d'avoir suffisamment d'échantillons afin de pouvoir représenter correctement les différentes classes dans les trois séries.

En ce qui concerne la pertinence d'un modèle non linéaire pour notre application, d'autres techniques de classification ont été utilisées (Doukouré *et al.* 2007). L'analyse factorielle discriminante, la régression logistique et les k plus proches voisins ont été testés sur les mêmes jeux de données et ont donné des performances peu satisfaisantes, de l'ordre de 65 % de bons classements sur le site de Valence. Les réseaux de neurones se sont donc avérés être une alternative donnant des résultats intéressants en termes de performance par rapport aux autres méthodes testées.

5.2.3 Performances et limites de l'étude

Dans le cadre de nos données, l'utilisation des RNA pour la classification nous a permis de classer correctement les échantillons avec ou sans détection de pesticides avec des performances de l'ordre de 80 %, et ceci en utilisant uniquement des données issues d'analyses classiques de la qualité de l'eau. Les analyses de pesticides étant extrêmement coûteuses, une telle méthode pourrait permettre d'améliorer sensiblement certaines campagnes d'échantillonnage en ciblant les ouvrages de captage potentiellement contaminés. Cette méthode pourrait ainsi offrir un compromis intéressant pour les gestionnaires entre les modèles mathématiques performants, mais délicats à mettre en œuvre à grande échelle et les indices de vulnérabilité, d'utilisation plus aisée mais de validité questionnable.

Les modèles mathématiques nécessitent en effet de nombreuses données d'entrée, variables dans l'espace et dans le temps, impliquant une très bonne connaissance du site d'étude et des pratiques culturales qui ne sont pas toujours accessibles. Les indices de vulnérabilité, bien que présentant différents degrés de complexité, définissent le plus souvent des grandes zones de vulnérabilité et n'ont pas pour objectif d'évaluer le risque pour un point précis.

L'avantage de notre méthode réside d'une part dans la facilité d'accès aux données d'entrée et d'autre part dans l'évaluation du risque pour un point donné duquel sont issues les variables d'entrée. Les performances ont été validées sur des séries de données indépendantes et sur plusieurs jeux de données. Cette méthode présente néanmoins plusieurs limites décrites ci-après.

Propriétés hydrochimiques des sites d'études

Les RNA fonctionnent comme une boîte noire. L'emprise sur l'apprentissage est ainsi minimale. Les variables d'entrée sélectionnées dans notre approche peuvent traduire un déséquilibre chimique dû à des contaminations, mais elles représentent avant tout la signature hydrochimique du site. L'apprentissage fonctionne par minimisation de l'erreur, en employant le chemin le plus court. Ainsi, si les différences de contamination inter-sites sont supérieures aux différences intra-sites, le moyen le plus simple de minimiser l'erreur est d'opposer les sites. L'apprentissage sera ainsi biaisé même si les performances peuvent paraître satisfaisantes (paragraphe 3.6.1.2, et 4.4.1). Il est donc recommandé, dans le cadre de cette approche, d'utiliser soit des données issues d'un même site aux propriétés hydrogéochimique uniformes, soit de s'assurer que les contaminations y sont équivalentes.

Données d'apprentissage

Les réseaux de neurones fonctionnent par apprentissage et ne sont pas capables d'extrapoler au-delà de ces données. Les données d'apprentissage doivent ainsi être représentatives des données de validation et de test. Il est donc nécessaire pour chaque application d'avoir suffisamment de données d'exemple provenant du même site d'étude.

Propriétés des différents pesticides

Notre étude ne tient pas compte des propriétés des différents pesticides. Il est reconnu que les propriétés physico-chimiques du composé jouent un rôle important dans les processus de transfert. Les différents pesticides vont évoluer différemment dans le sol en fonction de leur solubilité ou de leur coefficient d'adsorption par exemple. Notre approche évalue un risque global, indépendamment du composé ou même du fait qu'il ait été appliqué. Ceci peut entre autres expliquer les 20 % d'erreur, c'est-à-dire que la signature chimique de l'eau révèle une vulnérabilité potentielle aux pressions agricoles sans détection avérée. Dans les faits, nous ne savons pas si des pesticides ont été appliqués et quelles molécules ont été appliquées. Des études similaires limitées à un composé donné pourraient alors être intéressantes et renseigner sur les variables qui semblent les plus corrélées à telle ou telle molécule.

5.2.4 Conclusion

En conclusion, cette étude a permis de montrer que l'utilisation de variables chimiques dans l'évaluation du risque de contamination des eaux souterraines par les pesticides était pertinente. Les échantillons contaminés présentent en effet des caractéristiques chimiques qui traduisent leur vulnérabilité aux pressions agricoles. L'utilisation de ces variables dans la modélisation neuronale offre des perspectives intéressantes. Cette méthode a été testée sur plusieurs jeux de données, en utilisant différentes approches afin d'évaluer ses limites. L'ordre de grandeur des performances est le même pour les différentes approches, ce qui nous amène à dire que les variables chimiques pourraient permettre d'évaluer environ 80 % des risques de contamination de l'eau souterraine par les produits phytosanitaires. Les 20 % restants peuvent être attribués à des comportements atypiques des points de prélèvements, à des pollutions accidentelles ou à la non-application de produits phytosanitaires.

D'un point de vue temporel, les données chimiques ne permettraient pas *a priori* de rendre compte des variations de contamination au sein d'un même ouvrage. Cependant, la méthodologie utilisée pour cette approche et le nombre de données pour un même ouvrage ne nous ont pas permis d'approfondir cet aspect.

A l'issue de cette étude, plusieurs travaux complémentaires pourraient être menés afin de mieux comprendre le système et les limites de l'utilisation de variables chimiques dans l'évaluation du risque de contamination des eaux souterraines par les produits phytosanitaires :

- Il serait intéressant de mieux comprendre la nature des 20% d'erreurs, c'est-à-dire savoir si elles correspondent réellement à des comportements atypiques des points de prélèvement, tel que des contaminations accidentelles. Pour cela, l'application des réseaux de neurones sur des données théoriques issues de la modélisation pourrait être envisagée. La différence entre les performances sur des données expérimentales et des données théoriques permettrait d'évaluer si les limites sont dues à la méthode elle-même ou aux imprévus de terrain et d'expérimentation.
- Il serait également intéressant d'évaluer si l'ajout de certaines variables d'une autre nature permettrait d'améliorer les performances. En effet, des variables environnementales renseignant par exemple sur le type de sol ou les dates de

prélèvements pourraient apporter de nouvelles informations. L'ajout de telles variables pourrait peut-être permettre de rendre compte de l'ordre de grandeur des contaminations.

- L'application des réseaux de neurones sur un suivi temporel régulier d'un ouvrage présentant un très grand nombre de dates de prélèvement permettrait de confirmer ou d'infirmer la capacité des variables chimiques à traduire les variations et l'évolution de la contamination au sein d'un ouvrage.

Chapitre 6 Références bibliographiques

Aller, L., Bennett, T., Lehr, J. H. et Petty, R. J. (1987). DRASTIC: A standardized system for evaluating ground water pollution potential using hydrogeologic settings. *US Environmental Protection Agency EPA/600/2-85/018.*

Anderson, A. C. (1986). Environmental toxicology - biodegradation of xenobiotics. *Journal of Environmental Health*, (48) : 196-199.

Banton, O., Larocque, M., Lafrance, P., Montminy, M., et Gosselin, M. A. (1997). Développement d'un outil d'évaluation des pertes environnementales de pesticides: intégration d'un module PestiFlux au logiciel AgriFlux. INRS-EAU, Québec, Canada, 144 p

Banton, O. et Villeneuve, J.-P. (1989). Evaluation of groundwater vulnerability to pesticides: A comparison between the pesticide drastic index and the PRZM leaching quantities. *Journal of Contaminant Hydrology*, (4) : 285-296.

Barbash, J. E., Thelin, G. P., Kolpin, D. W., et Gilliom, R. J. (1999). Distribution of major herbicides in ground water of the United States. U.S. Geological Survey, Water-Ressources Investigation. Report 8056-61, 64 p.

Barbash, J.E., Thelin, G P., Kolpin, D.W. et Gilliom, R.J. (2001). Major Herbicides in Ground Water: Results from the National Water-Quality Assessment. *Journal of Environmental Quality*, (30) : 831-845.

Bebis, G. et Georgiopoulos, M. (1994). Feed-forward neural networks: Potentials, IEEE. *Potentials, IEEE*, (13) : 27-31.

Bedos, C., Cellier, P., Calvet, R., Barriuso, E. et Gabrielle, B. (2002). Mass transfer of pesticides into the atmosphere-by volatilization from soils and plants:overview. *Agronomie*, (22) : 21-33.

Ben-Hur, M., Letey, J., Farmer, W.J., Williams, C.F. et Nelson, S.D. (2003). Soluble and Solid Organic Matter Effects on Atrazine Adsorption in Cultivated Soils. *Soil Science Society American Journal*, (67) : 1140-1146.

Bockstaller, C. et Girardin, P. (2003). How to validate environmental indicators. *Agricultural Systems*, (76) : 639-653.

Bohlke, J.K. (2002). Groundwater Recharge and Agricultural Contamination. *Hydrogeology Journal*, (10) : 438-439.

Bowden, G.J., Dandy, G.C. et Maier, H.R. (2005a). Input determination for neural network models in water resources applications. Part 1--background and methodology. *Journal of Hydrology*, (301) : 75-92.

Bowden, G.J., Maier, H.R. et Dandy, G.C. (2005b). Input determination for neural network models in water resources applications. Part 2. Case study: forecasting salinity in a river. *Journal of Hydrology*, (301) : 93-107.

Bradley, A.P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, (30) : 1145-1159.

Burke, L.I. et Ignizio, J.P. (1992). Neural networks and operations research : An overview. *Computers & Operations Research*, (19) : 179-189.

Burow, K.R., Shelton, J.L., et Dubrovsky, N.M. (1998). Occurence of nitrate and pesticides in ground water beneath three agricultural land-use settings in the eastern San Joaquin Valley, California, 1993-1995. U.S. Geological Survey, Water-Ressources Investigation. Report 97-4284, 51 p.

Carsel, R.F, Mulkey, L.A., Lorber, M.N. et Badkin L.B. (1985). The Pesticide Root Zone Model (PRZM): a procedure for evaluating pesticide leaching threats to groundwater. *Ecologic modelling*, (30) :49-69.

Carter, A.D. (2000). Herbicide Movement in Soils: Principles, Pathways and Processes. *Weed Research*, (40) : 113-122.

Castellano, G., Fanelli, A.M. et Pelillo, M. (1997). An iterative pruning algorithm for feedforward neural networks: Neural Networks, IEEE Transactions on. *Neural Networks, IEEE Transactions on*, (8) : 519-531.

Cheng, B. et Titterington, D.M. (1994). Neural Networks: A Review from a Statistical Perspective. *Statistical Science*, (9) : 2-30.

de Bruyn, B. (2004). Étude de la vulnérabilité des eaux aux produits phytosanitaires: indicateur environnemental et modèle mécaniste en vue d'une meilleure gestion du bassin versant de la Leysse (Savoie). Thèse de doctorat. Université Joseph-Fourrier - Grenoble1.

de la Vaissière. R. (2006). Étude de l'aquifère néogène du Bas-Dauphiné - Apport de la géochimie et des isotopes dans le fonctionnement hydrogéologique du bassin de Valence (Drôme, Sud-Est de la France). Thèse de doctorat. Université d'Avignon et des Pays du Vaucluse, Avignon.

Deger, A.B., Gremm, T.J. et Frimmel, F.H. (2000). Problems and Solutions in Pesticide Analysis Using Solid-Phase Extraction (Spe). and Gas Chromatography Ion-Trap Mass Spectrometry Detection (Gc-Ms). *Acta Hydrochimica Et Hydrobiologica*, (28) : 292-299.

Devitt, E.C. et Wiesner, M.R. (1998). Dialysis Investigations of Atrazine-Organic Matter Interactions and the Role of a Divalent Metal. *Environmental Science & Technology*, (32) : 232-237.

Di, H.J., Aylmore, L.A.G. et Kookana, R.S. (1998). Degradation Rates of Eight Pesticides in Surface and Subsurface Soils Under Laboratory and Field Conditions. *Soil Science*, (163) : 404-411.

Doerfliger, N., Jeannin, P.Y. et Zwahlen, F. (1999). Water Vulnerability Assessment in Karst Environments: a New Method of Defining Protection Areas Using a Multi-Attribute Approach and Gis Tools (Epik Method). *Environmental Geology*, (39) : 165-176.

Doukoure, C., Banton, O. et Lafrance, P. (2007). Prediction of pesticide occurrence in domestic wells using chemical indicators. Proceeding of International Conference on WAter POllution in natural POrous media at different scales WAPO2 - Barcelona, Spain, April 10-13, 2007. Barcelona, 2007, pp. 115-120.

Dreyfus, G., Samuelides, M., Martinez, J., Gordon, M., Badran, F., Thiria, S., Hérault, L. (2004). Réseaux de neurones :Méthodologies et applications. Algorithmes, 2^{e} edition, 408 p.

Dubus, I.G., Brown, C.D. et Beulke, S. (2003). Sources of uncertainty in pesticide fate modelling. *The Science of The Total Environment*, (317) : 53-72.

Dunnivant, F..M., Jardine, P.M., Taylor, D.L. et Mccarthy, J.F. (1992). Transport of Naturally-Occurring Dissolved Organic-Carbon in Laboratory Columns Containing Aquifer Material. *Soil Science Society of America Journal*, (56) : 437-444.

Fagnan, N., Bourque, E., Michaud, Y., Lefebvre, R., Boisvert, E., Parent, M. et Martel, R. (1999). Hydrogéologie des comlexes deltad'ques sur la marge nord de la mer de Champlain, Québec. *Hydrogéologie*, (4) : 9-22.

Fitch, A. et Du, J. (1996). Solute Transport in Slay Media: Effect of Humic Acid. *Environmental Science & Technology*, (30) : 12-15.

Flood, I. et Kartam, N. (1994). Neural networks in civil engineering. I: Principles and understanding. *Journal of Computing in Civil Engineering.* (8) : 131-148.

FOOTPRINT (2006). La base de données FOOTPRINT PPDB, Base mise en place dans le cadre du projet européen FOOTPRINT (FP6-SSP-022704), http://www.eu-footprint.org/ppdb.html.

Fortin, V., Ouarda, T.B.M.J., Bobee, B. (1997). Comment on 'The use of artificial neural networks for the prediction of water quality parameters' by H.R. Maier and G.C. Dandy. *Water Resources Research*, (33) : 2423– 2424.

Foster, S.S.D.B(1987). Fundamental concepts in aquifer vulnerability, pollution risk and protection strategy. In: W. van Duijvenbooden, H.G. van Waegeningh (eds.), Vulnerability of soils and ground-water to pollution, Proceedings and information, TNO Committee on Hydrological Research, The Hague, No. 38.

Garbarini, D.R. et Lion, L.W. (1986). Influence of the nature of soil organics on the sorption of toluene and Trichloroethylene. *Environmental Science & Technology,* (20) : 1263-1269.

Giroux, I., Duchemin, M., et Roy, M. (1997). Contamination de l'eau par les pesticides dans les régions de cultures intensive du mais au Québec: Campagnes d'échantillonnage de 1994 et 1995. Ministère de l'Environnement et de la Faune, Direction des écosystèmes aquatiques, 54 p.

Giroux, I. (2003). Contamination de l'eau souterraine par les pesticides et les nitrates dans les régions en culture de pommes de terre. Ministère de l'Environnement et de la Faune, Direction des écosystèmes aquatiques, 23 p.

Gogu, R.C., Hallet, V. et Dassargues, A. (2003). Comparison of Aquifer Vulnerability Assessment Techniques. Application to the Neblon River Basin (Belgium). *Environmental Geology*, (44) : 881-892.

Gooddy, D.C., Bloomfield, J.P., Chilton, P.J., Johnson, A.C. et Williams, R.J. (2001). Assessing Herbicide Concentrations in the Saturated and Unsaturated Zone of a Chalk Aquifer in Southern England. *Ground Water*, (39) : 262-271.

Grass, B., Wenclawiak, B.W. et Rudel, H. (1994). Influence of Air Velocity, Air-Temperature, and Air Humidity on the Volatilization of Trifluralin From Soil. *Chemosphere*, (28) : 491-499.

Grathwohl, P. (1990). Influence of Organic-Matter From Soils and Sediments From Various Origins on the Sorption of Some Chlorinated Aliphatic-Hydrocarbons - Implications on Koc Correlations. *Environmental Science & Technology*, (24) : 1687-1693.

Gustafson, D.I. (1989). Groudwater Ubiquity Score: a simple method for assessing pesticide leachability. *Environmental Toxicology and Chemistry*, (8) : 339-357.

Guyon, I. et Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, (3) : 1157-1182.

Hamilton, P.A. et Helsel, D.R. (1995). Effects of agriculture on ground-water quality in five regions of the United States. *Ground Water*, (33) : 217-226.

Hecht-Nielsen, R. (1990). Neurocomputing. Addison-Wesley, Reading, MA. Kuo-Lin Hsu, Gupta, H. V. et Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process, 31 : 2517-2530.

Hennion, M.C. et Pichon, V. (1994). Solid-Phase Extraction of Polar Organic Pollutants From Water. *Environmental Science & Technology*, (28) : 576-583.

Herbst, M., Hardelauf, H., Harms, R., Vanderborght, J. et Vereecken, H. (2005). Pesticide fate at regional scale: Development of an integrated model approach and application. *Physics and Chemistry of the Earth, Parts A/B/C*, (30) : 542-549.

Heuvelink, G.B.M. et Pebesma, E.J. (1999). Spatial Aggregation and Soil Process Modelling. *Geoderma*, (89) : 47-65.

Heydens, W.F., Siglin, J.C., Holson, J.F. et Stegeman, S.D. (1996). Subchronic, Developmental, and Genetic Toxicology Studies With the Ethane Sulfonate Metabolite of Alachlor. *Fundamental and Applied Toxicology*, (33) : 173-181.

Huneau, F. (2000). Fonctionnement hydrogéologique et archives paléoclimatiques d'un aquifère profond méditerranéen. Etude géochimique et isotopique du bassin miocène de Valréas (Sud-Est de la France). Thèse de doctorat. Université d'Avignon et des Pays du Vaucluse, Avignon.

IFEN (2003). Les pesticides dans les eaux : Cinquième bilan annuel, Données 2001. Institut Français de l'Environnement, Paris, 29 p.

IFEN (2006). Les pesticides dans les eaux: données 2003 et 2004. Institut Français de l'Environnement, Paris, 40 p.

Istok, J.D. et Rautman, C.A. (1996). Probabilistic Assessment of Ground-Water Contamination: 2. Results of Case Study. *Ground Water*, (34) : 1050-1064.

Jury, W.A. et Fluhler, H. (1992). Transport of Chemicals Through Soil - Mechanisms, Models, and Field Applications. *Advances in Agronomy*, (47) : 141-201.

Jury, W.A., Spencer, W.F. et Farmer, W.J. (1983). Behavior assesment model for trace organics in soil: I. Model description. *Journal of Environmental Quality*, (12) : 558-564.

Kan, A.T., Fu, G.M. et Tomson, M.B. (1994). Adsorption-Desorption Hysteresis in Organic Pollutant and Soil Sediment Interaction. *Environmental Science & Technology*, (28) : 859-867.

Kerle, E.A., Jenkins, J.J., et Vogue, P.A. (1996). Understanding pesticide persistance and mobility for groundwater and surface water protection. Oregon State University.

Kolpin, D.W., Thurman, E.M. et Linhart, S.M. (1998). The Environmental Occurrence of Herbicides: the Importance of Degradates in Ground Water. *Archives of Environmental Contamination and Toxicology*, (35) : 385-390.

Kuo-Lin Hsu, Gupta, H.V. et Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process, (31) : 2517-2530.

Lafrance, P. et Banton, O. (1995). Implication of spatial variability of organic carbon on predicting pesticide mobility in soil. *Geoderma*, (65) : 331-338.

Lafrance, P., Banton, O., Campbell, P.G.C. et Villeneuve, J.P. (1990). A Complexation Adsorption Model Describing the Influence of Dissolved Organic-Matter on the Mobility of Hydrophobic Compounds in Groundwater. *Water Science and Technology*, (22) : 15-22.

Laird, D.A., Yen, P.Y., Koskinen, W.C., Steinheimer, T.R. et Dowdy, R.H. (1994). Sorption of Atrazine on Soil Clay Components. *Environmental Science & Technology*, (28) : 1054-1061.

Lalbat, F. (2006). Fonctionnement hydrodynamique de l'aquifère Micocène du bassin de Carpentras (Vaucluse, France). Thèse de doctorat. Université d'Avignon et des Pays du Vaucluse, Avignon.

Lasko, T.A., Bhagwat, J.G., Zou, K.H. et Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics: Clinical Machine Learning. *Journal of Biomedical Informatics*, (38) : 404-415.

Leonard, R.A. (1990). Pesticides in the soil environment: processes, impacts, and modeling. *In* H.H. Cheng (ed.). Pesticides in soil environment Processes, impacts, and modelling.

Leray, P. et Gallinari, P. (2001). De l'utilisation d'OBD pour la selection de variables dans les perceptrons multicouches. *Revue d'Intelligence Artificielle*, (15) : 373-391.

Lewis, K.A., Brown, C.D., Hart, A. et Tzilivakis, J. (2003). P-Ema (Iii): Overview and Application of a Software System Designed to Assess the Environmental Risk of Agricultural Pesticides. *Agronomie*, (23) : 85-96.

Lindström, R. (2005). Groundwater vulnerability assessment using process-based models. TRITA-LWR PhD Thesis 1022 - 36 p.

Loague, K., Lloyd, D., Nguyen, A., Davis, S.N. et Abrams, R.H. (1998a). A case study simulation of DBCP groundwater contamination in Fresno County, California 1. Leaching through the unsaturated subsurface. *Journal of Contaminant Hydrology*, (29) : 109-136.

Loague, K., Abrams, R.H., Davis, S.N., Nguyen, A. et Stewart, I.T. (1998b). A case study simulation of DBCP groundwater contamination in Fresno County, California 2. Transport in the saturated subsurface. *Journal of Contaminant Hydrology*, (29) : 137-163.

Loague, K. et Abrams, R.H. (2001). Stochastic-conceptual analysis of near-surface hydrological response. *Hydrological Processes*, (15) : 2715-2728.

Loague, K. et Corwin, D.L. (1998). Regional-scale assessment of non-point source groundwater contamination. *Hydrological Processes*, (12) : 957-965.

Maier, H.R. et Dandy, G.C. (1996). The use of artificial neural networks for the prediction of water quality parameters, (32) : 1013-1022.

Maier, H.R. et Dandy, G.C. (1997). Determining inputs for neural network models of multivariate time series, (12) : 353-368.

Maier, H.R. et Dandy, G.C. (2000a). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling and Software*, (15) : 101-124.

Maier, H.R. et Dandy, G.C. (2000b). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling and Software*, (15) : 101-124.

Maren, A., Harston, C., Pap, R. (1990). Handbook of Neural Computing Applications. Academic Press, San Diego, CA.

McCulloch, W. et Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity, (5) : 115-133.

McKenna, D.P., Bicki, T.J., Dey, W.S., Keefer, D.A., Mehnert, E., Panno, S.V., Ray, C., Wilson, S.D., and Schock, S.C. (1990). An Initial Evaluation of the Impact of Pesticides on

Groundwater in Illinois : Report to the Illinois Legislature . Report to the Illinois Legislature, 107 p.

Means, J.C., Wood, S.G., Hasset, J.J. et Banwart, W.L. (1982). Sorption of amino- and carboxy-subtitued polynuclear aromatic hydrocarbons by sediments and soils. *Environmental Science & Technology*, (16) : 93-98.

Merchant, J.W. (1994). Gis-Based Groundwater Pollution Hazard Assessment - a Critical-Review of the Drastic Model (Vol 60, Pg 1123, 1994). *Photogrammetric Engineering and Remote Sensing*, (60) : 1416.

Mishra, A., Ray, C. et Kolpin, D.W. (2004). Use of Qualitative and Quantitative Information in Neural Networks for Assessing Agricultural Chemical Contamination of Domestic Wells. *Journal of Hydrologic Engineering*, (9) : 502-511.

Mullins JA, Carsel RF, Scarbrough JE, Ivery AM. (1993). PRZM-2, a model for predicting pesticide fate in the crop root and unsaturated soil zones: Users manual for release 2.0. EPA/600/R-93/046. Technical Report. Environmental Research Laboratory, U.S. Environmental Protection Agency, Athens, GA.

Navarro, S., Vela, N., Jose Gimenez, M. et Navarro, G. (2004). Persistence of four s-triazine herbicides in river, sea and groundwater samples exposed to sunlight and darkness under laboratory conditions. *Science of The Total Environment*, (329) : 87-97.

Nelson, S.D., Letey, J., Farmer, W.J., Williams, C.F. et Ben-Hur, M. (1998). Facilitated Transport of Napropamide by Dissolved Organic Matter in Sewage Sludge-Amended Soil. *Journal of Environmental Quality*, (27) : 1194-1200.

Pacheco, F., van der Weijden, C.H. (1996). Contributions of water-rock interactions to the composition of groundwater in areas with a sizeable anthropogenic input : A case study of the waters of the Fundão area, central Portugal. Water Resources Research, (32) : 3553-3570.

Padovani, L., Trevisan, M. et Capri, E. (2004). A calculation procedure to assess potential environmental risk of pesticides at the farm level. *Ecological Indicators*, (4) : 111-123.

Pimentel, D. (1995). Amounts of Pesticides Reaching Target Pests - Environmental Impacts and Ethics. *Journal of Agricultural & Environmental Ethics* (8) : 17-29

Pionke, H.B., Urban, J.B. (1985). Effect of Agricultural Land Use on Ground-Water Quality in a Small Pennsylvania Watershed. *Ground Water* (23) : 98-80

Pussemier, L. (1999). SYPEP: A system for predicting the environmental impact of pesticides in Belgium. *Proceedings of the XI Symposium Pesticide Chemistry "Human and environmental exposure to xenobiotics"* 12-15 September 1999. Cremoana, Italy.

Rakotomalala R., 2005. "TANAGRA : un logiciel gratuit pour l'enseignement et la recherche", in Actes de EGC'2005, RNTI-E-3, vol. 2, p. 697-702.

Ramade F. (1998). *Dictionnaire encyclopédique des sciences de l'eau.* Ediscience international, Paris, 786 p.

Ray, C. et Klindworth, K.K. (2000). Neural Networks for Agrichemical Vulnerability Assessment of Rural Private Wells. *Journal of Hydrologic Engineering*, (5) : 162-171.

Refsgaard J.C., Butts M.B. (1999). Determination of grid-scale parameters in catchment modelling by upscaling local scale parameters. In Proceedings of the International Workshop of EurAgEng's Field of Interest on Soil and Water Modelling of transport processes in soils at various scales in time and space , Leuven; 650-665.

Reus, J., Leendertse, P., Bockstaller, C., Fomsgaard, I., Gutsche, V., Lewis, K., Nilsson, C., Pussemier, L., Trevisan, M. et van der Werf, H. (2002). Comparison and evaluation of eight pesticide environmental risk indicators developed in Europe and recommendations for future use. *Agriculture, Ecosystems & Environment*, (90) : 177-187.

Rennard, J.P. (2006). Réseaux neuronnaux – Une introductionaccompagnée d'un modèle Java. Paris, 282 p.

Rogers, L.L. et Dowla, F.U. (1994). Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling, (30) : 457-481.

Sahoo, G.B., Ray, C. et Wade, H.F. (2005). Pesticide prediction in ground water in North Carolina domestic wells using artificial neural networks, (183) : 29-46.

Sahoo, G.B., Ray, C., Mehnert, E. et Keefer, D.A. (2006). Application of artificial neural networks to assess pesticide contamination in shallow groundwater. *Science of The Total Environment*, (367) : 234-251.

Seidou, O., Ouarda, T.B.M.J., Bilodeau, L., Hessami, M., St.-Hilaire, A. et Bruneau, P. (2006). Modeling ice growth on Canadian lakes using artificial neural networks, (42).

Shukla, M.B., Kok, R., Prasher, S.O., Clark, G., Lacroix, R. (1996). Use of artificial neural networks in transient drainage design. Transactions of the ASAE, 39(1) : 119–124.

Spark, K.M. et Swift, R.S. (2002). Effect of soil composition and dissolved organic matter on pesticide sorption. *The Science of The Total Environment*, (298) : 147-161.

StatSoft, Inc. (2007). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/stathome.html.

Stenemo, F., Jorgensen, P.R. et Jarvis, N. (2005). Linking a one-dimensional pesticide fate model to a three-dimensional groundwater model to simulate pollution risks of shallow and deep groundwater underlying fractured till. *Journal of Contaminant Hydrology*, (79) : 89-106.

Stites, W. et Kraft, G.J. (2001). Nitrate and Chloride Loading to Groundwater from an Irrigated North-Central U.S. Sand-Plain Vegetable Field. *J Environ Qual*, (30) : 1176-1184.

Tariq, M.I., Afzal, S. et Hussain, I. (2004). Pesticides in shallow groundwater of Bahawalnagar, Muzafargarh, D.G. Khan and Rajan Pur districts of Punjab, Pakistan. *Environment International*, (30) : 471-479.

Teso, R.R., Poe, M.P., Younglove, T. et Mccool, P.M. (1996). Use of Logistic Regression and Gis Modeling to Predict Groundwater Vulnerability to Pesticides. *Journal of Environmental Quality*, (25) : 425-432.

Tessier, D.M. et Clark, J.M. (1995). Quantitative Assessment of the Mutagenic Potential of Environmental Degradative Products of Alachlor. *Journal of Agricultural and Food Chemistry*, (43) : 2504-2512.

Tiktak, A. (2000). Application of Pesticide Leaching Models to the Vredepeel Dataset Ii Pesticide Fate. *Agricultural Water Management*, 44, no. 1-3: 119-34.

Topp, E. et Smith, W. (1992). Sorption of the Herbicides Atrazine and Metolachlor to Selected Plastics and Silicone-Rubber. *Journal of Environmental Quality*, (21) : 316-317.

Triegel, E.K. et Guo, L. (1994). Overview of the fate of pesticides in the environment, water balance; runoff vs. leaching. In: *Mechanisms of pesticide movement into ground water* Lewis Publishers.

van der Werf, H..M. et Zimmer, C. (1998). An indicator of pesticide environmental impact based on a fuzzy expert system. *Chemosphere*, (36) : 2225-49.

van der Werf, H. M. G. (1996). Assessing the impact of pesticides on the environment. *Agriculture, Ecosystems & Environment*, (60) : 81-96.

Wagenet, R.J. et Hutson, J.L., (1987). LEACHM: A Finite-Difference Model for Simulating Water, Salt and Pesticide Movement in the Plant Root Zone, Continuum 2 Ithaca. New York State Resources Institute, Cornell University.

Wagenet, R.J. et Rao, P.S.C. (1990). Modeling pesticide Fate in soils. In: *Pesticides in the soil environment: processes, impacts, and modeling* . Soil Science Society of America.

Wauchope, R.D. (1978). The pesticide contente of surface water draining from agricultural field - a review. *Journal of Environmental Quality*, (7) : 459-472.

Wauchope, R.D., Yeh, S., Linders, J.B.H.J., Kloskowski, R., Tanaka, K., Rubin, B., Katayama, A., Kördel, W., Gerstl, Z., Lane, M. et Unsworth, J.B. (2002). Pesticide soil sorption parameters: theory, measurement, uses, limitations and reliability. *Pest Management Science*, (58) : 419-445.

Weber, J.C. (1994). Properties and behavior of pesticides in soil. In: *Mechanisms of pesticide movement into ground water* Lewis Publishers.

Worrall, F., Besien T. et Kolpin D.W. (2002). Groundwater vulnerability: interactions of chemical and site properties. *The Science of the Total Environment*, (299) : 131-143.

Yang, C.C., Prasher, S.O., Lacroix, R. et Kim, S.H. (2003). A multivariate adaptive regression splines model for simulation of pesticide transport in soils, (86) : 9-15.

Zhang, R., Hamerlinck, J.D., Gloss, S.P. et Munn, L. (1996). Determination of nonpoint-source pollution using GIS and numerical models. *Journal of Environmental Quality*, (25) : 411-418.

ANNEXE A - Points de prélèvement sur les sites d'étude

A.1. Site de Valence

A.2. Site de Carpentras-Valréas

A.3. Site de Portneuf

A.1. Site de Valence

N	X L2-é (m)	Y L2-é (m)	prof.	pH	cond. 25°C	F	Cl	NO2	Br	NO3	PO4	SO4	DOC	dia	DEA	atrazine	DET	terbuthyl azine	simazine	diuron	oxadixyl	somme pesticides
	(m)	(m)	(m)		(µS/cm)	mg/l								µ/l								
A01	811971	2004999	87	7.23	470	0.1	1.7	0.2	n.d.	5.0	n.d.	3.4	2.46	n.d.	n.d.	n.d.	n.d.	0.02	n.d.	n.d.	n.d.	0.02
A02	812839	2003125	59	6.9	818	n.d.	23.5	0.2	n.d.	82.5	n.d.	35.4	1.30	n.d.	0.025	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.025
A03	817238	2001858	20	7.09	769	n.d.	26.5	n.d.	n.d.	71.9	n.d.	25.6	1.87	n.d.	0.029	0.02	n.d.	n.d.	n.d.	n.d.	n.d.	0.049
A04	819266	2002728	34	6.8	749	n.d.	10.2	0.1	n.d.	56.0	n.d.	18.1	1.93	n.d.	n.d.	0.042	n.d.	n.d.	n.d.	n.d.	n.d.	0.042
A06	818196	1997146	48	6.9	690	n.d.	11.8	n.d.	n.d.	39.4	n.d.	21.3	1.86	n.d.	0.411	0.314	n.d.	n.d.	n.d.	n.d.	n.d.	0.725
A07	815299	1997342	120	7.33	565	n.d.	7.3	n.d.	n.d.	23.0	n.d.	15.8	1.86	n.d.	0.028	0.028	n.d.	n.d.	n.d.	n.d.	n.d.	0.056
A09	815987	1997101	92	7.26	453	n.d.	2.3	n.d.	n.d.	14.9	n.d.	4.0	1.24	n.d.	0.029	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.029
B03	816673	2001335	120	7.04	580	n.d.	10.8	n.d.	n.d.	46.4	n.d.	20.3	1.86	n.d.	0.061	0.077	n.d.	n.d.	n.d.	n.d.	n.d.	0.138
B08	815689	1994396	83	7.09	509	n.d.	3.2	n.d.	n.d.	21.3	n.d.	10.2	1.44	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
B09	813079	1986974	200	7.19	588	n.d.	34.8	0.2	n.d.	88.8	n.d.	30.5	2.07	n.d.	0.028	0.026	n.d.	n.d.	n.d.	n.d.	n.d.	0.054
C01	811573	1988057	80	6.87	700	n.d.	12.9	n.d.	n.d.	41.0	n.d.	23.9	2.38	n.d.	0.029	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.029
C02	812253	2000873	80	6.79	1232	n.d.	84.9	n.d.	n.d.	99.1	n.d.	59.2	3.33	n.d.	n.d.	n.d.	n.d.	n.d.	0.034	n.d.	n.d.	0.034
C03	812602	1991800	32	6.92	717	n.d.	14.5	n.d.	n.d.	37.8	n.d.	34.8	2.52	n.d.	0.031	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.031
C04	806824	1981901	57	6.99	524	n.d.	12.4	n.d.	n.d.	20.3	n.d.	16.3	1.52	n.d.	0.032	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.032
C05	804363	1982028	40	6.92	829	n.d.	18.8	n.d.	n.d.	56.7	n.d.	67.3	2.12	n.d.	0.171	0.123	n.d.	n.d.	n.d.	n.d.	n.d.	0.294
C06	809606	1981575	80	7.07	525	n.d.	5.9	n.d.	n.d.	11.4	n.d.	15.0	1.64	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
C07	809760	1982365	150	6.98	521	n.d.	3.2	n.d.	n.d.	6.4	n.d.	3.3	1.39	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.134	n.d.	0.134
F01	807093	1994870	40	6.85	766	n.d.	8.4	n.d.	n.d.	48.7	n.d.	22.3	0.98	n.d.	0.06	0.044	n.d.	n.d.	n.d.	n.d.	n.d.	0.104
F02	813943	1996161	220	7.61	563	n.d.	1.0	n.d.	n.d.	2.8	n.d.	4.8	1.25	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
F03	814163	1998070	180	7.26	656	n.d.	12.2	n.d.	n.d.	52.6	n.d.	21.7	1.23	n.d.	0.177	0.115	n.d.	n.d.	n.d.	n.d.	n.d.	0.292
F04	804459	1991272	151.5	7.15	618	n.d.	7.5	n.d.	n.d.	55.1	n.d.	14.1	0.73	n.d.	0.047	0.05	n.d.	n.d.	n.d.	n.d.	n.d.	0.097
F05	802552	1991731	80	7.42	459	n.d.	36.0	n.d.	n.d.	116.6	n.d.	48.2	1.08	n.d.	0.041	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.041
F06	805587	1979200	100	7.38	471	n.d.	10.3	n.d.	n.d.	78.3	n.d.	18.5	0.92	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
F07	812497	1978778	110	7.08	785	n.d.	22.0	0.2	n.d.	73.8	n.d.	26.7	1.61	n.d.	0.174	0.051	n.d.	n.d.	n.d.	n.d.	n.d.	0.225
F08	812358	1984002	60	7.8	453	n.d.	16.2	n.d.	n.d.	86.5	n.d.	30.6	1.04	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
F09	809431	2004612	80	7.05	530	n.d.	17.5	n.d.	n.d.	51.0	n.d.	36.1	1.23	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
F10	811581	2012781	150	7.16	515	n.d.	2.6	n.d.	n.d.	13.6	n.d.	5.0	0.86	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
F11	810726	2015671	110	7.34	715	n.d.	5.1	n.d.	n.d.	13.3	n.d.	8.5	1.00	n.d.	0.021	0.067	n.d.	n.d.	n.d.	n.d.	n.d.	0.088
G01	813492	2001746	63	7.33	438	n.d.	16.3	0.1	n.d.	55.5	n.d.	37.5	1.87	n.d.	n.d.	0.112	n.d.	0.07	n.d.	n.d.	n.d.	0.182
G02	819630	2007047	130	7.26	643	n.d.	10.1	n.d.	n.d.	64.7	n.d.	25.8	1.25	n.d.	0.173	0.214	n.d.	n.d.	n.d.	n.d.	n.d.	0.387
G04	812309	1993315	79	7	673	n.d.	12.4	n.d.	n.d.	28.5	n.d.	24.8	2.10	n.d.	0.064	0.051	n.d.	n.d.	n.d.	n.d.	n.d.	0.115
G05	807161	1983739	85	7.48	638	n.d.	11.6	n.d.	n.d.	48.8	n.d.	19.2	2.39	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
G06	804071	1984534	111	7	644	0.1	11.4	n.d.	n.d.	27.6	n.d.	19.3	2.11	n.d.	n.d.	0.023	n.d.	n.d.	n.d.	n.d.	0.023	0.046
G07	805928	1990996	70	7.22	526	0.0	0.7	n.d.	n.d.	n.d.	n.d.	6.1	2.10	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
G08	805928	1990990	37	7.26	446	n.d.	1.3	0.2	n.d.	5.8	n.d.	7.0	1.42	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
G09	804640	1990242	120	7.55	714	n.d.	31.2	n.d.	n.d.	85.1	n.d.	61.3	1.88	n.d.	0.438	0.353	n.d.	n.d.	n.d.	n.d.	n.d.	0.791
G10	805576	1989928	80	7.56	509	n.d.	4.3	n.d.	n.d.	22.5	n.d.	4.0	1.32	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
G11	819308	2007213	62	7.1	580	n.d.	4.6	n.d.	n.d.	21.6	n.d.	21.5	1.69	n.d.	0.08	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.08
H01	819730	2007087	160	7.48	593	n.d.	12.6	n.d.	n.d.	68.0	n.d.	5.5	0.91	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
H02	816922	2005648	153	7.82	409	n.d.	4.3	n.d.	n.d.	12.0	n.d.	5.4	0.87	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
H03	816932	2005546	50	7.39	831	n.d.	11.1	n.d.	n.d.	29.4	n.d.	14.8	1.13	n.d.	n.d.	n.d.	0.026	n.d.	n.d.	n.d.	n.d.	0.026
H04	813248	2007392	121.5	7.69	590	n.d.	18.3	n.d.	n.d.	66.9	n.d.	20.6	1.41	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
H05	805249	2011902	60	6.96	789	n.d.	16.2	0.1	n.d.	67.3	n.d.	31.0	1.13	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
H07	807865	1985197	100	7.01	677	n.d.	14.3	n.d.	n.d.	42.6	n.d.	27.7	1.06	n.d.	n.d.	n.d.	0.027	n.d.	n.d.	n.d.	n.d.	0.027
H08	807834	1987165	110	7.25	359	n.d.	9.5	n.d.	n.d.	23.5	n.d.	23.5	1.28	n.d.	n.d.	0.041	n.d.	n.d.	n.d.	n.d.	n.d.	0.041
H09	810054	1988286	100	7.75	346	n.d.	7.6	n.d.	n.d.	40.1	n.d.	10.7	0.93	n.d.	0.06	0.028	n.d.	n.d.	n.d.	n.d.	n.d.	0.088
H10	814616	2011706	70	6.98	539	n.d.	1.0	n.d.	n.d.	2.2	n.d.	0.7	0.74	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
H11	807224	1980851	100	6.97	645	n.d.	10.6	n.d.	n.d.	22.5	n.d.	24.1	1.18	n.d.	0.027	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.027

n.d. : non détecté - d. :détecté - / : non renseigné

N	X L2-é (m)	Y L2-é (m)	prof.	pH	cond. 25°C	F	Cl	NO2	Br	NO3	PO4	SO4	DOC	dia	DEA	atrazine	DET	terbuthyl azine	simazine	diuron	oxadixyl	somme pesticides
	(m)	(m)	(m)		(µS/cm)	mg/l								µ/l								
Q01	813805	2016036	80	6.81	604	0.1	5.8	n.d.	n.d.	11.0	n.d.	8.8	0.84	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Q02	813805	2016036	40	7.26	474	n.d.	7.6	n.d.	n.d.	15.9	n.d.	9.3	0.84	n.d.	0.052	n.d.	n.d.	n.d.	0.02	n.d.	n.d.	0.072
Q03	816177	2008357	56	7.3	364	n.d.	27.4	n.d.	n.d.	52.3	n.d.	32.3	1.44	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Q04	799253	2024155	92	7.26	428	0.2	25.2	n.d.	n.d.	37.0	n.d.	25.2	1.18	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Q05	800811	2022928	83	7.49	640	0.1	9.9	n.d.	n.d.	42.2	n.d.	14.4	0.88	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Q06	801554	2021058	126	6.98	694	n.d.	6.4	n.d.	n.d.	34.6	n.d.	11.9	0.96	n.d.	0.022	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.022
Q07	803392	2022691	122	6.97	570	n.d.	5.7	n.d.	n.d.	23.8	n.d.	7.3	0.76	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Q08	813430	2021694	60	7	711	n.d.	9.8	n.d.	n.d.	36.2	n.d.	12.5	1.01	n.d.	0.048	0.02	n.d.	n.d.	n.d.	n.d.	n.d.	0.068
Q09	815154	2022774	102	6.57	506	n.d.	16.8	n.d.	n.d.	42.6	n.d.	22.2	1.31	n.d.	0.169	0.207	n.d.	n.d.	n.d.	n.d.	n.d.	0.376
Q10	814286	2021728	87	/	480	n.d.	9.1	n.d.	n.d.	27.9	n.d.	15.1	0.85	n.d.	0.063	0.071	n.d.	n.d.	n.d.	n.d.	n.d.	0.134
Q11	805544	2018790	180	/	605	0.1	5.1	n.d.	n.d.	12.9	n.d.	3.3	0.96	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
R01	802806	2018752	38	5.46	573	0.1	17.7	n.d.	n.d.	63.6	n.d.	33.8	0.74	n.d.	0.203	0.478	n.d.	n.d.	0.061	n.d.	n.d.	0.742
R02	814441	2000429	100	7.18	420	0.1	10.6	n.d.	n.d.	9.2	n.d.	13.6	0.35	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
R03	809144	1989030	60	7.42	345	0.1	4.8	n.d.	n.d.	9.7	n.d.	5.6	0.39	n.d.	0.02	0.02	n.d.	n.d.	n.d.	n.d.	n.d.	0.04
S01	808140	2011220	156	5.83	566	0.1	6.7	n.d.	n.d.	26.4	n.d.	5.8	1.08	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.38	n.d.	0.38
S02	814509	2018167	74	6.4	509	0.1	10.4	n.d.	n.d.	26.4	n.d.	12.4	0.99	n.d.	0.033	0.08	n.d.	n.d.	n.d.	n.d.	n.d.	0.113
S03	813570	2017710	123	7.21	436	0.1	8.1	n.d.	n.d.	9.5	n.d.	10.2	0.65	n.d.	0.03	0.054	n.d.	n.d.	n.d.	n.d.	n.d.	0.084
S04	817054	2025009	50	7.38	379	n.d.	8.2	n.d.	n.d.	8.0	n.d.	16.9	0.63	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
S05	813513	2024820	70	7.35	536	n.d.	10.9	n.d.	n.d.	13.9	n.d.	13.9	0.82	n.d.	n.d.	0.029	n.d.	n.d.	n.d.	n.d.	n.d.	0.029
S06	813015	2023628	55	6.86	577	0.1	8.3	n.d.	n.d.	8.9	n.d.	11.9	0.71	n.d.	n.d.	0.04	n.d.	n.d.	n.d.	n.d.	n.d.	0.04
S07	815325	2023606	56	6.55	666	n.d.	9.9	n.d.	n.d.	7.3	n.d.	9.8	0.64	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
S08	805770	2023496	70	6.88	406	0.1	31.7	n.d.	n.d.	24.5	n.d.	21.2	0.61	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
S09	806412	2025079	215	6.5	426	n.d.	15.4	n.d.	n.d.	40.6	n.d.	16.8	0.66	n.d.	0.047	0.036	n.d.	n.d.	n.d.	0.02	n.d.	0.103
S10	805086	2025586	180	6.05	811	n.d.	35.0	n.d.	n.d.	62.4	n.d.	23.2	0.78	n.d.	0.041	0.034	n.d.	n.d.	n.d.	n.d.	n.d.	0.075
T01	811059	1989637	100	5.6	665	0.1	17.1	n.d.	n.d.	26.7	n.d.	14.0	0.53	n.d.	0.045	0.07	n.d.	n.d.	n.d.	n.d.	n.d.	0.115
T02	815272	2027693	122	6.43	555	n.d.	15.7	n.d.	n.d.	2.7	n.d.	8.9	0.44	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
T03	813455	2033097	100	5.76	486	n.d.	8.7	n.d.	1.3	6.9	n.d.	8.8	0.58	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
T04	809050	2031047	141	5.95	481	n.d.	6.3	n.d.	n.d.	11.2	n.d.	8.5	0.54	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
T05	803054	2025894	105	5.95	526	0.1	12.3	n.d.	n.d.	24.1	n.d.	15.5	0.53	n.d.	n.d.	0.024	n.d.	n.d.	n.d.	n.d.	n.d.	0.024
T06	803056	2031044	40	6.48	635	n.d.	24.1	n.d.	n.d.	59.2	n.d.	24.7	0.54	n.d.	0.057	0.034	n.d.	n.d.	n.d.	n.d.	n.d.	0.091
T07	807507	2028602	162	6.4	506	0.1	10.4	n.d.	n.d.	16.8	n.d.	9.8	0.60	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
T08	809260	2027971	205	5.98	550	0.1	63.0	n.d.	n.d.	15.3	n.d.	15.7	0.52	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
T09	805881	2027881	100	6.6	656	0.1	8.3	n.d.	n.d.	5.0	n.d.	2.6	0.51	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
T10	805734	2024605	50	5.93	428	0.1	6.2	n.d.	n.d.	26.0	n.d.	10.3	0.58	n.d.	n.d.	0.034	n.d.	n.d.	n.d.	n.d.	n.d.	0.034
T11	804624	2023638	129	6.4	657	n.d.	3.5	n.d.	n.d.	2.4	n.d.	11.3	0.50	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
B05	803836	1999864	26	7.01	697	n.d.	16.4	n.d.	n.d.	37.1	n.d.	28.9	2.20	n.d.	n.d.	0.022	n.d.	n.d.	n.d.	n.d.	0.07	0.092
B06	807381	2006288	0	7.01	531	n.d.	15.8	0.2	n.d.	75.3	n.d.	57.6	3.26	n.d.	n.d.	0.026	n.d.	n.d.	n.d.	n.d.	n.d.	0.026
B02	807559	2002415	45	6.86	866	n.d.	68.4	n.d.	n.d.	117.6	n.d.	46.8	2.63	n.d.	0.163	0.713	n.d.	n.d.	n.d.	n.d.	n.d.	0.876
B01	808265	2004962	33	6.84	838	n.d.	19.7	0.2	n.d.	77.4	n.d.	56.7	3.19	n.d.	0.03	0.118	n.d.	n.d.	n.d.	0.062	n.d.	0.21
G03	815780	2002900	13	7.02	660	0.2	22.2	n.d.	n.d.	62.1	n.d.	25.4	/	n.d.	0.345	1.11	n.d.	n.d.	0.107	n.d.	n.d.	1.562
A05	820025	2004345	16	6.8	724	n.d.	11.0	n.d.	n.d.	54.7	n.d.	19.5	2.05	n.d.	0.05	0.038	n.d.	n.d.	n.d.	n.d.	n.d.	0.088
B07	817861	1999470	28	7.03	827	n.d.	15.7	n.d.	n.d.	61.4	n.d.	24.2	1.72	n.d.	0.022	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.022
B11	805800	1992000	11	7.08	572	n.d.	13.7	n.d.	n.d.	43.7	n.d.	26.7	1.72	n.d.	0.045	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.045
B04	816218	2007766	0	6.95	925	n.d.	22.3	n.d.	n.d.	75.5	n.d.	64.5	2.58	n.d.	0.022	n.d.	n.d.	n.d.	0.055	n.d.	n.d.	0.077
B10	806707	1989378	12	7.26	915	n.d.	9.3	n.d.	n.d.	24.2	n.d.	19.5	2.64	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
H06	801318	1992327	12	7.05	639	n.d.	15.6	0.2	n.d.	44.0	n.d.	35.3	1.19	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
A08	815979	1997113	15	7.2	496	n.d.	4.7	n.d.	n.d.	19.9	n.d.	17.3	1.74	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

n.d. : non détecté - d. :détecté - / : non renseigné

A. 2. Site de Carpentras-Valrêas

N	X L2-é (m)	Y L2-é (m)	prof.	pH	cond. 25°C	F	Cl	NO2	Br	NO3	PO4	SO4	DOC	dia	DEA	atrazine	DET	terbuthylazine	simazine	diuron	oxadixyl	somme pesticides
	(m)	(m)	(m)		(µS/cm)	mg/l								µl								
C9	817769	1905542	/	7.03	683	0.19	33.76	n.d.	0.18	25.2	n.d.	82.97	0.99	n.d.	n.d.	n.d.	n.d.	0.021	n.d.	n.d.	n.d.	0.021
C10	817304	1904536	/	7.07	464	0.17	5.07	n.d.	n.d.	8.9	n.d.	5.42	0.73	n.d.	n.d.	n.d.	n.d.		n.d.	n.d.	n.d.	n.d.
C11	817627	1904222	150	7.28	522	0.08	11.34	n.d.	n.d.	24.1	n.d.	23.27	0.66	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
D1	817403	1902634	190	7.16	603	0.31	11.42	n.d.	n.d.	5.0	n.d.	32.18	1.15	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
D2	818235	1903652	/	6.91	727	0.18	11.09	n.d.	n.d.	2.8	n.d.	64.57	1.34	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
D3	817424	1902636	70	7.16	555	0.21	11.33	n.d.	n.d.	15.9	n.d.	18.05	1.33	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
D4	817564	1903525	160	7.11	568	0.22	10.84	n.d.	n.d.	8.7	n.d.	29.12	0.87	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
D5	818100	1905732	120	7.01	643	0.30	17.27	n.d.	n.d.	16.0	n.d.	56.48	0.96	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
D6	817526	1905093	90	7.03	463	0.20	15.42	n.d.	n.d.	8.9	n.d.	49.19	0.85	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
D7	817331	1903023	230	7.12	542	0.05	11.15	n.d.	n.d.	7.0	n.d.	31.41	0.71	n.d.	n.d.	n.d.	0.03	n.d.	n.d.	n.d.	n.d.	0.03
D8	817445	1902103	90	7.16	585	0.25	12.18	n.d.	n.d.	6.1	n.d.	31.68	1.24	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d
D9	816941	1901363	250	7.24	774	0.13	22.03	0.1	0.26	54.6	n.d.	128.69	1.08	n.d.	n.d.	n.d.	0.027	n.d.	n.d.	n.d.	n.d.	0.027
D10	816628	1900079	115	7.32	557	0.39	21.78	n.d.	n.d.	3.5	n.d.	86.87	0.90	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
E1	826143	1900421	40	7.23	736	1.95	13.85	n.d.	n.d.	17.4	n.d.	92.65	1.48	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
E2	825247	1900569	30	6.99	1256	1.19	20.05	n.d.	n.d.	13.1	n.d.	449.60	1.84	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
E3	824328	1901980	100	7.15	776	0.07	9.26	n.d.	n.d.	35.5	n.d.	177.32	1.40	n.d.	n.d.	n.d.	0.119	0.02	n.d.	n.d.	n.d.	0.139
E4	822647	1900440	130	7.08	979	0.29	43.19	n.d.	n.d.	25.8	n.d.	199.93	1.37	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
E5	822999	1901239	/	7.35	489	0.27	8.03	n.d.	n.d.	0.0	n.d.	34.61	1.00	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
E6	822435	1901612	40	7.02	940	0.19	31.48	n.d.	n.d.	5.1	n.d.	200.11	1.52	n.d.	n.d.	n.d.	0.282	n.d.	n.d.	n.d.	n.d.	0.282
E7	821670	1901080	120	7.18	846	0.24	29.91	n.d.	n.d.	16.6	n.d.	141.71	1.58	n.d.	n.d.	n.d.	0.288	n.d.	n.d.	n.d.	n.d.	0.288
E8	821385	1900568	240	7.23	634	0.20	15.00	n.d.	0.20	12.7	n.d.	74.82	1.99	n.d.	n.d.	n.d.	0.032	n.d.	n.d.	n.d.	n.d.	0.032
E9	820887	1900838	109	7.58	766	0.07	23.59	n.d.	n.d.	23.4	n.d.	109.23	1.74	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
E10	820308	1901393	100	7	945	0.11	29.65	n.d.	n.d.	31.0	n.d.	175.75	2.38	n.d.	n.d.	n.d.	0.064	n.d.	n.d.	n.d.	n.d.	0.064
E11	820006	1900550	130	7.02	1038	0.12	46.25	n.d.	0.20	113.8	n.d.	173.92	1.99	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
I1	819302	1899887	46	7.38	600	0.14	18.33	n.d.	n.d.	53.7	n.d.	66.69	1.33	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
I2	819138	1900301	80	7.7	490	0.27	18.31	n.d.	n.d.	19.5	n.d.	40.86	1.90	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
I3	824545	1901408	110	6.99	1819	0.74	10.76	0.74	n.d.	0.0	n.d.	926.54	1.88	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
I4	823999	1901397	/	6.87	1215	0.10	57.84	n.d.	n.d.	9.9	n.d.	226.31	3.23	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
I5	820802	1900270	100	7.3	577	0.16	10.74	n.d.	n.d.	7.4	n.d.	45.61	1.45	n.d.	n.d.	n.d.	0.028	n.d.	n.d.	n.d.	n.d.	0.028
I6	817715	1901698	100	7.6	474	0.15	7.20	n.d.	n.d.	5.9	n.d.	16.68	2.42	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
I7	816662	1902063	120	7.13	836	0.27	31.64	n.d.	n.d.	37.5	n.d.	125.67	1.88	n.d.	n.d.	n.d.	0.024	n.d.	n.d.	n.d.	n.d.	0.024
I8	816540	1900995	25	7.27	673	n.d.	20.68	n.d.	n.d.	46.7	n.d.	87.26	1.91	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
I9	815789	1900095	80	7.19	630	0.06	23.85	n.d.	n.d.	13.9	n.d.	94.93	1.49	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
I10	822050	1906231	99	7.22	510	n.d.	5.34	n.d.	n.d.	0.0	n.d.	18.80	1.59	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
I11	821819	1901902	220	7.18	609	0.22	9.24	0.18	n.d.	0.0	n.d.	40.73	1.97	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
L1	822386	1904429	90	7.23	586	0.26	6.96	n.d.	n.d.	0.0	n.d.	70.53	1.62	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
L2	821100	1905038	120	7.12	622	0.12	9.51	n.d.	n.d.	0.0	n.d.	42.57	1.99	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
L3	820464	1904515	130	6.93	871	n.d.	32.31	n.d.	n.d.	58.2	n.d.	113.81	2.67	n.d.	n.d.	n.d.	0.208	0.022	n.d.	n.d.	n.d.	0.23
L4	820489	1904414	130	7.15	595	0.14	6.98	n.d.	n.d.	0.0	n.d.	29.05	1.86	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
L5	820240	1903282	84	7.23	665	0.22	16.62	n.d.	n.d.	23.1	n.d.	93.83	1.48	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
L6	819111	1899219	80	7.23	560	0.16	13.33	n.d.	0.20	45.8	n.d.	69.07	1.10	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
L7	821453	1901889	80	7.44	555	n.d.	9.97	n.d.	n.d.	0.0	n.d.	39.85	1.82	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
L8	825636	1901502	/	7.23	1050	0.66	23.78	n.d.	n.d.	19.2	n.d.	271.55	1.87	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
L9	820587	1905765	110	7.14	864	0.26	23.59	n.d.	n.d.	6.7	n.d.	193.53	1.95	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
L10	819353	1903381	90	7.24	706	0.15	25.09	n.d.	n.d.	47.4	n.d.	115.31	1.21	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
L11	819713	1902633	80	7.32	667	0.18	14.80	n.d.	n.d.	11.2	n.d.	72.39	1.24	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

n.d. : non détecté - d. :détecté - / : non renseigné

N	X L2-é (m)	Y L2-é (m)	prof.	pH	cond. 25°C	F	Cl	NO2	Br	NO3	PO4	SO4	DOC	dia	DEA	atrazine	DET	terbuthyl azine	simazine	diuron	oxadixyl	somme pesticides
	(m)	(m)	(m)		(µS/cm)	mg/l								µ/l								
M1	820412	1903787	160	7.26	605	0.16	8.17	0.13	n.d.	0.0	n.d.	27.88	1.49	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
M2	821034	1904597	100	7.42	588	n.d.	13.28	n.d.	n.d.	0.0	n.d.	64.23	1.28	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
M3	819094	1902466	100	7.34	606	n.d.	11.61	0.11	n.d.	11.8	n.d.	71.13	1.48	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
M4	819384	1902048	15	7.52	957	0.15	46.50	n.d.	n.d.	11.5	n.d.	158.97	2.19	n.d.	n.d.	n.d.	0.04	0.027	n.d.	n.d.	n.d.	0.067
M5	819593	1903933	60	7.2	682	0.07	16.34	0.19	n.d.	3.7	n.d.	97.93	0.99	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
M6	819802	1902016	110	7.36	556	n.d.	7.80	n.d.	n.d.	8.9	n.d.	50.00	1.03	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
M7	818973	1900840	180	7.33	724	n.d.	22.63	0.13	n.d.	38.2	n.d.	101.30	1.44	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
M8	818963	1900960	15	7.16	1014	0.17	35.50	0.21	n.d.	22.0	n.d.	229.33	1.90	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
M9	818185	1901187	80-100	7.27	819	n.d.	33.18	n.d.	n.d.	97.9	n.d.	161.22	1.40	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
N1	809707	1938549	80	7.16	460	0.13	4.32	n.d.	n.d.	1.6	n.d.	7.58	1.22	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
N2	808007	1937797	70	7.23	497	0.12	3.75	n.d.	n.d.	0.0	n.d.	12.12	1.69	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
N3	811827	1937246	150	7.25	480	0.06	5.34	n.d.	n.d.	10.6	n.d.	8.61	2.36	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
N4	811217	1938839	80	7.11	649	0.10	14.88	n.d.	n.d.	21.4	n.d.	52.49	1.79	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
N5	810789	1937249	80	7.03	559	n.d.	8.69	n.d.	n.d.	14.2	n.d.	18.05	0.79	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
N6	810673	1936592	100	7.23	519	0.17	11.17	0.25	n.d.	22.7	n.d.	31.13	1.51	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d
N7	809734	1937327	15	7.22	573	n.d.	12.92	n.d.	n.d.	9.0	n.d.	30.33	1.17	n.d.	n.d.	n.d.	n.d.	n.d.	0.056	n.d.	n.d.	0.056
N8	809409	1936079	90	7.23	536	n.d.	14.55	n.d.	n.d.	31.3	n.d.	28.20	0.99	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
N9	810109	1936523	15	7.08	657	0.09	22.37	n.d.	n.d.	24.0	n.d.	65.55	1.33	n.d.	n.d.	n.d.	0.032	n.d	0.044	n.d.	n.d.	0.076
N10	809895	1935375	70	7.17	591	0.08	15.00	n.d.	n.d.	32.0	n.d.	50.50	0.89	n.d.	n.d.	n.d.	0.157	0.029	0.038	n.d.	n.d.	0.224
N11	808375	1934656	100	7.08	595	n.d.	15.90	n.d.	n.d.	30.0	n.d.	40.07	0.65	n.d.	n.d.	n.d.	0.18	0.029	n.d.	n.d.	n.d.	0.209
O1	809129	1934529	80	7.06	611	0.18	20.28	n.d.	n.d.	37.5	n.d.	60.07	0.96	n.d.	n.d.	n.d.	0.121	0.02	n.d.	n.d.	n.d.	0.141
O2	809573	1934099	120	7.15	535	n.d.	12.29	n.d.	n.d.	20.6	n.d.	15.09	1.71	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
O3	808768	1934202	80	7.16	496	n.d.	10.37	n.d.	n.d.	21.6	n.d.	20.11	1.51	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
O4	808408	1933602	63	7.24	416	n.d.	4.39	n.d.	n.d.	12.2	n.d.	3.22	1.51	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
O5	808474	1932340	30	7.38	571	0.13	19.48	n.d.	n.d.	0.0	n.d.	42.67	1.02	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
O6	807539	1932623	80	7.17	627	n.d.	19.78	n.d.	n.d.	33.9	n.d.	52.60	1.35	n.d.	n.d.	n.d.	n.d.	0.168	0.026	n.d.	n.d.	0.194
O7	807610	1931193	50	7.25	565	n.d.	14.58	n.d.	n.d.	10.4	n.d.	16.74	0.95	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
O8	807579	1932170	55	7.33	462	0.13	11.10	n.d.	n.d.	14.1	n.d.	25.18	1.21	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
O9	806251	1931203	/	7.17	659	0.17	14.81	n.d.	1.3	3.8	n.d.	55.35	1.46	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
O10	807729	1930542	120	7.4	503	n.d.	4.98	n.d.	n.d.	0.0	n.d.	5.88	1.87	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
O11	808541	1931442	45	7.13	616	n.d.	18.27	n.d.	n.d.	25.3	n.d.	20.78	0.85	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P1	809918	1932810	63	7.14	563	n.d.	11.50	n.d.	n.d.	34.6	n.d.	30.11	0.80	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P2	807316	1929092	100	7.47	442	0.08	4.92	n.d.	n.d.	0.0	n.d.	5.06	0.63	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P3	807585	1929575	80	7.02	483	n.d.	6.61	n.d.	n.d.	9.1	n.d.	6.03	0.99	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P4	806926	1929927	70	7.31	469	n.d.	7.70	n.d.	n.d.	0.7	n.d.	11.69	0.60	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P5	808443	1936717	55	6.96	654	0.07	13.87	n.d.	n.d.	28.6	n.d.	49.33	1.03	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P6	805569	1929806	196	7.52	441	0.15	6.35	n.d.	n.d.	0.0	n.d.	14.81	0.74	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P7	805053	1928608	90	7.44	443	0.09	5.46	n.d.	n.d.	0.0	n.d.	11.14	0.63	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P8	806901	1928482	100	7.61	478	0.20	6.60	n.d.	n.d.	0.0	n.d.	23.53	0.60	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P9	805903	1926850	120	7.5	571	0.08	8.85	n.d.	n.d.	0.0	n.d.	67.23	0.69	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P10	804722	1927151	80	7.47	499	0.10	13.28	n.d.	n.d.	0.0	n.d.	36.32	0.69	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P11	805760	1927671	50	7.57	556	0.06	24.83	n.d.	n.d.	0.0	n.d.	59.04	0.79	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
U1	804163	1920238	83	7.33	515	0.12	7.94	n.d.	n.d.	0.0	n.d.	40.41	0.75	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
U2	803590	1917746	78	7.24	546	0.08	13.78	n.d.	n.d.	0.0	n.d.	52.94	0.84	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
U3	802667	1919425	30	7.53	474	0.21	18.08	n.d.	n.d.	15.4	n.d.	41.12	0.81	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
U4	804728	1919139	40	7.47	503	0.13	6.41	n.d.	n.d.	0.0	n.d.	22.62	0.84	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
U5	804280	1918175	60	7.16	647	0.22	42.98	n.d.	n.d.	75.4	n.d.	64.66	1.05	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
U6	803995	1917028	30	7.57	808	0.09	33.59	n.d.	n.d.	60.2	n.d.	145.30	0.87	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
U7	804752	1919872	5	7.7	660	0.14	14.71	n.d.	n.d.	13.2	n.d.	82.07	1.23	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

n.d. : non détecté - d. :détecté - / : non renseigné

A.3. Site de Portneuf

N	UTM Nad83 (m)	UTM Nad83 (m)	prof.	pH	cond. 25°C	Cl	NO3	SO4	DOC	DEA	atrazine	Métribuzine	Linuron	somme pesticides
	(m)	(m)	(m)		(µS/cm)	(mg/l)				(µg/l/l)				
DK1	757537	5184227	12	6.05	39	0.70	0.34	6.90	1.674	0.02	n.d.	n.d.	n.d.	0.02
DK2	755705	5183497	12	5.65	146	13.74	47.47	23.72	0.594	0.02	0.02	n.d.	d.	0.04
DK3	756861	5184983	18	6.61	120	21.73	14.26	9.11	1.136	0.02	0.02	n.d.	n.d.	0.04
DK4	752697	5189002	80	6.48	159	10.86	60.10	29.81	0.241	n.d.	n.d.	0.4	n.d.	0.40
DK5	751473	5186799	/	6.92	59	3.00	15.89	7.88	< 0.05	0.02	n.d.	0.04	n.d.	0.06
DK6	755906	5192401	15	5.95	268	15.89	103.05	53.63	1.581	0.02	0.03	0.72	n.d.	0.77
DK7	751915	5187422	15	6.28	21	0.47	0.64	7.39	0.294	n.d.	n.d.	n.d.	n.d.	n.d.
DK8	752360	5188116	75	6.77	156	12.54	75.75	9.64	0.134	n.d.	n.d.	n.d.	d.	0
DK9	754912	5193219	15	5.35	165	29.59	33.44	26.26	2.5	0.02	0.03	0.04	n.d.	0.09
DK10	755181	5191485	12	7.05	187	36.33	21.51	11.42	3.095	0.03	n.d.	0.042	d.	0.07
DK11	755445	5191954	12	6.01	185	16.47	74.21	28.67	1.464	n.d.	n.d.	1.17	n.d.	1.17
DK12	753166	5189684	15	6.53	195	12.82	70.72	28.99	0.423	n.d.	0.02	0.12	d.	0.15
DK13	755288	5191727	/	6.06	223	15.51	56.67	47.66	2.411	0.15	0.09	0.09	d.	0.33
DK14	754764	5191362	/	5.54	65	2.29	13.71	19.54	0.48	0.04	0.05	0.04	d.	0.13
DK15	754705	5190948	15	6.05	179	11.48	69.26	34.50	0.336	0.02	n.d.	0.07	n.d.	0.09
DK16	754343	5190591	60	7.55	108	5.67	7.78	28.55	0.123	n.d.	n.d.	n.d.	n.d.	n.d.
DK17	753693	5190151	35	5.66	170	33.04	55.92	13.50	0.567	0.04	0.05	0.21	n.d.	0.30
DK18	754204	5183234	3	6.47	170	14.01	57.90	16.32	0.743	0.03	0.03	n.d.	d.	0.06
DK19	756214	5192901	/	6.95	268	13.61	76.19	32.75	0.2	n.d.	n.d.	0.10	d.	0.10
DK20	744917	5199246	14	6.11	33	2.87	3.15	6.35	0.905	0.02	0.03	n.d.	d.	0.05
DK21	745187	5191183	/	6.58	148	31.04	3.58	14.82	3.948	n.d.	n.d.	n.d.	n.d.	n.d.
DK22	745212	5191256	150	8.62	95	1.32	< 0.15	0.26	1.553	n.d.	n.d.	n.d.	d.	n.d.
DK23	745439	5190991	16	5.87	125	17.45	40.94	14.43	0.603	n.d.	n.d.	0.06	d.	0.06
DK24	744970	5195097	12	6.8	157	8.40	30.03	30.38	1.98	0.02	0.04	0.61	n.d.	0.67
DK25	756254	5192604	60	5.83	254	55.54	107.73	23.38	0.668	0.03	n.d.	0.15	d.	0.18
DK26	756445	5184770	20	5.67	133	11.02	55.45	18.74	0.381	0.03	0.04	n.d.	d.	0.07
DK27	754548	5183425	4.5	6.34	155	17.87	52.76	18.57	0.719	n.d.	n.d.	n.d.	n.d.	n.d.
DK28	751463	5186793	8	7.14	53	2.87	16.38	7.27	0.143	0.03	n.d.	0.06	d.	0.09
DK29	746355	5189244	10	6.27	53	4.13	13.20	5.68	0.108	n.d.	n.d.	n.d.	d.	n.d.
DK30	745986	5189793	8	5.71	145	16.12	60.44	16.80	2.382	n.d.	n.d.	0.04	n.d.	0.04
DK31	745345	5194819	18	5.52	147	14.37	62.75	21.85	1.403	n.d.	n.d.	0.8	d.	0.8
DK32	746117	5189478	10	5.5	164	16.48	62.34	26.80	1.059	0.03	0.04	0.05	d.	0.12
DK33	733888	5190280	30	5.96	209	11.43	88.91	27.29	1.312	0.03	0.04	n.d.	n.d.	0.07
DK34	733893	5193187	15	5.94	169	14.10	70.69	18.93	1.285	n.d.	n.d.	n.d.	d.	n.d.
DK35	723941	5177144	4	6.55	45	5.96	2.13	6.08	0.796	n.d.	n.d.	n.d.	n.d.	n.d.
DK36	726894	5181764	25	5.78	66	4.04	24.60	14.53	1.11	n.d.	n.d.	n.d.	d.	n.d.
DK37	732988	5192779	20	5.9	44	1.15	9.85	2.24	2.142	n.d.	0.05	n.d.	d.	0.05
DK38	734542	5194015	16	6.15	183	18.24	83.61	10.33	1.835	n.d.	n.d.	0.71	d.	0.71
DK39	730292	5186881	10	5.6	343	30.05	192.61	22.80	1.301	0.14	0.06	n.d.	d.	0.2
DK40	724086	5177146	15	5.83	244	25.41	108.95	19.33	1.994	0.04	0.02	n.d.	d.	0.06
DK41	722777	5176971	125	7.6	317	51.79	40.45	33.90	0.879	0.03	0.04	n.d.	n.d.	0.07
DK42	722778	5175135	15	5.79	97	5.97	27.40	23.02	0.826	n.d.	n.d.	n.d.	d.	n.d.
DK43	730911	5188474	32	6.25	261	20.51	126.95	32.32	0.551	n.d.	n.d.	0.07	n.d.	0.07
DK44	724845	5184014	/	6.51	100	4.65	14.50	8.06	0.895	n.d.	n.d.	n.d.	n.d.	n.d.
DK45	724379	5179712	10	5.98	215	17.12	94.06	9.49	2.044	0.06	0.02	n.d.	n.d.	0.08
DK46	722961	5177250	14	5.75	96	4.89	33.03	9.71	1.931	n.d.	n.d.	n.d.	d.	n.d.
DK47	721770	5175870	18	6.32	211	43.56	27.73	14.64	1.461	0.03	0.04	0.04	d.	0.11
DK48	723320	5175588	8	5.31	120	11.54	27.35	28.44	1.098	n.d.	n.d.	0.07	d.	0.07
DK49	731699	5191667	20	5.11	208	15.59	82.61	42.07	1.603	n.d.	0.01	0.08	n.d.	0.09
DK50	736914	5196197	5	6.34	173	19.49	20.78	26.21	3.153	n.d.	n.d.	0.12	d.	0.12

n.d. : non détecté - d. :détecté - / : non renseigné

ANNEXE B – Analyses non présentées

B.1. Seuils de classification

B.2. Classification à trois classes

B.3. Autres approches de classifications

B.1 Seuils de classification

Dans le document, les échantillons sont divisés en deux classes :

- la classe 1 qui correspond aux échantillons dont les concentrations en pesticides sont supérieures au seuil limite de quantification ;
- la classe 0 qui correspond aux échantillons dont les concentrations en pesticides sont inférieures au seuil limite de quantification.

D'autres limites de classes ont préalablement été testées afin d'évaluer si elles permettaient une meilleure discrimination des échantillons :

- seuils à 0.02 µg/l , 0.05µg/l et 0.1 µg/l/l pour la somme des pesticides → ces différents seuils présentaient tous des performances moins bonnes que pour la limite de quantification de l'ordre de 72%, 67% et 65% en moyenne respectivement pour les trois limites.

- seuil à 0.1 µg/l/l pour 1 composé ou à 0.5µg/l pour la somme des pesticides → ces seuils correspondent à la norme européenne et permettaient d'évaluer si les réseaux pouvaient nous permettre de détecter avec des performances satisfaisantes les échantillons présentant des concentrations en pesticides supérieures à la norme. Les performances pour cette approche étaient inférieures à 70% de bons classements. Il faut noter que le nombre d'échantillons dont les concentrations sont supérieures à la norme est relativement faible par rapport au nombre total de prélèvements, ne permettant peut-être pas un apprentissage optimal.

B.2 Classification à trois classes

Des systèmes de classification à trois classes ont également été testés :

- classe 1 : échantillons dont la concentration était supérieure à la norme européenne (0.1 µg/l/l pour 1 composé ou à 0.5µg/l pour la somme des pesticides)
- classe 0 : échantillons dont les concentrations en pesticides sont inférieures à la limite de quantification

- classe intermédiaire : échantillons dont les concentrations sont inférieure à la norme mais supérieure à la limite de quantification.

Les performances étaient satisfaisantes pour les deux classes extrêmes mais faibles pour la classe intermédiaire.

B.3 Autres approches de classifications

Communication présentée le 11 avril 2007 à Barcelone. *International Conference on WAter Pollution in naturl POrus media at different scale.WAPO2* (11-13 avril 2007)

PREDICTION OF PESTICIDE OCCURRENCE IN DOMESTIC WELLS USING CHEMICAL INDICATORS

C. Doukouré[1,2], O. Banton[1] and P. Lafrance[2]

(1) Hydrogeology Laboratory, University of Avignon, 33 rue Louis Pasteur 84000, France
Email: Cecile.doukoure@univ-avignon.fr, Olivier.banton@univ-avignon.fr
(2) Institut national de la recherche scientifique-Centre Eau, Terre et Environnement (INRS-ETE), 490 rue de la Couronne, Québec, QC, Canada G1K 9A9
Email: Cecile.doukoure@ete.inrs.ca, Pierre.lafrance@ete.inrs.ca

ABSTRACT

In this study, a method is proposed allowing the prediction of the potential presence of agricultural pesticides in well water. Wide scale sampling campaigns were carried out at three sites with different agronomic, pedologic and climatic characteristics. We thus gathered nearly 250 water samples on which the pesticides most frequently detected were measured as well as main chemical parameters. First, the use of Receiver Operating Curves (ROC) allowed us to highlight similar chemical characteristics for the samples for which pesticides were detected. Indeed, for the three sites, the same five parameters present discrimination capacities with regard to the detection or no detection of pesticides in the sample. This allows us to select input variable for further analysis. Secondly and starting from these indicators, four methods of classification with predictive approaches were tested. The prediction quality on validation data sets enabled us to correctly predict up to 80% the wells that showed detectable agricultural pesticides. Such methods could increase appreciably the performance of monitoring schemes by not targeting large zones to be controlled, but directly the specific wells that might be contaminated.

Key words: Groundwater contamination, pesticides, supervised classification, chemical indicators, regional survey

1. INTRODUCTION

It is now recognized that ground water is prone to pesticide contaminations as well in Europe than in North America. This trend is becoming alarming since this resource constitutes the main drinking water supply for rural population.

Whereas analyses to detect the presence of certain contaminants such as nitrates in private wells are frequent, pesticides analysis are fragmentary because of the costs involved in such surveys. The systematic sampling of the domestic wells on a broad scale is often limited by the high cost of the analyses and by the great diversity of compounds used in agriculture. Solute transport models are available to predict with appreciate adequacy the movement of pesticides from the land surface to ground water. But these models require a detailed description of the physical environment that is time and space-dependant and difficult to set up on a regional scale. In addition, these models cannot predict with accuracy for the pesticides in a given well.

Several studies have shown correlation between pesticides detection and some chemical parameters concentration such as nitrates or sulphates (Istok and Rautman 1996; Kolpin *et al.* 1998; Burow *et al.* 1998). But no studies have tried to combine several chemical parameters to predict pesticides occurrence in private wells.

2. OBJECTIVES

The main objective of this study is to predict the potential occurrence of pesticide in well water by using current and easily monitored chemical parameters defined here as indicators. Specifics task to meet this objectives are (1) to find some relevant parameters that could be related to the presence of pesticides; and (2) to test several classification methods to evaluate the ability of these indicators to discriminate wells showing detection and no detection of pesticides.

3. RESULTS AND DISCUSSION

3.1. Variable selection

To achieve this work, three wide-scale sampling campaigns were carried out between June and October 2005 at three sites with different agronomic, pedologic and climatic characteristics. We thus gathered 245 water samples on which the pesticides most frequently detected (97% of compounds detection on a regional base) were measured as well as main chemical parameters (Table 1).

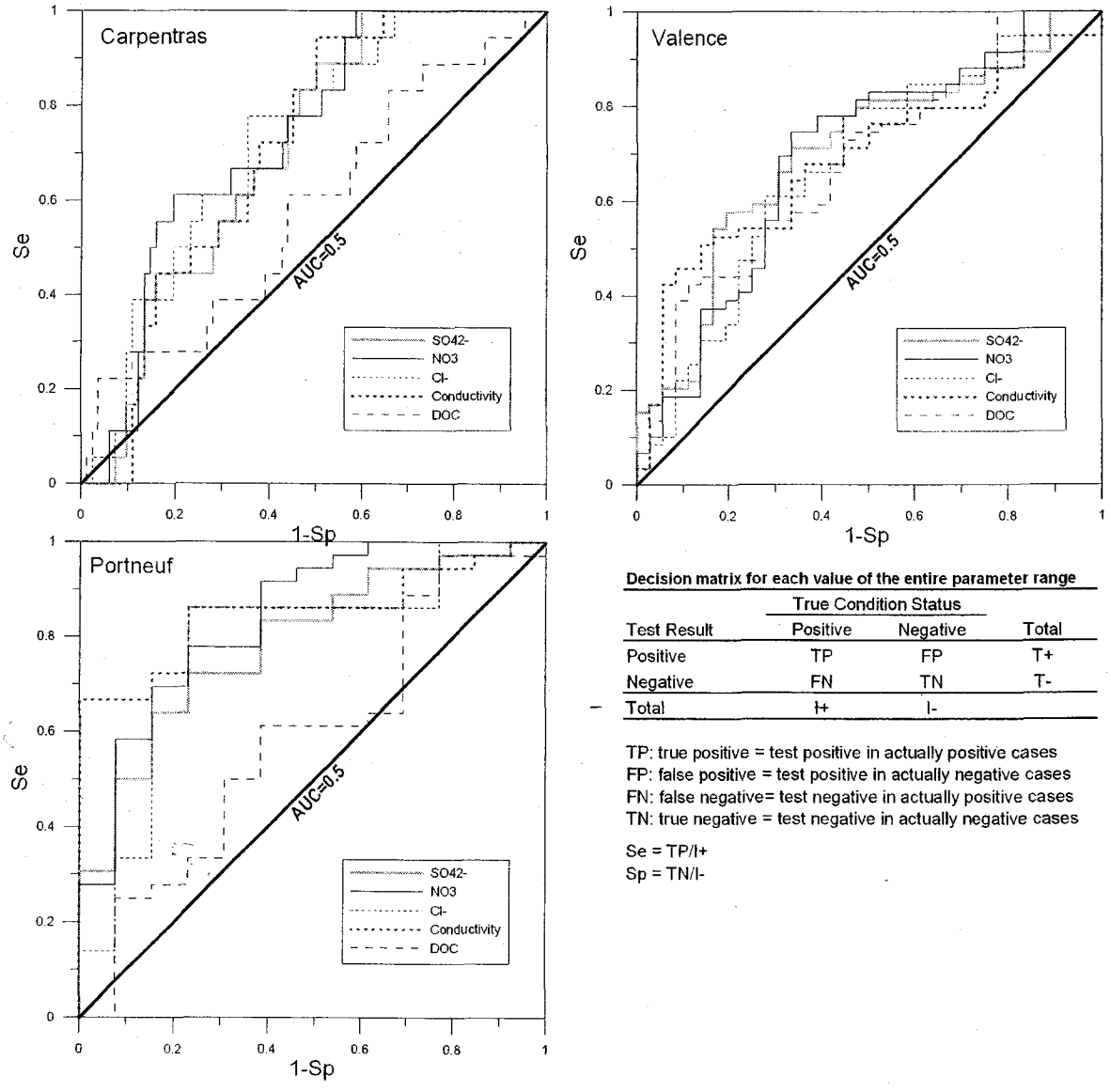
Table 12. Sampling sites characteristics

Sites	Sampling period	Number of samples		Pesticides	Other parameters
		Total	With pesticides		
Valence - France	June 2005	95	59	Atrazine, DEA, diuron, terbuthylazine, DET, simazine	Anions (NO3-, Cl-, SO42-,Br-, F-) DOC, temperature, pH, well depth, electrical conductivity
Carpentras - France	July 2005	100	18	Atrazine, DEA, diuron, terbuthylazine, DET, simazine	
Porneuf - Quebec	October 2005	50	45	Atrazine, DEA, metolachlore, metribuzine, linuron	

In order to determine which parameters can be used as discriminators, Receiver Operating Characteristics (ROC) curves were used. There is a number of reviews describing the proper use of ROC curves in statistical topics (Bradley 1997). It has been used in many scientific areas such as epidemiology or biomedical informatics as a means to separate a population into two subset.

In case of binary test when we use one value to separate a population into two subset, the accuracy is commonly assessed using measures of sensitivity Se and specificity Sp. The sensitivity represent the rate of true positive and the specificity represent the rate of true negative.For continuous tests, there is no particular value of sensitivity or specificity that characterizes the overall accuracy of the test, but rather an entire range of values that vary depending on what we use as the threshold for discretizing the test result. The ROC curve captures in a single graph the trade-off between a test sensitivity and specificity over its entire range. The ROC curve plots Se vs. (1 -Sp) of a test as the threshold varies over its entire range. Each data point on the plot represents a particular setting of the threshold, and each threshold setting defines a particular set of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) counts, and consequently a particular pair of Se and (1 - Sp) values (Fig.1).

The test performance is measured by the area under the ROC curve (AUC). AUC is interpreted as the average value of sensitivity for all possible values of specificity. An area of 1 represents a perfect test and an area of 0.5 (diagonal line) corresponds to random forecast. It is recommended to select the variables that present an AUC more than 0.50 that graphically represent curves upper the diagonal line.



Decision matrix for each value of the entire parameter range

	True Condition Status		
Test Result	Positive	Negative	Total
Positive	TP	FP	T+
Negative	FN	TN	T-
Total	I+	I-	

TP: true positive = test positive in actually positive cases
FP: false positive = test positive in actually negative cases
FN: false negative= test negative in actually positive cases
TN: true negative = test negative in actually negative cases

Se = TP/I+
Sp = TN/I-

Figure 1. ROC curves for parameters presenting AUC greater than 0.5

For the three sites, the ROC curves were applied for each parameter in order to identify those to use for the classification. The results show that for the three sites, five parameters (nitrate, chloride, sulphate, electrical conductivity and dissolved organic carbon) present capacities at discrimination (Fig.1). The other measured parameters are under the line of random prediction and will thus not be used for further analyses.

Among these five parameters and except from particular physics or geological characteristic (presence of gypsum), nitrates, sulphates and chlorides in high concentrations announce a contamination. According to the objective of the study, this demonstrates the pertinence of these three parameters as indicators for pesticide occurrence in well water. Being strongly related to these three parameters electrical conductivity is also found to be a good indicator. However information could be redundant and can even cause problems of colinearity during the analyses. Dissolved organic carbon also present an AUC superior than 0.5, but with a lower discriminative capacity as compared to the other. The ability of DOC to complex-certain pesticides and thus to affect their mobility (Lafrance *et al.* 1990) can justify its presence as an indicator.

3.2. Classification analysis

In order to investigate the ability of the fives selected indicators to predict pesticide occurrence in well water, several supervised classification methods were tested. The purpose of these methods is to predict the classes to which objects belong starting from descriptive features. For each one of the three sites, linear discriminant analysis, logistic regression and K-nearest-neighbour were tested. Results are shown for the cross-validation. Artificial Neural Networks (ANN) were also tested. As it needs numerous observations but no requirement about data distribution, observations of all the three sites were used.

For all analysis, output classes are detection and no detection of pesticides and input continuous variables are the five indicators previously selected. All input variables were scaled to the range [0,1] using a linear transformation making the minimum value zero and the maximum value 1.

Table 2 : Cross-validation accuracy for the three sites using three supervised classification techniques

		Sites		
Technique		Valence	Carpentras	Portneuf
Linear discriminant analysis	Accuracy	0.62	0.79	0.77
	Positive accuracy	0.38	0.01	0.86
	Negative accuracy	0.76	0.96	0.50
Logistic regression	Accuracy	0.61	0.79	0.72
	Positive accuracy	0.36	0	0.78
	Negative accuracy	0.74	0.96	0.53
K-nn	Accuracy	0.64	0.78	0.78
	Positive accuracy	0.46	0.07	0.88
	Negative accuracy	0.75	0.92	0.48

Table 3 : Accuracy of training, validation and test sets of a feed-forward back-propagation neural network

	Data set		
	Train	Verify	Test
Accuracy	0.81	0.85	0.78
Positive accuracy	0.80	0.88	0.77
Negative accuracy	0.82	0.82	0.79

Results for the three statistical methods (Table 2) present poor (acc. = 0.61) to good (acc. = 0.79) classification accuracy. But accuracy is not sufficient to illustrate the classifiers performance. Specific accuracy (positive and negative accuracies) that represents the rate of correct classification for each of the classes indicates that classification supports one of the classes. Indeed, by not exceeding 0.53 for the two classes, specific accuracy indicates that one class is always undervalued compare to the other. This is due to the bad repartition of the classes in the input data, or in the learning set. Indeed, the Carpentras data just have 20% of detection in the samples and the Portneuf data have 10% of no detection. This problem may be improved by test with data presenting better repartition of the two classes.

Table 3 presents the results for the feed-forward back-propagation neural network with one hidden layer. As it is recommended with dataset that present sufficient observations ({Hastie, Tibshirani, *et al.* 2001 #1410}), the three sites observations were randomly divided in a training set, a validation set and test set. The training set is used to fit the model; the validation set is used to estimate prediction error for model selection and the test set is used for assessment of the generalization error of the final chosen model. Classification presents a good accuracy for all the sets and does not support one of the classes (specific accuracy > 0.77).

The technique of the neural networks presents better results that the other statistical tools and enables us to predict up to 80% the wells in which pesticides were detected by using only the 5 selected parameters. That shows that the selected indicators are relevant and that the limit is the choice of the statistical tools used and the type of data. Indeed, for the learning step, data needs to present a sufficient number of observations and the output classes need to be equitably represented. Thus, by observing these conditions and by using neural networks, the five selected indicators can be used to discriminate wells with or without pesticides detection with a good accuracy.

4. CONCLUSION

Using the ROC curves, we have seen that some chemical parameters present an individual discriminative capacity. This capacity is however not sufficient for predicting pesticide occurrence in ground water. By combining them together and by using some statistical supervised methods, we can predict up to 80% the pesticide occurrence in well water. Such a method can appreciably increase the performance of monitoring schemes by identifying wells that might be contaminated by agricultural pesticides.

Acknowledgements: The authors would like to thank Rémi de la Vaissière and Frédéric Lalbat for the sampling campaigns in Valence and Carpentras areas and Pauline Fournier for the help with pesticides analysis. This study was fund by the Fonds québécois de la recherche sur la nature et les technologies (FQRNT, Gouvernement du Québec) and by the Rhône – Méditerranée & Corse Water Agency.

REFERENCES

Bradley, A.P. (1997): The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30: 1145-1159.

Burow, K.R., Shelton J.L., and Dubrovsky N.M. (1998): Occurence of nitrate and pesticides in ground water beneath three agricultural land-use settings in the eastern San Joaquin Valley, California, 1993-1995. U.S. Geological Survey, Report 97-4284. 51 p.

Hastie, T.,Tibshirani, R., and Friedman, J. (2001): *The elements of statistical learning*. Springer, New-York, 533pp.

Istok, J.D., and Rautman, C.A. (1996): Probabilistic assessment of ground-water contamination: 2. results of case study. *Ground Water*, 34: 1050-1064.

Lasko, T.A., Bhagwat, J.G., Zou, K.H., and Ohno-Machado, L. (2005): The use of receiver operating

characteristic curves in biomedical informatics: Clinical Machine Learning. *Journal of Biomedical Informatics*, 38: 404-415.
Kolpin, D.W., Thurman, E.M. and Linhart, S.M. (1998): The environmental occurrence of herbicides: the importance of degradates in ground water. *Archives of Environmental Contamination and Toxicology*, 35: 385-390.
Lafrance, P., Banton, O., Campbell, P.G.C. and Villeneuve, J.P. (1990): A complexation adsorption model describing the influence of dissolved organic-matter on the mobility of hydrophobic compounds in groundwater. *Water Science and Technology*, 22: 15-22.