



**HAL**  
open science

# Elaboration d'un modèle de prédiction de la phytodisponibilité du cadmium dans les sols agricoles : application à la contamination cadmiée du blé dur

Yoann Viala

► **To cite this version:**

Yoann Viala. Elaboration d'un modèle de prédiction de la phytodisponibilité du cadmium dans les sols agricoles : application à la contamination cadmiée du blé dur. Biodiversité et Ecologie. Université de Bordeaux, 2018. Français. NNT : 2018BORD0091 . tel-01895809

**HAL Id: tel-01895809**

**<https://theses.hal.science/tel-01895809>**

Submitted on 15 Oct 2018

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

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

# THÈSE

présentée à

## L'UNIVERSITÉ DE BORDEAUX

École doctorale 304 Sciences et environnement

par **Yoann VIALA**

Pour obtenir le grade de

### Docteur

Spécialité : **Géochimie et écotoxicologie**

\*\*\*\*\*

## ÉLABORATION D'UN MODÈLE DE PRÉDICTION DE LA PHYTODISPONIBILITÉ DU CADMIUM DANS LES SOLS AGRICOLES : Application à la contamination cadmiée du blé dur

\*\*\*\*\*

Soutenance prévue le : 27/06/2018

Devant la commission d'examen formée de :

M. Olivier <b>ATTEIA</b>	Pr., HDR	ENSEGID Bordeaux	Président
Mme Sophie <b>CORNU</b>	DR, HDR	CEREGE Aix Marseille	Rapporteur
M. Philippe <b>CAMBIER</b>	DR, HDR	UMR ECOSYS INRA/AgroParisTech	Rapporteur
Mme Marilyne <b>SOUBRAND</b>	MC	UMR GRESE Limoges	Examinatrice
M. Yann <b>SIVRY</b>	MC	IPGP Paris	Examineur
M. Guérolé <b>GRIGNON</b>		Arvalis-Institut du végétal	Invité
Mme Valérie <b>SAPPIN-DIDIER</b>	CR	UMR ISPA/INRA Bordeaux	Directrice de thèse
M. Christophe <b>NGUYEN</b>	DR, HDR	UMR ISPA/INRA Bordeaux	Co-directeur de thèse







## Résumé

Le cadmium (Cd) est un élément trace présent dans les sols agricoles qui contamine la chaîne alimentaire en étant prélevé par les plantes et accumulé dans les produits végétaux consommés. La biodisponibilité du Cd est un concept au cœur de l'évaluation des risques de transfert excessif du Cd du sol vers les plantes. Les plantes prélevant essentiellement l'ion  $\text{Cd}^{2+}$  dans la solution de sol, la biodisponibilité est fonction de deux principaux processus, la spéciation du Cd en solution (les différentes formes chimiques prises par le Cd en solution) et la distribution du  $\text{Cd}^{2+}$  entre la phase solide et la solution. L'objectif principal de ce travail a été d'élaborer des modèles simples à visée opérationnelle prédictifs de la concentration en  $\text{Cd}^{2+}$  dans la solution de sols agricoles, en modélisant soit la spéciation du Cd en solution de sol, soit la distribution phase solide-solution du  $\text{Cd}^{2+}$ , celle-ci permettant de renseigner en outre la capacité de la phase solide à réapprovisionner la solution de sol lors de l'absorption racinaire. Nous avons également recherché des modèles pour la prédiction des teneurs en  $\text{Cd}^{2+}$  retrouvées dans les grains de blé dur. Nous avons développé deux approches de modélisation. La première, statistique, permet de produire des modèles simples à visée opérationnelle. La seconde, géochimique, permet de comprendre les mécanismes dominants et donc de juger de la cohérence de modèles statistiques simples pour représenter des processus physico-chimiques complexes. Ces deux approches ont montré de manière cohérente que pour les sols agricoles faiblement contaminés, le Cd qui s'échange entre la phase solide et la solution est vraisemblablement sorbé faiblement et peut-être estimé par le Cd extrait par  $\text{NH}_4\text{NO}_3$  1 M minoré par une fraction fixée à des oxydes de manganèse. Les modélisations ont également montré l'importance du pH et de la teneur en Ca en solution comme variables contrôlant la solubilité de  $\text{Cd}^{2+}$ , probablement en raison de leur rôle à régir la disponibilité des sites de sorption et de complexation vis-à-vis du Cd. Le modèle statistique le plus performant pour prédire la teneur en Cd dans le grain de blé dur reprend également ces variables, suggérant ainsi que les modèles statistiques simples de prédiction de la spéciation en solution et de distribution sol-solution du Cd sont des modèles pertinents pour estimer la biodisponibilité et qu'ils peuvent permettre de classer des sols en fonction des risques de transfert du Cd du sol vers une culture. Par rapport à l'évolution du contexte réglementaire, le modèle statistique prédictif de la teneur en Cd dans le grain a montré par validation croisée qu'il pourrait discerner de façon assez fiable (88 %) des différences de  $0.05 \text{ mg Cd.kg}^{-1}$  de grain et que sa fiabilité serait moindre (65 %) pour des différences de  $0.025 \text{ mg Cd.kg}^{-1}$ .

**Mots clés :** Cadmium, Biodisponibilité, Modélisation, Spéciation, Sorption, Distribution



## ***Abstract***

Cadmium (Cd) is a trace element found in agricultural soils which can contaminate the food chain by being taken up by plants and accumulated in consumed plant products. The bioavailability of Cd is a concept at the centre of the risk assessment of Cd transfer from soil to plants. Plants, taking up essentially the free form of Cd ( $\text{Cd}^{2+}$ ) in the soil solution, bioavailability is a function of two main processes, the Cd speciation in solution (the different chemical forms taken by the Cd in solution) and the partitioning of  $\text{Cd}^{2+}$  between the solid phase and the solution. The main objective of this work was to develop simple predictive operational models of  $\text{Cd}^{2+}$  concentrations in agricultural soil solution, by modelling either Cd speciation in soil solution or  $\text{Cd}^{2+}$  solid-solution partitioning, the latter to further inform the ability of the phase to replenish the soil solution during root absorption. We also looked for models for the prediction of  $\text{Cd}^{2+}$  levels found in durum wheat grains. We have developed two modelling approaches. The first, statistical, allows to produce simple models for operational purposes. The second, geochemical, allows to understand the dominant mechanisms and thus to judge the coherence of simple statistical models to represent complex physicochemical processes. These two approaches have consistently shown that for poorly contaminated agricultural soils, the exchanged Cd between the solid phase and the solution is likely to be weakly sorbed and can be estimated by the Cd extracted by 1M  $\text{NH}_4\text{NO}_3$  minus a fraction attached to amorphous manganese oxides. Modelling also showed the importance of pH and Ca content in solution as variables controlling the solubility of  $\text{Cd}^{2+}$ , probably because of their role in controlling the sorption site availability and Cd complexation. The best-performing statistical model for predicting Cd content in durum wheat also picks up these variables, suggesting that simple statistical models for speciation in solution and soil-solution partitioning of Cd are relevant models to estimate bioavailability and that they can be used to classify soils according to the risks of the transfer of soil Cd to a crop. Compared to the evolution of the regulatory context, the statistical model predictive of the Cd content in the grain shown by cross validation that it could discern relatively reliably (88%) the differences of  $0.05 \text{ mg Cd.kg}^{-1}$  of grain and that its reliability would be less (65%) for differences of  $0.025 \text{ mg Cd.kg}^{-1}$ .

**Keywords** : Cadmium, Bioavailability, Modelling, Speciation, Sorption, Partitioning





## **Remerciements**

Cette première partie pour remercier toutes les personnes et organismes qui m'ont permis de réaliser ce doctorat et qui m'ont aidé au cours de ces années et sans lesquels ce travail n'aurait pas été possible. J'ai pris énormément de plaisir à travailler avec chacun d'entre vous et j'espère n'oublier personne !

Avec un profond respect, je tiens tout d'abord à remercier les rapporteurs et examinateurs, Olivier Atteia, Sophie Cornu, Philippe Cambier, Maryline Soubrand et Yann Sivry de m'avoir fait l'honneur d'évaluer ce travail ainsi que pour la discussion qui a suivie.

Merci également aux membres de mon comité de pilotage, Isabelle Lamy, Cédric Garnier et Jérôme Viers pour vos avis et conseils toujours très pertinents sur l'orientation de ce travail.

Un merci particulier à Benoît Méléard pour m'avoir fait confiance pour réaliser cette thèse mais aussi pour tes remarques justes et judicieuses au cours de ces années. Un grand merci également à Guénolé Grignon pour tes conseils sur le traitement de données et ta disponibilité que ce soit pour une question de stats ou pour ta participation au suivi et à la soutenance de cette thèse.

Merci également au département Environnement et Agronomie de l'INRA, à Arvalis - Institut du Végétal et à l'Agence Nationale de la Recherche pour avoir financé ces travaux.

Je tiens également à remercier toutes les personnes qui ont participé à la récolte d'échantillons de sols et de grains à travers différents projets de recherche (Aspitet, QUASAGRO, Cadur, CaDON, SOERE-PRO) et sans lesquels il aurait été difficile de réaliser ce travail.

Plus chaleureusement, un grand merci et toute ma reconnaissance à mes encadrants de thèse officiels Valérie Sappin-Didier et Christophe Nguyen mais aussi à Laurence Denaix et André Schneider qui ont pu également s'impliquer dans l'élaboration et l'évolution de ce travail. Merci de m'avoir fait confiance et m'avoir laissé l'opportunité de réaliser ce projet avec liberté. Valérie, je ne te serais jamais assez reconnaissant pour ta bienveillance et ta patience, dont tu as bien eu besoin pour m'encadrer, que ce soit en thèse ou en stage. Merci également d'avoir, à certains moments, stressé à ma place, ce qui m'a permis de me remettre en selle dans les moments de doute. Christophe, tu as dû également faire preuve d'une grande patience pendant ces quelques années, mais je souhaite te remercier pour ta rigueur de travail et ta capacité à prendre du recul, qui continuent de m'épater !

Laurence, merci d'avoir toujours été disponible et ravie de me voir pour « parler de science » et pour ton air déçu lorsque la question était purement administrative. Cette passion m'a fait le plus grand bien au cours de ces années et restera gravé dans ma mémoire dans les moments de doute. Et pour terminer André, le jeune retraité, avec qui j'ai pu découvrir la passionnante chimie du sol. Mon stage et les discussions interminables mais passionnantes sur le pouvoir tampon des sols, m'ont énormément appris en 6 mois et m'ont été plus que bénéfique pour réaliser ce travail. Merci à vous quatre pour votre encadrement et pour tout ce que vous m'avez enseigné et transmis, que ce soit professionnel ou personnel, pendant ces années avec vous.

Un grand merci à toutes les personnes qui se sont impliquées dans mon travail de thèse. Cécile Coriou et Sylvie Bussière, pour toute l'aide que vous m'avez apporté pendant mes années de manip, pour toutes les galères de machine et questions que vous avez supportées, et toutes les séries d'échantillons qui sont passées entre vos mains expertes, un énorme merci. Merci également pour votre présence toujours chaleureuse dans les labos qui ont rendu ces journées en blouse, beaucoup moins longues. Un merci particulier à Sylvie Millin pour ton aide sur les dosages très tardifs de carbone !

Merci également au reste de l'équipe BiogET, Jean-Yves Cornu, Thierry Robert, Stéphane Thunot et Valérie Nicaise pour votre bienveillance et pour avoir animé les réunions d'équipe !

J'aimerais remercier toutes les personnes d'ISPA, que j'ai pu croisé au cours de pauses cafés ou autre, que ce soit au B3, C2 ou même C1 et qui ont toujours été très chaleureuses et souriantes. Mention spéciale et un grand merci aux personnes avec qui j'ai pu jouer (et m'énerver) au volley : Tovo Alain M., Kiki (le vrai capitaine qui râle bien), Bylou, Amel (je fais soft en surnom), David, Thomas, Chipo et Delphine ! Merci de m'avoir permis de me défouler !

Les Tocards (je prends le parti du Grems sur l'orthographe), espèces aujourd'hui en voie de disparition à l'INRA (et c'est pas plus mal). Je sais pas si je dois vraiment vous remercier vis-à-vis du travail, parce qu'on en a perdu du temps avec toutes ces conneries, mais c'était du bon temps, donc ça va.

Fritoche, merci pour cette nonchalance qui te caractérise et qui t'as permis de me supporter au bureau. Sur cette même idée, La Vive, que je tiens à féliciter pour sa résilience à toute épreuve, bravo d'avoir enduré mes « râpages » pendant plus longtemps que les autres et les discussions sans fin avec Arthur dans le bureau. Mathi, qui nous a appris ce qu'était l'efficacité pour un moindre effort ! La Souille, merci pour ton énergie et ta grande gueule, ça me fatiguait bordel mais ça va me manquer ! Suze, la principale destinataire d'images ou de conneries animées (ça va pas s'arrêter de suite, t'inquiètes pas). Zaza, les discussions sur les hippies et les styles de vie alternatifs vont me manquer ! Le Grems, je ne me ferais jamais au fait que tu préfères R au fromage (et tu passeras le bonjour à Amandine). Et pour finir ce brave Kader (Gris-Poil et je tairais

les autres surnoms), jamais là les week-end mais toujours présent pour un petit café ou un barbeuc ! Merci à vous tous pour votre connerie, et on se revoit dans pas longtemps.

Un grand merci également à tous les non-permanents, encore présents dans l'unité ou déjà partis depuis longtemps qui n'ont pas été des tocards (je vous assure, c'est mieux), mais avec qui j'ai toujours pu discuter, rigoler et boire des coups et qui ont aussi énormément participé à la bonne ambiance générale au sein de l'unité, Bofang, Manon, Pietro, Marko, David, Mohamed, John, Julien, Evelyne, Rodolphe, Félix et tous les stagiaires, en particulier Fanny, Vincent et Inès pour leur participation à la récolte des données. Merci à tous pour ces bons moments partagés.

Merci également aux copains pour avoir toujours valorisé cette thèse même si c'était pas de la chimie et que ça parlait de radis, Michou, Simon, Vedov, Plot-plot, Juju, Mimix, Lélé, Caro et tous les gens du LCPO, beaucoup trop nombreux. Merci pour ces moments et les pétanques jusqu'à 3h à la lueur d'un spot de chantier.

Une phrase pour remercier la Laiterie, pour avoir toujours été ouverte, dès fois jusqu'à pas d'heures, pour boire une bière, jouer aux fléchettes, à la pétanque, aux cartes ou même rédiger dans des moments de doute.

Merci à mon meilleur ami, mes c\*\*\*\* qui sans jamais être présent était toujours là. Merci d'avoir partagé tes voyages et m'avoir fait te détester en voyant tes photos à mon bureau...

Et pour terminer, un grand merci à toute ma famille, pour avoir toléré mon « absence » pendant ces quelques années.

Merci à tous.



# Sommaire

<b>Introduction générale .....</b>	<b>1</b>
<b>Chapitre I. Étude bibliographique de la dynamique du Cd dans les sols .....</b>	<b>7</b>
1. Concept de la biodisponibilité .....	7
2. Nature des principaux constituants du sol impliqués dans la dynamique du Cd.....	8
2.1. Les phyllosilicates .....	9
2.2. Oxydes métalliques .....	10
2.3. Oxydes de fer .....	11
2.4. Oxydes de manganèse.....	12
2.5. Matières organiques du sol (MOS) .....	12
2.6. Minéraux carbonatés et sulfatés .....	14
3. Types de charges et réactivité des constituants du sol.....	15
4. Dynamique du Cd dans les sols .....	17
4.1. Spéciation du Cd en solution de sol.....	18
4.2. Facteurs influençant la spéciation du Cd .....	24
5. Estimation de la distribution sol-solution.....	26
5.1. Estimation de la fraction mobile du Cd.....	26
5.2. Estimation de la fraction libre de Cd en solution .....	31
6. Modélisation de la distribution sol-solution.....	32
6.1. Le coefficient de distribution .....	32
6.2. Modélisation empirique de la solubilité du Cd dans les sols.....	35
6.3. Modélisation géochimique .....	37
7. Objectifs et démarche de la thèse .....	41
<b>Chapitre II. Modélisation de la spéciation du Cd en solution.....</b>	<b>45</b>
1. Contexte .....	45
2. Matériels et méthodes.....	48
2.1. Sols et caractérisation de la solution de sol .....	48

2.2. Modélisation statistique de la fraction libre .....	49
2.3. Modélisation géochimique par CHEAQS Next® .....	53
<b>3. Résultats.....</b>	<b>57</b>
3.1. Propriétés des sols et des solutions de sol .....	57
3.2. Modélisation par régression statistique des concentrations en Cd <sup>2+</sup> de la solution de sol .....	60
3.3. Modélisation géochimique de la concentration en Cd <sup>2+</sup> de la solution de sol .....	65
3.4. Comparaison des approches pour la prédiction des concentrations en Cd <sup>2+</sup> en solution .....	69
<b>4. Discussion .....</b>	<b>72</b>
4.1. Quels paramètres de MOD pour la modélisation géochimique de la spéciation du Cd	72
4.2. Prédiction des concentrations en Cd <sup>2+</sup> en solution de sols agricoles .....	75
<b>5. Conclusion.....</b>	<b>76</b>
<b>Chapitre III. Étude de la distribution du Cd en sols agricoles et prédictions des concentrations en Cd<sup>2+</sup> en solution.....</b>	<b>79</b>
1. Contexte .....	79
2. Matériels et méthodes.....	83
2.1. Prélèvements des sols .....	83
2.2. Analyses des sols.....	83
2.3. Cd extrait par différentes solutions d'extraction .....	84
2.4. Récupération de la solution de sol.....	85
2.5. Modélisation de la concentration en Cd <sup>2+</sup> de la solution de sol par régressions linéaires multiples.....	86
2.6. Modélisation mécaniste de la distribution du Cd .....	87
<b>3. Résultats.....</b>	<b>92</b>
3.1. Caractéristiques des sols.....	92
3.2. Prédiction des concentrations en Cd <sup>2+</sup> et Cd total en solution par régressions linéaires multiples.....	96
3.3. Prédiction des concentrations de Cd <sup>2+</sup> par un modèle multisurfaces .....	101
3.4. Spéciation solide du Cd en sols agricoles non contaminés.....	105

4. Discussion .....	106
4.1. Cd extrait par extractions sélectives .....	106
4.2. Adéquation entre la quantité de Cd extrait et les surfaces de sorption à considérer dans un modèle géochimique .....	107
4.3. Convergences des approches statistiques et géochimiques pour la modélisation de la partition sol-solution du Cd en sols agricoles ? .....	109
5. Conclusion .....	112
<b>Chapitre IV. Modélisation statistique des teneurs en Cd dans le grain de blé dur</b> .....	<b>117</b>
<b>Partie 1 - Première élaboration de modèles statistiques prédictifs de la teneur en Cd dans le grain de blé dur</b> .....	<b>118</b>
1. Abstract.....	118
2. Introduction .....	119
3. Materials and methods.....	121
3.1. Sampling and preparation.....	121
3.2. Soil characteristics.....	122
3.3. Soil and soil solution extraction and analyses.....	122
3.4. DGT experiments .....	124
3.5. Testing statistical models from the literature .....	125
3.6. Statistical modelling.....	125
4. Results and discussion .....	126
4.1. Soil properties and Cd contents in plant materials .....	126
4.2. Is it possible to predict Cd content in durum wheat grains at maturity from shoot Cd content at the early heading stage?.....	129
4.3. Soil Cd extractions.....	129
4.4. Testing models from the literature for the prediction of grain Cd in durum wheat.....	131
4.5. Statistical models to predict Cd in durum wheat grain at maturity .....	133
5. Conclusion .....	138
<b>Partie 2 - Élargissement du domaine de variabilité des sols étudiés pour augmenter la robustesse des modèles</b> .....	<b>139</b>



1. Contexte .....	139
2. Matériels et méthodes.....	140
3. Résultats & discussion.....	141
3.1. Test du modèle de prédiction du Cd grain (Equation IV.8) sur un jeu de données indépendant du jeu de calibration.....	141
3.2. Recherche du meilleur modèle de prédiction des teneurs en Cd dans le grain à partir du jeu de données étendu à 48 sols.....	143
3.3. Analyse des erreurs de prédictions du Cd grains issues de la validation croisée.....	144
4. Conclusion.....	146
<b>Discussion générale .....</b>	<b>149</b>
1. Cohérence des modèles statistiques de prédiction des principaux processus impliqués dans la phytodisponibilité du Cd en sols agricoles et apport de la modélisation géochimique .....	149
2. Considérations sur les simplifications inhérentes aux modèles statistiques .....	152
3. Utilité et améliorations possible des modèles statistiques de prédictions .....	153
<b>Conclusion générale.....</b>	<b>157</b>
<b>Références bibliographiques.....</b>	<b>161</b>
Annexe 1: Protocole de récupération de la solution de sol .....	180
Annexe 2 : Protocole d'estimation de la fraction libre.....	182
Annexe 3 : Modèles statistiques obtenus au cours des différents chapitres .....	184
Annexe 4 : Article publié dans la revue Environmental Science and Pollution Research .....	187

## Liste des Figures

Figure I.1 : Biodisponibilité des métaux pour les organismes vivants, dont les plantes, à trois composantes : disponibilité environnementale, biodisponibilité environnementale et biodisponibilité toxicologique (repris de ISO 17402:2008). .....	8
Figure I.2 : Nomenclature des phyllosilicates (d'après Schulze, 1989) .....	10
Figure I.3 : Classification et principales propriétés chimiques des substances humiques .....	13
Figure I.4 : Représentation schématique de la procédure simplifiée proposée par van Zomeren and Comans (2007), pour la détermination des fractions d'acides humiques, d'acides fulviques et d'acides hydrophiles. DAX-8 (poly[méthacrylate de méthyle]) correspond à la résine utilisée pour adsorber les acides fulviques et les séparer des acides organiques de petite taille. ....	14
Figure I.5 : Schéma récapitulatif des principaux compartiments du sol et processus physico-chimiques et biologiques impliqués dans le devenir du Cd dans les sols.....	18
Figure I.6 : Répartition (%) entre les différentes formes du Cd en solution mesurées dans 64 solutions de sols en fonction du pH de la solution de sol (d'après Sauvé et al., 2000b).....	21
Figure I.7 : Différentes isothermes (Langmuir, Freundlich et Kd) utilisées pour modéliser la sorption du Cd sur une surface sorbante. ....	33
Figure I.8 : Représentation graphique du biais induit par le calcul du $K_d$ lorsque le compartiment considéré (q) dans le calcul n'est pas le compartiment mobile de Cd. ....	34
Figure I.9 : Descriptions des différents modèles de $K_d$ décrivant la distribution du métal dans un sol. (a) correspond à la représentation la plus complexe en prenant en compte l'adsorption et la désorption du métal sous sa forme libre ( $M^{2+}$ ) sur et de la phase solide et la cinétique d'association et de dissociation des complexes ML en solution. Les complexes en solution sont considérés comme ne se sorbant pas. Certains complexes sont inertes tout comme une partie du Cd de la phase solide et ne participent pas au passage de $Cd^{2+}$ en solution. Les réactions sont décrites à partir de constantes cinétiques d'ordre 1 décrivant des processus plus ou moins rapides de réapprovisionnement ou d'appauvrissement de la solution en $Cd^{2+}$ ( $k_{ads}$ , $k_{des}$ , $k_a$ , $k_d$ , $k_1$ , $k_{-1}$ ). (b-d) sont des visions plus simples à l'équilibre en considérant des compartiments de Cd plus ou moins labiles du sol et de la solution du sol. Le ratio du métal mobile du sol sur le métal libre en solution ( $K_d^{lab/free}$ ) ou la concentration totale en solution ( $K_d^{lab}$ ) ou pour finir le ratio du compartiment de Cd total du sol sur la concentration totale de Cd en solution ( $K_d^{tot}$ ). Adapté de Degryse et al., (2009) .....	35
Figure II.1 : Localisation des sols utilisés pour modéliser la spéciation du Cd en solution de sol. ....	48
Figure II.2 : Récapitulatif des scénarios testés pour la modélisation de la spéciation de Cd par CHEAQS Next® .....	54

Figure II.3 : Schéma récapitulatif de la démarche de modélisation des concentrations en Cd <sup>2+</sup> résumant les différentes approches utilisées. ....	57
Figure II.4 : Violin plots des caractéristiques pédogéochimiques des sols utilisés dans ce chapitre. Le point rouge correspond à la moyenne des individus, le box plot représente l'interquartile (25-75 %) et les segments le prolongeant correspondent à 1.5 fois l'interquartile. ....	57
Figure II.5 : Violin plots des caractéristiques chimiques mesurées en solutions de sol. Le point rouge correspond à la moyenne des individus, le box plot représente l'interquartile (25-75 %) et les segments le prolongeant correspondent à 1.5 fois l'interquartile. ....	59
Figure II.6 : Distribution des valeurs absolues des erreurs relatives (en pourcentage) obtenues par validation croisée des différents modèles statistiques corrigés de leur biais de prédiction (Équation II.14 à Équation II.19 et K'n médian). Le point rouge correspond à la moyenne des individus, le box plot représente l'interquartile (25-75 %) et les segments le prolongeant correspondent à 1.5 fois l'interquartile. La densité de probabilité de l'erreur relative est représentée de chaque côté du box plot. La droite pointillée rouge représente une erreur relative de 100 %. ....	64
Figure II.7 : Distribution comparée des prédictions de [Cd <sup>2+</sup> ] modélisées par CHEAQS Next en fonction de la nature de la MOD considérée. Les résultats issus de l'hypothèse d'une MOD composée de 100 % d'AF sont représentés en gris et en jaune pour l'hypothèse 50 % AF et 50 % AH. ....	65
Figure II.8 : Distribution comparée des prédictions des [Cd <sup>2+</sup> ] modélisées par CHEAQS Next et corrigée de leur biais éventuel en fonction du pourcentage de COD actif et pour 100 % d'AF. Le point rouge correspond à la moyenne des individus, le box plot représente l'interquartile (25-75 %) et les segments le prolongeant correspondent à 1.5 fois l'interquartile. La densité de probabilité de l'erreur relative est représentée de chaque côté du box plot. ....	67
Figure II.9 : Concentrations en Cd <sup>2+</sup> (log transformées) prédites sur l'hypothèse de 10 % de COD actif, dont 20 % d'AF, 72 % d'acide acétique et 8 % d'acide citrique, en fonction de celles mesurées. La ligne noire correspond au modèle linéaire entre valeurs observées et valeurs prédites, la ligne en pointillé rouge, l'équation y=x et les lignes vertes à l'intervalle de prédiction à 0.95). ....	67
Figure II.10 : Importance relative des variables pour la modélisation du nombre de sites de complexation obtenue par Random Forest. Les valeurs d'importance ont été recalculées en fixant leur somme à 100 %. ....	68
Figure II.11 : Concentrations en Cd <sup>2+</sup> (log transformées) prédites sur l'hypothèse d'un pourcentage de COD actif et d'AF dépendants du pH CaCl <sub>2</sub> , en fonction de celles mesurées. La ligne noire correspond au modèle linéaire entre valeurs observées et valeurs prédites, la ligne en pointillé rouge, l'équation y=x et les lignes vertes à l'intervalle de prédiction à 0.95. ....	69

Figure II.12 : Distributions des valeurs absolues des erreurs relatives corrigées de leur biais de prédiction pour les différentes approches de modélisation testées dans ce chapitre pour la prédiction des concentrations en $Cd^{2+}$ en solution. Le point rouge correspond à la moyenne des individus, le box plot représente l'interquartile (25-75 %) et les segments le prolongeant correspondent à 1.5 fois l'interquartile. La densité de probabilité de l'erreur relative est représentée de chaque côté du box plot.....	70
Figure II.13 : Schéma récapitulatif des résultats obtenus pour les 5 approches de modélisation testées dans le cadre de ce chapitre. Le résultat final, présenté en rouge, correspond à la moyenne des erreurs relatives en valeurs absolues obtenues par validation croisée et son coefficient de variation.....	71
Figure III.1 : Schéma récapitulatif des hypothèses faites pour la modélisation du compartiment solide du sol, surfaces de complexation et compartiments de Cd mobiles pour prédire les concentrations en $Cd^{2+}$ en solution.....	91
Figure III.2 : Concentrations en $Cd^{2+}$ (log transformées) prédites par l'équation V.4, en fonction de celles mesurées. La ligne noire correspond au modèle linéaire entre valeurs observées et valeurs prédites, la ligne en pointillé rouge, l'équation $y=x$ et les lignes vertes à l'intervalle de prédiction à 0.95).....	98
Figure III.3 : Violin plots des erreurs relatives corrigées du biais de prédiction pour les 8 meilleurs modèles statistiques à 1, 2, 3 et 4 régresseurs pour les prédictions des concentrations totales de Cd en solution et les concentrations en Cd libre ( $Cd^{2+}$ ) en solution. Le point rouge correspond à la moyenne des individus, le box plot représente l'interquartile (25-75 %) et la droite grise verticale correspond à 1.5 fois l'interquartile. La ligne en pointillé rouge représente un ratio de 1 entre valeurs prédites et valeurs observées montrant une prédiction égale à la valeur mesurée. Les ratios inférieurs à 1 indiquent que les valeurs prédites sont sous-estimées et inversement. Les couleurs distinguent le nombre de régresseurs utilisés pour la calibration des modèles : violet pour des modèles à 4 régresseurs, bleu, pour 3 régresseurs, vert pour 2 régresseurs et jaune pour les modèles à 1 régresseur. ....	100
Figure III.4 : Box plots des erreurs relatives de prédiction (prédit/observé) pour les concentrations en $Cd^{2+}$ en solution à partir des différentes hypothèses considérées lors de la modélisation multisurfaces de la distribution phase solide/solution du $Cd^{2+}$ dans le sol. Chaque couleur correspond à une famille d'hypothèses (H1 en rouge, H2 en vert, H3 en jaune, H4 en bleu et H5 en orange). Pour chacune de ces familles, l'extractant du Cd est mentionné ainsi que l'oxyde considéré et la façon dont est considérée la sorption non spécifique : par la CEC mesurée (approche a), par modélisation de la CEC à partir d'hypothèses sur la nature minéralogique des argiles (approche b).....	103

Figure III.5 : Distribution du Cd entre les phases solides modélisées par modèle géochimique WHAM à partir des teneurs en Cd obtenues par extraction au Pyrophosphate de sodium. Les argiles (échanges d'ions), la MOS et les oxydes métalliques amorphes du sol (H5).....	105
Figure IV.1: Comparison of grain Cd contents in this study and the corresponding predictions made by the model of Adams et al. (2004) (a), Brus et al. (2005) (b), Liu et al. (2015) (c) and Ran et al. (2016) (d) and the best model from this study (Equation IV.8) (e). The black line represents the regression between observed and predicted values; the red dashed line is the 1:1 line; the green dotted line represents the prediction interval at 95%. All grain Cd data are expressed at 15% water content.....	132
Figure IV.2: RMSE of predictions from the best model (Equation IV.8) for different sizes ( $n_v$ ) of validation subset from a Monte Carlo cross validation. The red line corresponds to the value of RMSE of the model parametrized with 27 soils ( $n_v = 0$ ) .....	135
Figure IV.3: Residual errors (observed minus predicted values) of Equation IV.8 for each soil in the dataset. All grain Cd data are expressed at 15% water content. ....	137
Figure IV.4 : Comparaison entre les teneurs en Cd dans le grain simulées par l'Equation IV.8 pour le jeu de données indépendant et les teneurs mesurées. La droite pointillée rouge représente la droite 1:1 correspondant à une adéquation parfaite entre valeurs observées et valeurs prédites. ....	141
Figure IV.5 : Distributions comparées de la capacité d'échange cationique (à gauche) et des teneurs on oxydes de manganèses amorphes (à droite) en fonction des deux jeux de données. Le jeu de données initial (27 sols) est représenté en jaune et le jeu de données additionnel (21 sols) est représenté en gris. ....	142
Figure IV.6 : Distribution cumulée des erreurs absolues ( observé-prédit ) de prédictions générées par l'Equation IV.10. Trois zones de couleur différentes sont définies à partir du projet de diminution de la valeur maximale autorisée dans les grains de blé (DG-SANCO, 2011). La zone verte correspond à une diminution du seuil actuel de $0.025 \text{ mg.kg}^{-1}$ (passage $0.2$ à $0.175 \text{ mg.kg}^{-1}$ ). La zone orangée pour une diminution du seuil de $0.05 \text{ mg.kg}^{-1}$ (passage de $0.2$ à $0.15 \text{ mg.kg}^{-1}$ ) et la zone rouge pour représenter les erreurs supérieures à ces deux paliers...	146

## Liste des Tableaux

Tableau I.1 : Classes granulométriques des terres fines .....	9
Tableau I.2 : Capacités d'échange cationique et surfaces spécifiques de différents minéraux secondaires d'après Sparks, 2003.....	16
Tableau I.3 : Valeurs de $pH_{PCN}$ pour différents minéraux secondaires. Les méthodes d'estimation différent (électrophorèse, titration) et par conséquent ne peuvent pas nécessairement être comparées. ....	17
Tableau II.1 : Récapitulatif des équations obtenues pour la prédiction des concentrations de $Cd^{2+}$ à partir d'un ligand moyen unique et les paramètres d'évaluation des modèles obtenus par validation croisée. $r^2_{aj}$ correspond au coefficient de détermination de la régression linéaire multiple, $VIF_{max}$ est la valeur maximale du facteur d'inflation de la variance, Ord.Orig. correspond à l'ordonnée à l'origine de la régression $\log_{10}(\text{observé}) = f(\log_{10}[\text{prédit}])$ et $r^2_{ord-pred}$ au coefficient de régression entre les valeurs observées et les valeurs prédites log-transformées.....	60
Tableau II.2 : Récapitulatif des équations obtenues pour la prédiction des concentrations de $Cd^{2+}$ à partir d'un ligand moyen unique et les paramètres d'évaluation des modèles obtenus par validation croisée. $r^2_{aj}$ correspond au coefficient de détermination de la régression linéaire multiple, $VIF_{max}$ est la valeur maximale du facteur d'inflation de la variance, Ord.Orig. correspond à l'ordonnée à l'origine de la régression $\log_{10}(\text{observé}) = f(\log_{10}[\text{prédit}])$ et $r^2_{ord-pred}$ au coefficient de régression entre les valeurs observées et les valeurs prédites log-transformées.....	62
Tableau III.1 : Résumé des méthodes d'extraction pour évaluer la fraction de Cd mobile des sols. ....	85
Tableau III.2 : Tableau récapitulatif des caractéristiques physico-chimiques des sols étudiés. N=105. ....	92
Tableau III.3 : Résumé des caractéristiques de solutions de sols étudiées. Les valeurs minimales, médianes et maximales du jeu de données complet (105 sols) sont renseignées. ....	93
Tableau III.4 : Résumé des teneurs en Cd extrait par extractions sélectives, extraction totale et des concentrations en Cd mesurées en solution de sol. Le nombre d'échantillons disponibles (N) pour chaque extraction est indiqué entre parenthèses. Les valeurs minimales, médianes et maximales obtenues sur les N valeurs sont reportées.....	95
Tableau III.5 : Teneurs et réactivité des phases réactives prises en compte dans cette étude. $CEC_{arg}$ correspond à la capacité d'échange cationique supposée de la phase argileuse. Les teneurs en AH ont été calculées selon Weng et al., (2001) (Équation III.3) et les termes FeO, AlO, MnO renvoient aux teneurs en Fe, Al et Mn mesurées dans un extrait Tamm. ....	96
Tableau III.6 : Récapitulatif des équations obtenues pour la prédiction des concentrations en $Cd^{2+}$ et $Cd_{ss}$ et les paramètres d'évaluation des modèles obtenus par validation croisée. ....	97

Table IV.1: Summary of the extraction methods used as single extractants to evaluate the phytoavailability of Cd for durum wheat. ....	122
Table IV.2: Models found in the literature to predict Cd contents in bread wheat grains. N is the number of observations used to parameterize the model; $r^2_{adj}$ : adjusted r-squared; $\alpha$ is the slope of the regression between observed and predicted values for each model; RMSE: root mean square error; $r^2$ : r-squared of the regression between observed and predicted values.....	125
<i>Table IV.3: Summary of the characteristics of the soils used in this study (a), of the Cd pools estimated by a single extraction and measured in soil solutions (b) and of Cd measured in different wheat organs, roots, shoots and grains (c).</i> .....	128
Table IV.4: Adjusted r-squared for the linear regressions between the grain Cd (log) contents and the Cd contents (log) recovered using different extraction methods.....	133
Table IV.5: Best models to predict Cd contents in durum wheat grains and associated statistics: $r^2_{adj}$ adjusted r-squared; RMSE: root mean square error; $RMSE_{LOOCV}$ root mean square error from the leave-one out cross-validation; $VIF_{max}$ : maximum value of the variance inflation factor; N number of observations. ....	134
Table IV.6 : Récapitulatif des paramètres pédogéochimiques principaux des sols du jeu de données complémentaire (ANR CaDON, N=21).....	140
Table IV.7 : Meilleurs modèles pour la prédiction des teneurs en Cd dans le grain de blé dur et paramètres statistiques pour leur évaluation : $r^2_{adj}$ de la régression multiple ; $RMSE_{CV}$ : racine de l'erreur quadratique moyenne obtenue par validation croisée ; $VIF_{max}$ : valeur maximale du facteur d'inflation de la variance ; N : nombre d'observations utilisées pour le paramétrage du modèle. La pente et l'ordonnée à l'origine concerne la régression des valeurs observées en fonction des valeurs prédites .....	145

## Liste des abréviations

$a_{\text{COOH}}$   $a_{\text{OH}}$  : abondance de sites des groupements carboxyliques et phénoliques

**AAS** : Atomic Absorption Spectrometry

**AES** : Atomic Emission Spectroscopy

**AF** : Acide Fulvique

**AH** : Acide Humique

**Al** : Aluminium

**AIO** : Oxyde d'aluminium amorphe

**ANR** : Agence Nationale de la Recherche

**ArgLF** : teneur sommée en argiles et limons fins

**ASV** : Anodic Stripping Voltammetry

**b** : pouvoir tampon du sol

**BCR** : Community Bureau of Reference

**BIC** : Bayesian Information Criteria

**Ca** : Calcium

**Cd** : Cadmium

**CdL** : Cadmium complexé en solution

**CD-MUSIC** : Charge Distribution – Multi-Site Complexation

**CEC** : Capacité d'Échange Cationique

**CEC<sub>Arg</sub>** : Capacité d'Échange Cationique associée aux argiles

**CEC<sub>MOS</sub>** : Capacité d'Échange Cationique associée à la matière organique du sol

**CHEAQS** : CHEmical Equilibria in AQUatic Systems

**CID** : Carbone Inorganique Dissout

**Co** : Cobalt

**COD** : Carbone Organique Dissout

**COS** : Carbone Organique du Sol

**Cu** : Cuivre

**CV** : Cross Validation

**DDL** : Diffuse Double Layer

**DG SANCO** : Direction générale de la santé et des consommateurs

**DGT** : Diffusive Gradient in Thin films

**DMT** : Donnan Membrane Technic

**DOC** : Dissolved Organic Carbon

**DPV** : Differential pulse voltammetry

**DTPA** : acide diéthylène triamine penta acétique

**DW** : Dry Weight

**EDTA** : acide Éthylène diamine tétra acétique

**EFSA** : Autorité Européenne de sécurité des aliments

**ET** : Élément Trace

**ETM** : Élément Trace Métallique

**EXAFS** : Extended X-Ray Absorption Fine Structure

**FAO** : Organisation pour l'alimentation et l'agriculture

**fCd** : Fraction libre de cadmium

**Fe** : Fer

**FeO** : Oxyde de fer amorphe

**FIAM** : Free Ion Activity Model

**GPS** : Global Positioning System

**GTLM** : Generalized Two Layers Model

**Hg** : Mercure

**HIB** : Humic Ion Binding

**HSAB** : Hard Soft Acid Base

**IARC** : International Agency for Research on Cancer

**ICP** : Inductively Coupled Plasma

**ISPA** : Interactions Sol-Plante-Atmosphère

**INRA** : Institut National de la Recherche Agronomique



**ISE** : Ion Selective Electrode

**ISO** : Organisation internationale de normalisation (International Organization for Standardization)

**K'n** : Produit de la constante de complexation conditionnelle du Cd avec la MOD et du nombre de ligands libre en solution

**kg** : kilogramme

**l** : litre

**LMWOA** : Low Molecular Weight Organic Acid

**log** : logarithme decimal

**LOOCV** : Leave One Out Cross Validation

**M** : Concentration molaire

**Max.** : Maximum

**MEB** : Microscope Électronique à Balayage

**Med.** : Médiane

**Min.** : Minimum

**Mg** : Magnésium

**mg** : milligramme

**ml** : millilitre

**Mn** : Manganèse

**MnO** : Oxyde de Manganèse Amorphe

**MOD** : Matière Organique Dissoute

**MOH** : Matière Organique Hydrophobe

**mol** : quantité de moles

**MOS** : Matière Organique du Sol

**MS** : Mass Spectrometry

**N** : nombre d'observations

**NF** : Norme Française

**ng** : nanogramme

**Ni** : Nickel

**NICA** : Non Idealised Competition Adsorption

**NIST** : National Institute of Standards and Technology

**nM** : nanomolaire

**Ord. Orig.** : Ordonnée à l'origine

**ORCHESTRA** : Object Representation of Chemical Speciation and Transport models

**Pb** : Plomb

**PCN** : Point de Charge Nulle

**pF** : logarithme de la pression négative

**pH** : potentiel Hydrogène

**pK** : Constante de complexation (log)

**pKa** : Constante d'acidité (log)

**pKs** : Constante de solubilité (log)

**PM** : Planetary Mill

**r<sup>2</sup><sub>aj</sub>** : Coefficient de détermination ajusté

**r<sup>2</sup><sub>obs/pred</sub>** : Coefficient de détermination observé/prédit

**RMQS** : Réseau de Mesures de la Qualité des Sols

**RMSE** : Root Mean Square Error

**S<sub>AH</sub>** : densité de sites des acides humiques

**SAX** : Spectrométrie d'Absorption de rayons X

**SCAMP** : Surface Chemistry Assemblage Model for Particles

**SH** : Substance Humique

**Si** : Silicium

**SOC** : Soil Organic Carbon

**SOM** : Soil Organic Matter

**ss** : Solution de sol (utilisé en indice)

**SWV** : Squarewave voltammetry

**T<sub>AH</sub>** : Teneur en AH

**T<sub>Arg</sub>** : Teneur en argiles

**TOC** : Total Organic Carbon

**UP** : Ultra-Pure

**VIF** : Facteur d'Inflation de la Variance

**V:m** : rapport volume masse

**WHAM** : Windermere Humic Aqueous Model

**WHC** : Water Holding Capacity

**XANES** : X-ray Absorption Near Edge Structure

**Zn** : Zinc

**µg** : microgramme

**µm** : micromètre





## **Introduction générale**

---

La contamination des sols par les éléments traces métalliques (ETM) constitue une des préoccupations majeures du débat sur la sécurité alimentaire et la qualité des produits végétaux récoltés, tant au niveau européen qu'au niveau mondial.

Les ETM sont définis comme des éléments dont les teneurs dans la croûte terrestre ne dépassent pas  $1 \text{ g.kg}^{-1}$  (Adriano, 2001). La présence de ces éléments traces dans les sols agricoles a donc une origine naturelle, provenant de l'altération des roches mères, mais a également des origines anthropiques (Adriano, 2001). En effet, les ETM dans les sols agricoles proviennent aussi des intrants agricoles comme les apports de fertilisants minéraux ou organiques, les amendements calcimagnésiens, les produits phytosanitaires et les dépôts atmosphériques. Ainsi, la contamination des sols agricoles, par les ETM, en France est en partie liée aux pratiques agricoles et leur distance des villes et des installations industrielles.

Parmi ces ETM, le Cd est un élément non essentiel, particulièrement étudié à cause de sa forte toxicité. Cet élément a été classé carcinogène pour les humains (IARC, 1993). Une exposition chronique à cet élément peut causer des dommages à différents organes, mêmes à faible exposition (insuffisance rénale et fragilité osseuse) (Järup, 2003).

En 2009, un inventaire relativement complet de l'état de contamination des horizons de surface des sols français a été réalisé et a montré que 95 % des horizons de sol étudiés ont une teneur en Cd total inférieure à  $0.9 \text{ mg.kg}^{-1}$  (Gis Sol, 2011). Malgré ces teneurs dans les sols relativement faibles, le Cd peut représenter une menace pour la santé humaine puisque le principal vecteur de contamination, excepté pour les fumeurs, se trouve être la consommation de produits alimentaires contaminés, la part la plus importante étant les végétaux (Satarug et al., 2003). Les concentrations du Cd dans les produits végétaux à destination de la chaîne alimentaire sont réglementées, ce qui conditionne la commercialisation des récoltes. En 2009, à la suite d'investigations toxicologiques réactualisées, l'Agence européenne de sécurité alimentaire (EFSA) recommande d'abaisser la dose hebdomadaire tolérable pour l'exposition humaine au Cd par voie alimentaire de 7 à  $2.5 \text{ } \mu\text{g.kg}^{-1}$  de poids corporel (EFSA, 2009).

Le blé fait partie des céréales parmi les plus produites et consommées à travers le monde (FAO, 2015). Parmi les sous-espèces de blé, le blé dur (*Triticum turgidum subs. durum*) a tendance à accumuler près de 10 fois plus de Cd que le blé tendre (*Triticum aestivum*) (Stolt et al., 2003). En

conséquence, la communauté européenne a travaillé sur un projet d'abaissement du seuil réglementaire pour les teneurs en Cd dans le blé de 0.2 à 0.1 mg.kg<sup>-1</sup>, initialement sans distinguer le blé tendre du blé dur puis dans un second temps, un projet de seuil spécifique au blé dur a été envisagé à 0.15 mg Cd.kg<sup>-1</sup> appliqué en deux temps avec un seuil intermédiaire à 0.175 mg Cd.kg<sup>-1</sup> (DG-SANCO, 2011). Ce projet de baisse importante a créé une forte tension dans la filière blé dur, car on estime que 24 % de la production française dépassent le seuil de 0.1 mg Cd.kg<sup>-1</sup> et 8 % dépassent celui de 0.15 mg Cd.kg<sup>-1</sup> (Projet ANR CaDON, Nguyen et al., 2017).

La teneur en Cd total du sol ne permet pas d'estimer la quantité de Cd prélevée par les racines, cette dernière étant liée à la fraction phytodisponible, c'est-à-dire biodisponible pour la plante, c'est-à-dire présente sous une forme assimilable par la plante : la forme libre Cd<sup>2+</sup> en solution. L'évaluation de la fraction du Cd du sol potentiellement disponible pour l'absorption racinaire est donc au cœur de l'évaluation du risque de contamination cadmiée du grain de blé dur. Cette disponibilité, variable selon les sols, a un effet 4 fois plus important sur les teneurs en Cd du grain que les différences d'accumulation entre variétés (Projet Cadur et CaDON, Nguyen et al., 2017). Le cadmium phytodisponible du sol est sous la dépendance très complexe de nombreux processus physico-chimiques dont les plus importants en sols agricoles sont la fixation plus ou moins réversible du Cd<sup>2+</sup> sur les phases solides du sol et la formation de complexes entre le Cd<sup>2+</sup> et des ligands organiques et inorganiques de la solution du sol (Lin et al., 2016).

Mon travail de thèse a donc eu pour objectif de prédire le plus simplement possible ces deux variables à partir d'un nombre restreint de mesures agronomiques (par exemple les variables issues d'analyses de terre, d'extractions sélectives, etc.) et d'identifier les paramètres pédogéochimiques influençant la disponibilité du métal en sols agricoles.

Pour cela nous avons décidé de nous appuyer sur des outils statistiques pour la prédiction de ces variables d'intérêt puisque cette approche reste la plus opérationnelle. Cependant, l'empirisme de cette approche fait que l'évaluation des mécanismes et l'interprétation des variables identifiées sont rendues difficiles.

Pour pallier ces points faibles, nous avons choisi d'évaluer les modèles statistiques par des méthodes de validation croisée pour cerner la robustesse et de les coupler à des modèles géochimiques dans le but d'explicitier les mécanismes et vérifier la cohérence des variables à la base des modèles statistiques.

En suivant cette démarche de dualité d'approches (modèles statistiques appuyés par la modélisation géochimique), nous avons, dans une première partie (Chapitre II), cherché à modéliser les concentrations en Cd<sup>2+</sup> en solution de sol car cette dernière est difficilement

accessible. Puis dans une seconde partie (Chapitre III), l'objectif a été de modéliser la distribution phase solide-solution du  $\text{Cd}^{2+}$  afin de considérer le possible réapprovisionnement de la solution par la phase solide. Enfin, la dernière partie (Chapitre IV) a cherché à modéliser de façon empirique les teneurs en Cd dans le grain de blé dur selon l'hypothèse que les concentrations en  $\text{Cd}^{2+}$  et sa distribution entre les phases solides et la solution de sol sont les variables majeures pour expliquer l'accumulation du Cd dans le grain de blé dur. Il est donc attendu que les variables prédictives pour estimer les teneurs en Cd dans le grain soient les mêmes que celles identifiées pour la prédiction des concentrations en  $\text{Cd}^{2+}$  en solution et de la distribution sol-solution du  $\text{Cd}^{2+}$









## **Chapitre I. Étude bibliographique de la dynamique du Cd dans les sols**

---

Les sols sont des milieux complexes et hétérogènes constitués de trois phases : solide (minérale et organique), liquide et gazeuse. Les constituants solides du sol constituent environ 50 % de son volume total tandis que l'eau et l'air vont occuper chacun entre 20 et 30 % du volume restant.

### **1. Concept de la biodisponibilité**

Les notions de biodisponibilité et de disponibilité sont au cœur de ce travail de thèse et nécessitent donc d'être définies. Les définitions fournies par l'ISO (Organisation Internationale de Standardisation) permettent de décomposer le concept de biodisponibilité en trois composantes : la disponibilité environnementale, la biodisponibilité environnementale et la biodisponibilité toxicologique (ISO 17402:2008). Ces trois concepts sont représentés Figure I.1.

La **disponibilité environnementale** peut être définie comme la fraction de métal du sol (sur la phase solide et dans la solution de sol) résultant de processus physico-chimiques de désorption et pouvant potentiellement être absorbée par un organisme vivant. Cette définition comprend donc la totalité du métal en solution et le Cd fixé de façon réversible sur les phases solides et pouvant donc se désorber. Cela permet de ne pas considérer la fraction de métal du sol inerte, ne pouvant être solubilisée que sur des temps très longs, non représentatifs d'un contexte agronomique. Ainsi, cette fraction de Cd du sol désorbable est définie comme mobile (Juste, 1988), c'est-à-dire capable de passer d'un compartiment du sol où il est fortement retenu vers celui où il est moins énergiquement retenu, le compartiment ultime étant la solution du sol.

Il s'agit donc d'une vision maximisant les précautions, du système sol, en considérant la fraction maximale de métal potentiellement absorbable par l'organisme vivant.

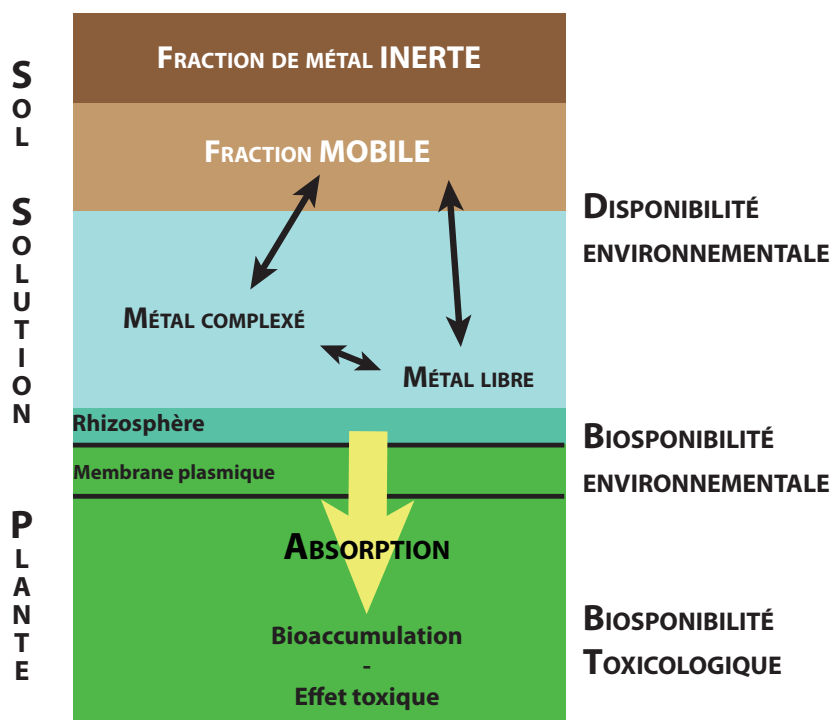


Figure I.1 : Biodisponibilité des métaux pour les organismes vivants, dont les plantes, à trois composantes : disponibilité environnementale, biodisponibilité environnementale et biodisponibilité toxicologique (repris de ISO 17402:2008).

La **biodisponibilité environnementale**, quant à elle, se définit comme la fraction de métal effectivement absorbée par l'organisme vivant, ici la plante. Cette fraction sera donc différente selon les organismes vivants considérés et la forme chimique du métal prélevée.

La sorption/désorption et la spéciation (formes chimiques de l'élément) sont au cœur de la disponibilité et biodisponibilité environnementales.

La **biodisponibilité toxicologique** correspond à la fraction de l'élément absorbée par l'organisme et ayant un effet toxique. Cet aspect ne sera pas abordé dans ce travail.

## 2. Nature des principaux constituants du sol impliqués dans la dynamique du Cd

Les principaux constituants solides du sol vont se répartir entre composés inorganiques et organiques dans une proportion proche des 90/10 % (v:v) et suivant un gradient de diamètres de particules allant des fractions les plus fines, dites argileuses (< 2 µm), aux fractions les plus grossières appelées fractions sableuses (0.05 à 2 mm). Les cinq fractions couramment étudiées en sciences du sol et en agronomie se répartissent comme indiqué Tableau I.1 et constituent la

terre fine du sol en opposition aux éléments grossiers (> 2 mm), comme les graviers, cailloux, pierres et blocs.

Tableau I.1 : Classes granulométriques des terres fines

Fractions	Diamètres des particules
Argiles granulométriques	< 2 $\mu\text{m}$
Limons fins	2-20 $\mu\text{m}$
Limons grossiers	20-50 $\mu\text{m}$
Sables fins	50-200 $\mu\text{m}$
Sables grossiers	200-2000 $\mu\text{m}$

Les constituants solides inorganiques, aussi appelés minéraux, peuvent être primaires ou secondaires. On les définit primaires lorsqu'ils n'ont pas été altérés chimiquement depuis leur dépôt ou leur cristallisation et secondaires lorsqu'ils proviennent de l'altération de ces minéraux primaires.

## 2.1. Les phyllosilicates

Les phyllosilicates sont des minéraux argileux secondaires provenant de l'altération des minéraux primaires silicatés (feldspaths, pyroxènes, amphiboles, etc.). Ces minéraux constituent la majeure partie de la fraction argileuse granulométrique. Ils sont connus pour largement influencer les propriétés physico-chimiques des sols grâce à leur petite taille, leur grande surface spécifique et leurs propriétés d'échanges cationiques exceptionnelles (Sparks, 2003).

Les phyllosilicates sont des silicates en feuillets composés de l'association de deux types de couches constituées de polyèdres élémentaires (tétraédriques et octaédriques) comportant des oxygènes ou des hydroxyles à leur sommet et divers cations en leur centre. La couche tétraédrique est constituée de tétraèdres comportant quatre oxygènes à leurs sommets et un cation central. Ce cation est principalement  $\text{Si}^{4+}$  ( $\text{SiO}_4$ )<sup>4-</sup>, fréquemment  $\text{Al}^{3+}$  ( $\text{AlO}_4$ )<sup>5-</sup>, et plus rarement  $\text{Fe}^{3+}$  ( $\text{FeO}_4$ )<sup>5-</sup>. Les tétraèdres sont liés entre eux par la mise en commun de trois sommets, les oxygènes basaux, le quatrième restant libre est appelé oxygène apical (Figure I.2). La couche octaédrique est formée par des octaèdres jointifs dont les sommets sont soit des oxygènes, soit des hydroxyles. Ces anions sont liés à un cation central trivalent ( $\text{Al}^{3+}$ ,  $\text{Fe}^{3+}$ ) ou divalent ( $\text{Fe}^{2+}$ ,  $\text{Mg}^{2+}$ ) en coordinance 6. Les octaèdres sont liés entre eux par les arêtes.

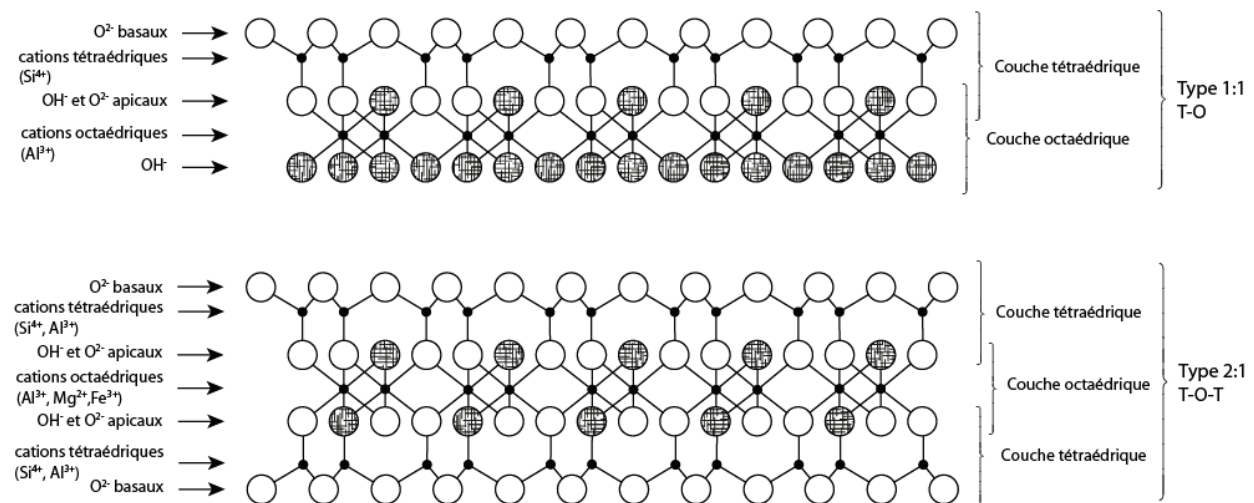


Figure 1.2 : Nomenclature des phyllosilicates (d'après Schulze, 1989)

La succession de la structure initiale tétraédrique peut aboutir à la formation de deux types de structures plus complexes, appelées feuillets, tétraédriques et octaédriques. L'arrangement de ces feuillets et l'espace interfoliaire caractérisent principalement deux types d'argiles minéralogiques existants, de type 1:1, ou de type 2:1.

Les argiles de type 1:1 consistent en un empilement d'un feuillet tétraédrique et d'un feuillet octaédrique articulés en 3 plans d'anions, le premier constitué par les oxygènes basaux de la couche tétraédrique entourant un atome de Si, le second constitué des  $O^{2-}$  apicaux de la couche tétraédrique et d'hydroxyles ( $OH^-$ ) provenant de la couche octaédrique et enfin un plan d' $OH^-$  qui ensemble entoure un atome d'Al de coordinence 6 (ex. kaolinite, serpentinite).

Lorsque deux couches tétraédriques entourent un feuillet octaédrique, elles forment une argile de type 2:1. Dans cette configuration, les couches tétraédriques peuvent inclure des cations  $Al^{3+}$  au lieu de  $Si^{4+}$ , et induire un excès de charges dans la structure qui sera comblé par des cations libres hydratés ou non dans l'espace interfoliaire ( $K^+$ ,  $Ca^{2+}$ ,  $Mg^{2+}$ ...). Ces substitutions de cations dans les différentes couches sont plus fréquentes que pour les argiles 1:1 et aboutissent donc à une plus grande variété d'argiles de type 2:1 (ex. mica, vermiculite, smectite, illite).

## 2.2. Oxydes métalliques

Lors de l'altération des minéraux primaires (principalement minéraux ferromagnésiens, puis argileux), des cations et anions sont relargués en solution et tendent à se recombinaer et précipiter pour former de nouveaux minéraux. Parmi ces cations, Al, Fe et Mn vont s'associer en oxydes, terme générique pour les hydroxydes, oxyhydroxydes et oxydes hydratés. Les processus et les conditions physico-chimiques des sols vont conditionner l'état cristallin des oxydes formés aboutissant soit à des oxydes cristallisés, soit à des oxydes amorphes (ordonnés à courte

distance), la présence de matière organique dans le milieu est un frein à la cristallisation des oxydes (Hénin, 1956).

Ces minéraux sont principalement présents dans la fraction granulométrique argileuse du sol, de la même façon que les phyllosilicates. Bien que leurs teneurs dans les sols soient généralement faibles, les oxydes métalliques vont jouer un rôle prépondérant dans un nombre important de processus physico-chimiques comme la sorption, correspondant à la fixation des métaux sur la phase solide (détaillée dans le Chapitre I.4.1.3.a) et l'oxydoréduction.

Dans les sols agricoles des régions tempérées, les oxydes de fer et de manganèse sont les espèces les plus retrouvées avec une prédominance des oxydes de fer. L'oxyde d'aluminium naturel le plus commun est la gibbsite ( $\text{Al}[\text{OH}]_3$ ) et est plus fréquemment rencontré dans des climats tropicaux et arides où l'altération des sols est la plus avancée. La boehmite (oxyhydroxyde de formule  $\gamma\text{-AlOOH}$ ) est également rencontrée naturellement, mais à des teneurs beaucoup plus faibles.

### **2.3. Oxydes de fer**

Principaux oxydes métalliques rencontrés dans les sols tempérés, les oxydes de fer sont de puissants pigments souvent responsables de la coloration des sols en brun rouge. La cristallisation des oxydes de fer est rapide. L'oxyde de fer a pour structure de base l'octaèdre, c'est-à-dire que chaque atome de Fe est entouré de 6 atomes d' $\text{O}^{2-}$  et  $\text{OH}^-$  (Sparks, 2003). La variété de ces oxydes est liée aux arrangements et liaisons des octaèdres entre eux. Une structure basée sur un arrangement hexagonal compact sera qualifiée de  $\alpha$ , tandis que les phases  $\gamma$  correspondent à un arrangement cubique. Ces dernières sont moins stables thermodynamiquement et le  $\text{Fe}^{3+}$  peut être remplacé par  $\text{Al}^{3+}$ ,  $\text{Mn}^{3+}$  et  $\text{Cr}^{3+}$  par substitution isomorphe, c'est-à-dire le remplacement d'un cation par un autre de charge identique ou différente.

La goethite est l'oxyde de fer le plus fréquemment retrouvé dans les sols non saturés. De formule  $\alpha\text{-FeO}(\text{OH})$  cet oxyhydroxyde est responsable de la couleur jaunâtre à brunâtre des sols. L'hématite, de formule  $\text{Fe}_2\text{O}_3$ , légèrement moins commune que la goethite, est souvent retrouvée en association avec cette dernière. Ces minéraux sont stables dans un milieu oxydant, c'est à dire dans des sols aérés, en milieu non ou peu acide. Les formes du fer présentes dans le sol sont principalement conditionnées par le pH et le potentiel d'oxydoréduction. Ainsi, en conditions réductrices, comme c'est le cas des sols saturés en eau, le  $\text{Fe}^{3+}$  peut être réduit en  $\text{Fe}^{2+}$  plus soluble et donc être transporté vers d'autres couches de sols où les conditions d'oxydoréduction permettront l'oxydation du  $\text{Fe}^{2+}$  et la formation d'autres oxydes, tel que la lépidocrocite ( $\gamma\text{-FeO}[\text{OH}]$ ), et la ferrihydrite ( $\text{Fe}_5\text{O}_7[\text{OH}], 4\text{H}_2\text{O}$ ), une forme amorphe.

## **2.4. Oxydes de manganèse**

La minéralogie du Mn dans les sols est complexe, en raison de la variété des structures minéralogiques et donc du nombre important d'oxydes et hydroxydes potentiellement formés dans lesquels les substitutions de  $Mn^{2+}$  et  $Mn^{3+}$  par  $Mn^{4+}$  sont fréquentes. Deux types de structures sont retrouvées dans les sols :

– les phyllomanganates à structure lamellaire arrangés sous forme d'une succession de feuillets, de base élémentaire un octaèdre  $MnO_6$ , lié entre eux par leurs arêtes. La birnessite ( $[Na,Ca] Mn_7O_{14}, 2.8H_2O$ ) et la vernadite ( $\delta-MnO_2$ ) sont les plus fréquentes dans les sols. De la même façon que pour les argiles, la succession de différents feuillets fait la diversité de ces oxydes de manganèse. Les phyllomanganates sont les oxydes de Mn les plus courants dans les sols.

– les structures tunnelaires ( $MnO_2, nH_2O$ ) consistent en une succession d'octaèdres  $MnO_6$  partageant des arêtes pour former des chaînes elles-mêmes liées entre elles par partage d'angle. Ces arrangements particuliers aboutissent à la formation de structure tunnel pouvant être occupée par des cations et des molécules d'eau. L'exemple le plus stable est la pyrolusite ou  $\beta-MnO_2$  cependant moins retrouvées dans les sols que la birnessite.

Rarement cristallisés, ces oxydes se retrouvent sous forme de revêtements sur d'autres minéraux ou de concrétions, particulièrement avec les oxydes de fer.

## **2.5. Matières organiques du sol (MOS)**

Les teneurs en MOS dans les sols varient généralement de 0.5 à 5 % en masse pour les sols minéraux à 80 % pour les sols organiques (histosols, tourbes). Dans le cas des sols agricoles, la MOS provient de l'héritage de pédogénèse et des apports par les résidus de cultures (litières aériennes et racinaires, engrais verts enfouis) et par les produits organiques fertilisants (fumier, composts, boues, etc.).

La fraction organique des sols se divise elle-même en deux constituants. Le premier constituant est la matière organique (MO) vivante, appelée édaphon, qui englobe la totalité de la biomasse animale (microfaune, mésofaune et macrofaune), végétale (végétaux supérieurs et inférieurs, microflore), fongique (champignons) et bactérienne. Le second constituant de la fraction organique est la MO non vivante qui englobe la totalité de la nécromasse animale (déjections, cadavres), végétale (résidus végétaux, exsudats), fongique (champignons morts) et bactérienne (bactéries mortes) et les produits dérivés de sa transformation. La MOS non vivante est donc

composée de matières non solubles (matière organique particulaire, précipitée et colloïdale) et solubles issues de sa transformation physique, chimique et biologique.

La matière organique dans les sols, tant physique que soluble (exsudats racinaires, carbohydrates), est transformée selon deux processus :

1. La minéralisation : La matière organique est consommée par différents organismes décomposeurs du sol (arthropodes, champignons, bactéries), éventuellement après fragmentation pour produire de la biomasse, des métabolites, des minéraux, du CO<sub>2</sub> et de l'eau.
2. L'humification qui est un processus tant physico-chimique que biologique qui aboutit à la formation de molécules de haut poids moléculaire par condensation et oxydation, qui sont récalcitrantes à la dégradation microbienne : les substances humiques (SH).

Ces dernières peuvent être définies comme un ensemble de substances organiques hétérogènes de poids moléculaires élevés, partiellement aromatiques colorant la solution de sol dans des teintes allant du jaune au noir. Les groupes fonctionnels majoritaires des SH sont les groupements carboxyliques (-COOH), phénoliques ( $\varphi$ -OH), amine (-NH<sub>2</sub>) et sulfhydriles (-SH) (Sigg et al., 2006). En fonction de leurs degrés de condensation et d'aromaticité croissants, les SH sont catégorisées en acides fulviques (AF), acides humiques (AH), et humine insoluble (Schnitzer and Khan 1972). La Figure I.3 résume cela.

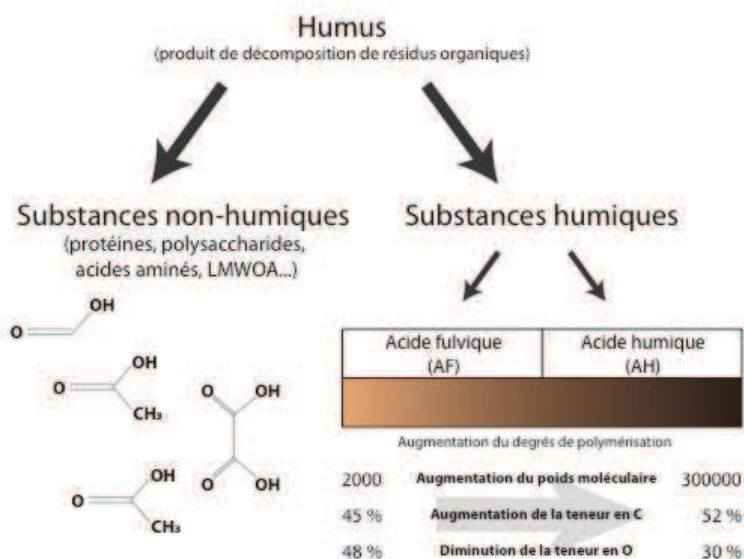


Figure I.3 : Classification et principales propriétés chimiques des substances humiques

La matière organique en solution, de taille inférieure à 0.45 ou 0.2  $\mu$ m, peut être définie comme la fraction dissoute de la matière organique (MOD). Les substances humiques en solution sont généralement considérées comme des colloïdes, c'est-à-dire des particules de tailles comprises



entre 1 et 200 nm ainsi distinctes des molécules en solution. Leurs larges surfaces spécifiques (surface par unité de masse) leur confèrent des capacités d'adsorption remarquables.

Sur le plan analytique, la quantité de MOD en solution est déterminée par la mesure du carbone organique dissous (COD). L'hypothèse générale reste que la matière organique est constituée à 50 % en masse de carbone (Wood, 1996). Différents protocoles d'extractions permettent la caractérisation de la MOD, en séparant l'humine des acides humiques (AH) et fulviques (AF) et en séparant via l'utilisation de résines, les AF des espèces non humiques. Un exemple de procédure proposée par van Zomeren and Comans, (2007) est schématisé Figure I.4.

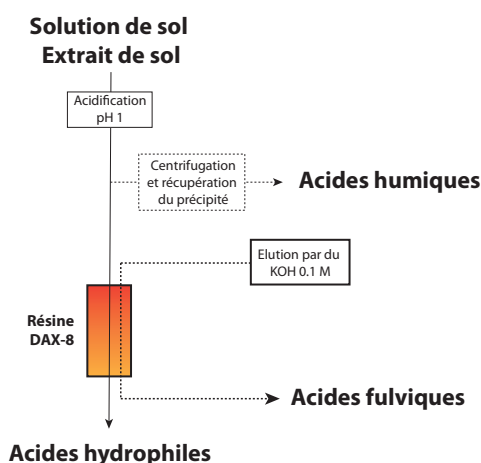


Figure I.4 : Représentation schématique de la procédure simplifiée proposée par van Zomeren and Comans (2007), pour la détermination des fractions d'acides humiques, d'acides fulviques et d'acides hydrophiles. DAX-8 (poly[méthacrylate de méthyle]) correspond à la résine utilisée pour adsorber les acides fulviques et les séparer des acides organiques de petite taille.

Tous les colloïdes en solution ne sont pas organiques. Ils peuvent être également de nature inorganique résultant des premiers stades de la précipitation d'une phase minérale, principalement des oxydes métalliques sous l'effet d'une sursaturation.

## 2.6. Minéraux carbonatés et sulfatés

Les minéraux carbonatés existent sous différentes formes. Les carbonates peuvent être présents sous forme de particules fines, de concrétions, de nodules et d'encroûtements. Dans les sols, la calcite ( $\text{CaCO}_3$ ) et la dolomie ( $\text{CaMg}[\text{CO}_3]_2$ ) sont les espèces les plus fréquentes. Ces minéraux peuvent se retrouver dans les sols directement sous forme de petits grains provenant de l'altération de la roche mère ou être issus des processus de dissolution/précipitation aboutissant à la formation de minéraux secondaires. En effet, tous les carbonates ont tendance à être

modérément solubles en solution de sol avec une solubilité plus forte à des pH faibles. Plus ces minéraux seront petits, plus ils seront solubles.

Les évaporites (dépôts ou roches riches en chlorures et sulfates), sont également fréquentes dans les sols. Elles sont facilement formées par précipitation lors d'épisodes d'évaporation ou de dessiccation de la solution de sol et ont une solubilité supérieure à la calcite.

Concernant les minéraux sulfatés, le gypse ( $\text{CaSO}_4$ ) est le plus abondant dans les sols. Sa provenance et son comportement sont similaires à la calcite.

Au cours de la formation des minéraux carbonatés et sulfatés, des éléments traces peuvent être incorporés au minéral. La proximité chimique du Cd avec le Ca conduit à des carbonates de cadmium associés aux carbonates de calcium.

### **3. Types de charges et réactivité des constituants du sol**

Les minéraux secondaires peuvent présenter deux types de charges, constantes ou variables.

**Les charges constantes**, ou permanentes, ne varient pas en fonction du pH et résultent de substitutions isomorphiques au sein du réseau cristallographique, c'est-à-dire obtenues lors de la formation du minéral. Les composants inorganiques du sol présentant majoritairement ce type de charges sont les **phyllosilicates** par substitution de  $\text{Si}^{4+}$  par  $\text{Al}^{3+}$  dans les couches tétraédriques et  $\text{Al}^{3+}$  par  $\text{Fe}^{2+}$  ou  $\text{Mg}^{2+}$  dans les couches octaédriques (type 2:1, smectite, montmorillonite vermiculite...) et les **phylломanganates** (présence de cations hétérovalents dans les feuillets, ou réduction du  $\text{Mn}^{4+}$  et  $\text{Mn}^{3+}$  en  $\text{Mn}^{2+}$ ). Ces substitutions induisent un déficit de charges et donc une charge globale négative à la surface du minéral.

**Les charges variables**, second type de charges, viennent du caractère acido-basique des groupements fonctionnels capables de se protoner ou déprotoner en fonction du pH (Équation I.1 et Équation I.2).



Le terme S peut correspondre à une structure minérale ou organique. La capacité de chaque groupement à se déprotoner est caractérisée par une constante d'acidité notée  $K_a$ . Dans le cas de l'Équation I.2, la constante d'acidité correspond au rapport des concentrations :

$$K_a = \frac{[H^+][S-O^-]}{[S-OH]} \quad \text{Équation I.3}$$

Aussi couramment exprimée par le  $pK_a$  :

$$pK_a = -\log_{10}(K_a) \quad \text{Équation I.4}$$

Le  $pK_a$  des groupements silanol (Si-OH) est relativement élevé ( $pK_a = 9.5$ ) ce qui indique que l'ionisation de ce groupement est faible pour des sols très alcalins et donc limitée dans un contexte environnemental, même agronomique. Par conséquent, les argiles de type 2:1, ayant plus de groupements silanol, vont moins contribuer aux charges variables que les argiles 1:1 ou les oxydes métalliques ( $pK_a$  de  $Al[OH_2]^+ = \sim 5$ ,  $[Al-OH-Si]_{0.5}^+ = \sim 7$ ).

La MOS contribue plus largement aux charges variables (~80-90% ; Sparks 2003) par déprotonation des groupes carboxyliques (R-COOH ;  $pK_a=4-6$ ) et phénoliques ( $\phi-OH$  ;  $pK_a=9-11$ ), majoritaires, ainsi que des groupes sulfurés (R-SO<sub>3</sub>H ;  $pK_a=2$ ) et sulfhydryles (R-SH ;  $pK_a<7$ ), minoritaires (Aiken, 1985). Les amines font exception à la règle puisqu'elles sont neutres à pH ambiant, et positivement chargées à pH basique (N-H ;  $8.5 < pH < 12.5$ ). Toutefois, ce groupe minoritaire ne peut en aucun cas inverser la charge négative de la MO qui est principalement déterminée par les charges négatives des groupes carboxyles et phénols.

Ces charges négatives (constantes ou variables) vont caractériser la capacité d'un sol à retenir par force électrostatique des atomes ou molécules chargées positivement, aussi appelée capacité d'échange cationique (CEC). Une valeur de CEC caractéristique pour chaque constituant du sol peut également être définie. Certaines valeurs de surface spécifique et de CEC de minéraux présentant principalement des charges permanentes sont regroupées Tableau I.2.

*Tableau I.2 : Capacités d'échange cationique et surfaces spécifiques de différents minéraux secondaires d'après Sparks, 2003.*

Minéral secondaire	Surface spécifique (m <sup>2</sup> .g <sup>-1</sup> )	CEC (cmol <sup>+</sup> .kg <sup>-1</sup> )
Kaolinite	7-30	2-15
Illite	100-175	10-50
Smectite	700-800	80-150
Vermiculite	50-800	10-200

L'ensemble de ces charges, négatives et positives, et leur surface spécifique permettent de déterminer la réactivité d'une surface en définissant une densité de charges (charges par unité de surface) et une charge globale de surface. On définit également le pH au point de charge nulle ( $pH_{PCN}$ ) correspondant au pH auquel la charge globale de surface est, en moyenne, nulle. Lorsque le pH du milieu est inférieur au  $pH_{PCN}$  d'un constituant, la charge globale de surface est considérée positive et lorsqu'elle dépasse cette valeur la charge de surface négative permettra la sorption de composés cationiques. Les valeurs de  $pH_{PCN}$  de minéraux secondaires sont

présentées Tableau I.3. Ainsi lorsque le pH du sol est supérieur à ce  $pH_{PCN}$ , les oxydes métalliques vont également contribuer à la CEC des sols.

*Tableau I.3 : Valeurs de  $pH_{PCN}$  pour différents minéraux secondaires. Les méthodes d'estimation diffèrent (électrophorèse, titration) et par conséquent ne peuvent pas nécessairement être comparées.*

Minéral	$pH_{PCN}$	Reference
$\gamma$ -AlOOH	8.2	Sparks (2003)
$\alpha$ -FeOOH	7.8	Sparks (2003)
$\alpha$ -Fe <sub>2</sub> O <sub>3</sub>	6.7	Sparks (2003)
$\delta$ -MnO <sub>2</sub>	~1.5	Murray et al. (1968)
$\beta$ -MnO <sub>2</sub>	7.2	Sparks (2003)
Sites de bordures des argiles	5-7	Sposito (2008)

Cependant, malgré une légère incertitude sur les valeurs de point de charge nulle, il est clair que les oxydes de manganèse vont présenter une charge de surface négative aux pH généralement rencontrés dans les sols, contrairement aux oxydes de fer ou aluminium. Ceci indique une réactivité plus importante de ces surfaces vis-à-vis des cations, comparée aux autres oxydes métalliques. Cette réactivité est renforcée par la taille typiquement nanométrique des phylломanganates (et notamment la vernadite) qui induit une augmentation de sa surface spécifique. En effet, non seulement la surface spécifique des phylломanganates est environ 5 à 10 fois supérieures à celle de la goethite ( $100 \text{ m}^2.\text{g}^{-1}$  vs  $10\text{-}20 \text{ m}^2.\text{g}^{-1}$ ), mais leur charge de surface permanente est également trois fois plus élevée que celle des smectites (argile gonflante,  $3 \text{ meq}.\text{g}^{-1}$  vs  $1 \text{ meq}.\text{g}^{-1}$ ).

Ces propriétés physico-chimiques tout à fait exceptionnelles des oxydes de manganèse se traduisent dans le milieu naturel par des affinités géochimiques remarquables avec nombre d'ETM.

## **4. Dynamique du Cd dans les sols**

La solution du sol peut être définie comme la phase liquide aqueuse d'un sol et ses solutés en équilibre avec la phase solide et gazeuse. Elle est au centre de nombreuses réactions physico-chimiques ou processus contrôlant la dynamique des métaux dans les sols ainsi que leur spéciation, c'est-à-dire leur forme chimique (Figure I.5). En particulier, il s'agit du milieu dans lequel les plantes vont prélever les ions, pour la majorité, nécessaires à leur développement et libérer leurs exsudats. La dynamique des ions dans l'environnement passe par cette phase

aqueuse. Dans le sol, les ions peuvent se sorber sur les surfaces des minéraux et de la matière organique, ou même précipiter si les conditions physico-chimiques le permettent. Également, ils peuvent être transportés vers des eaux souterraines ou les eaux de surfaces par lessivage.

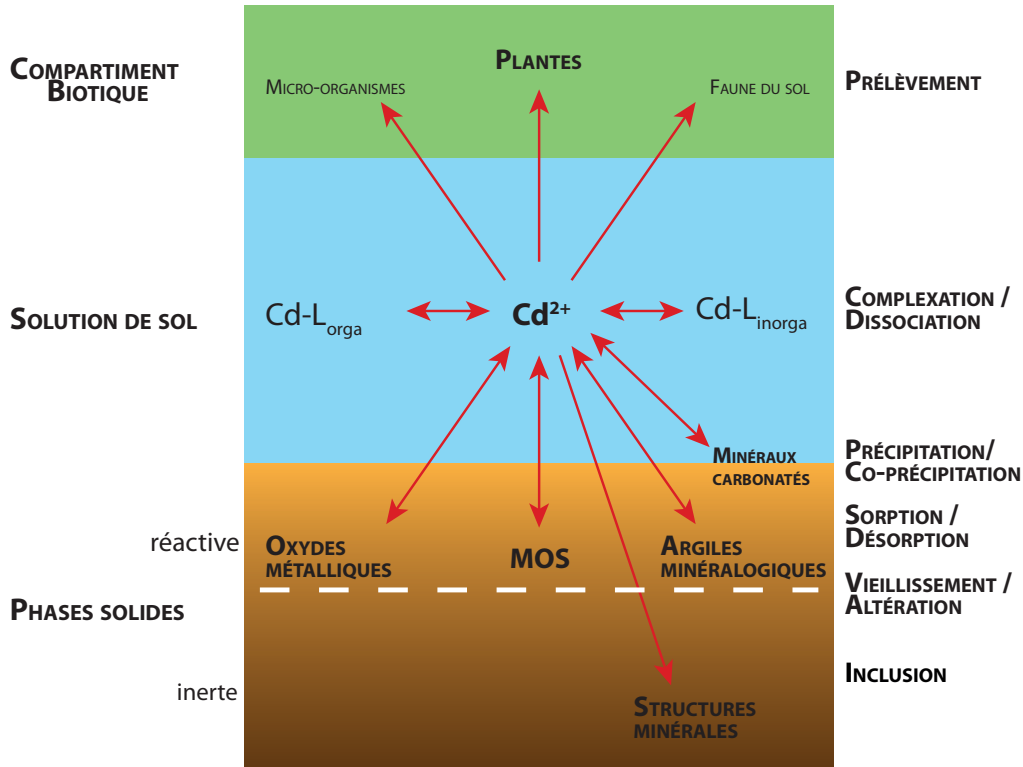


Figure I.5 : Schéma récapitulatif des principaux compartiments du sol et processus physico-chimiques et biologiques impliqués dans le devenir du Cd dans les sols.

## 4.1. Spéciation du Cd en solution de sol

### 4.1.1. Solution de sol

La solution de sol telle que définie dans ce chapitre et dans le reste du manuscrit correspond à l'eau porale représentée par deux types d'eau dans le sol, définie d'un point de vue opérationnel selon leur distance par rapport aux phases solides du sol (Di Bonito et al., 2008). En effet, les constituants du sol vont pouvoir retenir l'eau grâce à l'attraction générée par les groupements moléculaires des surfaces des constituants et la cohésion des molécules d'eau entre elles liée à la formation de liaisons hydrogènes. À mesure que l'on s'éloigne d'une surface, la force nécessaire pour mobiliser cette eau diminue, jusqu'à atteindre l'eau gravitaire dont la rétention par les surfaces solides est plus faible que la force de pesanteur. Les teneurs en eau retenues par les constituants du sol vont définir la capacité au champ qui est fonction de la texture et de la porosité des sols. Les phyllosilicates et les matières organiques du sol sont généralement les constituants du sol capables de retenir majoritairement l'eau.

La plante peut mobiliser jusqu'à l'eau capillaire (force de succion des racines supérieures aux forces de rétention des surfaces). D'un point de vue biogéochimique, cette eau proche des surfaces contrôle la diffusion dans la solution de sol de la fraction de Cd mobile sorbé tandis que l'eau gravitaire va gouverner le transport par convection et diffusion des solutés dans le sol.

#### 4.1.2. Spéciation du Cd en solution

La spéciation du métal en solution se définit comme l'ensemble des formes/espèces chimiques de cet élément dans un milieu naturel donné. Ainsi la composition de la solution, notamment par la présence de ligands, inorganiques ou organiques, va conditionner la spéciation du Cd via la formation de complexes plus ou moins stables. Il est généralement admis que la plante prélève le Cd principalement sous la forme ionique libre hydratée ( $Cd^{2+}$  ou  $Cd[H_2O]_6^{2+}$ ). Ainsi l'évaluation de la spéciation du Cd en solution apparaît comme essentielle pour la compréhension du transfert du Cd du sol vers les racines la plante.

En milieu aqueux, les formes chimiques du Cd peuvent être séparées en 2 grandes catégories

(1) Cd libre ou hydraté ( $Cd^{2+}$  ou  $Cd[H_2O]_6^{2+}$ )

(2) Cd complexé à des ligands inorganiques  $Cd_n(L_{inorg})_p^{2n-jp}$  ou organiques  $Cd_n(L_{org})_m^{2n-jp}$

D'un point de vue général, la réaction de complexation du  $Cd^{2+}$  et d'un ligand L (organique ou inorganique), libre, est décrite par la réaction d'équilibre suivante :



Ces réactions sont caractérisées par une constante thermodynamique de complexation à l'équilibre notée  $K_{CdL}$

$$K_{CdL} = \frac{[Cd_n(L)_p]}{[Cd]^n [L]^p} \quad \text{Équation I.6}$$

Avec les concentrations en Cd et L se référant aux métal et ligand libres en solution, c'est-à-dire pour L, non complexé à d'autres éléments. Plus cette constante est forte, plus le complexe est considéré comme stable. Il est également possible de caractériser le complexe d'un point de vue cinétique avec la notion de labilité. Un complexe labile se définit comme pouvant se dissocier/associer très rapidement.

Les complexes du Cd avec les ligands inorganiques présents en solution se font généralement avec les ions hydroxydes ( $OH^-$ ), chlorures ( $Cl^-$ ), nitrates ( $NO_3^-$ ), sulfates ( $SO_4^{2-}$ ), phosphates ( $PO_4^{3-}$ ) et carbonates ( $CO_3^{2-}$ ). Ces complexes ont été caractérisés et leurs constantes de complexation thermodynamiques avec le Cd sont connues et disponibles dans les bases de données thermodynamiques utilisées dans les logiciels de calculs de spéciation (Chapitre I.6.3). Les concentrations en ions hydroxydes et la formation de complexes  $Cd(OH)^+$ ,  $Cd(OH)_2, \dots$ , en solution

de sol sont principalement fonction du pH de la solution. Il est généralement considéré que, dans les sols alcalins, avec un pH de la solution de sol supérieur à 7, seuls les deux premiers produits d'hydrolyse du Cd se forment ( $\text{Cd}[\text{OH}]^+$ ,  $\text{Cd}[\text{OH}]_2$ ).

Les complexes du Cd formés avec des ligands de nature organique pouvant être différenciés suivant deux principales catégories :

(1) les ligands organiques de faibles poids moléculaires et de nature identifiée dont la réactivité vis-à-vis des cations est connue, comme les acides (mono-, di- ou tri-) carboxyliques aliphatiques (acide acétique, citrique ou oxalique...), les acides benzoïques plus ou moins substitués ou les acides aminés de type cystéine... Ces espèces sont généralement retrouvées en plus ou moins grandes concentrations dans les solutions de sol en fonction de l'usage des sols, des types de végétations les recouvrant et de la gestion des sols (apports d'amendements organiques pour les sols agricoles par exemple) (Strobel, 2001).

(2) les ligands organiques de fort poids moléculaire dont les structures sont très hétérogènes, la nature chimique exacte incertaine et dont la réactivité est plus difficilement quantifiable. Ces molécules, considérées comme des polyélectrolytes, possèdent un grand nombre de groupements fonctionnels différents dont la réactivité est très variable (Buffle et al., 1988). Les acides humiques et fulviques en font partie.

Ces deux sous catégories (1, 2) constituent les matières organiques dissoutes des sols (MOD). Les principaux complexes organiques de Cd retrouvés en solution sont principalement ceux formés avec les acides fulviques et les acides humiques présents en solution sous forme colloïdale et dans une moindre mesure ceux formés avec les molécules de plus faibles poids moléculaires.

Les contributions respectives des différentes formes du Cd à la concentration totale peuvent varier de façon importante en fonction du type de sol ou même de l'état de contamination des sols. Sauvé et al. (2000b) ont mesuré, sur 64 sols avec des caractéristiques pédogéochimiques variables, des teneurs en Cd total ( $0.1$  à  $38.1 \text{ mg.kg}^{-1}$ ) et en MOS ( $8.0$  à  $107.6 \text{ g C.kg}^{-1}$ ) différentes et des usages de sol variables (forestier, urbain, agricole), les concentrations en  $\text{Cd}^{2+}$  en solution. La spéciation moyennée par pH est représentée Figure I.6 avec une contribution du  $\text{Cd}^{2+}$  en solution s'échelonnant entre 10 et 40 % du Cd total en solution, du Cd complexé à la MOD allant de 30 à 70 % avec des concentrations en COD comprises entre  $1.4$  et  $180.9 \text{ mg C.kg}^{-1}$  et une contribution du Cd complexé aux espèces inorganiques en solution allant de 10 à 30 %.

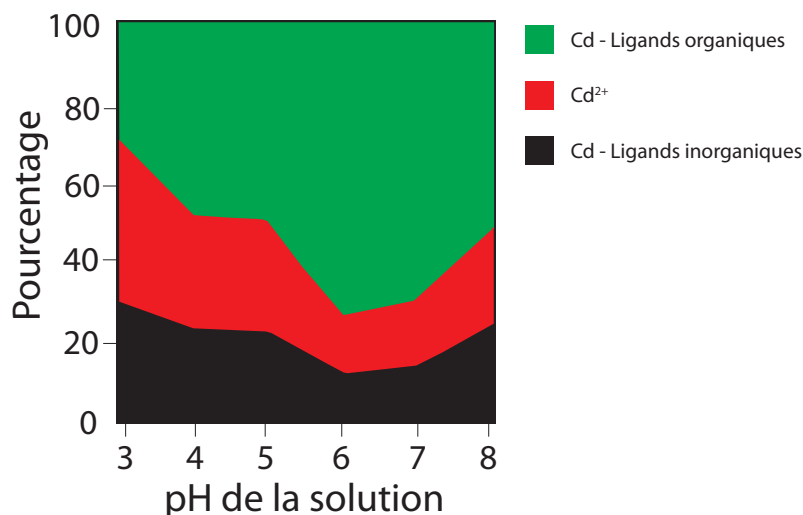


Figure I.6 : Répartition (%) entre les différentes formes du Cd en solution mesurées dans 64 solutions de sols en fonction du pH de la solution de sol (d'après Sauvé et al., 2000b).

L'affinité du Cd, ou d'autres éléments traces, pour les différents ligands en solution, c'est-à-dire leur capacité à former des complexes plus ou moins stables en solution, peut être approchée de façon globale par le principe HSAB (Hard/Soft Acid/Base) de Lewis, spécifiant qu'une espèce capable de perdre un électron est assimilée à une base et à un acide si l'espèce chimique peut accepter un électron. La notion de « douceur/dureté » d'un acide ou d'une base est estimée à partir des propriétés physico-chimiques des éléments. Ainsi, sa valence, son énergie d'ionisation et son rayon ionique permettent l'estimation du critère de Misono (Y) (Sposito, 2008). Si ce paramètre est fort (supérieur à 3), l'espèce est considérée comme molle, car polarisable, c'est-à-dire que son nuage d'électrons est facilement déformable. Dans le cas du Cd<sup>2+</sup>, ce paramètre est égal à 3.03, caractérisant cet ion comme un acide mou de Lewis. Le principe HSAB énonce que les acides durs vont avoir préférentiellement tendance à former des complexes avec des bases dures par interactions faibles (de type électrostatique), grâce à la faible polarisabilité de leurs structures électroniques et que les acides mous et les bases molles, comme le Cd, vont former de façon privilégiée des complexes par formation de liaisons covalentes. Ainsi, cette théorie explique la capacité du Cd à former des complexes plus stables que les cations majeurs, Ca, Mg (Y égal à 1.65 et 0.96 respectivement) notamment avec des bases intermédiaires à molles comme les sulfures, sulfates, thiocyanates, anilines ou les groupements thiols et amines des substances humiques.

Les structures électroniques des éléments traces métalliques (ETM) jouent également un rôle pour expliquer leur réactivité vis-à-vis des différents ligands, principalement en lien avec la formation de liaison  $\pi$ , plus stable (en plus d'une liaison sigma) grâce à une sous-couche  $d$  incomplète. Dans le cas du Cd, cette sous-couche est complète ( $[\text{Kr}].4d^{10}.5s^2$ ), ce qui



expliquerait l'affinité plus grande, comparée au Cd, des autres ETM (Cu, Pb) pour les constituants du sol.

### 4.1.3. Spéciation du Cd sur la phase solide

Les différents constituants organiques et inorganiques du sol présentés précédemment (Chapitre I.2) constituent des surfaces capables de fixer de manière plus ou moins réversible les éléments traces et donc réduire leur présence en solution. La capacité d'un élément à passer d'une surface solide à la solution de sol reflète sa mobilité. La fraction mobile de Cd, correspond donc à la fraction du Cd fixé aux différents constituants du sol **et** capable de passer en solution.

Les types de liaisons impliquées dans la fixation du Cd sur la phase solide vont donc être dépendants du constituant considéré, du type de sites et du type de charges. On définit généralement deux types de processus pour expliquer l'association du Cd avec la phase solide : la précipitation et l'adsorption.

D'autres processus conduisent à immobiliser le Cd comme l'inclusion dans les minéraux, mais ils ne seront pas explicités dans cette thèse, car généralement considérés comme irréversibles.

#### 4.1.3.a. L'adsorption/désorption du Cd

Dans les sols, il est généralement considéré que les processus d'adsorption et de désorption contrôlent les transferts du Cd entre le sol et la solution de sol (Loganathan et al., 2012). On distingue généralement deux types de mécanismes pour expliquer le processus d'adsorption. Le premier correspond à la formation de complexes par des liaisons covalentes/ioniques entre le métal et la surface, appelée chimisorption, ce sont des complexes de sphère interne. Le second correspond à des complexes de sphère externe résultant d'interactions électrostatiques entre la surface et le métal aussi appelé physisorption.

**Dans le cas de la physisorption**, aussi appelée échange d'ion ou adsorption non spécifique, des forces d'attraction électrostatiques de type Van der Waals sont mises en jeu en lien avec la polarisation des surfaces. Les complexes ainsi formés sont appelés de sphère externe, car dans le cas des cations sous forme libre, une partie de la sphère d'hydratation l'entourant va se retrouver entre la surface et le cation. Cette réaction est quasi instantanée et réversible, et a lieu, pour les cations, avec les surfaces chargées négativement.

Pour les pH de sols généralement étudiés (3-8), les principaux constituants portant une charge globale négative sont les phyllosilicates (charges fixes et variables), la MOS (charges variables) et dans une moindre mesure les oxydes, principalement, de manganèse (charges fixes et variables) compte tenu de leurs  $pH_{PCN}$  généralement faibles. Ce type de sorption n'est pas spécifique vis-à-vis de l'ion adsorbé. Cependant, la tendance qu'a une surface à former des

complexes de sphère externe va dépendre du potentiel ionique du cation considéré et plus particulièrement de son rayon ionique et de sa valence. Plus la valence est forte, plus la sorption non spécifique est favorisée. À valence égale, pour les cations divalents par exemple, les éléments ayant les rayons ioniques les plus forts, seront les plus favorablement sorbés par complexation de sphère externe (donc  $\text{Ca} > \text{Cd}$ ).

**Pour la chimisorption**, les liaisons impliquées sont beaucoup plus fortes que celle de la physisorption. Cette réaction se traduit par la perte de la sphère d'hydratation du métal et implique un échange d'électrons entre le groupement de surface et le métal, rendant cette réaction spécifique, c'est-à-dire que l'affinité sera fonction des configurations électroniques des éléments et plus difficilement réversible. Le processus de chimisorption se fait majoritairement sur les sites à charges variables des constituants du sol. Ainsi, tous les constituants du sol présentés précédemment présentant des sites de charges variables sont également capables de sorber spécifiquement les cations. La MOS, de par son nombre de sites de complexation et ces concentrations importantes dans le sol est généralement considérée comme la phase de sorption majoritaire du Cd par complexation spécifique, ensuite viennent les oxyhydroxydes métalliques (Fer, Aluminium, Manganèse) et pour finir les sites de complexation de bordures des phyllosilicates (Violante, 2013).

Les faibles concentrations du sol en Cd semblent également influencer le type de sorption. À fortes concentrations en Cd, par exemple dans le cas de sols contaminés, la sorption non spécifique serait dominante tandis que pour des sols faiblement contaminés, la sorption serait plus spécifique (Garcia-Miragaya et al., 1986; O'Connor et al., 1984).

#### **4.1.3.b. La précipitation/sorption du Cd avec les carbonates**

Il a été montré que même pour de forts pH, mais à de faibles concentrations en Cd, le produit de solubilité de l'otavite  $\text{CdCO}_3$ , n'est pas atteint et que ce minéral ne se forme a priori donc pas par précipitation dans ces conditions (Papadopoulos and Rowell, 1988). En sols alcalins non contaminés, la calcite ( $\text{CaCO}_3$ ) peut jouer un rôle important dans la sorption du Cd (Zachara et al., 1991). En effet, la fraction fine du calcaire total, souvent assimilée au calcaire actif, pourrait constituer une phase de sorption du Cd plutôt pour des pH élevés, supérieurs au point de charge nulle de la calcite proche de 8.2 (Pierre et al., 1990). De plus une forte affinité du Cd pour la surface de la calcite est attendue, puisque ce type de réaction tend vers la formation de  $\text{CdCO}_3$ , minéral plus stable que la calcite ( $\text{pKs CaCO}_3 = 8.48$  et  $\text{pKs CdCO}_3 = 12.0$ ) et que le rayon ionique de coordination 6 du Cd est proche de celui du Ca (0.095 et 0.100 nm respectivement) (Shannon, 1976). Cependant, cet exemple est typique de la définition d'une solution solide, soulignant la difficulté de distinguer les différents mécanismes de rétention du Cd sur la calcite (chimisorption, précipitation de surface, co-précipitation), puisqu'elle semble indiquer la

succession de ces différents mécanismes. À mesure que le pH de la solution de sol augmente, les processus d'adsorption non spécifique peuvent également prendre place, modifiant ainsi les contributions relatives des types de liaisons impliquées dans la sorption du Cd (Li et al., 2012).

## **4.2. Facteurs influençant la spéciation du Cd**

### **4.2.1. Le pH**

Le pH est généralement le paramètre physico-chimique dont l'effet sur la spéciation du Cd est le plus fort (Adriano, 2001).

L'un des effets du pH sur la sorption du Cd est lié à la modification des charges de surfaces variables du sol (Barrow, 1986; Naidu et al., 1994). En effet, à mesure que le pH de la solution de sol augmente, la densité de charges négatives augmente également en réponse à la dissociation des groupements hydroxyles de surface (*cf.* Chapitre I.3). Cette augmentation de la charge négative globale de la matrice solide du sol résulte dans une sorption privilégiée du Cd par physisorption.

À partir d'expériences de sorption/désorption successives, correspondant à la sorption de Cd ajouté ou natif du sol puis des lavages successifs pour désorber le métal, les auteurs ont montré que plus le pH augmentait, plus la quantité de Cd désorbé était faible (Gray et al., 1999). Cette irréversibilité de la désorption (après ajout de Cd) suppose que le Cd n'est plus adsorbé de façon non spécifique et donc qu'une augmentation du pH favoriserait l'adsorption de Cd de façon spécifique. Ceci pourrait être lié à la diminution de la compétition du proton et la complexation favorisée du premier produit d'hydrolyse du Cd ( $\text{CdOH}^+$ ) lié à la perte de sa sphère d'hydratation, lui permettant de s'approcher de la surface des particules résultant dans des interactions plus fortes et la formation d'une liaison plus stable (Hodgson et al., 1964). Des auteurs ont également supposé que l'hydrolyse du Cd pouvait être favorisée à la surface des particules grâce à un environnement entropique favorable. De plus, cette réaction d'hydrolyse a été supposée contrôlée par l'affinité de cette espèce avec la surface.

En solution, le pH joue également sur l'état de protonation des ligands organiques en fonction des constantes d'acidité des groupements fonctionnels considérés (Chapitre I.3). Plus le pH augmente moins les groupements seront protonés. Cette augmentation de la charge globale négative de surface augmente sa réactivité, favorisant la formation de complexes de sphères interne et externe.

Les interactions entre constituants du sol peuvent également être modifiées par le pH. Par exemple, la désorption de la MOS à pH alcalins en lien avec sa nature de polyélectrolyte chargé négativement (Curtin et al., 2016; Tipping and Woof, 1990) conduit à la présence plus importante de ligands en solution, favorisant la formation de complexes organiques de Cd et diminuant ainsi la quantité de Cd présent sous forme  $\text{Cd}^{2+}$ .

#### **4.2.2. Les cations compétiteurs**

La présence en solution de cations compétiteurs vis à vis du Cd va également influencer la concentration en  $Cd^{2+}$  en solution. Cette compétition peut être à la fois pour les sites de sorption diminuant en conséquence l'adsorption du Cd sur les différents constituants du sol, mais aussi pour les sites de complexation en solution.

On peut distinguer deux principaux types de cations pouvant avoir une forte influence sur la spéciation du Cd :

- les cations majeurs ( $Ca^{2+}$ ,  $Mg^{2+}$ ), présents à des concentrations beaucoup plus élevées que le Cd en solution de sol, surtout dans le contexte de sols non contaminés, et n'ayant pas une affinité particulière pour la sorption, et

- les autres éléments traces métalliques, pouvant être en concentrations plus importantes que le Cd mais ayant surtout une plus forte affinité ( $Zn^{2+}$ ,  $Cu^{2+}$ ,  $Pb^{2+}$ ) pour les sites de sorption. L'affinité des éléments traces pour les sites de sorption, généralement retrouvée dans la littérature, suit l'ordre suivant :  $Hg^{2+} > Pb^{2+} > Cu^{2+} > Zn^{2+} > Ni^{2+} \approx Co^{2+} > Cd^{2+}$  (McBride, 1994). On voit donc que le Cd montre l'affinité la plus faible pour les sites de sorption rendant la présence de ces cations compétiteurs capables de réduire la sorption du Cd sur le sol même en faibles concentrations (Christensen, 1987). Le calcium, souvent considéré comme le cation divalent majoritaire en solution a particulièrement été étudié comme compétiteur du Cd. Ainsi, de nombreux auteurs ont montré une inhibition de la sorption du Cd en présence de Ca (Christensen, 1984; Temminghoff et al., 1995; Tiller et al., 1979).

La compétition du Ca, ou des autres espèces cationiques en solution vont également impacter la complexation du Cd par les MOD ou les complexes inorganiques (Cao et al., 2006).

#### **4.2.3. Potentiel d'oxydo-réduction du sol**

L'influence du potentiel d'oxydo-réduction sur la spéciation du Cd n'a d'importance que dans les cas d'anoxie conduisant à des conditions réductrices. Pour les sols agricoles, le degré d'aération d'un sol, ou son niveau de dioxygène en solution, est majoritairement déterminé par sa texture, les pratiques culturales et les précipitations. En effet, l'irrigation ou de forts épisodes pluvieux, l'apport de MO biodégradables, la richesse des sols en argiles ou leur tassement sont des conditions favorisant l'anaérobie des sols. De plus, de forts épisodes pluvieux associés à une composition du sol riche en argiles ou MOS, peuvent favoriser la saturation en eau des sols et donc des épisodes d'anoxie. Le potentiel d'oxydoréduction est par exemple d'importance dans le cas de l'étude de la mobilité du Cd ou d'autres éléments dans les sols inondés de rizières (de Livera et al., 2011; Pan et al., 2015; Williams et al., 2012).

La baisse de la concentration en oxygène se traduit par une baisse du potentiel d'oxydo-réduction et donc d'une modification du nombre d'oxydation de certains éléments. Les plus impactés sont les éléments ayant une forte amplitude dans leur degré d'oxydation ou ceux présents en quantité importante dans le sol, comme l'azote, le phosphore, le soufre, le fer ou le manganèse (Husson, 2013).

Un changement du degré d'oxydation du Mn ou Fe peut induire la dissolution des oxydes composés de ces éléments, avec une dissolution plus facile des oxydes de manganèse comparée aux oxydes de fer (Suda and Makino, 2016) à un potentiel d'oxydo-réduction égal. Ainsi le Cd présent sur ces phases porteuses, peut se solubiliser et ensuite se redistribuer sur d'autres phases porteuses de façon plus ou moins spécifique modifiant ainsi la spéciation solide du Cd. Lorsque le milieu est considéré comme réducteur, le sulfate se réduit en sulfure, capable de former un précipité très peu soluble avec le Cd (CdS).

#### **4.2.4. Influence des micro-organismes et des racines**

Les micro-organismes présents dans les sols peuvent intervenir de façon directe via des processus physico-chimiques de biosorption, possibles à la fois sur la biomasse morte et vivante et des processus de bioaccumulation du Cd par la biomasse vivante. Ces deux processus tendent à diminuer les concentrations retrouvées en solution. Les micro-organismes peuvent également avoir une influence indirecte sur la solubilisation du Cd par la production de ligands anioniques capables de complexer le Cd en solution, par l'excrétion de protons augmentant la compétition pour les phases de sorption et pour les ligands complexant ou d'autres exsudats organiques complexant le Cd en solution (Deneux-Mustin et al., 2003).

Les racines des plantes modifient également la spéciation du Cd en solution par des mécanismes similaires : absorption du Cd<sup>2+</sup> qui se sorbe également sur les parois des cellules racinaires, excrétion de protons ou d'hydroxyles, sécrétion/exsudation de molécules complexantes.

### **5. Estimation de la distribution sol-solution**

#### **5.1. Estimation de la fraction mobile du Cd**

Nous avons vu dans les parties précédentes que la totalité du Cd du sol mesuré après dissolution complète des phases solides n'est pas représentative de la fraction mobile du Cd.

Afin d'étudier la distribution sol-solution du Cd et la spéciation solide du Cd, il convient donc d'essayer de déterminer la fraction du compartiment de Cd total qui est effectivement mobile et capable de s'échanger avec la solution dans un contexte donné.

Plusieurs méthodes sont reportées dans la littérature permettant d'estimer les différentes composantes du Cd mobile. Ces méthodes peuvent être catégorisées en deux grands

ensembles, le premier s'appuyant sur des analyses chimiques et la seconde sur des analyses physiques.

### **5.1.1. Les méthodes chimiques d'estimation de la fraction mobile de Cd dans les sols**

#### **5.1.1.a. Les extractions sélectives**

Ces méthodes d'estimation de la fraction mobile du Cd ne permettent pas à proprement parler d'estimer la spéciation solide du Cd dans les sols. Ces méthodes permettent la quantification d'une teneur en Cd sorbé sur le sol en fonction de la nature chimique de l'extractant à laquelle on fait correspondre des phases de sorption. Le principe consiste à mettre en contact une masse déterminée de sol avec une solution à un ratio V/m (volume/masse) fixé. Après agitation de la suspension obtenue pendant un temps déterminé, elle est centrifugée pour ensuite doser la quantité de Cd dans le surnageant. Le choix de l'extractant va donc conditionner la fraction de Cd estimée, et les mécanismes d'action supposés de cette solution sont utilisés pour interpréter les teneurs en Cd obtenues.

Parmi les extractants chimiques les plus utilisées, on retrouve les solutions salines ( $\text{Ca}[\text{NO}_3]_2$ ,  $\text{CaCl}_2$ ,  $\text{NH}_4\text{NO}_3$ ,  $\text{KNO}_3$  etc.) à des concentrations variables (Quevauviller et al., 1993; Sterckeman et al., 2006; Ure, 1996). Ces extractants chimiques sont supposés désorber le Cd par compétition des cations ajoutés pour les sites d'échanges et par la baisse de  $\text{Cd}^{2+}$  en solution résultant de la complexation du Cd en solution par les anions de l'extractant. Le type de sorption incriminé est ici l'échange d'ions, permettant alors d'estimer la fraction considérée comme la plus mobile du Cd du sol et donc la fraction de Cd du sol potentiellement assimilable par la plante (Kabata-Pendias, 2004).

Ensuite, par ordre croissant de force d'extraction, des molécules chélatantes peuvent être utilisées pour estimer les fractions de Cd retenue de façon plus forte dans les sols. Les complexants les plus communs sont l'éthylène diamine tétra-acétique (EDTA) ou l'acide diéthylène triamine penta-acétique (DTPA) (Sappin-Didier et al., 1997; Sterckeman et al., 2006). Ces molécules vont permettre la désorption non seulement du compartiment de Cd échangeable retenue par liaisons faibles, mais aussi le Cd sorbé de façon plus spécifique aux constituants du sol comme la MO ou les oxydes métalliques. Ces teneurs sont rarement corrélées aux teneurs retrouvées dans les plantes, mais permettent d'estimer la répartition du Cd sur les différentes phases solides du sol sans considération de sa participation au réapprovisionnement de la solution de sol. L'utilisation d'acide nitrique dilué ( $\text{HNO}_3$  0.43 M) est également répandue pour estimer la fraction de Cd réactif du sol (Sappin-Didier et al., 1997). Cet extractant fort va dissoudre partiellement les minéraux secondaires du sol et la MO permettant la solubilisation du Cd associé à ces constituants.

Pour terminer, certains extractants forts sont utilisés pour viser certaines fractions de Cd associé à des constituants particuliers. C'est le cas du mélange oxalate d'ammonium et acide oxalique à pH 3, supposé dissoudre les phases amorphes des minéraux des sols ou l'extraction à la dithionite pour la dissolution des phases amorphes et cristallines (McKeague and Day, 1966). Le pyrophosphate est quant à lui utilisé pour estimer la fraction de Cd associée à une partie de la MOS considérée comme facilement mobilisable (Garrett et al., 1998). La fraction acido-soluble du Cd, c'est-à-dire la fraction associée aux phases carbonatées peut également être estimée à partir d'une extraction à l'acétate de sodium acidifié à pH 5.5 par de l'acide acétique (Tessier et al., 1979).

Ces extractions n'ont de sélectif que le nom, et il est important de ne pas l'oublier lorsque ces extractions sont réalisées. Ce manque de sélectivité constitue une des limites de cette approche qui ne permet pas d'estimer avec certitude et en détail la spéciation du Cd sur les phases solides. Il convient également de ne pas oublier qu'un extractant appliqué sur le sol extrait le Cd qui est sorbé avec une force inférieure ou égale au pouvoir d'extraction de la solution chimique. Ainsi, pour un extractant « fort », une part du Cd extrait proviendra de la fraction faiblement liée et le reste correspondra à la fraction plus fortement liée ; il ne sera pas possible de les différencier sur la seule base de cette extraction (Lebourg et al., 1996; Rao et al., 2008; Ure, 1991).

Les principaux avantages des extractions chimiques sont opérationnels, avec une simplicité favorisant leur utilisation et leurs faibles coûts.

#### **5.1.1.b. Les extractions séquentielles**

Pour combler en partie le problème de sélectivité des extractions simples, il a été développé des procédures d'extractions séquentielles basées sur l'utilisation successive de différentes solutions d'extractions ayant une capacité de solubilisation du Cd croissante (Tessier et al., 1979).

De nombreuses procédures existent dans la littérature, variant selon le choix des différents extractants qui ne seront pas discutés ici (Cornu and Clozel, 2000; Laveuf et al., 2009; Quevauviller et al., 1993). Parmi ces différents protocoles, un a été normalisé par le BCR (Community Bureau of Reference) et a été décrit par Ure (1996).

Cette méthode d'extraction séquentielle permet d'identifier le Cd sorbé à différentes fractions supposées être : (a) échangeables, (b) acido-solubles c'est-à-dire liées aux phases carbonatées, (c) facilement réductibles, c'est-à-dire liées aux oxydes de manganèse, (d) facilement oxydables, liées à la MOS réactive, (e) moyennement réductibles, associées aux oxydes de fer et d'aluminium amorphes, (f) oxydables liées à la MOS, (g) difficilement réductibles (oxydes métalliques cristallisés) et (h) résiduelles correspondant au Cd associé aux phases minérales non réactives du sol.

Ces méthodes d'extractions chimiques bien qu'entachées d'approximation sont riches en informations si elles sont utilisées avec discernement. Elles sont une solution très prisée pour l'étude de la distribution du Cd sur les phases solides (Anju and Banerjee, 2011; Bakircioglu et al., 2011; Ren et al., 2015a; Schneider et al., 2016a).

#### **5.1.1.c. La méthode Diffusive Gradient in Thin Films (DGT)**

La méthode Diffusive Gradient in Thin Films (DGT), développée par Zhang et al. (2001), permet l'estimation de la fraction labile de Cd. L'outil DGT est un capteur passif, exposé pendant un temps défini au cours duquel, le métal va s'accumuler sur une résine très complexante, induisant un effet de puits infini à l'interface. Entre le sol et la résine, la succession d'un filtre (0.45 µm) et d'un gel de diffusion de porosité connue permet de contraindre l'accumulation à la forme libre du Cd, aux complexes inorganiques et organiques de petite taille capables de diffuser au sein du gel. La DGT peut être utilisée en solution de sol directement, mais elle nécessite alors un volume important de solution (~60 ml). Son déploiement à même le sol humide permet d'estimer, en plus, le réapprovisionnement de la fraction libre en solution par les complexes en solution et par la phase solide (Zhang et al., 2001). Cette méthode est souvent considérée capable de mimer le prélèvement du Cd ou d'autres éléments de la solution de sol par une racine, sous l'hypothèse (1) que la diffusion est le mécanisme de transport vers la racine largement dominant (pas ou très peu de convection), (2) que la concentration de l'ion libre à la surface racinaire est proche de zéro, comme dans le cas de la résine (forte capacité d'absorption racinaire par rapport à la diffusion du métal vers la racine) et (3) que l'état stationnaire prévaut pour le prélèvement racinaire. Ces hypothèses ne sont pas forcément vérifiées, surtout en sols très contaminés (Luo et al., 2014).

#### **5.1.1.d. La dilution isotopique**

La dilution isotopique est souvent considérée comme la méthode la plus attractive d'un point de vue conceptuel pour l'estimation de la fraction échangeable de Cd d'un sol. Cette méthode consiste à introduire dans un système sol-solution à l'état stationnaire, un isotope de l'élément étudié (stable ou radioactif ; Sivry et al., 2011) puis à suivre sa disparition de la solution au cours du temps. La disparition de l'isotope de la solution correspond à un transfert de la solution vers la phase solide du sol, c'est-à-dire un échange entre le traceur et l'élément stable adsorbé sur la phase solide.

L'isotope, ajouté à une suspension de sol à l'équilibre en faible quantité pour ne pas influencer sur l'importance de la sorption, va, de la même façon que l'élément tracé, se distribuer rapidement entre les phases solides et liquides (Fardeau, 1993). Il est donc possible d'estimer la quantité



d'éléments mobiles présents sur la phase solide susceptible, dans certaines conditions, de rejoindre la phase aqueuse après un temps  $t$  donné.

Cette méthode, bien que conceptuellement attrayante, est lourde méthodologiquement, car elle fait appel à l'utilisation d'isotopes. Elle ne peut pas être considérée comme opérationnelle en routine. Pour pallier cette limite, cette méthode a souvent été comparée à l'extraction sélective au  $\text{CaCl}_2$  à 1 M (Sterckeman et al., 2009; Young et al., 2000) montrant une bonne adéquation entre les valeurs mesurées par ces deux méthodes et concluant que l'extraction chimique par  $\text{CaCl}_2$  à 1 M des sols pouvait servir de proxy pour mimer la fraction isotopiquement échangeable (considérée généralement comme la fraction échangée en une minute).

### **5.1.2. Méthodes physiques pour la caractérisation de la spéciation solide du Cd**

Les méthodes physiques pour l'étude de la structure des constituants et des éléments sorbés à leurs surfaces sont nombreuses (D'Amore et al., 2005). Dans un premier temps, le fractionnement physique permet d'apporter une légère homogénéité à l'échantillon de sol et d'identifier les fractions granulométriques les plus à même de contenir la majorité des phases sorbant le Cd. Ce prétraitement est également nécessaire pour une majorité des méthodes de caractérisation.

Les méthodes spectroscopiques reposent sur les interactions entre un faisceau incident et l'échantillon. Le faisceau peut être soit des électrons, soit des ions, soit des rayons X avec pour chacun, des objectifs différents (analyse élémentaire, tomographie, spéciation). Le bombardement d'électrons sur une surface (par exemple microscopie électronique à balayage, MEB) permet, en couplant à des analyses en énergie, d'analyser les photons réémis et permettre l'analyse élémentaire qualitative ou semi-quantitative de la surface bombardée.

Le bombardement de rayons X d'énergies variables sur la surface constitue le second groupe de méthodes principalement utilisées pour l'étude de la spéciation des éléments sur les phases solides : les méthodes SAX (Spectroscopie d'Absorption de rayons X et méthodes associées EXAFS, XANES...) permettent d'identifier de façon qualitative uniquement, le type de complexes formés entre la surface et le Cd. L'analyse par diffraction de rayons X permet une analyse de la structure et de la quantité des différents minéraux cristallins et peut être utilisée pour l'identification des types d'argiles minéralogiques présents dans l'échantillon (Ren et al., 2015a). Toutes ces méthodes sont essentielles pour la compréhension des types de liaisons entre les métaux et les différents constituants des sols (Spadini et al., 1994; Vasconcelos et al., 2008). Elles favorisent le développement et l'affinement de modèles de complexation de surface permettant une formulation plus représentative des mécanismes réels (modèle CD-MUSIC).

Cependant, aucune de ces méthodes ne permet de déterminer de façon vraiment quantitative et de façon suffisamment précise les teneurs en éléments sorbés (principale limite pour l'étude de la spéciation solide du Cd en sols non contaminés).

Le couplage entre méthodes chimiques et méthodes physiques pour l'étude de la spéciation des métaux permet de valider les mécanismes sous-entendus par les extractions chimiques grâce à la vision moins quantitative, mais plus réaliste apportée par les méthodes physiques (Sivry et al., 2010).

## **5.2. Estimation de la fraction libre de Cd en solution**

Dans les sols, il y a un consensus sur le fait que la fraction libre du Cd dans la solution est la plus corrélée à l'adsorption d'une part et à la toxicité de cet élément pour les organismes vivants d'autre part (Campbell, 1995; Cancès et al., 2003; Ponthieu et al., 2016).

De nombreuses techniques ont été utilisées pour déterminer les concentrations en Cd<sup>2+</sup> en solution (Cornu et al., 2009; Kalis et al., 2007; Schneider, 2006; Sigg et al., 2006; Unsworth et al., 2006; Zhang and Young, 2005).

On distingue parmi elles, les méthodes électrochimiques, comme la potentiométrie (électrodes sélectives Ion Selective Electrode). Cette méthode n'est cependant pas applicable au contexte de l'étude de la fraction libre du Cd en solution de sols non contaminés à cause de faibles limites de détection pour le Cd, d'un manque de sélectivité et des interférences liées à la composition de la solution. Cette méthode n'a réussi à s'imposer que pour la mesure de la fraction libre de Cu avec une sensibilité pouvant atteindre jusqu'à 10<sup>-15</sup> M (Zhang et al., 2001). Une seconde méthode retrouvée dans la littérature est la voltampérométrie (Redissolution Anodique ASV ou méthodes impulsives comme la DPV ou SWV). Cette méthode n'est pas totalement liée à la seule forme libre du métal à l'équilibre, mais intègre également des processus cinétiques comme le transport du Cd complexé jusqu'à l'électrode de travail et sa dissolution dans la couche de diffusion de l'électrode. Parat et al., (2009) ont estimé la limite de détection de cette méthode pour le Cd à 2 10<sup>-09</sup> M.

L'emploi de résines échangeuses d'ions peut également permettre d'estimer la fraction libre du Cd en solution de sol. La méthode proposée par Schneider (2006), repose sur la quantification des formes cationiques du Cd (majoritairement Cd<sup>2+</sup>) capables de s'échanger sur une résine échangeuse de cations, saturée en Ca. Cependant, la présence d'autres espèces en solution, capables de désorber du Ca, et le fait que la variation de concentrations en solution du Cd est faible, impose l'utilisation d'un traceur radioactif du Cd (<sup>109</sup>Cd) qui est ajouté en faible quantité afin de suivre plus précisément la disparition du Cd en solution.

La connaissance de la radioactivité en solution et de la concentration en Ca avant et après contact ( $R$ ,  $r$  et  $C_{Ca,i}$  et  $C_{Ca,f}$  respectivement), permet d'estimer la fraction ionique libre de Cd échangé avec la résine ( $f_{Cd}$ ) grâce à la relation :

$$f_{Cd} = \frac{\left(\frac{R}{r} - 1\right) \times C_{Ca} \times V}{{}^V K_{Cd} \times R_{Ca,f} \times m} \times 100 \quad \text{Équation I.7}$$

où  ${}^V K_{Cd}$  est le coefficient de sélectivité de Vanselow pour l'échange Ca/Cd et  $R_{Ca,f}$  la quantité de Ca restant sur la résine.  ${}^V K_{Cd}$  est égal à 0.661 pour la résine utilisée (Schneider, 2006).  $R_{Ca,f}$  est calculé suivant l'équation suivante :

$$R_{Ca,f} = R_{Ca,i} + (C_{Ca,i} - C_{Ca,eq}) \times \frac{V}{m} \quad \text{Équation I.8}$$

avec  $R_{Ca,i}$  correspondant à la quantité de Ca initialement présente sur la résine.

Cette méthode est particulièrement adaptée lorsque les volumes de solution disponibles sont faibles.

La dernière méthode souvent reportée dans la littérature est la DMT (Donnan Membran Technique) (Cancès et al., 2003; Temminghoff et al., 2000; Weng et al., 2002b). Elle connaît un intérêt grandissant, car elle permet la mesure simultanée des concentrations de plusieurs espèces cationiques en solution avec des limites de détection relativement faibles. En bref, le principe est basé sur l'équilibre de Donnan entre une solution dite « donneur » (la solution de sol) et une solution « accepteur » de force ionique proche de celle de la solution « donneur ». Ces deux solutions sont séparées par une membrane échangeuse de cations. La charge négative de cette membrane permet le passage des composés cationiques suivant leur charge et leur rééquilibrage du côté de la solution « accepteur » (Temminghoff et al., 2000).

## **6. Modélisation de la distribution sol-solution**

### **6.1. Le coefficient de distribution**

Le coefficient de distribution représente un paramètre lié à des expériences de sorption/désorption. La Figure I.7 représente les différentes isothermes utilisées pour modéliser la sorption du Cd sur une surface sorbante en représentant le compartiment mobile du Cd (noté  $q$ ) en fonction de la concentration en solution (notée  $C$ ).

L'isotherme de Langmuir dont l'équation et la forme sont données Figure I.7, permet de représenter un système pour lequel la sorption du métal présente un palier ( $q_{max}$ ) qui est interprété comme correspondant à une saturation de la surface de sorption. Ce genre de modèle est

généralement utilisé dans les études visant à étudier la capacité de sorption d'une surface souvent pure et n'est pas forcément adapté dans l'étude de la sorption du Cd sur les phases porteuses du sol où la faible concentration du métal rend improbable une saturation. Il lui est généralement préféré les isothermes de Freundlich (Figure I.7), utilisées sur des gammes de teneurs en Cd total du sol plus représentatives de sols contaminés. Pour terminer, si l'on considère de faibles concentrations en Cd en solution, la variation du compartiment mobile  $q$  est suffisamment faible pour être considérée comme linéaire et la pente est associée au  $K_d$ .

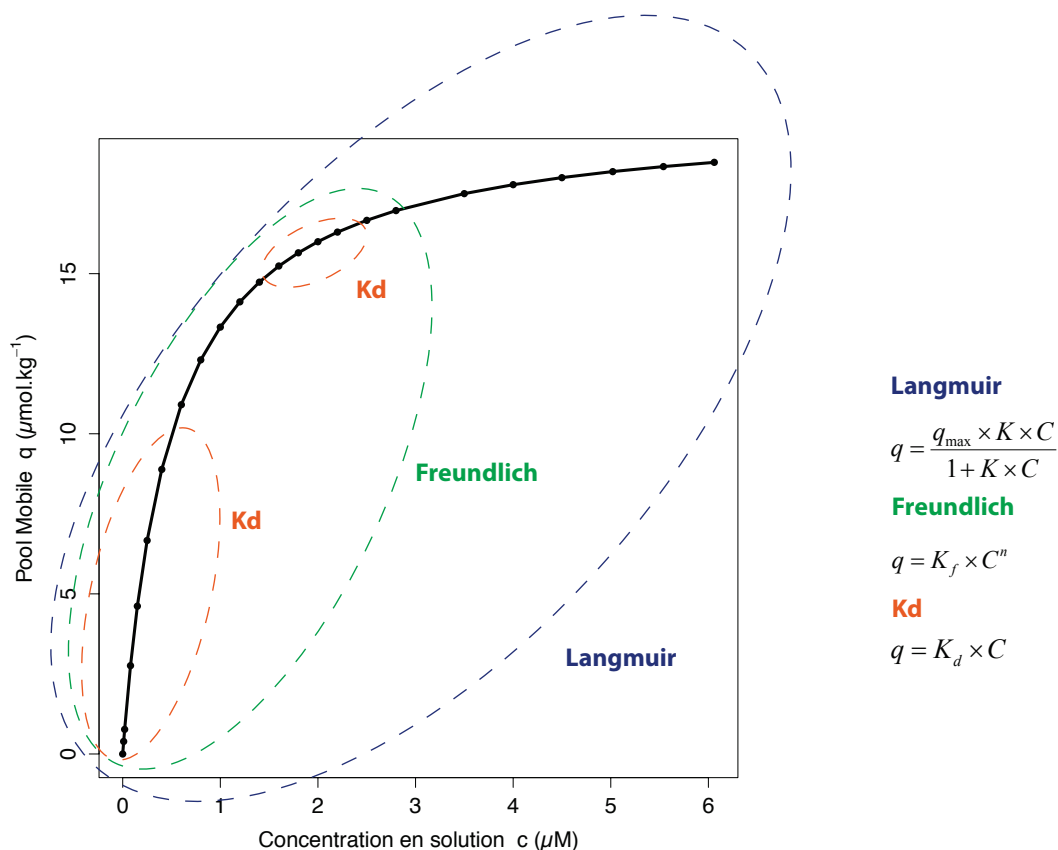


Figure I.7 : Différentes isothermes (Langmuir, Freundlich et Kd) utilisées pour modéliser la sorption du Cd sur une surface sorbante.

Le  $K_d$  retrouvé dans la littérature est généralement associé au rapport de la concentration en Cd du sol et de la concentration en solution (Figure I.7). Dans le meilleur des cas, il est déterminé comme étant la pente des isothermes de sorption dans le domaine de concentrations étudié. Cependant, cette méthode dynamique pour l'estimation du coefficient de partition entre le sol et la solution suppose une réversibilité de la sorption afin de compenser les variations solution. Cette réversibilité de la sorption a souvent été réfutée (Loganathan et al., 2012). Parfois, le  $K_d$  est calculé très simplement comme l'unique rapport entre le Cd sur la phase solide et le Cd en solution pour un rapport volume de solution:masse de sol fixé.

Dans tous les cas, le  $K_d$  représente la distribution à l'équilibre d'un métal en fonction de la détermination d'un compartiment de Cd dans le sol et de sa concentration en solution. Ce rapport est donc dépendant des méthodes utilisées pour caractériser les fractions de Cd dans le sol et en solution. Une des principales limites de cette approche est liée au biais induit sur le calcul du  $K_d$  lorsque le compartiment de Cd du sol utilisé n'est pas représentatif de la fraction mobile et que le  $K_d$  est calculé par l'unique rapport entre le Cd sur la phase solide et le Cd en solution (Figure I.8).

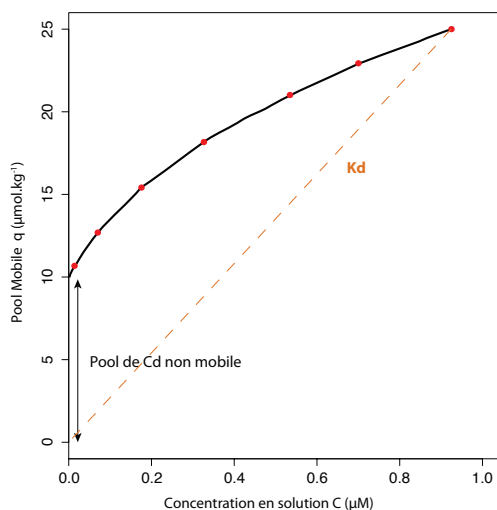


Figure I.8 : Représentation graphique du biais induit par le calcul du  $K_d$  lorsque le compartiment considéré ( $q$ ) dans le calcul n'est pas le compartiment mobile de Cd.

En effet, le  $K_d$  devient la corde de l'isotherme au lieu d'être la pente. Dans ces conditions, même si l'isotherme est linéaire et donc même si le rapport compartiment mobile:Cd en solution est constant (pente constante et donc  $K_d$  réel constant), la mesure expérimentale du  $K_d$  comme étant le rapport Cd total (et non uniquement mobile):Cd en solution va augmenter lorsque le Cd en solution baisse. Les valeurs de  $K_d$  mesurées vont donc être fonction des mécanismes considérés et de la méthode de mesure : méthodes possibles pour estimer les teneurs en Cd du sol (réactif ou non) et les concentrations en Cd en solution (libre ou non).

Degryse et al., (2009) ont décrit les mécanismes potentiels pouvant expliquer la distribution sol-solution du Cd et les coefficients de partition associés suivant les méthodes de détermination (Figure I.9).

La variabilité des valeurs retrouvées dans la littérature est très importante et potentiellement faussée si l'estimation du compartiment mobile n'est pas bonne.

À cela s'ajoute que la distribution du Cd en solution dépend de paramètres qui peuvent être très variables pour un sol agricole : teneur en matière organique soluble, teneur en cations compétiteurs (Ca notamment) et pH sont les principaux.

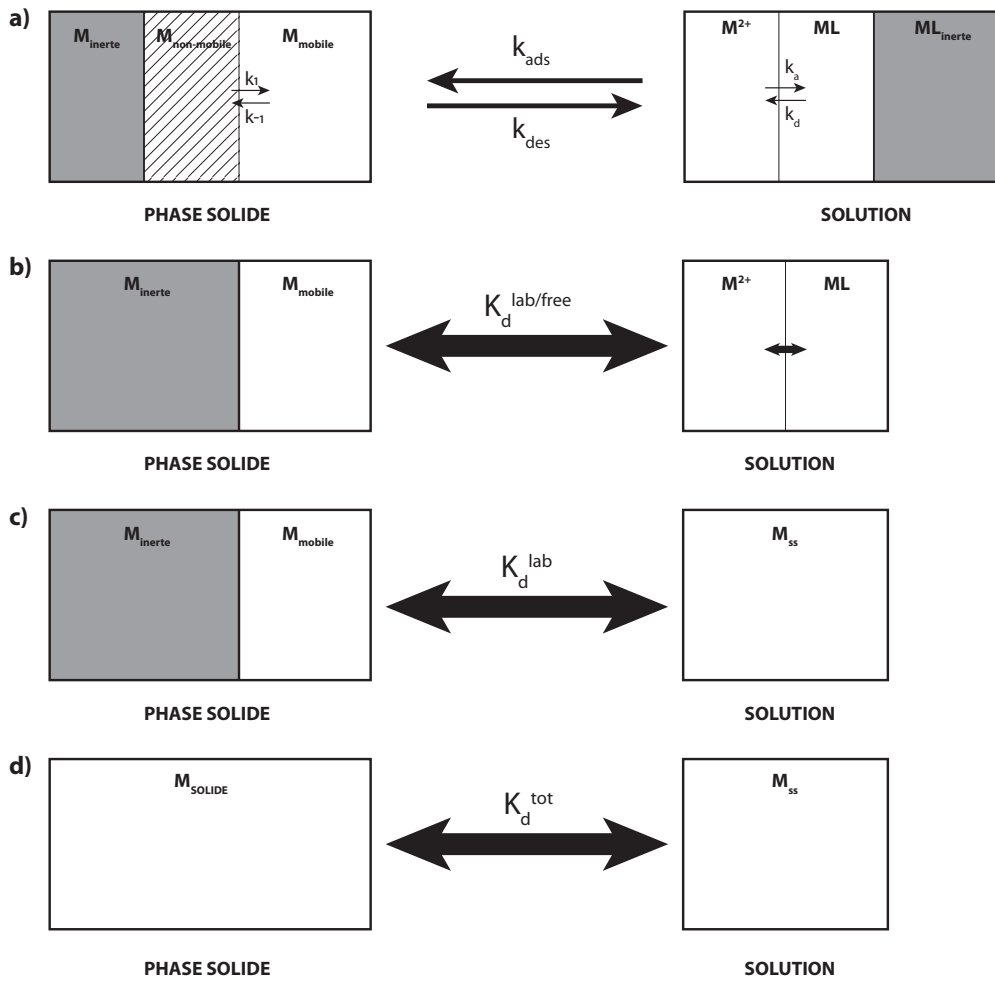


Figure 1.9 : Descriptions des différents modèles de  $K_d$  décrivant la distribution du métal dans un sol. (a) correspond à la représentation la plus complexe en prenant en compte l'adsorption et la désorption du métal sous sa forme libre ( $M^{2+}$ ) sur et de la phase solide et la cinétique d'association et de dissociation des complexes  $ML$  en solution. Les complexes en solution sont considérés comme ne se sorbant pas. Certains complexes sont inertes tout comme une partie du Cd de la phase solide et ne participent pas au passage de  $Cd^{2+}$  en solution. Les réactions sont décrites à partir de constantes cinétiques d'ordre 1 décrivant des processus plus ou moins rapides de réapprovisionnement ou d'appauvrissement de la solution en  $Cd^{2+}$  ( $k_{ads}$ ,  $k_{des}$ ,  $k_a$ ,  $k_d$ ,  $k_1$ ,  $k_{-1}$ ). (b-d) sont des visions plus simples à l'équilibre en considérant des compartiments de Cd plus ou moins labiles du sol et de la solution du sol. Le ratio du métal mobile du sol sur le métal libre en solution ( $K_d^{lab/free}$ ) ou la concentration totale en solution ( $K_d^{lab}$ ) ou pour finir le ratio du compartiment de Cd total du sol sur la concentration totale de Cd en solution ( $K_d^{tot}$ ). Adapté de Degryse et al., (2009)

## 6.2. Modélisation empirique de la solubilité du Cd dans les sols

Dérivée de la méthode précédente, une approche empirique permet également de modéliser la distribution du Cd. Elle consiste à établir des relations statistiques par régressions linéaires multiples entre une teneur du sol en Cd considéré mobile ( $Q$ ), les concentrations totales/libres du

Cd en solution (C), ou même le rapport  $K_d$  défini précédemment en fonction de variables explicatives du sol comme le pH, la teneur en MO du sol, les concentrations en carbone organique dissout (COD), la CEC, les teneurs en oxydes métalliques cristallisés ou amorphes, etc. (de Groot et al., 1998; de Vries et al., 2011; Degryse et al., 2003; Elzinga et al., 1999; Groenenberg et al., 2010b; Ivezić et al., 2012; Lee et al., 1996; Römkens et al., 2004; Sauvé et al., 2000a; Tipping et al., 2003).

Trois principaux types de relations sont retrouvées dans la littérature :

$$\log_{10}(C) = a + b \log_{10}(Q) + \sum c_i \log_{10}(Var_i) \quad \text{Équation I.9}$$

$$\log_{10}(Q) = a + b \log_{10}(C) + \sum c_i \log_{10}(Var_i) \quad \text{Équation I.10}$$

$$\log_{10}(K_d) = a + \sum b_i \log_{10}(Var_i) \quad \text{Équation I.11}$$

Avec, les lettres  $a$ ,  $b$  et  $c$  correspondant aux coefficients des variables explicatives ( $Var_i$ ). Le pH étant déjà log transformé, les valeurs de ce paramètre sont utilisées sans transformation dans les équations précédentes (Équation I.9 à Équation I.11).

Les Équation I.9 et Équation I.10 permettent de modéliser une sorption linéaire ou non linéaire du métal suivant que le coefficient  $b$  est égal ou non à 1. Ces deux équations sont équivalentes à l'équation de type Freundlich (Figure I.7) dont le paramètre  $n$  représentant les variations de l'intensité de liaison en fonction de la teneur en métal sorbé correspond au paramètre  $b$  des Équation I.9 et Équation I.10. Mathématiquement équivalentes, ces deux équations n'optimisent pas les erreurs sur le même terme (C ou Q) induisant des variations dans les coefficients de détermination.

Dans le cas du Cd, les paramètres généralement reportés dans ces modèles empiriques (peu importe l'approche) sont le pH du sol, les teneurs en Cd total (à défaut d'avoir estimé un compartiment de Cd réactif), les teneurs en MO, la CEC ou les teneurs en Mn, Fe total du sol (de Vries et al., 2011; Meers et al., 2005; Römkens and Salomons, 1998; Schröder et al., 2005).

Les principaux avantages de cette approche sont sa facilité de paramétrage et d'utilisation. Ce type de modélisation requiert cependant un certain nombre de données de paramétrages qui sont assez faciles à acquérir pour une utilisation dans un contexte régional, multipliant ainsi les études disponibles dans la littérature.

Le principal désavantage de cette approche est le caractère conditionnel des modèles aux données de paramétrage. En effet, ces modèles sont fortement dépendants du contexte dans lequel ils ont été calibrés, comme le type de sols utilisés (Tipping et al., 2003), les paramètres

incorporés aux modèles conditionnant celles qui seront identifiées, la méthode d'ajustement utilisée, et les méthodes de mesure des différentes variables explicatives. Il est donc primordial de bien considérer le domaine de validité des équations empiriques avant de les appliquer.

### **6.3. Modélisation géochimique**

Les modèles géochimiques sont des modèles de complexation de surface, élaborés dans le but de décrire l'interaction entre les ETM et les différentes phases réactives du sol.

La distribution à l'équilibre du Cd repose donc sur les spéciations du Cd en phase solide et dans la solution de sol. La complexité du système sol impose l'utilisation de codes de calculs de spéciation afin de correctement prendre en compte les différents effets combinés de différents paramètres physico-chimiques sur la spéciation du Cd. Ce type de modèles permet donc une représentation assez complète des mécanismes contrôlant la spéciation du Cd, permettant ainsi une vision quantitative des différentes formes de Cd dans les sols, indispensables pour la compréhension du transfert sol-solution du Cd.

#### **6.3.1. Modèles de complexation par les substances humiques**

Deux principaux modèles sont généralement confrontés dans la littérature lorsque la complexation des éléments traces par les substances humiques est étudiée, le modèle HIB (Humic Ion Binding, Tipping and Hurley, 1992) et le modèle de NICA-Donnan (Non-Ideal Competitive Adsorption, Benedetti et al., 1995; Kinniburgh et al., 1996) dont le principe général est énoncé ci-dessous.

Le **modèle HIB** développé par Tipping and Hurley, (1992) considère la sorption du proton et des cations sur les substances humiques, en supposant une distribution discrète des sites de sorption spécifique. Le modèle définit 8 sites différents, 4 assimilés à des groupements carboxyliques (types A) et 4 à des groupements phénoliques (Type B). Dans le but de minimiser les paramètres nécessaires au modèle, la description des types de sites (densités de sites et affinités relatives) a été structurée. Ainsi, les affinités de sites A et B sont égales et il y a deux fois plus de sites de type A que de type B. Les termes définissant les constantes de complexation du proton et du cation sont classiquement exprimés et comprennent un terme supplémentaire pour intégrer l'hétérogénéité des affinités des sites à l'intérieur des deux types.

Selon la valence de l'espèce sorbée considérée, la formation de complexes monodentés, bidentés et tridentés, est possible suivant une loi de probabilité permettant d'estimer la possibilité que deux ou trois sites de sorption, aléatoirement placés sur la surface, se retrouvent à une distance suffisamment faible (définie) pour permettre leur formation.



Le **modèle NICA** (Non-Ideal Competitive Adsorption) (Benedetti et al., 1995; Kinniburgh et al., 1996) utilise quant à lui une distribution continue pour représenter la diversité des sites de sorption. Cela permet de considérer les SH comme une multitude de ligands ayant chacun une affinité spécifique pour le cation considéré. L'hypothèse à la base de ce modèle est que le métal peut se complexer sur des sites capables de se déprotoner. Ce modèle considère deux sites de complexation (1 et 2), chacun défini par une densité de sites et une distribution d'affinités différentes rendant compte de l'hétérogénéité des substances humiques. Ces deux sites sont associés à des groupements carboxyliques et phénoliques.

Ces deux modèles intègrent également un terme permettant de prendre en compte la sorption non spécifique par la dissociation des protons des groupements carboxyliques et phénoliques en fonction du pH (charges variables) créant ainsi une charge de surface négative. Le modèle simple de Donnan considère que l'électroneutralité du système doit être conservée à tout moment par l'action des contre-ions chargés positivement. Ce modèle permet de définir une couche de Donnan, autour des substances humiques, isolée de la solution et dans laquelle les concentrations en contre-ions sont plus élevées.

Les paramètres génériques utilisés par ces deux modèles pour représenter la complexation du métal et du proton sur les AH et AF ont été dérivés à partir d'un nombre important de données obtenues dans des conditions expérimentales différentes (Lofts and Tipping, 1998; Milne et al., 2003, 2001; Tipping et al., 2011) et sont régulièrement mis à jour.

### **6.3.2. Interactions avec les oxyhydroxydes métalliques**

Les interactions des ETM avec les oxydes métalliques et plus particulièrement les groupements hydroxyles amphotériques de surfaces sont décrites par les équations de la loi d'action de masse auxquelles s'ajoute une correction des constantes d'affinités pour la prise en compte des effets électrostatiques causés par les charges portées par les groupements hydroxyles (Chapitre I.3). Ce terme de correction change entre les modèles présentés ci-dessous, mais ne fera pas l'objet d'une explication plus détaillée, car il n'est pas encore clair comment ces termes modifient les calculs (Di Bonito et al., 2018).

Ces modèles considèrent dans leur ensemble un système multicouches entre la surface et la solution de sol pour expliquer les différentes sorptions du métal à la surface du constituant (complexation de sphère interne, externe ou adsorption électrostatique).

À partir de là, plusieurs modèles ont été développés pour représenter ces interactions. Parmi eux, les plus utilisés sont les modèles DDL (Diffuse Double Layer, Stumm et al., (1970)) et son amélioration GTLM (Generalized Two Layer Model) par Dzombak et Morel, (1990) et le modèle CD-MUSIC (Charge Distribution - Multi-Site Ion Complexation ; Hiemstra & van Riemsdijk, 1996). Le premier (**DDL-GTLM**) considère un système à deux couches, la première, proche de la surface, dans laquelle les réactions sont assimilées à de la complexation spécifique, et la seconde, la couche de diffusion, dans laquelle des contre-ions peuvent s'accumuler pour compenser les charges des groupements hydroxyles de surface. Deux ionisations de surfaces différentes sont prises en compte pour représenter le caractère amphotérique des groupements hydroxyles, c'est-à-dire deux constantes de protonation. De plus, deux sites de surfaces sont considérés, un plus affine, mais moins abondant et le second moins affine et plus abondant. Le **modèle SCAMP** (Lofts and Tipping, 1998) implémenté dans le logiciel WHAM utilise les mêmes bases théoriques que le GTLM en considérant cependant 3 types de sites de surfaces d'affinités différentes, mais dont les abondances suivent la même logique, c'est-à-dire que les plus abondants sont les moins affines. En plus, ce modèle permet de prendre en compte la sorption du premier produit d'hydrolyse du Cd ( $\text{CdOH}^+$ ) ce qui n'est pas le cas du GTLM.

Le **modèle CD-MUSIC** développé par Hiemstra & van Riemsdijk, (1996) peut se décomposer en deux sous-modèles : le premier, CD (Charge Distribution) décrit l'interface sol-solution en distinguant 3 couches : (1) la surface, (2) une couche intermédiaire et (3) une troisième couche extérieure. Cette représentation permet de positionner les complexes de sphère interne en distribuant leur charge entre la couche de surface, et la couche intermédiaire et les complexes de sphère externe avec leur charge répartie entre la couche intermédiaire et externe. La distribution de ces charges est liée à la structure moléculaire du complexe de surface formé, information obtenue à partir de données d'analyse spectroscopique comme l'EXAF (Ponthieu et al., 2006). La partie MUSIC de ce modèle permet la prise en compte de la structure de la surface et plus particulièrement de la coordination des atomes d'oxygène de surface. La constante d'affinité du proton des différents groupements de surface est prédite à partir de la charge du complexe métal-surface et de la longueur de la liaison. La protonation est décrite à partir d'un modèle à 1 pK (Hiemstra and van Riemsdijk, 1996) aux pH généralement rencontrés dans les sols.

Pour les modèles GTLM et SCAMP, les paramètres génériques de complexation pour le proton et les métaux pour différents oxydes métalliques (Fe, Al, Mn, Si) (Dzombak and Morel, 1990; Lofts and Tipping, 1998; Tipping et al., 2011; Tonkin et al., 2004) sont disponibles pour une large gamme de conditions expérimentales et éprouvés dans la littérature. Cependant, l'approche du modèle CD-MUSIC plus avancée pourrait potentiellement être plus robuste.

### **6.3.3. Modélisation des interactions métal-argiles**

Les surfaces argileuses sont généralement modélisées comme une surface simple d'échanges cationiques en considérant l'attraction électrostatique de composés dans la couche de diffusion à partir des équations de Donnan (Tipping, 1994). Cette accumulation est donc supposée non spécifique et fonction de la charge et de la concentration des éléments. Les argiles sont caractérisées par leur abondance (teneur), leur surface spécifique (surface d'échange:unité de masse) et par une capacité d'échange cationique spécifique (CEC). Il est généralement considéré que cette CEC est comprise entre  $0.1 \text{ mol}^+.\text{kg}^{-1}$  et  $0.4 \text{ mol}^+.\text{kg}^{-1}$  (Ren et al., 2015a; Schneider et al., 2016a; Sparks, 2003; Weng et al., 2001).

Il faut souligner que, dans la modélisation de la sorption des cations par les argiles, les réactions de complexation sur les sites de bordures sont généralement omises, ce qui peut influencer les prédictions de la partition des métaux, et plus particulièrement pour les sols neutres à alcalins (Weng et al., 2001).

Les modèles de complexation spécifique ont également été utilisés pour décrire les interactions des ETM avec les argiles et plus particulièrement pour les sites de bordures à charges variables. Ces modèles se basent sur les mêmes hypothèses que les modèles développés pour la complexation avec les oxydes métalliques (CD-MUSIC, DDL) (Bradbury and Baeyens, 1997; Gu et al., 2014; Gu and Evans, 2007). Dans ces approches, l'échange d'ion n'est plus considéré comme une réaction non spécifique, mais caractérisé par une constante d'échange entre l'ion sorbé à la surface et l'élément en solution (Appelo and Postma, 2005).

De par la prise en compte d'un grand nombre de mécanismes, ces modèles semblent donc des solutions a priori robustes pour la prédiction de la distribution sol-solution à l'équilibre de métaux. Cependant, ils nécessitent de nombreuses données d'entrées. Pour l'exemple du Cd, il faut renseigner : la composition chimique de la solution (concentrations totales en éléments majeurs, anions, éléments traces), ainsi que les concentrations en AH et AF en solution ; pour les paramètres de la phase solide, les teneurs en Cd mobile ou réactif, les teneurs en argiles (idéalement minéralogiques, obtenues à partir de la composition minéralogique déterminée par diffraction des rayons X), les fractions réactives de substances humiques du sol, les teneurs en oxydes amorphes ou cristallins. Dans le cas où certaines de ces variables ne sont pas mesurées, il faut faire des hypothèses qui s'ajoutent à celles propres au modèle portant sur les paramètres génériques ajustables comme les surfaces spécifiques des oxydes métalliques amorphes ou cristallins, les constantes d'affinité du Cd pour les différentes surfaces considérées qui peuvent

varier selon le contexte... Toutes ces hypothèses constituent des sources potentielles d'erreurs de prédiction, qu'il est important de considérer.

## **7. Objectifs et démarche de la thèse**

En contexte agricole, les teneurs totales en Cd dans les horizons supérieurs sont généralement faibles (Tóth et al., 2016). Cependant, la contamination des produits de récoltes destinés à la consommation humaine doit être surveillée (Anses, 2011).

Le prélèvement du Cd par les racines des plantes se fait majoritairement sous sa forme libre :  $Cd^{2+}$  (Campbell, 1995). Ce prélèvement induit une diminution des concentrations en  $Cd^{2+}$  à l'interface entre les racines et la solution de sol. Cet abaissement est partiellement tamponné par la désorption du  $Cd^{2+}$  des phases solides, par dissociation des complexes en solution et par transport par convection et diffusion du Cd vers la racine. L'ensemble de ces processus permet de définir la fraction phytodisponible du Cd.

Un modèle semi-mécaniste a été développé par Lin et al., (2016) dans le but de prédire le prélèvement par la plante. Ce modèle, basé sur l'approche décrite par (Barber, 1995) pour les nutriments, considère les mécanismes de transport, de sorption/désorption et de complexation en solution ([SimTraces](#)). Les auteurs ont hiérarchisé les principaux paramètres influençant le prélèvement du Cd par les racines et ont identifié comme les plus importants, la concentration en  $Cd^{2+}$  initiale et la partition du Cd entre la solution de sol et le compartiment mobile du Cd du sol.

La concentration en  $Cd^{2+}$  dans la solution d'un sol conditionne non seulement la contamination des cultures, mais cette forme libre étant également biodisponible pour la composante biotique des sols, elle impacte aussi le fonctionnement biologique de ceux-ci. La mesure du  $Cd^{2+}$  dans la solution du sol pour les faibles concentrations des sols agricoles n'est pas aisée et requiert des méthodes difficilement transférables sur le plan opérationnel (méthode à base d'isotopes notamment). Prédire facilement la concentration en  $Cd^{2+}$  d'un sol donné à partir d'un nombre restreint de mesures plus simples à acquérir constitue donc un enjeu agroenvironnemental fort.

Les modèles statistiques sont à cet égard intéressants, car ils permettent de prédire le  $Cd^{2+}$  à partir de variables issues de l'analyse agronomique de terre (texture, pH, teneur en matière organique du sol, etc.). Cependant, ces modèles présentent généralement le désavantage d'avoir une variance importante : les variables explicatives et les coefficients associés varient assez largement en fonction du jeu de données utilisé pour construire les modèles. En effet, il est difficile d'estimer si le jeu de données utilisé est représentatif du contexte visé pour l'application du modèle. En outre, les modèles statistiques ne sélectionnent les variables explicatives que sur la base de la minimisation de l'erreur de prédiction et pas sur la pertinence des mécanismes géochimiques sous-jacents.

Le travail de thèse présenté ici est une contribution à l'élaboration de modèles de prédictions simples à visée opérationnelle de la concentration en  $\text{Cd}^{2+}$  dans la solution des sols agricoles pour évaluer le risque agroenvironnemental et plus particulièrement celui du transfert vers les organes végétaux consommés.

La première option, la plus simple, consiste à doser le Cd total en solution et à partir de cette mesure et de variables caractérisant la solution du sol, de prédire la concentration du  $\text{Cd}^{2+}$ . Les modèles sont alors des modèles prédictifs de la spéciation du Cd en solution. Cette option a fait l'objet du Chapitre II de ce travail.

La deuxième option consiste à prédire le  $\text{Cd}^{2+}$  en considérant l'équilibre du  $\text{Cd}^{2+}$  entre la phase solide et la solution. Les modèles prédictifs sont alors des modèles de spéciation du Cd sur la phase solide et en solution. Les variables explicatives de ces modèles concernent la fraction du Cd du sol échangeable avec la solution, et les variables du sol reliées à la spéciation du Cd en phase solide et en solution. Cette approche est plus complète que la précédente, car elle renseigne sur la capacité de la phase solide du sol à réapprovisionner le  $\text{Cd}^{2+}$  en solution lorsque ce dernier est prélevé par les racines. Cette approche est l'objet du Chapitre III.

Pour ces deux options de modélisation du  $\text{Cd}^{2+}$  en solution de sol, l'approche a été d'utiliser des modèles statistiques à visée opérationnelle, construits à partir d'un nombre réduit de variables explicatives facilement mesurables, dérivées le plus possible de l'analyse de terre. Pour asseoir la fiabilité de ces modèles statistiques, nous avons utilisé deux approches. La première, statistique, a consisté à tester la valeur prédictive des modèles par validation croisée. Le jeu de données initial est partitionné en deux sous-ensembles, l'un utilisé pour calibrer le modèle et l'autre pour tester ses prédictions et la procédure est répétée. Un des intérêts de cette méthode est de voir comment se comporte le modèle lorsqu'on induit des perturbations sur le jeu de données de calibration, ce qui permet de juger de la variance des modèles. La deuxième approche pour travailler sur la fiabilité des modèles statistiques a consisté à utiliser des modèles mécanistes géochimiques en parallèle pour modéliser la spéciation du  $\text{Cd}^{2+}$  en solution (Chapitre II) et en phase solide+solution (Chapitre III). De la sorte, il a été possible d'examiner si les variables sélectionnées par les modèles statistiques avaient un sens par rapport aux mécanismes dominants révélés par les modèles géochimiques.

Le  $\text{Cd}^{2+}$  est reconnu comme étant la principale espèce chimique du Cd absorbé par les racines. Cependant, il n'est pas direct que la concentration en  $\text{Cd}^{2+}$  estimée ponctuellement soit un bon prédicteur de l'accumulation du Cd dans l'organe végétal consommé, car cette concentration peut fluctuer au cours de la culture sous l'action des conditions de l'environnement incluant notamment les pratiques agricoles, la température, les précipitations, l'action de la plante. En outre, la physiologie de la plante joue également un rôle a priori important pour déterminer les teneurs en

Cd des tissus végétaux. Le Chapitre IV traite de cette question de la valeur d'un modèle statistique utilisant uniquement des variables du sol pour prédire la teneur en Cd d'un produit végétal, ici le grain de blé dur. Outre l'examen de la précision par rapport aux exigences des seuils réglementaires de ces modèles fondés sur une caractérisation ponctuelle du sol, nous avons examiné si les variables de ces modèles étaient bien cohérentes avec les variables prédictives du Cd<sup>2+</sup> identifiées dans les chapitres II et III.



## **Chapitre II. Modélisation de la spéciation du Cd en solution**

---

### **1. Contexte**

Dans le cadre de l'étude du transfert sol-plante, la disponibilité du Cd peut être définie comme le niveau d'exposition du système racinaire à cet élément, c'est-à-dire la quantité de Cd présente dans le sol potentiellement absorbable par les racines et résultant de processus physico-chimiques de désorption. La théorie de l'ion libre et ses améliorations (FIAM modifié, Biotic Ligand Model, Morel et al., 1993; Pagenkopf, 1983) supposent que la forme libre du métal ( $\text{Cd}^{2+}$ ) est la forme chimique majoritairement absorbée par les plantes dans la solution de sol et que les complexes labiles en solution jouent un rôle dans le réapprovisionnement en  $\text{Cd}^{2+}$  à proximité de la racine. De plus, Lin et al. (2016) ont montré, en sols agricoles non contaminés, que la concentration en  $\text{Cd}^{2+}$  initiale était le facteur le plus influent pour le prélèvement du Cd par la plante avec les propriétés de diffusion du sol (humidité, tortuosité) et le pouvoir tampon du sol vis-à-vis du Cd. **Par conséquent, l'étude de la spéciation du Cd en solution apparait comme primordiale pour la compréhension et la modélisation du transfert du Cd du sol vers la plante et accentue la nécessité de mesurer et de prédire le plus justement possible la concentration en Cd libre en solution.**

Dans la solution de sol, le Cd va former des complexes avec une grande variété de ligands organiques et inorganiques (Sauvé et al., 2000b). La formation de ces complexes en solution limite souvent l'exposition des plantes au Cd en diminuant la concentration en  $\text{Cd}^{2+}$ . La matière organique dissoute (MOD) joue un rôle essentiel dans la spéciation du Cd (Cornu et al., 2011; Sauvé et al., 2000b; Weng et al., 2002a). En effet, la MOD est généralement considérée comme le ligand majoritaire contrôlant la spéciation du Cd en solution de sol.

La caractérisation expérimentale de la MOD est basée sur les solubilités à différents pH des espèces organiques qui la composent. Elle se compose d'acides humiques (AH) et d'acides fulviques (AF), constituant les substances humiques (SH), d'une fraction hydrophile comprenant grossièrement des acides organiques de faibles poids moléculaires (LMWOA), des sucres, des acides aminés et d'une dernière fraction constituée de matière organique hydrophobe (MOH) (van Zomeren and Comans, 2007). En plus de leurs solubilités, les SH diffèrent également par



leur composition chimique. Ainsi, les AF ont une masse moléculaire plus faible que celle des AH, mais une abondance de groupes fonctionnels (principalement des sites carboxyliques, CO-OH, et phénoliques,  $\Phi$ -OH) plus importante (Dudal and Gérard, 2004). Ce sont ces groupements fonctionnels qui vont conditionner les propriétés acides et complexantes des SH et donc leur réactivité vis-à-vis des métaux.

Une multitude de facteurs peuvent influencer la MOD, tant la quantité retrouvée en solution que sa composition. Dans le cas de sols agricoles, ces principaux facteurs sont leurs propriétés pédogéochimiques, les espèces végétales cultivées et les pratiques agricoles employées (Chantigny, 2003; Raber et al., 1998). Le protocole expérimental utilisé pour la récupération de la solution de sol peut également influencer la MOD retrouvée en solution (Fest et al., 2008; Merckx et al., 2001; Raber et al., 1998). Il existe, dans la littérature, de nombreux protocoles de récupération de la solution de sol (Di Bonito et al., 2008), différant par l'état initial du sol (sec ou humide), par la nature de la solution de contact (eau ultra pure, solutions salines à différentes concentrations) (Degryse et al., 2009), du rapport du volume de solution sur la masse de sol utilisée (Yin et al., 2002) et de la méthode employée pour récupérer cette solution (dépression, centrifugation) (Bufflap and Allen, 1995; Coutelot et al., 2014). Ces différentes méthodes aboutissent à des solutions de sol de compositions possiblement différentes.

Il existe également différentes méthodes de mesure expérimentale de la fraction libre du Cd s'appuyant sur des principes différents (Schneider, 2006; Temminghoff et al., 2000; Van Leeuwen et al., 2005; Weng et al., 2002b; Xue and Sigg, 2002). Cependant, ces méthodes sont difficilement transposables en routine professionnelle et ne permettent pas la mesure rapide de cette variable d'importance, d'où l'intérêt de pouvoir prédire les concentrations en  $\text{Cd}^{2+}$  par modélisation à partir de mesures facilement accessibles.

La modélisation de la spéciation du Cd en solution par informatique a été rendue plus facile grâce au développement de modèles de calculs thermodynamiques. En effet, la complexité de la solution de sol, par la variété d'espèces la constituant et leurs interactions, nécessite des calculs nombreux. Le cas des complexes inorganiques est assez facilement traitable, car leur nature est accessible par analyses chimiques et leurs constantes thermodynamiques de stabilité sont disponibles dans les bases de données (Nordstrom et al., 1990; Smith and Martell, 1976).

Les interactions des métaux avec les SH sont plus difficilement modélisables à cause de la diversité des SH, c'est-à-dire la diversité des groupements fonctionnels qui composent leurs structures et conditionnent leur réactivité. Pour modéliser les interactions électrostatiques et spécifiques entre les ions et les SH, les modèles NICA-Donnan (Benedetti et al., 1996) et HIB ou WHAM (Tipping and Hurley, 1992) font partie des plus fiables et éprouvés dans la littérature (Guthrie et al., 2005; Kinniburgh et al., 1996; Lofts and Tipping, 2011; Ponthieu et al., 2016; Weng et al., 2002a).

Groenenberg et al. (2010a) ont identifié que pour des composés ayant une relative faible affinité avec la MOD, comme le Cd, la composition de cette dernière constitue l'un des paramètres les plus importants pour modéliser correctement leur spéciation. Ainsi, la quantité de carbone compris dans la MOD, la quantité active de MOD et les proportions des substances humiques et non humiques sont les points critiques à renseigner lorsque les paramètres génériques relatifs aux propriétés intrinsèques des MOD sont utilisés. L'environnement chimique joue également un rôle important dans le calcul de la spéciation des espèces en solution puisque la présence de cations compétiteurs ou de ligands en solution influence la constante de complexation conditionnelle d'un métal avec un ligand donné (Ringbom, 1963).

Il est admis que les MOD sont en moyennes constituées de 50 % de carbone (Dwane and Tipping, 1998), aboutissant à des concentrations en MOD égales à 2 fois celles du carbone organique dissous (COD). Le rapport MOD/COD actif peut cependant varier pour prendre en compte une part de carbone inerte dans les mesures de COD (Sjöstedt et al., 2010). La variation du rapport MOD/COD actif ou directement du pourcentage de COD actif aboutit à une gamme de variation importante du COD actif, entre 7 % et 100 % de la MOD (Bryan et al., 2002; Cornu et al., 2009; Ren et al., 2015b; Schneider et al., 2016b; Unsworth et al., 2006; Weng et al., 2002a). Plus récemment, les études ayant caractérisé la MOD ont souligné la part importante de la fraction hydrophile (Groenenberg et al., 2010a; Ren et al., 2015b; Schneider et al., 2016b). Cette fraction est constituée de composés non humiques, c'est-à-dire un mélange de petits acides organiques de faibles poids moléculaires, d'acides aminés, de sucres et de glucides (Merckx et al., 2001) jouant également un rôle dans la complexation des métaux en solution, en particulier les LMWOA définis comme un mélange de mono-, di- et tri- acides (*p. ex.* acides acétique, oxalique, citrique, malique...). Généralement la fraction de LMWOA représente 10 % de la concentration en COD mesurée en solution et provient majoritairement de trois différentes sources biotiques : les métabolites issus de la décomposition de la matière organique du sol (MOS), l'exsudation par les racines des plantes et les espèces mycorhizogènes du sol (Strobel, 2001).

La modélisation de la spéciation du Cd s'appuie donc sur un nombre important de paramètres d'entrée dont les plus difficiles à caractériser sont liés à la nature et à la réactivité de la MOD.

Ainsi l'objectif de ce travail a été de modéliser la fraction de Cd libre en solution de sols agricoles à partir d'un nombre minimal de mesures facilement réalisables afin de trouver le meilleur compromis entre facilité de mesures et justesse et robustesse de prédiction. **Dans un premier temps**, nous avons considéré un modèle statistique obtenu par régression multilinéaire de la fraction libre en fonction de la composition de la solution de sol en tant que régresseurs candidats. **Dans un second temps**, nous avons fait l'hypothèse d'un ligand moyen unique issu de la concentration en COD. Cette hypothèse sera traitée en supposant une réactivité unique, puis une

réactivité pouvant varier en fonction de caractéristiques du sol et de la solution du sol à identifier. **La troisième approche** a été basée sur la modélisation géochimique thermodynamique de la spéciation du Cd. Elle a été basée sur la complexation des cations en solution par une MOD constituée d'AH, d'AF et de LMWOA, dont la réactivité (nombre de sites et affinité) serait déduite d'une analyse simple du sol ou de la solution de sol.

Ces approches seront testées sur une base de données de sols agricoles obtenus au cours de différents projets de recherche.

## 2. Matériels et méthodes

### 2.1. Sols et caractérisation de la solution de sol

76 sols ont été utilisés pour modéliser la spéciation du Cd en solution de sol. Ces sols ont été sélectionnés en fonction de leurs caractéristiques pédogéochimiques afin de couvrir une large gamme de types de sols. Les échantillons proviennent de sols agricoles répartis dans toute la France (Figure II.1) obtenus au cours de différents projets de recherche : 23 sols prélevés en 1996 (ASPITET, Baize, 1997) ; 29 sols prélevés en 2011, 2013, 2015 (QUASAPROVE) ; 5 sols prélevés en 2016 (QUALIAGRO) ; 19 sols prélevés par Arvalis Institut du Végétal en 2016 (CaDON).

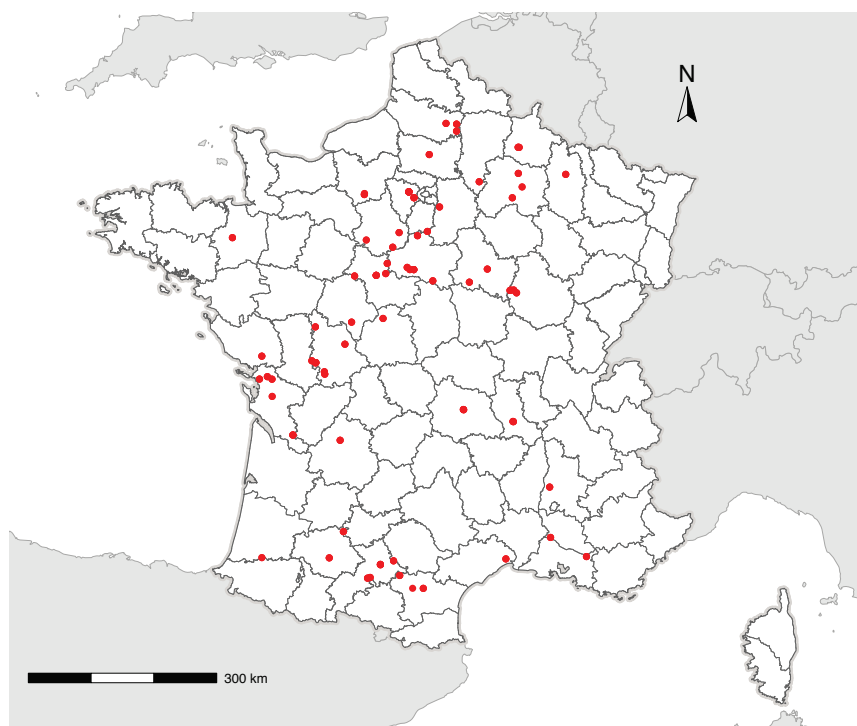


Figure II.1 : Localisation des sols utilisés pour modéliser la spéciation du Cd en solution de sol.

Les échantillons de sols ont été séchés à l'air puis tamisés à 2 mm. La teneur en eau résiduelle a été déterminée après séchage à 105 °C jusqu'à une masse constante (NF ISO 11465:1994). Les propriétés suivantes ont été mesurées sur les échantillons de sols séchés à l'air en accord avec les normes françaises. Le pH du sol a été mesuré dans l'eau distillée (ratio sol:solution de 1:5 en g:ml, NF ISO 10390:2005) et dans un extrait à 0.01 M CaCl<sub>2</sub> (ratio sol:solution de 1:10 en g:ml, Houba et al., 1990a). Les teneurs totales en CaCO<sub>3</sub> ont été quantifiées par la méthode de neutralisation acide (NF X 31-105). La fraction de carbone organique du sol (COS) a été mesurée par combustion sèche après correction du CaCO<sub>3</sub> (NF ISO 10694:1995). La texture du sol a été déterminée par la méthode de la pipette dite de Robinson (NF X 31-107) et la capacité d'échange cationique a été estimée par extraction au chlorure de cobaltihexamine au pH du sol (NF X 31-130). La teneur totale en Cd a été quantifiée après solubilisation par un mélange d'acide fluorhydrique et perchlorique (NF X 31-147). Toutes ces analyses ont été effectuées par le Laboratoire d'Analyses des Sols d'Arras (<http://www6.lille.inra.fr/las>).

**Afin de récupérer la solution de sol**, environ 25 g de sol séché à l'air ont été réhumidifiés à pH 1.7. La quantité d'eau ajoutée au sol sec est calculée à partir des équations proposées par Jong et al., (1983). Ces sols sont ensuite laissés à équilibrer pendant 24 h à une température de 20 °C. La solution de sol a été récupérée par déplacement grâce à l'ajout d'environ 10 g de solvant non miscible à l'eau (C<sub>2</sub>F<sub>3</sub>Cl<sub>3</sub>) avant d'être immédiatement centrifugée à 26500 g pendant une heure. La solution de sol a ensuite été filtrée à 0.2 µm (le protocole de la récupération de sol est renseigné en Annexe 1). Trois réplicats ont été effectués pour chaque sol.

Les propriétés chimiques suivantes ont été caractérisées sur les solutions de sol : pH, concentrations totales en Ca et Mg (notées Ca<sub>ss</sub> et Mg<sub>ss</sub>) par absorption atomique en flamme, concentrations totales en Cd et Cu (notées Cd<sub>ss</sub> et Cu<sub>ss</sub>) déterminées par spectrométrie d'absorption atomique en four et, fraction libre de Cd (f<sub>Cd</sub>) estimée via une méthode de résine échangeuse d'ions (méthode décrite Chapitre 1.5.2 et protocole fourni en Annexe 2). La composition anionique (Cl<sup>-</sup>, NO<sub>3</sub><sup>-</sup>, SO<sub>4</sub><sup>2-</sup>) a été mesurée par chromatographie ionique et les concentrations en carbone organique dissout (COD) et carbone inorganique dissout (CID) par combustion oxydative catalytique.

Pour chaque sol, la composition de la solution de sol variait légèrement entre les réplicats. Ainsi, chaque réplicat a été considéré comme un individu pour la modélisation de la spéciation en solution du Cd.

## ***2.2. Modélisation statistique de la fraction libre***

Différentes approches de modélisation ont été développées et sont décrites par la Figure II.3.

### 2.2.1. Modélisation des concentrations en Cd<sup>2+</sup> par régression multilinéaire

Les concentrations en Cd<sup>2+</sup> ont été régressées par régression multilinéaire en fonction de 1,2,3 ou 4 variables de la solution de sol choisie parmi les groupes suivants : (1) la concentration en Cd complexé (Cd<sub>CdL</sub>), calculée comme la différence entre la concentration totale de Cd en solution et la concentration en Cd<sup>2+</sup>, (2) le pH de la solution de sol, (3) les concentrations en Ca, Mg et Cu agissant comme des compétiteurs en solution vis-à-vis du Cd et (4) les concentrations en ligands, c'est-à-dire le COD et les ions carbonates, nitrates, sulfates et chlorures. Les modèles testés peuvent être listés comme suit :

$$\log_{10}(Cd^{2+}) = f(\log_{10}(CdL) + \left\{ \begin{array}{c} \emptyset \\ pH_{ss} \end{array} \right\} + \log_{10} \left\{ \begin{array}{c} \emptyset \\ Ca_{ss} \\ Mg_{ss} \\ Cu_{ss} \end{array} \right\} + \log_{10} \left\{ \begin{array}{c} \emptyset \\ COD \\ CO_{3ss} \\ NO_{3ss} \\ SO_{4ss} \\ Cl_{ss} \end{array} \right\})$$

La modélisation « log-log » permet de linéariser un modèle « multiplicatif » où les variables explicatives interviennent par leur produit et par leur puissance. Le produit permet de rendre compte d'interactions entre variables tandis que la puissance modélise la non-linéarité possible de l'effet d'une variable.

Afin d'éviter le surparamétrage, les modèles ont été limités à quatre régresseurs. Le coefficient de détermination de la régression linéaire multiple (noté r<sup>2</sup><sub>aj</sub>), le critère d'information bayésien (BIC) et le facteur d'inflation de la variance (VIF) ont été calculés afin de sélectionner la meilleure combinaison de régresseurs expliquant la concentration en Cd<sup>2+</sup> en solution.

### 2.2.2. Modélisation de la spéciation du Cd par un ligand unique dérivé de la MOD

#### 2.2.2.a. Modèle conceptuel de la réactivité du ligand moyen

Le Cd peut former de nombreux complexes en solution avec des ligands inorganiques et organiques. La principale hypothèse d'une partie de ce travail est que la modélisation de la spéciation du Cd pour l'étude du transfert sol-plante peut se simplifier en ne considérant qu'un seul complexe issu de la MOD (CdL) capable d'approximer les nombreux complexes en solution. Ainsi, nous considérons présents en solution le métal libre, Cd<sup>2+</sup> et CdL, résultant de l'association du ligand libre (de concentration L) avec Cd<sup>2+</sup> (Équation II.1)



Dont la constante thermodynamique d'association s'écrit :

$$K_{CdL} = \frac{[CdL]}{[Cd^{2+}] \times [L^{2-}]} \quad \text{Équation II.2}$$

En considérant également l'environnement chimique du Cd et donc la présence de cations compétiteurs en solution (hypothèse d'un ligand unique), la constante conditionnelle de complexation de cette réaction peut s'écrire :

$$K'_{CdL} = \frac{[CdL]}{[Cd^{2+}] \times [L]'} \quad \text{Équation II.3}$$

Avec  $[L]'$ , la concentration en ligands non liés au Cd s'écrivant :

$$[L]' = [L_{tot}] - [CdL] = [L^{2-}] + \sum [X_i L] - [CdL] \quad \text{Équation II.4}$$

avec  $X_i$  les compétiteurs du Cd pour la complexation avec le ligand  $L^{2-}$ .

La fraction libre de Cd, notée  $f_{Cd}$ , est donnée par l'équation :

$$f_{Cd} = \frac{[Cd^{2+}]}{[Cd_{ss}]} \quad \text{Équation II.5}$$

avec  $[Cd_{ss}]$  correspondant à la concentration totale de Cd en solution égale à :

$$[Cd_{ss}] = [Cd^{2+}] + [CdL] \quad \text{Équation II.6}$$

Alors :

$$f_{Cd} = \frac{[Cd^{2+}]}{[Cd^{2+}] + [CdL]} \quad \text{Équation II.7}$$

Ou également :

$$f_{Cd} = \frac{1}{1 + K'_{CdL} \times [L]'} \quad \text{Équation II.8}$$

Si l'on considère que le COD est le ligand majeur influençant la spéciation en solution du Cd alors :

$$[L]' = n \times [COD] \quad \text{Équation II.9}$$

Avec  $n$  correspondant au nombre de paires de sites par quantité de COD pouvant lier le Cd et  $[COD]$  la concentration en carbone organique dissous.

L'Équation II.8 devient donc :

$$K'_{CdL} \times n = \frac{1/f_{Cd} - 1}{[COD]} \quad \text{Équation II.10}$$

Le produit  $K'_{CdL} \cdot n$ , du nombre de paires de sites du ligand par la constante d'affinité conditionnelle du Cd avec le ligand peut être assimilé à la réactivité du COD.

### 2.2.2.b. Modélisation des concentrations en $Cd^{2+}$ en fonction d'un ligand unique de réactivité moyenne

La fraction libre  $f_{Cd}$  et la concentration en carbone organique dissous déterminées expérimentalement ont permis de calculer les valeurs expérimentales du produit  $K'_{CdL.n}$ , simplifié en  $K'n$ , et d'étudier sa variabilité.

Nous avons ensuite testé la valeur d'un modèle prédictif à partir d'un ligand unique dérivé du COD avec une réactivité unique (Figure II.3). Pour cela, la valeur médiane de  $K'n$  a été utilisée pour simuler la fraction libre de Cd pour l'ensemble des sols et pour calculer les erreurs d'estimations des concentrations en  $Cd^{2+}$ .

### 2.2.2.c. Modélisation des concentrations en $Cd^{2+}$ en fonction d'un ligand unique dérivé de la MOD de réactivité variable suivant l'environnement chimique

Dans ce cas, la réactivité du ligand unique,  $K'n$ , est autorisée à varier selon certaines propriétés de la solution de sol afin de rendre compte le fait que certaines de ces propriétés (pH, concentrations en Ca, etc.) peuvent modifier la compétition pour la complexation du Cd, la nature de la MOD et son abondance.

Le produit  $K'n$ , a donc été régressé par régression multilinéaire en fonction de paramètres de la solution de sol (Figure II.3). La liste condensée des modèles testés est la suivante :

$$\log_{10}(K'n) = f(pH_{ss} + \log_{10} \left\{ \begin{array}{c} \emptyset \\ Ca_{ss} \\ Mg_{ss} \\ Cu_{ss} \end{array} \right\} + \log_{10} \left\{ \begin{array}{c} \emptyset \\ CO_{3ss} \\ NO_{3ss} \\ SO_{4ss} \\ Cl_{ss} \end{array} \right\})$$

Afin d'éviter le surparamétrage, les modèles ont été limités à un maximum de 3 régresseurs au maximum. Le coefficient de détermination de la régression linéaire multiple, le critère d'information bayésien (BIC) et le facteur d'inflation de la variance (VIF) ont été calculés afin de sélectionner la meilleure combinaison de régresseurs expliquant  $K'n$ .

### 2.2.3. Comparaison des approches statistiques pour la prédiction des concentrations de $Cd^{2+}$ en solution

La valeur prédictive de chaque modèle a été testée par validation croisée selon la méthode *k-folds*, c'est-à-dire que 10 sous jeux de données de taille équivalente ont été aléatoirement formés, et chacun des modèles a été paramétré sur 9 sous jeux de données et testé sur le groupe restant.

Cette action est répétée pour que chacun des sous-jeux de données serve de test, et l'ensemble est reproduit 100 fois pour tester un nombre maximal de sous jeux de données. L'adéquation entre valeurs observées et valeurs prédites log-transformées a été évaluée par régression linéaire.

Différents critères ont permis l'évaluation de la qualité de prédiction des concentrations en  $\text{Cd}^{2+}$ .

- Le coefficient de détermination de la régression linéaire log-log entre valeurs observées et valeurs prédites, noté  $r^2_{\text{obs/pred}}$ . Ce coefficient permet d'évaluer la dispersion des erreurs de prédictions par rapport à la droite de régression.

- La pente et l'ordonnée à l'origine (Ord. Orig.) des régressions linéaires log-log entre valeurs observées et valeurs prédites ont également permis la détection d'un possible biais dans la prédiction des concentrations en  $\text{Cd}^{2+}$ .

Une fois le biais potentiel corrigé, les erreurs relatives définies par le rapport de la différence en valeur absolue entre valeur observée et valeur prédite sur la valeur observée (Équation II.11) sont analysées et permettent la comparaison des approches de prédiction des concentrations en  $\text{Cd}^{2+}$ .

$$\text{Erreur relative} = \frac{|[\text{Cd}^{2+}]_{\text{observée}} - [\text{Cd}^{2+}]_{\text{prédite}}|}{[\text{Cd}^{2+}]_{\text{observée}}} \quad \text{Équation II.11}$$

## 2.3. Modélisation géochimique par CHEAQS Next<sup>®</sup>

### 2.3.1. Description du modèle CHEAQS Next<sup>®</sup>

CHEAQS Next<sup>®</sup> (Release P2017.2, <http://www.cheaqs.eu/>) est un modèle de spéciation géochimique permettant le calcul des concentrations des différentes espèces chimiques présentes dans une solution à partir des concentrations totales en éléments mesurés en solution.

Les calculs de spéciation ont été effectués par le modèle CHEAQS Next<sup>®</sup> avec comme variables d'entrée : le pH de la solution de sol ( $\text{pH}_{\text{ss}}$ ), les concentrations totales de Ca, Mg, Cu et Cd en solution, les concentrations en anions inorganiques, notées  $\text{Cl}_{\text{ss}}$ ,  $\text{NO}_{3\text{ss}}$ ,  $\text{CO}_{3\text{ss}}$ ,  $\text{SO}_{4\text{ss}}$ , et la concentration en COD.

Dans CHEAQS Next<sup>®</sup>, la complexation organique est calculée par le modèle WHAM VII (Tipping et al., 2011), considérant qu'un cation est complexé à un AH ou un AF. Ce modèle prend également en compte la physisorption du cation et du proton, correspondant aux interactions d'attraction-répulsion entre les ions et la macromolécule organique chargée. Les constantes de complexation du H, Ca, Mg, Cu et Cd avec les ligands oxalate et citrate (NIST 46.6, 2001) ont été ajoutées à la base de données de CHEAQS Next<sup>®</sup>.



### 2.3.2. Modélisation géochimique de Cd<sup>2+</sup> par une MOD de réactivité constante pour tous les sols

Dans ces scénarios nous faisons l'hypothèse que la réactivité de la MOD est indépendante du type de sol et nous cherchons quelle réactivité générique explique le mieux les concentrations en Cd<sup>2+</sup> observées. Deux scénarios ont été testés afin de modéliser la spéciation du Cd en solution en tenant compte de la nature et de l'abondance des ligands qui conditionnent la réactivité de la MOD (Figure II.2).

Le premier (Scénario 1) cherche à tester l'influence des AH sur la prédiction de Cd<sup>2+</sup>. Il teste les compositions de MOD fréquemment reportées dans la littérature, c'est-à-dire une répartition 50/50 entre AH et AF (Weng et al., 2002a) ou une MOD uniquement composée d'AF (Unsworth et al., 2006). La valeur maximale pour le pourcentage de COD actif utilisée est celle généralement appliquée dans la bibliographie, il est de 50 % (Ponthieu et al., 2016; Unsworth et al., 2006). Cette valeur a été choisie pour tester l'influence des AH sur la spéciation Cd.

Le second scénario (Scénario 2) correspond à un factoriel équilibré faisant varier la quantité de COD actif de 10 à 50 % et la composition de la MOD séparée en substances humiques (AF uniquement) et substances non humiques (LMWOA) (Figure II.2). La fraction non humique est assimilée à une mixture de trois acides de faibles poids moléculaires, mono-, di- et tri- acides (acides acétique, oxalique, citrique) dont les proportions individuelles varient de 0 à 100 % (Figure II.2). Au total, 3305 combinaisons sont testées pour chaque solution de sol.

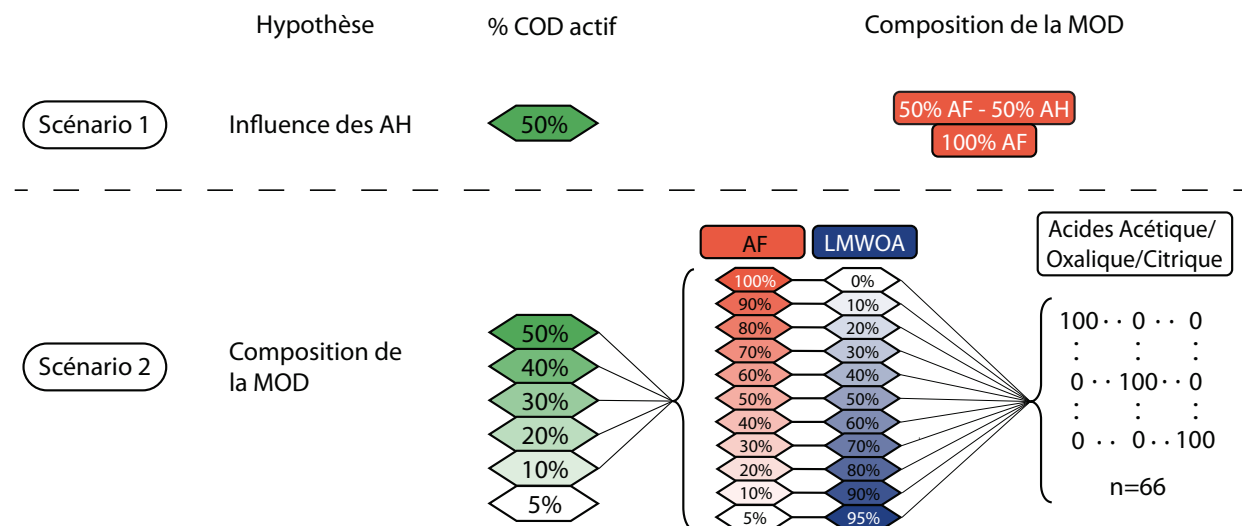


Figure II.2 : Récapitulatif des scénarios testés pour la modélisation de la spéciation de Cd par CHEAQS Next®

Pour chacune de ces combinaisons, les valeurs de Cd<sup>2+</sup> prédites ont été comparées par régressions linéaires par rapport aux concentrations mesurées. Le coefficient de détermination,

le biais, la pente de ces régressions et le calcul des erreurs relatives corrigées du biais de prédiction ont permis la sélection et la comparaison de ces modèles.

### 2.3.3. Modélisation géochimique de Cd<sup>2+</sup> en reliant la MOD aux propriétés du sol

Dans cette partie, la réactivité de la MOD a été autorisée à varier en fonction de propriétés du sol que nous avons cherché à identifier. Ainsi, parmi toutes les hypothèses testées dans le scénario 2, on sélectionne **pour chaque sol** les 5 % de combinaisons de composition de la MOD produisant les erreurs de prédictions de Cd<sup>2+</sup> les plus faibles. Parmi celles-ci, il est possible que des combinaisons % COD actif x % AF soient redondants en termes de nombre de sites phénoliques et carboxyliques. Les modalités (% COD actif et % AF) ont donc été traduites en nombre de sites (en mmol.L<sup>-1</sup>) en utilisant les concentrations en COD mesurées et les abondances de sites génériques utilisées par WHAM (Équation II.12 ; avec pour les sites carboxyliques : a<sub>COOH</sub> = 10.7 mmol.g<sup>-1</sup> C et pour les sites phénoliques : a<sub>φ-OH</sub> = 5.3 mmol.g<sup>-1</sup> C).

$$Sites = [COD] \times \%COD_{actif} \times \%AF \times (a_{CO-OH} + a_{\phi-OH}) \quad \text{Équation II.12}$$

Afin de relier la réactivité de la MOD aux propriétés du sol nous avons modélisé le nombre de sites en fonction des variables usuelles caractérisant les sols et par les mesures effectuées dans la solution de sol (v. partie 2.1) par forêt d'arbres de régression, ou *Random Forest* (Breiman, 2001). Un arbre de régression se décompose en nœuds, représentant la variable explicative testée, en branches, dont la hauteur représente l'importance de la variable testée à réduire la variance de la variable expliquée et se termine par des feuilles donnant les valeurs de la variable expliquée. Cette méthode de modélisation est ici basée sur l'établissement de 100 arbres de régression sur une fraction aléatoire du jeu de données et la prédiction sur la fraction restante. La construction de chaque nœud de l'arbre se fait donc sur un sous ensemble du jeu de données, mais aussi sur un sous-ensemble de variables explicatives tirées aléatoirement (cinq variables testées à chaque nœud). Cela permet de déduire l'importance des variables explicatives et de finalement prédire le nombre de sites, en moyennant l'ensemble des prédictions établies par les différents arbres.

Les variables explicatives du sol testées sont les pH du sol mesurés dans l'eau (rapport 1:5 m:v) et dans un extrait de CaCl<sub>2</sub> à 0.01 M (rapport 1:10 m:v), le carbone organique du sol, le pourcentage d'argiles granulométriques, la teneur en calcaire, la capacité d'échange cationique et les paramètres mesurés en solution de sol.

Afin de prédire les concentrations en Cd<sup>2+</sup> par modélisation géochimique (CHEAQS Next<sup>®</sup>), il faut renseigner les concentrations en COD, le pourcentage de COD actif et le pourcentage d'AF. À

partir du nombre de sites prédits par *Random Forest*, l'Équation II.13 permet de calculer le produit « %COD<sub>actif</sub> x %AF ».

$$\%COD_{actif} \times \%AF = \frac{Sites_{prédits}}{[COD] \times (a_{CO-OH} + a_{\Phi-OH})} \quad \text{Équation II.13}$$

Les pourcentages de COD actif et d'AF doivent être renseignés individuellement dans la modélisation géochimique, mais nous ne disposons que de la valeur du produit. Le pourcentage d'AF est alors assumé égal au pourcentage de COD actif et donc égal à la racine carrée du produit (Équation II.13). Les concentrations en Cd<sup>2+</sup> ont été calculées, via CHEAQS Next®, à partir de ces nouvelles hypothèses de composition de MOD et en assimilant la fraction non humique de la MOD restante à de l'acide acétique.

La totalité de la démarche de modélisation effectuée est résumée sur la Figure II.3. Dans son ensemble, elle consiste donc à modéliser les concentrations en Cd<sup>2+</sup> par 5 approches différentes, statistiques et mécanistiques, en faisant varier ou non la réactivité du ligand considéré comme le plus influent : la MOD.

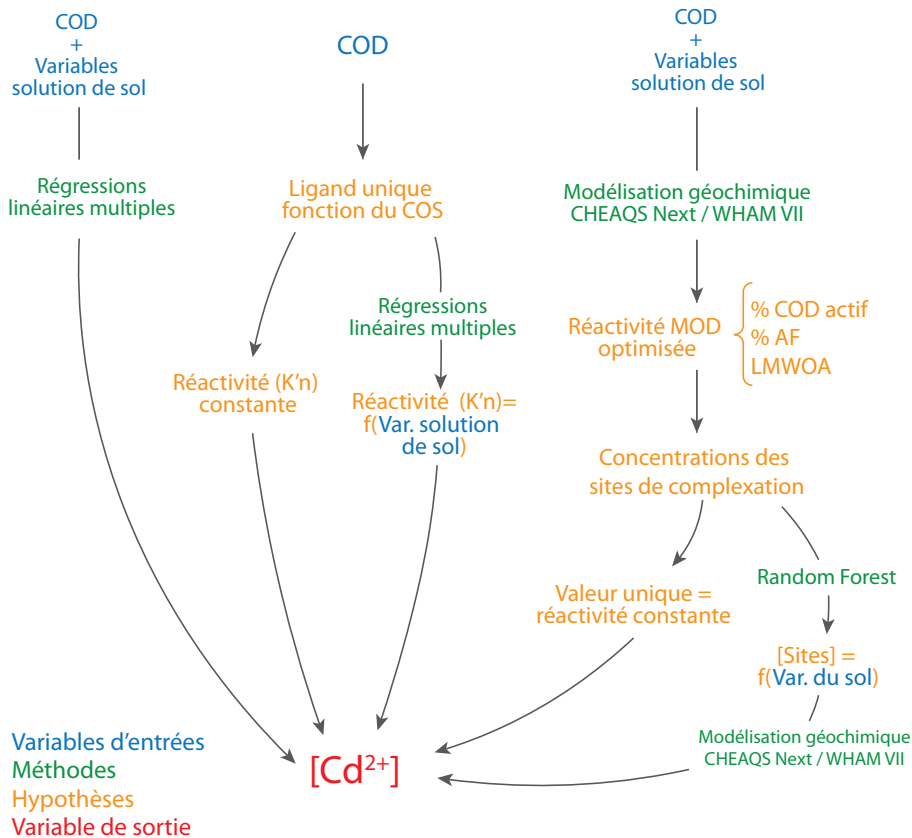


Figure II.3 : Schéma récapitulatif de la démarche de modélisation des concentrations en  $\text{Cd}^{2+}$  résumant les différentes approches utilisées.

### 3. Résultats

#### 3.1. Propriétés des sols et des solutions de sol

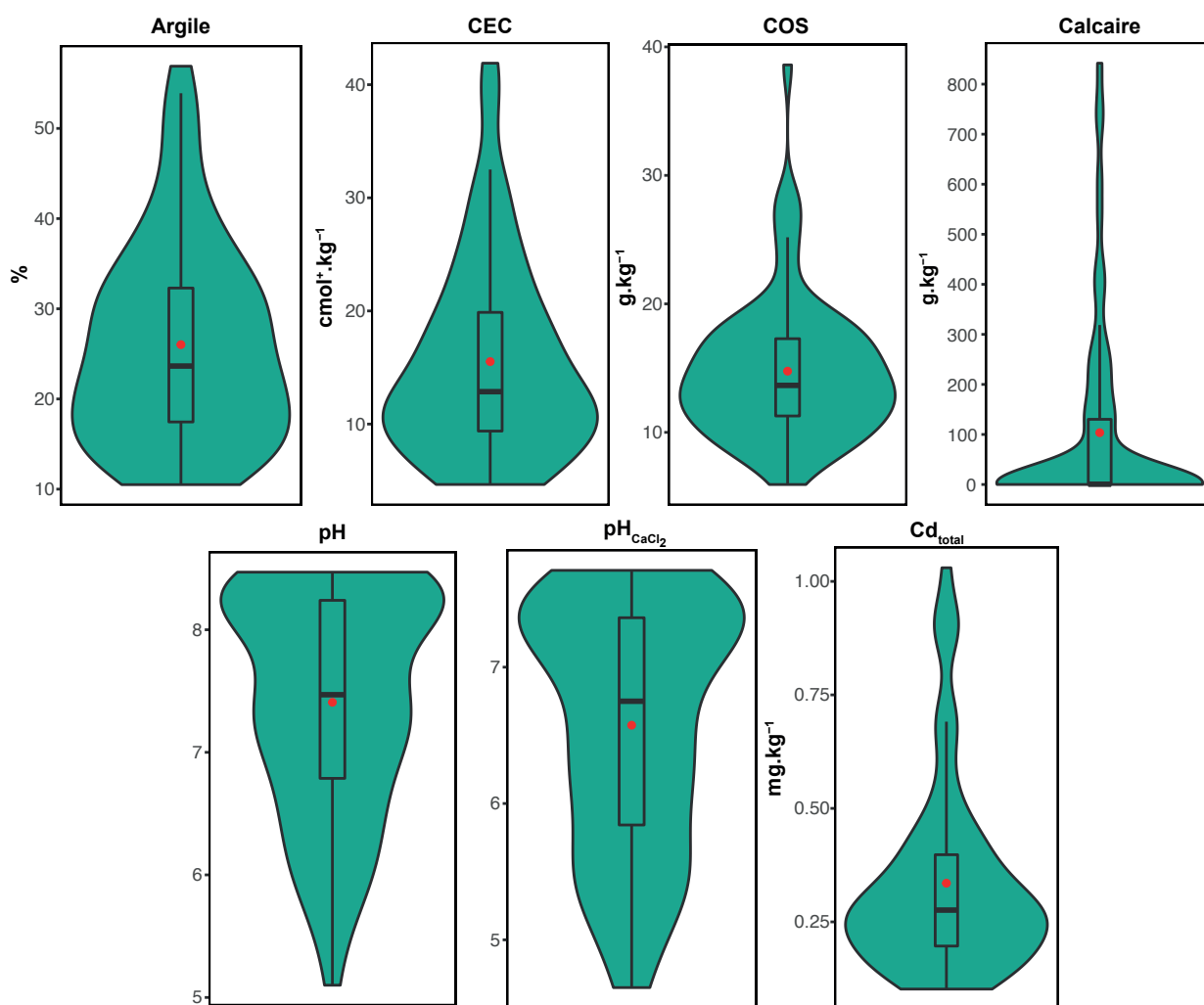


Figure II.4 : Violin plots des caractéristiques pédogéochimiques des sols utilisés dans ce chapitre. Le point rouge correspond à la moyenne des individus, le box plot représente l'interquartile (25-75 %) et les segments le prolongeant correspondent à 1.5 fois l'interquartile.

La granulométrie des sols étudiés couvre la majorité des textures rencontrées en sols agricoles, avec plus de 90 % des sols représentés entre les classes texturales argileuses et limoneuses (Figure II.4). Les sols sont majoritairement alcalins avec des pH allant de 5.1 à 8.5 et une médiane à 7.5 (Figure II.4). Les pH mesurés dans un extrait de  $\text{CaCl}_2$  à 0.01 M ( $\text{pH}_{\text{CaCl}_2}$ ) sont globalement plus faibles avec des valeurs se trouvant entre 4.7 et 7.7 et une médiane de 6.8 (Figure II.4). Ces gammes de variation couvrent également la gamme de pH généralement

rencontrés en sols agricoles. Cependant, avec seulement 20 % de sols ayant des pH inférieurs à 6.5, l'extrapolation des modèles de prédiction de  $\text{Cd}^{2+}$  pour des pH de sol acides devra être faite avec précaution. Les teneurs totales en Cd sont comprises entre 0.10 et 1.03  $\text{mg.kg}^{-1}$  ce qui est conforme avec les données reportées par le RMQS (Réseau de Mesure de la Qualité des Sols) dont 95 % des valeurs sont inférieures à 0.9  $\text{mg.kg}^{-1}$  (Gis Sol, 2011) (Figure II.4). Les sols sélectionnés ne montrent donc pas de contamination excessive en Cd. Les distributions et fréquences des autres caractéristiques de sols (CEC, teneurs en calcaire, carbone organique) sont présentées Figure II.4 et sont dans la gamme régulièrement retrouvée pour des sols agricoles (de Vries et al., 2011; Gis Sol, 2011; Ivezić et al., 2012).

Les distributions des caractéristiques des solutions de sol obtenues par centrifugation sont représentées Figure II.5. Les  $\text{pH}_{\text{ss}}$  s'échelonnent de 4.8 à 8.9 avec une médiane de 8.0. La concentration médiane de Ca en solution ( $\text{Ca}_{\text{ss}}$ ) (3.1 mM) est dans la gamme généralement retrouvée en sols agricoles avec un maximum à 26.9 mM pour le sol RETARB qui est un sol très calcaire avec une teneur totale en  $\text{CaCO}_3$  égale à 651  $\text{g.kg}^{-1}$  (Annexe IV.2). Après le Ca, le magnésium est le cation le plus concentré. Ces concentrations ( $\text{Mg}_{\text{ss}}$ ) vont de 0.06 à 1.44 mM. Le dernier cation étudié est le cuivre ( $\text{Cu}_{\text{ss}}$ ), car il montre une forte affinité avec les matières organiques (Evans, 1989) et peut donc jouer un rôle important de compétiteur pour les sites de complexation de la MOD. Sa concentration totale en solution varie entre 0.19 et 2.29  $\mu\text{M}$ . Les concentrations en carbone organique (COD) et inorganique ( $\text{CO}_{3\text{ss}}$ ) mesurées en solution vont de 59.0 à 952.4  $\text{mg C.l}^{-1}$  et 1.3 à 112.7  $\text{mg CO}_3^{2-}.\text{l}^{-1}$  respectivement. Les concentrations en COD et le pH de la solution de sol sont corrélés négativement ( $r = -0.33$ ,  $p\text{-value} = 9.114\text{e-}07$ , test de corrélation Bravais-Pearson), ce qui se traduit par des concentrations en COD plus faibles pour les sols ayant un pH élevé à l'exception de certains sols alcalins ayant des teneurs en COS importantes. Les concentrations en anions présents en solution s'étalent également sur de larges gammes de valeurs allant pour les nitrates ( $\text{NO}_{3\text{ss}}$ ) de 0.3 à 2500  $\text{mg.l}^{-1}$ , pour les sulfates ( $\text{SO}_{4\text{ss}}$ ) de 0.2 à 380  $\text{mg.l}^{-1}$  et pour les chlorures ( $\text{Cl}_{\text{ss}}$ ) de 1.6 à 81.9  $\text{mg.l}^{-1}$ . Les concentrations élevées de  $\text{NO}_{3\text{ss}}$  et  $\text{SO}_{4\text{ss}}$  peuvent être expliquées par un apport, peu avant le prélèvement des échantillons de sol, de fertilisants azotés, pour lesquels le soufre intervient également dans leurs compositions. Les concentrations en Cd ( $\text{Cd}_{\text{ss}}$ ) quantifiées en solution s'échelonnent sur deux ordres de grandeur (0.17 nM à 13.9 nM) avec des fractions libres du Cd allant de 12 à 95 %. Elles sont de l'ordre de celles couramment rencontrées en solutions de sols non contaminés (Sauvé et al., 2000b).

Toutes les variables mesurées en solution de sol sont reportées en Annexe IV.3.

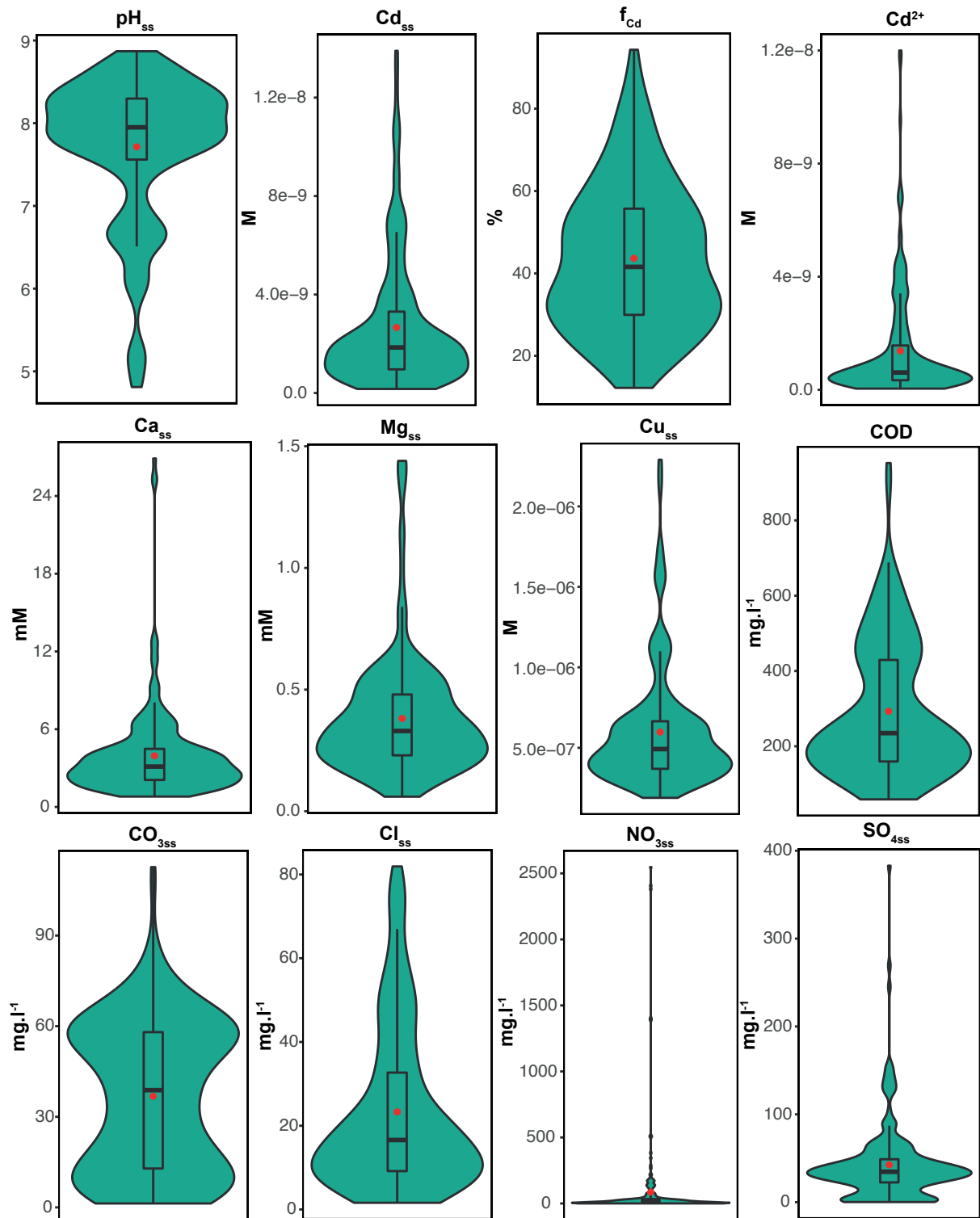


Figure II.5: Violin plots des caractéristiques chimiques mesurées en solutions de sol. Le point rouge correspond à la moyenne des individus, le box plot représente l'interquartile (25-75 %) et les segments le prolongeant correspondent à 1.5 fois l'interquartile.

## 3.2. Modélisation par régression statistique des concentrations en $Cd^{2+}$ de la solution de sol

### 3.2.1. Prédiction de $Cd^{2+}$ en solution par régression linéaire

Dans un premier temps, les concentrations en  $Cd^{2+}$  en solution ont été régressées en fonction des paramètres de la solution de sol.

Les 3 meilleurs modèles pour la prédiction des concentrations en  $Cd^{2+}$  sont résumés Tableau II.1 (Équation II.14, Équation II.15, Équation II.16). Ces modèles font intervenir en variables explicatives, la concentration en Cd complexé en solution (égale à la différence entre le  $Cd_{ss}$  et le  $Cd^{2+}$ ), le pH de la solution de sol, la concentration en COD et les concentrations en calcium et ions carbonates.

Tableau II.1 : Récapitulatif des équations obtenues pour la prédiction des concentrations de  $Cd^{2+}$  à partir d'un ligand moyen unique et les paramètres d'évaluation des modèles obtenus par validation croisée.  $r^2_{aj}$  correspond au coefficient de détermination de la régression linéaire multiple,  $VIF_{max}$  est la valeur maximale du facteur d'inflation de la variance, Ord.Orig. correspond à l'ordonnée à l'origine de la régression  $\log_{10}(\text{observé}) = f(\log_{10}[\text{prédit}])$  et  $r^2_{ord-pred}$  au coefficient de régression entre les valeurs observées et les valeurs prédites log-transformées.

Équation	Modèle	$r^2_{aj}$	$VIF_{max}$	Ord. Orig. obs/pred	Pente obs-pred	$r^2_{obs/pred}$
Équation II.14	$\log_{10}(Cd^{2+}) = -2.400^{***}$ $+ 0.630^{***} \log_{10}(Cd_{CdL})$ $- 0.251^{***} pH_{ss}$ $+ 0.365^{***} \log_{10}(COD)$	0.718	1.62	-0.08**	0.99***	0.710
Équation II.15	$\log_{10}(Cd^{2+}) = -0.093$ $+ 0.772^{***} \log_{10}(Cd_{CdL})$ $- 0.270^{***} pH_{ss}$	0.696	1.09	-0.06*	0.99***	0.690
Équation II.16	$\log_{10}(Cd^{2+}) = - 2.779^{***}$ $+ 0.618^{***} \log_{10}(Cd_{CdL})$ $- 0.760^{***} \log_{10}(CO_{3ss})$ $+ 0.546^{***} \log_{10}(Ca_{ss})$	0.744	2.00	-0.10***	0.98***	0.735

Significativité des paramètres: \*\*\*  $p$ -value < 0.001, \*\*  $p$ -value < 0.01, \*  $p$ -value < 0.05, ·  $p$ -value < 0.1

L'Équation II.14 explique 71.8 % de la variance totale des concentrations en  $Cd^{2+}$  et fait intervenir la concentration en Cd complexé, dont la contribution à la variance expliquée est de 28.8 %, le pH de la solution de sol, 25.7 %, et la concentration en COD, 17.3 %. L'Équation II.15 correspond

à un modèle dégradé de l'Équation II.14, expliquant 69.6 % de la variance totale et ne faisant intervenir que la concentration en Cd complexé, contribuant à 39.5 % de cette variance et le pH de la solution de sol, à 30.1 %. L'Équation II.16 explique 74.4 % de la variance totale des concentrations en Cd<sup>2+</sup> et fait intervenir les concentrations en Cd complexé (CdL) (33.6 %), en carbonates (36.1 %) et en calcium en solution (4.7 %). Les valeurs de VIF, inférieures ou égales à 2, ne montrent pas de fortes colinéarités entre les régresseurs. Il est également à noter que la concentration en carbonates log transformée est fortement corrélée négativement au pH de la solution de sol ( $r=0.84$ ,  $p\text{-value} = 2.2E-16$ , test de corrélation Bravais-Pearson) montrant que l'effet du pH recouvre peut-être également celui des carbonates et réciproquement (Équation II.16).

Les valeurs de pentes et d'ordonnées à l'origine de la droite obs/pred obtenues par validation croisée indiquent un léger biais de ces modèles dû à une ordonnée à l'origine significativement différente de zéro. La variation de la prédiction pour chaque échantillon en fonction du jeu de données utilisé pour le paramétrage du modèle a également été calculée et la moyenne pour chaque équation reste faible (~2.5 %). Les coefficients de détermination entre valeurs observées et valeurs prédites sont proches de 0.7 avec le meilleur modèle correspondant à l'Équation II.16, avec une valeur de 0.735. Ce dernier modèle apparaît donc comme l'alternative la plus précise et robuste pour la prédiction des concentrations en Cd<sup>2+</sup> par régression multilinéaire.

### 3.2.2. Prédiction de Cd<sup>2+</sup> sous l'hypothèse d'un ligand moyen unique

#### 3.2.2.a. Ligand unique de réactivité constante

On teste ici l'hypothèse que chaque gramme de COD a un nombre de sites et une affinité pour le Cd constants. Dans cette approche de modélisation,  $K'n$  correspond à la réactivité de la MOD, c'est-à-dire au produit du nombre de paires de sites de liaison avec Cd<sup>2+</sup> ( $L^2 \cdot g^{-1} C$ ) par la constante de stabilité conditionnelle ( $K'$ ) rapportée au carbone organique en solution (Chapitre II.2.2.2.a).  $K'n$  est relié à la fraction libre du Cd par l'Équation II.10. Ce produit varie entre 0.141  $l \cdot g^{-1} C$  et 64.4  $l \cdot g^{-1} C$  avec une valeur médiane de 5.83  $l \cdot g^{-1} C$ . Cette valeur médiane renvoie à la réactivité moyenne d'un ligand unique pour la gamme des sols étudiés. Pour tester l'hypothèse simplificatrice d'un unique ligand avec une réactivité par gramme de COD constante, la valeur médiane de  $K'n$  a été utilisée pour simuler les concentrations en Cd<sup>2+</sup> pour tous les sols en considérant qu'ils diffèrent juste par leur teneur en COD.

La régression linéaire entre les valeurs observées et valeurs prédites avec un  $K'n$  médian montre que le modèle est biaisé avec une ordonnée à l'origine significativement différente de zéro et une pente différente de 1 (2.932 et 1.319 respectivement, Tableau II.2). L'utilisation d'une valeur médiane de réactivité tend à surestimer les valeurs prédites de Cd<sup>2+</sup> pour les échantillons ayant



un K'n inférieur à la médiane et réciproquement à sous-estimer les concentrations en Cd<sup>2+</sup> pour des échantillons ayant un K'n supérieur à cette valeur. Ceci est cohérent avec le fait d'utiliser une valeur médiane unique ayant une variation assez forte (103.9 %) pour représenter la totalité des échantillons de sol. De plus, les prédictions obtenues par validation croisée varient d'en moyenne 20 % pour chaque échantillon attestant de la mauvaise robustesse de cette approche.

### 3.2.2.b. Ligand unique de réactivité fonction de son environnement chimique

Dans cette modélisation, on cherche à relier K'n à des propriétés de la solution du sol. Cela revient à prendre indirectement en compte la compétition pour la complexation en autorisant K' à varier. Comme on modélise K'n, on sous-entend que *n* peut également varier et donc que la nature de la MOD est aussi autorisée à fluctuer.

Trois modèles ont été sélectionnés, dont les équations sont résumées Tableau II.2 (Équation II.17, Équation II.18, Équation II.19).

Tableau II.2 : Récapitulatif des équations obtenues pour la prédiction des concentrations de Cd<sup>2+</sup> à partir d'un ligand moyen unique et les paramètres d'évaluation des modèles obtenus par validation croisée.  $r^2_{aj}$  correspond au coefficient de détermination de la régression linéaire multiple,  $VIF_{max}$  est la valeur maximale du facteur d'inflation de la variance, Ord.Orig. correspond à l'ordonnée à l'origine de la régression  $\log_{10}(\text{observé}) = f(\log_{10}[\text{prédit}])$  et  $r^2_{ord-pred}$  au coefficient de régression entre les valeurs observées et les valeurs prédites log-transformées.

Équation	Modèle	$r^2_{aj}$	$VIF_{max}$	Ord. Orig. obs/pred	Pente obs/pred	$r^2_{obs-pred}$
-	K'n médian	-	-	2.188***	1.237***	0.736
Équation II.17	$\log_{10}(K'n) = - 3.214$ $- 0.933^{***} \log_{10}(Ca_{ss})$ $+ 1.007^{***} \log_{10}(CO_{3ss})$	0.539	1.618	0.203***	1.023***	0.871
Équation II.18	$\log_{10}(K'n) = - 5.312$ $+ 0.431^{***} pH_{ss}$ $- 0.575^{***} \log_{10}(Ca_{ss})$	0.445	1.319	0.368***	1.041***	0.829
Équation II.19	$\log_{10}(K'n) = - 4.914$ $+ 0.343^{***} pH_{ss}$	0.369	-	0.179***	1.020***	0.787

Significativité des paramètres, \*\*\* *p-value* < 0.001, \*\* *p-value* < 0.01, \* *p-value* < 0.05, · *p-value* < 0.1

Le meilleur modèle expliquant  $K'n$ , l'Équation II.17, fait intervenir les concentrations en Ca et en carbonates en solution. Aucune colinéarité importante entre les régresseurs n'est détectée (VIF inférieur à 2). Ce modèle explique 53.9 % de la variance totale correspondant à 45.4 % expliqués par la concentration en carbonates et 8.5 % par les concentrations en Ca en solution. Le second modèle (Équation II.18) explique 44.5 % de la variance et fait intervenir le pH de la solution de sol, expliquant 40.5 % et la concentration en Ca en solution expliquant 4.0 %, sans colinéarité entre ces variables. La dernière équation correspond au modèle dégradé de l'Équation II.18 ne faisant intervenir que le pH de la solution de sol comme variable explicative de la réactivité d'un ligand unique assimilé à de la MOD. Le pH de la solution de sol seul explique 36.9 % de variance totale (Équation II.19).

En dépit d'une prédiction de qualité moyenne des valeurs de  $K'n$  ( $r^2_{aj}$  proches de 0.5), les valeurs calculées de  $Cd^{2+}$  peuvent être considérées en plutôt bonne adéquation avec les valeurs mesurées, et plus particulièrement pour l'Équation II.17 avec un coefficient de détermination de 0.871. Le biais observé est également plus important que précédemment, avec des ordonnées à l'origine (Ord. Orig.) significativement différents de zéro et des pentes restant cependant très proches de 1.

### 3.2.3. Comparaison des approches statistiques

La distribution des valeurs absolues des erreurs relatives des sept précédents modèles (Équation II.14 à Équation II.19 et  $K'n$  médian) sont comparées Figure II.6. Les erreurs relatives de tous les modèles présentés sont calculées à partir des validations croisées et après correction de la prédiction par un biais éventuel. Elles prennent donc en compte à la fois l'erreur liée au paramétrage du modèle statistique (robustesse) et l'erreur liée au biais résiduel (précision et justesse). Les distributions de ces erreurs sont représentées par des *violin plots* (Figure II.6) regroupant la distribution des erreurs, représentée par le *box plot*, et la densité de probabilité représentée de chaque côté du *box plot*. Plus les sections sont larges, plus forte est la probabilité qu'un échantillon de la population prenne cette valeur donnée. Ainsi, si la densité de probabilité est grande et proche de zéro, cela signifie que le modèle a produit une majorité des cas des erreurs relatives faibles après correction du biais et atteste donc de sa justesse et en partie de sa précision ainsi que sa robustesse.

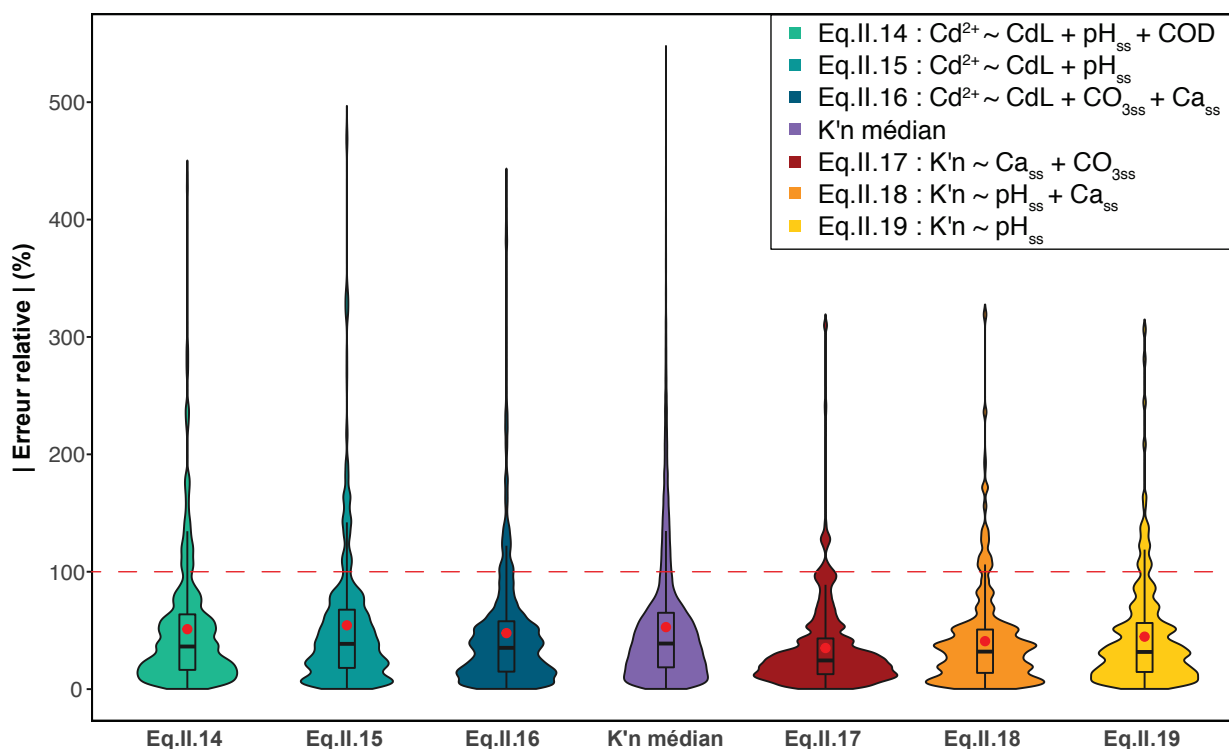


Figure II.6 : Distribution des valeurs absolues des erreurs relatives (en pourcentage) obtenues par validation croisée des différents modèles statistiques corrigés de leur biais de prédiction (Équation II.14 à Équation II.19 et K'n médian). Le point rouge correspond à la moyenne des individus, le box plot représente l'interquartile (25-75 %) et les segments le prolongeant correspondent à 1.5 fois l'interquartile. La densité de probabilité de l'erreur relative est représentée de chaque côté du box plot. La droite pointillée rouge représente une erreur relative de 100 %.

Pour toutes les approches, la majorité des valeurs absolues des erreurs relatives se situent entre 0 et 100 %, soit une surestimation ou sous-estimation de la concentration en  $Cd^{2+}$  d'un facteur 2. Cependant, la queue de la distribution est plus grande pour les Équation II.14 à Équation II.16 et l'approche du K'n médian, ce qui signifie des erreurs relatives globalement plus fortes comparées aux approches de modélisation de K'n en fonction des propriétés de la solution de sol (Équation II.17 à Équation II.19). Parmi ces trois dernières équations, on observe que les erreurs relatives sont plus proches de zéro pour une plus grande quantité d'échantillons dans le cas des prédictions obtenues à partir de l'Équation II.17. Le meilleur modèle se base donc sur un ligand moyen issu de la MOD dont la réactivité est variable suivant les concentrations en carbonates et le pH de la solution (effet confondu) et les concentrations en calcium en solution (Équation II.17).

### 3.3. Modélisation géochimique de la concentration en $\text{Cd}^{2+}$ de la solution de sol

#### 3.3.1. Influence relative des acides humiques

Pour tester l'influence relative des AH sur la qualité de la prédiction des concentrations en  $\text{Cd}^{2+}$ , la MOD est supposée être composée soit à 50/50 d'AH/AF, soit seulement d'AF (100 %).

La Figure II.7 représente la distribution des prédictions des concentrations de  $\text{Cd}^{2+}$  entre une MOD constituée exclusivement d'AF (en gris) et une MOD constituée de 50 % d'AH/50 % d'AF (en jaune). On observe bien que les deux distributions sont très similaires avec des prédictions qui ne sont pas significativement différentes si l'on considère des AH ou non dans le paramétrage de WHAM (Test T de Wilcoxon pour échantillons appariés,  $p\text{-value} = 0.472$ ).

Sur les sols testés, la simplification d'assimiler le COD actif uniquement à des AF est donc valide.

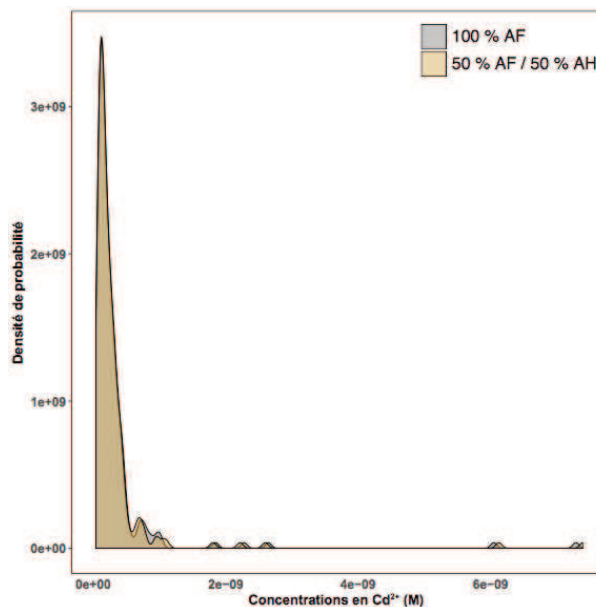


Figure II.7 : Distribution comparée des prédictions de  $[\text{Cd}^{2+}]$  modélisées par CHEAQS Next en fonction de la nature de la MOD considérée. Les résultats issus de l'hypothèse d'une MOD composée de 100 % d'AF sont représentés en gris et en jaune pour l'hypothèse 50 % AF et 50 % AH.

#### 3.3.2. Influence de la composition de la MOD ( $\text{COD}_{\text{actif}}$ + AF + LMWOA) sur les prédictions des concentrations de $\text{Cd}^{2+}$ en solution

Dans cette partie, l'effet de la réactivité de la MOD, c'est-à-dire des pourcentages de  $\text{COD}_{\text{actif}}$ , d'AF et de LMWOA, sur les prédictions des concentrations en  $\text{Cd}^{2+}$  a été testé en appliquant une réactivité identique à tous les sols. Pour rappel, la totalité des compositions possibles sont résumées Figure II.2.

D'un point de vue général, l'erreur relative générée par les différentes combinaisons est majoritairement expliquée par le nombre global de sites de complexation considérés, liés au pourcentage de COD actif et à la répartition de ces sites entre les AF et les LMWOA, lié au pourcentage d'AF (Anova,  $F$ -value de 6501.2 et 1730.6 respectivement,  $p$ -value  $< 2.2e^{-16}$ ). La composition de LMWOA a également une influence significative sur les erreurs générées ( $F$ -value = 22.0,  $p$ -value  $< 2.2e^{-16}$ ), mais très faible devant les effets des facteurs précédents comme l'atteste la  $F$ -value. Il est cependant à noter que dans l'ensemble, une composition de LMWOA avec des pourcentages d'acide acétique élevés a tendance à générer des erreurs relatives plus faibles. Ainsi, assimiler les LMWOA à de l'acide acétique améliorerait la qualité de prédiction des concentrations en  $Cd^{2+}$ .

De la même façon, l'analyse de l'effet du COD actif seul sur les erreurs relatives montre une tendance à l'augmentation de ces erreurs lorsque le pourcentage de COD actif augmente. Cette tendance est visible sur la Figure II.8 représentant, pour un pourcentage d'AF de 100 %, l'évolution de l'erreur relative de prédiction en fonction du pourcentage de COD actif. On note bien que l'erreur relative moyenne est plus faible pour l'hypothèse de 10 % COD actif (47.8 %) comparée aux autres modalités (68.5 % pour 20 % COD actif, 81.7 % pour 30 % COD actif, 90.1 % pour 40 % COD actif et 96.4 % pour 50 % COD actif). De la même façon, il est visible sur la Figure II.8 que la distribution des erreurs pour l'hypothèse 10 % COD et 100 % d'AF est la plus proche de zéro avec un coefficient de variation plus faible (domaine interquartile plus petit).

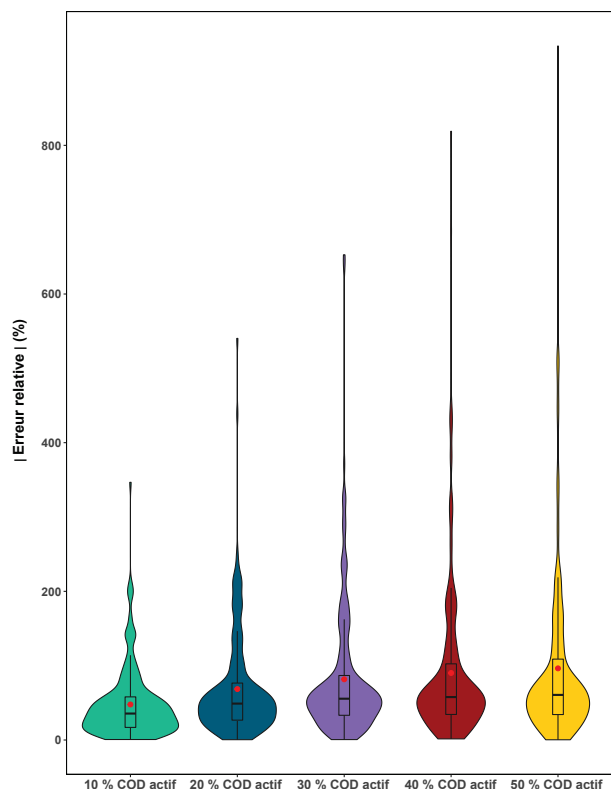


Figure II.8 : Distribution comparée des prédictions des  $[Cd^{2+}]$  modélisées par CHEAQS Next et corrigée de leur biais éventuel en fonction du pourcentage de COD actif et pour 100 % d'AF. Le point rouge correspond à la moyenne des individus, le box plot représente l'interquartile (25-75 %) et les segments le prolongeant correspondent à 1.5 fois l'interquartile. La densité de probabilité de l'erreur relative est représentée de chaque côté du box plot.

Parmi toutes les combinaisons de composition de la MOD testées, celle permettant la meilleure prédiction des concentrations en  $Cd^{2+}$ , pour l'ensemble des sols, correspond à 10 % de COD actif, dont 20 % sont constitués d'AF et 80 % de LMWOA dont la composition est 90 % d'acide acétique et 10 % d'acide citrique.

La Figure II.9 représente les données log transformées mesurées de  $Cd^{2+}$  en fonction des données log-transformées prédites en utilisant cette combinaison de paramètres de MOD et regroupe les paramètres statistiques ayant permis son évaluation. Le coefficient de détermination entre valeurs observées et valeurs prédites en log est de 0.875. Ce modèle est biaisé et tend à surestimer systématiquement les valeurs de  $Cd^{2+}$  prédites d'autant plus pour les faibles valeurs de  $\log_{10}([Cd^{2+}])$ . En effet, la pente du modèle observé=f(prédit) est égale à 1.18 et son ordonnée à l'origine est significativement différente de zéro (1.31).

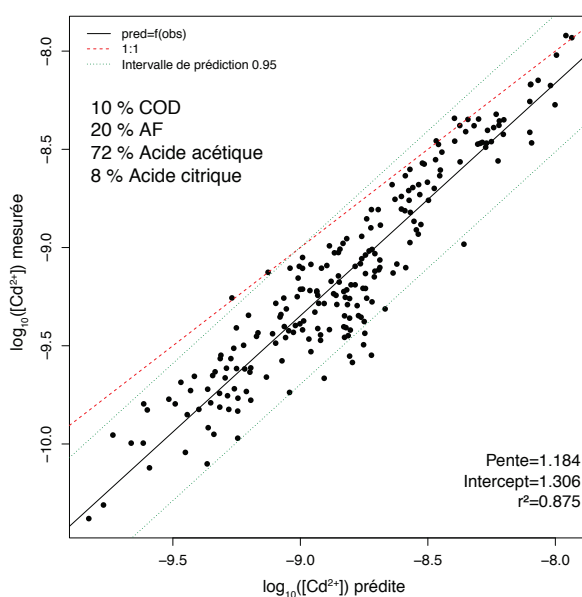


Figure II.9 : Concentrations en  $Cd^{2+}$  (log transformées) prédites sur l'hypothèse de 10 % de COD actif, dont 20 % d'AF, 72 % d'acide acétique et 8 % d'acide citrique, en fonction de celles mesurées. La ligne noire correspond au modèle linéaire entre valeurs observées et valeurs prédites, la ligne en pointillé rouge, l'équation  $y=x$  et les lignes vertes à l'intervalle de prédiction à 0.95).

### 3.3.3. Modélisation géochimique de Cd<sup>2+</sup> en reliant la MOD aux propriétés du sol

La modélisation géochimique précédente était fondée sur l'hypothèse d'une composition de MOD moyenne appliquée à tous les sols. Ici, nous avons tenté de relier cette composition de MOD aux propriétés du sol et de la solution du sol.

Après avoir sélectionné, pour chacun des sols, les réactivités de MOD qui produisent les erreurs de prédiction de Cd<sup>2+</sup> les plus faibles et après avoir converti les pourcentages de COD actif et d'AF en nombre de sites, ces derniers sont reliés aux propriétés du sol par *Random Forest*. Les variables identifiées comme étant les plus influentes sont le pH du sol mesuré dans un extrait de CaCl<sub>2</sub> à 0.01 M, la concentration en Ca en solution et en carbonates (Figure II.10). La totalité des variables reportées sur cette figure permet d'expliquer 65.1 % de la variance totale du nombre de sites de complexation.

Afin de prédire le nombre de sites de complexation, nous avons choisi de sélectionner les deux variables les plus influentes (pH CaCl<sub>2</sub> et Ca<sub>ss</sub>). Elles expliquent environ le même pourcentage de variance du nombre de sites de complexation que la totalité des variables étudiées (64.9 %), ce qui conforte l'idée de ne prédire les concentrations de Cd<sup>2+</sup> qu'à partir de ces deux variables. Afin de déterminer l'influence de ces deux variables sur le nombre de sites, une régression linéaire multiple a été effectuée donnant des coefficients positifs significatifs pour le pH CaCl<sub>2</sub> et la concentration en Ca<sub>ss</sub> ( $1.55 e^{-4 ***}$  et  $1.51 e^{-4 ***}$  respectivement,  $p\text{-value} < 6.4 e^{-7}$ ).

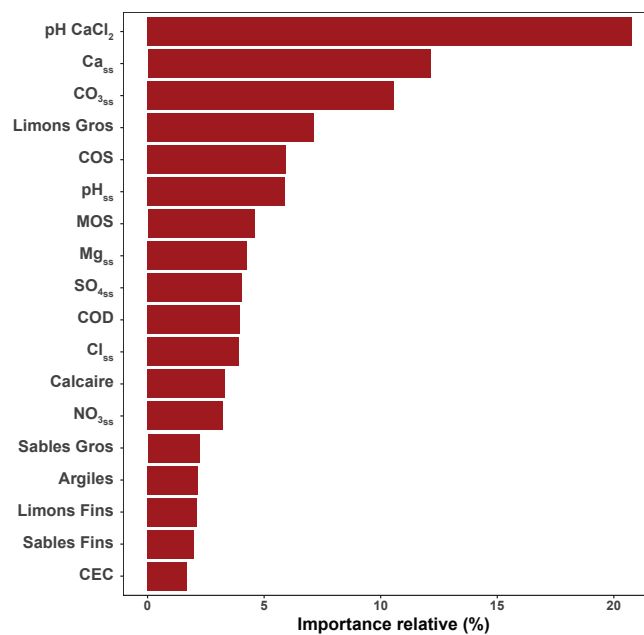


Figure II.10 : Importance relative des variables pour la modélisation du nombre de sites de complexation obtenue par *Random Forest*. Les valeurs d'importance ont été recalculées en fixant leur somme à 100 %.

L'adéquation entre valeurs observées et valeurs prédites par *Random Forest* en reliant le nombre de sites au pH CaCl<sub>2</sub> et aux concentrations en Ca en solution est présentée Figure II.11. Par rapport à l'application d'une réactivité de la MOD moyenne et unique pour tous les sols, la modulation de la réactivité de la MOD par le pH CaCl<sub>2</sub> du sol permet une prédiction précise ( $r^2=0.979$ ) et très peu biaisée des concentrations en Cd<sup>2+</sup> en solution avec une pente proche de 1 (1.012) et une ordonnée à l'origine significativement non différente de zéro (0.003,  $p$ -value = 0.971). Il tend encore à légèrement surestimer les valeurs de Cd<sup>2+</sup> en solution.

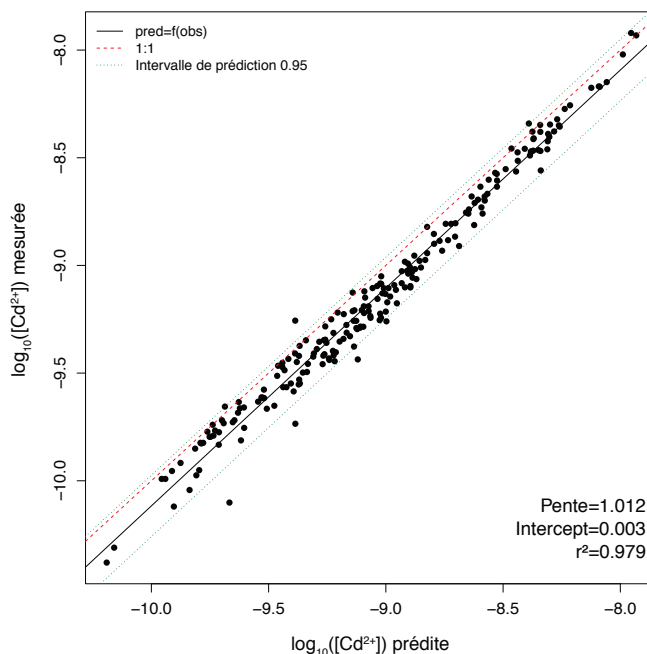


Figure II.11 : Concentrations en Cd<sup>2+</sup> (log transformées) prédites sur l'hypothèse d'un pourcentage de COD actif et d'AF dépendants du pH CaCl<sub>2</sub>, en fonction de celles mesurées. La ligne noire correspond au modèle linéaire entre valeurs observées et valeurs prédites, la ligne en pointillé rouge, l'équation  $y=x$  et les lignes vertes à l'intervalle de prédiction à 0.95.

### 3.4. Comparaison des approches pour la prédiction des concentrations en Cd<sup>2+</sup> en solution

La Figure II.12 regroupe la totalité des approches retenues permettant les prédictions de Cd<sup>2+</sup> les plus justes après correction de leur biais éventuel. Les approches statistiques avec l'Équation II.16, l'Équation II.17 et le K'n médian et les approches mécanistiques supposant une MOD de réactivité de 10 % COD actif – 100 % AF, 10 % COD actif – 20 % AF et 80 % LMWOA et une MOD de réactivité variable en fonction du pH CaCl<sub>2</sub> du sol et de la concentration en Ca en solution (notée RandomForest pH CaCl<sub>2</sub> + Ca<sub>ss</sub> sur la Figure II.12).



Globalement on observe des valeurs maximales d'erreurs relatives plus grandes pour les approches de modélisation purement statistique que pour les approches de modélisation géochimique et statistique. Les approches pouvant être qualifiées de plus simples (statistiques) génèrent des erreurs plus fortes pour une plus grande quantité d'échantillons. Ainsi, 37.9 % des individus pour l'hypothèse d'une réactivité médiane et 33.9 % des individus pour l'hypothèse d'une MOD composée de 10 % de COD actif dont 100 % d'AF génèrent des erreurs relatives supérieures à 50 %.

L'approche permettant de générer des erreurs de prédiction les plus faibles pour une majorité de sols correspond à une réactivité de MOD (déterminée par un pourcentage de COD actif et d'AF) dépendant du pH CaCl<sub>2</sub> du sol et de la concentration en Ca en solution. En effet, pour 98.6 % des sols, l'erreur relative générée par la prédiction est inférieure à 50 %. Les prédictions sont nettement meilleures que pour un modèle d'une MOD avec 10 % de COD actif dont 20 % d'AF et 80 % de LMWOA, mais permettent surtout une prédiction beaucoup moins biaisée (Figure II.11 et Figure II.9 respectivement) ce qui milite pour une représentation de la MOD plus réaliste dans les solutions de sol.

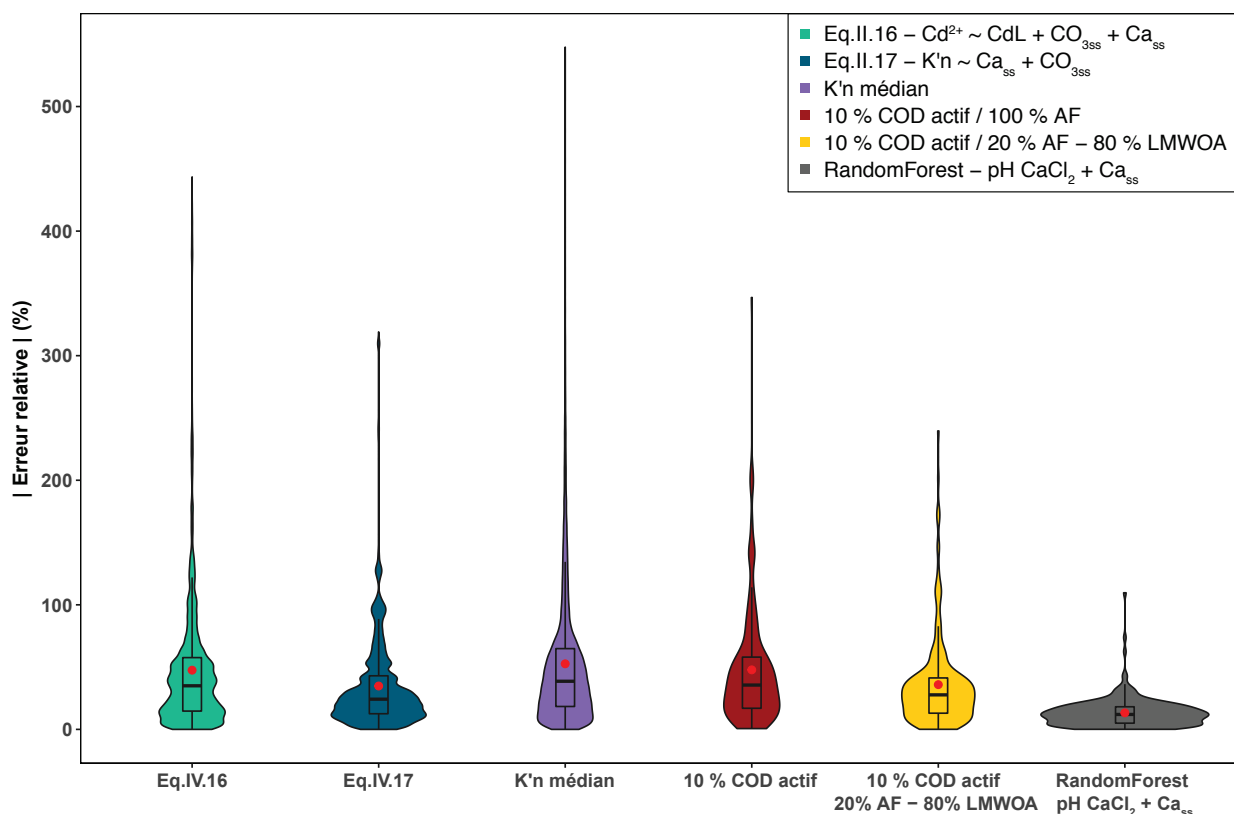


Figure II.12 : Distributions des valeurs absolues des erreurs relatives corrigées de leur biais de prédiction pour les différentes approches de modélisation testées dans ce chapitre pour la prédiction des concentrations en Cd<sup>2+</sup> en solution. Le point rouge correspond à la moyenne des individus, le box plot représente l'interquartile (25-75 %) et les segments le prolongeant correspondent à 1.5 fois l'interquartile. La densité de probabilité de l'erreur relative est représentée de chaque côté du box plot.

La Figure II.13 synthétise les meilleurs résultats obtenus par les 5 approches testées (Figure II.3). Elle récapitule donc, pour chaque approche, comment prédire les concentrations en  $\text{Cd}^{2+}$  et l'erreur moyenne estimée.

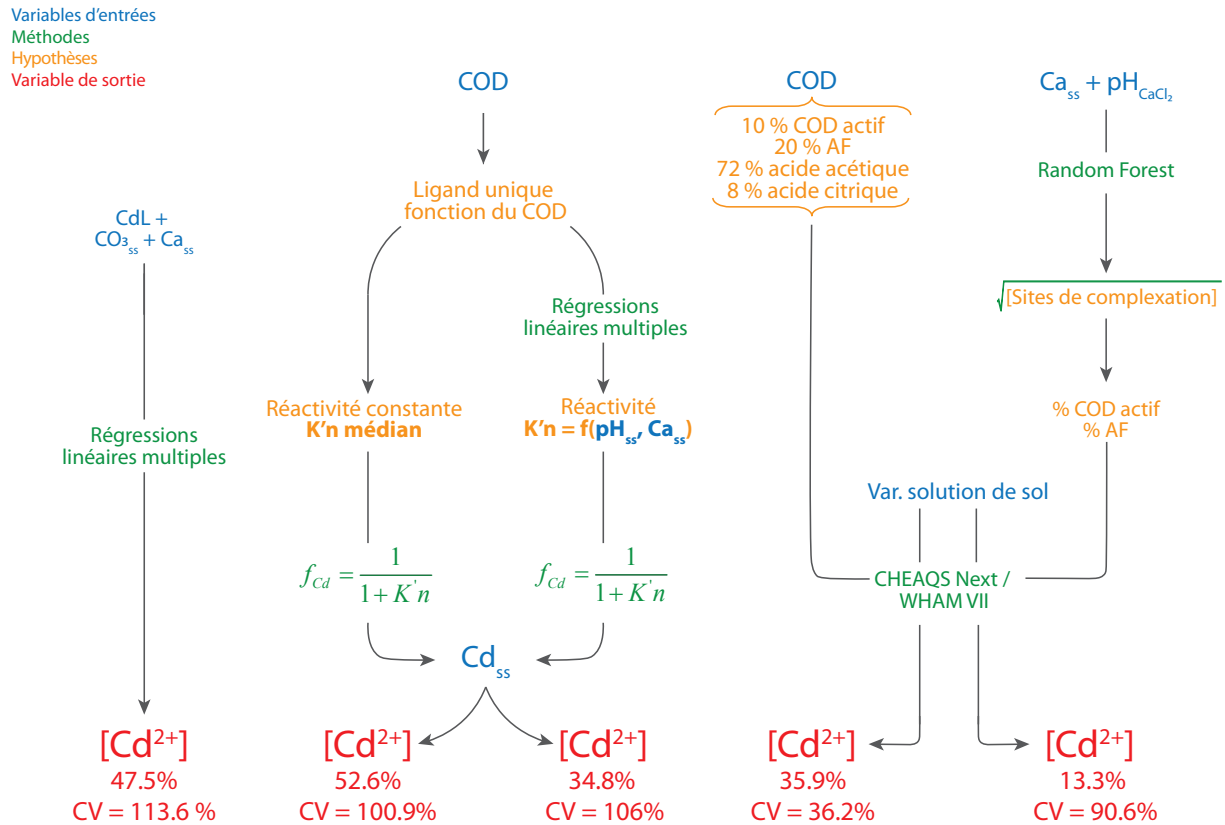


Figure II.13 : Schéma récapitulatif des résultats obtenus pour les 5 approches de modélisation testées dans le cadre de ce chapitre. Le résultat final, présenté en rouge, correspond à la moyenne des erreurs relatives en valeurs absolues obtenues par validation croisée et son coefficient de variation.

## **4. Discussion**

### **4.1. Quels paramètres de MOD pour la modélisation géochimique de la spéciation du Cd**

#### **4.1.1. Modélisation géochimique inverse pour la détermination de la réactivité de la MOD**

Considérer la présence d'AH dans la composition de la MOD ou la remplacer par des AF ne modifie pas significativement les prédictions des concentrations en  $\text{Cd}^{2+}$ . La constante de complexation du Cd étant légèrement plus élevée avec les AH qu'avec les AF, ce résultat indique que le Cd se complexe majoritairement avec les AF. Cette complexation favorisée du Cd avec les AF est probablement due à un nombre de sites de complexation (groupements carboxyliques et phénoliques) plus important sur les AF qui n'est jamais limitant dans le cadre de notre étude (même lorsque seulement 50 % d'AF sont considérés). Lorsque les sites de complexation du Cd sur les AF deviendront limitants, ce qui sera potentiellement le cas lorsque les concentrations en Cd seront plus importantes, lorsque les concentrations de cations compétiteurs en solution seront plus élevées et/ou lorsque les concentrations en COD actif seront plus faibles, les AH pourront influencer la complexation du Cd et ne devront pas être substitués à des AF.

Dans nos solutions de sol, les proportions d'AH sont faibles comparées à celle des AF puisqu'aucun précipité n'a été détecté (résultat non présenté) lors de la première étape de caractérisation de la MOD (Thurman and Malcolm, 1981). Un faible pourcentage d'AH en solution de sol est également reporté dans la bibliographie, où il représente de 0 à 4.2 % de la MOD, tandis que les AF vont en représenter de 15 à 48 % (Amery et al., 2009; Fest et al., 2008; Groenenberg et al., 2010a; Ren et al., 2015b; Schneider et al., 2016b). Cela pourrait être lié au fait que la réhumectation d'un sol sec semble favoriser la solubilisation d'une MOD plus hydrophile liée à la reprise de l'activité microbienne (Merckx et al., 2001), donc moins riche en AH. Une seconde hypothèse serait que la MOD ne serait pas en réalité une macromolécule, mais une supramolécule, c'est-à-dire un ensemble de plusieurs petites molécules organiques liées entre elles par des interactions intermoléculaires non covalentes, de type Van der Waals et liaisons hydrogènes (Piccolo, 2001). Ainsi, lorsque le sol est réhumecté et la solution est récupérée par centrifugation, la stabilité de ces interactions faibles pourrait être altérée, favorisant la dissociation de ces supramolécules en molécules de plus faibles poids moléculaires et plus hydrophiles expliquant ainsi les fractions hydrophiles, proches des AF, plus importantes mesurées en solution.

Donc, il apparaît raisonnable de substituer la totalité des substances humiques de la MOD à des AF dans un contexte de sols agricoles non contaminés, sans générer de fortes erreurs sur la prédiction du  $\text{Cd}^{2+}$ . Ceci est uniquement valable lorsque la quantité de sites de complexation

disponibles des AF n'est pas limitante, ce qui est le cas pour nos solutions de sol où les concentrations en Cd sont  $10^{-6}$  fois celles du COD.

La part de COD actif n'étant pas des AF a été considérée dans ce travail comme un mélange de LMWOA. La composition de ces LMWOA a peu d'influence sur la qualité de prédiction des concentrations de  $Cd^{2+}$  avec cependant, une prédiction globale meilleure pour des LMWOA assimilés à des mono-acides de type acide acétique. Cette faible influence de la composition peut également traduire une faible influence de la fraction non humique dans la spéciation du Cd. Cette conclusion est proche des conclusions reportées dans d'autres études où cette fraction a été considérée comme non significative pour la spéciation du Cd. Les concentrations en COD mesurées dans nos sols sont beaucoup plus élevées que celles retrouvées dans la bibliographie (facteur 5 à 10), aboutissant à des concentrations en LMWOA considérées dans le modèle géochimique plus fortes, ce qui pourrait expliquer cette influence légèrement plus importante sur la prédiction des concentrations en  $Cd^{2+}$  (Groenenberg et al., 2010a; Ren et al., 2015b; Schneider et al., 2016b).

La modélisation inverse de la complexation du Cd en solution de sol suggère donc un rôle prépondérant du nombre de sites de complexation des AF, et ce pour tous les pH de sol étudiés. Du fait de la prédominance des sols alcalins dans notre étude, cette conclusion est moins certaine pour les sols les plus acides. Cependant les conclusions similaires reportées dans la bibliographie où des sols plus acides ont été étudiés, permet d'accorder un degré de confiance assez fort à cette conclusion (Groenenberg et al., 2010a).

Les pourcentages de COD actif et d'AF, représentant le nombre de sites de complexation et leur répartition entre substances humiques (AF) et non humiques (LMWOA), sont donc les paramètres les plus importants pour la prédiction des concentrations de  $Cd^{2+}$  en solution de sol, en conditionnant la quantité de sites de complexation prise en compte par le modèle géochimique. Si l'on choisit d'appliquer une réactivité moyenne unique pour tous les sols, l'hypothèse 10 % de COD actif et 20 % d'AF permet les prédictions les plus précises, mais surestime fortement les concentrations en  $Cd^{2+}$  en solution pour les concentrations les plus faibles, c'est-à-dire pour les sols les plus alcalins. Cette réactivité unique peut donc être efficacement utilisée pour prédire les concentrations en  $Cd^{2+}$  en solution moyennant une correction de biais. Ce biais montre que l'on ne modélise pas correctement la MOD pour la totalité des sols, ce qui affaiblit la robustesse du modèle pour de futurs sols.

#### **4.1.2. Variabilité de la réactivité de la MOD en fonction des propriétés des sols et solutions de sol**

Pour améliorer les prédictions des concentrations en  $Cd^{2+}$ , la réactivité de la MOD, modélisée par approches statistique et géochimique, a été régressée en fonction des propriétés du sol afin de

pouvoir affiner sa prédiction à partir de variables, du sol ou de la solution du sol, facilement mesurables.

Les paramètres expliquant le mieux le produit  $K'n$ , approximant la réactivité de la MOD, sont les concentrations en carbonates et en calcium présents en solution (Équation II.17). L'influence positive des carbonates en solution sur le produit  $K'n$  doit être discutée. En effet les ligands en solution, autre que la MOD, ont logiquement une influence négative sur la constante conditionnelle de complexation de la MOD vis-à-vis du Cd. Cependant, les carbonates et le pH de la solution de sol, sont fortement corrélés et on ne peut pas savoir à qui attribuer l'effet observé sur  $K'n$ . La corrélation positive du pH de la solution du sol avec la valeur de  $K'n$  (Équation II.18), indique qu'à pH élevé, la réactivité de la MOD ( $K'n$ ) sera plus importante qu'à pH faibles. Une première explication serait que plus le pH est faible, plus les concentrations en protons en solution augmentent, ce qui se traduit par une compétition importante des protons avec le Cd et donc une complexation du Cd défavorisée induisant une constante conditionnelle de complexation du Cd par la MOD ( $K'$ ) plus faible. L'influence du pH peut également porter sur l'abondance de sites de complexation de la MOD. Ainsi, plus le pH augmenterait, plus le nombre de sites de complexation augmenterait également, se traduisant par une MOD de nature différente en fonction du pH. Cette observation est cohérente avec le fait qu'à pH plus élevé, l'activité microbienne favorise la dégradation des substances organiques augmentant ainsi les quantités d'AF (Roth et al., 2015) et donc le nombre de sites de complexation.

L'approche géochimique, en considérant les effets des différentes espèces en solution sur la constante conditionnelle de complexation du Cd avec la MOD, a permis d'identifier les effets des différentes variables du sol et de la solution de sol sur la quantité de sites impliqués dans la complexation. Ainsi l'effet positif du pH  $\text{CaCl}_2$  sur ce nombre de sites suggère bien qu'une quantité plus importante de sites de complexation est impliquée dans la complexation du Cd ce qui suggère que l'effet des carbonates sur  $K'n$  est en réalité un effet du pH de la solution de sol. Le pH peut également influencer le nombre de paires de sites réactifs en influençant la quantité de MOD solubilisée. Cependant, selon la littérature, deux effets opposés peuvent coexister, l'un positif, augmentant la solubilité du COD avec l'augmentation du pH, et l'autre négatif, la réduisant à mesure que le pH augmente. La MOD, généralement considérée comme un électrolyte polyfonctionnel de fort poids moléculaire, doit sa solubilité à sa densité de charge. En effet, plus une molécule sera chargée, plus les interactions entre les molécules du solvant (en solution de sol, l'eau) de types liaisons hydrogène et liaisons de Van der Waals, et le macro-ion permettront sa solvatation. Cette charge globale de surface dépend des constantes d'acidités de la molécule et donc du pH de la solution. Ainsi, la solubilité de la MOD et donc la quantité retrouvée en solution, est réduite à mesure que sa surface est protonée, c'est-à-dire à de faibles pH (Tipping and Woof, 1990). De plus, cet effet semble plus marqué dans le cas de solutions faiblement concentrées en Ca (Oste et al., 2002; Temminghoff et al., 1998). Cependant, l'effet opposé du

pH sur la quantité de MOD est visible sur notre jeu de données, tendance également retrouvée dans la bibliographie (de Troyer et al., 2014). L'explication pourrait être liée à une force de sorption de la MOD sur les surfaces du sol plus faibles à pH acide, qui combinée à une MOD protonée (pH de la solution de sol inférieur aux pKa des groupements acides de la MOD), favoriserait sa solubilisation (Kalbitz et al., 2000).

La corrélation négative des concentrations en Ca avec  $K'n$  peut s'expliquer par l'augmentation de la compétition du Ca en solution avec le Cd, favorisant la dissociation du Cd et sa présence sous forme libre en solution (Cao et al., 2006; Christensen and Christensen, 2000; John et al., 1988; Saar and Weber, 1979), baissant ainsi la constante conditionnelle de complexation. L'effet du Ca sur le nombre de sites par unité de MOD ( $n$ ) est beaucoup moins évident à expliquer d'un point de vue mécaniste et cette approche ne nous permet pas de cerner l'effet réel de Ca sur le nombre de sites. La régression multilinéaire du nombre de sites (indépendant des concentrations en MOD en solution) en fonction du pH  $\text{CaCl}_2$  et des concentrations en Ca en solution montre une influence positive de cette dernière variable sur la quantité de sites. Ceci pourrait indiquer que l'effet observé du Ca sur  $K'n$  serait à la fois lié à un effet négatif majoritaire sur la constante de complexation conditionnelle et un effet positif sur le nombre de sites. Ainsi, plus les concentrations en Ca augmenteraient, plus la quantité de sites serait élevée. Ceci peut être relié à une augmentation de la solubilisation de la matière organique du sol lorsque les concentrations en Ca en solution et le pH augmente (Curtin et al., 2016).

#### **4.2. Prédiction des concentrations en $\text{Cd}^{2+}$ en solution de sols agricoles**

L'approche la plus intéressante pour la prédiction des concentrations de  $\text{Cd}^{2+}$  fait appel à un modèle géochimique où la densité de sites de complexation est variable en fonction du pH  $\text{CaCl}_2$  du sol et de la concentration en Ca en solution. Elle permet les prédictions les plus précises avec un modèle peu biaisé et des erreurs relatives plus faibles que pour les autres approches (Figure II.12). Cette approche suggère que malgré la complexité de la MO du sol, il est possible de modéliser son comportement moyen à partir d'un nombre restreint de variables (pH  $\text{CaCl}_2$ ,  $\text{Ca}_{\text{ss}}$ ). Ici, environ 98 % des sols ont une erreur relative de moins de 50 % ce qui se traduit par des prédictions de  $\text{Cd}^{2+}$  excellentes vu le type de sol majoritairement utilisé dans ce travail, c'est-à-dire des sols alcalins faiblement contaminés en Cd. L'erreur quadratique moyenne de cette approche (RMSE = 0.127 unité log) sont cependant du même ordre de grandeur que celles généralement retrouvées dans le cadre d'études où la composition de la MOD a été expérimentalement déterminée (RMSE de 0.1-0.2 unité log) (Groenenberg et al., 2010a; Ren et al., 2015b) et largement inférieure à celle où la MOD a été assimilée à un pourcentage d'AF fixe

(0.26-0.67 unité log) (Ge et al., 2005; Weng et al., 2002a). Notre approche de modélisation simple apporte donc bien une plus-value par rapport à la littérature.

## **5. Conclusion**

Dans la littérature, la MOD est le ligand en solution reconnu comme influençant majoritairement la spéciation du Cd. Sa caractérisation est donc primordiale pour prédire correctement les concentrations en  $\text{Cd}^{2+}$  en solution et plus particulièrement la densité de sites de complexation liée à la fraction de COD actif en solution et au pourcentage d'AF composant cette fraction active. Nous avons montré que la prédiction de la composition de ces fractions est possible à partir de mesures simples de pH du sol (pH mesuré dans un extrait de  $\text{CaCl}_2$  à 0.01 M) et concentration en Ca en solution, et peut être utilisée en complément de variables mesurables facilement en solution de sol (concentrations en cations et anions majeurs, Cd, carbone organique et inorganique dissous et pH). Cette relation doit être validée et renforcée par la caractérisation de la composition de la MOD sur une plus grande variété de sols. Les variables de la composition de la solution utilisées pour paramétrer ce modèle géochimique peuvent également être simplifiées en étudiant leur influence sur la prédiction. De plus, ces prédictions sont dépendantes de la méthode d'extraction de la solution de sol. Dans cette étude, en raison de la réhumectation du sol sec, les conditions de récupération de cette solution ont pu favoriser la présence en grande quantité de composés organiques peu réactifs en comparaison à une solution de sol « à l'équilibre » au champ. Ainsi, il serait intéressant d'appliquer cette approche de modélisation à des solutions de sol récupérées, au champ, par bougie poreuse, ou par centrifugation de sols frais. Le ligand majoritaire étant la MOD, l'utilité de la composition anionique (sulfates, nitrates, chlorures) est questionnable et une analyse de sensibilité sur ces paramètres permettrait de potentiellement réduire les variables d'entrée nécessaires au paramétrage tout en conservant sa qualité de prédiction. La principale limite et voie d'amélioration de cette approche concerne sa précision qui peut avoir deux origines : la robustesse du modèle, c'est-à-dire sa dépendance au jeu de donnée de calibration et les hypothèses simplificatrices effectuées.

Cette approche de modélisation nécessiterait donc d'être testée sur un panel plus large de sols afin de voir s'il est possible de réduire l'erreur liée au jeu de données tout en gardant à l'esprit qu'un manque de précision doit être accepté et plus particulièrement lorsque les hypothèses posées sont simples.







## **Chapitre III. Étude de la distribution du Cd en sols agricoles et prédictions des concentrations en Cd<sup>2+</sup> en solution**

---

### **1. Contexte**

La biodisponibilité du cadmium (Cd) dans les sols est régie par une multitude de processus physiques, chimiques et biologiques souvent interdépendants, tels que la convection, la diffusion, le prélèvement par les racines et les réactions physico-chimiques comme la sorption/désorption et la précipitation/dissolution. Le Cd est considéré comme un métal relativement mobile dans les sols, c'est-à-dire ayant une forte propension à être solubilisé (Kabata-Pendias and Pendias, 2001). Le caractère mobile de cet élément conditionne sa disponibilité environnementale considérée comme la fraction maximale du Cd du sol (phase solide et solution) susceptible d'être absorbée par une cible biologique.

**Ainsi, la compréhension de la distribution du Cd entre le sol et la solution du sol et des paramètres physico-chimiques l'influençant est au cœur du processus d'évaluation du risque de contamination des cultures.**

La distribution sol-solution du Cd est largement liée aux phases réactives du sol et plus particulièrement les matières organiques du sol (MOS), les argiles minéralogiques et les oxydes métalliques amorphes, suivant différents processus physico-chimiques tels que l'adsorption non spécifique (échange d'ions), l'adsorption spécifique (complexation), et la précipitation (Sparks, 2003). Dans le cas de sols faiblement contaminés, ce dernier processus est largement minoritaire et n'affecte qu'indirectement la distribution du Cd (par exemple, via la diminution de la compétition par précipitation du Ca) (Tiller et al., 1979). La répartition du Cd entre les différentes phases réactives définit la spéciation en phase solide de cet élément et c'est donc la réactivité et l'abondance de ces différentes surfaces vis-à-vis du Cd qui conditionnent sa mobilité, c'est-à-dire sa distribution entre la phase solide et la solution de sol.

Dans le cas de la sorption non spécifique, le Cd est retenu sur les sites négativement chargés du sol par des liaisons électrostatiques formant avec la forme hydratée (Cd<sup>2+</sup>[H<sub>2</sub>O]<sub>6</sub>) des complexes de sphère externe supposés localisés dans la double couche de diffusion de la surface (McBride, 1994). Le Cd sorbé par adsorption non spécifique est soumis au processus d'échange d'ions et est donc

particulièrement sujet à la compétition avec les autres cations présents en solution, notamment les cations divalents.

La sorption spécifique du Cd est le second mécanisme influençant sa spéciation en phase solide. Dans ce cas, le Cd forme une liaison chimique covalente avec les sites négativement chargés ou neutres. Ce processus, aussi appelé complexation, permet la formation de complexes de sphères internes, moins échangeables pour lesquels la compétition est fonction de la constante d'affinité du métal avec la surface (McBride, 1994).

La sorption du Cd par les argiles est majoritairement du type sorption non spécifique par échange d'ions sur des sites présentant des charges négatives permanentes. Cependant, à des pH alcalins, le Cd peut également être sorbé de façon spécifique sur les groupements hydroxyles de bordures des feuillets d'argiles présentant des charges variables (Gu and Evans, 2007; Schindler et al., 1987; Serrano et al., 2009; Srivastava et al., 2005; Weng et al., 2001).

Les oxydes métalliques vont principalement sorber le Cd par adsorption spécifique entre les groupements hydroxyles de surface et le métal, avec seulement une faible fraction de Cd sorbé par adsorption non spécifique (Bradl, 2004), tandis que les groupements réactifs de la matière organique du sol (MOS) vont à la fois pouvoir complexer spécifiquement et non spécifiquement le Cd (Riffaldi and Levi-Minzi, 1975).

La complexation en solution est aussi un processus important influant la sorption du Cd. En effet, plus elle est importante, plus le Cd<sup>2+</sup> va se désorber de la phase solide en réponse à la baisse de Cd<sup>2+</sup> en solution générée par la complexation. Une présence importante de ligand en solution (incluant la matière organique soluble) va donc contribuer à augmenter le Cd total en solution même si la concentration de Cd<sup>2+</sup> baisse. Ce mécanisme est moindre en présence de cations compétiteurs du Cd pour la complexation tels que le cuivre ou le plomb par exemple. En effet, en s'associant aux ligands à la place du Cd, ces cations compétiteurs limitent la complexation du Cd<sup>2+</sup>, la concentration en Cd<sup>2+</sup> en solution diminue alors moins, avec pour conséquence une moindre désorption à partir de la phase solide. Ces mécanismes montrent que la modélisation de la sorption doit non seulement considérer la nature et l'abondance des phases solides porteuses, mais également la composition de la solution du sol qui régit la complexation.

Le pH est le paramètre le plus influent pour la sorption du Cd dans les sols (Buekers et al., 2007; Degryse et al., 2009), via son effet sur la protonation des groupements de surface (Naidu et al., 1994). La force ionique, et donc la composition de la solution de sol, ainsi que la concentration de Cd total dans le sol sont également attendues comme des paramètres influençant la quantité de Cd sorbé et le type de sorption (Collins et al., 2003; Naidu et al., 1994).

Degryse et al., (2009) ont résumé les modélisations possibles de la distribution sol-solution en considérant des fractions de Cd du sol ou de la solution de sol plus ou moins mobiles. Ces modélisations vont d'une vision simplifiée de la distribution, en considérant les teneurs et

concentrations totales du Cd du sol et de la solution du sol à une vision plus complète, mais également plus complexe qui considère les cinétiques de complexation en solution et d'adsorption sur les différents constituants du sol. Le compartiment mobile de Cd est défini comme la fraction de Cd du sol en équilibre avec le Cd présent en solution et donc susceptible de s'échanger réversiblement avec le Cd en solution. Afin d'estimer ce compartiment mobile, il existe deux principales méthodes. La première, la dilution isotopique, repose sur l'utilisation d'un traceur isotopique pour suivre l'équilibre dynamique du Cd entre la phase solide et la solution pour un laps de temps donné (Hamon et al., 2002). Cette technique apparaît comme conceptuellement la plus attractive, mais reste encore difficilement applicable en routine d'analyse de terre. La seconde méthode repose sur l'utilisation de solutions d'extraction permettant d'extraire différentes fractions du Cd qui sont cependant liées à la nature de l'extractant chimique choisi. Les extractions chimiques peuvent être considérées comme plus ou moins sélectives, suivant que la nature de l'extractant permet de récupérer plus ou moins spécifiquement le Cd lié à un type de phases réactives (par exemple l'extraction  $\text{CaCl}_2$  qui englobe la fraction échangeable et soluble de Cd). En combinant de manière séquentielle les extractants chimiques et en s'appuyant sur leurs propriétés de spécificité, il est possible de déduire une estimation de la répartition du Cd sur les différentes phases porteuses : échanges ions, complexation avec la MOS, les oxydes, sorption sur les carbonates (Tessier et al., 1979). Le principal avantage de l'estimation du compartiment mobile par extraction chimique, qui est aussi une sérieuse limite, est sa simplicité. En effet, les compartiments de Cd extraits sont difficilement interprétables d'un point de vue mécaniste, car l'action des solutions d'extractions n'est pas totalement spécifique (Degryse et al., 2009).

Pour décrire la distribution sol-solution du Cd, différentes approches existent : (1) l'établissement d'isothermes de sorption-désorption (Elzinga et al., 1999; Horn et al., 2004), (2) l'estimation de coefficients de partition définis comme le rapport entre la teneur en Cd total ou estimé mobile du sol et la concentration en solution pour un rapport volume:masse donné, (3) l'approche statistique qui prédit la concentration en Cd libre ( $\text{Cd}^{2+}$ ) ou total en solution à partir de variables explicatives pouvant être représentatives de mécanismes de sorption, de phases sorbantes ou d'espèces chimiques en solution, impliquées dans la mobilisation du Cd (de Vries et al., 2011; Groenenberg et al., 2010b; Ivezic et al., 2012; Römkens et al., 2004; Tipping et al., 2003) et (4) les modèles de spéciation (WHAM, Visual MINTEQ, ORCHESTRA, etc.) s'appuyant sur différents modèles d'adsorption décrivant la physico-chimie des surfaces du sol (Benedetti et al., 1995; Hiemstra and van Riemsdijk, 1996; Tipping et al., 2011; Tonkin et al., 2004...).

L'approche statistique est une méthode empirique reconnue pour sa simplicité. En effet, cette méthode s'appuie sur peu d'hypothèses et requiert une faible quantité (quelques dizaines) de données d'entrées pour être paramétrée avec succès (Groenenberg et al., 2010b). Les variables utilisées pour paramétrer ces modèles peuvent aller du pH du sol uniquement à une combinaison

de variables synthétisant les différentes phases porteuses du sol, incluant la CEC, les teneurs en carbone organique du sol ou les oxydes et les concentrations en carbone organique dissous en solution pour rendre compte de la complexation en solution (de Vries et al., 2011; Groenenberg et al., 2010b; Ivezić et al., 2012; Tipping et al., 2003). La principale faiblesse de cette approche est sa robustesse, c'est-à-dire la capacité du modèle à être utilisé dans un contexte autre que celui qui a servi au paramétrage. Les modèles développés à partir d'une large gamme de sols, variant dans leurs textures, leurs pH, leurs teneurs en MOS, en oxydes métalliques et en argiles minéralogiques constituent donc a priori les modèles les plus robustes. Malgré le développement récent de modèles statistiques génériques par Groenenberg et al. (2010b), en utilisant de larges bases de données de sols provenant du Royaume-Uni et des Pays-Bas, certains sols sont encore sous représentés, comme les sols alcalins, ou les sols riches en oxydes métalliques ou pauvres en MOS. Les sols agricoles français peuvent faire partie de ces catégories de sols sous représentés dans la bibliographie, ce qui souligne la nécessité de développer des modèles de prédictions des concentrations en Cd en solution (total ou libre) pour ces contextes particuliers.

Le développement de modèles plus mécanistes décrivant la complexation de surface a permis l'étude des interactions complexes entre les ions et les différents groupements fonctionnels de surface impliqués dans la sorption (Benedetti et al., 1995; Hiemstra and van Riemsdijk, 1996; Tipping and Hurley, 1992; Tonkin et al., 2004). L'assemblage de ces modèles unisurface en modèles multisurfaces permet donc de considérer la sorption dans son ensemble sur les principales phases réactives du sol en prenant en compte la compétition avec les autres éléments présents et les affinités variables du métal avec les surfaces. Ces modèles ont été largement utilisés pour étudier la distribution du Cd entre les différentes phases du sol (Groenenberg and Lofts, 2014), avec la limite couramment reportée de non-additivité de cette approche de modélisation, qui en considérant le sol comme une multitude de phases distinctes de propriétés uniques, ne prend pas en compte les interactions entre ces constituants (Alessi and Fein, 2010). La seconde limite est l'influence des paramètres d'entrées et le choix des phases considérées sur les concentrations en Cd<sup>2+</sup> prédites pour la solution de sol (Groenenberg and Lofts, 2014; Weng et al., 2001).

**Ainsi l'objectif de ce chapitre** est de modéliser la distribution du Cd en sols agricoles à partir de variables du sol et de sa solution, facilement mesurables à partir de deux approches complémentaires. **Dans une première approche, nous avons établi des modèles statistiques prédictifs des concentrations en Cd total ou libre (Cd<sup>2+</sup>)** à partir d'un nombre étendu de variables candidates. Ainsi, en plus des propriétés du sol généralement retrouvées dans la bibliographie (pH, MOS, carbone organique dissout [COD], capacité d'échange cationique [CEC], teneur en Cd total et teneur en Cd mobile mesurée par extraction HNO<sub>3</sub> ou EDTA), nous avons testé le Cd extrait par des sels neutres (CaCl<sub>2</sub>, NH<sub>4</sub>NO<sub>3</sub>) qui est a priori plus représentatif de la fraction échangeable. Nous avons également utilisé les teneurs en CaCO<sub>3</sub>, en oxydes métalliques amorphes et les

concentrations en Ca, Mg en solution en tant que compétiteurs majeurs vis-à-vis du Cd pour la sorption et la complexation (après le proton). Ces variables permettent de prendre en considération différents compartiments de Cd mobiles, les processus de sorption non spécifique, de complexation de surface et en solution ainsi que les compétitions. L'approche statistique minimalise les hypothèses de modélisation (un maximum de variables explicatives est testé) et le modèle est construit pour minimiser l'écart par rapport aux valeurs mesurées. Dans une seconde approche, nous avons mis en œuvre l'utilisation de la modélisation mécaniste multisurfaces en considérant différents compartiments de Cd mobile en lien avec différentes phases réactives (MOS, argiles minéralogiques et oxydes métalliques). Les modèles ont été examinés par rapport à la qualité de prédiction des concentrations en  $Cd^{2+}$ , ce qui a permis d'inférer certaines conclusions quant aux mécanismes majeurs régissant la mobilité du Cd dans les sols agricoles étudiés. L'approche de modélisation mécaniste géochimique est centrée sur les hypothèses faites quant aux mécanismes du transfert sol-solution du Cd. C'est donc une approche radicalement différente de la modélisation statistique centrée sur la minimisation de l'erreur. S'il y a une certaine cohérence au sujet des mécanismes suggérés par les modèles statistiques et ceux qui produisent les meilleures prédictions par la modélisation géochimique, la robustesse des modèles statistiques et géochimiques s'en trouve renforcée. Appuyer des modèles statistiques sur des mécanismes inférés à partir de la modélisation géochimique était également l'objectif de cette partie de notre travail.

## **2. Matériels et méthodes**

### **2.1. Prélèvements des sols**

Les sols ont été sélectionnés en fonction de leurs caractéristiques pédogéochimiques afin de couvrir une large gamme de sols utilisés en agriculture. Il s'agit, en partie, des mêmes sols que ceux utilisés dans le chapitre précédent. Ils proviennent de sols agricoles répartis dans toute la France et ont été obtenus au cours de différents projets de recherche : 25 sols prélevés en 1996 (ASPITET, Baize, 1997); 29 sols prélevés en 2011, 2013, 2015 (QUASAPROVE); 5 sols prélevés en 2016 (QUALIAGRO); 46 sols prélevés par Arvalis Institut du Végétal en 2013 et 2016 (CaDON, Cadur). Soit un total de 105 sols.

### **2.2. Analyses des sols**

Les échantillons de sols ont été séchés à l'air puis tamisés à 2 mm. La teneur en eau résiduelle a été déterminée après séchage à 105 °C jusqu'à une masse constante (NF ISO 11465:1994). Les propriétés suivantes ont été mesurées sur les échantillons de sols séchés à l'air en accord avec les normes françaises. Le pH du sol a été mesuré dans l'eau distillée (ratio sol:solution de 1:5, NF ISO 10390:2005) et dans un extrait à 0.01 M  $CaCl_2$  (ratio sol:solution de 1:10, Houba et al., 1990).

Les teneurs totales en  $\text{CaCO}_3$  ont été quantifiées par la méthode de neutralisation acide (NF X 31-105). La fraction de matière organique du sol (MOS) a été mesurée par combustion sèche après correction du  $\text{CaCO}_3$  (NF ISO 10694:1995). La texture du sol, et plus précisément les teneurs en argiles, limons fins, limons grossiers, sables fins et sables grossiers, a été déterminée par la méthode de la pipette dite de Robinson (NF X 31-107). La capacité d'échange cationique a été estimée par extraction au chlorure de cobalthexamine (NF X 31-130) au pH du sol. Les teneurs en oxydes métalliques amorphes (AlO, MnO, FeO) ont été estimées sur 39 sols par extraction au réactif de Tamm à l'obscurité (mélange équimolaire, 0.2 M, d'acide oxalique et d'oxalate d'ammonium) à un rapport volume de solution d'extraction sur masse de sol de 40.

Toutes ces analyses ont été effectuées par le Laboratoire INRA d'Analyses des Sols d'Arras (<http://www6.lille.inra.fr/las>).

### ***2.3. Cd extrait par différentes solutions d'extraction***

Les teneurs en Cd total du sol ont été quantifiées après dissolution complète du sol dans une solution constituée d'un mélange d'acides fluorhydrique et perchlorique (NF X 31-147) par ICP-MS (inductively coupled plasma - atomic emission spectrometry). Sur le même extrait, les teneurs en Fe et Mn totaux ont également été dosées, par ICP-AES (inductively coupled plasma – atomic emission spectrometry).

Cinq extractions différentes ont été effectuées de manière indépendante et donc non séquentielle sur sols séchés à l'air et tamisés à 2 mm. Elles sont décrites dans le Tableau III.1. À la fin des extractions, les suspensions de sol ont été centrifugées à 10395 g à 20 °C pendant 15 min et les surnageants ont été filtrés à 0.2  $\mu\text{m}$  et acidifiés à 2 % en volume par de l'acide nitrique. Pour chaque extractant chimique, trois répétitions ont été effectuées pour chaque sol. Les concentrations en Cd ont été déterminées dans chaque extrait par ICP-MS et le Fe, Al et Mn dans les extraits Tamm par ICP-AES.

Toutes les extractions chimiques n'ont pas pu être réalisées sur tous les sols sélectionnés. Ainsi, les extractions à l'eau et par une solution à 1 M  $\text{NH}_4\text{NO}_3$  ont été réalisées sur 57 sols, et les extractions aux réactifs de Tamm et Pyrophosphate sur 20 sols.

Nous avons retenu ces différentes extractions chimiques en fonction des compartiments de Cd mobile supposés extraits. L'eau ultra-pure (UP) a été utilisée pour estimer les compartiments de Cd solubilisables tandis que les solutions salines de  $\text{CaCl}_2$  et de  $\text{NH}_4\text{NO}_3$  sont supposées comprendre en plus le Cd complexé par adsorption non spécifique (échange d'ions). L'extraction Tamm a été réalisée pour estimer de façon additionnelle la fraction de Cd complexé aux minéraux secondaires amorphes dissous. Un tel extractant est supposé inclure également les compartiments de Cd plus faiblement liés aux constituants du sol. De la même façon, le pyrophosphate de sodium, va par chélation, estimer la part de Cd complexé à la MOS ainsi que les fractions du Cd plus faiblement lié.

Tableau III.1 : Résumé des méthodes d'extraction pour évaluer la fraction de Cd mobile des sols.

Solution d'extraction	Référence	Ratio sol/solution (m [g] : v [ml])	Durée d'agitation (h)
Extraction totale par HF/HClO <sub>4</sub>	NF X 31-147	-	-
0.1 M Na <sub>4</sub> P <sub>2</sub> O <sub>7</sub> (Pyrophosphate) à pH 10	Loveland and Digby, (1984)	1 : 100	16
0.2 M acide oxalique + 0.2 M d'oxalate d'ammonium (Réactif de Tamm) à pH 3 à l'obscurité	McKeague and Day, (1966)	1 : 40	4
1 M NH <sub>4</sub> NO <sub>3</sub> à pH 5	Symeonides and McRae, (1977)	1 : 2.5	1
0.01 M CaCl <sub>2</sub> au pH du sol	Houba et al., (1990 b)	1 : 10	2
Eau Ultra-Pure (UP) au pH du sol	Reportée par Lebourg et al., (1996)	1 : 5	1

#### 2.4. Récupération de la solution de sol

Environ 25 g de sol séché à l'air ont été réhumidifiés à pF 1.7, assimilé à la teneur en eau légèrement supérieure à la capacité de rétention maximale en eau du sol. Ces sols ont été ensuite laissés à équilibrer pendant 24 h à une température de 20 °C. La solution de sol a été récupérée par déplacement grâce à l'ajout d'environ 10 g de solvant non miscible à l'eau (C<sub>2</sub>F<sub>3</sub>Cl<sub>3</sub>) avant d'être immédiatement centrifugée à 26500 g pendant une heure. La solution de sol a ensuite été filtrée à 0.2 µm. Le protocole de cette méthode est fourni Annexe 1.

Les propriétés chimiques suivantes ont été caractérisées sur les solutions de sol : pH, concentrations en Ca et Mg total (notées Ca<sub>ss</sub> et Mg<sub>ss</sub>) mesurées par spectrométrie d'absorption atomique en flamme, concentrations en Cd et Cu total (notées Cd<sub>ss</sub> et Cu<sub>ss</sub>) déterminées par spectrométrie d'absorption atomique en four et fraction libre de Cd (notée f<sub>Cd</sub>) estimée via une méthode de résine échangeuse d'ions (méthode décrite Chapitre I.5.2 et protocole fourni Annexe 2) qui permet le calcul des concentrations en Cd libre (notée Cd<sup>2+</sup>). La composition en anions des solutions de sol (Cl<sup>-</sup>, NO<sub>3</sub><sup>-</sup>, SO<sub>4</sub><sup>2-</sup>) a été mesurée par chromatographie ionique et les concentrations en COD et carbone inorganique dissout (CID) par combustion oxydative catalytique.

Trois répétitions ont été effectuées pour chaque sol et les résultats obtenus pour les répétitions ont été moyennés.



## 2.5. Modélisation de la concentration en $Cd^{2+}$ de la solution de sol par régressions linéaires multiples

Le logarithme des concentrations en  $Cd^{2+}$  de la solution du sol résultant de la sorption et de la complexation en solution a été modélisé par régressions linéaires multiples en fonction du logarithme des variables relatives aux propriétés du sol et de la solution de sol en tant que régresseurs candidats.

Les variables explicatives candidates ont été classées dans différentes catégories suivant leur rôle sur les concentrations en  $Cd^{2+}$  en solution. Cela permet à la fois de minimiser a priori la redondance de certains régresseurs dans les modèles (et donc la multicollinéarité), mais aussi de tester des modèles qui ont un sens par rapport aux connaissances du fonctionnement du sol. Ainsi, 6 groupes de variables explicatives ont été créés et sont représentés sur l'équation générique suivante :

$$\log_{10}(Cd^{2+}) = f(a + b \log_{10} \begin{Bmatrix} Cd_{H_2O} \\ Cd_{CaCl_2} \\ Cd_{NH_4} \\ Cd_{Tamm} \\ Cd_{Pyro} \\ Cd_{tot} \end{Bmatrix} + c \begin{Bmatrix} \emptyset \\ pH_{sol} \\ pH_{CaCl_2} \\ pH_{ss} \\ \log_{10}(Calcaire) \end{Bmatrix} + d \log_{10} \begin{Bmatrix} \emptyset \\ Argile \\ ArgLF \\ CEC \end{Bmatrix} + e \log_{10} \begin{Bmatrix} \emptyset \\ MOS \\ FeO \\ AlO \\ MnO \end{Bmatrix} + f \log_{10} \begin{Bmatrix} \emptyset \\ Ca_{ss} \\ Mg_{ss} \\ Cu_{ss} \\ Fe_t \\ Mn_t \end{Bmatrix} + g \log_{10} \begin{Bmatrix} \emptyset \\ COD \\ CO_{3ss} \\ NO_{3ss} \\ SO_{4ss} \\ Cl_{ss} \end{Bmatrix})$$

La construction des modèles s'est faite de façon hiérarchique. La première catégorie de régresseurs testée représente la teneur en Cd total ou mobile du sol estimée par extractions chimiques (Tableau III.1). Cette catégorie de régresseurs est imposée dans la régression, c'est-à-dire que tous les modèles testés ont comporté un régresseur appartenant à ce groupe. Ensuite des modèles à deux, trois et quatre régresseurs ont été testés en sélectionnant une variable dans une, deux ou trois des autres groupes de l'équation générique. Les modèles étant limités à 4 régresseurs au maximum, le symbole  $\emptyset$  correspond au cas où aucun régresseur du groupe n'est sélectionné pour créer le modèle. Ces groupes de régresseurs sont : (i) l'acidité/alcalinité du sol (pH, calcaire), (ii) la capacité de sorption non spécifique (argiles, CEC, avec une variable *ArgLF* correspondant à la fraction granulométrique inférieure à 20  $\mu m$  des argiles + limons fins), (iii) la capacité de sorption spécifique (MOS, FeO, AlO, MnO), (iv) la complexation en solution (COD,  $CO_{3ss}$ ,  $NO_{3ss}$ ,  $Cl_{ss}$ ,  $SO_{4ss}$ ) et (v) les cations en solution impliqués dans la compétition pour la sorption et la complexation ( $Ca_{ss}$ ,  $Mg_{ss}$ ,  $Cu_{ss}$ ,  $Fe_t$ ,  $Mn_t$ ). Il est à noter que les teneurs en Fe et Mn total du sol ( $Fe_t$  et  $Mn_t$ ) ont été utilisées en tant que proxy des concentrations en Fe et Mn de la solution de sol. Au total, 5520 modèles différents ont été testés.

Les modèles ont été évalués à l'aide des coefficients de détermination de la régression linéaire multiple (noté  $r^2_{aj}$ ), de la régression entre valeurs observées et valeurs prédites log-transformées

(noté  $r^2_{\text{obs/pred}}$ ), du critère d'information bayésien (BIC) pour pénaliser le surparamétrage des modèles et du facteur d'inflation de la variance (VIF) permettant de détecter des cas de colinéarité entre régresseurs.

La valeur prédictive des meilleurs modèles sélectionnés a été testée par validation croisée selon la méthode *k-folds*, 10 sous-ensembles de taille équivalente sont créés, et chacun des sous-jeux sert successivement à tester le modèle qui aura été paramétré sur les 9 autres. Cette action est répétée 100 fois. L'adéquation entre valeurs observées et prédites log-transformées a été évaluée par régression linéaire.

Les modèles ont également été évalués par rapport à leurs erreurs relatives de prédictions, définies par le rapport des valeurs prédites sur les valeurs observées en valeurs naturelles (non transformées) (Équation III.1).

$$Erreur\ relative = \frac{[Cd^{2+}]_{prédites}}{[Cd^{2+}]_{observée}} \quad \text{Équation III.1}$$

## 2.6. Modélisation mécaniste de la distribution du Cd

Les concentrations en  $Cd^{2+}$  en solution ont été modélisées via un modèle géochimique multisurfaces WHAM 7 (Tipping et al., 2011) basé sur la mise en équilibre d'un système formé d'un compartiment solide constitué de différentes surfaces de réactivité et dont la taille (quantité de Cd sorbé) est estimée par extraction chimique, et d'un compartiment liquide composé de substances impliquées dans la spéciation du Cd en solution.

### 2.6.1. Compartiment en solution

Le modèle de complexation en solution inclus dans WHAM v7 a été utilisé pour modéliser le compartiment liquide du sol. Les concentrations en calcium ( $Ca_{ss}$ ), magnésium ( $Mg_{ss}$ ), cuivre ( $Cu_{ss}$ ), carbonates ( $CO_{3ss}$ ), chlorures ( $Cl_{ss}$ ), nitrates ( $NO_{3ss}$ ) et sulfates ( $SO_{4ss}$ ) et COD mesurées dans la solution de sol ainsi que son pH ont été utilisés comme données d'entrée pour le calcul de la spéciation du Cd en solution. Le modèle conceptuel WHAM VII a été utilisé pour décrire les interactions entre la MOD et les autres éléments présents en solution. La MOD a été supposée constituée de 50 % en masse de carbone et uniquement par des acides fulviques (AF), conformément aux résultats obtenus dans le Chapitre II.

La partie de MOD n'étant pas des substances humiques, a été assimilée à de l'acide acétique. La base de données générique de WHAM v7 a été utilisée pour les calculs de spéciations inorganiques et organiques (pour la complexation avec l'acide acétique). Les paramètres relatifs à la réactivité

des acides fulviques et à l'affinité des éléments pour ces acides n'ont pas été modifiés (Tipping et al., 2011).

### 2.6.2. Test de différentes surfaces réactives du sol

Afin de prédire la distribution du Cd entre la solution et la phase solide du sol, sans considérer une éventuelle dissolution ou formation de phases carbonatées, nous avons testé différentes hypothèses, résumées Figure III.1. Les hypothèses sont construites selon un modèle d'assemblage dans lequel les différents sous-modèles d'échange d'ions et de complexation de surface ont été progressivement additionnés en suivant la logique suivante :

- Hypothèse 1 : Échange d'ions
- Hypothèse 2 : Échange d'ions + Complexation par la MOS
- Hypothèse 3 : Échange d'ions + Complexation par un oxyhydroxyde amorphe (FeO **ou** AlO **ou** MnO)
- Hypothèse 4 : Échange d'ions + Complexation par la MOS + Complexation par FeO **ou** AlO **ou** MnO
- Hypothèse 5 : Échange d'ions + Complexation par la MOS + FeO + AlO + MnO

Avec : MOS (teneur du sol en matière organique), FeO (teneur du sol en oxyhydroxyde de fer), AlO (teneur du sol oxyhydroxyde d'aluminium), MnO (teneur du sol en oxyhydroxyde de manganèse).

Pour modéliser le compartiment de Cd mobile du sol pouvant s'échanger avec la solution, nous avons testé les teneurs en Cd issues des différentes extractions chimiques (Tableau III.1).

L'objectif est donc d'identifier la meilleure combinaison de surfaces réactives (Hypothèse 1 à 5), avec l'estimation du compartiment de Cd mobile (extractions chimiques) pour prédire les concentrations de Cd<sup>2+</sup> en solution.

#### 2.6.2.a. Modélisation de l'échange d'ions

Le modèle d'échange d'ions utilisé pour estimer la fraction de Cd sorbé de façon non spécifique aux surfaces d'échanges (accumulation électrostatique) est celui décrit à partir des équations de Donnan (Tipping, 1994) où l'accumulation de composés neutres et ioniques à la surface de la matrice solide du sol n'est dépendante que de la charge et du nombre de sites réactifs.

La modélisation de la surface échangeuse d'ions s'est faite de deux façons :

- a) en ne faisant aucune hypothèse explicite sur la nature des constituants et en travaillant sur une quantité globale de charges qui a été assimilée à la capacité d'échange cationique estimée par l'extraction par la cobaltihexamine (noté « CEC mesurée » dans la Figure III.1) ;

- b) en considérant que la surface échangeuse d'ions est constituée d'argiles et de matières organiques dont les teneurs fluctuent d'un sol à l'autre (noté « Argiles » dans la Figure III.1). Suivant leur nature minéralogique, les argiles ont une capacité d'échange d'ions variable. Or, nous ne disposons que de la teneur en argiles granulométriques. Nous avons donc utilisé la carte décrivant pour la France le type d'argiles minéralogiques dominant dans les horizons supérieurs suivant la localisation géographique (Gis Sol, 2011). À partir des coordonnées GPS des sols étudiés, nous avons pu attribuer à chacun d'eux un des 5 groupes dominants :
  - 1) Kaolinites avec une capacité d'échange fixée à  $0.1 \text{ mol}^+ \cdot \text{kg}^{-1}$
  - 2) Kaolinite et illites avec une capacité d'échange fixée à  $0.15 \text{ mol}^+ \cdot \text{kg}^{-1}$
  - 3) Illites, vermiculites, et chlorites avec une capacité d'échange fixée à  $0.2 \text{ mol}^+ \cdot \text{kg}^{-1}$
  - 4) Smectites et illites avec une capacité d'échange fixée à  $0.3 \text{ mol}^+ \cdot \text{kg}^{-1}$
  - 5) Smectites avec une capacité d'échange fixée à  $0.4 \text{ mol}^+ \cdot \text{kg}^{-1}$

### 2.6.2.b. Complexation de surface par la matière organique solide

La sorption du métal à la MOS a été décrite par le modèle de complexation spécifique de surface WHAM VII (Tipping et al., 2011). La MOS réactive a été supposée être constituée uniquement d'AH (Dijkstra et al., 2004; Weng et al., 2001). Le nombre de sites de complexation par les AH a été estimé à partir de la contribution de la MOS à la CEC du sol. Si l'on considère que la CEC totale du sol est principalement due aux argiles et la MOS, il est possible de différencier la charge portée par les MOS et celle portée par les argiles (Weng et al., 2001) :

La capacité d'échange cationique liée à la matière organique ( $CEC_{MOS}$ , en  $\text{mol}^+ \cdot \text{kg}^{-1}$  de sol) s'obtient à partir de la CEC totale mesurée :

$$CEC_{MOS} = CEC_{tot} - T_{Arg} \times CEC_{arg} \quad \text{Équation III.2}$$

Avec la  $CEC_{MOS}$ , la CEC attribuée à la MOS,  $CEC_{tot}$ , correspondant à la CEC estimée par la cobaltihexamine (en  $\text{mol}^+ \cdot \text{kg}^{-1}$  de sol), et  $CEC_{Arg}$ , étant la CEC du groupe d'argile du sol (en  $\text{mol}^+ \cdot \text{kg}^{-1}$  d'argile), et  $T_{Arg}$ , la teneur du sol en argile granulométrique (en  $\text{kg}$  d'argile  $\cdot \text{kg}^{-1}$  de sol).

Le modèle WHAM accepte comme variable d'entrée la teneur du sol en acides humiques réactifs ( $T_{AH}$ ) et non la CEC de la MOS. Pour une densité générique de sites des AH ( $S_{AH}$ ) de  $5.1 \text{ mol}^+ \cdot \text{g}^{-1}$  d'AH (Tipping et al., 2011), la teneur en AH,  $T_{AH}$ , en  $\text{g}$  AH  $\cdot \text{kg}^{-1}$  de sol est égale à :

$$T_{AH} = \frac{CEC_{MOS}}{S_{AH}} \quad \text{Équation III.3}$$

Le calcul montre que l'estimation de la teneur en MOS active ( $T_{AH}$ ) est conditionnée par une hypothèse concernant la charge globale des argiles ( $CEC_{arg}$ ).

Le reste des paramètres relatifs à la complexation du Cd par les AH n'a pas été modifié. En effet, ces paramètres ont été largement éprouvés dans la littérature et ont fourni des performances correctes sur une large gamme de conditions environnementales (Tipping et al., 2011).

### **2.6.2.c. Complexation de surface par les oxydes métalliques**

La sorption du Cd aux oxydes métalliques amorphes a été modélisée par le modèle SCAMP implémenté dans WHAM v7. Les teneurs en oxydes de fer, d'aluminium et de manganèse amorphes ont été estimées à partir des teneurs mesurées après extraction par le réactif de Tamm (Oxalate d'ammonium et acide oxalique), dont le protocole a été détaillé Tableau III.1.

Les paramètres génériques ont été utilisés pour le paramétrage de ces surfaces. Le sous-modèle relatif aux oxydes de fer suppose une surface spécifique de l'oxyde égale à  $600 \text{ m}^2.\text{g}^{-1}$  et égale à  $400 \text{ m}^2.\text{g}^{-1}$  pour les oxydes d'aluminium et de manganèse (Tipping et al., 2011).

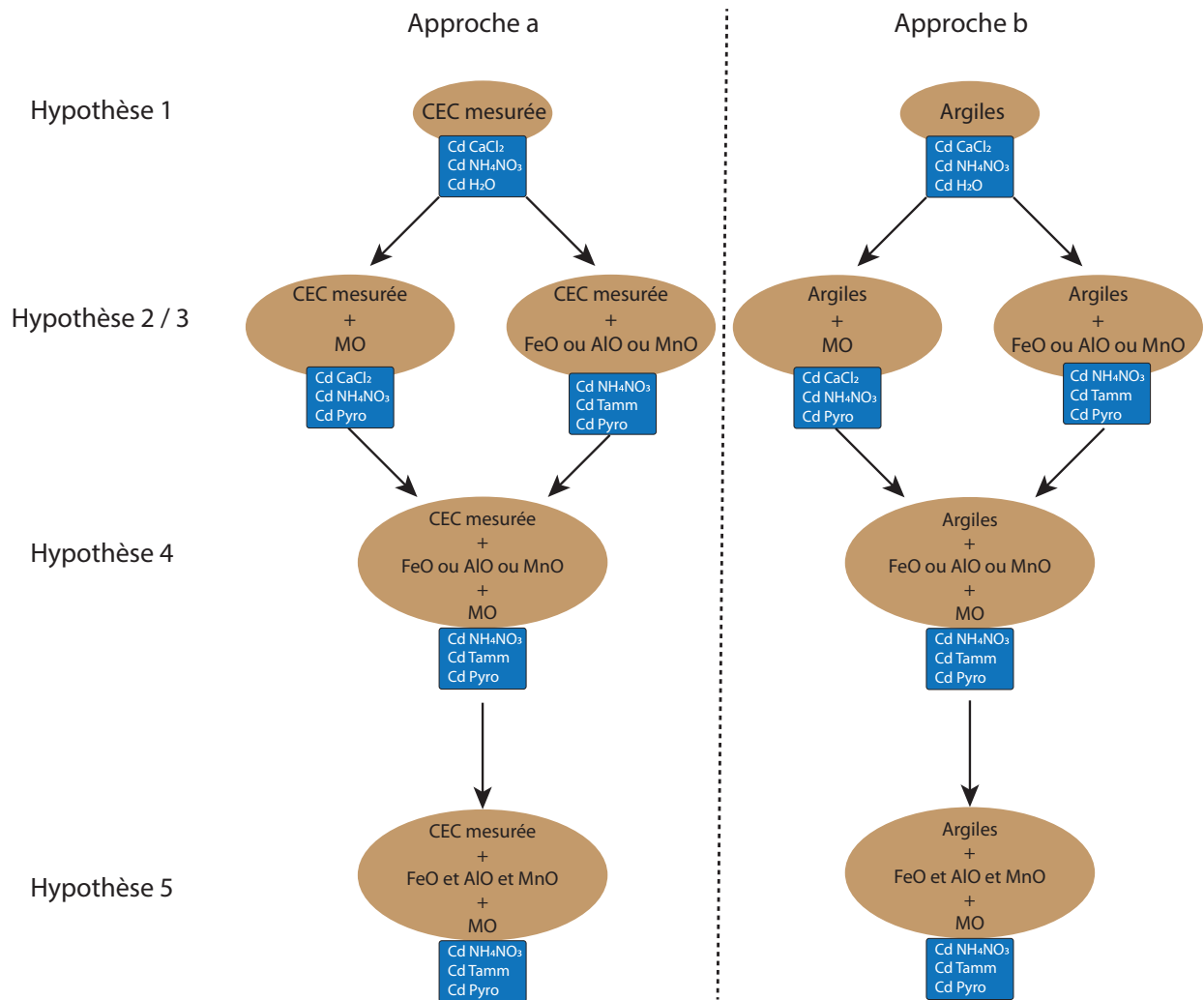


Figure III.1 : Schéma récapitulatif des hypothèses faites pour la modélisation du compartiment solide du sol, surfaces de complexation et compartiments de Cd mobiles pour prédire les concentrations en Cd<sup>2+</sup> en solution.

### 2.6.3. Évaluation des hypothèses testées par le modèle multisurfaces

La modélisation multisurfaces a été évaluée sur sa capacité à prédire le Cd<sup>2+</sup> en solution. Les prédictions issues des différentes hypothèses de modélisation (Figure III.1) ont été évaluées par différents critères. Nous avons tout d'abord examiné la qualité d'ajustement, via le coefficient de détermination  $r^2_{\text{obs}/\text{pred}}$  de la régression des valeurs log-transformées observées en fonction des valeurs log-transformées prédites. Ce critère permet d'évaluer la précision du modèle. La pente et l'ordonnée à l'origine de la droite de régression ont permis de détecter des biais éventuels (pente différente de 1 et ordonnée à l'origine différente de 0, évaluées par t-test). Les erreurs relatives de prédiction de la concentration en Cd<sup>2+</sup> par modélisation géochimique multisurfaces, ont été calculées par le rapport des concentrations prédites sur les concentrations mesurées (Équation III.1). Ceci permet de prendre en considération la capacité du modèle à sous-estimer ou surestimer la sorption

du Cd. Les valeurs d'erreurs relatives supérieures à 1 signifient une surestimation de la concentration en Cd<sup>2+</sup> en solution et donc une sous-estimation de la sorption. Inversement si l'erreur relative est inférieure à 1, la sorption est surestimée.

Les erreurs relatives pour valeurs naturelles non log-transformées ont été également calculées après correction de la prédiction par le biais éventuel déduit de la droite de régression  $obs=f(\text{prédit})$ .

### 3. Résultats

#### 3.1. Caractéristiques des sols

Les valeurs minimales, médianes et maximales des propriétés physico-chimiques des sols étudiés sont données Tableau III.2.

Les sols sélectionnés montrent une large variabilité sur une grande partie des propriétés physico-chimiques mesurées. La texture des sols est majoritairement entre la classe texturale des argiles et celle des limons avec peu de sols sableux (4 sols). Les sols sont en majorité alcalins, avec un pH médian de 8.1. Les teneurs en MOS sont caractéristiques des sols agricoles et des sols généralement observés dans la bibliographie, avec une gamme de variation cependant plus élevée. Elles sont comprises entre 10.8 et 77.2 g.kg<sup>-1</sup> avec une valeur médiane de 26.2 g.kg<sup>-1</sup>. Les valeurs de CEC sont également variées avec des valeurs s'étalant entre 3.4 et 41.9 cmol<sup>+</sup>.kg<sup>-1</sup> avec une médiane de 15.6 cmol<sup>+</sup>.kg<sup>-1</sup>. Les CEC observées sont significativement plus corrélées aux teneurs en argiles granulométriques ( $r = 0.84$ ,  $p\text{-value} < 0.001$ , test de corrélation Bravais-Pearson) qu'aux teneurs en MO du sol ( $r=0.46$ ,  $p\text{-value} < 0.001$ , test de corrélation Bravais-Pearson).

Les teneurs du sol en Fe total varient de 4.9 à 78.0 g.kg<sup>-1</sup> avec une valeur médiane de 22.3 g.kg<sup>-1</sup>. Les teneurs en Mn total sont comprises entre 0.12 et 7.36 g.kg<sup>-1</sup>. Les valeurs de Fe et Mn sont généralement liées aux oxydes métalliques cristallisés et amorphes présents dans les sols. Cette variabilité montre donc une gamme large de valeurs avec certains sols riches en oxydes de manganèse et de fer permettant de tester leur influence sur la dynamique du Cd.

Tableau III.2 : Tableau récapitulatif des caractéristiques physico-chimiques des sols étudiés. N=105.

	Argiles	Limons	Sables	MOS	Calcaire	CEC	pH	pH	Fe <sub>t</sub> <sup>*</sup>	Mn <sub>t</sub> <sup>*</sup>
	%			g.kg <sup>-1</sup>		cmol <sup>+</sup> .kg <sup>-1</sup>	eau	CaCl <sub>2</sub>	g.kg <sup>-1</sup>	g.kg <sup>-1</sup>
Min.	5.3	4.2	2.7	10.8	0.0	3.4	5.1	4.7	4.9	0.12
Méd.	26.0	48.7	18.8	26.2	31.0	15.6	8.1	7.1	22.3	0.53
Max.	56.9	83.5	62.8	77.2	841.9	41.9	8.5	7.7	78.0	7.36

24 données sont manquantes sur les variables annotées d'un astérisque.

### 3.1.1. Caractéristiques chimiques des solutions de sol

De la même façon que pour les caractéristiques pédogéochimiques, les solutions de sols obtenues par déplacement assisté par centrifugation montrent une proportion plus importante de solutions de sol alcalines, avec des pH mesurés allant de 4.9 à 8.8 et une médiane à 8.2. Les concentrations en cations majeurs ( $\text{Ca}_{\text{ss}}$ ,  $\text{Mg}_{\text{ss}}$ ) sont de l'ordre du millimolaire avec des valeurs allant de 0.8 à 25.8 mM et une médiane à 3.3 mM pour le Ca et des concentrations environ dix fois moindres pour le Mg (0.08-1.44 mM). Les valeurs élevées de Ca en solution sont liées aux sols très riches en calcaire. Les concentrations en cuivre total en solution ( $\text{Cu}_{\text{ss}}$ ) s'étalent entre 0.2  $\mu\text{M}$  et 2.2  $\mu\text{M}$ . Les concentrations en COD mesurées en solution sont plus élevées que celles généralement rencontrées dans la bibliographie (de Vries et al., 2011; Sauvé et al., 2000b) et du même ordre de grandeur que celles obtenues à partir de sols organiques (Tipping et al., 2003). Ceci pourrait être lié au protocole employé pour récupérer la solution de sol et plus particulièrement au faible rapport volume/sol, de l'ordre de la capacité de rétention en eau, mais aussi au fait que les sols aient été séchés à l'air avant réhumectation (Cornu et al., 2009). Les concentrations en carbonates mesurées varient de 1.6 à 108.9  $\text{mg CO}_3^{2-} \cdot \text{l}^{-1}$  avec une valeur médiane de 38.8  $\text{mg CO}_3^{2-} \cdot \text{l}^{-1}$ . Les carbonates, présents majoritairement dans les solutions de sols alcalins, ne sont pas corrélés aux teneurs en calcaire du sol. Ceci pourrait être lié à des réactivités différentes des calcaires en fonction des sols et au fait que la majorité du calcaire ne serait pas actif. Les concentrations en anions présents en solution s'étalent également sur de larges gammes de valeurs allant pour les nitrates ( $\text{NO}_{3\text{ss}}$ ) de 0.9 à 2445.1  $\text{mg} \cdot \text{l}^{-1}$ , pour les sulfates ( $\text{SO}_{4\text{ss}}$ ) de 0.6 à 383.0  $\text{mg} \cdot \text{l}^{-1}$  et pour les chlorures ( $\text{Cl}_{\text{ss}}$ ) de 1.8 à 79.2  $\text{mg} \cdot \text{l}^{-1}$ . Les concentrations élevées de  $\text{NO}_{3\text{ss}}$  et  $\text{SO}_{4\text{ss}}$  pourraient être expliquées par un apport, peu avant le prélèvement des échantillons de sol, de fertilisants azotés qui contenaient également du soufre.

Tableau III.3 : Résumé des caractéristiques de solutions de sols étudiées. Les valeurs minimales, médianes et maximales du jeu de données complet (105 sols) sont renseignées.

	$\text{pH}_{\text{ss}}$	$\text{Ca}_{\text{ss}}$	$\text{Mg}_{\text{ss}}$	$\text{Cu}_{\text{ss}}^*$	$\text{COD}^*$	$\text{CO}_{3\text{ss}}^*$	$\text{NO}_{3\text{ss}}$	$\text{SO}_{4\text{ss}}$	$\text{Cl}_{\text{ss}}$
		$\text{mmol} \cdot \text{l}^{-1}$		$\mu\text{mol} \cdot \text{l}^{-1}$		$\text{mg} \cdot \text{l}^{-1}$			
Min.	4.9	0.8	0.08	0.2	62.9	1.6	0.9	0.6	1.8
Méd.	8.2	3.3	0.27	0.5	233.7	38.8	23.2	32.3	13.9
Max.	8.8	25.8	1.44	2.2	917.4	108.9	2445.1	383.0	79.2

Les variables annotées d'un astérisque ont un nombre d'échantillons différent de la taille du jeu de données total. N = 78 pour le COD et  $\text{CO}_{3\text{ss}}$ , N = 76 pour  $\text{Cu}_{\text{ss}}$



### 3.1.2. Teneurs en Cd du sol obtenues par extractions sélectives et totale et concentrations en Cd mesurées en solution de sol

Les teneurs en Cd total obtenues par dissolution complète du sol dans un mélange d'acides fluorhydrique et perchlorique varient entre 0.08 et 1.03 mg.kg<sup>-1</sup> ce qui, comme dans le chapitre précédent, ne témoigne d'aucune contamination anormale (Gis Sol, 2011).

Les teneurs en Cd du sol extrait par les différentes méthodes sont résumées Tableau III.4. Comme attendu, les teneurs en Cd extraits par l'eau ultra-pure sont les plus faibles, représentant de 0.005 % à 0.96 % du Cd total du sol. Les teneurs en Cd extrait par les sels neutres, s'échelonnent de 0.01 % à 29.5 % du Cd total pour le CaCl<sub>2</sub> et entre 0.03 % et 37.7 % pour le NH<sub>4</sub>NO<sub>3</sub> avec des valeurs médianes respectivement à 0.29 % et 0.38 %, ce qui reste encore faible. Le principal mécanisme chimique expliquant le mode d'action de ces extractants est la désorption du Cd sorbé par échange d'ions en raison de la compétition pour les sites de sorption par les cations apportés (Ca<sup>2+</sup> et NH<sub>4</sub><sup>+</sup>). Les autres mécanismes chimiques pouvant expliquer l'extraction plus forte du NH<sub>4</sub>NO<sub>3</sub> seraient la diminution modérée du pH (pH de la solution d'extraction à une valeur d'environ 5) qui favoriserait la formation de complexe Cd-amine dans des sols alcalins (Gryschko et al., 2005).

Le réactif de Tamm a été choisi pour sa capacité à dissoudre les oxydes métalliques amorphes et à relarguer le Cd sorbé sur ces constituants. La solution de contact étant à pH 3, l'action du proton peut également induire une désorption du Cd complexé à d'autres constituants. Les teneurs en Cd obtenues varient entre 0.1 et 32.5 % avec en moyenne environ 9.5 % du Cd total extrait. Le pyrophosphate de sodium permet la plus importante extraction du Cd. Les teneurs extraites vont de 1 à 75 % du Cd total du sol. En plus du Cd faiblement lié aux constituants du sol (par sorption non spécifique), cet extractant peut induire une légère dissolution des oxydes métalliques amorphes et permet surtout l'extraction du Cd complexé à la MOS. La phase organique du sol semble donc être dans notre cas plus impliquée dans la complexation du Cd que les oxydes métalliques.

Les valeurs médianes des concentrations totales en Cd total en solution sont de l'ordre du nanomolaire, comprises entre 0.25 et 13 nM. Les concentrations en Cd<sup>2+</sup> en solution varient sur trois ordres de grandeur avec des valeurs entre 0.06 et 11.10 nM représentant entre 15.2 et 85.0 % du Cd<sub>ss</sub>, ce qui est cohérent avec les concentrations généralement rencontrées en solution de sols non contaminés (de Vries et al., 2011; Sauvé et al., 2000b).

Tableau III.4 : Résumé des teneurs en Cd extrait par extractions sélectives, extraction totale et des concentrations en Cd mesurées en solution de sol. Le nombre d'échantillons disponibles (N) pour chaque extraction est indiqué entre parenthèses. Les valeurs minimales, médianes et maximales obtenues sur les N valeurs sont reportées.

	Cd H <sub>2</sub> O (N=86)	Cd CaCl <sub>2</sub> (N=105)	Cd NH <sub>4</sub> N O <sub>3</sub> (N=86)	Cd Tamm (N=47)	Cd Pyrophosphate (N=47)	Cd total (N=105)	Cd <sub>ss</sub> (N=105)	Cd <sup>2+</sup> (N=105)
	µg.kg <sup>-1</sup>					mg.kg <sup>-1</sup>	nmol.l <sup>-1</sup>	
Min.	0.01	0.03	0.09	0.19	10.00	0.091	0.25	0.06
Méd.	0.30	1.55	1.72	26.50	73.32	0.284	1.58	0.53
Max.	1.45	44.50	56.95	133.94	197.23	1.030	13.00	11.10

### 3.1.3. Teneurs en phases réactives

Dans la bibliographie, les densités de charges généralement utilisées pour les argiles minéralogiques s'étalent entre 0.1 et 0.4 mol<sup>+</sup>.kg<sup>-1</sup> d'argile (Weng et al., 2001). Bien que cela soit souvent utilisé par simplification, appliquer une valeur unique de densité de charges d'argile peut fortement influencer la CEC imputable aux argiles, mais aussi par voie de conséquence celle de la MOS (Équation III.2). C'est pourquoi nous avons envisagé que les argiles minéralogiques des sols aient une réactivité moyenne pouvant varier entre 0.1 mol<sup>+</sup>.kg<sup>-1</sup> et 0.4 mol<sup>+</sup>.kg<sup>-1</sup> (Tableau III.5) en fonction de leur nature minéralogique que nous avons estimée à partir de leur origine géographique. À partir de la CEC du sol et de la part de celle-ci imputable aux argiles, nous avons calculé une estimation de la teneur des sols en acides humiques réactifs (AH) (Équation III.3). Les teneurs en AH calculées vont de 1.45 à 34.20 g AH.kg<sup>-1</sup> de sol (Tableau III.5) ce qui correspond à des pourcentages de MOS active (ratio des teneurs en AH sur les teneurs mesurées en MOS) allant de 6 % à 69.6 %. Ces pourcentages s'étalent sur une gamme plus large que ceux généralement reportés dans la bibliographie (Ren et al., 2015a; Weng et al., 2001). Ceci peut être dû en partie à une diversité plus importante des sols utilisés dans notre étude. Cependant, certains pourcentages de MOS active estimés prennent des valeurs peu vraisemblables (supérieurs à 80 %). Pour ces échantillons (10 sols), un pourcentage plafond de 63 % de MOS active a été utilisé, ce seuil correspondant à la valeur du 9<sup>e</sup> décile des données. Ces pourcentages d'activité supérieurs à 100 % pourraient être liés à la sous-estimation de la réactivité des argiles minéralogiques (CEC<sub>arg</sub>) si celles-ci étaient supérieures à 0.4 mol<sup>+</sup>.kg<sup>-1</sup> d'argile (Équation III.2).

Les teneurs en Fe, Al et Mn, provenant de la dissolution des oxydes métalliques amorphes, mesurées dans les extraits Tamm, varient entre 0.4 et 20.7 g Fe.kg<sup>-1</sup> pour les oxydes de fer, entre 0.29 et 12.8 g Al.kg<sup>-1</sup> pour les oxydes d'aluminium et entre 0.05 et 5.40 g Mn.kg<sup>-1</sup> pour les oxydes de manganèse (Tableau III.5). La totalité du Fe, Al, Mn extrait par réactif de Tamm a été attribuée à

des oxydes métalliques amorphes, négligeant donc la part de Fe, Al et Mn sorbée aux constituants impactés par l'extraction Tamm (Fe, Al, Mn échangeables ou solubilisables) considérée comme très faible (<1 %) (Rennert et al., 2017).

*Tableau III.5 : Teneurs et réactivité des phases réactives prises en compte dans cette étude.  $CEC_{arg}$  correspond à la capacité d'échange cationique supposée de la phase argileuse. Les teneurs en AH ont été calculées selon Weng et al., (2001) (Équation III.3) et les termes FeO, AlO, MnO renvoient aux teneurs en Fe, Al et Mn mesurées dans un extrait Tamm.*

	$CEC_{arg}$	AH	FeO	AlO	MnO
	mol.kg <sup>-1</sup>		g.kg <sup>-1</sup>		
Min.	0.1	1.45	0.40	0.29	0.05
Méd.	0.3	12.15	1.43	1.26	0.30
Max.	0.4	34.20	20.73	12.78	5.40

### **3.2. Prédiction des concentrations en $Cd^{2+}$ et Cd total en solution par régressions linéaires multiples**

Les concentrations en  $Cd^{2+}$  en solution ont été modélisées en fonction des caractéristiques du sol et de la solution de sol par régressions linéaires multiples dans le but d'identifier ce qui explique au mieux les différences de partition sol-solution du  $Cd^{2+}$  et donc les variations des concentrations en  $Cd^{2+}$  mesurées en solution en fonction des sols.

Les meilleurs modèles à 4, 3, 2 et 1 régresseurs (respectivement Équation III.4, Équation III.5, Équation III.6 et Équation III.7) sont présentés Tableau III.6.

Tableau III.6 : Récapitulatif des équations obtenues pour la prédiction des concentrations en  $Cd^{2+}$  et  $Cd_{ss}$  et les paramètres d'évaluation des modèles obtenus par validation croisée.

Équation	Modèle	$r^2_{aj}$	N	VIF max	Ord. Orig. obs/pred	Pente obs/pred	$r^2$ obs/pred
Équation III.4	$\log_{10}(Cd^{2+}) = -8.40^{***} + 0.38^{***}$ $\log_{10}(Cd_{NH_4NO_3}) - 0.26^{**} \text{ pH}$ $+ 0.52^{***} \log_{10}(Ca_{ss}) - 0.15 \cdot \log_{10}(MnO)$	0.828	47	2.33	-0.26 <sup>***</sup>	0.971 <sup>***</sup>	0.793
Équation III.5	$\log_{10}(Cd^{2+}) = -10.46^{***}$ $+ 0.58^{***} \log_{10}(Cd_{NH_4NO_3})$ $+ 0.53^{***} \log_{10}(Ca_{ss}) - 0.16 \cdot \log_{10}(AlO)$	0.772	47	1.16	-0.33 <sup>***</sup>	0.964 <sup>***</sup>	0.729
Équation III.6	$\log_{10}(Cd^{2+}) = -10.35^{***}$ $+ 0.98^{***} \log_{10}(Cd_{NH_4NO_3})$ $+ 0.47^{***} \log_{10}(Ca_{ss})$	0.755	86	1.16	-0.21 <sup>***</sup>	0.976 <sup>***</sup>	0.734
Équation III.7	$\log_{10}(Cd^{2+}) = -9.32^{***}$ $+ 0.63^{***} \log_{10}(Cd_{NH_4NO_3})$	0.695	86	-	-0.17 <sup>***</sup>	0.982 <sup>***</sup>	0.681
Équation III.8	$\log_{10}(Cd_{ss}) = -8.33^{***} + 0.53^{***}$ $\log_{10}(Cd_{NH_4NO_3}) - 0.95^{***} \log_{10}(ArgLF)$ $+ 0.46^{***} \log_{10}(Ca_{ss}) - 0.12 \cdot \log_{10}(MnO)$	0.845	47	1.28	-0.32 <sup>***</sup>	0.963 <sup>***</sup>	0.809
Équation III.9	$\log_{10}(Cd_{ss}) = -9.82^{***} + 0.55^{***}$ $\log_{10}(Cd_{NH_4NO_3}) + 0.41^{***} \log_{10}(Ca_{ss})$ $- 0.27^{**} \log_{10}(AlO)$	0.787	47	1.16	-0.23 <sup>***</sup>	0.974 <sup>***</sup>	0.765
Équation III.10	$\log_{10}(Cd_{ss}) = -9.65^{***} + 0.56^{***}$ $\log_{10}(Cd_{NH_4NO_3}) + 0.34^{***} \log_{10}(Ca_{ss})$	0.721	86	1.14	-0.18 <sup>***</sup>	0.979 <sup>***</sup>	0.703
Équation III.11	$\log_{10}(Cd_{ss}) = -8.91^{***} + 0.51^{***}$ $\log_{10}(Cd_{NH_4NO_3})$	0.676	86	-	-0.13 <sup>***</sup>	0.985 <sup>***</sup>	0.663

Significativité des paramètres : \*\*\*  $p$ -value < 0.001, \*\*  $p$ -value < 0.01, \*  $p$ -value < 0.05, ·  $p$ -value < 0.1. Comparaison des ordonnées à l'origine à 0 et la pente à 1.

Le modèle le plus performant (Équation III.4) pour expliquer les concentrations en  $Cd^{2+}$  en solution explique 82.8 % de variance totale et comporte quatre variables explicatives : les teneurs en Cd obtenues par extraction au  $NH_4NO_3$  à 1 M, le pH du sol, les concentrations en Ca en solution de sol et les teneurs en oxydes de manganèse amorphes. La contribution à la variance expliquée de ces régresseurs est respectivement de 38.7 %, 28.9 %, 13.0 % et 2.2 %. La valeur maximale de VIF de

cette équation est de 2.33 témoignant d'une légère corrélation négative entre le pH du sol et les teneurs en Cd extraites par  $\text{NH}_4\text{NO}_3$ . La quantité de Cd extraite par  $\text{NH}_4\text{NO}_3$  diminue donc à mesure que le pH augmente. En supposant que  $\text{NH}_4\text{NO}_3$  extrait principalement le Cd sorbé de manière non spécifique, cela pourrait indiquer que la part relative de la sorption non spécifique par rapport à la complexation diminue lorsque le pH du sol augmente.

La Figure III.2 représente la régression log-log entre les valeurs observées et les valeurs prédites par validation croisée *k-folds* de en solution. La pente, l'ordonnée à l'origine (Ord. Orig.) et la part de variance expliquée sont également reportées Tableau III.6. Ces valeurs montrent un très léger biais dans les prédictions avec une ordonnée à l'origine significativement différente de zéro (*p-value* < 0.001) et une pente significativement différente de 1 (*p-value* < 0.001).

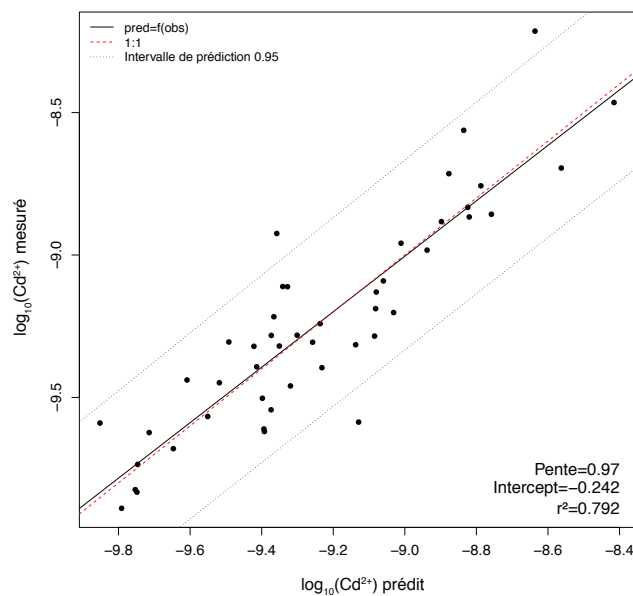


Figure III.2 : Concentrations en  $\text{Cd}^{2+}$  (log transformées) prédites par l'équation V.4, en fonction de celles mesurées. La ligne noire correspond au modèle linéaire entre valeurs observées et valeurs prédites, la ligne en pointillé rouge, l'équation  $y=x$  et les lignes vertes à l'intervalle de prédiction à 0.95).

Le meilleur modèle à 3 paramètres, comprend deux des régresseurs du modèle précédent (teneurs en Cd obtenues par extraction au  $\text{NH}_4\text{NO}_3$ , concentrations en Ca en solution) auquel s'ajoutent les teneurs en oxydes d'aluminium amorphes (Équation III.5). Ce modèle ne fait plus intervenir le pH du sol et explique un peu moins de 5 % de variance en moins par rapport au meilleur modèle. Pour ce modèle à trois variables, la contribution des oxydes d'aluminium amorphes à l'explication de la variance est faible et du même ordre de grandeur que celle des oxydes de manganèse identifiés dans le modèle à quatre variables (3 % de la variance expliquée), si bien que si ce régresseur est ôté (Équation III.6) le modèle est très peu dégradé et explique encore 75 % de la variance des concentrations en  $\text{Cd}^{2+}$  avec seulement deux variables explicatives. Le dernier modèle comporte uniquement le Cd obtenu par extraction au  $\text{NH}_4\text{NO}_3$  1 M comme variable explicative et la baisse du

$r^2_{\text{obs/pred}}$  par rapport aux modèles précédents n'est pas très importante ( $r^2_{\text{obs/pred}}=0.68$ ), ce qui illustre l'importance du compartiment de Cd extrait par le  $\text{NH}_4\text{NO}_3$  pour expliquer les concentrations en  $\text{Cd}^{2+}$  en solution.

Un biais faible, mais significatif est à noter sur ces quatre équations, lié aux ordonnées à l'origine significativement différentes de zéro ( $p\text{-value} < 0.001$ ) et aux pentes significativement différentes de 1.

Les concentrations en Cd total en solution ont également été prédites suivant la même approche (Tableau III.6, Équation III.8 à Équation III.11). Le modèle le plus performant (Équation III.8) identifie les mêmes régresseurs que l'Équation III.4 à l'exception du pH du sol qui est remplacé par la variable sommant teneurs en argiles granulométriques et en limons fins (*ArgLF*). Il faut noter la proportion élevée de la variance expliquée par le compartiment de Cd extrait par  $\text{NH}_4\text{NO}_3$  : 66.3 % comparé au 38.7 % pour l'Équation III.4 avec les régresseurs restants expliquant seulement 18.1 % de la variance totale. Cette proportion importante de variance expliquée par le compartiment de Cd  $\text{NH}_4\text{NO}_3$  est conservée sur les modèles plus simples à 3 et 2 paramètres. Il faut retenir que le compartiment de Cd extrait par  $\text{NH}_4\text{NO}_3$  demeure la variable explicative majeure devant les concentrations en Ca en solution et les oxydes d'aluminium ou de manganèse amorphes lorsqu'il s'agit de prédire la partition phase solide-solution du  $\text{Cd}^{2+}$  ou du Cd total.

La Figure III.3 regroupe la valeur absolue des erreurs relatives de tous les modèles présentés calculées à partir des validations croisées et après correction du biais éventuel de prédiction. Elles rendent compte essentiellement de l'erreur liée au paramétrage du modèle statistique (manque de précision) puisque le biais (manque de justesse) a été corrigé si nécessaire. Le modèle le plus robuste sera donc caractérisé par la plus faible variance des erreurs de prédictions lors de la validation croisée. Les distributions de ces erreurs sont représentées par des *violin plots* (Figure III.3) regroupant la distribution des erreurs, qui est représentée par le *box plot*, et la densité de probabilité des erreurs entourant le box plot. Un modèle robuste sera représenté par un *box plot* et une courbe de densité de probabilité des erreurs très étroites, centrée sur une valeur d'erreur la plus proche de 1.

En moyenne, nous observons que les concentrations en  $\text{Cd}^{2+}$  ou Cd total en solution sont mieux prédites par un modèle à 3 ou 4 régresseurs (Équation III.4, Équation III.5, Équation III.8, Équation III.9, représentés en violet et bleu sur la Figure III.3). La variance des erreurs relatives corrigées est légèrement plus faible pour la prédiction des concentrations en Cd total en solution comparé à la prédiction des concentrations en  $\text{Cd}^{2+}$ . Ce qu'il est intéressant de noter est la forme des distributions des erreurs relatives. Dans le cas des concentrations en Cd total, la courbe est de forme bimodale avec une partie des concentrations sous-estimées (rapport inférieur à 1) et une autre partie où les concentrations sont surestimées (rapport supérieur à 1). Les erreurs relatives des prédictions de

$Cd^{2+}$  sont plus de forme unimodale où le plus grand nombre d'erreurs relatives se trouve centré autour de 1. Ainsi, l'Équation III.4 semble être le meilleur modèle lié à la distribution normale centrée sur 1, valeur pour laquelle la prédiction est égale à l'observation. Pour un nombre réduit de sols (2-3 sols), ce modèle va cependant surestimer d'un facteur trois au maximum la valeur mesurée. Cependant, vu la faible différence de qualité entre les modèles à 4 et 3 variables, si l'on considère en plus comme critère de choix le principe de parcimonie qui nous oriente vers le modèle le plus économe en variables à renseigner, les modèles à 4 ou 3 variables sont proches d'être équivalents.

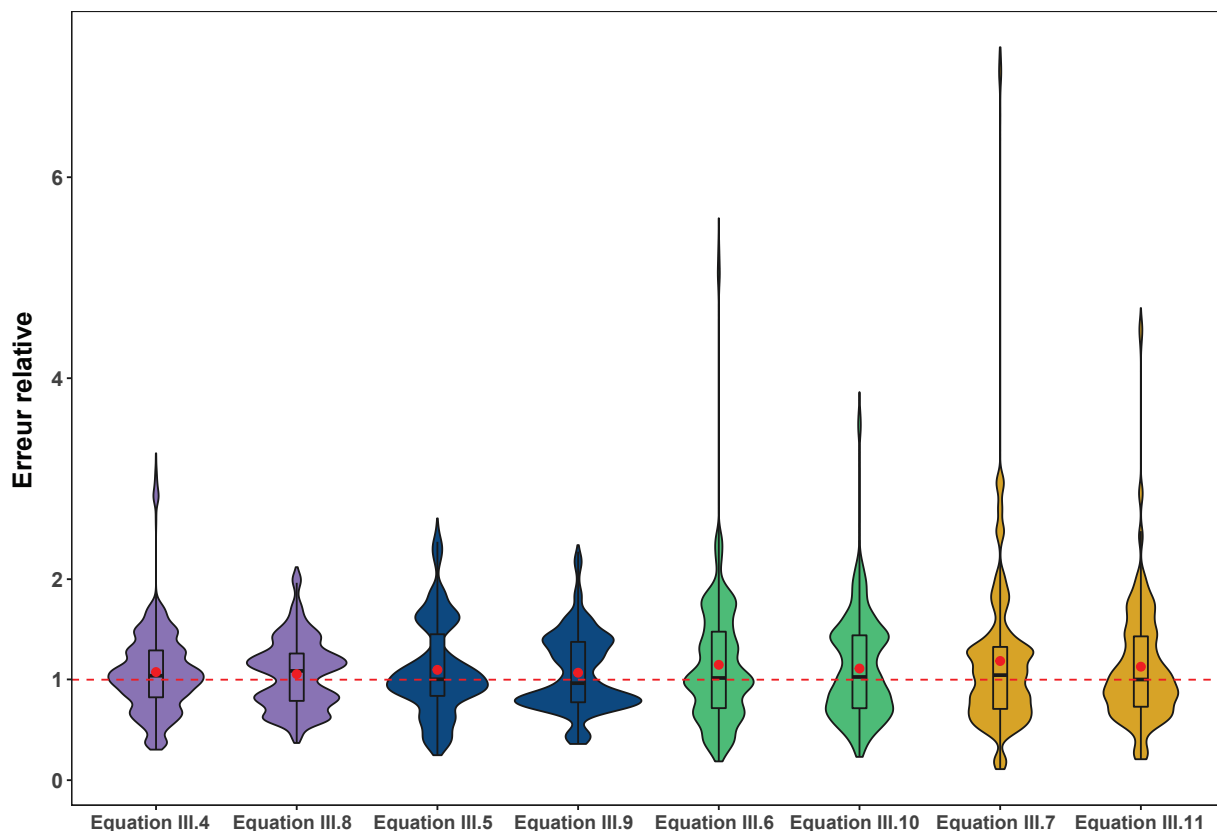


Figure III.3 : Violin plots des erreurs relatives corrigées du biais de prédiction pour les 8 meilleurs modèles statistiques à 1, 2, 3 et 4 régresseurs pour les prédictions des concentrations totales de Cd en solution et les concentrations en Cd libre ( $Cd^{2+}$ ) en solution. Le point rouge correspond à la moyenne des individus, le box plot représente l'interquartile (25-75 %) et la droite grise verticale correspond à 1.5 fois l'interquartile. La ligne en pointillé rouge représente un ratio de 1 entre valeurs prédites et valeurs observées montrant une prédiction égale à la valeur mesurée. Les ratios inférieurs à 1 indiquent que les valeurs prédites sont sous-estimées et inversement. Les couleurs distinguent le nombre de régresseurs utilisés pour la calibration des modèles : violet pour des modèles à 4 régresseurs, bleu, pour 3 régresseurs, vert pour 2 régresseurs et jaune pour les modèles à 1 régresseur.

### 3.3. Prédiction des concentrations de $Cd^{2+}$ par un modèle multisurfaces

Afin de prédire les concentrations en  $Cd^{2+}$  en solution de sol en considérant le compartiment solide du sol, nous avons également testé l'approche géochimique ou multisurfaces. Nous avons ainsi supposé que différents constituants du sol et différentes extractions chimiques pourraient permettre la prédiction de ces valeurs.

La Figure III.4 regroupe les erreurs relatives de prédiction de la totalité des hypothèses testées (résumées Figure III.1). Les box plots correspondent aux différentes hypothèses testées. Chaque hypothèse est notée de la façon suivante : hypothèse sur les phases réactives (de H1 à H5), extractant du Cd considéré et enfin hypothèse sur la sorption non spécifique (approche « a » ou « b ») :

- En rouge : Hypothèse 1 (Échange d'ions) (noté H1)
- En vert : Hypothèse 2 (Échange d'ions + Complexation par la MOS) (noté H2)
- En jaune : Hypothèse 3 (Échange d'ions + Complexation par oxyhydroxyde amorphe [FeO ou AlO ou MnO]) (noté H3)
- En bleu : Hypothèse 4 (Échange d'ions + Complexation par la MOS + Complexation par FeO ou AlO ou MnO) (noté H4)
- En orange : Hypothèse 5 (Échange d'ions + Complexation par la MOS + FeO + AlO + MnO) (noté H5)
- Approches « a » (sites de sorption non spécifiques calculés à partir de la CEC mesurée) et
- Approches « b » (sites de sorption non spécifiques calculés à partir des teneurs en argiles granulométriques estimées).

La totalité des hypothèses est classée par ordre croissant d'erreurs relatives médianes.

On note tout d'abord que pour une même hypothèse, les approches « a » et « b » génèrent des erreurs de prédiction similaires, et cela pour la quasi-totalité des hypothèses testées, avec seulement une différence notable dans les distributions des erreurs de prédictions pour quelques modalités de l'hypothèse 3 (Figure III.4).

À partir des box plots rangés par erreur relative médiane croissante, on peut distinguer les familles suivantes :

Les paramétrages générant les surestimations les plus fortes (médiane du rapport prédit/observé > 100) correspondent aux hypothèses 1 (en rouge) où seule la sorption non spécifique est considérée.



La concentration modélisée de  $\text{Cd}^{2+}$  est largement surestimée, indiquant donc que la modélisation de Cd sorbé est beaucoup trop faible pour expliquer les concentrations en  $\text{Cd}^{2+}$  en solution.

Les paramétrages générant des prédictions surestimées fortes (valeur médiane du rapport prédit/observé comprise entre 1 et 100) correspondent tous aux hypothèses 3 (en jaune) (sorption non spécifique + complexation par un oxyde) avec les extractions Tamm et pyrophosphate. Il faut toutefois noter que pour cette hypothèse, en fonction de l'extractant ou des phases réactives, l'erreur de prédiction peut être ponctuellement faible (voir paragraphe suivant). De façon générale, avec cette hypothèse H3, les concentrations de  $\text{Cd}^{2+}$  modélisées en solution sont encore surestimées, cependant moins que dans le cas de l'hypothèse 1. Les quantités de Cd extraites par les extractants Tamm et pyrophosphate sont fortes (entre 15 et 75 % du Cd total) et surestiment probablement le Cd de la phase solide qui s'échange réellement avec la solution. En outre modéliser la sorption par la sorption non spécifique et la complexation par un seul oxyde sous-estime vraisemblablement la quantité de Cd sorbé et donc surestime celle en solution.

Les paramétrages générant en moyenne des erreurs faibles où dans ces cas la modélisation du  $\text{Cd}^{2+}$  en solution est la plus proche des valeurs réelles, correspondent à deux familles d'hypothèses (Figure III.4) :

1/le Cd extrait au  $\text{NH}_4\text{NO}_3$  avec de la sorption non spécifique complétée par de la complexation par un oxyde métallique MnO ou AlO mais pas FeO (Hypothèses 3) et

2/le Cd extrait par le pyrophosphate à condition de considérer la MOS comme phase sorbante, seule (hypothèse H2) ou avec un (hypothèse H4) ou plusieurs (Hypothèse H5) oxydes. Dans le cas de l'extraction au pyrophosphate, les oxydes seuls sont donc peu influents, ce qui est important c'est d'inclure la MOS pour sorber par complexation le compartiment important de Cd supposé mobile par l'extraction pyrophosphate et ainsi de ne pas surestimer le  $\text{Cd}^{2+}$  en solution comme c'est le cas lorsque l'extraction du Cd par le pyrophosphate n'est pas associée à la MOS (hypothèse H3, cf. paragraphe précédent).

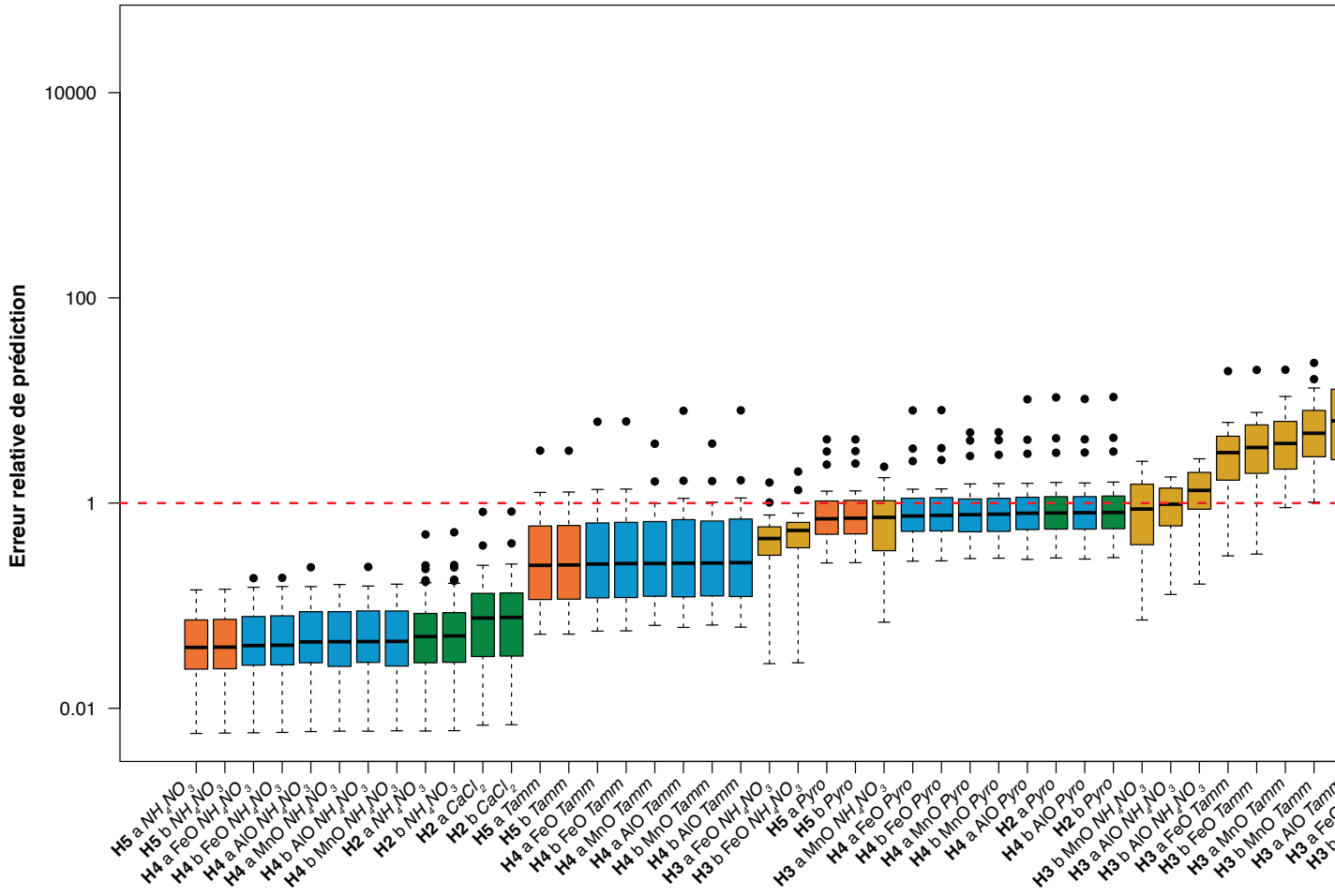


Figure III.4 : Box plots des erreurs relatives de prédiction (prédit/observé) pour les concentrations en  $\text{Cd}^{2+}$  en solution considérées lors de la modélisation multisurfaces de la distribution phase solide/solution du  $\text{Cd}^{2+}$  dans le sol. Chaque couleur (H1 en rouge, H2 en vert, H3 en jaune, H4 en bleu et H5 en orange). Pour chacune de ces familles, l'extractant du Cd est et la façon dont est considérée la sorption non spécifique : par la CEC mesurée (approche a), par modélisation de la minéralogie des argiles (approche b).

Ensuite viennent les paramétrages sous-estimant la concentration en  $\text{Cd}^{2+}$  en solution et donc surestimant la sorption du Cd et donc sous-estimant le  $\text{Cd}^{2+}$  en solution (troisième quartile des erreurs relatives inférieures à 1). Il s'agit quasi exclusivement des scénarios pour lesquels la MOS est considérée comme phase sorbante mais où le Cd supposé mobile est sous-estimé relativement à la capacité de sorption de la MOS. Dans ce cas le Cd mobile étant sous-estimé, la fraction du Cd mobile sorbé par la MOS est trop importante et le  $\text{Cd}^{2+}$  en solution est sous-estimé. Il s'agit des scénarios associant la MOS avec une extraction du Cd plus faible que celle du pyrophosphate (hypothèses H2, H4, H5 croisées avec les extractions  $\text{CaCl}_2$ ,  $\text{NH}_4\text{NO}_3$  et Tamm). De manière logique, le  $\text{Cd}^{2+}$  en solution est d'autant plus sous-estimé (et donc la sorption surestimée) que l'extractant est faible (erreur Tamm inférieure à l'erreur  $\text{CaCl}_2$  et  $\text{NH}_4\text{NO}_3$ ). On notera que le  $\text{Cd}^{2+}$  en solution est également sous-estimé pour le scénario où le Cd mobile est estimé par l'extraction à  $\text{NH}_4\text{NO}_3$  et la sorption par complexation uniquement lié à FeO (Hypothèse H3). Dans la mesure où le  $\text{Cd}^{2+}$  en solution est correctement estimé lorsque FeO est remplacé par MnO ou AIO (Figure III.4), on peut en déduire que les FeO vont complexer plus de Cd que MnO ou AIO ou que la complexation est plus forte.

En conclusion, nous avons pu observer que la sorption non spécifique seule sous-estime la sorption et surestime les concentrations en  $\text{Cd}^{2+}$  mesurées en solution et cela pour les compartiments de Cd mobiles estimés par trois extractants différents (eau UP,  $\text{CaCl}_2$  et  $\text{NH}_4\text{NO}_3$ ). La modélisation de la distribution sol-solution du Cd doit donc considérer la complexation spécifique en plus de la sorption non spécifique.

Dans ce cas, il est nécessaire que la taille du compartiment de Cd considéré mobile soit en adéquation avec le nombre et la nature des sites de sorption considérés dans le modèle. À titre d'illustration, si l'on considère un compartiment de Cd mobile très important comme celui extrait par le pyrophosphate, la MOS doit être incluse dans la modélisation, car la complexation par les oxydes seuls ne suffit pas à sorber le Cd et le Cd en solution est alors surestimé.

Parmi les différents paramétrages testés, les modèles permettant des prédictions les plus justes des concentrations en  $\text{Cd}^{2+}$  en solution se basent, soit sur une extraction  $\text{NH}_4\text{NO}_3$  1 M, de la sorption non spécifique, et de la complexation par un oxyde métallique amorphe (AIO, MnO) ou soit sur une extraction pyrophosphate ( $\text{Na}_4\text{P}_2\text{O}_7$ ) 0.1 M en considérant l'adsorption non spécifique et la complexation par la MOS avec ou sans oxydes métalliques amorphes. De manière synthétique, pour les sols agricoles étudiés, le partage phase solide-solution du  $\text{Cd}^{2+}$  peut se modéliser par les couples compartiments mobiles/phases réactives suivants :

Extraction pyrophosphate/sorption non spécifique CEC + sorption spécifique par la MOS

**ou**

Extraction par  $\text{NH}_4\text{NO}_3$ /sorption non spécifique sur la CEC et sorption spécifique par les oxydes de manganèse et d'aluminium amorphes.

### 3.4. Spéciation solide du Cd en sols agricoles non contaminés

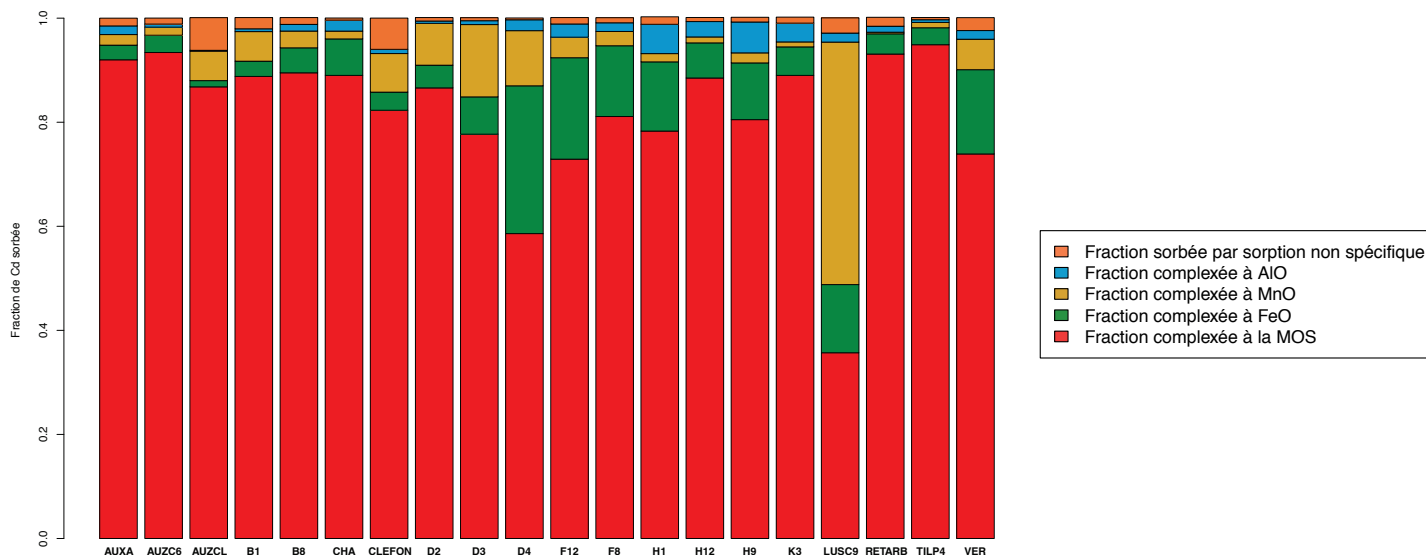


Figure III.5 : Distribution du Cd entre les phases solides modélisées par modèle géochimique WHAM à partir des teneurs en Cd obtenues par extraction au Pyrophosphate de sodium. Les argiles (échanges d'ions), la MOS et les oxydes métalliques amorphes du sol (H5).

L'hypothèse 5 est par construction l'hypothèse qui permet de considérer le maximum de phases de sorption : sorption non spécifique par la CEC (argiles + MOS), sorption spécifique par complexation par la MOS et les oxydes de Fe, Mn et Al amorphes. En considérant le compartiment mobile équivalent au Cd extrait par le pyrophosphate, les prédictions du  $\text{Cd}^{2+}$  en solution pour cette hypothèse sont proches de celles mesurées malgré une faible sous-estimation. Ce scénario peut donc être utilisé pour examiner la répartition du Cd sorbé entre les différents sites de sorption tel qu'il est modélisé par WHAM. La Figure III.5 représente la distribution du Cd entre les différentes phases réactives du sol considérées par l'Hypothèse 5 (Surface d'échange, oxydes d'aluminium, de manganèse et de fer amorphes et MOS) pour 20 des sols agricoles étudiés.

On note que la complexation du Cd par la MOS est majoritaire dans la quasi-totalité des sols, avec une fraction de Cd moyenne complexée par la MOS de 82 % [36 % - 95 %] puis vient la fraction complexée aux oxydes de fer en moyenne de 8.6 % [1.2 % - 8 %] et aux oxydes de Mn amorphes, avec une moyenne à 6.3 % [0.3 % - 46.6 %]. La fraction de Cd sorbé sur les oxydes d'aluminium amorphes est en moyenne plus importante sur les sols dont le pH est supérieur à 8.1 comparé aux sols neutres et acides (3.3 % contre 0.9 %), avec une gamme de variation

globale de 0.15 à 5.9 %. La fraction de Cd sorbé par échange d'ions semble relativement faible allant de 0.3 % à 6.3 % du Cd extrait par pyrophosphate de sodium. La répartition du Cd dans les différents compartiments de la Figure III.5 ne renseigne pas sur la nature du/des compartiments qui réapprovisionnent la solution du sol lorsque le  $\text{Cd}^{2+}$  dans cette dernière diminue suite au prélèvement racinaire. Formulé différemment, il n'est pas possible de savoir quels compartiments sont disponibles même si la fraction sorbée de manière non spécifique en fait sûrement totalement partie.

## **4. Discussion**

### **4.1. Cd extrait par extractions sélectives**

Les quantités de Cd extraites par les différents extractants nous renseignent déjà sur la distribution probable du Cd dans les sols suivant logiquement le type de réactions ciblées et la force d'extraction des différentes solutions. Les extractions faibles comme l'eau UP vont permettre la solubilisation du Cd faiblement lié aux phases solides du sol. En effet, l'eau UP a une force ionique très faible induisant un déséquilibre ionique entre la solution de contact et le sol. Ainsi le sol va désorber des éléments afin de rééquilibrer la solution.

Les solutions salines ( $\text{CaCl}_2$ ,  $\text{NH}_4\text{NO}_3$ ) permettent l'extraction du compartiment de Cd considéré comme échangeable (Houba et al., 1990) via la désorption du Cd par augmentation de la compétition par les cations ( $\text{Ca}^{2+}$  et  $\text{NH}_4^+$ ) apportés. Dans une moindre mesure, la présence plus importante des ligands  $\text{Cl}^-$  ou  $\text{NO}_3^-$  en solution favorise également la complexation du Cd en phase liquide ce qui favorise aussi la désorption par abaissement de la concentration du  $\text{Cd}^{2+}$  en solution. Nous avons observé que la désorption du Cd était légèrement supérieure lors de l'utilisation du  $\text{NH}_4\text{NO}_3$  comparé au  $\text{CaCl}_2$  et cela pourrait être attribué au fait que la majorité des sols étudiés sont alcalins. En effet, en sols neutres à alcalins, l'ammonium est rapidement transformé en amines qui favorisent la formation d'amino-complexes de Cd en solution, induisant une désorption du Cd plus importante (Gryschko et al., 2005). Par ailleurs, le pH de la solution de  $\text{NH}_4\text{NO}_3$ , plus faible que celui de la solution de  $\text{CaCl}_2$ , pourrait induire une désorption plus élevée de Cd en favorisant la protonation des sites de sorption. Cet effet pourrait être exacerbé par la désorption des protons sorbés faiblement aux constituants du sol par échange avec les ions ammonium, avec une acidification de la solution potentiellement plus forte pour les sols ayant un pouvoir tampon du proton élevé (différence entre pH du sol et pH  $\text{CaCl}_2$ ). Cette diminution du pH peut également favoriser la dissociation des complexes en solution et libérer du  $\text{Cd}^{2+}$  susceptible de se sorber sur les phases solides. L'extraction du Cd par  $\text{NH}_4\text{NO}_3$  1 M pourrait être également plus efficace que l'extraction  $\text{CaCl}_2$  0.01M, car l'ion ammonium est plus concentré (1 M) que le  $\text{Ca}^{2+}$  (0.01 M). En outre en sols alcalins, les concentrations en Ca en solution sont déjà élevées, proches de 0.01 M, rendant peu efficace l'extraction par le  $\text{CaCl}_2$  0.01M. Suivant ce

raisonnement, les différences de quantités de Cd extraites sont attendues être plus marquées pour les sols acides.

Le réactif de Tamm, constitué d'acide oxalique à 0.2 M et d'oxalate d'ammonium à 0.2 M, a extrait plus de Cd que les précédentes solutions. Ceci est lié à la dissolution des oxydes métalliques amorphes permettant la solubilisation du Cd complexé à ces surfaces. En plus de l'action sur les oxydes peu cristallisés, le réactif de Tamm via son faible pH (aux alentours de 3) pourrait solubiliser une partie de la MOS et également altérer les argiles minéralogiques (Beckett, 1989) en plus de dissoudre une fraction plus importante de minéraux carbonatés et de dissocier les complexes avec le Cd par compétition du proton.

Le pyrophosphate de sodium agit comme chélateur des cations qui maintiennent la MOS dans un état floculé et donc insoluble. En immobilisant ces cations, le pyrophosphate permet la solubilisation d'une partie de la MOS considérée comme facilement mobilisable (Garrett et al., 1998). La MOS solubilisée complexant le Cd, le pyrophosphate contribue à solubiliser également fortement le Cd. En outre, le pyrophosphate semble également capable de dissoudre les oxydes de manganèse permettant la solubilisation du Cd y étant sorbé sans cependant avoir d'effet sur les oxydes de fer amorphes (Beckett, 1989).

Ainsi les quantités de Cd solubilisées sont plus faibles pour l'eau UP < CaCl<sub>2</sub> ≈ NH<sub>4</sub>NO<sub>3</sub> < Tamm < Pyrophosphate, indiquant une distribution du Cd comme suit : solubilisable < échangeable < complexé par des oxydes métalliques amorphes < complexé par la MOS. La modélisation géochimique par le modèle WHAM prédisant correctement le Cd<sup>2+</sup> en solution prédit également qu'en moyenne la distribution du Cd s'organise par ordre d'importance selon l'ordre solubilisable < échangeable < complexé par des oxydes métalliques amorphes < complexé par la MOS (Figure III.5). Il y a donc bien cohérence entre les extractions chimiques et la modélisation géochimique concernant les estimations de la répartition du Cd sur les différentes phases sorbantes.

#### ***4.2. Adéquation entre la quantité de Cd extrait et les surfaces de sorption à considérer dans un modèle géochimique***

Afin de prédire correctement les concentrations en Cd<sup>2+</sup> en solution via l'utilisation de modèles géochimiques, nos résultats montrent qu'il est primordial que la taille du compartiment de Cd estimé mobile soit en adéquation avec les phases de sorption considérées et les mécanismes de sorption représentés. En effet, si le compartiment de Cd mobile considéré dans le modèle est trop important par rapport au nombre de sites de sorption modélisés et à leur affinité vis-à-vis du Cd, le modèle sous-estime la sorption du Cd et donc surestime les concentrations en Cd<sup>2+</sup>, et vice-versa. Suivant ce raisonnement, nous en déduisons que le NH<sub>4</sub>NO<sub>3</sub>, le CaCl<sub>2</sub> et même l'eau UP semblent extraire plus de Cd qu'il est possible d'en sorber sur les phases de sorption non

spécifique, définies par la CEC mesurée ou calculée à partir des teneurs en argiles et MOS (Figure III.4, hypothèse H1).

Lorsque seule la sorption non spécifique est considérée dans la modélisation, la quantité de Cd sorbé est donc sous-estimée. L'ordre de grandeur de la teneur en Cd du sol est  $0.1 \text{ mg.kg}^{-1}$  soit  $1.8 \cdot 10^{-4} \text{ cmol}^+.\text{kg}^{-1}$  de sol ce qui est très nettement inférieur à la CEC des sols étudiés (médiane de  $15.6 \text{ cmol}^+.\text{kg}^{-1}$ ). On peut donc écarter l'hypothèse que le nombre de sites de sorption non spécifiques soit limitant pour le Cd dans l'absolu. En revanche, par rapport au Cd, les teneurs en Ca sont très élevées, de l'ordre de 3 mM de Ca en solution (Tableau III.3). Si on considère un coefficient de partition du Ca aux environs de 100 (Sheppard et al., 2009), cela correspond à un ordre de grandeur de  $300 \text{ mmol.kg}^{-1}$  de Ca sorbé sur la phase solide soit,  $60 \text{ cmol}^+ \text{ Ca/kg}$  de sol. Les interactions électrostatiques de la sorption non spécifique se font au prorata des rapports de concentrations (Tipping et al., 2011) et le Cd étant minoritaire par rapport au Ca, les estimations précédentes montrent que la CEC du sol est très probablement dominée par le Ca et aussi probablement par Mg, et que le rapport des concentrations est très défavorable à la sorption du Cd. C'est probablement la raison pour laquelle la sorption du Cd est sous-estimée (et la concentration en  $\text{Cd}^{2+}$  en solution surestimée) lorsque seule la sorption non spécifique est considérée dans la modélisation.

De la même façon, un oxyde métallique amorphe (Mn, Al ou Fe) ajouté aux phases de sorption non spécifiques ne semble pas pouvoir suffire à sorber la totalité du Cd extrait par extraction Tamm ou pyrophosphate (Figure III.4, hypothèse H3). Ceci est cohérent avec le mode d'action du pyrophosphate expliqué précédemment qui suppose que cet extractant solubilise une partie de la MOS et donc le Cd qui y est complexé. Le mode d'action du réactif de Tamm, dissolvant les phases amorphes des minéraux secondaires du sol semble également extraire plus de Cd qu'il est possible d'en sorber sur des oxydes métalliques amorphes et sur la CEC. Cela peut provenir du faible pH de la solution Tamm (pH égal à 3), qui peut également décomplexer du Cd provenant d'autres constituants du sol par compétition du proton ou dissoudre une plus grande partie du calcaire du sol sur lequel le Cd est sorbé et non considéré dans la modélisation.

À l'inverse, lorsque l'on considère la sorption du Cd sur les MOS, si l'on ne met pas en regard le Cd extrait par pyrophosphate, le compartiment mobile de Cd sera trop faible pour estimer la distribution du Cd sur les phases solides et la concentration en  $\text{Cd}^{2+}$  en solution de sol sera sous-estimée d'autant plus que le compartiment de Cd est faible.

Parmi les paramétrages testés, deux semblent correctement expliquer les concentrations en  $\text{Cd}^{2+}$ . **Le premier** considère le compartiment de Cd extrait par pyrophosphate avec comme phases de sorption les surfaces d'adsorption non spécifiques, la MOS et les oxydes métalliques, ces derniers ayant une faible influence. **Le second** paramétrage simulant des valeurs justes de

$\text{Cd}^{2+}$  en solution comprend un compartiment de Cd extrait par une solution de  $\text{NH}_4\text{NO}_3$  à 1 M beaucoup plus faible que celui obtenu par le pyrophosphate, avec pour surface sorbant le Cd, des sites non spécifiques et un oxyde métallique amorphe (AIO, MnO) et donc sans considérer la complexation par la MOS.

Le fait de devoir ajouter une surface de complexation spécifique (AIO ou MnO) pour modéliser la répartition du compartiment en Cd extrait par  $\text{NH}_4\text{NO}_3$  pourrait signifier que cet extractant désorbe du Cd ne provenant pas de la fraction considérée comme échangeable. Ceci est probable si la diminution du pH est favorisée par la désorption de proton. Cette modification de pH pourrait modifier la charge globale de surface des oxydes métalliques amorphes et favoriser la désorption du Cd sorbé de façon spécifique. Cette explication permettrait d'expliquer l'hypothèse avec AIO puisque le pH de point de charge nulle ( $\text{pH}_{\text{PCN}}$ ) est élevé (aux alentours de 8). Cette explication est moins appropriée pour expliquer l'hypothèse avec MnO puisque le  $\text{pH}_{\text{PCN}}$  est beaucoup plus bas, donc les sites à charges variables devraient être chargés négativement même en considérant l'influence de la solution d'extraction. La conclusion sur les oxydes de manganèse reste cependant à débattre puisque des oxydes, considérées moins présentes dans les sols (pyrolusite ou  $\beta\text{-MnO}_2$ ) ont un  $\text{pH}_{\text{PCN}}$  plus élevé, aux alentours de 7.2 (Sparks, 2003).

#### ***4.3. Convergences des approches statistiques et géochimiques pour la modélisation de la partition sol-solution du Cd en sols agricoles ?***

Pour comprendre la spéciation du Cd sur la phase solide, c'est-à-dire sa distribution entre les différents constituants réactifs du Cd, nos résultats ont montré que l'utilisation d'un compartiment de Cd estimé par une extraction forte comme le pyrophosphate semblait la plus appropriée, car un compartiment mobile dérivé de cet extractant a permis de simuler correctement le  $\text{Cd}^{2+}$  en solution lorsque toutes les phases porteuses étaient considérées dans le modèle (non spécifique, complexation par la MOS, et par les oxydes amorphes). Dans la littérature, les études de distribution sol-solution considèrent généralement des compartiments de Cd extraits par solution d'EDTA, d' $\text{HNO}_3$  à différentes concentrations, ou même la fraction totale de Cd, avec des erreurs de prédictions des concentrations en solution (Cd total ou libre) (valeurs médianes des erreurs quadratiques moyennes, RMSE) allant de 0.3 à 0.6 (Bonten et al., 2008a; Dijkstra et al., 2009, 2004; Groenenberg et al., 2012; Gustafsson et al., 2003; Izquierdo et al., 2012; Khai et al., 2008; Marzouk et al., 2013; Rennert et al., 2017; Schröder et al., 2005). Ces travaux aboutissent à la même conclusion que la nôtre, à savoir que le Cd dans le sol est majoritairement complexé à la MOS.



Cependant, ce n'est pas forcément le Cd lié à la MOS qui régit le  $\text{Cd}^{2+}$  en solution en condition naturelle, même si ce compartiment Cd-MOS est majoritaire. Si l'on envisage une vision plus dynamique du système sol en prenant en considération le réapprovisionnement de la solution en  $\text{Cd}^{2+}$  par la phase solide lors du prélèvement du Cd par les racines des plantes et donc l'appauvrissement de la solution de sol en Cd, il est peu probable que les phases complexant le Cd de façon spécifique par formation de liaisons fortes de type covalent puissent participer de manière importante au réapprovisionnement de la solution de sol. Physiquement, le réapprovisionnement de la solution par la phase solide se fera vraisemblablement préférentiellement à partir des liaisons électrostatiques faibles du Cd avec les sites non spécifiques. Ainsi, dans le cadre de la modélisation de la biodisponibilité, conceptualiser la dynamique de la distribution sol-solution du Cd à partir d'un compartiment mobile relativement faible, extrait par le  $\text{NH}_4\text{NO}_3$  en y associant de la sorption non spécifique et de la sorption spécifique sur seulement un oxyde de manganèse ou d'aluminium amorphe serait probablement plus proche de la réalité qu'en considérant un compartiment biodisponible plus large extrait par le pyrophosphate associé à de la complexation à la MOS en plus de la sorption non spécifique.

Cette hypothèse est confortée par la modélisation statistique. Les modèles les plus performants pour prédire les concentrations en  $\text{Cd}^{2+}$  mesurées en solution comprennent notamment le Cd extrait par  $\text{NH}_4\text{NO}_3$  et les teneurs en oxydes de manganèse et/ou d'aluminium amorphes (Équation III.4, Équation III.5) en plus du pH du sol et des concentrations en Ca en solution. Entre les deux possibilités de paramétrage du modèle géochimique pour prédire correctement le  $\text{Cd}^{2+}$  en solution (Cd- $\text{NH}_4\text{NO}_3$  + sorption CEC + un oxyde Mn/Al et Cd-Pyrophosphate + sorption CEC + MOS), la modélisation statistique plaiderait pour le paramétrage basé sur le Cd extrait par  $\text{NH}_4\text{NO}_3$ .

Dans la modélisation statistique, l'influence négative du pH sur les concentrations en  $\text{Cd}^{2+}$  en solution (Équation III.4) est attendue, car largement documentée dans la littérature (de Vries et al., 2011; Groenenberg et al., 2010b; Ivezić et al., 2012; Rieuwerts et al., 2006; Römkens et al., 2004; Sauvé et al., 2000a, 2000b; Tipping et al., 2003). En effet, le pH joue un rôle important sur la majorité des processus influençant la distribution du Cd entre les phases solides et la solution de sol et sa fraction libre (Loganathan et al., 2012; Rieuwerts et al., 1998). Il a été montré que l'augmentation du pH induisait une augmentation du potentiel de surface négatif du sol favorisant la sorption du Cd. L'augmentation du pH favorise également la complexation spécifique du Cd sur les MOS, argiles ou oxydes métalliques amorphes au détriment de la sorption non spécifique (Christensen and Christensen, 2000; Tiller et al., 1979; Wang et al., 2009). Le pH en solution peut également influencer la spéciation en solution du Cd en favorisant la complexation avec les matières organiques dissoutes (Sauvé et al., 2000b). La légère corrélation négative entre le pH

du sol et la quantité de Cd extraite par  $\text{NH}_4\text{NO}_3$  pourrait expliquer son absence dans les modèles comportant moins de variables explicatives (Équation III.5 - Équation III.7) et la plus faible significativité de l'effet de cette variable dans nos modèles comparé aux travaux des auteurs précédents qui, pour la plupart, ont utilisé les teneurs en Cd total. Cette corrélation négative pH/Cd en solution est cohérente avec l'action du pH sur le type de sorption, puisqu'à pH faible, les teneurs en Cd extrait par  $\text{NH}_4\text{NO}_3$ , donc sorbé de façon non spécifique, sont plus importantes qu'à pH alcalins, où le Cd sera complexé de façon plus spécifique.

Toujours dans la modélisation statistique, la concentration en Ca en solution influence positivement la concentration en  $\text{Cd}^{2+}$  en solution. Cette influence positive est également cohérente avec le rôle connu de compétiteur de cet élément à la fois pour les sites de sorption et pour la complexation en solution du Cd (Chapitre II). Le pH des sols étudiés étant en moyenne plutôt élevé, les surfaces de sorption ont un potentiel de surface négatif plus important et la compétition Cd/Ca pour compenser ces charges négatives en est exacerbée. L'influence du Ca sur la distribution sol-solution du Cd est peu reportée dans la littérature et son influence significative dans notre étude pourrait être liée à la proportion plus importante de sols calcaires et alcalins.

Le dernier paramètre identifié par modélisation statistique est la teneur en oxydes de manganèse (Équation III.4) ou d'aluminium (Équation III.5) amorphes, avec une influence négative indiquant qu'ils limitent les concentrations en  $\text{Cd}^{2+}$  en solution. La présence de ces régresseurs dans les modèles statistiques, malgré une faible contribution à expliquer la variance (~2 %), reste significative. Les oxydes de manganèse amorphes dans les sols sont considérés comme des phases sorbantes importantes pour les éléments traces. Ceci est la conséquence de leur grande surface spécifique et leur pH de point de charge nulle très faible (Sparks, 2003), due à la réduction des cations  $\text{Mn}^{4+}$  en  $\text{Mn}^{2+}$  induisant un déficit de charges de surface indépendant du pH. Ces particularités induisent chez ces oxydes une plus grande capacité à sorber le Cd que les oxydes de fer ou d'aluminium.

Le cas des oxydes d'aluminium est moins évident. Dans les sols agricoles des régions tempérées, les oxydes d'aluminium ne sont théoriquement pas naturellement abondants (Sparks, 2003). Dans notre étude, la fraction d'AIO estimée par extraction Tamm pourrait représenter d'autres types de minéraux faiblement cristallisés également constitués d'aluminium comme des phyllosilicates. Il est généralement considéré que les sites de bordures de ces minéraux peuvent constituer, à des pH alcalins, des sites potentiels de complexation spécifique pour le Cd (Gu et al., 2014; Gu and Evans, 2007; Lackovic et al., 2004). Ainsi les extractions par le réactif de Tamm pourraient dissoudre une partie de ces sites de bordures (Deb, 1942) représentant ainsi une surface de complexation spécifique du Cd. Néanmoins, la fraction de sites de bordures sur les phyllosilicates est généralement faible comparée à la densité de sites basaux, majoritairement impliqués dans les processus de sorption non spécifique du Cd (Sparks, 2003). L'altération de

surface des phyllosilicates par la MOS comme proposée par Basile-Doelsch et al., (2015) permettrait aussi d'expliquer la contribution des oxydes d'aluminium amorphes pour prédire les concentrations en  $\text{Cd}^{2+}$  en complément des phases de sorption non spécifique. En effet, la MOS sorbée en fine couche à la surface des phyllosilicates pourrait permettre l'altération de ces minéraux et l'augmentation de sites de sorption à charges variables, cette altération pouvant être favorisée par la présence de défauts dans la structure des argiles.

La distribution du Cd entre la phase solide et la solution pourrait donc être correctement prédite à partir des teneurs en Cd extrait par  $\text{NH}_4\text{NO}_3$ , représentatif d'une fraction échangeable du Cd du sol, par une surface d'échange représentée par la valeur de CEC mesurée et par une surface de complexation assimilée à des oxydes d'aluminium ou de manganèse amorphes (Figure III.4). Les bonnes performances de prédiction, à la fois statistique et géochimique, des modèles fondés sur le Cd extrait par  $\text{NH}_4\text{NO}_3$  et la présence d'oxyde de Mn ou d'Al amorphes comme phase sorbante, laissent penser que ces variables sont bien représentatives des mécanismes de sorption du Cd dans les sols. Ainsi, les surfaces réactives participant réellement à l'approvisionnement de la solution de sol pourraient être résumées aux surfaces d'échanges non spécifiques (interactions électrostatiques en lien avec la CEC) et aux surfaces de complexation des oxydes amorphes (Mn, Al). Le compartiment de Cd réellement mobile du sol est alors plus proche des teneurs extraites par  $\text{NH}_4\text{NO}_3$  que par le pyrophosphate.

Il est bien sûr possible de modéliser géochimiquement le  $\text{Cd}^{2+}$  en solution en forçant l'intervention de la complexation par la matière organique et le Cd extrait par le pyrophosphate. Cependant, comme la matière organique du sol et le Cd extrait par le pyrophosphate n'apparaissent pas comme des régresseurs performants dans les modèles statistiques, sans que ce soit une preuve, cela suggère que la complexation du Cd par la MOS est trop forte pour significativement contribuer au  $\text{Cd}^{2+}$  retrouvée dans la solution de sol.

La limite de ce paramétrage ( $\text{NH}_4\text{NO}_3 + \text{AlO}/\text{MnO} + \text{surface d'échange}$ ) du modèle géochimique est visible pour deux sols, D4 et RETARB pour lesquels, la concentration en  $\text{Cd}^{2+}$  en solution est largement sous-estimée et donc la sorption surestimée. Ces sols sont respectivement riches en oxydes de fer amorphes ( $5.40 \text{ g.kg}^{-1}$ ) et en calcaire ( $841.9 \text{ g.kg}^{-1}$ ) et pourraient impliquer des mécanismes de sorption différents nécessitant l'utilisation d'un compartiment de Cd différent de celui extrait par une solution de  $\text{NH}_4\text{NO}_3$ .

## **5. Conclusion**

L'objectif de ce chapitre a été de prédire les concentrations en  $\text{Cd}^{2+}$  en solution à partir de variables facilement mesurables du sol et d'utiliser la modélisation géochimique pour tenter de

comprendre les phases solides qui seraient le plus à même de contribuer au Cd en solution. Les modélisations statistique et géochimique permettent de prédire assez bien la distribution sol-solution du Cd<sup>2+</sup> (RMSE = 0.170 pour l'approche géochimique et RMSE = 0.160 pour l'approche statistique) avec pour la modélisation statistique, un nombre de variables à mesurer restreint (3 avec [Cd NH<sub>4</sub>NO<sub>3</sub>, Ca<sub>ss</sub>, AIO] à 4 avec [Cd NH<sub>4</sub>NO<sub>3</sub>, pH, Ca<sub>ss</sub>, MnO]). Les bonnes performances des modèles statistiques testés par validation croisée et le domaine assez vaste des sols étudiés autorise a priori un degré de confiance plutôt satisfaisant concernant la robustesse des modèles statistiques qui de, par leur simplicité, seront privilégiés dans l'objectif de la prédiction. Il faut cependant garder en mémoire que les sols acides dans cette étude sont sous-représentés, mais c'est une situation classique en sols agricoles.

L'apport de la modélisation géochimique est intéressant, car parmi les scénarios performants pour prédire le Cd<sup>2+</sup> en solution, celui supposant que le Cd en équilibre avec la solution était essentiellement sorbé sur la CEC et les oxydes s'est avéré être également suggéré par les meilleurs modèles statistiques. La modélisation géochimique a donc permis de mieux comprendre les variables des meilleurs modèles statistiques et de renforcer la confiance dans la valeur prédictive de ces variables en y adossant des mécanismes. Il est tout de fois important de noter que ces conclusions reposent sur des hypothèses fortes concernant la réactivité des phases de sorption. En effet, le choix de la densité de charges des argiles se base sur le type d'argiles minéralogiques dominant et peut donc facilement être critiqué. Il serait donc important de valider les hypothèses de modélisation par des caractérisations physiques des phases de sorption du Cd et plus particulièrement des argiles, qui semblent jouer un rôle important dans la spéciation du Cd en sols agricoles. Par exemple, des analyses par diffraction de rayons X permettraient d'identifier les types d'argiles minéralogiques dominants de manière plus contrôlée et ainsi estimer les fractions d'AH de façon plus juste.

En perspective, deux approches peuvent être considérées pour estimer la fraction de Cd mobile du sol et la confronter à notre estimation. La première serait de réaliser des expériences de sorption/désorption en conditions contrôlées (concentrations en Ca dans la solution de contact équivalente à celles mesurées en solution de sol, et rapport volume:masse défini) avec différentes concentrations en Cd en solution, afin d'estimer le compartiment de Cd réellement mobile. Pour respecter la linéarité, les variations de Cd induites doivent être relativement proches de la concentration en Cd<sup>2+</sup> à l'équilibre du sol. Une fois ce compartiment de Cd mobile estimé pour des sols contrastés, il pourrait être régressé en fonction des paramètres pédogéochimiques des sols afin de pouvoir le prédire plus simplement et pour essayer d'identifier les constituants du sol qui l'expliquent.

Pour aller encore plus loin, il faudrait prendre en compte les cinétiques de sorption/désorption du Cd et des cations compétiteurs sur les différents constituants du sol dans un modèle géochimique multi-surfaces. Ce genre de modèles permettrait d'étudier comment les cinétiques de sorption

influent sur la contribution individuelle de chaque type de surfaces réactives au réapprovisionnement global en Cd de la solution de sol. Une étude récente, réalisée par Peng et al. (2018) considère cette approche mécaniste et dynamique de modélisation du réapprovisionnement car certaines cinétiques non prises en compte actuellement pourraient être limitantes. Cette approche cinétique est pertinente pour la prédiction du Cd prélevé par une culture. En effet, la hiérarchie des phases solides assurant le réapprovisionnement de la solution du sol en Cd<sup>2+</sup> doit être raisonnée par rapport aux échelles de temps du prélèvement par les racines.





## **Chapitre IV. Modélisation statistique des teneurs en Cd dans le grain de blé dur**

---

Dans ce chapitre, nous avons cherché à élaborer un modèle simple de prédiction des teneurs en Cd dans les grains de blé dur à partir de paramètres du sol. Pour rappel (cf. Chapitre I.7), l'objectif de cette partie est double : i) examiner la valeur d'un modèle statistique utilisant uniquement des variables du sol pour prédire la teneur en Cd du grain de blé dur par rapport aux seuils réglementaires actuel et en projet, ii) étudier si les variables d'un modèle prédictif du Cd dans le grain étaient bien cohérentes avec les variables prédictives du Cd<sup>2+</sup> en solution, identifiées dans les chapitres II et III, conformément au fait que cette forme du Cd est celle prélevée par les racines et la mieux corrélée à la phytodisponibilité.

Ce travail a été réalisé en deux phases dépendantes de la chronologie de deux projets de recherche.

La première phase porte sur l'élaboration de modèles statistiques prédictifs de la teneur en Cd du grain de blé dur à partir d'un premier jeu de données issues du projet Arvalis-Inra Cadur (2012-2014) qui était déjà disponible au début de mon travail de thèse. Ce jeu de données a permis de faire une première exploration de la démarche de modélisation. Dans une deuxième phase, le nombre de sols a été augmenté grâce aux enquêtes de terrain du projet ANR CaDON (Compréhension de la co-contamination du blé dur par le Cd et la mycotoxine deoxynivalenol). Les sols complémentaires ont été sélectionnés pour compléter le premier jeu de données afin de couvrir une gamme de situations la plus large possible.



## **Partie 1 - Première élaboration de modèles statistiques prédictifs de la teneur en Cd dans le grain de blé dur**

Ce premier travail de modélisation du Cd dans le grain de blé dur est présenté sous la forme de l'article tel que publié dans *Environmental Science and Pollution Research* en 2017.

Viala, Y., Laurette, J., Denaix, L., Gourdain, E., Méléard, B., Nguyen, C., Schneider, A., Sappin-Didier, V., 2017. Predictive statistical modelling of cadmium content in durum wheat grain based on soil parameters. *Environ. Sci. Pollut. Res.* <https://doi.org/10.1007/s11356-017-9712-z>

### ***Predictive statistical modelling of cadmium content in durum wheat grain based on soil parameters***

Yoann Viala<sup>a</sup>, Julien Laurette<sup>a</sup>, Laurence Denaix<sup>a</sup>, Emmanuelle Gourdain<sup>b</sup>, Benoit Méléard<sup>b</sup>, Christophe Nguyen<sup>a</sup>, André Schneider<sup>a</sup>, Valérie Sappin-Didier<sup>a</sup>

<sup>a</sup> ISPA, INRA, Bordeaux Sciences Agro, 33140, Villenave d'Ornon, France

<sup>b</sup> Arvalis-Institut du végétal, Station Expérimentale, 91720, Boigneville, France

#### **1. Abstract**

Regulatory limits on cadmium (Cd) content in food products are tending to become stricter, especially in cereals, which are a major contributor to dietary intake of Cd by humans. This is of particular importance for durum wheat, which accumulates more Cd than bread wheat. The contamination of durum wheat grain by Cd depends not only on the genotype but also to a large extent on soil Cd availability. Assessing the phytoavailability of Cd for durum wheat is thus crucial and appropriate methods are required. For this purpose, we propose a statistical model to predict Cd accumulation in durum wheat grain based on soil geochemical properties related to Cd availability in French agricultural soils with low Cd contents and neutral to alkaline pH (soils commonly used to grow durum wheat). The best model is based on the concentration of total Cd in the soil solution, the pH of a soil CaCl<sub>2</sub> extract, the cation exchange capacity (CEC), and on the content of manganese oxides (Tamm's extraction) in the soil. The model variables suggest a major influence of cadmium buffering power of the soil and of Cd speciation in solution. The model

successfully explains 88% of Cd variability in grains with, generally, below 0.02 mg Cd kg<sup>-1</sup> prediction error in wheat grain. Monte-Carlo cross-validation indicated that model accuracy will suffice for the European Community project to reduce the regulatory limit from 0.2 to 0.15 mg Cd.kg<sup>-1</sup> grain, but not for the intermediate step at 0.175 mg Cd.kg<sup>-1</sup>. The model will help farmers assess the risk that the Cd content of their durum wheat grain will exceed regulatory limits, and help food safety authorities test different regulatory thresholds to find a trade-off between food safety and the negative impact a too strict regulation could have on farmers.

Keywords: Cadmium, Durum wheat, Soil contamination, Statistical predictive model, chemical extraction

## **2. Introduction**

Cadmium is a non-essential, highly toxic metal, and was classified by the International Agency for Research on Cancer (IARC, 1993) as carcinogenic for humans. Chronic exposure to Cd, including through food, can cause damage to different organs as well as cancer (Cho et al., 2013; Jancic and Stosic, 2014; Nawrot et al., 2006). Cd occurs naturally in agricultural soils and its concentration generally depends on the parent material (Kabata-Pendias, 2004). Although Cd contents are usually low in agricultural soils, they sometimes tend to increase due to anthropogenic inputs, including atmospheric deposition and mineral fertilizers (Hutton, 1983). Despite these low or moderate contents, Cd can represent a threat to human health because the main pathway of exposure for the general population is consuming food made from contaminated plant products (Satarug et al., 2003). Therefore, in Europe, some regulatory limits have been established for maximum Cd content in some but not all foodstuffs (European Union, 2006). For cereal grains except rice and wheat, the maximum Cd content is 0.1 mg kg<sup>-1</sup> (15% water content). For rice and wheat, the limit is 0.2 mg.kg<sup>-1</sup> (European Union, 2006). Wheat is one of the most widely produced and consumed cereals worldwide (FAO, 2015) and, among wheat subspecies, durum wheat (*Triticum turgidum* subsp. *durum*) has been shown to accumulate much more Cd than bread wheat (Li et al., 1997; Stolt et al., 2003). Consequently, durum wheat, which is widely consumed as semolina and pasta, is an important contributor to the dietary intake of Cd by European populations. Following the recommendation of the European Food Safety Authority to reduce the dietary exposure of humans to Cd almost 3 fold (EFSA, 2009), the European Union proposed to reduce the maximum Cd content of durum wheat grain from 0.2 to 0.15 mg kg<sup>-1</sup> with an intermediate limit of 0.175 mg kg<sup>-1</sup> (DG-SANCO, 2011). These new limits call for simple predictive models to assess the risk of contamination of durum wheat grain by Cd.

Cadmium uptake by plants mainly depends on the plant genotype, but also on Cd availability in soils, which is driven by different mechanisms, and which, in turn, can be affected by agricultural

practices including crop rotation and tillage (Oliver et al., 1993). Total cadmium content in the soil is usually a poor indicator of the quantity of metal transferred to plants, as it is well established that only part of the total content can be transferred from the solid phases of the soil to the soil solution from which it is taken up by roots as free  $\text{Cd}^{2+}$  (Campbell, 1995). Exchanges of Cd between the soil solid phase and the soil solution are controlled by different mechanisms including ion exchange, specific adsorption or complexation involving chemical bonds of variable strength like Van der Waals interactions or covalent bonds. Clay minerals, organic matter, metallic oxides or oxi-hydroxides are the main constituents involved in the sorption of Cd onto the soil solid phase. In alkaline pH conditions, the availability of Cd for root uptake can decrease due to the precipitation of the metal, particularly with carbonates. In solution, Cd speciation is governed by complexation with organic and inorganic ligands. Protons and other cations (mainly  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ ) compete with  $\text{Cd}^{2+}$  for both sorption and complexation in solution (Del Castillo and Chardon, 1995; Naidu et al., 1994; Ross, 1994).

In France, durum wheat is usually cultivated on alkaline and calcareous soils with low Cd contents (median of  $0.3 \text{ mg Cd kg}^{-1}$  dry soil) (Arvalis, personal communication, <https://www.english.arvalisinstitutduvegetal.fr/index.html>). Although Cd phytoavailability could be expected to be lower due to the high pH values and carbonate contents, durum wheat crops can accumulate Cd in grains at levels close to the regulatory limit (Arvalis, personal communication). This means that a significant part of the soil Cd is probably solubilised and taken up by wheat, but is not taken into account by the statistical models used to predict the soil-solution partitioning of total or free Cd (Degryse et al., 2009). In the literature, plant available Cd is typically based on a single extraction or on sequential extractions, using chemical extractants that differ in their ability to selectively extract metals bound to different soil fractions (Harmsen et al., 2005; Lebourg et al., 1996; Meers et al., 2007; Tessier et al., 1979). However, these extractions cannot perfectly mimic resupply of  $\text{Cd}^{2+}$  from the solid phase to the soil solution during depletion resulting from root uptake. Yet, because of the initial low concentration of soluble  $\text{Cd}^{2+}$ , resupply is expected to play an important role in soils with alkaline pH and should therefore also be taken into account in the models. This can be accomplished by testing the cationic exchange capacity and the clay and organic matter contents as explanatory variables in the model used to predict Cd accumulation in the grain. Another possibility is to test the predictive value of the diffusion gradient in the thin film (DGT) technique developed by Zhang et al. (2001). DGT is a passive device that acts, like a root, as a sink for  $\text{Cd}^{2+}$ . It creates Cd depletion at the interface with the soil solution and can hence account for resupply from the solid phase. Finally, in the modelling approach, competition for sorption and complexation with soil cations can be included by considering the pH and the concentrations of Ca and Mg in the soil solution.

Most statistical models which predict Cd content in wheat grain from soil variables are designed for bread wheat (Adams et al., 2004; Baize et al., 2009; Brus et al., 2005; François et al., 2009;

Gray et al., 2001; Hough et al., 2003; Liu et al., 2015; Mench et al., 1997; Nan et al., 2002; Ran et al., 2016; Zogaj and Düring, 2016), whereas few are devoted to durum wheat (Garrett et al., 1998; Norvell et al., 2000). Furthermore, in the latter above-cited durum wheat models, the goodness of fit ( $r^2$  of 0.74 and 0.84, respectively) was likely linked to specific soil characteristics. Garrett et al. (1998) postulated that in their soils, Cd availability was mainly controlled by Cd-organic matter complexes and their model only included the Cd and the soil organic C extractable by pyrophosphate. In the study by Norvell et al. (2000), the authors reported highly variable soil salinity in their soils. The first goal of the present work was thus to extend the number of candidate variables to build a predictive model of the availability of Cd for durum wheat. In addition to classical extractions and soil characteristics (soil total Cd content, clay, loam and organic carbon, pH), we tested the concentration of Cd in the soil pore water, the free ionic fraction of Cd in solution, contents of calcareous and amorphous material, the Cd accumulated by DGT during a 48-h test, dissolved organic carbon, and finally concentrations of soluble Ca, Mg and K. These variables were tested in multiple linear models to account for soil-solution Cd partitioning, the buffering capacity of the soil, speciation in solution, the carbonated fraction of Cd, and competition with major cations in the soil. The second goal of this work was to assess the predictive value of the models by cross validation in order to estimate the error of prediction with respect to the current project of lowering the regulatory limits from 0.2 to 0.15 mg Cd/kg grain with an intermediate step at 0.175 mg Cd/kg (DG-SANCO, 2011).

### **3. Materials and methods**

#### **3.1. Sampling and preparation**

Nine French agricultural sites were chosen for this study. Each site corresponded to a cultivar comparison trial with three blocks, of 1000m<sup>2</sup>, per cultivar. Paired samples of topsoil and wheat (*Triticum turgidum* subsp. *Durum*, Miradoux) were collected at the heading stage (around 60 days after sowing) from each of the three blocks. The soil was sampled in the 0–30 cm ploughed layer where most of the root system developed. A total of 27-paired samples of soil and wheat were used. Soil samples were air dried and sieved at 2 mm. The residual water content was determined after drying at 105 °C until constant weight (NF ISO 11465:1994).

The plants were also sampled at maturity (around 90 days after sowing) to determine the grain Cd content and in the immediate vicinity of the previous sampling (soil and plants) so that soil conditions are assumed to be homogeneous. The plants were divided into shoots (stems and leaves) and grains for the last sampling. Four to five plants were collected for each sampling and a composite sample was made. Plant parts were carefully rinsed with permuted water and then with distilled water. These samples were dried at 60 °C for 48 hours to constant weight and ground (< 1 µm) using a Retch Planetary Ball Mill PM 400. Aliquots of powdered samples were digested

using a mixture of HNO<sub>3</sub> 69% (Aristar® for trace analysis, VWR Chemicals) and H<sub>2</sub>O<sub>2</sub> 30% (Ultratrace®, ppt-trace analysis grade, Scharlau) and the Cd in the extracts was quantified by furnace atomic absorption spectrometry (PinAAcle 900T, PerkinElmer). Grains sampled at maturity were air dried to a final constant water content of 15% as stated for the regulatory limit for Cd. The Cd contents of grains were determined by atomic absorption spectrometry after wet digestion (<http://www.capinov.fr/>).

### **3.2. Soil characteristics**

The following parameters were determined from air-dried soil subsamples according to French standardized procedures. Soil pH was determined in a distilled water (1:5 soil to solution ratio) and 0.01 M CaCl<sub>2</sub> extract (1:10 soil to solution ratio) (NF ISO 10390:2005). Total CaCO<sub>3</sub> content was quantified using the acid neutralisation method (NF X 31-105). The total soil organic carbon fraction (SOC) was determined by dry combustion after correction for carbonate (NF ISO 10694:1995). The soil texture was determined using the Robinson pipette method (NF X 31–107) and the cationic exchange capacity (CEC) was measured at soil pH using the cobaltihexamine chloride method (NF X 31–130). The concentration of dissolved organic carbon (DOC) was measured in the 0.01 M CaCl<sub>2</sub> extract by oxidative combustion (TOC-VCSH, Shimadzu).

### **3.3. Soil and soil solution extraction and analyses**

Total Cd soil content was quantified after solubilisation by fluorhydric and perchloric acids (NF X 31–147). After complete dissolution, the concentrations of Ca and Mg in the extracts were also analysed by ICP-AES (inductively coupled plasma atomic emission spectrometry) and in Cd by ICP-MS (inductively coupled plasma mass spectrometry).

Seven soil extractions were performed, the procedures, reagents and soil:solution ratios are detailed in Table IV.1. At the end of the extraction, the soil suspensions were centrifuged at 7,500 rpm at 20 °C for 15 min and the supernatants were filtered through a 0.2 µm cellulose acetate filter. The Cd was then quantified by ICP-MS, and Mn, Al and Fe on Tamm's extracts by ICP-AES. All trace metal analyses and soil characterisations were performed by the INRA soil-testing laboratory (<http://www6.lille.inra.fr/las>).

*Table IV.1: Summary of the extraction methods used as single extractants to evaluate the phytoavailability of Cd for durum wheat.*

Extracting solution	References	Soil:solution ratio (w:v)	Extraction time (h)
---------------------	------------	---------------------------	---------------------

HF/HClO <sub>4</sub> acid extraction (total extraction)	NF X 31-147	-	-
0.05M EDTA at pH 7	Ure, 1996	1:10	1
1M sodium acetate adjusted to pH 5 with acetic acid	Tessier et al., 1979	1:8	3
0.1M Na <sub>4</sub> P <sub>2</sub> O <sub>7</sub> (pyrophosphate) at pH 10	Loveland & Digby, 1984	1:100	16
0.2M oxalic acid + 0.2M ammonium oxalate (Tamm's reagent) at pH 3	McKeague & Day, 1966	1:100	4
1M NH <sub>4</sub> NO <sub>3</sub> at pH 5	Symeonides & McRae, 1977	1:2.5	1
0.01M CaCl <sub>2</sub> at soil pH	Houba et al., 1990	1:10	2
Ultra-pure (UP) water at soil pH	Reported by Lebourg et al., 1996	1:5	1

The soil pore solution was obtained by centrifugation as described by Schneider (2006). Briefly, soils were moisturised to pF 1.7, using the model from Jong et al. (1983) to calculate the right amount of ultra-pure water to add. After equilibrating overnight, approximately 12 ml of the non-water miscible organic solvent C<sub>2</sub>F<sub>3</sub>Cl<sub>3</sub> (ACS, Reag. Ph. Eur, Merck) were added to displace the soil solution during centrifugation at 13,000 rpm for 1 h. The aqueous supernatants were carefully separated from the solvent and filtered through a 0.2 µm cellulose acetate filter.

These solutions were used to determine the pH using a microelectrode (pHC4000-8, Radiometer analytical), and to quantify the major cations, Ca, Mg and K by flame-AAS, and Cd by furnace-AAS (Solaar M6, TJA Solutions). The free fraction of Cd (Cd<sup>2+</sup>) in these soil solutions was measured using the ion-exchange method developed by Schneider (2006). This method is based on the Ca/Cd exchange properties of a cation exchange resin (Amberlite IR-120, Aldrich) converted to Ca form. The soil pore-solution (around 2 ml) was mixed with the Ca-resin at a 500:1 v:w solution:solid ratio (determined precisely) and shaken at 20 °C for 24 h, after which 0.2 ml of the soil pore-solution pH was sampled to determine the concentration of Ca after contact. Next, a known amount of radioactive <sup>109</sup>Cd (Perkin-Elmer) was added to the resin:solution mixture. The vials were shaken at 20 °C for 24 h, after which the radioactivity remaining in solution was quantified in a 0.5 ml aliquot of solution by liquid scintillation counting (Tri-Carb 2100TR, Packard BioScience), after addition of 2.5 ml of scintillation cocktail (Instagel-Plus, Packard BioScience) and correction for quenching. The fraction of free Cd (F<sub>Cd</sub>) was calculated from the resin Vanselow selectivity coefficient for Cd/Ca exchange (<sup>v</sup>K<sub>Cd/Ca</sub>), using Equation IV.1 :

$$F_{Cd} = \frac{\left(\frac{R}{r} - 1\right) \times C_{Ca} \times V}{{}^v K_{Cd} \times R_{Ca,f} \times m} \times 100 \quad \text{Equation IV.1}$$

where  $R/r$  is the ratio of added ( $R$ ) to remaining ( $r$ ) radioactivity,  $C_{Ca}$  is the concentration of calcium in the solution after contact,  $V/m$  is the soil solution:resin ratio and  $R_{Ca,f}$  is the content of Ca on the resin after contact calculated by Equation IV.2.

$$R_{Ca,f} = R_{Ca,i} + (C_{Ca,i} - C_{Ca,eq}) \times \frac{V}{m} \quad \text{Equation IV.2}$$

Free Cd ( $Cd^{2+}$ ) concentrations were calculated using this fraction of free Cd and the concentration of total Cd in the soil pore-solution obtained by centrifugation, with Equation IV.3 :

$$[Cd^{2+}] = \frac{F_{Cd}}{100} \times [Cd]_{porewater} \quad \text{Equation IV.3}$$

### **3.4. DGT experiments**

DGT experiments were performed in the laboratory using fresh soils stored at 4 °C. Approximately 700 g of fresh soil were weighed and ultra-pure water was added to reach approximately 50% of the maximum water holding capacity ( $WHC_{max}$ ). The soils were incubated at 28 °C for 24 h in closed plastic containers. Next, the soils were moistened to 100% of their  $WHC_{max}$  and allowed to equilibrate at 28 °C in the dark for 24 h. Then, three DGT devices were applied to each soil, avoiding any free air space between the filter and the soil surface. The containers were closed and left at 28 °C in the dark for 48 h before collecting the DGTs. The 48-h period of contact between the DGT and the soil was chosen to obtain a quantifiable amount of Cd on the resin while minimising changes in the soil conditions during the contact (Harper et al., 1998). Each resin was individually bathed in 2 ml of 1 M  $HNO_3$  for 24 h to eluate the sorbed Cd. In addition, after the DGTs had been used, the soil solutions from the containers were sampled with Rhizon microcups (Rhizosphere Research Products, Wageningen) inserted horizontally. Cadmium in the soil solutions and in the resin eluates was quantified using furnace-AAS (PinAAcle 900T, PerkinElmer).

An indicator of the Cd buffering power of the soil can be derived from DGT data (Zhang et al., 2001). Indeed, the ratio ( $b$ ) between the calculated Cd concentrations at the interface between the gel and the resin of the DGT ( $C_{DGT}$ ) and the Cd concentrations in the soil solution obtained by the rhizons ( $C_{rhizon}$ , Equation IV.4) reflects the capacity of the solid phase to resupply the soil solution with Cd.

$$b = \frac{C_{DGT}}{C_{rhizon}} \quad \text{Equation IV.4}$$

If the resupply is fast, and the solid phase contains a large amount of mobile Cd, the concentration at the interface will not diminish significantly and  $b$  will be close to one. Conversely, if little is resupplied from the solid phase,  $C_{DGT}$  will tend to zero because the resin behaves as an infinite sink and  $b$  also tends to zero (Harper et al., 1998; Zhang et al., 2001).

### 3.5. Testing statistical models from the literature

We conducted an extensive search of the literature to find statistical models based on soil properties for the prediction of Cd contents in wheat grains that could be tested on our dataset. The following models were found for bread wheat (Adams et al., 2004; Baize et al., 2009; Brus et al., 2005; François et al., 2009; Gray et al., 2001; Hough et al., 2003; Liu et al., 2015; Mench et al., 1997; Nan et al., 2002; Ran et al., 2016; Zogaj and Düring, 2016) and for durum wheat (Garrett et al., 1998; Norvell et al., 2000). We discarded all the models which included soil variables not measured in our study. Finally, four models were retained and are detailed in Table IV.2. These models include total Cd, pH and in one model, soil organic matter, all these variables being highly relevant to explain Cd partitioning from soil to solution (Degryse et al., 2009; Sauvé et al., 2000a). Unfortunately, the selected models were developed for bread wheat, but we hypothesised that if the models are valid, the effect of the species (bread wheat versus durum wheat) would possibly affect the slope of the relationship between the predicted and the measured Cd content in grains, but not the strength of the relationship.

Table IV.2: Models found in the literature to predict Cd contents in bread wheat grains. *N* is the number of observations used to parameterize the model;  $r^2_{adj}$ : adjusted *r*-squared;  $\alpha$  is the slope of the regression between observed and predicted values for each model; RMSE: root mean square error;  $r^2$ : *r*-squared of the regression between observed and predicted values.

Model	Reference	N	$r^2_{adj}$	$\alpha$	RMSE (mg.kg <sup>-1</sup> )	$r^2$
$\log_{10}(Cd_{grain}) = 0.28 + 0.44 \log_{10}(Cd_{total}) - 0.18 \text{ pH}$	Adams et al., (2004)	162	0.49	1.81	0.033	0.42
$\log_{10}(Cd_{grain}) = 1.022 + 0.749 \log_{10}(Cd_{total}) - 0.257 \text{ pH} - 0.277 \log_{10}(\text{SOM})$	Brus et al., (2005)	84	NA	2	0.041	0.44
$\log_{10}(Cd_{grain}) = 1.386 + \log_{10}(Cd_{total}) - 0.279 \text{ pH}$	Liu et al., (2015b)	14	0.85	0.67	0.029	0.45
$\log_{10}(Cd_{grain}) = 0.703 + 1.04 \log_{10}(Cd_{total}) - 0.175 \text{ pH}$	Ran et al., (2016)	99	0.61	0.47	0.038	0.44

### 3.6. Statistical modelling

The content of Cd in durum wheat grain was modelled by multiple linear regressions with the soil variables as regressors. The models tested included one to a maximum of four regressors to avoid over-parameterization. The rationale used to select the regressors in a hierarchical way



was the following. For the one-regressor models, we selected variables that could reflect the soil available Cd, namely, the Cd in soil derived from the different extractions (Table IV.1). For the two-regressor models, we crossed the previous models with the soil pH, measured in water, in CaCl<sub>2</sub> extract, or in the soil pore solution. Indeed, the proton competes with Cd for both the complexation and the sorption of Cd and is therefore expected to be among the most influential parameters in soils (Adriano, 2001). The 3- and 4-regressor models crossed the 2-regressor models with one or two additional variables, respectively. The additional variables were related to sorption, speciation in solution and competition for sorption and complexation: SOC, contents of carbonates and amorphous oxides, CEC, texture, DOC, and cations and anions in solution. A total of 3,065 models were tested.

The models were individually tested by leave-one-out cross-validation (LOOCV). The models were parameterized on (N-1) observations and the remaining observation was used to calculate the error of prediction. The process was repeated N=27 times and the root mean square error of predictions (RMSE<sub>LOOCV</sub>) was then calculated for each model. Models were evaluated using the following criteria: adjusted multiple correlation coefficients ( $r^2_{adj}$ ) calculated from the multiple linear regressions and from the simple linear regression between predicted and observed values, the Bayesian information criterion (BIC) to penalise over-parameterization of the models, the RMSE<sub>LOOCV</sub> and the variance inflation factor (VIF) to detect multicollinearity between predictors. In order to test if the dataset used to parameterize was sufficiently big, a second cross-validation was performed based on the approach of Shao (Shao, 1993). This procedure involves the random selection of  $n_v$  data to validate the model, while the remaining data (N- $n_v$ ) are used to parameterize the model. The size of validation subset ( $n_v$ ) ranged from 1 to a maximum of 21. This left a minimum of 6 observations for parameterization of the largest models with 4 regressors. The case

where  $n_v=1$  corresponds to the leave-one-out cross-validation. There were up to 20 million  $\binom{N}{n_v}$  subsets and a maximum of  $10^5$  distinct possibilities were consequently randomly chosen and tested. The RMSE was calculated for each validation subset and compared to the RMSE of the corresponding model parameterized on the whole dataset. All statistics and data treatments were conducted using R software version 3.3.2. (R Core Team, 2016).

## **4. Results and discussion**

### **4.1. Soil properties and Cd contents in plant materials**

All the soils were alkaline with pH values ranging between 8.0 and 8.7 (Table IV.3.a). From field surveys (Arvalis, pers. comm.), this range was estimated to cover the 60% upper range of the French agricultural soils used to cultivate durum wheat. For durum wheat cropped in France, soil

pH generally ranges from 6 to 8.7, meaning that 40% of the cases where pH is between 6 and 8.0 were not tested in our models, so our results should be extrapolated with caution to these situations.

The total Cd contents of the soils we studied ranged from 0.14 to 1.00 mg.kg<sup>-1</sup> (Table IV.3.b), which we consider covered about 80% of the soils under durum wheat, for which the Cd content ranged between 0.09 and 1.44 mg kg<sup>-1</sup> soil according to the above-mentioned field surveys (Arvalis, pers. comm.).

The other soil characteristics, including texture, CEC, organic carbon and calcareous contents, were typical of French agricultural soils and displayed the necessary variability so that these characteristics could be used as explanatory variables (Table IV.3.a).

Plant Cd contents ranged from 0.125 to 0.652 mg.kg<sup>-1</sup> DW (dry weight) in the shoots and from 0.009 to 0.144 mg.kg<sup>-1</sup> in the grains at 15% water content (Table IV.3.c). These values are within the range of Cd contents previously reported in durum wheat grains (Adams et al., 2004; Li et al., 2011) and are below the maximum admissible content (0.2 mg.kg<sup>-1</sup>) set by the European Commission (European Union, 2006)

1 Table IV.3: Summary of the characteristics of the soils used in this study (a), of the Cd pools estimated by a single extraction (b) and of Cd measured in different wheat organs, roots, shoots and grains (c).

3 (a) Soil characteristics

Value	Soil pH	pH Ca Cl <sub>2</sub>	pH pore water	Granulometry			CEC cmol <sup>+</sup> /kg	SOC g/kg	CaCO <sub>3</sub> g/kg	DOC mg/l	Ca pore water mmol/l
				Clay	Silt	Sand					
					%						
Min.	8	6.8	8.2	18.6	15.2	11	10.7	9.21	27.1	33.9	2.02
Median	8.4	7.3	8.4	30.1	35.6	28.7	18.4	15.7	208.6	49.6	3.1
Max.	8.7	7.5	8.5	53	54	59.4	33.5	40.3	426	70.4	10.8

(b) Soil Cd content according to the extraction

Value	Cd HF μg/kg	Cd EDTA μg/kg	Cd sodium acetate μg/kg	Cd pyro- phosphate μg/kg	Cd Tamm μg/kg	Cd NH <sub>4</sub> N O <sub>3</sub> μg/kg	Cd CaCl <sub>2</sub> μg/kg	Cd UP water μg/kg	Cd DGT μg/kg	Cd pore water μg/l
Median	318	146	129.4	51.29	22.69	0.96	0.24	0.05	0.009	0.10
Max.	999	355	259.8	197.2	84.49	2.3	0.89	0.14	0.065	0.32

(c) Cd contents in different wheat organs

Value	Shoot Cd mg/kg DW	Grain Cd mg/kg <sup>a</sup>
Min.	0.125	0.009
Median	0.329	0.063
Max.	0.652	0.144

<sup>a</sup> Grain Cd contents are expressed at 15% water content

## **4.2. Is it possible to predict Cd content in durum wheat grains at maturity from shoot Cd content at the early heading stage?**

Early diagnosis of the contamination of the grain by Cd would be possible if the Cd content of wheat shoots at heading stage is correlated with that of grains at maturity. This correlation was positive and significant, but too weak to be used as a valid model of risk assessment ( $r^2 = 0.25$ ,  $p$ -value  $< 0.05$ ). From the plant physiological point of view, the weak correlation between the Cd in the shoot at heading and the Cd content of the grain at maturity could mean that the Cd accumulated in the grains does not always originate from remobilisation from shoots but could also be the result of direct uptake during post-heading growth. The balance between remobilisation and uptake for the allocation of Cd to the grain could mirror that of Zn, due to the fact that these two elements could behave similarly in plants (Khan et al., 2014). Remobilisation of Zn from vegetative tissues to the grain is favoured when the availability of Zn for root uptake is low (Sperotto, 2013). Therefore, assuming that Cd and Zn behave similarly, we introduced the indicator of the buffer power for Cd  $b$  in the regression model between the grain Cd at maturity and the Cd in the shoot at heading, and the goodness of fit increased greatly ( $r^2 = 0.78$ ,  $p$ -value  $< 0.001$ ,  $N=26$ ).

$$Cd_{grain} = -0.047^{**} + 0.242^{***} Cd_{shoots} + 0.230^{***} b - 0.405^{**} Cd_{shoots} \times b \quad \text{Equation IV.5}$$

The sign of the coefficients of Equation IV.5 are consistent with the above-mentioned hypotheses. The coefficients of  $Cd_{shoots}$  and of  $b$  are positive because they could be a source of Cd from remobilisation and uptake, respectively. The negative coefficient of the interaction between  $Cd_{shoots}$  and  $b$  could mean that the remobilisation of Cd from shoots to grains is greater in soils with low buffering power for Cd (low  $b$ -values). Conversely, when soils have high buffering power for Cd, grain Cd might mainly result from direct uptake and remobilisation could be minimum.

## **4.3. Soil Cd extractions**

In unpolluted agricultural soils with low Cd contents, the availability of Cd and not the root uptake capacity generally limits accumulation of the metal by the plant (Lin et al., 2016). Consequently, different extractions were tested to identify the one that best reflects the amounts of Cd potentially available to durum wheat: the amounts of Cd recovered in the soil pore- solution, by DGT and Rhizon samplers and by different chemical extractants applied separately and not sequentially (Table IV.3.b).

As expected, the least Cd was extracted by ultra-pure (UP) water, from 0.005 to 0.04% of the total Cd (HF/HClO<sub>4</sub> extraction) since water only extracts Cd weakly bound to the soil matrix. The

0.01 M CaCl<sub>2</sub> solution extracted three times more Cd than the ultra-pure water because this salt also extracted Cd sorbed onto the soil matrix by ion exchange due to the substitution of Cd by Ca (Houba et al., 1990). NH<sub>4</sub>NO<sub>3</sub> extracted on average 4.1 times more Cd than CaCl<sub>2</sub>. The chemical mechanisms reported to explain the action of 1 M NH<sub>4</sub>NO<sub>3</sub> are a moderate drop in pH values (around pH 5) and the formation of metal amine complexes, which is favoured in alkaline soils due to the dissociation of NH<sub>4</sub><sup>+</sup> (Gryschko et al., 2005). The high concentration of NH<sub>4</sub><sup>+</sup> strongly competes with cations for sorption (Barrow, 1986) and consequently favours Cd desorption. NH<sub>4</sub>NO<sub>3</sub> extraction also estimates the content of exchangeable-Cd but extracted more exchangeable-Cd due to the higher molarity of the competitive cation NH<sub>4</sub><sup>+</sup> (Table IV.1). Tamm's reagent was chosen for its ability to dissolve amorphous metallic oxides and to release the Cd sorbed on these constituents into solution. This extractant retrieved from 0.08% to 21.3% of the total Cd, much less than the Cd extracted by the sodium-pyrophosphate solution, which ranged from 2.8% to 47.4% of the total Cd. In addition to the Cd weakly sorbed onto the soil, sodium-pyrophosphate enables the extraction of Cd complexed with organic matter and can slightly dissolve amorphous metallic oxides. This shows that in our cases, complexation of Cd by the solid organic matrix was likely to dominate. This is in agreement with the literature on the modelling of the soil-solution partitioning of Cd which clearly shows that soil organic carbon is a much more frequent significant regressor than oxides (de Vries et al., 2011; Degryse et al., 2009; Groenenberg et al., 2010b; Sauvé et al., 2000a).

Sodium acetate/acetic acid at pH 4.5 released more Cd, from 25% to 55% of the total Cd, than the previous extractants. This can be explained by the fact that, in addition to the weakly sorbed Cd, this extractant dissolved Cd-carbonates, which were likely an important fraction because of the high pH and calcareous contents of the soils. Finally, the 0.05 M EDTA solution extracted the most Cd, i.e., from 35% to 64% of the total Cd. EDTA is a strong complexing ligand anion which can chelate the Cd complexed with organic matter and with metallic oxides (Ure, 1996), like sodium pyrophosphate but with a greater extraction strength.

The soil pore water solutions collected by centrifugation contained higher concentrations of Cd than the solutions collected by the Rhizons, 118 versus 56.5 ng Cd.l<sup>-1</sup>, respectively. One possible explanation for this difference could be the length of the period of equilibration after rewetting the soil: overnight for the centrifugation method compared to four days for the Rhizon method. Rhizons extract water poorly held by the soil constituents, mainly the gravitational water of the macro porosity, whereas centrifugation also extracts the capillary water. Due to the buffering capacity of the solid phase and due to soil tortuosity, four days after rewetting may not have been long enough to reach equilibrium between the Cd concentration in the capillary water and in the water in the macroporosity. Furthermore, rewetting a soil generally results in a transient burst of microbial activity accompanied by the release of soluble organic matter (Merckx et al., 2001), which can complex and therefore desorb part of the Cd sorbed onto the solid phase. The soils

used for the extraction by centrifugation were incubated wet for a shorter time than the soils used for Rhizon extraction. Consequently, the former might have more soluble organic ligands that may have desorbed more Cd (Cornu et al., 2007).

The free fractions of Cd measured in soil solutions obtained by centrifugation ranged from 18% to 78% with a median of 44% corresponding to Cd<sup>2+</sup> concentrations ranging from 14.53 to 147.50 ng.l<sup>-1</sup>. The high values obtained for the free fractions were not expected because of the alkaline pH values. They might be explained by the high concentrations of competitive cations including Ca compared to the concentration of ligands.

#### ***4.4. Testing models from the literature for the prediction of grain Cd in durum wheat***

The models from the literature tested on our dataset (Adams et al., 2004; Brus et al., 2005; Liu et al., 2015; Ran et al., 2016) explained less than half the variability of the observed grain Cd contents, and underestimation by the models was not systematic, as we had expected from the fact that they were parameterized for bread wheat (Figure IV.1 a to d). This shows that the soil total Cd, pH, and soil organic matter can enable a rough estimate of how Cd in grain is expected to change when these variables vary, but that the parameterization in the literature was not robust. Several explanations are possible. First, the range of data used for their calibration may not cover the soils tested in our work. Indeed, the model proposed by Brus et al. (2005) was parameterized on soils with SOM contents ranging from 1.4 to 42 g.kg<sup>-1</sup> and tested on soils with SOM contents ranging 15.9 to 69.8 g.kg<sup>-1</sup>. The accumulation of Cd in grain also depends on the plant genotype, which is not explicitly taken into account in the models in the literature, as they were parameterized for bread wheat, whereas we tested them on durum wheat. Furthermore, within the same subspecies (bread wheat or durum wheat), differences in accumulation between cultivars can reach a x2 to x3 factor (Perrier et al., 2016; Zimmerl et al., 2014). Finally, in models only based on total Cd, pH and soil organic carbon, the speciation and the buffer capacity of Cd in the soil are poorly taken into account. Total Cd is unlikely to always be fully mobile. Only pH and soil organic carbon could be partly linked with the speciation of Cd in solution but these variables also reflect sorption and desorption.

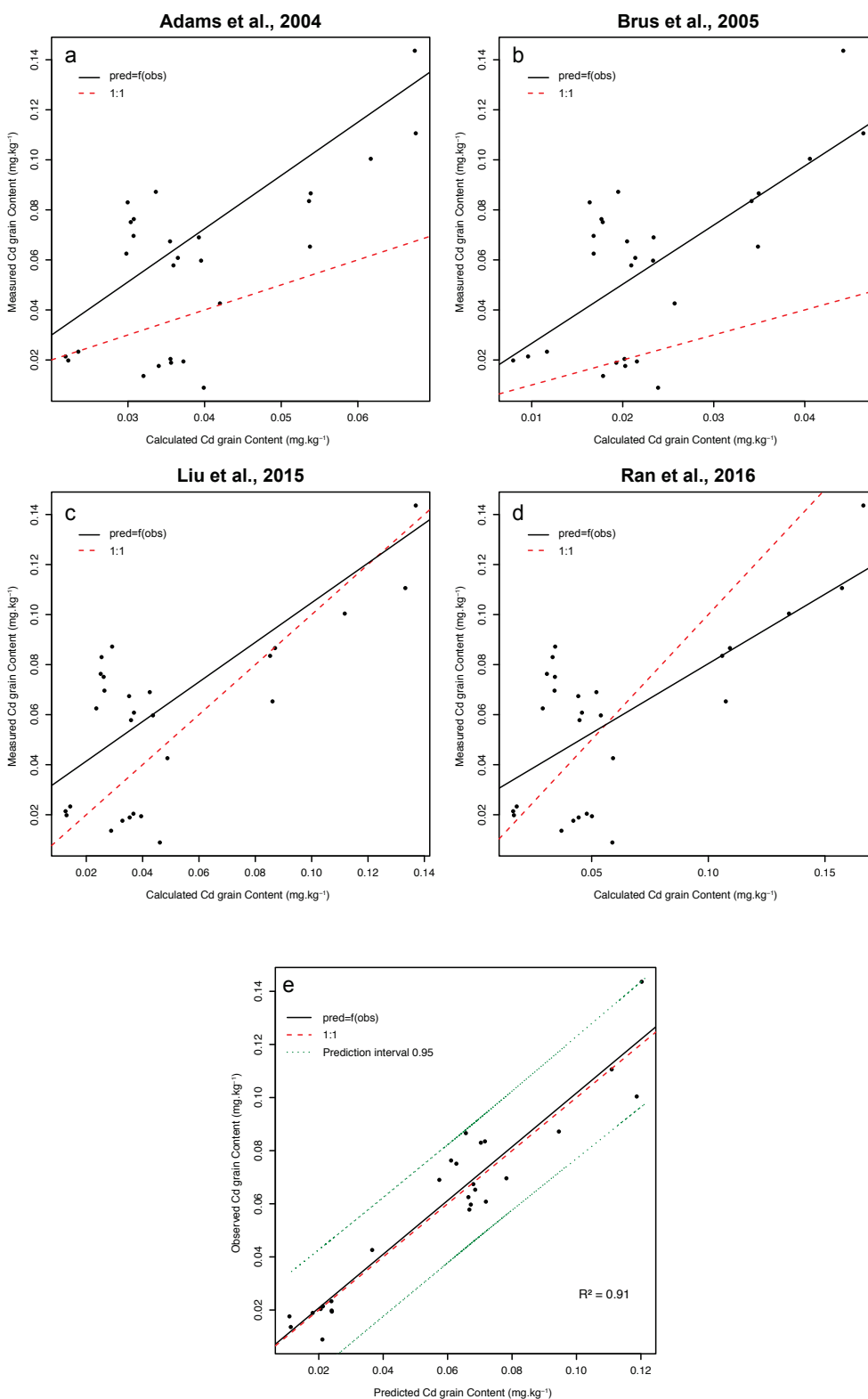


Figure IV.1: Comparison of grain Cd contents in this study and the corresponding predictions made by the model of Adams et al. (2004) (a), Brus et al. (2005) (b), Liu et al. (2015) (c) and Ran et al. (2016) (d) and the best model from this study (Equation IV.8) (e). The black line represents the regression between observed and predicted values; the red dashed line is the 1:1 line; the green dotted line represents the prediction interval at 95%. All grain Cd data are expressed at 15% water content

### **4.5. Statistical models to predict Cd in durum wheat grain at maturity**

Among the one-variable models, those with HF/HClO<sub>4</sub>, EDTA, sodium acetate and Tamm extracted Cd showed poor goodness-of-fit with r-squared values around 0.15 (Table IV.4). The Cd recovered in the soil pore-water by centrifugation or by the Rhizon samplers and even the concentration of the free ion Cd<sup>2+</sup> did not produce good models (Table IV.4).

Intermediate goodness of fit was obtained for the one-variable models based on the Cd recovered by the DGT ( $r^2 = 0.40$ ,  $p$ -value < 0.001) or extracted by water, 0.01 m CaCl<sub>2</sub> or pyrophosphate solutions (Table IV.4).

*Table IV.4: Adjusted r-squared for the linear regressions between the grain Cd (log) contents and the Cd contents (log) recovered using different extraction methods.*

Extraction method	$r^2_{adj}$
HF/HClO <sub>3</sub> acids	0.176*
0.05M EDTA	0.153*
Sodium acetate at pH 5	0.133*
Pyrophosphate at pH 10	0.579***
Tamm's reagent	0.164*
1 M NH <sub>4</sub> NO <sub>3</sub> at pH 5	0.624***
0.01 M CaCl <sub>2</sub>	0.543***
Ultra-pure water	0.490***
DGT	0.397***
Pore-water by centrifugation	0.467***
Free Cd (Cd <sup>2+</sup> )	0.493***

Statistical significance: \*\*\* for  $p$ -value < 0.001, \*\* for  $p$ -value < 0.01, \* for  $p$ -value < 0.05

The 1 M NH<sub>4</sub>NO<sub>3</sub> solution was the best extractant, reflecting Cd bioavailability for durum wheat in the soils studied here ( $r^2 = 0.62$ ,  $p$ -value < 0.001). The effectiveness of 1 M NH<sub>4</sub>NO<sub>3</sub> extraction in correlating with Cd accumulation in the plant has also been reported for wheat, carrot, spinach and lettuce (Knoche et al., 1997 reported in Gryschko et al., 2005). As a whole, even if a particular one-variable model can give a satisfactory goodness of fit, the robustness of these simple models



is -or can be expected to be- low because the complexity of the mechanisms governing the phytoavailability of Cd in soils.

The best two-variable model included Cd from the DGT and the CaCl<sub>2</sub> pH ( $r^2 = 0.66$ , Equation IV.6, Table IV.5), which showed that adding the CaCl<sub>2</sub> pH to the Cd from the DGT notably improved the predictive value of the model. One possible explanation could be that, in alkaline soil, roots can reduce the pH in the rhizosphere to increase the availability of nutrients (Hinsinger et al., 2003) which is not taken into account if only the Cd DGT is used in the model.

When the number of regressors was increased to three, the  $r^2$  of the best model was 0.79 (Equation IV.7, Table IV.5).

The model with the best predictive value included the total Cd in the soil solution obtained by centrifugation, the pH measured in the CaCl<sub>2</sub> extract, the CEC and the amount of Mn extracted by Tamm's reagent (Equation IV.8, Table IV.5). This model explained 88% of the total variance of the grain Cd contents. The variance inflation factor (VIF) did not identify strong correlations between the regressors following the rule of thumb that suggests examining collinearity for value above 2.

*Table IV.5: Best models to predict Cd contents in durum wheat grains and associated statistics:  $r^2_{adj}$  adjusted r-squared; RMSE: root mean square error;  $RMSE_{LOOCV}$  root mean square error from the leave-one out cross-validation;  $VIF_{max}$ : maximum value of the variance inflation factor; N number of observations.*

Eq.	Model	$r^2_{adj}$	RMSE (mg.kg <sup>-1</sup> )	$RMSE_{LOOCV}$	$VIF_{max}$	N
Equation IV.6	$\log_{10}(Cd_{grain}) = 7.359 + 0.697^{***}$ $\log_{10}(Cd_{DGT})$ $- 1.014^{***} \text{ pH CaCl}_2$	0.66	0.0219	0.0254	2.2	26
Equation IV.7	$\log_{10}(Cd_{grain}) = -3.584^{***} + 0.339^*$ $\log_{10}(Cd^{2+})$ $+ 1.722^{***} \log_{10}(CEC) - 0.864^{***}$ $\log_{10}(Mn_{Tamm})$	0.79	0.0184	0.022	3.0	27
Equation IV.8	$\log_{10}(Cd_{grain}) = -1.979 + 0.553^{***}$ $\log_{10}(Cd_{pore\ water}) - 0.347^* \text{ pH CaCl}_2$ $+ 1.630^{***} \log_{10}(CEC)$ $- 0.685^{***} \log_{10}(Mn_{Tamm})$	0.88	0.0102	0.0125	2.2	27

Statistical significance: \*\*\* for  $p$ -value < 0.001, \*\* for  $p$ -value < 0.01, \* for  $p$ -value < 0.05

Figure IV.1.e shows the regression between observed grain Cd contents and predictions made by Equation IV.8. The slope is close to one (1.012) with a non-significant intercept ( $p$ -value = 0.9) and 91% of the variance was explained, showing that the model was accurate and not biased. Equation IV.8 had a RMSE of 0.013 mg.kg<sup>-1</sup>, which is quite low and accurate for the prediction of grain Cd contents whose values ranged from 0.09 to 0.144 mg. kg<sup>-1</sup>.

Figure IV.2 presents the RMSE values for each different size of the subset tested for validation ( $n_v$ ). Equation IV.8, when parameterized with N = 27 soils, produced a low value of RMSE (Table IV.5), represented by the red line in Figure IV.2. When the size of the parameterizing subset decreased ( $n_v$  increases), the prediction errors of the models increased. However, no excessive increase in the RMSE was observed with a validation subset of less than 13 soils (Figure IV.2). This suggests that when the model is parameterized with 14 soils and tested on the remaining 13, the model uncertainty remains acceptable. Consequently, 26 soils were considered to be enough to select variables to produce a robust model.

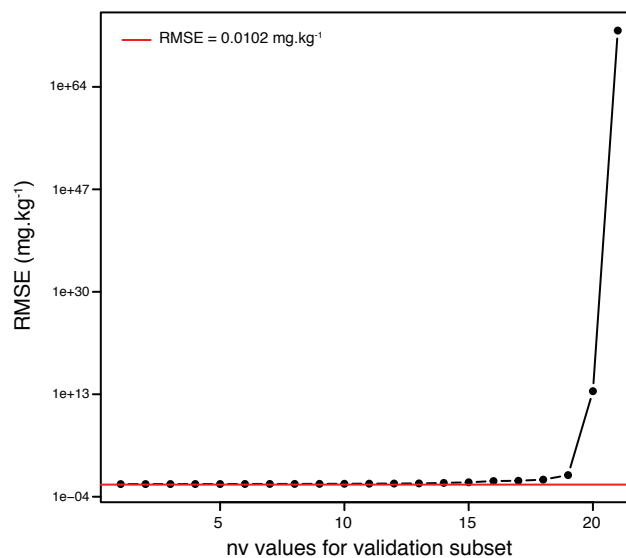


Figure IV.2: RMSE of predictions from the best model (Equation IV.8) for different sizes ( $n_v$ ) of validation subset from a Monte Carlo cross validation. The red line corresponds to the value of RMSE of the model parameterized with 27 soils ( $n_v = 0$ )

The contributions of each regressor of Equation IV.8 model to the total sum of squares were 38% for the concentrations of Cd in the pore water, 5.7% for pH CaCl<sub>2</sub>, 43.8% for the CEC and 12.5% for the Mn Tamm's contents. It is worth noting that the two main contributors are the initial availability of Cd in the soil solution (pore water Cd) and the CEC, which can be interpreted as the buffering capacity of the solid phases by ion exchange. Therefore, the importance of the CEC in the model is in agreement with the fact that, among the extractions tested, those that gave the best correlations with grain Cd were those that used low strength extractants. The importance of

the CEC in the model is also consistent with the mechanistic modelling of Cd phytoavailability, which showed that desorption is generally the main mechanism buffering Cd in the soil solution compared to dissociation of the complex, the latter being limited by kinetics (Lin et al., 2016). Therefore, the two main explanatory variables of the model with their positive coefficients could illustrate the fact that plant roots take up Cd from the soil solution, which is in part replenished by the solid phase. However, it is generally assumed that only the free Cd ion is taken up by plants (Campbell, 1995); this would mean that the total Cd in solution and the CaCl<sub>2</sub> pH in Equation IV.8 can be interpreted as variables representing the free Cd<sup>2+</sup> concentration in the soil solution because Cd speciation is strongly governed by pH. Following this rationale, we examined a model derived from Equation IV.8 by replacing the total Cd in solution and the CaCl<sub>2</sub> pH by the free Cd concentrations (Equation IV.7). The performances of the resulting model were significantly lower ( $F_{2,025}=18.05$ ,  $p\text{-value} < 0.001$  for the model comparison) with an r-squared that decreased from 0.88 to 0.79. This could mean that the total Cd in solution and the pH did not only represent the free Cd in solution. The dissociation of Cd complexes might have helped to buffer Cd<sup>2+</sup> availability during uptake of the metal by roots. As indicated above, the contribution of the complexes to the buffering of Cd<sup>2+</sup> in the solution is likely to have been much less than buffering by the CEC since the drop in the r-squared between Equation IV.7 and Equation IV.8 was only 0.09, whereas removing the CEC from Equation IV.8 results in a drop of 0.32 ( $r^2 = 0.56$ ,  $RMSE = 0.033 \text{ mg.kg}^{-1}$ ).

There was no improvement in the overall goodness-of-fit of Equation IV.8 when the DOC was added to the regression ( $r^2=0.87$ ,  $RMSE = 0.01 \text{ mg.kg}^{-1}$ ). This could mean that due to the relatively high buffering capacity of the solid phase (CEC) and therefore, the limited contribution of Cd complexes to Cd uptake, the concentration of organic Cd complexes associated with the DOC was not limiting. A second possible hypothesis is that the contributing complexes were not organic but inorganic, including those with HCO<sub>3</sub> (McLaughlin et al., 1999), maybe because the latter were more labile.

In Equation IV.8, when the pH measured in a 0.01M CaCl<sub>2</sub> was replaced by the pH H<sub>2</sub>O or by the pH of the soil pore water extracted by centrifugation, the  $r^2$  dropped from 0.88 to 0.84 with RMSE equal to 0.015  $\text{mg.kg}^{-1}$  for both. The slightly better predictive value of the CaCl<sub>2</sub> pH can be explained as follows: the CaCl<sub>2</sub> pH reflects the proton buffering power of the soil because it gives the amount of protons desorbed due to Ca competition (Miller and Kissel, 2010) and is therefore lower than the pH of H<sub>2</sub>O or the pH of the soil pore water obtained by centrifugation. The pH which influences Cd uptake by roots is the pH of the rhizosphere, which is affected by excretion of organic acids or protons by the roots resulting from the differential rates of cation and anion uptake and/or from micronutrient mobilisation, particularly in alkaline soils (Nye, 1981), like in our study. The pH of the rhizosphere can also change due to fluctuating environmental conditions, including drying of the soil during the grain-filling period which takes place in late spring and early summer

(June-July). Compared to the pH of H<sub>2</sub>O, or to the pH of the soil pore water, because it takes into account the proton buffering capacity of the soil, the pH CaCl<sub>2</sub> is probably a better proxy of the mean rhizosphere pH. Finally, regarding pH, it should be noted that the range of variation in our study was narrow (8.0-8.7) and predictions should be used with caution when the soils are more acidic.

Manganese extracted from amorphous materials by Tamm's reagent explained more than 12% of the total variability in the grain Cd (*p-value* <0.001). Its negative coefficient shows that Mn oxides reduce Cd availability for root uptake (Wenzel et al., 1996). Mn oxides are one of the most reactive sorbing phases for Cd (Dong, 2000) due to their large specific surface areas and their low zero-point charge pH, the pH value at which the oxides are mainly negatively charged and not associated with cations (Feng et al., 2007). Mn oxides have already been shown to be an efficient stabilizing amendment for Cd, Pb and Zn in contaminated soils by significantly reducing metal mobility and uptake by plants (Sappin-Didier et al., 1997).

With absolute errors of around 0.02 mg.kg<sup>-1</sup> (Figure IV.3), the best model (Equation IV.8) was able to discriminate grain Cd contents of 0.15 mg.kg<sup>-1</sup> from those of 0.2 mg.kg<sup>-1</sup>. Therefore, the model could be useful for the project of the European Union to lower the regulatory limit of durum wheat grain from 0.2 to 0.15 mg Cd.kg<sup>-1</sup> grain (DG-SANCO, 2011). However, the model is not sufficiently accurate to distinguish the intermediate step of 0.175 mg kg<sup>-1</sup> between 0.2 and 0.15 mg Cd kg<sup>-1</sup>, which underlines the need to produce more mechanistic models to help set realistic regulatory limits.

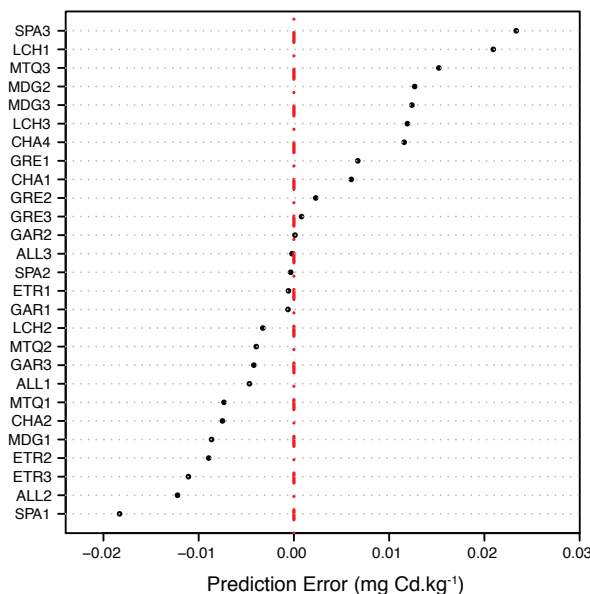


Figure IV.3: *Residual errors (observed minus predicted values) of Equation IV.8 for each soil in the dataset. All grain Cd data are expressed at 15% water content.*

## **5. Conclusion**

Predictions of Cd contents in durum wheat grains were obtained from a regression model including the concentration of the Cd in the soil solution obtained by centrifugation, the soil CaCl<sub>2</sub> pH, CEC and Tamm's Mn oxide content. The predictive model, parametrized on 27 agricultural soils, explains 88% of the variation in Cd in durum wheat grains and has a reasonably low RMSE value of around 0.01 mg.kg<sup>-1</sup>. The regressor variables of the model suggest that Cd availability in the alkaline soils tested here was controlled by the initial concentration of Cd in the soil solution and by CaCl<sub>2</sub> pH and by the CEC, which can be interpreted as the buffering capacity of the solid phase. Amorphous manganese content appears to be the main soil phase restraining Cd availability.

This model will be useful to detect cases where grain Cd exceeds the planned regulatory limit for durum wheat of 0.15 mg Cd.kg<sup>-1</sup> compared to the current limit of 0.20 mg Cd.kg<sup>-1</sup>. However, the model is not sufficiently accurate for the intermediate step of 0.175 mg Cd.kg<sup>-1</sup> and the limited predictive accuracy of this model as well as the limited accuracy of the models taken from the literature should be taken into account when revising the regulatory limits. Nevertheless, this simple predictive model will be useful not only for farmers to assess the risk that the Cd contents in their durum wheat grain exceed the regulatory limit, but also for food safety authorities to test different regulatory thresholds to identify a trade-off between food safety and the negative impact on farmers of a lower limit.

## **Acknowledgments**

The authors are grateful to Sylvie Bussière and Cécile Coriou for their technical assistance and want to thank Dr. Jean-Yves Cornu for useful discussions and his comments. This work was supported by funding from *Arvalis-Institut du végétal* (Cadur project) and from the French National Institute for Agricultural Research.

## **Partie 2 - Élargissement du domaine de variabilité des sols étudiés pour augmenter la robustesse des modèles**

### **1. Contexte**

Dans cette seconde partie du travail, le jeu de données disponible pour élaborer des modèles de prédiction du Cd dans le grain a été élargi de manière à compléter le jeu de données initial par des données acquises plus récemment et choisies pour couvrir la majorité des sols agricoles cultivés en blé dur. Le but était alors d'augmenter la robustesse des modèles et aussi leur précision. En effet, le meilleur modèle identifié précédemment présente une relativement bonne capacité de prédiction des teneurs dans le grain (RMSE égale à  $0.0102 \text{ mg Cd.kg}^{-1}$ ), mais sa précision est insuffisante pour son utilisation dans le cadre de la révision des seuils maximums de Cd autorisés dans les grains. Pour une diminution progressive du seuil actuel de  $0.2 \text{ mg Cd.kg}^{-1}$  dans le grain à  $0.175$  puis  $0.15 \text{ mg Cd.kg}^{-1}$ , l'erreur du modèle sur les prédictions est encore trop importante, car du même ordre de grandeur que la diminution du seuil réglementaire.

La robustesse des modèles précédemment établis, c'est-à-dire leur valeur prédictive pour de nouveaux sols, peut également être discutée, car il est difficile d'apprécier si les sols ayant servi à les calibrer sont représentatifs de la majorité des sols cultivés en blé dur. La dépendance de ce type de modèles aux caractéristiques du jeu de données utilisé pour leur élaboration est une limite bien connue de ce genre de modèle et elle est à prendre en considération lorsque ces équations sont appliquées à d'autres contextes (Bonten et al., 2008b; Groenenberg et al., 2012). La robustesse des modèles est généralement testée en utilisant pour la prédiction un jeu de données indépendant, c'est-à-dire n'ayant pas servi à paramétrer le modèle, et en étudiant les erreurs de prédictions générées.

Dans cette seconde phase du travail, la valeur prédictive du meilleur modèle a d'abord été testée sur les nouveaux sols. Ensuite, nous avons utilisé l'ensemble des données pour réitérer la démarche de recherche de modèle statistique sur un domaine de sols plus vaste et représentatif des situations réelles en espérant augmenter la précision des modèles évaluée par validation croisée.

## 2. Matériels et méthodes

Le jeu de données initial a été complété par 21 cas supplémentaires pour lesquels nous disposons de l'échantillon de terre associé à l'échantillon de grain sur lequel la teneur en Cd avait été déterminée. Ces échantillons ont été choisis parmi les échantillons du programme ANR CaDON dans lequel des campagnes de prélèvements appariés terre+grain ont été réalisées par Arvalis et FranceAgrimer auprès d'agriculteurs volontaires. Par rapport aux contraintes de calendrier, nous avons choisi les échantillons uniquement parmi ceux de la campagne de prélèvement 2016. Un total de 21 sols supplémentaires ont été sélectionnés afin d'élargir les types de sols et de couvrir le domaine correspondant aux sols cultivés en blé dur. Le nombre de sols retenus a été optimisé par rapport à la charge de travail et par rapport au coût financier généré par les analyses qui ont été sous-traitées.

Les méthodes de mesure des caractéristiques des sols sur ce jeu de données sont les mêmes que celles présentées dans l'article (section Chapitre IV.3.1). Les caractéristiques de ce nouveau jeu de données sont présentées Table IV.6. Ces données additionnelles permettent de travailler sur une plus grande variabilité du pH du sol qui s'échelonne entre 6.1 et 8.4 avec une valeur médiane de 8.0. Il est à noter qu'en France, les sols utilisés pour la culture du blé dur sont en grande majorité alcalins (Arvalis Institut du végétal, communication personnelle), justifiant que la quantité de sols acides choisie soit plus faible.

La modélisation statistique a été effectuée de façon hiérarchique comme décrite dans l'article, c'est-à-dire en intégrant les régresseurs 1 à 1 à partir de groupes de variables suivant leur effet supposé sur la teneur en Cd du grain.

Table IV.6 : Récapitulatif des paramètres pédogéochimiques principaux des sols du jeu de données complémentaire (ANR CaDON, N=21)

	Argiles	MOS	Calcaire	MnO	CEC	pH	pH	pH <sub>ss</sub>	Cd <sub>t</sub>	Cd <sub>ss</sub>	Cd <sub>grain</sub>
	%	g.kg <sup>-1</sup>	g.kg <sup>-1</sup>	g.kg <sup>-1</sup>	cmol <sup>+</sup> .kg <sup>-1</sup>	eau	CaCl <sub>2</sub>		mg.kg <sup>-1</sup>	nM	mg.kg <sup>-1</sup>
Min.	16.1	13.1	0.0	0.051	8.3	6.1	5.2	6.7	0.127	0.25	0.026
Méd.	28.1	25.1	31.0	0.305	17.7	8.0	7.3	7.9	0.259	0.96	0.079
Max.	53.9	46.1	442.0	0.908	37.7	8.5	7.6	8.2	0.542	5.10	0.300

Les paramètres et groupes de paramètres sont notés sur l'équation suivante.

$$\log_{10}(Cd_{grain}) = f\left(\log_{10}\left\{\begin{matrix} Cd^{2+} \\ Cd_{ss} \\ Cd_{CaCl_2} \\ Cd_t \end{matrix}\right\} + \left\{\begin{matrix} \emptyset \\ pH \\ pH_{CaCl_2} \\ pH_{ss} \\ \log_{10}(Calcaire) \end{matrix}\right\} + \log_{10}\left\{\begin{matrix} \emptyset \\ Argile \\ ArgLF \\ CEC \end{matrix}\right\} + \log_{10}\left\{\begin{matrix} \emptyset \\ MOS \\ Al_{Tamm} \\ Mn_{Tamm} \\ Fe_{Tamm} \end{matrix}\right\} + \log_{10}\left\{\begin{matrix} \emptyset \\ Ca_{ss} \\ Mg_{ss} \end{matrix}\right\} + \log_{10}\left\{\begin{matrix} \emptyset \\ NO_{3ss} \\ SO_{4ss} \\ Cl_{ss} \end{matrix}\right\}\right)$$

La différence par rapport à la démarche décrite dans l'article (section Chapitre IV.3.6) repose sur le fait que le groupe de variables représentant l'acidité/alcalinité du sol (pH, pH CaCl<sub>2</sub>, pH<sub>ss</sub> et calcaire) n'est plus forcé en tant que deuxième régresseur dans le choix des variables et que la variable Calcaire, montrant une corrélation avec le pH du sol (r = 0.66, p-value < 0.01, test de corrélation de Bravais-Pearson), a été incluse dans ce groupe de variables.

### 3. Résultats & discussion

#### 3.1. Test du modèle de prédiction du Cd grain (Equation IV.8) sur un jeu de données indépendant du jeu de calibration

Le modèle établi sur 27 données pour prédire le Cd grain (Equation IV.8) a été testé sur un jeu de données de 21 nouveaux sols. La Figure IV.4 présente les valeurs observées en fonction des prédictions. Il apparaît que la précision n'est pas excellente sans être très mauvaise, mais le principal défaut est une sous-estimation de prédiction très marquée.

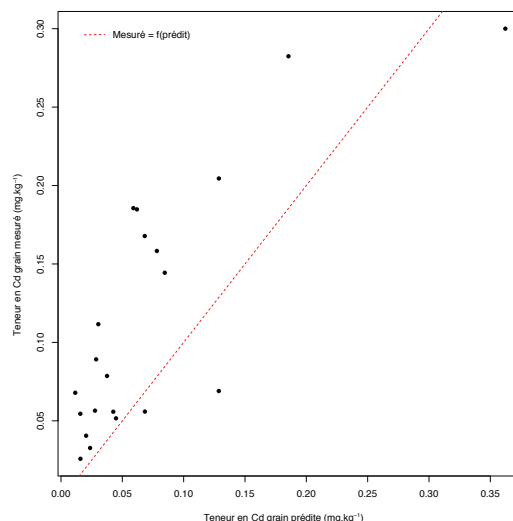


Figure IV.4 : Comparaison entre les teneurs en Cd dans le grain simulées par l'Equation IV.8 pour le jeu de données indépendant et les teneurs mesurées. La droite pointillée rouge représente la droite 1:1 correspondant à une adéquation parfaite entre valeurs observées et valeurs prédites.



Lorsque ce même modèle ( $Cd_{ss}$ , pH  $CaCl_2$ , CEC, MnO) est recalibré sur 43-44 sols et testé sur les 4-5 restants par validation croisée  $k=10$ -folds (100 répétitions), le biais n'existe plus, mais la précision n'est pas beaucoup améliorée (Equation IV.9 versus Equation IV.8, Table IV.7). La disparition du biais illustre que le jeu de donnée initial de 27 sols manquait de représentativité. En comparant les coefficients du modèle entre la calibration sur 27 sols et celle sur 48 sols, nous constatons que seuls les coefficients de la CEC et surtout de MnO varient. Ces deux variables n'étant pas corrélées ( $r = 0.29$ ,  $p$ -value = 0.05, test de corrélation de Bravais-Pearson), cela indique que les deux jeux de données sont différents concernant la distribution des valeurs de CEC et MnO (Figure IV.5). Effectivement, pour ces deux variables, le jeu de 21 données supplémentaires complète mieux la distribution de ces deux variables. En conséquence, les coefficients de ces variables dans les modèles sont fortement affectés par l'addition de nouvelles données (Equation IV.8, Equation IV.10).

Cette partie de nos résultats démontre et rappelle l'importance du jeu de données de calibration vis-à-vis de la robustesse des modèles statistiques de prédiction. Le fait d'avoir élargi le domaine couvert par le modèle pour la CEC et les teneurs en MnO ne prouve pas que ce domaine soit au final parfaitement représentatif des sols cultivés en blé dur. Cette information est difficile à obtenir, plus spécifiquement pour MnO, car ce n'est pas une variable analysée de façon systématique, car peu proposée par les laboratoires d'analyses de terre.

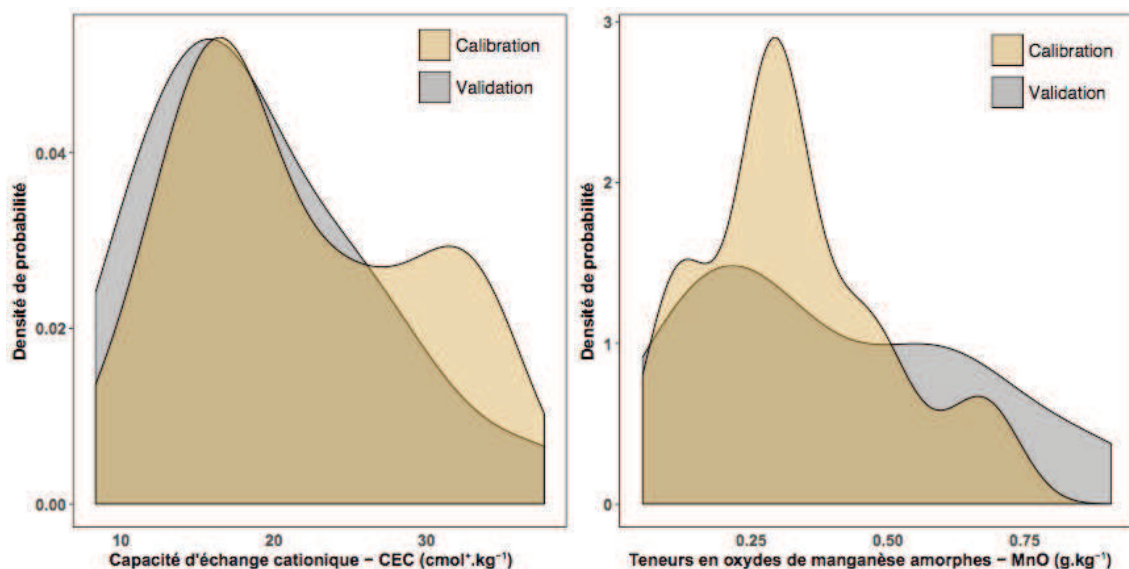


Figure IV.5 : Distributions comparées de la capacité d'échange cationique (à gauche) et des teneurs en oxydes de manganèse amorphes (à droite) en fonction des deux jeux de données. Le jeu de données initial (27 sols) est représenté en jaune et le jeu de données additionnel (21 sols) est représenté en gris.

### **3.2. Recherche du meilleur modèle de prédiction des teneurs en Cd dans le grain à partir du jeu de données étendu à 48 sols**

Nous avons utilisé les deux jeux de données précédents pour tester à nouveau la totalité des combinaisons de variables afin d'identifier les variables expliquant le mieux les teneurs en Cd dans le grain sur un plus grand nombre de sols, ayant des propriétés physico-chimiques légèrement plus variables.

Sur un total de 48 sols, le modèle permettant les meilleures prédictions des teneurs en Cd dans le grain est construit à partir des concentrations totales de Cd en solution ( $Cd_{ss}$ ), des teneurs en calcaire du sol, de la capacité d'échange cationique (CEC) et des teneurs en Mn extrait par le réactif de Tamm, assimilée aux oxydes de manganèse amorphes (MnO) (Equation IV.10, Table IV.7). Ce modèle correspond à une variable près au même modèle que précédemment ( $Cd_{ss}+CEC+MnO$ ), mais le pH a été remplacé par le Calcaire. Ces deux variables sont corrélées et nous pouvons donc considérer que l'effet du calcaire est en grande partie inclus dans celui du pH  $CaCl_2$  et inversement. Le fait que ce soit le calcaire qui soit sélectionné à la place du pH peut avoir deux explications. La première, mathématique, serait la gamme de variation légèrement plus large du  $\log_{10}(\text{calcaire})$  (3 unités) comparée à celle du pH (2.4 unités). La deuxième explication serait d'ordre géochimique. Le calcaire aurait un effet négatif sur les teneurs en Cd du grain en raison de sa capacité à sorber fortement le métal, au même titre que les oxydes de manganèse amorphes (coefficients négatifs dans l'Equation IV.10). À pH alcalins, la précipitation, ou coprécipitation, du Cd avec les ions carbonates et calcium présents en solution, pourrait être invoquée. Cependant, les faibles concentrations en Cd en solution, retrouvées par rapport aux concentrations en Ca en solution et en ions carbonates ainsi que les produits de solubilité de la calcite et de l'otavite, respectivement égaux à 8.34 et 11.39 (Fujino et al., 1976), ne favorisent pas cette hypothèse pour nos sols. Par contre, il a été observé que le Cd à faibles concentrations pourrait être capable de se sorber à la surface de la calcite et former une phase ayant un comportement proche d'un complexe de sphère interne, c'est-à-dire dont la désorption est faible (Li et al., 2012; Papadopoulos and Rowell, 1988; Zachara et al., 1991).

Dans le modèle, le cadmium en solution et le pH (masqué par le calcaire) définiraient la concentration en  $Cd^{2+}$  en solution, la forme phytodisponible.

Dans ce modèle plus performant (Equation IV.10, Table IV.7), nous retrouvons que le coefficient de la CEC est positif, favorisant donc l'accumulation du Cd dans le grain. Cela peut s'expliquer par le fait que la CEC permet le réapprovisionnement de la solution du sol en  $Cd^{2+}$  lors du prélèvement racinaire par désorption du métal des sites de sorption à faible capacité de rétention, notamment les sites non spécifiques de sorption. Nous pouvons expliquer l'influence du rôle négatif des oxydes de Mn sur le Cd dans le grain en nous référant à ce que nous avons conclu au Chapitre II sur la modélisation mécaniste de la sorption du  $Cd^{2+}$ . Nous avons conclu que le

$Cd^{2+}$  de la phase solide, échangeable avec la solution, était très probablement en majorité le Cd faiblement lié sur la phase solide, notamment sur les sites non spécifiques. Il alors est possible que la CEC estimée par la cobaltihexamine surestime le nombre de sites échangeables pour le Cd en considérant également des sites de complexation spécifiques dépendant du pH. Dans ce cas, les oxydes de manganèse sont nécessaires au modèle pour réduire le nombre de sites estimés par la CEC pour l'échange phase solide solution du  $Cd^{2+}$  lors du réapprovisionnement (effet négatif de MnO).

L'importance du réapprovisionnement de la solution de sol pour prédire la teneur en Cd du grain est également illustrée par le fait que seule, la concentration en  $Cd^{2+}$  en solution ne permettent pas de prédire correctement les teneurs en Cd dans le grain, même s'il est généralement considéré que seule la fraction libre est prélevée (Campbell, 1995).

L'augmentation de l'erreur résiduelle du meilleur modèle lorsque le jeu de données de calibration augmente de 27 à 48 (0.0102 à 0.0345) n'est pas totalement surprenante. En effet, l'augmentation du nombre de sols s'est aussi accompagnée d'une augmentation du domaine de variabilité des variables explicatives avec en conséquence une plus forte probabilité d'avoir des situations plus marginales surtout pour un modèle simple à 4 variables qui n'est pas capable d'expliquer la totalité de l'information.

### ***3.3. Analyse des erreurs de prédictions du Cd grains issues de la validation croisée***

Les erreurs générées par l'Equation IV.10, correspondant à la différence en valeurs absolues entre les teneurs en Cd, dans le grain, observées et les teneurs prédites, sont représentées Figure IV.6.

La précision avec laquelle un modèle prédit les teneurs en Cd dans le grain va conditionner son utilisation. Le projet de révision de la réglementation européenne de la teneur en Cd du blé dur (DG-SANCO, 2011) prévoyait une baisse du seuil 0.2 à 0.15 mg Cd/kg de grain avec une étape intermédiaire à 0.175 mg/kg. Pour qu'un modèle prédictif puisse être utile dans ce contexte, il faut que l'erreur de prédiction soit inférieure à 0.025 pour permettre de discriminer les valeurs 0.15, de 0.175 et 0.175 de 0.2. En effet, seule une erreur inférieure à 0.025 permet de savoir si une valeur prédite entre 0.15 et 0.175 est supérieure à 0.15 ou inférieure à 0.175.

*Table IV.7 : Meilleurs modèles pour la prédiction des teneurs en Cd dans le grain de blé dur et paramètres statistiques pour leur évaluation :  $r^2_{adj}$  de la régression multiple ;  $RMSE_{CV}$  : racine de l'erreur quadratique moyenne obtenue par validation croisée ;  $VIF_{max}$  : valeur maximale du facteur d'inflation de la variance ;  $N$  : nombre d'observations utilisées pour le paramétrage du modèle. La pente et l'ordonnée à l'origine concerne la régression des valeurs observées en fonction des valeurs prédites*

Eq.	Model	$r^2_{adj}$	$RMSE_{CV}$ (mg.kg <sup>-1</sup> )	$VIF_{max}$	$R^2_{obs/pred}$	Pente	Ord. Orig.	N
Equation IV.8	$\log_{10}(Cd_{grain}) = -1.979$ $+ 0.553^{***} \log_{10}(Cd_{ss})$ $- 0.347^* \text{pH CaCl}_2$ $+ 1.630^{***} \log_{10}(CEC)$ $- 0.685^{***} \log_{10}(MnO)$	0.878	0.010	2.20	0.847	0.948	-0.067	27
Equation IV.9	$\log_{10}(Cd_{grain}) = 4.068^{***}$ $+ 0.489^{***} \log_{10}(Cd_{ss})$ $- 0.346^{***} \text{pH CaCl}_2$ $+ 1.097^{***} \log_{10}(CEC)$ $- 0.343^{**} \log_{10}(MnO)$	0.654	0.044	2.03	0.616	0.930	-0.083	48
Equation IV.10	$\log_{10}(Cd_{grain}) = 1.39$ $+ 0.426^{***} \log_{10}(Cd_{ss})$ $- 0.303^{***} \log_{10}(\text{Calcaire})$ $+ 1.145^{***} \log_{10}(CEC)$ $- 0.596^{***} \log_{10}(MnO)$	0.811	0.0345	1.96	0.778	0.965	-0.04	48

Statistical significance: \*\*\* for  $p$ -value < 0.001, \*\* for  $p$ -value < 0.01, \* for  $p$ -value < 0.05

La même condition s'applique pour l'intervalle 0.175-0.2 mg Cd.kg<sup>-1</sup> de grain. L'Equation IV.10 génère des erreurs s'échelonnant de 2.10<sup>-07</sup> à 0.1636 mg Cd.kg<sup>-1</sup>. La validation croisée indique que 64.5 % d'entre elles sont inférieures à 0,025 ce qui confère au modèle une fiabilité de 64.5 % pour pouvoir évaluer le risque de dépassement du seuil réglementaire dans le contexte d'une baisse progressive de 0.2 à 0.175 puis 0.15 mg Cd.kg<sup>-1</sup> de grain. En omettant le seuil intermédiaire et en raisonnant sur une baisse de 0.2 à 0.15 mg Cd/kg de grain, la condition sur l'erreur est d'être inférieure à 0.05 et le taux de fiabilité du modèle augmente alors à 88 %.

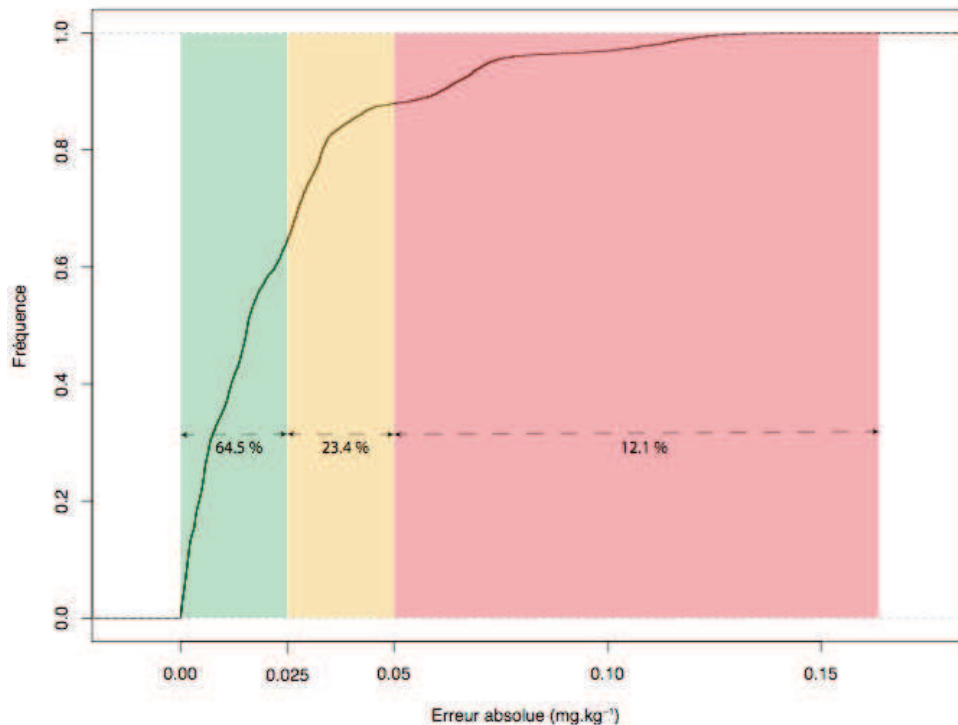


Figure IV.6 : Distribution cumulée des erreurs absolues (observé-prédit) de prédictions générées par l'Equation IV.10. Trois zones de couleur différentes sont définies à partir du projet de diminution de la valeur maximale autorisée dans les grains de blé (DG-SANCO, 2011). La zone verte correspond à une diminution du seuil actuel de  $0.025 \text{ mg.kg}^{-1}$  (passage  $0.2$  à  $0.175 \text{ mg.kg}^{-1}$ ). La zone orangée pour une diminution du seuil de  $0.05 \text{ mg.kg}^{-1}$  (passage de  $0.2$  à  $0.15 \text{ mg.kg}^{-1}$ ) et la zone rouge pour représenter les erreurs supérieures à ces deux paliers.

## 4. Conclusion

La première partie de ce chapitre a permis la calibration d'un modèle prédictif des teneurs en Cd dans les grains de blé dur à partir de 27 sols. Ce modèle statistique ne nécessite de renseigner que 4 variables identifiées comme les plus représentatives de la phytodisponibilité du Cd. La validation sur jeu de données indépendant a permis de nuancer la robustesse du modèle initial et de montrer l'importance du jeu de données de calibration vis-à-vis de cette robustesse. Ceci souligne la nécessité de paramétrer les modèles statistiques, en fonction du contexte, sur des gammes de propriétés suffisamment étendues. Cependant, plus le jeu de données de calibration est grand, plus l'erreur du modèle risque d'être importante puisque le risque est grand de tomber sur des sols dont le fonctionnement est particulier (par exemple lié à un apport récent de MO fraîche ou d'engrais modifiant la phytodisponibilité du Cd).

Cette seconde partie a également permis de recalibrer le modèle sur un nombre de données plus grand, et nous a permis d'identifier des paramètres proches de ceux identifiés précédemment.

Les concentrations en Cd total en solution, les teneurs en calcaire du sol, en oxydes de manganèse amorphes et la capacité d'échange cationique, constituent donc les variables à renseigner pour prédire correctement les teneurs dans le grain. Même si les performances de ce modèle ne sont pas mauvaises, il reste que pour 35.5 % des sols, il ne sera pas possible de discriminer les baisses progressives (de 0.2 à 0.175 puis à 0.15 mg Cd.kg<sup>-1</sup>). Par rapport au contexte réglementaire, il serait intéressant de tester des modèles logistiques d'évaluation de la probabilité qu'à un sol d'induire une teneur en Cd dans le grain, supérieure à un seuil fixé. Ces modèles prédisent non plus la teneur en Cd du grain, mais la probabilité que cette teneur dépasse un seuil donné. Ces modèles pourraient avoir de meilleures performances et devraient être testés. Nous ne l'avons pas fait par manque de temps, mais aussi parce que le nombre d'observations est faible (n=48) ce qui conduit à ce que le nombre de cas positifs (supérieurs au seuil) soit très faible.

Pour terminer, il est important de considérer que le transfert sol-plante intègre un nombre important de processus, la spéciation du Cd dans le sol et la solution, le réapprovisionnement, le prélèvement et également la répartition du Cd dans la plante. Certains d'entre eux n'étant pas dépendant du sol, mais de la plante, les modèles statistiques centrés sur les variables du sol ne permettront jamais de prédire parfaitement les teneurs dans le grain ce qui souligne l'intérêt de développer des modèles de transfert sol-plante semi-mécanistique ou mécaniste couplant à la fois les mécanismes propres au sol et ceux relatifs à la physiologie de la plante.



## ***Discussion générale***

---

### ***1. Cohérence des modèles statistiques de prédiction des principaux processus impliqués dans la phytodisponibilité du Cd en sols agricoles et apport de la modélisation géochimique***

L'étude de la disponibilité du Cd est au cœur de l'évaluation de la phytodisponibilité de cet élément. Pour rappel, la disponibilité et la phytodisponibilité sont deux concepts très proches, le premier étant défini comme la fraction maximale du Cd du sol (phase solide et solution) susceptible d'être absorbée par une cible biologique et le second comme la quantité de Cd qui est effectivement prélevée par la plante durant un temps d'exposition donné.

Nous avons observé à travers les différents chapitres que l'équilibre phases solides/solution, la complexation en solution et l'accumulation du Cd dans le grain de blé dur pouvaient être prédits statistiquement à partir d'un nombre restreint (4-5) de variables. Vu la complexité réelle de la géochimie du Cd que l'on peut apprécier en examinant la structure conceptuelle des modèles géochimiques, il est assez remarquable de constater que le comportement global du système dans le contexte étudié se résume par un modèle mathématique de 3 à 4 variables sachant que la gamme de sols couverte est tout de même large. Une explication pourrait être que nous avons travaillé sur des sols peu contaminés. Il est imaginable que si le Cd était plus abondant dans le sol, les phases pour lesquelles la sorption du Cd est significative seraient plus nombreuses. Les modèles statistiques additif log-log établis dans ce travail correspondent en valeurs naturelles à des modèles multiplicatifs qui traduisent bien les interactions et la non-linéarité du comportement du système telle qu'on peut la constater dans les modèles mécanistes géochimiques.

Il est également remarquable que parmi les modèles statistiques élaborés, certaines variables étaient de manière récurrente communes entre les modèles : le pH, la concentration en Ca, les oxydes de manganèse. Dans le modèle de prédiction du Cd dans le grain (Chapitre IV), en considérant que la variable Calcaire qui est corrélée au pH représente ce dernier, le pH est une



variable retrouvée dans les 3 modèles statistiques. Le coefficient de cette variable est toujours négatif, ce qui montre que plus les pH des sols augmentent, plus les teneurs en  $\text{Cd}^{2+}$  en solution ou en Cd dans le grain diminuent. Cette observation est conforme à ce qui est largement documenté dans la littérature. Les explications sont multiples et les plus importantes sont liées au fait que l'augmentation du pH favorise la déprotonation i) des sites de surface favorisant la sorption du  $\text{Cd}^{2+}$  et ii) des ligands favorisant la complexation du métal en solution.

La concentration du Ca en solution a également été identifiée dans les modèles statistiques à la fois pour la spéciation du  $\text{Cd}^{2+}$  en solution et pour la distribution entre les phases solides et la solution de sol. La distribution sol-solution est dépendante de la complexation du Cd en solution, mais cette dernière n'est généralement pas le mécanisme majeur gouvernant la distribution du Cd. En conséquence, la présence de la concentration du Ca en solution dans le modèle statistique de distribution ne peut pas être uniquement attribuée à l'effet de la complexation et conformément à ce que modélise les modèles géochimiques, le Ca intervient également directement sur la sorption du Cd sur les phases solides en tant que compétiteur. Le même raisonnement tient pour le pH qui est également une variable commune à nos modèles statistiques pour la spéciation du  $\text{Cd}^{2+}$  en solution et pour sa distribution entre les phases solides et la solution de sol.

Bien que le Ca soit clairement considéré dans la littérature comme très influent vu son statut de compétiteur majeur du Cd, la teneur en Ca dans la solution de sol n'est généralement pas retrouvée dans les modèles statistiques de distribution ( $K_d$ ) ou de phytodisponibilité. La récurrence de cette variable dans les modèles que nous avons développés est donc un élément original par rapport à la littérature. Ceci est probablement lié à la nature alcaline et calcaire des sols étudiés. En effet, la variation des concentrations en Ca en sols alcalins, plus grande que celle généralement retrouvée en sols acides, peut favoriser l'identification de cette variable dans les modèles statistiques.

Dans le modèle de prédiction du Cd dans le grain, le calcaire du sol pourrait aussi être considéré représentatif des concentrations en Ca en solution de sol puisqu'une partie du calcaire total est considérée comme actif donc facilement dissociable. Cependant le coefficient négatif du calcaire dans la régression n'est pas cohérent avec cette hypothèse. Si le calcaire représentait l'effet du Ca, il devrait avoir un coefficient positif puisque le Ca en tant que compétiteur vis-à-vis du Cd pour la sorption ou la complexation en solution augmente la concentration en  $\text{Cd}^{2+}$  en solution et donc théoriquement le prélèvement par la plante et l'accumulation dans le grain. Ceci plaide en faveur du fait que dans le modèle de prédiction du Cd dans le grain, le calcaire total du sol est une variable probablement plus représentative du pH en solution que des concentrations en Ca, ou au moins que l'effet du pH serait plus important que celui du Ca en solution.

La troisième variable récurrente de nos modèles statistiques est la teneur en oxydes de manganèse (MnO) amorphes, que l'on retrouve dans les modèles statistiques de distribution sol-solution et de teneurs en Cd dans le grain. Le coefficient négatif est cohérent avec la modélisation géochimique qui considère MnO comme une phase de sorption, probablement spécifique, diminuant les concentrations en  $\text{Cd}^{2+}$  en solution et donc par voie de conséquence les teneurs retrouvées dans le grain. Les oxydes métalliques sont effectivement généralement considérés comme des phases de sorption importantes du Cd sans pour autant être systématiquement identifiées dans les modèles statistiques. Par exemple, de Groot et al. (1998) ont identifié les oxydes d'aluminium amorphes dans un modèle statistique de prédiction du coefficient de partition  $K_d$ . La modélisation géochimique de la distribution sol-solution du  $\text{Cd}^{2+}$  nous a montré que les oxydes d'aluminium (AIO) pouvaient être substitués aux MnO en donnant quasiment les mêmes qualités de prédiction. Il est difficile de savoir si dans les modèles statistiques la sélection des MnO plutôt que celle des AIO a une réalité géochimique. Cela pourrait être le cas puisque selon la littérature, le point de charges nulles des MnO (1.5, Murray et al., 1968) comparé à celui des AIO (8.2, Sparks, 2003) indique qu'ils sont sûrement plus dissociés que les AIO.

Comme cela a été exposé ci-dessus, les modèles géochimiques ont permis de donner un sens et une interprétation mécaniste aux variables identifiées par les modèles statistiques. La cohérence des deux approches (statistique et géochimique) a été plus loin. En effet, lors de la modélisation géochimique de la spéciation du  $\text{Cd}^{2+}$  en solution (Chapitre II), nous avons fait l'hypothèse que la MOD était le ligand majoritaire en solution contrôlant la spéciation du Cd et par modélisation inverse, nous avons trouvé que le Ca et le pH étaient les principaux paramètres expliquant la variabilité de la réactivité de la MOD en solution vis à vis du Cd. La modélisation statistique de cette réactivité (modélisation de  $K'n$ ) a également identifié le Ca et le pH comme variables explicatives. De même, l'approche géochimique de modélisation de la distribution sol-solution du Cd (Chapitre III) a montré que l'on pouvait correctement prédire le  $\text{Cd}^{2+}$  en solution à partir d'un compartiment échangeable correspondant au Cd extrait par  $\text{NH}_4\text{NO}_3$  et deux surfaces de sorption, une surface d'échange cationique non spécifique et une surface de complexation spécifique par des oxydes de manganèse, ce qui donne un sens au « meilleur » modèle statistique prédictif de la distribution construit à partir du Cd extrait par  $\text{NH}_4\text{NO}_3$ , du pH, du Ca en solution et de la teneur en MnO.

## **2. Considérations sur les simplifications inhérentes aux modèles statistiques**

Un des objectifs de ce travail était de développer des modèles statistiques prédictifs simples, mais assez robustes mettant en œuvre un nombre restreint de variables. Cela suppose de faire des simplifications assez drastiques par rapport au fonctionnement réel du système modélisé. Concernant la robustesse des modèles, nous avons largement utilisé la validation croisée pour tester nos modèles statistiques afin d'examiner leurs performances de prédiction suivant une légère modification du jeu de donnée de calibration réalisée par échantillonnage. En aucun cas, les résultats des validations croisées ne permettent cependant de statuer de la robustesse de nos modèles pour des contextes sortant du domaine des sols étudiés. Ce domaine est entre autre délimité par les propriétés des sols qui nous ont servis et qui ont été présentées comme élément de bonne utilisation des modèles développés.

Le domaine de validité de nos modèles est également lié à une simplification majeure que nous avons faite concernant la notion de dynamismes. Nous avons travaillé sur des sols qui lors de leur prélèvement au champ étaient relativement à l'équilibre. Cependant, les sols agricoles peuvent subir des apports, pouvant modifier cet état d'équilibre auquel ils ont été prélevés. En effet, l'apport de fertilisants minéraux ou organiques peuvent modifier la disponibilité du Cd dans les sols (Bolan and Duraisamy, 2003; Grant and Sheppard, 2008) et modifier la nature de la MOD en solution (apports de MO fraîche de composition différente de la MOD supposée essentiellement composée d'acides fulviques) ou la spéciation du Cd en solution par complexation ou précipitation du Cd avec les ligands inorganiques ajoutés en grandes quantités dans les sols (par exemple la chaux). Dans la littérature, ces interventions ont un effet avéré sur l'accumulation du Cd par les cultures (Gao et al., 2010). En dehors des interventions humaines, les conditions de température et de précipitation modifient aussi continuellement la disponibilité des éléments du sol. En outre, la teneur en Cd dans le grain résulte de l'intégration temporelle de la disponibilité du Cd dans le sol au cours de la culture, de la croissance de la plante et de l'écophysologie de la répartition du métal entre les organes. Cette teneur a été modélisée à partir de variables du sol mesurées à l'équilibre. Le modèle statistique de prédiction du Cd dans le grain de blé dur développé dans ce travail n'a pas pris en compte les interventions agronomiques diverses opérées sur les cultures ni la physiologie de la plante et présente cependant des performances correctes. Cela confirme que les différences de disponibilité du Cd entre sols sont en général plus importantes que les variations induites par les la physiologie de la plante pour expliquer les variations de teneurs du contaminant dans le grain. Il est possible également que les modifications de disponibilité du Cd liées au pratiques aient assez peu de conséquences une fois intégrées sur la durée de la culture en raison d'une certaine capacité de pouvoir tampon du sol.

Concernant la non prise en compte explicite de la dynamique des processus pour la phytodisponibilité, on peut penser que certaines variables statiques peuvent partiellement rendre compte de phénomènes dynamiques. Le modèle statistique de prédiction du Cd dans le grain ne fait pas uniquement intervenir le Cd en solution, mais également la CEC comme variable ayant un effet positif. La CEC a deux effets antagonistes, elle réduit le Cd en solution, mais assure aussi le réapprovisionnement de la solution du sol en Cd<sup>2+</sup> lors du prélèvement racinaire (pouvoir tampon pour le Cd<sup>2+</sup>). Dans le modèle statistique, le Cd en solution étant déjà identifié comme variable explicative, la CEC interviendrait donc pour rendre compte du réapprovisionnement de la solution du sol et rendrait ainsi compte d'un processus dynamique.

### **3. Utilité et améliorations possible des modèles statistiques de prédictions**

Malgré toute la complexité des mécanismes régissant la teneur en Cd du grain de blé dur, un modèle statistique assez simple de 4 variables prédictives relatives aux propriétés du sol présente tout de même une fiabilité de 65 % ou 88 % selon que l'on veuille discriminer des seuils à 0.025 ou 0.05 mg Cd.kg<sup>-1</sup> près. De la même manière, il ne faut pas plus de 4 variables pour prédire correctement la concentration en Cd<sup>2+</sup> ou en Cd total dans la solution du sol avec des erreurs pour la plupart assez faibles (environ 25 % en médiane pour les modèles de prédiction à 3 et 4 paramètres). Cela permet de comparer de manière relative des sols agricoles vis-à-vis du risque de transfert vers les organismes vivants y compris les cultures (prédiction de la disponibilité environnementale) sans pour autant pouvoir chiffrer dans l'absolu l'accumulation du Cd par ces organismes.

Ces conclusions issues de notre travail montrent que malgré la complexité du sol et des mécanismes qui y régissent la dynamique du Cd, cette dernière peut être modélisée assez simplement de manière statistique. Les modèles géochimiques mécanistes semblent confirmer la pertinence d'un comportement global statistiquement modélisable. Nos résultats plaident donc pour la constitution de bases de données permettant aux modèles statistiques de poursuivre leur apprentissage et de gagner en robustesse pour des contextes toujours plus larges. Concrètement, cela signifie que les analyses de terre doivent être élargies pour inclure des variables propres à la dynamique des ETM dans les sols. En effet, au-delà du cas particulier du Cd, la contamination des sols par les ETM, est une problématique à la fois environnementale et sanitaire qui ne se résoudra pas dans les quelques années à venir. Cependant, compte tenu de la complexité de la problématique, les approches statistiques, comme nous en avons développé et qui se perfectionnent par auto-apprentissage sont nécessaires pour fournir des outils de diagnostic assez rapidement. Ils doivent s'appuyer sur les modèles mécanistes qui même s'ils

sont plus complexes et éloignés de l'opérationnel, demeurent essentiels pour comprendre et pour donner du sens aux modèles statistiques.

Pour illustrer ce besoin d'accumuler des données pour renforcer les modèles, nous rappelons que la fiabilité du modèle de prédiction du Cd dans le grain de blé n'est que de 65 % si l'on souhaite discriminer des seuils réglementaires ne différant que de  $0.025 \text{ mg Cd.kg}^{-1}$ . Il y a donc un manque de cohérence entre les autorités sanitaires qui souhaite réguler les teneurs en Cd à des pas de variations de  $0.025 \text{ mg Cd.kg}^{-1}$  les outils d'évaluation qui sont assez peu précis pour ces variations. Il est nécessaire d'accumuler des données par les plans de surveillance et que ceux-ci doivent inclure des mesures assez simples sur les sols appariés aux échantillons de grains. Il sera alors possible de savoir jusqu'où les modèles statistiques peuvent gagner en fiabilité, en précision, en robustesse et donc d'identifier les limites de modèles aussi simples.

Il est important de garder à l'esprit que l'accumulation de données provenant de contextes différents (dans le but d'augmenter la robustesse des modèles) pourrait également tendre vers une diminution de leur précision et vers une augmentation de leur biais. En effet, l'élargissement de la base de données de calibration à d'autres contextes pourrait favoriser la nécessité de prendre en compte de nouvelles variables influentes et la disparition de certaines autres. Les modèles statistiques reposant sur des effets multiplicatifs tels que nous les avons développés font l'hypothèse forte d'une relation de type puissance ( $y=a.x^b$ ) entre la variable à expliquer et une variable explicative particulière. Cela permet de représenter une réponse linéaire ou courbe, mais dans ce cas avec un seul rayon de courbure ce qui est assez générique pour des domaines d'études assez restreints. Pour des contextes très larges, où la relation pourrait être plus complexe, nos modèles seraient inadaptés. Dans ce cas, une possibilité serait de diviser le large domaine d'étude en sous-domaines, où les modèles puissance peuvent s'appliquer. La difficulté est de trouver les variables et les seuils de segmentation du domaine. Une autre approche moins contraignante, consisterait à utiliser des méthodes statistiques sans a priori sur la relation entre la variable prédite et les variables explicatives (Breiman, 2001). Les arbres de classification/régression sont des approches de ce type. Dans cette approche, la prédiction d'une variable  $y$  se fait sur la base d'une combinaison de choix par rapport aux valeurs des variables explicatives  $x$  (si  $x_i < s_i$  alors si  $x_j < s_j$  alors si... Alors  $y=y_k$ ).

Le défaut de ces arbres est d'avoir une variance forte, c'est-à-dire que les variables entrant dans les choix, leur ordre et les seuils de choix dépendent beaucoup du jeu de données. Pour pallier ce problème, les forêts d'arbres aléatoires (« random forests », Breiman (2001)) sont intéressantes. La prédiction se fait à partir de la valeur médiane issue des prédictions d'un grand nombre d'arbres, tous différents, construits à partir d'un échantillonnage de données sur le jeu initial et en sélectionnant au hasard un nombre déterminé de variables explicatives candidates.

Ces algorithmes, sans a priori sur la nature de la relation mathématique, se sont déjà montrés très performants dans la littérature et sont particulièrement adaptés à la modélisation statistique des systèmes complexes comme ceux qui régissent les concentrations en  $\text{Cd}^{2+}$  en solution ou dans les plantes. Cette approche apprend en permanence en fonction de nouvelles données et plaide également pour l'acquisition de données appariées sol-plante.



## **Conclusion générale**

---

L'objectif principal de ce travail de thèse a été l'élaboration de modèles de prédictions simples à visée opérationnelle de la concentration en  $\text{Cd}^{2+}$  dans la solution des sols agricoles pour évaluer le risque de transfert du Cd vers les organes végétaux consommés, en particulier le grain du blé dur.

Dans un premier temps (Chapitre II), nous avons cherché à prédire les concentrations en  $\text{Cd}^{2+}$  à partir de variables de la solution de sol, car cette dernière est facilement échantillonnable et donc la prédiction est a priori assez directe. Nous avons élaboré un modèle statistique simple pour lequel la médiane des erreurs relatives de prédictions par validation croisée des concentrations en  $\text{Cd}^{2+}$  est très correcte, aux environs de 13 %. Ce modèle est basé sur la prédiction de la réactivité de la MOD, considérée comme le ligand majoritaire en solution, à partir des concentrations en Ca en solution et du pH  $\text{CaCl}_2$  du sol. De plus, en faisant l'hypothèse que la complexation du Cd avec la MOD contrôlait la spéciation du métal en solution, la modélisation géochimique a montré qu'au-delà de la concentration de la MOD, ce qui déterminant était la disponibilité des groupements pour la complexation du Cd. La modélisation géochimique a montré que cette disponibilité était principalement gouvernée par le pH qui régit la protonation des groupements complexants et par le Ca en solution qui intervient comme compétiteur du Cd pour la complexation. Ces mécanismes ont permis de comprendre que le modèle statistique soit fondé sur la concentration en carbonates (corrélée au pH de la solution) et en calcium en solution.

Dans un second temps (Chapitre III), nous avons cherché à prédire la concentration en  $\text{Cd}^{2+}$  dans la solution du sol à partir de son équilibre avec les constituants de la phase solide sur lesquels le métal se sorbe. Dans la perspective de comparer des sols vis-à-vis de la biodisponibilité du Cd, cette approche est plus complète, car en plus de la concentration en  $\text{Cd}^{2+}$  en solution, elle s'intéresse également au Cd mobile sorbé sur la phase solide, capable de tamponner en partie la baisse de  $\text{Cd}^{2+}$  en solution lors du prélèvement par des organismes biologiques, comme une culture. À nouveau, nous avons pu élaborer un modèle statistique assez simple (modèle à 4 variables) générant des erreurs relatives par validation croisée tout à fait acceptables (~25 % en médiane). Les variables explicatives de ce modèle sont le compartiment de Cd (échangeable) extrait par une solution de  $\text{NH}_4\text{NO}_3$  à 1 M, le pH du sol, les teneurs en oxydes de manganèse (MnO) et la concentration en Ca en solution. La pertinence de ces variables a été démontrée par



la modélisation géochimique puisque le  $\text{Cd}^{2+}$  en solution est correctement prédit par un modèle mécaniste fondé, entre autres, sur le Cd extrait par une solution de  $\text{NH}_4\text{NO}_3$  associé au compartiment de Cd sorbé sur la phase solide par sorption non spécifique sur la CEC et spécifique sur les oxydes de Manganèse MnO. Le modèle géochimique inclut également la compétition par le proton (effet du pH) et la compétition par le Ca en solution. Ces résultats permettent de conclure que dans le contexte des sols faiblement contaminés, le partage du  $\text{Cd}^{2+}$  entre les phases solides et la solution est gouverné par i) le Cd faiblement sorbé, probablement essentiellement de manière non spécifique (extrait par  $\text{NH}_4\text{NO}_3$ ), ii) par la protonation des sites de surfaces et des ligands en solution, iii) par la compétition du Ca pour les sites de sorption et pour la complexation. La modélisation géochimique a aussi permis de comprendre que le Cd extrait par  $\text{NH}_4\text{NO}_3$  surestimait probablement le Cd véritablement en équilibre avec la solution ce qui nécessitait de considérer les MnO comme une phase de sorption corrigeant cette surestimation. Le Cd extrait par  $\text{NH}_4\text{NO}_3$  corrigé par sa sorption sur les oxydes de manganèse amorphes serait donc un proxy pour l'estimation du compartiment de Cd réellement mobile du sol dans la prédiction de la biodisponibilité.

Dans la dernière partie (Chapitre IV), nous avons cherché à prédire les teneurs en Cd dans le grain de blé dur (Cd effectivement absorbé et alloué au grain) à partir de variables du sol par modélisation statistique. Un modèle basé sur les concentrations en Cd total en solution, sur la CEC du sol, sur la teneur en calcaire du sol et sur les teneurs en oxydes de manganèse amorphes du sol ont permis la prédiction des teneurs en Cd dans le grain de blé dur, avec des erreurs relatives par validation croisée relativement faibles, environ 25 % d'erreur en médiane. Nous avons estimé que la fiabilité de ce modèle était de 65 % pour discerner une différence minimale de teneurs dans le grain de  $0.025 \text{ mg Cd.kg}^{-1}$  et de 88 % pour une différence de  $0.05 \text{ mg Cd.kg}^{-1}$ . Le modèle montre donc ses limites par rapport au projet de baisse intermédiaire du seuil réglementaire pour le Cd dans le grain de blé dur de  $0.2$  à  $0.175 \text{ mg Cd.kg}^{-1}$ , mais pourrait être utilisé pour discerner la conformité du blé dur entre le seuil actuel de  $0.2 \text{ mg Cd.kg}^{-1}$  et le seuil envisagé de  $0.15 \text{ mg Cd.kg}^{-1}$ .

Le modèle statistique est construit à partir du Cd total en solution, de la CEC (compartiment de Cd échangeable), de la teneur en oxydes de manganèse amorphes, du calcaire qui est corrélé au pH et du Ca en solution. À partir des résultats des deux précédents chapitres, l'identification de ses variables par le modèle statistique suggère que la phytodisponibilité du Cd pour le grain de blé dur en contexte de faible contamination est en grande partie gouvernée par la sorption-désorption du Cd à partir d'un compartiment facilement échangeable et que le pH et le Ca modulent la disponibilité des sites de sorption et celle des groupements complexants en solution.

La première perspective de ce travail porte sur la poursuite du test de nos modèles statistiques. Les bonnes performances par validation croisée, la pertinence des variables retenues suggérée par la modélisation géochimique encourage à accumuler davantage de données pour étendre les évaluations des modèles, affiner leur paramétrage et examiner leurs limites. Comme indiqué dans la discussion générale, nous pensons qu'il serait important d'ajouter de nouvelles déterminations dans les analyses de terre (les variables retenues dans ce travail) et d'organiser la collecte de données de terrain pour coupler les analyses de terre avec les analyses des produits végétaux appariés. Dans ce contexte, le mouvement actuel de science participative est très favorable à l'accumulation de données qui reste cependant dépendante de financements pour les analyses.

Une seconde perspective concerne le modèle de prédiction du  $\text{Cd}^{2+}$  en solution à partir de la distribution phase solide-solution (Chapitre III). Ce modèle fournit la concentration en  $\text{Cd}^{2+}$  en solution du sol ainsi que sa corrélation avec un proxy du compartiment mobile de Cd correspondant au produit entre le Cd extrait par  $\text{NH}_4\text{NO}_3$ , le pH, le Ca en solution et la teneur en MnO (aux puissances des variables près). En faisant l'hypothèse qu'aux faibles concentrations de Cd, la sorption est linéaire, le rapport entre le proxy du compartiment mobile de Cd et la concentration en  $\text{Cd}^{2+}$  en solution fournit un proxy du pouvoir tampon du sol pour  $\text{Cd}^{2+}$ . Ces variables et proxys peuvent renseigner un modèle de transport réactif fondé sur l'approche développé par Barber (1995) tel que celui qui est implémenté dans le modèle de transfert sol-plante Simtraces ([www6.inra.fr/simtraces](http://www6.inra.fr/simtraces)). Ce modèle de transfert sol-plante permet d'associer les mécanismes géochimiques et physiologiques pour comprendre et prédire l'accumulation d'ETM par les organes végétaux consommés. Notre travail permet donc de renseigner deux des variables d'entrées du modèle (concentration initiale en  $\text{Cd}^{2+}$  en solution et proxy du pouvoir tampon) à partir de mesures simples sur le sol.

Une dernière perspective porte sur l'utilisation des modèles statistiques de prédiction du  $\text{Cd}^{2+}$  en solution à partir de l'analyse de la solution de sol (Chapitre II) ou à partir de la distribution phase solide-solution (Chapitre III). Les variables explicatives de ces modèles sont cohérentes avec celles qui ont été retenues pour prédire le Cd dans le grain de blé dur. Cela signifie que les modèles statistiques de prédiction du  $\text{Cd}^{2+}$  ont une valeur prédictive de la phytodisponibilité, au moins pour le blé dur. Les modèles statistiques de prédiction du  $\text{Cd}^{2+}$  développés dans notre travail ne peuvent pas prédire l'accumulation du Cd par une plante, car, même implicitement, ils ne considèrent pas les spécificités du végétal. Ils peuvent par contre être utilisés pour classer des sols vis-à-vis du risque de transfert du Cd vers une culture. Il faudrait donc en parallèle étudier pour ces sols l'accumulation du Cd par un même type de culture et examiner si le classement obtenu sur la prédiction du  $\text{Cd}^{2+}$  en solution est cohérent avec celui résultant de la teneur en Cd de la plante.



## Références bibliographiques

---

- Adams, M.L., Zhao, F.J., McGrath, S.P., Nicholson, F.A., Chambers, B.J., 2004. Predicting Cadmium Concentrations in Wheat and Barley Grain Using Soil Properties. *J. Environ. Qual.* 33, 532. <https://doi.org/10.2134/jeq2004.0532>
- Adriano, D.C., 2001. Trace Elements in Terrestrial Environments, Environmental Chemistry. Springer New York, New York, NY. <https://doi.org/10.1007/978-0-387-21510-5>
- Aiken, G.R., 1985. Humic substances in soil, sediment, and water : geochemistry, isolation, and characterization. Wiley.
- Alessi, D.S., Fein, J.B., 2010. Cadmium adsorption to mixtures of soil components: Testing the component additivity approach. *Chem. Geol.* 270, 186–195. <https://doi.org/10.1016/j.chemgeo.2009.11.016>
- Amery, F., Vanmoorleghem, C., Smolders, E., 2009. Adapted DAX-8 fractionation method for dissolved organic matter (DOM) from soils: Development, calibration with test components and application to contrasting soil solutions. *Eur. J. Soil Sci.* 60, 956–965. <https://doi.org/10.1111/j.1365-2389.2009.01166.x>
- Anju, M., Banerjee, D.K., 2011. Associations of cadmium, zinc, and lead in soils from a lead and zinc mining area as studied by single and sequential extractions. *Environ. Monit. Assess.* 176, 67–85. <https://doi.org/10.1007/s10661-010-1567-4>
- Anses, 2011. Étude de l'alimentation totale française 2 (EAT 2) Tome 1 - Contaminants inorganiques, minéraux, polluants organiques persistants, mycotoxines et phyto-estrogènes, Anses, Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail.
- Appelo, C.A.J., Postma, D., 2005. Geochemistry, groundwater and pollution. Second edition. A.A. Balkema Publishers, Leiden, The Netherlands.
- Baize, D., 1997. Teneurs totales en métaux lourds dans les sols français premiers résultats du programme ASPITET. *Courr. l'Environnement l'INRA* # 22 37–46.
- Baize, D., Bellanger, L., Tomassone, R., Denis, B., Lise, B., Richard, T., 2009. Relationships between concentrations of trace metals in wheat grains and soil. *Agron. Sustain. Dev.* 29, 297–312. <https://doi.org/10.1051/agro:2008057>
- Bakircioglu, D., Kurtulus, Y.B., Ibar, H., 2011. Investigation of trace elements in agricultural soils by BCR sequential extraction method and its transfer to wheat plants. *Environ. Monit. Assess.* 175, 303–314. <https://doi.org/10.1007/s10661-010-1513-5>
- Barber, S.A., 1995. Soil nutrient bioavailability: A mechanistic approach, 2nd Edition.
- Barrow, N.J., 1986. Reaction of Anions and Cations with Variable-Charge Soils. *Adv. Agron.* 38, 183–230. [https://doi.org/10.1016/S0065-2113\(08\)60676-8](https://doi.org/10.1016/S0065-2113(08)60676-8)
- Basile-Doelsch, I., Balesdent, J., Rose, J., 2015. Are interactions between organic compounds

- and nanoscale weathering minerals the key drivers of carbon storage in soils? *Environ. Sci. Technol.* 49, 3997–3998. <https://doi.org/10.1021/acs.est.5b00650>
- Beckett, P.H.T., 1989. The Use of Extractants in Studies on Trace Metals in Soils, Sewage Sludges, and Sludge-Treated Soils. pp. 143–176. [https://doi.org/10.1007/978-1-4612-3532-3\\_3](https://doi.org/10.1007/978-1-4612-3532-3_3)
- Benedetti, M.F., Milne, C.J., Kinniburgh, D.G., Van Riemsdijk, W.H., Koopal, L.K., 1995. Metal Ion Binding to Humic Substances: Application of the Non-Ideal Competitive Adsorption Model. *Environ. Sci. Technol.* 29, 446–457. <https://doi.org/10.1021/es00002a022>
- Benedetti, M.F., Van Riemsdijk, W.H., Koopal, L.K., 1996. Humic substances considered as a heterogeneous Donnan gel phase. *Environ. Sci. Technol.* 30, 1805–1813. <https://doi.org/10.1021/es950012y>
- Bolan, N.S., Duraisamy, V.P., 2003. Role of inorganic and organic soil amendments on immobilisation and phytoavailability of heavy metals: a review involving specific case studies. *Aust. J. Soil Res.* 533–555. <https://doi.org/10.1071/SR02122>
- Bonten, L.T.C., Groenenberg, J.E., Weng, L., van Riemsdijk, W.H., 2008a. Use of speciation and complexation models to estimate heavy metal sorption in soils. *Geoderma*. <https://doi.org/10.1016/j.geoderma.2008.06.005>
- Bonten, L.T.C., Römkens, P.F.A.M., Brus, D.J., 2008b. Contribution of Heavy Metal Leaching from Agricultural Soils to Surface Water Loads. *Environ. Forensics* 9, 252–257. <https://doi.org/10.1080/15275920802122981>
- Bradbury, M.H., Baeyens, B., 1997. A mechanistic description of Ni and Zn sorption on Namontmorillonite. Part II: modelling. *J. Contam. Hydrol.* 27, 223–248. [https://doi.org/10.1016/S0169-7722\(97\)00008-9](https://doi.org/10.1016/S0169-7722(97)00008-9)
- Bradl, H.B., 2004. Adsorption of heavy metal ions on soils and soils constituents. *J. Colloid Interface Sci.* <https://doi.org/10.1016/j.jcis.2004.04.005>
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brus, D.J., De Gruijter, J.J., Römkens, P.F.A.M., 2005. Probabilistic quality standards for heavy metals in soil derived from quality standards in crops. *Geoderma* 128, 301–311. <https://doi.org/10.1016/j.geoderma.2005.04.014>
- Bryan, S.E., Tipping, E., Hamilton-Taylor, J., 2002. Comparison of measured and modelled copper binding by natural organic matter in freshwaters. *Comp. Biochem. Physiol. - C Toxicol. Pharmacol.* 133, 37–49. [https://doi.org/10.1016/S1532-0456\(02\)00083-2](https://doi.org/10.1016/S1532-0456(02)00083-2)
- Buekers, J., Van Laer, L., Amery, F., Van Buggenhout, S., Maes, A., Smolders, E., 2007. Role of soil constituents in fixation of soluble Zn, Cu, Ni and Cd added to soils. *Eur. J. Soil Sci.* 58, 1514–1524. <https://doi.org/10.1111/j.1365-2389.2007.00958.x>
- Bufflap, S.E., Allen, H.E., 1995. Comparison of pore water sampling techniques for trace metals. *Water Res.* 29, 2051–2054.

- Buffle, J., Chalmers, R.A., Masson, M.R., Midgley, D., 1988. Complexation reactions in aquatic systems : an analytical approach. E. Horwood.
- Campbell, P.G.C., 1995. Interactions between trace metals and aquatic organisms: a critique of the free-ion activity model, in: Tessier, A., Turner, D.R. (Eds.), *Metal Speciation and Bioavailability in Aquatic Systems*. Wiley, pp. 45–102.
- Cancès, B., Ponthieu, M., Castrec-Rouelle, M., Aubry, E., Benedetti, M.F., 2003. Metal ions speciation in a soil and its solution: Experimental data and model results. *Geoderma* 113, 341–355. [https://doi.org/10.1016/S0016-7061\(02\)00369-5](https://doi.org/10.1016/S0016-7061(02)00369-5)
- Cao, J., Xue, H., Sigg, L., 2006. Effects of pH and Ca competition on complexation of cadmium by fulvic acids and by natural organic ligands from a river and a lake. *Aquat. Geochemistry* 12, 375–387. <https://doi.org/10.1007/s10498-006-9004-6>
- Chantigny, M.H., 2003. Dissolved and water-extractable organic matter in soils: a review on the influence of land use and management practices. *Geoderma* 113, 357–380. [https://doi.org/10.1016/S0016-7061\(02\)00370-1](https://doi.org/10.1016/S0016-7061(02)00370-1)
- Cho, Y.A., Kim, J., Woo, H.D., Kang, M., 2013. Dietary Cadmium Intake and the Risk of Cancer: A Meta-Analysis. *PLoS One* 8, 1–8. <https://doi.org/10.1371/journal.pone.0075087>
- Christensen, J.B., Christensen, T.H., 2000. The effect of pH on the complexation of Cd, Ni and Zn by dissolved organic carbon from leachate-polluted groundwater. *Water Res.* 34, 3743–3754. [https://doi.org/10.1016/S0043-1354\(00\)00127-5](https://doi.org/10.1016/S0043-1354(00)00127-5)
- Christensen, T.H., 1987. Cadmium soil sorption at low concentrations: V. Evidence of competition by other heavy metals. *Water. Air. Soil Pollut.* 34, 293–303. <https://doi.org/10.1007/BF00193777>
- Christensen, T.H., 1984. Cadmium soil sorption at low concentrations: I. Effect of time, cadmium load, pH, and calcium. *Water. Air. Soil Pollut.* 21, 105–114. <https://doi.org/10.1007/BF00163616>
- Collins, R.N., Merrington, G., McLaughlin, M.J., Morel, J.L., 2003. Organic ligand and pH effects on isotopically exchangeable cadmium in polluted soils. *Soil Sci. Soc. Am. J.* 67, 112–121.
- Cornu, J.Y., Denaix, L., Schneider, A., Pellerin, S., 2007. Temporal evolution of redox processes and free Cd dynamics in a metal-contaminated soil after rewetting. *Chemosphere* 70, 306–314. <https://doi.org/10.1016/j.chemosphere.2007.06.003>
- Cornu, J.Y., Parat, C., Schneider, A., Authier, L., Dauthieu, M., Sappin-Didier, V., Denaix, L., 2009. Cadmium speciation assessed by voltammetry, ion exchange and geochemical calculation in soil solutions collected after soil rewetting. *Chemosphere* 76, 502–508. <https://doi.org/10.1016/j.chemosphere.2009.03.016>
- Cornu, J.Y., Schneider, A., Jezequel, K., Denaix, L., 2011. Modelling the complexation of Cd in soil solution at different temperatures using the UV-absorbance of dissolved organic matter. *Geoderma* 162, 65–70. <https://doi.org/10.1016/j.geoderma.2011.01.005>
- Cornu, S., Clozel, B., 2000. Extractions séquentielles et spéciation des éléments trace

- métalliques dans les sols naturels : analyse critique. *Etudes Gest. des Sols* 7, 179–189.
- Coutelot, F., Sappin-Didier, V., Keller, C., Atteia, O., 2014. Comparison of soil solution sampling techniques to assess metal fluxes from contaminated soil to groundwater. *Environ. Monit. Assess.* 186, 8929–8941. <https://doi.org/10.1007/s10661-014-4055-4>
- Curtin, D., Peterson, M.E., Anderson, C.R., 2016. pH-dependence of organic matter solubility: Base type effects on dissolved organic C, N, P, and S in soils with contrasting mineralogy. *Geoderma* 271, 161–172. <https://doi.org/10.1016/j.geoderma.2016.02.009>
- D'Amore, J.J., Al-Abed, S.R., Scheckel, K.G., Ryan, J.A., 2005. Methods for Speciation of Metals in Soils. *J. Environ. Qual.* 34, 1707. <https://doi.org/10.2134/jeq2004.0014>
- de Groot, A.C., Peijnenburg, W.J.G.M., van den Hoop, M.A.G.T., Ritsema, R., van Venn, R.P.M., 1998. Heavy Metals in Dutch field soils: an experimental and theoretical study on equilibrium partitioning.
- de Livera, J., McLaughlin, M.J., Hettiarachchi, G.M., Kirby, J.K., Beak, D.G., 2011. Cadmium solubility in paddy soils: Effects of soil oxidation, metal sulfides and competitive ions. *Sci. Total Environ.* 409, 1489–1497. <https://doi.org/10.1016/j.scitotenv.2010.12.028>
- de Troyer, I., Merckx, R., Amery, F., Smolders, E., 2014. Factors Controlling the Dissolved Organic Matter Concentration in Pore Waters of Agricultural Soils. *Vadose Zo. J.* 13, 0. <https://doi.org/10.2136/vzj2013.09.0167>
- de Vries, W., McLaughlin, M.J., Groenenberg, J.E., 2011. Transfer functions for solid-solution partitioning of cadmium for Australian soils. *Environ. Pollut.* 159, 3583–3594. <https://doi.org/10.1016/j.envpol.2011.08.006>
- Deb, B.C., 1942. The estimation of free iron oxides in soils and clays and their removal. *J. Soil Sci. I.*
- Degryse, F., Broos, K., Smolders, E., Merckx, R., 2003. Soil solution concentration of Cd and Zn can be predicted with a CaCl<sub>2</sub> soil extract. *Eur. J. Soil Sci.* 54, 149–158. <https://doi.org/10.1046/j.1365-2389.2003.00503.x>
- Degryse, F., Smolders, E., Parker, D.R., 2009. Partitioning of metals (Cd, Co, Cu, Ni, Pb, Zn) in soils: concepts, methodologies, prediction and applications - a review. *Eur. J. Soil Sci.* 60, 590–612. <https://doi.org/10.1111/j.1365-2389.2009.01142.x>
- Del Castilho, P., Chardon, W.J., 1995. Uptake of soil cadmium by three field crops and its prediction by a pH-dependent Freundlich sorption model. *Plant Soil* 171, 263–266. <https://doi.org/10.1007/BF00010280>
- Deneux-Mustin, S., Roussel-Debet, S., Mustin, C., Henner, P., Munier-Lamy, C., Colle, C., Berthelin, J., Garnier-Laplace, J., Leyval, C., 2003. Mobilité et transfert racinaire des éléments en traces : influence des micro-organismes du sol. Éditions TEC & DOC.
- DG-SANCO, 2011. Summary record of the standing committee on the food chain and animal health in Bruxelles on 23 november 2011. SANCO-Ares 2011:1436040.
- Di Bonito, M., Breward, N., Crout, N., Smith, B., Young, S., 2008. Overview of Selected Soil Pore

- Water Extraction Methods for the Determination of Potentially Toxic Elements in Contaminated Soils Operational and Technical Aspects. *Environ. Geochemistry* 213–249. <https://doi.org/10.1016/B978-0-444-53159-9.00010-3>
- Di Bonito, M., Lofts, S., Groenenberg, J.E., 2018. Models of Geochemical Speciation: Structure and Applications, 2nd ed, *Environmental Geochemistry: Site Characterization, Data Analysis and Case Histories*. Elsevier B.V. <https://doi.org/10.1016/B978-0-444-63763-5.00012-4>
- Dijkstra, J.J., Meeussen, J.C.L., Comans, R.N.J., 2009. Evaluation of a generic multisurface sorption model for inorganic soil contaminants. *Environ. Sci. Technol.* 43, 6196–6201. <https://doi.org/10.1021/es900555g>
- Dijkstra, J.J., Meeussen, J.C.L., Comans, R.N.J., 2004. Leaching of Heavy Metals from Contaminated Soils: An Experimental and Modeling Study. *Environ. Sci. Technol.* 38, 4390–4395. <https://doi.org/10.1021/es049885v>
- Dong, D., 2000. Adsorption of Pb and Cd onto metal oxides and organic material in natural surface coatings as determined by selective extractions: new evidence for the importance of Mn and Fe oxides. *Water Res.* 34, 427–436. [https://doi.org/10.1016/S0043-1354\(99\)00185-2](https://doi.org/10.1016/S0043-1354(99)00185-2)
- Dudal, Y., Gérard, F., 2004. Accounting for natural organic matter in aqueous chemical equilibrium models: a review of the theories and applications. *Earth-Science Rev.* 66, 199–216. <https://doi.org/10.1016/j.earscirev.2004.01.002>
- Dwane, G.C., Tipping, E., 1998. Testing a humic speciation model by titration of copper-amended natural waters. *Environ. Int.* 24, 609–616. [https://doi.org/10.1016/S0160-4120\(98\)00046-4](https://doi.org/10.1016/S0160-4120(98)00046-4)
- Dzombak, D.A., Morel, F., 1990. Surface complexation modeling : hydrous ferric oxide. Wiley.
- EFSA, 2009. Cadmium in food - Scientific opinion of the Panel on Contaminants in the Food Chain. *EFSA J.* 7, 980. <https://doi.org/10.2903/j.efsa.2009.980>
- Elzinga, E.J., Van Grinsven, J.J.M., Swartjes, F.A., 1999. General purpose Freundlich isotherms for cadmium, copper and zinc in soils. *Eur. J. Soil Sci.* 50, 139–149. <https://doi.org/10.1046/j.1365-2389.1999.00220.x>
- European Union (2006). Commission regulation (EC) No. 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foodstuffs. *Off J Eur Union L* 364:5–24
- Evans, L.J., 1989. Chemistry of metal retention by soils. *Environ. Sci. Technol.* 23, 1046–1056. <https://doi.org/10.1021/es00067a001>
- FAO, 2015. *FAO Statistical Pocketbook 2015*, Food and Agriculture Organization of the United Nations. <https://doi.org/978-92-5-108802-9>
- Fardeau, J.C., 1993. Le phosphore assimilable des sols : sa représentation par un modèle fonctionnel à plusieurs compartiments. *Agronomie* 13, 317–331.
- Feng, X.H., Zhai, L.M., Tan, W.F., Liu, F., He, J.Z., 2007. Adsorption and redox reactions of heavy metals on synthesized Mn oxide minerals. *Environ. Pollut.* 147, 366–373. <https://doi.org/10.1016/j.envpol.2006.05.028>
- Fest, E.P.M.J., Temminghoff, E.J.M., Comans, R.N.J., van Riemsdijk, W.H., 2008. Partitioning of



- organic matter and heavy metals in a sandy soil: Effects of extracting solution, solid to liquid ratio and pH. *Geoderma* 146, 66–74. <https://doi.org/10.1016/j.geoderma.2008.05.005>
- François, M., Grant, C., Lambert, R., Sauvé, S., 2009. Prediction of cadmium and zinc concentration in wheat grain from soils affected by the application of phosphate fertilizers varying in Cd concentration. *Nutr. Cycl. Agroecosystems* 83, 125–133. <https://doi.org/10.1007/s10705-008-9204-0>
- Fujino, O., Kumagai, T., Shigematsu, T., Matusi, M., 1976. Coprecipitation of Cd with calcite. *Bull. Inst. Chem. Res.* 54, 312–319.
- Gao, X., Brown, K.R., Racz, G.J., Grant, C.A., 2010. Concentration of cadmium in durum wheat as affected by time, source and placement of nitrogen fertilization under reduced and conventional-tillage management. *Plant Soil* 337, 341–354. <https://doi.org/10.1007/s11104-010-0531-y>
- Garcia-Miragaya, J., Cardenas, R., Page, A.L., 1986. Surface loading effect of Cd and Zn sorption by kaolinite and montmorillonite from low concentration solutions. *Water, Air, Soil Pollut.* 27, 181–190.
- Garrett, R.G., MacLaurin, A.I., Gawalko, E.J., Tkachuk, R., Hall, G.E.M., 1998. A prediction model for estimating the cadmium content of durum wheat from soil chemistry. *J. Geochemical Explor.* 64, 101–110. [https://doi.org/10.1016/S0375-6742\(98\)00023-5](https://doi.org/10.1016/S0375-6742(98)00023-5)
- Ge, Y., MacDonald, D., Sauvé, S., Hendershot, W., 2005. Modeling of Cd and Pb speciation in soil solutions by WinHumicV and NICA-Donnan model. *Environ. Model. Softw.* 20, 353–359. <https://doi.org/10.1016/j.envsoft.2003.12.014>
- Gis Sol, 2011. Synthèse sur l'état des sols de France.
- Grant, C.A., Sheppard, S.C., 2008. Fertilizer impacts on cadmium availability in agricultural soils and crops. *Hum. Ecol. Risk Assess.* 14, 210–228. <https://doi.org/10.1080/10807030801934895>
- Gray, C.W., McLaren, R.G., Roberts, A.H.C., 2001. Cadmium concentrations in some New Zealand wheat grain. *New Zeal. J. Crop Hortic. Sci.* 29, 125–136. <https://doi.org/10.1080/01140671.2001.9514170>
- Gray, C.W., McLaren, R.G., Roberts, A.H.C., Condon, L.M., 1999. Solubility, sorption and desorption of native and added cadmium in relation to properties of soils in New Zealand. *Eur. J. Soil Sci.* 50, 127–137. <https://doi.org/10.1046/j.1365-2389.1999.00221.x>
- Groenenberg, J.E., Dijkstra, J.J., Bonten, L.T.C., de Vries, W., Comans, R.N.J., 2012. Evaluation of the performance and limitations of empirical partition-relations and process based multisurface models to predict trace element solubility in soils. *Environ. Pollut.* 166, 98–107. <https://doi.org/10.1016/j.envpol.2012.03.011>
- Groenenberg, J.E., Koopmans, G.F., Comans, R.N.J., 2010a. Uncertainty analysis of the nonideal competitive adsorption - Donnan model: Effects of dissolved organic matter variability on predicted metal speciation in soil solution. *Environ. Sci. Technol.* 44, 1340–

1346. <https://doi.org/10.1021/es902615w>

- Groenenberg, J.E., Lofts, S., 2014. The use of assemblage models to describe trace element partitioning, speciation, and fate: A review. *Environ. Toxicol. Chem.* 33, 2181–2196. <https://doi.org/10.1002/etc.2642>
- Groenenberg, J.E., Römkens, P.F.A.M., Comans, R.N.J., Luster, J., Pampura, T., Shotbolt, L., Tipping, E., de Vries, W., 2010b. Transfer functions for solid-solution partitioning of cadmium, copper, nickel, lead and zinc in soils: Derivation of relationships for free metal ion activities and validation with independent data. *Eur. J. Soil Sci.* 61, 58–73. <https://doi.org/10.1111/j.1365-2389.2009.01201.x>
- Gryschko, R., Kuhnle, R., Terytze, K., Breuer, J., Stahr, K., 2005. Soil Extraction of Readily Soluble Heavy Metals and As with 1 M NH<sub>4</sub>NO<sub>3</sub>-Solution - Evaluation of DIN 19730. *J. Soils Sediments* 5, 101–106. <https://doi.org/10.1065/jss2004.10.119>
- Gu, X., Evans, L.J., 2007. Modelling the adsorption of Cd(II), Cu(II), Ni(II), Pb(II), and Zn(II) onto Fithian illite. *J. Colloid Interface Sci.* 307, 317–325. <https://doi.org/10.1016/j.jcis.2006.11.022>
- Gu, X., Sun, J., Evans, L.J., 2014. The development of a multi-surface soil speciation model for Cd (II) and Pb (II): Comparison of two approaches for metal adsorption to clay fractions. *Appl. Geochemistry* 47, 99–108. <https://doi.org/10.1016/j.apgeochem.2014.05.014>
- Gustafsson, J.P., Pechová, P., Berggren, D., 2003. Modeling Metal Binding to Soils: The Role of Natural Organic Matter. *Environ. Sci. Technol.* 37, 2767–2774. <https://doi.org/10.1021/es026249t>
- Guthrie, J.W., Hassan, N.M., Salam, M.S.A., Fafous, I.I., Murimboh, C.A., Murimboh, J., Chakrabarti, C.L., Grégoire, D.C., 2005. Complexation of Ni, Cu, Zn, and Cd by DOC in some metal-impacted freshwater lakes: A comparison of approaches using electrochemical determination of free-metal-ion and labile complexes and a computer speciation model, WHAM V and VI. *Anal. Chim. Acta* 528, 205–218. <https://doi.org/10.1016/j.aca.2004.10.003>
- Hamon, R.E., Bertrand, I., McLaughlin, M.J., 2002. Use and abuse of isotopic exchange data in soil chemistry. *Aust. J. Soil Res.* 40, 1371–1381. <https://doi.org/10.1071/SR02046>
- Harmsen, J., Rulkens, W., Eijsackers, H., 2005. Bioavailability: Concept for understanding or tool for predicting? *L. Contam. Reclam.* 13, 161–171.
- Harper, M.P., Davison, W., Zhang, H., Tych, W., 1998. Kinetics of metal exchange between solids and solutions in sediments and soils interpreted from DGT measured fluxes. *Geochim. Cosmochim. Acta* 62, 2757–2770. [https://doi.org/10.1016/S0016-7037\(98\)00186-0](https://doi.org/10.1016/S0016-7037(98)00186-0)
- Hénin, S., 1956. Quelques oxydes et hydroxydes considérés du point de vue pédologique. *Bull. du Groupe français des argiles* 8, 1–10. <https://doi.org/10.3406/argil.1956.928>
- Hiemstra, T., van Riemsdijk, W.H., 1996. A Surface Structural Approach to Ion Adsorption: The Charge Distribution (CD) Model. *J. Colloid Interface Sci.* 179, 488–508. <https://doi.org/10.1006/jcis.1996.0242>
- Hinsinger, P., Plassard, C., Tang, C., Jaillard, B., 2003. Origins of root-mediated pH changes in

- the rhizosphere and their responses to environmental constraints: A review. *Plant Soil* 248, 43–59. <https://doi.org/10.1023/A:1022371130939>
- Hodgson, J.F., Tiller, K.G., Fellows, M., 1964. The Role of Hydrolysis in the Reaction of Heavy Metals with Soil-Forming Materials. *Soil Sci. Soc. Am. J.* 28, 42. <https://doi.org/10.2136/sssaj1964.03615995002800010026x>
- Horn, A.L., Düring, R.A., Gäth, S., 2004. Sorption of Cd in soils: Pedotransfer functions for the parameters of the freundlich sorption isotherm. *Water, Air, Soil Pollut. Focus* 4, 61–71. <https://doi.org/10.1023/B:WAFO.0000028345.79442.2e>
- Houba, V.J.G., Novozamsky, I., Lexmond, T.M., van der Lee, J.J., 1990. Applicability of 0.01 M CaCl<sub>2</sub> as a single extraction solution for the assessment of the nutrient status of soils and other diagnostic purposes. *Commun. Soil Sci. Plant Anal.* 21, 2281–2290. <https://doi.org/10.1080/00103629009368380>
- Hough, R.L., Young, S.D., Crout, N.M.J., 2003. Modelling of Cd, Cu, Ni, Pb and Zn uptake, by winter wheat and forage maize, from a sewage disposal farm. *Soil Use Manag.* 19, 19–27. <https://doi.org/10.1111/j.1475-2743.2003.tb00275.x>
- Husson, O., 2013. Redox potential (Eh) and pH as drivers of soil/plant/microorganism systems: A transdisciplinary overview pointing to integrative opportunities for agronomy. *Plant Soil* 362, 389–417. <https://doi.org/10.1007/s11104-012-1429-7>
- Hutton, M., 1983. Sources of cadmium in the environment. *Ecotoxicol. Environ. Saf.* 7, 9–24. [https://doi.org/10.1016/0147-6513\(83\)90044-1](https://doi.org/10.1016/0147-6513(83)90044-1)
- IARC, 1993. Cadmium and cadmium compounds. *IARC Monogr. Eval. Carcinog. Risks to Humans* 58, 119–237. <https://doi.org/10.1002/14356007.a04>
- ISO 17402:2008, n.d. Soil quality - Requirements and guidance for the selection and application of methods for the assessment of bioavailability of contaminants in soil and soil materials.
- Ivezić, V., Almás, Á.R., Singh, B.R., 2012. Predicting the solubility of Cd, Cu, Pb and Zn in uncontaminated Croatian soils under different land uses by applying established regression models. *Geoderma* 170, 89–95. <https://doi.org/10.1016/j.geoderma.2011.11.024>
- Izquierdo, M., Tye, A.M., Chenery, S.R., 2012. Sources, lability and solubility of Pb in alluvial soils of the River Trent catchment, U.K. *Sci. Total Environ.* 433, 110–122. <https://doi.org/10.1016/j.scitotenv.2012.06.039>
- Jancic, S.A., Stosic, B.Z., 2014. *Cadmium Effects on the Thyroid Gland*, 1st ed, Vitamins and Hormones. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-800095-3.00014-6>
- Järup, L., 2003. Hazards of heavy metal contamination. *Br. Med. Bull.* 68, 167–182. <https://doi.org/10.1093/bmb/ldg032>
- John, J., Salbu, B., Gjessing, E.T., Bjørnstad, H.E., 1988. Effect of pH, humus concentration and molecular weight on conditional stability constants of cadmium. *Water Res.* 22, 1381–1388. [https://doi.org/10.1016/0043-1354\(88\)90094-2](https://doi.org/10.1016/0043-1354(88)90094-2)
- Jong, R. De, Campbell, C.A., Nicholaichuk, W., 1983. Water retention equations and their

- relationship to soil organic matter and particle size distribution for disturbed samples. *Can. J. Soil Sci.* 63, 291–302. <https://doi.org/10.4141/cjss83-029>
- Juste, C., 1988. Appréciation de la mobilité et de la biodisponibilité des éléments en traces du sol (in French, with English summary). *Sci. du Sol* 26, 103–112.
- Kabata-Pendias, A., 2004. Soil-plant transfer of trace elements - An environmental issue. *Geoderma* 122, 143–149. <https://doi.org/10.1016/j.geoderma.2004.01.004>
- Kabata-Pendias, A., Pendias, H., 2001. Trace elements in soils and plants. CRC Press.
- Kalbitz, K., Solinger, S., Park, J.-H., Michalzik, B., Matzner, E., 2000. Controls on the Dynamics of Dissolved Organic matter in Soils: a Review. *Soil Sci.* 165, 277–304. <https://doi.org/10.1097/00010694-200004000-00001>
- Kalis, E.J.J., Temminghoff, E.J.M., Van Riemsdijk, W.H., Weng, L., 2007. Measuring Free Metal Ion Concentrations in Multicomponent Solutions Using the Donnan Membrane Technique Measuring Free Metal Ion Concentrations in Multicomponent Solutions Using the Donnan Membrane Technique. *Anal. Chem.* 40, 955–955. <https://doi.org/10.1021/ac0615403>
- Khaj, N.M., Öborn, I., Hillier, S., Gustafsson, J.P., 2008. Modeling of metal binding in tropical Fluvisols and Acrisols treated with biosolids and wastewater. *Chemosphere* 70, 1338–1346. <https://doi.org/10.1016/j.chemosphere.2007.09.046>
- Khan, M.A., Castro-Guerrero, N., Mendoza-Cozatl, D.G., 2014. Moving toward a precise nutrition: preferential loading of seeds with essential nutrients over non-essential toxic elements. *Front. Plant Sci.* 5, 1–7. <https://doi.org/10.3389/fpls.2014.00051>
- Kinniburgh, D.G., Milne, C.J., Benedetti, M.F., Pinheiro, J.P., Filius, J., Koopal, L.K., Van Riemsdijk, W.H., 1996. Metal Ion Binding by Humic Acid: Application of the NICA-Donnan Model. *Environ. Sci. Technol.* 30, 1687–1698. <https://doi.org/10.1021/es950695h>
- Lackovic, K., Wells, J.D., Johnson, B.B., Angove, M.J., 2004. Modeling the adsorption of Cd(II) onto kaolinite and Mulloorina illite in the presence of citric acid. *J. Colloid Interface Sci.* 270, 86–93. <https://doi.org/10.1016/j.jcis.2003.09.009>
- Laveuf, C., Cornu, S., Baize, D., Hardy, M., Josiere, O., Drouin, S., Bruand, A., Juillot, F., 2009. Zinc Redistribution in a Soil Developed from Limestone During Pedogenesis. *Pedosphere* 19, 292–304. [https://doi.org/10.1016/S1002-0160\(09\)60120-X](https://doi.org/10.1016/S1002-0160(09)60120-X)
- Lebourg, A., Sterckeman, T., Ciesielski, H., Proix, N., 1996. Intérêt de différents réactifs d'extraction chimique pour l'évaluation de la biodisponibilité des métaux en traces du sol pour l'évaluation de la biodisponibilité des métaux. *Agronomie* 16, 201–215.
- Lee, S.Z., Allen, H.E., Huang, C.P., Sparks, D.L., Sanders, P.F., Peijnenburg, W.J.G.M., 1996. Predicting soil-water partition coefficients for cadmium. *Environ. Sci. Technol.* 30, 3418–3424. <https://doi.org/10.1021/es9507933>
- Li, X., Ziadi, N., Bélanger, G., Cai, Z., Xu, H., 2011. Cadmium accumulation in wheat grain as affected by mineral N fertilizer and soil characteristics. *Can. J. Soil Sci.* 91, 521–531. <https://doi.org/10.4141/cjss10061>

- Li, Y.M., Chaney, R.L., Schneiter, A.A., Miller, J.F., Elias, E.M., Hammond, J.J., 1997. Screening for low grain cadmium phenotypes in sunflower, durum wheat and flax. *Euphytica* 94, 23–30. <https://doi.org/10.1023/A:1002996405463>
- Li, Z., Hofmann, A., Wolthers, M., Thomas, P., 2012. Reversibility of cadmium sorption to calcite revisited. *J. Colloid Interface Sci.* 368, 434–442. <https://doi.org/10.1016/j.jcis.2011.09.085>
- Lin, Z., Schneider, A., Sterckeman, T., Nguyen, C., 2016. Ranking of mechanisms governing the phytoavailability of cadmium in agricultural soils using a mechanistic model. *Plant Soil* 399, 89–107. <https://doi.org/10.1007/s11104-015-2663-6>
- Liu, K., Lv, J., He, W., Zhang, H., Cao, Y., Dai, Y., 2015. Major factors influencing cadmium uptake from the soil into wheat plants. *Ecotoxicol. Environ. Saf.* 113, 207–213. <https://doi.org/10.1016/j.ecoenv.2014.12.005>
- Lofts, S., Tipping, E., 2011. Assessing WHAM/Model VII against field measurements of free metal ion concentrations: Model performance and the role of uncertainty in parameters and inputs. *Environ. Chem.* 8, 501–516. <https://doi.org/10.1071/EN11049>
- Lofts, S., Tipping, E., 1998. An assemblage model for cation binding by natural particulate matter. *Geochim. Cosmochim. Acta* 62, 2609–2625. [https://doi.org/10.1016/S0016-7037\(98\)00183-5](https://doi.org/10.1016/S0016-7037(98)00183-5)
- Loganathan, P., Vigneswaran, S., Kandasamy, J., Naidu, R., 2012. Cadmium Sorption and Desorption in Soils: A Review. *Crit. Rev. Environ. Sci. Technol.* 42, 489–533. <https://doi.org/10.1080/10643389.2010.520234>
- Loveland, P.J., Digby, P., 1984. The extraction of Fe and Al by 0.1 M pyrophosphate solutions: a comparison of some techniques. *J. Soil Sci.* 35, 243–250. <https://doi.org/10.1111/j.1365-2389.1984.tb00280.x>
- Luo, J., Cheng, H., Ren, J., Davison, W., Zhang, H., 2014. Mechanistic insights from DGT and soil solution measurements on the uptake of Ni and Cd by radish. *Environ. Sci. Technol.* 48, 7305–7313. <https://doi.org/10.1021/es500173e>
- Marzouk, E.R., Chenery, S.R., Young, S.D., 2013. Predicting the solubility and lability of Zn, Cd, and Pb in soils from a minespoil-contaminated catchment by stable isotopic exchange. *Geochim. Cosmochim. Acta* 123, 1–16. <https://doi.org/10.1016/j.gca.2013.09.004>
- McBride, M.B., 1994. *Environmental Chemistry of Soils*. Oxford University Press Inc, New-York.
- McKeague, J.A., Day, J.H., 1966. Dithionite- and Oxalate- Extractable Fe and Al As Aids in Differentiating Various Classes of Soils. *Can. J. Soil Sci.* 46, 13–22. <https://doi.org/10.4141/cjss66-003>
- McLaughlin, M.J., Parker, D.R., Clarke, J.M., 1999. Metals and micronutrients - Food safety issues. *F. Crop. Res.* 60, 143–163. [https://doi.org/10.1016/S0378-4290\(98\)00137-3](https://doi.org/10.1016/S0378-4290(98)00137-3)
- Meers, E., Du Laing, G., Unamuno, V., Ruttens, A., Vangronsveld, J., Tack, F.M.G., Verloo, M.G., 2007. Comparison of cadmium extractability from soils by commonly used single extraction protocols. *Geoderma* 141, 247–259. <https://doi.org/10.1016/j.geoderma.2007.06.002>

- Meers, E., Unamuno, V., Vandegheuchte, M., Vanbroekhoven, K., Geebelen, W., Samson, R., Vangronsveld, J., Diels, L., Ruttens, A., Du Laing, G., Tack, F., 2005. Soil-solution speciation of Cd as affected by soil characteristics in unpolluted and polluted soils. *Environ. Toxicol. Chem.* 24, 499–509. <https://doi.org/10.1897/04-231R.1>
- Mench, M., Baize, D., Mocquot, B., 1997. Cadmium availability to wheat in five soil series from the Yonne district, Burgundy, France. *Environ. Pollut.* 95, 93–103. [https://doi.org/10.1016/S0269-7491\(96\)00078-4](https://doi.org/10.1016/S0269-7491(96)00078-4)
- Merckx, R., Brans, K., Smolders, E., 2001. Decomposition of dissolved organic carbon after soil drying and rewetting as an indicator of metal toxicity in soils. *Soil Biol. Biochem.* 33, 235–240. [https://doi.org/10.1016/S0038-0717\(00\)00135-8](https://doi.org/10.1016/S0038-0717(00)00135-8)
- Miller, R.O., Kissel, D.E., 2010. Comparison of Soil pH Methods on Soils of North America. *Soil Sci. Soc. Am. J.* 74, 310. <https://doi.org/10.2136/sssaj2008.0047>
- Milne, C.J., Kinniburgh, D.G., Tipping, E., 2001. Generic NICA-Donnan model parameters for proton binding by humic substances. *Environ. Sci. Technol.* 35, 2049–2059. <https://doi.org/10.1021/es000123j>
- Milne, C.J., Kinniburgh, D.G., van Riemsdijk, W.H., Tipping, E., 2003. Generic NICA-Donnan Model Parameters for Metal-Ion Binding by Humic Substances. *Environ. Sci. Technol.* 37, 958–971. <https://doi.org/10.1021/es0258879>
- Morel, F., Hering, J.G., Morel, F., 1993. Principles and applications of aquatic chemistry. Wiley.
- Murray, D.J., Healy, T.W., Fuerstenau, D.W., 1968. The Adsorption of Aqueous Metal on Colloidal Hydrous Manganese Oxide. *Adsorpt. From Aqueous Solut.* 74–81. <https://doi.org/10.1021/ba-1968-0079.ch007>
- Naidu, R., Bolan, N.S., Kookana, R.S., Tiller, K.G., 1994. Ionic-strength and pH effects on the sorption of cadmium and the surface charge of soils. *Eur. J. Soil Sci.* 45, 419–429. <https://doi.org/10.1111/j.1365-2389.1994.tb00527.x>
- Nan, Z., Li, J., Zhang, J., Cheng, G., 2002. Cadmium and zinc interactions and their transfer in soil-crop system under actual field conditions. *Sci. Total Environ.* 285, 187–195. [https://doi.org/10.1016/S0048-9697\(01\)00919-6](https://doi.org/10.1016/S0048-9697(01)00919-6)
- Nawrot, T., Plusquin, M., Hogervorst, J., Roels, H.A., Celis, H., Thijs, L., Vangronsveld, J., Van Hecke, E., Staessen, J.A., 2006. Environmental exposure to cadmium and risk of cancer: A prospective population-based study. *Lancet Oncol.* 7, 119–126. [https://doi.org/10.1016/S1470-2045\(06\)70545-9](https://doi.org/10.1016/S1470-2045(06)70545-9)
- NF ISO 10390 International Organization for Standardization (ISO), 2005. Soil quality – Soil pH determination.
- NF ISO 10694 International Organization for Standardization (ISO), 1995. Soil quality. Determination of organic and total carbon after dry combustion (elementary analysis).
- NF ISO 11465 International Organization for Standardization (ISO), 1994. Soil quality -- Determination of dry matter and water content on a mass basis -- Gravimetric method.

- NF X 31–105 standard Association Française de Normalisation (AFNOR), 2014. Soil quality - Determination of carbonate content - Volumetric method.
- NF X 31-107 standard Association Française de Normalisation (AFNOR), 1983. Soil Quality - Particle size determination by sedimentation - Pipette method.
- NF X 31-130 standard Association Française de Normalisation (AFNOR), 1999. Soil quality - Chemical methods - Determination of cationic exchange capacity (CEC) and extractible cations.
- NF X 31–147 standard Association Française de Normalisation (AFNOR), 1996. Soil quality. Soils, sediments. Total solubilizing by acid attack.
- Nguyen, C., Gourdain, E., Grignon, G., Barrier-Guillot, B., Méléard, B., 2017. Contamination of Durum Wheat by Cadmium : Lessons from the Field. Conference: Icobte conference 2017, ETH Zurich, Switzerland, July 2017. DOI10.13140/RG.2.2.27701.37603. <https://doi.org/10.13140/RG.2.2.27701.37603>
- Nordstrom, D.K., Plummer, L.N., Langmuir, D., Busenberg, E., May, H.M., Jones, B.F., Parkhurst, D.L., 1990. Revised Chemical Equilibrium Data for Major Water—Mineral Reactions and Their Limitations. pp. 398–413. <https://doi.org/10.1021/bk-1990-0416.ch031>
- Norvell, W. a., Wu, J., Hopkins, D.G., Welch, R.M., 2000. Association of Cadmium in Durum Wheat Grain with Soil Chloride and Chelate-Extractable Soil Cadmium. *Soil Sci. Soc. Am. J.* 64, 2162. <https://doi.org/10.2136/sssaj2000.6462162x>
- Nye, P.H., 1981. Changes of pH across the rhizosphere induced by roots. *Plant Soil* 61, 7–26. <https://doi.org/10.1007/BF02277359>
- O'Connor, G.A., O'Connor, C., Cline, G.R., 1984. Sorption of Cadmium by Calcareous Soils: Influence of Solution Composition. *Soil Sci. Soc. Am. J.* 48, 1244–1247.
- Oliver, D.P., Schultz, J.E., Tiller, K.G., Merry, R.H., 1993. The effect of crop rotations and tillage practices on cadmium concentration in wheat grain. *Aust. J. Agric. Res.* 44, 1221–1234. <https://doi.org/10.1071/AR9931221>
- Oste, L.A., Temminghoff, E.J.M., Van Riemsdijk, W.H., 2002. Solid-solution partitioning of organic matter in soils as influenced by an increase in pH or Ca concentration. *Environ. Sci. Technol.* 36, 208–214. <https://doi.org/10.1021/es0100571>
- Pagenkopf, G.K., 1983. Gill surface interaction model for trace-metal toxicity to fishes: role of complexation, pH, and water hardness. *Environ. Sci. Technol.* 17, 342–347. <https://doi.org/10.1021/es00112a007>
- Pan, Y., Koopmans, G.F., Bonten, L.T.C., Song, J., Luo, Y., Temminghoff, E.J.M., Comans, R.N.J., 2015. In-situ measurement of free trace metal concentrations in a flooded paddy soil using the Donnan Membrane Technique. *Geoderma* 241–242, 59–67. <https://doi.org/10.1016/j.geoderma.2014.11.003>
- Papadopoulos, P., Rowell, D.L.C.N.-P., 1988. The reactions of cadmium with calcium-carbonate surfaces. *J. Soil Sci.* 39, 23-36 ST-The reactions of cadmium with calcium-

- Parat, C., Cornu, J.Y., Schneider, A., Authier, L., Sapin-Didier, V., Denaix, L., Potin-Gautier, M., 2009. Comparison of two experimental speciation methods with a theoretical approach to monitor free and labile Cd fractions in soil solutions. *Anal. Chim. Acta* 648, 157–161. <https://doi.org/10.1016/j.aca.2009.06.052>
- Peng, L., Liu, P., Feng, X., Wang, Z., Cheng, T., Liang, Y., Lin, Z., Shi, Z., 2018. Kinetics of heavy metal adsorption and desorption in soil: Developing a unified model based on chemical speciation. *Geochim. Cosmochim. Acta* 224, 282–300. <https://doi.org/10.1016/j.gca.2018.01.014>
- Perrier, F., Yan, B., Candaudap, F., Pokrovsky, O.S., Gourdain, E., Meleard, B., Bussi re, S., Coriou, C., Robert, T., Nguyen, C., Cornu, J.Y., 2016. Variability in grain cadmium concentration among durum wheat cultivars: impact of aboveground biomass partitioning. *Plant Soil* 404, 307–320. <https://doi.org/10.1007/s11104-016-2847-8>
- Piccolo, A., 2001. The Supramolecular Structure of Humic Substances. *Soil Sci.* 166, 810–832. <https://doi.org/10.1097/00010694-200111000-00007>
- Pierre, A., Lamarche, J.M., Merciei, R., Foissy, A., Persello, J., 1990. Calcium as potential determining ion in aqueous calcite suspensions. *J. Dispers. Sci. Technol.* 11, 611–635. <https://doi.org/10.1080/01932699008943286>
- Ponthieu, M., Juillot, F., Hiemstra, T., van Riemsdijk, W.H., Benedetti, M.F., 2006. Metal ion binding to iron oxides. *Geochim. Cosmochim. Acta* 70, 2679–2698. <https://doi.org/10.1016/j.gca.2006.02.021>
- Ponthieu, M., Pourret, O., Marin, B., Schneider, A.R., Morvan, X., Conreux, A., Canc s, B., 2016. Evaluation of the impact of organic matter composition on metal speciation in calcareous soil solution: Comparison of Model VI and NICA-Donnan. *J. Geochemical Explor.* 165, 1–7. <https://doi.org/10.1016/j.gexplo.2016.01.017>
- Quevauviller, P., Griepink, B., Rauret, G., 1993. Single and sequential extraction in sediments and soils. *Int. J. Environ. Anal. Chem.* 51, 231–235. <https://doi.org/10.1080/03067319308027629>
- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Raber, B., K gel-Knabner, I., Stein, C., Klem, D., 1998. Partitioning of polycyclic aromatic hydrocarbons to dissolved organic matter from different soils. *Chemosphere* 36, 79–97. [https://doi.org/10.1016/S0045-6535\(97\)00352-4](https://doi.org/10.1016/S0045-6535(97)00352-4)
- Ran, J., Wang, D., Wang, C., Zhang, G., Zhang, H., 2016. Heavy metal contents, distribution, and prediction in a regional soil-wheat system. *Sci. Total Environ.* 544, 422–431. <https://doi.org/10.1016/j.scitotenv.2015.11.105>
- Rao, C.R.M., Sahuquillo, A., Lopez Sanchez, J.F., 2008. A Review of the Different Methods Applied in Environmental Geochemistry For Single and Sequential Extraction of Trace Elements in Soils and Related Materials, Water, Air, and Soil Pollution.



<https://doi.org/10.1007/s11270-007-9564-0>

- Ren, Z.L., Sivry, Y., Dai, J., Tharaud, M., Cordier, L., Benedetti, M.F., 2015a. Multi-element stable isotopic dilution and multi-surface modelling to assess the speciation and reactivity of cadmium and copper in soil. *Eur. J. Soil Sci.* 66, 973–982. <https://doi.org/10.1111/ejss.12298>
- Ren, Z.L., Tella, M., Bravin, M.N., Comans, R.N.J., Dai, J., Garnier, J.M., Sivry, Y., Doelsch, E., Straathof, A., Benedetti, M.F., 2015b. Effect of dissolved organic matter composition on metal speciation in soil solutions. *Chem. Geol.* 398, 61–69. <https://doi.org/10.1016/j.chemgeo.2015.01.020>
- Rennert, T., Rabus, W., Rinklebe, J., 2017. Modelling the concentrations of dissolved contaminants (Cd, Cu, Ni, Pb, Zn) in floodplain soils. *Environ. Geochem. Health* 39, 331–344. <https://doi.org/10.1007/s10653-016-9859-4>
- Rieuwerts, J.S., Ashmore, M.R., Farago, M.E., Thornton, I., 2006. The influence of soil characteristics on the extractability of Cd, Pb and Zn in upland and moorland soils. *Sci. Total Environ.* 366, 864–875. <https://doi.org/10.1016/j.scitotenv.2005.08.023>
- Rieuwerts, J.S., Thornton, I., Farago, M.E., Ashmore, M.R., 1998. Factors influencing metal bioavailability in soils: Preliminary investigations for the development of a critical loads approach for metals. *Chem. Speciat. Bioavailab.* 10, 61–75. <https://doi.org/10.3184/095422998782775835>
- Riffaldi, R., Levi-Minzi, R., 1975. Adsorption and desorption of Cd on humic acid fraction of soils. *Water, Air, Soil Pollut.* 5, 179–184.
- Ringbom, A., 1963. *Complexation in analytical chemistry*. John Wiley & Sons, Ltd.
- Römken, P.F.A.M., Groenenberg, J.E., Bonten, L.T.C., de Vries, W., Bril, J., 2004. Derivation of partition relationships to calculate Cd, Cu, Ni, Pb, Zn solubility and activity in soil solutions 75.
- Römken, P.F.A.M., Salomons, W., 1998. Cd, Cu and Zn solubility in arable and forest soils: Consequences of land use changes for metal mobility and risk assessment. *Soil Sci.* 163, 859–871. <https://doi.org/10.1097/00010694-199811000-00003>
- Ross, S., 1994. *Toxic metals in soil-plant systems*. Wiley.
- Roth, V.-N., Dittmar, T., Gaupp, R., Gleixner, G., 2015. The Molecular Composition of Dissolved Organic Matter in Forest Soils as a Function of pH and Temperature. *PLoS One* 10, e0119188. <https://doi.org/10.1371/journal.pone.0119188>
- Saar, A., Weber, H., 1979. Complexation of cadmium(II) with water- and soil-derived fulvic acids: effect of pH and fulvic acid concentration. *Can. J. Chem.* 57, 1–6.
- Sappin-Didier, V., Mench, M., Gomez, A., Masson, P., 1997. Évaluation par des extractions sélectives de l'immobilisation du Cd après l'apport de matériaux inorganiques dans deux terres polluées. *Comptes Rendus l'Académie des Sci. - Ser. III - Sci. la Vie* 320, 413–419. [https://doi.org/10.1016/S0764-4469\(97\)85030-5](https://doi.org/10.1016/S0764-4469(97)85030-5)
- Satarug, S., Baker, J.R., Urbanjapol, S., Haswell-Elkins, M., Reilly, P.E.B., Williams, D.J., Moore,

- M.R., 2003. A global perspective on cadmium pollution and toxicity in non-occupationally exposed population. *Toxicol. Lett.* 137, 65–83. [https://doi.org/10.1016/S0378-4274\(02\)00381-8](https://doi.org/10.1016/S0378-4274(02)00381-8)
- Sauvé, S., Hendershot, W.H., Allen, H.E., 2000a. Solid-Solution Partitioning of Metals in Contaminated Soils: Dependence on pH, Total Metal Burden, and Organic Matter. *Environ. Sci. Technol.* 34, 1125–1131. <https://doi.org/10.1021/es9907764>
- Sauvé, S., Norvell, W.A., McBride, M.B., Hendershot, W.H., 2000b. Speciation and Complexation of Cadmium in Extracted Soil Solutions. *Environ. Sci. Technol.* 34, 291–296. <https://doi.org/10.1021/es990202z>
- Schindler, P.W., Liechti, P., Westall J.C., 1987. Adsorption of copper cadmium and lead from aqueous solution to the kaolinite water interface. *Netherlands J. Agric. Sci.* 35, 219–230.
- Schneider, A., 2006. Adaptation of the ion exchange method for the determination of the free ionic fraction of cadmium in solution. *J. Environ. Qual.* 35, 394–401. <https://doi.org/10.2134/jeq2005.0094>
- Schneider, A.R., Cancès, B., Ponthieu, M., Sobanska, S., Benedetti, M.F., Pourret, O., Conreux, A., Calandra, I., Martinet, B., Morvan, X., Gommeaux, M., Marin, B., 2016a. Lead distribution in soils impacted by a secondary lead smelter: Experimental and modelling approaches. *Sci. Total Environ.* 568, 155–163. <https://doi.org/10.1016/j.scitotenv.2016.06.001>
- Schneider, A.R., Ponthieu, M., Cancès, B., Conreux, A., Morvan, X., Gommeaux, M., Marin, B., Benedetti, M.F., 2016b. Influence of dissolved organic matter and manganese oxides on metal speciation in soil solution: A modelling approach. *Environ. Pollut.* 213, 618–627. <https://doi.org/10.1016/j.envpol.2016.03.010>
- Schnitzer, M., Khan, S.U., 1972. *Humic Substances in the Environment*. Marcel Dekker Inc., New York.
- Schröder, T.J., Hiemstra, T., Vink, J.P.M., Van der Zee, S.E.A.T.M., 2005. Modeling of the solid-solution partitioning of heavy metals and arsenic in embanked flood plain soils of the rivers rhine and meuse. *Environ. Sci. Technol.* 39, 7176–7184. <https://doi.org/10.1021/es048154s>
- Serrano, S., O'Day, P. a., Vlassopoulos, D., García-González, M.T., Garrido, F., 2009. A surface complexation and ion exchange model of Pb and Cd competitive sorption on natural soils. *Geochim. Cosmochim. Acta* 73, 543–558. <https://doi.org/10.1016/j.gca.2008.11.018>
- Shannon, R.D., 1976. Revised Effective Ionic Radii and Systematic Studies of Interatomic Distances in Halides and Chalcogenides. *Acta Crystallogr. - Sect. A* 32, 751–767.
- Shao, J., 1993. Linear model selection by cross-validation. *J. Am. Stat. Assoc.* 88, 496–494. <https://doi.org/10.1080/01621459.1993.10476299>
- Sheppard, S., Long, J., Sanipelli, B., 2009. Solid/liquid partition coefficients (K<sub>d</sub>) for selected soils and sediments at Forsmark and Laxemar-Simpevarp.
- Sigg, L., Black, F., Buffle, J., Cao, J., Cleven, R., Davison, W., Galceran, J., Gunkel, P., Kalis, E., Kistler, D., Martin, M., Noël, S., Nur, Y., Odzak, N., Puy, J., Van Riemsdijk, W.H.,

- Temminghoff, E.J.M., Tercier-Waeber, M.L., Toepperwien, S., Town, R.M., Unsworth, E., Warnken, K.W., Weng, L., Xue, H., Zhang, H., 2006. Comparison of analytical techniques for dynamic trace metal speciation in natural freshwaters. *Environ. Sci. Technol.* 40, 1934–1941. <https://doi.org/10.1021/es051245k>
- Sivry, Y., Munoz, M., Sappin-Didier, V., Riotte, J., Denaix, L., de Parseval, P., Destrigneville, C., Dupre, B., 2010. Multimetallic contamination from Zn-ore smelter: solid speciation and potential mobility in riverine floodbank soils of the upper Lot River (SW France). *Eur. J. Mineral.* 22, 679–691. <https://doi.org/10.1127/0935-1221/2010/0022-2060>
- Sivry, Y., Riotte, J., Sappin-Didier, V., Munoz, M., Redon, P.O., Denaix, L., Dupré, B., 2011. Multielementary (Cd, Cu, Pb, Zn, Ni) stable isotopic exchange kinetic (SIEK) method to characterize polymetallic contaminations. *Environ. Sci. Technol.* 45, 6247–6253. <https://doi.org/10.1021/es2006644>
- Sjöstedt, C.S., Gustafsson, J.P., Köhler, S.J., 2010. Chemical Equilibrium Modeling of Organic Acids, pH, Aluminum, and Iron in Swedish Surface Waters. *Environ. Sci. Technol.* 44, 8587–8593. <https://doi.org/10.1021/es102415r>
- Smith, R.M., Martell, A.E., 1976. Inorganic Ligands, in: *Critical Stability Constants*. Springer US, Boston, MA, pp. 1–129. [https://doi.org/10.1007/978-1-4757-5506-0\\_1](https://doi.org/10.1007/978-1-4757-5506-0_1)
- Spadini, L., Manceau, A., Schindler, P.W., Charlet, L., 1994. Structure and Stability of Cd<sup>2+</sup> Surface Complexes on Ferric Oxides. *J. Colloid Interface Sci.* 168, 73–86. <https://doi.org/10.1006/jcis.1994.1395>
- Sparks, D.L., 2003. *Environmental Soil Chemistry - Second Edition*, Environmental Soil Chemistry. Academic Press. <https://doi.org/10.1016/B978-012656446-4/50004-9>
- Sperotto, R.A., 2013. Zn/Fe remobilization from vegetative tissues to rice seeds: should I stay or should I go? Ask Zn/Fe supply! *Front. Plant Sci.* 4, 1–4. <https://doi.org/10.3389/fpls.2013.00464>
- Sposito, G., 2008. *The Chemistry of Soils*. Oxford University Press, New York.
- Srivastava, P., Singh, B., Angove, M., 2005. Competitive adsorption behavior of heavy metals on kaolinite. *J. Colloid Interface Sci.* 290, 28–38. <https://doi.org/10.1016/j.jcis.2005.04.036>
- Sterckeman, T., Carignan, J., Srayeddin, I., Baize, D., Cloquet, C., 2009. Availability of soil cadmium using stable and radioactive isotope dilution. *Geoderma* 153, 372–378. <https://doi.org/10.1016/j.geoderma.2009.08.026>
- Sterckeman, T., Douay, F., Baize, D., Fourier, H., Proix, N., Schwartz, C., 2006. Trace elements in soils developed in sedimentary materials from Northern France. *Geoderma* 136, 912–929. <https://doi.org/10.1016/j.geoderma.2006.06.010>
- Stolt, J.P., Sneller, F.E.C., Bryngelsson, T., Lundborg, T., Schat, H., 2003. Phytochelatin and cadmium accumulation in wheat. *Environ. Exp. Bot.* 49, 21–28. [https://doi.org/10.1016/S0098-8472\(02\)00045-X](https://doi.org/10.1016/S0098-8472(02)00045-X)
- Strobel, B.W., 2001. Influence of vegetation on low-molecular-weight carboxylic acids in soil

- solution—a review. *Geoderma* 99, 169–198. [https://doi.org/10.1016/S0016-7061\(00\)00102-6](https://doi.org/10.1016/S0016-7061(00)00102-6)
- Suda, A., Makino, T., 2016. Functional effects of manganese and iron oxides on the dynamics of trace elements in soils with a special focus on arsenic and cadmium: A review. *Geoderma* 270, 68–75. <https://doi.org/10.1016/j.geoderma.2015.12.017>
- Symeonides, C., McRae, S.G., 1977. The assessment of plant-available cadmium in soils. *J. Environ. Qual.* 6, 120. <https://doi.org/10.2134/jeq1977.00472425000600020003x>
- Temminghoff, E.J.M., Plette, A.C.C., Van Eck, R., Van Riemsdijk, W.H., 2000. Determination of the chemical speciation of trace metals in aqueous systems by the Wageningen Donnan Membrane Technique. *Anal. Chim. Acta* 417, 149–157. [https://doi.org/10.1016/S0003-2670\(00\)00935-1](https://doi.org/10.1016/S0003-2670(00)00935-1)
- Temminghoff, E.J.M., van der Zee, S.E.A.T.M., de Haan, F.A.M., 1995. Speciation and calcium competition effects on cadmium sorption by sandy soil at various pHs. *Eur. J. Soil Sci.* 46, 649–655. <https://doi.org/10.1111/j.1365-2389.1995.tb01361.x>
- Temminghoff, E.J.M., Van der Zee, S.E.A.T.M., de Haan, F.A.M., 1998. Effects of dissolved organic matter on the mobility of copper in a contaminated sandy soil. *Eur. J. Soil Sci.* <https://doi.org/10.1046/j.1365-2389.1998.4940617.x>
- Tessier, A., Campbell, P.G.C., Bisson, M., 1979. Sequential Extraction Procedure for the Speciation of Particulate Trace Metals. *Anal. Chem.* 51, 844–851. <https://doi.org/10.1021/ac50043a017>
- Thurman, E.M., Malcolm, R.L., 1981. Preparative isolation of aquatic humic substances. *Environ. Sci. Technol.* 15, 463–6. <https://doi.org/10.1021/es00086a012>
- Tiller, K.G., Nayyar, V.K., Clayton, P.M., 1979. Specific and non-specific sorption of cadmium by soil clays as influenced by zinc and calcium. *Aust. J. Soil Res.* 17, 17–28. <https://doi.org/10.1071/SR9790017>
- Tipping, E., 1994. WHAM - A chemical equilibrium model and computer code for waters, sediments, and soils incorporating a discrete site/electrostatic model of ion-binding by humic substances. *Comput. Geosci.* 20, 973–1023. [https://doi.org/10.1016/0098-3004\(94\)90038-8](https://doi.org/10.1016/0098-3004(94)90038-8)
- Tipping, E., Hurley, M.A., 1992. A unifying model of cation binding by humic substances. *Geochim. Cosmochim. Acta* 56, 3627–3641. [https://doi.org/10.1016/0016-7037\(92\)90158-F](https://doi.org/10.1016/0016-7037(92)90158-F)
- Tipping, E., Lofts, S., Sonke, J.E., 2011. Humic Ion-Binding Model VII: A revised parameterisation of cation-binding by humic substances. *Environ. Chem.* 8, 225–235. <https://doi.org/10.1071/EN11016>
- Tipping, E., Rieuwerts, J., Pan, G., Ashmore, M.R., Lofts, S., Hill, M.T.R., Farago, M.E., Thornton, I., 2003. The solid-solution partitioning of heavy metals (Cu, Zn, Cd, Pb) in upland soils of England and Wales. *Environ. Pollut.* 125, 213–225. [https://doi.org/10.1016/S0269-7491\(03\)00058-7](https://doi.org/10.1016/S0269-7491(03)00058-7)
- Tipping, E., Woof, C., 1990. Humic substances in acid organic soils: modelling their release to

- the soil solution in terms of humic charge. *J. Soil Sci.* 41, 573–586. <https://doi.org/10.1111/j.1365-2389.1990.tb00227.x>
- Tonkin, J.W., Balistrieri, L.S., Murray, J.W., 2004. Modeling sorption of divalent metal cations on hydrous manganese oxide using the diffuse double layer model. *Appl. Geochemistry* 19, 29–53. [https://doi.org/10.1016/S0883-2927\(03\)00115-X](https://doi.org/10.1016/S0883-2927(03)00115-X)
- Tóth, G., Hermann, T., Da Silva, M.R., Montanarella, L., 2016. Heavy metals in agricultural soils of the European Union with implications for food safety. *Environ. Int.* 88, 299–309. <https://doi.org/10.1016/j.envint.2015.12.017>
- Unsworth, E.R., Warnken, K.W., Zhang, H.A.O., Gunkel, P., Kalis, E., Leeuwen, H.P.V.A.N., Noe, Ä.P., Town, R.M., Weng, L., 2006. Model Predictions of Metal Speciation in Freshwaters Compared to Measurements by In Situ Techniques 40, 1942–1949.
- Ure, A.M., 1996. Single extraction schemes for soil analysis and related applications. *Sci. Total Environ.* 178, 3–10. [https://doi.org/10.1016/0048-9697\(95\)04791-3](https://doi.org/10.1016/0048-9697(95)04791-3)
- Ure, A.M., 1991. Trace Element Speciation in Soils, Soil Extracts and Solutions. *Mikrochim. Acta* 2, 49–57.
- Van Leeuwen, H.P., Town, R.M., Buffle, J., Cleven, R.F.M.J., Davison, W., Puy, J., Van Riemsdijk, W.H., Sigg, L., 2005. Dynamic speciation analysis and bioavailability of metals in aquatic systems. *Environ. Sci. Technol.* 39, 8545–8556. <https://doi.org/10.1021/es050404x>
- van Zomeren, A., Comans, R.N.J., 2007. Measurement of Humic and Fulvic Acid Concentrations and Dissolution Properties by a Rapid Batch Procedure. *Environ. Sci. Technol.* 41, 6755–6761. <https://doi.org/10.1021/es0709223>
- Vasconcelos, I.F., Haack, E.A., Maurice, P.A., Bunker, B.A., 2008. EXAFS analysis of cadmium(II) adsorption to kaolinite. *Chem. Geol.* 249, 237–249. <https://doi.org/10.1016/j.chemgeo.2008.01.001>
- Violante, A., 2013. Elucidating Mechanisms of Competitive Sorption at the Mineral/Water Interface, *Advances in Agronomy*. Elsevier. <https://doi.org/10.1016/B978-0-12-405942-9.00003-7>
- Wang, Y., Tang, X., Chen, Y., Zhan, L., Li, Z., Tang, Q., 2009. Adsorption behavior and mechanism of Cd ( II ) on loess soil from China 172, 30–37. <https://doi.org/10.1016/j.jhazmat.2009.06.121>
- Weng, L., Temminghoff, E.J.M., Lofts, S., Tipping, E., van Riemsdijk, W.H., 2002a. Complexation with Dissolved Organic Matter and Solubility Control of Heavy Metals in a Sandy Soil Complexation with Dissolved Organic Matter and Solubility Control of Heavy Metals in a Sandy Soil 36, 4804–4810. <https://doi.org/10.1021/es0200084>
- Weng, L., Temminghoff, E.J.M., Van Riemsdijk, W.H., 2002b. Aluminum speciation in natural waters: Measurement using Donnan membrane technique and modeling using NICA-Donnan. *Water Res.* 36, 4215–4226. [https://doi.org/10.1016/S0043-1354\(02\)00166-5](https://doi.org/10.1016/S0043-1354(02)00166-5)
- Weng, L., Temminghoff, E.J.M., Van Riemsdijk, W.H., 2001. Contribution of individual sorbents

- to the control of heavy metal activity in sandy soil. *Environ. Sci. Technol.* 35, 4436–4443. <https://doi.org/10.1021/es010085j>
- Wenzel, W.W., Blum, W.E.H., Brandstetter, A., Jockwer, F., Köchl, A., Oberförster, M., Oberländer, H.E., Riedler, C., Roth, K., Vladeva, I., 1996. Effects of soil properties and cultivar on cadmium accumulation in wheat grain. *Zeitschrift für Pflanzenernährung und Bodenkd.* 159, 609–614. <https://doi.org/10.1002/jpln.1996.3581590613>
- Williams, P.N., Zhang, H., Davison, W., Zhao, S., Lu, Y., Dong, F., Zhang, L., Pan, Q., 2012. Evaluation of in situ DGT measurements for predicting the concentration of Cd in Chinese field-cultivated rice: Impact of soil Cd:Zn ratios. *Environ. Sci. Technol.* 46, 8009–8016. <https://doi.org/10.1021/es301195h>
- Wood, S.A., 1996. The role of humic substances in the transport and fixation of metals of economic interest (Au, Pt, Pd, U, V). *Ore Geol. Rev.* 11, 1–31. [https://doi.org/10.1016/0169-1368\(95\)00013-5](https://doi.org/10.1016/0169-1368(95)00013-5)
- Xue, H., Sigg, L., 2002. A Review of Competitive Ligand-Exchange-Voltammetric Methods for Speciation of Trace Metals in Freshwater. pp. 336–370. <https://doi.org/10.1021/bk-2002-0811.ch018>
- Yin, Y., Impellitteri, C.A., You, S.J., Allen, H.E., 2002. The importance of organic matter distribution and extract soil:solution ratio on the desorption of heavy metals from soils. *Sci. Total Environ.* 287, 107–119. [https://doi.org/10.1016/S0048-9697\(01\)01000-2](https://doi.org/10.1016/S0048-9697(01)01000-2)
- Young, S.D., Tye, A., Carstensen, A., Resende, L., Crout, N., 2000. Methods for determining labile cadmium and zinc in soil. *Eur. J. Soil Sci.* 51, 129–136. <https://doi.org/10.1046/j.1365-2389.2000.00286.x>
- Zachara, J.M., Cowan, C.E., Resch, C.T., 1991. Sorption of divalent metals on calcite. *Geochim. Cosmochim. Acta* 55, 1549–1562. [https://doi.org/10.1016/0016-7037\(91\)90127-Q](https://doi.org/10.1016/0016-7037(91)90127-Q)
- Zhang, H., Young, S.D., 2005. Characterizing the availability of metals in contaminated soils. II. The soil solution. *Soil Use Manag.* 21, 459–467. <https://doi.org/10.1079/SUM2005349>
- Zhang, H., Zhao, F.J., Sun, B., Davison, W., Mcgrath, S.P., 2001. A new method to measure effective soil solution concentration predicts copper availability to plants. *Environ. Sci. Technol.* 35, 2602–2607. <https://doi.org/10.1021/es000268q>
- Zimmerl, S., Lafferty, J., Buerstmayr, H., 2014. Assessing diversity in *Triticum durum* cultivars and breeding lines for high versus low cadmium content in seeds using the CAPS marker usw47. *Plant Breed.* 133, 712–717. <https://doi.org/10.1111/pbr.12218>
- Zogaj, M., Düring, R.A., 2016. Plant uptake of metals, transfer factors and prediction model for two contaminated regions of Kosovo. *J. Plant Nutr. Soil Sci.* 179, 630–640. <https://doi.org/10.1002/jpln.201600022>

# Annexe 1: Protocole de récupération de la solution de sol

---

## Principe de la méthode

La méthode consiste à déplacer par forte centrifugation la colonne d'eau d'une suspension de sol et la séparer de la matrice solide à l'aide d'un solvant apolaire aprotique, donc fortement hydrophobe.

## Matériels nécessaires

- Coupelles en aluminium
- Tubes en polypropylène à centrifuger + Bouchons à ailettes
- Seringues BD, 10mL Luer-Lok Tip. Référence : 300912
- Filtres VWR International : "25mm Syringe Filter, 0.2µm, Cellulose Acetate Membrane". N°514-0060
- Flacons Snaptwist, 6.5mL
- Eau ultra pure (18.2 MΩ)
- 1,1,2-Trichloro-1,2,2-trifluoroéthane

## Contenu du mode opératoire

### 1 – Mesure de l'humidité résiduelle du sol (NF ISO 11465 :1993)

- Faire sécher deux coupelles en aluminium dans une étuve à 105°C, puis les laisser refroidir pendant 45min. Peser l'ensemble des coupelles et noter la masse  $m_0$ .
- Peser environ 10 g de sol séché à l'air et tamisé à 2mm dans une coupelle en aluminium. Peser le récipient fermé et noter la masse initiale ( $m_1$ ).
- Faire sécher l'ensemble dans une étuve à 105°C jusqu'à masse constante (~48h). Noter la masse finale ( $m_2$ ), après 45min hors de l'étuve.

$$\%HR = \frac{m_1 - m_2}{m_2 - m_0} \times 100$$

### 2- Récupération des solutions de sol :

- Peser 25g équivalent sec (tenir compte de l'humidité résiduelle) de sol dans un tube à centrifuger.

Noter la masse de sol et la masse totale (tube + bouchon + sol).

- Ajouter le volume V d'eau UP (résistivité 18.2Ω) nécessaire pour humidifier la masse m sol à pF 1.7, correspondant à une force de succion d'environ 50cm, c'est-à-dire à la capacité de rétention du sol.

Afin de minimiser la modification des caractéristiques de la SOM et de la stabilité des agrégats, il est conseillé d'humecter le sol par capillarité lente i.e. par dessous.

Le volume d'eau ajoutée est calculé en fonction de la capacité maximale de rétention en eau, préalablement mesurée ou obtenue par fonction de pédotransfert (De Jong et al., 1983).

- Laisser 24h au repos à température constante. (Rester constant entre les différents sols).
- Ajouter environ 10mL de solvant (1,1,2-Trichloro-1,2,2-trifluoroéthane).

- Immédiatement après ajout du solvant, centrifuger à 13000 rpm pendant 1h. (6 tubes à la fois)
- À l'aide d'une seringue 10mL récupérer la phase surnageante en veillant à ne pas prélever du solvant. Filtrer la solution à 0.2 $\mu$ m dans un flacon Snaptwist préalablement taré. Noter la masse de solution récupérée.



## Annexe 2 : Protocole d'estimation de la fraction libre

---

### Principe de la méthode

La solution de sol est mise en contact avec une résine échangeuse d'ion de type Amberlite. Le cadmium présent en solution sous sa forme libre va s'échanger avec le calcium présent sur la résine suivant un coefficient spécifique à ce couple d'élément appelé coefficient de Vanselow ( $V_{Ca/Cd}$ ).

L'utilisation d'un traceur radioactif ( $^{109}\text{Cd}$ ), se comportant de la même façon que  $\text{Cd}^{2+}$ , va permettre de suivre l'évolution de la concentration en  $\text{Cd}^{2+}$  en solution de manière plus sensible en comptant le  $^{109}\text{Cd}$  restant en solution.

### Matériels nécessaires

- Amberlite IR-120(plus) cation exchange resin (Na form; Aldrich, St. Louis, MO)
- Snaptwist Scintillation Vials 6.5m, 16mm\*57mm PolyEthylène Tube
- Tubes en PC 5mL
- Agitateur à rouleaux
- Tubes 12mL en PP pour dilutions et mesures Ca avant/après contact
- Compteur à scintillation liquide
- Matériel de radioprotection (écran, plateau à rebord,...) et appareil de détection (MIP 10 sonde X) de rayonnements ionisants
- Solution de  $\text{Ca}(\text{NO}_3)_2$  Suprapur
- Solution de NaOH
- Solution d' $\text{HNO}_3$  79%
- Liquide de scintillation Perkin Elmer
- Solution de  $^{109}\text{Cd}$  Perkin Elmer Amersham (CUSR74246)
- Eau ultra pure (18.2 $\Omega$ )

### Contenu du mode opératoire

#### 1 – Préparation de la résine :

- Peser 100g de résine sèche et la mettre en suspension dans 1L d'eau UP
- Acidifier la suspension de résine avec  $\text{HNO}_3$  à un pH de 0.5 puis rincer à l'eau UP.
- Alcaliniser avec NaOH à un pH de 11 puis rincer à l'eau UP.

Répéter l'opération deux fois.

- Faire tremper la résine dans une solution de  $\text{Ca}(\text{NO}_3)_2$  0.5M pendant 2h puis rincer à l'eau UP au-dessus d'un tamis nylon 200 $\mu\text{m}$ .

Répéter l'opération 4 fois

- Répéter ensuite ce cycle 4 fois avec une solution de  $\text{Ca}(\text{NO}_3)_2$  0.1M.
- Rincer abondamment à l'eau UP puis sécher dans un four à convection à 40°C.

#### 2 – Préparation de la solution radioactive de marquage :

- A partir de la solution commerciale de  $^{109}\text{Cd}$ , préparer un volume suffisant de solution de marquage pour tracer les N échantillons. L'activité dans 10 $\mu\text{L}$ , volume injecté dans la suspension, de cette solution doit être d'environ 150 Bq.
- Le volume initial de solution commerciale prélevée doit être faible (~100 $\mu\text{L}$ )

### 3 – Contact solution du sol-résine :

- Mettre en contact environ 2mL de solution de sol (obtenue par centrifugation ou bougie poreuse) avec une masse connue de résine d'environ 4mg. Augmenter la masse de résine pour des fortes concentrations en Ca en solution tout en respectant le rapport V/m de 0.5.
- Après un temps de contact de 24h, prélever 0.5 mL de solution dans un flacon à scintillation Snaptwist en veillant à ne pas récupérer de résine. Mettre de côté pour analyse de la concentration en Ca en solution
- Dans la solution restante (avec la résine), ajouter 10 $\mu$ L de  $^{109}\text{Cd}$  de la solution de marquage puis laisser 24h sur un agitateur à rouleaux (~40 rpm)

**Note : le rapport volume masse (solution de sol/résine) doit être le plus proche possible entre les contacts avec et sans radioactivité pour un échantillon.**

### 4 – Comptage de la radioactivité :

- Après 24h de contact, prélever 0.5 mL de solution dans un flacon à scintillation Snaptwist en veillant à ne pas récupérer de résine
- Ajouter 2.5 mL de liquide scintillant Instagel-Plus
- Préparer 5 blancs de mesure (0.5mL d'eau mQ + 2.5mL d'Instagel-plus) et 5 solutions contenant 10 $\mu$ L de solution injectée + 490 $\mu$ L d'eau mQ et 2.5mL d'Instagel-plus
- Compter les différents échantillons

### 5 – Mesure du calcium après contact :

- Après 24h, prélever 0.5 mL de la solution non radioactive et la diluer avec une solution de lanthane à 50 g/L afin d'obtenir une dilution par 50 de l'échantillon, un volume final de 10 mL et une concentration finale en La de 1 g/L
- Analyser par AAS flamme la concentration en Ca après contact.

## Annexe 3 : Modèles statistiques obtenus au cours des différents c

Les tableaux suivants sont un extrait des sorties de modélisation statistique. Le meilleur modèle ainsi qu la détermination ne variant pas de plus de 5 % de la valeur maximale sont présentés. Les signes des coefficients

### Chapitre II : Prédiction des concentrations en Cd<sup>2+</sup> en solution ou des K'n permettant le calcul des co MOD)

Modèle	r <sup>2</sup> obs/pred	r <sup>2</sup> aj	RMSE	BIC	pmax	VII
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +DOC+Ca <sub>ss</sub> +CO <sub>3ss</sub>	0.74978718	0.74626306	0.2545964	42.79088369	0.06986021	2.457
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +DOC+CaMg <sub>ss</sub> +CO <sub>3ss</sub>	0.74848568	0.74494322	0.25523214	43.92189101	0.06021686	2.387
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +Ca <sub>ss</sub> +CO <sub>3ss</sub>	0.7458876	0.74351272	0.25527216	40.77772792	2.02E-10	1.99
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +CaMg <sub>ss</sub> +CO <sub>3ss</sub>	0.74427102	0.74188103	0.25605905	42.16017677	4.04E-10	1.952
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +pH <sub>ss</sub> +DOC+SO <sub>4ss</sub>	0.74435017	0.74074947	0.25695298	47.47719158	0.0008323	1.689
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +pH <sub>ss</sub> +Cu <sub>ss</sub> +SO <sub>4ss</sub>	0.74106569	0.73741873	0.25827526	50.26012735	0.00364039	1.224
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +pH <sub>ss</sub> +CaMg <sub>ss</sub> +SO <sub>4ss</sub>	0.73566675	0.73194374	0.26058103	54.75882405	0.04365766	1.515
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +pH <sub>ss</sub> +Ca <sub>ss</sub> +SO <sub>4ss</sub>	0.73560415	0.73188026	0.26061839	54.81044251	0.04497874	1.500
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +pH <sub>ss</sub> +SO <sub>4ss</sub>	0.73055554	0.72803737	0.26207665	53.54938964	6.65E-07	1.3
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +pH <sub>ss</sub> +DOC+CaMg <sub>ss</sub>	0.727645	0.72380902	0.26518296	61.27609151	0.01719358	1.836
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +pH <sub>ss</sub> +DOC+Mg <sub>ss</sub>	0.72730549	0.72346472	0.26544416	61.54767648	0.02000144	1.686
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +pH <sub>ss</sub> +DOC+Ca <sub>ss</sub>	0.72707573	0.72323173	0.26548797	61.73127468	0.02216482	1.832
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +DOC+Cu <sub>ss</sub> +CO <sub>3ss</sub>	0.72565355	0.72178952	0.26710871	62.86430054	0.00825886	1.64
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +pH <sub>ss</sub> +DOC+NO <sub>3ss</sub>	0.72444696	0.72056593	0.26648625	63.8209737	0.07382756	1.776
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +pH <sub>ss</sub> +DOC	0.72027161	0.71765732	0.26746319	61.71497664	4.34E-05	1.624
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +DOC+Mg <sub>ss</sub> +CO <sub>3ss</sub>	0.72058837	0.71665299	0.26906049	66.85247235	0.07882976	1.676
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +DOC+CO <sub>3ss</sub>	0.71649725	0.71384769	0.26977788	64.63677194	3.59E-05	1.62
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +pH <sub>ss</sub> +CaMg <sub>ss</sub>	0.71492113	0.71225684	0.26918813	65.84537228	0.00037136	1.429
Cd <sup>2+</sup> ~ Cd <sub>cdL</sub> +pH <sub>ss</sub> +Ca <sub>ss</sub>	0.71401093	0.71133814	0.26964279	66.54029376	0.00053473	1.434

Modèle	r <sup>2</sup> obs/pred	r <sup>2</sup> aj	RMSE	BIC	pmax	VIF ma
K'n ~Ca <sub>ss</sub> +CO <sub>3</sub> <sub>ss</sub>	0.86073933	0.55396621	0.18733277	178.1522185	1.30E-15	1.6
K'n ~CaMg <sub>ss</sub> +CO <sub>3</sub> <sub>ss</sub>	0.85945218	0.5488392	0.18817235	180.746623	4.76E-15	1.571
K'n ~pH <sub>ss</sub> +CaMg <sub>ss</sub> +NO <sub>3</sub> <sub>ss</sub>	0.84597182	0.5252696	0.19800193	196.7154887	1.68E-07	1.502
K'n ~pH <sub>ss</sub> +Ca <sub>ss</sub> +NO <sub>3</sub> <sub>ss</sub>	0.84322915	0.51842127	0.19975678	199.9667325	4.59E-07	1.494
K'n ~pH <sub>ss</sub> +CaMg <sub>ss</sub> +SO <sub>4</sub> <sub>ss</sub>	0.83016439	0.49571868	0.20663783	210.42339	0.00030703	1.506
K'n ~pH <sub>ss</sub> +Ca <sub>ss</sub> +SO <sub>4</sub> <sub>ss</sub>	0.83007512	0.49549811	0.20669687	210.5226551	0.00032341	1.485
K'n ~pH <sub>ss</sub> +Cu <sub>ss</sub> +SO <sub>4</sub> <sub>ss</sub>	0.81959234	0.47001841	0.21255794	204.9725966	9.22E-05	1.094
K'n ~pH <sub>ss</sub> +CaMg <sub>ss</sub>	0.8190433	0.46551834	0.21378606	219.2171308	3.63E-07	1.310
K'n ~pH <sub>ss</sub> +Ca <sub>ss</sub>	0.81782527	0.46252564	0.21454052	220.4846168	6.93E-07	1.319

### Chapitre III : Modèles statistiques de distribution du Cd<sup>2+</sup>

Modèle	r <sup>2</sup> obs/pred	r <sup>2</sup> aj	RMSE	BIC	pmax	VIF
Cd <sup>2+</sup> ~Cd NH <sub>4</sub> +pH <sub>ss</sub> +ArgLF+Ca <sub>ss</sub>	0.83806115	0.83206342	0.1900954	-24.2845344	0.00623936	2.319
Cd <sup>2+</sup> ~Cd NH <sub>4</sub> +pH <sub>ss</sub> +Argile+Ca <sub>ss</sub>	0.83779029	0.83178252	0.18977241	-24.1408086	0.00672387	2.328
Cd <sup>2+</sup> ~Cd NH <sub>4</sub> +pH+MnO+Ca <sub>ss</sub>	0.83983872	0.82839863	0.16901842	-23.4690423	0.05901908	2.334
Cd <sup>2+</sup> ~Cd NH <sub>4</sub> +pH <sub>ss</sub> +MOS+Ca <sub>ss</sub>	0.8344529	0.82832153	0.19153627	-22.3893576	0.01691009	2.320
Cd <sup>2+</sup> ~Cd NH <sub>4</sub> +pH <sub>ss</sub> +CEC+Ca <sub>ss</sub>	0.83240683	0.82619967	0.19393277	-21.332956	0.02987005	2.336
Cd <sup>2+</sup> ~Cd NH <sub>4</sub> +pH <sub>ss</sub> +COS+Ca <sub>ss</sub>	0.83204732	0.82582685	0.19296263	-21.1486716	0.03302962	2.320
Cd <sup>2+</sup> ~Cd NH <sub>4</sub> +pH <sub>ss</sub> +Ca <sub>ss</sub>	0.82229413	0.81795984	0.19741353	-20.7485351	5.25E-07	2.315
Cd <sup>2+</sup> ~Cd NH <sub>4</sub> +Calcaire+MnO+Ca <sub>ss</sub>	0.82856985	0.81632484	0.1746921	-20.2732969	0.08539028	2.08
Cd <sup>2+</sup> ~Cd H <sub>2</sub> O+pH+CEC+MnO	0.81544498	0.80226248	0.18156335	-16.8060266	0.00731796	2.440
Cd <sup>2+</sup> ~Cd Pyro+pH CaCl <sub>2</sub> +NO <sub>3</sub> <sub>ss</sub> +Ca <sub>ss</sub>	0.81019519	0.7966377	0.17984719	-15.4877412	0.00012348	1.420
Cd <sup>2+</sup> ~Cd NH <sub>4</sub> +pH <sub>ss</sub> +AlO+Ca <sub>ss</sub>	0.80862534	0.79495572	0.18371532	-15.1006107	0.03861831	1.629
Cd <sup>2+</sup> ~Cd H <sub>2</sub> O+pH+CEC+AlO	0.80015637	0.78588182	0.18786145	-13.0654087	0.04859733	1.961
Cd <sup>2+</sup> ~Cd Pyro+pH <sub>ss</sub> +NO <sub>3</sub> <sub>ss</sub> +Ca <sub>ss</sub>	0.7955028	0.78089586	0.19183098	-11.9835131	1.99E-06	1.405

Modèle	r <sup>2</sup> obs/pred	r <sup>2</sup> aj	RMSE	BIC	pmax	VII
Cd <sub>ss</sub> -Cd NH <sub>4</sub> +ArgLF+MnO+Ca <sub>ss</sub>	0.85494738	0.84458647	0.14491465	-38.4229713	0.07020877	1.275
Cd <sub>ss</sub> -Cd NH <sub>4</sub> +ArgLF+AlO+Ca <sub>ss</sub>	0.85308517	0.84259126	0.14555872	-37.8234209	0.09733799	1.475
Cd <sub>ss</sub> -Cd H <sub>2</sub> O+Calcaire+CEC+FeO	0.83233615	0.82036016	0.15285116	-31.6143405	0.00125315	1.933
Cd <sub>ss</sub> -Cd H <sub>2</sub> O+Calcaire+AlO+Mn <sub>t</sub>	0.8309984	0.81892686	0.16106724	-31.2408286	0.0812079	1.841
Cd <sub>ss</sub> -Cd H <sub>2</sub> O+Calcaire+CEC+MnO	0.82627217	0.81386304	0.15911296	-29.944488	0.0027975	2.070
Cd <sub>ss</sub> -Cd H <sub>2</sub> O+pH+CEC+MnO	0.82500728	0.8125078	0.15830365	-29.6035251	0.00280909	2.440
Cd <sub>ss</sub> -Cd NH <sub>4</sub> +Argile+FeO+Ca <sub>ss</sub>	0.82438361	0.81183958	0.15438472	-29.4363176	0.02053552	1.640
Cd <sub>ss</sub> -Cd NH <sub>4</sub> +NO <sub>3</sub> <sub>ss</sub> +AlO+Ca <sub>ss</sub>	0.82164365	0.80890392	0.16005249	-28.7086879	0.01836349	1.552
Cd <sub>ss</sub> -Cd NH <sub>4</sub> +Argile+MnO+Ca <sub>ss</sub>	0.82147044	0.80871833	0.15511344	-28.6630648	0.03046595	1.230
Cd <sub>ss</sub> -Cd H <sub>2</sub> O+pH+CEC+AlO	0.8208743	0.80807961	0.16019467	-28.5063861	0.03972544	1.961
Cd <sub>ss</sub> -Cd H <sub>2</sub> O+pH+CEC+FeO	0.8190832	0.80616057	0.15907141	-28.0387612	0.00601703	1.979
Cd <sub>ss</sub> -Cd NH <sub>4</sub> +NO <sub>3</sub> <sub>ss</sub> +MnO+Ca <sub>ss</sub>	0.81657253	0.80347057	0.16069801	-27.391004	0.01779625	1.448
Cd <sub>ss</sub> -Cd H <sub>2</sub> O+Calc+CEC+AlO	0.81434683	0.80108589	0.16465394	-26.8241411	0.09763064	1.914

#### Chapitre IV : Modèles statistiques de prédiction de la teneur en Cd dans les grains de blé dur


Modèle	r <sup>2</sup> obs/pred	r <sup>2</sup> aj	RMSE	BIC	pmax	VII
Cd grain-Cd <sub>ss</sub> +Calcaire+CEC+MnO	0.82351523	0.81120234	0.03422531	-27.2117745	2.24E-05	1.275
Cd grain-Cd CaCl <sub>2</sub> +CEC+MnO+NO <sub>3</sub> <sub>ss</sub>	0.82032137	0.80778565	0.03602845	-26.3508816	0.00520503	1.475
Cd grain-Cd <sub>ss</sub> +pH+CEC+MnO	0.81669977	0.80391138	0.04181363	-25.393014	0.00112337	1.933
Cd grain-Cd CaCl <sub>2</sub> +CEC+MnO+SO <sub>4</sub> <sub>ss</sub>	0.81331555	0.80029106	0.03491231	-24.5148877	0.01291633	1.640
Cd grain-Cd CaCl <sub>2</sub> +CEC+AlO+Mg <sub>ss</sub>	0.81189882	0.79877548	0.04471748	-24.1519951	0.00641419	2.070
Cd grain-Cd CaCl <sub>2</sub> +CEC+MnO+Mg <sub>ss</sub>	0.80933457	0.79603233	0.03608196	-23.5020648	0.02152887	1.841
Cd grain-Cd <sup>2+</sup> +Calcaire+CEC+MnO	0.80796156	0.79456353	0.03506492	-23.1576484	0.00014937	2.070
Cd grain-Cd CaCl <sub>2</sub> +CEC+Mg <sub>ss</sub> +SO <sub>4</sub> <sub>ss</sub>	0.80047702	0.78655681	0.02953821	-21.3224229	0.02648829	1.640
Cd grain-Cd CaCl <sub>2</sub> +Argile+MnO+NO <sub>3</sub> <sub>ss</sub>	0.79961068	0.78563003	0.03646627	-21.1144558	0.00434399	1.475
Cd grain-Cd <sup>2+</sup> +pH+CEC+MnO	0.7994198	0.78542583	0.04706426	-21.0687542	0.00920946	2.070
Cd grain-Cd <sub>ss</sub> +Calcaire+Argile+MnO	0.79885676	0.78482351	0.04595189	-20.9342042	1.15E-05	1.275
Cd grain-Cd CaCl <sub>2</sub> +pH <sub>ss</sub> +CEC+Mg <sub>ss</sub>	0.79683218	0.78265768	0.03161554	-20.4534816	0.04152969	2.070

***Annexe 4 : Article publié dans la revue Environmental  
Science and Pollution Research***

---

## RESEARCH ARTICLE

# Predictive statistical modelling of cadmium content in durum wheat grain based on soil parameters

Yoann Viala<sup>1</sup>  · Julien Laurette<sup>1</sup> · Laurence Denaix<sup>1</sup> · Emmanuelle Gourdain<sup>2</sup> · Benoit Méleard<sup>2</sup> · Christophe Nguyen<sup>1</sup> · André Schneider<sup>1</sup> · Valérie Sappin-Didier<sup>1</sup>

Received: 31 March 2017 / Accepted: 4 July 2017 / Published online: 15 July 2017  
© Springer-Verlag GmbH Germany 2017

**Abstract** Regulatory limits on cadmium (Cd) content in food products are tending to become stricter, especially in cereals, which are a major contributor to dietary intake of Cd by humans. This is of particular importance for durum wheat, which accumulates more Cd than bread wheat. The contamination of durum wheat grain by Cd depends not only on the genotype but also to a large extent on soil Cd availability. Assessing the phytoavailability of Cd for durum wheat is thus crucial, and appropriate methods are required. For this purpose, we propose a statistical model to predict Cd accumulation in durum wheat grain based on soil geochemical properties related to Cd availability in French agricultural soils with low Cd contents and neutral to alkaline pH (soils commonly used to grow durum wheat). The best model is based on the concentration of total Cd in the soil solution, the pH of a soil CaCl<sub>2</sub> extract, the cation exchange capacity (CEC), and the content of manganese oxides (Tamm's extraction) in the soil. The model variables suggest a major influence of cadmium buffering power of the soil and of Cd speciation in solution. The model successfully explains 88% of Cd variability in grains with, generally, below 0.02 mg Cd kg<sup>-1</sup> prediction error in wheat grain. Monte Carlo cross-validation indicated that model accuracy will suffice for the European Community project to reduce the regulatory limit from 0.2 to 0.15 mg Cd kg<sup>-1</sup> grain, but not for the intermediate step at 0.175 mg

Cd kg<sup>-1</sup>. The model will help farmers assess the risk that the Cd content of their durum wheat grain will exceed regulatory limits, and help food safety authorities test different regulatory thresholds to find a trade-off between food safety and the negative impact a too strict regulation could have on farmers.

**Keywords** Cadmium · Durum wheat · Soil contamination · Statistical predictive model · Chemical extraction

## Introduction

Cadmium is a non-essential, highly toxic metal, and was classified by the International Agency for Research on Cancer (IARC 1993) as carcinogenic for humans. Chronic exposure to Cd, including through food, can cause damage to different organs as well as cancer (Cho et al. 2013; Jancic and Stosic 2014; Nawrot et al. 2006). Cd occurs naturally in agricultural soils, and its concentration generally depends on the parent material (Kabata-Pendias 2004). Although Cd contents are usually low in agricultural soils, they sometimes tend to increase due to anthropogenic inputs, including atmospheric deposition and mineral fertilizers (Hutton 1983). Despite these low or moderate contents, Cd can represent a threat to human health because the main pathway of exposure for the general population is consuming food made from contaminated plant products (Satarug et al. 2003). Therefore, in Europe, some regulatory limits have been established for maximum Cd content in some but not all foodstuffs (European Union 2006). For cereal grains except rice and wheat, the maximum Cd content is 0.1 mg kg<sup>-1</sup> (15% water content). For rice and wheat, the limit is 0.2 mg kg<sup>-1</sup> (European Union 2006). Wheat is one of the most widely produced and consumed cereals worldwide (FAO 2015), and, among wheat subspecies, durum wheat (*Triticum turgidum* subsp. *durum*) has been shown to

Responsible editor: Philippe Garrigues

✉ Yoann Viala  
yoann.viala@inra.fr

<sup>1</sup> ISPA, INRA, Bordeaux Sciences Agro, 33140 Villenave d'Ornon, France

<sup>2</sup> ARVALIS-Institut du végétal, Station Expérimentale, 91720 Boigneville, France

accumulate much more Cd than bread wheat (Li et al. 1997; Stolt et al. 2003). Consequently, durum wheat, which is widely consumed as semolina and pasta, is an important contributor to the dietary intake of Cd by European populations. Following the recommendation of the European Food Safety Authority to reduce the dietary exposure of humans to Cd almost threefold (EFSA 2009), the European Union proposed to reduce the maximum Cd content of durum wheat grain from 0.2 to 0.15 mg kg<sup>-1</sup> with an intermediate limit of 0.175 mg kg<sup>-1</sup> (DG SANCO 2011). These new limits call for simple predictive models to assess the risk of contamination of durum wheat grain by Cd.

Cadmium uptake by plants mainly depends on the plant genotype but also on Cd availability in soils, which is driven by different mechanisms, and which, in turn, can be affected by agricultural practices including crop rotation and tillage (Oliver et al. 1993). Total cadmium content in the soil is usually a poor indicator of the quantity of metal transferred to plants, as it is well established that only part of the total content can be transferred from the solid phases of the soil to the soil solution from which it is taken up by roots as free Cd<sup>2+</sup> (Campbell 1995). Exchanges of Cd between the soil solid phase and the soil solution are controlled by different mechanisms including ion exchange, specific adsorption, or complexation involving chemical bonds of variable strength like Van der Waals interactions or covalent bonds. Clay minerals, organic matter, metallic oxides, or oxi-hydroxides are the main constituents involved in the sorption of Cd onto the soil solid phase. In alkaline pH conditions, the availability of Cd for root uptake can decrease due to the precipitation of the metal, particularly with carbonates. In solution, Cd speciation is governed by complexation with organic and inorganic ligands. Protons and other cations (mainly Ca<sup>2+</sup> and Mg<sup>2+</sup>) compete with Cd<sup>2+</sup> for both sorption and complexation in solution (Del Castillo and Chardon 1995; Naidu et al. 1994; Ross 1994).

In France, durum wheat is usually cultivated on alkaline and calcareous soils with low Cd contents (median of 0.3 mg Cd kg<sup>-1</sup> dry soil) (Arvalis, personal communication, <https://www.english.arvalisinstitutduvegetal.fr/index.html>). Although Cd phytoavailability could be expected to be lower due to the high pH values and carbonate contents, durum wheat crops can accumulate Cd in grains at levels close to the regulatory limit (Arvalis, personal communication). This means that a significant part of the soil Cd is probably solubilized and taken up by wheat, but is not taken into account by the statistical models used to predict the soil-solution partitioning of total or free Cd (Degryse et al. 2009). In the literature, plant available Cd is typically based on a single extraction or on sequential extractions, using chemical extractants that differ in their ability to selectively extract metals bound to different soil fractions (Harmsen et al. 2005;

Lebourg et al. 1996; Meers et al. 2007; Tessier et al. 1979). However, these extractions cannot perfectly mimic resupply of Cd<sup>2+</sup> from the solid phase to the soil solution during depletion resulting from root uptake. Yet, because of the initial low concentration of soluble Cd<sup>2+</sup>, resupply is expected to play an important role in soils with alkaline pH and should therefore also be taken into account in the models. This can be accomplished by testing the cationic exchange capacity and the clay and organic matter contents as explanatory variables in the model used to predict Cd accumulation in the grain. Another possibility is to test the predictive value of the diffusion gradient in the thin film (DGT) technique developed by Zhang et al. (2001). DGT is a passive device that acts, like a root, as a sink for Cd<sup>2+</sup>. It creates Cd depletion at the interface with the soil solution and can hence account for resupply from the solid phase. Finally, in the modelling approach, competition for sorption and complexation with soil cations can be included by considering the pH and the concentrations of Ca and Mg in the soil solution.

Most statistical models which predict Cd content in wheat grain from soil variables are designed for bread wheat (Adams et al. 2004; Baize et al. 2009; Brus et al. 2005; François et al. 2009; Gray et al. 2001; Hough et al. 2003; Liu et al. 2015; Mench et al. 1997; Nan et al. 2002; Ran et al. 2016; Zogaj and Düring 2016), whereas few are devoted to durum wheat (Garrett et al. 1998; Norvell et al. 2000). Furthermore, in the latter above-cited durum wheat models, the goodness of fit ( $r^2$  of 0.74 and 0.84, respectively) was likely linked to specific soil characteristics. Garrett et al. (1998) postulated that in their soils, Cd availability was mainly controlled by Cd-organic matter complexes and their model only included the Cd and the soil organic C extractable by pyrophosphate. In the study by Norvell et al. (2000), the authors reported highly variable soil salinity in their soils. The first goal of the present work was thus to extend the number of candidate variables to build a predictive model of the availability of Cd for durum wheat. In addition to classical extractions and soil characteristics (soil total Cd content, clay, loam, and organic carbon, pH), we tested the concentration of Cd in the soil pore water, the free ionic fraction of Cd in solution, contents of calcareous and amorphous material, the Cd accumulated by DGT during a 48-h test, dissolved organic carbon, and finally concentrations of soluble Ca, Mg, and K. These variables were tested in multiple linear models to account for soil-solution Cd partitioning, the buffering capacity of the soil, speciation in solution, the carbonated fraction of Cd, and competition with major cations in the soil. The second goal of this work was to assess the predictive value of the models by cross-validation in order to estimate the error of prediction with respect to the current project of lowering the regulatory limits from 0.2 to 0.15 mg Cd/kg grain with an intermediate step at 0.175 mg Cd/kg (DG SANCO 2011).



## Materials and methods

### Sampling and preparation

Nine French agricultural sites were chosen for this study. Each site corresponded to a cultivar comparison trial with three blocks, of 1000 m<sup>2</sup>, per cultivar. Paired samples of topsoil and wheat (*T. turgidum* subsp. *durum*, Miradoux) were collected at the heading stage (around 60 days after sowing) from each of the three blocks. The soil was sampled in the 0–30-cm ploughed layer where most of the root system developed. A total of 27 paired samples of soil and wheat were used. Soil samples were air dried and sieved at 2 mm. The residual water content was determined after drying at 105 °C until constant weight (NF ISO 11465:1994).

The plants were also sampled at maturity (around 90 days after sowing) to determine the grain Cd content and in the immediate vicinity of the previous sampling (soil and plants) so that soil conditions are assumed to be homogeneous. The plants were divided into shoots (stems and leaves) and grains for the last sampling. Four to five plants were collected for each sampling, and a composite sample was made. Plant parts were carefully rinsed with permuted water and then with distilled water. These samples were dried at 60 °C for 48 h to constant weight and ground (<1 µm) using a Retsch Planetary Ball Mill PM 400. Aliquots of powdered samples were digested using a mixture of HNO<sub>3</sub> 69% (Aristar® for trace analysis, VWR Chemicals) and H<sub>2</sub>O<sub>2</sub> 30% (Ultraprace®, ppt-trace analysis grade, Scharlau), and the Cd in the extracts was quantified by furnace atomic absorption spectrometry (PinAAcle 900T, PerkinElmer). Grains sampled at maturity were air dried to a final constant water content of 15% as stated for the regulatory limit for Cd. The Cd contents of grains were determined by atomic absorption spectrometry after wet digestion (<http://www.capinov.fr/>).

### Soil characteristics

The following parameters were determined from air-dried soil subsamples according to French standardized procedures. Soil pH was determined in a distilled water (1:5 soil to solution ratio) and 0.01 M CaCl<sub>2</sub> extract (1:10 soil to solution ratio) (NF ISO 10390:2005). Total CaCO<sub>3</sub> content was quantified using the acid neutralisation method (NF X 31–105:2014). The total soil organic carbon fraction (SOC) was determined by dry combustion after correction for carbonate (NF ISO 10694:1995). The soil texture was determined using the Robinson pipette method (NF X 31–107:1983), and the cationic exchange capacity (CEC) was measured at soil pH using the cobaltihexamine chloride method (NF X 31–130:1999). The concentration of dissolved organic carbon (DOC) was measured in the 0.01 M CaCl<sub>2</sub> extract by oxidative combustion (TOC-VCSH, Shimadzu).

### Soil and soil solution extraction and analyses

Total Cd soil content was quantified after solubilisation by fluorhydric and perchloric acids (NF X 31–147:1996). After complete dissolution, the concentrations of Ca and Mg in the extracts were also analysed by ICP-AES (inductively coupled plasma atomic emission spectrometry) and in Cd by ICP-MS (inductively coupled plasma mass spectrometry).

Seven soil extractions were performed; the procedures, reagents, and soil/solution ratios are detailed in Table 1. At the end of the extraction, the soil suspensions were centrifuged at 7500 rpm at 20 °C for 15 min and the supernatants were filtered through a 0.2-µm cellulose acetate filter. The Cd was then quantified by ICP-MS, and Mn, Al, and Fe on Tamm's extracts by ICP-AES. All trace metal analyses and soil characterisations were performed by the INRA soil testing laboratory (<http://www6.lille.inra.fr/las>).

The soil pore solution was obtained by centrifugation as described by Schneider (2006). Briefly, soils were moisturized to pF 1.7, using the model from Jong et al. (1983) to calculate the right amount of ultra-pure water to add. After equilibrating overnight, approximately 12 ml of the non-water miscible organic solvent C<sub>2</sub>F<sub>3</sub>Cl<sub>3</sub> (ACS, Reag. Ph. Eur, Merck) was added to displace the soil solution during centrifugation at 13,000 rpm for 1 h. The aqueous supernatants were carefully separated from the solvent and filtered through a 0.2-µm cellulose acetate filter.

These solutions were used to determine the pH using a microelectrode (pHC4000-8, Radiometer Analytical), and to quantify the major cations Ca, Mg, and K by flame-AAS, and Cd by furnace-AAS (Solaar M6, TJA Solutions). The free fraction of Cd (Cd<sup>2+</sup>) in these soil solutions was measured using the ion exchange method developed by Schneider (2006). This method is based on the Ca/Cd exchange properties of a cation exchange resin (Amberlite IR120, Aldrich) converted to Ca form. The soil pore solution (around 2 ml) was mixed with the Ca resin at a 500:1 v/w solution/solid ratio (determined precisely) and shaken at 20 °C for 24 h, after which 0.2 ml of the soil pore solution was sampled to determine the concentration of Ca after contact. Next, a known amount of radioactive <sup>109</sup>Cd (Perkin-Elmer) was added to the resin/solution mixture. The vials were shaken at 20 °C for 24 h, after which the radioactivity remaining in solution was quantified in a 0.5-ml aliquot of solution by liquid scintillation counting (Tri-Carb 2100TR, Packard BioScience), after addition of 2.5 ml of scintillation cocktail (Insta-Gel Plus, Packard BioScience) and correction for quenching. The fraction of free Cd (F<sub>Cd</sub>) was calculated from the resin Vanselow selectivity coefficient for Cd/Ca exchange (<sup>V</sup>K<sub>Cd/Ca</sub>), using Eq. (1)

$$F_{Cd} = \frac{\left(\frac{R}{r} - 1\right) \times C_{Ca} \times V}{{}^V K_{Cd} \times R_{Ca,f} \times m} \times 100 \quad (1)$$

**Table 1** Summary of the extraction methods used as single extractants to evaluate the phytoavailability of Cd for durum wheat

Extracting solution	References	Soil/solution ratio (w/v)	Extraction time (h)
HF/HClO <sub>4</sub> acid extraction (total extraction)	NF X 31–147	–	–
0.05 M EDTA at pH 7	Ure 1996	1:10	1
1 M sodium acetate adjusted to pH 5 with acetic acid	Tessier et al. 1979	1:8	3
0.1 M Na <sub>4</sub> P <sub>2</sub> O <sub>7</sub> (pyrophosphate) at pH 10	Loveland and Digby 1984	1:100	16
0.2 M oxalic acid + 0.2 M ammonium oxalate (Tamm's reagent) at pH 3	McKeague and Day 1966	1:100	4
1 M NH <sub>4</sub> NO <sub>3</sub> at pH 5	Symeonides and McRae 1977	1:2.5	1
0.01 M CaCl <sub>2</sub> at soil pH	Houba et al. 1990	1:10	2
Ultra-pure (UP) water at soil pH	Reported by Lebourg et al. 1996	1:5	1

where  $R/r$  is the ratio of added ( $R$ ) to remaining ( $r$ ) radioactivity,  $C_{Ca}$  is the concentration of calcium in the solution after contact,  $V/m$  is the soil solution/resin ratio, and  $R_{Ca,f}$  is the content of Ca on the resin after contact calculated by Eq. (2).

$$R_{Ca,f} = R_{Ca,i} + (C_{Ca,i} - C_{Ca,eq}) \times \frac{V}{m} \quad (2)$$

Free Cd ( $Cd^{2+}$ ) concentrations were calculated using this fraction of free Cd and the concentration of total Cd in the soil pore solution obtained by centrifugation, with Eq. (3):

$$[Cd^{2+}] = F_{Ca/100} \times [Cd]_{porewater} \quad (3)$$

## DGT experiments

DGT experiments were performed in the laboratory using fresh soils stored at 4 °C. Approximately 700 g of fresh soil was weighed, and ultra-pure water was added to reach approximately 50% of the maximum water holding capacity ( $WHC_{max}$ ). The soils were incubated at 28 °C for 24 h in closed plastic containers. Next, the soils were moistened to 100% of their  $WHC_{max}$  and allowed to equilibrate at 28 °C in the dark for 24 h. Then, three DGT devices were applied to each soil, avoiding any free air space between the filter and the soil surface. The containers were closed and left at 28 °C in the dark for 48 h before collecting the DGTs. The 48-h period of contact between the DGT and the soil was chosen to obtain a quantifiable amount of Cd on the resin while minimising changes in the soil conditions during the contact (Harper et al. 1998). Each resin was individually bathed in 2 ml of 1 M HNO<sub>3</sub> for 24 h to eluate the sorbed Cd. In addition, after the DGTs had been used, the soil solutions from the containers were sampled with Rhizon microcups (Rhizosphere Research Products, Wageningen) inserted horizontally. Cadmium in the soil solutions and in the resin eluates was quantified using furnace-AAS (PinAAcle 900T, PerkinElmer).

An indicator of the Cd buffering power of the soil can be derived from DGT data (Zhang et al. 2001). Indeed, the ratio ( $b$ ) between the calculated Cd concentrations at the interface between the gel and the resin of the DGT ( $C_{DGT}$ ) and the Cd concentrations in the soil solution obtained by the rhizons ( $C_{rhizon}$ , Eq. (4)) reflects the capacity of the solid phase to resupply the soil solution with Cd.

$$b = \frac{C_{DGT}}{C_{rhizon}} \quad (4)$$

If the resupply is fast, and the solid phase contains a large amount of mobile Cd, the concentration at the interface will not diminish significantly and  $b$  will be close to 1. Conversely, if little is resupplied from the solid phase,  $C_{DGT}$  will tend to be 0 because the resin behaves as an infinite sink and  $b$  also tends to be 0 (Harper et al. 1998; Zhang et al. 2001).

## Testing statistical models from the literature

We conducted an extensive search of the literature to find statistical models based on soil properties for the prediction of Cd contents in wheat grains that could be tested on our dataset. The following models were found for bread wheat (Adams et al. 2004; Baize et al. 2009; Brus et al. 2005; François et al. 2009; Gray et al. 2001; Hough et al. 2003; Liu et al. 2015; Mench et al. 1997; Nan et al. 2002; Ran et al. 2016; Zogaj and Düring 2016) and for durum wheat (Garrett et al. 1998; Norvell et al. 2000). We discarded all the models which included soil variables not measured in our study. Finally, four models were retained and are detailed in Table 2. These models include total Cd and pH, and in one model, soil organic matter, all these variables being highly relevant to explain Cd partitioning from soil to solution (Degryse et al. 2009; Sauvé et al. 2000). Unfortunately, the selected models were developed for bread wheat, but we hypothesized that if the models are valid, the effect of the species

**Table 2** Models found in the literature to predict Cd contents in bread wheat grains

Model	Reference	<i>N</i>	$r^2_{adj}$	$\alpha$	RMSE	$r^2$
$\log_{10}(\text{Cd}_{\text{grain}}) = 0.28 + 0.44 \log_{10}(\text{Cd}_{\text{total}}) - 0.18 \text{ pH}$	Adams et al. 2004	162	0.49	1.81	0.033	0.42
$\log_{10}(\text{Cd}_{\text{grain}}) = 1.022 + 0.749 \log_{10}(\text{Cd}_{\text{total}}) - 0.257 \text{ pH} - 0.277 \log_{10}(\text{SOM})$	Brus et al. 2005	84	NA	2	0.041	0.44
$\log_{10}(\text{Cd}_{\text{grain}}) = 1.386 + \log_{10}(\text{Cd}_{\text{total}}) - 0.279 \text{ pH}$	Liu et al. 2015	14	0.85	0.67	0.029	0.45
$\log_{10}(\text{Cd}_{\text{grain}}) = 0.703 + 1.04 \log_{10}(\text{Cd}_{\text{total}}) - 0.175 \text{ pH}$	Ran et al. 2016	99	0.61	0.47	0.038	0.44

*N* number of observations used to parameterize the model,  $r^2_{adj}$  adjusted *r*-squared,  $\alpha$  slope of the regression between observed and predicted values for each model, *RMSE* root mean square error,  $r^2$  *r*-square of the regression between observed and predicted values

(bread wheat versus durum wheat) would possibly affect the slope of the relationship between the predicted and the measured Cd content in grains, but not the strength of the relationship.

### Statistical modelling

The content of Cd in durum wheat grain was modelled by multiple linear regressions with the soil variables as regressors. The models tested included one to a maximum of four regressors to avoid over-parameterisation. The rationale used to select the regressors in a hierarchical way was the following. For the one-regressor models, we selected variables that could reflect the soil available Cd, namely, the Cd in soil derived from the different extractions (Table 1). For the two-regressor models, we crossed the previous models with the soil pH, measured in water, in CaCl<sub>2</sub> extract, or in the soil pore solution. Indeed, the proton competes with Cd for both the complexation and the sorption of Cd and is therefore expected to be among the most influential parameters in soils (Adriano 2001). The three- and four-regressor models crossed the two-regressor models with one or two additional variables, respectively. The additional variables were related to sorption, speciation in solution, and competition for sorption and complexation: SOC, contents of carbonates and amorphous oxides, CEC, texture, DOC, and cations and anions in solution. A total of 3065 models were tested.

The models were individually tested by leave-one-out cross-validation (LOOCV). The models were parameterized on (*N* – 1) observations, and the remaining observation was used to calculate the error of prediction. The process was repeated *N* = 27 times, and the root mean square error of predictions (RMSE<sub>LOOCV</sub>) was then calculated for each model. Models were evaluated using the following criteria: adjusted multiple correlation coefficients ( $r^2_{adj}$ ) calculated from the multiple linear regressions and from the simple linear regression between predicted and observed values, the Bayesian information criterion (BIC) to penalize over-parameterisation of the models, and the RMSE<sub>LOOCV</sub> and the variance inflation factor (VIF) to detect multicollinearity between predictors.

In order to test if the dataset used to parameterize was sufficiently big, a second cross-validation was performed based on the approach of Shao (1993). This procedure involves the random selection of  $n_v$  data to validate the model, while the remaining data (*N*– $n_v$ ) are used to parameterize the model. The size of validation subset ( $n_v$ ) ranged from 1 to a maximum of 21. This left a minimum of six observations for parameterisation of the largest models with four regressors. The case where  $n_v$  = 1 corresponds to the leave-one-out cross-validation. There were up to 20 million (*N*  $n_v$ ) subsets, and a maximum of 10<sup>5</sup> distinct possibilities were consequently randomly chosen and tested. The RMSE was calculated for each validation subset and compared to the RMSE of the corresponding model parameterized on the whole dataset. All statistics and data treatments were conducted using R software version 3.3.2. (R Core Team 2017).

## Results and discussion

### Soil properties and Cd contents in plant materials

All the soils were alkaline with pH values ranging between 8.0 and 8.7 (Table 3 (a)). From field surveys (Arvalis, pers. comm.), this range was estimated to cover the 60% upper range of the French agricultural soils used to cultivate durum wheat. For durum wheat cropped in France, soil pH generally ranges from 6 to 8.7, meaning that 40% of the cases where pH is between 6 and 8.0 were not tested in our models, so our results should be extrapolated with caution to these situations.

The total Cd contents of the soils we studied ranged from 0.14 to 1.00 mg kg<sup>-1</sup> (Table 3 (b)), which we consider covered about 80% of the soils under durum wheat, for which the Cd content ranged between 0.09 and 1.44 mg kg<sup>-1</sup> soil according to the above-mentioned field surveys (Arvalis, pers. comm.).

The other soil characteristics, including texture, CEC, organic carbon, and calcareous contents, were typical of French agricultural soils and displayed the necessary variability so that these characteristics could be used as explanatory variables (Table 3 (a)).

**Table 3** Summary of the characteristics of the soils used in this study (a), of the Cd pools estimated by a single extraction and measured in soil solutions (b), and of Cd measured in different wheat organs, roots, shoots and grains (c)

(a) Soil characteristics														
Value	Soil pH	pH CaCl <sub>2</sub>	pH pore water	Granulometry			CEC cmol <sup>+</sup> /kg	SOC g/kg	CaCO <sub>3</sub> g/kg	DOC mg/l	Ca pore water mmol/l	Tamm extractions		
				Clay %	Silt	Sand						Al g/kg	Fe g/kg	Mn mg/kg
Min.	8	6.8	8.2	18.6	15.2	11	10.7	9.21	27.1	33.9	2.02	0.29	0.54	0.11
Median	8.4	7.3	8.4	30.1	35.6	28.7	18.4	15.7	208.6	49.6	3.11	1.15	1.10	0.30
Max.	8.7	7.5	8.5	53	54	59.4	33.5	40.3	426	70.4	10.88	2.81	1.79	0.69
(b) Soil Cd content according to the extraction														
Value	Cd HF μg/kg	Cd EDTA μg/kg	Cd sodium acetate μg/kg	Cd pyro-phosphate μg/kg	Cd Tamm μg/kg	Cd NH <sub>4</sub> NO <sub>3</sub> μg/kg	Cd CaCl <sub>2</sub> μg/kg	Cd UP water μg/kg	Cd DGT μg/kg	Cd pore water μg/l	Cd rhizon μg/l	Cd <sup>2+</sup> μg/l		
Min.	135	63.7	70.05	10.18	0.19	0.3	0.03	0.01	0.003	0.032	0.014	0.014		
Median	318	146	129.4	51.29	22.69	0.96	0.24	0.05	0.009	0.101	0.041	0.040		
Max.	999	355	259.8	197.2	84.49	2.3	0.89	0.14	0.065	0.327	0.184	0.148		
(c) Cd contents in different wheat organs														
Value	Shoot Cd mg/kg DW	Grain Cd mg/kg <sup>a</sup>												
Min.	0.125	0.009												
Median	0.329	0.063												
Max.	0.652	0.144												

<sup>a</sup> Grain Cd contents are expressed at 15% water content

Plant Cd contents ranged from 0.125 to 0.652 mg kg<sup>-1</sup> DW (dry weight) in the shoots and from 0.009 to 0.144 mg kg<sup>-1</sup> in the grains at 15% water content (Table 3 (c)). These values are within the range of Cd contents previously reported in durum wheat grains (Adams et al. 2004; Li et al. 2011) and are below the maximum admissible content (0.2 mg kg<sup>-1</sup>) set by the European Commission (European Union 2006).

### Is it possible to predict Cd content in durum wheat grains at maturity from shoot Cd content at the early heading stage?

Early diagnosis of the contamination of the grain by Cd would be possible if the Cd content of wheat shoots at the heading stage is correlated with that of grains at maturity. This correlation was positive and significant, but too weak to be used as a valid model of risk assessment ( $r^2 = 0.25$ ,  $p$  value <0.05). From the plant physiological point of view, the weak correlation between the Cd in the shoot at heading and the Cd content of the grain at maturity could mean that the Cd accumulated in the grains does not always originate from remobilisation from shoots but could also be the result of direct uptake during post-heading growth. The balance between remobilisation and uptake for the allocation of Cd to the grain could mirror that of Zn, due to the fact that these two elements could behave similarly in plants (Khan et al. 2014). Remobilisation of Zn from vegetative tissues to the grain is favoured when the availability of Zn for root uptake is low (Sperotto 2013). Therefore, assuming that Cd and Zn behave similarly, we introduced the indicator of the buffer power for Cd  $b$  in the regression model between the grain Cd at maturity and the Cd in the shoot at heading, and the goodness of fit increased greatly ( $r^2 = 0.78$ ,  $p$  value <0.001,  $N = 26$ ).

$$\text{Cd}_{\text{grain}} = -0.047^{**} + 0.242^{***} \text{Cd}_{\text{shoots}} + 0.230^{***} b - 0.405^{**} \text{Cd}_{\text{shoots}} \times b \quad (5)$$

The sign of the coefficients of Eq. (5) are consistent with the above-mentioned hypotheses. The coefficients of Cd<sub>shoots</sub> and of  $b$  are positive because they could be a source of Cd from remobilisation and uptake, respectively. The negative coefficient of the interaction between Cd<sub>shoots</sub> and  $b$  could mean that the remobilisation of Cd from shoots to grains is greater in soils with low buffering power for Cd (low  $b$  values). Conversely, when soils have high buffering power for Cd, grain Cd might mainly result from direct uptake and remobilisation could be minimum.

### Soil Cd extractions

In unpolluted agricultural soils with low Cd contents, the availability of Cd and not the root uptake capacity generally

limits accumulation of the metal by the plant (Lin et al. 2016). Consequently, different extractions were tested to identify the one that best reflects the amounts of Cd potentially available to durum wheat: the amounts of Cd recovered in the soil pore solution, by DGT and Rhizon samplers and by different chemical extractants applied separately and not sequentially (Table 3 (b)).

As expected, the least Cd was extracted by ultra-pure (UP) water, from 0.005 to 0.04% of the total Cd (HF/HClO<sub>4</sub> extraction), since water only extracts Cd weakly bound to the soil matrix. The 0.01 M CaCl<sub>2</sub> solution extracted three times more Cd than the ultra-pure water because this salt also extracted Cd sorbed onto the soil matrix by ion exchange due to the substitution of Cd by Ca (Houba et al. 1990). NH<sub>4</sub>NO<sub>3</sub> extracted on average 4.1 times more Cd than CaCl<sub>2</sub>. The chemical mechanisms reported to explain the action of 1 M NH<sub>4</sub>NO<sub>3</sub> are a moderate drop in pH values (around pH 5) and the formation of metal amine complexes, which is favoured in alkaline soils due to the dissociation of NH<sub>4</sub><sup>+</sup> (Gryshko et al. 2005). The high concentration of NH<sub>4</sub><sup>+</sup> strongly competes with cations for sorption (Barrow 1986) and consequently favours Cd desorption. NH<sub>4</sub>NO<sub>3</sub> extraction also estimates the content of exchangeable Cd but extracted more exchangeable Cd due to the higher molarity of the competitive cation NH<sub>4</sub><sup>+</sup> (Table 1). Tamm's reagent was chosen for its ability to dissolve amorphous metallic oxides and to release the Cd sorbed on these constituents into solution. This extractant retrieved from 0.08 to 21.3% of the total Cd, much less than the Cd extracted by the sodium pyrophosphate solution, which ranged from 2.8 to 47.4% of the total Cd. In addition to the Cd weakly sorbed onto the soil, sodium pyrophosphate enables the extraction of Cd complexed with organic matter and can slightly dissolve amorphous metallic oxides. This shows that in our cases, complexation of Cd by the solid organic matrix was likely to dominate. This is in agreement with the literature on the modelling of the soil solution partitioning of Cd which clearly shows that soil organic carbon is a much more frequent significant regressor than oxides (De Vries et al. 2011; Degryse et al. 2009; Groenenberg et al. 2010; Sauvé et al. 2000).

Sodium acetate/acetic acid at pH 4.5 released more Cd, from 25 to 55% of the total Cd, than the previous extractants. This can be explained by the fact that, in addition to the weakly sorbed Cd, this extractant dissolved Cd carbonates, which were likely an important fraction because of the high pH and calcareous contents of the soils. Finally, the 0.05 M EDTA solution extracted the most Cd, i.e., from 35 to 64% of the total Cd. EDTA is a strong complexing ligand anion which can chelate the Cd complexed with organic matter and with metallic oxides (Ure 1996), like sodium pyrophosphate but with a greater extraction strength.

The soil pore water solutions collected by centrifugation contained higher concentrations of Cd than the solutions

collected by the Rhizons, 118 versus 56.5 ng Cd l<sup>-1</sup>, respectively. One possible explanation for this difference could be the length of the period of equilibration after rewetting the soil: overnight for the centrifugation method compared to 4 days for the Rhizon method. Rhizons extract water poorly held by the soil constituents, mainly the gravitational water of the macro porosity, whereas centrifugation also extracts the capillary water. Due to the buffering capacity of the solid phase and due to soil tortuosity, 4 days after rewetting may not have been long enough to reach equilibrium between the Cd concentration in the capillary water and in the water in the macroporosity. Furthermore, rewetting a soil generally results in a transient burst of microbial activity accompanied by the release of soluble organic matter (Merckx et al. 2001), which can complex and therefore desorb part of the Cd sorbed onto the solid phase. The soils used for the extraction by centrifugation were incubated wet for a shorter time than the soils used for Rhizon extraction. Consequently, the former might have more soluble organic ligands that may have desorbed more Cd (Cornu et al. 2007).

The free fractions of Cd measured in soil solutions obtained by centrifugation ranged from 18 to 78% with a median of 44% corresponding to Cd<sup>2+</sup> concentrations ranging from 14.53 to 147.50 ng l<sup>-1</sup>. The high values obtained for the free fractions were not expected because of the alkaline pH values. They might be explained by the high concentrations of competitive cations including Ca compared to the concentration of ligands.

#### Testing models from the literature for the prediction of grain Cd in durum wheat

The models from the literature tested on our dataset (Adams et al. 2004; Brus et al. 2005; Liu et al. 2015; Ran et al. 2016) explained less than half the variability of the observed grain Cd contents, and underestimation by the models was not systematic, as we had expected from the fact that they were parameterized for bread wheat (Fig. 1a–d). This shows that the soil total Cd, pH, and soil organic matter can enable a rough estimate of how Cd in grain is expected to change when these variables vary, but that the parameterisation in the literature was not robust. Several explanations are possible. First, the range of data used for their calibration may not cover the soils tested in our work. Indeed, the model proposed by Brus et al. (2005) was parameterized on soils with SOM contents ranging from 1.4 to 42 g kg<sup>-1</sup> and tested on soils with SOM contents ranging 15.9 to 69.8 g kg<sup>-1</sup>. The accumulation of Cd in grain also depends on the plant genotype, which is not explicitly taken into account in the models in the literature, as they were parameterized for bread wheat, whereas we tested them on durum wheat. Furthermore, within the same subspecies (bread wheat or durum wheat), differences in accumulation between cultivars can reach a ×2 to ×3 factor (Perrier et al. 2016;

Zimmerl et al. 2014). Finally, in models only based on total Cd, pH, and soil organic carbon, the speciation and the buffer capacity of Cd in the soil are poorly taken into account. Total Cd is unlikely to always be fully mobile. Only pH and soil organic carbon could be partly linked with the speciation of Cd in solution, but these variables also reflect sorption and desorption.

#### Statistical models to predict Cd in durum wheat grain at maturity

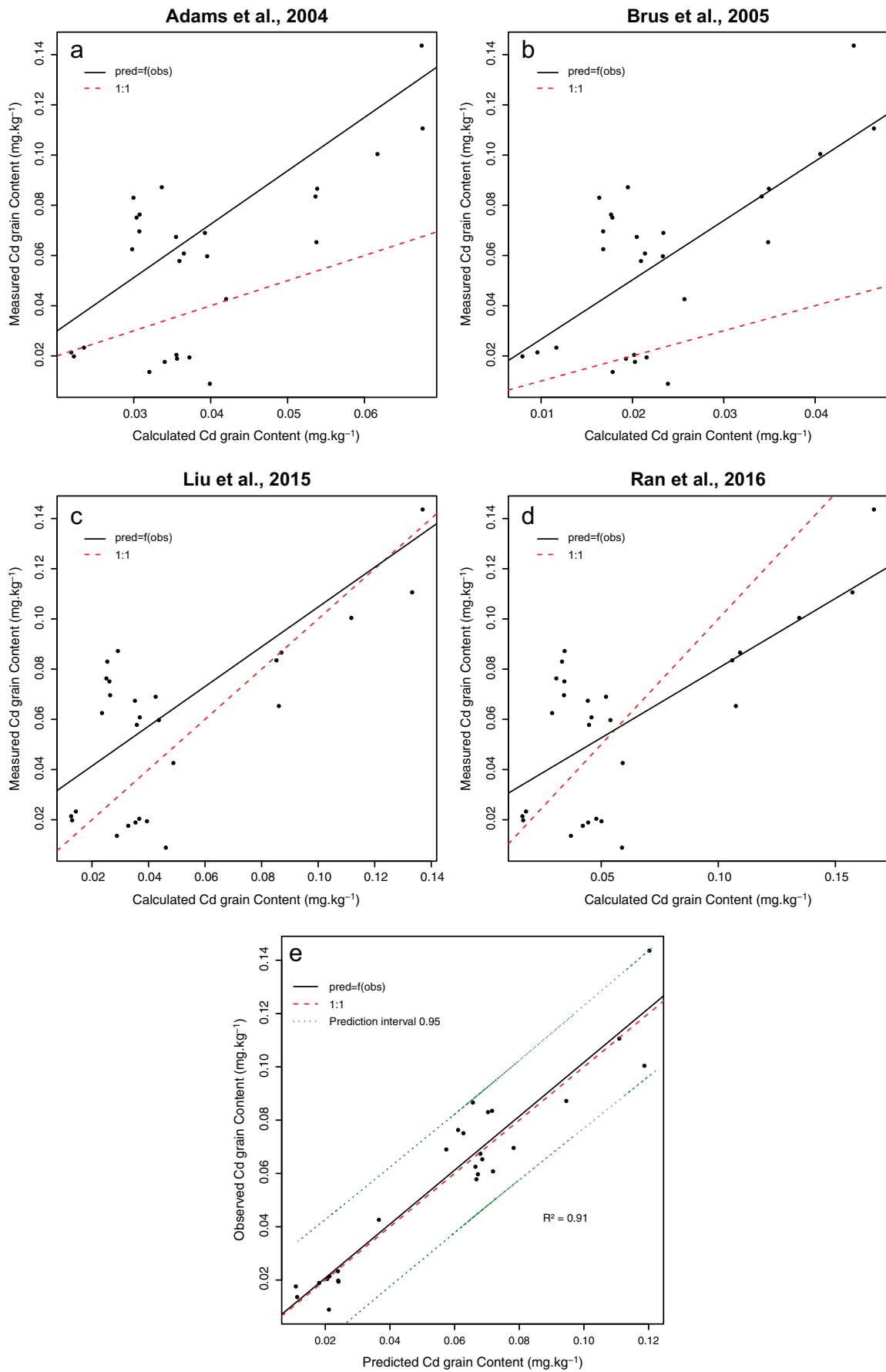
Among the one-variable models, those with HF/HClO<sub>4</sub>, EDTA, sodium acetate, and Tamm extracted Cd showed poor goodness of fit with *r*-squared values around 0.15 (Table 4). The Cd recovered in the soil pore water by centrifugation or by the Rhizon samplers and even the concentration of the free ion Cd<sup>2+</sup> did not produce good models (Table 4). Intermediate goodness of fit was obtained for the one-variable models based on the Cd recovered by the DGT (*r*<sup>2</sup> = 0.40, *p* value <0.001) or extracted by water, 0.01 m CaCl<sub>2</sub>, or pyrophosphate solutions (Table 4). The 1 M NH<sub>4</sub>NO<sub>3</sub> solution was the best extractant, reflecting Cd bioavailability for durum wheat in the soils studied here (*r*<sup>2</sup> = 0.62, *p* value <0.001). The effectiveness of 1 M NH<sub>4</sub>NO<sub>3</sub> extraction in correlating with Cd accumulation in the plant has also been reported for wheat, carrot, spinach, and lettuce (Gryschko et al. 2005). As a whole, even if a particular one-variable model can give a satisfactory goodness of fit, the robustness of these simple models is—or can be expected to be—low because of the complexity of the mechanisms governing the phytoavailability of Cd in soils.

The best two-variable model included Cd from the DGT and the CaCl<sub>2</sub> pH (*r*<sup>2</sup> = 0.66, Eq. (6), Table 5), which showed that adding the CaCl<sub>2</sub> pH to the Cd from the DGT notably improved the predictive value of the model. One possible explanation could be that, in alkaline soil, roots can reduce the pH in the rhizosphere to increase the availability of nutrients (Hinsinger et al. 2003) which is not taken into account if only the Cd DGT is used in the model.

When the number of regressors was increased to three, the *r*<sup>2</sup> of the best model was 0.79 (Eq. (7), Table 5).

The model with the best predictive value included the total Cd in the soil solution obtained by centrifugation, the pH measured in the CaCl<sub>2</sub> extract, the CEC, and the amount of Mn extracted by Tamm's reagent (Eq. (8), Table 5). This

**Fig. 1** Comparison of grain Cd contents in this study and the corresponding predictions made by the model of Adams et al. (2004) (a), Brus et al. (2005) (b), Liu et al. (2015) (c), and Ran et al. (2016) (d) and the best model from this study (Eq. (8)) (e). The black line represents the regression between observed and predicted values; the red dashed line is the 1:1 line; and the green dotted line represents the prediction interval at 95%. All grain Cd data are expressed at 15% water content



**Table 4** Adjusted  $r^2$ -square for the linear regressions between the grain Cd (log) contents and the Cd contents (log) recovered using different extraction methods

Extraction method	$r^2_{adj}$
HF/HClO <sub>3</sub> acids	0.176*
0.05 M EDTA	0.153*
Sodium acetate at pH 5	0.133*
Pyrophosphate at pH 10	0.579***
Tamm's reagent	0.164*
1 M NH <sub>4</sub> NO <sub>3</sub> at pH 5	0.624***
0.01 M CaCl <sub>2</sub>	0.543***
Ultra-pure water	0.490***
DGT	0.397***
Pore water by centrifugation	0.467***
Free Cd (Cd <sup>2+</sup> )	0.493***

\* $p$  value <0.05; \*\* $p$  value <0.01; \*\*\* $p$  value <0.001 (statistical significance)

model explained 88% of the total variance of the grain Cd contents. The variance inflation factor (VIF) did not identify strong correlations between the regressors following the rule of thumb that suggests examining collinearity for values above 2.

Figure 1e shows the regression between observed grain Cd contents and predictions made by Eq. (8). The slope is close to 1 (1.012) with a non-significant intercept ( $p$  value = 0.9), and 91% of the variance was explained, showing that the model was accurate and not biased. Equation (8) had a RMSE of 0.013, which is quite low and accurate for the prediction of grain Cd contents whose values ranged from 0.09 to 0.144 mg kg<sup>-1</sup>.

Figure 2 presents the RMSE values for each different size of the subset tested for validation ( $n_v$ ). Equation (8), when parameterized with  $N = 27$  soils, produced a low value of RMSE (Table 5), represented by the red line in Fig. 2. When the size of the parameterizing subset decreased ( $n_v$  increases), the prediction errors of the models increased. However, no excessive increase in the RMSE was observed with a validation subset of less than 13 soils (Fig. 2). This suggests that when the model is parameterized with 14 soils and tested on

the remaining 13, the model uncertainty remains acceptable. Consequently, 26 soils were considered to be enough to select variables to produce a robust model.

The contributions of each regressor of Eq. (8) to the total sum of squares were 38% for the concentrations of Cd in the pore water, 5.7% for pH CaCl<sub>2</sub>, 43.8% for the CEC, and 12.5% for the Mn Tamm's contents. It is worth noting that the two main contributors are the initial availability of Cd in the soil solution (pore water Cd) and the CEC, which can be interpreted as the buffering capacity of the solid phases by ion exchange. Therefore, the importance of the CEC in the model is in agreement with the fact that, among the extractions tested, those that gave the best correlations with grain Cd were those that used low-strength extractants. The importance of the CEC in the model is also consistent with the mechanistic modelling of Cd phytoavailability, which showed that desorption is generally the main mechanism buffering Cd in the soil solution compared to dissociation of the complex, the latter being limited by kinetics (Lin et al. 2016). Therefore, the two main explanatory variables of the model with their positive coefficients could illustrate the fact that plant roots take up Cd from the soil solution, which is in part replenished by the solid phase. However, it is generally assumed that only the free Cd ion is taken up by plants (Campbell 1995); this would mean that the total Cd in solution and the CaCl<sub>2</sub> pH in Eq. (8) can be interpreted as variables representing the free Cd<sup>2+</sup> concentration in the soil solution because Cd speciation is strongly governed by pH. Following this rationale, we examined a model derived from Eq. (8) by replacing the total Cd in solution and the CaCl<sub>2</sub> pH by the free Cd concentrations (Eq. (7)). The performances of the resulting model were significantly lower ( $F_{2,025} = 18.05$ ,  $p$  value <0.001, for the model comparison) with an  $r^2$ -square that decreased from 0.88 to 0.79. This could mean that the total Cd in solution and the pH did not only represent the free Cd in solution. The dissociation of Cd complexes might have helped to buffer Cd<sup>2+</sup> availability during uptake of the metal by roots. As indicated above, the contribution of the complexes to the buffering of Cd<sup>2+</sup> in the solution is likely to have been much less than buffering by the

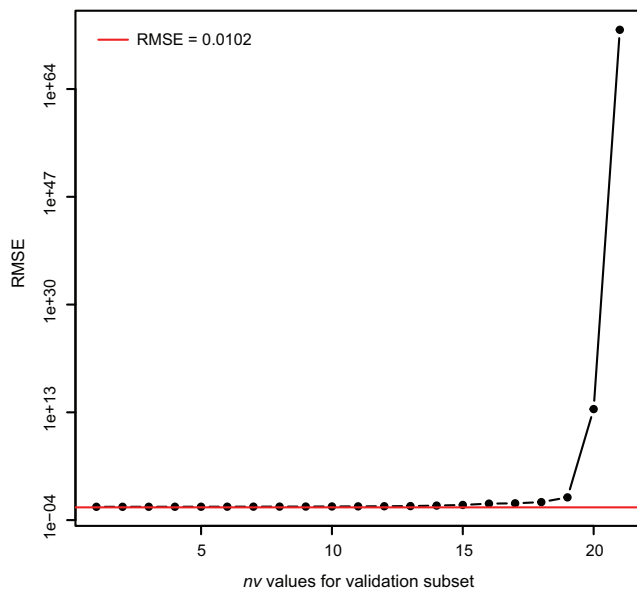
**Table 5** Best models to predict Cd contents in durum wheat grains and associated statistics

Equation	Model	$r^2_{adj}$	RMSE	RMSE <sub>LOOCV</sub>	VIF	$N$
[6]	$\log_{10}(\text{Cd}_{\text{grain}}) = 7.359 + 0.697^{***}$ $\log_{10}(\text{Cd}_{\text{DGT}}) - 1.014^{***}$ pH CaCl <sub>2</sub>	0.66	0.0219	0.0254	2.2	26
[7]	$\log_{10}(\text{Cd}_{\text{grain}}) = -3.584^{***} + 0.339^*$ $\log_{10}(\text{Cd}^{2+}) + 1.722^{***}$ $\log_{10}(\text{CEC}) - 0.864^{***}$ $\log_{10}(\text{Mn}_{\text{Tamm}})$	0.79	0.0184	0.022	3.0	27
[8]	$\log_{10}(\text{Cd}_{\text{grain}}) = -1.979 + 0.553^{***}$ $\log_{10}$ $(\text{Cd}_{\text{pore water}}) - 0.347^*$ pH CaCl <sub>2</sub> + 1.630 <sup>***</sup> $\log_{10}(\text{CEC}) - 0.685^{***}$ $\log_{10}(\text{Mn}_{\text{Tamm}})$	0.88	0.0102	0.0125	2.2	27

$r^2_{adj}$  adjusted  $r^2$ -square, RMSE root mean square error, RMSE<sub>LOOCV</sub> root mean square error from the leave-one-out cross validation, VIF variance inflation factor,  $N$  number of observations

\* $p$  value <0.05; \*\* $p$  value <0.01; \*\*\* $p$  value <0.001 (statistical significance)





**Fig. 2** RMSE of predictions from the best model (Eq. (8)) for different sizes ( $n_v$ ) of validation subset from a Monte Carlo cross-validation. The red line corresponds to the value of RMSE of the model parametrized with 27 soils ( $n_v = 0$ )

CEC since the drop in the  $r$ -square between Eqs. (7) and (8) was only 0.09, whereas removing the CEC from Eq. (8) results in a drop of 0.32 ( $r^2 = 0.56$ , RMSE = 0.033).

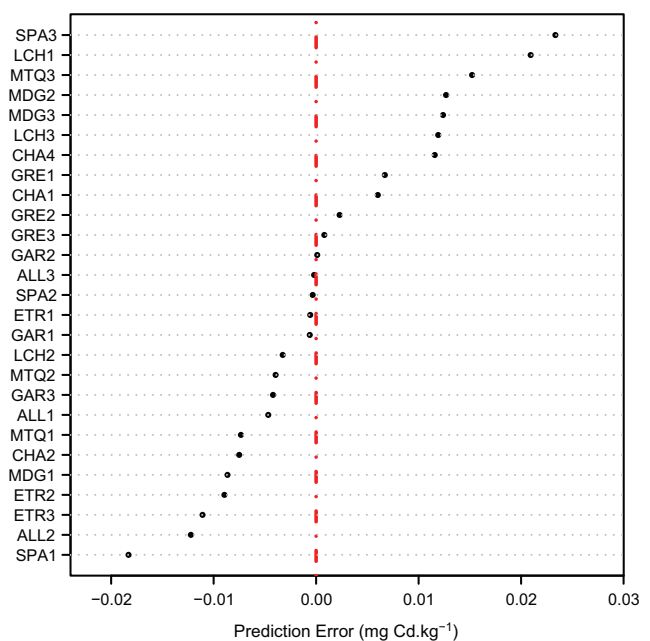
There was no improvement in the overall goodness of fit of Eq. (8) when the DOC was added to the regression ( $r^2 = 0.87$ , RMSE = 0.01). This could mean that due to the relatively high buffering capacity of the solid phase (CEC) and therefore the limited contribution of Cd complexes to Cd uptake, the concentration of organic Cd complexes associated with the DOC was not limiting. A second possible hypothesis is that the contributing complexes were not organic but inorganic, including those with  $\text{HCO}_3^-$  (McLaughlin et al. 1999), maybe because the latter were more labile.

In Eq. (8), when the pH measured in 0.01 M  $\text{CaCl}_2$  was replaced by the pH of  $\text{H}_2\text{O}$  or by the pH of the soil pore water extracted by centrifugation,  $r^2$  dropped from 0.88 to 0.84 with RMSE equal to 0.015 for both. The slightly better predictive value of the  $\text{CaCl}_2$  pH can be explained as follows: the  $\text{CaCl}_2$  pH reflects the proton buffering power of the soil because it gives the amount of protons desorbed due to Ca competition (Miller and Kissel 2010) and is therefore lower than the pH of  $\text{H}_2\text{O}$  or the pH of the soil pore water obtained by centrifugation. The pH which influences Cd uptake by roots is the pH of the rhizosphere, which is affected by excretion of organic acids or protons by the roots resulting from the differential rates of cation and anion uptake and/or from micronutrient mobilisation, particularly in alkaline soils (Nye 1981), like in our study. The pH of the rhizosphere can also change due to fluctuating environmental conditions, including drying of the soil during the grain filling period which takes place in late spring and early summer (June–July). Compared to the pH of

$\text{H}_2\text{O}$ , or to the pH of the soil pore water, because it takes into account the proton buffering capacity of the soil, the pH  $\text{CaCl}_2$  is probably a better proxy of the mean rhizosphere pH. Finally, regarding pH, it should be noted that the range of variation in our study was narrow (8.0–8.7) and predictions should be used with caution when the soils are more acidic.

Manganese extracted from amorphous materials by Tamm's reagent explained more than 12% of the total variability in the grain Cd ( $p$  value <0.001). Its negative coefficient shows that Mn oxides reduce Cd availability for root uptake (Wenzel et al. 1996). Mn oxides are one of the most reactive sorbing phases for Cd (Dong 2000) due to their large specific surface areas and their low zero-point charge pH, the pH value at which the oxides are mainly negatively charged and not associated with cations (Feng et al. 2007). Mn oxides have already been shown to be an efficient stabilising amendment for Cd, Pb, and Zn in contaminated soils by significantly reducing metal mobility and uptake by plants (Sappin-Didier et al. 1997).

With absolute errors of around  $0.02 \text{ mg kg}^{-1}$  (Fig. 3), the best model (Eq. (8)) was able to discriminate grain Cd contents of  $0.15 \text{ mg kg}^{-1}$  from those of  $0.2 \text{ mg kg}^{-1}$ . Therefore, the model could be useful for the project of the European Union to lower the regulatory limit of durum wheat grain from 0.2 to  $0.15 \text{ mg Cd kg}^{-1}$  grain (DG SANCO 2011). However, the model is not sufficiently accurate to distinguish the intermediate step of  $0.175 \text{ mg kg}^{-1}$  between 0.2 and  $0.15 \text{ mg Cd kg}^{-1}$ , which underlines the need to produce more mechanistic models to help set realistic regulatory limits.



**Fig. 3** Residual errors (observed minus predicted values) of Eq. (8) for each soil in the dataset. All grain Cd data are expressed at 15% water content

## Conclusions

Predictions of Cd contents in durum wheat grains were obtained from a regression model including the concentration of the Cd in the soil solution obtained by centrifugation, the soil CaCl<sub>2</sub> pH, CEC, and Tamm's Mn oxide content. The predictive model, parametrized on 27 agricultural soils, explains 88% of the variation in Cd in durum wheat grains and has a reasonably low RMSE value of around 0.01 mg kg<sup>-1</sup>. The regressor variables of the model suggest that Cd availability in the alkaline soils tested here was controlled by the initial concentration of Cd in the soil solution and by CaCl<sub>2</sub> pH and by the CEC, which can be interpreted as the buffering capacity of the solid phase. Amorphous manganese content appears to be the main soil phase restraining Cd availability.

This model will be useful to detect cases where grain Cd exceeds the planned regulatory limit for durum wheat of 0.15 mg Cd kg<sup>-1</sup> compared to the current limit of 0.20 mg Cd kg<sup>-1</sup>. However, the model is not sufficiently accurate for the intermediate step of 0.175 mg Cd kg<sup>-1</sup> and the limited predictive accuracy of this model as well as the limited accuracy of the models taken from the literature should be taken into account when revising the regulatory limits. Nevertheless, this simple predictive model will be useful not only for farmers to assess the risk that the Cd contents in their durum wheat grain exceed the regulatory limit but also for food safety authorities to test different regulatory thresholds to identify a trade-off between food safety and the negative impact on farmers of a lower limit.

**Acknowledgements** The authors are grateful to Sylvie Bussi re and C cile Coriou for their technical assistance and want to thank Dr. Jean-Yves Cornu for useful discussions and his comments. This work was supported by funding from *Arvalis-Institut du v g tal* (Cadur project) and from the French National Institute for Agricultural Research.

## References

- Adams ML, Zhao FJ, McGrath SP, Nicholson FA, Chambers BJ (2004) Predicting cadmium concentrations in wheat and barley grain using soil properties. *J Environ Qual* 33:532–541. doi:10.2134/jeq2004.0532
- Adriano DC (2001) Trace elements in terrestrial environments, environmental chemistry. Springer, New York. doi:10.1007/978-0-387-21510-5
- Baize D, Bellanger L, Tomassone R, Denis B, Lise B, Richard T (2009) Relationships between concentrations of trace metals in wheat grains and soil. *Agron Sustain Dev* 29:297–312. doi:10.1051/agro:2008057
- Barrow NJ (1986) Reaction of anions and cations with variable-charge soils. *Adv Agron* 38:183–230. doi:10.1016/S0065-2113(08)60676-8
- Brus DJ, De Gruijter JJ, R mkens PFAM (2005) Probabilistic quality standards for heavy metals in soil derived from quality standards in crops. *Geoderma* 128:301–311. doi:10.1016/j.geoderma.2005.04.014
- Campbell PGC, (1995) Interactions between trace metals and aquatic organisms: a critique of the free-ion activity model In: Tessier A, Turner DR (Eds.) Metal speciation and bioavailability in aquatic systems. Wiley, New Jersey, pp. 45–102
- Cho YA, Kim J, Woo HD, Kang M (2013) Dietary cadmium intake and the risk of cancer: a meta-analysis. *PLoS One* 8:1–8. doi:10.1371/journal.pone.0075087
- Cornu JY, Denaix L, Schneider A, Pellerin S (2007) Temporal evolution of redox processes and free Cd dynamics in a metal-contaminated soil after rewetting. *Chemosphere* 70:306–314. doi:10.1016/j.chemosphere.2007.06.003
- DG SANCO (2011) Summary record of the standing committee on the food chain and animal health held in Bruxelles on 23 november 2011. SANCO – Ares 2011:1436040
- De Vries W, McLaughlin MJ, Groenenberg JE (2011) Transfer functions for solid-solution partitioning of cadmium for Australian soils. *Environ Pollut* 159:3583–3594. doi:10.1016/j.envpol.2011.08.006
- Degryse F, Smolders E, Parker DR (2009) Partitioning of metals (Cd, Co, Cu, Ni, Pb, Zn) in soils: concepts, methodologies, prediction and applications—a review. *Eur J Soil Sci* 60:590–612. doi:10.1111/j.1365-2389.2009.01142.x
- Del Castillo P, Chardon WJ (1995) Uptake of soil cadmium by three field crops and its prediction by a pH-dependent Freundlich sorption model. *Plant Soil* 171:263–266. doi:10.1007/BF00010280
- Dong D (2000) Adsorption of Pb and Cd onto metal oxides and organic material in natural surface coatings as determined by selective extractions: new evidence for the importance of Mn and Fe oxides. *Water Res* 34:427–436. doi:10.1016/S0043-1354(99)00185-2
- European Union (2006) Commission regulation (EC) no. 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foodstuffs. *Off J Eur Union L* 364:5–24
- EFSA (2009) Cadmium in food—scientific opinion of the Panel on Contaminants in the Food Chain. *EFSA J* 7:980. doi:10.2903/j.efsa.2009.980
- FAO (2015). *FAO Statistical Pocketbook 2015*. ISBN: 978-92-5-108802-9
- Feng XH, Zhai LM, Tan WF, Liu F, He JZ (2007) Adsorption and redox reactions of heavy metals on synthesized Mn oxide minerals. *Environ Pollut* 147:366–373. doi:10.1016/j.envpol.2006.05.028
- Fran ois M, Grant C, Lambert R, Sauv  S (2009) Prediction of cadmium and zinc concentration in wheat grain from soils affected by the application of phosphate fertilizers varying in Cd concentration. *Nutr Cycl Agroecosystems* 83:125–133. doi:10.1007/s10705-008-9204-0
- Garrett RG, MacLaurin AI, Gawalko EJ, Tkachuk R, Hall GEM (1998) A prediction model for estimating the cadmium content of durum wheat from soil chemistry. *J Geochemical Explor* 64:101–110. doi:10.1016/S0375-6742(98)00023-5
- Gray CW, McLaren RG, Roberts AHC (2001) Cadmium concentrations in some New Zealand wheat grain. *New Zeal J Crop Hortic Sci* 29: 125–136. doi:10.1080/01140671.2001.9514170
- Groenenberg JE, R mkens PFAM, Comans RNJ, Luster J, Pampura T, Shotbolt L, Tipping E, De Vries W (2010) Transfer functions for solid-solution partitioning of cadmium, copper, nickel, lead and zinc in soils: derivation of relationships for free metal ion activities and validation with independent data. *Eur J Soil Sci* 61:58–73. doi:10.1111/j.1365-2389.2009.01201.x
- Gryschko R, Kuhnle R, Terytze K, Breuer J, Stahr K (2005) Soil extraction of readily soluble heavy metals and as with 1 M NH<sub>4</sub>NO<sub>3</sub>-solution—evaluation of DIN 19730. *J. Soils Sediments* 5:101–106. doi:10.1065/jss2004.10.119
- Harmsen J, Rulkens W, Eijssackers H (2005) Bioavailability: concept for understanding or tool for predicting? *L Contam Reclam* 13:161–171
- Harper MP, Davison W, Zhang H, Tych W (1998) Kinetics of metal exchange between solids and solutions in sediments and soils interpreted from DGT measured fluxes. *Geochim Cosmochim Acta* 62:2757–2770. doi:10.1016/S0016-7037(98)00186-0

- Hinsinger P, Plassard C, Tang C, Jaillard B (2003) Origins of root-mediated pH changes in the rhizosphere and their responses to environmental constraints: a review. *Plant Soil* 248:43–59. doi:10.1023/A:1022371130939
- Houba VJG, Novozamsky I, Lexmond TM, van der Lee JJ (1990) As a single extraction solution for the assessment of the nutrient status of soils and other diagnostic purposes. *Commun Soil Sci Plant Anal* 21:2281–2290. doi:10.1080/00103629009368380
- Hough RL, Young SD, Crout NMJ (2003) Modelling of Cd, Cu, Ni, Pb and Zn uptake, by winter wheat and forage maize, from a sewage disposal farm. *Soil Use Manag* 19:19–27. doi:10.1111/j.1475-2743.2003.tb00275.x
- Hutton M (1983) Sources of cadmium in the environment. *Ecotoxicol Environ Saf* 7:9–24. doi:10.1016/0147-6513(83)90044-1
- IARC (1993) Cadmium and cadmium compounds. IARC Monogr Eval Carcinog Risks to Humans 58:119–237. doi:10.1002/14356007.a04
- Jancic SA, Stosic BZ (2014) Cadmium effects on the thyroid gland, 1st ed. *Vitamins and Hormones* Elsevier Inc. doi:10.1016/B978-0-12-800095-3.00014-6
- Jong RD, Campbell CA, Nicholaichuk W (1983) Water retention equations and their relationship to soil organic matter and particle size distribution for disturbed samples. *Can J Soil Sci* 63:291–302. doi:10.4141/cjss83-029
- Kabata-Pendias A (2004) Soil-plant transfer of trace elements—an environmental issue. *Geoderma* 122:143–149. doi:10.1016/j.geoderma.2004.01.004
- Khan MA, Castro-Guerrero N, Mendoza-Cozatl DG (2014) Moving toward a precise nutrition: preferential loading of seeds with essential nutrients over non-essential toxic elements. *Front Plant Sci* 5:1–7. doi:10.3389/fpls.2014.00051
- Lebourg A, Sterckeman T, Ciesielski H, Proix N (1996) Intérêt de différents réactifs d'extraction chimique pour l'évaluation de la biodisponibilité des métaux en traces du sol pour l'évaluation de la biodisponibilité des métaux. *Agronomie* 16:201–215
- Li X, Ziadi N, Bélanger G, Cai Z, Xu H (2011) Cadmium accumulation in wheat grain as affected by mineral N fertilizer and soil characteristics. *Can J Soil Sci* 91:521–531. doi:10.4141/cjss10061
- Li YM, Chaney RL, Schneider AA, Miller JF, Elias EM, Hammond JJ (1997) Screening for low grain cadmium phenotypes in sunflower, durum wheat and flax. *Euphytica* 94:23–30. doi:10.1023/A:1002996405463
- Lin Z, Schneider A, Sterckeman T, Nguyen C (2016) Ranking of mechanisms governing the phytoavailability of cadmium in agricultural soils using a mechanistic model. *Plant Soil* 399:89–107. doi:10.1007/s11104-015-2663-6
- Liu K, Lv J, He W, Zhang H, Cao Y, Dai Y (2015) Major factors influencing cadmium uptake from the soil into wheat plants. *Ecotoxicol Environ Saf* 113:207–213. doi:10.1016/j.ecoenv.2014.12.005
- Loveland PJ, Digby P (1984) The extraction of Fe and Al by 0.1 M pyrophosphate solutions: a comparison of some techniques. *J Soil Sci* 35:243–250. doi:10.1111/j.1365-2389.1984.tb00280.x
- McKeague JA, Day JH (1966) Dithionite- and oxalate- extractable Fe and Al as aids in differentiating various classes of soils. *Can J Soil Sci* 46:13–22. doi:10.4141/cjss66-003
- McLaughlin MJ, Parker DR, Clarke JM (1999) Metals and micronutrients—food safety issues. *F Crop Res* 60:143–163. doi:10.1016/S0378-4290(98)00137-3
- Meers E, Du Laing G, Unamuno V, Ruttens A, Vangronsveld J, Tack FMG, Verloo MG (2007) Comparison of cadmium extractability from soils by commonly used single extraction protocols. *Geoderma* 141:247–259. doi:10.1016/j.geoderma.2007.06.002
- Mench M, Baize D, Mocquot B (1997) Cadmium availability to wheat in five soil series from the Yonne district, Burgundy. *France Environ Pollut* 95:93–103. doi:10.1016/S0269-7491(96)00078-4
- Merckx R, Brans K, Smolders E (2001) Decomposition of dissolved organic carbon after soil drying and rewetting as an indicator of metal toxicity in soils. *Soil Biol Biochem* 33:235–240. doi:10.1016/S0038-0717(00)00135-8
- Miller RO, Kissel DE (2010) Comparison of soil pH methods on soils of North America. *Soil Sci Soc Am J* 74:310. doi:10.2136/sssaj2008.0047
- Naidu R, Bolan N, Kookana R, Tiller K (1994) Ionic-strength and pH effects on the sorption of cadmium and the surface charge of soils. *Eur J Soil Sci* 45:419–429. doi:10.1111/j.1365-2389.1994.tb00527.x
- Nan Z, Li J, Zhang J, Cheng G (2002) Cadmium and zinc interactions and their transfer in soil-crop system under actual field conditions. *Sci Total Environ* 285:187–195. doi:10.1016/S0048-9697(01)00919-6
- Nawrot T, Plusquin M, Hogervorst J, Roels HA, Celis H, Thijs L, Vangronsveld J, Van Hecke E, Staessen JA (2006) Environmental exposure to cadmium and risk of cancer: a prospective population-based study. *Lancet Oncol* 7:119–126. doi:10.1016/S1470-2045(06)70545-9
- NF ISO 10390, International Organization for Standardization (ISO) (2005) Soil quality-soil pH determination
- NF ISO 10694, International Organization for Standardization (ISO), 1995 Soil quality. Determination of organic and total carbon after dry combustion (elementary analysis)
- NF ISO 11465, International Organization for Standardization (ISO) (1994) Soil quality-determination of dry matter and water content on a mass basis-gravimetric method
- NF X 31–105, standard Association Française de Normalisation (AFNOR) (2014) Soil quality-determination of carbonate content-volumetric method
- NF X 31–107, standard Association Française de Normalisation (AFNOR) (1983) Soil quality-particle size determination by sedimentation-Pipette method
- NF X 31–130, standard Association Française de Normalisation (AFNOR) (1999) Soil quality-chemical method - determination of cationic exchange capacity (CEC) and extractible cations
- NF X 31–147, standard Association Française de Normalisation (AFNOR) (1996). Soil quality. Soils, sediments. Total solubilizing by acid attack
- Norvell WA, Wu J, Hopkins DG, Welch RM (2000) Association of cadmium in durum wheat grain with soil chloride and chelate-extractable soil cadmium. *Soil Sci Soc Am J* 64:2162. doi:10.2136/sssaj2000.6462162x
- Nye PH (1981) Changes of pH across the rhizosphere induced by roots. *Plant Soil* 61:7–26. doi:10.1007/BF02277359
- Oliver DP, Schultz JE, Tiller KG, Merry RH (1993) The effect of crop rotations and tillage practices on cadmium concentration in wheat grain. *Aust J Agric Res* 44:1221–1234. doi:10.1071/AR9931221
- Perrier F, Yan B, Candaudap F, Pokrovsky OS, Gourdain E, Meleard B, Bussi ere S, Coriou C, Robert T, Nguyen C, Cornu JY (2016) Variability in grain cadmium concentration among durum wheat cultivars: impact of aboveground biomass partitioning. *Plant Soil* 404:307–320. doi:10.1007/s11104-016-2847-8
- Ran J, Wang D, Wang C, Zhang G, Zhang H (2016) Heavy metal contents, distribution, and prediction in a regional soil-wheat system. *Sci Total Environ* 544:422–431. doi:10.1016/j.scitotenv.2015.11.105
- Ross S (1994) Toxic metals in soil-plant systems. Wiley
- Sappin-Didier V, Mench M, Gomez A, Masson P (1997)  valuation par des extractions s lectives de l'immobilisation du Cd apr s l'apport de mat riaux inorganiques dans deux terres pollu es. *Comptes Rendus l'Acad mie des Sci - Ser III - Sci la Vie* 320:413–419. doi:10.1016/S0764-4469(97)85030-5
- Satarug S, Baker JR, Urbenjapol S, Haswell-Elkins M, Reilly PEB, Williams DJ, Moore MR (2002) A global perspective on cadmium pollution and toxicity in non-occupationally exposed population. *Toxicol Lett* 137:65–83. doi:10.1016/S0378-4274(02)00381-8
- Sauv  S, Hendershot W, Allen HE (2000) Solid-solution partitioning of metals in contaminated soils: dependence on pH, total metal burden,

- and organic matter. *Environ. Sci. Technol.* 34:1125–1131. doi:10.1021/es9907764
- Schneider A (2006) Adaptation of the ion exchange method for the determination of the free ionic fraction of cadmium in solution. *J Environ Qual* 35:394–401. doi:10.2134/jeq2005.0094
- Shao J (1993) Linear model selection by cross-validation. *J Am Stat Assoc* 88:496–494. doi:10.1080/01621459.1993.10476299
- Sperotto RA (2013) Zn/Fe remobilization from vegetative tissues to rice seeds: should I stay or should I go? Ask Zn/Fe supply! *Front Plant Sci* 4:1–4. doi:10.3389/fpls.2013.00464
- Stolt JP, Sneller FEC, Bryngelsson T, Lundborg T, Schat H (2003) Phytochelatin and cadmium accumulation in wheat. *Environ Exp Bot* 49:21–28. doi:10.1016/S0098-8472(02)00045-X
- Symeonides C, McRae SG (1977) The assessment of plant-available cadmium in soils. *J Environ Qual* 6:120. doi:10.2134/jeq1977.00472425000600020003x
- Tessier A, Campbell PGC, Bisson M (1979) Sequential extraction procedure for the speciation of particulate trace metals. *Anal Chem* 51: 844–851. doi:10.1021/ac50043a017
- Ure AM (1996) Single extraction schemes for soil analysis and related applications. *Sci Total Environ* 178:3–10. doi:10.1016/0048-9697(95)04791-3
- Wenzel WW, Blum WEH, Brandstetter A, Jockwer F, Köchl A, Oberförster M, Oberländer HE, Riedler C, Roth K, Vladeva I (1996) Effects of soil properties and cultivar on cadmium accumulation in wheat grain. *Z Pflanzenernähr Bodenkd* 159:609–614. doi:10.1002/jpln.1996.3581590613
- Zhang H, Zhao FJ, Sun B, Davison W, Mcgrath SP (2001) A new method to measure effective soil solution concentration predicts copper availability to plants. *Environ Sci Technol* 35:2602–2607. doi:10.1021/es000268q
- Zimmerl S, Lafferty J, Buerstmayr H (2014) Assessing diversity in *Triticum durum* cultivars and breeding lines for high versus low cadmium content in seeds using the CAPS marker usw47. *Plant Breed* 133:712–717. doi:10.1111/pbr.12218
- Zogaj M, Düring RA (2016) Plant uptake of metals, transfer factors and prediction model for two contaminated regions of Kosovo. *J Plant Nutr Soil Sci* 179:630–640. doi:10.1002/jpln.201600022