Mécanismes de rétention du cuivre dans les sols : évaluation statistique des approches macroscopiques et spectroscopiques

Thomas Lenoir

▶ **To cite this version:**

Thomas Lenoir. Mécanismes de rétention du cuivre dans les sols : évaluation statistique des approches macroscopiques et spectroscopiques. Sciences de la Terre. Université de Grenoble, 2011. Français. NNT : 2011GRENU028 . tel-00649483

HAL Id: tel-00649483

https://theses.hal.science/tel-00649483

Submitted on 8 Dec 2011

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.

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **TERRE, UNIVERS, ENVIRONNEMENT**

Arrêté ministériel : 7 août 2006

Présentée par

Thomas LENOIR

Thèse dirigée par **Alain MANCEAU**

préparée au sein de **l'Institut des Sciences de la Terre**
dans **l'École Doctorale Terre, Univers, Environnement**

Mécanismes de rétention du cuivre dans les sols : Évaluation statistique des approches macroscopiques et spectroscopiques

Thèse soutenue publiquement le **« 3 octobre 2011 »**,
devant le jury composé de :

Mme Cécile DELOLME
Ingénieur Chercheur, ENTPE, Examinateur
M. Alain MANCEAU
Directeur de Recherche, ISTerre, Directeur de Thèse
M. Laurent MICHOT
Directeur de Recherche, LEM, Rapporteur
M. Michel SCHLEGEL
Ingénieur Chercheur, CEA, Rapporteur
M. Alexandre SIMIONOVICI
Professeur, ISTerre, Examinateur
M. Henri VAN DAMME
Professeur, IFSTTAR, Président

SOMMAIRE

Introduction générale 5

Chapitre 1
Généralités sur le cuivre 9
1) Présentation et propriétés physiques du cuivre 11
2) Utilisations 13
3) Propriétés sanitaires 15
3.1) Formes bio-disponibles 15
3.2) Essentialité du cuivre 17
3.3) Homéostasie du cuivre 19
3.3.1) Chez l'homme et les animaux 19
3.3.2) Chez les plantes 19
3.4) Mécanismes de toxicité 21
3.5) Effets macroscopiques d'un excès de cuivre 22
3.6) Seuils de toxicité 23
3.6.1) Micro-organismes aquatiques et terrestres (protozoaires, champignons, bactéries, virus) 23
3.6.2) Organismes aquatiques 23
3.6.2.1) Invertébrés aquatiques 23
3.6.2.2) Vertébrés aquatiques 25
3.6.2.3) Plantes aquatiques 25
3.6.3) Organismes terrestres 25
3.6.3.1) Invertébrés terrestres 25
3.6.3.2) Vertébrés terrestres 26
3.6.3.3) Plantes terrestres 26
3.7) Conclusions sur la toxicité du cuivre 26
4) Quantités et formes du cuivre dans l'environnement 29
4.1) Dans les roches 29
4.2) Dans l'atmosphère 30
4.3) Dans les eaux 30
4.4) Dans les sols et sédiments 31

Chapitre 2
Mécanismes de rétention du cuivre par les constituants des sols 33
1) Introduction 35
2) Rappels sur les différents mécanismes de fixation 37
2.1) Complexation de sphère externe (CSE) 38
2.2) Complexation de sphère interne (CSI) 39
2.3) Complexation multinucléaire de surface (CMS) et précipitation de surface 40
2.4) Précipitation homogène et coprécipitation 40
2.5) Diffusion dans le réseau cristallin 42
3) Processus de l'adsorption 43
3.1) Capacité d'échange cationique 43
3.2) Isothermes d'adsorption 43
3.2.1) Classification 43
3.2.2) Isotherme de Langmuir 44
3.2.3) Isotherme de Freundlich 45
3.3) La matière minérale 46

3.3.1) Modèles de Complexation de Surface (MCS) 46
3.3.1.1) Approche théorique 46
3.3.1.2) Limites des MCS 50
3.3.2) Approche spectroscopique 50
3.3.3) Description des surfaces minérales dans les sols et mécanismes d'interactions Cu-matière minérale 51
3.3.3.1) Les phyllosilicates 51
3.3.3.2) Les oxy(hydroxy)des de fer 56
3.3.3.3) Les oxy(hydroxy)des de manganèse 58
3.3.3.4) Les autres oxydes 59
3.3.3.5) Les carbonates 60
3.3.4) Conclusions sur les mécanismes de fixation avec la matière minérale 61
3.4) La matière organique 61
3.4.1) Structure et propriétés 62
3.4.2) Approches macroscopiques 65
3.4.2.1) Comparaison des approches électrostatiques 65
3.4.2.2) Modèle WHAM - modélisation par distribution discrète 66
3.4.2.3) Modèle NICA - modélisation par distribution continue 67
3.4.2.5) Problèmes liés aux modélisations des courbes de titrages protoniques de la MO 68
3.4.3) Approche spectroscopique 70
3.4.4) Mécanismes de rétention Cu-MO 70

Chapitre 3
Détermination de la spéciation du cuivre par couplage µ-XRF, µ-XANES et ACP 73
Application au porphyre cuprifère de Mamut (Malaisie) et au bassin de rétention et d'infiltration d'eaux pluviales de Django Reinhardt (France) 73
1) Principe de l'approche 75
2) Présentation des sites 79
2.1) Porphyre cuprifère de Mamut 79
2.1.1) Données géographiques, climatiques et géologiques 79
2.1.2) Problèmes environnementaux 81
2.1.2.1) Accidents d'exploitation 81
2.1.2.2) "Acid Mine Drainage" 81
2.1.3) Echantillonnage 82
2.2) Bassin d'infiltration et de rétention Django Reinhardt (DjR) 83
2.2.1) Données géographiques et fonctionnelles 83
2.2.2) Problèmes environnementaux 85
2.2.3) Echantillonnage 86
3) Méthodes 87
3.1) Préparation des échantillons et observations préliminaires 87
3.2) Mesures par rayonnement X 87
3.2.1) Caractéristiques du microspectromètre 10.3.2 87
3.2.2) Cartographie de µ-XRF 88
3.2.3) Spectroscopies µ-XANES et µ-XRF 88
3.3) Analyse en Composantes Principales 90
3.3.1) Cascade des valeurs propres 90
3.3.2) Indicateur IND 91
3.3.3) Test de Fisher statistique 91
3.3.4) Approche itérative 92

4) Résultats et discussion 93
4.1) Site de Mamut 93
4.1.1) Analyses chimiques 93
4.1.2) Cartographie μ-XRF 95
4.1.3) Spectroscopies μ-XANES et μ-XRF 95
4.1.4) Analyse en composantes Principales 99
4.1.4.1) Cascade des valeurs propres 99
4.1.4.2) IND 100
4.1.4.3) F-test 100
4.1.4.4) Approche itérative 101
4.2) Site de Django Reinhardt 104
4.2.1) Analyses chimiques 104
4.2.2) Cartographie μ-XRF 104
4.2.3) Spectroscopies μ-XANES et μ-XRF 105
4.2.4) Analyses en Composantes Principales 107
4.2.4.1) Cascade des valeurs propres 107
4.2.4.2) IND 108
4.2.4.3) F-test 108
4.2.4.4) Approche itérative 110
4.3) Comparaison entres les deux sites de prélèvement 110

Chapitre 4
Evaluation statistique des critères ACP usuels utilisés en Spectroscopie d'Absorption X 113
Proposition d'un nouveau critère : l'indicateur NSS 113
1) Introduction 115
2) Evaluation de l'ACP en spectroscopie d'absorption X – Cascade des valeurs propres 119
2.1) Principe 119
2.2) Résultats 119
2.2.1) XANES 1 et 2 119
2.2.2) EXAFS 1 et 2 120
2.3) Conclusion 121
3) Mesure du bruit expérimental et génération sur des données synthétiques 123
3.1) Introduction 123
3.2) XANES 124
3.2.1) Evaluation du bruit 124
3.2.1.1) Principe 124
3.2.1.2) Résultats et discussion 125
3.2.2) Génération de bruit sur les données synthétiques 127
3.3) EXAFS 129
3.3.1) Evaluation du bruit 129
3.3.1.1) Principe 129
3.3.1.2) Résultats 129
3.3.1.3) Discussion 134
3.3.2) Génération de bruit sur les données synthétiques 137
3.4) Conclusion 138
4) Evaluation statistique de la fonction IND et du F-Test 139
4.1) Principe 139
4.2) Résultats et discussions 140
4.2.1) XANES 1 140
4.2.1.1) Evaluation de la fonction IND 140

4.2.1.2) Evaluation du F-test (α = 5%) 141
4.2.2) XANES 2 142
4.2.2.1) Evaluation de la fonction IND 142
4.2.2.2) Evaluation du F-test (α = 5%) 144
4.2.3) Conclusions sur le XANES 145
4.2.4) EXAFS 1 145
4.2.4.1) Evaluation de la fonction IND 145
4.2.4.2) Evaluation du F-Test (α = 5%) 147
4.2.5) EXAFS 2 148
4.2.5.1) Evaluation de la fonction IND 148
4.2.5.2) Evaluation du F-Test (α = 5%) 149
4.2.6) Conclusions sur l'EXAFS 150
4.3) Conclusion générale 150
5) Proposition et évaluation d'un nouveau critère ACP : l'indicateur NSS 151
5.1) Introduction 151
5.2) Construction de l'indicateur 151
5.3) Application aux données expérimentales 152
5.3.1) XANES 153
5.3.2) EXAFS 155
5.3.3) Discussion 155
5.4) Evaluation statistique 155
5.4.1) XANES 156
5.4.1.1) XANES 1 156
5.4.1.2) XANES 2 158
5.4.1.3) Conclusions sur le XANES 159
5.4.2) EXAFS 160
6) Conclusions 163

Conclusions et Perspectives 165

Annexe 1
Principes de la spectroscopie d'absorption X 169

Annexe 2
Semi-empirical proton binding constants for natural organic matter 173

Annexe 3
Number of Independant Parameters in the Potentiometric Titration of Humic Substances 207

Annexe 4
Convergence-Optimized Procedure for Applying the NICA-Donnan Model to Potentiometric Titrations of Humic Substances 227

Bibliographie 251

Introduction générale

On dénomme communément 'métaux lourds' (ML) les éléments métalliques et métalloïdes naturels qui possèdent une masse volumique élevée, typiquement supérieure à 5 grammes par cm^3 à l'état métallique, et qui sont nocifs pour les êtres vivants au delà d'un certain seuil de concentration sous certaines formes chimiques. Ce sont des éléments chimiques non radiogéniques, donc ils ne se transforment pas. En revanche, ils sont susceptibles de changer de degré d'oxydation et de forme cristallochimique, c'est-à-dire de spéciation, et donc d'être plus ou moins mobiles et assimilables par les êtres vivants en fonction de leur état. On peut citer comme exemple le cas du chrome qui est très soluble et cancérigène sous sa forme la plus oxydée (Cr^{6+}), et peu soluble et non toxique sous forme d'ion chromique (Cr^{3+}). Depuis l'arrêté du 2 février 1998 relatif aux installations classées pour la protection de l'environnement, neuf ML font l'objet de contraintes environnementales visant à limiter leur rejet dans les eaux : le chrome, le nickel, le cuivre, le zinc, l'arsenic, le cadmium, l'étain, le mercure et le plomb.

Les ML sont présents dans tous les compartiments environnementaux (atmosphère, sols, sédiments, eaux de surface, aquifères...), du fait de phénomènes naturels (éruption volcanique, érosion physique, altération chimique...), ou bien d'activités humaines. Dans le premier cas, et en dehors des gisements miniers et autres enrichissements géochimiques particuliers, ils sont le plus souvent à l'état de traces (typiquement quelques mg/kg), alors que dans le second cas, leur concentration est supérieure, souvent de plusieurs ordres de grandeur (plusieurs g/kg de zinc dans des friches industrielles ou des sols urbains). Sur les surfaces émergées, les ML pénètrent dans les écosystèmes par retombées atmosphériques ou par transport en solution. La spéciation de plusieurs ML, dont l'arsenic et le zinc, dans chacun des ces milieux est désormais assez bien documentée (Manceau et al. 2002; Morin et Calas 2006).

Les sols et sédiments sont des lieux d'interactions et d'échanges dynamiques entre le milieu aqueux, le monde organo-minéral, la biosphère et l'atmosphère. Chaque compartiment est à lui seul un objet complexe, et donc étudié le plus souvent de façon séparée par des approches généralement disciplinaires (chimie des solutions, minéralogie, biogéochimie, microbiologie, physiologie et écotoxicologie du vivant...). Or, une bonne compréhension de la dynamique des éléments traces, et en particulier des polluants métalliques, à l'interface entre ces trois compartiments nécessite de coupler, autant que possible, les outils et d'intégrer les concepts propres à chacune des communautés scientifiques (Manceau 2005).

Ces processus de transfert et de rétention peuvent ainsi être étudiés par une approche macroscopique, en utilisant des modèles géochimiques qui prennent en compte les effets du pH, de la force ionique, du potentiel redox, et de la nature, forme et concentration des éléments présents en solution. Cette approche macroscopique est directement exploitable lorsqu'on s'intéresse à des systèmes simples et homogènes, comme un sable (Ramos et al. 1994). Cependant, un sol est un milieu complexe qui ne peut pas être décrit comme la juxtaposition des ses composants. Par ailleurs, les observations macroscopiques sont le résultat de phénomènes élémentaires à l'échelle moléculaire dont les mécanismes intrinsèques sont difficilement appréhendables à partir de mesures globales.

Un des moyens d'accéder à ces mécanismes est alors d'utiliser des outils d'analyse sensibles à l'état atomique et moléculaire de la matière. C'est le cas notamment de la spectroscopie d'absorption X (XAS) qui est une sonde locale de l'environnement chimique et structural d'un élément donné, comme un ML. Néanmoins, si cette méthode permet de connaître le mécanisme élémentaire de rétention d'un ML au sein de la matrice hôte, l'information recueillie est la moyenne de toutes les formes du ML dans le volume analysé. Or, dans un sol, les ML sont le plus souvent présents sous plusieurs formes organiques et inorganiques. L'utilisation d'un faisceau X focalisé sur quelques micromètres (μXAS) permet de contourner, ou au moins d'atténuer, cette difficulté. Une première étape consiste alors à imager la distribution chimique du ML à l'échelle microscopique par micro-fluorescence X sur rayonnement synchrotron (μSXRF). Sa spéciation dans chacune de ces associations est ensuite déterminée par spectroscopie μXAS. Le volume analysé est alors de quelques centaines à plusieurs milliers de μm^3, ce qui accroît la probabilité d'isoler le signal de phases métallifères pures, qu'elles soient majoritaires ou minoritaires dans la totalité de l'échantillon. Les spectres μXAS sont ensuite traités par Analyse en Composantes Principales (ACP) pour estimer le nombre de phases porteuses présentes dans l'échantillon. Une fois dénombrées, ces phases sont identifiées par comparaison avec les signatures spectrales d'espèces pures, puis leurs proportions à l'échelle macroscopique sont obtenues en reconstruisant le spectre XAS de poudre représentatif de la totalité de l'échantillon par une combinaison linéaire des spectres d'espèces pures.

L'objectif du présent travail de thèse a été de contribuer à la caractérisation, à l'échelle spatiale microscopique, des formes chimiques et structurales du cuivre dans des matériaux de surface de matrices contaminées.

En 1983, il a été estimé que 66% des émissions de cuivre dans l'environnement résultaient d'activités anthropiques (Nriagu 1989). Parmi les sources d'émissions humaines,

on compte les fours de fusion, les fonderies, les centrales thermiques, les sources de combustion telles que les installations d'incinération. Les rejets de cuivre dans le sol proviennent essentiellement de l'agriculture où il est utilisé comme fongicide, de l'irrigation par des eaux usées non clarifiées, de l'épandage d'amendements organiques divers, dont les boues de station d'épuration, et de terres de recouvrement des exploitations minières.

En cohérence avec cette constatation, cette étude a été focalisée sur la détermination des formes du cuivre aux alentours d'une exploitation cuprifère et dans un bassin d'infiltration d'eaux pluviales.

Ce manuscrit comporte quatre parties :

- La première partie traite de généralités sur le cuivre. Elle présente en particulier sa toxicité vis-à-vis des différents organismes vivants, justifiant ainsi l'étude de cet élément en tant que contaminant. Les formes du cuivre le plus souvent présentes dans l'environnement sont également évoquées.
- La seconde partie examine les mécanismes de rétention du cuivre par les principaux constituants du sol. Au travers de l'étude approfondie du modèle géochimique NICA-Donnan, elle illustre notamment la nécessité de coupler les approches macroscopiques et moléculaires, comme cela a été mentionné précédemment.
- La troisième partie tente de déterminer les formes chimiques sous lesquelles le cuivre est présent dans les deux matrices étudiées. Elle met à jour les difficultés rencontrées au niveau de la détermination par ACP du nombre d'espèces cuprifères présentes.
- Dans la quatrième et dernière partie, cette difficulté est contournée par la mise au point d'un nouvel indicateur ACP. Il sera d'abord évalué statistiquement sur un jeu de données synthétiques, puis appliqué à des données expérimentales, dont celles obtenues dans ce travail.

Chapitre 1

Généralités sur le cuivre

1) Présentation et propriétés physiques du cuivre

Le cuivre est un métal utilisé depuis la préhistoire. Il y a plus de 10000 ans, il était déjà exploité dans le nord de l'Irak, probablement à partir de pépites de cuivre natif. Les premiers objets en cuivre remontent à plus de 6000 ans. Il était utilisé pour la fabrication d'armes, d'outils et de bijoux. L'extraction du cuivre sous forme de minerai a commencé 5000 ans avant JC (Emsley 2003). Plus tard, il a été découvert qu'associé à l'étain il formait un alliage beaucoup moins malléable et qui pouvait être affilé. Le bronze était né. L'Age du Bronze s'étend d'environ 3000 à 1000 avant JC (Zoltán et al. 2009). Dans l'antiquité, le cuivre a commencé à être utilisé pour ses propriétés colorantes et fongiques : les civilisations grecques et romaines se servaient du vert de gris (acétate de cuivre) comme pigment et comme médicament. Au moyen âge, l'excellente conductivité thermique du cuivre métallique était utilisée dans les casseroles et bouilloires, puis dans l'imprimerie comme plaques de gravures. Plus tard, il a été utilisé pour gainer les coques des bateaux en bois pour empêcher la fixation des crustacés (Emsley 2003).

Il tire son nom du latin *Cyprium aes* qui a pour traduction 'le métal de Chypre', principal producteur dans l'antiquité. Son symbole chimique, Cu, provient de son nom latin.

Le cuivre est un métal de la première série de transition. Il fait partie du groupe 11 (IB) dans le tableau périodique des éléments. Il est malléable, ductile, et résiste à la corrosion de l'air et de l'eau mais est attaqué par les acides. Son numéro atomique est égal à 29 et sa masse molaire est de 63,57 $g.mol^{-1}$ (Isabey 1933). Les deux isotopes naturels sont le ^{63}Cu (69%) et le ^{65}Cu (31%) (Emsley 2003). A l'état métallique (Cu^0), il a une couleur rouge-orangée, sa structure cristalline est cubique face centrée de paramètre de maille a=3,61 Å (Fig. 1.1) et de densité 8.9 $g.cm^{-3}$. Sa densité supérieure à 5 $g.cm^{-3}$ le classe dans la famille des métaux lourds (Flemming et Trevors 1989). Sa configuration électronique est $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^1$ (Isabey 1933 ; Alloway 1995; Cotton et al. 1999). Suivant la règle classique de remplissage des orbitales atomiques (règle de Klechkowski), la configuration devrait être $3d^9 4s^2$. Mais la configuration électronique de type $s^1 d^{10}$ est énergétiquement plus favorable que la configuration $s^2 d^9$. Du fait de la capacité des électrons de la sous couche d à se délocaliser dans le réseau métallique, ses températures de fusion et de vaporisation sont élevées (Tableau 1.1). Les formes les plus courantes à l'état naturel étant sont les ions cupriques Cu^{2+} ($3d^9$) (Fig. 1.1) et cupreux Cu^+ ($3d^{10}$).

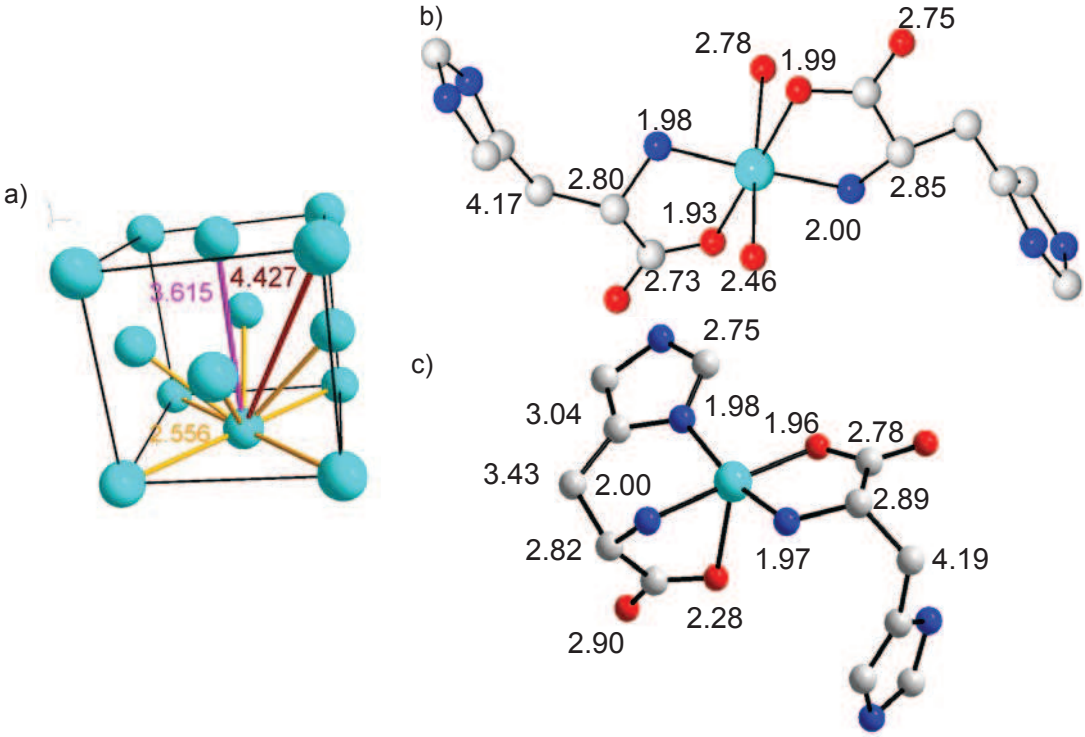


Figure 1.1 : a) Structure cristalline du cuivre métallique et distances interatomiques des atomes dans la maille. b) Complexe Cu(II)-histidine stable à pH<5 (Evertsson 1969). c) Complexe Cu(II)-histidine au pH physiologique de 7,4 (Deschamps et al. 2004). Ce complexe intervient dans la régulation (homéostasie) et le transport du cuivre à travers les membranes cellulaires (Sarkar 1999). Les atomes de cuivre sont représentés en bleu turquoise, les atomes d'azote en bleu foncé, les atomes d'oxygène en rouge et les atomes de carbone en gris clair. Les atomes d'hydrogène ne sont pas représentés dans les complexes.

Tableau 1.1 : Principales propriétés physico-chimiques du cuivre aux conditions de pression et température ordinaires

N° atomique	29
Masse molaire	63,5 $g.mol^{-1}$
Densité	8,9 $g.cm^{-3}$
T° de fusion	1084°C
T° de vaporisation	2567°C
Conductivité électrique	59,6 10^5 $S.m^{-1}$
Conductivité thermique	401 $W.m^{-1}K^{-1}$

2) Utilisations

Les principaux gisements de cuivre actuels sont au Chili, en Indonésie, aux Etats-Unis, en Australie et au Canada (80% de la quantité mondiale). Le cuivre s'y trouve généralement sous forme de sulfures (e.g. la chalcopyrite $CuFeS_2$ et la chalcocite Cu_2S). La production mondiale de cuivre est de l'ordre de 15 MT/an et les réserves exploitables connues sont de 300 MT environ. Environ 2 MT/an sont actuellement recyclés, et les déblais des porphyres cuprifères sont une source potentielle de cuivre. Vingt mille tonnes par an sont ainsi produites dans la mine de Manza au Chili (Emsley 2003). L'oxydation des sulfures de cuivre par certaines bactéries comme *Thiobacillus ferroxidans* conduit à la libération de l'ion soluble Cu^{2+} (Mathur et al. 2005; Watling 2006).

Du fait de la grande conductivité du cuivre métallique, 60% de sa production entre dans la confection de câbles ou circuits électriques. 20% sont utilisés dans le bâtiment pour la plomberie et les toitures, 15% dans des machines industrielles comme les échangeurs thermiques, les 5% restant servant à la fabrication des alliages. Si le bronze est l'alliage le plus anciennement connu, le cuivre forme avec le zinc un autre alliage, le laiton. C'est un alliage résistant à l'aspect lustré utilisé notamment dans la production d'armes. Le nickel peut également être adjoint au cuivre. Les cupronickels entrent dans la composition des pièces de monnaie ainsi que dans la confection de matériels marins (pompes, hélices, unité de désalinisation) du fait de leur résistance à la corrosion. Enfin, les bronzes d'aluminium ont un aspect doré et brillant qui est utilisé en ornementation. Ainsi, bien que relativement peu visible, le cuivre est un élément omniprésent dans notre entourage : électricité, tuyauterie, outillage, etc….

De manière encore plus discrète, le cuivre est utilisé par l'ensemble des organismes vivants. C'est un cofacteur enzymatique dans la chaine de transport des électrons qui conduit au stockage d'énergie sous forme d'adénosine triphosphate (ATP) (Sarkar 1999). Paradoxalement à cette essentialité, il possède des caractéristiques fongiques. La bouillie bordelaise, mélange de sulfate de cuivre hydraté et de chaux, est depuis sa découverte à la fin du XIXème siècle le principal fongicide contre le mildiou de la vigne. Il entre également dans la composition des peintures antifouling utilisées pour prévenir la fixation des algues sur les coques de bateaux. (Flemming et Trevors 1989).

Le cuivre est donc indispensable à la vie mais des effets toxiques lui sont imputables. Plusieurs questions se posent alors : quelles sont les formes biodisponibles ?, à quoi sert-il ?,

quel est son devenir dans les organismes ?, pourquoi peut il être toxique ?, quels sont ses méfaits ?, quand peut-on alors parler de toxicité ?

3) Propriétés sanitaires

3.1) Formes bio-disponibles

La biodisponibilité d'un élément se définit comme la quantité de cet élément qui est effectivement assimilée par un organisme. L'évaluation de la biodisponibilité ne peut pas se faire à partir de la seule connaissance des quantités totales de cuivre dans le sol (Wightwick et al. 2008). Elle dépend de la spéciation de l'élément et des conditions physico-chimiques du milieu qui sont susceptibles de modifier cette spéciation vers des formes chimiques plus ou moins assimilables (Alloway 1995).

La forme divalente hydratée, $Cu(H_2O)_n^{2+}$, est particulièrement labile, comme le montre la relation entre la toxicité du cuivre et l'activité en solution de l'ion Cu^{2+} (Luo et al. 2006). Comme ses voisins chimiques divalents (Zn et Ni), l'ion cuprique solvaté est généralement représenté par un octaèdre. Or sa configuration électronique est de type d^9 et lorsqu'il est sous coordinence octaédrique, on observe un accroissement de la stabilisation d'énergie lié à la déformation spontanée de la structure connu sous le nom d'effet Jahn-Teller (Pokrovsky et al. 2008 ; Frank et al. 2009; Smirnov et Trostin 2009). Cet effet s'explique grâce aux orbitales moléculaires. Dans un champ de ligands octaédriques, les cinq orbitales 3d du cuivre se subdivisent en deux groupes d'orbitales : les orbitales triplement dégénérées T_{2g} (d_{xy}, d_{xz} et d_{yz}) et les orbitales doublement dégénérées E_g ($d_{x^2-y^2}$ et d_{z^2}). L'ion Cu(II) étant en configuration $3d^9$, l'orbitale E_g (de plus haute énergie) contient un électron célibataire. Cet électron peut alors se placer dans l'un ou l'autre des niveaux $d_{x^2-y^2}$ et d_{z^2} entrainant une levée de dégénérescence (Fig. 1.2). Cette configuration électronique est à l'origine d'une déformation structurale qui se caractérise par une contraction de l'octaèdre dans le plan équatorial et une élongation perpendiculairement ($d(Cu\text{-}O_{eq})<d(Cu\text{-}O_{ax})$). Dans le complexe hydraté $Cu(H_2O)_6$, les distances mesurées par spectroscopie EXAFS entre le cuivre central et les quatre molécules d'eau équatoriales sont de 1,97 ± 0,01 Å tandis que les deux molécules d'eau apicales sont à 2,24 ± 0,03 Å (Korshin et al. 1998). Néanmoins, il n'est pas certain que cette représentation soit tout à fait exacte. Une autre représentation octaédrique donne pour les liaisons équatoriales deux distances Cu-H_2O, l'une à 1,97 Å et l'autre à 2,00 Å, et une distance axiale de 2,32 Å (Farquhar et al. 1997). Récemment, c'est la représentation octaédrique elle-même qui a été mise en doute. Des calculs théoriques ont montré qu'en solution les coordinations à 5 et 6 molécules d'eau avaient une stabilité équivalente, suggérant que les deux formes coexisteraient (Bryantsev et al. 2008). Il a d'abord été cru que le

complexe de cuivre solvaté en coordinence cinq subissait de fréquentes transformations (de l'ordre de la picoseconde) passant d'une structure pyramidale carrée à une structure bipyramidale triangulaire avec les cinq distances Cu-H_2O égales à 1,96 Å dans la première sphère d'hydratation, la seconde sphère de coordination étant composée de six molécules d'eau à des distances comprises entre 3,02 et 3,44 Å (Pasquarello et al. 2001). Cet édifice pentaédrique a été confirmé par la suite par deux autres études, mais elles ont proposé des distances Cu-H_2O différentes. Dans l'une, le plan équatorial est déformé et deux distances Cu-O_{eq}, équivalentes aux incertitudes près, ont été déterminées (1,98 ± 0,03 et 1,95 ± 0,03 Å) et la distance Cu-O_{ax} = 2,35 ± 0,05 Å (Benfatto et al. 2002; Frank et al. 2005). Dans l'autre, le plan équatorial reste régulier (Cu-O_{eq} = 2,00 Å) et Cu-O_{ax} = 2,45 Å. Dans cette seconde étude, la seconde sphère de coordination est composée de 8 molécules d'eau (Amira et al. 2005).

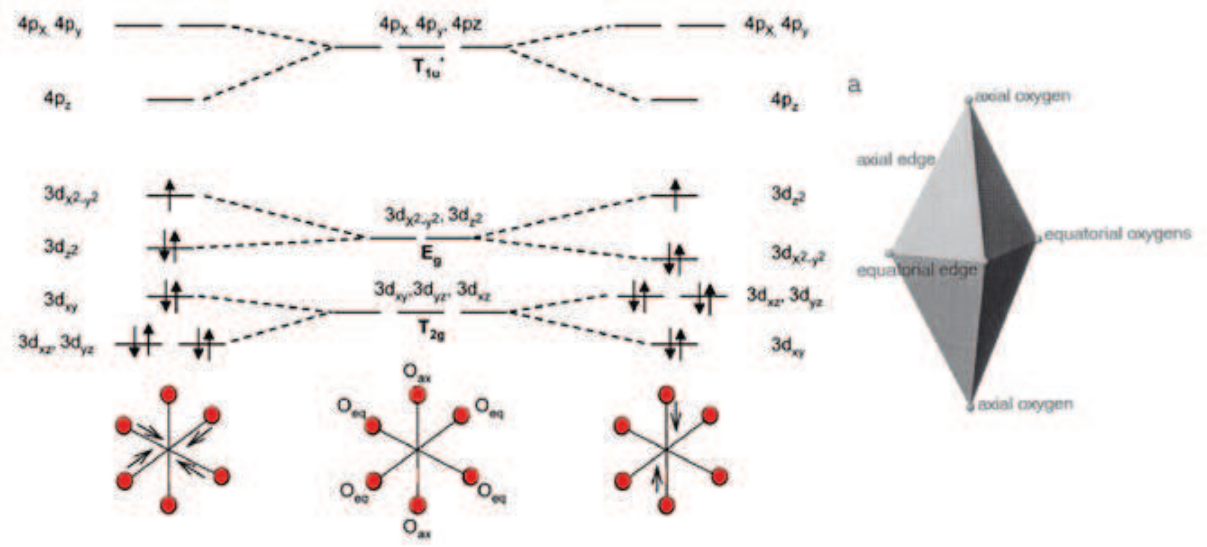


Figure 1.2 : Principe de l'effet Jahn-Teller et schéma d'un octaèdre de cuivre typique avec quatre courtes distances dans le plan équatorial et deux longues perpendiculairement (Cheah et al. 2000; Matynia 2009).

L'ion libre hydraté $Cu(H_2O)_n^{2+}$ n'est pas le seul à être biodisponible. Aujourd'hui il semble assez bien établi que l'assimilation des éléments chimiques par les cellules vivantes se fait par complexation à des petites molécules organiques, comme par exemple les phytosidérophores dans le cas du fer (Roberts et al. 2004). Ainsi, la toxicité du cuivre augmente lorsqu'il est associé à certains ligands liposolubles ce qui facilite le passage à travers les membranes cellulaires (Ahsanullah et Florence 1984) et dans le corps humain, il passerait la barrière intestinale complexé à des petits acides aminés (Marceau et al. 1970) provenant des sécrétions gastro-intestinales comme la salive et les sucs gastriques (Gollan et Deller 1973). Enfin, compte tenu de la faible abondance de composés solubles de cuivre

monovalent dans les milieux oxygénés, peu de données existent sur la biodisponibilité de Cu(I). Néanmoins, on peut noter que chez les êtres humains, le cuivre pénètre dans les cellules sous forme monovalente associé à une protéine (hCTR) et qu'une fois dans le cytoplasme, toujours sous forme monovalente, il serait associé en majorité avec un polypeptide, le glutathion (Freedman et al. 1989; Zhou et Gitschier 1997 ; Sarkar 1999).

3.2) Essentialité du cuivre

Le caractère essentiel du cuivre vient du fait qu'il est incorporé dans un grand nombre d'enzymes et de protéines structurelles. Son rôle dans l'activité redox des enzymes est une conséquence de sa capacité à intervenir en tant qu'intermédiaire dans les transferts d'électrons. Ainsi, le cuivre est présent dans des enzymes impliquées notamment dans la respiration cellulaire, la défense contre les radicaux libres, les fonctions de neurotransmission, la biosynthèse des tissus conjonctifs et le métabolisme du fer dans les cellules. C'est un cofacteur qui confère aux enzymes leur structure allostérique appropriée à leur activité catalytique. On peut par exemple citer la cytochrome c oxydase qui intervient dans la production d'énergie. Le cuivre confère ses propriétés oxydoréductrices à cette coenzyme intermédiaire de la chaine respiratoire pour permettre la réoxydation des coenzymes réduites au cours du cycle de Krebs, dont la finalité est de produire les intermédiaires énergétiques qui serviront à la production d'adénosine triphosphate (ATP). Toujours au chapitre des mécanismes de production d'énergie, la plastocyanine est une protéine du photosystème I du dispositif photosynthétique des plantes qui intervient les processus de transfert électrons (Bishop 1964; Linder et Hazegh-Azam 1996). La superoxyde dismutase protège contre les radicaux libres via la dismutation catalytique du superoxyde (O_2^-) en dioxygène et peroxyde d'hydrogène (H_2O_2), alors que la céruloplasmine est impliquée dans le métabolisme du fer (oxydoréduction Fe^{2+}/Fe^{3+}) et le transport du cuivre dans le sang. (Linder et Hazegh-Azam 1996 ; Emsley 2003). Le site actif de l'hémocyanine, qui est la protéine de transport de l'oxygène chez certains invertébrés (mollusques, crustacés), contient deux atomes de cuivre (Rose et Bodansky 1920). Cette forme oxydée du cuivre donne au sang des ces animaux une couleur bleu-verte. L'hémocyanine constitue l'équivalent chez ces animaux de l'hémoglobine chez les vertébrés qui contient du fer oxydé donnant au sang sa couleur rouge. Aucun autre élément chimique de la classification ne fournit à ces protéines les propriétés redox au cœur de leur fonction que le cuivre assure.

Le cuivre joue également un rôle important dans l'activation et la répression de la transcription des gènes. Ainsi, dans la levure, il est associé à une protéine (ACE1) activateur

de la transcription de la métallothionéine, qui joue un rôle de détoxication de l'organisme contre les éléments traces métalliques (Dameron et al. 1991).

Le cuivre est présent dans tous les tissus humains (Tableau 1.2) et l'Organisation Mondiale de la Santé recommande un apport journalier minimum de 30 $\mu m_{Cu}/kg_{poids}$ pour un adulte, et de 40 $\mu m_{Cu}/kg_{poids}$ pour un nourrisson (WHO 1973). La carence en cuivre peut entrainer, chez l'homme et les animaux, des anémies, des lésions cardiovasculaires, et des anomalies osseuses (Rucker et al. 1969; WHO 1973; Schoenemann et al. 1990).

Une quantité de cuivre comprise entre 5 et 20 ppm est nécessaire dans le tissu des plantes terrestres pour une croissance normale. Une quantité inférieure à 4 ppm est considérée comme insuffisante et une quantité supérieure à 20 ppm est considérée comme excessive (Nriagu 1979; Robson et Reuter 1981; Alloway 1995; Stevenson 1986). Les symptômes habituellement rencontrés lors d'une carence sont une croissance réduite (Lipman et Mackinney 1931; Sommer 1931), une déformation des jeunes feuilles, une nécrose du méristème apical, une chlorose et la fanaison des jeunes pousses (Fig. 1.3). Chez les citronniers ou les orangers, cette carence entraîne une chute des jeunes feuilles connue sous le terme de mort estivale (Rahimi et Bussler 1973 ; Gupta 1979; Hopkins 2003).

Tableau 1.2 : Quantités ubiquitaires du cuivre dans le corps humain (Emsley 2003).

Sang	1 mg .kg^{-1}
Os	1-25 mg .kg^{-1}
Tissus	2-10 mg .kg^{-1}
Quantité totale dans le corps	70 mg .kg^{-1}

Figure 1.3 : Blé atteint d'une carence en cuivre (Reid 2006)

3.3) Homéostasie du cuivre

3.3.1) Chez l'homme et les animaux

C'est par les voies digestives que le cuivre est principalement absorbé dans le corps (Van Campen et Mitchell 1965). Une fois la paroi intestinale franchie, le cuivre est transporté dans le sang sous forme divalente et complexée (ou chélatée) en majorité à la céruloplasmine (90%) et dans une moindre mesure à l'albumine et des acides aminés dont l'histidine principalement (Fig. 1.1c) (Neumann et Sass-Kortsak 1967). Le cuivre est réduit sous forme monovalente à son entrée dans les cellules, puis est associé principalement au glutathion dans le cytoplasme, avant sa chélation éventuelle par les métallothionéines chargées de son stockage. En sus du glutathion, des protéines chaperonnes (dont la fonction est d'assister d'autres protéines dans leur maturation) assurent le transfert du cuivre à la cytochrome c oxydase (Cox17), à la superoxyde dismutase (Ccs1), et aux ATPases cuprifères (ATOX1). Ce serait sous cette dernière forme qu'il sortirait de la cellule puis serait excrété par la bile. (Bremner 1987; Amaravadi et al. 1997; Culotta et al. 1997; Klomp et al. 1997; Sarkar 1999).

3.3.2) Chez les plantes

Les taux d'absorption du cuivre sont parmi les plus bas comparés à ceux des autres éléments essentiels, et varient suivant les espèces et les variétés (Gilkes 1981). Dans la rhizosphère, c'est-à-dire dans le volume d'interaction des racines, le cuivre est généralement majoritairement complexé à la matière organique du sol (exsudats racinaires, substances humiques…), mais également sous forme minérale dans une proportion qui dépend de la nature du milieu. Les quantités absorbées par les plantes sont corrélées à la quantité d'ions libres en solution à proximité des racines (Quartacci et al. 2009), et donc au pH. Quand celui-ci décroit, l'activité de l'ion libre $[Cu(OH_2)_n]^{2+}$ s'accroît du fait de la dissolution des minéraux et de la protonation des groupements fonctionnels (carboxyles, amines, phénols…) de la matière organique. C'est à des pH<5 que le cuivre est le plus mobile et soluble. Comme le montre le diagramme de spéciation dans l'eau pure (Fig. 1.4), à pH>7 il précipite sous la forme $Cu(OH)_2$ en solution aqueuse, et n'est pratiquement plus mobile (Adriano 1986). Le cuivre dissous est en compétition avec les autres éléments chimiques pour son absorption racinaire, dont Ca^+, K^+, NH_4^+, et principalement Zn^{2+} (Alloway 1995). Le cuivre étant un oligo-élément, toutes les plantes possèdent un niveau de base de tolérance. Contrairement à certains autres métaux, dont le zinc par exemple (Verbruggen et al. 2009), on ne connait néanmoins pas la ou les protéines et donc les formes qui permettent au cuivre de passer la

paroi des racines. On suppose que l'absorption métallique est un processus actif de la part de la plante, et qu'elle est favorisée par les exsudats racinaires (Quartacci et al. 2009).

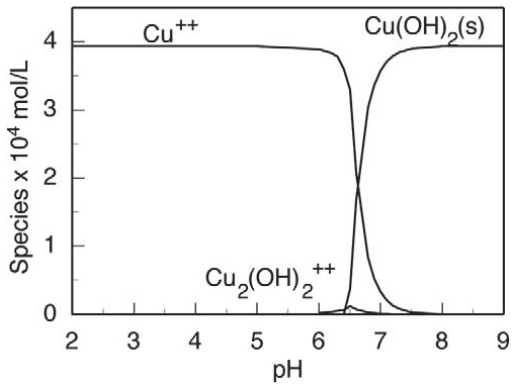


Figure 1.4 : Spéciation du cuivre divalent en fonction du pH. $[Cu]_{tot}$ = 3,94.10^{-4} mol/L dans 0,1 M $NaNO_3$ (Peacock et Sherman 2005).

En particulier, il a été montré que l'absorption de cuivre augmentait dans les sols pauvres en fer, cette augmentation étant attribuée à l'augmentation des phytosidérophores dans les exsudats (Komarek et al. 2010). Une fois cette barrière passée, les réseaux de transport du cuivre dans les cellules des plantes sont similaires à ceux des cellules animales (Clemens 2001). Le cuivre entre dans la cellule via un transporteur COPT1. Puis il est pris en charge par des chaperonnes qui le distribuent aux protéines qui en ont besoin et à des phytochélatines, qui sont des métalloprotéines de stockage proches des métallothionéines (Grotz et Guerinot 2006). Il est également complexé à des acides aminés, dont un des rôles est de limiter la présence d'ions libres hydratés particulièrement toxiques dans le cytosol. C'est également complexé à des molécules organiques, dont la nicotianamine, qu'il est exporté via xylème dans les tissus foliaires. Cet acide aminé permet la chélation du cuivre sous forme octaédrique stable grâce à ses six groupements fonctionnels, même dans des environnements acides dont le pH est de l'ordre de 3 (Rellan-Alvarez et al. 2008; Irtelli et al. 2009; Mijovilovich et al. 2009; Trampczynska et al. 2010).

Quand les quantités de cuivre dépassent les besoins nécessaires, les stratégies des plantes varient selon les espèces. En fonction des quantités absorbables on distingue alors les plantes tolérantes des plantes hyperaccumulatrices qui peuvent accumuler plus de 1000 $mg.kg^{-1}_{matière\ sèche}$ dans leurs tissus aériens (Ali et al. 2002). Deux processus principaux de défense ont été décrits : l'exclusion quand il y a réduction de l'absorption racinaire, et l'accumulation lorsque la barrière endodermique des racines est franchie (Baker et Walker 1990). Comme illustration du mécanisme d'exclusion, on peut citer la formation de nanoparticules de cuivre métallique autour des racines de *Phragmites australis* et de l'*Iris*

pseudoacorus (Manceau et al. 2008). En cas d'accumulation, dans la grande majorité des cas, la partie en excès du cuivre qui pénètre n'est pas transférée vers les parties aériennes mais reste localisée dans les racines. Soit le métal est fixé à la paroi des cellules où il serait alors associé à des groupements thiols -SH ou hydroxyles -OH (Liu et Kottke 2004; Lou et al. 2004; Sahi et al. 2007; Komarek et al. 2010), soit il est stocké à l'intérieur des cellules, dans le cytosol associé aux métallothionéines ou dans les vacuoles associé aux phytochélatines et à des acides organiques (Clemens 2001; Kupper et al. 2009). Les parois et vacuoles seraient les principaux sites d'accumulation du cuivre (Arru et al. 2004; Lou et al. 2004). En sus de ces mécanismes, il a été montré que les plantes pouvaient répondre à ces agressions par un changement dans l'expression de leurs gènes qui conduit à la synthèse de protéines spécifiques (Cuypers et al. 2005), et il a été mis en évidence la présence de précipités de cuivre dans les différents compartiments cellulaires (Liu et Kottke 2004; Sahi et al. 2007; Manceau et al. 2008). Ces précipités se retrouvent tout d'abord dans les vacuoles et sur la paroi des cellules, puis à mesure que les quantités augmentent, ils se retrouvent dans les autres compartiments. Il se peut néanmoins que le cuivre ne reste pas stocké dans les racines, et il a été observé des accumulations dans les tiges (dans les tissus du xylème), les cellules des pousses, les épines et trichomes abaxiaux (Arru et al. 2004). Il semblerait que, contrairement aux racines, où Cu serait lié principalement à des composés soufrés, il serait préférentiellement lié à des ligands azotés/oxygénés comme les acides organiques dans les feuilles (Sahi et al. 2007 ; Kupper et al. 2009).

3.4) Mécanismes de toxicité

Quand la concentration en cuivre dans les organismes est trop importante, il interfère avec de nombreux processus physiologiques. Cette toxicité biochimique découle de la haute réactivité de cet élément. Il est capable de se lier fortement aux molécules riches en électrons, en raison de ses propriétés oxydoréductrices. Cette réactivité peut entrainer par exemple une dépolarisation accompagnée d'un dysfonctionnement physiologique des récepteurs et des transporteurs moléculaires cellulaires, ou encore sa fixation sur des sites actifs de macromolécules (ADN…) contenant des thiols (-SH), des carboxyles ($-COO^-$), ou des noyaux imidazoles ($C_3H_4N_2$) (Nriagu 1979; Alt et al. 1990). Ainsi, en se fixant aux groupements thiols des protéines des membranes cellulaires, il provoque la péroxydation des lipides insaturés (Devos et al. 1989). On peut également observer la production de radicaux libres suivant la réaction suivante (réaction de Fenton) : $Cu^+ + H_2O_2 \rightarrow Cu^{2+} + OH^* + OH^-$. La production excessive de ces radicaux libres peut entrainer des réactions d'oxydoréduction

en chaine (stress oxydant) qui aboutissent à la perte de l'intégrité cellulaire (Goldstein et Czapski 1986).

Lors de la réaction de photosynthèse, l'excès de cuivre perturbe le photosystème II, soit par insertion dans la Phéo (molécule de chlorophylle sans atome central de magnésium), soit par substitution des ions Mg^{2+} dans la chlorophylle elle-même (Cedeno-Maldonado et al. 1970; Kupper et al. 2009).

3.5) Effets macroscopiques d'un excès de cuivre

Les dégâts provoqués par un excès de cuivre sont non négligeables. Chez l'homme, par exemple, c'est notamment chez les personnels travaillant dans les vignobles où le cuivre est utilisé comme fongicide sous forme de sulfate que les effets d'une sur-absorption ont pu être observés: goût métallique, douleurs épigastriques, maux de tête, vomissements, diarrhées, insuffisance hépatique. Des insuffisances rénale et hépatique peuvent survenir de manière secondaire. Il a également pu être constaté des hémorragies gastro-intestinales (Pimentel et Menezes 1975; Suciu et al. 1981; O'Donohue et al. 1993; Komarek et al. 2010).

Chez les plantes, dans les parties aériennes, les effets d'une toxicité métallique sont sensiblement les mêmes que ceux de la déficience, i.e. nécrose (chlorose + enroulement) et dépérissement des jeunes pousses (Foy et al. 1978) qui résultent d'une inhibition du processus photosynthétique (Irtelli et al. 2009). Dans les parties souterraines, un excès de cuivre se traduit par la réduction du développement racinaire (Brams et Fiskell 1971; Daniels et al. 1972). Ces modifications morphologiques, sous-tendues par des altérations physiologiques, diminuent l'absorption des macronutriments et entrainent un retard de croissance.

Dans les sols et les eaux, il a été montré que le cuivre avait une influence sur le nombre et la biodiversité des organismes (Yamamoto et al. 1985; Komarek et al. 2010). Cette toxicité pour la microfaune et la microflore ralentit l'absorption et la biominéralisation des détritus organiques.

Qualitativement, on peut écrire que le cuivre est excessivement toxique pour le biote aquatique. Les limites de tolérance toxicologique des mammifères sont généralement 10 à 100 fois supérieures à celles des poissons et des invertébrés marins, et environ 1000 fois supérieures à celles de certaines algues (Flemming et Trevors 1989). L'objet des paragraphes suivants est d'apporter une information plus quantitative sur les seuils de toxicité pour les organismes vivants.

3.6) Seuils de toxicité

3.6.1) Micro-organismes aquatiques et terrestres (protozoaires, champignons, bactéries, virus)

Le cuivre est particulièrement toxique pour les micro-organismes. Dans les eaux douces naturelles, l'ajout de $CuSO_4$ à des doses infimes (0,001 $mg_{Cu}.L^{-1}$) perturbe les processus de biominéralisation (Albright et al. 1972). Sur un sol limoneux amandé avec des boues d'épuration, l'émission de CO_2 est perturbée à des seuils inferieurs à 10 $mg_{Cu}/kg_{sol\ sec}$ (Chang et Broadbent 1981). Utilisé dans l'eau potable, il entraine, à des doses variant de 0,025 à 0,05 $mg_{Cu}.L^{-1}$, la mort des bactéries coliformes (Domek et al. 1984).

La phytotoxicité du cuivre varie cependant énormément d'une espèce à l'autre. Ainsi l'addition de sulfate de cuivre (de 0,005 à 0,01 $mg_{Cu}.L^{-1}$) réduit la biomasse alguaire en favorisant les chlorophycées vis-à-vis des cyanobactéries (Elder et Horne 1978; Effler et al. 1980), et dans un sol de rizière, l'addition de cuivre (1600 $mg_{Cu}/kg_{sol\ sec}$) provoque l'augmentation de la population des champignons tout en diminuant la biodiversité (Yamamoto et al. 1985).

Cette phytotoxicité dépend également du milieu. Il est nécessaire d'ajouter 1000 $mg_{Cu}/kg_{sol\ sec}$ à un sol riche en matière organique (tourbe) contre seulement 55 à un sol sableux pour observer une diminution dans la décomposition de l'acide glutamique (Haanstra et Doelman 1984).

3.6.2) Organismes aquatiques

3.6.2.1) Invertébrés aquatiques

La sensibilité au cuivre des invertébrés aquatiques est également très variable (Fig. 1.5). Pour les invertébrés d'eaux douces, les concentrations létales pour 50% de la population après 48h d'exposition (48h LC_{50}) varient de 5 $\mu g_{Cu}.L^{-1}$ chez certains crustacés (*Daphnia sp*.) à 6 400 $\mu g_{Cu}.L^{-1}$ chez certains arthropodes (larves de phryganes) (Hodson 1979). Pour les invertébrés d'eaux de mer, les 96h LC_{50} mesurées sont comprises entre 29 $\mu g_{Cu}.L^{-1}$ pour les pétoncles et 9400 $\mu g_{Cu}.L^{-1}$ pour le crabe violoniste (Devi 1987; Nelson et al. 1988). Dans un même phylum, les mollusques, et une même classe, les gastéropodes, ces concentrations varient selon les genres d'animaux puisqu'avec du sulfate de cuivre, la 96h LC_{50} est de 39 $\mu g_{Cu}.L^{-1}$ pour *Physia integra*, contre 1700 $\mu g_{Cu}.L^{-1}$ pour *Campeloma decisum* (Arthur et Leonard 1970).

Elle varie également avec les paramètres physico-chimiques du milieu ainsi que la biologie de l'animal. Chez les annélides, dans la classe des clitella, pour un nitrate de cuivre, les 96h LC_{50} d'*Oligochaete Lumbriculus variegatus* varient de 130 $\mu g_{Cu}.L^{-1}$ à pH 6,6 à 500 $\mu g_{Cu}.L^{-1}$ à 8.3 (Schubauer-Berigan et al. 1993). Chez les arthropodes, dans la classe des branchiopodes, la 48h CL_{50} de *Ceriodaphnia dubia* augmente de 35 à 79 $\mu g_{Cu}.L^{-1}$ quand la dureté de l'eau passe de 94 à 170 mg de $CaCO_3$ (Belanger et al. 1989). Comme autres facteurs influençant la toxicité du cuivre, on peut citer la température, la salinité, l'âge de l'animal, etc…, en plus bien entendu de la forme du cuivre utilisée pour réaliser les tests écotoxicologiques (Gauss et al. 1985; Snell et al. 1991; Ozoh 1992).

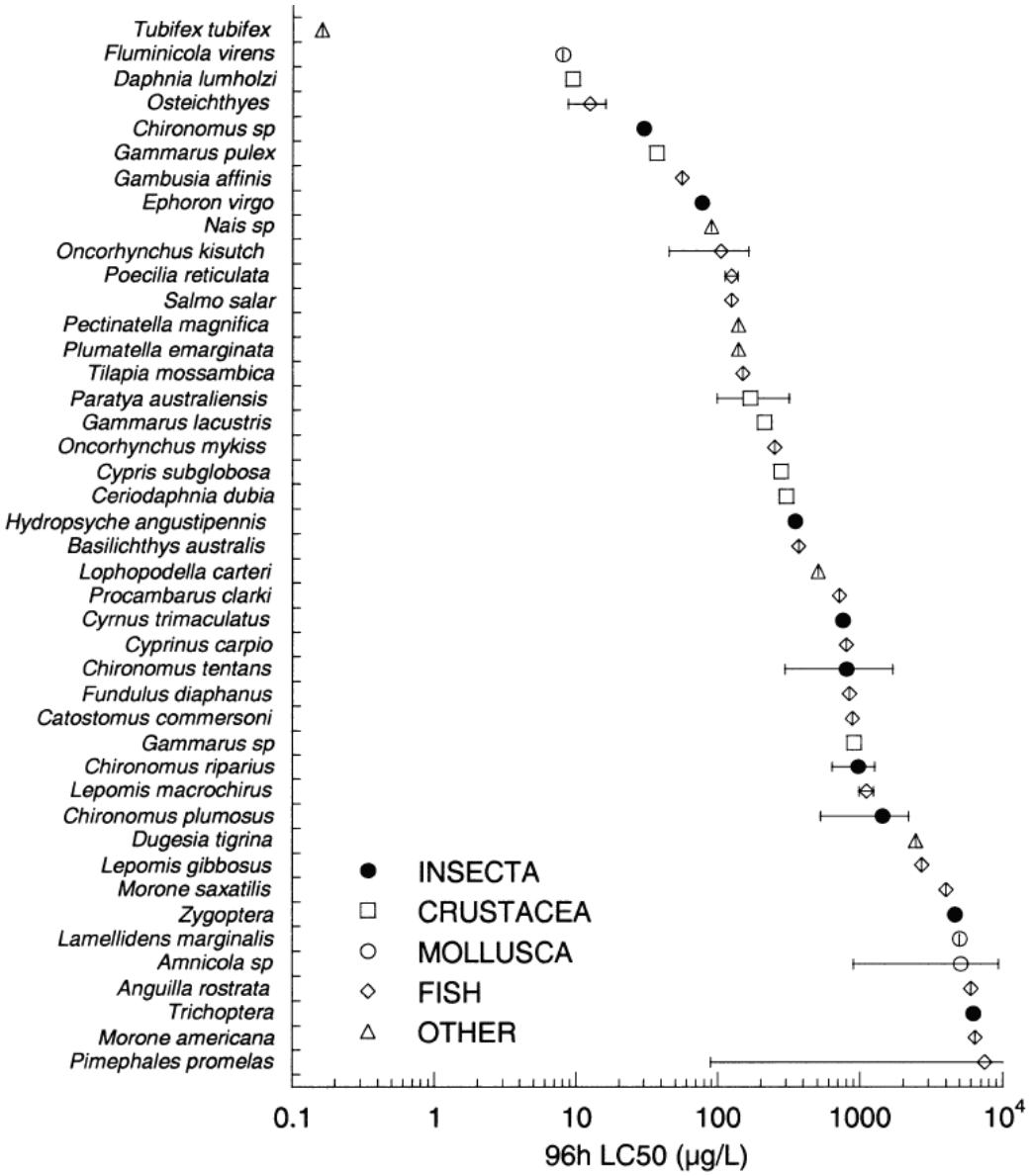


Figure 1.5: Variabilité de la toxicité aigue du cuivre chez les organismes aquatiques (Van der Geest et al. 2000).

3.6.2.2) Vertébrés aquatiques

De la même manière, les vertébrés aquatiques présentent une sensibilité variable au cuivre (Fig. 1.5). Elle dépend en particulier de la surface des branchies (et donc de l'âge), du taux respiratoire, du courant, etc… Plus ces paramètres sont élevés, plus le cuivre est absorbé (Hodson 1979). Les études de laboratoire menées avec différents sels de cuivre (sulfate, chlorure et nitrate) ont montré que dans les eaux douces, les 96h LC_{50} variaient de moins de 10 $\mu g_{Cu}.L^{-1}$ pour l'omble arctique à plus de 10000 $\mu g_{Cu}.L^{-1}$ pour le cyprin doré (Mayer et Ellersieck 1986; Buhl et Hamilton 1990). Des effets non létaux (retard de croissance, reprotoxicité) ont été observés chez la truite saumonée à partir de 30 $\mu g_{Cu}.L^{-1}$. Dans les eaux salées, les seuils de toxicité varient de 60 $\mu g_{Cu}.L^{-1}$ (96h LC_{50}) chez le saumon royal à 19000 (48h LC_{50}) pour le choquemort *Fundulus heteroclitus* (McKim et Benoit 1971; Burton et Fisher 1990; Hamilton et Buhl 1990). Parmi les différents facteurs modulant cette toxicité, on retrouve le pH, la salinité, la température, la dureté, l'alcalinité, la présence de MO et l'âge des individus (McNulty et al. 1994; Erickson et al. 1996).

3.6.2.3) Plantes aquatiques

En milieu de culture, pour un même sel de cuivre (chlorure de cuivre II), au bout de 96h, 50 % de la concentration effective (96h EC_{50}) des populations d'algues qui présentent des retards de croissance sont observés pour des quantités de cuivre comprises entre 50 (*Chlamydomonas bullosa*) et 500 (*Dunaliella minuta*) $\mu g_{Cu}.L^{-1}$. Cette variabilité se retrouve également avec le sulfate de cuivre II. (Visviki et Rachlin 1991; Schafer et al. 1994; Visviki et Rachlin 1994).

3.6.3) Organismes terrestres

3.6.3.1) Invertébrés terrestres

Des effets toxiques ont été observés en laboratoire chez les lombrics. Sur 56 jours, jusqu'à environ 30 $mg_{Cu}/kg_{sol\ sec}$, aucun effet n'a été observé, tandis que l'EC_{50} de la production des cocons se situe vers 60 $mg_{Cu}/kg_{sol\ sec}$ (Spurgeon et al. 1994). Sur 14 jours, à partir de 500 $mg_{Cu}/kg_{sol\ sec}$, des réductions de taille ont été observées sur des individus adultes et une concentration de 1000 $mg_{Cu}/kg_{sol\ sec}$ est létale (Martin 1986).

Cette toxicité dépend à nouveau de la spéciation du cuivre, les sels (sulfates et chlorates) étant beaucoup plus toxiques que les oxydes en raison de leur plus grande solubilité (Malecki et al. 1982). Elle dépend également du type de sol, et en particulier de la teneur en

matière organique du fait de la grande affinité du cuivre pour celle-ci : le cuivre est plus bio-disponible dans les sols pauvres en MO que dans les sols riches (Streit 1984; Donkin et Dusenbery 1993).

3.6.3.2) Vertébrés terrestres

De nombreux animaux (e.g. lapins, cochons…) peuvent tolérer de très fortes concentrations (ex. 300 à 800 $\mu g_{Cu}.g^{-1}_{\text{matière sèche de nourriture}}$) dans leur régime alimentaire sans présenter de signe de toxicité (Flemming et Trevors 1989). Les doses létales chez les animaux (LD_{50}) sont comprises entre 10 et 1000 $mg_{Cu}/kg_{\text{poids corporel}}$. Ces doses sont fonction de l'espèce et du sel de cuivre utilisé (WHO 1998). Ce sont les sels de cuivre II, agents particulièrement irritants, qui sont généralement responsables des cas d'intoxication. Les formes les plus solubles, sulfate de cuivre II et chlorure de cuivre II, ont de nouveau une plus grande toxicité que les formes peu solubles, comme l'hydroxyde de cuivre II et l'oxyde de cuivre I (Tableau 1.3).

3.6.3.3) Plantes terrestres

Comme précédemment, les effets varient avec l'espèce végétale et le type de sol. Ainsi, des effets sur le développement racinaire ont été observés avec du ray-grass anglais aux alentours de 1000 $mg_{Cu}/kg_{\text{sol sec}}$ sur un sol argilo-sableux (Jarvis 1978) alors que les rendements de haricots mange-tout sur un sable limoneux ont diminué dès 15 $mg_{Cu}/kg_{\text{sol sec}}$ extraits avec 0,1 M d'EDTA, soit 90% de la teneur totale en cuivre (Walsh et al. 1972; Lo et Yang 1999).

3.7) Conclusions sur la toxicité du cuivre

C'est dans le monde végétal, aquatique et pour les micro-organismes que le cuivre présente une toxicité véritablement remarquable. L'utilisation de cuivre à des fins pesticides (purification de l'eau, bouille bordelaise…) a conduit a un enrichissement spectaculaire de certains milieux (Pietrzak et McPhail 2004). Cette partie nous a permis de montrer que la toxicité du cuivre ne dépendait pas uniquement de sa concentration totale dans le milieu, mais que les variables physico-chimiques et biologiques devaient être intégrées dans l'estimation de l'impact réel d'une contamination sur le vivant. Ce sont principalement la forme solvatée $Cu(H_2O)_n^{2+}$ et les petits complexes organiques capables de traverser les membranes cellulaires qui sont les plus biodisponibles. Tout facteur capable de modifier l'abondance de ces espèces

labiles est susceptible d'augmenter ou de diminuer la toxicité du cuivre. Le pH, le potentiel d'oxydoréduction (Eh), et la composition et la concentration des solutés (i.e. l'activité de tous les ions dans la phase aqueuse) sont de ce point de vue des paramètres déterminants, sur lesquels il est possible d'intervenir pour atténuer l'effet d'une contamination.

Tableau 1.3 : Toxicité aigue par voie orale de différents sels de cuive

Sel de cuivre	Solubilité (g/L à 25°C)	Espèce	Dose Létale 50% ($mg_{Cu}/kg_{poids\ corporel}$)*
Acétate de cuivre (II) monohydrate $Cu(CH_3COO)_2$		rat	226
		souris	509
Carbonate de cuivre (II) $Cu(CO_3^-)$		rat	82
		souris	165
Carbonate de cuivre (II) hydroxydé $Cu_3(CO_3)_2(OH)_2$		rat	409
		lapin	91
Chlorure de cuivre (II) $CuCl_2$	706	rat	66
		souris	90
		cochon d'inde	15
Hydroxyde de cuivre (II) $Cu(OH)_2$	$2{,}9\ 10^{-3}$	rat	651
Nitrate de cuivre (II) trihydraté $Cu(NO_3)_2$		rat	247
Oxyde de cuivre (I) Cu_2O	Insoluble	rat	417
Oxychlorure de cuivre $(Cu(OH)_2)_3CuCl_2$		rat	417-476
Sulfate de cuivre $CuSO_4$	143 (0°C)	rat	244
		souris	20
		lapin	50

*(WHO 1998)

4) Quantités et formes du cuivre dans l'environnement

Le cuivre est présent dans tous les milieux : eaux, air, sols et sédiments (Tableau 1.4). C'est le 26ème élément le plus abondant sur la Terre et sa concentration dans la croute est estimée à 70 mg .kg^{-1}. Il s'associe à la matière organique, aux oxydes de fer et de manganèse, et aux argiles, pour l'essentiel (Alloway 1995). C'est un des éléments trace les moins mobiles qui tend à rester « fixé» du fait principalement de sa forte affinité pour la matière organique vivante ou en décomposition, dont la quantité est en très large excès dans la biosphère.

Tableau 1.4 : Synthèse des différentes concentrations ubiquitaires du cuivre dans les fonds géochimiques

Milieu		Concentration
Air		5-50 ng/m^3
Eaux	Eaux de mer	0,05-12 $\mu g.L^{-1}$
	Eaux douces	5-30 $\mu g.L^{-1}$
Sols	Sols	2-250 mg .kg^{-1}
	Sédiments	2-5000 mg .kg^{-1}
	Roches volcaniques	13-90 mg .kg^{-1}
	Roches sédimentaires	5-39 mg .kg^{-1}
	Croute terrestre	24-70 mg .kg^{-1}

(Bowen 1969; Forstner et Wittman 1979; Meranger et al. 1979; Nriagu 1979; Flemming et Trevors 1989; Abollino et al. 1995; Alloway 1995)

4.1) Dans les roches

Naturellement, Cu est présent en quantité plus importante dans les basaltes que dans les granites (90 mg .kg^{-1} vs 15 mg .kg^{-1} en moyenne), et se trouve en très faible quantité dans les carbonates (5-20 mg .kg^{-1}). La plus grande abondance du cuivre dans les roches ignées basiques est contrôlée en partie par les processus de différentiation magmatique durant la cristallisation. Dans la chambre magmatique, les silicates de magnésium suivis des minéraux ferreux et calciques cristallisent en premier, laissant un mélange alcalin peu miscible, riche en aluminium, silicates et quartz. A mesure que la cristallisation avance, des circulations d'eau chargées en sulfures détritiques (provenant par exemple de l'activité hydrothermale) saturent le liquide en soufre qui précipite sous forme de sulfures métalliques, dont la covellite (CuS), la chalcocite (Cu_2S), la chalcopyrite ($CuFeS_2$) et la bornite (Cu_5FeS_4) (Alloway 1995; Thomas 2010). Ce sont les formes principales des gisements cuprifères. L'oxydation secondaires des sulfure conduit à la formation de carbonates comme la malachite ($Cu_2[(OH)_2(CO_3)]$) et l'azurite ($Cu_3[(OH)(CO_3)]$), et d'oxydes comme la ténorite (CuO) (Thomas 2010). Il existe également des sulfosels comme l'énargite (Cu_3AsS_4) et la tennantite ($(Cu,Fe)_{12}As_4S_{13}$), mais ces espèces minérales sont peu fréquentes.

4.2) Dans l'atmosphère

Le cuivre n'existant pas à l'état gazeux dans les conditions physico-chimiques de la Terre, il est forcément associé aux aérosols. Ces poussières peuvent être générées par l'activité volcanique, la dégradation de la matière organique (ex. feux de forêts), et les embruns marins, cependant les émissions de cuivre sont essentiellement d'origine anthropique (Nriagu 1989). On peut citer à titre d'exemple les fours de fusion des fonderies et les centrales thermiques. Le cuivre est libéré dans l'atmosphère en association avec des particules de matière sous forme d'oxyde, de sulfate, de carbonate ou adsorbé à la MO. Il s'élimine par dépôt gravitaire à sec et entrainement par les précipitations. Sa concentration dans l'atmosphère non polluée est de l'ordre de la dizaine de nano-grammes par mètre cube (Tableau 1. 4). Dans une zone contaminée, elle est grandement liée à la proximité de points d'émission (Kim et Fergusson 1994).

4.3) Dans les eaux

Le cuivre est introduit dans les eaux de manière passive, par lessivage des sols, boues de station d'épuration et déchets divers, et active, par introduction de fongicide cuprique. Il est concentré principalement dans les fractions particulaires et colloïdales, le plus souvent à l'état divalent, mais occasionnellement sous forme monovalente dans les solutions et effluents riches en sulfures, cyanures et fluorures. Sa concentration dans les eaux naturelles est de l'ordre de la dizaine de $\mu g.L^{-1}$ (Tableau 1.4), celle-ci étant plus élevée dans les eaux salées que dans les eaux douces (Flemming et Trevors 1989).

A pH basique et forte concentration, le cuivre peut précipiter (Fig. 1.6a). Les formes précipitées prépondérantes sont l'hydroxyde de cuivre ($Cu(OH)_2$), la ténorite (CuO), la malachite ($Cu_2[(OH)_2(CO_3)]$) et l'azurite ($Cu_3[(OH)(CO_3)]$) (Stumm et Morgan 1996). Lorsqu'ils sont présents, comme c'est souvent le cas dans les eaux contaminées et les effluents industriels, les phosphates peuvent également précipiter avec le cuivre, et dans les milieux anaérobiques le cuivre précipite sous forme de sulfures.

Dans les eaux naturelles, il est rarement libre mais plutôt complexé aux molécules organiques en suspension ou adsorbé sur la surface des particules minérales (Fig. 1.6b). Cette affinité pour la MO s'explique, d'une part, par la déprotonation à pH acide des groupements carboxyliques et des fonctions alcools adjacentes portées sur un carbone aliphatique en position α et, d'autre part, par l'excellent accord stérique entre le chélate COOH-COH-R et la distance $Cu\text{-}O_{eq}$ de 1.97 Å (Manceau et Matynia 2010).

4.4) Dans les sols et sédiments

Dans les sols non contaminés, les concentrations en Cu sont de l'ordre de 2 à 250 mg/kg avec une moyenne de 30 mg/kg, dans les sédiments d'eaux douces elles varient typiquement de 16 à 5000 mg/kg et dans les sédiments marins de 2 à 740 mg/kg (Tableau 1.4). Les pollutions au cuivre proviennent surtout des exploitations minières (Luo et al. 2006), de l'agriculture (Brun et al. 2001) et des boues des stations d'épuration (Ashworth et Alloway 2007).

Les sols étant constitués en grande majorité de matière organique, d'argiles (phyllosilicates) et d'oxy(hydroxy)des de fer et de manganèse, qui tous possèdent des sites réactionnels donneurs d'électrons, comme les groupements COO^-, OH^-, capables de créer des liaisons fortes avec l'ion Cu^{2+} en comblant la lacune électronique dans la couche d^9. Le cuivre peut s'associer préférentiellement aux oxydes de manganèse et à la matière organique, comme le suggère la dépendance en pH des isothermes d'adsorption de Cu sur chacune de ces phases (Fig. 1.6b). Cependant, si le sol est argileux ou argilo-limoneux et pauvre en MO, la majorité du cuivre pourra très bien être fixée sur les argiles en raison de leur abondance. Les mécanismes de rétention du cuivre sur ces constituants essentiels sont détaillés dans le chapitre suivant.

a) b)

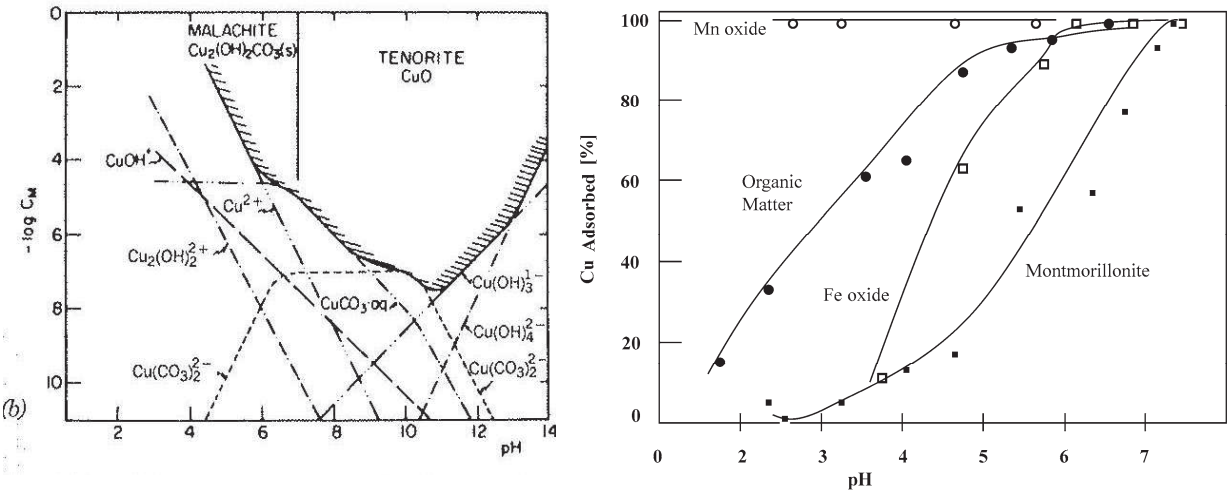


Figure 1.6 : a) Diagramme de solubilité du cuivre, $[Cu]_{Tot.}$= $[Carbonate]_{Tot.}$= 0,002 M (Stumm et Morgan 1996). b) Complexation (MO) et adsorption des ions Cu^{2+} sur les différents constituants du sol en fonction du pH. Les oxydes de manganèse et la matière organique ont une très grande affinité pour le cuivre (Bradl 2004).

Chapitre 2

Mécanismes de rétention du cuivre par les constituants des sols

1) Introduction

Les éléments chimiques des sols et des sédiments sont partagés entre la phase solide, et les phases liquides et gazeuses du milieu poral. Le cuivre n'étant pas volatile, son transport s'effectue dans la seule phase liquide, sous forme dissoute, colloïdale ou éventuellement sous forme particulaire à l'état solide. La fraction dissoute est la plus labile et biodisponible. La phase solide est un assemblage complexe et hétérogène de minéraux, de petites molécules organiques et de composés organiques plus ou moins polymérisés issus de la matière vivante. Comme cela a été dit au chapitre précédent, les minéraux proviennent essentiellement de l'altération des roches (on parle de minéraux secondaires) ou ont été introduits dans l'environnement par les activités humaines, voire formés par des organismes vivants (biominéralisation). Ces constituants ont le plus souvent une taille nanométrique, et donc une très grande surface spécifique. Ils sont rarement isolés les uns des autres, autrement le sol n'aurait pas de cohésion, mais consolidés sous la forme d'agrégats nanoporeux organo-minéraux. Du fait de leurs propriétés amphotériques, les groupements fonctionnels de surface, portés par les constituants organiques et inorganiques, ont une charge électrostatique nette positive ou négative qui varie avec la composition de la solution, notamment l'activité des protons (pH). Comme le montrent les isothermes d'adsorption de la figure 1.6b, l'ion Cu^{2+} peut être retenu par au moins un type de composant quelles que soient les conditions physico-chimiques du milieu. Au cours du temps les espèces labiles se fixent durablement, généralement par précipitation ou incorporation dans le réseau cristallin des minéraux (Han et al. 2001; Manceau et al. 2002).

2) Rappels sur les différents mécanismes de fixation

D'un point de vue mécanistique, on distingue classiquement : la complexation de sphère externe (adsorption non spécifique), la complexation de sphère interne (adsorption spécifique), la complexation multinucléaire de surface, la substitution chimique isomorphique au sein de la structure, la précipitation homogène et la coprécipitation en phase aqueuse (Manceau et al. 2002). Plusieurs mécanismes de rétention peuvent coexister au sein d'un même système, avec une importance relative qui dépend des conditions physico-chimiques du milieu et de la nature des constituants. La complexité des sols fait qu'il est généralement difficile de distinguer ces mécanismes, d'autant que l'on ne détecte que les états finaux et non les processus réactionnels eux-mêmes. C'est pourquoi le terme «sorption » est souvent employé dans la littérature anglo-saxonne pour décrire la séquestration des ions à l'interface liquide / solide (Xu et al. 1994; Hettiarachchi et al. 2000; Panfili et al. 2005; Kirpichtchikova et al. 2006). Néanmoins, on distingue souvent l'*ad*sorption sur une surface (2 dimensions) de l'*ab*sorption dans un espace à trois dimensions (Figs. 2.1 et 2.2).

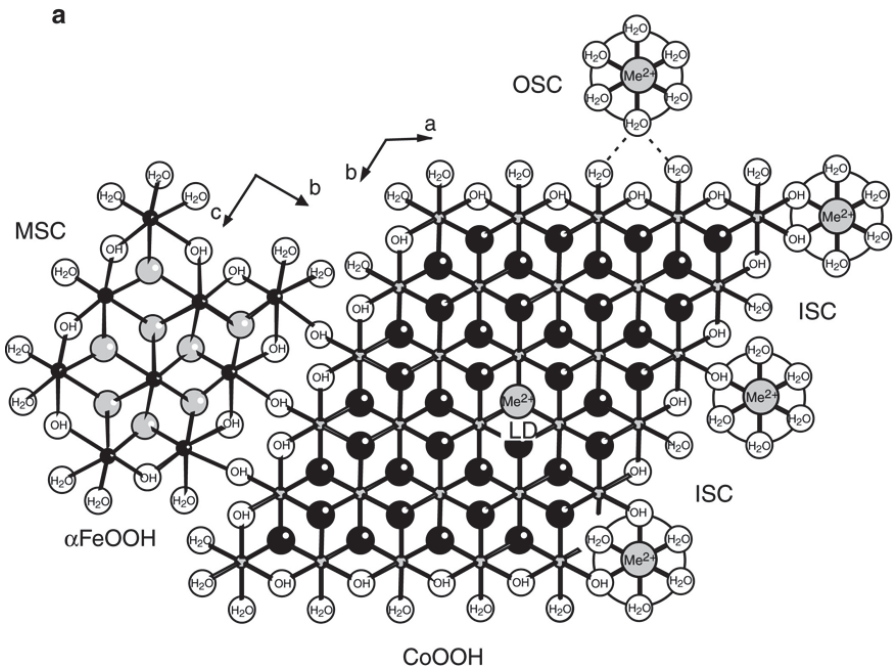


Figure 2.1: Représentation schématique de quatre des cinq principaux processus de sorption d'un élément chimique. Projection dans le plan *ab* de la structure CoOOH. La complexation de surface multinucléaire (CSM) est représentée par le processus de croissance cristalline de la goethite précipitée en condition hétérogène. Trois types de complexation de sphère interne (CSI) sont projetés : monodentée à un cation de la surface (milieu droit), bidentée à un cation de la surface (haut droit), bidentée à deux cations de la surface (bas droit). En haut, représentation de la complexation de sphère externe (CSE) et au centre de la diffusion dans le réseau cristallin (LD) (Manceau et al. 2002).

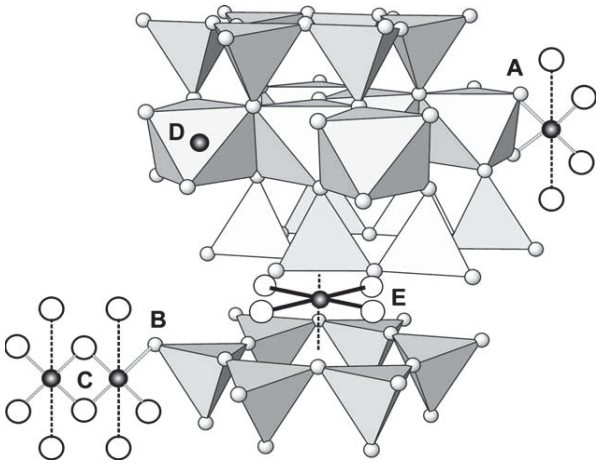


Figure 2.2 : Illustration des sites de sorption possibles du cuivre sur une argile. A et B sont des exemples de coordination de sphère interne, C est un complexe de surface multinucléaire, D illustre le phénomène d'occupation (LD) des sites octaédriques vacants et E est un complexe de sphère externe dans l'espace interfoliaire (Furnare et al. 2005).

2.1) Complexation de sphère externe (CSE)

Dans ce mécanisme de rétention, l'ion adsorbé conserve sa sphère d'hydratation et est retenu sur la surface chargée par des interactions électrostatiques. Il s'agit d'une physisorption également qualifiée de non-spécifique. Ce phénomène peut se produire à la surface des minéraux et des molécules organiques, mais il survient surtout dans l'espace interfoliaire des argiles 2/1 (Fig. 2.2). Cet espace est situé dans un champ électrostatique permanent, donc indépendante du pH, qui résulte de substitutions isomorphiques par des cations de charge inférieure à celle des cations du réseau cristallin (e.g. substitution de Al^{3+} ou Fe^{3+} par Mg^{2+} dans la couche octaédrique des smectites dioctaédriques et substitution de Si^{4+} par Al^{3+} ou Fe^{3+} dans la couche tétraédrique de nombreux phyllosilicates). Récemment, à la surface de la muscovite, phyllosilicate 2/1 du groupe des micas, une distinction a été faite selon que l'ion hydraté est au contact de la surface ou bien séparée de celle-ci par une couche d'eau (Lee et al. 2010). Dans le premier cas, la distance entre l'ion adsorbé et la surface est de l'ordre de 5 à 10 Å contre 4 à 5 Å dans le second (Fig. 2.3).

Ce type de liaison est faible et les ions retenus de la sorte sont facilement échangeables, par exemple en augmentant la force ionique ou la charge électrique des ions du soluté (e.g., Na^{+} vs. Ca^{2+} ou Cu^{2+}). Dans les sols acides, la fraction des métaux retenus de cette façon peut être d'une dizaine de pourcent mais, en général, elle est de l'ordre de quelques pourcents, au plus (Manceau et al. 2002; Scheinost et al. 2002).

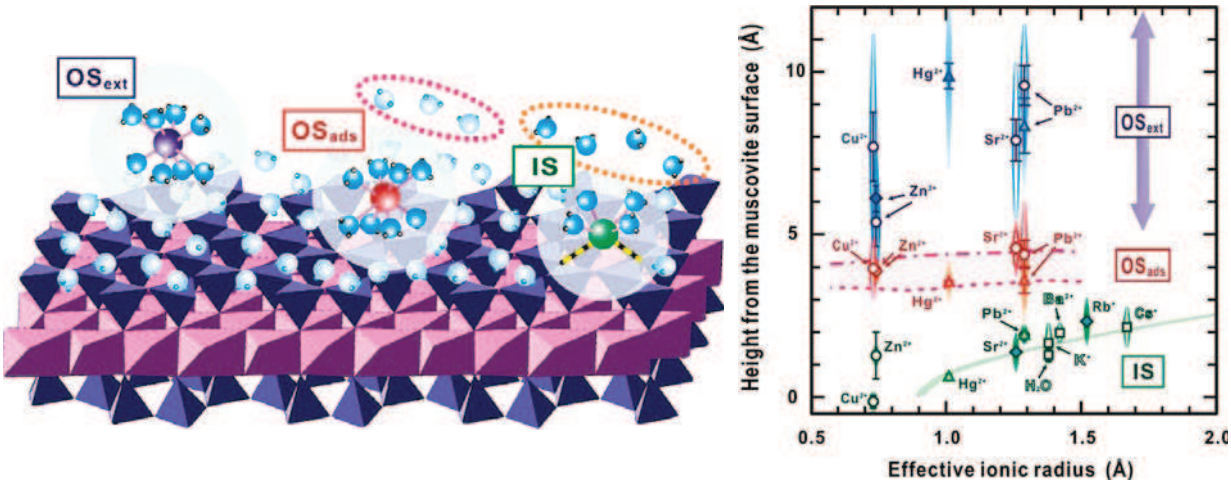


Figure 2.3 : A gauche, représentation schématique des spéciations possibles des cations à l'interface muscovite (00l)-solution. Dans les phénomènes de complexation de sphères externes, le cation adsorbé, par définition, garde sa sphère d'hydratation, mais il est fait une différence selon que la surface conserve sa couche d'hydratation (CSE_{ext} écrite OS_{ext} sur la figure) ou non (CSE_{abs} écrite OS_{ads}). Lors d'une complexation de sphère interne (IS), le cation perd une partie de sa sphère d'hydratation pour se lier aux atomes oxygènes de la surface. Les distances cations-surface (à droite) varient en fonction du type de complexation et du rayon ionique des cations (Lee et al. 2010).

2.2) Complexation de sphère interne (CSI)

Dans ce mécanisme, les ions perdent une partie de leurs molécules d'eau d'hydratation (Figs. 2.1, 2 .2 et 2.3). S'ils se lient aux liaisons pendantes des ligands de surface, généralement des oxygènes, il y a établissement d'une liaison chimique et on parlera de chimisorption. Ces liaisons pendantes résultent dans le cas d'un solide inorganique de la rupture du réseau cristallin en surface. Le polyèdre de coordination de l'adatome, un octaèdre ou un tétraèdre dans la plupart des cas, peut partager un ligand (complexe monodenté), deux ligands (complexe bidenté) ou trois ligands (complexe tridenté) avec les polyèdres de surface de la phase adsorbante. Si on adopte une description structurale du phénomène, on parlera de partage d'un sommet, d'une arête ou d'une face entre les deux entités polyédriques (Manceau et al. 2002). Ces structures géométriques sont représentées sur la figure 2.4.

Les cations alcalins dont l'enthalpie d'hydratation est faible perdent facilement leur sphère d'hydratation et s'adsorbent au contact des oxygènes de surface sans pour autant créer de liaison chimique (physisorption). C'est le cas du potassium (ΔH_{hydr} = -322 kJ/mol) dans l'espace interfoliaire des phyllosilicates (Fig. 2.3). Comme la force électrostatique est inversement proportionnelle au carré de la distance qui sépare les charges, cet adatome en sphère interne est également très stable malgré l'absence de liaison chimique (Sposito 2008).

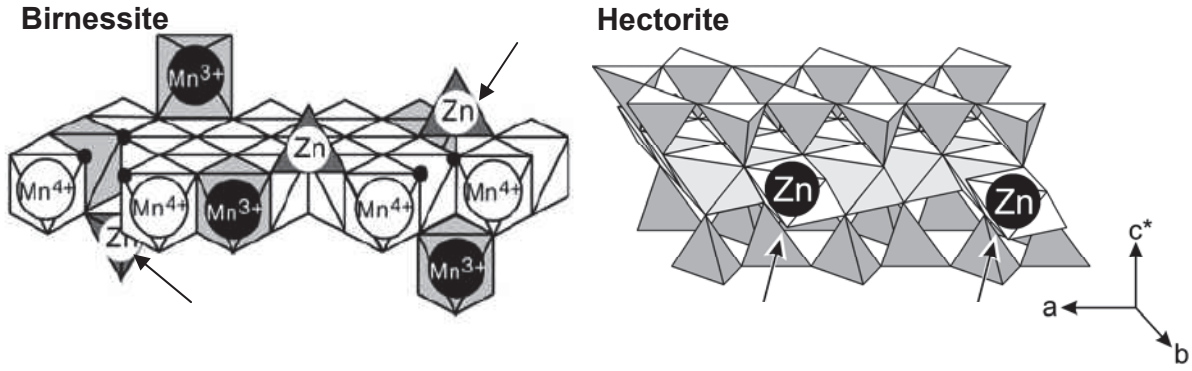


Figure 2.4 : Illustrations des partages polyédriques. Dans le cas de la birnessite hexagonale, phyllomanganate de formule type $Mn^{2+}_{0.05}Mn^{3+}_{0.12}(Mn^{4+}_{0.74}Mn^{3+}_{0.10}\square_{0.17})O_{1.7}(OH)_{0.3}$, le zinc adsorbé est ici tetracoordonné et lié à trois atomes d'oxygène de surface à l'aplomb d'une lacune octaédrique (Manceau et al. 2002). Dans l'hectorite, smectite trioctaédrique de formule générale $R_{0.33}^{+}(Mg_{2.67}Li_{0.33})Si_4O_{10}(OH)_2$, le zinc peut, sous certaines conditions, s'adsorber sous forme octaédrique sur les cotés (ici 010) de la couche (Mg,Li), elle-même formée d'octaèdres. Il y a alors partage d'arêtes entre les octaèdres de surface et celui de l'adatome (Schlegel et al. 2001b).

2.3) Complexation multinucléaire de surface (CMS) et précipitation de surface

Lorsque les adatomes ne sont pas isolés mais leurs polyèdres condensés en surface, on parle de complexation multinucléaire. Deux cas peuvent se présenter selon qu'il y a continuité du réseau des anions, comme dans une épitaxie (Fig. 2.5), ou que les adatomes possèdent leur propre couche d'anion différenciée de celle de la surface, mais néanmoins orientée cristallographiquement par des liaisons hydrogène ou des forces de Van der Waals. Les empilements compacts de ces deux cas de figure peuvent être représentés ainsi : AcBcAcBcAcBaC, pour le premier, et AcBcAcB-AcBaC pour le second. Les complexes multinucléaires sont les nuclei d'une nouvelle phase bi ou tridimensionnelle qui croitra si la solution est sursaturée vis-à-vis de cette nouvelle phase. La nucléation hétérogène est probablement le processus majeur de formation des minéraux dans les sols (Manceau et al. 2002). Les précipités de surface sont généralement peu solubles et donc assez stables en termes de remobilisation des éléments traces.

2.4) Précipitation homogène et coprécipitation

La précipitation homogène en solution se produit idéalement lorsque le produit de l'activité des espèces ioniques excède le produit de solubilité de la phase solide. En effet, thermodynamiquement, si la création d'un cristal libère de l'énergie, la création d'une interface solide-liquide en nécessite. Dans le cadre de la théorie de la nucléation classique homogène, pour que la germination ait lieu, il faut que la concentration des ions soit

suffisamment importante dans des conditions de pression et de température données, pour que l'énergie libérée par la mise en ordre des atomes soit supérieure à la tension superficielle. Ces précipités peuvent ensuite se déposer physiquement sur les grains ou les parois des pores de la matrice au cours des cycles d'humectation-déshydratation. Les dépôts en surface des vides poraux sont parfois dénommés 'cutanes'. Il est assez rare qu'un élément trace précipite, ou alors la contamination est telle que c'est un élément majeur. Le plus souvent il coprécipite avec un élément majeur, comme Mg ou Al pour former une solution solide dont la solubilité est toujours plus faible que celle des pôles purs. De cette manière, les éléments traces sont fréquemment incorporés dans le réseau cristallin d'une phase précipitée. C'est le cas par exemple d'argiles néoformées cuprifères. Cette forme d'immobilisation est durable en dehors de changement physico-chimique du milieu.

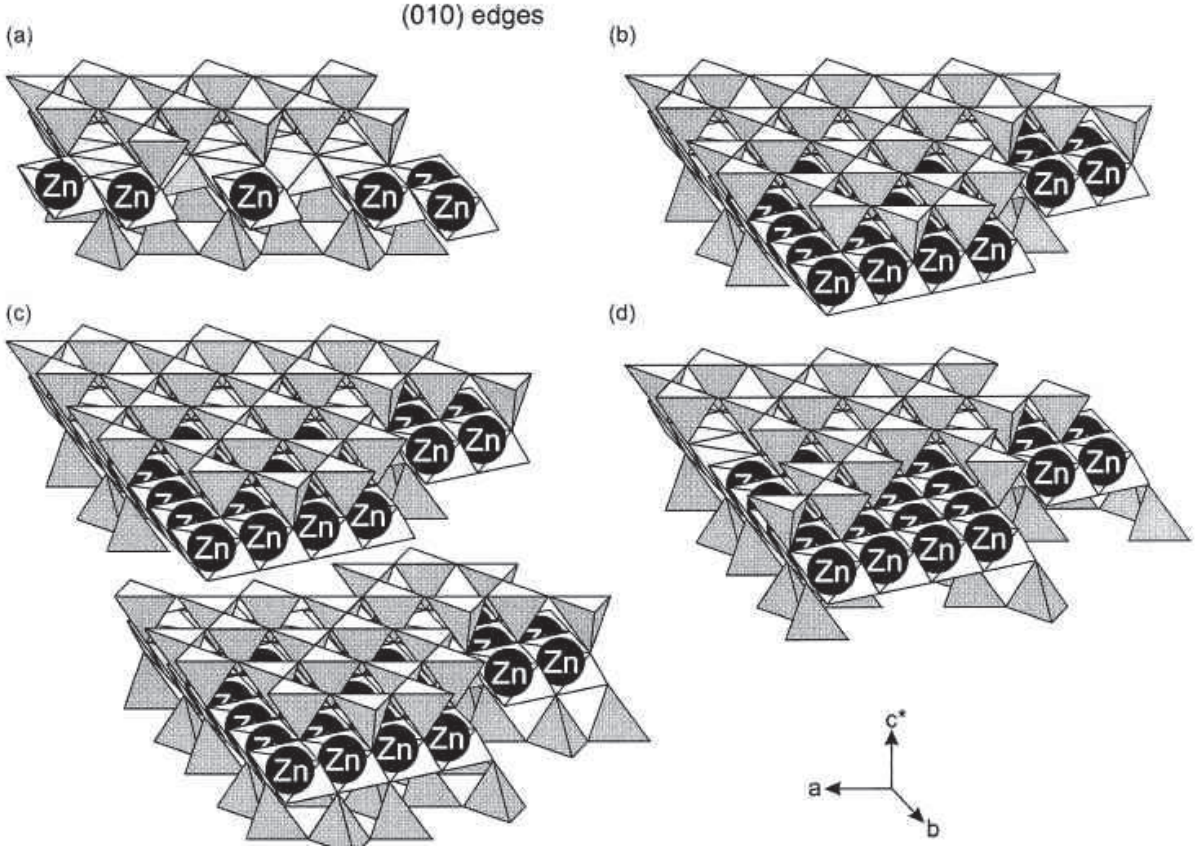


Figure 2.5 : Modèles structuraux de sorption du zinc sur les bords de l'hectorite dans une solution (I = 0.3M $NaNO_3$) saturée en zinc (520 μM) en fonction de la concentration en silicium de la solution $[Si]_{aq}$ et du temps. a) $[Si]_{aq}$ = 30-60 μM, t = 96h, des petits polymères de zinc se forment sur les bords en continuité des octaèdres (Mg,Li). b,c,d) $[Si]_{aq}$ = 530 μM, t = 9h, les cations Zn forment une couche octaédrique en continuité de la couche (Mg,Li) et sont liés aux feuillets de silicium tétraédriques suivant des empilements TO (b), TO et TOT (c) ou uniquement TOT (d) avec de gros défauts dans les feuillets tétraédriques (Schlegel et al. 2001a).

2.5) Diffusion dans le réseau cristallin

Un cation chimisorbé, voire physisorbé (cas du lithium), peut diffuser dans la structure cristalline et combler une lacune cationique (Fig. 2.6), voire remplacer un cation du substrat (Manceau et al. 2002). Ce mécanisme dépend de l'électronégativité, et donc du rapport charge/rayon de l'élément incorporé, de sa taille par rapport à celle du réseau cristallin, ainsi que de la structure de la phase hôte (densité et accessibilité des lacunes). Généralement, la différence entre les rayons ioniques des éléments substituant et substitué ne doit pas dépasser environ 15 %, et la charge ne doit pas différer de plus d'une unité (Alloway 1995). Au final, la substitution isomorphique d'un élément trace dans une phase solide infinie, qu'elle s'opère par coprécipitation homogène ou hétérogène ou bien par incorporation dans un minéral préexistant, est le mécanisme principal d'atténuation de la contamination au cours du temps (Vespa et al. 2010).

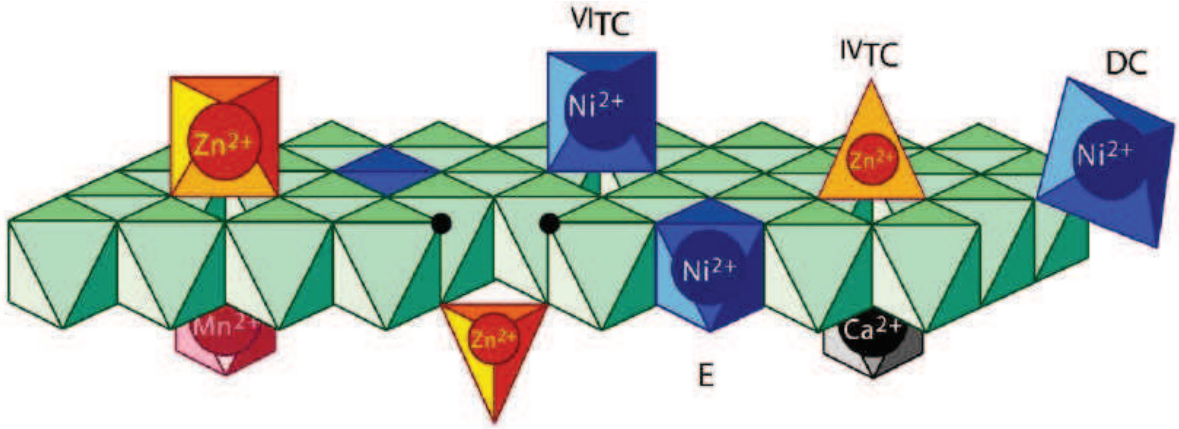


Figure 2.6: Représentation structurale des mécanismes de sorption du nickel et du zinc sur de la birnessite. Le déficit de charge qui résulte de l'existence de lacunes octaédriques peut être compensé par des protons ou des cations en position interfoliaire ou foliaire et octa ou tetracoordonnés. TC illustre le partage de trois sommets au dessus d'un site vacant de la couche de manganèse de structure 'brucitique' (complexe tridenté), E illustre le partage d'arêtes résultant de la diffusion d'un cation Ni^{2+} dans une lacune de Mn^{4+}, et DC illustre le partage de deux sommets (complexe bidenté) sur les cotés du feuillet (Manceau et al. 2007).

3) Processus de l'adsorption

3.1) Capacité d'échange cationique

La capacité d'échange cationique (CEC) est une mesure macroscopique qui exprime la quantité totale de cations qu'un sol peut retenir sur son complexe d'échange. Elle résulte des excès et déficits de charges permanentes présents sur les surfaces. Cette grandeur opérationnelle varie en fonction du pH suite à la présence de charges surfaciques liées aux réactions acido-basiques des groupements fonctionnels.

3.2) Isothermes d'adsorption

A la surface des minéraux, les sites protonables et déprotonables sont essentiellement des groupements hydroxyles –OH. Ces fonctions sont portées par les groupements carboxyliques et phénoliques à la surface de la matière organique sur laquelle on trouve également des groupements amines (-NH2) et sulfhydryles (-SH).

Les réactions d'adsorption de l'ion Cu^{2+} sur ces sites peuvent être écrites de la façon suivante, le symbole ≡ *S* représentant le groupement de surface protonable et déprotonable :

$$\equiv S - H + Cu^{2+} \rightleftarrows \equiv S - Cu^{+} + H^{+}$$

Ces équilibres sont régis par les lois de la thermodynamique, c'est-à-dire que l'énergie totale est conservée et que l'entropie du système augmente. En combinant ces deux concepts, il vient alors la fonction d'énergie libre de Gibbs (G) qui suit la loi suivante : pour qu'une réaction ait lieu, les produits de la réaction doivent avoir une enthalpie libre inferieure à celle des réactants, soit $\Delta G<0$ (Stumm et Morgan 1996).

Expérimentalement, ces équilibres chimiques sont décrits par des isothermes d'adsorption qui mesurent l'activité des espèces dans la phase aqueuse (C) et sur la surface (Q).

3.2.1) Classification

Quatre formes majeures d'isothermes pH ont été identifiées (Fig. 2.7). L'isotherme S a une pente faible au départ qui s'accroit quand la concentration de l'adsorbat augmente. A faible concentration l'adsorbant a une plus faible affinité pour l'adsorbat que la solution. Typiquement, l'adsorption des cations sur les constituants des sols a cette forme lorsqu'ils sont complexés en solution ou sont en compétition forte avec d'autres ions pour la surface,

comme les protons. Cette forme a été utilisée pour décrire l'adsorption du cuivre sur des marnes argileuses (Hinz 2001). L'isotherme L, de forme concave, a une pente plus élevée à faible qu'à haute concentration. Cette forme est observée quand l'adsorbant a une très forte affinité pour l'adsorbat. Celle-ci décroit au fur et à mesure que la surface est recouverte. L'isotherme H est une forme extrême de l'isotherme L traduisant un mécanisme d'adsorption sur des sites de très haute énergie, comme par exemple la diffusion de Co^{3+}, Ni^{2+}, et Cu^{2+} dans les sites lacunaires des oxydes de manganèse (Fig. 2.6). Enfin, dans l'isotherme C la pente est indépendante de la concentration de l'adsorbat ce qui traduit une constante d'association fixe. Cette isotherme permet en général de décrire l'adsorption de substances organiques sur les surfaces minérales, comme les acides aminés sur la montmorillonite calcique (Hinz 2001; Sposito 2008). Plusieurs modèles de modélisation des isothermes existent. Les deux plus utilisés sont le modèle de Langmuir et celui de Freundlich (Harter 1983).

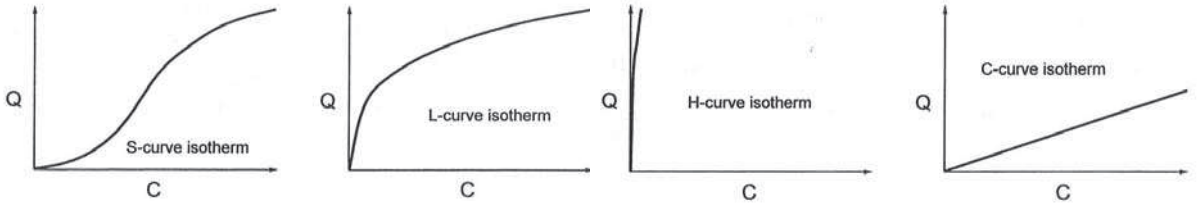


Figure 2.7 : Les quatre formes principales d'isothermes (Sposito 2008). C (en mol/L) est la mesure de l'activité des espèces dans la phase aqueuse et Q (mol/kg) est la mesure de l'activité des espèces sur la surface.

3.2.2) Isotherme de Langmuir

A l'origine, ce modèle a été développé pour décrire l'adsorption à température constante d'un gaz en couche mononucléaire sur une surface. Il a par la suite été étendu aux solutés. Dans le cas simplifié, (1), d'une surface plane avec un seul type de site de surface protoné en nombre fini, (2), d'une réaction réversible, (3), où chaque groupement n'interagit qu'avec une seule molécule ou adatome, (4), qui de plus est immobile, alors l'énergie d'adsorption est la même pour tous les sites et ne dépend pas du taux de couverture. En reprenant la notation $\equiv SH$ pour décrire le site de surface protoné, il vient :

$$\equiv SH \rightleftarrows S^- + H^+ \quad (1)$$

$$\equiv SH + Cu^{2+} \rightleftarrows SCu^+ + H^+ \quad (2)$$

En supposant que les activités des espèces en surface sont proportionnelles à leur concentration, et en appliquant la loi de conservation de la masse, il vient la constante d'équilibre:

$$K_{Cu}^{s} = \frac{\left[SCu^{+} \right]\left[H^{+} \right]}{\left[SH \right]\left[Cu^{2+} \right]}$$

K^s_{Cu} est défini comme la constante de formation du complexe de surface. En faisant l'approximation que dans les conditions habituelles de pH et de température [SH] >> [S^-], alors la concentration maximum des sites de surfaces S_T est donnée par :

$$S_T = \left[SCu^{+} \right] + \left[SH \right] + \left[S^{-} \right]$$
$$S_T \approx \left[SCu^{+} \right] + \left[SH \right]$$

En définissant Γ_{Cu} = [SCu^+]/masse de l'adsorbant, Γ_{max} = S_T/masse de l'adsorbant, et $K^s_{Cu}/[H^+] = K_{ads}$ on peut alors écrire :

$$\Gamma_{Cu} = \frac{\Gamma_{\max} K_{ads(pH)} \left[Cu^{2+} \right]}{1 + K_{ads(pH)} \left[Cu^{2+} \right]}$$

K_{ads} est la constante d'adsorption à une concentration en proton donnée. Elle est directement reliée à la fonction de Gibbs par la relation:

$$K_{ads} = \exp(-\frac{\Delta G^{0}_{ads}}{RT})$$

où ΔG^0_{ads} est l'énergie libre d'adsorption, R la constante des gaz parfaits et T la température en Kelvin.

Expérimentalement, il est possible de mesurer Γ_{Cu} qui représente le taux d'adsorption en fonction de la concentration en [Cu^{2+}] dans la solution et d'en déduire les paramètres Γ_{max} et K_{ads}. Néanmoins, les hypothèses posées précédemment sont extrêmement fortes et il est peu représentatif de ce qui se passe réellement dans les sols, du fait notamment de la très grande hétérogénéité chimique et structurale des surfaces les plus réactives.

3.2.3) Isotherme de Freundlich

Les mesures expérimentales d'adsorption des éléments traces sur les constituants des sols sont mieux décrites avec l'équation de Bemmelen-Freundlich que celle de Langmuir (Boudesocque et al. 2007; Sposito 2008). L'équation s'écrit :

$$\Gamma = m\left[A \right]^{n}$$

Comme précédemment, Γ est la quantité d'adsorbat sur la surface. Ce paramètre s'exprime en mol.kg^{-1} tandis que [A] est la quantité totale d'adsorbat dans le système exprimée en mol.L^{-1}. m est la constante de Freundlich et n une mesure de la non linéarité. Quand la température augmente, m diminue et n augmente. Après transformation logarithmique, l'équation s'écrit:

$$\log \Gamma = \log m + n \log[A]$$

Ainsi, en traçant log Γ vs log [A], on obtient une droite de pente n dont la valeur à l'origine est log m (Fig. 2.8).

Contrairement à l'équation de Langmuir, cellede Freundlich ne donne pas d'informations directes sur la capacité d'adsorption de la surface ni sur l'énergie d'adsorption. Néanmoins, les constantes m et n peuvent être interprétées comme des indicateurs relatifs respectivement de la capacité et de l'affinité (Slejko 1985). A faible taux d'adsorption, les données sont correctement reproduites par les deux modèles. Enfin, aucune de ces deux équations ne donne d'information sur les mécanismes réels d'adsorption aux échelles atomique et moléculaire (Warren et Haack 2001).

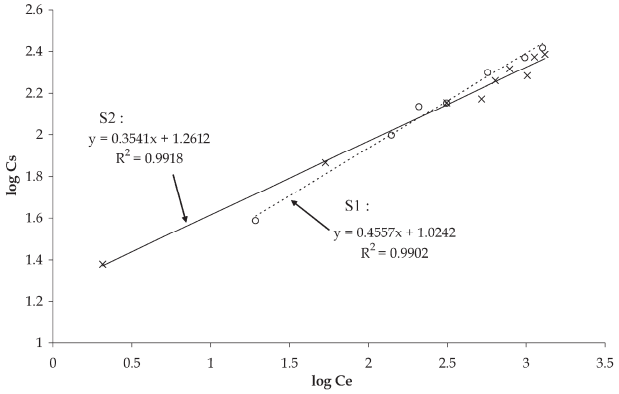


Figure 2.8 : Modélisation des isothermes d'adsorption du cuivre dans deux sols de vignobles avec le modèle de Freundlich. Ce = [Cu^{2+}], $C_s = \Gamma_{Cu}$, T= 293K, pH= 6, I = 0,1M (Boudesocque et al. 2007).

3.3) La matière minérale

3.3.1) Modèles de Complexation de Surface (MCS)

3.3.1.1) Approche théorique

Des modèles de complexation de surface (MCS) ont été développés pour tenter d'obtenir des informations mécanistiques sur les processus de sorption à l'interface minéral-solution, à partir des données macroscopiques (Stumm et Morgan 1996). Dans le cas des constituants des sols, les hypothèses très fortes posées au § 3.2.2 ne sont plus valables. Il en résulte que si l'on peut toujours raisonner en termes de constante d'équilibre (lois de conservation de la masse), ces constantes sont apparentes (notées K_{app}) au sens où elles dépendent alors de la charge de la surface qui est la somme arithmétique :

a) De la charge des groupements fonctionnels -OH, -SH… dont l'état d'ionisation dépend des paramètres de la solution. Ces états sont quantifiés par des constantes

d'affinité K_a. La charge de surface « protonique » σ_H et son point de charge nulle, c'est-à-dire la valeur de pH pour laquelle la charge nette est nulle, sont obtenus par titrage potentiométrique, c'est-à-dire en mesurant l'isotherme d'adsorption de l'ion H^+.

b) De la charge permanente σ_0 provenant des substitutions isomorphiques.

c) De la charge des adatomes dans la sphère externe (σ_{OS}) et dans la sphère interne (σ_{IS}). Il est possible de différencier ces deux dernières charges en faisant varier la force ionique de la solution.

Au final, la charge nette est égale à $\sigma_N = \sigma_0 + \sigma_H + \sigma_{IS} + \sigma_{OS}$. Elle s'exprime en Coulomb par mètre carré ($C.m^{-2}$).

L'énergie libre d'adsorption ΔG_{ads} peut alors s'écrire comme la somme d'une composante chimique résultant des lois de conservation de masse (composante dite intrinsèque ΔG_{int}) et d'une composante résultant de la charge de surface (composante dite électrostatique ΔG_{elec}). Cette énergie électrostatique correspond au travail nécessaire pour amener un ion de la solution vers le site réactif du sorbant via un potentiel ψ_0 à l'interface entre les deux phases (Stumm et Morgan 1996). En combinant $\Delta G_{ads} = -RT\ln K_{app}$, $\Delta G_{int} = -RT\ln K_{int}$ et $\Delta G_{elec} = \Delta Z F \psi_0$ (F constante de Faraday), il vient la relation entre les constantes intrinsèque et apparente : $K_{app} = K_{int}\exp(-\Delta Z F \psi_0 / RT)$.

Dans l'idéal, en connaissant ce paramètre K_{int} et en recoupant avec la nature de la surface il est alors possible, d'une part, de déterminer le(s) type(s) de sites réactifs impliqué(s) dans la réaction de sorption de l'adatome et, d'autre part, de connaitre la réactivité de la surface indépendamment des conditions physico-chimiques du milieu, et donc de prévoir son comportement dans d'autres conditions. Les réactions de surface sont en compétition les unes avec les autres et les phénomènes d'adsorption majeurs sont (Stumm et Morgan 1996) :

a) Equilibres acide-base

$$S-OH + H^+ \rightleftarrows S-OH_2^+ \text{ et } S-OH + OH^- \rightleftarrows S-O^- + H_2O$$

b) Liaisons simples

$$S-OH + M^{z+} \rightarrow S-OM^{(z-1)} + H^+$$

$$2S-OH + M^{z+} \rightleftarrows (S-O)_2 M^{(z-2)} + 2H^+$$

$$S-OH + M^{z+} + H_2O \rightleftarrows S-OMOH^{(z-2)} + 2H^+$$

c) Echange de ligands

$$S-OH+L^- \rightleftarrows S-L+OH^- \text{ et } 2S-OH+L^- \rightleftarrows S_2-L^+ +2OH^-$$

d) Formation de complexes de surface ternaires

$$S-OH+L^- +M^{z+} \rightleftarrows S-L-M^{z+} +OH^- \text{ et } S-OH+L^- +M^{z+} \rightleftarrows S-OM-L^{(z-2)} +H^+$$

Dans ces modèles les sites d'échange cationique ont une charge unitaire et sont considérés comme des sites de sorption de cations analogues aux sites déprotonables situés en bordure du réseau cristallin. La réaction d'échange est décrite comme la réaction d'adsorption de cations M^{z+} sur des sites de charge permanente S_p^- suivant l'équation :

$$S_p^- + M^{z+} \Leftrightarrow S_p M^{(z-1)+}$$

et est définie par une constante d'équilibre intrinsèque (Dzombak et Hudson 1995).

Pour déterminer K_{int}, il est d'abord nécessaire de déterminer ψ_0. Ce potentiel diminue à mesure que la distance à la surface augmente. Helmholtz en 1879 a assimilé le milieu à un condensateur dont la plaque négative serait la surface et la plaque positive serait située à une distance égale au diamètre du cation compensateur. Cette théorie fut modifiée par Gouy et Chapman en 1923, qui ont postulé que la seconde plaque n'était pas plane mais formée d'une couche diffuse, où les ions sont en équilibre sous l'action conjointe du champ électrique et de l'agitation thermique. Mais cette approche n'est pas satisfaisante à des distances extrêmement proches de la surface : la concentration ionique devient alors trop importante. C'est l'assimilation des ions à des points, et non à des sphères, qui provoque cet artefact. Stern a alors proposé de diviser l'interface en deux régions, la première est une couche compacte d'ions adsorbés à la surface et la seconde est la double couche diffuse (DCD). Cette couche diffuse présente un plan de cisaillement. Entre la surface et ce plan, la solution est fortement liée à la surface, tandis qu'au-delà elle est plus mobile. On appelle charge électrocinétique ou potentiel ζ la charge au niveau du plan de cisaillement. A partir de cette approche trois modèles principaux sont utilisés, le modèle de la double couche diffuse, le modèle à capacité constante (CC) et le modèle de la triple couche (TL) (Rudzinski et al. 1997; Bradl 2004; Richter et al. 2005). Ils sont illustrés à la figure 2.9 et les relations charges/potentiels détaillées dans le tableau 2.1.

Dans le modèle DCD, la couche diffuse comporte une sous couche près de la surface dans laquelle le potentiel est constant. Les complexations de sphère interne et externe ont lieu dans cette couche (Schlegel 2000). Puis le potentiel décroit à mesure que la distance à la surface augmente selon la loi de Gouy-Chapman. Dans le modèle CC, le potentiel décroit

linéairement jusqu'à 0. La charge de surface est liée au potentiel par le biais d'une capacitance spécifique C_1. L'intérêt de ce modèle est limité, C_1 étant dépendant de la force ionique. Il s'agit en réalité d'une approximation du modèle DCD dans le cas de hautes forces ioniques (Schlegel 2000).

Tableau 2.1 : Relations charge/potentiel des trois modèles de complexation de surface (MCS) couramment utilisés pour modéliser l'adsorption des ions à l'interface liquide/solide (Stumm et Morgan 1996)

DCD	CC	TC
		$\sigma_d = -0.1174\sqrt{I}\sinh(zF\psi_d/2RT)$
$-\sigma_0 = \sigma_d = -0.1174\sqrt{I}\sinh(zF\psi_d/2RT)$	$\sigma_0 = C_1\psi_0$	$\sigma_0 = (\psi_0 - \psi_\beta)C_1$
		$\sigma_0 + \sigma_\beta = (\psi_\beta - \psi_d)C_2 = -\sigma_d$

F : constante de Faraday (96490 $C.mol^{-1}$) ; Z valence de l'ion adsorbé, ψ (V) ; R : constante des gaz parfaits (8,314 $J.mol^{-1}K^{-1}$) ; *T* : température (K) ; C_1($F.m^{-2}$) et $C_2 = 0,2$ $F.m^{-2}$.

Enfin, dans le modèle TC la couche diffuse est divisée en trois couches. Dans les deux couches les plus proches de la surface les potentiels décroissent linéairement. Puis le potentiel décroit de la même manière que dans le modèle DCD. Dans ce modèle, les complexations de sphère interne ont lieu dans la couche la plus proche de la surface, et les complexations de sphère externe dans la seconde couche (Warren et Haack 2001).

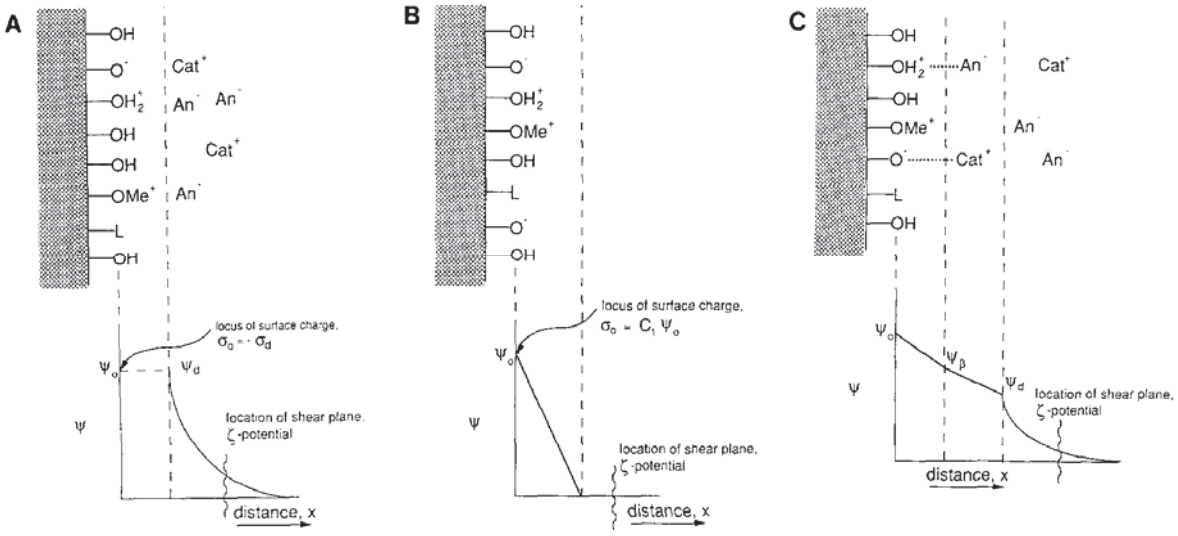


Figure 2.9 : Représentation schématique des propriétés électriques de l'interface solide/liquide. A- Modèle de la double couche diffuse (DCD), B- Modèle à capacité constante (CC), C-Modèle de la triple couche (TC) (Warren et Haack 2001).

3.3.1.2) Limites des MCS

Bien qu'en théorie ces modèles soient capables de simuler l'adsorption d'espèces aqueuses en solution en fonction du pH, de la concentration de l'adsorbat, et de la force ionique, il apparaît qu'ils sont d'utilisation limitée. Tout d'abord ces modèles impliquent que toutes les réactions chimiques sont à l'équilibre, or les ordres de grandeur des durées pour atteindre l'équilibre sont de l'ordre du mois (Schlegel 2000). Par ailleurs, ces modèles ne prennent pas en compte la possibilité de précipitation des cations métalliques avec des espèces dissoutes comme Si et Al. Enfin, ces modèles se heurtent à la complexité physico-chimique des constituants naturels et aux informations limitées contenues dans les données expérimentales (Buffle 1990). Il a ainsi été montré sur une phase apparemment aussi simple que la kaolinite que les modèles ne pouvaient pas conclure sur les mécanismes d'adsorption du cuivre sur cette surface (Lund et al. 2008; Peacock et Sherman 2005). Or pour se rapprocher des conditions réelles, il est indispensable d'augmenter la complexité des systèmes étudiés, notamment pour prendre en compte l'existence de complexes secondaires MO-matière minérale (Hizal et Apak 2006) et la formation possible de complexes ternaires entre les groupements de surface, l'adatome et d'autres espèces de la solution (Maqueda et al. 2002).

3.3.2) Approche spectroscopique

Pour connaître la nature réelle des interactions chimiques à l'interface liquide-solide, il est nécessaire de changer d'échelle d'observation. Les techniques spectroscopiques permettent de sonder ces interactions à l'échelle élémentaire, et ainsi de contraindre les modèles chimiques macroscopiques. Plusieurs techniques de spectroscopie de surface existent (LEED, UPS, XPS, Auger…), seulement, et sauf cas particulier, les mesures sont effectuées habituellement sous vide, et donc sur des surfaces déshydratées. En revanche, les spectroscopies fondées sur l'absorption des rayons X (XAS) permettent un examen in situ, et de plus avec une sensibilité souvent en dessous de la centaine de mg/kg, et donc à relativement bas taux de couverture des surfaces (Brown et al. 1999). Elles sont particulièrement pertinentes pour l'étude des interactions élements traces-surface. Les principes de la spectroscopie d'absorption X sont maintenant bien connus et rapidement décrits à l'annexe 1 de ce manuscrit.

Dans le cas du cuivre, compte tenu de la déformation de l'octaèdre suite à l'effet Jahn-Teller, les deux oxygènes apicaux sont difficilement détectables par EXAFS. Par cette

technique, le nombre de voisins oxygène effectivement observé est donc généralement d'environ 4 même lorsque la coordinence est octaédrique (Manceau et al. 2000).

3.3.3) Description des surfaces minérales dans les sols et mécanismes d'interactions Cu-matière minérale

3.3.3.1) Les phyllosilicates

Structure

Les phyllosilicates sont des minéraux à structure lamellaire dans laquelle les feuillets sont formés par l'empilement de couches tétraédriques 'T' de SiO_4 et de couches octaédriques 'O' de cations en général de l'aluminium $Al(OH)_6$, du fer ou du magnésium. Les tétraèdres sont associés entre eux par trois sommets et forment en surface (plan basal) des cavités hexagonales, ou plus exactement di-trigonales en raison de leur déformation (Fig. 2.10). Les oxygènes apicaux sont partagés avec les octaèdres de la couche O. Ces derniers sont reliés par des arêtes. Si tous les sites octaédriques sont occupés, chaque octaèdre possédant alors six octaèdres voisins, la couche O est dite trioctaédrique. Pour des raisons de stœchiométrie de la structure, les cations sont alors très majoritairement divalents (Mg^{2+}, Fe^{2+}…).

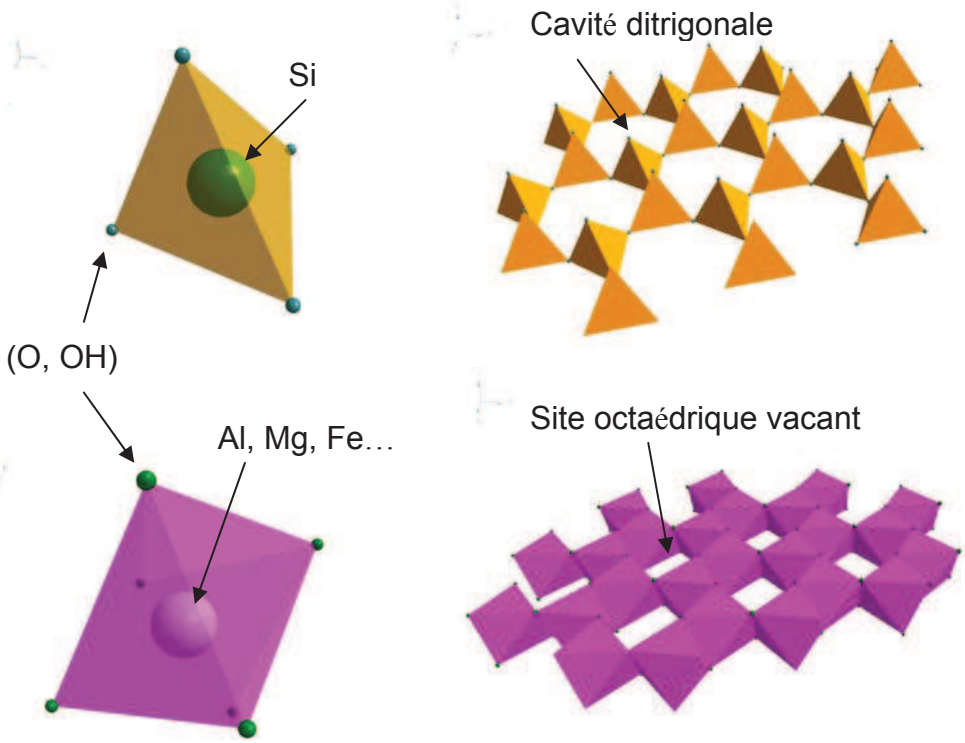


Figure 2.10 : En haut, tétraèdre de $Si(O,OH)_4$ et agencement des tétraèdres SiO_4 représentés sous forme polyédrale. En bas, octaèdre $(Al,Fe,Mg…)(O,OH)_6$ et agencement des octaèdres sous forme polyédrale. Dans cet exemple, deux sites octaédriques sur trois sont occupés, la couche est alors dioctaédrique. Quand tous les sites sont occupés, la couche est trioctaédrique.

Hors des fonds océaniques, ce cas, bien que non négligeable, reste tout de même marginal et, en général, les cations sont principalement trivalents (Al^{3+}, Fe^{3+}…). Deux sites octaédriques sur trois sont alors occupés et la couche est dite dioctaédrique. La distance entre deux feuillets successifs est appelée distance inter-réticulaire et l'espace entre deux feuillets constitue l'espace interfoliaire. Les différents cations octaédriques ont des rayons ioniques distincts et donc la taille des octaèdres varie. L'ajustement des dimensions des couches O et T se fait par le biais de modifications structurales. Il s'agit d'un aplatissement de la couche O et d'une rotation avec basculement des tétraèdres de la couche T, conduisant à un abaissement de la symétrie des cavités hexagonales (Schlegel 2000).

En fonction du rapport nombre de couche tétraédrique sur nombre de couche octaédrique, plusieurs familles structurales sont définies (Fig. 2.11) :

- groupe 1/1 : La couche O est liée à une seule couche T. C'est le groupe des kaolinites $Al_2Si_2O_5(OH)_4$ (dioctaédrique) ou des serpentines $(Mg,Fe)_3Si_2O_5(OH)_4$ (trioctaédrique). Lorsqu'il y a une substitution chimique dans une couche, elle est compensée en général par une substitution dans l'autre couche (Moore et Reynolds 1997). La charge de surface est faible (voire nulle) et ces matériaux sont peu réactifs. La CEC des kaolinites est de l'ordre d'une dizaine meq/100g.

- groupe 2/1 : La couche O est prise en sandwich entre deux couches T. Ce groupe est subdivisé suivant la charge *z* du feuillet.

Quand la charge est nulle ($z = 0$), c'est alors le groupe des talcs (trioctaédrique) et des pyrophyllites (dioctaédrique). Il n'y a pas de substitution dans les couches T et O et il n'y a pas de cations compensateurs dans l'espace interfoliaire.

Quand la charge foliaire est proche de -1, c'est alors le groupe des micas, comme la muscovite (dioctaédrique) et la biotite (trioctaédrique). La charge provient généralement de la substitution d'un cation Si^{4+} par un cation Al^{3+} dans la couche T mais des substitutions ($Al^{3+} \rightarrow R^{2+}$, R pour Fe ou Mg par ex.) peuvent également survenir dans la couche octaédrique. La charge est neutralisée par un cation compensateur (le plus souvent K^+) situé dans la cavité di-trigonale. La CEC des micas est faible, de l'ordre de 10 à 50 meq/100 g (Osman et Suter 2000).

Quand la charge est comprise entre -0.3 et -0.9 on distingue alors les illites ($z \sim -0.9$), les vermiculites ($-0.9 > z > -0.6$) et les smectites ($-0.6 > z > -0.3$). Les illites sont des phases mal définies, assez proches des micas (et d'ailleurs souvent classées dans la même famille) qui présenteraient moins de substitutions alumineuses dans la couche T et moins de cations compensateurs K^+ interfoliaires (Moore et Reynolds 1997). Chez les smectites et les

vermiculites, les charges sont variables et sont induites par des substitutions tétraédriques ou octaédriques. Les smectites sont classifiées suivant la nature, la localisation du déficit de charge du feuillet et la nature du cation octaédrique dominant (Schlegel 2000). Les feuillets sont chargés négativement et ont tendance à se repousser. La charge est compensée par un cation plus ou moins hydraté en position interfoliaire qui induit des interactions attractives avec les deux feuillets. Les smectites et les vermiculites présentent des CEC comprises entre 80 et 150 meq/100 g (Baize 2000). Il faut noter que des empilements successifs de feuillets de smectites, illites et vermiculites peuvent survenir. On parle alors d'interstratifiés.

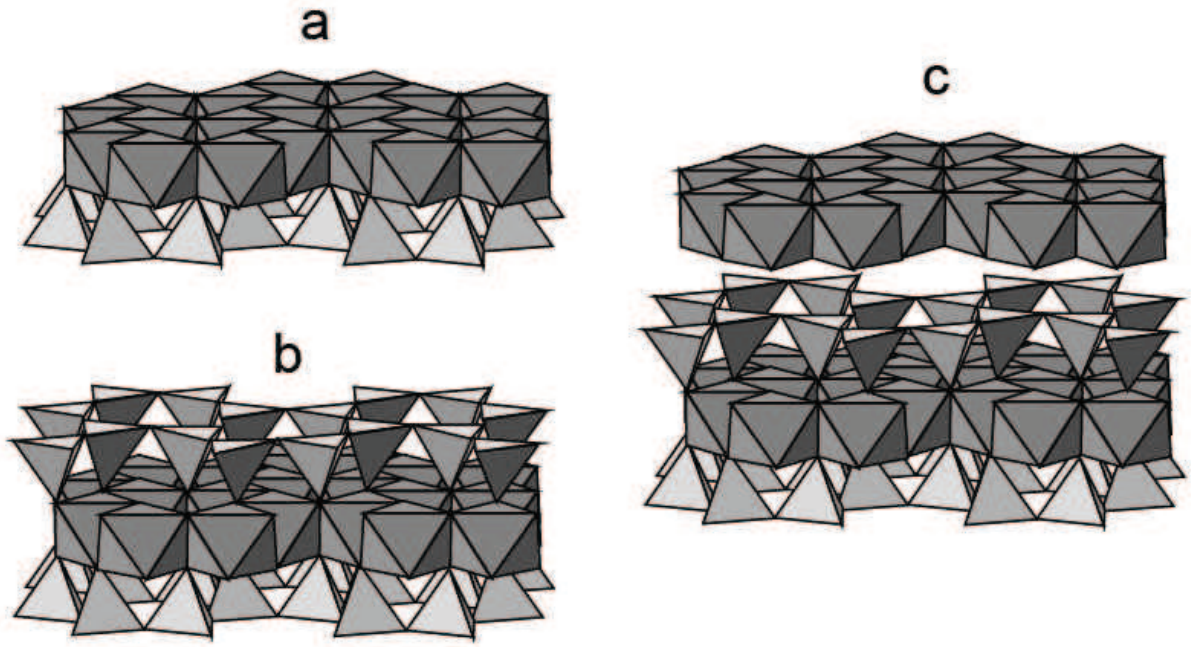


Figure 2.11 : Agencement des couches octaédriques et tétraédriques dans les phyllosilicates. a) feuillet de type 1/1 (kaolinite, serpentine, b) feuillet de type 2/1 (talc, pyrophyllite, illite vermiculite, smectite), c) feuillet 2/1 + couche octaédrique (chlorite) (Ferrage 2004).

Enfin, quand les cations compensateurs de charge ne sont pas isolés mais condensés et forment un feuillet octaédrique, il s'agit alors de chlorites. Bien que les phénomènes de substitutions soient communs dans les chlorites, les cations octaédriques majeurs sont Mg^{2+}, Fe^{2+}, Al^{3+} et Fe^{3+}. La CEC de ces matériaux est de l'ordre de 10 à 40 meq/100 g (Baize 2000).

Mécanismes de fixation de Cu

Kaolinite

A la surface de la kaolinite, à I = 0.1M et pH = 4,4 puis 5,8 et 6,5 avec des taux de couverture de surface respectif de 7,6%, 37,1% et 42,6%, plusieurs processus d'adsorption auraient lieu (Peacock et Sherman 2005). Tout d'abord à faible taux de recouvrement, il se

produirait une complexation de sphère externe entre Cu^{2+} hydraté et la surface (pH = 4,4). Ensuite pour le taux supérieur de 37,1% (pH=5,8), une complexation de sphère interne aurait lieu ; le cluster de cuivre serait plan carré mononucléaire bidenté $(\equiv AlOH)_2Cu(OH)_2^0$ avec des distances Cu-O égales à 1,99, 2,06, 2,07 et 2.08 Å et des distances Cu-Al de l'ordre de 3,70-3,71 Å, soit un angle 'Cu-O-Al' de l'ordre de 160°. Enfin, en augmentant le taux de recouvrement à 42.6% (pH = 6,5), toujours en complexation de sphère interne, le cuivre s'adsorberait également sous forme d'un cluster binucléaire de Cu (d_{Cu-Cu} = 2,94 Å) tridenté $(\equiv Al_3O(OH)_2)Cu_2(OH)_3^0$ avec la couche alumineuse, qui est le prélude d'une précipitation de surface. Concernant cette étude, on notera que les auteurs se sont arrêtés au nombre de voisins oxygène du cuivre déterminé par EXAFS (i.e. 4) et n'ont pas étudié la possibilité d'existence de clusters de cuivre en coordinence 6.

Smectites

Pour des rapports Cu/montmorillonite compris entre 10 et 150 $mmol.kg^{-1}$, l'adsorption des cations sur ces argiles gonflantes est influencée par la composition et la concentration de l'électrolyte et par le pH de la solution. Sur les vermiculites et smectites, l'adsorption augmente à bas pH quand la force ionique diminue. Cette dépendance est caractéristique d'une adsorption sur les sites d'échange cationique (Schlegel 2000). Le cuivre hydraté forme alors un complexe de sphère externe dans l'espace interfoliaire (Morton et al. 2001) (Fig. 2.12).

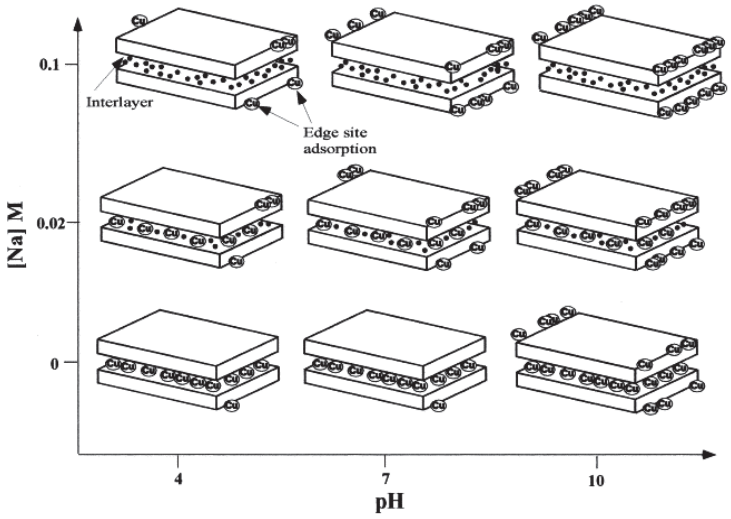


Figure 2.12 : Diagramme schématique illustrant les effets du pH et de la force ionique (ici les cations de l'électrolyte sont des Na^+) sur la distribution des sites d'adsorption du cuivre dans une smectite. Quand la force ionique augmente, il se développe une compétition entre les cations de l'électrolyte et les ions Cu^{2+} à la faveur des premiers pour les sites d'échange, ces derniers allant se fixer sur les sites acido-basiques situés sur les bords de la surface. Quand le pH augmente, les sites –OH en bordure des feuillets se déprotonent laissant des places libres pour la fixation des cations métalliques (Morton et al. 2001).

A haute force ionique, les cations de l'électrolyte, alors en fort excès, occupent les sites d'échange préférentiellement aux atomes de cuivre qui eux se fixent majoritairement sur les liaisons pendantes situées en bordure des couches octaédriques et tétraédriques (Strawn et al. 2004). Ce mécanisme est renforcé à haut pH alors que les sites de surface se déprotonent (Fig. 2.12).

A un pH = 4,2 et une force ionique de 0,1 M, avec un rapport Cu/montmorillonite de 21 $mmol.kg^{-1}$ en plus de l'échange cationique, il a été observé un mécanisme de complexation de sphère interne sur les sites de bordure (Morton et al. 2001). A un pH proche de 6 pour une force ionique de 0,1 M, il se forme un cluster de cuivre binucléaire avec une distance Cu-Cu de l'ordre de 2,65 Å (Strawn et al. 2004) à 3,00 Å (Morton et al. 2001). La différence entre ces deux résultats a été attribuée à la différence rapport Cu/montmorillonite (100 vs 25,7 $mmol.kg^{-1}$). Dans les deux cas, les distances $Cu\text{-}O_{éq}$ sont égales à 1,95-1,96 Å. Dans l'étude la plus récente, il a été proposé que ce cluster partage un oxygène avec une couche tétraédrique des smectites (Fig. 2.13). Cependant ce mécanisme est hypothétique car aucune paire Cu-Si n'a été détectée.

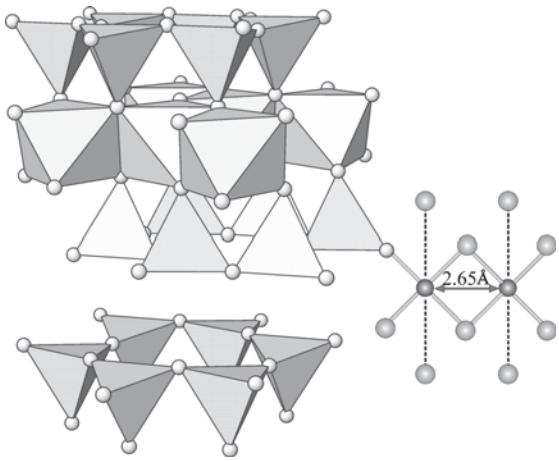


Figure 2.13 : Illustration du possible mécanisme de complexation du cuivre sur les smectites. Le cluster est binucléaire, complexé en sphère interne en partageant un atome d'oxygène avec un tétraèdre de Si (Strawn et al. 2004).

Vermiculites

L'adsorption du cuivre à la surface des vermiculites est semblable à celle des smectites. Dans des échantillons naturels, les ions Cu^{2+} hexacoordonnés occupent l'espace interfoliaire (Ildefonse et al. 1986). A haute force ionique (I = 0,1 M), pH = 6 et un rapport Cu/minéral de l'ordre de 40 $mmol.kg^{-1}$, l'octaèdre de cuivre partage une arête de la couche octaédrique avec des distances Cu-O de 1,96 Å (3,8 voisins) et Cu-Al/Si de 3,02 Å (2,6 voisins) (Fig. 2.14).

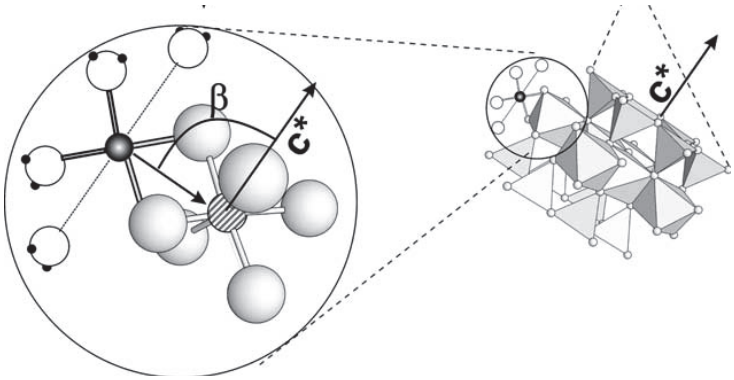


Figure 2.14 : Spéciation du cuivre sur la vermiculite à haute force ionique (I = 0,1M). Les sphères blanches représentent les molécules d'eau, les sphères grises les molécules d'oxygène, la sphère hachurée est un atome d'aluminium et la sphère noire l'atome de cuivre (Furnare et al. 2005).

Micas

Il a été montré par réflectivité des rayons X sur des monocristaux de muscovite que le cuivre hydraté pouvait former deux types de complexes de sphère externe, l'un à 4 Å de la surface basale, et l'autre à 8 Å. (Lee et al. 2010). Ces résultats montrent que le cuivre à une affinité pour la surface basale de la muscovite supérieure à celle du potassium.

Par ailleurs, pour des taux de recouvrement de surface de 1 atome de cuivre sur 9 Å² quand la force ionique est élevée (I = 0,1 M, pH = 5,4), un mécanisme de complexation de sphère interne a pu être observé. Dans la muscovite, quatre atomes d'oxygène à une distance de 1,95-1,98 Å, un atome de cuivre à une distance de 2,61-2,68 Å et quatre atomes de Si/Al à des distances de 3,09-3,26 Å ont été détectés par EXAFS. Ces clusters binucléaires de cuivre s'adsorberaient sur les bords de la surface. Dans la biotite, quatre atomes d'oxygène à une distance de 1,95 Å et un atome de cuivre à une distance de 3,11 Å ont été observés, ce qui ne permet de conclure que le cuivre est réellement en surface (Farquhar et al. 1996; Farquhar et al. 1997). La faible densité des sites de bordure résultant de la grande surface basale des micas rend difficile l'observation des phénomènes de complexation de sphère interne.

3.3.3.2) Les oxy(hydroxy)des de fer

Structure

Les oxydes (hématite, Fe_2O_3) et oxyhydroxydes (FeOOH) de fer (OHF) sont constitués d'octaèdres $Fe(O,OH)_6$ associés par les faces (Fe_2O_3), arêtes et sommets (Fe_2O_3 et FeOOH) (Fig. 2.15). Les substitutions les plus fréquentes et abondantes (jusque ~30%) se font avec l'aluminium, et sont donc isovalentes entre les ions Fe^{3+} et Al^{3+} (Schwertmann et Carlson 1994), si bien que ces minéraux n'ont pas de charge fixe, mais uniquement variable

en fonction du pH. La goethite (α-FeOOH) est un oxyhydroxyde cristallisé tres abondant, tandis que la ferrihydrite est une forme mal cristallisée, et donc métastable, d'oxyhydroxyde de fer obtenue par hydrolyse des ions ferriques ou ferreux en solution, et qui avec le temps se transforme le plus souvent en goethite et hématite (Manceau 2011).

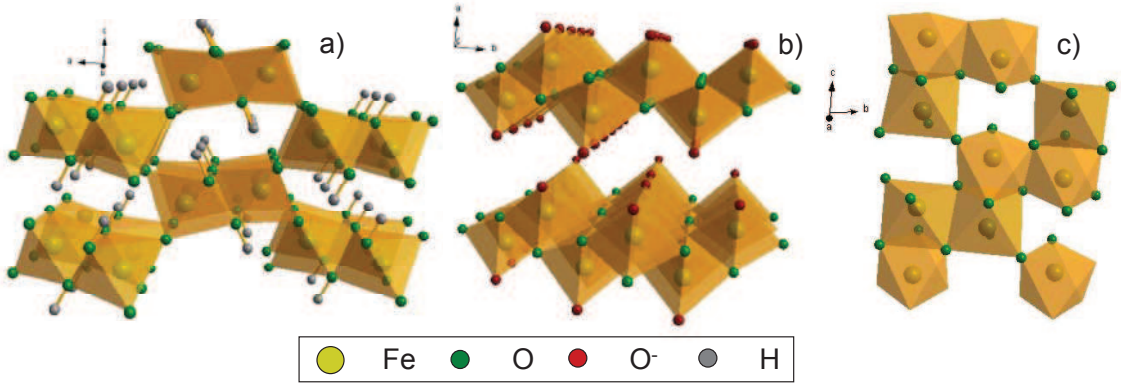


Figure 2.15 : Représentation de la structure d'oxy(hydroxy)des de fer bien cristallisés. a) goethite α-FeOOH, b) lépidocrocite γ-FeOOH sous sa forme déprotonée et c) hématite α-Fe_2O_3. Les atomes de fer sont représentés en taupe, les atomes d'oxygène en vert, les atomes d'oxygène déprotonés en rouge (b) et les atomes d'hydrogène en blanc (a).

Mécanismes de fixation du cuivre

La goethite et l'hématite peuvent coprécipiter avec une grande variété de cations métalliques dont Cu^{2+} et les séquestrer dans leur structure cristalline (Manceau et al. 2000). Dans la goethite ([Cu] = 2630 ppm), le cuivre se substitue aux atomes de fer sous la forme d'un octaèdre déformé. Les distances Cu-O sont de 1,98 Å et 2,11 Å, et Cu-Fe de 3,00 Å, 3,23 Å et 3,48 Å dans la goethite synthétique. Dans la goethite naturelle, ces paires atomiques sont à 1,94 Å et 2,12 Å et à 3,05 Å, 3,26 Å et 3,50 Å. Ces différences peuvent s'expliquer par l'existence d'impuretés (comme Al^{3+}) dans la goethite naturelle.

La complexation en sphère interne de Cu sur la lépidocrocite a été décrite (Parkman et al. 1999). Dans cette étude, les atomes de Cu sont entourés de quatre atomes d'oxygène à 1,96 Å, de deux autres à 2,43 Å, et de plusieurs cations métalliques (Fe ou Cu) à 3,04 Å et 3,67 Å pour des rapports Cu/minéral de 68 $mmol.kg^{-1}$ et pH = 4,9. Ces distances interatomiques ont été attribuées respectivement au partage d'une arête et d'un sommet entre les octaèdres de fer et de cuivre.

Toujours selon cette étude, pour des conditions initiales différentes (pH = 4,7 et Cu/goethite = 42 $mmol.kg^{-1}$), il a été observé sur la goethite que le cuivre était chimisorbé sous forme octaédrique déformée par l'effet Jahn Teller, avec quatre atomes d'oxygène à 1,94 Å, deux à 2,41 Å et deux cations Fe/Cu à 2,92 Å. Ces mécanismes ont été attribués à un partage d'arête entre les octaèdres de fer et de cuivre. Plus récemment, sur de la goethite, de la

lépidocrocite et de l'hématite, pour des pH de 4,7 à 6,4 (de 13,8 à 52,3% de couverture de surface), dans des solutions à I = 0,1 M, il a été proposé (Peacock et Sherman 2004) :

- un cluster binucléaire tridenté en coordinence 4 $Cu_2O(OH)_5^{3-}$ partageant trois coins avec les octaèdres $Fe(O,OH)_6$.
- un complexe mononucléaire en coordinence 4 bidenté $Cu(OH)_4^{2-}$ partageant deux coins avec les octaèdres $Fe(O,OH)_6$.

Les distances entre les atomes varient suivant le cristal et le pH. Comme dans l'étude des interactions du cuivre avec la kaolinite menée par ces auteurs, la possibilité d'existence de clusters de cuivre en coordinence 6 n'a pas été envisagée.

3.3.3.3) Les oxy(hydroxy)des de manganèse

Structure

Les oxydes de manganèse (OM) de formule type MnO_2 sont composés d'octaèdres MnO_6 liés par les arêtes et sommets. On distingue deux grandes familles, les phyllomanganates et les tectomanganates (Post 1999; Tonkin et al. 2004). Les premiers ont une structure lamellaire constituée de feuillets empilés d'octaèdres associés par les arêtes. L'espèce la plus fréquente est la birnessite, et sa forme désordonnée la vernadite (δ-MnO_2). Les charges permanentes proviennent de l'existence de lacunes de Mn^{4+} dans les feuillets pour les formes hexagonales stables à des pH < 7 et de substitutions Mn^{4+}/Mn^{3+} pour les formes tricliniques stables à des pH > 8. Les tectomanganates ont une structure en tunnel de taille comprise entre 1x1 (pyrolusite) et 3x3 (todorokite) (Fig. 2.16). Ils n'ont pas de lacunes octaédriques, mais souvent des substitutions Mn^{4+}/Mn^{3+} qui génèrent également une charge compensée par des cations à l'intérieur des canaux.

Mécanismes de fixation du cuivre

Dans les phyllomanganates hexagonaux, les atomes de cuivre divalents occupent les mêmes sites cristallographiques que Zn dans la chalcophanite, c'est-à-dire au dessus ou au dessous des lacunes des octaèdres Mn^{4+}. Ils sont sous forme d'octaèdres déformées et partagent trois coins avec les octaèdres de manganèse de la couche (Manceau et al. 1987; Manceau et al. 2000; Sherman et Peacock 2010). Ces octaèdres sont formés de deux atomes d'oxygène et de deux molécules d'eau dans le plan équatorial et d'un atome d'oxygène et d'une molécule d'eau dans l'axe, les 3 atomes d'oxygène étant partagés avec la surface de la birnessite. Dans le plan équatorial la distance Cu-O est de 1,96 Å et dans l'axe, de 2,23 Å. Ces

résultats ont été obtenus sur des échantillons naturels et en laboratoire avec des rapports Cu/minéral de l'ordre de 1000 mmol.kg^{-1}. Récemment, il a été montré que pour des pH élevés (pH = 8), Cu pouvait être incorporé dans la couche brucitique en occupant un site Mn octaédrique vacant (Sherman et Peacock 2010).

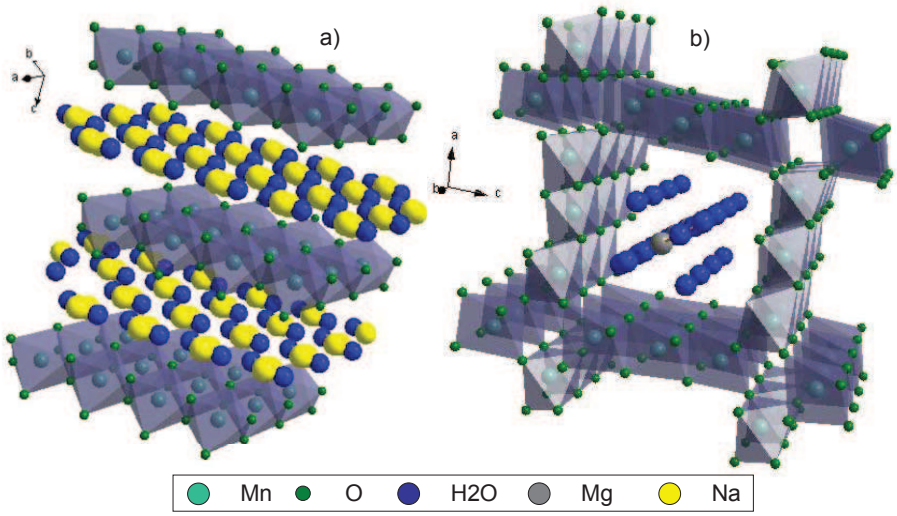


Figure 2.16 : Représentation de la structure d'oxydes de manganèse bien cristallisés. a) birnessite triclinique, b) todorokite. Dans la birnessite triclinique, la charge foliaire est compensée par des alcalins ou des alcalino-terreux en position interfoliaire (a). Dans le tunnel de la todorokite, les cations compensateurs sont le plus souvent des ions Mg^{2+} et Ca^{2+} (b).

La spéciation du cuivre a également été étudiée dans la pyrolusite (MnO_2). Pour des rapports Cu/minéral supérieurs à 10000 mmol.kg^{-1} et jusqu'à un pH inférieur à 5,73, l'environnement local serait le suivant : 4 atomes d'oxygène à 1,95 Å, un atome de cuivre à 2,65 Å et 4 atomes de Mn/Cu à 3,34 Å. Pour des pH supérieur à 5,63, Cu précipiterait sous forme de $Cu(OH)_2$, et les deux formes cohabiteraient à la surface (Chang et al. 2005).

Enfin, dans un échantillon naturel de lithiophorite ($(Li,Al)MnO_2(OH)_2$), le cuivre se trouve dans la couche alumineuse (hydrargillite), il se substitue à (Li,Al) et la distance Cu-Al est égale à 2,96 Å (Manceau et al. 1990).

3.3.3.4) Les autres oxydes

L'objet de cette section est de faire une rapide synthèse bibliographique des mécanismes de rétention du cuivre à la surface des oxydes d'aluminium et de silicium sur lesquels peu de travaux ont été consacrés.

Oxydes d'aluminium

La spectroscopie d'absorption X a permis de mettre en évidence des complexations de sphère interne et externe avec les oxydes d'aluminium (Weesner et Bleam 1997). A la surface

de la boémite (γ-AlOOH), à pH = 6,5, pour des taux de couverture inférieurs à 0,2 mmol/m^2, un complexe de sphère interne se forme. Le cuivre divalent est ligandé à 4 atomes d'oxygène avec une distance Cu-O = 1,94 Å. Quand la charge de surface augmente, le signal se modifie et la formation d'un complexe de sphère externe de type $Cu(OH)_N(H2O)_X^{(2-N)+}$ a été proposé pour expliquer ce changement.

Adsorbé sur l'alumine (γ-Al_2O_3) à un taux de 0.60 µmol/m² le cuivre est entouré d'un Al à 2,85 Å et de 0,3 Cu à 2,96 Å. La première distance est attribuable à des partages d'arêtes avec les groupements AlOH et la seconde à la formation d'un complexe multinucléaire (Cheah et al. 2000).

Sur la gibbsite ($Al(OH)_3$) à pH = 5,48 pour un rapport Cu/minéral de 937 mmol.kg^{-1}, l'environnement local du cuivre est le suivant : 4 O à 2,04 Å, 1,2 Cu à 2,64 Å et 4,5 Al/Cu à 3,30 Å. (Chang et al. 2005).

Hydroxydes de silicium

Dans une première étude (Cheah et al. 2000), pour des pH compris entre 4,4 et 6,4 environ et des chargements de 0,6-0,7 µmol/m², il semblerait que le cuivre précipite sous la forme $Cu(OH)_2$ à la surface de la silice amorphe (SiO_2 amorphe). Dans une étude plus récente (Chang et al. 2005) pour des pH > 5,5 et un rapport Cu/minéral de 335 mmol.kg^{-1}. Il y apparait que le cuivre s'adsorberait bien à la surface, mais le signal obtenu serait un mélange entre le cuivre adsorbé et le cuivre précipité sous forme $Cu(OH)_2$. Les distances seraient les mêmes que dans la gibbsite et la pyrolusite, à savoir quatre atomes d'oxygène à 1,92 Å, entre 1 atome de cuivre à 2,66 Å et 5 atomes de Si/Cu à 3,31 Å.

3.3.3.5) Les carbonates

Structure

La calcite est un assemblage d'octaèdres CaO_6 et d'entités CO_3 planes. Les distances Ca-O sont égales à 2,36 Å (Reeder et al. 1999). Les métaux sont fréquemment associés à la calcite (Schosseler et al. 1999).

Mécanismes de fixation du cuivre

Sa surface se reconstruit en permanence entrainant l'incorporation rapide des éléments adsorbés (Cheng et al. 1998). Sur la surface la distance $Cu\text{-}O_{eq}$ est de 1,95 Å et le cuivre est ligandé à des groupements CO_3 (Fig. 2.17). Dans les sphères de coordination suivantes, les

simulations montrent que O, C et Ca seraient localisés à 2,95, 3,30 et 3,90 Å (Elzinga et Reeder 2002).

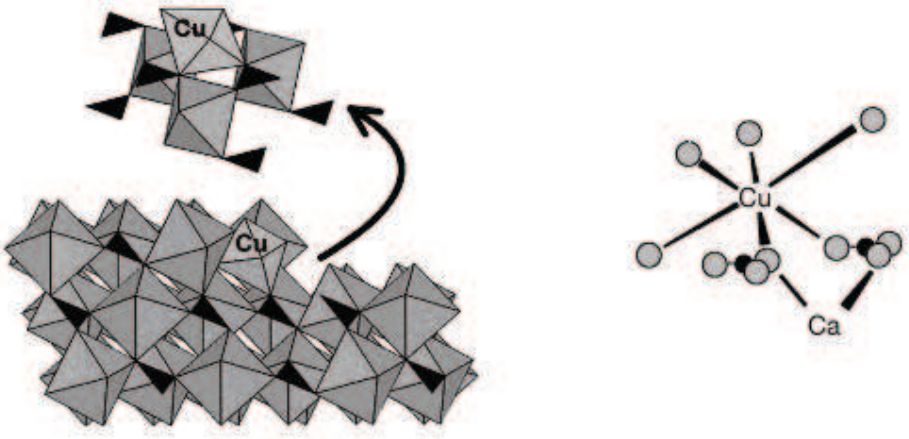


Figure 2.17 : Configuration possible de l'adsorption du cuivre sur la calcite. En première sphère de coordination, $R_{Cu\text{-}Oeq}$ = 1,95Å. Dans les sphères suivantes, O, C et Ca sont localisés à 2,95, 3,30 et 3,90 Å (Elzinga et Reeder 2002; Elzinga et al. 2006).

3.3.4) Conclusions sur les mécanismes de fixation avec la matière minérale

Les résultats précédents montrent que le cuivre peut s'associer de nombreuses manières avec les constituants du sol. Heureusement dans beaucoup de ces études les quantités de cuivre mobilisées normalisées par les quantités de sorbants sont largement supérieures à celles rencontrées dans les fonds géochimiques, et les quelques travaux sur échantillons naturels prouvent que la complexation de sphère externe est le processus majeur d'interaction. Localement lorsque ces quantités relatives de cuivre en solution ont pu être importantes, il est possible de trouver des processus de diffusion dans le réseau cristallin.

3.4) La matière organique

Il a été vu précédemment que le cuivre se complexait également avec la matière organique (MO) des sols et des sédiments. Comme pour les minéraux, l'adsorption des cations métalliques à la surface de la MO naturelle peut être décrite à partir de la modélisation des isothermes. Néanmoins, la structure de la MO est bien moins connue et présente en théorie beaucoup plus d'hétérogénéités que celle des minéraux. En particulier, une des différences majeures entre la surface réactive de la MO et celle des minéraux provient de la charge et de la nature des groupements fonctionnels, carboxyles, amines et phénoliques pour

la MO et hydroxyles pour les minéraux, ces derniers pouvant être protonés ($S\text{-}OH^{2+}$) ou déprotonés ($S\text{-}O^{-}$), (Fig. 2.9). Par exemple, les groupements de la goethite se déprotonent vers pH = 9 et la surface peut présenter une charge nette positive, négative ou nulle de pH 3 à 10-11 (Hayes et al. 1990). En revanche, les groupements réactifs de la MO sont neutres lorsqu'ils sont protonés (COOH, NH2, Ψ-OH), autrement négatifs. Les groupements COOH se déprotonant à des pH très bas, il en résulte que la MO est négativement chargée des pH = 2-3 et possède donc, en plus de sa capacité à former des complexes avec les cations métalliques, une réactivité vis-à-vis des ions positifs renforcée par l'effet électrostatique (Warren et Haack 2001). Du fait de sa polyfonctionalité, les modèles de complexation de surface présentés précédemment ne sont plus applicables.

Une partie du travail scientifique de ce projet de thèse a été d'essayer de déterminer et de comprendre la sensibilité d'une approche macroscopique à la quantification des interactions Cu-MO.

3.4.1) Structure et propriétés

La MO peut être divisée en deux catégories : les substances humiques et "non humiques". Les substances non humiques sont constituées par des composés qui n'ont pas, ou partiellement, été transformés depuis leur synthèse par les organismes vivants (molécules biologiques) ; on peut citer les acides aminés, les lipides, les lignines, les carbohydrates… (Alloway 1995; Dudal et Gerard 2004). La transformation chimique et biochimique de la matière organique aboutit, dans sa phase ultime à la libération d'ions, de CO_2, d'H_2O, SO4, NO3… (on parle de minéralisation), ou sinon à la formation transitoire (à l'échelle des grands cycles naturels) de substances humiques (SH), qui sont des polymères organiques constitués de carbone, d'hydrogène, d'oxygène, d'azote, de phosphore et de soufre (on parle d'humification). Si elles ne sont pas minéralisées, ces SH peuvent par enfouissement se transformer en charbon et hydrocarbures gazeux (méthane) ou liquide (pétrole). Elles peuvent être subdivisées en trois catégories selon leur solubilité et leur poids moléculaire: les humines, de masse moléculaire comprise entre 100 000 et 1 000 000 Da et insolubles à tout pH; les acides humiques (AH), de masse moléculaire comprise entre 10 000 et 300 00 Da et insolubles à pH<2, et les acides fulviques (AF), de masse moléculaire 500-5000 Da et solubles dans toutes les conditions de pH (Dudal et Gerard 2004; Leenheer 2009; Matynia 2009). Les propriétés physico-chimiques des SH varient, notamment, avec leur poids

moléculaire (Cabaniss et al. 2000). Ainsi, les ET sont davantage biodisponibles lorsqu'ils sont complexés aux AF (Fig. 2.18).

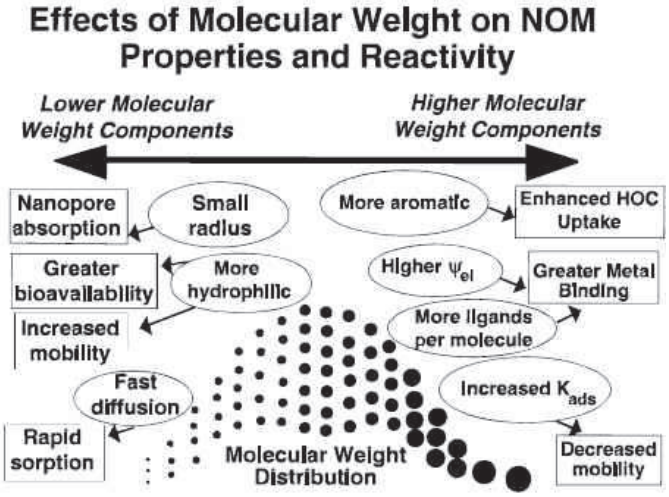


Figure 2.18 : Incidence du poids moléculaire sur les propriétés de la MO pour des compositions chimiques équivalentes (Cabaniss et al. 2000).

La structure des humines est mal connue et elles sont en général considérées comme un mélange de matière organique et d'argile liées par des liaisons covalentes ou hydrogènes (Dudal et Gerard 2004). Chimiquement les AF et AH sont des macromolécules contenant 50 à 70% de noyaux aromatiques substitués par des groupements hydroxyles et carboxyles (phénol, benzoate, phthalate, salicylate,…), et de 30 à 50% de chaines peptidiques et de polysaccharides (Fig. 2.19). La densité de sites fonctionnels (i.e. nombre de site rapporté au nombre d'atomes de carbone) est plus grande pour les AF que pour les AH (Matynia 2009).

Les courbes de titrage protonique (Q_{HT} = f(pH)) sur un intervalle pH compris entre 3 et 10 environ montre que la distribution des groupements réactifs est bimodale (Fig. 2.20). Le premier pic de réactivité, centré vers pH = 3-5, est attribué aux groupements carboxyliques portés par des carbones aromatiques ou aliphatiques et le second vers pH = 8-10 aux groupements phénoliques. L'abondance des deux types de familles de sites réactifs est égale au nombre de protons (en $mmol_C$) libérés par gramme de matière sèche à bas pH (sites acides, $Q_{1max,H}$) et haut pH (sites basiques, $Q_{2max,H}$). La mesure des valeurs de $Q_{imax,H}$ et de pK_{Hi} est difficile pour trois raisons. Tout d'abord, les plateaux attendus à chacune des extrémités des courbes de titrage qui définissent les états de protonation initiaux et finaux sont en dehors de l'intervalle de mesure. Ensuite, une fois les groupements acides déprotonés, l'état de protonation initial des groupements basiques est masqué partiellement par des groupements neutres dont les pK_H sont compris entre 5 et 8 (e.g. phosphate H_2PO_4, amine NH_2, benzenethiol C_6H_6S, anilium $C_6H_5NH_3^+$…) (Takacs et al. 1999).

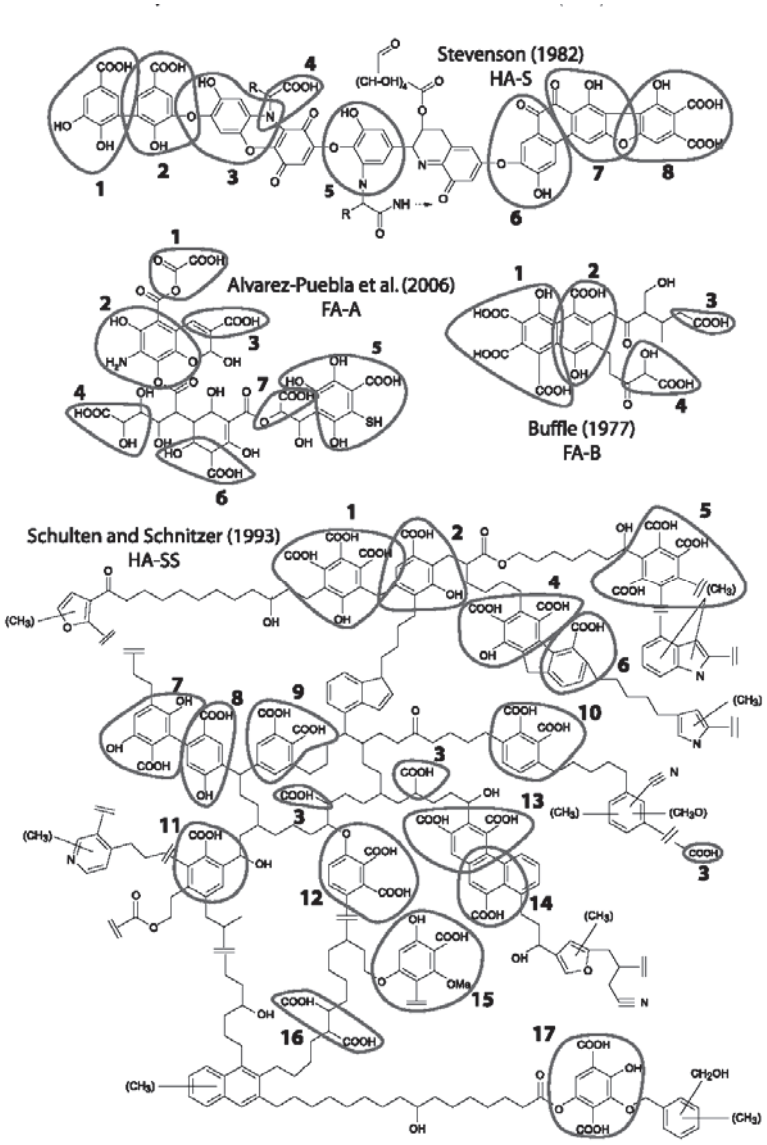


Figure 2.19 : Modèles chimiques et structuraux d'acides humiques et fulviques (Matynia et al. 2010)

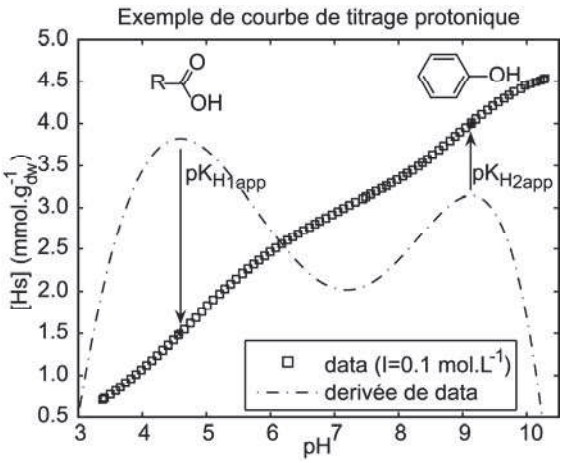


Figure 2.20 : Courbe expérimentale de titrage protonique FH22 (Milne et al. 2001) (carrés) mesurée pour une force ionique I = 0,1 mol.L^{-1} et exprimée en nombre de protons libérés par gramme de matière sèche en fonction du pH. Les constantes apparentes d'acidité $pK_{H1app.}$ et $pK_{H2app.}$ sont données par les valeurs pH des points d'inflexion (maxima de la dérivée).

Enfin, comme avec les minéraux, les valeurs pH des points d'inflexion des courbes correspondent à des pK_{Hi} apparents ($pK_{Hi\text{app.}}$) car la MO étant chargée négativement, elle retient les protons à sa surface par effet électrostatique, celui-ci variant avec la force ionique *I*. Les constantes d'acidité réelles, dites intrinsèques (pK_{Hi}), sont alors déduites des valeurs mesurées à partir d'un modèle électrostatique.

3.4.2) Approches macroscopiques

Pour déduire des informations structurales des observations macroscopiques, trois principaux modèles ont été développés pour modéliser la distribution des sites réactionnels et les mécanismes macroscopiques d'adsorption des cations à la surface de la MO (Warren et Haack 2001; Dudal et Gerard 2004). D'un coté on trouve le modèle NICA-Donnan et le modèle WHAM pour lesquels l'approche électrostatique associée est le modèle de Donnan (Tipping 1998), et de l'autre le modèle électrostatique de la sphère imperméable (Avena et al. 1999).

3.4.2.1) Comparaison des approches électrostatiques

Dans le modèle de la sphère imperméable, les substances humiques sont modélisées par des sphères imperméables et la charge est située à l'extérieur (Tipping et Hurley 1992; Avena et al. 1999). Les ions pouvant pénétrer à l'intérieur des substances humiques, cette porosité est prise en compte en optimisant la taille des particules durant la modélisation (Saito et al. 2005). Dans le modèle de Donnan, la structure des substances humiques est assimilée à un gel (Marinsky et Ephraim 1986). Cette phase est séparée de la solution par le volume de Donnan qui contient les contre-ions qui neutralisent la charge du gel (Fig. 2.21). Le modèle de Donnan est conceptuellement proche de ceux développés pour les MCS. Il s'agit d'un modèle double couche issu de la théorie de Gouy-Chapman dans lequel les paramètres intrinsèques sont déduits suivant la formule $K_{app} = K_{int}\exp(-\Delta ZF\psi/RT)$, avec ψ le potentiel de la phase mobile du gel. La différence majeure vient du calcul de ψ. Pour les minéraux, il est possible de mesurer expérimentalement la charge de surface σ et de relier ψ et σ. Ceci n'est plus possible avec la matière organique, du fait de sa charge négative permanente. Selon la théorie, pour connaitre les activités des contre- et co-ions à proximité de la surface, il est requis une intégration du potentiel par rapport à la distance de la surface, ainsi qu'une coupure de la couche diffuse à une distance arbitraire (Stumm et Morgan 1996). Ce calcul est complexe compte tenu de la géométrie des substances humiques. Dans le modèle de Donnan, la

géométrie est postulée sphérique et le système couche diffuse/électrolyte est simplifié : les co-ions (charges négatives) sont supposés être complètement absents de la sphère. Cette hypothèse conduit à la définition d'un volume de Donnan (V_D) adjacent à la surface dans lequel les charges positives sont accumulées (Tipping 1998). Ce volume, qui exprime la quantité d'eau présente dans le matériau, est soumis à un potentiel électrostatique moyen uniforme ψ_D (Kinniburgh et al. 1996). Ce potentiel relie l'activité protonique dans le volume de Donnan c_{DH} et en solution c_H suivant la formule $c_{DH} = \chi.c_H$, où χ est le facteur de Boltzmann avec $\chi = \exp(-e\psi_D/kT)$, dans lequel T est la température, e la charge de l'électron, et k la constante de Boltzmann.

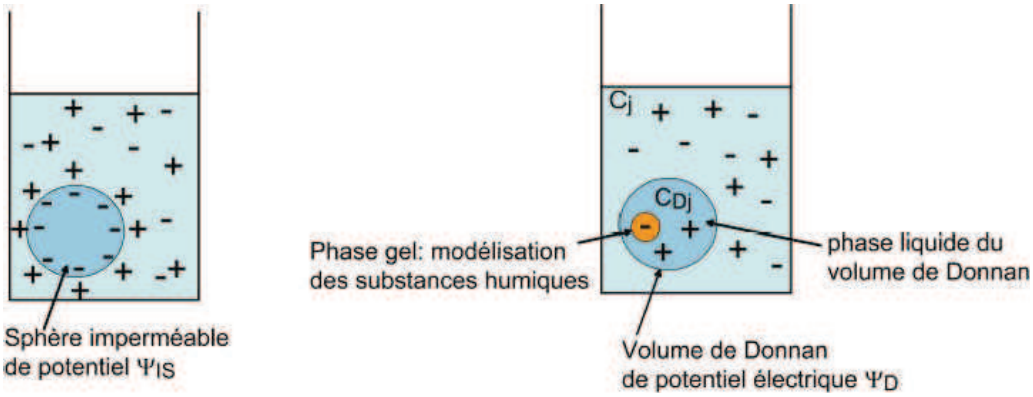


Figure 2.21 : Représentation schématique du modèle de la sphère imperméable à gauche et du modèle de Donnan à droite.

Expérimentalement, V_D suit une loi empirique qui est fonction de la force ionique I et d'un paramètre ajustable b : $\log(V_D) = b(1\text{-}\log I)\text{-}1$ (Benedetti et al. 1996; Kinniburgh et al. 1996). Quand le potentiel ψ_D et le volume de Donnan sont correctement choisis, les courbes de titrage dépendantes de la force ionique convergent toutes vers une même courbe indépendante de l'effet électrostatique appelée Master Curve (MC, Fig. 2.22). Cette courbe intrinsèque permet d'accéder aux propriétés intrinsèques du matériau. En d'autres termes, pour modéliser les propriétés acido-basiques des substances humiques, la charge de surface ne pouvant être calculée, on s'affranchit de ce paramètre en modélisant simultanément les courbes de titrages protoniques mesurées à différentes forces ioniques. Une fois la MC connue, la charge de surface est alors déduite de la formule $pH_s = pH_{bulk}\text{-}0.434(F\psi_D/RT)$, qui provient directement de la définition de l'énergie libre (Dewit et al. 1993; Saito et al. 2005).

3.4.2.2) Modèle WHAM - modélisation par distribution discrète

Le modèle WHAM suppose une distribution discrète des sites réactifs présents à la surface de la MO. Compte tenu du caractère bimodal évoqué précédemment, les distributions

sont centrées sur les valeurs pH pour lesquelles 50% des sites carboxyliques et phénoliques sont déprotonés (constantes d'équilibre acide-base moyennes pK_A et pK_B). Un étalement $\Delta pK_{A\ ou\ B}$ est ensuite associé à chacun des deux sites. Pratiquement, chaque distribution est décrite par quatre constantes d'équilibre « secondaires ». La densité des sites phénoliques ($Q_{2max,H}$) est imposée égale à la moitié de la densité des sites carboxylique ($Q_{1max,H}$).

Les interactions surface-cation avec chacun des deux types de sites sont quant à elles décrites avec trois paramètres supplémentaires $pK_{MA\ et\ MB}$ et ΔLK_1. On notera que deux rapports théoriques f_{prB} et f_{prT}, fixés car calculés à partir de considérations géométriques, déterminent la densité des sites de complexation bidentée et tridentée (Tipping 1998). Ils traduisent la proportion de sites qui sont suffisamment proches les uns des autres pour former des complexes multidentés (Dudal et Gerard 2004).

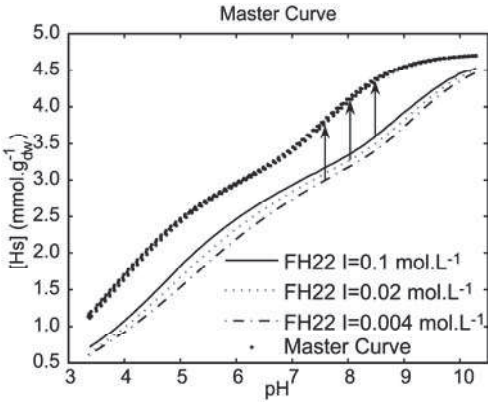


Figure 2.22 : Courbes expérimentales de titrage protonique (Milne et al. 2001) mesurées à différentes forces ioniques et la courbe intrinsèque (Master Curve) associée exprimées en nombre de protons relâchés par gramme de matière sèche de la surface vers la solution en fonction du pH.

3.4.2.3) Modèle NICA - modélisation par distribution continue

Les propriétés intrinsèques du matériau sont obtenues cette fois-ci en modélisant la MC à partir de l'équation de NICA (Milne et al. 2001). Dans le cas de la MO, la polyfonctionalité des sites acides et basiques s'exprime :

$$Q_H = Q_{1\max,H}\frac{(K_{H1}[H_s])^{m_1}}{1+(K_{H1}[H_s])^{m_1}} + Q_{2\max,H}\frac{(K_{H2}[H_s])^{m_2}}{1+(K_{H2}[H_s])^{m_2}}$$

$[H_s]$ est l'activité protonique à la surface de la MO. $Q_{Himax,H}$ est le nombre total de sites déprotonables de chaque type, K_{Hi} la valeur moyenne de leur constante d'affinité et m_{Hi} décrit l'étalement de la distribution des sites d'affinités. Dans cette expression, l'hétérogénéité des sites réactionnels est décrite selon une approche statistique continue basée sur une distribution

de Sips (Fig. 2.23). Dans l'Approximation de Condensation, les constantes moyennes d'affinité et leurs distributions peuvent être déduites directement à partir de la dérivée première de la MC (Koopal et al. 2005). A ces six paramètres ajustables, il faut ajouter Q_0, l'état de protonation initial du matériau.

En supplément, deux constantes d'équilibre intrinsèques associées à chacun des types de site, et deux paramètres (dits de Hill) reflétant l'hétérogénéité des modes de complexations sont nécessaires pour décrire les complexes métal-MO (Koopal et al. 2005).

3.4.2.5) Problèmes liés aux modélisations des courbes de titrages protoniques de la MO

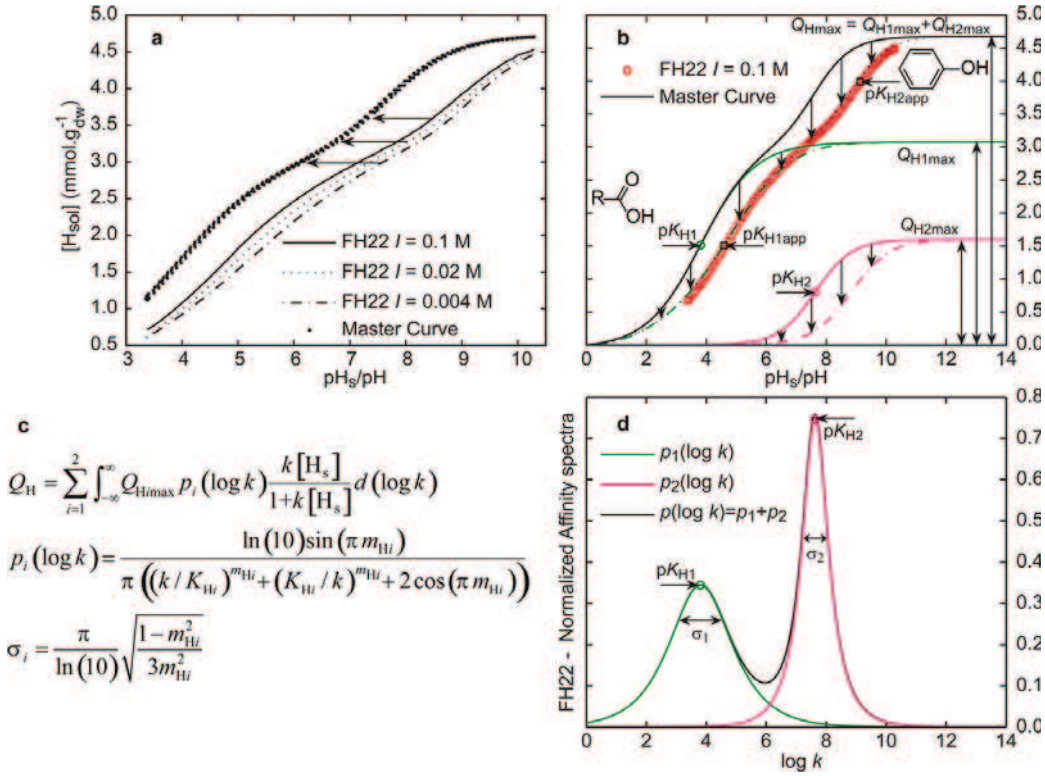
Plusieurs modèles existent donc pour décrire la dépendance en pH de la charge de surface de la matière organique. Les limites et difficultés de mise en œuvre de ces modèles sont mises en perspectives par la disparité des résultats (Tableau 2.2), même quand les approches électrostatiques sont identiques. En particulier, sur un jeu de données publié d'une cinquantaine de titrages (Milne et al. 2001), les constantes d'équilibre acide-base (pK_{H1}) déterminées avec le modèle NICA-Donnan sont particulièrement basses comparées à celles obtenues avec les autres modèles. Durant ce travail de thèse, nous avons cherché à comprendre d'où venaient ces différences, et s'il était possible de faire converger ces valeurs vers celles obtenues avec les autres modèles.

Un premier article scientifique (Annexe 2) a permis de montrer, à partir d'une approche structurale semi-empirique et du calcul des corrélations entre les variables ajustées, que (1) certains paramètres étaient corrélés et les solutions non uniques et (2) que les constantes d'équilibre acide-base déterminées avec le modèle NICA-Donnan étaient vraisemblablement sous-estimées (Matynia et al. 2010).

Dans un second article (Annexe 3), nous avons démontré mathématiquement par une Analyse en Composante Principale sur ce même jeu de données (Milne et al. 2001) que le nombre de paramètres contenus dans ce modèle était supérieur d'au moins une unité au nombre de points indépendants d'une courbe de titrage protonique dans l'intervalle de mesure. Cet excès de paramètres explique la non-unicité des solutions mathématiques du modèle NICA-Donnan (Lenoir et Manceau 2010).

Enfin, dans un troisième article (Annexe 4) nous avons calculé à partir du modèle NICA seul les densités de groupements carboxyliques et phénoliques. En contraignant alors ces paramètres dans le modèle NICA-Donnan, il a alors été possible de faire converger les

pK_{H1} du jeu de données vers des valeurs supérieures et proches de celles obtenues avec les autres modèles (Lenoir et al. 2010).



$$Q_H = \sum_{i=1}^{2} \int_{-\infty}^{\infty} Q_{Himax}\, p_i(\log k)\, \frac{k[H_s]}{1+k[H_s]}\, d(\log k)$$

$$p_i(\log k) = \frac{\ln(10)\sin(\pi m_{Hi})}{\pi\left((k/K_{Hi})^{m_{Hi}} + (K_{Hi}/k)^{m_{Hi}} + 2\cos(\pi m_{Hi})\right)}$$

$$\sigma_i = \frac{\pi}{\ln(10)}\sqrt{\frac{1-m_{Hi}^2}{3m_{Hi}^2}}$$

Figure 2.23 : (a) Données expérimentales de titrage protonique (pH) de l'acide fulvique FH 22 (Milne et al. 2001) mesurée à à trois forces ioniques et la « Master Curve » associée (pH_s). (b) Quantité de protons relâchés en solution [H_{sol}] par grammes de matière sèche en fonction du pH (carrés rouges) et meilleure reconstruction de cette donnée (points noirs). Les points d'inflexion de la donnée sont les constantes de dissociation apparentes des groupements carboxyliques et phénoliques (pK_{Hiapp}). La courbe en noir représente la MC associée aux données obtenue avec la code FIT (Kinniburgh 1999). Les points d'inflexion de la MC sont les constantes de dissociation intrinsèques pK_{Hi}. Les flèches verticales montrent l'influence des effets électrostatiques : ils retiennent les protons à la surface. Ces effets électrostatiques ne modifient pas les quantités de sites réactifs, seulement les pK. (c, d) Distributions de Sips associées aux groupements carboxyliques et phénoliques et normalisés par $Q_{Himax,H}$ (Lenoir et al. 2010).

Tableau 2.2 : Constantes d'équilibre acide-base et densités de sites obtenues avec plusieurs modèles. (Matynia et al. 2010).

		log $K_{H,COOH}$	log $K_{H,Ph\text{-}OH}$	$Q_{H,COOH}$	$Q_{H,Ph\text{-}OH}$	Method of analysis	Electrostatic interaction
Milne et al. (2001)[a]	HA	3.09 ± 0.51	7.98 ± 0.96	3.17 ± 0.89	2.66 ± 1.37	N-D	Donnan model
	FA	2.65 ± 0.43	8.60 ± 1.06	5.66 ± 1.25	2.57 ± 1.94		
Plaza et al. (2005)	HA	3.74 ± 1.09	7.68 ± 0.60	3.3 ± 0.3	1.3 ± 0.8	N-D	Donnan model
	FA	2.83 ± 0.36	7.02 ± 0.18	4.3 ± 0.6	2.3 ± 0.2		
Drosos et al. (2009)	HA	3.62 ± 0.19	8.54 ± 0.48	4.16 ± 0.58	1.34 ± 0.27	N-D	Donnan model
Tipping (1998)	HA	4.10 ± 0.16	8.80 ± 0.23	3.3 ± 0.5	1.7 ± 0.2	Model VI	Donnan model
	FA	3.20 ± 0.13	9.40 ± 0.78	4.8 ± 0.7	2.4 ± 0.4		
Gustafsson (2001)	HA	4.10 ± 0.20	8.95 ± 0.15	-	-	Stockholm Humic Model	Donnan model
	FA	3.50 ± 0.40	8.75 ± 0.30				
Avena et al. (1999)	HA	3.50 - 4.50	7.50 - 8.50	-	-	Impermeable Sphere	Impermeable Sphere
Ritchie and Perdue (2003)	HA	4.38 ± 0.13	9.72 ± 0.23	5.20 ± 0.52	0.73 ± 0.24	Modified Henderson-Hasselbalch Model	-
	FA	3.80 ± 0.11	9.78 ± 0.48	7.04 ± 0.79	0.64 ± 0.21		

3.4.3) Approche spectroscopique

Comme avec la matière minérale, les approches macroscopiques appliquées à la MO ne permettent pas d'obtenir d'informations structurales fiables sur les mécanismes de fixation du cuivre sur ce constituant. Il est donc nécessaire de changer à nouveau d'échelle d'observation. Bien que les groupements réactifs soient constitués d'éléments légers, comme le carbone, l'oxygène et l'azote, et que les propriétés de symétrie du système cristallin n'existent plus, la spectroscopie d'absorption X reste une technique performante pour obtenir ce type d'information mécanistique.

3.4.4) Mécanismes de rétention Cu-MO

Sur la base de simulations, il a tout d'abord été montré que pour un pH compris entre 4,8 et 6,3 et des concentrations entre 990 et 11000 µg $Cu.g^{-1}_{poids\ sec}$, le cuivre divalent et la matière organique formaient un complexe fort, de type chélate à cinq atomes, avec les groupements carboxyles et amines ou alcool. Une structure moyenne de type $Cu(glutamate)_2$ a été proposée (Fig. 2.24). En première sphère de coordination, on trouve quatre atomes d'oxygène/azote équatoriaux à une distance de 1,92-1,95 Å. En seconde sphère de coordination, sont situés deux à quatre atomes de carbone à une distance de 2,76-2,86 Å. Une troisième sphère de coordination composée de deux à quatre atomes de carbone/oxygène a été également détectée par spectroscopie EXAFS (Karlsson et al. 2006; Karlsson et al. 2008).

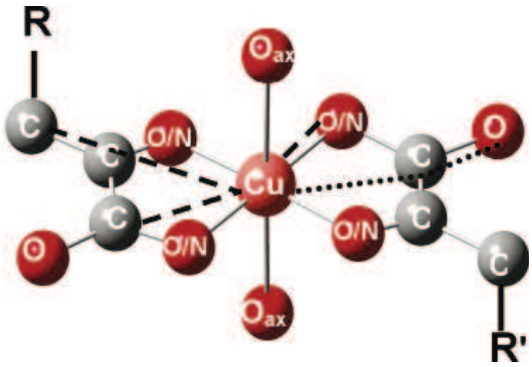


Figure 2.24 : Modèle de complexation Cu^{2+}-MO du type chélate double $Cu(glutamate)_2$. R et R' représentent des chaines carbonées (Karlsson et al. 2006).

Plus récemment, et en accord avec les modèles récents de structure des MO obtenus par spectroscopies infrarouge et RMN, à un pH compris entre 4,5 et 5,5, le chélate a cinq atomes a été confirmé à basse concentration ([Cu] <300 µg $Cu.g^{-1}_{poids\ sec}$) mais il a été possible de différencier les ligands N (de l'amine) et O (de l'alcool) grâce à l'étude de composés de référence judicieusement choisis et un complexe $Cu(malate)_2$ a été énoncé (Fig. 2.25). A plus

haute concentration, les cations Cu^{2+} excédentaires forment un chélate à six atomes de type Cu-(malonate)$_{1-2}$ (Manceau et Matynia 2010). Il a été proposé que les sites COOH-COH-R, caractérisés par une fonction alcool portée par un carbone α à proximité d'un groupement carboxylique, avaient une plus forte affinité pour le cuivre que les sites dicarboxyliques COOH-CH-COOH car les cycles chélates à 5 chaînons sont généralement plus stables que ceux à 6 chaînons. Ces chélations se produisent du fait de l'accord stérique entre les liaisons du cycle et la dimension du plan équatorial de l'ion Cu^{2+} en coordinence 6. Cette adéquation permet d'expliquer l'affinité du cuivre pour la matière organique.

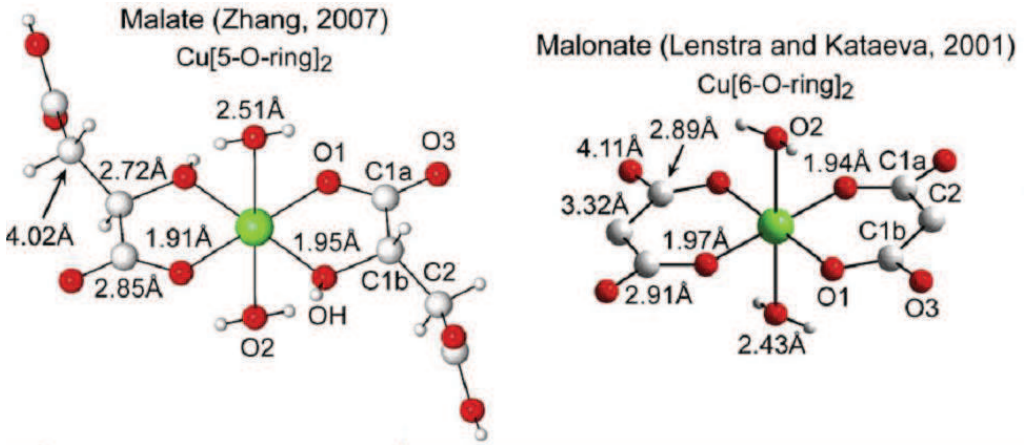


Figure 2.25 : Complexes Cu-MO à basse (malate) et haute (malonate) concentrations en cuivre (Manceau et Matynia 2010).

Chapitre 3

Détermination de la spéciation du cuivre par couplage µ-XRF, µ-XANES et ACP.

Application au porphyre cuprifère de Mamut (Malaisie) et au bassin de rétention et d'infiltration d'eaux pluviales de Django Reinhardt (France)

1) Principe de l'approche

A partir de la simulation des signaux, la spectroscopie d'absorption X possède de nombreux atouts pour l'étude de la spéciation des éléments traces dans les sols car elle permet d'obtenir des informations structurales. Malheureusement, sa principale limite est inhérente à la nature des échantillons naturels. En effet, les sols et les sédiments sont des milieux hétérogènes et un élément est généralement présent sous plusieurs formes (systèmes multiphasiques). Comme le signal mesuré reflète l'environnement moyen de l'élément étudié, l'XAS de poudre est donc peu sensible aux formes minoritaires, alors masquées par les formes majoritaires. Ce phénomène est accentué lorsque ces formes minoritaires sont constituées d'atomes légers. Ainsi, le cuivre ayant une forte affinité pour la matière organique, et celle-ci étant hétérogène, l'application de cette technique à cet élément est délicate.

Cette difficulté peut être contournée en complétant les mesures d'XAS de poudre en couplant par des mesures de micro-fluorescence X (μ-XRF), de micro-spectroscopie X (μ-XAS), et de micro-diffraction X (μ-XRD). La micro-fluorescence X donne accès à la répartition des éléments chimiques majeurs et en traces et donc à leurs associations. Mais des conclusions déduites de cette seule technique pourraient se révéler infondées, voire fausses. Supposons, par exemple, que l'on observe sur une carte de fluorescence X une association Cu-Fe. Il est tentant de conclure que le cuivre est inclus dans une seule espèce ferrifère, par exemple la chalcopyrite $CuFeS_2$. Or ce minéral peut être recouvert d'oxyhydroxyde de fer issu de l'oxydation du sulfure et dans lequel une fraction des atomes de Cu aurait été piégée. Dans le minéral primaire Cu est un élément majeur, tandis que dans le minéral secondaire c'est un élément mineur ou trace selon sa concentration. Les atomes de cuivre n'ayant pas le même environnement structural dans chacun des cas, cette ambiguïté peut être levée par μ-XAS, d'autant que la fraction de Cu contenue dans chacune des phases porteuses du système binaire devrait normalement varier d'un point d'analyse à l'autre facilitant ainsi leur discrimination (Fig. 3.1).

Dans les sols, la taille des hétérogénéités peut descendre jusqu'au nanomètre et les signaux obtenus par μ-XAS restent alors fondamentalement des signaux moyens, cependant il est tout de même possible à cette échelle d'obtenir le signal d'espèces pures dont la taille est supérieure à celle du faisceau (Fig. 3.2) et la contribution des espèces minoritaires est ponctuellement renforcée, ce qui facilite leur identification. Dans la pratique, la spectroscopie

µ-XANES, dont le temps d'acquisition est relativement court (une vingtaine de minutes), permet d'analyser un grand nombre de points. Lorsqu'un spectre µ-XANES est suffisamment différent des précédents, un spectre µ-EXAFS est enregistré. Sur l'ensemble des signaux µ-XANES obtenus, les proportions des espèces sont variables. L'Analyse en Composantes Principales (ACP) permet alors de déterminer le nombre d'espèces pures présentes dans la matrice. Une opération mathématique associée, la transformation cible (Target Transformation en anglais), permet de déterminer si le spectre d'une espèce connue (référence) fait partie des composantes du jeu de données et donc de déterminer si cette espèce est présente ou non dans le système étudié. Dans le cas d'espèces pour lesquelles il n'existe pas de référence, le couplage avec la µ-XRD fournit des informations sur la structure cristalline éventuelle de la phase porteuse. Une fois le nombre et la forme des espèces connues, la spectroscopie d'absorption X de poudre permet de quantifier la spéciation de l'élément étudié, c'est-à-dire de déterminer la proportion de chacune des espèces dans le volume analysé. Cette démarche analytique est largement utilisée pour étudier la spéciation des éléments traces dans les sols naturels (Manceau et al. 2002; Manceau et al. 2004; Sarret et al. 2004; Panfili et al. 2005; Isaure et al. 2005; Kirpichtchikova et al. 2006).

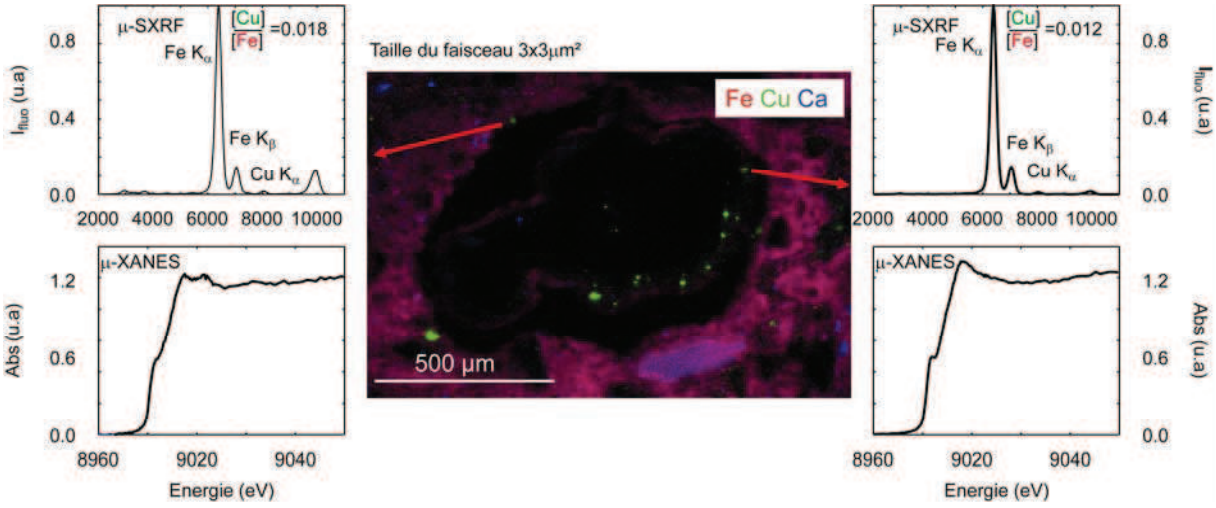


Figure 3.1: Cartographie µ-XRF et spectres µ-XRF et µ-XANES de deux spots. Pour chacun des spots le rapport de l'intensité de fluorescence da la raie K_α du cuivre sur l'intensité de fluorescence de la raie K_α du fer ($I_{fluo.}K_\alpha[Cu]/ I_{fluo.}K_\alpha[Fe]$) est proche, pourtant les signatures spectrales µ-XANES au seuil K du cuivre sont très différentes. Le fer est en rouge, le cuivre en vert et le calcium en bleu.

Toutefois, cette méthode d'identification et de quantification présente quelques limites. Il faut tout d'abord que les fractions de chaque espèce varient d'un point d'analyse à l'autre et que les signatures spectrales des espèces soient suffisamment différentes (au sens ou les différences doivent être significatives vis-à-vis du bruit présent sur les données). Ensuite, son succès dépend de la base de données des spectres XAS de référence, qui doit être la plus

complète possible. Enfin, le choix du nombre et des spectres de référence à utiliser pour la reconstruction du spectre de poudre se fait à partir de l'analyse des résultats de l'ACP. A l'heure actuelle, les indicateurs usuels associés à cette technique présentent des comportements atypiques (Beauchemin et al. 2002) et les jeux de données étudiés étant relativement limités (une trentaine de spectres au maximum) le choix est souvent fait à partir de l'appréciation de l'expérimentateur. Or pour une évaluation la plus rigoureuse possible de l'ensemble des formes de l'élément, il est indispensable de multiplier les observations puis de traiter l'information par une approche statistique fiable.

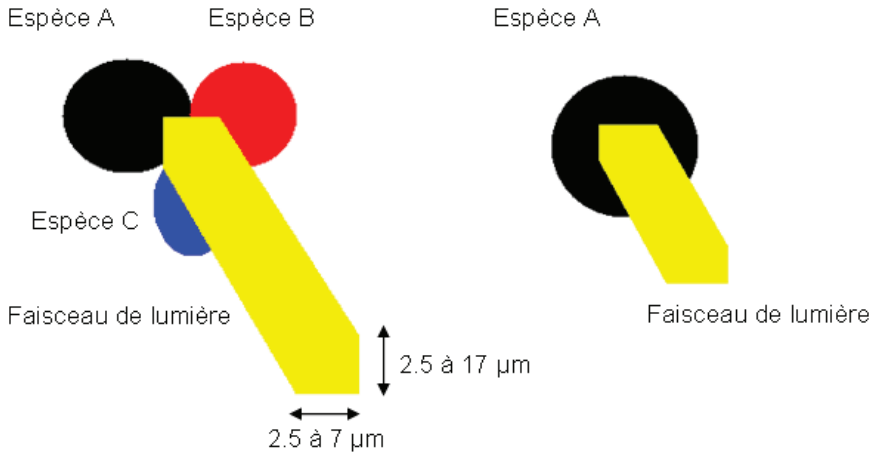


Figure 3.2 : Principe du µ-XAS. Dans cette technique la taille du faisceau X est de l'ordre de quelques dizaines de µm². Dans le cas de gauche, le signal obtenu est la moyenne pondérée des signaux des espèces A B et C. Dans le cas de droite le signal obtenu est celui de l'espèce A.

L'objet de la suite de ce chapitre est de présenter, au travers de l'étude de cas expérimentaux représentatifs des pollutions au cuivre dans les sols (mine de cuivre et rejets anthropiques en milieu périurbain), l'application de cette approche couplée qui mettra en évidence les difficultés rencontrées pour la détermination du nombre de phases cuprifères.

2) Présentation des sites

2.1) Porphyre cuprifère de Mamut

2.1.1) Données géographiques, climatiques et géologiques

La mine de cuivre de Mamut est située dans l'Etat du Sarawak (Malaisie orientale) de l'île de Bornéo à une distance de soixante kilomètres de Kota Kinabalu (capitale de la province de Sabah) sur le flanc sud est du Mont Kinabalu (4101 m), et à une altitude de 1300 mètres (Fig. 3.3). La province de Sabah présente un climat de type équatorial. Les températures varient en moyenne de 23 à 31°C au cours de l'année. De mars à septembre, le climat est sec et chaud. La période humide dure de novembre à février. La pluviométrie annuelle est de 2400 mm. L'humidité atmosphérique est relativement importante sur les basses terres, généralement entre 85 et 95% (Lanson 2005).

Les réserves cuprifères de Mamut sont génétiquement associées au plutonisme Miocène supérieur du Mont Kinabalu. Elles sont localisées dans l'un des nombreux centres minéralisés, situés le long d'une zone tectonisée d'orientation nord ouest – sud est parcourant la province de Sabah, de part et d'autre du massif du Kinabalu. On y trouve de nombreuses failles orientées Nord-Nord Est. Ces structures se sont développées dans une séquence sédimentaire clastique et serpentinisée, dont le dépôt s'étale de l'Eocène au Miocène. Quelques strates éparses de roches volcaniques (tufs en fines couches) s'intercalent également dans la séquence sédimentaire. Ces roches subissent au Miocène supérieur une intrusion par des porphyres adamellitique et granodioritique, riches en potasse. Le cuivre et l'or sont dispersés. Ils sont localisés dans la séquence sédimentaire encaissante comme dans les filons porphyriques. Globalement, les minerais sont de nature plutôt quartzique. Les minéraux primaires dominants dans les dépôts de Mamut sont la pyrite (FeS_2), la pyrrhotite (FeS) et la chalcopyrite ($CuFeS_2$). On y trouve également en moindres quantités de la sphalérite (ZnS), de la galène (PbS), ainsi que de l'or (Au) et de l'argent (Ag) natifs (Imai 2000; Kosaka et Wakita 1978). Le gisement présentait ainsi à l'origine une réserve totale de 179 millions de tonnes de minerai à 0,476 pourcents de cuivre et 0,5 grammes par tonne d'or. Ceci équivaut, après raffinage du concentré métallique, à 24 000 tonnes de cuivre, 12 tonnes d'argent et 1,5 tonne d'or extraites par an. L'exposition constante aux forces érosives de la pluie et du vent provoque la présence de taux anormalement élevés de divers métaux dans tout le système hydrographique environnant. Dès la fin des années 60, ces anomalies géochimiques sont repérées dans les sols, les sédiments et les eaux. Suite à cette découverte un consortium minier

japonais a mené une campagne de prospection, puis le site a été exploité selon la méthode conventionnelle dite de l'«open pit».

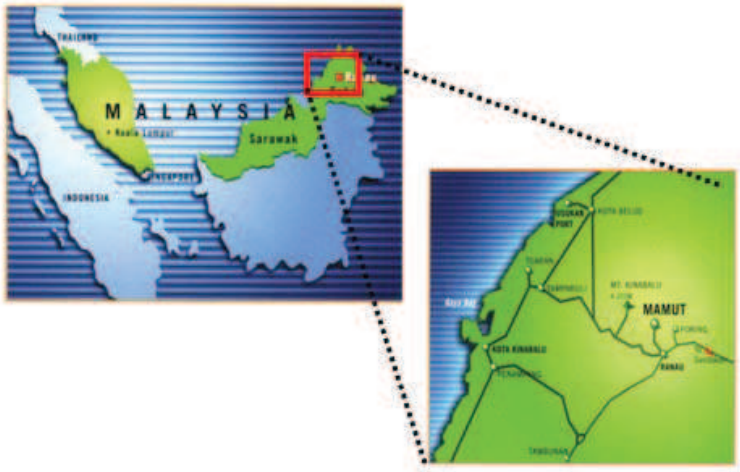


Figure 3.3: Situation géographique générale du porphyre cuprifère de Mamut. Il est situé sur l'île de Bornéo, en Malaisie, à une distance de soixante kilomètres de Kota Kinabalu (capitale de la province de Sabah) sur le flanc sud est du Mont Kinabalu (4101 m), à une altitude de 1300 mètres.

La fosse est quasi circulaire, avec un diamètre d'environ 1200 m et une profondeur de 500 m. Les déchets solides ('tailing'), principalement de fins fragments de roches infra-millimétriques voire micrométriques résultant des divers stages du traitement du minerai, ont été transportés du pit vers une décharge ('dam') rectangulaire d'une surface de 44 hectares située à une altitude de 400 m, à une dizaine de km, via une canalisation. La quantité totale de ces boues chargées en métaux lourds et stockées dans la décharge est estimée à 160 millions de tonnes. Elles contiennent de grandes quantités de métaux lourds, dont Cu, Ni, Zn et Co, principalement présents sous forme de minéraux sulfatés (Jopony et al. 1987), donc facilement solubilisables. La mine a été mise en activité en 1975 et elle a fermé définitivement en 1999 (Fig. 3.4).

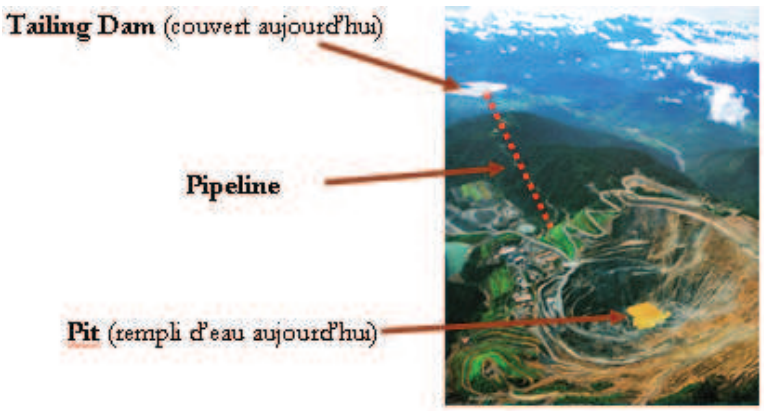


Figure 3.4 : Vue aérienne de la mine.

2.1.2) Problèmes environnementaux

2.1.2.1) Accidents d'exploitation

La mine de Mamut a provoqué, au début de son exploitation, deux pollutions accidentelles majeures. En 1975, un jour de fortes précipitations, une fuite sur la conduite liant la fosse à la décharge a contaminé 1000 hectares de rizières. En 1977, également un jour de précipitations exceptionnelles, les déchets miniers de la décharge ont été lessivés par les eaux météoriques. Les terres agricoles s'en trouvèrent une nouvelle fois souillées. Suite à ces deux accidents, le contrôle du comportement érosif des déchets miniers de la décharge et les caractéristiques techniques du pipeline furent remis en question. La décharge fut redessinée et un réseau de canaux bétonnés construit pour le drainage des eaux de pluie. Par ailleurs, en vue de minimiser la quantité de matériaux transportés vers la rivière par les eaux de ruissellement, un barrage de béton a été construit. Enfin, les zones de décharge ont été végétalisées ce qui a encore diminué le risque de nouveaux accidents en stabilisant les boues cuprifères. Jusqu'à sa fermeture en 1999, aucun nouvel incident n'a été recensé.

Néanmoins, même si des mesures ont été prises et la mine fermée, le risque d'accident n'est toujours pas nul et les populations voisines s'inquiètent de l'exfiltration éventuelle hors du barrage de ces 160 millions de tonnes de produits toxiques. Il est possible également que les parois du barrage finissent par céder. Dans ce cas, on assisterait à une catastrophe similaire à celle ayant eu lieu à Baia Mare en Roumanie en janvier 2000, et qui a détruit les écosystèmes de la rivière Tisza et largement pollué le Danube avec des composés du cyanure (Souren 2000). Par ailleurs, la décharge est aujourd'hui partiellement asséchée et des poussières de taille micrométrique sont dispersées par le vent dans le milieu environnant.

2.1.2.2) "Acid Mine Drainage"

Outre ces pollutions majeures mais anciennes et ponctuelles et les contaminations actuelles par voie éolienne s'ajoute une contamination régulière des sols par un autre phénomène appelé drainage acide ("Acid Mine Drainage", AMD) lié à l'oxydation biotique et abiotique des minerais sulfurés exposés aux intempéries, suivie de la libération dans les eaux de drainage de grandes quantités d'acide sulfurique (SO_4^{2-} et H^+) et de métaux. Selon la composition du minéral, il peut s'agir de fer ou de métaux lourds, tels que le cuivre, le plomb et le mercure. Sur le site minier, les fortes teneurs en cuivre et fer se remarquent par la présence de précipités secondaire verts, bleus, ocres, rouges, oranges ou jaunes (Fig. 3.5). La

dissolution oxydative de la pyrite et des autres sulfures se déroule en plusieurs étapes selon les réactions suivantes :

$$FeS_{2\,(s)} + \frac{7}{2} O_2 + H_2O \rightarrow Fe^{2+} + 2\,SO_4^{2-} + 2H^+ \qquad (1)$$

Le Fe(II) dissout peut être oxydé en Fe(III) :

$$Fe^{2+} + \frac{1}{4} O_2 + H^+ \rightarrow Fe^{3+} + \frac{1}{2} H_2O \qquad (2)$$

Le Fe(III) peut aussi oxyder la pyrite, et produire du Fe(II) :

$$FeS_{2\,(s)} + 14\,Fe^{3+} + 8\,H_2O \rightarrow 15\,Fe^{2+} + 2\,SO_4^{2-} + 16H^+ \qquad (3)$$

En fonction du degré d'oxydation du fer, divers solides secondaires peuvent précipiter (Fortey 2004). Les eaux de percolation qui s'écoulent de la mine abandonnée ont des pH inférieurs à 3,5 provoquant une baisse de la qualité des eaux (Murtezda et Lee 1985). La rivière Mamut, dans laquelle se déversent les eaux de ruissellement de la mine et la rivière Lohan située à proximité de la décharge (Fig. 3.6), présentent les caractéristiques ordinaires d'un cours d'eau subissant des dégradations causées par les AMD (Fig. 3.5) : pH acide, conductivité élevée, forte concentration de sulfates et précipités d'hydroxydes de fer.

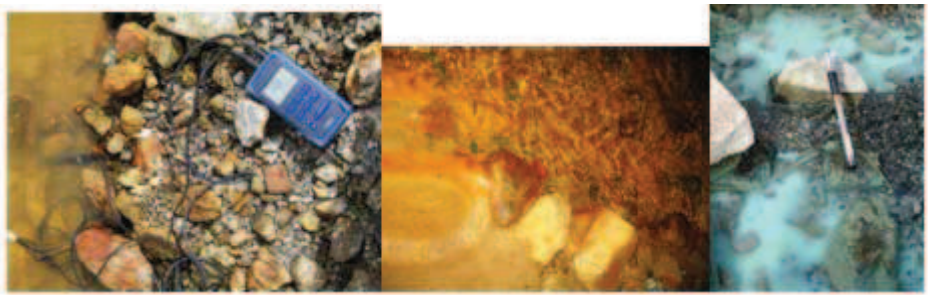


Figure 3.5 : Effets des AMD autour de la mine : dépôts d'oxyhydroxydes de fer en orange et de sulfate de cuivre en bleu. Le pH des eaux à proximité immédiate de la mine est inférieur à 3,5.

2.1.3) Echantillonnage

Les échantillons de sol ont été prélevés au mois de décembre de l'année 2005 le long des bassins versants de la mine et de la décharge. Les points B1, B3, B4, B5 sont situés dans des rizières en activité le long des rivières Lohan et Bongkud. En particulier B1, B3 et B4 sont caractéristiques de la pollution résultant de la décharge. Les points A3, C3, D3 sont situés près des rivières Mamut et Langanan, dans des zones où il est prévu de développer la culture du riz et sont représentatifs de l'évolution spatiale de la pollution résultant de la mine. Les points T1, S1, S2 sont situés à proximité de la décharge dans des zones éventuellement cultivables. Enfin, les échantillons D2 et A1 sont des points de contrôle assez éloignés des

deux sources de métaux qui devraient permettre d'évaluer l'extension de la contamination (Fig. 3.6). Les matrices de sols ont été prélevées au couteau et à la tarière sur des profondeurs de 30 cm. Des répliquats ont été faits dans un rayon d'une cinquantaine de cm.

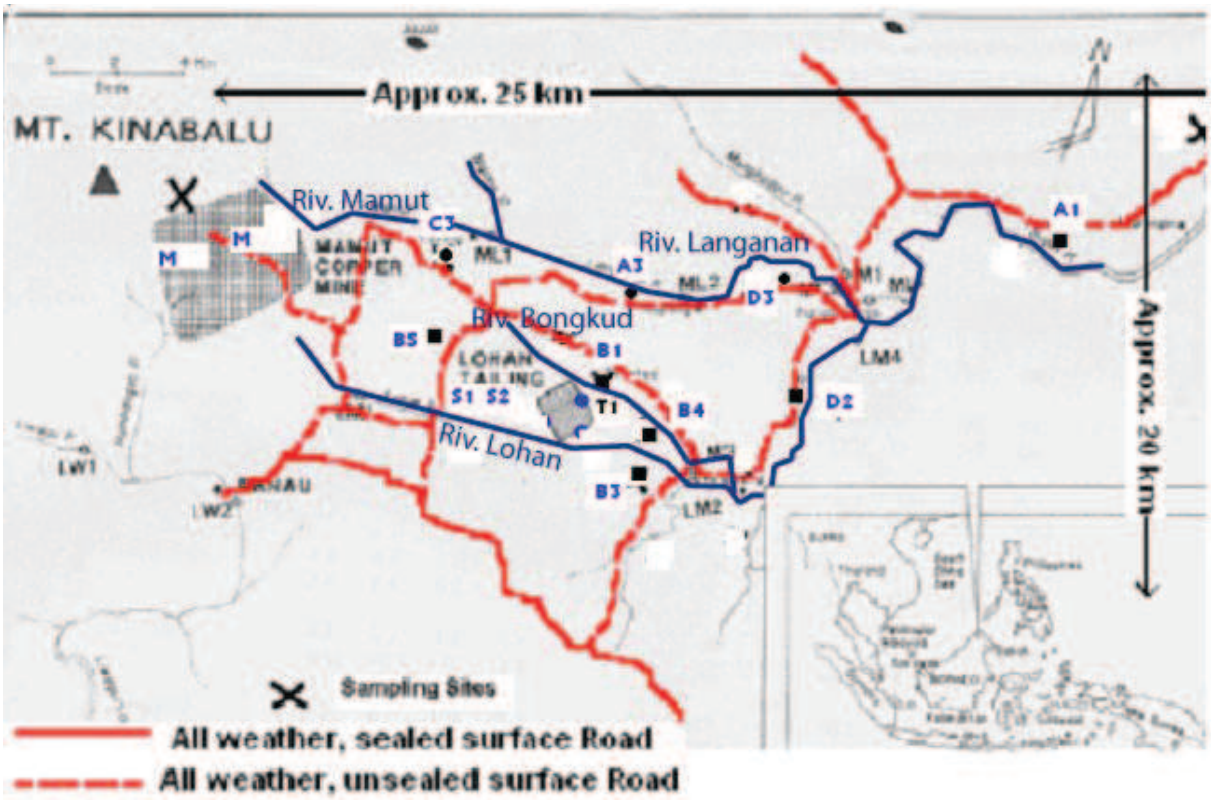


Figure 3.6 : Aire de l'étude et emplacement des échantillonnages. Les points ont été choisis de façon à caractériser, d'une part, l'étendue de la pollution due à la mine dans le système hydrographique Mamut-Langanan (C3 / A3 / D3) et, d'autre part, l'étude de la pollution provenant de la décharge dans le système hydrographique Lohan-Bongkud (B1 / B3 / B4 / D2). Le point A1 se trouve à la confluence des deux systèmes.

2.2) Bassin d'infiltration et de rétention Django Reinhardt (DjR)

2.2.1) Données géographiques et fonctionnelles

Le drainage et l'accumulation dans les exutoires des rejets urbains par temps de pluie constituent une des sources principales de pollution produite par les agglomérations. Les techniques classiques de réseau posent de nombreux problèmes techniques et environnementaux. La concentration urbaine et son développement conduisent à collecter et transporter des quantités d'eaux de plus en plus importantes (accroissement des volumes et des débits de pointe) qui provoquent des inondations chroniques et sévères. De plus, des recherches antérieures ont montré que les eaux de ruissellement, ainsi que celles qui ont transité dans les réseaux séparatifs ou unitaires, sont très chargées en métaux lourds, en hydrocarbures et parfois aussi en autres composés organiques naturels ou de synthèse

essentiellement portés par la phase particulaire et colloïdale (Chebbo 1992; Ellis 1991), si bien que les rejets urbains par temps pluvieux sont de plus en plus souvent mis en cause dans la pollution des cours d'eau servant d'exutoire (Chocat 1997). Pour maîtriser ces rejets, de nombreuses villes ont développé des méthodes de gestion des eaux pluviales reposant sur des techniques alternatives aux réseaux de conduites, fondées sur la collecte et l'infiltration des eaux de ruissellement. Ces techniques s'inscrivent dans une mouvance de contrôle à la source et de gestion dite « verte » ou « douce » des eaux pluviales en milieu urbain. Leurs qualités sont nombreuses. L'infiltration des eaux dans des zones semi-naturelles diminue les risques d'inondation et contribue à la recharge des nappes souterraines. Les polluants organiques et inorganiques sont concentrés à la surface des ouvrages, ce qui facilite leur dégradation et leur gestion, et non plus dispersés dans la nature. Enfin, les zones humides végétalisées sont bien acceptées de la population car elles s'intègrent facilement dans le « paysage » urbain.

Le site d'étude est un bassin de rétention et d'infiltration d'eaux pluviales situé sur la commune de Chassieu dans la banlieue Est de Lyon (France) et appelé Django Reinhardt (Fig 3.7). Ce bassin s'étend sur 2 ha et peut contenir environ 30 000 m^3 d'eau. Il draine une zone industrielle d'une surface de 185 ha. Il est composé de deux bassins:

- un bassin de décantation imperméable qui sert à décanter une partie de la charge sédimentaire transportée par les eaux de ruissellement et à transporter les volumes d'eaux à infiltrer,
- le bassin d'infiltration en lui-même, constitué du sous-sol en place, perméable.

Le substrat naturel du site est un dépôt fluvioglaciaire datant du quaternaire (Winiarski et al. 2006). La nappe d'eaux souterraines est à environ 13 m de profondeur sous la surface du bassin d'infiltration. L'infiltration des eaux y est rapide : le sol (ou sédiment) en place a une conductivité hydraulique saturée variant de 7.10^{-3} à 9.10^{-3} m/s (BURGEAP 1995). Les infiltrations successives des eaux pluviales sont à l'origine du dépôt d'une couche de sédiments fins de plusieurs cm à la surface du bassin. Cette couche peut limiter les vitesses d'infiltration grâce au phénomène de colmatage (Barraud et al. 2002; Lassabatere et al. 2010).

Ce bassin est un site d'étude pilote de l'Observatoire de Terrain en Hydrologie Urbaine (OTHU). A ce titre il est fortement instrumenté. Les débits entrants, la pluviométrie, la turbidité et le pH des eaux entrantes, la hauteur d'eau dans le bassin, etc… sont suivis en continu. Des piézomètres installés en amont et en aval du bassin permettent le suivi de la qualité des eaux atteignant la nappe. Les eaux entrant dans ce bassin sont de deux types : les eaux de ruissellement qui sont la conséquence d'épisodes pluvieux, et les eaux dites de temps secs qui constituent un apport continu d'eau (eaux de procès ou de lavage, etc.). Certaines

caractéristiques de ce bassin le distinguent des autres bassins existant dans la région lyonnaise : sa grande taille, le type industriel du bassin versant qu'il draine et sa (récente) colonisation végétale naturelle (Badin 2009).

Figure 3.7 : Situation des bassins de rétention et d'infiltration de la zone industrielle située au sud de la ville de Chassieu à l'est de l'agglomération lyonnaise. Les zones de prélèvement sont situées à l'entrée du bassin dans une zone plus sollicitée par l'infiltration et au fond dans une zone moins sollicitée (Badin 2009).

2.2.2) Problèmes environnementaux

Ces techniques de rétention/infiltration soulèvent néanmoins de nombreuses questions. La principale concerne leurs performances réelles vis-à-vis de la préservation durable des écosystèmes et l'efficacité de ces « Eco-infrastructures » doit être validée scientifiquement, en particulier pour ce qui concerne le rôle épuratoire des zones humides (Walsh 2000; Datry et al. 2004). En effet, les quantités de charges polluantes que les eaux pluviales sont susceptibles de véhiculer incitent à la prudence car elles peuvent conduire à l'accumulation durable de polluants en sub-surface, voire à leur migration dans les nappes souterraines. Autrement dit, la protection des cours d'eau par la collecte et la gestion des eaux de surface ne

doit pas se faire au détriment de la qualité des sols et sédiments et des eaux souterraines. Par ailleurs, même si l'efficacité des phénomènes de décantation et de piégeage de la pollution dans les couches superficielles du sol des ouvrages semble être importante (Mikkelsen et al. 1994; Pitt et al. 1999; Dechesne 2002; Badin et al. 2008), leur seuil de saturation (e.g. capacité épuratoire) n'est pas connu, le devenir des produits de curage éventuels n'est toujours pas réglé et le bilan global jamais réalisé (Ruban et al. 2005).

2.2.3) Echantillonnage

Les échantillons ont été prélevés en juillet 2006. Certaines zones du bassin d'infiltration étant plus sollicitées du fait de leur proximité du bassin de rétention, deux zones ont été retenues (Fig. 3.7). L'une sollicitée à chaque épisode pluvieux au niveau de l'entrée du bassin (zone A) et l'autre plus au fond du bassin (zone B) uniquement sollicitée en cas de pluie soutenue. Dans les deux zones, le « sol », au sens générique du terme, est constitué d'une couche de sédiment de 3 à 5 cm puis d'un dépôt fluvio-glaciaire dont les nombreux cailloux et galets donnent une forte incohésion au milieu. Il est riche en matières organiques/hydrocarbures et la végétation assez bien développée. Etonnement, alors que la zone A est d'ordinaire plus humide que la zone B, au moment des prélèvements de l'eau s'infiltrait en B alors que A était complètement sèche. Les prélèvements ont porté sur du sol et des échantillons de plantes (Fig. 3.8). Les échantillons ont été prélevés à la pelle et à la main sur des profondeurs de 0-5 cm à 0-10 cm. Des répliquats ont été faits dans un rayon d'une cinquantaine de cm.

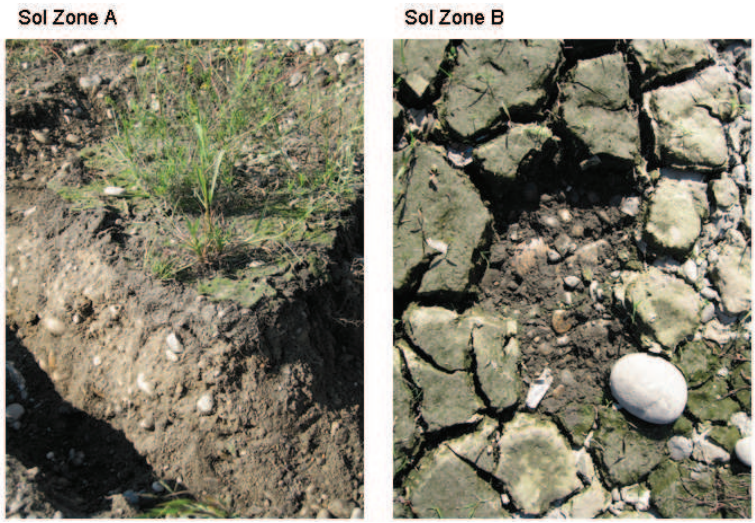


Figure 3.8 : Sol des zone A et B. L'horizon de surface est constitué d'une couche de sédiments de 3 à 5 cm puis d'un dépôt fluvio-glaciaire. Dans la partie profonde (>5-10cm) non humide, les nombreux cailloux et galets donnent une forte incohésion au milieu.

3) Méthodes

3.1) Préparation des échantillons et observations préliminaires

Les échantillons ont été congelés (-18°C) dès réception au laboratoire pour stopper les changements éventuels de spéciation, liés notamment au contact avec l'oxygène de l'air. Une partie des échantillons a été utilisée pour doser les éléments traces métalliques (Cu, Ni, Mn, Zn) et le fer par spectrométrie de masse avec plasma à couplage inductif (ICP-MS).

Les autres échantillons ont été lyophilisés puis des blocs de sol non-perturbé imprégnés dans de la résine époxy, découpés et amincis à 30 μm. Ces lames minces ont tout d'abord été observées à la loupe binoculaire, pour caractériser l'échantillon dans sa globalité (présence de matière organique, sable, argile…, texture) et identifier des zones d'intérêt. On recherche plus particulièrement les zones racinaires, ou rhizosphères, car l'acidification du milieu et l'excrétion de molécules organiques complexantes liées à l'activité métabolique des racines et des micro-organismes associés sont susceptibles de modifier la spéciation des métaux, et l'interface sol-racine est le point d'entrée principal du cuivre minéral dans la biosphère. On recherche aussi les précipités d'oxydes de fer et de manganèse, repérables visuellement à leur couleur, car ces minéraux secondaires ont une forte affinité pour le cuivre.

Sur chacun des deux sites, deux lames minces ont été étudiées (Mamut : C3 et B4 ; DjR : A et B). Les noms correspondent à la zone d'échantillonnage.

3.2) Mesures par rayonnement X

Les lames minces ont été étudiées par μ-XRF, μ-XAS (μ-XANES + μ-EXAFS) et μ-XRD à l'Advanced Light Source de Berkeley, et la spéciation moyenne par XAS de poudre a été obtenue sur la ligne FAME de l'ESRF de Grenoble. Cependant, les mesures de μ-EXAFS, μ-XRD et XAS ne sont pas incluses dans le manuscrit car leur analyse n'est pas achevée. Les mesures ont toutes été effectuées à température ambiante.

3.2.1) Caractéristiques du microspectromètre 10.3.2

Les principales caractéristiques de la ligne de microspectrocopie X 10.3.2 de l'ALS sont les suivantes (Marcus et al. 2004):

- la gamme d'énergie s'étend de 2,5 à 17 keV ;

- le monochromateur est constitué de deux cristaux de silicium taillés selon le plan réticulaire Si(111) ;

- le détecteur de fluorescence est constitué de 7 cristaux semi-conducteurs de germanium ;

- la taille du faisceau sur l'échantillon est ajustable de 2,5 x 2,5 μm² (fentes au minimum) à 17 (H) x 7 (V) μm² (fentes ouvertes).

Les spécificités de cette ligne, associées à l'intensité lumineuse fournie par l'anneau de stockage d'électrons (1,9 GeV, 400 mA), permettent d'obtenir pour le cuivre une limite de détection de l'ordre du ppm en μ-XRF et de la centaine de ppm en μ-XAS.

3.2.2) Cartographie de μ-XRF

Les cartographies de μ-XRF ont été faites en deux étapes. Les lames minces ont tout d'abord été scannées avec un pas de 20 x 20 μm à une énergie d'excitation de 10 keV de façon à visualiser à grande échelle la distribution du cuivre (seuil K = 8979 eV) et du zinc (seuil = 9649 eV). Puis un nombre restreint de régions représentatives de l'ensemble ont été scannées à haute résolution (5 x 5 μm) à deux énergies de part et d'autre du seuil K d'excitation du cuivre (+/- 20 eV) pour s'affranchir par différence de la contamination dans la raie de fluorescence K_α du cuivre (8040 eV) de la raie parasite K_β du nickel (8265 eV) et du signal élastique (diffusion Rayleigh) et inélastique (effet Compton).

Ponctuellement, sur les spots étudiés par μ-XANES, les spectres μ-XRF ont été enregistrés à une énergie d'excitation de 10 keV.

3.2.3) Spectroscopies μ-XANES et μ-XRF

Les spectres μ-XANES au seuil K du cuivre ont été enregistrés en mode de détection par fluorescence. Les conditions d'enregistrement sont détaillées au tableau 3.1. Pour faciliter les opérations de normalisation, le domaine en énergie des spectres de DjR a été augmenté vis-à-vis de celui des spectres de Mamut. Par ailleurs, pour éviter de saturer le détecteur il a parfois été utilisé des filtres en aluminium. Les données brutes ont été corrigées du temps mort du détecteur, calibrées en énergie et normalisées. La calibration se fait en fixant le maximum de la dérivée première du spectre d'absorption XANES du cuivre métallique à 8980.48 eV (Kraft et al. 1996). La normalisation se fait en soustrayant le bruit de fond avant le seuil et en divisant le spectre obtenu par la valeur de l'intensité de l'absorption au milieu de la première oscillation EXAFS. L'ensemble des données calibrées et normalisées est représenté à la figure 3.9 Au total 99 spectres de μ-XANES ont été enregistrés sur les lames

minces du site C3 de Mamut, 79 sur celles du site B4, 59 sur le site A de DjR A et 114 sur le site B.

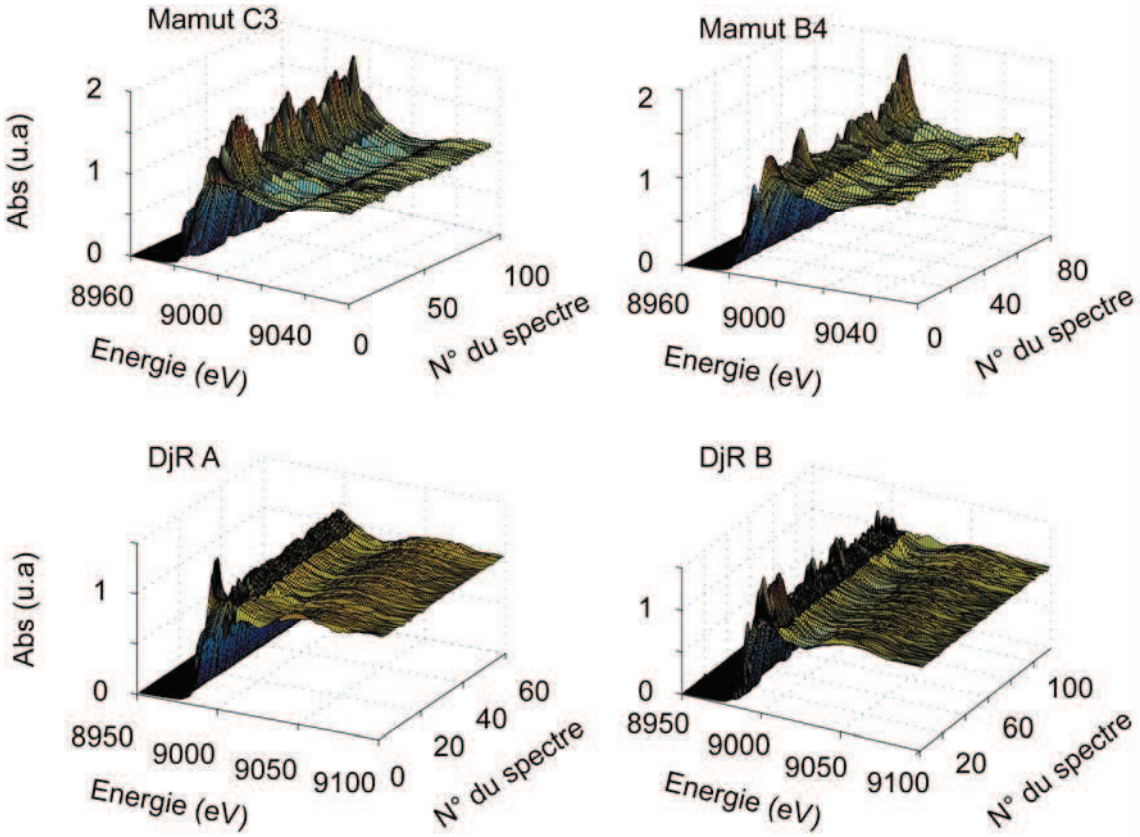


Figure 3.9 : Ensemble des spectres µ-XANES normalisés enregistrés autour du seuil K du cuivre étudié dans le cadre de ce travail.

Tableau 3.1 : Caractéristiques d'enregistrement des spectres µ-XANES des sites de Mamut et DjR. Pour les spots très riches en cuivre, le temps de comptage a été diminué (2 à 3 secondes dans les zones 8960-9010-9060 eV).

Mamut				
Energie (eV)	8920	8960	9010	9060
Nbre de points par région d'énergie	9	103	28	
Temps de comptage par points (s)	2	6	6	

DjR					
Energie (eV)	8880	8949	9011	9079	9179
Nbre de points par région d'énergie	15	125	35	21	
Temps de comptage par points (s)	2	5	5	3	

Dimensions du faisceau : V= 5µm H= 5 à 16 µm

3.3) Analyse en Composantes Principales

L'ACP est une technique d'algèbre linéaire qui consiste à orthogonaliser un espace vectoriel. La taille de l'espace orthogonalisé est égale au nombre de composantes vectorielles indépendantes de l'espace initial. Pratiquement, on construit une matrice expérimentale m*n à partir des données μ-XANES. Chaque colonne de cette matrice représente un spectre, et chaque ligne contient la valeur de l'absorbance (normalisée) à une énergie donnée. Cette matrice est alors décomposée dans une base orthonormée m*n de vecteurs propres e_i (1<i<n), et chaque spectre du jeu de données peut alors s'écrire comme une combinaison linéaire des vecteurs propres. L'importance d'un vecteur propre e_i dans l'espace initial est donnée par la valeur propre qui lui est associée, λ_i.

$$\underbrace{\begin{bmatrix} data_{11} & \cdots & data_{n1} \\ \vdots & \ddots & \vdots \\ data_{1m} & \cdots & data_{nm} \end{bmatrix}_{m\times n}}_{\text{data}} = \underbrace{\begin{bmatrix} e_{11} & \cdots & e_{n1} \\ \vdots & \ddots & \vdots \\ e_{1m} & \cdots & e_{nm} \end{bmatrix}}_{\text{vecteurs propres}} \times \underbrace{\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}^{1/2}}_{\text{valeurs propres}} \times \underbrace{\begin{bmatrix} \rho_{e_1,data_1} & \cdots & \rho_{e_1,data_n} \\ \vdots & \ddots & \vdots \\ \rho_{e_n,data_1} & \cdots & \rho_{e_n,data_n} \end{bmatrix}}_{\text{correlations}}$$

L'utilisation de l'ensemble des composantes (vecteurs propres) permet de reproduire exactement le jeu de données original. Les valeurs d'un spectre reconstruit avec *j* composantes sont calculées suivant la formule :

$$rec(data)_{jcp} = \sum_{i=1}^{j} \lambda_i e_i \rho_{e_i,data}$$

Dans un système expérimental, les composantes peuvent être séparées en deux groupes : les composantes primaires qui contiennent des informations structurales et les autres qui représentent le bruit et autres erreurs. Dans le cas où les proportions d'espèces sont variables dans le système, le paradigme est le suivant : le nombre d'espèces qui forment le système est égal au nombre de composantes (cp) primaires données par l'ACP (Beauchemin et al. 2002; Manceau et al. 2002). Dans le domaine de la spectroscopie d'absorption X, trois critères sont généralement utilisés pour déterminer le nombre de cp: la cascade des valeurs propres, l'indicateur IND, et un test statistique de Fisher sur les valeurs propres.

3.3.1) Cascade des valeurs propres

Le principe de ce critère est simple puisqu'il consiste à classer par ordre décroissant les valeurs propres. Dans le cas idéal d'un espace vectoriel de dimension n construit à partir

de la combinaison linéaire de r vecteurs (r<n), la valeur propre λ_{r+1} est strictement nulle. Dans la réalité λ est toujours supérieur à zéro du fait du bruit qui est différent pour chaque donnée expérimentale. Ces bruits sont vus comme des composantes indépendantes par l'ACP. Si le rapport signal/bruit est suffisant et la variabilité spectrale bonne, alors on peut néanmoins espérer observer une cassure dans la cascade des vp.

Une autre approche plus quantitative consiste à exprimer ces valeurs propres en termes de variance :

$$Variance = \frac{\lambda_j}{\sum_{j=1}^{n} \lambda_j}$$

Cette variance mesure la capacité d'un vecteur propre à décrire les données. On peut fixer (de manière arbitraire en général) une valeur limite qui départagera les composantes primaires et secondaires.

3.3.2) Indicateur IND

IND est une fonction empirique qui dépend du nombre *u* de composantes utilisées pour la reconstruction, des valeurs propres λi, de la taille n du jeu de données (nombre de colonnes) et du nombre de points m d'interpolation en énergie (nombre de lignes). Elle a été développée par Malinowski (1991) et s'écrit sous la forme :

$$IND(u) = \left(\frac{\sum_{j=u+1}^{t} \lambda_j}{m(n-u)^5} \right)^{1/2}$$

Cette fonction est censée atteindre un minimum lorsque le nombre correct de composantes primaires est employé pour reconstruire le jeu de données.

3.3.3) Test de Fisher statistique

Ce test est basé sur le calcul de valeurs propres réduites (REV) :

$$REV_j = \frac{\lambda_j}{(m-j+1)(n-j+1)}$$

Toujours selon Malinowski (Malinowski 1991), les REV non significatives, i.e associées aux composantes secondaires (au bruit), sont statistiquement égales. Il définit une valeur moyenne

des REV non significatives (REV^0_{pool}) à laquelle il associe une variance. En partant du principe que les REV secondaires sont distribuées normalement, il utilise un test de Fisher pour tester l'égalité des variances. L'hypothèse nulle $H_0 : REV_k = REV^0_{pool}$ teste si la variance associée aux *k* dernières composantes est égale à la variance des cp secondaires, tandis que l'hypothèse $H_1 : REV_n > REV^0_{pool}$ indique si, à un niveau de signification α donné, la composante *n-k*+1 a une contribution plus importante que l'erreur expérimentale. En général, α = 5%.

3.3.4) Approche itérative

Le nombre *j* de composantes est également souvent déterminé par comparaison visuelle entre les données et les reconstructions ACP avec *j* cp (Van Damme et al. 2010). Compte tenu de la taille de nos jeux de données, cette approche directe n'est pas possible. Pour contourner cette difficulté, plusieurs pistes ont été explorées. En particulier on a imaginé une approche par étapes : on effectue l'ACP sur un jeu de données restreint (typiquement une quinzaine de spectres présentant de grandes similitudes). On sélectionne visuellement le nombre de composantes. Puis on teste alors les spectres exclus du sous-ensemble de données un par un par Target Transformation (TT). Quand un spectre n'est pas reconstruit on le rajoute aux composantes déterminées et on crée ainsi un nouveau jeu de données qui peut à nouveau être réduit par ACP. La TT est une technique directement associée à l'ACP qui consiste à vérifier si une donnée est reconstruite par combinaison linéaire des composantes sélectionnées, c'est-à-dire primaires. Mathématiquement elle s'exprime suivant la formule :

$$\underbrace{\begin{bmatrix} sp_1 \\ \vdots \\ sp_m \end{bmatrix}_{m\times 1}}_{\text{SP}} = \underbrace{\begin{bmatrix} e_{11} & \cdots & e_{v1} \\ \vdots & \ddots & \vdots \\ e_{1m} & \cdots & e_{vm} \end{bmatrix}_{m\times v}}_{\text{v composantes primaires}} \times \begin{bmatrix} e_{11} & \cdots & e_{1m} \\ \vdots & \ddots & \vdots \\ e_{v1} & \cdots & e_{vm} \end{bmatrix}_{v\times m} \times \underbrace{\begin{bmatrix} data_1 \\ \vdots \\ data_m \end{bmatrix}_{m\times 1}}_{\text{data}}$$

Le principe est de comparer les vecteurs SP et data et s'il y a superposition, alors cela signifie que le vecteur 'data' appartient bien à l'espace propre réduit.

4) Résultats et discussion

4.1) Site de Mamut

4.1.1) Analyses chimiques

Le tableau 3.2 récapitule les résultats des analyses chimiques. Dans le système hydrographique Lohan-Bongkud, les points B5 en amont de la décharge et B1 à proximité mais de l'autre coté de la rivière Bongkud ne sont pas pollués au cuivre ([Cu] ~ 30-40 ppm). En revanche, les prélèvements B3 et B4 situés en aval présentent des taux élevés en concentration métallique supérieure d'environ un ordre de grandeur à celles des fonds géochimiques habituels (Kabata-Pendias et Pendias 1992). La contamination au point B3 est cependant moins importante car il est situé sur la rive de la rivière Lohan qui est opposée à la décharge. Le long de la rivière Mamut, C3 présente également des teneurs en Cu élevées ([Cu] ~ 600 ppm), contrairement à D3 ([Cu] ~ 100 ppm). Cette différence s'explique par l'apport d'eau « propre » au niveau du confluant Mamut\Langanan et par des eaux de ruissellement non contaminées. Il est probable que les analyses chimiques du site A3 révèleraient des concentrations intermédiaires entre C3 et D3. Enfin, en ce qui concerne les points de contrôle, D2 présente des valeurs presque normales ([Cu] ~ 70-80 ppm), alors que la contamination ne semble pas s'être propagée jusqu'au site A1. En première conclusion de ces analyses, on peut donc avancer que le bassin versant est contaminé sur 15-20 km en aval de l'activité minière, et que la contamination n'a pas atteint les zones urbaines le long du réseau hydrographique, notamment la ville de Ranau.

Par ailleurs, la concentration en cuivre diminue avec la profondeur de prélèvement. L'atténuation entre 0-10 cm et 10-30 cm est de l'ordre de 5 à 10 ppm aux points B1, B5, D3 et D2, et de plusieurs centaines de ppm aux points B3 et B4. Ce résultat s'explique par l'apport de surface du cuivre.

Enfin au point B4, où les concentrations en manganèse, fer, nickel, cuivre et zinc ont été mesurées, les coefficients de corrélation $R_{[Cu]/[Zn]} = 0,99$, $R_{[Cu]/[Mn]} = 0,84$, $R_{[Cu]/[Ni]} = -0,85$ et $R_{[Cu]/[\%Fe]} = 0,61$ semblent montrer que les phases porteuses de zinc et manganèse contiennent du cuivre tandis que les phases porteuses de nickel sont différentes des phases porteuses du cuivre. Cependant cette interprétation est à considérer avec précaution car le nombre de points d'analyse est seulement de 7.

Tableau 3.2 : Résultats des analyses chimiques

Nom de code et profondeur de prélèvement	Cu (ppm)	Zn (ppm)	Fe (%)	Mn (ppm)	Ni (ppm)
A1 1 0-30 cm	48	49	Non mes	Non mes	Non mes
A1 2 0-30 cm	47	17	Non mes	Non mes	Non mes
A1 3 0-30 cm	45	26	Non mes	Non mes	Non mes
B1 1 0-10cm	80	Non mes	Non mes	Non mes	Non mes
B1 2 10-30cm	20	Non mes	Non mes	Non mes	Non mes
B1 3 0-10cm	30	Non mes	Non mes	Non mes	Non mes
B1 4 10-30cm	20	Non mes	Non mes	Non mes	Non mes
B1 5 0-10cm	30	Non mes	Non mes	Non mes	Non mes
B1 6 10-30cm	25	Non mes	Non mes	Non mes	Non mes
B3 1 0-10cm	590	Non mes	Non mes	Non mes	Non mes
B3 2 10-30cm	80	Non mes	Non mes	Non mes	Non mes
B3 3 0-10cm	150	Non mes	Non mes	Non mes	Non mes
B3 4 10-30cm	70	Non mes	Non mes	Non mes	Non mes
B3 5 0-10cm	580	Non mes	Non mes	Non mes	Non mes
B3 6 10-30cm	215	Non mes	Non mes	Non mes	Non mes
B4 1 0-5cm	840	170	4.7	685	400
B4 1 5-10cm	35	80	3.0	395	735
B4 2 10-25cm	40	90	5.3	585	1135
B4 3 0-10cm	1445	240	7.5	930	245
B4 4 10-30cm	185	105	4.2	595	765
B4 5 0-10cm	920	200	6.1	865	540
B4 6 10-30cm	270	120	7.1	760	1080
B5 1 0-10cm	30	Non mes	Non mes	Non mes	Non mes
B5 2 10-30cm	25	Non mes	Non mes	Non mes	Non mes
B5 3 0-10cm	40	Non mes	Non mes	Non mes	Non mes
B5 4 10-30cm	30	Non mes	Non mes	Non mes	Non mes
B5 5 0-10cm	25	Non mes	Non mes	Non mes	Non mes
B5 6 10-30cm	25	Non mes	Non mes	Non mes	Non mes
C3 1 0-30 cm	775	90	3.1	385	Non mes
C3 2 0-18 cm	385	45	2.2	240	Non mes
C3 3 0-15 cm	625	55	3.1	330	Non mes
D2 1 0-10 cm	80	13	Non mes	Non mes	Non mes
D2 2 10-30 cm	55	15	Non mes	Non mes	Non mes
D2 3 0-10 cm	69	17	Non mes	Non mes	Non mes
D2 4 10-15 cm	63	17	Non mes	Non mes	Non mes
D2 5 0-10 cm	84	15	Non mes	Non mes	Non mes
D2 6 10-30 cm	50	14	Non mes	Non mes	Non mes
D3 1 0-10 cm	102	91	Non mes	Non mes	Non mes
D3 2 10-30 cm	91	76	Non mes	Non mes	Non mes
D3 3 0-10 cm	76	397	Non mes	Non mes	Non mes
D3 4 10-25 cm	110	271	Non mes	Non mes	Non mes
D3 5 0-10 cm	73	102	Non mes	Non mes	Non mes
D3 6 10-30 cm	67	81	Non mes	Non mes	Non mes

4.1.2) Cartographie μ-XRF

La carte XRF de la figure 3.10 est une cartographie chimique parmi un total de 70 enregistrées, et contient l'essentiel des associations chimiques observées. Le cuivre est présent à la fois dans la racine de riz coupée transversalement et dans la matrice pédologique, cependant sa distribution n'est pas homogène. On peut dénombrer au moins cinq associations géochimiques, qui pourraient représenter cinq espèces chimiques différentes. Le cuivre est associé au calcium dans la racine (1) et dans les débris organiques de la matrice. A priori il est coordonné à des atomes d'oxygène ou de soufre, voire d'azote portés par des groupements amines. Les taches vertes intenses (2) correspondent à des grains riches en cuivre, comme du cuivre natif ou un autre cristal cuprifère. Dans la matrice fine argileuse, le cuivre est associé au fer, soit en quantités bien inférieures (rose 3), soit dans des proportions du même ordre de grandeur que celles du fer (vert 4). La première association pourrait correspondre à une argile ferrifère et cuprifère, et la seconde à un mélange d'oxyhydroxyde de fer (ex., goethite) cuprifère et de cuivre associé à la MO. Le cuivre est associé au manganèse dans zones bleutées. Il a été montré par μ-XRD que l'oxyde de Mn était un phyllomanganate turbostratique (i.e., vernadite).

4.1.3) Spectroscopies μ-XANES et μ-XRF

Comme cela a été dit précédemment, un spectre est soit la signature spectrale d'une espèce pure, soit la somme pondérée de plusieurs signatures d'espèces distinctes. Il a été remarqué que certains spectres enregistrés sur des points différents des lames minces de B4 et C3 étaient semblables. Comme il est peu probable que des points d'analyse distants contiennent les mêmes espèces dans les mêmes proportions, ces spectres sont probablement ceux d'espèces pures. Deux des trois signatures caractéristiques observées ont été identifiées par comparaison avec les spectres XANES de composés connus :

- Celle de grains de Cu^0 sous forme de spots très riches en Cu dont la taille peut atteindre la centaine de μm². Leurs spectres présentent un pré pic à 8983 eV, un sommet dédoublé avec des maxima à 8995 et 9004 eV, suivi d'une oscillation à 9027 eV (Fig. 3.11). Certains des spots contiennent du zinc dans des proportions semblables à celle du laiton ([Zn]/[Cu] = 0,5). Les spectres XANES du cuivre métallique et du laiton étant identiques, seule l'analyse de l'EXAFS permet de faire la différence (Matynia 2009).

- Celle de la chalcopyrite sous forme de spots riches en Cu et Fe et dans lesquels le calcium est absent. Leur taille peut également atteindre la centaine de µm². Sa signature spectrale présente un prépic à 8981 eV, un pic intense à 8988 eV et un sommet de nouveau dédoublé mais dont les maxima à 9002 et 9012 eV sont arrondis. (Fig. 3.12).

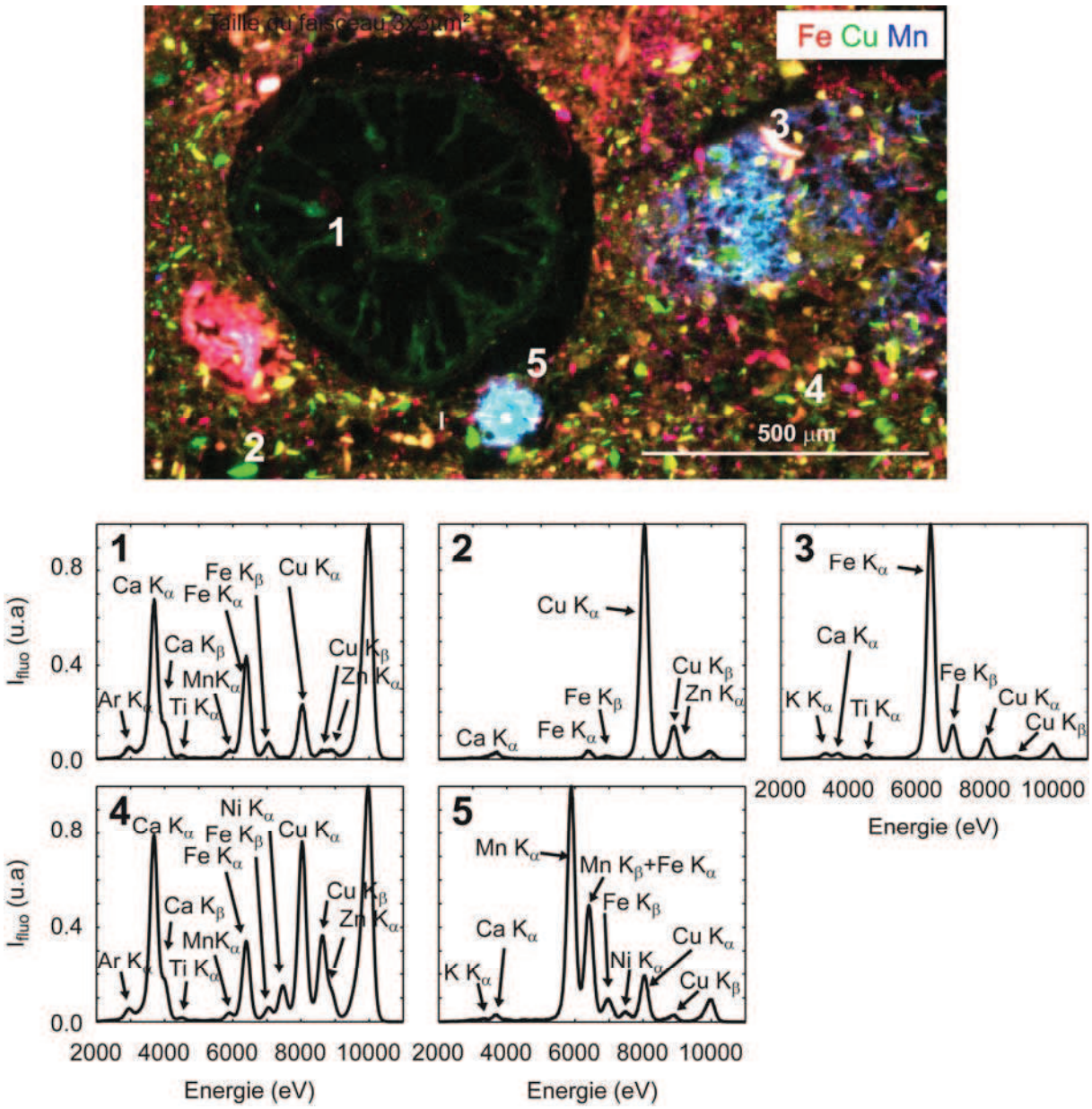


Figure 3.10 : Cartographie µ-XRF d'une lame mince de l'échantillon C3. Elle image la distribution élémentaire dans une racine de riz et dans la rhizosphère environnante. Le cuivre est présent dans la racine et dans le sol mais sa distribution n'est pas uniforme. Le fer est en rouge, le cuivre en vert et le manganèse en bleu.

- La troisième espèce est également très riche en Cu et sans calcium, mais de rapport Cu/Fe variable. Bien que non définie, elle affiche les caractéristiques d'un cuivre en coordinance octaédrique (Fig. 3.13). Le petit prépic à 8979 eV distinctif de la signature spectrale de la

transition électronique de l'orbitale électronique 1s vers l'orbitale 3d est difficilement visible mais l'intense raie blanche asymétrique dont le sommet est à 8998 eV démontre que le cuivre est sous forme divalente (Manceau et Matynia, 2010). Dans la zone en énergie de cette raie (8985 à 9000 eV), la différence entre les transitions des orbitales 1s vers $4p_{xy}$ et $4p_z$ est de 5 eV ce qui permet d'estimer la distance Cu-O_{ax} à 2,30 Å suite à la levée de dégénérescence résultant de l'effet Jan-Teller (Boudesocque et al. 2007; Manceau et Matynia, 2010).

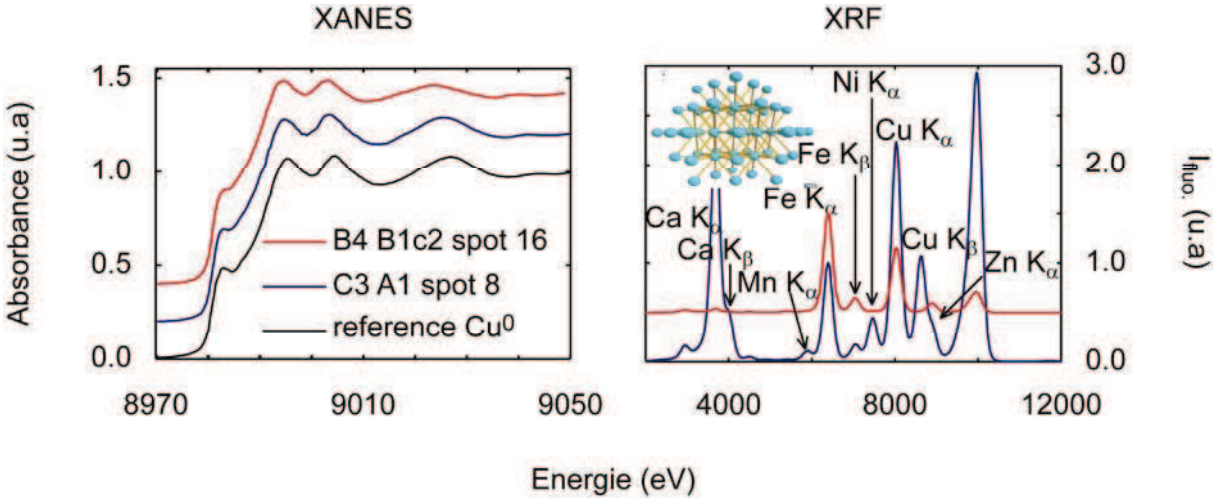


Figure 3.11 : Identification du cuivre sous forme Cu^0 par spectroscopie µ-XANES et µ-XRF. Les spectres de micro-fluorescence X sont normalisés par $I_{fluo}(K_\alpha[Fe])$. Dans un cas Cu° est associé à la MO (spot riche en Ca) et pas dans l'autre.

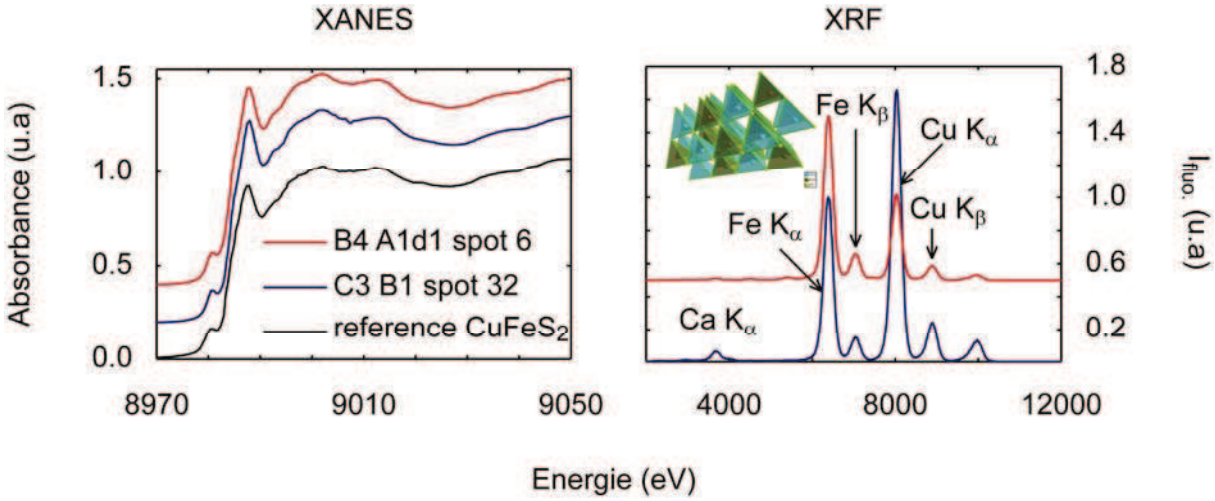


Figure 3.12 : Identification de la chalcopyrite ($CuFeS_2$)

Ces formes pures ont été observées sur des spots riches en cuivre. D'autres points d'analyse ont des formes continument variables qui indiquent qu'ils proviennent de mélanges d'espèces. Ceci est illustré à la figure 3.14 avec l'exemple de la troisième espèce pure, dont la nature est inconnue. L'intensité de la raie blanche (maximum à 8998 eV) est anti-corrélée à la

hauteur du prépic à 8984 eV, montrant ainsi que ces spectres sont la signature d'un mélange d'au moins deux formes différentes du cuivre, l'une divalente et la seconde monovalente (prépic à 8983 eV caractéristique de la transition électronique 1s→4p de Cu^{+}) (Liu et Robota 1993 ; Manceau et Matynia, 2010).

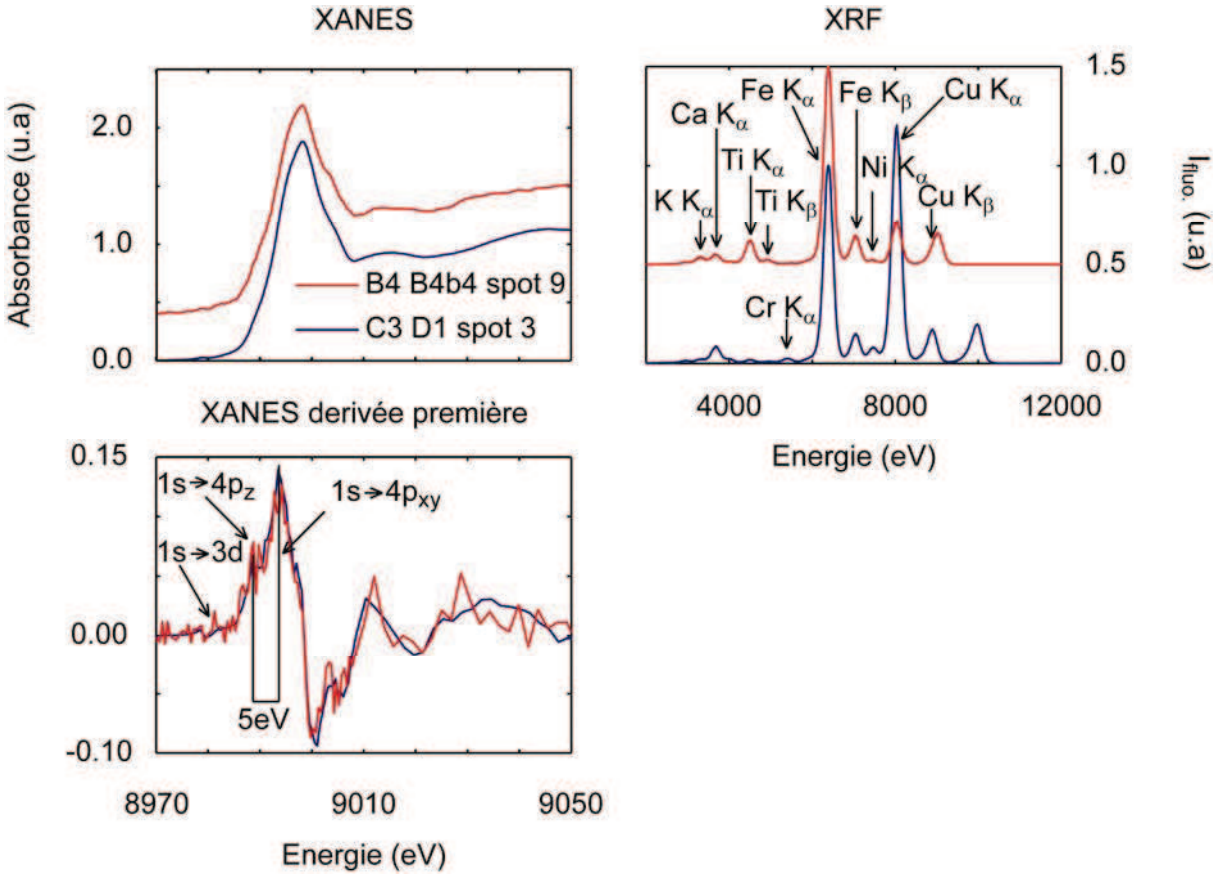


Figure 3.13 : Mise en évidence d'une espèce pure cuprifère dans laquelle le cuivre est divalent et en coordinance octaédrique avec des distances Cu-O_{ax} = 2,30 Å.

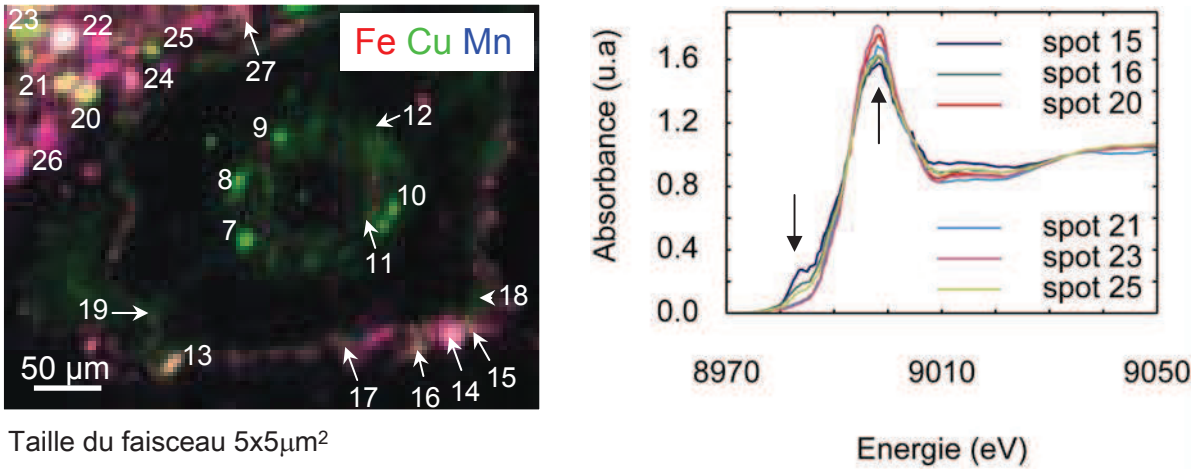


Figure 3.14 : Variabilité des spectres de mélange enregistrés sur une même zone.

Par ailleurs, il également été observé que les formes du cuivre dans le fond matriciel ou associées à la matière organique évoluaient au cours du temps sur un même point d'analyse, du fait de la photoréduction sous faisceau X du cuivre divalent en cuivre monovalent (Fig. 3.15). La forme non identifiée ne montrant pas d'évolution sous faisceau, cette remarque montre que les spectres de la figure 3.14 sont un mélange binaire entre la forme non identifiée et un complexe Cu-MO photoréduit.

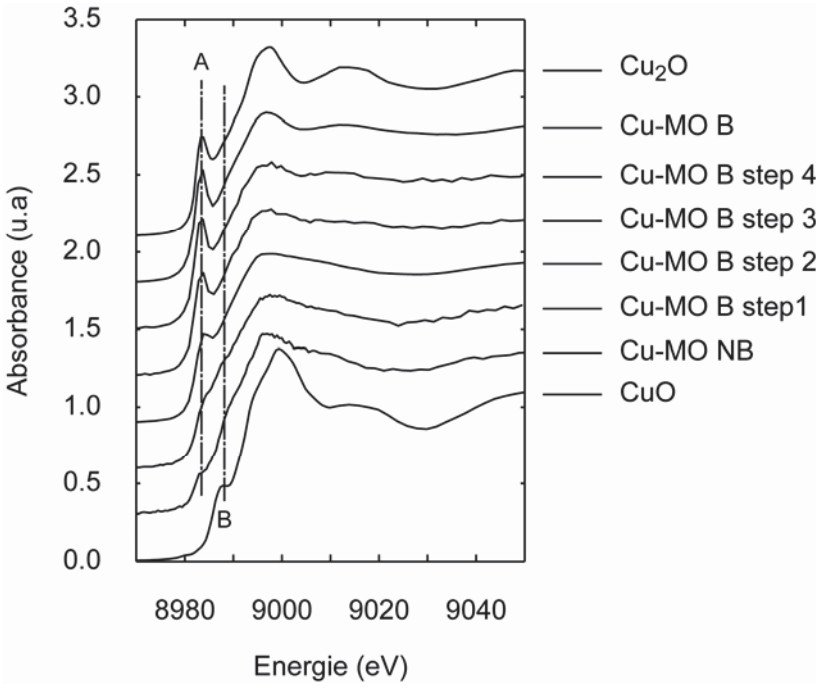


Figure 3.15 : Photoréduction du cuivre associé à la matière organique sous l'effet du rayonnement X. Les spectres Cu-MO vont de l'espèce 'non brulée' (NB) à l'espèce 'brulée' (B) en passant par divers stade d'évolution (B step 1 à 4). Pour pouvoir comparer les transitions électroniques du XANES, on trace en bas le spectre de CuO (Cu divalent) et en haut, le spectre Cu_2O (Cu monovalent)**.**

4.1.4) Analyse en composantes Principales

4.1.4.1) Cascade des valeurs propres

Les cascades des valeurs propres ont été déterminées pour chacun des deux jeux de données, c'est-à-dire les 99 spectres µ-XANES du site C3 et les 79 du site B4, (Fig. 3.16). Dans les deux cas, il apparait une décroissance 'continue' des valeurs propres sans qu'aucune discontinuité ne soit observable. La première valeur propre a une importance considérable (environ 99%) et avec trois composantes, plus de 99,9% de la variance est expliquée.

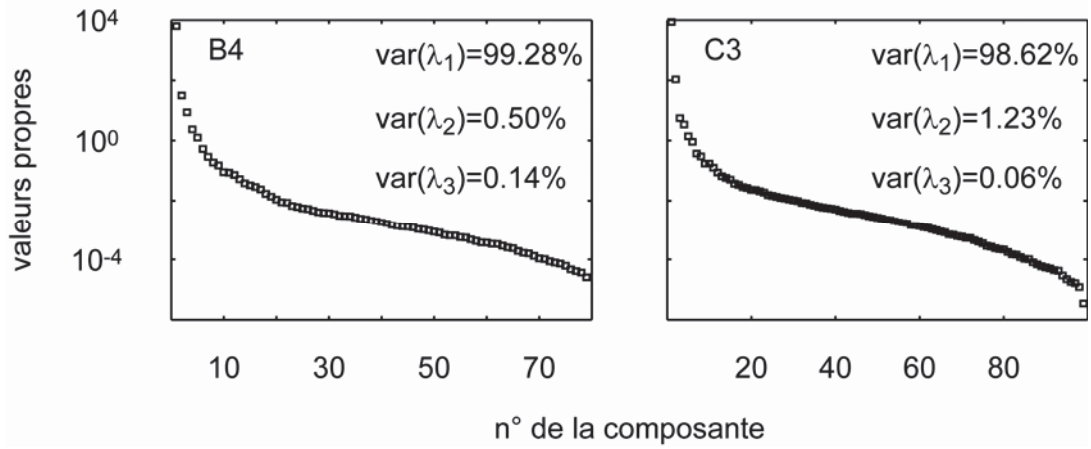


Figure 3.16 : Cascades des valeurs propres de la série de spectres μ-XANES du site de Mamut (Malaisie).

4.1.4.2) IND

Les résultats sont présentés à la figure 3.17. Cette fonction présente une forme en U dont le minimum est atteint avec 22 composantes primaires pour B4 contre 31 pour C3.

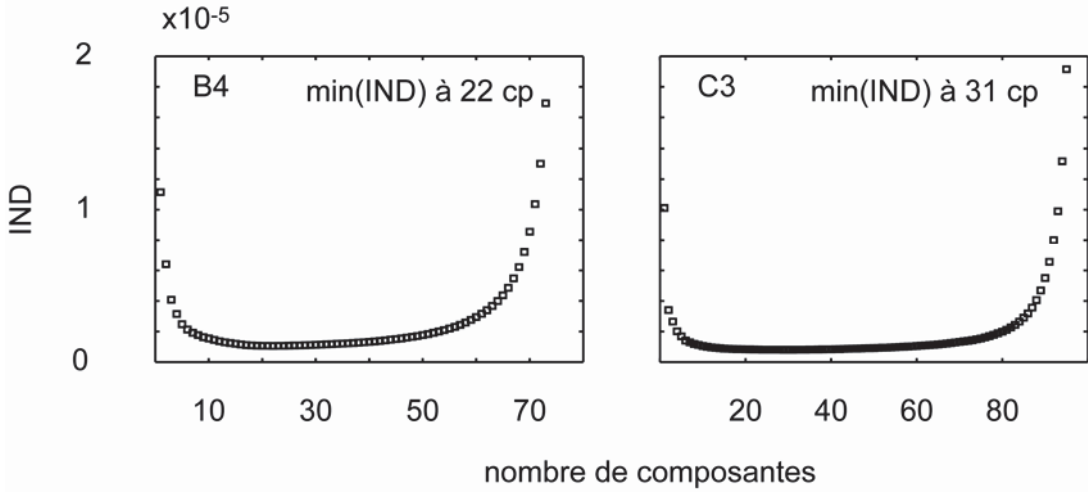


Figure 3.17 : Valeurs de la fonction IND en fonction du nombre de composantes utilisées dans la reconstruction des données.

4.1.4.3) F-test

Les résultats sont présentés à la figure 3.18 pour un niveau de signification α de 5%. Les courbes ont une forme sigmoïdale croissante. Cependant, certaines reconstructions à *j* cp ont un niveau de signification supérieur à celles à *j*+1 cp. Ce comportement atypique a été observé dans d'autres études (Beauchemin et al. 2002). Ce paramètre donne 13 composantes primaires pour C3, contre 18 pour B4.

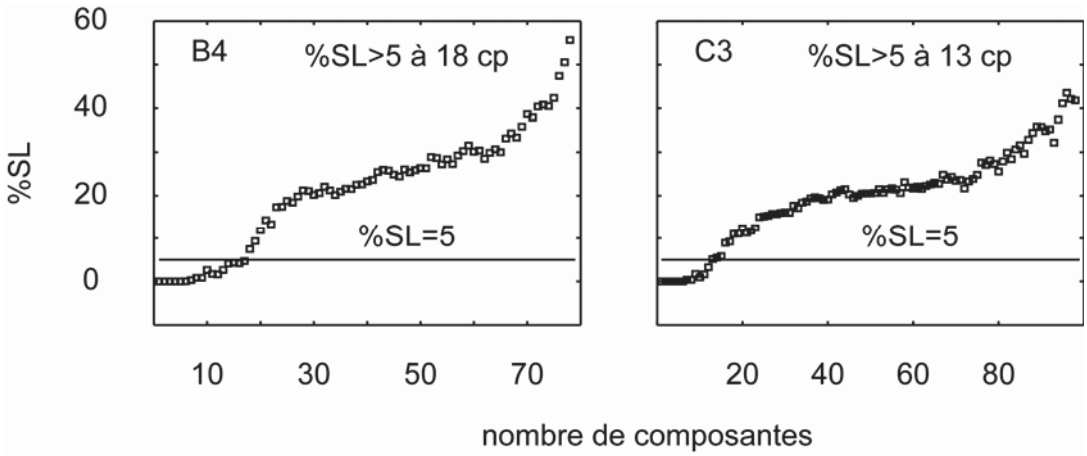


Figure 3.18 : Résultats du test de Fisher sur les valeurs propres réduites en fonction du nombre de composantes utilisées dans la reconstruction des données.

4.1.4.4) Approche itérative

L'approche itérative telle que décrite plus haut et appliquée au site B4 a permis de conclure que 8 composantes suffisaient à reconstruire les données du site B4 et du site C3. Si on travaille dans l'espace propre de l'ACP, les composantes n'ont pas véritablement de sens physique (Fig. 3.19).

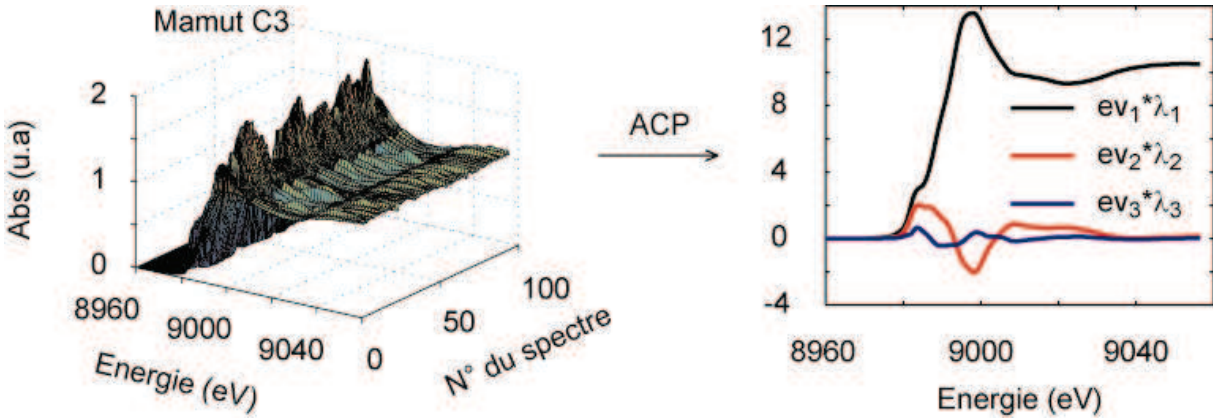


Figure 3.19 : Les trois premières composantes principales obtenues par ACP sur les spectres µ-XANES du site C3 pondérées par leur valeur propre. Si la première composante ressemble effectivement à un spectre XANES, les composantes suivantes n'ont pas de signification physique.

En utilisant la rotation varimax (Kaiser 1958), on peut néanmoins tenter de donner un sens à ces composantes. C'est une rotation orthogonale dans laquelle le nombre de variables indépendantes corrélées avec un axe factoriel (composante principale) est maximisé. Avec cette rotation, il est possible d'obtenir huit composantes orthogonales qui ressemblent à des spectres expérimentaux contenus dans le jeu de données (Fig. 3.20).

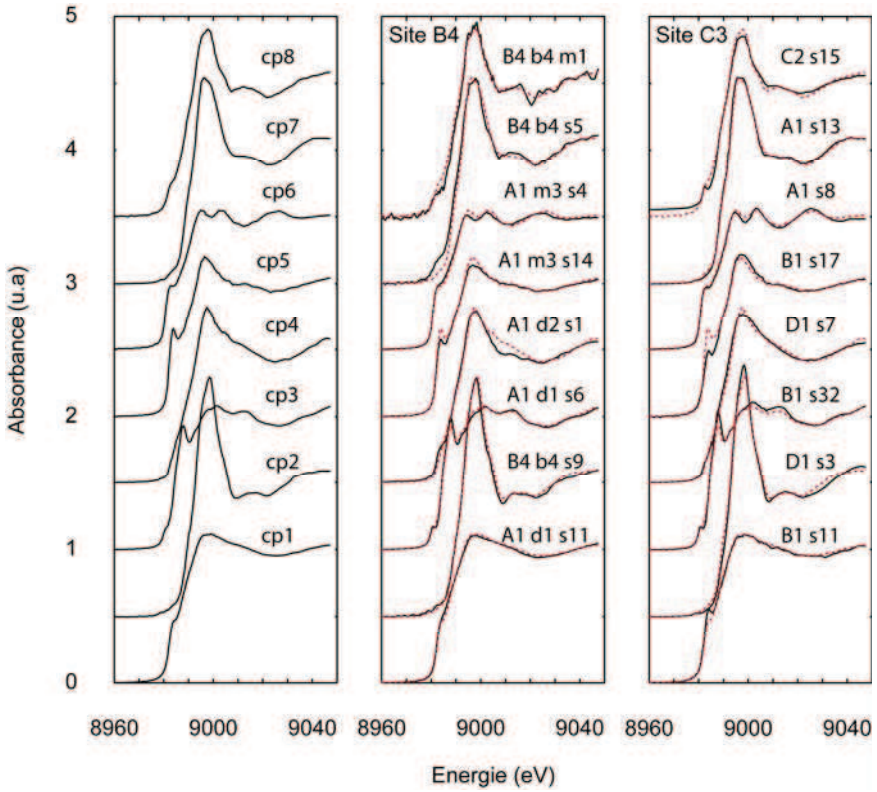


Figure 3.20 : Huit premières composantes déterminées de manière itérative. Non seulement elles sont proches de spectres expérimentaux individuels, mais de plus elles permettent de reconstruire de façon satisfaisante l'ensemble des 99+79=178 spectres.

Malheureusement, ce chiffre de 8 majore le nombre d'espèces qui n'est pas connu précisément, et les huit spectres repérés ne sont que des candidats potentiels à la signature spectrale d'espèces pures. Ainsi, les composantes cp2, cp7 et cp8 de la figure 3.20 sont proches, et il est peu probable qu'elles correspondent chacune à des formes du cuivre réellement différentiables par spectroscopie XANES. Ce propos est illustré également à la figure 3.21 qui montre les reconstructions de deux spectres (C3A1 s9 et C3A1 s11) avec, d'une part, les huit composantes déterminées de manière itérative et, d'une part, les huit et onze premières composantes déterminées par ACP sur l'ensemble des 26 spectres µ-XANES de la carte µ-XRF C3A1. Dans le premier cas, les spectres sont reconstruits de manière satisfaisante puisque leur forme est bien reproduite à l'exception du bruit. Dans le second cas, les 8 premières cp ne sont pas suffisantes pour reconstruire l'asymétrie du maximum la raie blanche du spectre C3A1 s9. Onze cp sont alors nécessaires pour reproduire ce maximum, mais alors le bruit du spectre C3A1 s11 est également reproduit, ce qui n'est pas désiré.

Si dans notre approche itérative nous avions considéré pour commencer l'ensemble des spectres de la lame C3A1, nous aurions probablement déduit un nombre de cp supérieur à 8, ce qui montre que l'approche itérative est largement dépendante à la fois du bruit présent sur

les données et du « chemin emprunté » pour réduire la dimension. Dans notre cas nous avons eu de la chance car les familles de spectres obtenus par lame mince du site B4 présentaient des caractères binaires (Fig. 3.14) et les pollutions en C3 et B4 étant similaires, il a été possible de retrouver les spectres du site C3.

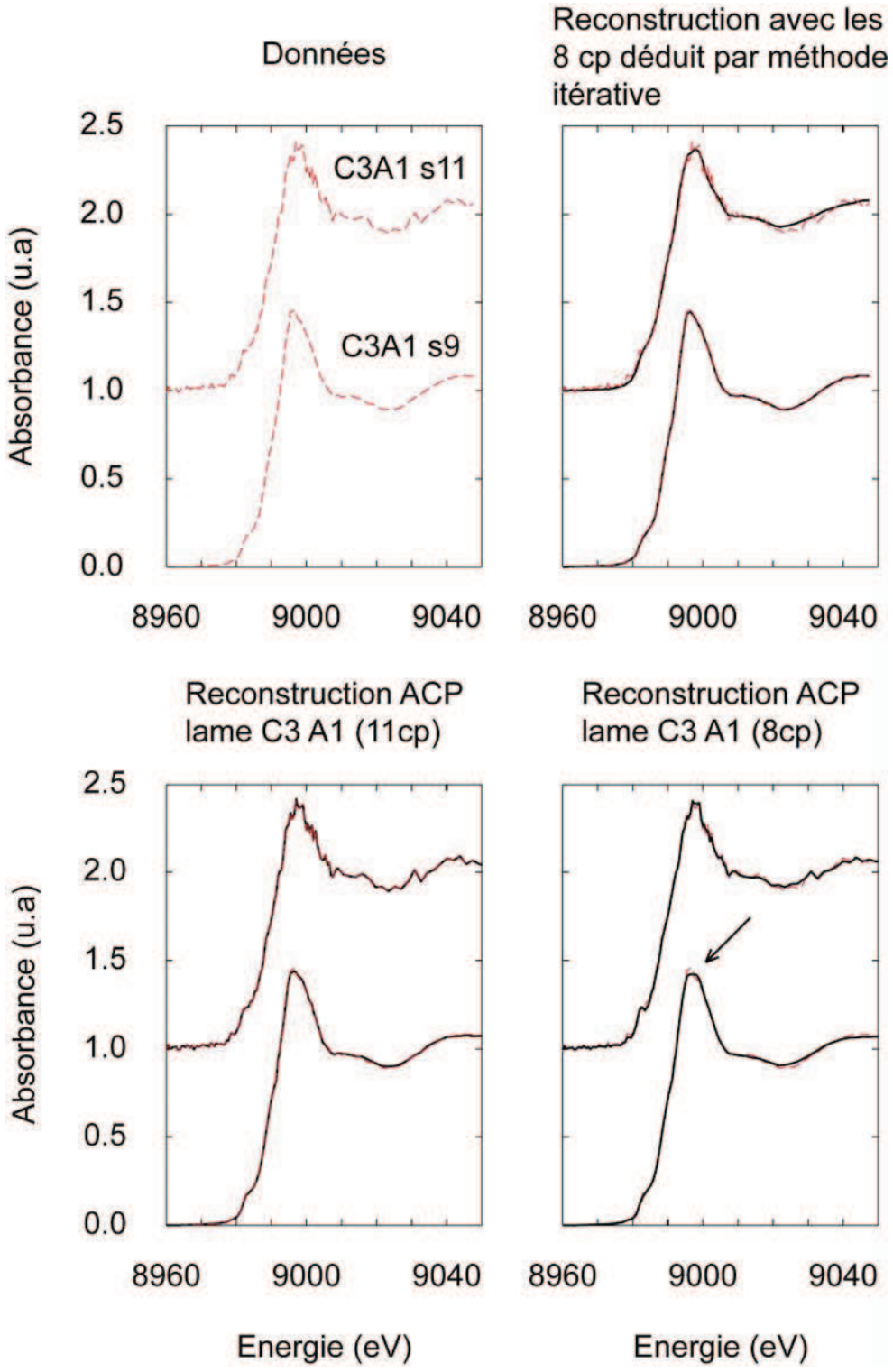


Figure 3.21 : Reconstruction des spectres C3A1 s11 et s9 à partir de 8 composantes déterminées par l'approche itérative, et à partir des 8 et 11 premières composantes obtenues par ACP sur l'ensemble des spectres de la lame C3 A1.

4.2) Site de Django Reinhardt

4.2.1) Analyses chimiques

Les analyses chimiques des prélèvements n'ont pas encore été effectuées au moment de l'écriture de ce manuscrit, mais d'autres analyses effectuées au moment de notre prélèvement mais à l'entrée du bassin (environ une dizaine de mètre de la zone A) donnent des concentrations en chrome de 181 mg/kg, en nickel de 116 mg/kg, en cuivre de 232 mg/kg, en zinc de 2354 mg/kg et en plomb de 165 mg/kg. Par ailleurs, au cours d'un stage de recherche sur l'étude des interactions plantes-Mn, Cu et Zn, il a été clairement montré que la concentration des métaux dans les plantes décroissait avec l'éloignement par rapport à l'entrée du bassin. Cet effet est à corréler avec la topologie du site qui montre que l'entrée du bassin est plus souvent inondée par les eaux usées (Fig. 3.22).

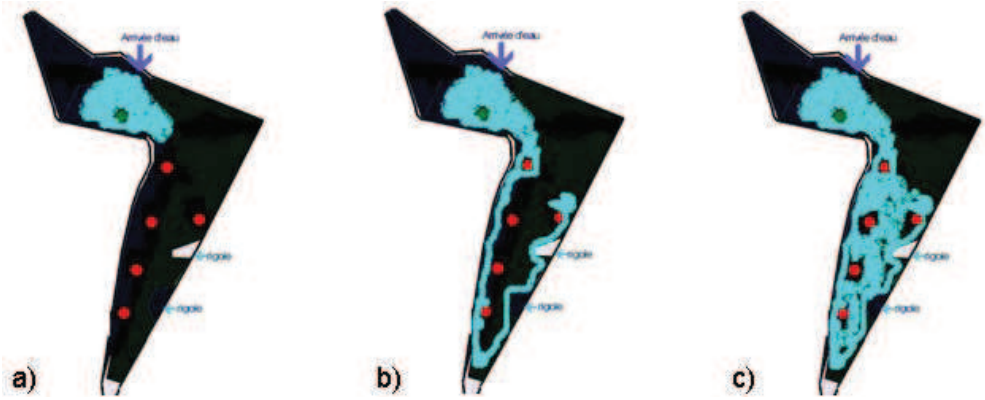


Figure 3.22 : Aire de répartition des eaux de surface en fonction de l'intensité des pluies a) faible pluie, b) pluie modérée et c) pluie soutenue. La zone en entrée du bassin est la première à être inondée, expliquant ainsi le gradient des concentrations en polluants métalliques observé sur les prélèvements des végétaux.

4.2.2) Cartographie μ-XRF

La figure 3.23 montre des cartographies μ-XRF d'un échantillon de sol de la zone A et un autre de la zone B. Dans la zone A, le cuivre est associé avec des débris organiques et avec le fer et le zinc dans la matrice fine. Comme à Mamut, Cu est également associé au manganèse. Dans la zone B, la matrice argileuse est moins riche en Cu, celui-ci étant alors présent dans des spots concentrés. L'analyse détaillée de la carte ne met pas en évidence d'association Cu-Mn. Ces observations coïncident avec le fonctionnement du bassin. La zone A est très végétalisée en surface. C'est une zone humide souvent marécageuse (au moins en période hivernale). Ceci explique l'abondance des racines et débris végétaux. La zone B est un sol plus sec ou la couverture végétale est plutôt éparse. Les grains y sont transportes gravitairement et déposés, puis s'altèrent chimiquement très lentement car il y a peu d'eau. Ils

sont par contre altérés dans la zone humide, libérant du Cu^{2+} qui est refixe par la matrice organo-minérale. Le cas A se rapproche de Mamut.

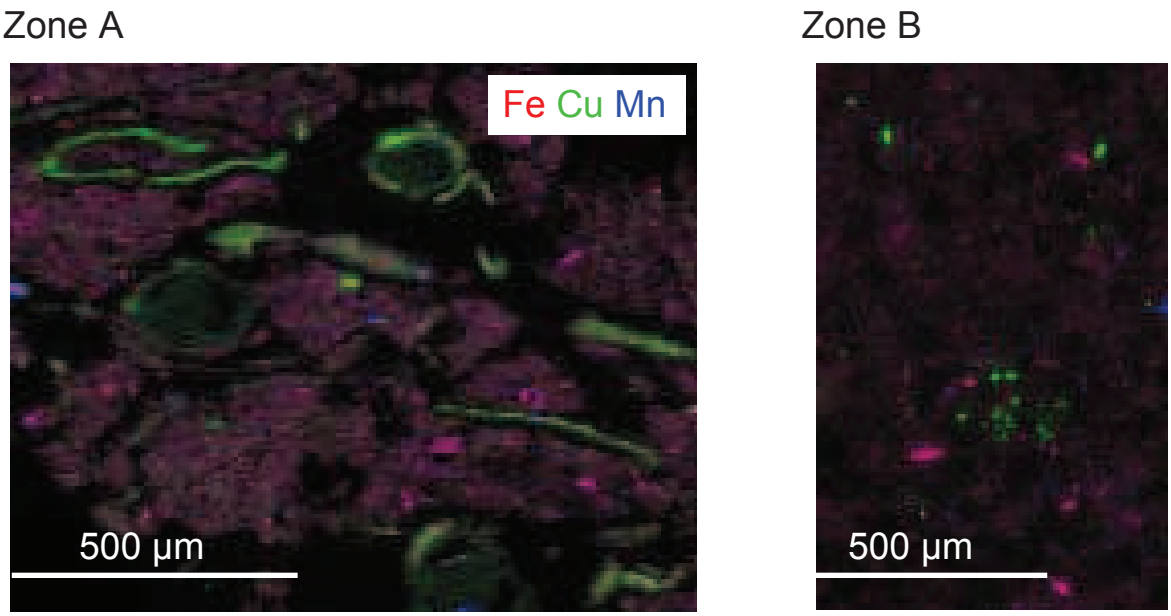


Figure 3.23 : Cartographie de micro-fluorescence X d'une région de lame mince de la zone A (gauche) et de la zone B (droite).

4.2.3) Spectroscopies µ-XANES et µ-XRF

Plusieurs formes présentes à différents endroits dans les échantillons A et B ont été repérées. Pour commencer, Cu^0 et la chalcopyrite ($CuFeS_2$) se retrouvent en A et B (Fig. 3.24). Les mesures de XRF de ces spots présentent toutes une raie K_α(Cu) très intense. Dans le cas de Cu^0, l'un des deux spectres (B2HGs7) montre la présence de zinc en quantité comparable à celle du cuivre, alors que le zinc est absent du second spectre (A32HDs3). Il est possible que B2HGs7 soit du laiton. En revanche, A32HDs3 est du cuivre métallique pur. En ce qui concerne la chalcopyrite, du zinc est présent sur l'un des deux spectres XRF (A32HGs11) et absent dans l'autre (B2BDr2s11).

Une première espèce non identifiée très riche en cuivre est également présente dans les deux zones (Fig. 3.24). L'intensité de la raie K_α(Cu) est similaire à celle du fer dans l'une (A32HDs4) et environ deux fois plus importante dans l'autre (B2BDr2s13). Comme dans le cas des deux spectres de chalcopyrite, l'un des spectres montre la présence de Zn (A32HDs4). Les deux spectres XANES ressemblent à celui de la chalcopyrite en moins résolu. Par analogie avec le cuivre métallique pour lequel il a été montré que l'amplitude des structures

XANES et oscillations EXAFS diminuait avec la cristallinité, l'espèce non identifiée serait un sulfure de Cu et de Fe moins bien cristallisé que la chalcopyrite.

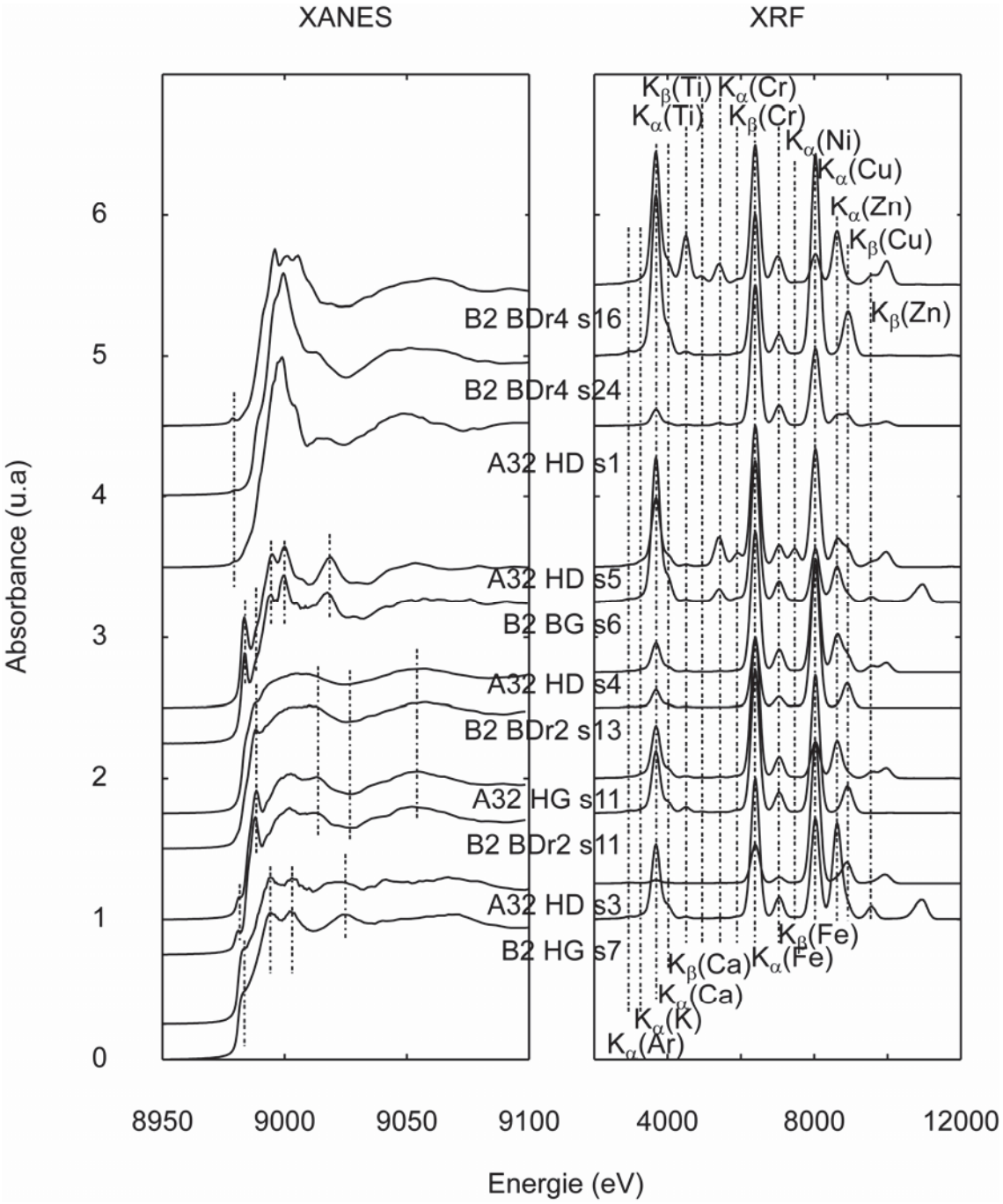


Figure 3.24 : Identifications d'espèces pures présentes dans les zones A et B du site de DjR à partir de leur spectre XANES et XRF. Sur 7 espèces, quatre sont présentes dans les deux sites. Du bas vers le haut, on identifie le cuivre métal/laiton et la chalcopyrite. La troisième espèce est probablement une forme de chalcopyrite mal cristallisée. La quatrième espèce devrait être un alliage cuivre(I)-chrome. Les 5ème et 6ème espèces seraient des oxydes de cuivre(II), et dans la septième le cuivre est divalent et probablement associé à des éléments comme le titane, le chrome ou le zinc.

Une seconde espèce non identifiée a également été observée dans les zones A et B (Fig. 3.24). Son spectre XANES a la forme de celui du cuivre métal. Par ailleurs, la première transition vers 8983 eV est très creusée comme dans Cu_2O (Manceau et Matynia, 2010). Du calcium, du chrome, du fer, du cuivre et du zinc sont détectés par XRF. Sur le spot B2BGs6, il a été observé que les intensités K_α(Cr) et K_α(Cu) étaient corrélées, alors que le spot A32HDs5 possède en plus du nickel. Dans la zone industrielle proche du bassin se trouvent des usines fabriquant des alliages de cuivre. Il se peut que cette espèce soit un alliage de cuivre-chrome utilisé notamment pour coupler les propriétés de conductivité du cuivre avec une résistance aux déformations.

Trois autres espèces n'ont pas été identifiées (Fig. 3.24). La première a été trouvée dans la zone A (A32HDs1), et les deux autres dans la zone B (B2BDr4 s24 et s16). Leur forme générale est celle du cuivre divalent, qui de plus serait en coordinence six aux points A32HDS1 et B2BDr4s24. B2BDr4s16 a une forme moins commune. L'analyse XRF révèle une association Cu-Ca. Le zinc est absent au point B2BDr4s24, alors que B2BDr4s16 possède des quantités relativement élevées de titane, de chrome et de zinc. Malheureusement une seule occurrence du spectre B2BDr4s16 a été trouvée, ce qui empêche de conclure sur la nature probable des associations chimiques.

Une telle abondance de formes pures n'a pas été remarquée lors de l'étude sur le porphyre cuprifère de Mamut. En Malaisie, le cuivre est principalement transporté en solution sous forme dissoute et les espèces sont majoritairement secondaires. Dans le bassin, les grains « primaires » sont apportés de manière régulière par les activités industrielles. Ils se transforment sur le site mais il existe néanmoins des formes secondaires (dans la matrice fine et organique) qui ne sont pas des formes héritées.

4.2.4) Analyses en Composantes Principales

4.2.4.1) Cascade des valeurs propres

Les valeurs propres des jeux de données (59 spectres μ-XANES pour DjR A et 115 pour DjR B) sont présentées sous forme de cascade à la figure 3.25. Comme précédemment, les vp décroissent continument sans discontinuité clairement observée. La première vp a une importance très élevée (environ 99% de la variance) et les deux premières vp expliquent à elles seules plus de 99,99% de la variance du système.

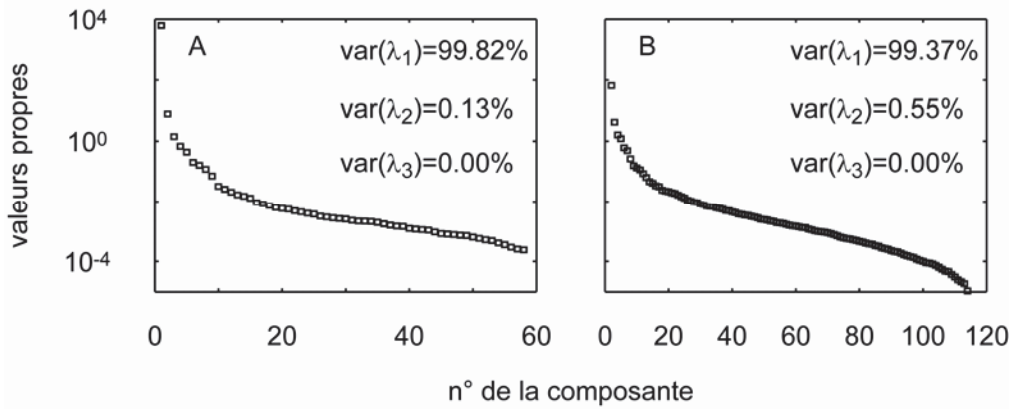


Figure 3.25 : Cascades des valeurs propres des séries de spectres µ-XANES des deux zones d'études du site de DjR.

4.2.4.2) IND

Les fonctions présentent une forme en U caractéristique, avec un minimum à cp = 15 pour DjR A et cp = 31 pour DjR B (Fig. 3.26).

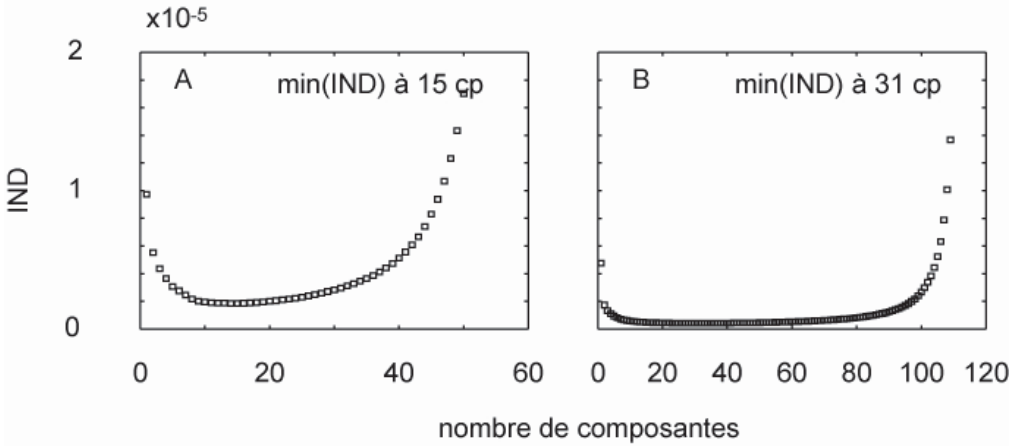


Figure 3.26 : Valeurs de la fonction IND en fonction de nombre de composantes utilisées dans la reconstruction.

4.2.4.3) F-test

Comme pour les données de Mamut, les courbes de pourcentage de signification en fonction du nombre de cp (Fig. 3.27) ont une forme sigmoïdale croissante et d'autre part le caractère atypique du test (Beauchemin et al. 2002). Cet indicateur donne 9 composantes pour DjR A et 16 pour DjR B.

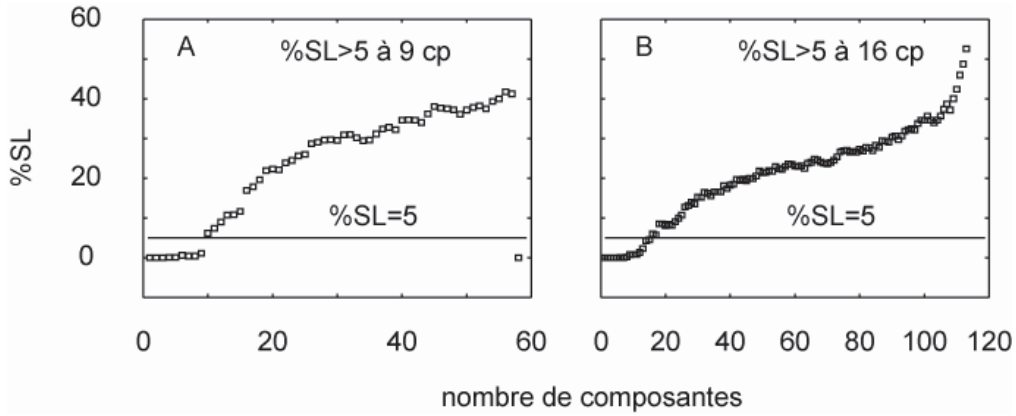


Figure 3.27 : Résultats du test de Fisher sur les valeurs propres réduites en fonction de nombre de composantes utilisées dans la reconstruction.

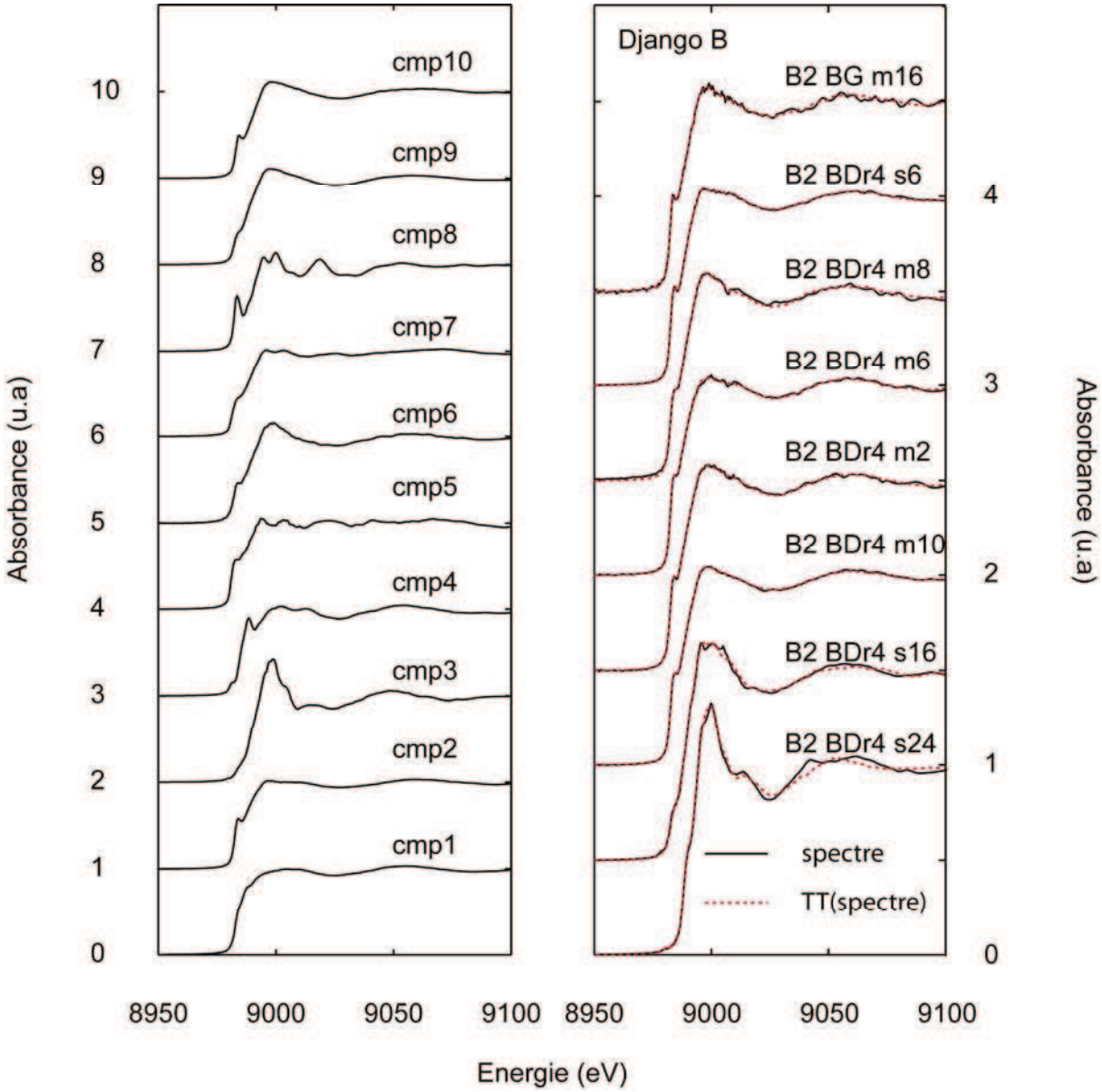


Figure 3.28 : Ensemble des dix premières composantes principales (cmp 1 à 10) calculées sur le jeu de données de spectres XANES du site de prélèvement DjR A (à gauche) et résultats de la TT avec ces composantes sur des spectres du site DjR B (à droite).

4.2.4.4) Approche itérative

A partir d'une approche itérative sur les spectres de la zone A, le nombre de composantes principales (notées cmp pour les différencier des cp du site de Mamut) significatives a été évalué à 10 (Fig. 3.28). Comme pour Mamut, elles sont toutes proches de spectres expérimentaux. Certaines composantes sont très similaires, comme les cmp 5 et 7 qui ressemblent toutes les deux au spectre XANES de Cu°. Par ailleurs les composantes semblables aux formes « matrices » cmp 2, 6, 9 et 10 montrent des similitudes dans la position du prépic (8983 eV) et du sommet de la raie blanche (8997 eV). Cette approche n'a pas pu être menée sur les spectres du site de prélèvement B en raison du trop grand nombre de données, mais les composantes de la zone A permettent de reconstruire l'ensemble des spectres de la zone B, à l'exception de ceux semblables à B2BDr4s24 et B2BDr4s16 (Fig. 3.28). Ceci signifie que ces espèces ont disparu de A, probablement par altération chimique dans cette zone humide.

4.3) Comparaison entres les deux sites de prélèvement

Le nombre de phases cuprifères ne parait pas infini. Tout d'abord, le cuivre métallique, sous forme native ou de laiton, et la chalcopyrite ont été clairement identifiés. On a vu également que le cuivre s'associait de manière préférentielle à la matière organique et que cette forme se réduisait sous faisceau. En termes de nombre d'espèces, on peut donc dire qu'il existe deux formes Cu-MO dans nos systèmes. Le cuivre s'associe également aux oxydes de manganèse, aux oxyhydroxydes de fer et aux argiles. Enfin, et comme cela a été vu, en considérant que la concentration en cuivre de certaines phases est véritablement importante, on peut imaginer l'existence de phases non déterminées dans lesquelles le cuivre est une majeure. En majorant par 2, par sécurité, le nombre de phases inconnues observées, on peut alors estimer une fourchette dans laquelle les résultats obtenus par l'ACP paraissent crédibles. Ce nombre de cp est compris entre 5 et 12 pour le site de Mamut et entre 5 et 14 pour le site de DjR.

Ce raisonnement nous permet d'affirmer d'une part, que la cascade des valeurs propres et le calcul des variances associé ne permettent pas de déduire le nombre de composantes principales. Les fortes variances observées résultent des corrélations entre les spectres. Elles valent au minimum 85% pour le jeu de Mamut C3, 90% pour Mamut B4, 92% pour DjR A et 94% pour DjR B. D'autre part, les résultats donnés par la fonction IND et par le F-test sont surestimés. Le nombre de composantes indiqué par IND est corrélé à $R^2 = 93\%$

avec la taille du jeu de données, ce qui montre que le nombre de spectres dans le système est un facteur prépondérant pour expliquer le comportement de ce paramètre. Le résultat du test de Fisher ne semble pas dépendre de la taille du jeu de données, le coefficient de détermination entre le nombre de composantes donné par ce paramètre et le nombre de spectres étant de R^2 = 28%. Les résultats obtenus par l'approche itérative sont consistants avec les estimations, malheureusement en terme de nombre de composantes ils sont suffisants pour reconstruire des systèmes, mais pas forcement nécessaires.

Néanmoins, les composantes déterminées sur les deux sites peuvent être comparées, notamment par TT (Fig 3.29). On peut tester si les 8 composantes déterminées par approche itérative sur le site B4 de Mamut appartiennent aussi à l'espace défini par les 10 composantes principales egalement obtenues par approche itérative du site A de DjR.

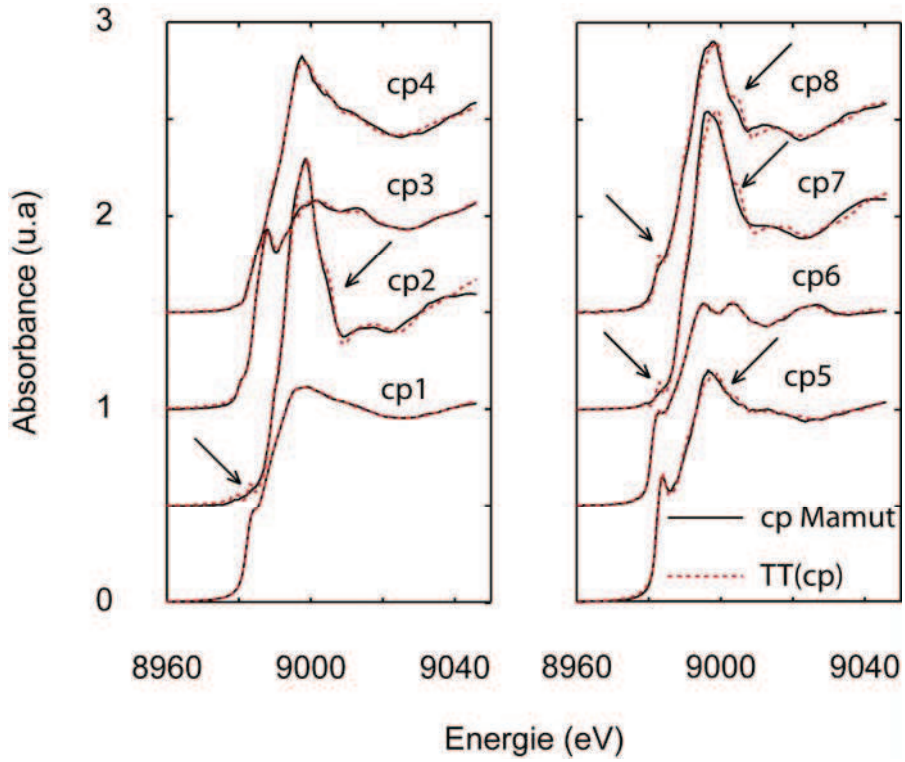


Figure 3.29 : Composantes principales (cp 1 à 8) du site de Mamut déterminées par la méthode itérative (trait plein noir) et testées par Target Transformation avec les dix composantes de DjR (pointillé rouge). Les flèches noires pointent les structures mal reconstruites.

Il apparaît clairement sur la figure 3.29 que les formes proches cp2, cp 7 et cp 8, définies au § 4.1.4.4, n'appartiennent pas à cet espace. A l'inverse, et logiquement car elles ont été repérées sur tous les sites, la chalcopyrite et Cu° (cp3, cp 6) en font bien partie. Pour les espèces de la matrice (cp 1, 4 et 5), il apparait que cp 1 et 4 appartiennent à l'espace, tandis que cp5 n'en fait pas partie. Après examen attentif, il s'agit de la forme la plus réduite

sous faisceau qui n'a pas été recherchée sur le site de DjR. Cette similitude entre les formes matricielles des deux sites, si elle est bien réelle, montre qu'il existe une certaine 'universalité' dans ces formes secondaires que l'on retrouve dans des contextes pédo-géochimiques totalement différents. Néanmoins, une conclusion de ce type ne peut pas être posée à ce stade. En effet, il a largement été montré que les composantes déterminées par ACP étaient extrêmement dépendantes de la qualité (au sens rapport signal sur bruit) des données et qu'il n'existait pas à l'heure actuelle de méthode véritablement précise pour déduire le nombre de ces composantes. Or la similitude entre les espèces Cu-matrices dans des milieux (Mamut en Malaisie et Django Reinhardt près de Lyon) extrêmement contrastés géochimiquement, peut provenir d'une surestimation de la dimension des espaces propres qui engendre, à cause du bruit, la reconstruction de structures fictives.

A ce stade, tant pour l'identification que pour la quantification des espèces cuprifères, il apparaît donc indispensable de développer des critères fiables permettant de déterminer le nombre précis de composantes primaires lors d'une ACP. Cette question est abordée au chapitre suivant.

Chapitre 4

Evaluation statistique des critères ACP usuels utilisés en Spectroscopie d'Absorption X

Proposition d'un nouveau critère : l'indicateur NSS

1) Introduction

Au vu des résultats du chapitre précédent, il apparaît clairement que, sur nos jeux de données expérimentales, aucun des paramètres usuels (cascade des vp, fonction IND et F-Test) ne donne le nombre réel de composantes primaires. Il est alors légitime : (1) de se demander si l'ACP est bien une méthode adaptée pour déterminer la dimension d'un système linéaire de signaux XAS ; (2) de chercher à comprendre pourquoi les indicateurs usuels donnent des résultats biaisés ; et (3) de déterminer dans quelle mesure l'ensemble de ces données peut tout de même être exploité de manière quantitative.

La stratégie choisie pour répondre à ces questions est d'évaluer statistiquement ces indicateurs. Cette évaluation statistique se fera de la manière suivante. Tout d'abord, nous allons créer des jeux de données artificielles en combinant linéairement et de manière aléatoire des spectres synthétiques (fig. 4.1). Bien qu'aucune donnée EXAFS expérimentale n'ait véritablement été traitée dans ce travail, la méthode ACP étant utilisée à la fois en spectroscopie XANES et EXAFS (Panfili 2004; Panfili et al. 2005; Van Damme et al. 2010; Vespa et al. 2010) les deux types de signaux seront étudiés.

Les spectres XANES théoriques de sept espèces cuprifères, $Cu_2S_{(P\ 43212)}$ quadratique (Janosi 1964), $Cu_2S_{(P63\ mmc)}$ 'high chalcocite' (Villars et al. 2006), CuS (Villars et al. 2006), Cu_2O (Niggli 1922), Cu métal (Suh et al. 1988), $CuFeS_2$ (Kratz et Fuess 1989), et CuO (Massarotti et al. 1998), ont été calculés *ab initio* au seuil K du cuivre avec le logiciel FDMNES (Joly 2001). Le domaine en énergie couvert par un spectre XANES étant variable d'une étude à l'autre, deux systèmes ont été générés :

- L'un dont la gamme en énergie est restreinte à la proximité du seuil de -20 à +60 eV (XANES 1). Pour ces données les transitions quadripolaires ont été autorisées afin de reproduire au mieux la signature de la transition électronique 1s→3d dans le prépic. Les spectres générés sont corrélés au maximum à hauteur de 97,3 % (CuS vs $Cu_2S_{(P43212)}$) et au minimum à hauteur de 89,0 % (Cu^0 vs CuO).
- L'autre dont la gamme en énergie s'étend de -40 à +150 eV (XANES 2). Pour ce second jeu, les transitions quadripolaires ont été interdites. Les spectres générés sont corrélés au maximum à 99 % (CuS vs $Cu_2S_{(P43212)}$) et au minimum à 97 % (Cu^0 vs CuO).

Les spectres EXAFS théoriques de six des sept espèces cuprifères précédentes (Cu_2S, CuS, Cu_2O, Cu métal, $CuFeS_2$ et CuO) ont été construits avec le logiciel FEFF v. 7.02 (Ankudinov et Rehr 1997). Ces spectres ont été représentés dans les espaces $\chi(k)$ (EXAFS 1)

et $k^3\chi(k)$ (EXAFS 2). Les spectres de la série EXAFS (1) sont corrélés au maximum à hauteur de 84 % (CuS vs Cu_2S) et au minimum à hauteur de 0 % (CuO vs Cu_2O), tandis que les spectres de la série EXAFS (2) sont corrélés au maximum à hauteur de 66 % (CuS vs Cu_2S) et au minimum à 0 % ($CuFeS_2$ et CuO).

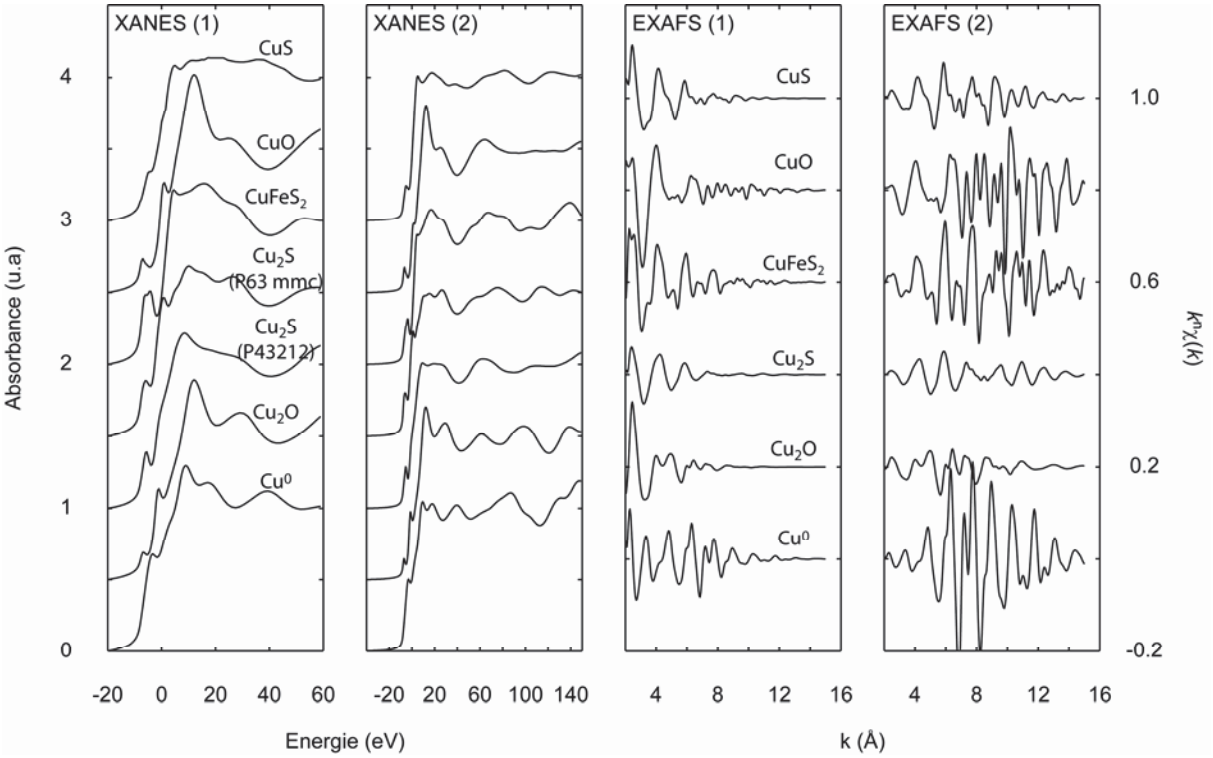


Figure 4.1 : Spectres XAS (XANES et EXAFS) théoriques utilisés pour la création de systèmes pseudo-naturels permettant l'évaluation des critères ACP. Le domaine en énergie relatif au seuil K d'absorption est de -20 à +60 eV pour la série XANES 1 et de -40 à +150 eV pour la série XANES 2. Les spectres EXAFS sont représentés dans l'espace $\chi(k)$ (EXAFS 1) ou dans l'espace $k^3\chi(k)$ (EXAFS 2).

Pour chacune des spectroscopies, les spectres de référence peuvent alors être mélangés (mélange aléatoire pondéré) de manière à reproduire au mieux un système naturel (Fig. 4.2). Les paramètres qui varient dans la construction d'un mélange sont le nombre d'espèces initiales v (2 à 7 dans le cas du XANES et 2 à 6 dans le cas de l'EXAFS), et le nombre de spectres de mélanges t. Cette approche aléatoire (Matsumoto et Nishimura 1998) permet de créer autant de systèmes pseudo-naturels qu'il est souhaité et il devient alors possible de tester statistiquement les indicateurs ACP en fonction du couple (v, t).

Nous vérifierons, tout d'abord, sur ces systèmes synthétiques et à l'aide de la cascade des valeurs propres si l'ACP est une méthode véritablement efficace pour déterminer la dimension d'un système linéaire de données XAS sans bruit experimental.

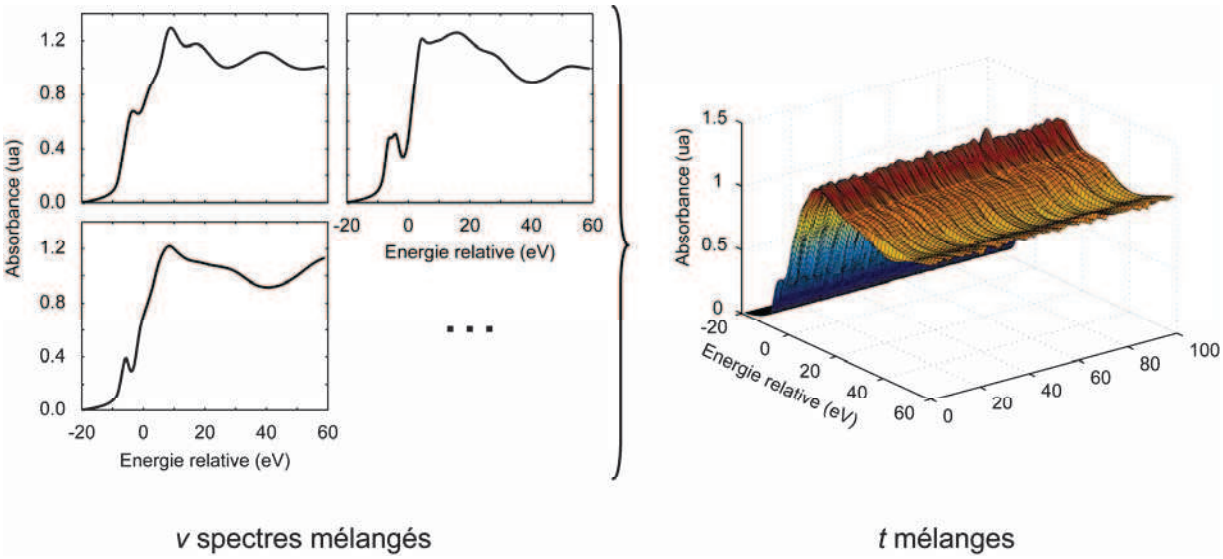


Figure 4.2 : Principe de la création des systèmes multiphasiques synthétiques illustré avec les spectres de la série XANES 1. A partir de *v* spectres initiaux mélangés de manière aléatoire, on calcule *t* spectres de mélanges. Ces *t* spectres forment alors un système synthétique. Dans cet exemple, $v = 7$ et $t = 100$.

Suite aux résultats du chapitre précédent, il apparaît que le bruit (statistique ou non) présent dans les données est un des facteurs qui influe le plus sur les résultats. Dans un second temps, nous déterminerons à l'aide de méthodes de filtrage classiques les niveaux de bruit présents sur nos données expérimentales, puis nous les incorporerons dans nos systèmes synthétiques pour évaluer les performances de la fonction IND et du F-test.

Enfin, nous proposerons un nouvel indicateur dont nous détaillerons la construction et le testerons statistiquement de la même manière que les autres indicateurs. Il sera ensuite appliqué aux données de Mamut et de Django Reinhardt (DjR).

2) Evaluation de l'ACP en spectroscopie d'absorption X – Cascade des valeurs propres

2.1) Principe

Pour v variant de 2 à 7 en XANES et 2 à 6 en EXAFS, et t variant de 10 à 105, on crée des systèmes synthétiques multiphasiques (v, t) pour lesquels on calcule les valeurs propres déduites de l'ACP. On note le numéro de la première composante pour laquelle la valeur propre (vp) λ est nulle. Pour chaque couple (v, t) donné, on répète l'opération 100 fois, ce qui permet de déterminer la moyenne statistique. Le principe est détaillé à la figure 4.3 sur des mélanges obtenus à partir des données synthétiques XANES 1 pour le couple $v = 6$ et $t = 40$. Dans cet exemple, la première composante pour laquelle $\lambda = 0$ est toujours la cp n°7.

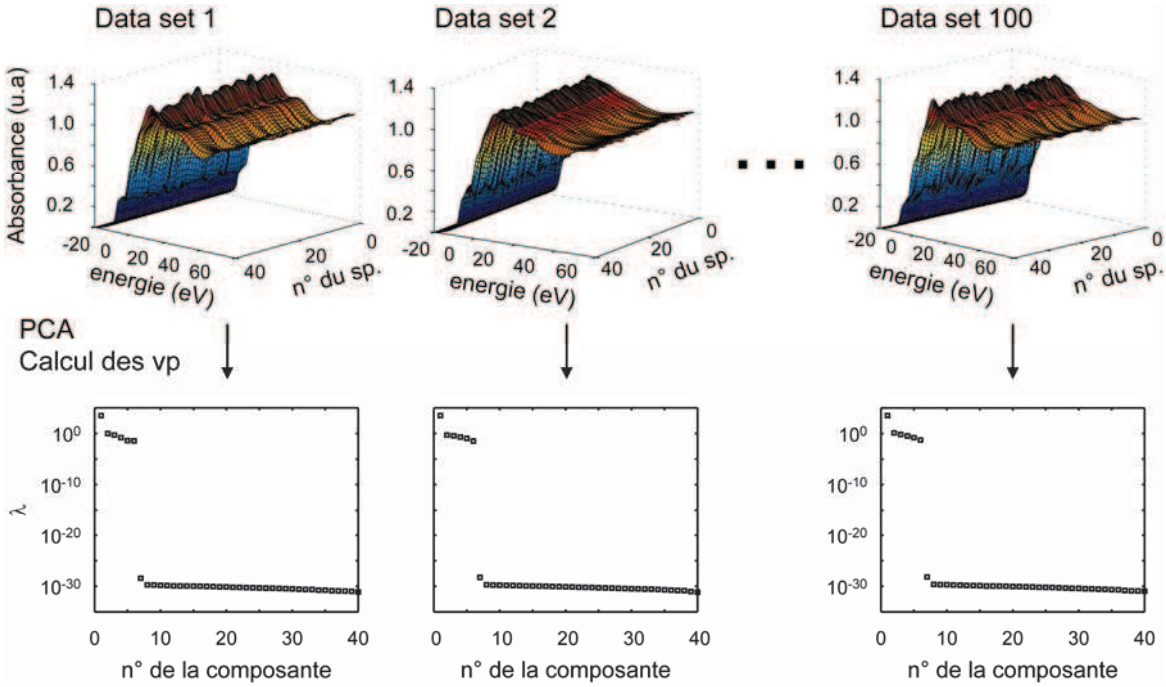


Figure 4.3 : Principe de l'approche statistique pour le calcul du numéro de la première composante dont la valeur propre (vp) est nulle. Dans cet exemple on mélange initialement $v = 6$ spectres provenant du jeu de données synthétiques XANES 1 en $t = 40$ combinaisons de manière aléatoire. Sur le graphique, il apparaît que la $\lambda(cp7) > \lambda(cp8) \neq 0$, mais il s'agit simplement d'un artefact de calcul ($\lambda(cp7) < 10^{-25}$) exacerbé par une représentation graphique semi-logarithmique.

2.2) Résultats

2.2.1) XANES 1 et 2

Les résultats obtenus à partir des mélanges (v, t) des données synthétiques XANES 1 sont représentés à la figure 4.4. Il apparaît que le numéro de la première valeur propre nulle est égal à v+1 et $\sigma = 0$ même quand, dans un même jeu de données, certains spectres sont très

corrélés. Le maximum des corrélations a été mesuré à $R^2 = 1 - 6{,}92\ 10^{-11}$. Les résultats sont identiques pour les mélanges obtenus à partir des données synthétiques XANES 2, excepté le maximum des corrélations qui cette fois est $R^2 = 1 - 3{,}38\ 10^{-11}$.

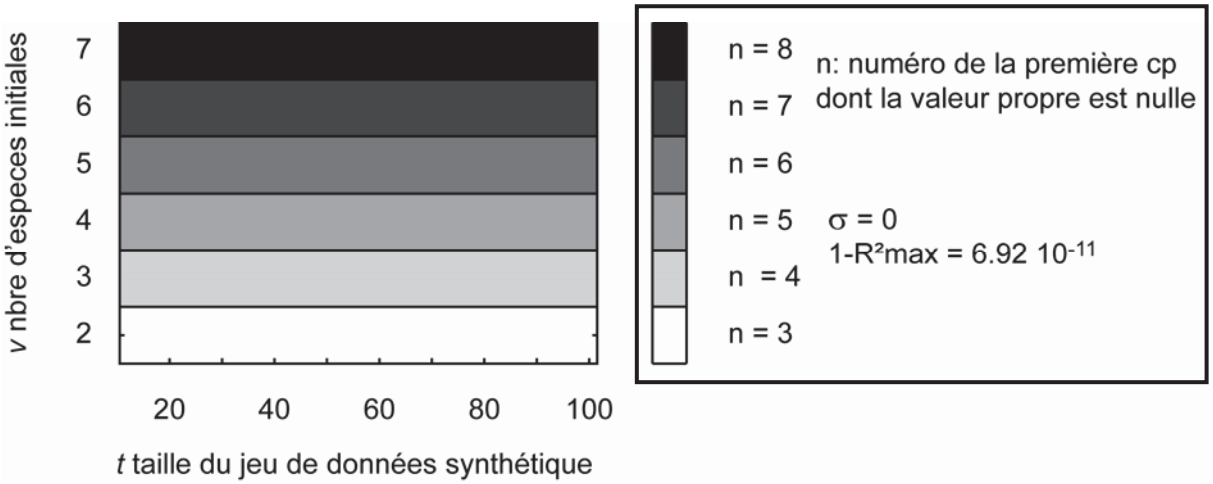


Figure 4.4 : Résultats statistiques sur 100 tirages du calcul de la première vp nulle lorsque l'ACP est appliquée sur des mélanges de 10 à 105 spectres issus des données synthétiques XANES 1. Seule la valeur moyenne est représentée car l'erreur est nulle ($\sigma = 0$).

2.2.2) EXAFS 1 et 2

Les résultats obtenus à partir des mélanges (v, t) des données synthétiques EXAFS 1 sont détaillés à la figure 4.5. Comme précédemment, la première composante pour laquelle $\lambda = 0$ est la cp v+1 avec un maximum de corrélations mesuré de $R^2 = 1 - 2{,}96\ 10^{-11}$. Les résultats sont les mêmes pour les mélanges obtenus à partir des données EXAFS 2. Le maximum des corrélations est de $R^2 = 1 - 1{,}74\ 10^{-11}$.

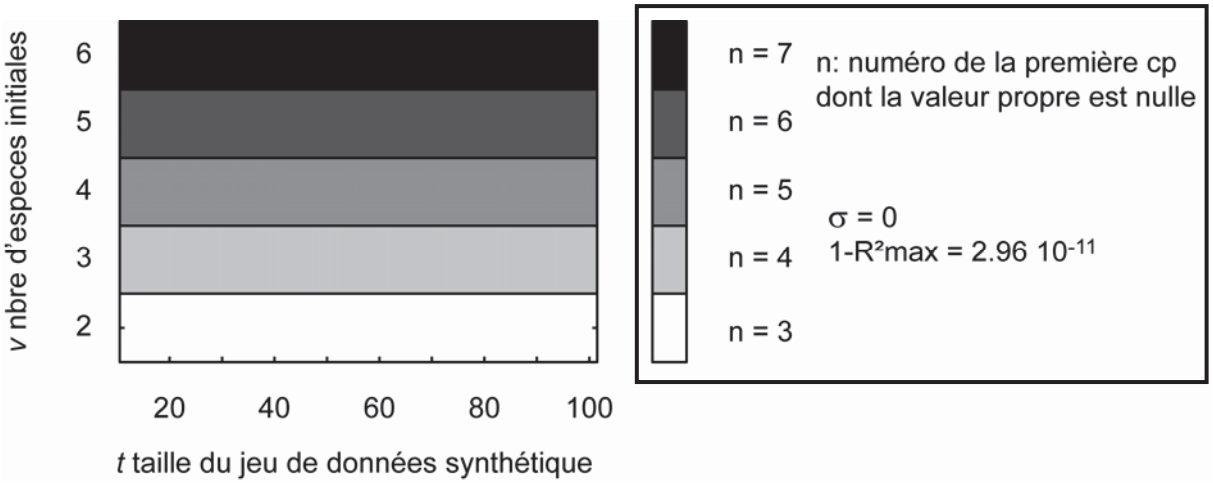


Figure 4.5 : Résultats statistiques (100 tirages) du calcul de la première vp nulle lorsque l'ACP est appliquée sur des mélanges ($10 \leq t \leq 105$) obtenus à partir des données synthétiques EXAFS 1. Seule la valeur moyenne est représentée, car l'erreur est nulle ($\sigma = 0$).

2.3) Conclusion

Indépendamment du type de données XAS, pour un mélange (v, t), la première vp nulle est λ_{v+1}, ce qui montre que l'ACP est adaptée à l'analyse de jeux de données XAS à plusieurs composantes, pour autant que le signal soit pur. Ces résultats étaient prévisibles au regard de la définition mathématique de l'ACP. Pourtant, lorsque ce critère est appliqué sur des jeux de données expérimentales, la cassure des valeurs propres (Fig. 4.3) disparait (Figs. 3.16 et 3.25). Ceci confirme que les bruits statistiques et non statistiques des signaux expérimentaux sont 'vus' par l'ACP comme autant de composantes indépendantes. La forte proximité (au sens du coefficient de corrélations) dans les jeux de données multiphasiques de XAS est responsable de l'importance de la première valeur propre qui écrase les autres composantes principales (cp) et les rend difficiles à distinguer des composantes secondaires. Pour évaluer de façon réaliste les deux autres critères (IND et F-Test) il est donc indispensable d'ajouter artificiellement du bruit sur les jeux de données synthétiques à des taux similaires de ceux des données expérimentales.

3) Mesure du bruit expérimental et génération sur des données synthétiques

3.1) Introduction

Il existe deux types de bruit sur des données expérimentales : le bruit statistique et le bruit systématique. Ce bruit systématique est dépendant de l'opérateur. Il résulte en général de l'acquisition et du traitement des données, principalement la normalisation de l'avant et de l'après seuil. L'estimation de ce bruit systématique est très difficile et les approches de traitement empiriques. Il est généralement approché par des fonctions proportionnelles au bruit statistique sur lesquelles sont appliqués des facteurs correcteurs (IXS-Standards-and-Criteria-Committe, 2000). Compte tenu de la taille de nos observables (une centaine de spectres par jeu de données), nous avons choisi de considérer ce bruit comme faisant partie du bruit statistique.

Par définition, le niveau de bruit statistique sur des données expérimentales peut être calculé en soustrayant une fonction χ' lissée du bruit expérimental aux données χ. La fonction lissée χ' est obtenue en filtrant les données, le rapport recommandant soit l'utilisation d'un polynôme dont le degré est adapté au nombre de points de la donnée (Dent et al. 1992; Filipponi 1995), soit en utilisant un filtre passe bas (IXS-Standards-and-Criteria-Committe, 2000). En chaque point de la donnée, il est alors possible de définir :

$$\varepsilon_i = \chi_i - \chi_i'$$

L'erreur statistique moyenne ε_{stat} de l'ensemble du spectre est alors définie comme étant l'écart-type de l'ensemble des ε_i.

Ensuite, sur des zones où le bruit est statistiquement constant, on peut déterminer le rapport signal sur bruit :

$$S/N = \frac{1}{npt}\sqrt{\sum_{i=1}^{npt} \chi_i^2 / \varepsilon_{stat}^2}$$

Dans notre cas, l'utilisation brute de cette formule est difficile. La statistique qui permet de décrire l'occurrence d'interaction entre les rayons X incidents et la section efficace d'interaction d'un atome est de type Poissonnienne (IXS-Standards-and-Criteria-Committe, 2000), ce qui signifie que l'incertitude sur le nombre de photons mesurés est proportionnelle à la racine carré de ce nombre de photons. Or les spectres XAS sont normalisés pour s'affranchir de la concentration en atome cible dans l'échantillon. Cette normalisation de l'absorbance n'est pas préjudiciable dans le cas du XANES puisqu'une partie essentielle du

signal final oscille autour de l'unité, et il est alors possible d'appliquer la formule en prenant en compte le caractère Poissonnien du bruit. En revanche, dans le cas de l'EXAFS cette normalisation est préjudiciable. Les signaux EXAFS normalisés, non seulement oscillent autour de zéro, mais de plus l'absorbance normalisée $\chi(k)$ est généralement multipliée par une puissance du vecteur d'onde ($k^n\chi(k)$, n variant de 0 à 3). Il n'est alors plus possible de prendre en compte le caractère Poissonnien du bruit et d'appliquer la formule *S*/*N* précédente.

3.2) XANES

3.2.1) Evaluation du bruit

3.2.1.1) Principe

Pour mesurer le niveau de bruit (rapport S/N) de nos données XANES normalisées, et lui associer une distribution, il a été procédé de la manière suivante (Fig. 4.6) :

1) définition de la zone filtrée ; conformément aux recommandations du Comité de Standardisation de la Société Internationale de Spectroscopie d'Absorption X (IXS-Standards-and-Criteria-Committe, 2000), le rapport S/N a été mesuré au-delà du maximum d'absorption car celui-ci peut être considéré en bonne approximation comme constant et le signal oscille autour de l'unité ;

2) filtrage des données par un filtre adapté et définition du résidu (ε_i) ;

3) mesure de $\varepsilon_i/data_i^2$ pour l'ensemble des points filtrés et détermination de l'écart-type σ des variables aléatoires $[\varepsilon_i/data_i^2]$;

4) calcul de $\rho_i = \sigma(data_i)^{0.5}$ puis du rapport signal sur bruit, *npt* étant le nombre de points :

$$S/N = \frac{1}{npt}\sqrt{\sum_{i=1}^{npt} data_i^2 / \rho_i}$$

Pour valider cette approche numérique, deux techniques de filtrage ont été utilisées:

- filtrage de Butterworth

Les filtres de Butterworth sont communément utilisés en traitement du signal (Phoong et al. 1995). Ils sont caractérisés par une amplitude qui est maximale et constante dans la bande passante et monotone à l'extérieur. On cherche à filtrer la descente de la raie blanche ainsi que la première oscillation EXAFS, soit des signaux d'une période de l'ordre d'une vingtaine d'eV (Fig. 4.6). Par sécurité, on choisit de filtrer les signaux dont les périodes sont

inférieures à 10 eV, soit par un filtre passe bas avec une fréquence de coupe de 0,1 eV^{-1}. On note S/N_{butt} le taux de bruit obtenu avec cette technique.

- filtrage par ondelettes

Cette technique de filtrage a été utilisée en chimie pour supprimer le bruit de données spectrales (Barclay et al. 1997). Elle a également été utilisée pour déterminer les oscillations EXAFS de spectres enregistrés en mode transmission (Shao et al. 1998). Les signaux sont projetés sur deux composantes et un seuillage doux (soft-thresholding) est appliqué (Donoho 1995). On note S/N_{wav} le taux de bruit obtenu avec cette technique.

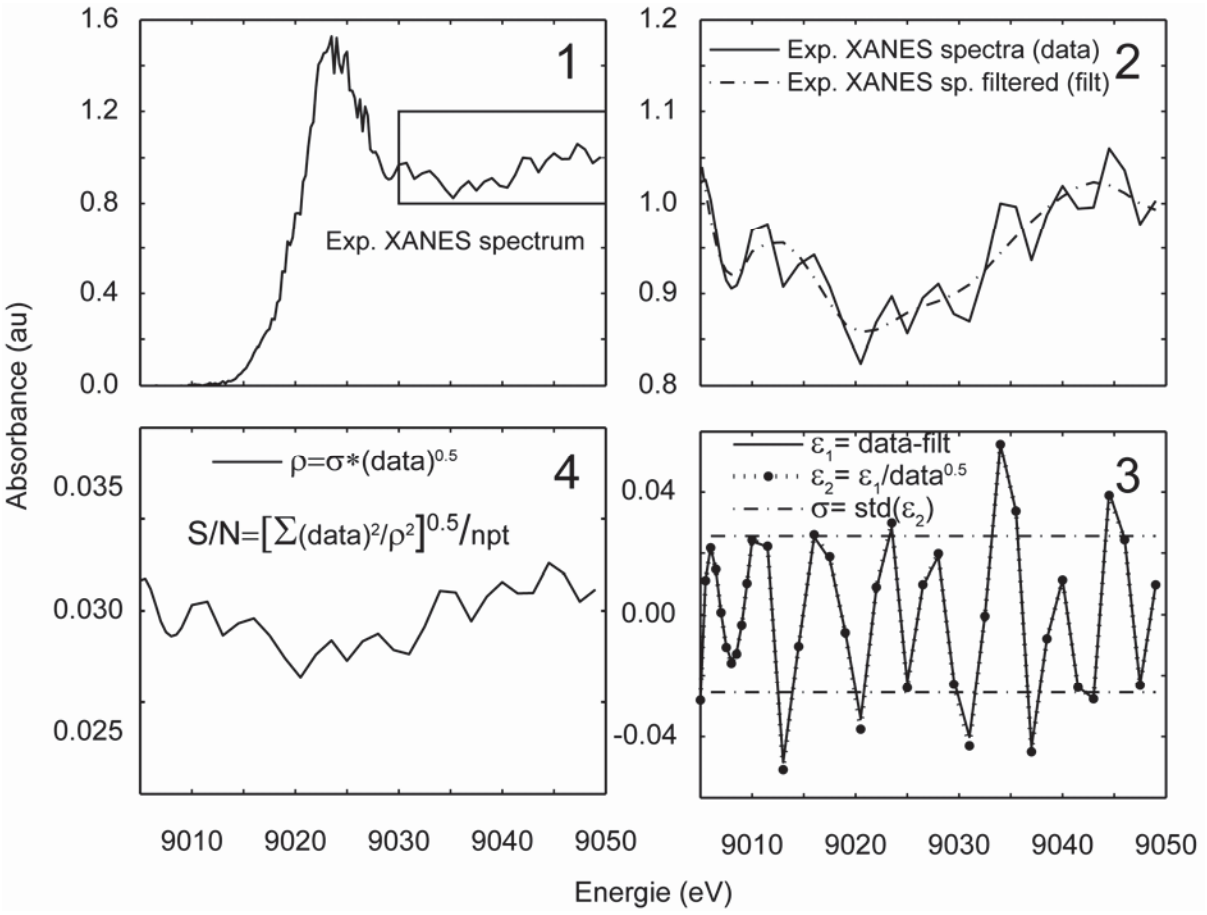


Figure 4.6 : Etapes du calcul du niveau de bruit. Pour plus de visibilité, un spectre XANES expérimental très bruité du site Mamut C3 a été choisi.

3.2.1.2) Résultats et discussion

Les distributions statistiques des rapports S/N calculés avec les deux approches sont comparées à des lois normales à la figure 4.7. Les deux types de filtrage proposent une distribution statistique similaire. Dans le cas de Mamut (C3 et B4), une description statistique normale est pertinente car les données sont proches de la droite de Henry. Dans le cas de DjR, il existe une dérive quel que soit le filtrage utilisé, la probabilité d'obtenir une valeur S/N > 40

expérimentale étant inférieure à la probabilité théorique obtenue à partir d'une distribution gaussienne. Sur les deux sites DjR A et B, seulement 10 % des données ont un bruit non-statistique détectable par analyse numérique, et par la suite on considèrera que la description normale est néanmoins acceptable, bien qu'erronée en toute rigueur.

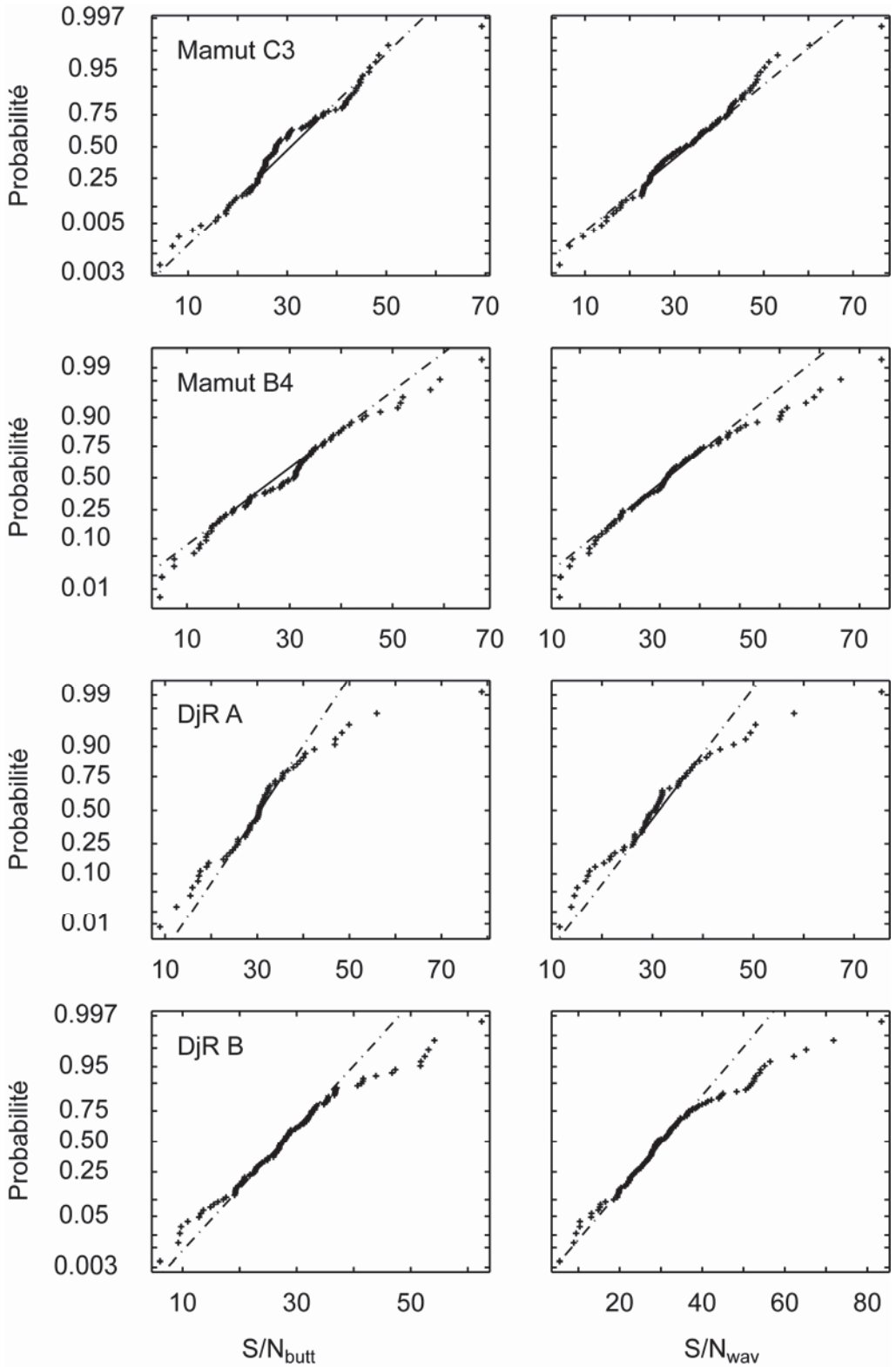


Figure 4.7 : Description des rapports S/N expérimentaux en termes de probabilité sur un papier de Gauss et comparaison à la droite de Henry.

Les valeurs des rapports S/N calculés avec les deux approches sur les sites C3 et B4 de Mamut et A et B de DjR sont synthétisées dans le tableau 4.1. En termes de moyennes et d'écarts-types, non seulement les deux types de filtrage donnent des résultats similaires pour un même jeu de données, mais de plus les quatre jeux eux-mêmes présentent des niveaux de bruits semblables : en moyenne S/N = 31,09 (± 1,82) avec un écart-type moyen associé de 12,18 (± 2,11) environ.

Tableau 4.1 : Statistiques descriptives des rapports S/N des données μ-XANES expérimentales

	Moyenne	Ecart-type	Valeur maximale	Valeur minimale
Mamut C3				
S/N_{butt}	30.08	10.56	69.29	4.31
S/N_{wav}	32.73	11.98	76.36	4.30
Mamut B4				
S/N_{butt}	29.08	12.66	67.35	4.75
S/N_{wav}	33.80	16.66	85.51	4.93
Django Reinhard A				
S/N_{butt}	31.24	11.14	78.76	8.87
S/N_{wav}	31.33	11.15	75.46	11.60
Django Reinhard A				
S/N_{butt}	28.43	9.98	62.45	6.04
S/N_{wav}	32.05	13.29	83.44	5.48
Synthèse	31.09 (1.82)	12.18 (2.11)	/	/

La dernière question qui se pose est la suivante : pour une donnée, les deux filtrages sont-ils équivalents pour évaluer le rapport S/N ? La réponse donnée à la figure 4.8 est oui : les coefficients de corrélation (seuil de confiance égal à 5%) entre les rapports S/N déterminés par les filtrages de Butterworth et par ondelettes sont de 92% (Mamut C3), 91% (Mamut B4), 98% (DjR A) et 79% (DjR B).

3.2.2) Génération de bruit sur les données synthétiques

Le bruitage d'un mélange XANES synthétique (v, t) se fait de la manière suivante :

1) calcul aléatoire : (1) d'un rapport S/N_G calculé selon une loi normale $N(31.09,1.82)$ et (2) d'un écart-type σ_G calculé selon une loi normale $N(12.18,2.11)$;

2) génération aléatoire de t rapports S/N suivant la loi normale $N(S/N_G,\sigma_G)$;

3) spectre par spectre, calcul du σ défini au § 3.2.1 (Fig. 4.6) suivant la formule :

$$\sigma = \frac{1}{(S/N)npt}\sqrt{\sum_{i=1}^{npt} data_i}$$

4) génération des *npt* variables aléatoires ε_2 suivant une loi $N(0,\sigma)$, puis calcul de $\varepsilon_1=\varepsilon_2*data_i^{0.5}$et addition de ε_1 au signal.

Une illustration du résultat de la procédure est donnée à la figure 4.9.

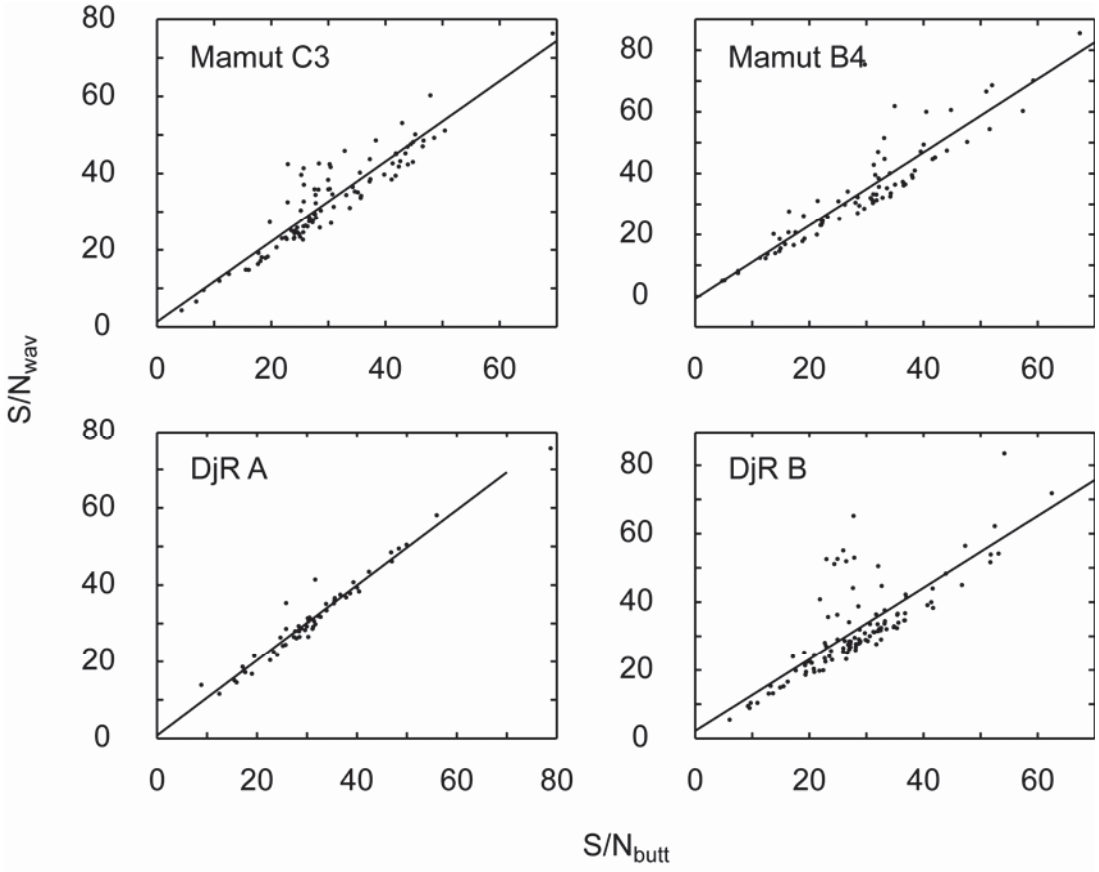


Figure 4.8 : Comparaison entre les rapports S/N_{butt} et S/N_{wav}. Les coefficients de corrélation sont de 92% (Mamut C3), 91% (Mamut B4), 98% (DjR A) et 79% (DjR B).

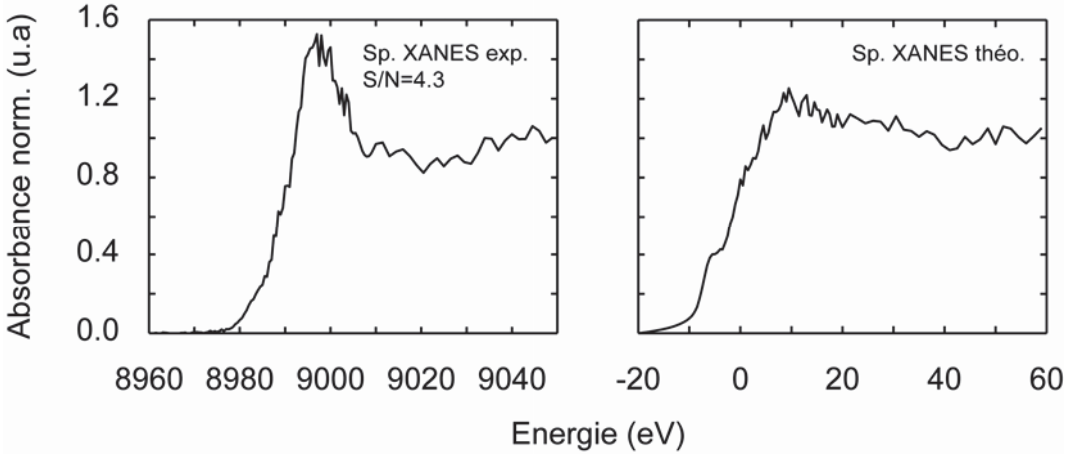


Figure 4.9 : Comparaison visuelle entre une donnée expérimentale possédant un niveau de bruit S/N=4.3 (à gauche) et une donnée théorique sur laquelle le même niveau de bruit a été introduit (à droite). Pour une meilleure perception visuelle, la donnée choisie était volontairement très bruitée.

3.3) EXAFS

3.3.1) Evaluation du bruit

3.3.1.1) Principe

Aucune donnée EXAFS enregistrée sur les sites de Mamut et DjR n'ayant été étudiée dans ce travail de thèse, les niveaux de bruit de spectres EXAFS ont été calculés à partir de données publiées dans la littérature. Il s'agit de spectres EXAFS enregistrés soit au seuil K du zinc (Isaure et al. 2002; Panfili et al. 2005; Van Damme et al. 2010), soit au seuil K du cuivre (Manceau et Matynia 2010) et sur lesquels l'ACP a déjà été appliquée dans ces articles pour évaluer le nombre d'espèces pures présentes dans ces systèmes. Les conditions d'enregistrement des données sont précisées dans les articles originaux.

L'objectif étant de mesurer des taux de bruit expérimentaux pour les reporter sur des spectres théoriques, il est important que pour l'ensemble des spectres, l'amplitude des oscillations EXAFS soit comparable et indépendante de k. Or les termes d'amplitude, comme le nombre de voisins (N_i), leur distance (R_i), le désordre σ_i (agitation thermique et désordre statique), le libre parcours moyen du photoélectron et le facteur de réduction d'amplitude (S_0) sont des paramètres structuraux et électroniques qui dépendent de la nature de l'espèce chimique. De plus, pour une même espèce leur dépendance en k fait que le signal s'amortit aux hautes valeurs de k.

L'approche retenue pour mesurer le niveau de bruit sur les jeux de données EXAFS ($k^3\chi(k)$ et $\chi(k)$) est la suivante. Les spectres sont d'abord normalisés par leur enveloppe. Cette opération permet d'obtenir des données qui oscillent entre -1 et 1. Les spectres renormalisés sont alors filtrés avec un algorithme par ondelettes (Shao et al. 1998) et les résultantes déterminées par soustraction. Sur l'ensemble d'un jeu de données, les résultantes sont alors moyennées (moyenne et écart-type) en fonction de k. La procédure est décrite sur un spectre à la figure 4.10.

3.3.1.2) Résultats

<u>Jeux de données EXAFS A</u>

Il s'agit de 29 spectres EXAFS enregistrés au seuil K du zinc résultants du mélange de 3 espèces pures zincifères (Van Damme et al. 2010). Les spectres sont représentés à la figure 4.11.

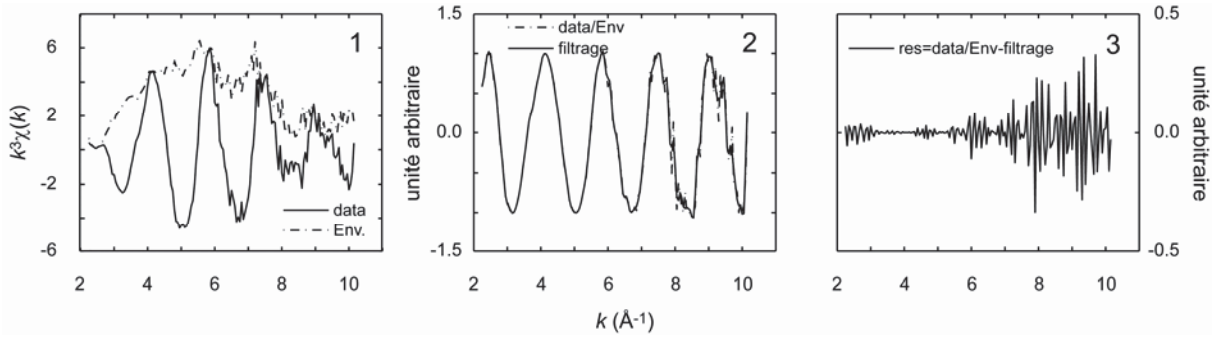


Figure 4.10 : Etapes du calcul du niveau de bruit sur un spectre EXAFS expérimental. La première étape consiste à calculer l'enveloppe (Env.) du signal (data). Il est alors renormalisé par cette enveloppe (data/Env) puis filtré avec un filtrage par ondelettes (filtrage). La soustraction des deux traces donne la résultante (res). Ici il s'agit d'un spectre représenté dans l'espace $k^3\chi(k)$

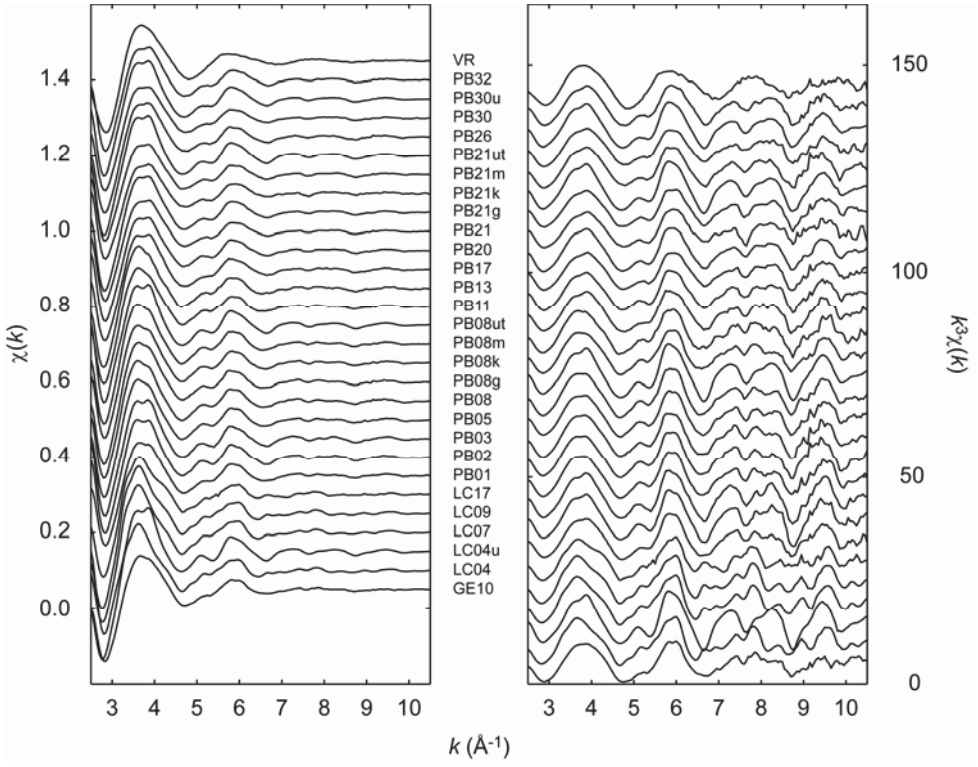


Figure 4.11 : Données expérimentales, à gauche dans l'espace $\chi(k)$ et à droite dans l'espace $k^3\chi(k)$, du jeu de données de spectres EXAFS A enregistrés au seuil K du zinc.

Les moyennes des résultantes en fonction de k oscillent autour de 0 pour les deux types de représentation, $\chi(k)$ et $k^3\chi(k)$ (Fig. 4.12). Elles ont des amplitudes faibles, de l'ordre de 0,005, pour $k < 7$ Å^{-1}. Elles augmentent d'un ordre de grandeur à $k \approx 7$-8 Å^{-1}. Dans l'espace $\chi(k)$, l'amplitude redécroît à $k > 8$ Å^{-1}. Les écarts-types moyens suivent la même évolution. Dans l'espace $k^3\chi(k)$, les écarts-types augmentent avec k de 0,0 jusqu'à 0,2-0,3 environ. Il n'y a pas de corrélation entre la valeur absolue des moyennes et les écarts-types (R = 0,42 dans l'espace $\chi(k)$ et R = 0,48 en $k^3\chi(k)$).

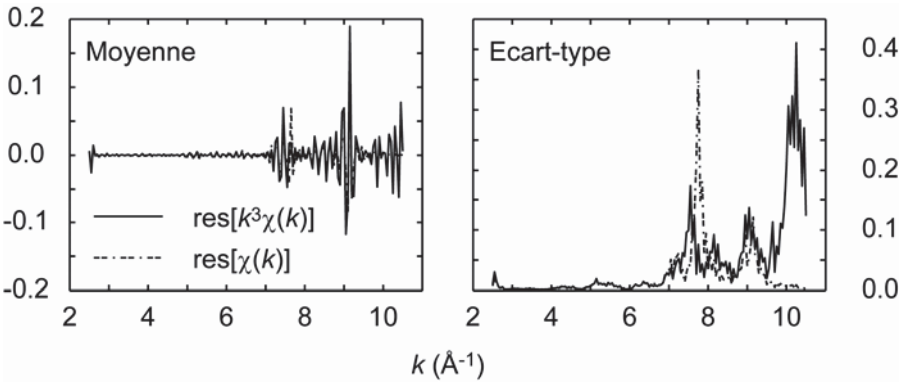


Figure 4.12 : A gauche, moyennes en fonction de *k* des résultantes de l'opération de calcul du bruit sur le jeu de données EXAFS A. En pointillé résultats dans l'espace $\chi(k)$, en trait plein, dans l'espace $k^3\chi(k)$. A droite, écarts-types des résultantes en fonction de *k*.

Jeux de données EXAFS B

Il s'agit de 27 spectres EXAFS enregistrés au seuil K du cuivre résultants du mélange de 3 espèces pures cuprifères (Manceau et Matynia 2010). Ils sont représentés à la figure 4.13.

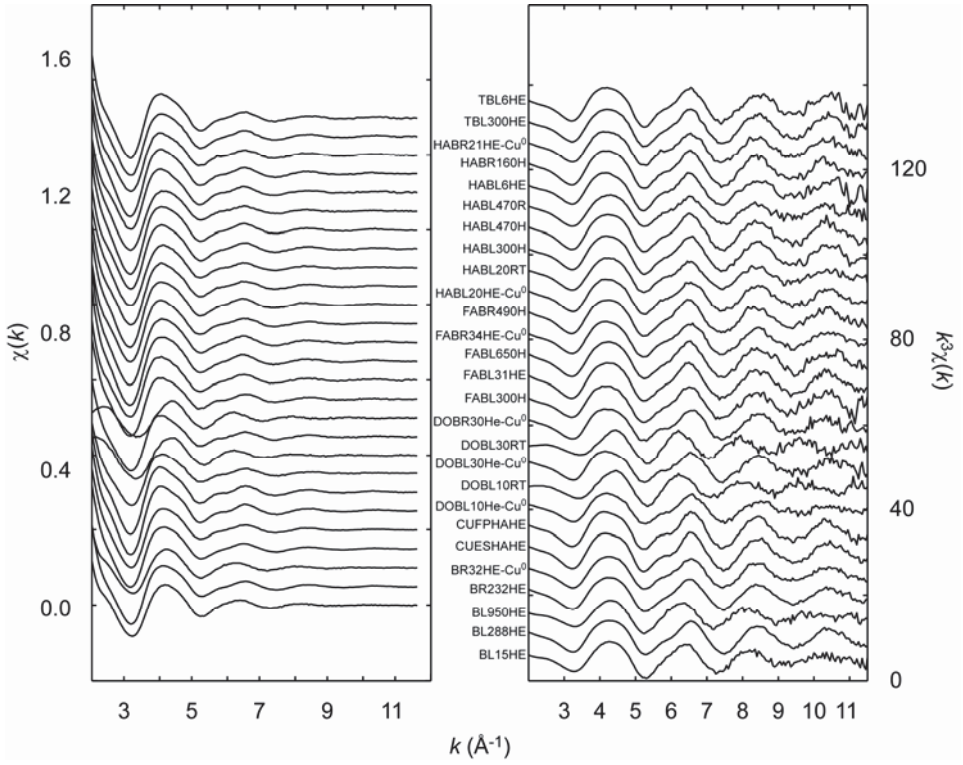


Figure 4.13 : Données expérimentales du jeu de données EXAFS B représentées à gauche dans l'espace $\chi(k)$ et à droite dans l'espace $k^3\chi(k)$.

Les moyennes et les écarts-types des résidus en fonction de *k* sont représentés à la figure 4.14. Dans l'espace $\chi(k)$, les moyennes tendent à augmenter (en valeur absolue) jusqu'à atteindre 0.2 puis décroissent brutalement à 10,5 Å^{-1}. Dans l'espace $k^3\chi(k)$, les moyennes des résidus sont de l'ordre de 0,005 jusqu'à $k \approx 7$ Å^{-1}, puis elles augmentent d'un ordre de grandeur pour les *k* supérieurs. Dans l'espace $\chi(k)$, les écarts-types augmentent puis chutent

brutalement autour de 10,5 $Å^{-1}$ tandis que dans l'espace $k^3\chi(k)$, les écarts-types augmentent avec k de 0,0 jusqu'à 0,3 environ. Les coefficients de corrélation entre les écarts-types et les valeurs absolues des moyennes sont égaux à 68 % dans l'espace $\chi(k)$ et à 48 % dans l'espace $k^3\chi(k)$.

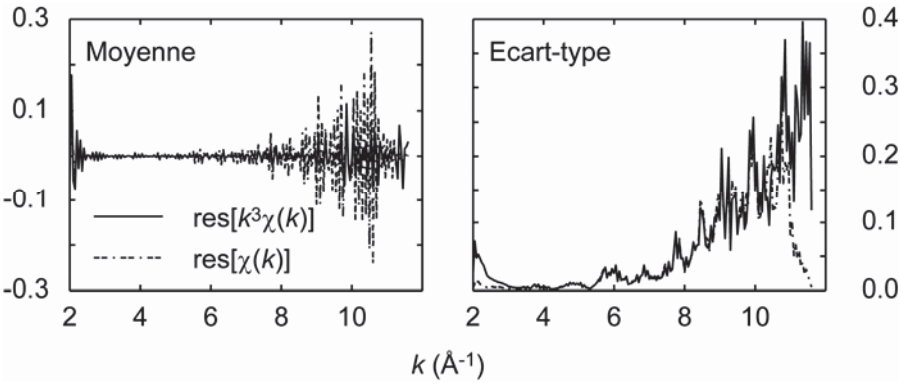


Figure 4.14 : A gauche, moyennes des résultantes de l'opération de calcul du bruit sur le jeu de données EXAFS B. En pointillé, résultats dans l'espace $\chi(k)$, en trait plein, dans l'espace $k^3\chi(k)$. A droite, écarts-types des résultantes.

Jeux de données EXAFS C

Il s'agit de 13 spectres EXAFS enregistrés au seuil K du zinc résultant du mélange de 3 espèces pures zincifères (Isaure et al. 2002). Ils sont représentés à la figure 4.15.

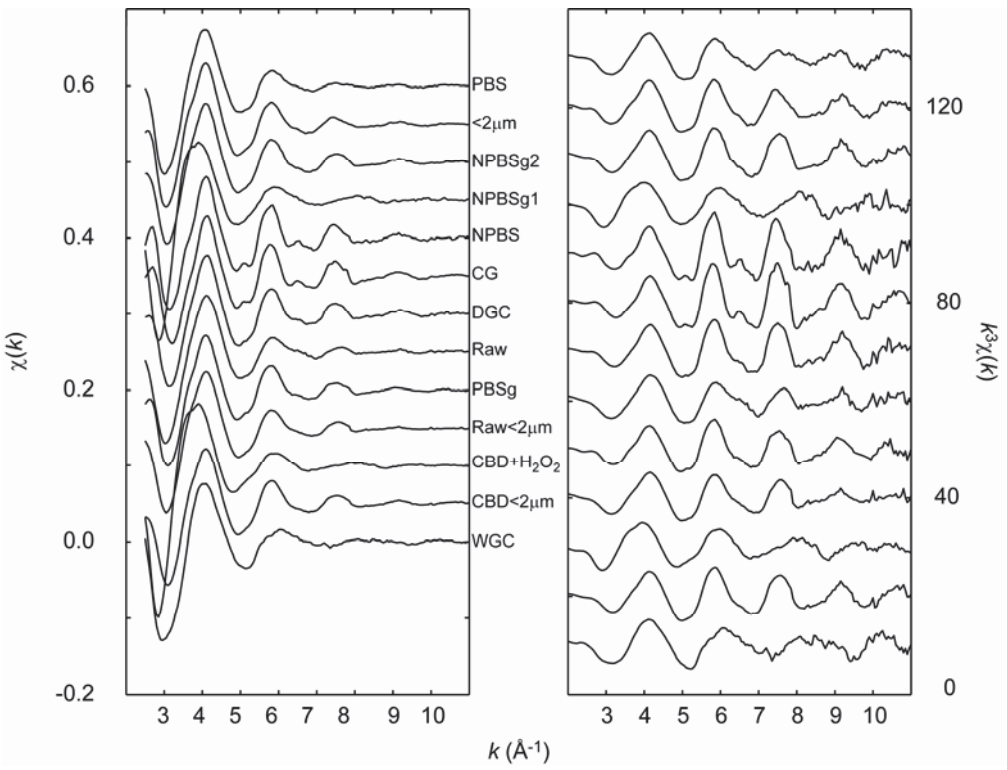


Figure 4.15 : Données expérimentales du jeu de données EXAFS C représentées à gauche dans l'espace $\chi(k)$ et à droite dans l'espace $k^3\chi(k)$.

Sur l'ensemble du jeu de données, les moyennes et les écarts-types des résidus en fonction de k sont représentés à la figure 4.16. Dans l'espace $\chi(k)$ les moyennes augmentent jusqu'à 0,05 à k = 9 Å^{-1} environ puis décroissent. Dans l'espace $k^3\chi(k)$, les moyennes des résidus sont de l'ordre de 0,005 jusqu'à k de l'ordre de 6 Å^{-1}. Puis elles augmentent jusqu'environ 0,5 aux k supérieurs. Dans l'espace $\chi(k)$, les écarts-types des résidus augmentent puis chutent autour de 9 Å^{-1}, tandis qu'ils augmentent régulièrement jusqu'à 0,2-0,3 dans l'espace $k^3\chi(k)$. Les coefficients de corrélation entre les écarts-types et les valeurs absolues des moyennes sont de 78 % dans l'espace k vs $\chi(k)$ et de 76 % dans l'espace k vs $k^3\chi(k)$.

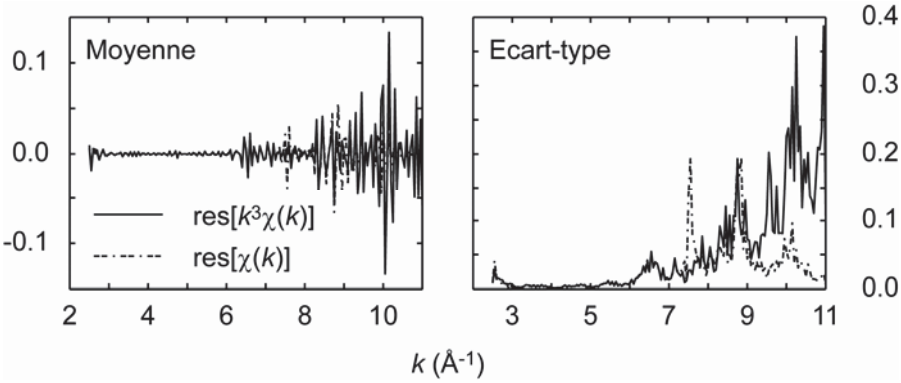


Figure 4.16 : A gauche, moyennes des résultantes de l'opération de calcul du bruit sur le jeu de données EXAFS C. En pointillé, résultats dans l'espace $\chi(k)$, en trait plein, dans l'espace $k^3\chi(k)$. A droite, écarts-types des résultantes.

Jeux de données EXAFS D

Il s'agit de 31 spectres EXAFS enregistrés au seuil K du zinc résultant du mélange de 5 espèces pures zincifères (Fig. 4.17) dont une est non détectable par ACP car constante en concentration dans les matrices étudiées (Panfili et al. 2005). Dans les deux espaces (Fig. 4.18), les moyennes des résultantes sont de l'ordre de 0,005 quand k < 7 Å^{-1} et augmentent autour de 0,03-0,04 pour les k supérieurs. Dans l'espace $\chi(k)$ les écarts-types des résidus augmentent puis chutent autour des k = 7 Å^{-1}. Dans l'espace $k^3\chi(k)$ les écarts-types des résidus augmentent des bas k vers les hauts k jusqu'à 0,25 environ. Les coefficients de corrélation entre les écarts-types et les valeurs absolues des moyennes sont égaux 61 % dans l'espace $\chi(k)$ et à 73 % dans l'espace $k^3\chi(k)$.

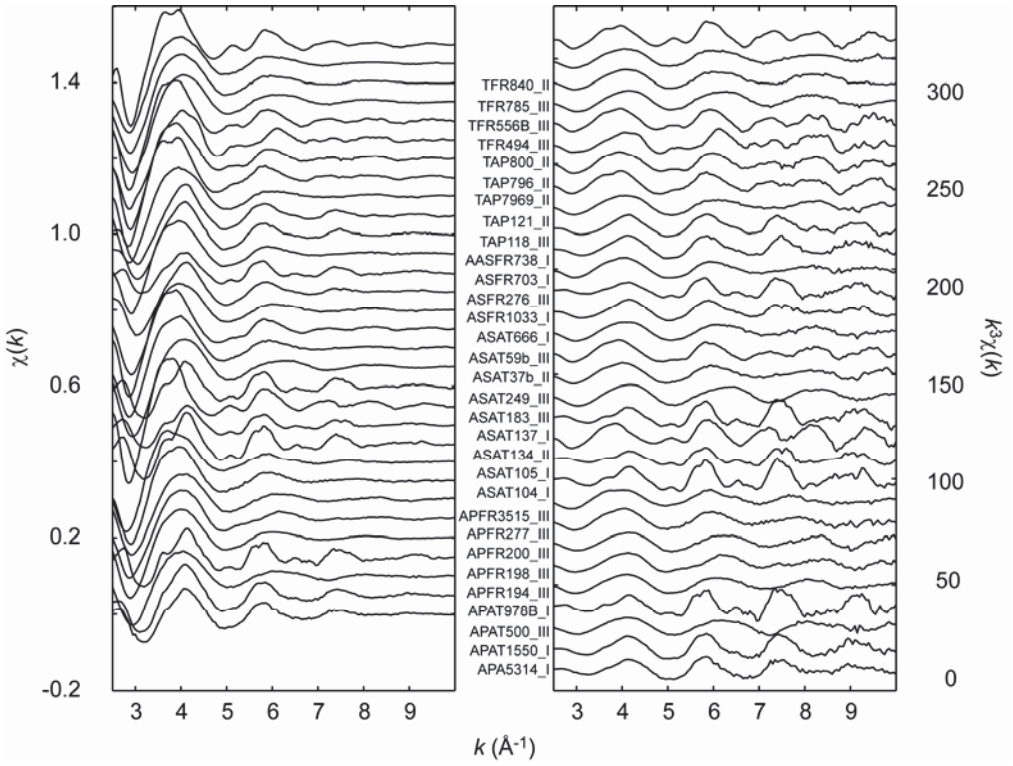


Figure 4.17 : Données expérimentales du jeu de données EXAFS D représentées à gauche dans l'espace $\chi(k)$ et à droite dans l'espace $k^3\chi(k)$.

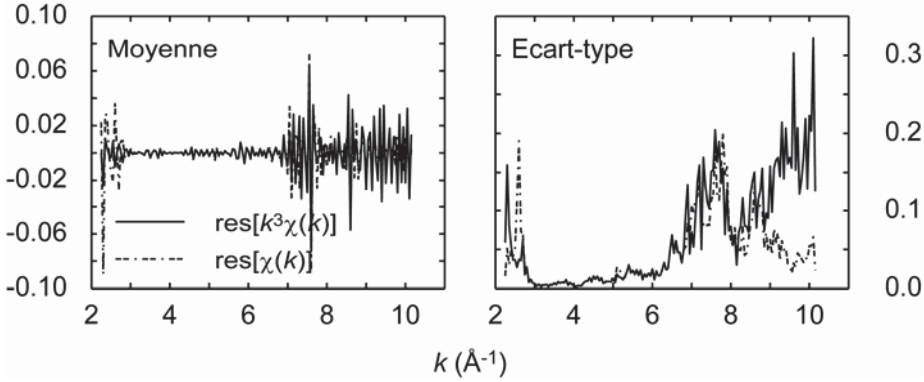


Figure 4.18: A gauche, moyennes des résultantes de l'opération de calcul du bruit sur le jeu de données EXAFS D. En pointillé, résultats dans l'espace $\chi(k)$, en trait plein, dans l'espace $k^3\chi(k)$. A droite, écarts-types des résultantes

3.3.1.3) Discussion

Hors saut d'absorption, les sections efficaces de photo-ionisation décroissent quand E augmente (Yeh et Lindau 1985), et la probabilité d'interaction entre un photon et un électron de l'atome est de type Poissonnienne, ce qui signifie que le bruit est proportionnel à la racine de la probabilité d'interaction. Soit, si on traduit le rapport S/N d'un spectre EXAFS en termes de section efficace d'interaction σ, il est égal à √σ, avec σ décroissante quand E augmente ce qui signifie que relativement à l'intensité du signal, le bruit augmente avec E. L'augmentation monotone observée dans l'espace $k^3\chi(k)$ est donc en accord avec la

diminution du signal vis-à-vis du bruit d'un spectre EXAFS. En revanche, les ruptures observées en $\chi(k)$ ne sont pas physiquement cohérentes. Cette opération est donc écartée et les niveaux de bruits seront mesurés à partir des spectres représentés dans l'espace $k^3\chi(k)$.

Pour les quatre jeux de données et les deux types de représentations, les valeurs des corrélations ne montrent pas de véritables liens entre moyenne et écart-type. Cependant, si on ne regarde pas point par point, mais plutôt d'un point de vue général, il ressort clairement sur les graphes que l'écart-type suit l'augmentation de la moyenne. Or, si le filtrage est bien effectué, les valeurs des spectres différences devraient osciller autour de 0 indépendamment de k. La moyenne étant entouré d'une incertitude, plus l'incertitude est grande, plus la moyenne mesurée de n mesures pourra être différente de la vraie valeur moyenne. Pour la suite, on considère que seule l'augmentation de l'écart-type est significative, les variations de la moyenne pouvant être assimilées à cette augmentation.

Pour la suite, on approche l'écart-type par une fonction de k du type $f(k) = a\exp^{(bk)}$, avec a et b ajustés au jeu de données (Fig. 4.19). Il apparaît graphiquement que cette famille de fonction reprend assez bien le comportement général des écarts-types des résidus en fonction de k.

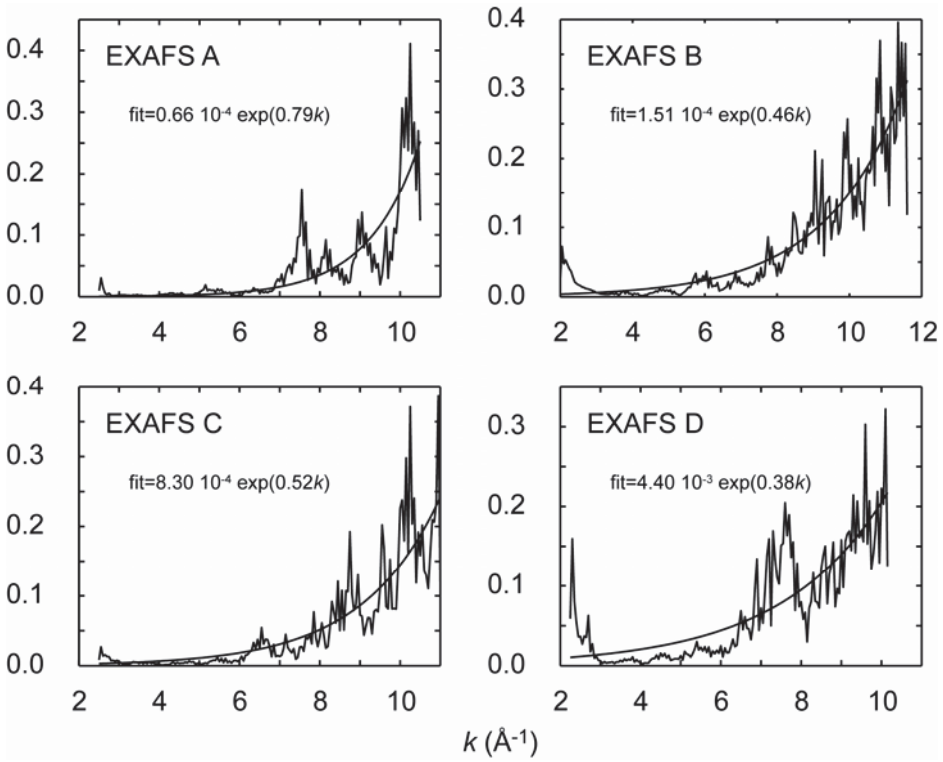


Figure 4.19 : Approximation des écarts-types en fonction de k des résidus des spectres expérimentaux par une fonction du type $f(k) = ae^{bk}$.

Malheureusement, il apparait que la seule définition de cette fonction de fit au travers des paramètres *a* et *b* n'est pas suffisante. En effet, pour certains couples *a* et *b* correspondant aux intervalles déterminés expérimentalement, on obtient des valeurs de f(*k*) = $a\exp^{(bk)}$ largement supérieures à celles mesurées, en particulier à haut *k*. Par exemple, on montre à la figure 4.20 quand *a* est fixé à 0,004 pour des valeurs de *b* supérieures à 0,34, on arrive à de valeurs de 10 à 15 pour la fonction de fit à haut *k* (*k* = 10 Å^{-1}) alors que les valeurs attendues sont de l'ordre de 0,2-0,4. Pour éviter cet artefact, on lie les paramètres *a* et *b* suivant une relation linéaire correspondant à la régression entre les valeurs de *a* et *b* mesurées expérimentalement, *b* = -55,50*a* + 0,61 (Fig. 4.20). Ces paramètres ainsi fixés, la fonction f(*k*) n'atteint plus de valeurs incohérentes et reste de l'ordre de 0,2-0,3 quand *k* = 10 Å^{-1}.

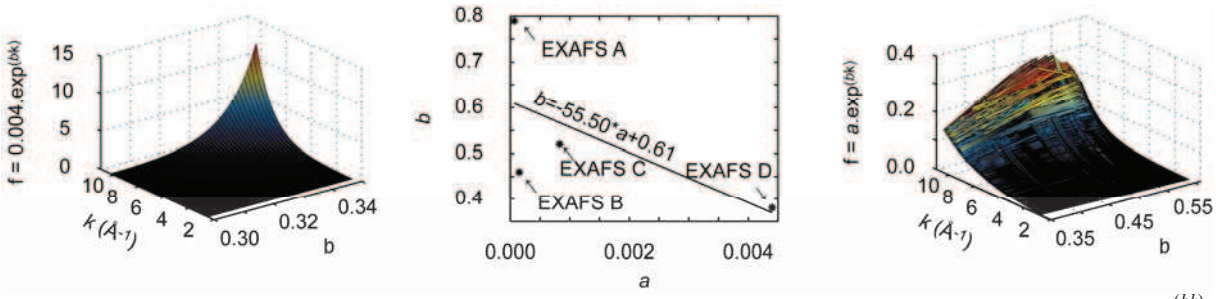


Figure 4.20 : A gauche, illustration du comportement indésirable de la fonction de fit f(*k*) = $a\exp^{(bk)}$, avec *a* = 0,004 et *b* compris entre 0,30 et 0,35. Au centre on lie les paramètres *a* et *b* suivant la relation de régression linéaire entre les deux paramètres. A droite, illustration du comportement de la fonction de fit avec les paramètres a et b liés linéairement,.

Pour finir avec la mesure du bruit sur les spectres EXAFS, on évalue l'écart statistique (i.e. la moyenne et l'écart-type) entre les résidus des jeux de données expérimentaux et leur fonction de fit associée (Fig. 4.21).

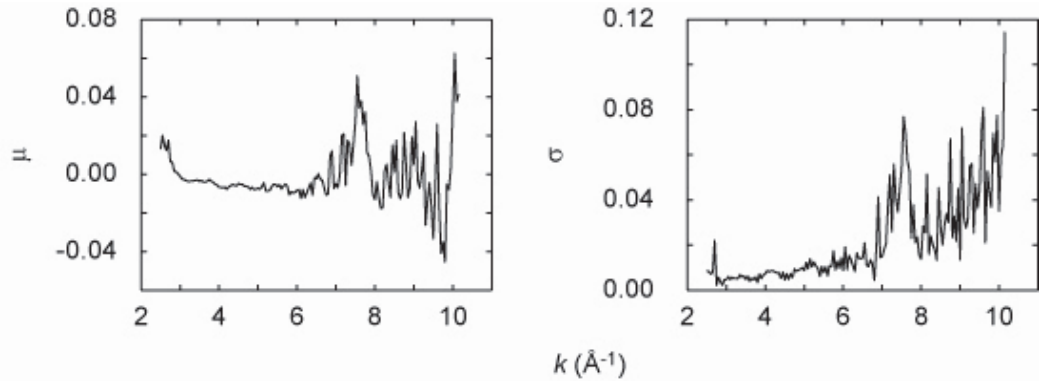


Figure 4.21 : Ecart statistique en fonction de k des écarts-types des résidus résultant des opérations de filtrages vis-à-vis des fonctions optimales f(*k*) = $a\exp^{(bk)}$ associées. A gauche, les moyennes, à droite les écarts-types.

3.3.2) Génération de bruit sur les données synthétiques

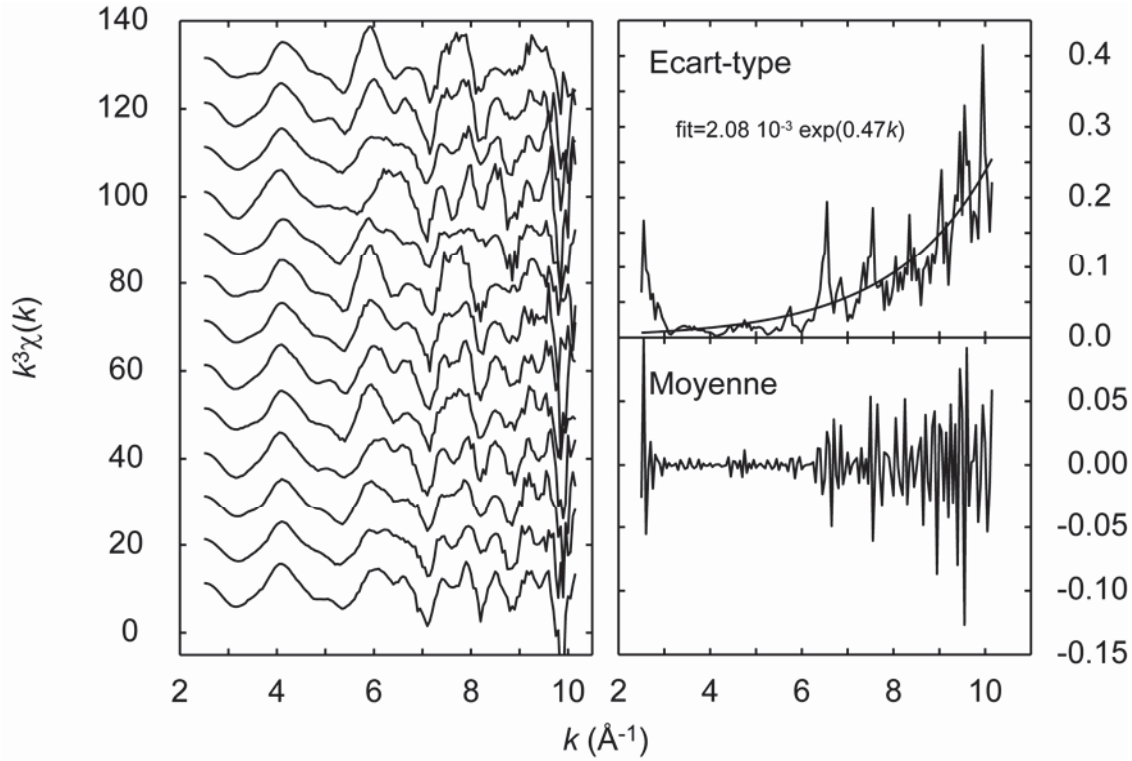


Figure 4.22 : A gauche, résultats de l'opération de bruitage d'une série de mélange de trois spectres EXAFS théoriques et à droite mesure du bruit généré sur ce mélange suivant la méthode définie au § 3.3.1.

Dans un jeu de données EXAFS théoriques, les spectres sont bruités de la manière suivante :

1) pour chaque k, à partir des paramètres $\mu(k)$ et $\sigma(k)$ déterminés à la figure 4.21, génération aléatoire suivant des lois $N(\mu(k),\sigma(k))$ de l'écart entre les résidus des jeux de données théoriques et leur fonction de fit associée ;

2) détermination aléatoire suivant une loi uniforme du paramètre a de la fonction de fit des résidus $f(k) = a \exp^{(bk)}$. L'intervalle de a est déduit des données expérimentales et varie de 0,5 10^{-4} à 5.10^{-3}. Le paramètre b est déduit suivant la relation de la figure 4.20, b = -55,50a + 0,61 ;

3) addition de la fonction de fit et de l'écart, ce signal représentant le bruit moyen sur l'ensemble des spectres du jeu de données ;

4) le bruit de chacun des spectres de la série est alors engendré par une loi normale centrée dont l'écart-type est le bruit moyen de la série multiplié par un facteur 1.5. En effet, pour un k donné, le bruit est moyen, il ne s'agit donc pas d'un écart-type. Les tables statistiques de la loi normale centrée réduite donnent P(X) < 75% pour X ~ 0,67, soit pour une loi normale centrée

non réduite, x/σ ~ 0,67 et d'où σ ~ 1,5*x. L'ensemble est ajouté aux spectres théoriques en $k^3\chi(k)$ normalisés par leur enveloppe.

5) les traces obtenues sont 'dénormalisées ' par les enveloppes de manière à revenir dans l'espace $k^3\chi(k)$.

Les résultats de cette procédure de bruitage et son évaluation sont illustrés à la figure 4.22 sur un jeu de données de 13 spectres résultant du mélange de trois spectres théoriques. Dans cet exemple, les conditions initiales du bruitage déterminées aléatoirement sont a = 0,0023 et b = 0,49. La procédure de calcul du bruit montre que les écarts-types varient de 0 à 0,4 et leur dépendance en k s'approche bien par une fonction du type $f(k) = ae^{(bk)}$ avec a = 0,00208 et b = 0,47. Les moyennes des résidus en fonctions de k sont du même ordre de grandeur que celles des données expérimentales (Figs. 4.12, 4.14, 4.16 et 4.18) justifiant a posteriori l'approximation consistant à ne pas prendre en compte ce paramètre dans la mesure du bruit.

3.4) Conclusion

Les procédures de bruitage sur les différents types de signaux étant définies, il est maintenant possible de tester statistiquement la fonction IND et le test statistique F-Test comme cela a été fait sur la cascade des vp. De plus, ces évaluations nous permettront d'obtenir des indications qualitatives quant à la qualité de nos méthodes de bruitage.

4) Evaluation statistique de la fonction IND et du F-Test

4.1) Principe

On commence par mélanger t fois ($10 \leq t \leq 105$) v spectres XAS théoriques. $2 \leq v \leq 6$ pour l'EXAFS et $3 \leq v \leq 7$ pour le XANES. On obtient alors une première série de spectres sur lesquels on génère du bruit. Cette opération est renouvelée 100 fois. On calcule pour chacun des 100 tirages du couple (v, t) le numéro de la cp pour laquelle la fonction IND est minimale (min(IND). On calcule également le numéro de la première cp à partir de laquelle le niveau de signification du F-test est supérieur à 5% (n[SF>5%]). Les numéros moyens de cp sont alors calculés sur les 100 tirages. Ce principe est décrit graphiquement à la figure 4.23.

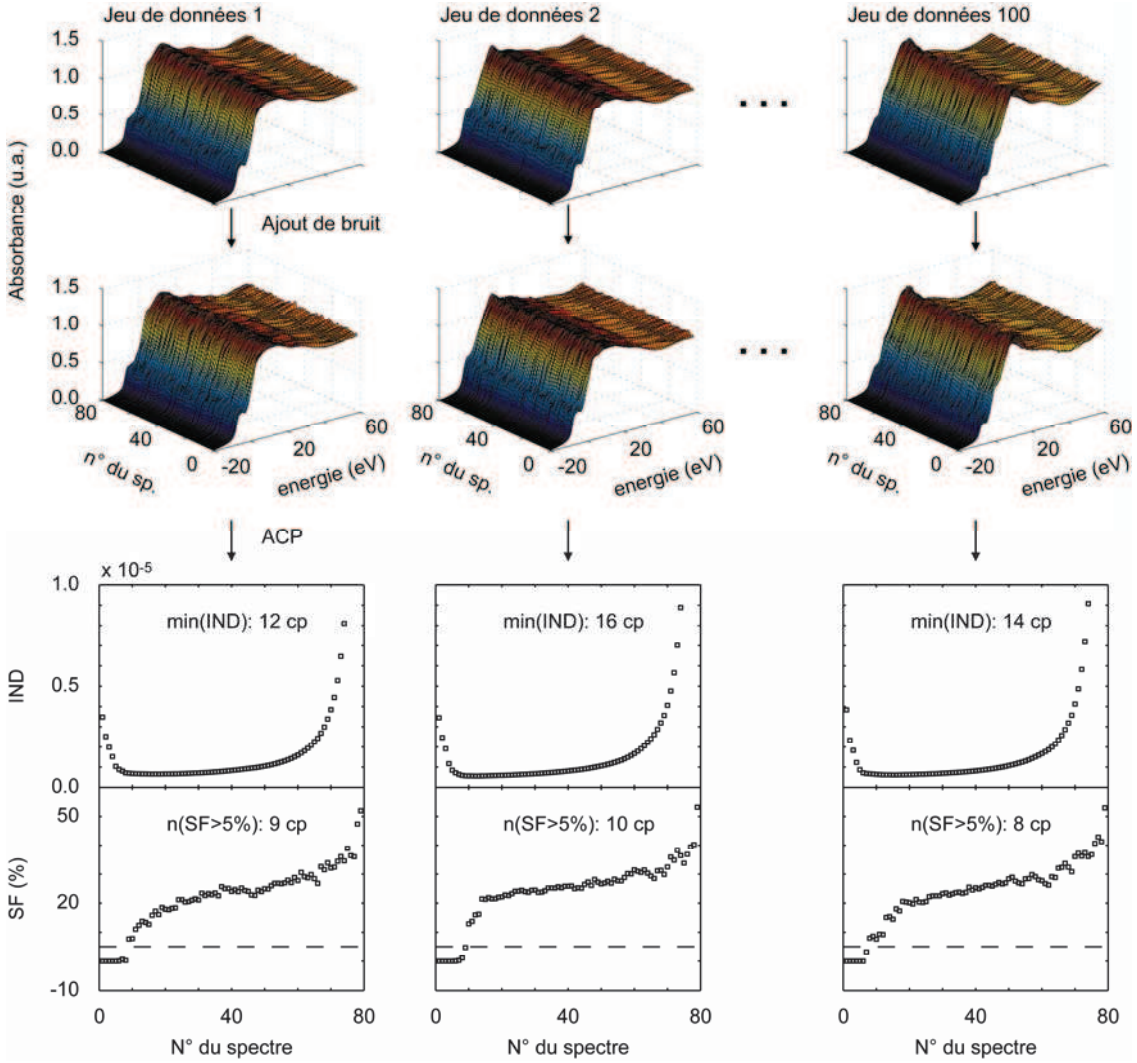


Figure 4.23 : Principe de l'approche statistique pour l'évaluation de la fonction IND et du F-test sur un mélange (5,80) de spectres théoriques XANES 1.

Dans cet exemple, on construit une première série de $t = 80$ spectres multicomposants qui sont un mélange aléatoire de $v = 5$ spectres XANES 1 théoriques. Le numéro de la composante pour laquelle la fonction IND est minimale est 12 et celui pour lequel le niveau de signification du F-Test (SF) est supérieur à 5% est 9. En répétant l'opération 100 fois, on mesure <min(IND)>= 14,34 et σ=3,69, et <n(SF>5%)> = 9,51 et σ = 2,09 pour le couple (5,80).

4.2) Résultats et discussions

4.2.1) XANES 1

4.2.1.1) Evaluation de la fonction IND

-Résultats

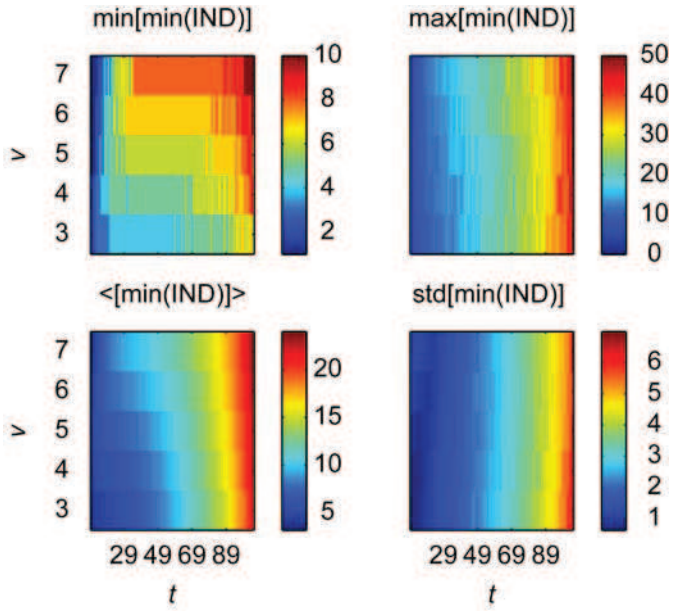


Figure 4.24 : En fonction de v le nombre de composantes dans le système théorique ($3 \leq v \leq 7$) et t la taille du système ($10 \leq t \leq 105$), description statistique (valeurs minimales et maximales, valeurs moyennes et écarts-types associés pour 100 tirages) des valeurs minimales de la fonction IND calculées pour des jeux de données générés à partir des spectres de la série XANES 1.

Les valeurs minimales de l'indicateur min(IND) pour 100 tirages de chacune des conditions $3 \leq v \leq 7$ et $10 \leq t \leq 105$ varient de 1 à 10 (fig. 4 .24). Elles augmentent avec v et t, mais sont assez constantes sur de large intervalle de t. Les valeurs maximales de min(IND) varient de 6 à 50. Elles dépendent de t et dans une moindre mesure de v (essentiellement quand t est inférieur à 30-40). Pour un v donné, la moyenne de l'indicateur min(IND) (<min(IND)>) augmente de 3 à 24 avec t, et cette dépendance est plus rapide lorsque v

augmente. L'écart-type de <min(IND)> est quasiment indépendant de *v*, mais augmente également avec *t*. Il vaut 0,6 pour *t* = 10-20 et 7 pour *t* = 100.

-Discussion

Les valeurs minimales de IND obtenues avec les systèmes théoriques sont proches des valeurs expérimentales des données C3 et B4 de Mamut. Pour C3, pour lequel la taille du jeu de données est de 99, min(IND)=31 ce qui entre parfaitement dans la description statistique du paramètre. En effet, l'approche itérative a montré que le nombre de cp était inférieur ou égale à 8 et l'évaluation statistique du couple (*v* = 7, *t* = 99) donne une valeur moyenne de 20,9 associée à un écart-type de 4,9 et des valeurs minimales et maximales de 9 et 42. Pour le système B4 (*t* = 79), la même observation est faite. La valeur expérimentale de min(IND) est de 22, et l'évaluation statistique du couple (*v* = 7, *t* = 79) donne <[min(IND)]> = 15,1(3,6), min[min(IND)] = 8 et max[min(IND)] = 33.

4.2.1.2) Evaluation du F-test (α = 5%)

- Résultats

Pour l'ensemble des jeux de données (*v*, *t*) obtenus à partir du mélange de spectres XANES 1, les résultats de l'évaluation statistique du F-test avec un seuil de signification supérieur à α = 5% sont synthétisés à la figure 4.25. Les valeurs minimales des 100 tirages sont comprises entre 2 et 9 et augmentent avec *v* et *t*. De la même manière, les valeurs maximales dépendent du couple (*v*, *t*). Elles sont comprises entre 5 et 24. Les valeurs moyennes varient de 2.4 à 12.0 et croissent avec *t* pour toutes les valeurs de *v*. En fonction de *v*, quand *t* est inférieur à 30, elles décroissent puis quand *t* supérieur à 30, elles augmentent. L'écart-type associé est relativement restreint puisque compris entre 0,5 et 2,6. Bien que les comportements de l'écart-type associé en fonction de *v* et en fonction de *t* ne soient pas monotones, on peut écrire qu'il dépend principalement de *t* et croit avec ce paramètre dès qu'il est supérieur à 30 environ.

- Discussion

En comparant les valeurs théoriques avec les valeurs expérimentales, il apparait que ces résultats sont assez cohérents avec les données C3 de Mamut puisque n[SF>5%] = 13. En effet, pour les couples (*v*, 99) avec *v* variant de 6 à 7, les valeurs théoriques varient de 10,7 à 11,9 avec un écart-type associé de l'ordre de 2,3. En revanche, cette remarque est moins vraie

avec les données B4 de Mamut pour lesquelles n[SF>5%] = 18. Pour le couple (7, 79), la valeur moyenne est de 11,2(2,0). La valeur maximale pour ce couple est égale à 19. Ce jeu de données expérimentales donne donc des résultats plus défavorables que le jeu de données théoriques. Pour l'expliquer, on peut émettre l'hypothèse, d'une part que l'erreur systématique non décrite l'approche théorique est plus importante sur le jeu B4 que sur le jeu C3, et d'autre part que le paramètre IND est moins sensible à cette erreur systématique que le F-test (Malinowski 1991).

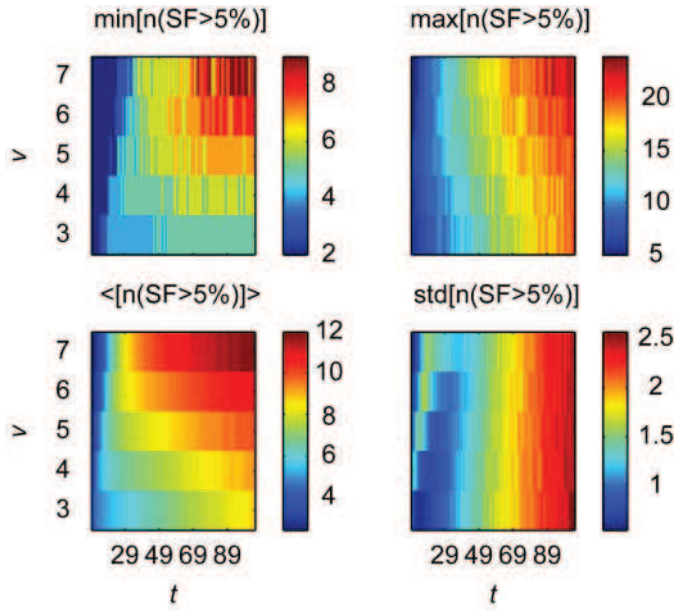


Figure 4.25 : En fonction de *v* le nombre de composantes dans le système théorique ($3 \leq v \leq 7$) et *t* la taille du système ($10 \leq t \leq 105$), distribution statistique des valeurs n[SF>5%] calculées pour des jeux de données générés à partir des spectres XANES 1.

4.2.2) XANES 2

4.2.2.1) Evaluation de la fonction IND

- Résultats

Pour l'ensemble des jeux de données (*v*, *t*) obtenus à partir du mélange de spectres XANES 2, les résultats de l'évaluation statistique de la valeur minimale la fonction IND sont synthétisés à la figure 4.26. Les valeurs min[min(IND)] varient de 1 à 10 et augmentent avec *v* et *t*. La dépendance en t est relative, puisque sur de larges intervalles de t ces valeurs minimales sont constantes. Les valeurs max[min(IND)] varient de 6 à 33 et sont croissantes avec *v* et *t*. <min(IND)> est monotone croissantes en fonction des paramètres *v* et *t*. Elles varient de 3,15 à 16,95. Les écarts-types associés varient de 0,5 à 5,0. Ils dépendent essentiellement du paramètre *t*.

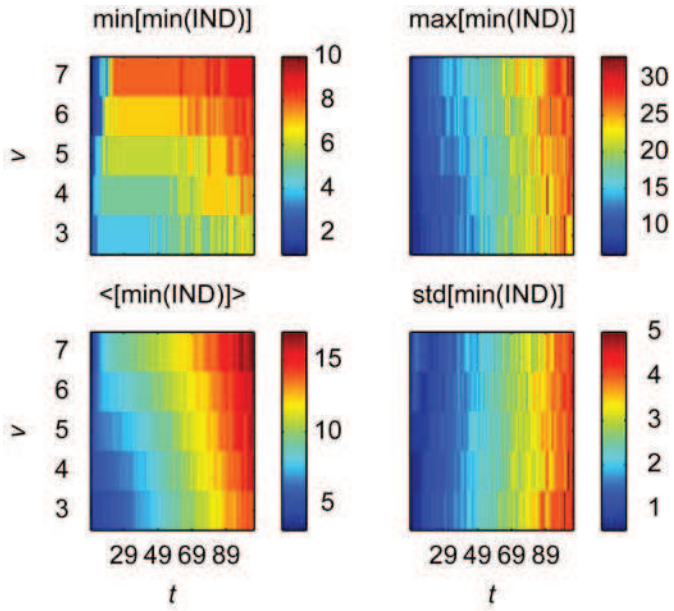


Figure 4.26 : En fonction de *v* le nombre de composantes dans le système théorique ($3 \leq v \leq 7$) et *t* la taille du système ($10 \leq t \leq 105$), distribution statistique des valeurs minimales de la fonction IND calculées pour des jeux de données générés à partir des spectres XANES 2.

- Discussion

Pour les données du site A de DjR ($t = 59$), min(IND) = 15. La valeur théorique de ce paramètre pour le couple (7,59) est 11,2(2,1). L'approche itérative a montré que le nombre maximal de cp était égal à 10. En extrapolant linéairement pour les valeurs de *v* de 8 à 10, et en considérant que l'écart-type associé est indépendant de *v*, il vient les valeurs de 11,6 pour le couple (8,59), 12,0 pour le couple (9,59) et 12,4 pour le couple (10,59). Ces résultats confirment que les valeurs expérimentales maximales entrent assez bien dans la description statistique du paramètre.

Dans le cas de la série de données B de DjR ($t = 114$), min (IND) = 31. La valeur théorique pour ce paramètre pour le couple (7, 105) est égale à 17(4). L'approche itérative a permis de montrer que le nombre de cp dans ce système était au maximum de 11. De la même manière que précédemment, en extrapolant linéairement les écarts-types et en considérant qu'ils sont seulement dépendants de *t*, on obtient pour les couples (*v*, $t = 114$) un écart-type associé de 4,5. En extrapolant les moyennes (extrapolation linéaire dans la dimension de *v* et parabolique dans la dimension de *t*), il vient la valeur 18,1 pour le couple (11, 114). En prenant alors en compte la valeur maximale du couple (7, 105), i.e. 33, et en considérant que cette valeur augmente quand *t* augmente, il apparaît donc que la valeur min(IND) trouvée pour le jeu de données expérimental DjR B, bien qu'élevée, rentre tout de même dans la description statistique.

4.2.2.2) Evaluation du F-test ($\alpha = 5\%$)

- Résultats

Pour l'ensemble des jeux de données (v, t) obtenus à partir du mélange de spectres XANES 2, les résultats de l'évaluation statistique du F-test pour un seuil de signification supérieur à α = 5% sont synthétisés à la figure 4.27. min[min(IND)] varie de 2 à 9. Il augmente avec v et t mais est constant sur de larges intervalles de t. max[min(IND)] varie 4 à 26 et croît avec v et avec t. <min(IND)> varie de 2,56 à 12,45 et croît en fonction de t et en fonction de v quand t est supérieur à 30 environ. Les écarts-types varient de 0.5 à 2.9. Ils croissent avec t mais sont relativement indépendants de v, dès que t est suffisamment grand i.e. de l'ordre de 30.

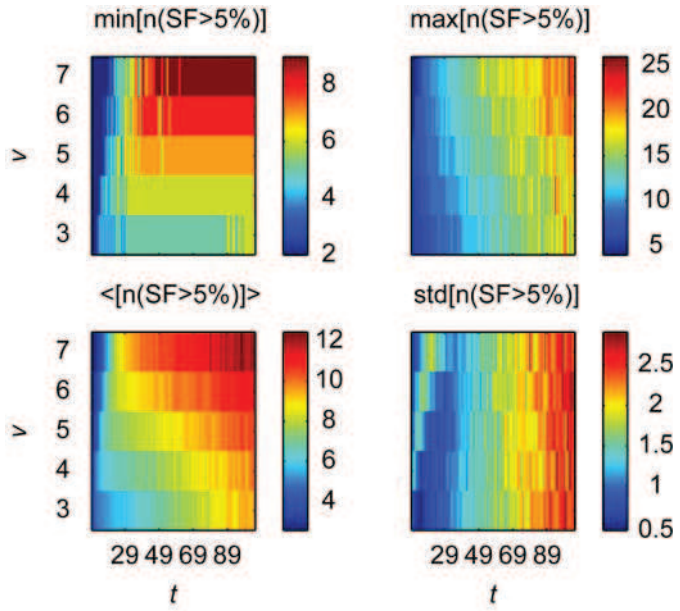


Figure 4.27 : En fonction de v le nombre de composantes dans le système théorique ($3 \leq v \leq 7$) et t la taille du système ($10 \leq t \leq 105$), distribution statistique des valeurs n[SF>5%] calculées pour des jeux de données générés à partir des spectres XANES 2.

- Discussion

Expérimentalement, pour les données de DjR A, n[SF>5%] = 9. En comparant avec les résultats théoriques, pour le couple (7, 59) il vient la valeur moyenne de 10,7(1,5). En extrapolant linéairement, il vient pour le couple (10, 59) la valeur de 12,2. En considérant que l'écart-type ne dépend que de t, on peut associer une valeur de 1.7 à la valeur moyenne. Les valeurs minimales théoriques de ce test pour des valeurs de t de l'ordre de 60 varient entre 7 et 9. Les résultats expérimentaux de DjR A sont donc légèrement inférieurs aux résultats théoriques. Pour les données DjR B, n[SF>5%] = 16, et la valeur moyenne théorique du couple (7, 105) est de 12,3(2,6). En extrapolant linéairement, il vient pour le couple (7, 114)

la valeur de 12,5 associée à un écart-type de l'ordre de 2,5 à 3. Comme précédemment avec IND, on peut alors estimer la valeur moyenne théorique du couple (10, 114) à 14,5, l'écart-type restant inchangé. Cette fois, les résultats expérimentaux entrent bien dans la description statistique.

4.2.3) Conclusions sur le XANES

D'une manière générale, on peut affirmer que les résultats statistiques obtenus sur les systèmes de données théoriques sont cohérents avec les valeurs maximales des nombres d'espèces chimiques dans les systèmes expérimentaux.

Quantitativement, si les valeurs sont les mêmes dans les deux gammes d'énergie avec le F-Test, les valeurs du paramètre min(IND) sont moins importantes dans la gamme d'énergie la plus large. Qualitativement les comportements des deux paramètres testés (IND et F-test) sont indépendants de l'intervalle en énergie du système théorique.

Enfin, il apparait clairement que ces tests ne sont pas pertinents pour évaluer le nombre réel de composantes principales d'un ensemble de spectres XANES, et ceci indépendamment de leur étendue en énergie avant et après seuil. Tout d'abord les valeurs moyennes ne correspondent pas aux valeurs attendues, ensuite les écarts-types associés sont tellement élevés qu'il n'est pas possible de restreindre le champ des possibilités du paramètre *v*.

4.2.4) EXAFS 1

4.2.4.1) Evaluation de la fonction IND

- Résultats

Pour l'ensemble des jeux de données (*v*, *t*) obtenus à partir du mélange de spectres EXAFS 1, la description statistique de la valeur minimale de la fonction IND est synthétisée à la figure 4.28. max[min(IND)]varie de 2 à 6, augmente avec *v* quand *t* est fixé, elle est constante sur des plages de *t* à *v* fixé. min[min(IND)] est compris entre 3 et 36, croît avec *t* et tend à décroitre avec *v*. <min(IND)> varie entre 2,0 et 13,8 et augmente avec *t* pour une valeur de *v* fixée. Quand *t* est fixé et inférieur à 90 environ, il augmente avec *v*. En revanche cette tendance s'inverse quand $t > 90$. Les valeurs des écarts-types varient de 0,1 à 6,0. Elles augmentent avec *t* quand *v* est fixé et diminuent avec *v* pour un *t* fixé.

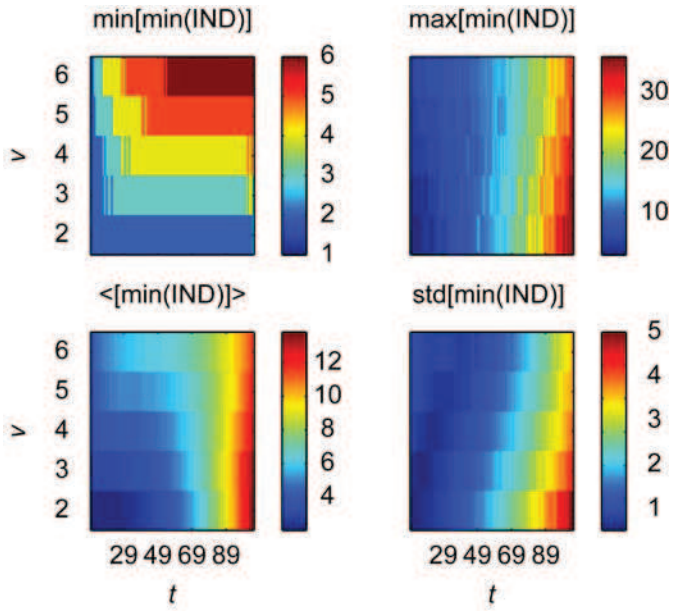


Figure 4.28 : En fonction de v le nombre de composantes dans le système théorique ($2 \leq v \leq 6$) et t la taille du système ($10 \leq t \leq 105$), description statistique des valeurs minimales de la fonction IND calculées pour des jeux de données générés à partir des spectres EXAFS 1.

- Discussion

Les résultats théoriques et expérimentaux dans l'espace $\chi(k)$ de ce paramètre, sont comparés au tableau 4.2.

Tableau 4.2 : Comparaison entre les résultats théoriques et expérimentaux du calcul de min(IND) – espace χk)

Résultats expérimentaux				*Résultats théoriques*				
	v	t	min(IND)	(v, t)	min.	max.	moy.	Std.
EXAFS A	3	29	6	(3,29)	3	7	3.2	0.4
EXAFS B	3	27	7	(3,27)	3	6	3.1	0.4
EXAFS C	3	13	3	(3,13)	2	4	3.0	0.1
EXAFS D	5	31	8	(5,31)	4	8	5.0	0.3

La comparaison montre que, hormis dans le cas où t est faible (jeu de données EXAFS C), les valeurs expérimentales sont plus élevées que les valeurs théoriques moyennes. Elles sont à la limite supérieure de la validité statistique i.e. égales (voire supérieures) à la valeur maximale obtenue statistiquement. Deux explications peuvent être proposées : (1) comme dans le cas du jeu de données XANES DjR B, il peut simplement s'agir soit d'une coïncidence, (2) le mode de bruitage des données dans l'espace $\chi(k)$ n'est pas adapté.

4.2.4.2) Evaluation du F-Test ($\alpha = 5\%$)

- Résultats

Pour l'ensemble des jeux de données (v, t) obtenus à partir du mélange de spectres EXAFS 1, les résultats de l'évaluation statistique du F-test pour un seuil de signification supérieur à $\alpha = 5\%$ sont synthétisés à la figure 4.29. Les valeurs minimales varient de 2 à 7. Elles augmentent en fonction de t mais sont constantes sur de larges plages de ce paramètre. Elles sont monotones croissantes en fonction de v dès que $t > 20$. Les valeurs maximales augmentent avec t et v et sont comprises entre 3 et 14. Les valeurs moyennes sont comprises entre 3 et 7, croissent avec v et sont relativement indépendantes de t. Les écarts-types varient entre 0,0 et 1,3 et augmentent avec v et avec t.

- Discussion

Les résultats expérimentaux et théoriques sont détaillés au tableau 4.3. Toujours hormis le cas où t est petit, cette comparaison montre que les résultats expérimentaux sont toujours supérieurs aux résultats théoriques moyens et à la limite des moyennes maximales statistiques obtenues. Il est peu probable que sur 2 tests différents, 3 jeux de données soient tous à la limite supérieure des descriptions statistiques. Cela signifie donc que le mode de bruitage qui consiste à bruiter les spectres dans l'espace $k^3\chi(k)$ des spectres puis à revenir dans l'espace $\chi(k)$ ne simule pas la réalité.

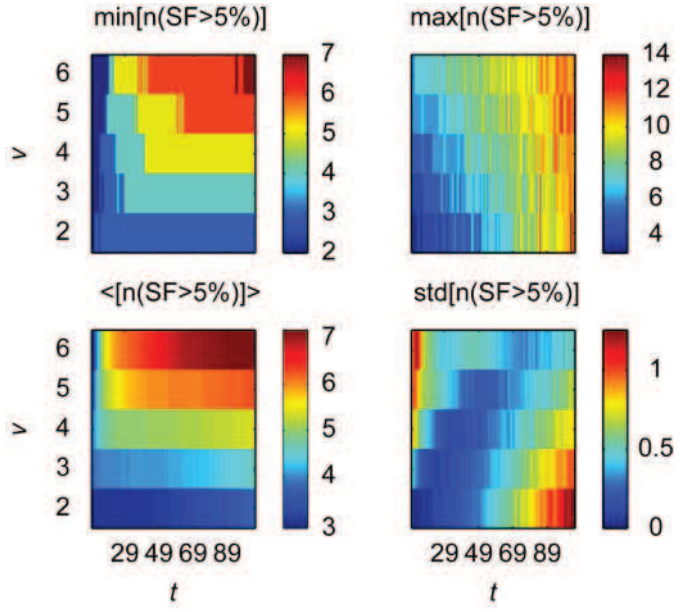


Figure 4.29 : En fonction de v le nombre de composantes dans le système théorique ($2 \leq v \leq 6$) et t la taille du système ($10 \leq t \leq 105$), description statistique des valeurs pour lesquelles n[SF>5%] calculées pour des jeux de données générés à partir des spectres EXAFS 1.

Tableau 4.3 : Comparaison entre les résultats théoriques et expérimentaux du calcul de n(SF>5%) – espace χk)

Résultats expérimentaux				*Résultats théoriques*				
	v	*t*	n(SF>5%)	(*v*, *t*)	min.	max.	moy.	Std.
EXAFS A	3	29	6	(3,29)	4	5	4	0.1
EXAFS B	3	27	6	(3,27)	3	6	4	0.1
EXAFS C	3	13	4	(3,13)	2	4	3.9	0.3
EXAFS D	5	31	8	(5,31)	4	7	5.8	0.5

4.2.5) EXAFS 2

4.2.5.1) Evaluation de la fonction IND

-Résultats

Pour l'ensemble des jeux de données (*v*, *t*) obtenus à partir du mélange de spectres EXAFS 2, les résultats de l'évaluation statistique de la valeur minimale la fonction IND sont synthétisés à la figure 4.30. min[min(IND)] varie de 2 à 20 et croît avec *v* et *t*. max[min(IND)] est compris entre 4 et 81, croît avec *t* et tend à décroitre avec *v*. <min(IND)> varie de 1,8 à 41,1, croît avec *t* et est relativement indépendant de *v*. Les écarts-types sont compris entre 0,5 et 13,3. Ils croissent avec *t* et décroissent avec *v*.

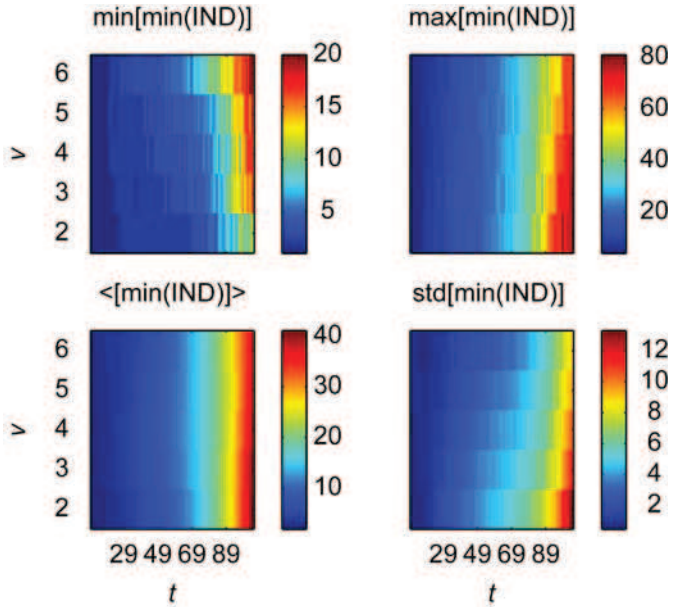


Figure 4.30 : En fonction de *v* le nombre de composantes dans le système théorique ($2 \leq v \leq 6$) et *t* la taille du système ($10 \leq t \leq 105$), description statistique des valeurs minimales de la fonction IND calculées pour des jeux de données générés à partir des spectres EXAFS 2.

- Discussion

La synthèse de la comparaison entre les résultats expérimentaux et théoriques est présentée tableau 4.4. Dans tous les cas, les résultats expérimentaux sont cohérents avec la description statistique des données théoriques.

Tableau 4.4 : Comparaison entre les résultats théoriques et expérimentaux du calcul de min(IND) – espace $k^3\chi k$)

Résultats expérimentaux				*Résultats théoriques*				
	v	*t*	min(IND)	(*v*, *t*)	min.	max.	moy.	Std.
EXAFS A	3	29	3	(3,29)	2	10	4.1	1.6
EXAFS B	3	27	3	(3,27)	2	11	3.7	1.4
EXAFS C	3	13	3	(3,13)	2	6	2.3	0.6
EXAFS D	5	31	6	(5,31)	2	12	4.2	1.3

4.2.5.2) Evaluation du F-Test ($\alpha = 5\%$)

- Résultats

Pour l'ensemble des jeux de données (*v*, *t*) obtenus à partir du mélange de spectres EXAFS 2, les résultats de l'évaluation statistique du F-test pour un seuil de signification supérieur à $\alpha = 5\%$ sont synthétisés à la figure 4.31. Les valeurs minimales varient de 2 à 5. Elles dépendent de (*v*, *t*) et tendent à être monotones croissantes avec ces deux paramètres. Le comportement des valeurs maximales est le suivant : monotone croissant avec *t*, et symétrique autour de *v* = 4. Elles varient de 4 à 26. Le comportement des valeurs moyennes est similaire, elles varient de 2,2 à 9,1. Enfin les écarts-types associés sont monotones croissants avec v et t et sont compris entre 0,4 et 4,5.

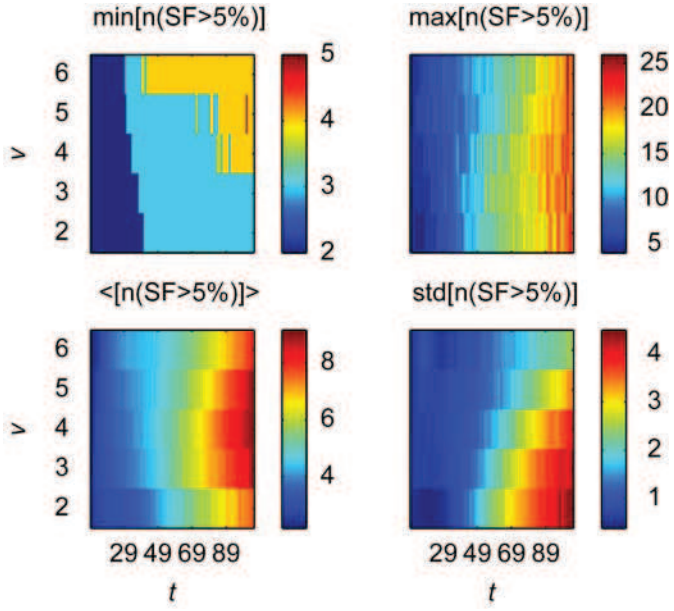


Figure 4.31 : En fonction de *v* le nombre de composantes dans le système théorique ($2 \leq v \leq 6$) et *t* la taille du système ($10 \leq t \leq 105$), description statistique des valeurs pour lesquelles n[SF>5%] calculées pour des jeux de données générés à partir des spectres EXAFS 2.

- Discussion

La synthèse de la comparaison entre les résultats expérimentaux et théoriques est présentée tableau 4.5. Dans tous les cas, les résultats expérimentaux sur le paramètre n(SF>5%) sont cohérents avec la description statistique des données théoriques.

Tableau 4.5 : Comparaison entre les résultats théoriques et expérimentaux du calcul de n(SF>5%) – espace $k^3\chi k$)

Résultats expérimentaux				*Résultats théoriques*				
	v	*t*	n(SF>5%)	(*v*, *t*)	min.	max.	moy.	Std.
EXAFS A	3	29	4	(3,29)	2	7	3.7	0.7
EXAFS B	3	27	4	(3,27)	2	7	3.6	0.7
EXAFS C	3	13	3	(3,13)	2	5	2.8	0.6
EXAFS D	5	31	4	(5,31)	3	7	4.1	0.7

4.2.6) Conclusions sur l'EXAFS

La comparaison entre les résultats expérimentaux et statistiques des paramètres min(IND) et de n(SF>5%) a permis de montrer que le mode de bruitage proposé n'était pas adapté pour simuler le bruit sur des spectres représentés dans l'espace $\chi(k)$. En revanche les résultats obtenus dans l'espace $k^3\chi(k)$ sont cohérents et laissent à penser que la méthode est bien adaptée à cette représentation des données.

Enfin, comme dans le cas du XANES, pour des mélanges dont on ignore le nombre réel de cp, ces deux tests présentent un écart type trop important et ne permettent malheureusement pas d'évaluer la dimension de l'espace principal.

4.3) Conclusion générale

En conclusion de ces calculs il apparaît que les paramètres min(IND) et n(SF>5%), habituellement utilisés avec l'ACP, ne sont pas adaptés pour déterminer la dimension de l'espace principal associé à des mélanges de spectres XAS, en raison principalement de leur trop grande sensibilité au bruit qui a pour effet d'étaler les écarts-types. L'objet du prochain paragraphe est de proposer un nouvel indicateur dont la valeur sera moins sensible au bruit.

5) Proposition et évaluation d'un nouveau critère ACP : l'indicateur NSS

5.1) Introduction

L'objet de cette dernière partie est de proposer un nouveau critère de détermination du nombre de composantes primaires dans l'ACP de spectres XAS. Pour commencer, la construction de cet indicateur sera détaillée sur la série de spectres Mamut C3. Ensuite, dans une seconde partie, il sera appliqué et discuté sur les données expérimentales. Enfin, il sera évalué statistiquement de la même manière que les deux autres indicateurs ACP. Cette évaluation est justifiée. Il vient d'être montré que les systèmes théoriques étaient de bons modèles des systèmes expérimentaux. Elle sera faite sur les deux séries de XANES théoriques et sur des données EXAFS théoriques définies dans l'espace $k^3\chi(k)$.

5.2) Construction de l'indicateur

La genèse du principe de l'indicateur NSS se trouve dans les calculs faits pour la détermination du nombre de cp dans les courbes de titrage protonique (Lenoir et al. 2010). La première étape consiste à dé-bruiter les données expérimentales à l'aide d'un filtre adapté puis de calculer la somme des carrés normalisée (NSS) entre les données expérimentales et les données filtrées.

$$NSS = \sum (y_i - y_i^{fit})^2 / \sum y_i^2$$

On obtient pour chaque donnée une valeur NSS_{fv}. Parallèlement, en appliquant l'ACP au jeu de données filtrées, on peut mesurer, pour chaque spectre de ce jeu de données, la distance NSS_{jcp} avec sa reconstruction par ACP avec j composantes. A chaque ajout d'une nouvelle composante, la reconstruction s'améliore, les distances entre les données et leurs reconstructions diminuent et il vient $NSS_{1cp}>NSS_{2cp}>NSS_{3cp}\ldots$

En faisant ensuite, spectre par spectre, le rapport entre NSS_{fv} et NSS_{jcp} (j variant de 1 à la taille du jeu de donnée), on obtient une nouvelle valeur OG qui dépend de la donnée (n° de la donnée dans le jeu) et du nombre *j* de composantes utilisées dans la reconstruction ACP. En calculant pour l'ensemble du jeu de données les valeurs moyennes et les écarts-types associés en fonction du nombre *j*, nous avons remarqué l'apparition d'un saut dans les valeurs des écarts-types. Ces étapes sont illustrées à la figure 4.32 avec le jeu de données expérimentales Mamut C3. Un saut apparait entre la $5^{ème}$ et la $6^{ème}$ cp. En travaillant avec une échelle

logarithmique, et en mesurant Δlog[σ(OG)], i.e. les différences une à une des valeurs de log[σ(OG)], il est alors possible de faire apparaitre un pic à la 6ème cp.

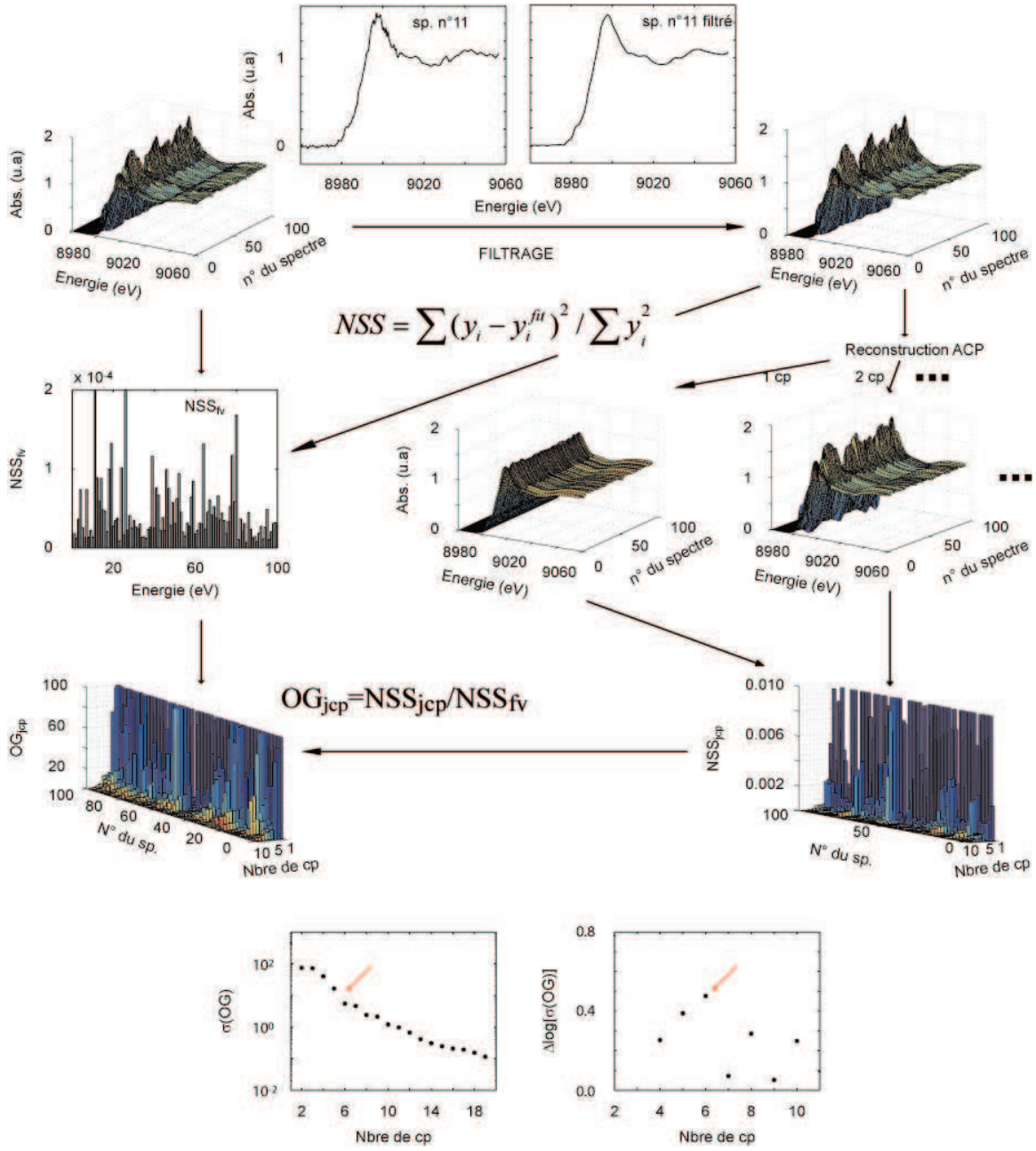


Figure 4.32 : Principe du calcul de l'indicateur NSS appliqué au jeu de données Mamut C3.

5.3) Application aux données expérimentales

Ce saut est apparu pour l'ensemble des données XAS expérimentales : les spectres XANES Mamut B4, DjR A et DjR B (Fig. 4.33) et les spectres EXAFS (Fig. 4.34).

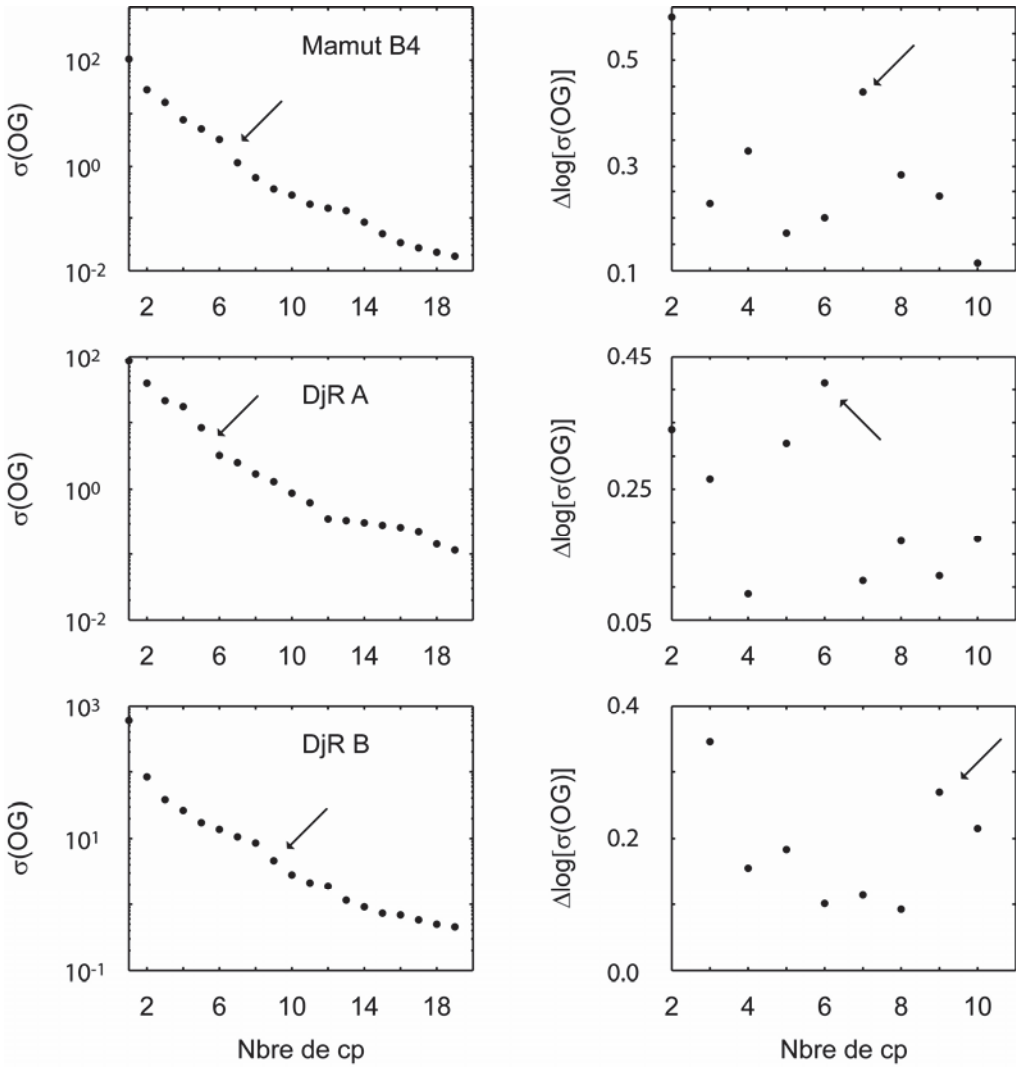


Figure 4.33 : Résultats de l'indicateur NSS sur l'ensemble des données expérimentales XANES étudiées. A gauche représentation de σ(OG) en fonction du nombre de cp retenues dans l'espace principal, à droite représentation Δlog[σ(OG)] en fonction du nombre de cp.

5.3.1) XANES

σ(OG) décroît en général fortement entre les reconstructions à 1 cp et 2 cp car la première composante explique environ 99% du signal. Ensuite, on peut observer des cassures dans cette décroissance. Dans le cas de Mamut C3, les trois premières apparaissent entre 5 et 6 cp, 7 et 8 cp et 9 et 10 cp. La différence est la plus importante pour les valeurs de log(σ) entre 5 et 6 cp. Dans le cas de Mamut B4, une cassure apparaît entre la 6ème et la 7ème cp, puis une seconde entre la 13ème et la 14ème cp. Encore une fois, la différence entre les valeurs de log(σ) est plus importante dans la première cassure. Sur le jeu de données XANES de DjR A, trois cassures apparaissent, la plus importante entre la 5ème et la 6ème cp, la seconde entre la 12ème et la 13ème cp et la dernière entre la 17ème et la 18ème cp. Enfin, 2 cassures apparaissent

sur le jeu de données DjR B. L'une entre le 8ème et la 9ème cp et l'autre entre la 12ème et 13ème cp avec toujours la différence sur log(σ) plus importante dans le premier cas. Il faut néanmoins noter que si l'on considère seulement les différences de log(σ), pour ce jeu de données expérimentales, elle est la plus importante entre 2 et 3 cp, sans qu'il n'apparaisse de cassure.

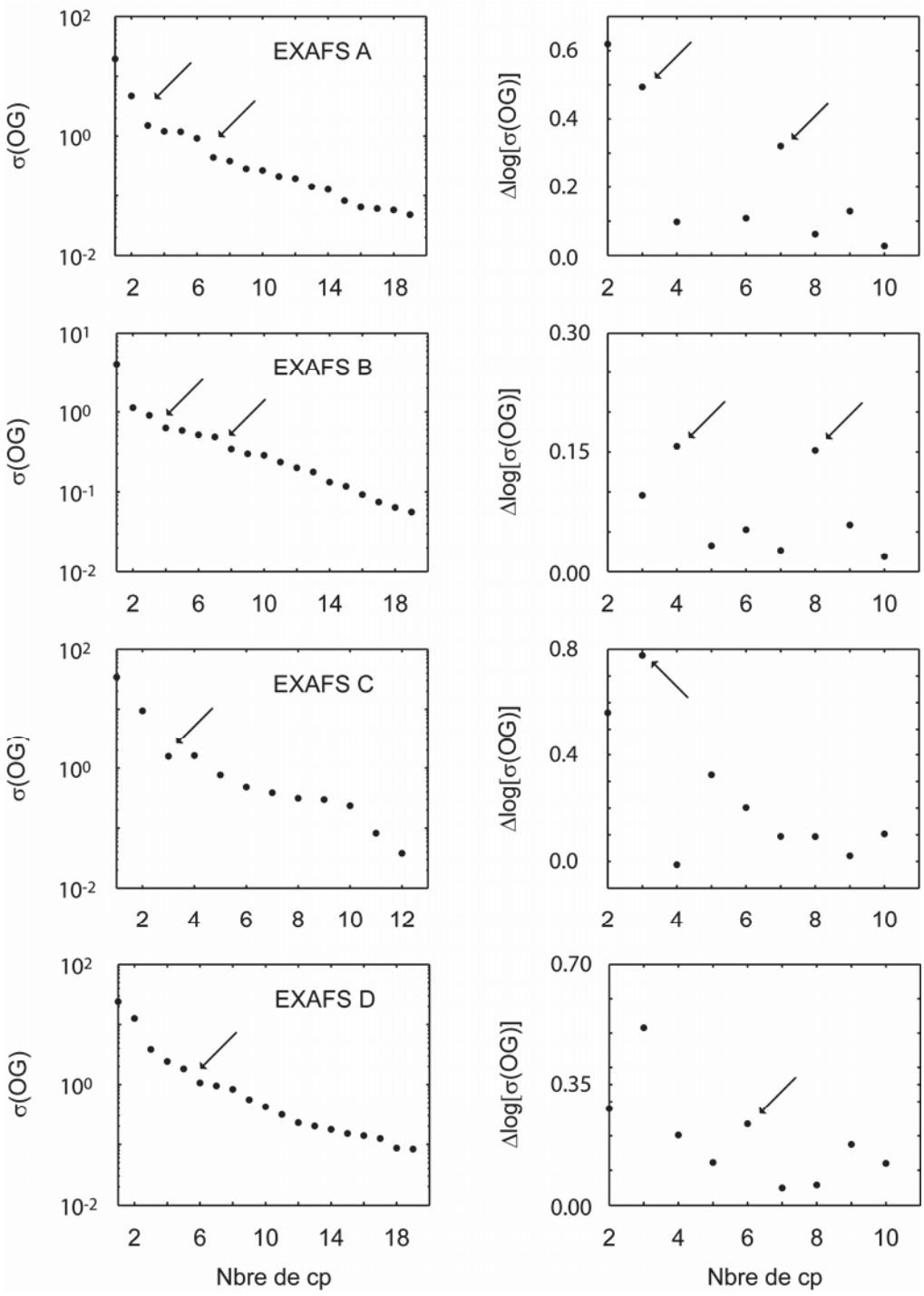


Figure 4.34 : Résultats de l'indicateur NSS sur l'ensemble des données expérimentales EXAFS étudiées. A gauche représentation de σ(OG) en fonction du nombre de cp retenues dans l'espace principal, à droite représentation Δlog[σ(OG)] en fonction du nombre de cp.

5.3.2) EXAFS

Dans le cas du jeu de données EXAFS A, une première cassure apparait entre la 2nde et la 3ème cp, puis une seconde entre la 6ème et la 7ème cp. Dans le jeu de données EXAFS B, une première cassure apparaît entre la cp3 et la cp 4, puis il y a une seconde cassure entre la cp 7 et la cp 8. Dans ces deux cas, comme dans le cas des données XANES, c'est la première cassure qui produit la différence la plus importante entre les valeurs de log(σ). Dans le jeu de données EXAFS C, on observe une cassure nette et unique entre la 2ème cp et la 3ème cp. Cette cassure est d'autant plus remarquable que le caractère monotone décroissant de σ(OG) observé jusque là n'est plus vérifié et que $\sigma(OG_{3cp}) < \sigma(OG_{4cp})$. Enfin dans le jeu de données EXAFS D on observe la cassure entre la 5ème et la 6ème cp. Néanmoins, et comme dans le cas de Django B, la différence la plus importante entre $\sigma(OG_{jcp})$ et $\sigma(OG_{j+1cp})$ n'a pas lieu pour *j* égal 5 mais pour $j = 2$.

5.3.3) Discussion

A partir des estimations faites au chapitre 3, l'étude des données expérimentales XANES montre que les valeurs des cp pour lesquelles ces cassures ont lieu sont cohérentes avec les nombres d'espèces pures dans les différents systèmes.

L'étude des données expérimentales EXAFS semble confirmer cette cohérence, puisque les valeurs obtenues concordent dans 2 cas sur 4 (jeu de données EXAFS A et C) et diffèrent de une unité pour les deux autres cas (jeu de données EXAFS B et D), vis-à-vis du nombre d'espèces pures déterminées dans les études d'où proviennent ces données (Panfili et al. 2005; Isaure et al. 2002; Manceau et Matynia 2010; Van Damme et al. 2010).

A ce stade, il convient donc d'évaluer statistiquement si ces cassures sont bien en rapport avec le nombre de spectres individuels, ou 'espèces', contenus dans le système.

5.4) Evaluation statistique

A la différence des indicateurs min(IND) et n(SF>5%), il est difficile d'évaluer statistiquement une cassure :

- quelle tolérance prendre sur la notion de pente ?

- à partir de quel saut peut on considérer qu'il y a ou pas une rupture de pente ?

Comme cela a été remarqué, le delta dans le log(σ) qui accompagne la rupture de pente est en général (mais ce n'est pas toujours le cas) plus important que les autres. Nous

avons donc choisi de tester statistiquement la valeur j+1 de la cp pour laquelle la différence [log(σ(OG_{jcp}))-log(σ(OG_{j+1cp}))] est maximale. Cette valeur j+1 a été testée statistiquement de la même manière que les autres indicateurs ACP, à savoir par des mélanges de spectres XANES et EXAFS théoriques, ces derniers uniquement dans l'espace $k^3\chi(k)$ puisque les tests précédents ont démontré que la procédure de bruitage était inadaptée dans l'espace $\chi(k)$.

5.4.1) XANES

5.4.1.1) XANES 1

- Résultats

Pour des données théoriques dont la gamme en énergie varie de -20 à +60 eV relativement au seuil, les résultats statistiques de l'évaluation de l'indicateur NSS en fonction du couple (v, t), avec $3 < v < 7$ et $10 < t < 105$, sont donnés à la figure 4.35.

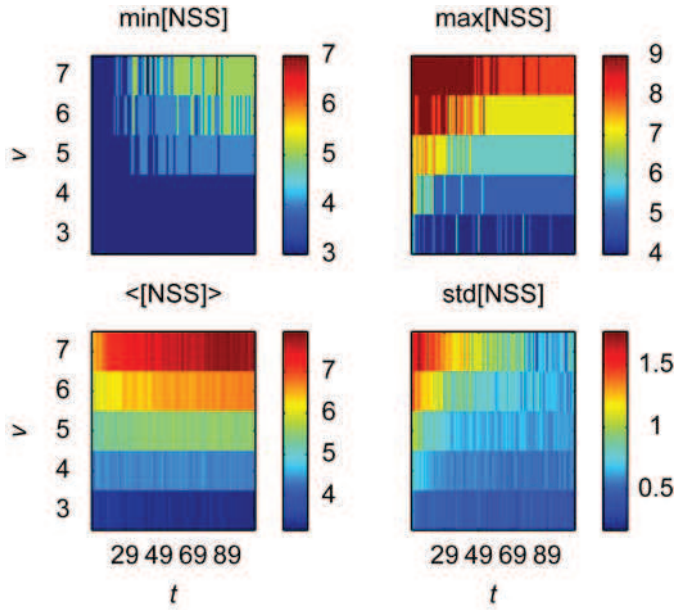


Figure 4.35 : Valeurs statistiques de l'indicateur NSS pour des jeux de données issus du mélange de spectres théorique XANES 1.

Les valeurs minimales varient de 3 à 7 et croissent avec v et t. Les valeurs maximales sont comprises entre 4 et 9. Elles augmentent avec v mais diminuent avec t. Les valeurs moyennes varient entre 3.2 et 8.0. Elles tendent à ne dépendre que de v et croissent avec ce paramètre. Quand v=3, elles diminuent légèrement avec t de 3.5 à 3.2. Quand v = 4, elles sont indépendantes de t et varient autour de 4,4. Pour v = 5, elles croissent de 5,1 à 5,5 avec t dans l'intervalle $10<t<30$ puis se stabilisent autour de 5,5 au delà. Pour v = 6, elles augmentent de 6,0 à 6,7 pour t compris entre 10 et 70 puis se stabilisent autour de 6,7. Enfin, quand v = 7,

elles croissent de 6,5 à 7,7 pour t inférieur à 80 puis se stabilise autour de 7,7. En moyenne, et pour des hautes valeurs de t, les résultats théoriques tendent à être supérieurs de moins de une unité au nombre réel de composantes dans le système. Les valeurs des écarts-types associés dépendent du couple (v, t). Elles diminuent avec t quand v est fixé et augmentent avec v pour un t fixé. Elles varient de 0,2 à 1,8, les écarts les plus importants se retrouvant pour des hautes valeurs de v et des faibles valeurs de t.

En conclusion de ces calculs, nous avons démontré que pour des données XANES théoriques dont la gamme en énergie est comprise entre -20 et +60 eV relativement au seuil, l'indicateur NSS est un paramètre efficace pour l'évaluation du nombre de composantes principales dans les systèmes. Le résultat moyen tend à être supérieur d'une unité au nombre réel de cp et permet de proposer au minimum 5 cp dans le système Mamut C3 et 6 cp dans le système Mamut B4.

- Discussion

Pour évaluer la pertinence de ces résultats, on compare dans le tableau 4.6 les résultats expérimentaux de l'indicateur NSS avec les statistiques théoriques des indicateurs min(IND) et n(SF>5%). Bien que, pour Mamut C3, la valeur <min(IND)> soit élevée au regard de la moyenne théorique (31 vs 19,8 ± 5,1) comme la valeur <n(SF>5%)> pour le site de Mamut B4 (18 vs 10,4 ± 2,1), dans tous les cas, compte-tenu de l'incertitude associée aux résultats moyens théoriques du NSS, les résultats expérimentaux entrent dans la description statistique théorique.

Tableau 4.6 : Comparaison entre les résultats NSS expérimentaux des sites Mamut C3 et B4 avec les résultats statistiques théoriques min(IND) et n(SF>5%)

Résultats expérimentaux	*Résultats théoriques*			
Mamut C3	(v, t)=(5,99)	min(IND)	n(SF>5%)	NSS
t = 99	min	8	6	4
NSS = 6 → v = 5	max	41	20	6
min(IND) = 31	moy.	19.8	9.9	5.5
n(SF>5%) = 13	std.	5.1	2.4	0.6
Mamut B4	(v, t)=(6,79)	min(IND)	n(SF>5%)	NSS
t = 79	min	7	8	4
NSS = 7 → v = 6	max	27	19	7
min(IND) = 22	moy.	14.6	10.4	6.7
n(SF>5%) = 18	std.	3.6	2.1	0.7

5.4.1.2) XANES 2

- Résultats

Pour des données théoriques dont la gamme en énergie varie de -40 à +150 eV relativement au seuil, les résultats statistiques de l'évaluation de l'indicateur NSS en fonction du couple (v, t), avec $3 < v < 7$ et $10 < t < 105$, sont donnés à la figure 4.36.

Les valeurs minimales varient de 3 à 8 et croissent avec v et t. Pour des valeurs de v de 6 et 7, on remarque le comportement asymptotique en fonction de t vers 7 ($v = 6$) et 8 ($v = 7$). Les valeurs max comprises entre 9 et 4, croissent avec v et décroissent avec t. Pour des valeurs de v de 5, 6 et 7, on remarque le comportement asymptotique en fonction de t vers la valeur de 6 ($v = 5$), de 7 ($v = 6$) et de 8 ($v = 7$). En moyenne, l'indicateur NSS est relativement stable vis-à-vis du paramètre t et tend à ne dépendre que de v. Quand $v = 3$, pour des valeurs de t de 10 à 50, il décroît de 3,6 à 3,3 puis oscille autour de cette valeur indépendamment de t. Quand $v = 4$, il est indépendant de t et varie autour de 4,6. Pour $v = 5$, pour t compris entre 10 et 40 il croît de 5,5 à 5,9 puis oscille autour de cette valeur. Pour $v = 6$, pour t compris entre 10 et 50 il croît de 6,2 à 7,0 puis oscille autour de cette valeur. Enfin, quand $v = 7$, pour t compris entre 10 et 40 il croit de 7 à 8 puis est strictement égal à 8 pour les valeurs de t supérieures. Les valeurs des écarts-types associés varient de 0,0 à 1,9, les écarts les plus importants se retrouvant pour des hautes valeurs de v et des faibles valeurs de t. Elles diminuent avec t quand v est fixé mais le comportement à t fixé n'est pas monotone croissant. Quand t est suffisamment grand (de l'ordre de 50) alors std(NSS) décroît en fonction de v pour atteindre la valeur remarquable de 0 quand $v = 8$.

Comme précédemment, ces calculs démontrent que pour des données XANES théoriques dont la gamme en énergie est comprise entre -40 et +150 eV relativement au seuil, l'indicateur NSS est un paramètre rigoureux pour l'évaluation du nombre de composantes principales dans les systèmes avec un résultat qui tend à être supérieur d'une unité au nombre réelle de cp. On conclue que le nombre de cp dans le système DjR A serait au minimum de 5 et de 8 dans le système DjR B

- Discussion

On compare dans le tableau 4.7 les résultats expérimentaux de l'indicateur NSS avec les statistiques théoriques des indicateurs min(IND) et n(SF>5%). Comme précédemment, l'incertitude théorique sur NSS étant faible, les résultats expérimentaux entrent tous dans les

descriptions statistiques théoriques bien qu'une valeur expérimentale (min(IND) pour DjR B) soit néanmoins élevée au regard de la moyenne théorique (31 vs 18,1 ± 4,4).

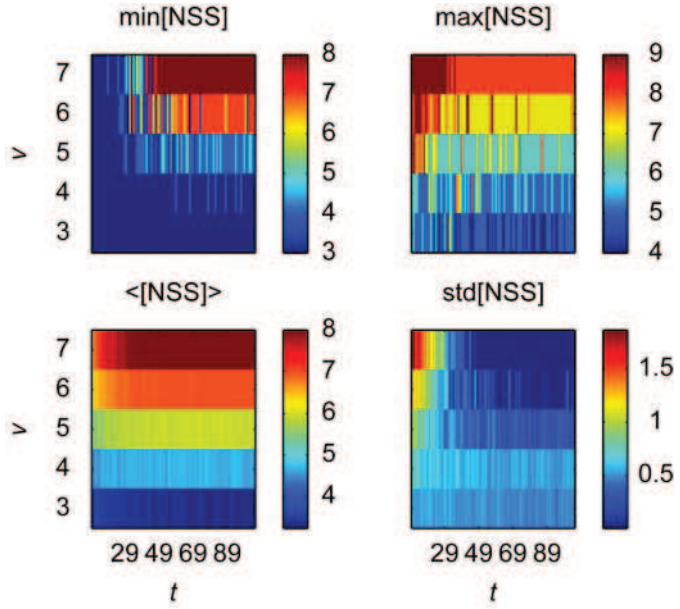


Figure 4.36 : Valeurs statistiques de l'indicateur NSS pour des jeux de données issus du mélange de spectres théorique XANES 2.

Tableau 4.7 : Comparaison entre les résultats NSS expérimentaux des sites DjR A et DjR B avec les résultats statistiques théoriques min(IND) et n(SF>5%)

Résultats expérimentaux	*Résultats théoriques*			
DjR A	(v, t)=(5,59)	min(IND)	n(SF>5%)	NSS
t = 59	min	7	7	4
NSS = 6 → v = 5	max	19	14	6
min(IND) = 15	moy.	10.3	9.0	5.9
n(SF>5%) = 9	std.	2.3	1.7	0.4
DjR B	(v, t)=(8,114)	min(IND)*	n(SF>5%)*	NSS**
t = 114	min	10	10	9
NSS = 9 → v = 8	max	33	23	9
min(IND) = 31	moy.	18.1	13.2	9
n(SF>5%) = 16	std.	4.4	2.9	0

*résultats obtenus après extrapolations polynomiales de degré 1 suivant v et t
** extrapolation impossible, résultats déduits suivant l'observation visuelle des comportements à la figure 4.36.

5.4.1.3) Conclusions sur le XANES

Ces calculs montrent que l'indicateur NSS est un paramètre performant pour la détermination du nombre de cp dans un mélange de spectres XANES théoriques. Plus le domaine en énergie des données initiales est grand, plus ce paramètre est précis.

Par ailleurs, les résultats NSS obtenus sur les données expérimentales sont inférieurs aux résultats maximums acceptables déterminés au précédent chapitre avec l'approche

itérative et les résultats expérimentaux sont cohérents avec les descriptions statistiques théoriques des paramètres min(IND) et n(SF>5%).

L'ensemble de ces remarques permet de proposer 5 composantes principales pour le site de Mamut C3, 6 pour le site de Mamut B4, 5 pour le site DjR A et 8 pour le site DjR B.

5.4.2) EXAFS

Les résultats statistiques de l'évaluation de l'indicateur NSS sur des mélanges issus des spectres EXAFS 2 sont donnés à la figure 4.37. Du fait de l'écart en terme de vp entre la première composante et la seconde qui peut faire apparaître un saut important sur σ(OG), le paramètre *v* varie de 3 à 6 et le paramètre *t* de 10 à 105.

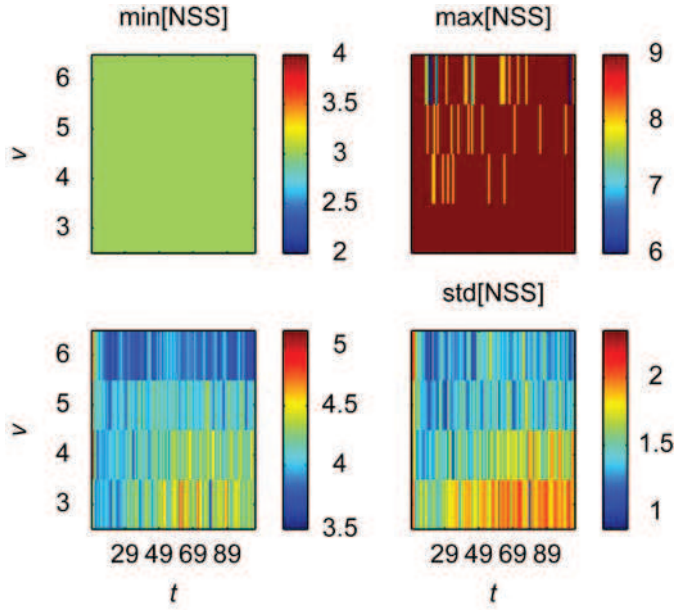


Figure 4.37 : Valeurs statistiques de l'indicateur NSS pour des jeux de données issus du mélange de spectres théorique EXAFS 2.

Contrairement aux résultats précédents obtenus en spectroscopie XANES il apparait clairement que l'indicateur NSS n'est pas pertinent pour mesurer directement le nombre de cp présentes dans les spectres EXAFS multicomposants. Pourtant l'étude des spectres expérimentaux a montré que des cassures étaient observables pour des valeurs de σ(OG) proches du nombre *j* de cp dans le système. Cette observation est confirmée en affichant les résultats d'un jeu de données théoriques pour lequel $v = 6$ et $t = 30$ (Fig. 4.38). Sur ce jeu de données, on retrouve bien un pic pour cp = 7, mais le pic principal est à cp = 3, en accord avec la figure 4.37. Néanmoins, la présence de cette cassure entre la $j^{ème}$ et la j+$1^{ème}$ cp confirme que l'étude de la courbe σ(OG) peut néanmoins donner des informations. Malheureusement,

il est impossible de tester statistiquement ces cassures et donc de mesurer la précision sur cet indicateur avec la spectroscopie EXAFS. Il reste donc indispensable de mesurer le nombre de cp par d'autres méthodes (visuelles notamment), cet indicateur NSS pouvant tout de même servir à faire converger le faisceau de présomptions vers le nombre de cp réel.

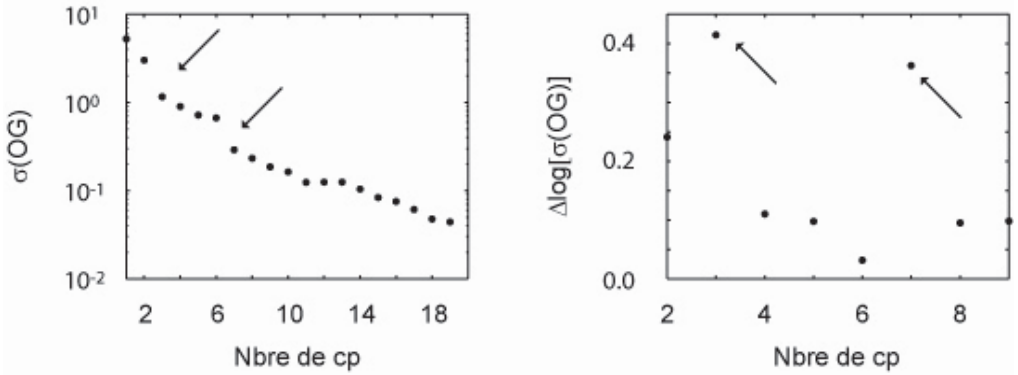


Figure 4.38 : à gauche, valeurs de σ(OG) en fonction du nombre de cp utilisées dans la reconstruction ACP d'un jeu de données (v = 6 et t = 30) EXAFS théoriques représentées dans l'espace $k^3\chi(k)$.

6) Conclusions

En conclusion de ce chapitre, nous pouvons dire que l'Analyse en Composantes Principales est effectivement une approche mathématique très puissante pour déterminer le nombre de cp présentes dans un système expérimental, et par inférence le nombre d'espèces pures contenues dans un mélange. Néanmoins, cette force est aussi sa faiblesse, puisque le bruit en général, statistique ou non, est perçu comme des composantes indépendantes. Dans le domaine de la spectroscopie d'absorption X, la proximité au sens des corrélations des signatures spectrales d'un système multicomposants empêche de déterminer par la mesure des valeurs propres le nombre de cp, et les indicateurs développés à l'origine pour l'étude des systèmes chimiques (IND, F-test) ne sont pas pertinents pour l'étude des données d'absorption X (EXAFS et XANES). Au mieux, grâce à l'évaluation statistique dont ils ont fait l'objet dans ce chapitre, ils peuvent servir à vérifier la pertinence d'un résultat déduit d'une autre méthode d'évaluation plus robuste.

L'indicateur NSS proposé est fiable sur les données XANES si, d'une part, elles sont définies sur un domaine en énergie suffisamment grand après le seuil (typiquement 150 eV) et si, d'autre part, la taille des jeux de données, c'est-à-dire le nombre d'observables, est relativement conséquente (typiquement une quarantaine de spectres). Si l'une de ces deux conditions n'est pas réunie, cet indicateur fournit le nombre de cp avec une incertitude de l'ordre d'une unité.

Compte tenu de la forme réellement très spécifique du bruit sur les données EXAFS normalisées, il n'a pas été possible d'évaluer statistiquement la fiabilité de cet indicateur, et ainsi d'en déterminer son domaine d'application. L'analyse de données expérimentales montre qu'il peut exister deux solutions de l'indicateur NSS, et éventuellement davantage.

Conclusions et Perspectives

Le cuivre, de part sa configuration électronique d^9 à l'état divalent, est un élément chimique très réactif capable de se lier fortement aux structures et molécules organiques donneuses d'électrons. Sa grande réactivité chimique est mise à profit par l'ensemble des êtres vivants pour, notamment, stocker l'énergie produite dans les mitochondries (cytochrome) et les chloroplastes (plastocyanine), mais elle peut également entrainer des désordres fonctionnels lorsqu'il est en excès. Relativement peu toxique pour les animaux terrestres, le cuivre est particulièrement nocif pour les micro-organismes et les végétaux, principalement pour ceux vivants en milieu aquatique.

C'est sous sa forme libre hydratée et liée à des petites molécules organiques que l'ion Cu^{2+} est le plus biodisponible. Dans les sols, il est relativement peu mobile en raison de sa capacité à réagir avec un grand nombre de constituants organiques et inorganiques. De nombreuses études ont été réalisées pour quantifier les équilibres entre les formes solubles et insolubles et évaluer ses formes de rétention (e.g. Benedetti et al. 1995; Koopal et al. 2005 ; Matynia et al. 2010). Mais les modélisations thermodynamiques sont surparamétrées dans l'espace des observables, si bien que les mesures macroscopiques ne permettent pas, à elles seules, de caractériser avec exactitude la réactivité physico-chimique de cet élément, et encore moins ses mécanismes réels d'interaction avec les constituants du milieu aux échelles atomiques et moléculaires (Lenoir et al. 2010 ; Lenoir et Manceau, 2010). L'utilisation de techniques structurales, comme la spectroscopie d'absorption X (XAS), permet d'accéder à ce type d'information. Il a ainsi été montré sur de la matière organique de sol que l'affinité du cuivre pour cette fraction pouvait s'expliquer par la formation de chélates Cu-MO du fait de l'adéquation géométrique entre la taille des sites réactifs COOH-COH-R de la matière organique et celle du plan équatorial formé par l'ion Cu^{2+} et les groupements hydroxyles (Manceau et Matynia 2010).

Néanmoins, les sols sont des milieux hétérogènes dans lequel le cuivre est présent sous plusieurs formes et la spectroscopie X, qui est sensible à son environnement chimique et structural moyen, permet difficilement de vérifier sur des échantillons bruts la pertinence des observations effectuées sur des échantillons fractionnés. Cette difficulté peut être contournée en couplant des techniques dites « micro » dans lesquelles la taille latérale du faisceau X est de l'ordre de quelques μm^2. Ainsi, la micro-fluorescence X donne accès à la répartition et à l'association des éléments chimiques au sein de la matrice pédologique et la micro-

spectroscopie d'absorption X dispense des informations structurales in-situ. Bien que les hétérogénéités des sols puissent être de l'ordre du nanomètre, s'il existe à cette échelle une variabilité spécifique des formes de l'élément étudié, alors leurs proportions varient, sauf exception, d'un point d'analyse à l'autre, ce qui se traduit par un changement du signal mesuré par micro-XAS. L'analyse en Composantes Principales (ACP) permet alors en théorie de dénombrer les espèces pures contenues dans la matrice hétérogène.

Ces techniques microscopiques ont été utilisées pour caractériser les formes du cuivre dans des rizières contaminées par une exploitation minière de chalcopyrite ($CuFeS_2$) située en Malaisie et dans le bassin d'infiltration des eaux de ruissellement Django-Reinhardt situé en bordure Est de l'agglomération lyonnaise. En plus de l'association habituelle Cu-MO, qui est omniprésente dans ces formations, du cuivre métallique et de la chalcopyrite, diverses formes cristallines localement spécifiques ont été repérées. Malheureusement, pour chacun des deux sites, l'ensemble des formes cuprifères n'a pas pu être déterminé, leur nombre total n'ayant pu être que majoré par ACP.

A partir d'une approche statistique sur des modèles spectraux représentatifs des données expérimentales, il a été montré que cette indétermination pouvait s'expliquer par le bruit expérimental, la taille du jeu de données, le nombre assez élevé d'espèces pures dans ces formations et la faible fiabilité des indicateurs ACP existants (Malinowski 1991). Ce problème a été surmonté par la mise au point d'un nouvel indicateur ACP dénommé NSS. Son applicabilité a été évaluée statistiquement sur des mélanges de composition connue: dans le cas le plus défavorable, la précision sur le nombre d'espèces est de ± 1.

Maintenant que ce critère existe, la prochaine étape serait d'identifier, aussi précisément que possible, chacune des espèces cuprifères présentes dans ces deux milieux, ceci dans le but d'apprécier au mieux les risques environnementaux. A cette fin, deux techniques complémentaires pourraient être mises en œuvre : la micro-diffraction X, de façon à préciser la minéralogie des phases cristallines cuprifères, et l'EXAFS polarisé sur des films minces orientés, de façon à préciser les sites cristallographiques d'insertion du cuivre dans le réseau ou à la surface des phyllosilicates. Par ailleurs, l'approche mathématique développée dans ce travail peut permettre, moyennant quelques développements supplémentaires, de quantifier statistiquement la présence ou non d'une espèce pure, dont on aurait le spectre XAS, dans une série de spectres de mélanges analysée par ACP. L'indicateur SPOIL (Malinowski 1991), actuellement utilisé à cet effet, est purement qualitatif (Manceau et al. 2002).

Enfin, il serait intéressant d'évaluer le domaine de validité du critère NSS sur d'autres systèmes de complexité croissance. A l'instar de ce qui a été fait sur le cuivre, ceci est possible en couplant, sur d'autres éléments chimiques, des calculs spectraux *ab initio* à l'approche inférentielle employée.

ANNEXE 1

Principes de la spectroscopie d'absorption X

La spectroscopie d'absorption est une technique analytique locale qui sonde l'environnement électronique et structural d'un atome absorbeur. Elle est basée sur l'absorption des rayons X par la matière qui résulte d'une interaction entre les charges électriques d'un atome cible et le champ électromagnétique dont est constituée la lumière. L'absorption d'un photon X correspond à l'excitation d'un électron de l'atome vers un état électronique lié vide ou vers le continuum.

La méthode consiste à mesurer l'absorbance μx de l'échantillon en fonction de l'énergie du rayonnement incident. En mode transmission, μx est défini par la loi de Beer-Lambert : $I = I_0 \exp(-\mu x)$. I_0 et I représentent respectivement l'intensité du faisceau incident et l'intensité du faisceau transmis, x, l'épaisseur de l'échantillon, μ le coefficient d'absorption linéaire. En mode de détection par fluorescence, μx est proportionnel au rapport I_f/I_0 ou I_f représente l'intensité de fluorescence (processus de désexcitation radiatif) des éléments excités.

Plusieurs phénomènes physiques sont à considérer dans un spectre d'absorbance (Fig. A1) et on appelle seuil d'absorption l'augmentation brutale de l'absorbance, à une valeur d'énergie donnée :

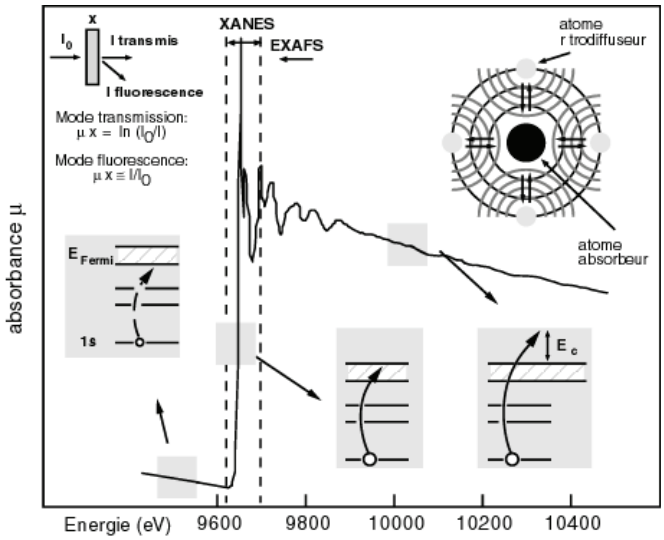


Figure A1 : Principe physique de la spectroscopie d'absorption des rayons X, exemple d'un spectre au seuil K du zinc (Panfili 2004).

- la première région située avant le seuil d'absorption proprement dit se caractérise par une décroissance monotone de l'absorbance en fonction de l'énergie croissante des rayons X. L'énergie incidente est alors inférieure à l'énergie de liaison des électrons de l'atome cible et les rayons X ne sont donc pas absorbés par l'atome cible.

- la deuxième région (Fig. A2) située quelques eV autour du seuil d'absorption est appelée XANES. Le seuil traduit l'accession possible par les électrons excités à des états vides situés juste au dessus du niveau de Fermi. Le seuil est accompagné de structures plus ou moins fines reflétant l'existence d'états inoccupés. L'analyse de cette zone du spectre fournit des informations sur la structure électronique de l'atome étudié. Lorsque l'énergie incidente atteint l'énergie d'ionisation E_0, l'électron quitte l'atome et le photoélectron assimilé à une onde électromagnétique se propage dans le continuum interatomique.

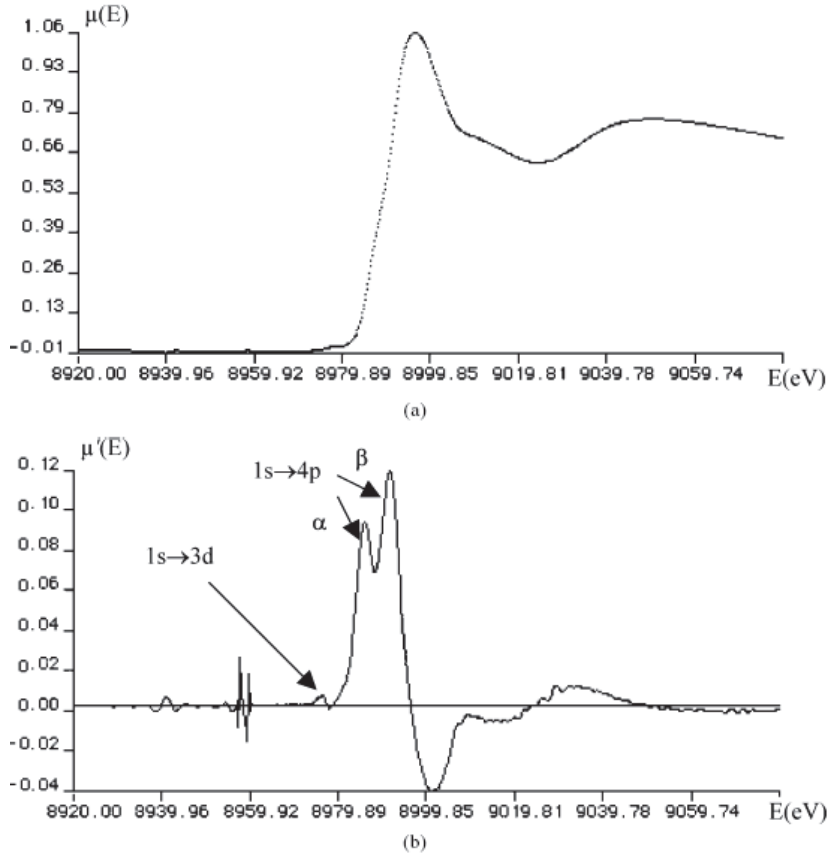


Figure A2 : En haut, spectre XANES au seuil K du cuivre d'un système Cu-sol naturel, en bas la dérivée première du spectre par rapport à l'énergie. Le petit pré-pic entre 8975 et 8980 eV correspond à la transition électronique 1s→3d, les pics α et β correspondent à la transition 1s→4p (Boudesocque et al. 2007). Dans le cas de Cu^{2+}, à cause de l'effet Jahn Teller, il y a une levée de dégénérescence des états $4p_z$ (pic α) et $4p_{x,y}$ (pic β) qui se traduit expérimentalement par le dédoublement du maximum de la dérivée première.

- la troisième région spectrale (Fig. A3), dénommée EXAFS, s'étend environ de E_0 + 50 eV à E_0 + 1000 eV. Dans ce domaine d'énergie, l'électron est envoyé dans le milieu avec

une énergie résiduelle E égale en première approximation à l'énergie du photon moins celle du seuil d'absorption E_0. A proximité du seuil, l'énergie cinétique du photoélectron est faible et son libre parcours moyen est donc grand (plusieurs dizaines d'angströms), si bien que l'onde associée a une grande probabilité d'être diffusée plusieurs fois par les atomes proches voisins (phénomène de diffusion multiple). Aux hautes énergies transférées, l'énergie cinétique du photoélectron est grande et son libre parcours moyen de ce fait limité à moins d'une dizaine d'angströms : c'est le domaine de la diffusion simple. La rétrodiffusion, simple ou multiple, du photoélectron module le coefficient d'absorption μ de l'atome émetteur, ce qui se traduit sur le spectre d'absorption par des oscillations dont l'amplitude et la fréquence dépendent de la nature (numéro atomique), du nombre (coordination) et de la distance des atomes situés à proximité de l'atome central. La précision sur les mesures de distance entre l'atome absorbeur et ses voisins est de l'ordre de 0.02 Å (Farquhar et al. 1997), et sur le nombre de voisins de 20%.

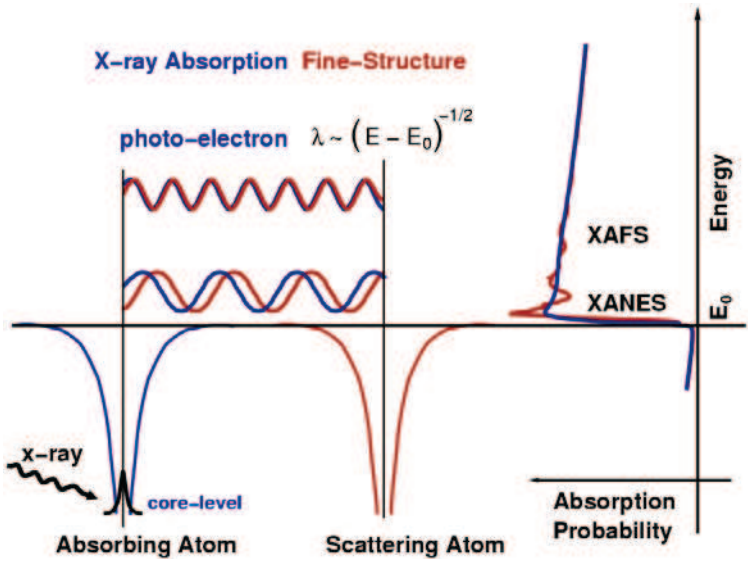


Figure A3: Les oscillations XAFS proviennent de la rétrodiffusion par les atomes voisins de l'onde électronique émise par un atome cible (Newville 2004).

Le signal EXAFS est décrit par une somme de sinusoïdes amorties, chacune d'elles associée à une couche atomique entourant l'atome absorbeur, et définie comme un ensemble N_i d'atomes identiques (i) tous situés à la même distance (R_i) de l'atome central (c). L'équation EXAFS s'écrit :

$$\chi(k) = -S_0^2 \sum_i \frac{N_i}{kR_i^2} \left|\phi_i(\pi,k)\right| e^{-2\sigma_i^2 k^2} e^{\frac{-2R_i}{\lambda(k)}} \sin\left[2kR_i + \varphi_i(k) + 2\delta_c(k)\right]$$

où,

$k = \sqrt{\frac{4\pi^2 m_e}{h^2}(h\nu - E_0)}$ est le module du vecteur d'onde, m_e la masse de l'électron, $h\upsilon$ l'énergie du photon incident et E_0 l'énergie d'ionisation de l'élément absorbant ($\hbar = h/2\pi$ avec h, la constante de Planck) ;

S_0^2 : facteur d'échelle, terme de réduction d'amplitude dû aux effets multiélectroniques ;

$|\phi_i(\pi,k)|$: amplitude de rétrodiffusion de l'onde électronique par l'atome voisin de type i ;

σ_i : facteur de Debye-Waller dans l'approximation d'une distribution Gaussienne. Il comprend un terme de désordre statique (distribution des distances) et dynamique (agitation thermique) ;

λ: libre parcours moyen de l'électron ;

$\delta_c(k)$: fonction de déphasage de l'onde électronique par l'atome central ;

$\varphi_i(k)$: fonction de déphasage de l'onde électronique par l'atome rétrodiffuseur.

Les fondements théoriques et les équations qui en découlent ont été largement abordés dans la littérature et ne seront pas détaillés ici (Teo 1986 ; Koningsberger et Prins 1998).

ANNEXE 2

Semi-empirical proton binding constants for natural organic matter

Available online at www.sciencedirect.com
ScienceDirect
Geochimica et Cosmochimica Acta 74 (2010) 1836–1851
Geochimica et Cosmochimica Acta
www.elsevier.com/locate/gca

Semi-empirical proton binding constants for natural organic matter

Anthony Matynia [a,b], Thomas Lenoir [a,c], Benjamin Causse [d,e], Lorenzo Spadini [d], Thierry Jacquet [b], Alain Manceau [a,*]

[a] Mineralogy & Environments Group, LGCA, Université Joseph Fourier and CNRS, 38041 Grenoble Cedex 9, France
[b] Phytorestore – Site et Concept, 7 impasse Milord, 75018 Paris, France
[c] Laboratoire Central des Ponts et Chaussées (LCPC), Route de Bouaye, BP 4129, 44341 Bouguenais Cedex, France
[d] Environmental Geochemistry Group, LGIT, Université Joseph Fourier and CNRS, 38041 Grenoble Cedex 9, France
[e] LSE-ENTPE, Université de Lyon, 69518 Vaulx en Velin Cedex, France

Received 7 July 2009; accepted in revised form 17 December 2009; available online 25 January 2010

Abstract

Average proton binding constants ($K_{H,i}$) for structure models of humic (HA) and fulvic (FA) acids were estimated semi-empirically by breaking down the macromolecules into reactive structural units (RSUs), and calculating $K_{H,i}$ values of the RSUs using linear free energy relationships (LFER) of Hammett. Predicted log $K_{H,COOH}$ and log $K_{H,Ph\text{-}OH}$ are 3.73 ± 0.13 and 9.83 ± 0.23 for HA, and 3.80 ± 0.20 and 9.87 ± 0.31 for FA. The predicted constants for phenolic-type sites (Ph-OH) are generally higher than those derived from potentiometric titrations, but the difference may not be significant in view of the considerable uncertainty of the acidity constants determined from acid–base measurements at high pH. The predicted constants for carboxylic-type sites agree well with titration data analyzed with Model VI (4.10 ± 0.16 for HA, 3.20 ± 0.13 for FA; Tipping, 1998), the Impermeable Sphere model (3.50–4.50 for HA; Avena et al., 1999), and the Stockholm Humic Model (4.10 ± 0.20 for HA, 3.50 ± 0.40 for FA; Gustafsson, 2001), but differ by about one log unit from those obtained by Milne et al. (2001) with the NICA-Donnan model (3.09 ± 0.51 for HA, 2.65 ± 0.43 for FA), and used to derive recommended generic values. To clarify this ambiguity, 10 high-quality titration data from Milne et al. (2001) were re-analyzed with the new predicted equilibrium constants. The data are described equally well with the previous and new sets of values ($R^2 \geqslant 0.98$), not necessarily because the NICA-Donnan model is overparametrized, but because titration lacks the sensitivity needed to quantify the full binding properties of humic substances. Correlations between NICA-Donnan parameters are discussed, but general progress is impeded by the unknown number of independent parameters that can be varied during regression of a model fit to titration data. The high consistency between predicted and experimental $K_{H,COOH}$ values, excluding those of Milne et al. (2001), gives faith in the proposed semi-empirical structural approach, and its usefulness to assess the plausibility of proton stability constants derived from simulations of titration data.
© 2010 Elsevier Ltd. All rights reserved.

1. INTRODUCTION

Humic (HA) and fulvic (FA) acids are the two most abundant and acid–base reactive fractions of natural organic matter. They are complex and heterogeneous assemblages of dominantly carboxyl (COOH) and phenol (Ph-OH) functional groups issued from the breakdown of bacterial, algal, and/or higher plant organic material. The acid–base equilibria of these groups, which controls HA and FA ion binding properties, is measured usually by potentiometric titration and expressed numerically by the two proton binding constants log $K_{H,COOH}$ and log $K_{H,Ph\text{-}OH}$, and the two associated site densities $Q_{H,COOH}$ and $Q_{H,Ph\text{-}OH}$. The density of

* Corresponding author. Tel.: +33 4 76 82 80 15; fax: +33 4 76 82 81 01.
E-mail addresses: Alain.Manceau@obs.ujf-grenoble.fr, Manceau@ujf-grenoble.fr (A. Manceau).

0016-7037/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.gca.2009.12.022

protonatable acidic and basic groups can be obtained directly from the height of the two ascending S-shaped curves with pH, as for any polyelectrolyte. However, the proton dissociation constant of each group cannot be determined simply from the pH at an inflection point, because this pH does not coincide with the stoichiometric equivalence point of the acid–base titration. Since protons are electrostatically retained by COO^- and Ph-O^- groups, more base needs to be added to remove them from the negatively charged organic molecule than from a neutral molecule. Consequently, inflection points represent apparent (also called 'conditional') pK values ($pK_{H,iapp}$) which are shifted to higher pH relative to intrinsic binding constants ($pK_{H,i}$) by a value not directly accessible by experiment, and instead estimated by computing the proton consumption (ΔQ) as a function of pH with an electrostatic model.

Three main empirical models have been developed to describe electrostatic interactions: the Impermeable Sphere (IS) model, the Donnan model, and the Stockholm Humic Model (SHM) (Tipping, 2002). In the IS model, humic substances (HS) are viewed as rigid spheres or cylinders, having charge assumed to be localized on their exteriors (Tipping et al., 1990; Tipping and Hurley, 1992; Bartschat et al., 1992; De Wit et al., 1993; Milne et al., 1995). Because ions can penetrate HS, the softness of the organic molecules is accounted for by optimizing the particles size during the data regression (Kinniburgh et al., 1996; Saito et al., 2005). Using this model, Avena et al. (1999) obtained log $K_{H,i}$ values for humic acids in the range 3.5–4.5 (carboxylic-type) and 7.5–8.5 (phenolic-type) (Table 1).

In the Donnan model, HS are considered to form a gel phase separated from the bulk water phase by a 'Donnan volume', which contains counterions that neutralize the gel (Marinsky et al., 1982; Marinsky and Ephraim, 1986). This electrostatic model has been implemented in Model VI (Tipping, 1998), and the NICA (non-ideal competitive adsorption) model (Koopal et al., 1994; Benedetti et al., 1995) under the name NICA-Donnan (N-D; Kinniburgh et al., 1996). From the simulation of nineteen titration curves, Tipping (1998) obtained log $K_{H,COOH}$ of 4.10 ± 0.16 (HA) and 3.20 ± 0.13 (FA), and log $K_{H,Ph\text{-}OH}$ of 8.80 ± 0.23 (HA) and 9.40 ± 0.78 (FA) (Table 1). Milne et al. (2001) extended the dataset analyzed by Tipping (1998) to a total of 48 titration curves, and analyzed the expanded dataset with the N-D model in two ways. In the first, every titration curve was best-fit individually to calculate average N-D values and their standard deviations, and in the second, the model fits for HA and FA were optimized by simultaneous fitting of 15 HA curves and 11 FA curves to derive generic N-D values (Table 1). The average (2.65 ± 0.43) and generic (2.34) log $K_{H,COOH}$ values for FA are lower than the value obtained with Model VI (3.20 ± 0.13), and this difference reaches one order of magnitude for HA (3.09 ± 0.51 and 2.93 with N-D vs. 4.10 ± 0.16 with Model VI). Since these values are intrinsic proton affinity constants, i.e., specific properties of the HS that are independent of solution parameters, such as pH, salt concentration, and solid/liquid ratio, these large differences between model approaches are unlikely. The differences do not seem, however, uniquely inherent to the models used in the two studies, because values closer to those obtained with Model VI than the generic values were obtained in two other studies with N-D (Table 1; Plaza et al., 2005; Drosos et al., 2009). In summary, the generic N-D values reported by Milne et al. (2001) seem low for carboxylic-type groups.

In the Stockholm Humic Model, humic materials also are considered to have a gel-like structure, but the coulombic charges inside the gel are screened with a modified Stern model (Gustafsson, 2001). Following the philosophy of Tipping (1998) and Milne et al. (2001), a new set of generic parameters was produced by Gustafsson (2001). The new recommended values for carboxylic-type sites, 4.10 ± 0.20

Table 1
Compilation of proton binding constants (log $K_{H,i}$ and σ) and site densities ($Q_{H,i}$ and σ) for humic and fulvic acids.

		log $K_{H,COOH}$	log $K_{H,Ph\text{-}OH}$	$Q_{H,COOH}$	$Q_{H,Ph\text{-}OH}$	Method of analysis	Electrostatic interaction
Milne et al. (2001)[a]	HA	3.09 ± 0.51	7.98 ± 0.96	3.17 ± 0.89	2.66 ± 1.37	N-D	Donnan model
	FA	2.65 ± 0.43	8.60 ± 1.06	5.66 ± 1.25	2.57 ± 1.94		
Plaza et al. (2005)	HA	3.74 ± 1.09	7.68 ± 0.60	3.3 ± 0.3	1.3 ± 0.8	N-D	Donnan model
	FA	2.83 ± 0.36	7.02 ± 0.18	4.3 ± 0.6	2.3 ± 0.2		
Drosos et al. (2009)	HA	3.62 ± 0.19	8.54 ± 0.48	4.16 ± 0.58	1.34 ± 0.27	N-D	Donnan model
Tipping (1998)	HA	4.10 ± 0.16	8.80 ± 0.23	3.3 ± 0.5	1.7 ± 0.2	Model VI	Donnan model
	FA	3.20 ± 0.13	9.40 ± 0.78	4.8 ± 0.7	2.4 ± 0.4		
Gustafsson (2001)	HA	4.10 ± 0.20	8.95 ± 0.15	–	–	Stockholm	Donnan model
	FA	3.50 ± 0.40	8.75 ± 0.30			Humic Model	
Avena et al. (1999)	HA	3.50–4.50	7.50–8.50	–	–	Impermeable Sphere	Impermeable Sphere
Ritchie and Perdue (2003)	HA	4.38 ± 0.13	9.72 ± 0.23	5.20 ± 0.52	0.73 ± 0.24	Modified Henderson–	–
	FA	3.80 ± 0.11	9.78 ± 0.48	7.04 ± 0.79	0.64 ± 0.21	Hasselbalch Model	

[a] The recommended generic values are log $K_{H,COOH} = 2.93$ (HA) and 2.34 (FA), and log $K_{H,Ph\text{-}OH} = 8.00$ (HA) and 8.60 (FA). All constants are intrinsic except those from Ritchie and Perdue (2003), which were not corrected for electrostatic effects (i.e., they are conditional values). Q values have units of $mmol_c/g_{dw}$, which are equivalent to units of meq_c/g_{dw} for protons.

for HA and 3.50 ± 0.40 for FA, are within precision identical to those derived with Model VI.

The log $K_{H,COOH}$ values for HS also were evaluated using infrared spectra. From the analysis of 24 aliphatic monocarboxylates, Cabaniss and McVey (1995) established a linear relationship between log $K_{H,COOH}$ and the asymmetric COO^- stretch frequency. This correlation was used by Hay and Myneni (2007) to estimate the acidity constants of a series of HS, including references from the International Humic Substance Society (IHSS). This approach yielded values in the range of 3.3–3.5, again in disagreement with the generic N-D values of 2.93 for HA and 2.34 for FA (Table 1).

The large variability among the proton binding constants of carboxylic-type and, to a lesser extent of phenolic-type, sites for HS underscores the need for an independent evaluation of their validity. This question is addressed here in two successive and complementary steps. First, average log $K_{H,COOH}$ and $K_{H,Ph\text{-}OH}$ values are calculated semi-empirically from structure models of FA and HA using linear free energy relationships (LFER) of Hammett (Hepler, 1963; Ren, 1998). Second, 10 high-quality titration data from Milne et al. (2001) were simulated with the N-D model using successively the predicted LFER $K_{H,i}$ values and the recommended generic $K_{H,i}$ values. Results show that this same dataset can be described statistically equally well with $K_{H,i}$ values that differ typically by one order of magnitude or more. Thus, the variability of titration constants reported in the literature partly results from the great flexibility of the N-D model.

Since all N-D parameters cannot be obtained reliably by regression analysis of titration data, due merely to the lack of sensitivity of the titration method (Westall et al., 1995), some parameters need to be estimated independently by other techniques or conceptual approaches. For example, the Donnan volume can be derived from peak elution times measured by size exclusion chromatography, and the number of phenolic-type sites by ^{13}C NMR (Christl and Kretzschmar, 2001; Drosos et al., 2009). Attempts to determine the Donnan volume by viscometry were less successful, however (Benedetti et al., 1996; Avena et al., 1999). The present study extends this communal effort to improve the applicability of the N-D model by estimating proton binding constants on a structural basis.

2. STRUCTURE MODELS FOR HA AND FA

Four main structure models have been proposed over the last three decades (Fig. 1). Their main chemical characteristics are given in Table 2 and summarized below.

2.1. HA model of Stevenson (HA-S; Stevenson, 1982)

This model is based on the "modified lignin theory" (Waksman, 1938), which considers HS to be degradation products of lignin. The structure is characterized by a high polycyclicity (atomic ratio of aromatic to aliphatic carbons $C_{aro}/C_{ali} = 1.36$), the majority of carboxyl ($COOH_{aro}/COOH_{ali} = 4.0$) and all hydroxyl groups branching off aromatic rings. The aromatic rings are linked by C–C, C–O–C, and C–N–C bonds.

2.2. HA model of Schulten and Schnitzer (HA-SS; Schulten and Schnitzer, 1993)

This model is derived from ^{13}C NMR, analytical chemistry, pyrolysis, and oxidative degradation data. The macromolecular unit contains 368 C atoms compared to 73 in the HA-S model, but it counts a lower proportion of aromatic carbons ($C_{aro}/C_{ali} = 0.57$), which are now linked by aliphatic chains. Consistent with the lower C_{aro}/C_{ali} atomic ratio, HA-SS has relatively fewer phenolic groups ($OH_{aro}/C_{tot} = 0.021$ vs. 0.11) and hydroxyls are almost evenly distributed between aliphatic and aromatic carbons ($OH_{aro}/OH_{ali} = 1.14$). The molar ratio of COOH to total carbon ($COOH/C_{tot}$) is similar in the two models (0.078 vs. 0.068), and carboxyl groups still are dominantly attached to aromatic rings ($COOH_{aro}/COOH_{ali} = 4.8$).

2.3. FA model of Buffle (FA-B; Buffle, 1977)

With only 27 carbons, this is the smallest macromolecule. It consists of a naphthalene core substituted by COOH and OH groups, and branched with two aliphatic chains terminated by a carboxyl. This molecule has the highest $COOH_{tot}/C_{tot}$ (0.22) and $COOH_{aro}/C_{aro}$ (0.4) ratios, with almost one in two aromatic C–H bonds substituted by a COOH.

2.4. FA model of Alvarez-Puebla et al. (FA-A; Alvarez-Puebla et al., 2006)

This FA model is characterized by a high proportion of chain aliphatic and alicyclic carbon ($C_{aro}/C_{ali} = 0.32$), to which the majority of functional groups are connected ($OH_{aro}/OH_{ali} = 0.50$; $COOH_{aro}/COOH_{ali} = 0.2$). This model is the only one to incorporate in its structure sulfhydryl (–SH) and amine ($–NH_2$) functional groups.

2.5. Comparison between the HA and FA models

The HA models have a higher proportion of aromatic carbons than the FA models: $C_{aro}/C_{ali} = 1.36$ and 0.57 for HA-S and HA-SS, vs. 0.59 and 0.32 for FA-B and FA-A, respectively. However, the FA models have a higher proportion of COOH to total carbon (0.16–0.22 vs. 0.068–0.078), and thus a higher acidity with site density ($Q_{H,COOH}$) values of 6.8–9.4 vs. 3.4-4.5 $mmol_c/g$. Although the distinction is less conclusive for phenolic hydroxyls ($Q_{H,Ph\text{-}OH}$), FA have a total proton-reactive site density ($Q_{H,tot}$) well above that of HA: 11.6–12.5 vs. 5.7–8.2 $mmol_c/g$ (Table 2). Overall, the nature and amounts of reactive sites in structure models are in good agreement with pH-based estimates from titration (Table 1).

2.6. Carboxyl structures (NOM model of Myneni and co-authors)

Previously, HA and FA were represented as arbitrary hydrocarbon macromolecules substituted by carboxyl and hydroxyl groups. Currently, the tendency is to represent HA and FA as assemblies of well-defined chemical entities.

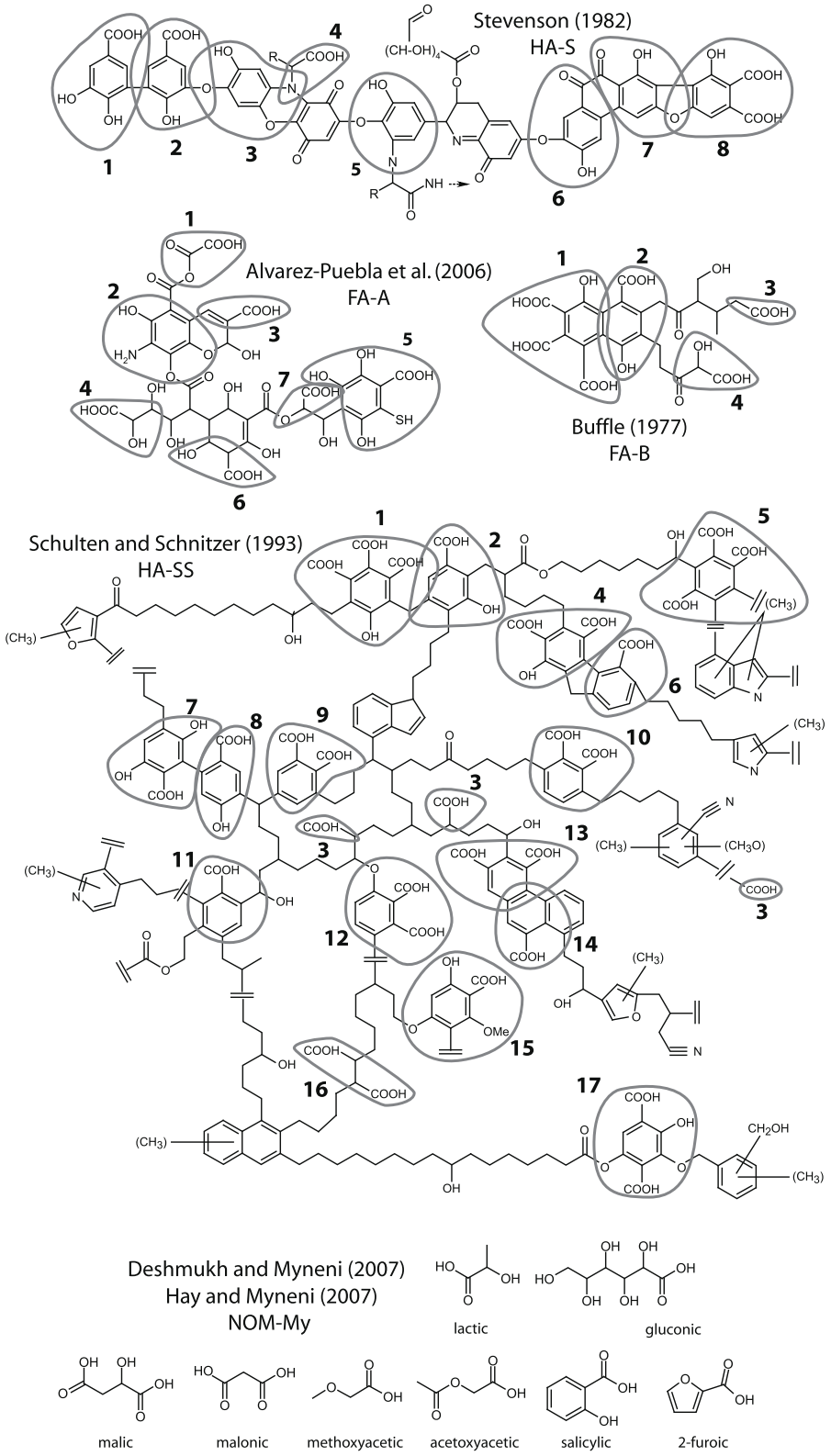


Fig. 1. Structure models for humic and fulvic acids. Circled areas delimit the reactive structural units (RSUs).

Using total reflection Fourier-transform infrared spectroscopy and aliphatic and aromatic carboxylate model reference materials, the following generic types of carboxyl environments have been identified in natural organic matter

Table 2
Chemical characteristics of structure models for humic and fulvic acids.

	$\frac{O}{C}$	$\frac{C_{aro}}{C_{ali}}$	$\frac{COOH_{aro}}{COOH_{ali}}$	$\frac{OH_{aro}}{OH_{ali}}$	$\frac{COOH_{aro}}{C_{aro}}$	$\frac{OH_{aro}}{C_{aro}}$	$\frac{COOH_{tot}}{C_{tot}}$	$\frac{OH_{tot}}{C_{tot}}$	$Q_{H,COOH}$ ($mmol_c/g$)	$Q_{H,Ph\text{-}OH}$ ($mmol_c/g$)	$Q_{H,tot}$ ($mmol_c/g$)
HA-S $C_{73}H_{49}N_3O_{32}$	0.44	1.36	4.0	–	0.095	0.19	0.068	0.11	3.4	4.8	8.2
HA-SS $C_{368}H_{501}N_{23}O_{89}$	0.24	0.57	4.8	1.14	0.179	0.06	0.078	0.021	4.5	1.2	5.7
FA-B $C_{27}H_{26}O_{18}$	0.66	0.59	2.0	1.0	0.40	0.20	0.22	0.074	9.4	3.1	12.5
FA-A $C_{37}H_{35}NO_{32}S$	0.86	0.32	0.2	0.50	0.083	0.33	0.16	0.11	6.8	4.8	11.6

(NOM) by Hay and Myneni (2007): aliphatic acids containing a hydroxyl (similar to lactate, malate, gluconate), ether/ester (–O–/–O–CO–, e.g., methoxyacetate and acetoxyacetate) or carboxylate (e.g., malonate) substituent on the α-carbon, aromatic acids (e.g., salicylate), and O-heterocycle acids (e.g., furancarboxylate). Aliphatic carboxyls with –OH, –OR, or –COOH for H substitution on the α-carbon, and furan-type and salicylate-type aromatic carboxylates functionalities were confirmed by two-dimensional NMR spectroscopy (Deshmukh et al., 2007). This technique also highlighted the importance of α-substituted alicyclic carboxyls, which are indistinguishable from α-substituted aliphatic carboxyls in infrared spectroscopy.

3. CALCULATION OF SEMI-EMPIRICAL LOG $K_{H,I}$

3.1. Method

Empirical proton binding constants were calculated in three steps. First, each structure model was broken down to chemically characterizable entities, called reactive structural units (RSUs). Second, $K_{H,i}$ values of the RSUs were taken from the NIST thermodynamic database (Smith and Martell, 2004), or calculated using basic chemical concepts derived from the Hammett equation (Hepler, 1963; Ren, 1998). Third, the average $K_{H,i}$ value of the whole macromolecule was calculated as the weighted sum of all individual values. Calculations were performed at an ionic strength of $I = 0.1$ M because the NIST database contains more entries for this value. The RSUs were defined as follows:

(i). Each RSU contains only aliphatic or aromatic hydrocarbons, and at least one acid–base functionality.
(ii). Unreactive naphthalene and fluorene cores, aliphatic chains, and phenyl groups on aromatics were replaced by $–CH_3$ (denoted Me). This chemical substitution is supported by the similarities of log $K_{H,COOH}$, (i), for 3-(2methyl-2-propyl)benzoic acid (4.20) and 3-methylbenzoic (4.27) and (ii) for 1-naphthoic acid (3.67) and 2,3-methylbenzoic acid (3.76).
(iii). Ester groups on aromatics (Ph–CO–O–R) were replaced by $–CH_3$. This substitution is justified by the similarity of log $K_{H,Ph\text{-}OH}$ for 2-acetylphenol (9.94) and 2-methylphenol (10.09).
(iv). Ether groups (R–O–R) were replaced by methoxy groups ($–O–CH_3$, denoted OMe). This substitution is supported by the similarity of log $K_{H,COOH}$ for 3-phenoxybenzoic acid (3.95) and 3-methoxybenzoic acid (3.82).
(v). Secondary amines on aromatics (Ph–NH–R) were considered to be primary amines ($Ph–NH_2$). This approximation is justified by the similarity of log $K_{H,COOH}$ for 2-aminobenzoic acid (4.78) and 2-(2-hydroxyethylamino)benzoic acid (4.82).

3.2. The RSUs

The nature and number of RSUs generated in all structure models are represented in Table 3. HA-S can be decomposed into eight RSUs (Fig. 1). The two first RSUs (RSU1 and RSU2) are molecular remnant of lignin, from which the model is derived. HA-SS can be described by seventeen RSUs. All RSUs contain at least one COOH group, and half of them contain at least two. Thus, none contains only hydroxyls, in contrast to RSU3, RSU5, RSU6, and RSU7 in HA-S. Acetic acid is the most abundant RSU; it is repeated three times. FA-B is described by four RSUs, two small aliphatic acids (acetic acid and lactic acid), and two substituted benzene-type rings. FA-A can be described by seven RSUs, five non-aromatic and two benzene-type aromatic. A proton is replaced by NH_2 in one aromatic ring, and by SH in the other. Hence, more RSUs are needed to describe HA than FA, and apart from acetic and lactic acids, each RSU occurs in one model only. The limited number of common structural elements among the first four structure models highlights the complexity and diversity of HS.

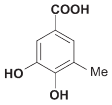
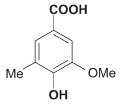
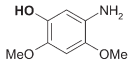
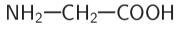
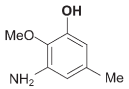
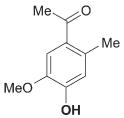
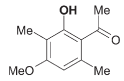
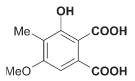
Eight RSUs were used for the NOM model: lactic acid, gluconic acid, malic acid, malonic acid, methoxyacetic acid, acetoxyacetic acid, salicylic acid, and 2-furoic acid (also named furan-2-carboxylic acid).

3.3. Log $K_{H,i}$ values of the RSUs

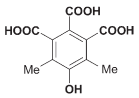
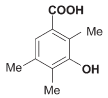
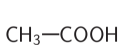
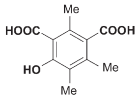
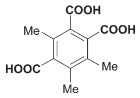
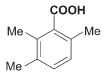
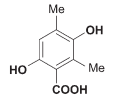
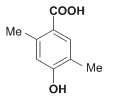
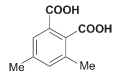
Ten RSUs out of the 34 defined previously are referenced in the NIST database: acetic and lactic acid, glycine, 2-oxopropanoic acid (FA-A_RSU1), acrylic acid (FA-A_RSU3), methoxyacetic acid (FA-A_RSU7), 3-hydroxypropanoic acid (FA-A_RSU6), but-1,4-dicarboxylic acid (HA-SS_RSU16), and the two aromatic HA-SS_RSU6

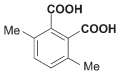
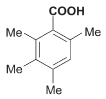
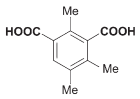
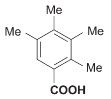
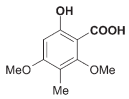
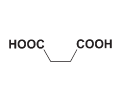
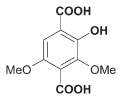
Table 3
Reactive structural units and calculated proton binding constants at 0.1 M ionic strength of the structure models for humic and fulvic acids.

HA-S (Stevenson, 1982)

								
	RSU1	RSU2	RSU3	RSU4	RSU5	RSU6	RSU7	RSU8
log $K_{H,COOH}$	4.45	4.33	NC	2.33[a]	NC	NC	NC	3.45
log $K_{H,OH}$	9.07	9.47	10.18	NC	10.32	8.35	10.61	14.7[b]
Multiplicity	1	1	1	1	1	1	1	1
Average		log $K_{H,COOH} = 3.64$			log $K_{H,OH} = 9.67$			

HA-SS (Schulten and Schnitzer, 1993)

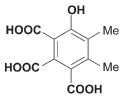
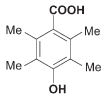
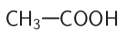
									
	RSU1	RSU2	RSU3	RSU4	RSU5	RSU6	RSU7	RSU8	RSU9
log $K_{H,COOH}$	4.32	3.99	4.56[a]	3.07	3.51	3.31[a]	2.97	4.35	4.08
log $K_{H,OH}$	9.472	10.72	NC	13.47[b]	NC	NC	11.28[b]	9.78	NC
Multiplicity	1	1	3	1	1	1	1	1	1

								
	RSU10	RSU11	RSU12	RSU13	RSU14	RSU15	RSU16	RSU17
log $K_{H,COOH}$	3.82	3.47[a]	3.76	3.61	4.00[a]	2.99	4.62[a]	2.82
log $K_{H,OH}$	NC	NC	NC	NC	NC	13.23[b]	NC	13.59[b]
Multiplicity	1	1	1	1	1	1	1	1
Average		log $K_{H,COOH} = 3.81$			log $K_{H,OH} = 9.99$			

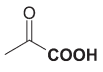
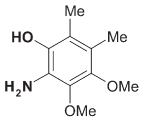
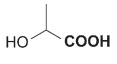
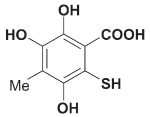
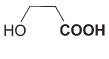
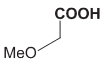
(*continued on next page*)

Table 3 (*continued*)

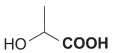
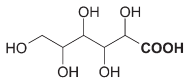
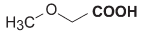
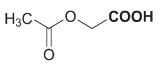
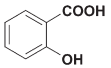
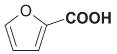
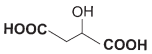
FA-B (Buffle, 1977)

	 RSU1	 RSU2	 RSU3	RSU4
log $K_{H,COOH}$	3.87	3.76	4.56[a]	3.67[a]
log $K_{H,OH}$	13.17[b]	10.08	NC	NC
Multiplicity	1	1	1	1
Average	log $K_{H,COOH} = 3.97$		log $K_{H,OH} = 10.08$	

FA-A (Alvarez-Puebla et al., 2006)

	 RSU1	 RSU2	RSU3	 RSU4	 RSU5	 RSU6	 RSU7
log $K_{H,COOH}$	2.26[a]	NC	4.09[a]	3.67[a]	2.37	4.40[a]	3.32[a]
log $K_{H,SH}$	NC	NC	NC	NC	8.96	NC	NC
log $K_{H,OH}$	NC	10.86[b]	NC	NC	10.33	NC	NC
log $K_{H,NH2}$	NC	5.27	NC	NC	NC	NC	NC
Multiplicity	1	1	1	1	1	1	1
Average		log $K_{H,COOH} = 3.63$			log $K_{H,OH} = 9.65$		

NOM-My (Deshmukh et al., 2007; Hay and Myneni, 2007)

	 Lactic acid	 Gluconic acid	Malonic acid	 methoxy-acetic acid	 Acetoxy-acetic acid	 Salicylic acid	 2-furoic acid	 Malic acid
log $K_{H,COOH}$	3.67	3.44	3.96[c]	3.32	2.81	2.80	2.98	3.96[c]
log $K_{H,OH}$	NC	NC	NC	NC	NC	13.4 [b]	NC	NC
Multiplicity	1	1	1	1	1	1	1	1
Average		log $K_{H,COOH} = 3.37$			log $K_{H,OH} = NC$			

[a] Reference from NIST database (Smith and Martell, 2004).

[b] Value excluded from the calculation. The cutoff for log $K_{H,OH}$ was set at 10.80. For this value, one third of hydroxyl groups are deprotonated at pH 10.5, which is the upper limit of titration measurements.

[c] Average of the log $K_{H,COOH}$ for the two groups. NC, not calculated.

and HA-SS_RSU11 (Table 3). The log $K_{H,i}$ values of the other RSUs were calculated by taking the known constant of the closest structural analog and adding to this value the Δ log K offset induced by the chemical substituents present in the RSU but absent in the analog (Table EA-1). The inductive modification of the chemical properties of organic molecules by substituents is described conceptually as a mesomeric or resonance effect. Details of the calculations are given in Electronic annex data, and results reported in Table 3. The method is illustrated below with HA-SS_RSU17, a benzene molecule substituted by two *para*-COOH (noted $COOH^1$ and $COOH^2$), one *ortho*-OH, and two (3,5)-O-CH_3 (noted OMe) (Fig. 2).

The proton binding constant of this benzenedicarboxylate can be calculated from the log $K_{H,i}$ values of 2-hydroxybenzoic acid (2.80) and 3-hydroxybenzoic acid (3.99), corrected by the dissociation offset induced by the addition of two OMe and a second COOH group on the ring (Fig. 2). Starting from 2-hydroxybenzoic acid, the Δlog K offset due to the two OMe substituents in (3,5) C positions on the benzene ring is two times the difference between the log $K_{H,COOH}$ for 3-methoxybenzoic acid and benzoic acid : $2 \times (3.82 - 4.01) = -0.38$. The shift in log $K_{H,i}$ induced by a second COOH can be obtained from the binding constants (one per hydroxyl) of benzene-1,4-dicarboxylic acid: $3.38 + 4.15$. The Δlog K associated with this substituent is $0.5 \times (4.15 + 3.38) - 4.01 = -0.25$. Thus, (OMe,COOH)-substituted 2-hydroxybenzoic acid has a calculated log $K_{H,COOH}$ of $2.80 - 0.38 - 0.25 = 2.17$.

Similarly, the log $K_{H,COOH}$ of (OMe,COOH)-substituted 3-hydroxybenzoic acid is obtained from the difference in proton binding constants of 2-methoxybenzoic acid and benzoic acid ($3.87 - 4.01 = 0.14$), and the difference in the mean binding constant of benzenedicarboxylate and benzoic acid (-0.25). Note that 2-methoxybenzoic acid has to be taken instead of 3-methoxybenzoic acid because the two OMe substituents are now in (2,6) C positions relative to the carboxyl group (Fig. 2). Completing the algebra yields log $K_{H,COOH} = 2.82$ for RSU17. Calculated and experimental constants generally coincide within 0.1–0.2 log units (SI).

3.4. Results

The log $K_{H,COOH}$ values calculated by the substituent approach at 0.1 M ionic strength are 3.64 (HA-S), 3.81 (HA-SS), 3.97 (FA-B), and 3.63 (FA-A). The average is 3.73 ± 0.13 for HA, and 3.80 ± 0.20 for FA. The log $K_{H,Ph\text{-}OH}$ values are 9.67 (HA-S), 9.99 (HA-SS), 10.08 (FA-B), and 9.65 (FA-A), and the averages are 9.83 ± 0.23 (HA) and 9.87 ± 0.31 (FA). Their dependence on ionic strength (I) can be estimated (i) experimentally from the variation of pK_a with I for model compounds or (ii) empirically from the Davies equation (Davis, 1962; Stumm and Morgan, 1996):

$$pK_a(I) = pK_a(I = 0) + 2 \log \gamma_I$$

$$\log \gamma = -AZ^2 \left(\frac{\sqrt{I}}{1 + \sqrt{I}} - BI \right)$$

where γ is the activity coefficient, A the Debye-Hückel constant (0.49 at 298 K), Z the ion charge, and B an adjustable parameter between 0.2 and 0.3. Fig. 3 shows that experimental pK_a varies by 0.28 for carboxylates and 0.35 for phenol in the $0 \leqslant I \leqslant 2.0$ interval, and is a minimum at $I = 0.5$. The Davies equation gives the same result (Fig. 3). From these results:

$$\log K_{H,i} = \log K_{H,i}^{RSU(I=0.1\ M)} \pm 0.2$$

The uncertainty is small relative to the heterogeneity of binding properties of HS, and the validity of models (e.g., Donnan, Impermeable Sphere; Avena et al., 1999) used to describe electrostatic interactions in titration measurements.

The predicted proton binding constants of carboxylic groups are essentially the same for the four HS models, differing only by a few tenths of log units. In our calculation, the NOM-My model has a lower constant (3.4), but this value should be considered cautiously, if it is even meaningful, because the proportions of the constitutive RSUs in this model are unknown. Also, the RSUs from NOM-My are short-chain carboxylates with higher O/C ratios (0.84 on average) than those in the four macromolecular models (Table 2). This excess of electron acceptor increases the acidity of the acid–base reactive sites, as shown with gluconic acid (3.44) and its oxygen-depleted analog hexanoic acid (4.63), and with acetoxyacetic (2.81) and methoxyacetic (3.32) acids, and their oxygen-depleted analog acetic acid (4.56). The RSUs identified by IR and NMR are connected in humic and fulvic acids by unsubstituted aliphatic and alicyclic chains (Hay and Myneni, 2007), with the consequence of lowering the O/C ratio of the whole molecule, and thus the acidity of carboxyl groups. The relationship between the O/C ratio and acidity of carboxylates is consistent with the higher measured acidity of FA relative to HA (Table 1). However, this difference is not confirmed by the structural approach since HA has a predicted log $K_{H,COOH}$ of 3.73 ± 0.13 and FA 3.80 ± 0.20.

The average predicted acidity of carboxylic groups (3.77 ± 0.20) agrees remarkably well with the average experimental value of 3.58 ± 0.34 for HA + FA calculated from the data by Tipping (1998), Gustafsson (2001), Plaza et al. (2005), and Drosos et al. (2009) (Table 1). In contrast, the HA (3.09 ± 0.51) and FA (2.65 ± 0.43) values derived by Milne et al. (2001) from the NICA-Donnan titration model for proton binding to HS are abnormally much lower, and thus likely underestimated. The average predicted basicity of hydroxyl groups (9.85 ± 0.33) is higher than the titration estimates (8.41 ± 0.53), but this difference is probably insignificant because titration curves often have their second inflection point at higher pH than pH_{max}. For example, this is the case for 25 data out of the 48 analyzed by Milne et al. (2001). Determining a meaningful $K_{H,Ph\text{-}OH}$ value for these HS is fraught with considerable difficulty and imprecision, because only 50% of the phenolic-type sites were deprotonated at most at pH_{max}. To help clarify further the reasons for the discrepancy between the NICA-Donnan and predicted log $K_{H,i}$ values, potentiometric titration data from Milne et al. (2001) were re-analyzed with the new constants.

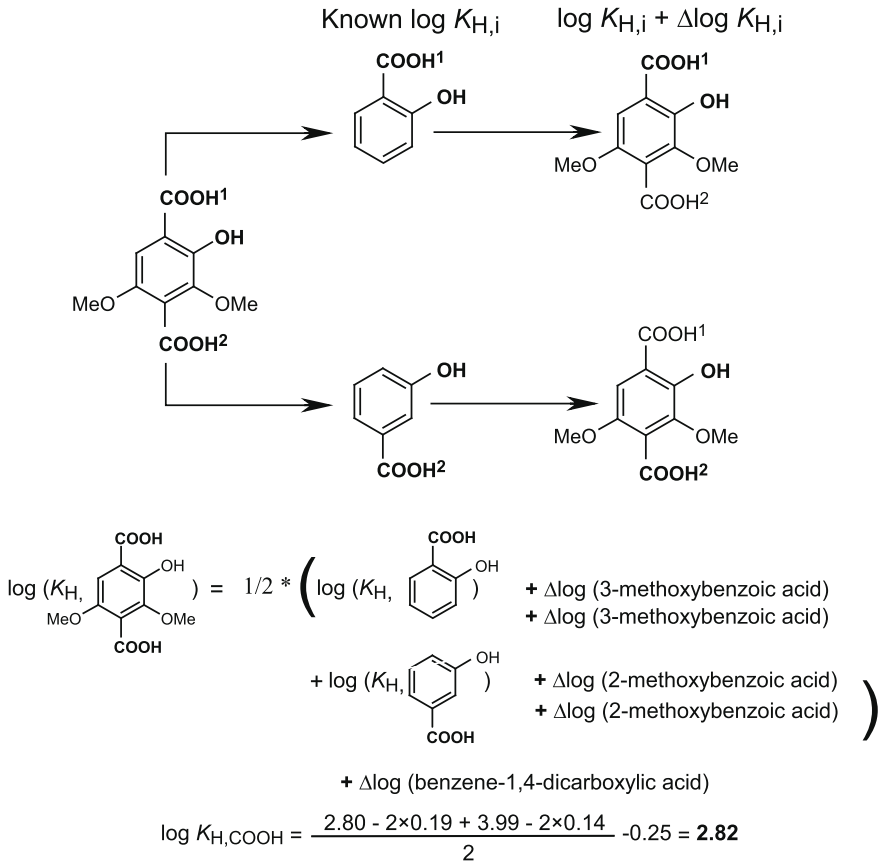


Fig. 2. Calculation of the predicted proton binding constant for HA-SS_RSU17. Calculations for the other RSUs are given in Electronic annex.

4. SIMULATION OF TITRATION DATA

This analysis was performed on five sets of HA and five sets of FA data, out of the 48 analyzed by Milne et al. (2001). The criteria retained for choosing a titration data were that (i) it was measured at several ionic strengths, (ii) in a pH interval of at least seven units (~[3.0–3.5; 10.0–10.5]), and (iii) at least one log $K_{H,i}$ (as determined by Milne et al. (2001)) was close to a generic N-D value or distant from a predicted value. Thus, all data are high-quality and none of them had extreme proton binding constants in the previous study.

In the NICA-Donnan model, the distribution of proton binding constants resulting from structural heterogeneity of carboxylic and phenolic groups is represented by two pseudo-Gaussians (Benedetti et al., 1996; Kinniburgh et al., 1996, 1999; Koopal et al., 2005). In total, titration data are simulated generally with eight parameters: the two proton binding constants (log $K_{H,COOH}$, log $K_{H,Ph\text{-}OH}$) and their associated distribution widths ($m_{H,COOH}$, $m_{H,Ph\text{-}OH}$), the densities of proton-binding sites ($Q_{H,COOH}$, $Q_{H,Ph\text{-}OH}$), an empirical parameter b, which accounts for the dependence of the Donnan volume on ionic strength, and the initial protonation state or charge (IC) defined as $IC = Q_{H,COOH} + Q_{H,Ph\text{-}OH} - Q_0$, with Q_0 adjustable (Kinniburgh, 1999).

In a first simulation (hereafter referred to as 'Gen' simulation) the 10 sets of HA and FA data were fit by fixing log $K_{H,COOH}$ and log $K_{H,\ Ph\text{-}OH}$ to their recommended generic values from the NICA-Donnan model (Milne et al., 2001), and optimizing the six other parameters with the FIT code (Kinniburgh et al., 1999). For HA, the two generic constants are 2.93 and 8.0, and for FA 2.34 and 8.60, respectively. This first fit strategy provided good agreement between data and theory, with R^2 values >0.99 and RMSE < 0.12 (Fig. 4, Table 4). The mean parameter values from the individual fits are all statistically identical to those obtained by Milne et al. (2001), which engenders confidence in our calculations. More precisely, $\langle b(HA)\rangle = 0.52 \pm 0.17$ in the new simulation vs. 0.51 ± 0.19 in Table 2 of Milne et al. (2001), $\langle b(FA)\rangle = 0.68 \pm 0.06$ vs. 0.63 ± 0.16, $\langle m_{H,COOH}(HA)\rangle = 0.48 \pm 0.10$ vs. 0.55 ± 0.13, $\langle m_{H,COOH}(FA)\rangle = 0.39 \pm 0.06$ vs. 0.41 ± 0.09, $\langle m_{H,Ph\text{-}OH}(HA)\rangle = 0.30 \pm 0.10$ vs. 0.43 ± 0.21, $\langle m_{H,Ph\text{-}OH}(FA)\rangle = 0.52 \pm 0.27$ vs. 0.57 ± 0.21, $\langle Q_{H,COOH}(HA)\rangle = 3.54 \pm 0.63$ vs. 3.17 ± 0.89, $\langle Q_{H,COOH}(FA)\rangle = 5.91 \pm 1.55$ mmol$_c$/g vs. 5.66 ± 1.25, $\langle Q_{H,Ph\text{-}OH}(HA)\rangle = 2.77 \pm 0.28$ vs. 2.66 ± 1.37, and $\langle Q_{H,Ph\text{-}OH}(FA)\rangle = 2.16 \pm 1.48$ vs. 2.57 ± 1.94. The initial charge varies from -0.58 to 1.09 mmol$_c$/g in the five HA, and -0.31 to 1.26 mmol$_c$/g in the five FA. These variations are comparable to the variability of the $Q_{H,COOH}$ values reported by Milne et al. (2001) between different humic substances: $\sigma = 0.89$ for HA and $\sigma = 1.25$ for FA, thus essentially negligible. A negative IC value is, however, plausible, as it can be considered to be derived from the dissociation of H^+ from the material, for example

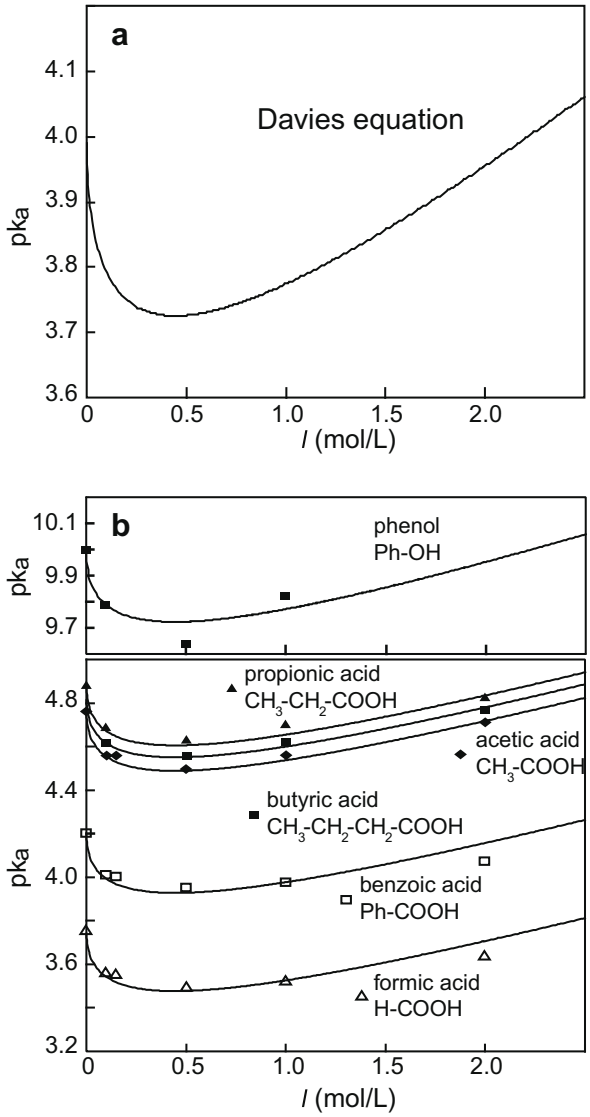


Fig. 3. (a) Dependence of pK_a on ionic strength as calculated with the Davies equation ($pK_a = 4.0$ at $I = 0$ and $B = 0.27$). (b) pK_a for simple organic acids in the $0 \leqslant I \leqslant 2.0$ interval from the NIST database. The points are experimental values and the lines are the Davies equations calculated for $B = 0.27$ and the pK_a value of each organic acid at $I = 0$. The experimental dependence of pK_a values on ionic strength is well reproduced theoretically by the Davies equation. In both cases (i.e., theory and measurement), the acidity constant is minimum for $I = 0.5$ and $\Delta pK_a = 0.28$, except for $\Delta pK_{exp.}$ for phenol (0.35). Ph stands for benzene ring.

consecutively to the addition of salt. Therefore, the initial charge was added to the best-fit value of the abundance of acidic groups when it was negative. In contrast, positive *IC* values were not added to the fitted $Q_{H,COOH}$ values, because they are chemically unrealistic.

In a second simulation (referred to as 'Stru' simulation) of the same dataset, $\log K_{H,COOH}$ and $\log K_{H,Ph\text{-}OH}$ were fixed to their predicted values (3.73 for HA and 3.80 FA, and 9.83 for HA and 9.87 for FA), and the six other parameters best-fit. Thus, the numbers of adjustable parameters were the same in the two approaches. All these fits gave $R^2 > 0.98$ and RMSE < 0.17, which is satisfactory (Fig. 4, Table 4). Four data were reproduced better than previously (HH15, HH18, FH20, and FH 22), one similarly (HH20), and five less well (HH09, HH11, FH02, and FH05). These differences are minor, meaning that the two fit strategies are equivalent.

The 'Stru' best-fit values obviously have different arithmetic means than the 'Gen' values, but the ranges of the two sets of parameter values generally overlap, except for $Q_{H,Ph\text{-}OH}$ and $m_{H,COOH}$ which are systematically higher in the new simulation by a factor of approximately two. More important, the results seem to follow a pattern in the two approaches, with systematic trends among some parameter values from the 10 sets of titration curves. This is verified by calculating the correlations for $Q_{H,i}$, $m_{H,i}$, and b between the Gen and Stru regressions. The $Q_{H,COOH}$ (Stru) and $Q_{H,COOH}$ (Gen) values from the 10 series of titrations are correlated to 91%, the two $Q_{H,Ph\text{-}OH}$ values to 94%, the two $m_{Ph\text{-},OH}$ to 79%, and b to 81%. Thus, the numerical differences among titration data analyzed with the same minimization procedure are probably meaningful, but the values inaccurate. The relative accuracy of the parameter values among different humic substances can be verified by calculating the $Q_{H,COOH}$ (HA) to $Q_{H,COOH}$ (FA) ratios, because the FAs are known to have a higher amount of carboxylic-type groups than the HAs (Section 2.5 and Table 2). This ratio is 0.62 in the Stru simulation and 0.60 in the Gen simulation, in agreement with the values of 0.66 reported by Ritchie and Perdue (2003), 0.69 by Tipping (1998), and 0.67 by Gondar et al. (2005).

The inaccuracy of the N-D parameter values can be demonstrated by calculating now the correlation coefficients of all best-fit values from the Gen and Stru regressions taken together. Calculation shows that the Stru + Gen values of $Q_{H,COOH}$ and $Q_{H,Ph\text{-}OH}$ are anticorrelated to 76%, and $Q_{H,COOH}$ and $m_{H,Ph\text{-}OH}$ correlated to 77%. Logically, $Q_{H,Ph\text{-}OH}$ and $m_{H,Ph\text{-}OH}$ are anticorrelated to 69%. The two equivalent Gen and Stru regressions were obtained by fixing $\log K_{H,i}$ either to their generic or predicted values. We therefore can ask if it is possible to change their values while keeping the same fit quality? The answer is obtained by calculating the correlations between $\log K_{H,i}$ and the other parameter values from the Stru + Gen results: $\log K_{H,COOH}$ is correlated with $m_{H,COOH}$ to 86% and with $Q_{Ph\text{-}OH}$ to 66%, whereas $\log K_{H,Ph\text{-}OH}$ is correlated with $m_{H,COOH}$ to 78%.

The correlations of $\log K_{H,COOH}$ with the other parameter values are shown graphically in Fig. 5 with the HH18 data. The same fit quality ($R^2 > 0.999$) can be obtained by co-varying $Q_{H,i}$ and $\log K_{H,COOH}$ in the adjustment procedure (Fig. 5a). Hence, the value of $Q_{H,COOH}$ can be decreased, that of $Q_{H,Ph\text{-}OH}$ increased, and $\log K_{H,COOH}$ adjusted by a few tenths of units with no significant change in the fit of the model to the data. Similarly, a variation of $\log K_{H,COOH}$ in the data fit can be compensated by adjusting $m_{H,i}$ and b (Fig. 5b). All these covariances are due primarily to the fact that the 'data window' of titration measurements is narrower than the 'spectral window' of the model (Westall et al., 1995).

In summary, a single dataset can be fit equally well with the NICA-Donnan model using variable combinations of

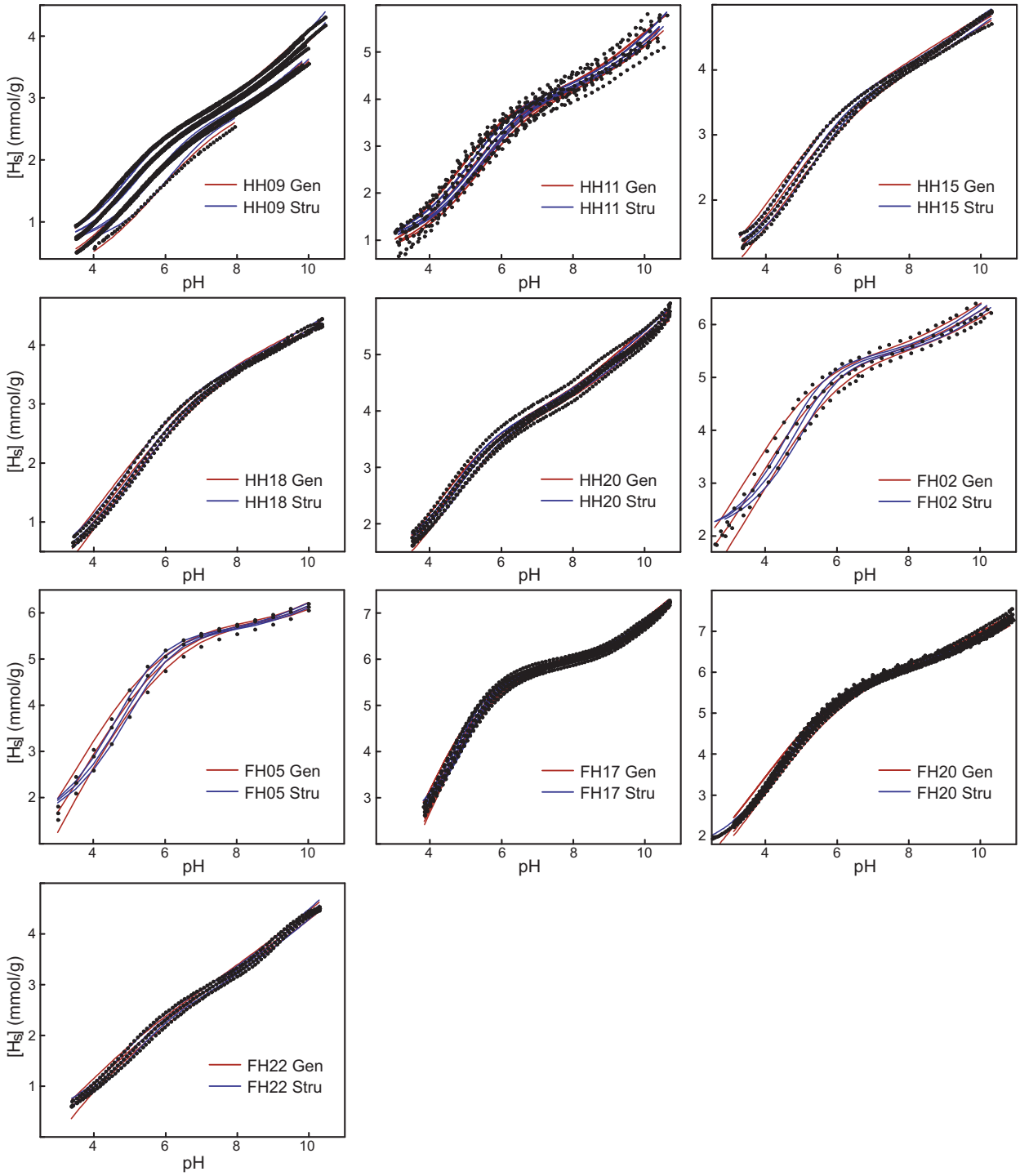


Fig. 4. Generic NICA-Donnan and structural models fits to titration data from Milne et al. (2001) measured at several ionic strengths. Black points are data and continuous lines are fits. HH codes for humic acid, and FH for fulvic acid.

{log $K_{H,COOH}$, $Q_{H,i}$} and {$Q_{H,i}$, $m_{H,i}$} numerical values. Therefore, differences in proton affinities and site densities reported in the literature (Table 1) result partly from the use of loosely constrained data fits, which obviously differ among studies. Results in Table 4 show that even regressions employing six freely varied parameters (b, $m_{H,i}$, $Q_{H,i}$, Q_0), instead of eight as in the N-D model (log $K_{H,i}$ were fixed), yield ambiguous values when the parameters are not independent. One pair of strongly correlated parameters suffices to render a seemingly good agreement between experimental data and theoretical model meaningless. Thus, some parameters need to be estimated by other

Table 4
Generic and best-fit parameter values from the simulations of HA and FA titration data previously studied by Milne et al. (2001)[a].

	RMSE[b]	R^{2b}	b^c	$Q_{H,COOH}$[d]	log $K_{H,COOH}$	$m_{H,COOH}$[e]	$Q_{H,Ph\text{-}OH}$	log $K_{H,Ph\text{-}OH}$	$m_{H,Ph\text{-}OH}$	IC[f]
Optimal parameter values of proton binding data for HA and FA. Values from Table 2 of Milne et al. (2001)										
HH										
Min	0.0120	0.9547	0.21	1.93	1.99	0.38	0.76	6.06	0.14	–
Max	0.1748	0.9999	0.84	4.73	3.90	0.89	5.39	10.06	0.86	–
Mean	0.0600	0.9944	0.51	3.17	3.09	0.55	2.66	7.98	0.43	–
Std	0.0416	0.0096	0.19	0.89	0.51	0.13	1.37	0.96	0.21	–
FH										
Min	0.0159	0.9802	0.29	2.64	2.00	0.27	0.55	7.19	0.17	–
Max	0.1529	0.9997	0.94	8.76	3.81	0.65	7.77	10.91	0.96	–
Mean	0.0814	0.9941	0.63	5.66	2.65	0.41	2.57	8.60	0.57	–
Std	0.0359	0.0058	0.16	1.25	0.43	0.09	1.94	1.06	0.21	–
Generic N-D parameter values. Values from Table 4 of Milne et al. (2001)										
HH			0.49	3.15	2.93	0.50	2.55	8.00	0.26	–
FH			0.57	5.88	2.34	0.38	1.86	8.60	0.53	–
Optimal parameter values from the 'Gen' simulation										
HH09	0.0221	0.9994	0.34	2.57	2.93[g]	0.51	2.81	8.00[g]	0.35	0.01
HH11	0.1192	0.9913	0.33	3.90	2.93	0.63	3.20	8.00	0.36	−0.58
HH15	0.0395	0.9987	0.61	3.26	2.93	0.45	2.82	8.00	0.21	−0.22
HH18	0.0415	0.9987	0.68	4.07	2.93	0.34	2.50	8.00	0.18	1.09
HH20	0.0676	0.9938	0.63	3.91	2.93	0.48	2.53	8.00	0.39	−0.38
Mean	0.0580	0.9964	0.52	3.54	–	0.48	2.77	–	0.30	−0.02
Std	0.0379	0.0036	0.17	0.63	–	0.10	0.28	–	0.10	0.65
FH02	0.0912	0.9950	0.61	5.46	2.34[g]	0.44	1.84	8.60[g]	0.39	−0.31
FH05	0.0892	0.9960	0.72	7.10	2.34	0.36	0.60	8.60	0.91	1.16
FH17	0.0548	0.9977	0.73	7.28	2.34	0.47	1.84	8.60	0.64	1.25
FH20	0.0954	0.9963	0.62	6.26	2.34	0.37	1.88	8.60	0.45	0.04
FH22	0.0537	0.9978	0.72	3.45	2.34	0.33	4.63	8.60	0.21	1.26
Mean	0.0679	0.9966	0.68	5.91	–	0.39	2.16	–	0.52	0.68
Std	0.0207	0.0012	0.06	1.55	–	0.06	1.48	–	0.27	0.75
Optimal parameter values from the 'Stru' simulation										
HH09	0.0564	0.9963	0.42	2.25	3.73[h]	0.82	5.84	9.83[h]	0.24	−0.52
HH11	0.1646	0.9835	0.56	3.87	3.73	0.72	5.66	9.83	0.30	−0.85
HH15	0.0242	0.9995	0.67	2.60	3.73	0.74	5.42	9.83	0.13	−0.68
HH18	0.0130	0.9999	0.69	2.81	3.73	0.53	4.26	9.83	0.15	0.17
HH20	0.0639	0.9939	0.70	3.05	3.73	0.89	6.16	9.83	0.18	−1.16
Mean	0.0644	0.9946	0.61	2.92	–	0.74	5.47	–	0.20	−0.61
Std	0.0599	0.0067	0.12	0.61	–	0.13	0.73	–	0.07	0.49

(*continued on next page*)

Table 4 (*continued*)

	RMSE[b]	R^2[b]	b[c]	$Q_{H,COOH}$[d]	log $K_{H,COOH}$	$m_{H,COOH}$[e]	$Q_{H,Ph\text{-}OH}$	log $K_{H,Ph\text{-}OH}$	$m_{H,Ph\text{-}OH}$	IC[f]
FH02	0.1636	0.9838	0.74	5.26	3.80[h]	0.87	3.29	9.87[h]	0.32	−2.09
FH05	0.1233	0.9923	0.79	5.49	3.80	0.78	2.18	9.87	0.33	−1.49
FH17	0.0659	0.9967	0.78	5.66	3.80	0.98	3.88	9.87	0.34	−2.03
FH20	0.0815	0.9973	0.89	5.56	3.80	0.67	3.67	9.87	0.25	−1.66
FH22	0.0365	0.9990	0.72	1.40	3.80	0.82	7.78	9.87	0.15	0.02
Mean	0.0942	0.9938	0.78	4.67	–	0.82	4.16	–	0.28	−1.45
Std	0.0499	0.0061	0.06	1.84	–	0.11	2.13	–	0.08	0.86

[a] For consistency, the original names of titration data were preserved, i.e., HHx and FHx, where the first letter stands for humic (H) or fulvic (F) acid, the second for proton, and x for the sample number.
[b] Merit of fit as in Kinniburgh et al. (1999) and Milne et al. (2001).
[c] $b = (\log V_D)/(1 - \log I)$, with V_D the Donnan volume and I the ionic strength (Kinniburgh et al., 1996).
[d] $Q_{H,i}$, the site densities, have units of mmol$_c$/g$_{dw}$.
[e] $m_{H,i}$ is a measure of the apparent distribution of $K_{H,i}$ values.
[f] If negative, the initial charge was added to $Q_{H,COOH}$.
[g] Values fixed to those of the generic NICA-Donnan model.
[h] Values fixed to those of the structural model.

means than titration measurements alone (Christl and Kretzschmar, 2001; Drosos et al., 2009). In the absence of complementary data, the applicability of a model can be improved tangibly by constraining the model fit to converge towards chemically and physically realistic values following an operationally defined minimization procedure, as proposed by Matynia (2009) for the N-D model.

5. APPLICATION OF THE RSU APPROACH TO OTHER POLYACIDS

The RSU approach developed in this work can be applied to other polyelectrolyte molecules. Two examples are shown in Fig. 6. The first is lignin, a precursor of HS, which has predicted acidity constants of 4.0 and 9.8 in close proximity to those for HA and FA. A major difference between the two types of natural polymers is the absence of aromatic carboxyls in the structure model of lignin, and the predominance of phenolic groups, which are many times more numerous than carboxylic groups (Merdy et al., 2002). The second is the carboxyl-rich alicyclic molecules (CRAM), which are the most

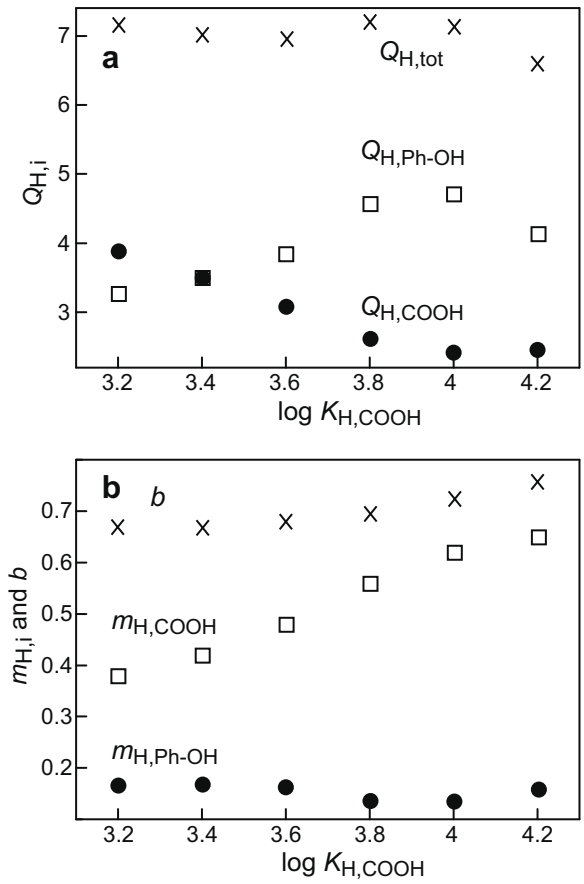


Fig. 5. Correlations between log $K_{H,COOH}$ and NICA-Donnan parameters obtained by varying incrementally the proton binding constant from 3.2 to 4.2 and optimizing $Q_{H,i}$, $m_{H,i}$ and b. Log $K_{Ph\text{-}OH}$ was fixed to 9.83. $Q_{H,COOH}$ is a little more sensitive to a variation of log $K_{H,COOH}$ than $Q_{H,Ph\text{-}OH}$ is, meaning that $Q_{H,COOH}$ and log $K_{H,COOH}$ have a higher degree of correlation (71% vs. 66%). $Q_{H,i}$ is in mmol$_c$/g, $m_{H,i}$ and b are unitless.

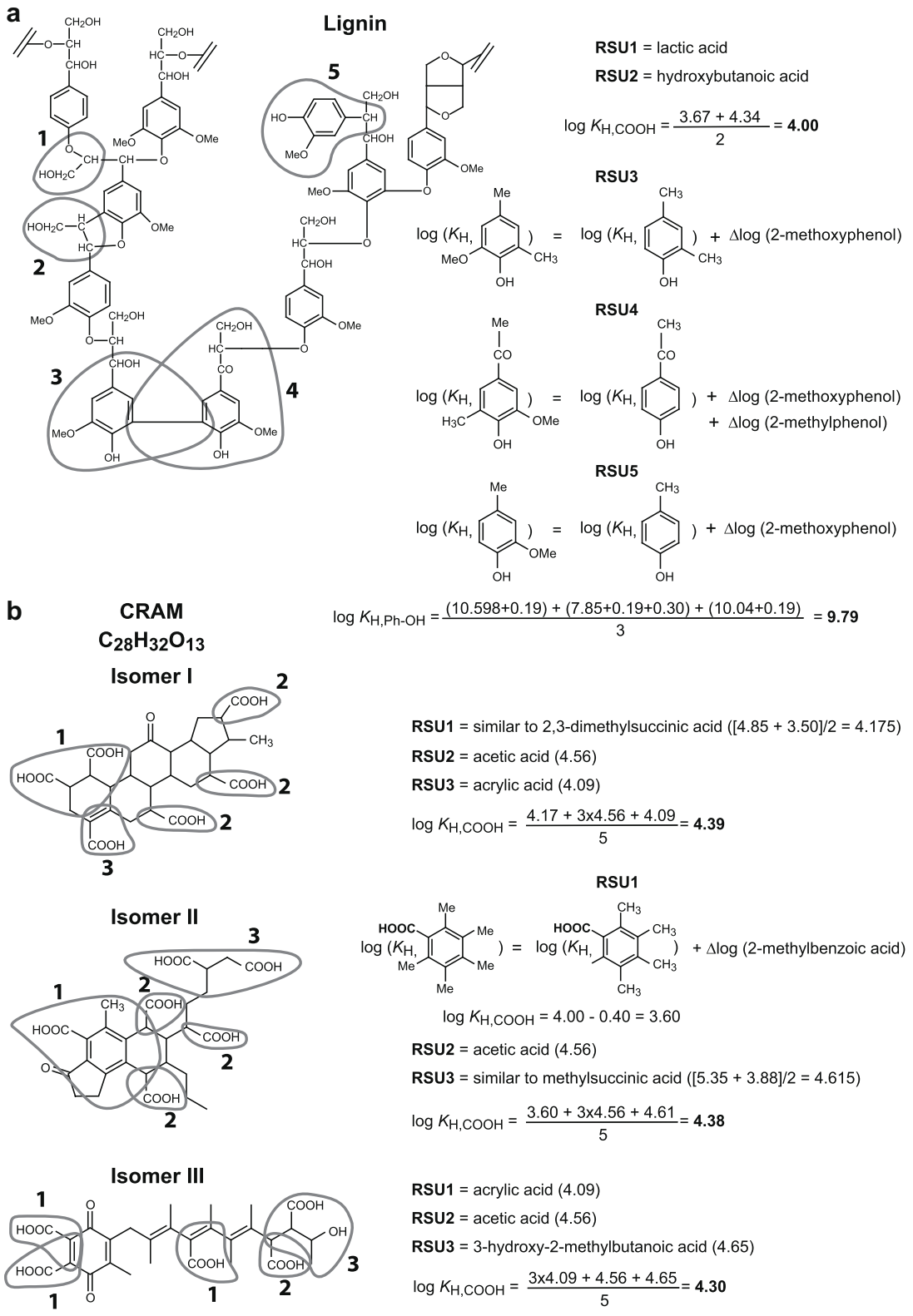


Fig. 6. Calculation of the predicted proton binding constants for the refractory components lignin (a) from natural organic matter (model after Merdy et al., 2002) and carboxyl-rich alicyclic molecules (b, CRAM) from dissolved organic matter (models after Hertkorn et al., 2006). Isomers I and II, which are most abundant, have multiple fused aromatic rings and a high ratio of substituted carboxyl groups. Circled areas delimit the reactive structural units (RSUs).

abundant identified component of DOM in marine and fresh waters (Hertkorn et al., 2006; Lam et al., 2007). Their key features that are characteristic of the majority of CRAM structures are a cyclic terpenoid backbone with a high degree of carboxylation shown as isomer I in Fig. 6b. These carboxylated alicyclic structures have carboxyl to aliphatic carbon ratios of approximately 1:2 to 1:7 and a predicted acidity constant of 4.3–4.4.

6. CONCLUSION

Characterizing the acid–base reactivity of NOM by potentiometric titration is challenging because model parameters are not easily resolvable by regularization methods leading to convergence problems. Although the NICA-Donnan model has been shown in a number of publications to be able to describe proton and metal cation binding of humic substances fairly well mathematically, some researchers reported difficulties in obtaining reliable parameter values caused by overfitting (Christl and Kretzschmar, 2001). The semi-empirical chemical substituent method presented in this study provides a rationale to estimate intrinsic proton affinity constants of NOM on a structural basis. Although some of the published structural models used here were not created expressly to describe the acid–base chemistry of humic substances, the predicted acidity constants agree with the majority of those obtained from experiments. In addition, the differences in log K values among these structural models are generally small in comparison to the standard deviations of constants derived from titration. We expect that the RSU approach will become more precise as progress in spectroscopic techniques and analytical chemistry continue. The RSU approach also should help constrain minimization algorithms of titration data and achieve convergence towards plausible fit parameters.

ACKNOWLEDGMENTS

This manuscript benefitted from the constructive comments of Drs. Michal Borkovec and Dimitri A. Sverjensky, and two anonymous reviewers. This research was funded by the Association Nationale de la Recherche Technique (ANRT) through the awarding of a Ph.D. fellowship to A. Matynia, and the Région Ile-de-France through the ARITT program (Aide Régionale Ile-de-France à l'Innovation et aux Transferts de Technologie).

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gca.2009.12.022.

REFERENCES

Alvarez-Puebla R. A., Valenzuela-Calahorro C. and Garrido J. J. (2006) Theoretical study on fulvic acid structure, conformation and aggregation: a molecular modelling approach. *Sci. Total Environ.* **358**, 243–254.

Avena M. J., Koopal L. K. and van Riemsdijk W. H. (1999) Proton binding to humic acids: electrostatic and intrinsic interactions. *J. Colloid Interf. Sci.* **217**, 37–48.

Bartschat B. M., Cabaniss S. E. and Morel F. M. M. (1992) Oligoelectrolyte model for cation binding by humic substances. *Environ. Sci. Technol.* **26**, 284–294.

Benedetti M. F., Milne C. J., Kinniburgh D. G., van Riemsdijk W. H. and Koopal L. K. (1995) Metal ion binding to humic substances: application of the non-ideal competitive adsorption model. *Environ. Sci. Technol.* **29**, 446–457.

Benedetti M. F., van Riemsdijk W. H. and Koopal L. K. (1996) Humic substances considered as a heterogeneous Donnan gel phase. *Environ. Sci. Technol.* **30**, 1805–1813.

Buffle J. A. E. (1977) Les substances humiques et leurs interactions avec les ions minéraux. In *Conference Proccedings de la Commission d'Hydrologie Appliquée de l'A.G.H.T.M*, Université d'Orsay.

Cabaniss S. E. and McVey L. F. (1995) Aqueous infrared carboxylate absorbances: aliphatic monocarboxylates. *Spectrochim. Acta A* **51**, 2385–2395.

Christl I. and Kretzschmar R. (2001) Relating ion binding by fulvic and humic acids to chemical composition and molecular size. 1. Proton binding. *Environ. Sci. Technol.* **35**, 2505–2511.

Davis C. W. (1962) *Ion Association*. Butterworths.

Deshmukh A. P., Pacheco C., Hay M. B. and Myneni S. C. B. (2007) Structural environments of carboxyl groups in natural organic molecules from terrestrial systems. Part 2. 2D NMR spectroscopy. *Geochim. Cosmochim. Acta* **71**, 3533–3544.

De Wit J. C. M., van Riemsdijk W. H. and Koopal L. K. (1993) Proton binding to humic substances. 1. Electrostatic effects. *Environ. Sci. Technol.* **27**, 2005–2014.

Drosos M., Jerykiewicz M. and Deligiannakis Y. (2009) H-binding groups in lignite *vs.* soil humic acids: NICA-Donnan and spectroscopic parameters. *J. Colloid Interf. Sci.* **332**, 78–84.

Gondar D., Lopez R., Fiol S., Antelo J. M. and Arce F. (2005) Characterization and acid–base properties of fulvic and humic acids isolated from two horizons of an ombrotrophic peat bog. *Geoderma* **126**, 367–374.

Gustafsson J. P. (2001) Acid–base properties and metal complexation of humic substances with the Stockholm Humic Model. *J. Colloid Interf. Sci.* **244**, 102–112.

Hay M. B. and Myneni S. C. B. (2007) Structural environments of carboxyl groups in natural organic molecules from terrestrial systems. Part 1. Infrared spectroscopy. *Geochim. Cosmochim. Acta* **71**, 3518–3532.

Hepler L. G. (1963) Effects of substituents on acidities of organic acids in water: thermodynamic theory of Hammett equation. *J. Am. Chem. Soc.* **85**, 3089.

Hertkorn N., Benner R., Frommberger M., Schmitt-Kopplin P., Witt M., Kaiser K., Kettrup A. and Hedges J. I. (2006) Characterization of a major refractory component of marine dissolved organic matter. *Geochim. Cosmochim. Acta* **70**, 2990–3010.

Kinniburgh D. G. (1999) *FIT User Guide*. British Geological Survey, Keyworth, England.

Kinniburgh D. G., Milne C. J., Benedetti M. F., Pinheiro J. P., Filius J., Koopal L. K. and van Riemsdijk W. H. (1996) Metal ion binding by humic acid: application of the NICA-Donnan model. *Environ. Sci. Technol.* **30**, 1687–1698.

Kinniburgh D. G., van Riemsdijk W. H., Koopal L. K., Borkovec M., Benedetti M. F. and Avena M. J. (1999) Ion binding to natural organic matter: competition, heterogeneity, stoichiometry and thermodynamic consistency. *Colloids Surf. A* **151**, 147–166.

Koopal L. K., van Riemsdijk W. H., de Wit J. C. M. and Benedetti M. F. (1994) Analytical isotherm equations for multicomponent adsorption to heterogeneous surfaces. *J. Colloid Interf. Sci.* **166**, 51–60.

Koopal L. K., Saito T., Pinheiro J. P. and van Riemsdijk W. H. (2005) Ion binding to natural organic matter: general considerations and the NICA-Donnan model. *Colloids Surf. A* **265**, 40–54.

Lam B., Baer A., Alaee M., Lefebvre B., Moser A., Willimas A. and Simpson A. J. (2007) Major structural components in freshwater dissolved organic matter. *Environ. Sci. Technol.* **41**, 8240–8247.

Matynia A. (2009) Mécanisme de rétention du cuivre sur de la tourbe végétalisée. *Ph.D. Thesis*, University of Grenoble.

Marinsky J. A., Gupta S. and Schindler P. (1982) The interaction of Cu(II) ion with humic-acid. *J. Colloid Interf. Sci.* **89**, 401–411.

Marinsky J. A. and Ephraim J. H. (1986) A unified physicochemical description of the protonation and metal-ion complexation equilibria of natural organic-acids (humic and fulvic acids). 1. Analysis of the influence of polyelectrolyte properties on protonation equilibria in ionic media – fundamental concepts. *Environ. Sci. Technol.* **20**, 349–354.

Milne C. J., Kinniburgh D. G., De Wit J. C. M., van Riemsdijk W. H. and Koopal L. K. (1995) Analysis of proton binding by a peat humic-acid using a simple electrostatic model. *Geochim. Cosmochim. Acta* **59**, 1101–1112.

Milne C. J., Kinniburgh D. G. and Tipping E. (2001) Generic NICA-Donnan model parameters for proton binding by humic substances. *Environ. Sci. Technol.* **35**, 2049–2059.

Merdy P., Guillon E., Dumonceau J. and Aplincourt M. (2002) Characterisation of a wheat straw cell wall residue by various techniques. A comparative study with a synthetic and an extracted lignin. *Anal. Chim. Acta* **459**, 133–142.

Plaza C., Senesi N., Polo A. and Brunetti G. (2005) Acid–base properties of humic and fulvic acids formed during composting. *Environ. Sci. Technol.* **39**, 7141–7146.

Ren T. (1998) Substituent effects in dinuclear paddlewheel compounds: electrochemical and spectroscopic investigations. *Coord. Chem. Rev.* **175**, 43–58.

Ritchie J. D. and Perdue E. M. (2003) Proton-binding study of standard and reference fulvic acids, humic acids, and natural organic matter. *Geochim. Cosmochim. Acta* **67**, 85–96.

Saito T., Nagasaki S., Tanaka S. and Koopal L. K. (2005) Electrostatic interaction models for ion binding to humic substances. *Colloids Surf. A* **265**, 104–113.

Schulten H. R. and Schnitzer M. (1993) A state of the art structural concept for humic substances. *Naturwissenschaften* **80**, 29–30.

Smith S. M. and Martell A. E. (2004) Critically selected stability constants of metal complexes database. Standard reference database 46, version 8.0, NIST, Gaithersburg.

Stumm W. and Morgan J. J. (1996) *Aquatic Chemistry*, third ed. Wiley Interscience, Wiley, New York.

Stevenson F. J. (1982) *Humus Chemistry: Genesis, Composition, Reactions*. New York.

Tipping E. and Hurley M. A. (1992) A unifying model of cation binding by humic substances. *Geochim. Cosmochim. Acta* **56**, 3627–3641.

Tipping E. (1998) Humic ion-binding model VI: an improved description of the interactions of protons and metal ions with humic substances. *Aquat. Geochem.* **4**, 3–48.

Tipping E., Reddy M. M. and Hurley M. A. (1990) Modeling electrostatic and heterogeneity effects on proton dissociation from humic substances. *Environ. Sci. Technol.* **24**, 1700–1705.

Tipping E. (2002) *Cation Binding by Humic Substances*. Cambridge University Press, Cambridge.

Waksman S. A. (1938) *Humus: Origin, Chemical Compositions, and Importance in Nature*. Williams and Wilkins, Baltimore.

Westall J. C., Jones J. D., Turner G. D. and Zachara J. M. (1995) Models for association of metal ions with heterogeneous environmental sorbents. 1. Complexation of Co(II) by leonardite humic acid as a function of pH and $NaClO_4$ concentration. *Environ. Sci. Technol.* **29**, 951–959.

Associate editor: Dimitri Sverjensky

Electronic Annex

Semi-empirical proton binding constants for natural organic matter

Anthony Matynia[1,2], Thomas Lenoir[1,3], Benjamin Causse[4,5], Lorenzo Spadini[4], Thierry Jacquet[2], and Alain Manceau[1]

[1] Mineralogy & Environments Group, LGCA, Université Joseph Fourier and CNRS, 38041 Grenoble Cedex 9 France
[2] Phytorestore – Site et Concept, 7 impasse Milord, 75018 Paris, France
[3] Laboratoire Central des Ponts et Chaussées (LCPC), Route de Bouaye, BP 4129, 44341 Bouguenais cedex, France
[4] Environmental Geochemistry Group, LGIT, Université Joseph Fourier and CNRS, 38041 Grenoble cedex 9 France
[5] LSE-ENTPE, Université de Lyon, 69518 Vaulx-en-Velin cedex, France

1. Calculation of $\Delta\log K_{H,i}$

This calculation quantifies mesomeric and inductive effects of a substituent on the log $K_{H,i}$ of a six-carbon aromatic ring. Our approach, derived conceptually from linear free energy relations of Hammett (Hepler, 1963; Ren, 1998), was applied to describe how acidity constants vary with the type and position of the substituent on the ring. Adjusted values were calculated for carboxyl, hydroxyl, methoxy, methyl, amine, and thiol functional groups attached in ortho, meta, and para positions on the ring, and then combined to calculate the unknown proton dissociation constants of the reactive structural units (RSUs).

The additivity of pK values results from the relationship between the Gibbs free energy and the equilibrium constant :

A+B = C+D $\quad \Delta G_1 = G_C + G_D - G_A - G_B = -RT(\ln K_1)$

E+F = G+H $\quad \Delta G_2 = G_H + G_G - G_E - G_F = -RT(\ln K_2)$

$\Delta G_{1+2} = \Delta G_1 + \Delta G_2 = -RT(\ln K_1 + \ln K_2)$

1.1. Carboxyl group

The $\Delta\log K_{H,COOH}$ value for substituent A (A = hydroxyl, methoxy, and methyl) was obtained by subtracting the constant of benzoic acid ($\log K_{H,\,COOH}$= 4.01) from that of the A-substituted species (Table EA-1).

(i) Hydroxyl substituent. The $\log K_{H,\,COOH}$ of 2-, 3-, and 4-hydroxybenzoic acids are shifted by -1.21, -0.02, and 0.36 log units relative to benzoic acid, respectively. Consideration of chemically complex ligands, with known acidity constants, show that $\Delta\log K_{H,COOH}$ values can be approximated as additive. For example, log *K* for 2, 3, 4-trihydroxylbenzoic is 3.02, and the predicted value obtained by adding

$\Delta\log K_{H,COOH}$ for 2-, 3- and 4-hydroxybenzoic acid to log $K_{H,COOH}$ of benzoic acid is 4.01 - 1.21 - 0.02 + 0.36 = 3.14.

(ii) Methoxy (ph-O-CH_3) substituent. The log $K_{H,COOH}$ of 2-, 3-, and 4-methoxybenzoic acids are shifted by -0.14, -0.19 and 0.25 log units relative to benzoic acid, respectively. The additivity rule also applies for this substituent. For example, the acidity constant of 2,3-dimethoxybenzoic acid is 3.50, and the predicted value obtained by adding $\Delta\log K_{H,COOH}$ for 2- and 3- methoxybenzoic acid to log $K_{H,COOH}$ of benzoic acid is 4.01 - 0.14 - 0.19 = 3.68.

(iii) Methyl substituent. The log $K_{H,COOH}$ for 2- or 3- or 4-methylbenzoic acids are shifted by -0.10, 0.27, and 0.37 log units relative to benzoic acid, respectively. Predicted and experimental constants for benzoic acids with several -CH_3 substituents differ by more than 0.2 log units. From the cross-comparison with more complex ligands, the values retained for this substituent are -0.40, 0.15, and 0.25. With these values, 2,3,4,6- tetramethylbenzoic acid has an experimental constant of 3.47 and a predicted constant of 4.01 - 0.40 + 0.15 + 0.25 - 0.40 = 3.61. Deviation from the additivity rule for this substituent has little impact on the predicted acidity constants of HA and FA because most methyl-substituted benzoic acid reactive structural units are referenced in the NIST database.

(iv) Carboxyl substituent. Since di- and tri-carboxylates are polyacids, their mean acidities were used in the calculations. The shift for 1,2- and 1,3-di-carboxylic acids is 3.84 - 4.01 = -0.17 relative to benzoic acid, and for 1,4-di-carboxylic acid is 3.76 - 4.01 = - 0.25 log units. The validity of the additivity rule for a second carboxyl substituent can be verified only with 4-methylbenzene-1,2-dicarboxylic acid, because the NIST database contains only one acid with four dicarboxylates. Its average acidity constant is 4.15 and the predicted constant is 4.01 - 0.17 + 0.25 = 4.09. The known constants of the three tricarboxylic acid isomers (1,2,3; 1,2,4; 1,3,5) are 3.99, 3.71, and 3.74, and their $\Delta\log K_{H,COOH}$ values -0.02, -0.30, and -0.27 log units (Table EA-1).

1.2. Other functional groups (-OH, -NH_2 or –SH)

The $\Delta\log K_{H,OH}$, $\Delta\log K_{H,NH2}$, and $\Delta\log K_{H,SH}$ values were calculated by taking phenol (log $K_{H,OH}$ = 9.79), aniline (log $K_{H,NH2}$ = 4.64) and benzenethiol (log $K_{H,SH}$= 6.46) as references.

Table EA-1. $\Delta\log K_{H,i}$ values used for the calculation of proton binding constants by the chemical substituent approach.

Missing substituent	$\Delta\log K_{H,COOH}$	Missing substituent	$\Delta\log K_{H,OH}$
2-hydroxybenzoic acid	-1.21	2-methylphenol	0.30
3-hydroxybenzoic acid	-0.02	3-methylphenol	0.30
4-hydroxybenzoic acid	0.36	4-methylphenol	0.25
2-methoxybenzoic acid	-0.14	2-hydroxyphenol[a]	-0.53
3-methoxybenzoic acid	-0.19	3-hydroxyphenol[b]	0.39
4-methoxybenzoic acid	0.25	4-hydroxyphenol[b]	0.83
2-methylbenzoic acid	-0.40	2-carboxylicphenol[c]	3.61
3-methylbenzoic acid	0.15	3-carboxylicphenol	-0.17
4-methylbenzoic acid	0.25	4-carboxylicphenol	-0.81
Benzene-1,2-dicarboxylic acid	-0.17	2-methoxyphenol	0.19
Benzene-1,3-dicarboxylic acid	-0.17	3-methoxyphenol	0.07
Benzene-1,4-dicarboxylic acid	-0.25	4-methoxyphenol	0.17
Benzene-1,2,3-tricarboxylic acid	-0.02	2-mercaptobenzoic	-0.55

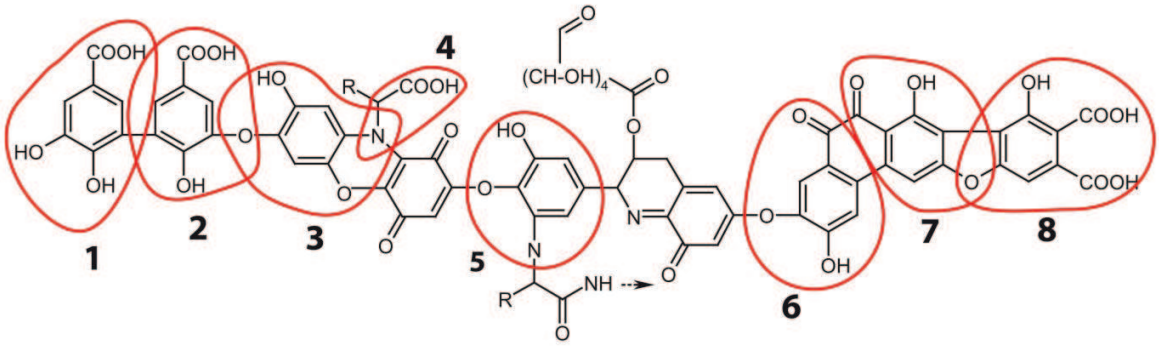
Benzene-1,2,4-tricarboxylic acid	-0.30	2-aminophenol	0.08
Benzene-1,3,5-tricarboxylic acid	-0.27	3-aminophenol	0.03
		1,2,4-trihydroxyphenol[d]	0.41
		2-acetylphenol	0.15
		4-acetylphenol	-1.94

Missing substituent	**$\Delta\log K_{H,SH}$**	**Missing substituent**	**$\Delta\log K_{H,NH2}$**
2-methoxybenzenethiol	0.43	2-hydroxyaniline	0.10
3-methoxybenzenethiol	-0.08	3-methylaniline	0.26
4-methoxybenzenethiol	0.32	2-methoxyaniline	0.08
3-carboxylicbenzenethiol	1.63	3-methoxyaniline	-0.26

[a] Only one binding constant has been considered, because the second is questionable (13.3).[b] Average of the two binding constants.[c] This value is questionable (occurs in parenthesis in the NIST database), thus the effect of adjacent COOH on the acidity of OH was dismissed.[d] Two constants considered, because the third is questionable.

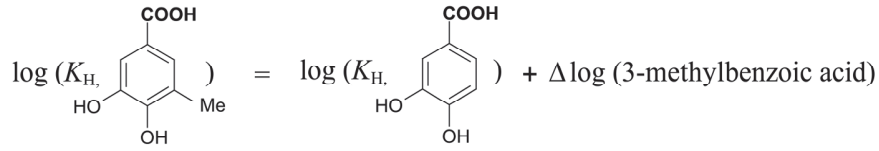
2. Calculations of log $K_{H,i}^{RSU}$ values

2.1. HA-S model



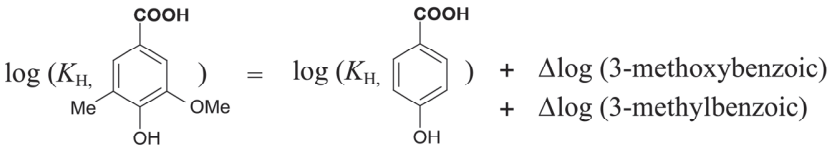
2.1.1. log $K_{H,COOH}$

RSU1



$$\log K_{H,COOH} = 4.30 + 0.15 = 4.45$$

RSU2

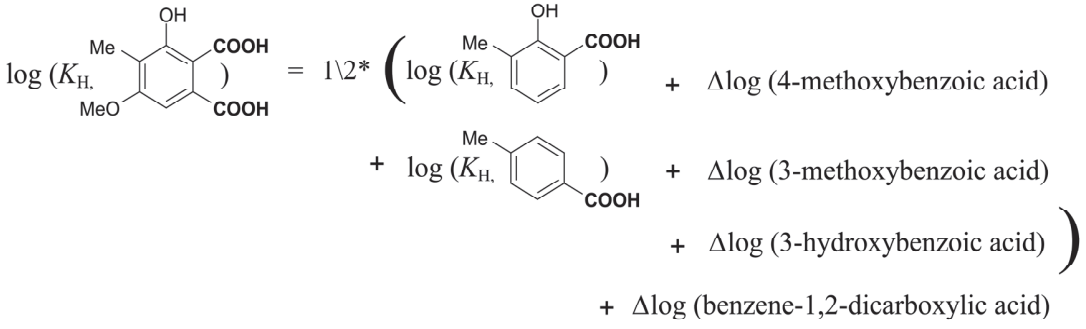


$$\log K_{H,COOH} = 4.37 - 0.19 + 0.15 = 4.33$$

RSU4

Same structure as glycine (2.33).

RSU8



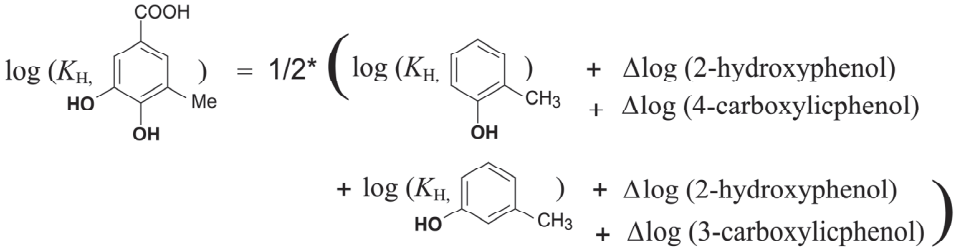
$$\log K_{\mathrm{H,COOH}} = \frac{2.83 + 0.25 + 4.37 - 0.19 - 0.02}{2} - 0.17 = 3.45$$

Average

$$\log K_{\mathrm{H,COOH}} = \frac{4.45 + 4.33 + 2.33 + 3.45}{4} = 3.64$$

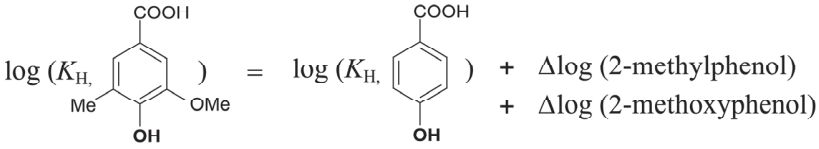
2.1.2. log $K_{\mathrm{H,Ph\text{-}OH}}$

RSU1



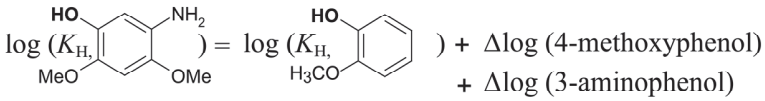
$$\log K_{\mathrm{H,Ph\text{-}OH}} = \frac{10.09 - 0.53 - 0.81 + 10.10 - 0.53 - 0.17}{2} = 9.07$$

RSU2



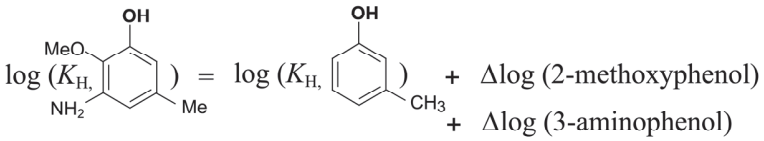
$$\log K_{\mathrm{H,Ph\text{-}OH}} = 8.98 + 0.30 + 0.19 = 9.47$$

RSU3



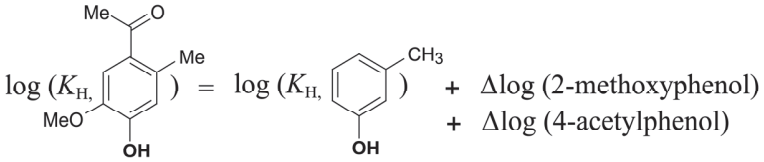
$$\log K_{\mathrm{H,Ph\text{-}OH}} = 9.98 + 0.17 + 0.03 = 10.18$$

RSU5



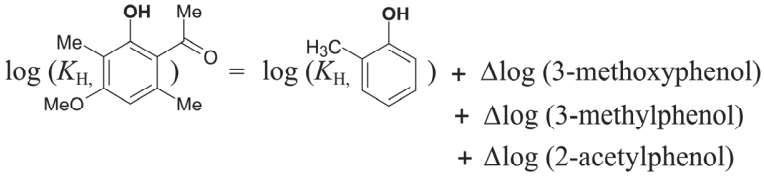
$$\log K_{H,Ph\text{-}OH} = 10.10 + 0.19 + 0.03 = 10.32$$

RSU6



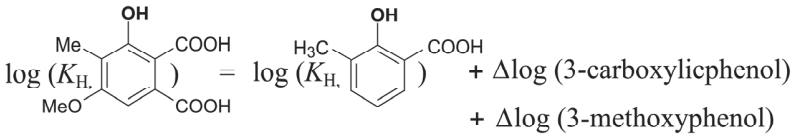
$$\log K_{H,Ph\text{-}OH} = 10.10 + 0.19 - 1.94 = 8.35$$

RSU7



$$\log K_{H,Ph\text{-}OH} = 10.09 + 0.07 + 0.30 + 0.15 = 10.61$$

RSU8



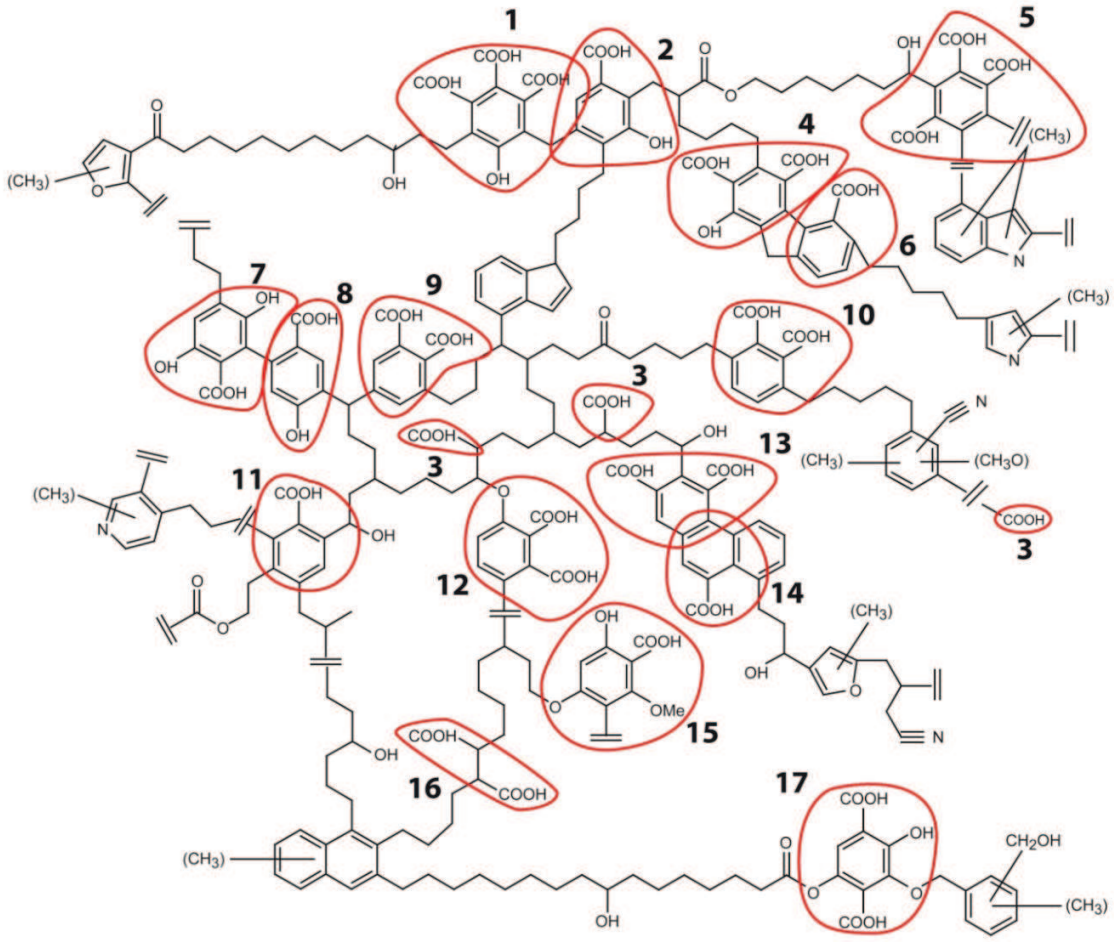
$$\log K_{H,Ph\text{-}OH} = 14.80 - 0.17 + 0.07 = 14.70$$

This value is too high (i.e., above the cutoff set at 10.80) and was dismissed in the calculation of the average.

Average

$$\log K_{H,Ph\text{-}OH} = \frac{9.07 + 9.47 + 10.18 + 13.32 + 8.35 + 10.61}{6} = 9.67$$

2.2. HA-SS model

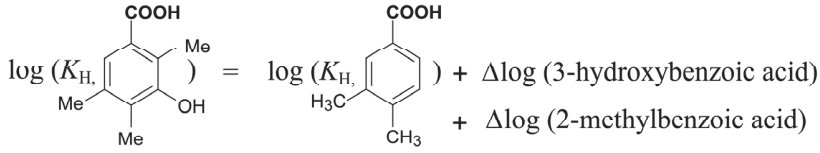


2.2.1. log $K_{H,COOH}$

RSU1

$$\log K_{H,COOH} = \frac{4.20 - 0.02 + 4.31 + 0.36 + 4.20 - 0.02}{3} - 0.02 = 4.32$$

RSU2

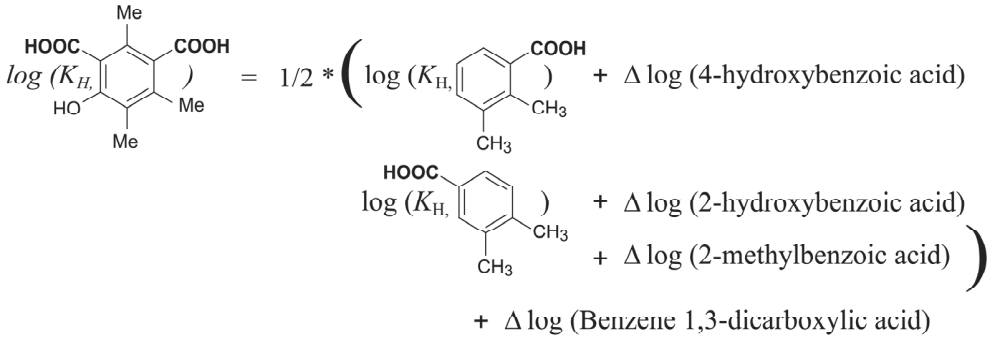


$$\log K_{H,COOH} = 4.41 - 0.40 - 0.02 = 3.99$$

RSU3

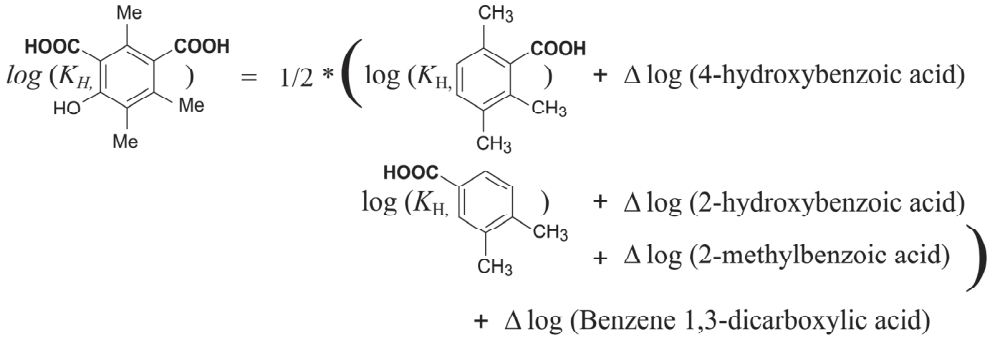
Same structure as acetic acid (4.56)

RSU4



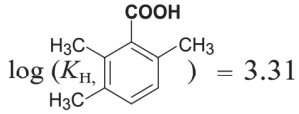
$$\log K_{H,COOH} = \frac{3.31 + 0.36 + 4.41 - 0.40 - 1.21}{2} - 0.17 = 3.07$$

RSU5

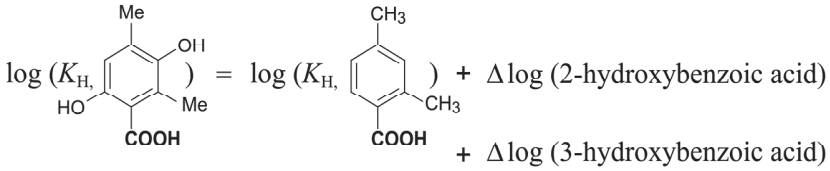


$$\log K_{H,COOH} = \frac{3.73 + 0.15 + 3.99 + 0.25 + 3.31}{3} - 0.30 = 3.51$$

RSU6

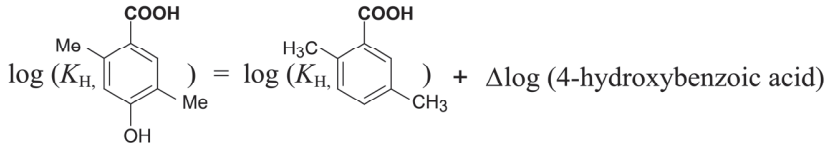


RSU7



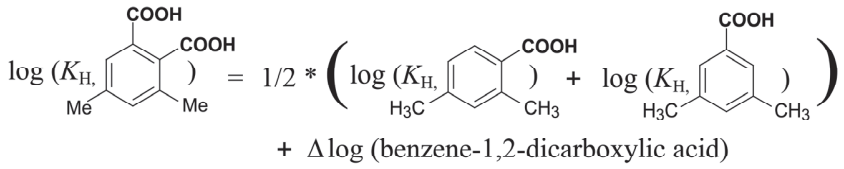
$$\log K_{H,COOH} = 4.20 - 1.21 - 0.02 = 2.97$$

RSU8



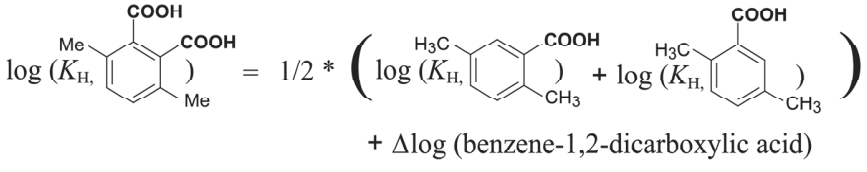
$$\log K_{H,COOH} = 3.99 + 0.36 = 4.35$$

RSU9



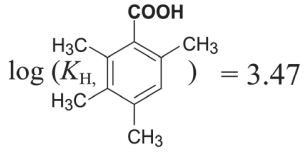
$$\log K_{H,COOH} = \frac{4.31 + 4.20}{2} - 0.17 = 4.08$$

RSU10

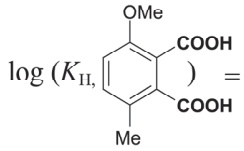
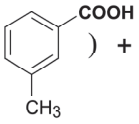


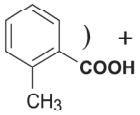
$$\log K_{H,COOH} = \frac{3.99 + 3.99}{2} - 0.17 = 3.82$$

RSU11



RSU12

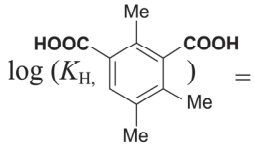
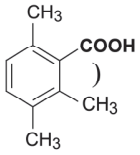
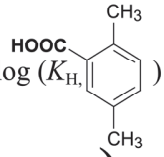
log () = 1/2 * (log () + Δlog (2-methoxybenzoic acid)

log () + Δlog (3-methoxybenzoic acid))

+ Δlog (benzene-1,2-dicarboxylic acid)

$$\log K_{H,COOH} = \frac{4.27 - 0.14 + 3.91 - 0.19}{2} - 0.17 = 3.76$$

RSU13

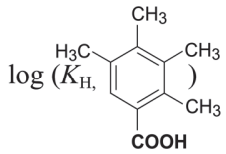
log () = 1/2 * (log () + log ()

+ Δlog (4-methylbenzoic acid))

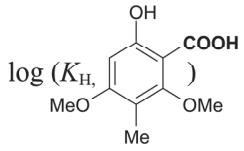
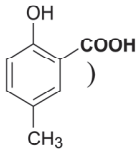
+ Δlog (benzene-1,3-dicarboxylic acid)

$$\log K_{H,COOH} = \frac{3.31 + 3.99 + 0.25}{2} - 0.17 = 3.61$$

RSU14

log () = 4.00

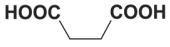
RSU15

log () = log () + Δlog (2-methoxybenzoic acid)

+ Δlog (4-methoxybenzoic acid)

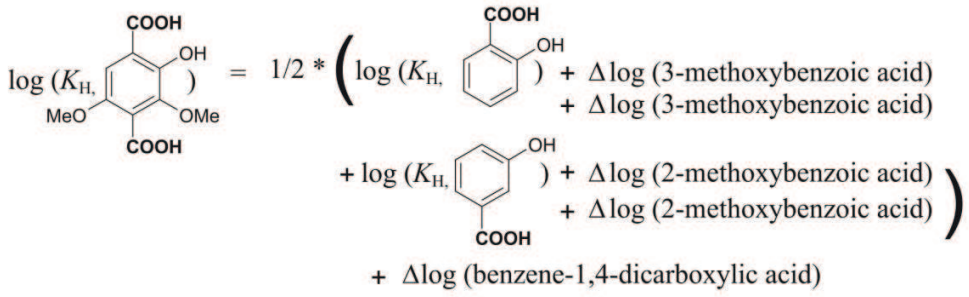
$$\log K_{H,COOH} = 2.88 - 0.14 + 0.25 = 2.99$$

RSU16



$$\log K_{H,COOH} = \frac{5.24 + 3.99}{2} = 4.62$$

RSU17



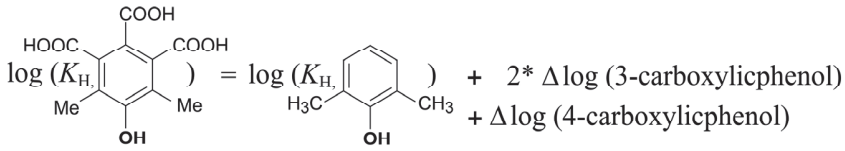
$$\log K_{H,COOH} = \frac{2.80 - 0.19 - 0.19 + 3.99 - 0.14 - 0.14}{2} - 0.25 = 2.82$$

Average

$$\log K_{H,COOH} = \frac{4.32 + 3.99 + 4.56 \times 3 + 4.62 + 3.07 + 3.51 + 3.31 + 2.97 + 4.35 + 4.08 + 3.82 + 3.47 + 3.76 + 3.61 + 4.00 + 2.99 + 2.82}{19} = 3.81$$

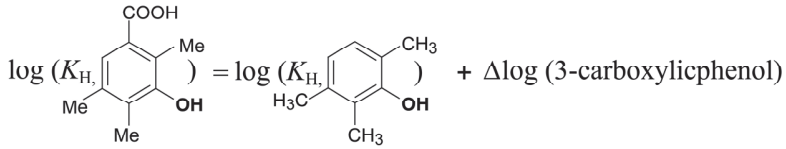
2.2.2. log $K_{H,Ph\text{-}OH}$

RSU1



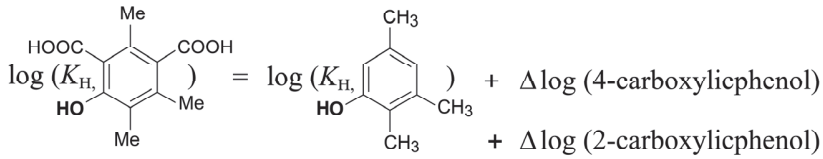
$$\log K_{H,Ph\text{-}OH} = 10.62 - 2 \times 0.17 - 0.81 = 9.47$$

RSU2



$$\log K_{H,Ph\text{-}OH} = 10.89 - 0.17 = 10.72$$

RSU4

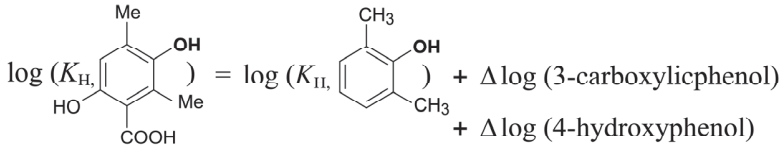


$$\log K_{H,Ph\text{-}OH} = 10.67 - 0.81 + 3.61 = 13.47$$

This value was dismissed.

RSU7

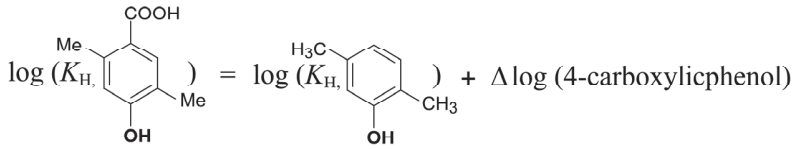
The effect of adjacent COOH on the acidity of OH was not considered.



$$\log K_{\mathrm{H,Ph\text{-}OH}} = 10.62 - 0.17 + 0.83 = 11.28$$

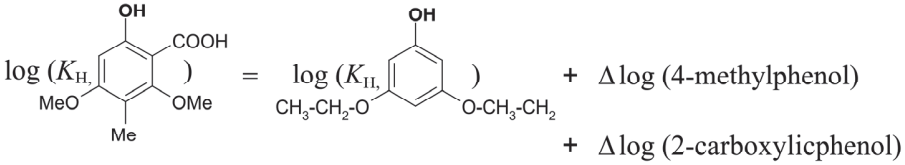
This value was dismissed.

RSU8



$$\log K_{\mathrm{H,Ph\text{-}OH}} = 10.60 - 0.81 = 9.79$$

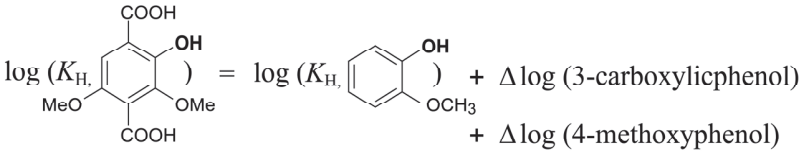
RSU15



$$\log K_{\mathrm{H,Ph\text{-}OH}} = 9.37 + 0.25 + 3.61 = 13.23$$

This value was dismissed.

RSU17

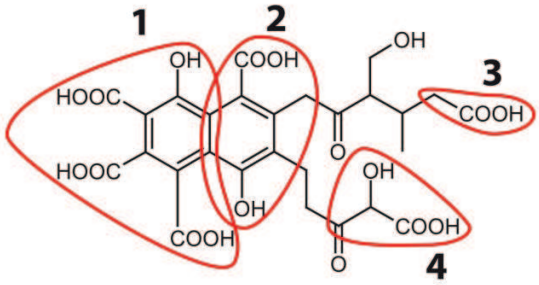


$$\log K_{\mathrm{H,Ph\text{-}OH}} = 9.98 - 0.17 + 0.17 + 3.61 = 13.59$$

Average

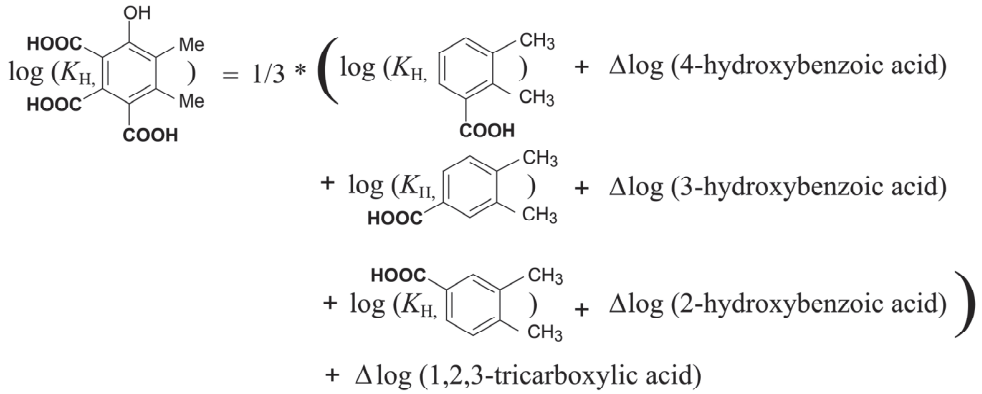
$$\log K_{\mathrm{H,Ph\text{-}OH}} = \frac{9.47 + 10.72 + 9.79}{3} = 9.99$$

2.3. FA-B model



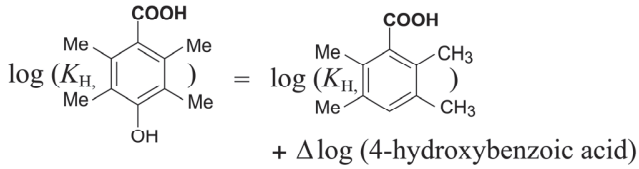
2.3.1. log $K_{H,COOH}$

RSU1



$$\log K_{H,COOH} = \frac{3.73 + 0.36 + 4.41 - 0.02 + 4.41 - 1.21}{3} - 0.02 = 3.87$$

RSU2



$$\log K_{H,COOH} = 3.40 + 0.36 = 3.76$$

RSU3

Same structure as acetic acid (4.56).

RSU4

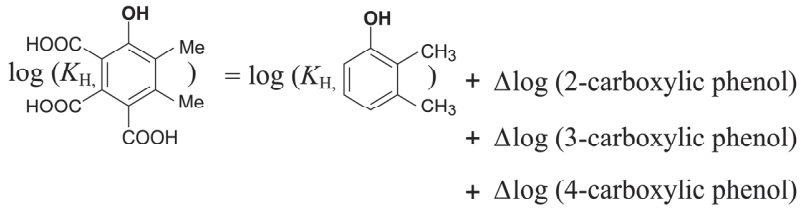
Same structure as lactic acid (3.67).

Average

$$\log K_{H,COOH} = \frac{3.87 + 3.76 + 4.56 + 3.67}{4} = 3.97$$

2.3.2. log $K_{H,Ph\text{-}OH}$

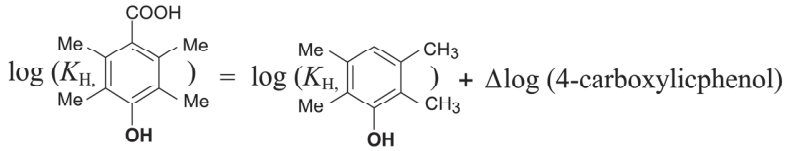
RSU1



$$\log K_{H,Ph\text{-}OH} = 10.54 + 3.61 - 0.17 - 0.81 = 13.17$$

This value was dismissed.

RSU2

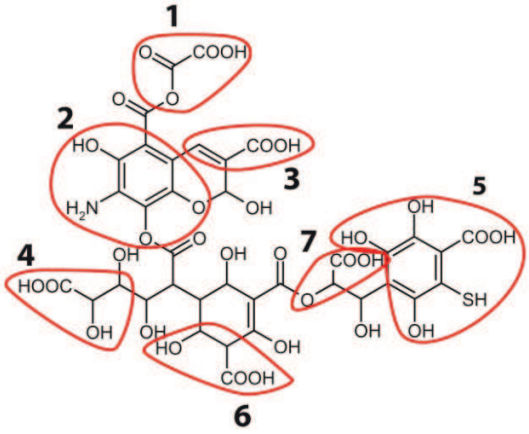


$\log K_{\mathrm{H,Ph\text{-}OH}} = 10.89 - 0.81 = 10.08$

Average

$\log K_{\mathrm{H,Ph\text{-}OH}} = 10.08$

2.4. FA-A model



2.4.1. log $K_{\mathrm{H,COOH}}$

RSU1

Structure similar to pyruvic acid (2.26).

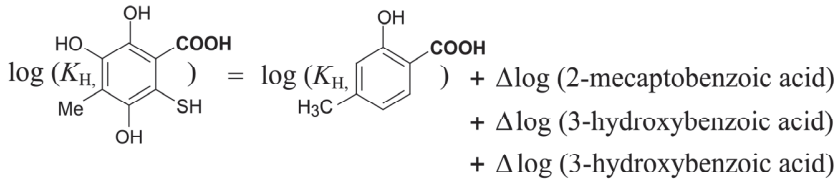
RSU3

Same structure as acrylic acid (4.09)

RSU4

Same structure as lactic acid (3.67).

RSU5



$\log K_{\mathrm{H,COOH}} = 2.96 - 2 \times 0.02 - 0.55 = 2.37$

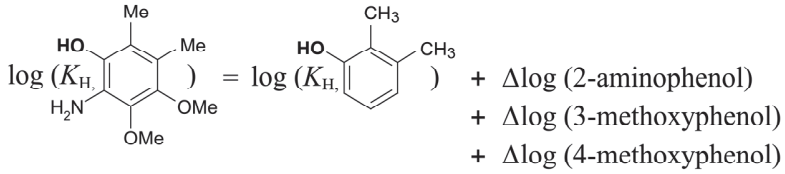
RSU6

Same structure as 3-hydroxypropanoic acid (4.40)

RSU7
Same structure as methoxyacetic acid (3.32)

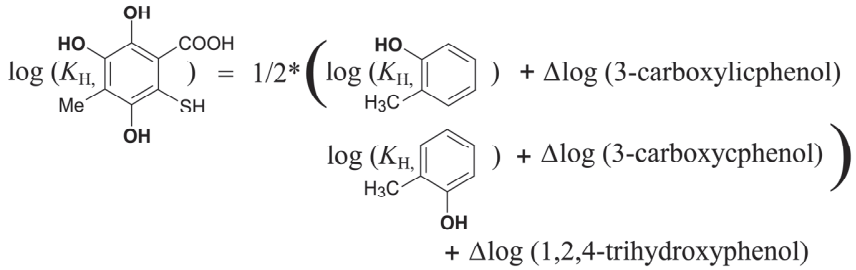
2.4.2. log $K_{H, Ph\text{-}OH}$

RSU2



$$\log K_{H,Ph\text{-}OH} = 10.54 + 0.08 + 0.07 + 0.17 = 10.86$$

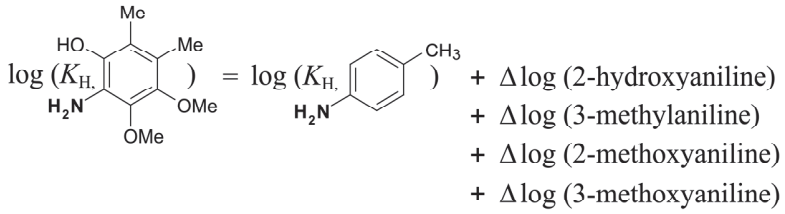
RSU5
The acidity of SH could not be calculated because some of the binding constants are unknown. The OH group next to COOH deprotonates at pH ≥ 13.



$$\log K_{H,Ph\text{-}OH} = \frac{10.09 - 0.17 + 10.09 - 0.17}{2} + 0.41 = 10.33$$

2.4.3. log $K_{H, NH2}$

RSU2

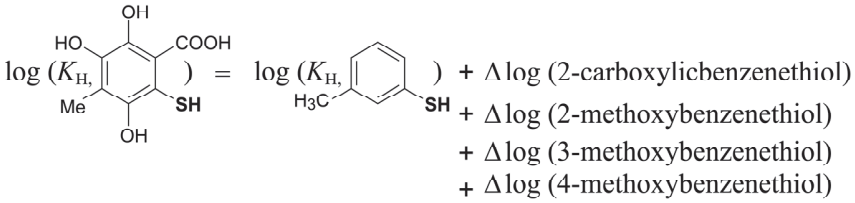


$$\log K_{H,NH_2} = 5.09 + 0.10 + 0.26 + 0.08 - 0.26 = 5.27$$

This value was included in the calculation of the average value of log $K_{H,COOH}$.

2.4.4. log $K_{H, SH}$

RSU5



$$\log K_{H,SH} = 6.66 + 1.63 + 0.43 - 0.08 + 0.32 = 8.96$$

The OH groups were replaced by OMe, because the constants of the methylated species are known. The average value was included in the total calculation of log $K_{H,Ph\text{-}OH}$.

Average

$$\log K_{H,COOH} = \frac{2.26 + 4.40 + 4.09 + 3.67 + 3.32 + 2.37 + 5.27}{7} = 3.63$$

$$\log K_{H,Ph\text{-}OH} = \frac{10.33 + 8.96}{2} = 9.65$$

3. References

Hepler L.G. (1963) Effects of substituents on acidities of organic acids in water: thermodynamic theory of Hammett equation. *J. Am. Chem. Soc*., **85**, 3089.

Ren T. (1998) Substituent effects in dinuclear paddlewheel compounds: electrochemical and spectroscopic investigations. *Coord. Chem. Rev.*, **175**, 43-58.

ANNEXE 3

Number of Independant Parameters in the Potentiometric Titration of Humic Substances

Langmuir
Article
pubs.acs.org/Langmuir
© 2010 American Chemical Society

Number of Independent Parameters in the Potentiometric Titration of Humic Substances

Thomas Lenoir[†,‡] and Alain Manceau*[,†]

[†]Mineralogy & Environments Group, LGCA, Université Joseph Fourier and CNRS, 38041 Grenoble Cedex 9, France, and [‡]Laboratoire Central des Ponts et Chaussées (LCPC), Route de Bouaye, BP 4129, 44341 Bouguenais Cedex, France

Received September 10, 2009. Revised Manuscript Received November 16, 2009

With the advent of high-precision automatic titrators operating in pH stat mode, measuring the mass balance of protons in solid-solution mixtures against the pH of natural and synthetic polyelectrolytes is now routine. However, titration curves of complex molecules typically lack obvious inflection points, which complicates their analysis despite the high-precision measurements. The calculation of site densities and median proton affinity constants (pK) from such data can lead to considerable covariance between fit parameters. Knowing the number of independent parameters that can be freely varied during the least-squares minimization of a model fit to titration data is necessary to improve the model's applicability. This number was calculated for natural organic matter by applying principal component analysis (PCA) to a reference data set of 47 independent titration curves from fulvic and humic acids measured at $I = 0.1$ M. The complete data set was reconstructed statistically from pH 3.5 to 9.8 with only six parameters, compared to seven or eight generally adjusted with common semi-empirical speciation models for organic matter, and explains correlations that occur with the higher number of parameters. Existing proton-binding models are not necessarily overparametrized, but instead titration data lack the sensitivity needed to quantify the full set of binding properties of humic materials. Model-independent conditional pK values can be obtained directly from the derivative of titration data, and this approach is the most conservative. The apparent proton-binding constants of the 23 fulvic acids (FA) and 24 humic acids (HA) derived from a high-quality polynomial parametrization of the data set are $pK_{H,COOH}(FA) = 4.18 \pm 0.21$, $pK_{H,Ph-OH}(FA) = 9.29 \pm 0.33$, $pK_{H,COOH}(HA) = 4.49 \pm 0.18$, and $pK_{H,Ph-OH}(HA) = 9.29 \pm 0.38$. Their values at other ionic strengths are more reliably calculated with the empirical Davies equation than any existing model fit.

Introduction

Potentiometric titration is the method of choice for measuring equilibrium proton-binding constants in colloids and materials surface science, biology, and biochemistry. In environmental science, the titration method is commonly used to characterize the acid−base properties of natural organic matter (NOM). NOM is a complex, heterogeneous assemblage that includes many polyelectrolyte groups, such as carboxyl (COOH), phenolic (Ph−OH), amine, sulfhydryl, phosphate, and alcohol functional groups.[1] The first two types of ligands predominate and are the only ones considered in the modeling of titration data. Their amounts per unit mass of NOM, expressed as the total concentration of protons denoted by Q_{H1} and Q_{H2}, and their proton-binding constants K_{H1} and K_{H2} are the main intrinsic properties of a humic substance derived from a titration experiment. These two properties are also the most important because they determine the ion-binding capacity of NOM for hard and intermediate metals (e.g., Al(III), Fe(III), Cu(II), and Zn(II)).[2] For example, pK_H values are used to determine the extent of competition for binding sites between metal cations and protons at a given pH.

Proton-binding constants for fulvic (FA) and humic (HA) acids vary by at least one order of magnitude ($2.6 \le pK_{H1}(FA) \le 3.5$; $3.1 \le pK_{H1}(HA) \le 4.1$; $7.0 \le pK_{H2}(FA) \le 9.4$; and $7.7 \le pK_{H2}(HA) \le 8.9$) (Table 1). The large variability in pK_{Hi} obtained by titration either is real and results from the variability in composition and functionality of the two types of humic materials or is within the uncertainty and accuracy of the titration method, in which case model fits would not necessarily provide a realistic chemical description of the binding properties of humic substances. As an example, a variation of one pK unit on a data set causes an error in the acidity constant of carboxylic-type groups ($pK_{H1} \sim 3$) of $\sim^1/_3 = 33\%$ in log units. If we can determine the number of independent parameters that can be fit to the titration data, then we may be able to clarify the reason for the variability in pK_{Hi} values and advance the interpretation of acid−base potentiometric measurements of complex macromolecular NOM. We addressed this fundamental but still unresolved problem by using a principal component analysis (PCA) algorithm to calculate the number of uncorrelated abstract components that maximize the variance in a large data set of independent titration curves.[3,4] The number of statistically significant components obtained by this correlation-based analysis is equal to the number of independent parameters needed to reconstruct titration curves. To help follow the presentation and discussion of the PCA results, a brief technical description of the titration experiments and data analysis is first given below.

Derivation of Proton-Binding Parameters from Titration Data

Titration curves are usually acquired between pH 3 and 10 because of the difficulty in precisely measuring the variation in acidity at lower and higher pH and the risk of altering the original material at extreme pH.[5] Over this pH range, the distribution of

*Corresponding author. E-mail: alain.manceau@obs.ujf-grenoble.fr.

(1) Takács, M.; Alberts, J. J.; Egeberg, P. K. *Environ. Intern.* **1999**, *25*, 315–323.
(2) Smith, D. S.; Bella, R. A.; Kramerb, J. R. *Compar. Biochem. Phys.* **2002**, *C133*, 65–74.
(3) Malinowski, E. R. *Anal. Chem.* **1977**, *49*, 612–617.
(4) Malinowski, E. R. *Factor Analysis in Chemistry*, 2nd ed.; Wiley: New York, 1991.
(5) Santos, E. B. H.; Esteves, V. I.; Rodrigues, J. P. C.; Duarte, A. C. *Anal. Chim. Acta* **1999**, *392*, 333–341.

Table 1. Compilation of Intrinsic pK_{Hi} Values for Humic Materials as Determined by Potentiometric Titration

		pK_{H1} (COOH groups)		pK_{H2} (Ph−OH groups)	
electrostatic model	humics	mean	std dev	mean	std dev
NICA-Donnan	FA[a]	2.65	0.43	8.60	1.06
	HA[a]	3.09	0.51	7.98	0.96
	FA[b]	2.83	0.36	7.02	0.18
	HA[b]	3.74	1.09	7.68	0.60
	HA[c]	3.62	0.19	8.54	0.48
model VI[d]	FA	3.20	0.13	9.40	0.78
	HA	4.10	0.16	8.80	0.23
Stockholm humic model[e]	FA	3.50	0.40	8.75	0.30
	HA	4.10	0.20	8.95	0.15

[a] Milne et al.[21] [b] Plaza et al.[44] [c] Drosos et al.[45] [d] Tipping.[15] [e] Gustafsson.[18]

acid−base reactive sites in NOM is fundamentally bimodal, with the first maximum of the derivative obtained at pH 3−5 from the deprotonation of carboxylic-type groups and the second obtained at pH 8−10 from phenolic-type groups (Figure 1a).[6] Because protons tend to be electrostatically retained by COO^- and $Ph-O^-$ groups, more base has to be added to remove them from the negatively charged NOM than from a neutral surface. Consequently, inflection points on titration curves are apparent pK values (pK_{Hiapp}) that are shifted to higher pH relative to intrinsic binding constants (pK_{Hi}) by a value not directly accessible by experiment but instead estimated by fitting the data to an electrostatic model (Table 1, Figure 1b,c). In the standard Donnan-type equilibrium model originally developed by Marinsky and coworkers,[7,8] the ionic strength dependence of the surface charge is captured simply with a single electrostatic interaction parameter b.[9−12] Similarly, for two reasons the concentrations of carboxylic- (Q_{H1}) and phenolic-type (Q_{H2}) groups are not obtained directly from titration measurements but instead by fitting the data to an ion-binding model such as the NICA model.[6,13−15] First, the low-pH plateau of titration curves, which defines the initial charge of the reactant, is below pH 3, and the high-pH plateau, which defines the final protonation state ($Q_{H1} + Q_{H2}$) of NOM, is above pH 10. As Westall et al.[16] stated, the "spectral window" is wider than the "data window". Second, the final protonation state of carboxylic-type groups (Q_{H1}), which is also the initial protonation state of phenolic-type groups, is masked by minor circumneutral groups with pK_H values between 5 and 8, such as phosphate H_2PO_4, amine NH_2, thiol R−SH, and anilium $C_6H_5NH_3^+$.[1]

In total, a minimum of seven parameters are required to fit a single titration curve of NOM to the combined NICA-Donnan (N-D) model: two pK_{Hi} values, two affinity distribution parameters (m_{Hi}) accounting for the variability in binding strength (i.e., in equilibrium constants), two Q_{Hi} values, and b.[17] When the

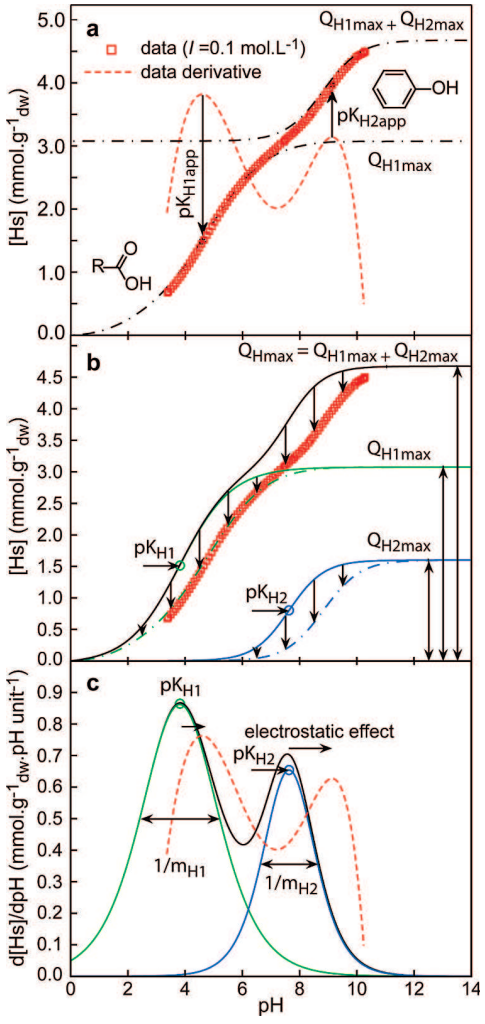


Figure 1. Main physicochemical parameters for natural organic matter derived from a titration curve obtained by measuring the number of protons released in solution [Hs] per gram of dry matter. (a) The red dotted line is the first derivative. Inflection points are the apparent (also called conditional) dissociation constants for carboxylic- and phenolic-type sites. The dashed lines in black are the optimal bimodal (R−COOH, Ph−OH) fit obtained with the N-D FIT code.[11] (b) Same simulation as previously but without an electrostatic effect. The new pK_H values are intrinsic values. The inclusion of an electrostatic correction does not change Q_{Himax}. Vertical arrows point to the direction in which the deprotonation curves shift because of electrostatic effects. (c) First derivatives of the previous deprotonation curves and modeling of the pseudo-normal affinity distribution with two Gaussians of $\sigma = 1/m_{Hi}$ width, as in the NICA model.[13,17]

(6) Dudal, Y.; Gérard, F. *Earth Sci. Rev.* **2004**, *66*, 199–216.
(7) Marinsky, J. A.; Ephraim, J. H. *Environ. Sci. Technol.* **1986**, *20*, 349–354.
(8) Marinsky, J. A.; Gupta, S.; Schindler, P. *J. Colloid Interface Sci.* **1982**, *89*, 412–426.
(9) Benedetti, M. F.; Van Riemsdijk, W. H.; Koopal, L. K. *Environ. Sci. Technol.* **1996**, *30*, 1805–1813.
(10) Kinniburgh, D. G.; Milne, C. J.; Benedetti, M. F.; Pinheiro, J. P.; Filius, J.; Koopal, L. K.; Van Riemsdijk, W. H. *Environ. Sci. Technol.* **1996**, *30*, 1687–1698.
(11) Kinniburgh, D. G. *FIT User Guide*; British Geological Survey: Keyworth, England, 1999.
(12) Avena, M. J.; Koopal, L. K.; Van Riemsdijk, W. H. *J. Colloid Interface Sci.* **1999**, *217*, 37–48.
(13) Koopal, L. K.; Van Riemsdijk, W. H.; De Wit, J. C. M.; Benedetti, M. F. *J. Colloid Interface Sci.* **1994**, *166*, 51–60.
(14) Benedetti, M. F.; Milne, C. J.; Kinniburgh, D. G.; Van Riemsdijk, W. H.; Koopal, L. K. *Environ. Sci. Technol.* **1995**, *29*, 446–457.
(15) Tipping, E. *Aquat. Geochem.* **1998**, *4*, 3–47.
(16) Westall, J. C.; Jones, J. D.; Turner, G. D.; Zachara, J. M. *Environ. Sci. Technol.* **1995**, *29*, 951–959.
(17) Koopal, L. K.; Saito, T.; Pinheiro, J. P.; Riemsdijk, W. H. V. *Colloids Surf., A* **2005**, *265*, 40–54.

initial charge is treated as an adjustable parameter,[10] the total number is eight. In model VI[15] and the Stockholm humic model,[18] Q_{H2} is deduced from Q_{H1}, thus reducing the number of degrees of freedom by one. Because titration curves for humic substances are rather featureless, increasing monotonically with pH, and the range of data is at best over seven pH units, the question arises as to the number of independent parameters that can be uniquely fit to the data.[17,19,20] As will be shown below by PCA, the limit is six but may be even lower if the data window is smaller.

PCA Theory

PCA is an eigenvector-based multivariate analysis tool that reduces a set of overdetermined vectors (titration curves) to a subset of linear independent orthogonal vectors, called principal or abstract components (PCs), needed to describe the complete set of data within the error. The conversion from titration curves to PCs can be expressed as the matrix equation $\mathbf{D} = \mathbf{EV^{1/2}W^T}$

$$\underbrace{\begin{bmatrix} data_{11} & \cdots & data_{1n} \\ \vdots & \ddots & \vdots \\ data_{m1} & \cdots & data_{mn} \end{bmatrix}_{u*v}}_{data} = \underbrace{\begin{bmatrix} e_{11} & \cdots & e_{1n} \\ \vdots & \ddots & \vdots \\ e_{m1} & \cdots & e_{mn} \end{bmatrix}}_{e_{i(1\le i\le n)}} \times \underbrace{\begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix}^{1/2}}_{\lambda_i} \times \underbrace{\begin{bmatrix} \rho_{e_1,data_1} & \cdots & \rho_{e_1,data_n} \\ \vdots & \ddots & \vdots \\ \rho_{e_n,data_1} & \cdots & \rho_{e_n,data_n} \end{bmatrix}}_{correlations}$$

where $\mathbf{D}(m \times n)$ is the data matrix whose columns are titration curves and rows are data points (i.e., [Hs] values in Figure 1a), $\mathbf{E}(m \times n)$ is a column-orthogonal matrix (eigenvectors), $\mathbf{V}(n \times n)$ is a diagonal matrix (eigenvalues λ), and $\mathbf{W^T}(n \text{ x } n)$ is the transpose of an $n \times n$ orthogonal matrix. The eigenvectors are the PCs, and the eigenvalues represent variance. They are used to rank PCs according to their importance in reproducing data. Thus, the result of applying PCA to titration data is a set of orthogonal PCs equal to the number of curves (n), and the analysis seeks the number of significant PCs ($j < n$) that is necessary and sufficient to reconstruct all data within error:

$$rec(data)_{jpc} = \sum_{i=1}^{j} \lambda_i e_i \rho_{e_i,\, data}$$

Because j is the number of uncorrelated linear combinations of PCs that maximizes the variance in the data set and hence explains the variation among all observables, its value represents the number of independent parameters contained in a titration curve.

The fit quality of the reconstructions was evaluated with the normalized sum-square ($NSS_{ipc} = \sum_i (y_{exp} - y_{fit})^2/y_{exp}^2$) parameter (Figure S1 in Supporting Information). Because NSS decreases when the number of component fits (i) increases, a best-fit criterion had to be defined to estimate j. We considered that $i = j$ when $NSS_{ipc} \approx NSS_{data}$, that is, when the precision of the reconstruction was comparable to the dispersion of the data points. In fact, the reconstruction by PCA can be seen as a noise-filtering technique because the least important components ($i > j$) represent noise and other errors.

Analytical Treatment

Statistically, the precision of j increases with n and m, that is, with the total number of data points. To maximize n and in the meantime to be able to discuss the PCA results in the context of proton-binding constants from the literature, the extended data set analyzed by Milne et al.[21] was chosen for this study. For consistency, the original names of the titration data were preserved (i.e., HHx and FHx, where the first letter stands for humic (H) or fulvic (F) acid, the second letter stands for a proton, and x stands for the sample number). Because all curves were not measured over the same pH interval, there is a trade off between optimizing n and m. Enlarging the common pH interval by extrapolation of the data points is not an option because the statistical analysis would lose robustness and is likely to compromise the results.[4] The pH interval retained is [3.5−9.8] and comprises 23 FA and 24 HA curves all at $I = 0.1$ M (Figures S2 and S3 in Supporting Information). No data that satisfied the two conditions were deliberately omitted. Replicate measurements (denoted HH/FHx_y) were considered to be independent. The validity of this hypothesis was verified a posteriori. PCA was conducted on three data subsets (FA, HA, and FA + HA) over two pH intervals ([3.5−5.5] and [3.5−9.8]) for each group. The basis data were interpolated on two uniform grids starting at pH 3.5:

- ΔpH = 0.02 for 3.5 ≤ pH ≤ 5.5 (101 points)
- ΔpH = 0.1 for 3.5 ≤ pH ≤ 9.8 (62 points)

The primary aim of the parametrization is to obtain the best possible analytical expression for the titration data, this unified representation being necessary for statistical analysis. Although Nederlof et al.[22] recommended using a spline function to parametrize titration data on a common pH grid, a polynomial interpolation was preferred for two reasons. First, polynomials are more suitable for PCA because they can be expressed as a sum of orthogonal components. Second, pK_{Hiapp} is calculated simply by derivation of the polynomial. Finally, all titration curves were translated to the same origin of [3.5, 0.0] before numerical analysis. Thus, the real number of independent parameters is $j + 1$. Interpolations were performed with MATLAB 7 and PCA with the homemade Labview software described in Manceau et al.[23]

The normalized sum-square difference between data and the polynomial fit (NSS_{inter}) became quite sensitive to experimental errors above the sixth order (Tables S1 and S2). Therefore, this order was retained for the interpolation (i.e., $NSS_{inter} = NSS_{data}$). Upon inspecting Tables S1 and S2 and Figures S2 and S3, the reader may easily convince himself that neither unwanted nor unacceptable errors were introduced by these interpolations, thus possible errors are negligible. The precision of [Hs] derived from NSS_{inter} varies from 10^{-7} to 10^{-5} (Table S2), and the uncertainty in pK_{H1app} deduced from these values is 0.057 ± 0.040 for FA and 0.026 ± 0.014 for HA. These calculated uncertainties are similar to experimental errors reported in the literature.[24,25] In addition, the pK_{H1app} values calculated for all FA and all HA followed a normal distribution, as shown numerically in Table S3 and graphically in Figure 2. This result provides a stringent test of the quality of our polynomial parametrization and also demonstrates

(18) Gustafsson, J. P. *J. Colloid Interface Sci.* **2001**, *244*, 102–112.
(19) Kinniburgh, D. G.; Van Riemsdijk, W. H.; Luuk, K.; Koopal, C.; Borkovec, M.; Benedetti, M. F.; Avena, M. J. *Colloids Surf., A* **1999**, *151*, 147–166.
(20) Christl, I.; Milne, C. J.; Kinniburgh, D. G.; Kretzschmar, R. *Environ. Sci. Technol.* **2001**, *35*, 2512–2517.
(21) Milne, C. J.; Kinniburgh, D. G.; Tipping, E. *Environ. Sci. Technol.* **2001**, *35*, 2049–2059.
(22) Nederlof, M. M.; Van Riemsdijk, W. H.; Koopal, L. K. *Environ. Sci. Technol.* **1994**, *28*, 1037–1047.
(23) Manceau, A.; Marcus, M. A.; Tamura, N. Quantitative Speciation of Heavy Metals in Soils and Sediments by Synchrotron X-ray Techniques. In *Applications of Synchrotron Radiation in Low-Temperature Geochemistry and Environmental Science*; Fenter, P. A., Rivers, M. L., Sturchio N. C., Sutton, S. R., Eds.; Reviews in Mineralogy and Geochemistry; Geochemical Society and Mineralogical Society of America: Washington, DC, 2002; Vol. *49*, pp 341−428.
(24) Cabaniss, S. E.; Mcvey, I. F. *Spectrochim. Acta, Part A* **1995**, *51*, 2385–2395.
(25) Koort, E.; Herodes, K.; Pihl, V.; Leito, I. *Anal. Bioanal. Chem.* **2004**, *379*, 720–729.

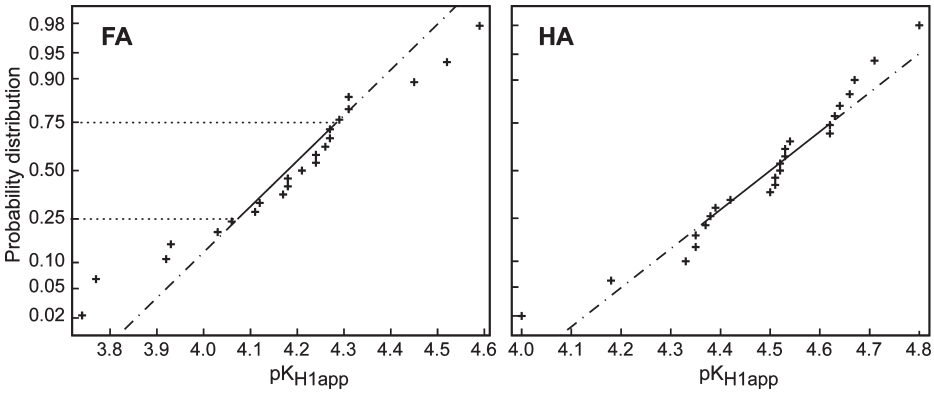


Figure 2. Probability distribution of pK_{H1app} values represented in the form of a cumulative distribution function. The lines are the cumulative functions for a random variable (i.e., normal distribution) with mean values of 4.18 and $\sigma = 0.21$ (FA = FH subdata set) and 4.49 and $\sigma = 0.18$ (HA = HH subdata set). The straight lines represent the middle half of the distribution or the interquartile range, defined as the spread in the distribution between the first and third quartiles (dashed-dotted lines) with respect to the median. For the most part, the pK_{H1app} values follow a normal distribution, meaning that all titration curves are independent data.

Table 2. PCA Results in the [3.5−5.5] pH Interval

FA	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}
min	2.64×10^{-7}	5.77×10^{-6}	7.75×10^{-8}	1.30×10^{-11}
max	8.30×10^{-5}	2.30×10^{-3}	1.76×10^{-4}	4.93×10^{-7}
$NSS_{inter} < NSS_{jpc}$		22	16	0
HA	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}
min	1.67×10^{-7}	2.08×10^{-5}	3.15×10^{-7}	1.54×10^{-9}
max	2.49×10^{-4}	9.95×10^{-3}	1.36×10^{-4}	1.99×10^{-7}
$NSS_{inter} < NSS_{jpc}$		23	12	0
FA + HA	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}
min	1.67×10^{-7}	2.17×10^{-5}	9.34×10^{-8}	5.46×10^{-12}
max	2.49×10^{-4}	1.67×10^{-2}	1.99×10^{-4}	4.96×10^{-7}
$NSS_{inter} < NSS_{jpc}$		47	35	0

our initial assumption about the lack of correlation between multiple measurements of the same sample in the data set. The distribution probability of pK_{H2app} could not be calculated because 25 curves have their second inflection point at pH >9.8 (i.e., only 50% of phenolic-type sites had been deprotonated).

Apparent pK_{Hi} Values

The mean values of the apparent proton-binding constants are pK_{H1app}(FA) = 4.18 ± 0.21, pK_{H1app}(HA) = 4.49 ± 0.18, pK_{H2app}(FA) = 9.29 ± 0.33, and pK_{H2app}(HA) = 9.29 ± 0.38 (Table S3). These values coincide with those reported in the literature: $4.20 \le pK_{H1app} \le 4.28$ and $9.27 \le pK_{H2app} \le 9.80$ for NOM at $I = 0.1$ M;[1] pK_{H1app}(FA) = 3.80 ± 0.11, pK_{H1app}(HA) = 4.38 ± 0.13, pK_{H2app}(FA) = 9.78 ± 0.48, and pK_{H2app}(HA) = 9.72 ± 0.23 for NOM standards from the International Humic Substances Society (IHSS) at $I = 0.1$ M using the NICA model (called the modified Henderson−Hasselbalch in the source article);[26] pK_{H1app}(HA) = 4.60 ± 0.17 and pK_{H2app}(HA) = 8.87 ± 0.23 for vermicompost humic acids at $I = 0.1$ M;[27] and pK_{H1app} = 3.3/4.8 (two-component fit) and pK_{H2app} = 9.60 for dissolved organic matter at $I = 0.01$ M.[28] They are also consistent with the compilation of Perdue,[29] who showed that carboxylic- and phenolic-type groups have a Gaussian distribution of pK values centered at $pK_{H1app} = 4.5$ and $pK_{H2app} = 10$.

PCA Results over the [3.5−5.5] pH Interval

As mentioned previously, NSS_{ipc} decreases continuously when n increases. PCA results (Tables 2 and S4) show that the reconstructed signal falls progressively within the noise level for $2 < n < 3$ because $NSS_{3pc} < NSS_{inter} \approx NSS_{2pc} < NSS_{1pc}$. For $n = 2$, NSS_{ipc} and NSS_{inter} have the same scale class (hereafter noted $\langle NSS \rangle$) when all data are taken together. However, an examination of individual reconstructions (Table S4) shows that 16 out of the 23 FA (70% of data), 12 out of the 24 HA (50%), and 35 out of 47 FA + HA (75%) do not satisfy the $NSS_{npc} \approx NSS_{inter}$ criterion. Thus, the minimum number of PCs required to reproduce the data set within error in the [3.5−5.5] pH interval is three.

With two PCs, fewer FAs (30%) are reproduced than HAs (50%). This result is consistent with the difference in variability of the pK_{H1app} values for the two humic materials noted previously (σ(FA) = 0.21 and σ(HA) = 0.18). It agrees also with the generic values of the site distribution parameter m_{Hi} in the N-D model because $m_{H1(FA)} = 0.38$ and $m_{H1(HA)} = 0.50$[21] with m_{Hi} being inversely proportional to heterogeneity (Figure 1c).

PCA Results over the [3.5−9.8] pH Interval

In the [3.5−9.8] pH interval, every j increment decreases $\langle NSS_{npc} \rangle$ by about one order of magnitude up to $j = 6$ (Tables 3 and S5). For $j = 6$, $\langle NSS_{npc} \rangle \approx 0$ because a sixth-degree polynomial is used to interpolate titration curves. A sixth-order

(26) Ritchie, J. D.; Perdue, E. M. *Geochim. Cosmochim. Acta* **2003**, *67*, 85–96.
(27) Masini, J. C.; Abate, G.; Lima, E. C.; Hahn, L. C.; Nakamura, M. S.; Lichtig, J.; Nagatomy, H. R. *Anal. Chim. Acta* **1998**, *364*, 223–233.
(28) Lu, Y.; Allen, H. E. *Water Res.* **2002**, *36*, 5083–5101.
(29) Perdue, E. In *Humic Substances in Soil, Sediment, and Water: Geochemistry, Isolation, and Characterization*; Aiken, G. R., McKnight, D. M., Wershaw, R. L., McCarthy, P., Eds.; John Wiley and Sons: New York, 1985; Vol. *1*, pp 493−526.

Table 3. PCA Results in the [3.5−9.8] pH Interval

FA	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}	NSS_{4pc}	NSS_{5pc}	NSS_{6pc}
min	2.89×10^{-7}	3.90×10^{-5}	3.14×10^{-6}	2.64×10^{-7}	2.74×10^{-8}	5.35×10^{-9}	2.18×10^{-17}
max	3.55×10^{-5}	1.40×10^{-2}	3.27×10^{-4}	1.40×10^{-4}	1.61×10^{-5}	5.84×10^{-6}	1.31×10^{-16}
$NSS_{inter} < NSS_{jpc}$		23	23	21	10	3	0
HA	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}	NSS_{4pc}	NSS_{5pc}	NSS_{6pc}
min	4.09×10^{-7}	8.70×10^{-5}	2.17×10^{-6}	8.78×10^{-7}	4.86×10^{-7}	$4.98 \times 10^{-1}0$	2.82×10^{-17}
max	7.83×10^{-5}	3.50×10^{-5}	7.95×10^{-4}	2.68×10^{-4}	5.84×10^{-5}	3.67×10^{-6}	2.54×10^{-16}
$NSS_{inter} < NSS_{jpc}$		24	24	19	9	1	0
FA + HA	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}	NSS_{4pc}	NSS_{5pc}	NSS_{6pc}
min	2.89×10^{-7}	1.23×10^{-4}	1.98×10^{-5}	9.96×10^{-7}	3.56×10^{-8}	$4.95 \times 10^{-1}0$	3.98×10^{-17}
max	7.83×10^{-5}	5.29×10^{-2}	5.19×10^{-3}	5.77×10^{-4}	6.12×10^{-5}	5.73×10^{-6}	2.61×10^{-16}
$NSS_{inter} < NSS_{jpc}$		47	47	45	27	4	0

polynomial has seven degrees of freedom, but here this number is reduced to six because all curves were normalized to the same origin. For the three data sets, $\langle NSS_{inter} \rangle \approx \langle NSS_{4pc} \rangle$. With four PCs, 10 FA (44%), 9 HA (38%), and 27 FA + HA (57%) fail to satisfy the reconstruction criterion. With five PCs, all curves meet the criterion except for three FAs (13%), one HA (4%), and four FA + HAs (9%). Upon closer examination, in fact it appears that the three FAs (FH19, 20, and 21) have an NSS_{5pc}/NSS_{inter} ratio ≤3.14, which is small on the logarithmic scale (Table S5). Therefore, we conclude that j(FA) = 5. The HA that is not reproduced correctly is HH12_2. Its NSS_{5pc}/NSS_{inter} ratio is $3.12 \times 10^{-6}/4.09 \times 10^{-7} = 7.6$, a relatively high value. However, this curve has the lowest NSS_{inter} value of the 24 curves in the HA data set, meaning that the quality of this fit is probably too high and thus the error in this data is underestimated. We conclude that this data is extreme in terms of the experimental error, thus j(HA) = 5. Lastly, $1.1 < NSS_{5pc}/NSS_{inter} < 3.8$ for three out of the four FA + HAs (FH211, FH22, and HH8), which is small, and the fourth, which has the highest ratio (6.9), is again HH12_2. Thus, j(FA+HA) = 5.

In summary, the number of significant independent PCs accounting for the variance in the titration data is five from pH 3.5 to 9.8 and three from pH 3.5 to 5.5. The number of three at low pH means that the reactivity of carboxylic-type groups can be described (i) with a continuous probability distribution of binding constants, such as a normal distribution with a maximum density at the mean (pK_{Hi}), (ii) a pK_{Hi} ($1/m_{Hi}$) distribution equal to σ at the 68% confidence level, and (iii) a total concentration of sites (Q_{Hi}) equal to the integrated area (Figure 1c). The lower sensitivity of titration data to phenolic-type groups (two degrees of freedom instead of three) is explained by the fact that the experimental pK_{H2app} values (9.29 ± 0.33 for FA and 9.29 ± 0.38 for HA, Table S3) are close to the upper limit of the titration curves (pH_{max} 9.8). Clearly, the titration data lack sensitivity to pK_{H2}. Thus, the higher variability of phenolic constants reported in the literature may not always be real, and pK_{H2} values are likely underestimated for $pH_{max} < 10$.

Concluding Remarks

Because of the bimodal character of the proton affinity for NOM, two Gaussian functions are typically needed to describe the full acid−base properties of the reactant. By adding electrostatic effects and the initial charge offset,[30] seven to eight parameters are optimized to simulate a single data set. The number of unknown parameters exceeds by at least one the number of independent parameters (six) that can be freely varied during data fitting and up to three if the pH span is insufficient. Thus, the flexibility of proton-binding models is in general too high to provide robust chemical solutions because a single data set can be fit with an infinite number of numerical combinations. Obviously, some constraints have to be enforced to decrease correlations between parameters, as in model VI[15] and the Stockholm humic model;[18] otherwise, the results obtained represent essentially subjective choices. This does not mean that the existing models are overparametrized but that the titration data lack sensitivity to quantify the full binding properties of humic substances and that independent scientific and statistical knowledge has to be applied to constrain parameters in order to obtain meaningful values from measurements.

The considerable covariance between fit parameters has often been mentioned in the literature (e.g., refs 16 and 19), but the maximum number of parameters that can be refined was always unknown. In fact, modelers are cautioned against the existence of correlations in the NICA-Donnan User Guide:[11] "High correlations indicate that a minimum is being sought in a deep flat-bottomed valley: there is no problem finding a low point but you may have to wander down the valley a long way to find the exact lowest point." The PCA results show that the task at hand is statistically impossible unless one decreases the number of adjustable parameters in the least-squares fit of the experimental data. Sometimes the electrostatic b parameter of the N-D model is determined independently by measuring data at multiple ionic strengths and calculating a master curve.[31,32] Still, "in practice, it is not easy to derive a unique master curve," states Kinniburgh et al.[19] This difficulty can be explained by the linear relationship between the logarithms of the Donnan volume (V_D) and the electrolyte concentration in the Donnan electrostatic model as $\log V_D = b(1 - \log I) - 1$. Therefore, varying I does not solve the problem, and if it did, the fit model to the master curve still would be underconstrained by one degree in most cases.

The only way to improve tangibly the applicability of potentiometric titration models for NOM is to gather independent estimates of some proton-binding parameters by other means in order to reduce further the number of parameters adjusted by regression analysis. In a study of the binding of Cu(II) and Pb(II) to FA and HA, the number of phenolic-type sites was determined by ^{13}C NMR and the Donnan volume was estimated from peak elution times measured by size exclusion chromatography.[20,33]

(30) Milne, C. J.; Kinniburgh, D. G.; De Wit, J. C. M.; Van Riemsdijk, W. H.; Koopal, L. K. *Geochim. Cosmochim. Acta* **1995**, *59*, 1101–1112.

(31) De Wit, J. C. M.; Van Riemsdijk, W. H.; Nederlof, M. M.; Kinniburgh, D. G.; Koopal, L. K. *Anal. Chim. Acta* **1990**, *232*, 189–207.

(32) De Wit, J. C. M.; Van Riemsdijk, W. H.; Koopal, L. K. *Environ. Sci. Technol.* **1993**, *27*, 2005–2014.

(33) Christl, I.; Kretzschmar, R. *Environ. Sci. Technol.* **2001**, *35*, 2505–2511.

Avena et al.[12] investigated the possibility of using hydrodynamic radii measured by viscometry as the Donnan radius and came to the conclusion that the thus-obtained Donnan volume was too small to account for ionic strength effects on the surface charge (or the gel volumes used in the N-D model were unrealistically high, which is the same). More recently, semi-empirical pK_{Hi} values were estimated from structural models using the linear free-energy relationships (LFER) of Hammett and were used to reduce the number of fitting parameters.[34] The predicted pK_{Hi} values obtained by this approach are 3.73 ± 0.13 and 9.83 ± 0.23 for HA and 3.80 ± 0.20 and 9.87 ± 0.31 for FA. The predicted values for carboxylic-type groups compare well with results from potentiometric titrations analyzed using model VI and the Stockholm humic model (Table 1) but are about one log unit higher than the generic N-D values.[21] In contrast, phenolic-type sites have titration constants that are all lower than the predicted values, in part because it is complicated to measure the proton balance at high pH, as discussed previously. Given the difficulty in safely deducing intrinsic pK values from titration data only, publishing apparent (i.e., conditional) values obtained directly from the first derivative[35,36] of the titration data seems to be the most conservative approach. Then their dependence on ionic strength can be estimated empirically from the Davies equation[37,38] with reasonably good precision because log K_H typically varies by only 0.3 to 0.4 unit in the $0 \leq I \leq 2.0$ interval.[34]

The accuracy of acidic−basic parameters also depends on the accuracy of the titration measurement. Existing models suppose that protonation equilibrium is reached between subsequent aliquot additions of acid and base, which may not be the case when the initially rapid proton uptake is followed by a slower diffusion process. Suggestive evidence for kinetic effects are provided by titration hysteresis.[39−43] Lastly, systematically publishing the data from titration measurements in electronic annexes also would be helpful in the communal effort to improve the applicability of the existing electrostatic models and in turn the reliability of published intrinsic constants. Because the intrinsic constants are a property of the humic substance alone, they should not vary with solution composition nor the unpredictability of a minimization code.[19]

Acknowledgment. We thank Dr. M. F. Benedetti for providing the titration database analyzed in this work and Drs. A. Matynia and B. Causse for fruitful scientific discussions.

Supporting Information Available: Method used to calculate normalized sum-square values (*NSS*), plots of all titration curves with best-fit sixth-order polynomials, numerical results of all polynomial fits and coefficients of the sixth-order polynomial fits, pK_{Hiapp} values for all titration curves, and numerical PCA results. This material is available free of charge via the Internet at http://pubs.acs.org.

(34) Matynia, A.; Lenoir, T.; Causse, B.; Spadini, L.; Jacquet, T.; Manceau, A. *Geochim. Cosmochim. Acta*, 2010, doi: 10.1016/j.gca.2009.12.022.
(35) Prélot, B.; Charmas, R.; Zarzycki, P.; Thomas, F.; Villiéras, F.; Piasecki, W.; Rudzinski, W. *J. Phys. Chem.* **2002**, *106*, 13280–13286.
(36) Charmas, R.; Zarzycki, P.; Villiéras, F.; Thomas, F.; Prélot, B.; Piasecki, W. *Colloids Surf., A* **2004**, *244*, 9–17.
(37) Davies, C. W. *Ion Association*; Butterworths: Washington, DC, 1962.
(38) Stumm, W.; Morgan, J. J. *Aquatic Chemistry*, 3rd ed. Wiley: New York, 1996.
(39) Marshall, S. J.; Young, S. D.; Gregson, K. *Eur. J. Soil Sci.* **1995**, *46*, 471–480.
(40) Duc, M.; Adekola, F.; Lefevre, G.; Fedoroff, M. *J. Colloid Interface Sci.* **2006**, *303*, 49–55.
(41) Lefevre, G.; Duc, M.; Fedoroff, M. Accuracy in the Determination of Acid-Base Properties of Metal Oxides Surfaces. In *Surface Complexation Modeling*; Lützenkirchen, J., Ed.; Elsevier: Amsterdam, 2006; pp 35−66.
(42) Cooke, J. D.; Hamilton-Taylor, J.; Tipping, E. *Environ. Sci. Technol.* **2007**, *41*, 465–470.
(43) Zarzycki, P.; Rosso, K. M. *Langmuir* **2009**, *25*, 6841–6848.
(44) Plaza, C.; Senesi, N.; Polo, A.; Brunetti, G. *Environ. Sci. Technol.* **2005**, *39*, 7141–7146.
(45) Drosos, M.; Jerykiewicz, M.; Deligiannakis, Y. *J. Colloid Interface Sci.* **2009**, *332*, 78–84.

Supporting Information

Number of Independent Parameters in Potentiometric Titration of Humic Substances

Thomas Lenoir, and Alain Manceau*

Mineralogy & Environments Group, LGCA,
Université Joseph Fourier and CNRS, 38041 Grenoble cedex 9 France

E-mail for correspondence : Alain.Manceau@ujf-grenoble.fr

Number of pages: 12

Number of Figures: 3

Number of Tables: 5

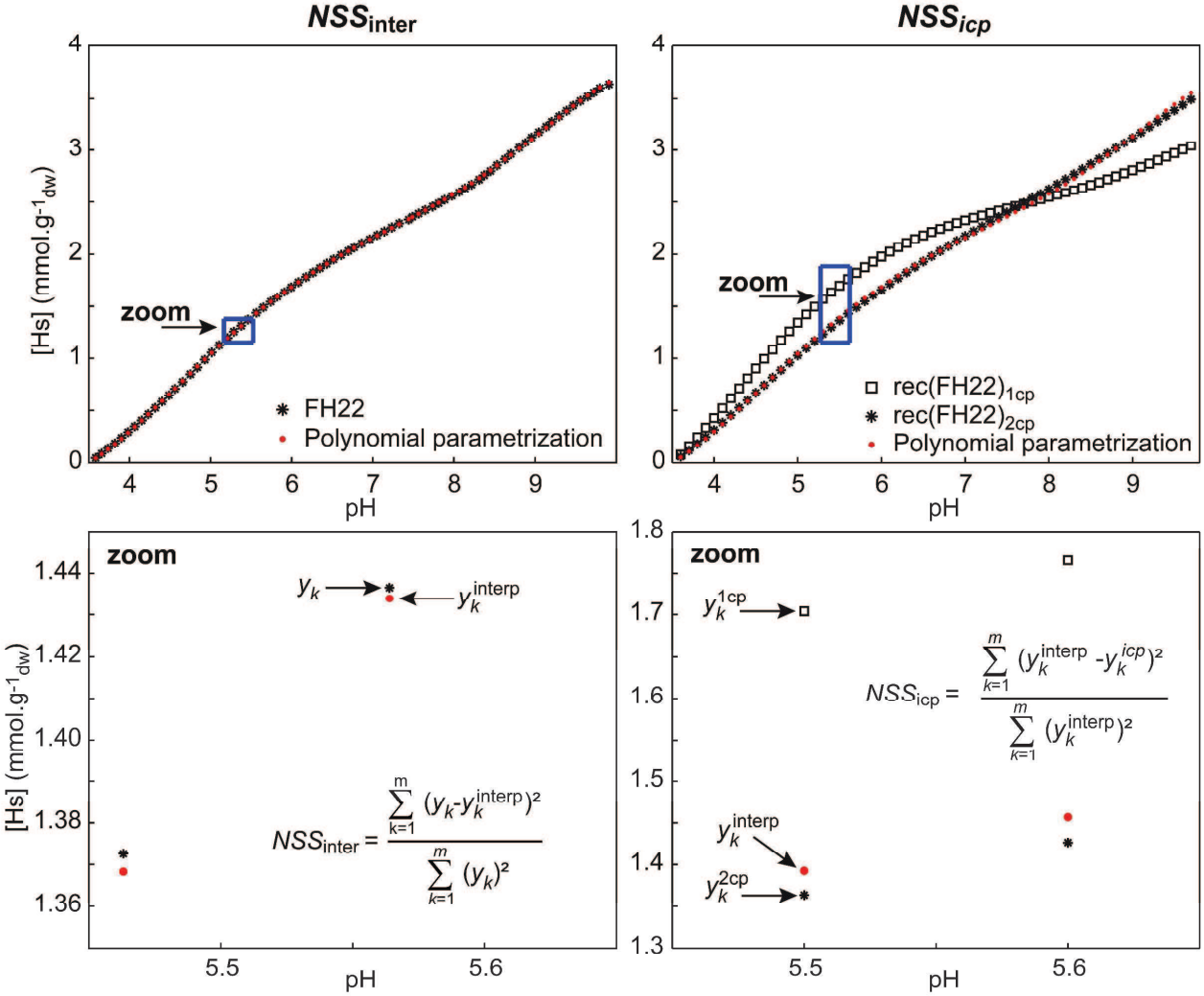


Figure S1. Calculation method of NSS_{inter} and NSS_{ipc}, taking FH22 as an example. For clarity, the PCA reconstruction is shown with one (i =1) and two (i =2) abstract components.

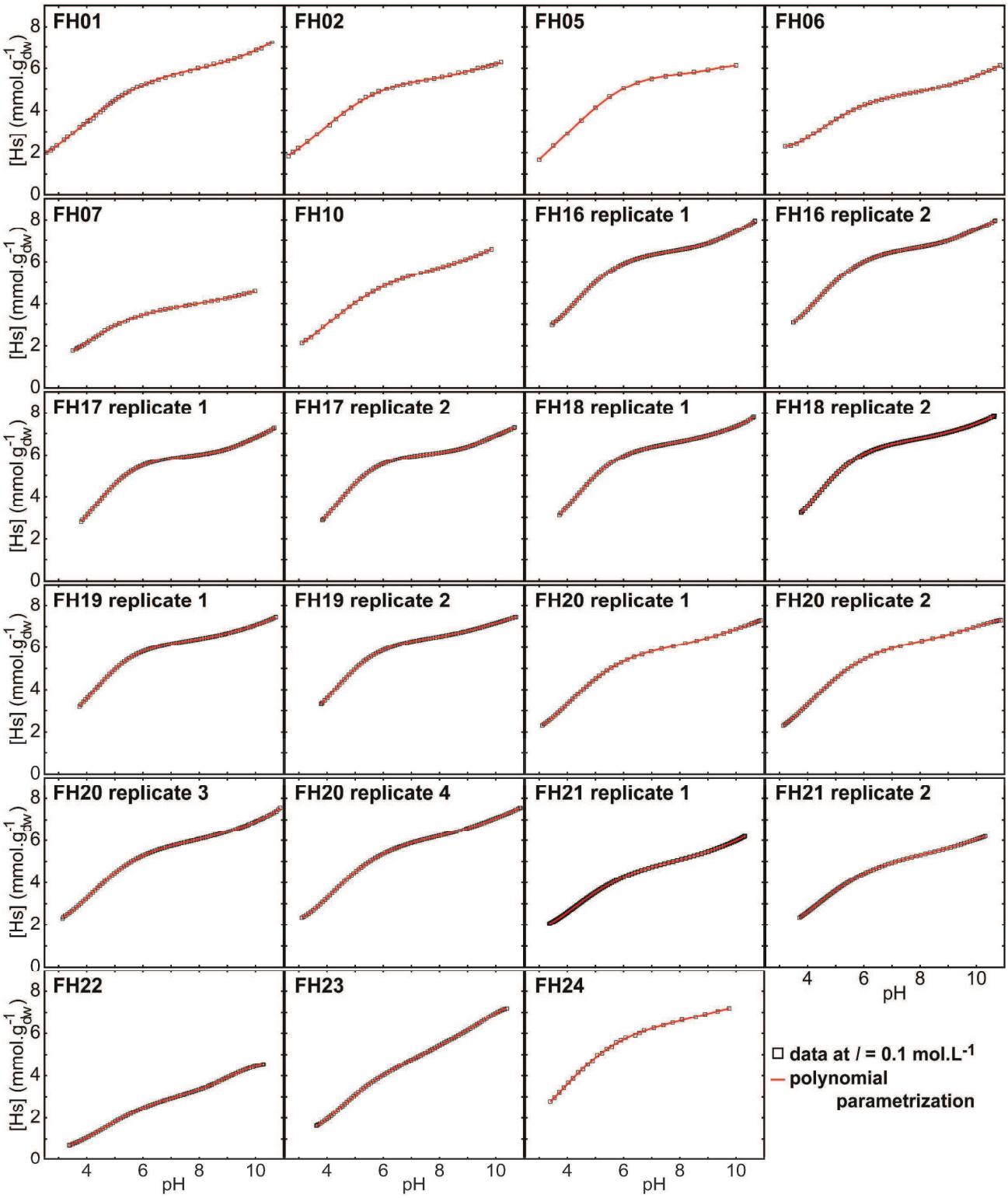


Figure S2. Titration dataset for fulvic acids (FH)[1] used for the PCA, and sixth order polynomial parametrization. Replicates are considered as single data. [Hs] is the amount of protons released per gram of dry matter.

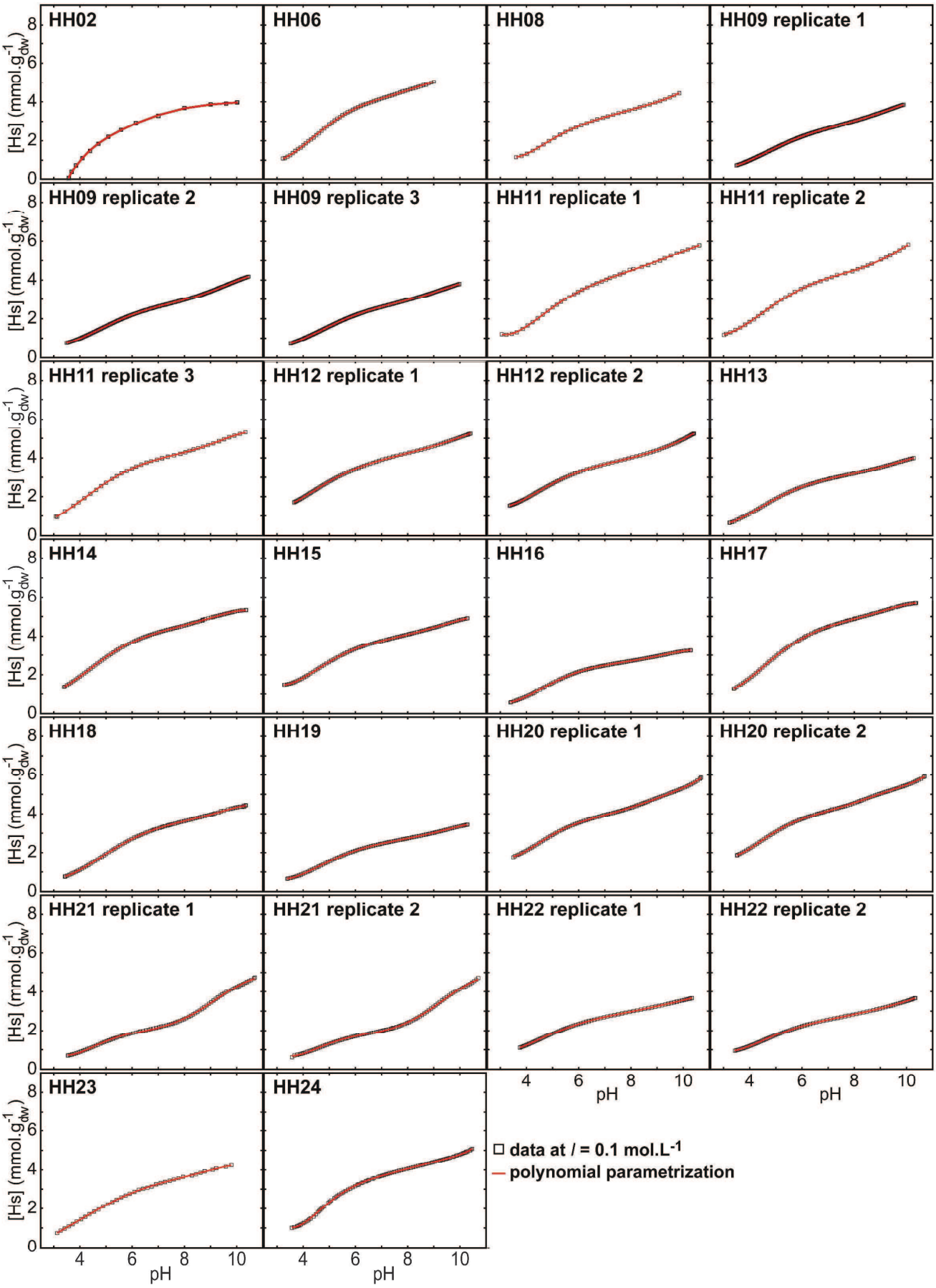

HH02
HH06
HH08
HH09 replicate 1
HH09 replicate 2
HH09 replicate 3
HH11 replicate 1
HH11 replicate 2
HH11 replicate 3
HH12 replicate 1
HH12 replicate 2
HH13
HH14
HH15
HH16
HH17
HH18
HH19
HH20 replicate 1
HH20 replicate 2
HH21 replicate 1
HH21 replicate 2
HH22 replicate 1
HH22 replicate 2
HH23
HH24
[Hs] (mmol.g-1dw)
pH
data at I = 0.1 mol.L-1
polynomial parametrization

Figure S3. Titration dataset for humic acids (HH)[1] used for the PCA (except HH02 and HH06), and sixth order polynomial parametrization. Replicates are considered as single data. [H] is the amount of protons released per gram of dry matter. HH02 was excluded because the quality of this data is low, as reported by Milne et al.[1], and HH06 has a $pH_{max} < 9.8$.

Table S1. Normalized sum-square (NSS_{inter}) fit values of the polynomial regressions to titration data with $pH_{max} \geq 9.0$. The fit quality increases with the degree of the polynomial (d°). However, Table S2 shows that the function wanders between data points above the sixth order, meaning that the reconstruction becomes sensitive to experimental errors.

FH	d°1	d°2	d°3	d°4	d°5	d°6	d°7	d°8	d°9	d°10
01	4.75E-03	7.33E-04	2.66E-04	6.27E-05	3.32E-05	3.31E-05	1.93E-05	1.89E-05	1.29E-05	1.03E-05
02	6.22E-03	5.66E-04	2.20E-04	8.37E-05	2.89E-05	2.82E-05	1.01E-05	8.50E-06	8.39E-06	7.59E-06
05	1.01E-02	4.95E-04	2.60E-04	5.19E-05	2.32E-05	1.13E-05	1.27E-06	5.28E-07	5.22E-07	1.59E-07
06	2.10E-03	7.64E-04	2.57E-04	1.03E-04	1.19E-05	5.73E-06	5.28E-06	5.18E-06	4.94E-06	3.63E-06
07	3.90E-03	5.31E-04	7.12E-05	4.44E-05	1.94E-05	1.48E-05	1.22E-05	1.13E-05	1.12E-05	1.01E-05
10	2.88E-03	3.00E-04	8.90E-05	2.00E-05	3.29E-06	3.06E-06	2.94E-06	2.84E-06	2.54E-06	2.54E-06
16_1	3.64E-03	1.46E-03	8.56E-05	5.72E-05	5.93E-06	1.50E-06	1.49E-06	1.47E-06	3.93E-07	3.41E-07
16_2	3.81E-03	1.22E-03	9.76E-05	6.95E-05	3.07E-06	5.80E-07	5.60E-07	5.28E-07	1.24E-07	5.11E-08
17_1	4.15E-03	1.93E-03	5.87E-05	5.39E-05	9.93E-06	1.98E-06	1.25E-06	1.16E-06	1.06E-07	1.05E-07
17_2	4.59E-03	1.91E-03	7.36E-05	6.84E-05	7.97E-06	1.45E-06	9.08E-07	6.90E-07	7.88E-08	7.85E-08
18_1	3.93E-03	1.23E-03	3.29E-05	3.18E-05	1.21E-05	4.70E-07	4.59E-07	4.57E-07	1.59E-07	1.52E-07
18_2	4.47E-03	1.17E-03	4.04E-05	3.83E-05	9.47E-06	1.07E-06	9.18E-07	8.89E-07	8.37E-07	8.37E-07
19_1	3.52E-03	1.22E-03	3.91E-05	3.22E-05	1.14E-05	1.23E-06	6.12E-07	5.46E-07	2.45E-07	1.89E-07
19_2	3.74E-03	1.02E-03	4.41E-05	3.70E-05	8.39E-06	8.72E-07	2.11E-07	2.05E-07	1.12E-07	7.58E-08
20_1	6.08E-03	8.43E-04	1.72E-04	8.64E-05	4.70E-06	1.15E-06	9.58E-07	5.14E-07	2.67E-07	2.35E-07
20_2	7.17E-03	6.57E-04	1.78E-04	1.02E-04	2.21E-06	8.32E-07	7.50E-07	2.61E-07	1.69E-07	1.67E-07
20_3	3.76E-03	1.03E-03	1.31E-04	6.64E-05	1.08E-05	3.22E-06	2.34E-06	2.16E-06	1.27E-06	1.21E-06
20_4	3.84E-03	8.09E-04	1.34E-04	8.48E-05	5.60E-06	7.73E-07	6.47E-07	6.34E-07	2.88E-07	2.86E-07
21_1	2.84E-03	5.29E-04	1.14E-04	3.01E-05	1.87E-06	3.99E-07	3.35E-07	2.47E-07	1.81E-07	1.41E-07
21_2	3.24E-03	3.39E-04	4.06E-05	1.35E-05	8.01E-07	3.24E-07	1.99E-07	1.78E-07	7.54E-08	7.50E-08
22	1.03E-03	4.62E-04	2.28E-04	2.24E-04	1.05E-05	5.57E-06	3.32E-06	2.84E-06	1.78E-06	1.76E-06
23	8.88E-04	2.69E-04	6.27E-05	5.96E-05	9.47E-06	6.51E-06	1.68E-06	6.57E-07	5.89E-07	5.88E-07
24	4.89E-03	3.17E-04	5.66E-05	4.01E-05	2.21E-05	2.16E-05	2.09E-05	2.08E-05	2.01E-05	1.93E-05

HH	d°1	d°2	d°3	d°4	d°5	d°6	d°7	d°8	d°9	d°10
02	3.00E-02	2.56E-03	6.26E-04	2.19E-04	1.00E-04	9.15E-05	3.91E-05	1.58E-05	1.41E-05	1.14E-06
06	4.31E-03	2.80E-04	2.46E-04	5.28E-05	1.26E-05	1.22E-05	6.69E-06	6.43E-06	5.00E-06	4.07E-06
08	1.85E-03	3.90E-04	2.15E-04	5.81E-05	1.23E-05	3.58E-06	3.19E-06	3.05E-06	2.84E-06	2.30E-06
09_1	1.30E-03	2.97E-04	1.20E-04	4.15E-05	1.08E-06	8.26E-07	4.14E-07	3.47E-07	3.22E-07	2.80E-07
09_2	8.91E-04	5.36E-04	1.55E-04	1.01E-04	1.33E-06	1.13E-06	2.59E-07	2.48E-07	2.47E-07	2.30E-07
09_3	1.36E-03	2.97E-04	1.19E-04	4.87E-05	1.63E-06	1.29E-06	2.12E-07	1.88E-07	1.23E-07	1.00E-07
11_1	2.60E-03	5.48E-04	5.42E-04	2.00E-04	2.55E-05	2.04E-05	1.67E-05	1.66E-05	1.56E-05	1.11E-05
11_2	2.26E-03	7.69E-04	5.56E-04	7.68E-05	1.28E-05	1.09E-05	4.87E-06	4.73E-06	4.51E-06	4.44E-06
11_3	4.38E-03	6.65E-04	2.24E-04	1.10E-04	9.11E-06	7.38E-06	5.13E-06	4.88E-06	4.86E-06	4.85E-06
12_1	2.15E-03	3.86E-04	4.08E-05	1.96E-05	2.80E-06	2.21E-06	6.89E-07	6.89E-07	2.42E-07	2.41E-07
12_2	1.50E-03	7.25E-04	1.60E-04	4.28E-05	3.73E-06	6.83E-07	6.41E-07	4.40E-07	4.39E-07	4.38E-07
13	4.59E-03	5.86E-04	2.61E-04	8.72E-05	3.30E-06	3.09E-06	3.04E-06	1.75E-06	1.60E-06	1.54E-06
14	4.29E-03	3.05E-04	1.01E-04	8.15E-05	4.52E-06	4.30E-06	2.34E-06	2.32E-06	2.25E-06	1.74E-06
15	2.56E-03	3.45E-04	2.26E-04	1.03E-04	3.71E-06	8.04E-07	4.48E-07	3.46E-07	3.40E-07	3.40E-07

16	6.34E-03	6.78E-04	3.23E-04	1.91E-04	5.87E-06	5.36E-06	4.10E-06	1.02E-06	8.25E-07	7.54E-07
17	5.72E-03	3.28E-04	1.68E-04	8.46E-05	1.67E-06	1.57E-06	1.30E-06	9.51E-07	9.25E-07	8.70E-07
18	4.80E-03	3.36E-04	2.85E-04	8.59E-05	1.18E-05	9.53E-06	8.09E-06	8.02E-06	7.44E-06	6.08E-06
19	3.36E-03	4.91E-04	2.84E-04	1.08E-04	1.89E-06	1.37E-06	1.26E-06	4.82E-07	4.70E-07	4.09E-07
20_1	9.62E-04	6.51E-04	7.04E-05	6.56E-05	3.42E-05	6.15E-06	9.82E-07	9.63E-07	9.62E-07	7.45E-07
20_2	1.33E-03	4.71E-04	6.33E-05	6.20E-05	3.75E-05	4.71E-06	1.66E-06	1.31E-06	9.03E-07	8.47E-07
21_1	5.28E-03	1.57E-03	8.14E-04	4.91E-04	8.91E-05	6.26E-05	1.35E-05	3.51E-06	3.50E-06	1.34E-06
21_2	8.46E-03	1.88E-03	7.57E-04	4.44E-04	8.76E-05	7.55E-05	1.93E-05	7.12E-06	4.71E-06	1.70E-06
22_1	2.43E-03	3.14E-04	5.38E-05	2.70E-05	2.61E-06	4.84E-07	4.60E-07	3.55E-07	1.77E-07	1.61E-07
22_2	1.70E-03	4.83E-04	1.64E-04	5.33E-05	4.42E-06	1.51E-06	7.98E-07	4.82E-07	4.65E-07	3.18E-07
23	3.91E-03	1.38E-04	8.90E-05	3.60E-05	2.41E-05	2.13E-05	2.05E-05	2.04E-05	2.04E-05	1.92E-05
24	5.64E-03	8.85E-04	3.16E-04	1.74E-04	7.16E-05	1.26E-05	1.05E-05	1.05E-05	6.79E-06	4.16E-06

Table S2. The marginal decline of NSS_{inter} from Table S1 with the polynomial order was calculated by smoothing each polynomial with a spline function and calculating NSS_{spline} between the two functions. If NSS_{inter} decreased monotonously with the order power of the polynomial, the marginal improvement of the data fit should be consistently smaller and smaller. Starting with the seventh degree, NSS_{inter} is increased by a greater margin than with the previous order, which is unjustified in the absence of experimental errors. Thus, above the sixth order the polynomial parametrization becomes quite sensitive to data noise. The first NSS_{spline} minimum for each titration curve is in green and the absolute minimum in red.

FH	d°1	d°2	d°3	d°4	d°5	d°6	d°7	d°8	d°9	d°10
01	4.75E-03	7.33E-04	2.66E-04	6.29E-05	3.28E-05	3.27E-05	**2.73E-05**	3.02E-05	3.30E-05	4.39E-05
02	6.22E-03	5.66E-04	2.20E-04	8.40E-05	2.80E-05	2.75E-05	**2.26E-05**	4.10E-05	3.17E-05	**1.45E-05**
05	1.01E-02	4.95E-04	2.60E-04	5.19E-05	2.28E-05	1.15E-05	**3.12E-06**	4.80E-06	5.50E-06	**2.55E-05**
06	2.10E-03	7.64E-04	2.57E-04	1.02E-04	1.21E-05	**5.75E-06**	5.80E-06	7.06E-06	1.17E-05	**4.99E-06**
07	3.90E-03	5.31E-04	7.12E-05	4.44E-05	1.94E-05	1.48E-05	1.22E-05	1.13E-05	1.12E-05	**1.00E-05**
10	2.88E-03	3.00E-04	8.90E-05	2.00E-05	3.32E-06	3.08E-06	2.97E-06	2.89E-06	2.80E-06	**2.79E-06**
16_1	3.64E-03	1.46E-03	8.56E-05	5.69E-05	5.65E-06	**1.57E-06**	1.63E-06	1.94E-06	**1.40E-05**	1.90E-05
16_2	3.81E-03	1.22E-03	9.76E-05	6.92E-05	3.02E-06	**5.92E-07**	6.09E-07	**5.48E-07**	3.57E-06	6.67E-06
17_1	4.15E-03	1.93E-03	5.87E-05	5.40E-05	9.53E-06	2.11E-06	1.38E-06	**1.19E-06**	7.44E-06	7.18E-06
17_2	4.59E-03	1.91E-03	7.36E-05	6.85E-05	7.89E-06	1.49E-06	1.27E-06	**8.46E-07**	6.82E-06	6.60E-06
18_1	3.93E-03	1.23E-03	3.29E-05	3.17E-05	1.15E-05	**1.39E-06**	1.56E-06	**1.37E-06**	6.07E-06	5.04E-06
18_2	4.47E-03	1.17E-03	4.04E-05	3.82E-05	9.31E-06	1.76E-06	1.02E-06	**9.08E-07**	1.38E-06	1.31E-06
19_1	3.52E-03	1.22E-03	3.91E-05	3.23E-05	1.10E-05	1.86E-06	**6.12E-07**	8.64E-07	1.07E-06	**3.83E-07**
19_2	3.74E-03	1.02E-03	4.41E-05	3.71E-05	8.20E-06	1.15E-06	3.76E-07	**2.59E-07**	5.36E-07	**1.39E-07**
20_1	6.08E-03	8.43E-04	1.72E-04	8.56E-05	5.61E-06	**1.11E-06**	1.25E-06	2.03E-06	**4.39E-07**	6.02E-07
20_2	7.17E-03	6.57E-04	1.78E-04	1.01E-04	4.53E-06	1.57E-06	**8.80E-07**	6.39E-06	2.30E-06	2.06E-06
20_3	3.76E-03	1.03E-03	1.31E-04	6.60E-05	1.01E-05	**3.58E-06**	5.11E-06	**2.84E-06**	1.15E-05	1.37E-05
20_4	3.84E-03	8.09E-04	1.34E-04	8.42E-05	5.90E-06	**7.64E-07**	9.73E-07	1.34E-06	6.89E-06	6.43E-06
21_1	2.84E-03	5.29E-04	1.14E-04	3.00E-05	1.82E-06	4.00E-07	3.40E-07	2.45E-07	**2.27E-07**	4.38E-07
21_2	3.24E-03	3.39E-04	4.06E-05	1.35E-05	7.89E-07	3.24E-07	2.06E-07	1.99E-07	1.14E-07	**1.00E-07**
22	1.03E-03	4.62E-04	2.28E-04	2.24E-04	1.06E-05	5.32E-06	3.30E-06	**3.02E-06**	4.03E-06	3.21E-06
23	8.88E-04	2.69E-04	6.27E-05	5.95E-05	9.55E-06	6.49E-06	**2.41E-06**	2.71E-06	1.84E-06	**1.72E-06**
24	4.89E-03	3.17E-04	5.66E-05	4.01E-05	2.21E-05	2.16E-05	2.09E-05	2.08E-05	2.01E-05	**1.93E-05**

HH	d°1	d°2	d°3	d°4	d°5	d°6	d°7	d°8	d°9	d°10
02	3.00E-02	2.56E-03	6.26E-04	2.19E-04	1.00E-04	9.15E-05	3.89E-05	1.59E-05	**1.41E-05**	6.39E-05
06	4.31E-03	2.80E-04	2.46E-04	5.28E-05	1.26E-05	1.22E-05	6.52E-06	6.38E-06	**5.45E-06**	5.78E-06
08	1.85E-03	3.90E-04	2.15E-04	5.81E-05	1.23E-05	3.58E-06	3.19E-06	3.05E-06	2.84E-06	**2.30E-06**

09_1	1.30E-03	2.97E-04	1.20E-04	4.15E-05	1.08E-06	8.26E-07	4.14E-07	3.47E-07	3.21E-07	**2.80E-07**
09_2	8.91E-04	5.36E-04	1.55E-04	1.01E-04	1.38E-06	1.11E-06	2.58E-07	**2.55E-07**	2.65E-07	**2.41E-07**
09_3	1.36E-03	2.97E-04	1.19E-04	4.87E-05	1.63E-06	1.29E-06	2.11E-07	1.88E-07	1.22E-07	**1.00E-07**
11_1	2.60E-03	5.48E-04	5.42E-04	1.99E-04	2.57E-05	2.07E-05	**1.82E-05**	1.91E-05	**1.61E-05**	5.11E-05
11_2	2.26E-03	7.69E-04	5.56E-04	7.69E-05	1.25E-05	1.07E-05	**5.21E-06**	5.27E-06	5.86E-06	**4.94E-06**
11_3	4.38E-03	6.65E-04	2.24E-04	1.09E-04	9.06E-06	7.28E-06	5.15E-06	**5.00E-06**	5.00E-06	5.18E-06
12_1	2.15E-03	3.86E-04	4.08E-05	1.95E-05	2.81E-06	2.21E-06	**9.31E-07**	9.38E-07	5.42E-07	**4.76E-07**
12_2	1.50E-03	7.25E-04	1.60E-04	4.27E-05	3.64E-06	6.68E-07	6.39E-07	**6.14E-07**	6.44E-07	**5.90E-07**
13	4.59E-03	5.86E-04	2.61E-04	8.71E-05	3.30E-06	3.10E-06	3.05E-06	1.88E-06	1.58E-06	**1.54E-06**
14	4.29E-03	3.05E-04	1.01E-04	8.14E-05	4.51E-06	4.24E-06	2.36E-06	2.35E-06	**2.25E-06**	3.90E-06
15	2.56E-03	3.45E-04	2.26E-04	1.03E-04	3.67E-06	8.16E-07	4.75E-07	**4.29E-07**	4.63E-07	4.63E-07
16	6.34E-03	6.78E-04	3.23E-04	1.91E-04	5.82E-06	5.38E-06	4.05E-06	1.59E-06	**9.66E-07**	1.36E-06
17	5.72E-03	3.28E-04	1.68E-04	8.44E-05	1.74E-06	1.61E-06	1.32E-06	1.23E-06	**1.02E-06**	1.52E-06
18	4.80E-03	3.36E-04	2.85E-04	8.57E-05	1.18E-05	9.39E-06	8.12E-06	**7.99E-06**	9.02E-06	2.36E-05
19	3.36E-03	4.91E-04	2.84E-04	1.08E-04	1.86E-06	1.39E-06	1.26E-06	**6.94E-07**	7.81E-07	**4.35E-07**
20_1	9.62E-04	6.51E-04	7.04E-05	6.55E-05	3.27E-05	**9.87E-06**	1.67E-05	1.47E-05	1.42E-05	**6.74E-06**
20_2	1.33E-03	4.71E-04	6.33E-05	6.20E-05	3.61E-05	**9.02E-06**	1.53E-05	7.90E-06	1.59E-06	**8.91E-07**
21_1	5.28E-03	1.57E-03	8.14E-04	4.93E-04	8.11E-05	6.20E-05	**6.01E-05**	1.50E-04	1.45E-04	7.90E-05
21_2	8.46E-03	1.88E-03	7.57E-04	4.45E-04	8.06E-05	7.34E-05	**6.61E-05**	1.65E-04	2.64E-04	1.62E-04
22_1	2.43E-03	3.14E-04	5.38E-05	2.70E-05	2.54E-06	4.82E-07	4.60E-07	3.87E-07	2.73E-07	**1.68E-07**
22_2	1.70E-03	4.83E-04	1.64E-04	5.32E-05	4.28E-06	1.52E-06	8.59E-07	4.78E-07	**4.76E-07**	9.88E-07
23	3.91E-03	1.38E-04	8.90E-05	3.60E-05	2.41E-05	2.13E-05	2.06E-05	**2.05E-05**	2.05E-05	**1.95E-05**
24	5.64E-03	8.85E-04	3.16E-04	1.74E-04	7.02E-05	1.31E-05	1.06E-05	**1.05E-05**	1.77E-05	**4.16E-06**

Table S3. Coefficients of the sixth order polynomial fits to data and apparent proton binding constants obtained from the maxima of the first. Twenty-five curves have only one maximum in the $3.5 \leq pH \leq 9.8$ interval.

FH	Coeff d°6	Coeff d°5	Coeff d°4	Coeff d°3	Coeff d°2	Coeff d°1	Coeff d°0	pK_{H1app}	pK_{H2app}
01	2.07E-08	-1.45E-06	3.70E-05	-4.42E-04	2.54E-03	-5.81E-03	6.18E-03	3.74	
02	-6.77E-08	1.49E-06	-1.26E-06	-1.96E-04	1.73E-03	-4.50E-03	5.22E-03	3.77	9.48
05	-3.92E-07	1.42E-05	-2.02E-04	1.42E-03	-5.22E-03	1.09E-02	-9.17E-03	3.93	9.26
06	1.67E-07	-8.26E-06	1.67E-04	-1.73E-03	9.57E-03	-2.59E-02	2.89E-02	4.45	
07	2.94E-07	-1.29E-05	2.33E-04	-2.18E-03	1.10E-02	-2.74E-02	2.78E-02	4.18	
10	7.49E-08	-4.00E-06	8.50E-05	-9.05E-04	4.98E-03	-1.24E-02	1.29E-02	4.03	
16_1	2.74E-07	-1.34E-05	2.64E-04	-2.67E-03	1.44E-02	-3.73E-02	3.94E-02	4.21	
16_2	2.10E-07	-1.09E-05	2.27E-04	-2.39E-03	1.33E-02	-3.53E-02	3.81E-02	4.27	
17_1	4.71E-07	-2.24E-05	4.31E-04	-4.26E-03	2.25E-02	-5.83E-02	6.03E-02	4.24	
17_2	4.23E-07	-2.06E-05	4.07E-04	-4.10E-03	2.20E-02	-5.79E-02	6.06E-02	4.29	
18_1	5.83E-07	-2.65E-05	4.91E-04	-4.71E-03	2.43E-02	-6.25E-02	6.46E-02	4.27	
18_2	4.97E-07	-2.31E-05	4.37E-04	-4.27E-03	2.25E-02	-5.85E-02	6.13E-02	4.31	
19_1	5.04E-07	-2.32E-05	4.33E-04	-4.16E-03	2.14E-02	-5.45E-02	5.57E-02	4.18	
19_2	4.43E-07	-2.08E-05	3.96E-04	-3.88E-03	2.03E-02	-5.23E-02	5.44E-02	4.24	
20_1	1.35E-07	-6.87E-06	1.40E-04	-1.45E-03	7.87E-03	-1.99E-02	2.05E-02	4.06	
20_2	8.54E-08	-4.99E-06	1.12E-04	-1.24E-03	7.01E-03	-1.82E-02	1.92E-02	4.12	9.81
20_3	2.18E-07	-1.03E-05	1.99E-04	-1.97E-03	1.03E-02	-2.60E-02	2.67E-02	4.11	
20_4	1.73E-07	-8.64E-06	1.74E-04	-1.79E-03	9.68E-03	-2.49E-02	2.61E-02	4.17	
21_1	1.40E-07	-6.78E-06	1.34E-04	-1.36E-03	7.39E-03	-1.91E-02	2.04E-02	4.26	
21_2	1.15E-07	-5.80E-06	1.19E-04	-1.25E-03	6.93E-03	-1.83E-02	1.98E-02	4.31	
22	-1.85E-07	5.49E-06	-5.44E-05	1.37E-04	9.13E-04	-5.15E-03	7.37E-03	4.59	9.13
23	2.62E-07	-1.28E-05	2.51E-04	-2.53E-03	1.37E-02	-3.67E-02	3.90E-02	4.52	9.24
24	1.70E-07	-8.27E-06	1.62E-04	-1.60E-03	8.32E-03	-2.01E-02	1.99E-02	3.92	8.82

HH	Coeff d°6	Coeff d°5	Coeff d°4	Coeff d°3	Coeff d°2	Coeff d°1	Coeff d°0	pK_{H1app}	pK_{H2app}
02	-3.12E-07	1.45E-05	-2.75E-04	2.76E-03	-1.54E-02	4.66E-02	-5.83E-02		
06	-1.77E-07	3.74E-06	-4.17E-06	-4.54E-04	4.27E-03	-1.39E-02	1.62E-02	4.35	8.40
08	4.46E-07	-1.97E-05	3.57E-04	-3.40E-03	1.77E-02	-4.72E-02	5.05E-02	4.67	
09_1	-6.07E-08	1.20E-06	4.26E-06	-2.67E-04	2.38E-03	-7.93E-03	9.58E-03	4.64	9.28
09_2	-3.37E-08	6.73E-08	2.45E-05	-4.63E-04	3.45E-03	-1.10E-02	1.32E-02	4.71	9.62
09_3	-6.62E-08	1.42E-06	7.48E-07	-2.39E-04	2.27E-03	-7.75E-03	9.52E-03	4.66	9.26
11_1	1.40E-07	-7.28E-06	1.52E-04	-1.63E-03	9.26E-03	-2.57E-02	2.82E-02	4.62	9.67
11_2	-1.26E-07	3.63E-06	-2.98E-05	-4.37E-05	1.68E-03	-6.66E-03	8.91E-03	4.51	9.88
11_3	-9.93E-08	2.62E-06	-1.65E-05	-1.07E-04	1.61E-03	-5.27E-03	5.89E-03	4.18	9.43
12_1	9.05E-08	-4.63E-06	9.55E-05	-1.00E-03	5.59E-03	-1.48E-02	1.58E-02	4.33	
12_2	1.52E-07	-7.22E-06	1.41E-04	-1.42E-03	7.67E-03	-2.02E-02	2.18E-02	4.39	
13	3.13E-08	-2.38E-06	6.12E-05	-7.37E-04	4.47E-03	-1.24E-02	1.32E-02	4.38	9.69
14	-4.86E-08	3.74E-07	2.40E-05	-4.89E-04	3.61E-03	-1.08E-02	1.22E-02	4.35	9.05
15	1.54E-07	-7.87E-06	1.61E-04	-1.68E-03	9.35E-03	-2.55E-02	2.80E-02	4.52	9.36
16	4.67E-08	-3.47E-06	8.75E-05	-1.04E-03	6.30E-03	-1.80E-02	1.98E-02	4.52	9.19
17	-3.63E-08	-2.69E-07	3.89E-05	-6.74E-04	4.84E-03	-1.47E-02	1.68E-02	4.50	9.16
18	-1.27E-07	3.99E-06	-4.09E-05	8.68E-05	1.01E-03	-5.34E-03	7.74E-03	4.80	9.39
19	4.77E-08	-3.15E-06	7.62E-05	-8.94E-04	5.42E-03	-1.56E-02	1.75E-02	4.63	9.57
20_1	4.76E-07	-2.12E-05	3.85E-04	-3.62E-03	1.85E-02	-4.77E-02	4.98E-02	4.42	
20_2	5.17E-07	-2.29E-05	4.11E-04	-3.83E-03	1.94E-02	-4.93E-02	5.10E-02	4.37	8.66
21_1	3.15E-07	-1.56E-05	3.08E-04	-3.07E-03	1.64E-02	-4.38E-02	4.66E-02	4.51	9.29
21_2	2.14E-07	-1.13E-05	2.32E-04	-2.39E-03	1.30E-02	-3.54E-02	3.82E-02	4.53	9.46
22_1	1.39E-07	-6.61E-06	1.29E-04	-1.30E-03	7.06E-03	-1.90E-02	2.05E-02	4.54	
22_2	1.18E-07	-5.69E-06	1.12E-04	-1.14E-03	6.27E-03	-1.69E-02	1.83E-02	4.53	
23	-1.68E-07	5.93E-06	-8.17E-05	5.54E-04	-1.95E-03	4.17E-03	-3.94E-03	4.00	8.81
24	7.96E-07	-3.55E-05	6.46E-04	-6.15E-03	3.20E-02	-8.48E-02	8.99E-02	4.62	

Table S4. NSS_{jpc} for $1 \leq j \leq 3$ and $3.5 \leq pH \leq 5.5$ compared to NSS_{inter}. Data for which $NSS_{jpc} < NSS_{inter}$ are in bold.

FH	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}
01	6.79E-05	7.16E-04	**1.80E-05**	2.77E-08
02	8.30E-05	8.62E-04	**3.43E-06**	1.77E-08
05	3.29E-05	2.47E-04	6.34E-05	**4.93E-07**
06	1.19E-05	1.33E-03	2.57E-05	**7.27E-09**
07	5.06E-05	**2.00E-05**	4.32E-06	1.87E-07
10	1.19E-05	2.18E-04	2.10E-05	**5.69E-08**
16_1	1.62E-06	5.77E-06	3.88E-06	**1.37E-08**
16_2	3.10E-07	4.85E-05	**7.75E-08**	5.55E-08
17_1	4.14E-06	6.37E-05	4.26E-05	**2.24E-08**
17_2	2.70E-06	1.95E-04	2.71E-05	**6.87E-08**
18_1	2.26E-06	1.50E-04	2.33E-05	**1.10E-08**
18_2	4.36E-07	3.38E-04	7.34E-06	**2.51E-10**
19_1	3.13E-06	6.95E-05	4.41E-05	**1.26E-08**
19_2	2.30E-06	3.88E-05	2.19E-05	**1.86E-10**
20_1	1.97E-06	2.36E-04	**1.85E-06**	7.20E-09
20_2	1.79E-06	9.79E-05	9.21E-06	**1.30E-11**
20_3	4.17E-06	1.42E-04	**1.04E-07**	2.84E-08
20_4	1.15E-06	2.97E-05	3.54E-06	**7.69E-09**

21_1	4.84E-07	4.69E-05	2.40E-05	**3.98E-08**
21_2	2.64E-07	1.51E-04	4.56E-05	**1.68E-08**
22	1.52E-05	1.64E-03	1.76E-04	**1.32E-07**
23	1.27E-05	2.30E-03	8.25E-05	**2.49E-07**
24	3.60E-05	6.64E-04	**2.59E-07**	7.19E-08

HH	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}
02	4.89E-04			
06	4.61E-05			
08	6.54E-06	6.13E-03	1.26E-05	**8.13E-08**
09_1	4.58E-06	4.08E-05	3.87E-05	**1.93E-08**
09_2	5.47E-06	4.82E-04	6.49E-05	**3.59E-09**
09_3	5.72E-06	7.95E-05	4.90E-05	**1.20E-08**
11_1	1.83E-05	2.44E-04	2.13E-05	**5.74E-08**
11_2	1.41E-05	2.67E-04	**8.46E-06**	4.31E-08
11_3	1.34E-05	2.08E-03	**9.23E-06**	3.69E-08
12_1	2.92E-06	8.75E-04	3.29E-06	**3.47E-08**
12_2	1.67E-07	3.18E-04	8.92E-06	**7.49E-09**
13	2.20E-05	6.17E-04	**1.82E-06**	2.90E-09
14	3.83E-06	9.24E-04	**1.51E-06**	1.80E-08
15	2.05E-06	4.16E-05	**3.15E-07**	3.61E-09
16	3.67E-05	2.24E-05	**6.83E-07**	1.99E-07
17	9.14E-06	1.85E-04	**3.46E-06**	2.14E-08
18	1.52E-05	3.45E-04	1.33E-04	**1.54E-09**
19	4.48E-06	3.26E-04	1.91E-05	**6.72E-09**
20_1	1.18E-05	1.52E-04	8.86E-05	**3.82E-08**
20_2	1.35E-05	1.34E-04	1.13E-04	**8.80E-08**
21_1	1.53E-04	4.23E-03	**1.36E-04**	1.14E-08
21_2	2.49E-04	4.18E-03	**8.46E-05**	4.08E-08
22_1	1.35E-06	6.48E-05	1.57E-06	**4.11E-08**
22_2	3.62E-06	2.08E-05	**1.04E-06**	3.65E-08
23	2.45E-05	2.40E-03	**3.29E-06**	3.93E-08
24	1.04E-04	9.95E-03	**8.72E-06**	5.39E-08

FH+HH	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}
FH01	6.79E-05	1.47E-03	**1.83E-05**	7.54E-08
FH02	8.30E-05	1.70E-03	**4.29E-06**	2.05E-09
FH05	3.29E-05	6.81E-04	5.15E-05	**3.44E-07**
FH06	1.19E-05	5.92E-04	**1.29E-06**	2.00E-08
FH07	5.06E-05	2.73E-04	**1.18E-05**	1.63E-07
FH10	1.19E-05	6.80E-04	1.47E-05	**9.99E-08**
FH16_1	1.62E-06	1.97E-04	1.23E-05	**2.28E-08**
FH16_2	3.10E-07	3.15E-05	5.01E-06	**7.38E-08**
FH17_1	4.14E-06	1.34E-04	7.12E-05	**6.42E-08**
FH17_2	2.70E-06	5.89E-05	5.86E-05	**1.40E-07**
FH18_1	2.26E-06	5.27E-05	5.14E-05	**4.15E-11**
FH18_2	4.36E-07	6.50E-05	3.02E-05	**1.25E-08**
FH19_1	3.13E-06	3.70E-04	6.31E-05	**1.15E-09**
FH19_2	2.30E-06	1.12E-04	4.31E-05	**4.55E-09**
FH20_1	1.97E-06	7.43E-04	**4.95E-07**	1.65E-08
FH20_2	1.79E-06	4.57E-04	3.97E-06	**2.05E-09**
FH20_3	4.17E-06	5.74E-04	**4.05E-07**	3.61E-08
FH20_4	1.15E-06	2.92E-04	**3.30E-07**	1.19E-08

FH21_1	4.84E-07	5.76E-05	8.71E-06	**5.20E-08**
FH21_2	2.64E-07	2.17E-05	1.98E-05	**2.71E-08**
FH22	1.52E-05	8.10E-04	8.47E-05	**1.10E-07**
FH23	1.27E-05	1.28E-03	1.98E-05	**2.10E-07**
FH24	3.60E-05	1.43E-03	**2.94E-07**	1.09E-07
HH08	6.54E-06	1.17E-02	**9.34E-08**	5.31E-08
HH09_1	4.58E-06	1.14E-03	1.02E-04	**2.05E-09**
HH09_2	5.47E-06	2.73E-03	1.01E-04	**3.31E-08**
HH09_3	5.72E-06	1.43E-03	1.10E-04	**9.97E-09**
HH11_1	1.83E-05	2.12E-03	5.16E-05	**1.46E-08**
HH11_2	1.41E-05	2.82E-04	7.32E-05	**7.75E-08**
HH11_3	1.34E-05	2.71E-04	3.11E-05	**4.53E-09**
HH12_1	2.92E-06	2.77E-05	2.73E-05	**7.53E-08**
HH12_2	1.67E-07	1.61E-04	8.19E-06	**4.88E-08**
HH13	2.20E-05	5.53E-05	2.70E-05	**4.49E-10**
HH14	3.83E-06	3.45E-05	3.45E-05	**4.39E-09**
HH15	2.05E-06	1.34E-03	8.26E-06	**5.46E-12**
HH16	3.67E-05	1.21E-03	**8.55E-06**	1.46E-07
HH17	9.14E-06	3.39E-04	5.26E-05	**3.22E-08**
HH18	1.52E-05	2.29E-03	1.99E-04	**1.04E-07**
HH19	4.48E-06	2.35E-03	4.46E-05	**3.66E-08**
HH20_1	1.18E-05	1.41E-03	3.71E-05	**3.20E-07**
HH20_2	1.35E-05	6.41E-04	3.70E-05	**4.96E-07**
HH21_1	1.53E-04	8.76E-03	1.93E-04	**1.73E-07**
HH21_2	2.49E-04	8.75E-03	**1.31E-04**	5.59E-08
HH22_1	1.35E-06	1.48E-03	2.05E-05	**3.98E-08**
HH22_2	3.62E-06	1.22E-03	2.15E-05	**3.79E-08**
HH23	2.45E-05	4.53E-04	1.15E-04	**2.32E-08**
HH24	1.04E-04	1.67E-02	**7.76E-05**	3.35E-09

As mentioned previously, HH02 is a poor data [1], and HH06 has a $pH_{max} < 9.8$.

Table S5. NSS_{jpc} for $1 \leq j \leq 6$ and $3.5 \leq pH \leq 9.8$ compared to NSS_{inter}. Data for which $NSS_{jpc} < NSS_{inter}$ are in bold.

FH	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}	NSS_{4pc}	NSS_{5pc}	NSS_{6pc}
01	3.55E-05	4.53E-04	8.98E-05	7.99E-05	**8.16E-06**	1.02E-06	6.86E-17
02	2.55E-05	2.02E-04	9.98E-05	9.98E-05	**6.73E-06**	6.86E-07	8.13E-17
05	1.22E-05	1.32E-03	3.27E-04	5.31E-05	1.61E-05	**2.38E-06**	2.18E-17
06	5.08E-06	1.44E-03	1.42E-04	1.40E-04	**3.17E-06**	1.11E-06	9.54E-17
07	1.55E-05	1.41E-04	2.81E-05	2.62E-05	**5.39E-06**	1.05E-06	1.31E-16
10	2.50E-06	1.58E-03	4.40E-05	4.32E-05	**1.73E-06**	3.26E-07	6.55E-17
16_1	8.55E-07	6.30E-04	3.49E-05	7.35E-06	7.26E-06	**4.94E-07**	5.18E-17
16_2	4.68E-07	4.37E-04	2.95E-05	2.59E-05	7.07E-06	**2.23E-07**	4.80E-17
17_1	2.28E-06	2.56E-03	1.40E-04	**8.65E-07**	8.47E-07	5.35E-09	4.80E-17
17_2	1.76E-06	2.19E-03	1.13E-04	2.01E-05	**1.39E-06**	4.89E-08	5.38E-17
18_1	4.79E-07	1.01E-03	2.35E-05	2.30E-05	8.39E-06	**7.23E-08**	5.08E-17
18_2	2.89E-07	7.84E-04	8.13E-05	7.23E-05	6.49E-06	**1.15E-08**	3.65E-17
19_1	1.32E-06	1.77E-03	6.60E-05	2.58E-05	6.47E-06	**1.81E-07**	5.86E-17
19_2	9.75E-07	1.17E-03	1.33E-05	7.32E-06	6.70E-06	**3.25E-07**	5.37E-17
20_1	9.77E-07	3.90E-05	3.41E-05	1.92E-05	**2.74E-08**	1.32E-08	6.91E-17
20_2	9.60E-07	6.39E-05	6.35E-05	**2.64E-07**	4.22E-08	8.52E-09	5.08E-17

20_3	9.87E-07	4.54E-05	2.67E-05	2.67E-05	**2.88E-07**	2.28E-07	5.80E-17
20_4	7.56E-07	1.04E-04	9.45E-06	2.81E-06	**2.93E-08**	2.67E-08	6.04E-17
21_1	3.54E-07	1.95E-03	3.14E-06	2.73E-06	2.71E-06	1.11E-06	**1.05E-16**
21_2	3.60E-07	1.35E-03	1.51E-04	3.22E-05	9.88E-07	4.35E-07	**7.55E-17**
22	4.93E-06	1.40E-02	1.97E-04	1.74E-05	1.44E-05	5.84E-06	**6.53E-17**
23	5.98E-06	1.36E-02	4.35E-05	2.17E-05	**5.80E-06**	3.01E-08	3.26E-17
24	2.18E-05	1.49E-04	1.43E-04	9.03E-05	**1.07E-05**	6.12E-07	5.36E-17

HH	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}	NSS_{4pc}	NSS_{5pc}	NSS_{6pc}
02	1.09E-04						
06	1.16E-05						
08	3.69E-06	1.09E-03	6.12E-04	1.26E-05	5.48E-06	**3.67E-06**	1.01E-16
09_1	8.01E-07	1.05E-03	1.27E-04	6.50E-05	**5.28E-07**	2.97E-07	1.27E-16
09_2	1.12E-06	1.90E-03	1.69E-04	3.89E-05	9.12E-06	**4.98E-10**	1.40E-16
09_3	1.31E-06	9.82E-04	1.54E-04	6.10E-05	**5.29E-07**	4.77E-07	1.19E-16
11_1	2.51E-05	2.23E-04	2.00E-04	3.34E-05	**3.09E-06**	5.79E-08	6.88E-17
11_2	9.23E-06	1.29E-04	1.11E-04	5.34E-05	5.30E-05	**3.03E-07**	6.35E-17
11_3	5.17E-06	1.45E-03	4.78E-04	7.42E-06	**5.04E-06**	3.46E-08	8.68E-17
12_1	2.62E-06	3.78E-04	8.88E-05	1.29E-05	**2.06E-06**	1.23E-06	1.27E-16
12_2	4.09E-07	1.33E-04	1.33E-04	1.18E-05	1.03E-05	3.12E-06	**1.39E-16**
13	2.88E-06	8.42E-04	9.95E-05	**1.48E-06**	1.23E-06	7.62E-08	1.64E-16
14	4.77E-06	1.51E-03	1.19E-04	**1.41E-06**	1.41E-06	7.89E-07	8.80E-17
15	7.84E-07	2.31E-04	5.78E-06	4.87E-06	**4.86E-07**	1.08E-07	1.27E-16
16	5.25E-06	1.74E-03	1.01E-04	8.19E-05	9.25E-06	**2.46E-06**	1.25E-16
17	1.65E-06	1.82E-03	2.17E-06	**8.78E-07**	8.61E-07	5.78E-07	3.99E-17
18	4.59E-06	6.02E-04	4.41E-04	5.53E-05	**3.17E-06**	2.91E-07	7.80E-17
19	1.45E-06	1.35E-04	8.64E-05	3.07E-06	3.07E-06	**1.22E-07**	2.52E-16
20_1	5.48E-06	3.20E-04	1.22E-04	4.36E-05	2.06E-05	**9.85E-07**	1.29E-16
20_2	6.06E-06	1.89E-04	1.87E-04	6.59E-05	5.84E-05	**7.38E-07**	7.40E-17
21_1	6.92E-05	2.74E-02	1.23E-04	**3.13E-05**	6.88E-06	1.57E-06	7.98E-17
21_2	7.83E-05	3.50E-02	1.64E-04	**2.83E-05**	5.96E-06	5.32E-07	9.13E-17
22_1	5.87E-07	9.87E-05	5.85E-05	1.08E-05	3.80E-06	**3.02E-07**	1.95E-16
22_2	1.52E-06	8.70E-05	1.70E-05	6.55E-06	**1.30E-06**	1.17E-06	2.54E-16
23	2.08E-05	4.78E-04	1.86E-04	1.40E-04	**1.98E-05**	5.32E-07	8.00E-17
24	1.56E-05	1.48E-03	7.95E-04	2.68E-04	**2.50E-06**	2.40E-07	2.82E-17

FH+HH	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}	NSS_{4pc}	NSS_{5pc}	NSS_{6pc}
FH01	3.55E-05	2.61E-04	1.49E-04	1.31E-04	**4.18E-06**	1.82E-06	8.32E-17
FH02	2.55E-05	1.72E-03	1.63E-04	7.67E-05	**1.91E-06**	2.35E-07	8.33E-17
FH05	1.22E-05	3.97E-03	1.66E-04	3.23E-05	2.57E-05	**3.22E-06**	4.66E-17
FH06	5.08E-06	1.23E-04	7.30E-05	3.90E-05	6.92E-06	**5.40E-07**	1.11E-16
FH07	1.55E-05	4.11E-04	6.87E-05	3.04E-05	**9.15E-06**	1.64E-06	1.43E-16
FH10	2.50E-06	2.03E-04	1.08E-04	9.40E-05	**8.10E-07**	7.94E-07	6.92E-17
FH16_1	8.55E-07	2.91E-03	1.11E-04	1.60E-05	7.06E-06	**4.86E-07**	5.59E-17
FH16_2	4.68E-07	2.44E-03	5.18E-05	3.38E-05	9.96E-06	**9.66E-08**	6.04E-17
FH17_1	2.28E-06	6.27E-03	3.46E-04	7.11E-05	**9.35E-07**	1.00E-08	6.11E-17
FH17_2	1.76E-06	5.67E-03	2.68E-04	1.19E-04	3.15E-06	**1.61E-07**	5.51E-17
FH18_1	4.79E-07	3.66E-03	4.55E-05	3.13E-05	5.30E-06	**3.15E-08**	6.81E-17
FH18_2	2.89E-07	3.13E-03	4.71E-05	4.36E-05	2.08E-06	**1.70E-08**	4.48E-17
FH19_1	1.32E-06	5.04E-03	1.92E-04	1.14E-05	8.54E-06	**6.21E-08**	6.04E-17
FH19_2	9.75E-07	4.00E-03	6.73E-05	1.83E-05	5.67E-06	**3.18E-07**	6.21E-17
FH20_1	9.77E-07	1.06E-03	4.93E-05	4.26E-05	**2.52E-07**	9.92E-08	6.84E-17
FH20_2	9.60E-07	9.63E-04	3.83E-05	2.92E-05	**3.56E-08**	4.95E-10	5.53E-17

FH20_3	9.87E-07	6.87E-04	6.72E-05	3.12E-05	**7.07E-07**	5.06E-07	6.01E-17
FH20_4	7.56E-07	4.08E-04	2.15E-05	1.91E-05	**1.97E-07**	8.34E-08	6.28E-17
FH21_1	3.54E-07	2.43E-04	2.28E-05	2.26E-05	2.19E-06	1.35E-06	**1.32E-16**
FH21_2	3.60E-07	1.55E-04	1.34E-04	1.76E-05	3.76E-07	**3.34E-07**	9.51E-17
FH22	4.93E-06	8.03E-03	6.23E-05	1.03E-05	9.42E-06	5.22E-06	**1.06E-16**
FH23	5.98E-06	7.70E-03	1.98E-05	1.66E-05	7.88E-06	**3.13E-09**	5.88E-17
FH24	2.18E-05	1.21E-03	1.60E-04	1.57E-04	**1.74E-05**	1.47E-07	6.18E-17
HH08	3.69E-06	6.48E-03	5.53E-04	2.03E-04	6.15E-06	5.73E-06	**1.34E-16**
HH09_1	8.01E-07	6.68E-03	2.85E-05	5.61E-06	2.85E-06	**5.08E-07**	1.19E-16
HH09_2	1.12E-06	8.60E-03	3.71E-05	2.88E-05	1.92E-05	**2.42E-08**	1.67E-16
HH09_3	1.31E-06	6.54E-03	5.03E-05	4.19E-06	3.56E-06	**6.54E-07**	1.36E-16
HH11_1	2.51E-05	3.77E-03	3.04E-04	**1.58E-06**	2.37E-07	1.43E-08	7.21E-17
HH11_2	9.23E-06	2.63E-03	6.08E-05	5.23E-05	5.11E-05	**6.42E-07**	6.01E-17
HH11_3	5.17E-06	2.17E-04	4.49E-05	3.67E-05	**4.22E-06**	2.31E-07	1.09E-16
HH12_1	2.62E-06	1.04E-03	5.85E-05	3.04E-05	**5.34E-07**	5.19E-07	1.45E-16
HH12_2	4.09E-07	2.11E-03	2.81E-05	9.62E-06	8.91E-06	2.84E-06	**1.60E-16**
HH13	2.88E-06	4.78E-04	4.91E-05	6.92E-06	**9.81E-07**	6.34E-09	1.65E-16
HH14	4.77E-06	2.33E-04	1.09E-04	1.84E-05	**1.98E-06**	1.52E-06	9.57E-17
HH15	7.84E-07	1.39E-03	1.39E-04	1.11E-05	1.17E-06	**1.59E-07**	1.39E-16
HH16	5.25E-06	2.99E-04	2.24E-04	4.76E-05	**2.41E-06**	1.91E-06	1.71E-16
HH17	1.65E-06	4.72E-04	3.90E-04	**9.96E-07**	9.39E-07	7.48E-07	3.98E-17
HH18	4.59E-06	2.90E-03	9.00E-04	1.45E-05	1.00E-05	**1.56E-07**	8.38E-17
HH19	1.45E-06	2.38E-03	2.91E-04	2.75E-05	2.98E-06	**3.63E-08**	2.61E-16
HH20_1	5.48E-06	3.60E-03	9.76E-05	3.22E-05	2.68E-05	**3.81E-07**	1.58E-16
HH20_2	6.06E-06	2.31E-03	8.53E-05	6.34E-05	6.12E-05	**3.68E-08**	1.17E-16
HH21_1	6.92E-05	4.38E-02	3.91E-03	8.20E-05	**6.64E-06**	1.90E-06	1.03E-16
HH21_2	7.83E-05	5.29E-02	5.19E-03	9.63E-05	**6.53E-06**	3.52E-07	1.03E-16
HH22_1	5.87E-07	2.36E-03	2.31E-04	2.85E-06	1.70E-06	**1.48E-07**	1.91E-16
HH22_2	1.52E-06	3.48E-03	4.51E-05	3.04E-06	1.90E-06	**1.03E-06**	2.58E-16
HH23	2.08E-05	1.36E-03	2.53E-04	1.58E-04	**5.26E-06**	2.22E-06	1.27E-16
HH24	1.56E-05	2.35E-03	1.47E-03	5.77E-04	2.72E-05	**1.02E-08**	6.48E-17

Reference

(1) Milne, C. J.; Kinniburgh, D. G.; Tipping, E., *Environ. Sci. Technol.* **2001,** *35*, 2049-2059.

ANNEXE 4

Convergence-Optimized Procedure for Applying the NICA-Donnan Model to Potentiometric Titrations of Humic Substances

Environ. Sci. Technol. **2010,** *44,* 6221–6227

Convergence-Optimized Procedure for Applying the NICA-Donnan Model to Potentiometric Titrations of Humic Substances

THOMAS LENOIR,[†,‡] ANTHONY MATYNIA,[†] AND ALAIN MANCEAU*[,†]

Mineralogy & Environments Group, LGCA, Université Joseph Fourier and CNRS, 38041 Grenoble cedex 9, France, and Laboratoire Central des Ponts et Chaussées (LCPC), Route de Bouaye, BP 4129, 44341 Bouguenais cedex, France

Received September 18, 2009. Revised manuscript received June 30, 2010. Accepted July 7, 2010.

Despite the high success of the NICA-Donnan (N-D) model to describe the interaction of protons and metal ions with natural organic matter, the large number of fit parameters is a major hindrance to its capacity to provide unique numerical solutions. This well-known difficulty is reflected in the unusually low value of the generic proton binding constant for carboxylic-type groups of fulvic acid ($pK_{H1} = 2.34$), and to some extent of humic acid (2.93), and by the considerable covariance of the other generic N-D parameters. In some studies, the number of parameters obtained by regression is reduced by estimating some values independently with other techniques. Alternatively, the applicability of the model can be improved by devising a rigorous simulation procedure, which constrains the model-fit to converge toward chemically and physically realistic values. A procedure based on three successive iterations is proposed, and the solution is shown to be stable and invariant with the initial set of parameter values. The new generic parameters, in particular pK_{H1}(FA) = 3.54 and pK_{H1}(HA) = 3.87, derived from the same data set as the previous generic parameters, are in better agreement with literature data.

Introduction

The acid−base properties of fulvic (FA) and humic (HA) acids are characterized generally by measuring the amount of protons released in solution [H_{sol}] against pH in potentiometric titrations. The pH values of the two inflection points are apparent (also called "conditional") pK values that are shifted by 0.5 to 1.0 units relative to intrinsic pK values, because protons are retained at the surface of organic matter by the net negative charge from dissociated groups (*1*). Because intrinsic constants are the properties of a humic substance needed to predict acid−base reactivity in the environment, electrostatic models were developed to calculate their values numerically by simulating Coulombic interactions (*2*−*9*). A problem inherent to any electrostatic model is the impossibility of finding a unique mathematical solution, even when titration curves are measured at several ionic strengths (I) to vary the local electrostatic potential at the binding sites (*6, 10*−*13*). Thus, it can be questioned if the extent to which the large variability in proton binding constants reported in the literature for the carboxylic-type and phenolic-type groups (pK_{Hi}) of FA and HA, sometimes obtained with the same Donnan-type electrostatic model, is real or biased (Table S1 in the Supporting Information). For example, the average (2.65 ± 0.43) and generic (2.34, Table S2) NICA-Donnan pK_{H1} values of carboxylic-type groups for FA derived from the analysis of a large data set (*14*) are much lower than other values reported in the literature for the same type of humic substance, e.g., 3.20 obtained with Model VI (*7*), 3.50 with the Stockholm Humic Model (*9*), and 3.80 predicted semiempirically (*15*).

The first aim of this investigation was to test the robustness of the analytical solutions used to derive the generic NICA-Donnan parameters. As will be shown, several pairs of parameters are strongly correlated, so that some can be increased and others decreased with no significant change in quality of a model-fit to data. As mentioned previously, evidence for multiple covariances in the analysis of titration data is not new (*11, 13, 15, 16*), and modelers are cautioned against this possibility in the NICA-Donnan (N-D) User Guide (*10*): "High correlations indicate that a minimum is being sought in a deep flat-bottomed valley: there is no problem finding a low point but you may have to wander down the valley a long way to find the exact lowest point". Thus, the second aim is to establish an optimal fitting procedure, which constrains the N-D minimization algorithm to converge toward realistic solutions, hence reducing model bias. Harmonizing practices among laboratories should also improve the reliability of fit results. Following the philosophy of Tipping (*7*), Milne and co-workers (*14*), and Gustafsson (*9*), a new set of generic parameters derived from the computational analysis of the same data set as in Milne et al. (*14*) is produced. The new N-D pK_{Hi} values compare better with results from Model VI (*7*) and the Stockholm Humic Model (*9*), and since the mathematical convergence is well constrained in the new regression procedure, the same solution is found irrespective of the starting values.

NICA-Donnan Model

Only a brief introduction to the N-D model is given here; more detailed accounts can be found in the following references: *1, 4, 6, 8, 12, 17*−*21*. In the N-D model, pK_{Hi} and Q_{Hi} are determined by fitting a master-curve (MC, $Q_H = f(\text{pH})$) corrected for Coulombic effects and generated from a titration data set at several ionic strengths ($Q_{HT} = f(\text{pH})$). The MC is calculated by

$$Q_H = Q_{HT} - V_D(c_{DH} - c_H) \quad (1)$$

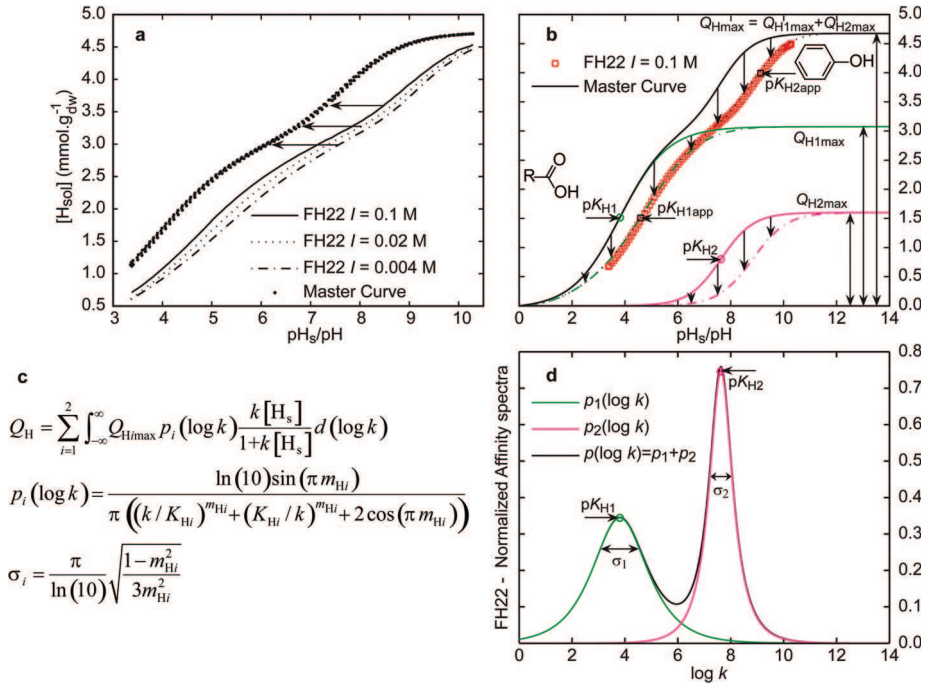
where V_D is the Donnan volume adjacent to the organic surface within which counterions are accumulated, and c_{DH} and c_H are the concentrations of protons in the Donnan volume and in the bulk solution, respectively. The two concentrations are related by $c_{DH} = \chi c_H$, where χ is the Boltzmann factor given by the Boltzmann equation $\chi = \exp(-e\Psi_D/kT)$. The Donnan approximation assumes that Ψ_D is constant within V_D and drops to zero at the boundary between the molecule and the solution (*8*). The Donnan volume is linearly correlated to the electrolyte concentration, and the following empirical relation was postulated (*6*):

$$\log V_D = b(1 - \log I) - 1 \quad (2)$$

where b is an adjustable parameter. Thus, when the Donnan volume and its electric potential (Ψ_D) are chosen correctly, all

* Corresponding author e-mail: Alain.Manceau@obs.ujf-grenoble.fr.
† Université Joseph Fourier and CNRS.
‡ Laboratoire Central des Ponts et Chaussées (LCPC).

10.1021/es1015313 © 2010 American Chemical Society
Published on Web 07/21/2010



$$Q_{\mathrm{H}} = \sum_{i=1}^{2} \int_{-\infty}^{\infty} Q_{\mathrm{H}i\max}\, p_i(\log k)\, \frac{k[\mathrm{H_s}]}{1+k[\mathrm{H_s}]}\, d(\log k)$$

$$p_i(\log k) = \frac{\ln(10)\sin(\pi m_{\mathrm{H}i})}{\pi\left((k/K_{\mathrm{H}i})^{m_{\mathrm{H}i}} + (K_{\mathrm{H}i}/k)^{m_{\mathrm{H}i}} + 2\cos(\pi m_{\mathrm{H}i})\right)}$$

$$\sigma_i = \frac{\pi}{\ln(10)}\sqrt{\frac{1-m_{\mathrm{H}i}^2}{3m_{\mathrm{H}i}^2}}$$

FIGURE 1. (a) Titration data (pH) and master-curve (pH_s) for the fulvic acid FH22 (*14*) calculated from measurements at three ionic strengths. (b) Amount of protons released in solution ($[H_{sol}]$) per gram of dry matter against pH (red squares) and best-fit proton binding parameters. Inflection points of data are the apparent (also called conditional) dissociation constants for carboxylic- and phenolic-type sites (pK_{Hiapp}). The black line is the optimal bimodal (R-COOH, Ph−OH) fit obtained with the N-D FIT code (*10*). The new pK_{Hi} are intrinsic values. Vertical arrows point to the direction in which the deprotonation curves shift because of electrostatic effects. The inclusion of an electrostatic correction does not change $Q_{Hi\max}$. (c,d) Intrinsic affinity Sips distributions for R-COOH (p_1, σ_1, pK_{H1}) and Ph−OH (p_2, σ_2, pK_{H2}) groups normalized to $Q_{Hi\max}$.

Q_{HT} curves merge in a single Q_{H} proton binding curve independent of electrostatic effect (Figure 1a,b). Practically, this is done by optimizing b.

Intrinsic acid−base parameters are obtained by modeling the MC with the NICA equation recast for protons (*10, 18, 22*):

$$Q_{\mathrm{H}} = Q_{\mathrm{H1max}}\frac{(K_{\mathrm{H1}}[\mathrm{H_s}])^{m_{\mathrm{H1}}}}{1+(K_{\mathrm{H1}}[\mathrm{H_s}])^{m_{\mathrm{H1}}}} + Q_{\mathrm{H2max}}\frac{(K_{\mathrm{H2}}[\mathrm{H_s}])^{m_{\mathrm{H2}}}}{1+(K_{\mathrm{H2}}[\mathrm{H_s}])^{m_{\mathrm{H2}}}} \quad (3)$$

$Q_{\mathrm{H}i\max}$ is the total number of available proton binding sites within each distribution, $K_{\mathrm{H}i}$ is the median value of the affinity distribution for protons, $[\mathrm{H_s}]$ is the concentration of protons at the surface of the humic particle, and $m_{\mathrm{H}i}$ defines the width of the affinity distribution. In this expression, the binding heterogeneity is considered to be continuous for the two types of sites and is described by a Sips distribution (Figure 1c,d) (*1, 19, 21*). The Q_{H} proton binding equilibrium (eq 3) is obtained by integrating the Sips function with a Langmuir local isotherm for protons. The dependence on m of this convolution is shown graphically in Figure S1. The assumption of a continuous affinity distribution in the N-D model generates fewer systematic errors (*12, 22*) than representing the heterogeneity with discrete $\Delta pK_{\mathrm{H}i}$ values centered on the median $pK_{\mathrm{H}i}$ value, as in the WHAM model (*7*).

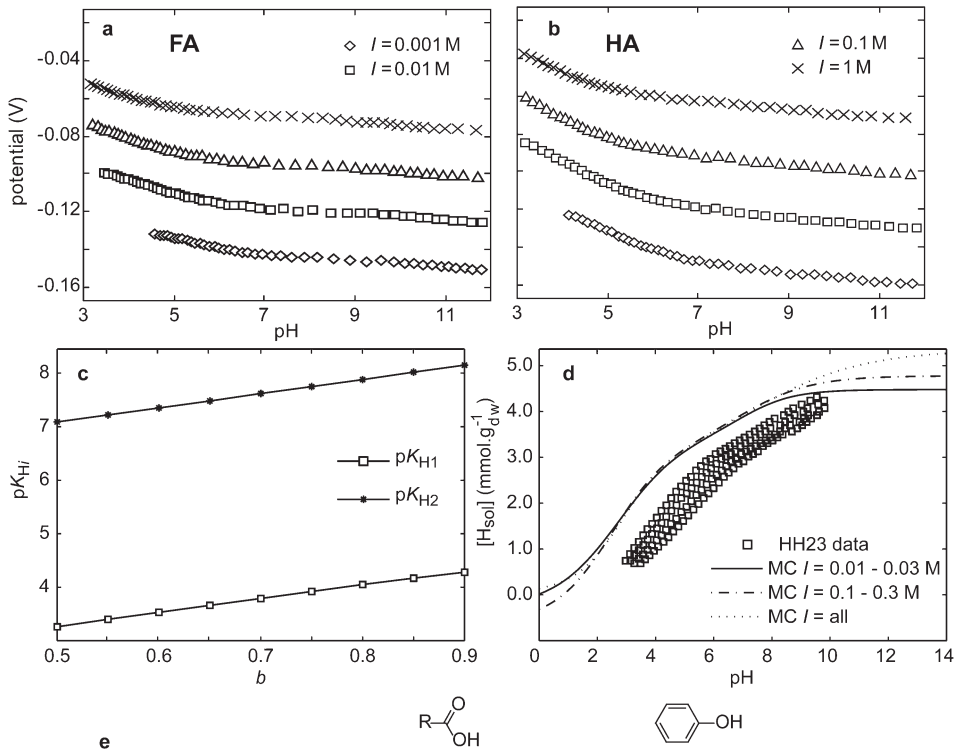
Generic NICA-Donnan Parameter Values

The N-D model has been applied to a titration data set of 25 fulvic acids (FA) and 23 humic acids (HA) titration curves to establish generic parameter values for humic substances (*14*). Below, the results from the previous study summarized in Table S2 are examined first qualitatively, then semiquantitatively, and finally statistically. For consistency, the original code name was preserved when discussing a specific titration data set, i.e., HH*x* and FH*x*, where the first letter stands for humic (H) or fulvic (F) acid, the second for proton, and *x* for the sample number. All precisions are given as 1σ standard deviation.

Physical Meaning. The generic Donnan potentials at the particles surface ($\Psi_s = \Psi_D$ in the Donnan model) as a function of pH near the binding sites (pH_s) were calculated from the generic MCs and $pH_s = pH - 0.434(F\Psi_s/RT)$ (*4, 8*). At constant pH, Ψ_s decreases with I, whereas it increases when pH decreases at constant I (Figure 2a,b). While these variations agree with other studies, the $\Psi_s(I)$ curves are all parallel at $pH < 5-6$ instead of converging to pH_{min}. Convergence has been obtained previously with both the impermeable sphere (*4, 8*) and Donnan (*6*) models. In addition, $\Psi_s(I = 1\ \mathrm{M}) \approx -30$ mV and $\Psi_s(I = 10^{-3}\ \mathrm{M}) \approx -120$ mV at pH 3, instead of the more realistic 0 to −60 mV range (*4*). These two problems question the validity of the best-fit values used to deduce the N-D generic parameters (*14*).

Evidence for Correlations between Parameters. The b parameter was chosen because of its physical relationship with Ψ_s. The b values ranged from 0.2 to 0.9 in the FA+HA series, which translates into $0.55 \le V_D \le 2.23$ L/kg at $I =$ 0.1 M after eq 2. The large variation of water content in the Donnan volume is expected to modify appreciably the proton activity, and hence the reactivity of the humic substance. The relationship between b and $pK_{\mathrm{H}i}$ is illustrated in Figure 2c with the fulvic acid FH22. This data set can be fit equally ($R^2 > 0.996$) with variable combinations of $\{b, pK_{\mathrm{H}i}\}$, however the $pK_{\mathrm{H2}} - pK_{\mathrm{H1}}$ difference is constant, and thus independent of b. These correlations were obtained by fitting simultaneously the titration curves at $I = 0.1$, $I = 0.02$, and $I = 0.004$ M for $0.5 \le b \le 0.9$ with 0.05 increments, and fixing the $Q_{\mathrm{H}i\max}$ and $m_{\mathrm{H}i}$ parameters to their generic values (Table S2).

Generalized Correlations. The previous approach is extended next to the complete FA + HA data set at all measured I values and the seven main N-D parameters,



e

	b	Q_{H1max}	pK_{H1}	m_{H1}	Q_{H2max}	pK_{H2}	m_{H2}	R^2
I = 0.01 - 0.03	0.33	3.76	2.74	0.39	1.01	7.05	0.49	0.9996
I = 0.1 - 0.3	0.36	3.82	2.49	0.44	1.56	7.05	0.36	0.9991
I = all	0.32	2.55	2.77	0.57	2.80	7.07	0.22	0.9982

FIGURE 2. (a,b) Electrostatic potentials as a function of pH calculated at four ionic strengths from the generic N-D parameters of Milne et al. (*14*). The generic MCs for FA and HA were calculated from Table 4 of Milne et al., then Ψ_s from $Q_H = f(pH_s)$ using eq 13 of Avena et al. (*8*). (c) Correlations between pK_{Hi} and b for the FH22 titration data as analyzed with the FIT code (*10*) and enforcing the $R^2 \geq 0.996$ reconstruction criterion recommended by Milne et al. (*14*). (d) Master-curves calculated by the FIT code for three combinations of HH23 measurements at I = 0.01, 0.03, 0.1, and 0.3. (e) Best-fit N-D parameters for the three MCs in (d). The three MCs are superimposed in the "data window", but not in the "fit window", also called "spectral window" (*16*).

TABLE 1. Correlation Matrix of Optimal Parameter Values Determined by Milne et al. (*14*)

	b	Q_{H1max}	pK_{H1}	m_{H1}	Q_{H2max}	pK_{H2}	m_{H2}	RMSE	R^2
b	1.00	0.14	0.58	0.02	−0.24	0.39	0.08	−0.33	0.32
Q_{H1max}	-	1.00	−0.61	−0.62	−0.15	0.30	0.43	0.43	−0.23
pK_{H1}	-	-	1.00	0.48	−0.21	−0.13	−0.13	−0.67	0.51
m_{H1}	-	-	-	1.00	0.32	−0.15	−0.55	−0.47	0.35
Q_{H2max}	-	-	-	-	1.00	0.49	−0.69	0.15	−0.01
pK_{H2}	-	-	-	-	-	1.00	−0.45	0.03	0.05
m_{H2}	-	-	-	-	-	-	1.00	0.33	−0.31
RMSE	-	-	-	-	-	-	-	1.00	−0.87
R^2	-	-	-	-	-	-	-	-	1.00

namely $Q_{Hi max}$, m_{Hi}, pK_{Hi}, and b, by calculating the correlation matrix for all best-fit values from Table 2 of Milne et al. (*14*) (Table S3). The pK_{H1} and b values are correlated to 58%, pK_{H1} and RMSE anticorrelated to 67%, $Q_{Hi max}$ and m_{Hi} to 62−69%, Q_{H1max} and pK_{H1} to 61%, and m_{H1} and m_{H2} to 55% (Table 1). The two goodness-of-fit parameters RMSE and R^2 are strongly anticorrelated (87%), meaning that one is sufficient to assess the fit quality for this data set. In the following only R^2 will be used, because RMSE depends on the degrees of freedom of the regression and this number is different in this study and in Milne et al. (*14*).

Influence of Ionic Strength on Correlations. The MC approach allows separating electrostatic and heterogeneity effects by removing the shift in pH due to Coulombic interactions. This is verified in Figure 2d which shows that the three MCs calculated independently from three combinations of measurements at I = 0.01, 0.03, 0.1, and 0.3 are essentially superimposed from pH 3 to 10. However, the heights of the plateaus on each side of the experimental pH interval vary, and thus do Q_{H1max} and Q_{H2max}. Calculation of the intrinsic parameters for the three MCs (Figure 2e) shows that Q_{H1max} is anticorrelated to pK_{H1} (62%) and m_{H1} (95%), and Q_{H2max} to m_{H2} (98%), in agreement with the correlation matrix (Table 1). The full score of parameter correlations makes it possible to increase one parameter value, while decreasing another, with no perceptible change in R^2. In addition, the system is loosely constrained because the fit parameters derived from the measurements at the four I

TABLE 2. Steps of the Convergence-Optimized Fit Procedure

data at $I = 0.1$ M	Step 1[a]	Step 2	Step 3
fixed parameters	pK_{H1}, pK_{H2}	b_{min}, Q_0, Q_{H1max}, Q_{H2max}	Q_0, Q_{H1max}, Q_{H2max}, m_{H1}, m_{H2}
optimized parameters	Q_0, Q_{H1max}, Q_{H2max}, m_{H1}, m_{H2}	pK_{H1}, pK_{H2}, m_{H1}, m_{H2}	pK_{H1}, pK_{H2}, b
determined parameters	Q_0, Q_{H1max}, Q_{H2max}	m_{H1}, m_{H2}	pK_{H1}, pK_{H2}, b

[a] Step 1: Donnan model is turned off.

taken together are not the average of the best-fit parameters determined separately at $I = [0.001, 0.03]$ and $I = [0.1, 0.3]$.

In the next section, a minimization procedure is proposed to increase the robustness of acid−base parameters derived from titration measurements. This procedure was established from the FA + HA titration curves measured at $I = 0.1$ M, whose first derivative presented two maxima, one for pK_{H1app} and a second for pK_{H2app}. In total, 6 FA and 13 HA satisfied the $pH_{max} > pK_{H2app}$ condition (Tables S4 and S5). For consistency, the new calculations were performed with the same FIT code software (*10*) as in the previous (*14*) study.

Convergence-Optimized Fit Procedure

Step 1. For a given ionic strength, the initial charge (IC) and $Q_{Hi max}$ are independent of electrostatic effects (i.e., b), as shown in Figure 1b and Table 1. Therefore, the total amounts of sites can be determined at constant I by choosing $pK_{H1} = pK_{H1app}$, $pK_{H2} = pK_{H2app}$, and $b = 0$. The pK_{Hiapp} values are obtained directly from the derivative of the data, and b is set equal to zero by turning off the Donnan model (*10*). Then $Q_{Hi max}$ and m_{Hi} are optimized, and IC indirectly from the equation $IC = Q_{H1max} + Q_{H2max} - Q_0$, with Q_0 adjustable. It is recommended to start out with the variables set to their generic values (Table S2). This step is analogous to the frequently used "pH-based" procedure (in comparison to the "model-based" methods) followed to estimate carboxyl and phenolic contents (*23*).

Step 2. If pK_{Hi}, m_{Hi}, and b are treated simultaneously as adjustable parameters, then b is optimized first by the N-D FIT code to generate the master curve, from which pK_{Hi} and m_{Hi} are deduced subsequently. Tests showed that the mathematical solution is nonphysical (Table S6). A way to prevent this pitfall is to fix b to the lowest possible value (b_{min}) over the meaningful interval [0.2−0.9], and to optimize pK_{Hi} and m_{Hi}. Because the system is already strongly constrained by IC and $Q_{Hi max}$, the N-D equation may have solutions only for certain values of b in this interval. Therefore, b_{min} may vary from one system to another.

Step 3. b and pK_{Hi} are optimized with m_{Hi} fixed. To verify that the minimum reached is the exact lowest point of the fit-procedure, m_{Hi} and pK_{Hi} can be optimized iteratively with b fixed to its previous optimal value.

The three steps are synthesized in Table 2. Optimal fits obtained with the FIT code (*10*) are shown in Figure S2, and results are detailed in Table S7, summarized in Table 3, and discussed next.

Evaluation of the New Optimal N-D Parameters

IC. The average value of IC is 0.46 ± 0.47 mmol/g for FA and −0.32 ± 0.52 mmol/g for HA. This data set had been adjusted already for the initial charge (*14*), and so the slight deviations from zero result from differences in the two fit procedures. Our two intervals of variation are close in value to the variability of Q_{H1max} obtained by Milne et al. (*14*) ($\sigma = 1.25$ for FA and 0.89 for HA), which engenders confidence in our calculations. The negative IC values were added to the best-fit Q_{H1max} values (Table 3), but not the positive IC values, because they are chemically unrealistic (*15*) and may be caused by errors in the calibration of pH during measurement.

$Q_{Hi max}$. Fulvic acids contain more carboxyl-type groups ($Q_{H1mean} = 5.08 \pm 1.46$ vs 3.20 ± 0.91), but similar amounts of phenolic-type groups as humic acids ($Q_{H2mean} = 2.05 \pm 2.19$ vs 2.30 ± 1.22, Tables 3, S8). The same conclusion was drawn by Tipping (*7*) with Model VI, Gustafsson (*9*) with the Stockholm Humic Model, Milne et al. (*14*) with N-D, and Ritchie and Perdue (*23*) with a modified Henderson−Hasselbalch model, meaning that site densities are reliably quantified regardless of the fit method. The low covariance of this parameter is also seen in Tables S8 and S9, as $Q_{Hi mean}$ values obtained in this study and in Milne et al. (*14*) for the same data set differ by less than 1 mmol/g_{dw}.

TABLE 3. New Generic N-D Values Obtained with the Convergence-Optimized Fit Procedure

	b	Q_{H1max}	pK_{H1}	m_{H1}	Q_{H2max}	pK_{H2}	m_{H2}	R^2
FA								
min	0.83	3.23	3.10	0.43	0.31	7.98	0.23	0.99806
max	0.95	6.53	4.12	0.92	6.37	8.86	1.35	0.99981
mean	0.89	5.08	3.54	0.54	2.05	8.37	0.65	0.99935
std. dev.	0.05	1.46	0.40	0.19	2.19	0.31	0.37	0.00064
HA								
min	0.70	1.73	3.04	0.46	1.04	7.58	0.14	0.99180
max	0.93	4.82	4.15	1.42	4.46	8.66	0.60	0.99987
mean	0.79	3.20	3.87	0.67	2.30	8.23	0.40	0.99822
std. dev.	0.07	0.91	0.29	0.26	1.22	0.37	0.14	0.00300

pK_{H1} and pK_{H2}. The average proton carboxylic constants obtained here (pK_{H1}[FA] = 3.54, pK_{H1}[HA] = 3.87) are in agreement with those obtained with Model VI (pK_{H1}[FA] = 3.20, pK_{H1}[HA] = 4.10) and the Stockholm Humic Model (pK_{H1}[FA] = 3.50, pK_{H1}[HA] = 4.10), but are 0.5 to 1.0 log unit higher than with the N-D model (pK_{H1}[FA] = 2.74/2.83, pK_{H1}[HA] = 3.22/3.74) (Table S1). As for phenolic constants, the new values (pK_{H2}[FA] = 8.37 ± 0.31, pK_{H2}[HA] = 8.23 ± 0.37) are similar to those obtained by Milne et al. (*14*) (pK_{H2}[FA] = 8.37 ± 0.83, pK_{H2}[HA] = 7.98 ± 0.96) and Avena et al. (*8*) (pK_{H2}[HA] = 7.50 − 8.50), but are at the lower end of the variation interval reported by Tipping (*7*) (pK_{H2}[FA] = 9.40 ± 0.78, pK_{H2}[HA] = 8.80 ± 0.23) and Gustafsson (*9*) (pK_{H2}[FA] = 8.75 ± 0.30, pK_{H2}[HA] = 8.95 ± 0.15) (Table S1). A higher variability of pK_{H2} relative to pK_{H1} does not imply that phenolic-type groups have a larger affinity distribution than carboxylic-type groups, because pK_{H2app} is close to the upper limit of titration curves ($pH_{max} \sim 10$). In some studies $pH_{max} < pK_{H2app}$, stretching even more the apparent heterogeneity of binding sites.

m_{H1} and m_{H2}. The new affinity distributions are $m_{H1} = 0.54 \pm 0.19$ for FA and 0.67 ± 0.26 for HA, and $m_{H2} = 0.65 \pm 0.38$ for FA and 0.40 ± 0.14 for HA (Table 3). The average heterogeneity of the proton binding for the same data is lower than in the previous study (*14*): m_{H1}[FA] = 0.46 ± 0.10, m_{H1}[HA] = 0.50 ± 0.07, m_{H2}[FA] = 0.53 ± 0.24, m_{H2}[HA] = 0.43 ± 0.21 (Table S10). The higher σ of m_{H1}[HA] (0.26 vs 0.07) and m_{H2}[FA] (0.38 vs 0.24) in this study is discussed in the Supporting Information. The two studies consistently show that carboxylic groups are more heterogeneous in FA than in HA (m_{H1}[FA] < m_{H1}[HA]), and phenolic groups more heterogeneous in HA (m_{H2}[HA] < m_{H2}[FA]) (Figure 3). Plazza et al. (*24*) reached the same conclusion (Table S10), but Tipping (*7*) determined that carboxylic and phenolic groups were both more heterogeneous in FA. This ambiguity again

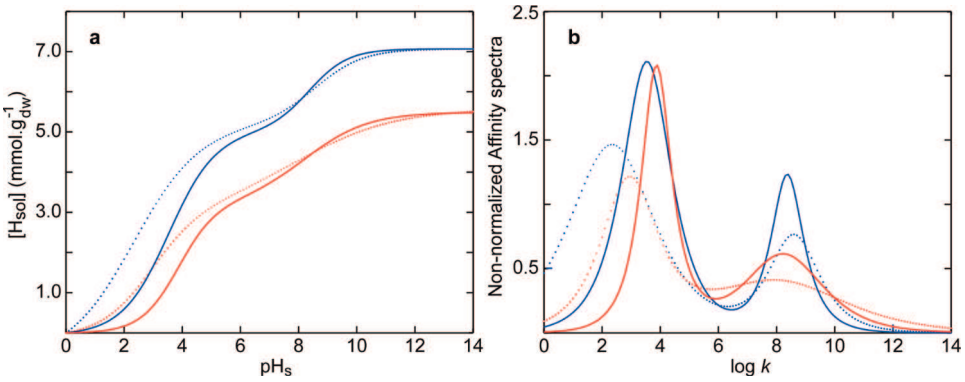


FIGURE 3. Master-curves (a) and affinity distribution of deprotonable sites (b) for FA (blue) and HA (red) calculated from the generic N-D parameters (dotted lines) (*14*), and results from the convergence-optimized fit procedure (full lines). The two approaches yield similar $Q_{H,max}$ and pK_{H2}. In this study, the average proton binding constants of carboxylic groups (pK_{H1}) are shifted by 0.94 log units for HA and 1.20 for FA relative to Milne's generic values, and the affinity distributions are narrower. The difference in heterogeneity of the proton binding is seen on the MCs by the difference in slope.

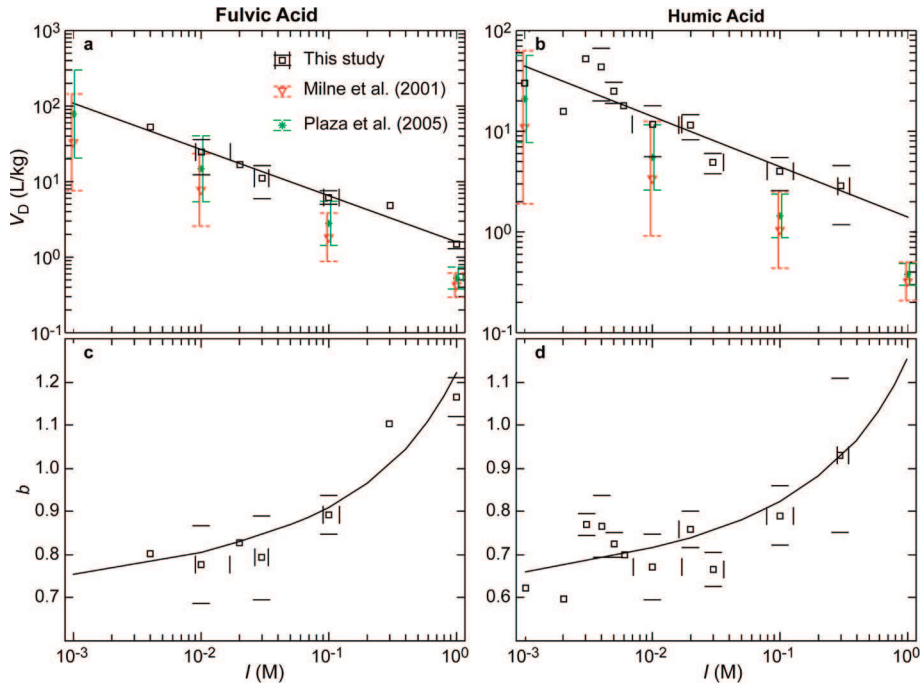


FIGURE 4. (a,b) Donnan volumes (V_D) with uncertainties against ionic strength calculated for the 6 FA and 13 HA titration curves at each ionic strength with the convergence-optimized fit procedure (black squares), and comparison with results from Milne et al. (*14*) (red triangles), and Plaza et al. (*24*) (green stars). For the two last studies, $\sigma(V_D,I)$ was calculated from $\sigma(b,I)$ (given in the main text) using eq 2. An error on I was introduced when the ionic strength varied during the titration. The regression equations (solid lines) of the new parameter values are $V_D[FA] = 1.67I^{-0.60}$, and $V_D[HA] = 1.42I^{-0.50}$. (c,d) Plots of the optimized b values against I (data from Table S11). The generic b values (solid lines) obtained from the regressed V_D values and eq 2 are: $b(I = 0.001\ M) = 0.75$, $b(I = 0.01\ M) = 0.81$, $b(I = 0.1\ M) = 0.91$, $b(I = 1\ M) = 1.22$ for FA, and $b(I = 0.001\ M) = 0.66$, $b(I = 0.01\ M) = 0.72$, $b(I = 0.1\ M) = 0.82$, $b(I = 1\ M) = 1.15$ for HA.

reflects the considerable uncertainty in the proton binding parameters obtained at high pH from titration data.

***b* and V_D at $I = 0.1$ M.** The new b values, and hence V_D values, are systematically higher than those previously published: b[FA] = 0.89 ± 0.05 and b[HA] = 0.79 ± 0.07 in this study compared to b[FA] = 0.67 and b[HA] = 0.43 in Benedetti et al. (*6*), b[FA] = 0.63 ± 0.16 and b[HA] = 0.51 ± 0.19 in Milne et al. (*14*), and b[FA] = 0.72 ± 0.15 and b[HA] = 0.58 ± 0.11 in Plaza et al. (*24*). However, in all cases b[FA] > b[HA], indicating that FA has a larger Donnan volume than HA at this ionic strength.

***b* and V_D at Other *I* Values.** The b parameter is usually considered independent of I, and is evaluated from the simulation of titration curves measured at different I using eq 2. However, the correlation between I and V_D in the log−log scale shows that the uncertainty on V_D should increase when I decreases, which is confirmed by the analysis of published results (Figure 4a,b). The hypothesis of b constant also partly explains the anomalous shape of the electrostatic potentials in the generic approach (*14*) (Figure 2a,b). V_D is more accurately deduced by optimizing b at every ionic strength, while keeping the other intrinsic parameters constant, as described in the convergence-optimized fit procedure (Figures S3−S5). Introducing variable $b(I)$ did not increase the total number of parameters because $Q_{H,max}$ were determined separately by a potential-independent regression of a NICA fit to data. When applied to the 6 FA and 13 HA, this strategy yields a variation of V_D of 41% for FA and 72% for HA at I = 0.1 M, and of 98% and 105% at I = 0.01 M (Figure 4a,b, Table S11). When b is constant, the precision on V_D for the same

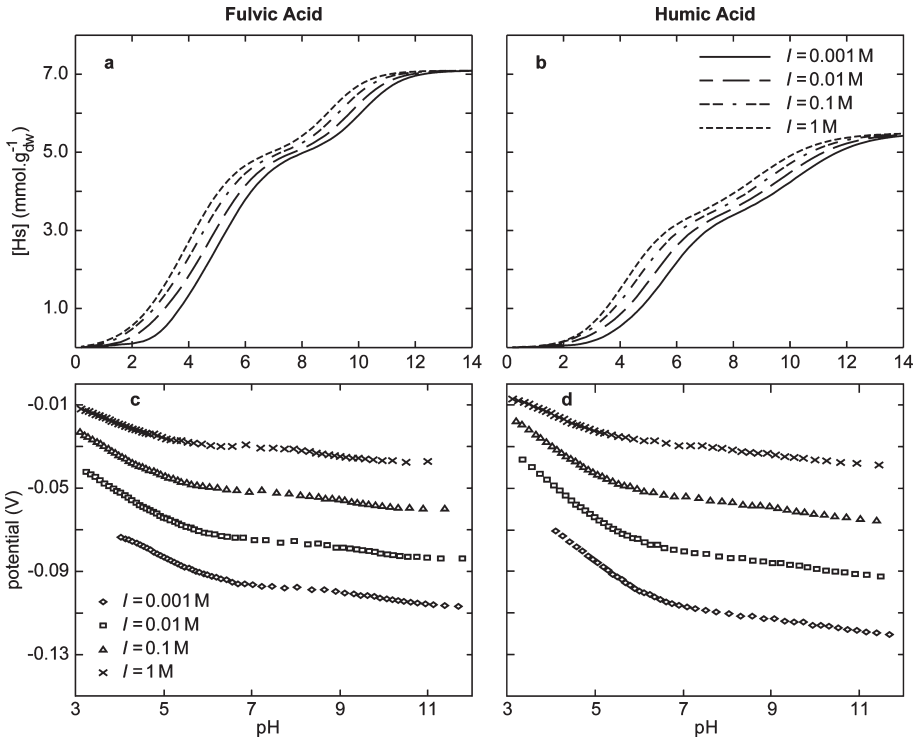


FIGURE 5. Generic titration curves (a,b) and surface potentials (c,d) for proton binding to FA and HA at four concentrations of background electrolyte.

data set decreases to 161% for FA and 211% for HA at $I = 0.1$ M, and to 269% for FA and 373% for HA at $I = 0.01$ M. Consequently, the precision on V_D is more immune to data processing in the new procedure and, in addition, can be increased by calculating its value at every ionic strength by a regression analysis of $V_D = f(I)$. The new procedure also allows the calculation of generic $b(I)$ values from the regressed values of mean $V_D(I)$ and eq 2 (Figures 4c,d, S5).

Lastly, generic titration curves and surface potentials can be calculated from the new generic N-D model parameters derived from the convergence-optimized fit procedure (Figure 5). The profiles of the Donnan potentials arising from the deprotonated groups decrease with pH as expected, but also converge now to values comprised between −10 and −70 mV at pH 3, in contrast to the previous (*14*) study (Figure 2), but in agreement with results by de Wit et al. (*4*), Benedetti et al. (*6*), and Avena et al. (*8*).

Concluding Remarks

Model bias is a serious problem in acid−base titrations, and has led in the past to incorrect pK values. Effects of model bias cannot be suppressed because there are more adjustable parameters than independent data points (*13*), but they can be minimized by adopting an iterative procedure in which a small number of parameters are calculated independently of the others in each cycle of the iteration.

In the N-D model, the total likelihood of eq 3 can be reduced by initially fixing $b = 0$. Using this procedure, the same solution is found irrespective of the starting values, and bias is minimized because three out of eight parameters are independent of b: Q_0, Q_{H1max}, and Q_{H2max}. Finally, the total covariance of the system cannot be reduced by measuring titration data at various ionic strengths because log I is correlated to log V_D (Figure 4a,b). Since the Donnan volume mainly depends on the ionic strength, its value at a desired I can be obtained by optimizing b at several I, then calculating the $V_D(I)$ values with eq 2, and regressing the $V_D = f(I)$ data points. Figure 4 shows that this procedure helps decrease the error on V_D, which otherwise can acquire unrealistic values when $Q_{HT} = f(\text{pH})$ multi data are corefined (*8*).

Acknowledgments

This manuscript benefitted from the constructive comments of two anonymous reviewers.

Supporting Information Available

Detailed calculations of the new optimal parameters obtained by the convergence-optimized fit procedure, plots of Sips functions, titration data with fits, and $V_D = f(I)$ and $b = f(I)$ results for each titration data. This material is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) De Wit, J. C. M.; van Riemsdijk, W. H.; Nederlof, M. M.; Kinniburgh, D. G.; Koopal, L. K. Analysis of ion binding on humic substances and the determination of intrinsic affinity distributions. *Anal. Chim. Acta* **1990**, *232*, 189–207.

(2) Marinsky, J. A.; Ephraim, J. H. A unified physicochemical description of the protonation and metal-ion complexation equilibria of natural organic-acids (humic and fulvic acids). 1. Analysis of the influence of polyelectrolyte properties on protonation equilibria in ionic media - Fundamental concepts. *Environ. Sci. Technol.* **1986**, *20*, 349–354.

(3) Bartschat, B. M.; Cabaniss, S. E.; Morel, F. M. M. Oligoelectrolyte model for cation binding by humic substances. *Environ. Sci. Technol.* **1992**, *26*, 284–294.

(4) De Wit, J. C. M.; van Riemsdijk, W. H.; Koopal, L. K. Proton binding to humic substances. 1. Electrostatic effects. *Environ. Sci. Technol.* **1993**, *27*, 2005–2014.

(5) Milne, C. J.; Kinniburgh, D. G.; de Wit, J. C. M.; van Riemsdijk, W. H.; Koopal, L. K. Analysis of metal-ion binding by a peat humic acid using a simple electrostatic model. *J. Colloid Interface Sci.* **1995**, *175*, 448–460.

(6) Benedetti, M. F.; van Riemsdijk, W. H.; Koopal, L. K. Humic substances considered as a heterogeneous Donnan gel phase. *Environ. Sci. Technol.* **1996**, *30*, 1805–1813.

(7) Tipping, E. Humic ion-binding model VI: An improved description of the interactions of protons and metal ions with humic substances. *Aquat. Geochem.* **1998**, *4*, 3–47.

(8) Avena, M. J.; Koopal, L. K.; van Riemsdijk, W. H. Proton binding to humic acids: Electrostatic and intrinsic interactions. *J. Colloid Interface Sci.* **1999**, *217*, 37–48.
(9) Gustafsson, J. P. Modeling the acid-base properties and metal complexation of humic substances with the Stockholm Humic Model. *J. Colloid Interface Sci.* **2001**, *244*, 102–112.
(10) Kinniburgh, D. *FIT user guide*; British Geological Survey: Keyworth, England, 1999.
(11) Christl, I.; Kretzschmar, R. Relating ion binding by fulvic and humic acids to chemical composition and molecular size. 1. Proton binding. *Environ. Sci. Technol.* **2001**, *35*, 2505–2511.
(12) Koopal, L. K.; Saito, T.; Pinheiro, J. P.; van Riemsdijk, W. H. Ion binding to natural organic matter: General considerations and the NICA-Donnan model. *Colloids Surf., A* **2005**, *265*, 40–54.
(13) Lenoir, T.; Manceau, A. Number of independent parameters in the potentiometric titration of humic substances. *Langmuir* **2010**, *26*, 3998–4003.
(14) Milne, C. J.; Kinniburgh, D. G.; Tipping, E. Generic NICA-Donnan model parameters for proton binding by humic substances. *Environ. Sci. Technol.* **2001**, *35*, 2049–2059.
(15) Matynia, A.; Lenoir, T.; Causse, B.; Spadini, L.; Jacquet, T.; Manceau, A. Semi-empirical proton binding constants for natural organic matter. *Geochim. Cosmochim. Acta* **2010**, *74*, 1836–1851.
(16) Westall, J. C.; Jones, J. D.; Turner, G. D.; Zachara, J. M. Models for association of metal ions with heterogeneous environmental sorbents. 1. Complexation of Co(II) by leonardite humic acid as a function of pH and $NaClO_4$ concentration. *Environ. Sci. Technol.* **1995**, *29*, 951–959.
(17) Benedetti, M. F.; van Riemsdijk, W. H.; Koopal, L. K.; Kinniburgh, D. G.; Gooddy, D. C.; Milne, C. J. Metal ion binding by natural organic matter: From the model to the field. *Geochim. Cosmochim. Acta* **1996**, *60*, 2503–2513.
(18) Kinniburgh, D. G.; Milne, C. J.; Benedetti, M. F.; Pinheiro, J. P.; Filius, J.; Koopal, L. K.; van Riemsdijk, W. H. Metal ion binding by humic acid: Application of the NICA-Donnan model. *Environ. Sci. Technol.* **1996**, *30*, 1687–1698.
(19) Garces, J. L.; Mas, F.; Cecilia, J.; Companys, E.; Galceran, J.; Salvador, J.; Puy, J. Complexation isotherms in metal speciation studies at trace concentration levels. Voltammetric techniques in environmental samples. *Phys. Chem. Chem. Phys.* **2002**, 3764–3773.
(20) Saito, T.; Nagasaki, S.; Tanaka, S.; Koopal, L. K. Electrostatic interaction models for ion binding to humic substances. *Colloids Surf., A* **2005**, *265*, 104–113.
(21) Orsetti, S.; Andrade, E. M.; Molina, F. V. Modeling ion binding to humic substances: Elastic polyelectrolyte network model. *Langmuir* **2010**, *26*, 3134–3144.
(22) Kinniburgh, D.; van Riemsdijk, W. H.; Koopal, L. K.; Benedetti, M. F. Ion binding to humic substances: Measurements, models, and mechanisms. In *Adsorption of Metals by Geomedia*; Jenne, E. A., Ed.; Elsevier: San Diego, CA, 1998; pp 484−521.
(23) Ritchie, J. D.; Perdue, E. M. Proton-binding study of standard and reference fulvic acids, humic acids, and natural organic matter. *Geochim. Cosmochim. Acta* **2003**, *67*, 85–96.
(24) Plaza, C.; Senesi, N.; Polo, A.; Brunetti, G. Acid-base properties of humic and fulvic acids formed during composting. *Environ. Sci. Technol.* **2005**, *39*, 7141–7146.

ES1015313

Supporting information for

Convergence-Optimized Procedure for Applying the NICA-Donnan Model to Potentiometric Titrations of Humic Substances

THOMAS LENOIR[1,2], ANTHONY MATYNIA[1], AND ALAIN MANCEAU[1]

[1]Mineralogy & Environments Group, LGCA, Université Joseph Fourier and CNRS, 38041 Grenoble cedex 9, France

[2]Laboratoire Central des Ponts et Chaussées (LCPC), Route de Bouaye, BP 4129, 44341 Bouguenais cedex, France

Number of pages: 14

Number of Tables: 11

Number of Figures: 5

Table S1. Compilation of intrinsic pK_{Hi} values for humic substances.

	Method	Humic	pK_{H1}		pK_{H2}	
		substance	average	std. dev.	average	std. dev.
NICA-Donnan	Direct titration[a]	FA	2.65 (2.74)	0.43 (0.57)	8.60 (8.37)	1.06 (0.83)
		HA	3.09 (3.22)	0.51 (0.54)	7.98 (7.98)	0.96 (0.96)
	Direct titration[b]	FA	2.83	0.36	7.02	0.18
		HA	3.74	1.09	7.68	0.60
	Direct titration[c]	HA	3.62	0.19	8.54	0.48
This study	Direct titration	FA	3.54	0.40	8.37	0.31
		HA	3.87	0.29	8.23	0.37
Model VI[d]	Direct titration	FA	3.20	0.13	9.40	0.78
		HA	4.10	0.16	8.80	0.23
Stockholm Humic Model[e]	Direct titration	FA	3.50	0.40	8.75	0.30
		HA	4.10	0.20	8.95	0.15
Other	RSU[f]	FA	3.80	0.20	9.87	0.31
		HA	3.73	0.13	9.83	0.23
	Direct titration[g]	FA	-	-	-	-
		HA	3.50 - 4.50	-	7.50 - 8.50	-
	IR[h]	FA	3.40 - 3.50	-	-	-
		HA	3.40 - 3.50	-	-	-

[a] (*1*). [b] (*2*). [c] (*3*). [d] (*4*). [e] (*5*). [f] (*6*). [g] (*7*). [h] (*8*). In parenthesis, results of Milne et al. (*1*) for the same reduced dataset as the one studied in this work.

Table S2. Optimal and generic N-D values determined by Milne et al. (*1*).

	b	Q_{H1max}	pK_{H1}	m_{H1}	Q_{H2max}	pK_{H2}	m_{H2}
FA							
min	0.29	2.64	2.00	0.27	0.55	7.19	0.17
max	0.94	8.76	3.81	0.65	7.77	10.91	0.96
std. dev.	0.16	1.25	0.43	0.09	1.94	1.06	0.21
mean	0.63	5.66	2.65	0.41	2.57	8.60	0.57
generic	0.57	5.88	2.34	0.38	1.86	8.60	0.53
HA							
min	0.21	1.93	1.99	0.38	0.76	6.06	0.14
max	0.84	4.73	3.90	0.89	5.39	10.06	0.86
std. dev.	0.19	0.89	0.51	0.13	1.37	0.96	0.21
mean	0.51	3.17	3.09	0.55	2.66	7.98	0.43
generic	0.49	3.15	2.93	0.50	2.55	8.00	0.26

Table S3. Optimal parameter values for the proton binding dataset studied by Milne et al. (*1*) fitted to each titration curve individually at all available *I* values, and used to calculate the correlation matrix reported in Table 1 of the main article. $Q_{Hi\max}$, m_{Hi}, and pK_{Hi} listed below are for a subdataset of 14 FA and 13 HA, not the 25 FA and 23 HA from the original dataset of Milne et al. (*1*), because some parameters were fixed to their generic NICA-Donnan values and, thus, cannot be used for statistical analysis.

Dataset	*RMSE*	R^2	*b*	Q_{H1max}	pK_{H1}	m_{H1}	Q_{H2max}	pK_{H2}	m_{H2}
FH02	0.0975	0.9946	0.64	5.64	2.28	0.39	2.62	9.43	0.45
FH05	0.0734	0.9977	0.72	5.79	2.89	0.45	0.55	8.15	0.85
FH06	0.0417	0.9987	0.57	5.15	2.35	0.32	1.46	8.56	0.67
FH07	0.0159	0.9997	0.57	3.83	2.46	0.45	2.21	9.18	0.32
FH09	0.1507	0.9802	0.42	5.02	2.00	0.41	0.94	7.19	0.90
FH10	0.0879	0.9966	0.48	5.92	2.20	0.30	0.92	7.19	0.96
FH16	0.0610	0.9980	0.65	6.66	2.40	0.42	1.74	8.37	0.69
FH17	0.0505	0.9981	0.73	5.96	2.80	0.55	1.93	8.64	0.60
FH18	0.0604	0.9977	0.71	6.59	2.64	0.48	6.55	10.91	0.36
FH19	0.0570	0.9974	0.70	6.32	2.49	0.49	2.35	9.26	0.41
FH20	0.0690	0.9979	0.74	6.20	2.68	0.38	2.12	9.10	0.45
FH21	0.1083	0.9929	0.29	5.34	2.00	0.34	5.05	9.45	0.41
FH22	0.0228	0.9996	0.70	3.12	3.81	0.48	1.60	7.63	0.71
FH23	0.0692	0.9983	0.32	2.64	2.43	0.65	7.77	7.30	0.17
HH06	0.0464	0.9987	0.35	4.40	2.67	0.48	0.76	6.06	0.86
HH08	0.0254	0.9994	0.56	3.16	3.19	0.52	2.33	7.97	0.38
HH09	0.0219	0.9994	0.34	2.33	2.89	0.54	4.25	8.81	0.25
HH11	0.1112	0.9926	0.21	4.25	2.31	0.44	1.95	7.20	0.53
HH13	0.0161	0.9997	0.62	2.32	3.36	0.65	3.27	8.75	0.20
HH14	0.0278	0.9995	0.72	4.58	3.38	0.42	0.92	7.75	0.59
HH16	0.0202	0.9994	0.76	2.57	3.89	0.54	0.88	8.02	0.51
HH17	0.0264	0.9997	0.70	4.73	3.44	0.49	1.23	7.57	0.50
HH18	0.0120	0.9999	0.69	2.94	3.76	0.55	2.40	8.07	0.24
HH19	0.0196	0.9995	0.68	2.68	3.85	0.47	1.06	8.14	0.57
HH20	0.0905	0.9940	0.67	3.76	2.74	0.42	5.39	10.06	0.23
HH21	0.0807	0.9956	0.84	1.95	3.90	0.52	3.23	8.3	0.58
HH23	0.0442	0.9982	0.32	2.55	2.77	0.57	2.80	7.07	0.22

Table S4. Titration data (I = 0.1 M) used to establish the convergence-optimized fit procedure. All have $pH_{max} > pK_{H2app}$.

Dataset	pK_{H1app}	pK_{H2app}
FH02	3.77	9.48
FH05	3.93	9.26
FH20	4.11	9.81
FH22	4.59	9.13
FH23	4.52	9.24
FH24	3.92	8.82
HH06	4.35	8.40
HH09	4.67	9.39
HH11	4.44	9.66
HH13	4.38	9.69
HH14	4.35	9.05
HH15	4.52	9.36
HH16	4.52	9.19
HH17	4.50	9.16
HH18	4.80	9.39
HH19	4.63	9.57
HH20	4.40	8.66
HH21	4.52	9.38
HH23	4.00	8.81

Table S5. Optimal parameter values reproduced from Table 2 of Milne et al. (*1*) for the same subdataset as the one analyzed by the convergence-optimized procedure in Table S6. Values obtained with the limited and full dataset are similar, showing that the data selection used for the convergence-optimized analysis did not introduce any statistical bias.

Subdataset	b	Q_{H1max}	pK_{H1}	m_{H1}	Q_{H2max}	pK_{H2}	m_{H2}	R^2
FH02	0.64	5.64	2.28	0.39	2.62	9.43	0.45	0.99460
FH05	0.72	5.79	2.89	0.45	0.55	8.15	0.85	0.99770
FH20	0.74	6.20	2.68	0.38	2.12	9.10	0.45	0.99790
FH22	0.70	3.12	3.81	0.48	1.60	7.63	0.71	0.99960
FH23	0.32	2.64	2.43	0.65	7.77	7.30	0.17	0.99830
FH24	0.49	6.57	2.37	0.39	1.94	8.60	0.53	0.99680
HH06	0.35	4.40	2.67	0.48	0.76	6.06	0.86	0.99870
HH09	0.34	2.33	2.89	0.54	4.25	8.81	0.25	0.99940
HH11	0.21	4.25	2.31	0.44	1.95	7.20	0.53	0.99260
HH13	0.62	2.32	3.36	0.65	3.27	8.75	0.20	0.99970
HH14	0.72	4.58	3.38	0.42	0.92	7.75	0.59	0.99950
HH15	0.65	3.68	2.95	0.38	1.97	8.00	0.26	0.99850
HH16	0.76	2.57	3.89	0.54	0.88	8.02	0.51	0.99940
HH17	0.70	4.73	3.44	0.49	1.23	7.57	0.50	0.99970
HH18	0.69	2.94	3.76	0.55	2.40	8.07	0.24	0.99990
HH19	0.68	2.68	3.85	0.47	1.06	8.14	0.57	0.99950
HH20	0.67	3.76	2.74	0.42	5.39	10.06	0.23	0.99400
HH21	0.84	1.95	3.90	0.52	3.23	8.30	0.58	0.99560
HH23	0.32	2.55	2.77	0.57	2.80	7.07	0.22	0.99820

FH	b	Q_{H1max}	pK_{H1}	m_{H1}	Q_{H2max}	pK_{H2}	m_{H2}	R^2
min	0.32	2.64	2.28	0.38	0.55	7.30	0.17	0.99460
max	0.74	6.57	3.81	0.65	7.77	9.43	0.85	0.99960
mean	0.60	4.99	2.74	0.46	2.77	8.37	0.53	0.99748
std. dev.	0.16	1.68	0.57	0.10	2.55	0.83	0.24	0.00168
HH	b	Q_{H1max}	pK_{H1}	m_{H1}	Q_{H2max}	pK_{H2}	m_{H2}	R^2
min	0.21	1.95	2.31	0.38	0.76	6.06	0.20	0.99260
max	0.84	4.73	3.90	0.65	5.39	10.06	0.86	0.99990
mean	0.58	3.29	3.22	0.50	2.32	7.98	0.43	0.99805
std. dev.	0.20	0.98	0.54	0.07	1.43	0.96	0.21	0.00241

Table S6. Optimal N-D parameters obtained in Step 2 of the proposed procedure with b adjusted simultaneously with pK_{Hi} and m_{Hi}. In the FIT algorithm, V_D is optimized first independently of the initial estimate of the binding constant. Therefore, in some cases the minimization algorithm converges to a local minimum with p$K_{Hi} \approx$ p$K_{Hi\text{initial}}$, which in the proposed procedure is pK_{Hi} = p$K_{Hi\text{app}}$ (Step 1). In other cases, V_D diverges to positive infinity (i.e., $V_D > \sim$10 L/kg, sample names in red), which is equivalent to $I = 0$ (no electrostatic effect), and p$K_{Hi} \geq$ p$K_{Hi\text{app}}$, which is meaningless.

Dataset	pK_{H1app}	pK_{H2app}	b	V_D [a]	Q_0	Q_{H1max}	pK_{H1}	m_{H1}	Q_{H2max}	pK_{H2}	m_{H2}
FH02	3.77	9.48	1.17	2.15E+01	6.66	4.78	3.54	0.44	1.14	9.04	0.53
FH05	3.93	9.26	1.33	4.67E+01	6.21	5.86	3.79	0.41	0.31	9.04	1.21
FH20	4.11	9.81	1.38	5.83E+01	7.90	4.98	4.02	0.45	1.87	9.61	0.46
FH22	4.59	9.13	5.44	7.47E+09	4.98	3.25	4.59	0.39	1.78	9.12	0.54
FH23	4.52	9.24	1.39	6.15E+01	9.60	2.37	4.45	0.71	6.37	9.01	0.22
FH24	3.92	8.82	1.31	4.23E+01	7.37	6.40	3.76	0.40	0.84	8.56	0.61
HH06	4.35	8.40	2.89	6.12E+04	5.61	3.73	4.35	0.51	1.51	8.40	0.46
HH09	4.67	9.39	5.93	7.28E+10	4.66	2.51	4.67	0.44	1.95	9.39	0.43
HH11	4.44	9.66	9.50	9.98E+17	6.88	3.11	4.56	0.45	3.13	9.90	0.21
HH13	4.38	9.69	4.16	2.05E+07	4.70	2.96	4.38	0.45	1.74	9.69	0.36
HH14	4.35	9.05	7.10	1.61E+13	5.56	4.34	4.35	0.39	1.04	9.04	0.55
HH15	4.52	9.36	6.30	4.00E+11	5.96	2.23	4.52	0.58	2.67	9.36	0.24
HH16	4.52	9.19	5.23	2.82E+09	3.65	2.27	4.52	0.53	1.22	9.19	0.35
HH17	4.50	9.16	5.99	9.73E+10	5.99	4.63	4.50	0.42	1.18	9.16	0.48
HH18	4.80	9.39	3.68	2.25E+06	4.83	3.44	4.80	0.43	1.28	9.39	0.44
HH19	4.63	9.57	8.21	2.61E+15	4.13	2.12	4.63	0.51	1.75	9.57	0.32
HH20	4.40	8.66	0.96	8.22E+00	7.08	1.29	4.13	1.24	4.46	8.03	0.20
HH21	4.52	9.38	1.38	5.63E+01	5.71	1.25	4.49	0.72	3.98	9.24	0.42
HH23	4.00	8.81	0.69	2.38E+00	5.95	2.73	3.02	0.46	4.05	7.57	0.14

[a] in L/kg.

Table S7. Optimal NICA-Donnan parameters values obtained with the convergence-optimized fit procedure at I = 0.1 M.

Dataset	b	Q_0	Q_{H1max} sim.[a]	pK_{H1}	m_{H1}	Q_{H2max}	pK_{H2}	m_{H2}	IC	Q_{H1max}[b]	R^2
FH02	0.86	6.66	4.78	3.10	0.47	1.14	8.49	0.56	-0.74	5.52	0.99958
FH05	0.95	6.21	5.86	3.34	0.44	0.31	8.47	1.35	-0.04	5.90	0.99957
FH20	0.90	7.90	4.98	3.50	0.52	1.87	8.86	0.49	-1.04	6.02	0.99806
FH22	0.88	4.98	3.25	4.12	0.45	1.78	8.31	0.59	0.05	3.25	0.99962
FH23	0.83	9.60	2.37	3.93	0.92	6.37	8.11	0.23	-0.86	3.23	0.99976
FH24	0.94	7.37	6.40	3.26	0.43	0.84	7.98	0.67	-0.13	6.53	0.99953
HH06	0.93	5.61	3.73	3.89	0.63	1.51	7.63	0.53	-0.37	4.10	0.99981
HH09	0.81	4.66	2.51	4.15	0.53	1.95	8.48	0.48	-0.20	2.71	0.99922
HH11	0.77	6.88	3.11	3.76	0.66	3.13	8.53	0.32	-0.65	3.76	0.99180
HH13	0.77	4.70	2.96	3.74	0.54	1.74	8.66	0.40	0.00	2.96	0.99982
HH14	0.90	5.56	4.34	3.79	0.46	1.04	8.22	0.60	-0.19	4.53	0.99971
HH15	0.76	5.96	2.23	3.95	0.77	2.67	8.36	0.26	-1.06	3.29	0.99980
HH16	0.76	3.65	2.27	3.98	0.67	1.22	8.29	0.39	-0.16	2.43	0.99983
HH17	0.87	5.99	4.63	3.87	0.50	1.18	8.26	0.54	-0.18	4.82	0.99977
HH18	0.79	4.83	3.44	4.14	0.52	1.28	8.41	0.50	-0.11	3.55	0.99977
HH19	0.75	4.13	2.12	4.09	0.64	1.75	8.62	0.35	-0.26	2.39	0.99987
HH20	0.76	7.08	1.29	3.87	1.42	4.46	7.68	0.21	-1.33	2.62	0.99227
HH21	0.70	5.71	1.25	4.09	0.93	3.98	8.31	0.46	-0.48	1.73	0.99534
HH23	0.70	5.95	2.73	3.04	0.46	4.05	7.58	0.14	0.84	2.73	0.99981

[a] Best-fit value. [b] Best-fit value + initial charge (see main text for explanation).

The heterogeneity parameters of phenolic-type groups for FH05 and of carboxylic-type groups for HH20 are both higher than 1, which is non-physical (Figure S1) and also non-chemical because a monoacid (m = 1) cannot lose more than one proton. For these data, some best-fit parameter values are unrealistic because the experimental error is high. The high m_{H2} (FH05) value is balanced out by a low Q_{H2max} value of 0.31 mmol/g. HH20 has a Q_{H2max} value unusually high (4.46 mmol/g) and a ΔpK unusually small (3.8). In Milne et al. (*1*), Q_{H2max} and pK_{H2} are both too high: 5.39 mmol/g and 10.06, respectively. All best-fit values were included, however, in the calculations of the means and standard deviations, because they give an indication of the overall precision of the proposed method, which depends on the covariance of the fit parameters and data quality. In the study by Van Zomeren et al. (*9*), the upper bound of m was fixed to 1, suggesting that these authors faced similar problems in their data analysis.

Table S8. Proton binding site densities for the same FA+HA dataset as obtained in this study (TS) and by Milne et al. (*1*).

	Q_{H1max}		Q_{H2max}			Q_{H1max}		Q_{H2max}	
FA	**TS**	**Milne**	**TS**	**Milne**	**HA**	**TS**	**Milne**	**TS**	**Milne**
mean	5.08	4.99	2.05 (1.19)	2.77 (1.77)	mean	3.20	3.29	2.30	2.32
std. dev.	1.46	1.68	2.19 (0.65)	2.55 (0.77)	std. dev.	0.91	0.98	1.22	1.43
Differences[a]					**Differences**				
min	0.04		0.18		min	0.05		0.05	

max	0.59	1.48	max	1.14	2.30
mean	0.19	0.77	mean	0.38	0.90

[a] Differences calculated for each titration curve. Results obtained without the FH23 curve are in parenthesis, see Table S9 for details.

Table S9. Site density obtained with the generic (*1*) and the convergence-optimized procedure. Results expressed in mmol/g_{dw}.

	Q_{H1max}			Q_{H2max}		
Dataset	**This study**	**Milne et al.**	**Difference**	**This study**	**Milne et al.**	**Difference**
FH02	5.52	5.64	0.12	1.14	2.62	1.48
FH05	5.90	5.79	0.11	0.31	0.55	0.24
FH20	6.02	6.20	0.18	1.87	2.12	0.25
FH22	3.25	3.12	0.13	1.78	1.60	0.18
FH23	3.23	2.64	0.59	6.37	7.77	1.40
FH24	6.53	6.57	0.04	0.84	1.94	1.10
mean	5.08	4.99		2.05	2.77	
std. dev.	1.46	1.68		2.19	2.55	
HH06	4.10	4.40	0.30	1.51	0.76	0.75
HH09	2.71	2.33	0.38	1.95	4.25	2.30
HH11	3.76	4.25	0.49	3.13	1.95	1.18
HH13	2.96	2.32	0.64	1.74	3.27	1.53
HH14	4.53	4.58	0.05	1.04	0.92	0.12
HH15	3.29	3.68	0.39	2.67	1.97	0.70
HH16	2.43	2.57	0.14	1.22	0.88	0.34
HH17	4.82	4.73	0.09	1.18	1.23	0.05
HH18	3.55	2.94	0.61	1.28	2.4	1.12
HH19	2.39	2.68	0.29	1.75	1.06	0.69
HH20	2.62	3.76	1.14	4.46	5.39	0.93
HH21	1.73	1.95	0.22	3.98	3.23	0.75
HH23	2.73	2.55	0.18	4.05	2.8	1.25
mean	3.20	3.29		2.30	2.32	
std. dev.	0.91	0.98		1.22	1.43	

Table S10. Average proton binding heterogeneity for FA+HA as derived from the N-D model.

	Method	**Humic**	m_{H1}		m_{H2}	
		substance	**average**	**std. dev.**	**average**	**std. dev.**
Milne et al. (*1*)	Direct titration	FA	0.41 (0.46)	0.09 (0.10)	0.57 (0.53)	0.21 (0.24)
		HA	0.55 (0.50)	0.13 (0.07)	0.43 (0.43)	0.21 (0.21)
Plaza et al. (*2*)	Direct titration	FA	0.45	0.05	0.47	0.42
		HA	0.57	0.01	0.21	0.03
This study	Direct titration	FA	0.54	0.19	0.65	0.37
		HA	0.67	0.26	0.40	0.14

In parenthesis, results of Milne et al. (*1*) for the same reduced dataset as the one studied in this work.

Table S11. Optimal *b* values as a function of *I*.

Dataset	*I* (M)	uncertainty on *I* (%)		*b*	V_D (L)	R^2
		min	max			
FH02	1	-	-	1.20	1.57	0.9966
	0.1	-	-	0.86	5.16	
	0.01	-	-	0.77	20.74	
FH05	1	-	-	1.13	1.36	0.9977
	0.1	-	-	0.95	7.78	
	0.01	-	-	0.84	32.72	
FH20	0.1	9.40%	-	0.90	6.32	0.9981
	0.03	-	12.33%	0.86	14.92	
	0.01	-	68.00%	0.85	34.57	
FH22	0.1	-	20.00%	0.88	5.86	0.9990
	0.02	-	-	0.83	17.09	
	0.004	-	-	0.80	53.19	
FH23	0.1	5.00%	-	0.83	4.57	0.9974
	0.03	3.33%	-	0.68	5.23	
	0.01	10.00%	-	0.65	8.99	
FH24	0.3	-	0.90%	1.10	4.78	0.9982
	0.1	-	3.30%	0.94	7.53	
	0.03	-	12.93%	0.84	12.90	
HH06	0.1	-	-	0.93	7.28	0.9991
	0.01	-	-	0.70	12.18	
	0.001	-	-	0.62	30.09	
HH09	0.3	6.00%	16.67%	1.06	4.05	0.9987
	0.1	21.00%	-	0.81	4.23	
	0.01	20.00%	70.00%	0.66	9.70	
	0.002	-	-	0.59	15.83	
HH11	0.1	-	12.50%	0.77	3.48	0.9929
	0.03	-	12.00%	0.66	4.75	
	0.01	-	12.00%	0.60	6.16	
HH13	0.1	-	20.00%	0.77	3.39	0.9995
	0.02	10.00%	-	0.72	8.73	
	0.006	-	-	0.70	17.79	
HH14	0.1	-	10.00%	0.90	6.22	0.9996
	0.02	-	-	0.82	16.56	
	0.004	-	-	0.82	59.60	
HH15	0.1	-	20.00%	0.76	3.37	0.9989
	0.02	-	-	0.72	8.95	
	0.005	-	-	0.70	20.75	
HH16	0.1	-	20.00%	0.76	3.31	0.9992
	0.02	-	-	0.74	10.05	
	0.003	-	-	0.75	45.00	
HH17	0.1	-	10.00%	0.87	5.47	0.9990
	0.02	-	-	0.81	15.57	
	0.003	-	-	0.79	59.68	
HH18	0.1	-	-	0.79	3.77	0.9997
	0.02	20.00%	-	0.75	10.85	
	0.005	-	-	0.75	29.01	
HH19	0.1	-	20.00%	0.75	3.21	0.9990
	0.02	-	-	0.73	9.37	
	0.004	-	-	0.71	26.67	
HH20	0.1	-	24.00%	0.76	3.36	0.9940
	0.03	-	18.33%	0.68	5.05	

	0.01	19.00%	-	0.72	14.55	
HH21	0.1	-	27.00%	0.71	2.57	0.9947
	0.03	-	20.00%	0.71	6.26	
	0.01	30.00%	-	0.78	21.88	
HH23	0.3	-	0.33%	0.80	1.67	0.9977
	0.1	-	2.00%	0.70	2.49	
	0.03	-	6.67%	0.62	3.57	
	0.01	-	40.00%	0.58	5.38	

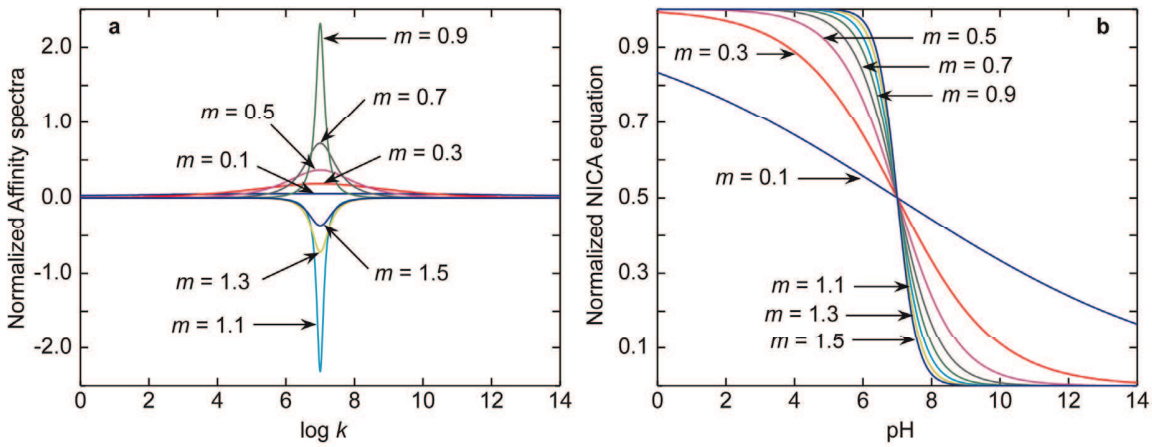


Figure S1. (a) Variation with *m* of an affinity distribution centered at log *k* = 7 and modeled with a Sips function. For *m* = 1, the distribution is a Dirac delta function, above it is negative.
(b) Normalized NICA isotherms obtained by convoluting a Langmuir local isotherm with the Sips functions in (a).

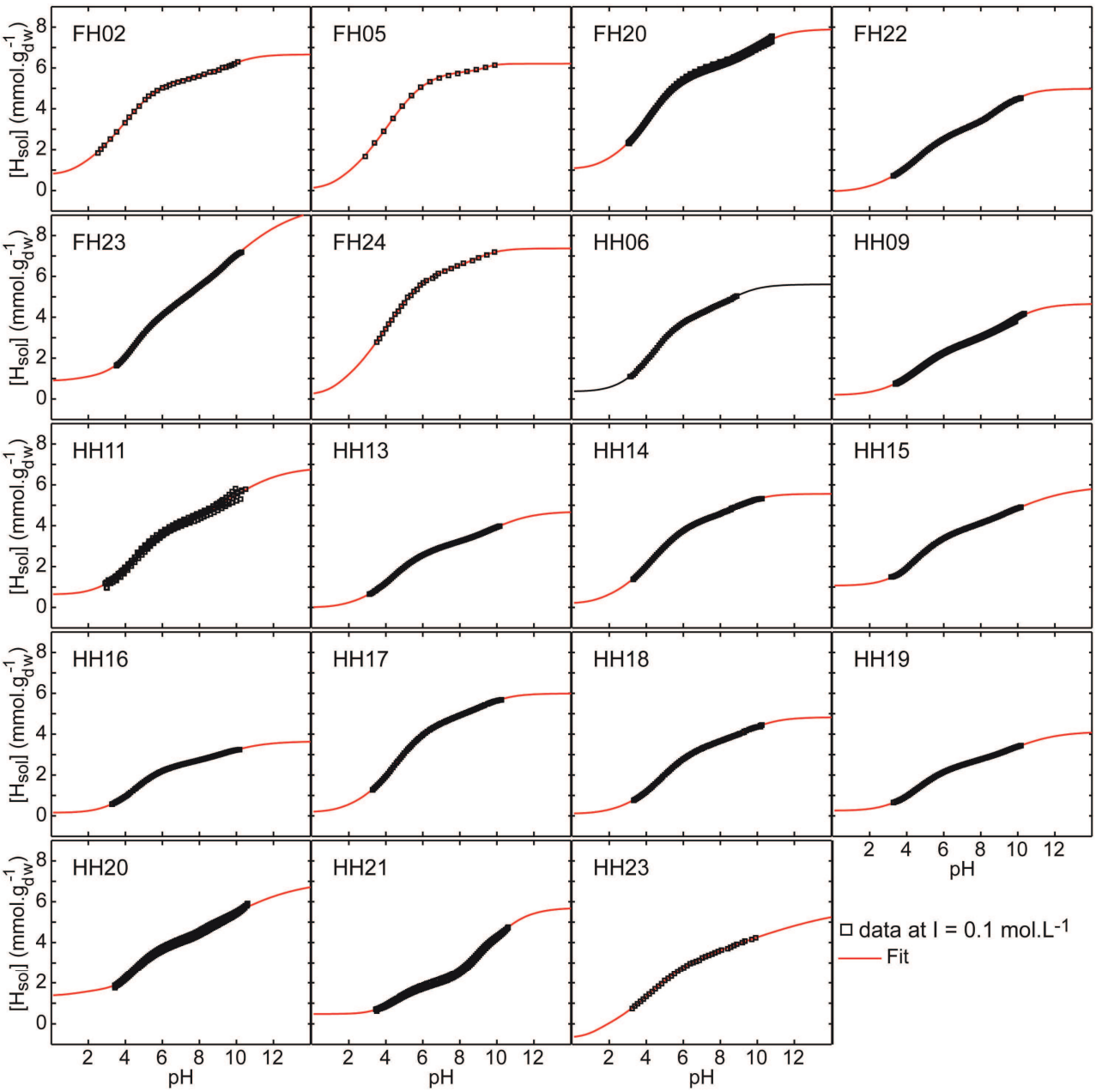


Figure S2. Optimal fits of titration data at I = 0.1 M obtained with the convergence-optimized fit procedure. *IC* was calculated afterwards from $IC = Q_{H1max} + Q_{H2max} - Q_0$, and the final Q_{Himax} values corrected from *IC*.

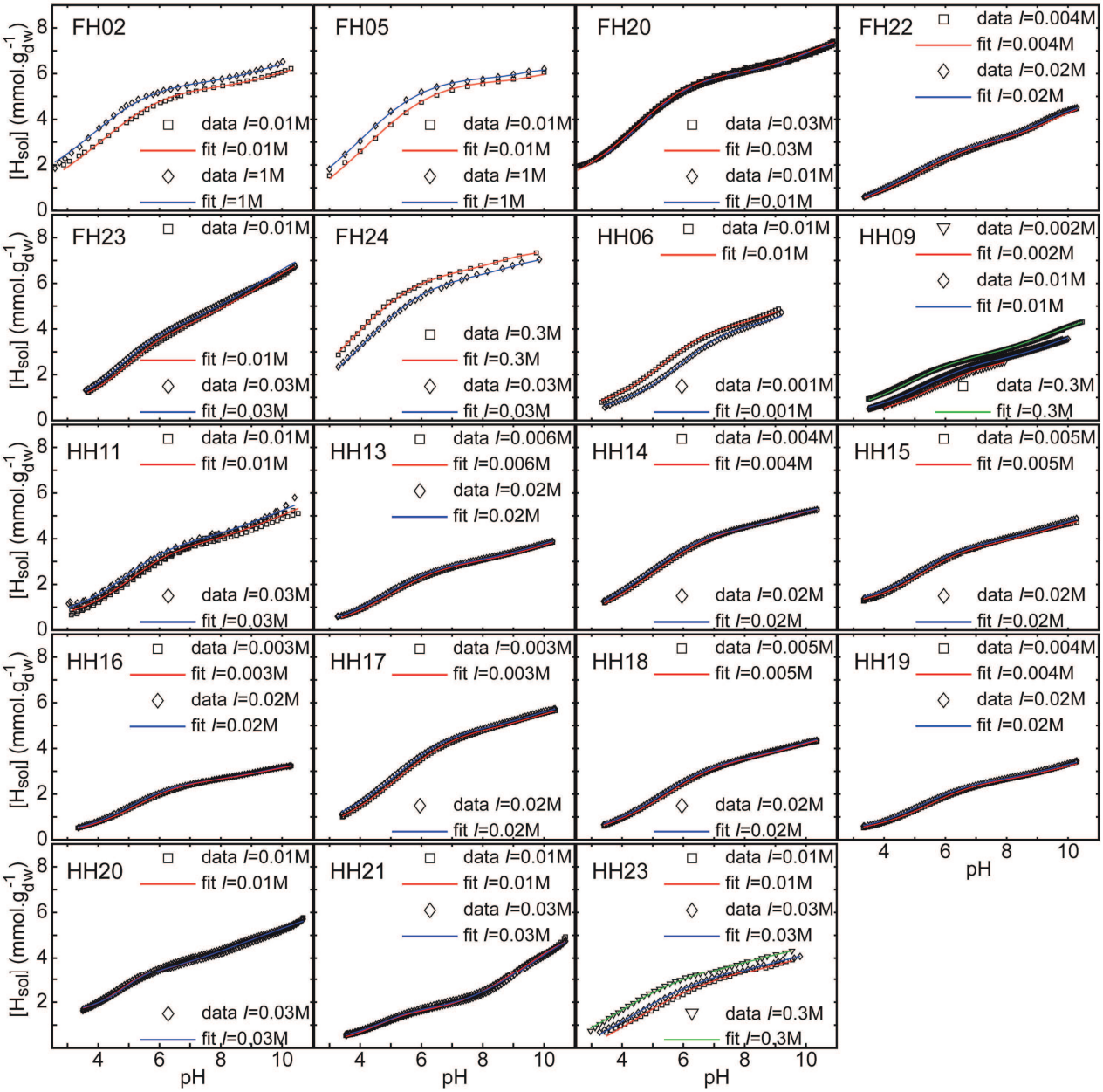


Figure S3. Optimal fits of titration data for *I* values different from 0.1 M.

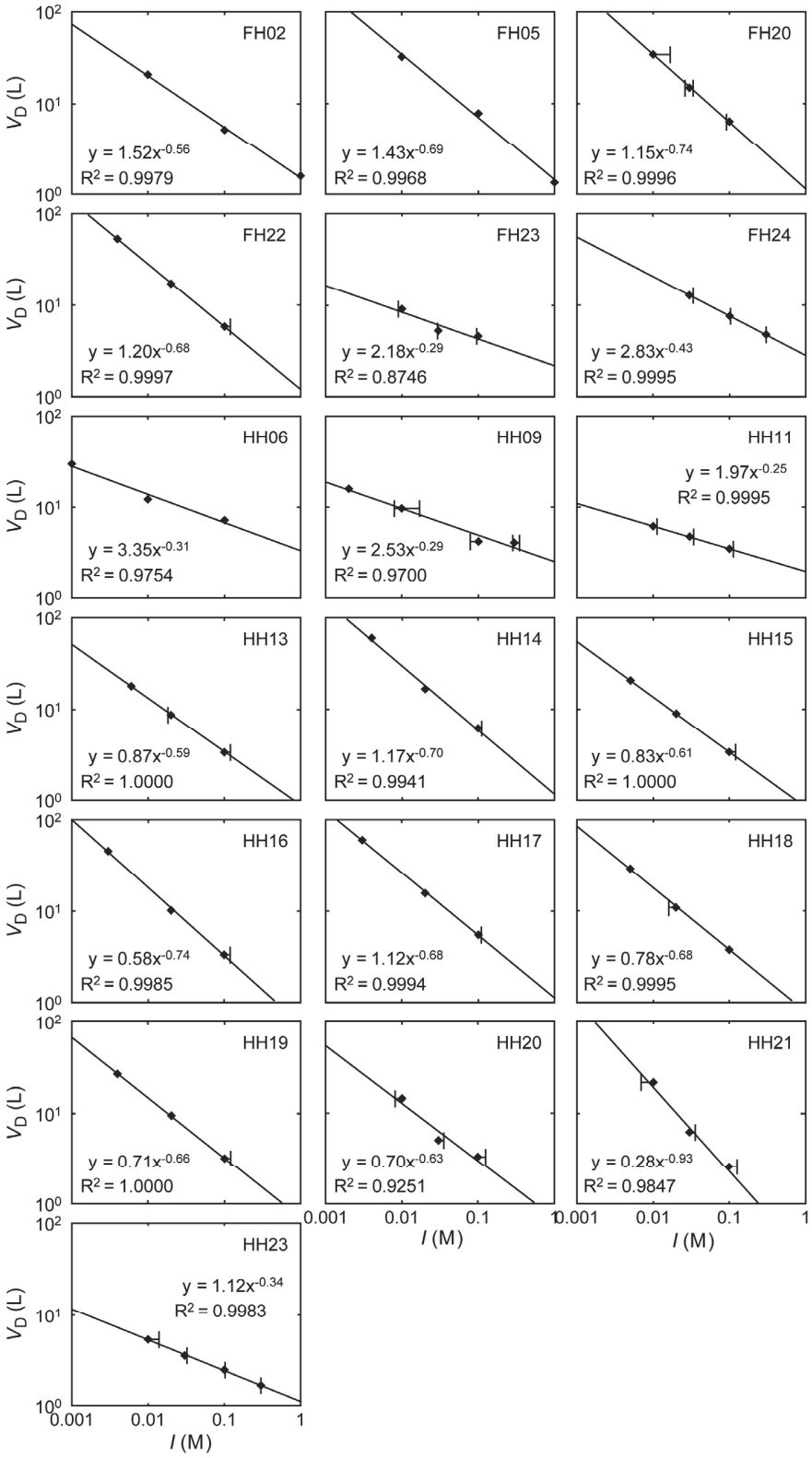


Figure S4. Relationships between the Donnan volume (V_D) of the humic substances and the ionic strength. For each data, the dependence on ionic strength can be written as linear functions of I and V_D in the log-log scale.

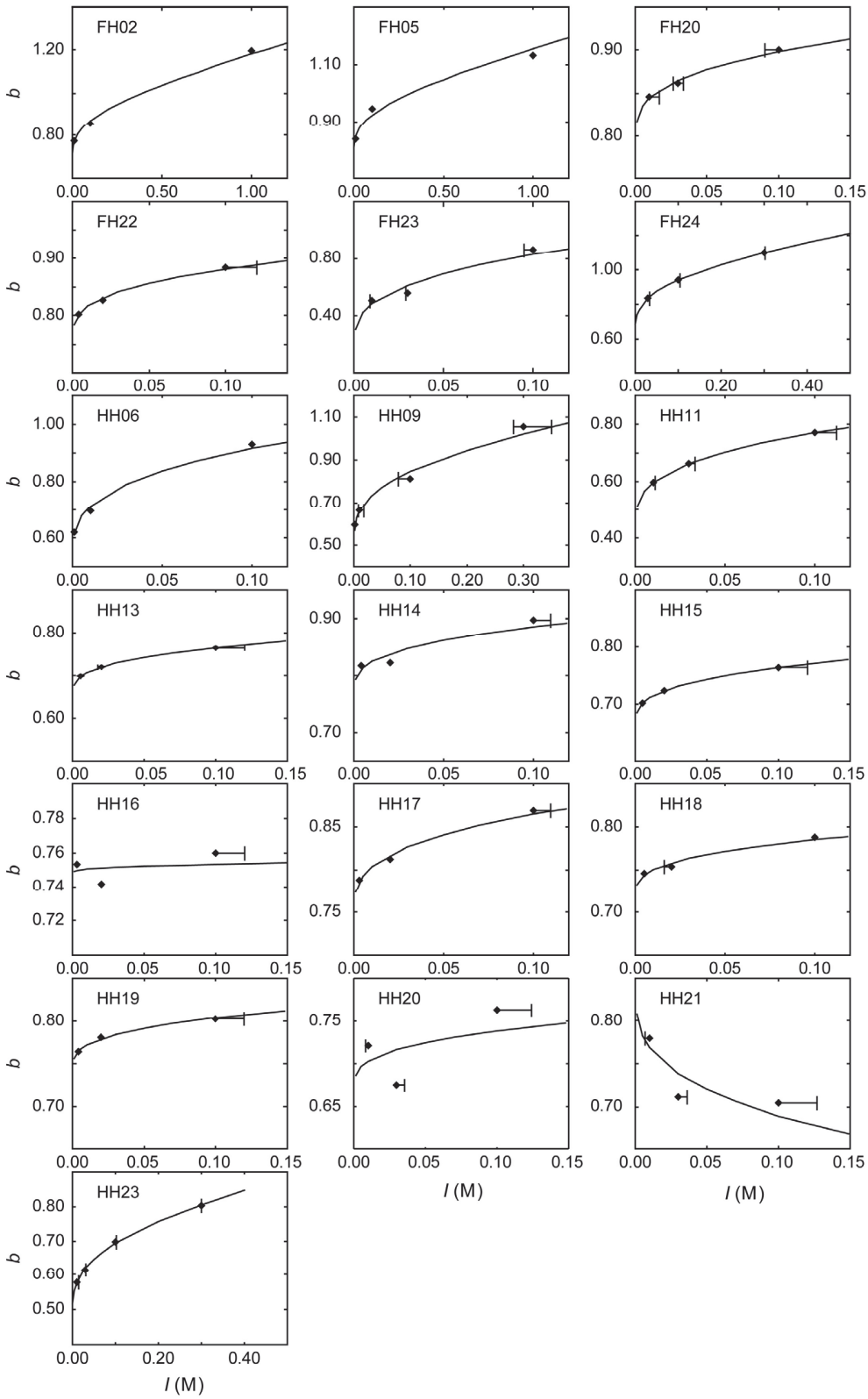


Figure S5. Relationships between the *b* parameter of the humic substances and the ionic strength. The solid lines were calculated from the regression analyses in Figure S4 and $b = \left(1 + \log V_{\mathrm{D}}^{\mathrm{R}}\right)/\left(1 - \log I\right)$, with $V_{\mathrm{D}}^{\mathrm{R}}$ the regressed Donnan volume.

Cited References

1. Milne, C. J.; Kinniburgh, D. G.; Tipping, E. Generic NICA-Donnan model parameters for proton binding by humic substances. *Environ. Sci. Technol.* **2001**, *35*, 2049-2059.
2. Plaza, C.; Senesi, N.; Polo, A.; Brunetti, G. Acid-base properties of humic and fulvic acids formed during composting. *Environ. Sci. Technol.* **2005**, *39*, 7141-7146.
3. Drosos, M.; Jerykiewicz, M.; Deligiannakis, Y. H-binding groups in lignite vs. soil humic acids: NICA-Donnan and spectroscopic parameters. *J. Colloid Interf. Sci.* **2009**, *332*, 78-84.
4. Tipping, E. Humic Ion-Binding Model VI: An improved description of the interactions of protons and metal ions with humic substances. *Aquatic Geochem.* **1998**, *4*, 3-47.
5. Gustafsson, J. P. Modeling the acid-base properties and metal complexation of humic substances with the Stockholm Humic Model. *J. Colloid Interf. Sci.* **2001**, *244*, 102-112.
6. Matynia, A.; Lenoir, T.; Causse, B.; Spadini, L.; Jacquet, T.; Manceau, A. Semi-empirical proton binding constants for natural organic matter. *Geochim. Cosmochim. Acta* **2010**, *74*, 1836-1851.
7. Avena, M. J.; Koopal, L. K.; van Riemsdijk, W. H. Proton binding to humic acids: Electrostatic and intrinsic interactions. *J. Colloid Interf. Sci.* **1999**, *217*, 37-48.
8. Hay, M. B.; Myneni, S. C. B. Structural environments of carboxyl groups in natural organic molecules from terrestrial systems. Part 1: Infrared spectroscopy. *Geochim. Cosmochim. Acta* **2007**, *71*, 3518-3532.
9. Van Zomeren, A.; Costa, A.; Pinheiro, J. P.; Comans, R. N. J. Proton binding properties of humic substances originating from natural and contaminated materials. *Environ. Sci. Technol.* **2009**, *43*, 1393-1399.

Bibliographie

Abollino, O., Aceto, M., Sacchero, G., Sarzanini, C., et Mentasti, E. (1995). "Determination of copper, cadmium, iron, manganese, nickel and zinc in Antarctic sea water. Comparison of electrochemical and spectroscopic procedures." *Analytica Chimica Acta*, 305(1-3), 200-206.

Adriano, D. C. (1986). *Trace elements in the terrestrial environment*, Springer-Verlag., New York.

Ahsanullah, M., et Florence, T. M. (1984). "Toxicity of copper to the marine amphipod Allorchestes-compressa in the presence of water-and lipid-soluble ligands." *Marine Biology*, 84(1), 41-45.

Albright, L. J., Wentworth, J. W., et Wilson, E. M. (1972). "Technique for measuring metallic salt effects upon the indigenous heterotrophic microflora of a natural water." *Water Research*, 6(12), 1589-1596.

Ali, N. A., Bernal, M. P., et Ater, M. (2002). "Tolerance and bioaccumulation of copper in Phragmites australis and Zea mays." *Plant and Soil*, 239(1), 103-111.

Alloway, B. J. (1995). *Heavy Metals in Soils*, Blackie Academic & Professional, London

Alt, E. R., Sternlieb, I., et Goldfischer, S. (1990). "The cytopathology of metal overload." *International Review of Experimental Pathology*, 31, 165-188.

Amaravadi, R., Glerum, D. M., et Tzagoloff, A. (1997). "Isolation of a cDNA encoding the human homolog of COX17, a yeast gene essential for mitochondrial copper recruitment." *Human Genetics*, 99(3), 329-333.

Amira, S., Spangberg, D., et Hermansson, K. (2005). "Distorted five-fold coordination of Cu2+(aq) from a Car-Parrinello molecular dynamics simulation." *Physical Chemistry Chemical Physics*, 7(15), 2874-2880.

Ankudinov, A. L., et Rehr, J. J. (1997). "Relativistic calculations of spin-dependent x-ray-absorption spectra." *Physical Review B*, 56(4), R1712-R1715.

Arru, L., Rognoni, S., Baroncini, M., Bonatti, P. M., et Perata, P. (2004). "Copper localization in Cannabis sativa L. grown in a copper-rich solution." *Euphytica*, 140(1-2), 33-38.

Arthur, J. W., et Leonard, E. N. (1970). "Effects of copper on *Gammarus pseudolimnaeus*, *Physa integra*, and *Campeloma decisum* in soft water." *Journal of Fisheries Research Board of Canada*, 27, 1277-1283.

Ashworth, D. J., et Alloway, B. J. (2007). "Complexation of copper by sewage sludge-derived dissolved organic matter: Effects on soil sorption behaviour and plant uptake." *Water Air and Soil Pollution*, 182(1-4), 187-196.

Avena, M. J., Koopal, L. K., et van Riemsdijk, W. H. (1999). "Proton binding to humic acids: Electrostatic and intrinsic interactions." *Journal of Colloid and Interface Science*, 217(1), 37-48.

Badin, A. L. (2009). "Répartition et influence de la matière organique et des microorganismes sur l'agrégation et le relargage de polluants dans des sédiments issus de l'infiltration d'eaux pluviales urbaines." Institut National des Sciences Appliquées de Lyon, Lyon.

Badin, A. L., Faure, P., Bedell, J. P., et Delolme, C. (2008). "Distribution of organic pollutants and natural organic matter in urban storm water sediments as a function of grain size." *Science of the Total Environment*, 403(1-3), 178-187.

Baize, D. (2000). *Guide des analyses en pédologie*, INRA, Paris.

Baker, A. J. M., et Walker, P. L. (1990). "Ecophysiology of metal uptake by tolerant plants." Heavy metal tolerance in plants: evolutionary aspects. , A. J. Shaw, ed., CRC press, Boca Raton, Florida, pp155-177.

Barclay, V. J., Bonner, R. F., et Hamilton, I. P. (1997). "Application of wavelet transforms to experimental spectra: Smoothing, denoising, and data set compression." *Analytical Chemistry*, 69(1), 78-90.

Barraud, S., Gibert, J., Winiarski, T., et Bertrand-Krajewski, J. L. (2002). "Implementation of a monitoring system to measure impact of stormwater runoff infiltration." *Water Science and Technology*, 45(3), 203-210.

Beauchemin, S., Hesterberg, D., et Beauchemin, M. (2002). "Principal Component Analysis Approach for Modeling Sulfur K-XANES Spectra of Humic Acids." *Soil Sci Soc Am J*, 66(1), 83-91.

Belanger, S. E., Farris, J. L., et Cherry, D. S. (1989). "Effects of diet, water hardness, and population source on acute and chronic copper toxicity to *Ceriodaphnia dubia*." *Archives of Environmental Contamination and Toxicology*, 18, 601-611.

Benedetti, M. F., VanRiemsdik, W. H., et Koopal, L. K. (1996). "Humic substances considered as a heterogeneous donnan gel phase." *Environmental Science & Technology*, 30(6), 1805-1813.

Benedetti, M. F., Milne, C. J., Kinniburgh, D. G., VanRiemsdijk, W. H., and Koopal, L. K. (1995). "Metal-Ion Binding to Humic Substances - Application of the Nonideal Competitive Adsorption Model." *Environmental Science & Technology*, 29(2), 446-457.

Benfatto, M., D'Angelo, P., Della Longa, S., et Pavel, N. V. (2002). "Evidence of distorted fivefold coordination of the Cu^{2+} aqua ion from an x-ray-absorption spectroscopy quantitative analysis." *Physical Review B*, 65(17), 174205.

Bishop, N. I. (1964). "Site of action of copper in photosynthesis." *Nature*, 204, 401-402.

Boudesocque, S., Guillon, E., Aplincourt, M., Marceau, E., et Stievano, L. (2007). "Sorption of Cu(II) onto vineyard soils: Macroscopic and spectroscopic investigations." *Journal of Colloid and Interface Science*, 307(1), 40-49.

Bowen, J. E. (1969). "Absorption of copper, zinc and manganese by sugarcane leaf tissue." *Plant Physiology*, 44, 255-261.

Bradl, H. B. (2004). "Adsorption of heavy metal ions on soils and soils constituents." *Journal of Colloid and Interface Science*, 277(1), 1-18.

Brams, E. A., et Fiskell, J. G. A. (1971). "Copper accumulation in citrus roots and desorption with acid." *Soil Science Society of America Proceedings*, 35, 772-775.

Bremner, I. (1987). "Involvement of Metallothionein in the Hepatic-Metabolism of Copper." *Journal of Nutrition*, 117(1), 19-29.

Brown, G. E., Henrich, V. E., Casey, W. H., Clark, D. L., Eggleston, C., Felmy, A., Goodman, D. W., Gratzel, M., Maciel, G., McCarthy, M. I., Nealson, K. H., Sverjensky, D. A., Toney, M. F., et Zachara, J. M. (1999). "Metal oxide surfaces and their interactions with aqueous solutions and microbial organisms." *Chemical Reviews*, 99(1), 77-174.

Brun, L. A., Maillet, J., Hinsinger, P., et Pépin, M. (2001). "Evaluation of copper availability to plants in copper-contaminated vineyard soils." *Environmental Pollution*, 111(2), 293-302.

Bryantsev, V. S., Diallo, M. S., van Duin, A. C. T., et Goddard Iii, W. A. (2008). "Hydration of Copper(II): New Insights from Density Functional Theory and the COSMO Solvation Model." *The Journal of Physical Chemistry A*, 112(38), 9104-9112.

Buffle, J. (1990). *Complexation Reactions in Aquatic Systems: An Analytical Approach*, Ellis Horwood, New York.

Buhl, K. J., et Hamilton, S. J. (1990). "Comparative toxicity of inorganic contaminants released by placer mining to early life stages of salmonids." *Ecotoxicology and Environmental Safety*, 20, 325-342.

BURGEAP. (1995). "Etude de la nappe de l'Est Lyonnais." Le Grand Lyon - Direction de l'eau.

Burton, D. T., et Fisher, D. J. (1990). "Acute toxicity of cadmium, copper, zinc, ammonia, 3,3'-dichlorobenzidine, 2,6-dichloro-4-nitroaniline, methylene chloride, and 2,4,6-trichlorophenol to juvenile grass shrimp and killifish. ." *Bulletin of Environmental Contamination and Toxicology*, 44, 776-783.

Cabaniss, S. E., Zhou, Q. H., Maurice, P. A., Chin, Y. P., et Aiken, G. R. (2000). "A log-normal distribution model for the molecular weight of aquatic fulvic acids." *Environmental Science & Technology*, 34(6), 1103-1109.

Cedeno-Maldonado, A., Swader, J. A., et Heath, R. (1970). "The cupric ion as an inhibitor of photosynthetic electron transport in isolated chloroplast." *Plant Physiology*, 50, 698-701.

Chang, T. W., Wang, M. K., et Jang, L. Y. (2005). "An extended X-ray absorption spectroscopy study of copper(II) sorption by oxides." *Geoderma*, 129(3-4), 211-218.

Cheah, S. F., Brown, G. E., et Parks, G. A. (2000). "XAFS study of Cu model compounds and Cu2+ sorption products on amorphous SiO2, gamma-Al2O3, and anatase." *American Mineralogist*, 85(1), 118-132.

Chebbo, G. (1992). "Solides des rejets pluviaux urbains, caractérisation et traitabilité," Paris.

Cheng, L., Sturchio, N. C., Woicik, J. C., Kemner, K. M., Lyman, P. F., et Bedzyk, M. J. (1998). "High-resolution structural study of zinc ion incorporation at the calcite cleavage surface." *Surface Science*, 415(1-2), L976-L982.

Chocat, B. (1997). "Aménagement urbain et hydrologie." *La Houille Blanche*(7), 12-19.

Clemens, S. (2001). "Molecular mechanisms of plant metal tolerance and homeostasis." *Planta*, 212(4), 475-486.

Cotton, F. A., Wilkinson, G., Murillo, C. A., et Bochmann, M. (1999). *Advanced inorganic chemistry*, John Wiley & Sons Ltd, New York.

Culotta, V. C., Klomp, L. W. J., Strain, J., Casareno, R. L. B., Krems, B., et Gitlin, J. D. (1997). "The copper chaperone for superoxide dismutase." *Journal of Biological Chemistry*, 272(38), 23469-23472.

Cuypers, A., Koistinen, K. M., Kokko, H., Karenlampi, S., Auriola, S., et Vangronsveld, J. (2005). "Analysis of bean (Phaseolus vulgaris L.) proteins affected by copper stress." *Journal of Plant Physiology*, 162(4), 383-392.

Dameron, C. T., Winge, D. R., George, G. N., Sansone, M., Hu, S., et Hamer, D. (1991). "A Copper Thiolate Polynuclear Cluster in the Ace1 Transcription Factor." *Proceedings of the National Academy of Sciences of the United States of America*, 88(14), 6127-6131.

Daniels, R. R., Struckmeyer, B. E., et Peterson, L. A. (1972). "Copper toxicity in Phaseolus vulgaris L. as influenced by iron nutrition. I. An anatomical study." *Journal of the American Society for Horticultural Science*, 97, 249-254.

Datry, I., Malard, F., et Gibert, J. (2004). "Dynamics of solutes and dissolved oxygen in shallow urban groundwater below a stormwater infiltration basin." *Science of the Total Environment*, 329(1-3), 215-229.

Dechesne, M. (2002). "Connaissance et modélisation du fonctionnement des bassins d'infiltration d'eaux de ruissellement urbain pour l'évaluation des performances techniques et environnementales sur le long terme," Lyon (France).

Dent, A. J., Farrugia, L. J., Orpen, A. G., et Stratford, S. E. (1992). "Kinetic and Structural Information on Metal Cluster Rearrangement in Solution from Exafs Spectra." *Journal of the Chemical Society-Chemical Communications* (19), 1456-1457.

Deschamps, P., Kulkarni, P. P., et Sarkar, B. (2004). "X-ray Structure of Physiological Copper(II)-Bis(L-histidinato) Complex." *Inorganic Chemistry*, 43(11), 3338-3340.

Devi, V. U. (1987). "Heavy metal toxicity to fiddler crabs, Uca annulipes Latreille and Uca triangularis (Milne Edwards): tolerance to copper, mercury, cadmium, and Zinc. ." *Bulletin of Environmental Contamination and Toxicology* , 39, 1020-1027.

Devos, C. H. R., Schat, H., Vooijs, R., et Ernst, W. H. O. (1989). "Copper-Induced Damage to the Permeability Barrier in Roots of Silene-Cucubalus." *Journal of Plant Physiology*, 135(2), 164-169.

Dewit, J. C. M., Vanriemsdijk, W. H., et Koopal, L. K. (1993). "Proton Binding to Humic Substances .1. Electrostatic Effects." *Environmental Science & Technology*, 27(10), 2005-2014.

Domek, M. J., LeChevallier, M. W., Cameron, S. C., et McFeters, G. A. (1984). "Evidence for the role of copper in the injury process of coliform bacteria in drinking water." *Applied and Environmental Microbiology*, 48(2), 289-293.

Donkin, S. G., et Dusenbery, D. B. (1993). "A soil toxicity test using the nematode Caenorhabditis elegans and an effective method of recovery." *Archives of Environmental Contamination and Toxicology*, 25, 145-151.

Donoho, D. L. (1995). "De-Noising by Soft-Thresholding." *Ieee Transactions on Information Theory*, 41(3), 613-627.

Dudal, Y., et Gerard, F. (2004). "Accounting for natural organic matter in aqueous chemical equilibrium models: a review of the theories and applications." *Earth-Science Reviews*, 66(3-4), 199-216.

Dzombak, D.A. et Hudson, R.J. (1995). Ion exchange : the contributions of diffuse layer sorption and surface complexation. *Aquatic surface chemistry*. C.P. Huang, C.R. O'Melia, et J.J. Morgan, eds. American Chemical society, Washington D.C.

Effler, S. W., Litten, S., Field, S. D., Tong-Ngork, T., Hale, F., Meyer, M., et Quirk, M. (1980). "Whole lake responses to low level copper sulfate treatment." *Water Research*, 14(10), 1489-1499.

Elder, J. F., et Horne, A. J. (1978). "Copper cycles and CuSO4 algicidal capacity in two California lakes." *Environmental Management*, 2(1), 17-30.

Ellis, B. (1991). "Urban runoff quality in the UK: problems, prospects and procedures." *Applied Geography*, 11(3), 187-200.

Elzinga, E. J., et Reeder, R. J. (2002). "X-ray absorption spectroscopy study of Cu2+ and Zn2+ adsorption complexes at the calcite surface: Implications for site-specific metal incorporation preferences during calcite crystal growth." *Geochimica et Cosmochimica Acta*, 66(22), 3943-3954.

Elzinga, E. J., Rouff, A. A., et Reeder, R. J. (2006). "The long-term fate of Cu2+, Zn2+, and Pb2+ adsorption complexes at the calcite surface: An X-ray absorption spectroscopy study." *Geochimica et Cosmochimica Acta*, 70(11), 2715-2725.

Emsley, J. (2003). *Nature's building blocks*, Oxford University Press Inc. , New York.

Erickson, R. J., Benoit, D. A., Mattson, V. R., Nelson, H. P., et Leonard, E. N. (1996). "The effects of water chemistry on the toxicity of copper to fathead minnows." *Environmental Toxicology and Chemistry*, 15(2), 181-193.

Evertsson, B. (1969). "The crystal structure of bis-l-histidinecopper(II) dinitrate dihydrate." *Acta Crystallographica Section B*, 25(1), 30-41.

Farquhar, M. L., Charnock, J. M., England, K. E. R., et Vaughan, D. J. (1996). "Adsorption of Cu(II) on the (0001) plane of mica: A REFLEXAFS and XPS study." *Journal of Colloid and Interface Science*, 177(2), 561-567.

Farquhar, M. L., Vaughan, D. J., Hughes, C. R., Charnock, J. M., et England, K. E. R. (1997). "Experimental studies of the interaction of aqueous metal cations with mineral substrates: Lead, cadmium, and copper with perthitic feldspar, muscovite, and biotite." *Geochimica et Cosmochimica Acta*, 61(15), 3051-3064.

Ferrage, E. (2004). "Etude expérimentale de l'hydratation des smectites par simulation des raies 00ℓ de diffraction des rayons X. Implications pour l'étude d'une perturbation thermique sur la minéralogie de l'argilite du site Meuse-Haute Marne.," Université Joseph Fourier - Grenoble 1., Grenoble.

Filipponi, A. (1995). "Statistical errors in X-ray absorption fine-structure data analysis." *Journal of Physics: Condensed Matter*, 7(48), 9343.

Flemming, C. A., et Trevors, J. T. (1989). "Copper toxicity and chemistry in the environment: a review." *Water, Air, & Soil Pollution*, 44(1), 143-158.

Forstner, U., et Wittman, G. T. W. (1979). *Metal pollution in the aquatic environment*, Spinger-Verlag, Berlin.

Fortey, N. J. (2004). "Jambor, J.L., Blowes, D.W. et Ritchie, A.I.M. (Eds) Environmental Aspects of Mine Wastes. Mineralogical Association of Canada, Short Course Series Volume 31, 2003, 430 pp. ISBN 0-921294-31-X." *Mineralogical Magazine*, 68(2), 415-417.

Foy, C. D., Chaney, R. L., et White, M. C. (1978). "The physiology of metal toxicity in plants." *Annual Review of Plant Physiology and Plant Molecular Biology*, 29, 511-566.

Frank, P., Benfatto, M., Hedman, B., et Hodgson, K. O. (2009). "The XAS model of dissolved Cu(II) and its significance to biological electron transfer." *Journal of Physics: Conference Series*, 190(1), 012059.

Frank, P., Benfatto, M., Szilagyi, R. K., D'Angelo, P., Longa, S. D., et Hodgson, K. O. (2005). "The Solution Structure of [Cu(aq)]2+ and Its Implications for Rack-Induced Bonding in Blue Copper Protein Active Sites." *Inorganic Chemistry*, 44(6), 1922-1933.

Freedman, J. H., Ciriolo, M. R., et Peisach, J. (1989). "The Role of Glutathione in Copper-Metabolism and Toxicity." *Journal of Biological Chemistry*, 264(10), 5598-5605.

Furnare, L. J., Vailionis, A., et Strawn, D. G. (2005). "Polarized XANES and EXAFS spectroscopic investigation into copper(II) complexes on vermiculite." *Geochimica et Cosmochimica Acta*, 69(22), 5219-5231.

Gauss, J. D., Woods, P. E., Winner, R. W., et Skillings, J. H. (1985). "Acute toxicity of copper to three life stages of Chironomus tentans as affected by water hardness-alkalinity." *Environmental Pollution (Series A)*, 37, 149-157.

Gilkes, R. J. (1981). Copper in soils and plants, J. F. Loneragan, Robson, A. D., Graham, R. D., eds., Academic Press., New York, 97-117.

Goldstein, S., et Czapski, G. (1986). "The role and mechanism of metal ions and their complexes in enhancing damage in biological systems or in protecting these systems from the toxicity of O2-." *Journal of Free Radicals in Biology Medicine*, 2(1), 3-11.

Gollan, J. L., et Deller, D. J. (1973). "Studies on Nature and Excretion of Biliary Copper in Man." *Clinical Science*, 44(1), 9-15.

Grotz, N., et Guerinot, M. L. (2006). "Molecular aspects of Cu, Fe and Zn homeostasis in plants." *Biochimica Et Biophysica Acta-Molecular Cell Research*, 1763(7), 595-608.

Gupta, U. C. (1979). *Copper in the Environment. Part I: Ecological Cycling*, J. O. Nriagu, ed., John Wiley & Sons, New York, 255-288.

Haanstra, L., et Doelman, P. (1984). "Glutamic acid decomposition as a sensitive measure of heavy metal pollution in soil." *Soil Biology and Biochemistry*, 16, 595-600.

Hamilton, S. J., et Buhl, K. J. (1990). "Safety assessment of selected inorganic elements to fry of chinook salmon (*Oncorhynchus tshawytscha*)." *Ecotoxicology and Environmental Safety*, 20, 307-324.

Han, F. X., Banin, A., et Triplett, G. B. (2001). "Redistribution of heavy metals in arid-zone soils under a wetting-drying cycle soil moisture regime." *Soil Science*, 166(1), 18-28.

Harter, R. D. (1983). "Effect of Soil-Ph on Adsorption of Lead, Copper, Zinc, and Nickel." *Soil Science Society of America Journal*, 47(1), 47-51.
Hayes, K.F., Redden, G., Ela, W., et Leckie, J.O. (1991). "Surface Complexation Models: An Evaluation of Model Parameter Estimation Ising FITEQL and Oxide Mineral Titration Data." *Journal of Colloid and Interface Science*, 142(2), 448-469.
Hettiarachchi, G. M., Pierzynski, G. M., et Ransom, M. D. (2000). "In situ stabilization of soil lead using phosphorus and manganese oxide." *Environmental Science & Technology*, 34(21), 4614-4619.
Hinz, C. (2001). "Description of sorption data with isotherm equations." *Geoderma*, 99(3-4), 225-243.
Hizal, J., et Apak, R. (2006). "Modeling of copper(II) and lead(II) adsorption on kaolinite-based clay minerals individually and in the presence of humic acid." *Journal of Colloid and Interface Science*, 295(1), 1-13.
Hodson, P. V., Borgmann, U. et Shear, H. (1979). Copper in the environment: Part 2 Health Effects, J. O. Nriagu, ed., John Wiley & Sons Ltd, Toronto, pp307-372.
Hopkins, W. G. (2003). *Physiologie végétale*, S. Rambour, translator, De Boeck.
Ildefonse, P., Manceau, A., Prost, D., et Groke, M. C. T. (1986). "Hydroxy-Cu-Vermiculite Formed by the Weathering of Fe-Biotites at Salobo, Carajas, Brazil." *Clays and Clay Minerals*, 34(3), 338-345.
Imai, A. (2000). "Genesis of the Mamut porphyry copper deposit, Sabah, East Malaysia." *Resource Geology*, 50(1), 1-23.
Irtelli, B., Petrucci, W. A., et Navari-Izzo, F. (2009). "Nicotianamine and histidine/proline are, respectively, the most important copper chelators in xylem sap of Brassica carinata under conditions of copper deficiency and excess." *Journal of Experimental Botany*, 60(1), 269-277.
Isabey, J. (1933). *"Cuivre." Traité de chimie minérale Tome VIII Elements des terres rares Cuivre -Argent-Or-Mercure*, P. Pascal et P. Baud, eds., Masson et Compagnie, Paris, 1180 pages.
Isaure, M.-P., Laboudigue, A., Manceau, A., Sarret, G., Tiffreau, C., Trocellier, P., Lamble, G., Hazemann, J.-L., et Chateigner, D. (2002). "Quantitative Zn speciation in a contaminated dredged sediment by [mu]-PIXE, [mu]-SXRF, EXAFS spectroscopy and principal component analysis." *Geochimica et Cosmochimica Acta*, 66(9), 1549-1567.
Isaure, M. P., Manceau, A., Geoffroy, N., Laboudigue, A., Tamura, N., et Marcus, M. A. (2005). "Zinc mobility and speciation in soil covered by contaminated dredged sediment using micrometer-scale and bulk-averaging X-ray fluorescence, absorption and diffraction techniques." *Geochimica et Cosmochimica Acta*, 69(5), 1173-1198.
IXS-Standards-and-Criteria-Committe. (2000). "Error Reporting Recommendations: A Report of the Standards and Criteria Committee." IXS Standards and Criteria Committe.
Janosi, A. (1964). "La structure du sulfure cuivreux quadratique." *Acta Crystallographica*, 17(3), 311-312.
Jarvis, S. C. (1978). "Copper uptake and accumulation by perennial ryegrass grown in soil and solution culture." *Journal of the Science Food Agriculture*, 29, 12-18.
Joly, Y. (2001). "X-ray absorption near-edge structure calculations beyond the muffin-tin approximation." *Physical Review B*, 63(12), 125120.
Jopony, M., Usup, G., et Mohamed, M. (1987). "Particle size Distribution of copper mine tailings from Lohan Ranau Sabah and its relationship with heavy metal content." *Pertanika*, 10(1), 37-40.
Kabata-Pendias, A., et Pendias, H.-L. U. (1992). *Trace elements in soils and plants*, CRC Press (Boca Raton), London.

Kaiser, H. (1958). "The varimax criterion for analytic rotation in factor analysis." *Psychometrika*, 23(3), 187-200.
Karlsson, T., Elgh-Dalgren, K., et Skyllberg, U. (2008). "Modeling Copper(II) Complexation in a Peat Soil Based on Spectroscopic Structural Information." *Soil Science Society of America Journal*, 72(5), 1286-1291.
Karlsson, T., Persson, P., et Skyllberg, U. (2006). "Complexation of Copper(II) in Organic Soils and in Dissolved Organic Matter − EXAFS Evidence for Chelate Ring Structures." *Environmental Science & Technology*, 40(8), 2623-2628.
Kim, N. D., et Fergusson, J. E. (1994). "The Concentrations, Distribution and Sources of Cadmium, Copper, Lead and Zinc in the Atmosphere of an Urban-Environment." *Science of the Total Environment*, 144, 179-189.
Kinniburgh, D. G. (1999). "FIT User Guide." British Geological Survey, Keyworth, UK.
Kinniburgh, D. G., Milne, C. J., Benedetti, M. F., Pinheiro, J. P., Filius, J., Koopal, L. K., et VanRiemsdijk, W. H. (1996). "Metal ion binding by humic acid: Application of the NICA-Donnan model." *Environmental Science & Technology*, 30(5), 1687-1698.
Kirpichtchikova, T. A., Manceau, A., Spadini, L., Panfili, F., MarcuS, M. A., et Jacquet, T. (2006). "Speciation and solubility of heavy metals in contaminated soil using X-ray microfluorescence, EXAFS spectroscopy, chemical extraction, and thermodynamic modeling." *Geochimica et Cosmochimica Acta*, 70(9), 2163-2190.
Klomp, L. W. J., Lin, S. J., Yuan, D. S., Klausner, R. D., Culotta, V. C., et Gitlin, J. D. (1997). "Identification and functional expression of HAH1, a novel human gene involved in copper homeostasis." *Journal of Biological Chemistry*, 272(14), 9221-9226.
Komarek, M., Cadkova, E., Chrastny, V., Bordas, F., et Bollinger, J. C. (2010). "Contamination of vineyard soils with fungicides: A review of environmental and toxicological aspects." *Environment International*, 36(1), 138-151.
Koningsberger, D., et Prins, R. (1998). *X-Ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS, and XANES.*, Wiley, New York.
Koopal, L. K., Saito, T., Pinheiro, J. P., et van Riemsdijk, W. H. (2005). "Ion binding to natural organic matter: General considerations and the NICA-Donnan model." *Colloids and Surfaces a-Physicochemical and Engineering Aspects*, 265(1-3), 40-54.
Korshin, G. V., Frenkel, A. I., et Stern, E. A. (1998). "EXAFS Study of the Inner Shell Structure in Copper(II) Complexes with Humic Substances." *Environmental Science & Technology*, 32(18), 2699-2705.
Kosaka, H., and Wakita, K. (1978). "Some Geologic Features of Mamut-Porphyry-Copper-Deposit, Sabah, Malaysia." *Economic Geology*, 73(5), 618-627.
Kraft, S., Stumpel, J., Becker, P., et Kuetgens, U. (1996). "High resolution x-ray absorption spectroscopy with absolute energy Calibration for the determination of absorption edge energies." *Review of Scientific Instruments*, 67(3), 681-687.
Kratz, T., et Fuess, H. (1989). "Simultane Strukturbestimmung von Kupferkies und Bornit an einem Kristall." *Zeitschrift Für Kristallographie*, 186, 167.
Kupper, H., Gotz, B., Mijovilovich, A., Kupper, F. C., et Meyer-Klaucke, W. (2009). "Complexation and Toxicity of Copper in Higher Plants. I. Characterization of Copper Accumulation, Speciation, and Toxicity in Crassula helmsii as a New Copper Accumulator." *Plant Physiology*, 151(2), 702-714.
Lanson, M. (2005). "Impact environnemental de l'exploitation du porphyre cuprifère de la mine de Mamut en Malaisie." GDR TRANSMET.
Lassabatere, L., Angulo-Jaramillo, R., Goutaland, D., Letellier, L., Gaudet, J. P., Winiarski, T., et Delolme, C. (2010). "Effect of the settlement of sediments on water infiltration in two urban infiltration basins." *Geoderma*, 156(3-4), 316-325.

Lee, S. S., Fenter, P., Park, C., Sturchio, N. C., et Nagy, K. L. (2010). "Hydrated Cation Speciation at the Muscovite (001)-Water Interface." *Langmuir*, 26(22), 16647-16651.
Leenheer, J. A. (2009). "Systematic Approaches to Comprehensive Analyses of Natural Organic Matter " *Annals of Environmental Science*, 3, 1-130.
Lenoir, T., et Manceau, A. (2010). "Number of Independent Parameters in the Potentiometric Titration of Humic Substances." *Langmuir*, 26(6), 3998-4003.
Lenoir, T., Matynia, A., et Manceau, A. (2010). "Convergence-Optimized Procedure for Applying the NICA-Donnan Model to Potentiometric Titrations of Humic Substances." *Environmental Science & Technology*, 44(16), 6221-6227.
Linder, M. C., et Hazegh-Azam, M. (1996). "Copper biochemistry and molecular biology." *American Journal of Clinical Nutrition*, 63, 797-811.
Lipman, C. B., et Mackinney, G. (1931). "Proof of the essential nature of copper for higher green plants." *Journal of Physiology*, 6, 593-599.
Liu, D. H., et Kottke, I. (2004). "Subcellular localization of copper in the root cells of Allium sativum by electron energy loss spectroscopy (EELS)." *Bioresource Technology*, 94(2), 153-158.
Liu, D. J., et Robota, H. J. (1993). "In-Situ Xanes Characterization of the Cu Oxidation-State in Cu-Zsm-5 During No Decomposition Catalysis." *Catalysis Letters*, 21(3-4), 291-301.
Lo, I. M. C., et Yang, X. Y. (1999). "EDTA Extraction of Heavy Metals from Different Soil Fractions and Synthetic Soils." *Water, Air, & Soil Pollution*, 109(1), 219-236.
Lou, L. Q., Shen, Z. G., et Li, X. D. (2004). "The copper tolerance mechanisms of Elsholtzia haichowensis, a plant from copper-enriched soils." *Environmental and Experimental Botany*, 51(2), 111-120.
Lund, T. J., Koretsky, C. M., Landry, C. J., Schaller, M. S., et Das, S. (2008). "Surface complexation modeling of Cu(II) adsorption on mixtures of hydrous ferric oxide and kaolinite." *Geochemical Transactions*, 9.
Luo, X.-s., Zhou, D.-m., et Wang, Y.-j. (2006). "Free cupric ions in contaminated agricultural soils around a copper mine in eastern Nanjing City, China." *Journal of Environmental Sciences*, 18(5), 927-931.
Malecki, M. R., Neuhauser, E. F., et Loehr, R. C. (1982). "The effect of metals on the growth and reproduction of Eisenia foetida (Oligochaeta, Lumbricidae)." *Pedobiologia*, 24, 129-137.
Malinowski, E. R. (1991). *Factor analysis in chemistry*, John Wiley &Sons, Inc, New York, pp350.
Manceau, A. (2005). "Dynamique et spéciation moléculaire des éléments traces (Zn, Cu, Pb et Hg) à l'interface sol-matière organique-plante." Agence Nationale de la Recherche, Grenoble.
Manceau, A. (2011). "Critical evaluation of the revised akdalaite model for ferrihydrite." *American Mineralogist*, 96(4), 521-533.
Manceau, A., Buseck, P. R., Miser, D., Rask, J., et Nahon, D. (1990). "Characterization of Cu in Lithiophorite from a Banded Mn Ore." *American Mineralogist*, 75(5-6), 490-494.
Manceau, A., Lanson, M., et Geoffroy, N. (2007). "Natural speciation of Ni, Zn, Ba, and As in ferromanganese coatings on quartz using X-ray fluorescence, absorption, and diffraction." *Geochimica et Cosmochimica Acta*, 71(1), 95-128.
Manceau, A., Llorca, S., et Calas, G. (1987). "Crystal-Chemistry of Cobalt and Nickel in Lithiophorite and Asbolane from New-Caledonia." *Geochimica et Cosmochimica Acta*, 51(1), 105-113.
Manceau, A., Marcus, M. A., et Tamura, N. (2002). "Quantitative speciation of heavy metals in soils and sediments by synchrotron X-ray techniques." *Applications of Synchrotron*

Radiation in Low-Temperature Geochemistry and Environmental Sciences, 49, 341-428.

Manceau, A., Marcus, M. A., Tamura, N., Proux, O., Geoffroy, N., et Lanson, B. (2004). "Natural speciation of Zn at the micrometer scale in a clayey soil using X-ray fluorescence, absorption, and diffraction." *Geochimica et Cosmochimica Acta*, 68(11), 2467-2483.

Manceau, A., et Matynia, A. (2010). "The nature of Cu bonding to natural organic matter." *Geochimica et Cosmochimica Acta*, 74(9), 2556-2580.

Manceau, A., Nagy, K. L., Marcus, M. A., Lanson, M., Geoffroy, N., Jacquet, T., et Kirpichtchikova, T. (2008). "Formation of metallic copper nanoparticles at the soil-root interface." *Environmental Science & Technology*, 42(5), 1766-1772.

Manceau, A., Schlegel, M. L., Musso, M., Sole, V. A., Gauthier, C., Petit, P. E., et Trolard, F. (2000). "Crystal chemistry of trace elements in natural and synthetic goethite." *Geochimica et Cosmochimica Acta*, 64(21), 3643-3661.

Maqueda, C., Morillo, E., et Undabeytia, T. (2002). "Cosorption of glyphosate and copper(II) on goethite." *Soil Science*, 167(10), 659-665.

Marceau, N., Aspin, N., et Sass-Kortsak, A. (1970). "Absorption of copper 64 from gastrointestinal tract of the rat." *American Journal of Physiology*, 218, 377-383.

Marcus, M. A., MacDowell, A. A., Celestre, R., Manceau, A., Miller, T., Padmore, H. A., et Sublett, R. E. (2004). "Beamline 10.3.2 at ALS: a hard X-ray microprobe for environmental and materials sciences." *Journal of Synchrotron Radiation*, 11(3), 239-247.

Marinsky, J. A., et Ephraim, J. (1986). "A Unified Physicochemical Description of the Protonation and Metal-Ion Complexation Equilibria of Natural Organic-Acids (Humic and Fulvic-Acids) .1. Analysis of the Influence of Polyelectrolyte Properties on Protonation Equilibria in Ionic Media - Fundamental-Concepts." *Environmental Science & Technology*, 20(4), 349-354.

Martin, N. A. (1986). "Toxicity of pesticides to Allolobophora caliginosa (Oligochaeta: Lumbricidae)." *New Zealand Journal of Agricultural Research*, 29, 699-706.

Massarotti, V., Capsoni, D., Bini, M., Altomare, A., et Moliterni, A. G. G. (1998). "X-ray powder diffraction ab initio structure solution of materials from solid state synthesis : the copper oxide case." *Zeitschrift für Kristallographie*, 213(5), 259-265.

Mathur, R., Ruiz, J., Titley, S., Liermann, L., Buss, H., et Brantley, S. (2005). "Cu isotopic fractionation in the supergene environment with and without bacteria." *Geochimica et Cosmochimica Acta*, 69(22), 5233-5246.

Matsumoto, M., et Nishimura, T. (1998). "Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator." *ACM: Transactions on Modeling and Computer Simulation*, 8(1), 3-30.

Matynia, A. (2009). "Mécanisme de rétention du cuivre sur de la tourbe végétalisée," Université Joseph Fourier - Grenoble I, Grenoble.

Matynia, A., Lenoir, T., Causse, B., Spadini, L., Jacquet, T., et Manceau, A. (2010). "Semi-empirical proton binding constants for natural organic matter." *Geochimica et Cosmochimica Acta*, 74(6), 1836-1851.

Mayer, F. L., et Ellersieck, M. R. (1986). "Manual of acute toxicity: Interpretation and data base for 410 chemicals and 66 species of freshwater animals." *Resource Publication 160*, US Department of Interior, Fish and Wildlife Service, Washington, DC.

McKim, J. M., et Benoit, D. A. (1971). "Effects of long-term exposures to copper on survival, growth, and reproduction of brook trout (*Salvelinus fontinalis*)." *Journal of the Fisheries Research Board of Canada*, 28, 655-662.

McNulty, H. R., Anderson, B. S., Hunt, J. W., Turpen, S. L., et Singer, M. M. (1994). "Age-Specific Toxicity of Copper to Larval Topsmelt Atherinops-Affinis." *Environmental Toxicology and Chemistry*, 13(3), 487-492.

Meranger, J. C., Subramanian, K. S., et Chalifoux, C. (1979). "A national survey for cadmium, chromium, copper, lead, zinc, calcium, and magnesium in Canadian drinking water supplies." *Environmental Science & Technology*, 13(6), 707-711.

Mijovilovich, A., Leitenmaier, B., Meyer-Klaucke, W., Kroneck, P. M. H., Gotz, B., et Kupper, H. (2009). "Complexation and Toxicity of Copper in Higher Plants. II. Different Mechanisms for Copper versus Cadmium Detoxification in the Copper-Sensitive Cadmium/Zinc Hyperaccumulator Thlaspi caerulescens (Ganges Ecotype)." *Plant Physiology*, 151(2), 715-731.

Mikkelsen, P. S., Weyer, G., Berry, C., Walden, Y., Colandini, V., Poulsen, S., Grotehusmann, D., et Rohlfing, R. (1994). "Pollution from Urban Stormwater Infiltration." *Water Science and Technology*, 29(1-2), 293-302.

Milne, C. J., Kinniburgh, D. G., et Tipping, E. (2001). "Generic NICA-Donnan model parameters for proton binding by humic substances." *Environmental Science & Technology*, 35(10), 2049-2059.

Moore, D. M., et Reynolds, R. C. J. (1997). *X-Ray diffraction and the Identification and Analysis of Clay Minerals*, Oxford university press, New York.

Morin, G., et Calas, G. (2006). "Arsenic in soils, mine tailings, and former industrial sites." *Elements*, 2(2), 97-101.

Morton, J. D., Semrau, J. D., et Hayes, K. F. (2001). "An X-ray absorption spectroscopy study of the structure and reversibility of copper adsorbed to montmorillonite clay." *Geochimica et Cosmochimica Acta*, 65(16), 2709-2722.

Murtezda, M., et Lee, Y. H. (1985). "Heavy metals in rice field soil of Lohan Valley, Sabah." *Sumber*, 1, 179-187.

Nelson, D. A., Miller, J. E., et Calabrese, A. (1988). "Effect of heavy metals on bay scallops, surf clams, and blue mussels in acute and long-term exposures." *Archives of Environmental Contamination and Toxicology*, 17, 595-600.

Neumann, P. Z., et Sass-Kortsak, A. (1967). "The state of copper in human serum: evidence for an amino acid-bound fraction." *Journal of Clinical Investigation*, 46(4), 646-58.

Newville. (2004). "Fundamentals of XAFS." Consortium for Advanced Radiation Sources University of Chicago, Chicago, IL, Chicago.

Niggli, P. (1922). "Die Kristallstruktur Einiger Oxyde I." *Zeitschrift Für Kristallographie* 57, 253.

Nriagu, J. O. (1979). *Copper in the environment: Part 1 Ecological cycling*, John Wiley & Sons Ltd, New York.

Nriagu, J. O. (1989). "A global assessment of natural sources of atmospheric trace metals." *Nature*, 338(6210), 47-49.

O'Donohue, J. W., Reid, M. A., Varghese, A., Portmann, B., et Williams, R. (1993). "Case report: Micronodular cirrhosis and acute liver failure due to chronic copper self-intoxication." *European Journal of Gastroenterology and Hepatology*, 5, 561-562.

Osman, M. A., et Suter, U. W. (2000). "Determination of the Cation-Exchange Capacity of Muscovite Mica." *Journal of Colloid and Interface Science*, 224(1), 112-115.

Ozoh, P. T. E. (1992). "The effect of temperature and salinity on copper body-burden and copper toxicity to Hediste (Nereis) diversicolor." *Environmental Monitoring and Assessment*, 21, 11-17.

Panfili, F. (2004). "Etude de l'évolution de la spéciation du zinc dans la phase solide d'un sédiment de curage contaminé, induit par phytostabilisation," Université de Provence - Aix Marseille I, Marseille.

Panfili, F., Manceau, A., Sarret, G., Spadini, L., Kirpichtchikova, T., Bert, V., Laboudigue, A., Marcus, M. A., Ahamdach, N., et Libert, M.-F. (2005). "The effect of phytostabilization on Zn speciation in a dredged contaminated sediment using scanning electron microscopy, X-ray fluorescence, EXAFS spectroscopy, and principal components analysis." *Geochimica et Cosmochimica Acta*, 69(9), 2265-2284.

Parkman, R. H., Charnock, J. M., Bryan, N. D., Livens, F. R., et Vaughan, D. J. (1999). "Reactions of copper and cadmium ions in aqueous solution with goethite, lepidocrocite, mackinawite, and pyrite." *American Mineralogist*, 84(3), 407-419.

Pasquarello, A., Petri, I., Salmon, P. S., Parisel, O., Car, R., Toth, E., Powell, D. H., Fischer, H. E., Helm, L., et Merbach, A. E. (2001). "First Solvation Shell of the Cu(II) Aqua Ion: Evidence for Fivefold Coordination." *Science*, 291(5505), 856-859.

Peacock, C. L., et Sherman, D. M. (2004). "Copper(II) sorption onto goethite, hematite and lepidocrocite: a surface complexation model based on ab initio molecular geometries and EXAFS spectroscopy." *Geochimica et Cosmochimica Acta*, 68(12), 2623-2637.

Peacock, C. L., et Sherman, D. M. (2005). "Surface complexation model for multisite adsorption of copper(II) onto kaolinite." *Geochimica et Cosmochimica Acta*, 69(15), 3733-3745.

Phoong, S. M., Kim, C. W., Vaidyanathan, P. P., et Ansari, R. (1995). "A new class of 2-channel biorthogonal filter banks and wavelet bases." *IEEE Transactions on Signal Processing*, 43(3), 649-665.

Pietrzak, U., et McPhail, D. C. (2004). "Copper accumulation, distribution and fractionation in vineyard soils of Victoria, Australia." *Geoderma*, 122(2-4), 151-166.

Pimentel, J. C., et Menezes, A. P. (1975). "Liver granulomas containing copper in vineyard sprayer's lung: A new etiology of hepatic granulomatosis." *American Review of Respiratory Disease*, 111, 189-195.

Pitt, R., Clark, S., et Field, R. (1999). "Groundwater contamination potential from stormwater infiltration practices." *Urban Water*, 1, 217-236.

Pokrovsky, O. S., Viers, J., Emnova, E. E., Kompantseva, E. I., et Freydier, R. (2008). "Copper isotope fractionation during its interaction with soil and aquatic microorganisms and metal oxy(hydr)oxides: Possible structural control." *Geochimica et Cosmochimica Acta*, 72(7), 1742-1757.

Post, J. E. (1999). "Manganese oxide minerals: Crystal structures and economic and environmental significance." *Proceedings of the National Academy of Sciences of the United States of America*, 96(7), 3447-3454.

Quartacci, M. F., Irtelli, B., Gonnelli, C., Gabbrielli, R., et Navari-Izzo, F. (2009). "Naturally-assisted metal phytoextraction by Brassica carinata: Role of root exudates." *Environmental Pollution*, 157(10), 2697-2703.

Rahimi, A., et Bussler, W. (1973). "[Physiological conditions for the development of copper deficiency symptoms]." *Zeitschrift für Pflanzenernährung und Bodenkunde*, 136, 25-32.

Ramos, L., Hernandez, L., et Gonzales, M.J. (1994). "Sequential Fractionation of Copper, Lead, Cadmium and Zinc in Soils from or near Donana-National-Park." Journal of environmental Quality, 23(1), 50-57.

Reeder, R. J., Lamble, G. M., et Northrup, P. A. (1999). "XAFS study of the coordination and local relaxation around Co2+, Zn2+, Pb2+, and Ba2+ trace elements." *American Mineralogist*, 84(7-8), 1049-1060.

Reid, K. (2006). "Carences en oligo-éléments dans le blé d'automne." http://www.omafra.gov.on.ca Ministère de l'Agriculture, de l'Alimentation et des Affaires Rurales, Toronto, Ontario Canada.

Rellan-Alvarez, R., Abadia, J., et Alvarez-Fernandez, A. (2008). "Formation of metal-nicotianamine complexes as affected by pH, ligand exchange with citrate and metal exchange. A study by electrospray ionization time-of-flight mass spectrometry." *Rapid Communications in Mass Spectrometry*, 22(10), 1553-1562.

Richter, A., Brendler, V., et Nebelung, C. (2005). "Blind prediction of Cu(II) sorption onto goethite: Current capabilities of diffuse double layer model." *Geochimica et Cosmochimica Acta*, 69(11), 2725-2734.

Roberts, L. A., Pierson, A. J., Panaviene, Z., et Walker, E. L. (2004). "Yellow stripe1. Expanded roles for the maize iron-phytosiderophore transporter." *Plant Physiology*, 135(1), 112-120.

Robson, A. D., et Reuter, D. J. "Diagnosis of copper deficiency and toxicity." *Copper in soils and plants: Proceedings of the Golden Jubilee International Symposium*, Murdoch University, Perth, Western Australia, pp 287-312.

Rose, W. C., et Bodansky, M. (1920). "biochemical studies on marine organisms." *Journal of Biological Chemistry*, 44(1), 99-112.

Ruban, V., Larrarte, F., Berthier, M., Favreau, L., Sauvourel, Y., Letellier, L., Mosisni, M. L., et Raimbault, G. (2005). "Quantitative and qualitative hydrologic balance for a suburban watershed with a separate sewer system (Nantes, France)." *Water Science and Technology*, 51(2), 231-238.

Rucker, R. B., Parker, H. E., et Rogler, J. C. (1969). "Effect of copper deficiency on chick bone collagen and selected bone enzymes." *Journal of Nutrition*, 98, 57-63.

Rudzinski, W., Charmas, R., Cases, J. M., Francois, M., Villieras, F., et Michot, L. J. (1997). "Calorimetric effects and temperature dependence of simple ion adsorption at oxide-electrolyte interface: A theoretical analysis based on the triple-layer complexation model." *Langmuir*, 13(3), 483-495.

Sahi, S. V., Israr, M., Srivastava, A. K., Gardea-Torresdey, J. L., et Parsons, J. G. (2007). "Accumulation, speciation and cellular localization of copper in Sesbania drummondii." *Chemosphere*, 67(11), 2257-2266.

Saito, T., Nagasaki, S., Tanaka, S., et Koopal, L. K. (2005). "Electrostatic interaction models for ion binding to humic substances." *Colloids and Surfaces A-Physicochemical and Engineering Aspects*, 265(1-3), 104-113.

Sarkar, B. (1999). "Treatment of Wilson and Menkes diseases." *Chemical Reviews*, 99(9), 2535-2544.

Sarret, G., Balesdent, J., Bouziri, L., Garnier, J. M., Marcus, M. A., Geoffroy, N., Panfili, F., et Manceau, A. (2004). "Zn speciation in the organic horizon of a contaminated soil by micro-x-ray fluorescence, micro- and powder-EXAFS spectroscopy, and isotopic dilution." *Environmental Science & Technology*, 38(10), 2792-2801.

Schafer, H., Hettler, H., Fritsche, U., Pitzen, G., Roderer, G., et Wenzel, A. (1994). "Biotests Using Unicellular Algae and Ciliates for Predicting Long-Term Effects of Toxicants." *Ecotoxicology and Environmental Safety*, 27(1), 64-81.

Scheinost, A. C., Kretzschmar, R., Pfister, S., et Roberts, D. R. (2002). "Combining selective sequential extractions, x-ray absorption spectroscopy, and principal component analysis for quantitative zinc speciation in soil." *Environmental Science & Technology*, 36(23), 5021-5028.

Schlegel, M. (2000). "De l'adsorption du cobalt et du zinc sur l'hectorite et le quartz, à la nucléation hétérogène de phyllosilicates ", Université Joseph Fourier - Grenoble I, Grenoble.

Schlegel, M. L., Manceau, A., Charlet, L., Chateigner, D., et Hazemann, J. L. (2001a). "Sorption of metal ions on clay minerals. III. Nucleation and epitaxial growth of Zn

phyllosilicate on the edges of hectorite." *Geochimica et Cosmochimica Acta*, 65(22), 4155-4170.

Schlegel, M. L., Manceau, A., Charlet, L., et Hazemann, J. L. (2001b). "Adsorption mechanisms of Zn on hectorite as a function of time, pH, and ionic strength." *American Journal of Science*, 301(9), 798-830.

Schoenemann, H. M., Failla, M. L., et Steele, N. C. (1990). "Consequences of severe copper deficiency are independent of dietary carbohydrate in young pigs." *American Journal of Clinical Nutrition*, 52, 147-154.

Schosseler, P. M., Wehrli, B., et Schweiger, A. (1999). "Uptake of Cu2+ by the calcium carbonates vaterite and calcite as studied by continuous wave (CW) and pulse electron paramagnetic resonance." *Geochimica et Cosmochimica Acta*, 63(13-14), 1955-1967.

Schubauer-Berigan, M. K., Dierkes, J. R., Monson, P. D., et Ankley, G. T. (1993). "pH-dependent toxicity of Cd, Cu, Ni and Zn to *Ceriodaphnia dubia*, *Pimephales promelas*, *Hyalella azteca* and *Lumbriculus variegatus*." *Environmental Toxicology and Chemistry*, 12, 1261-1266.

Schwertmann, U., et Carlson, L. (1994). "Aluminum Influence on Iron Oxides: XVII. Unit-Cell Parameters and Aluminum Substitution of Natural Goethites." *Soil Science Society of America Journal*, 58(1), 256-261.

Shao, X. G., Shao, L. M., et Zhao, G. W. (1998). "Extraction of extended X-ray absorption fine structure information from the experimental data using the wavelet transform." *Analytical Communications*, 35(4), 135-137.

Sherman, D. M., et Peacock, C. L. (2010). "Surface complexation of Cu on birnessite (delta-MnO2): Controls on Cu in the deep ocean." *Geochimica et Cosmochimica Acta*, 74(23), 6721-6730.

Slejko, F. L. (1985). *Adsorption Technology: A Step-by-Step Approach to process Evaluation and Application*, Marcel Dekker, New York.

Smirnov, P., et Trostin, V. (2009). "Structural parameters of Cu^{2+} aqua complexes in aqueous solutions of its salts." *Russian Journal of General Chemistry*, 79(8), 1591-1599.

Snell, T. W., Moffat, B. D., Janssen, C., et Persoone, G. (1991). "Acute toxicity tests using rotifers: IV. Effects of cyst age, temperature, and salinity on the sensitivity of *Branchionus calyciflorus*." *Ecotoxicology and Environmental Safety*, 21, 308-317.

Sommer, A. L. (1931). "Copper as an essential to plant growth." *Plant Physiology*, 6, 339-345.

Souren, A. W. M. G. (2000). "Weeping rivers: Cyanide Spills in Romania." Newsletter of the Geochemical Society.

Sposito, G. (2008). *The Chemistry of Soils*, Oxford University Press, New York.

Spurgeon, D. J., Hopkin, S. P., et D.T., J. (1994). "Effects of cadmium, copper, lead and zinc on growth, reproduction and survival of the earthworm Eisenia fetida (Savigny): Assessing the environmental impact of point-source metal contamination in terrestrial ecosystems." *Environmental Pollution*, 84, 123-130.

Stevenson, F. J. (1986). *Cycles of soil: Carbon, nitrogen, phosphorus, sulfur, micronutrients*, John Wiley and Sons Ltd, New York.

Strawn, D. G., Palmer, N. E., Furnare, L. J., Goodell, C., Amonette, J. E., et Kukkadapu, R. K. (2004). "Copper sorption mechanisms on smectites." *Clays and Clay Minerals*, 52(3), 321-333.

Streit, B. (1984). "Effects of high copper concentrations on soil invertebrates (earthworms and oribatid mites): experimental results and a model." *Oecologia*, 64, 381-388.

Stumm, W., et Morgan, J. J. (1996). *Chemical Equilibria and Rates in Natural Waters*, John Wiley & Sons, New York.

Suciu, I., Prodan, L., Lazar, V., Ilea, E., Cocîrla, A., Olinici, L., Paduraru, A., Zagreanu, O., Lengyel, P., Gyrffi, L., et Andru, D. (1981). "Research on copper poisoning." *La Medicina del Lavoro* 3, 190-197.

Suh, I.-K., Ohta, H., et Waseda, Y. (1988). "High-temperature thermal expansion of six metallic elements measured by dilatation method and X-ray diffraction." *Journal of Materials Science*, 23(2), 757-760.

Takacs, M., Alberts, J. J., et Egeberg, P. K. (1999). "Characterization of natural organic matter from eight Norwegian surface waters: Proton and copper binding." *Environment International*, 25(2-3), 315-323.

Teo, B. K. (1986). *EXAFS: Basic Principles and Data Analysis*, Springer-Verlag, Berlin.

Thomas, P. (2010). "La mine de cuivre du Cap Garonne, le Pradet (Var) : visite de la mine et géologie du gisement." http://planet-terre.ens-lyon.fr, Dequincey.,O., Lyon.

Tipping, E. (1998). "Humic ion-binding model VI: An improved description of the interactions of protons and metal ions with humic substances." *Aquatic Geochemistry*, 4(1), 3-48.

Tipping, E., et Hurley, M. A. (1992). "A Unifying Model of Cation Binding by Humic Substances." *Geochimica et Cosmochimica Acta*, 56(10), 3627-3641.

Tonkin, J. W., Balistrieri, L. S., et Murray, J. W. (2004). "Modeling sorption of divalent metal cations on hydrous manganese oxide using the diffuse double layer model." *Applied Geochemistry*, 19(1), 29-53.

Trampczynska, A., Kupper, H., Meyer-Klaucke, W., Schmidt, H., et Clemens, S. (2010). "Nicotianamine forms complexes with Zn(II) in vivo." *Metallomics*, 2(1), 57-66.

Van Campen, D. R., et Mitchell, E. A. (1965). "Absorption of Cu^{64}, Zn^{65}, Mo^{99},and Fe^{59} from ligated segments of the rat gastrointestinal tract." *Journal of Nutrition*, 86, 120-124.

Van Damme, A., Degryse, F., Smolders, E., Sarret, G., Dewit, J., Swennen, R., et Manceau, A. (2010). "Zinc speciation in mining and smelter contaminated overbank sediments by EXAFS spectroscopy." *Geochimica et Cosmochimica Acta*, 74(13), 3707-3720.

Van der Geest, H. G., Greve, G. D., Kroon, A., Kuijl, S., Kraak, M. H. S., et Admiraal, W. (2000). "Sensitivity of characteristic riverine insects, the caddisfly Cyrnus trimaculatus and the mayfly Ephoron virgo, to copper and diazinon." *Environmental Pollution*, 109(2), 177-182.

Verbruggen, N., Hermans, C., et Schat, H. (2009). "Molecular mechanisms of metal hyperaccumulation in plants." *New Phytologist*, 181(4), 759-776.

Vespa, M., Lanson, M., et Manceau, A. (2010). "Natural Attenuation of Zinc Pollution in Smelter-Affected Soil." *Environmental Science & Technology*, 44(20), 7814-7820.

Villars, P., Cenzual, K., Daams, J., Gladyshevskii, R., Shcherban, O., Dubenskyy, V., Melnichenko-Koblyuk, N., Pavlyuk, O., Stoiko, S., Sysa, L.: *Cu2S*. Villars, P., Cenzual, K. (ed.). SpringerMaterials - The Landolt-Börnstein Database (http://www.springermaterials.com).

Visviki, I., et Rachlin, J. W. (1991). "The Toxic Action and Interactions of Copper and Cadmium to the Marine Alga Dunaliella-Minuta, in Both Acute and Chronic Exposure." *Archives of Environmental Contamination and Toxicology*, 20(2), 271-275.

Visviki, I., et Rachlin, J. W. (1994). "Acute and Chronic Exposure of Dunaliella-Salina and Chlamydomonas-Bullosa to Copper and Cadmium - Effects on Growth." *Archives of Environmental Contamination and Toxicology*, 26(2), 149-153.

Walsh, C. J. (2000). "Urban impacts on the ecology of receiving waters: a framework for assessment, conservation and restoration." *Hydrobiologia*, 431(2), 107-114.

Walsh, L. H., Erhardt, W. H., et Seibel, H. D. (1972). "Copper toxicity in snapbeans (*Phaseolus vulgaris* L.). ." *Journal of Environmental Quality*, 1, 197-200.

Warren, L. A., et Haack, E. A. (2001). "Biogeochemical controls on metal behaviour in freshwater environments." *Earth-Science Reviews*, 54(4), 261-320.
Watling, H. R. (2006). "The bioleaching of sulphide minerals with emphasis on copper sulphides - A review." *Hydrometallurgy*, 84(1-2), 81-108.
Weesner, F. J., et Bleam, W. F. (1997). "X-Ray Absorption and EPR Spectroscopic Characterization of Adsorbed Copper(II) Complexes at the Boehmite (AlOOH) Surface." *Journal of Colloid and Interface Science*, 196(1), 79-86.
WHO. (1973). "Trace elements in human nutrition; a report of a WHO Expert Committee." *532*, World Health Organization, Geneva.
WHO. (1998). "Environmental Health criteria 200 Copper." World Health Organization, Geneva.
Wightwick, A. M., Mollah, M. R., Partington, D. L., et Allinsono, G. (2008). "Copper fungicide residues in Australian vineyard soils." *Journal of Agricultural and Food Chemistry*, 56(7), 2457-2464.
Winiarski, T., Bedell, J. P., Delolme, C., et Perrodin, Y. (2006). "The impact of stormwater on a soil profile in an infiltration basin." *Hydrogeology Journal*, 14(7), 1244-1251.
Xu, Y. P., Schwartz, F. W., et Traina, S. J. (1994). "Sorption of Zn-2+ and Cd-2+ on Hydroxyapatite Surface." *Environmental Science & Technology*, 28(8), 1472-1480.
Yamamoto, H., Tatsuyama, K., et Uchiwa, T. (1985). "Fungal flora of soil polluted with copper." *Soil Biology and Biochemistry*, 17(6), 785-790.
Yeh, J.-J., et Lindau, I. (1985). "Atomic Subshell Photoionization Cross Sections and Asymmetry Parameters: 1<Z<103." *Atomic Data and Nuclear Data Tables*, 32(1).
Zhou, B., et Gitschier, J. (1997). "hCTR1: A human gene for copper uptake identified by complementation in yeast." *Proceedings of the National Academy of Sciences of the United States of America*, 94(14), 7481-7486.
Zoltán, S., Attila, D., Torsten, W. V., Sebastien, P., Jan, K., Szabolcs, L., Odblac, ssy, Mária, B., Chuan-Chou, S., et Ernst, H. (2009). "Bronze Age volcanic event recorded in stalagmites by combined isotope and trace element studies." *Rapid Communications in Mass Spectrometry*, 23(6), 801-808.