

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

Université Mohamed Khider – Biskra
Faculté des Sciences et de la technologie
Département : Génie civil et Hydraulique
Ref :



جامعة محمد خيضر بسكرة
كلية العلوم و التكنولوجيا
قسم : الهندسة المدنية و الري
المرجع:

Thèse présentée en vue de l'obtention
Du diplôme de
Doctorat en sciences

Spécialité : Sciences Hydrauliques

**Modélisation de l'efficacité de la floculation
Effets des caractéristiques de la phase solide
et des paramètres opératoires**

Présentée par :
Fateh SEKIOU
Maître-assistant A

Soutenue publiquement le 26/01/2013

Devant le jury composé de :

Dr. DEBABECHE Mahmoud	Grade	Professeur	Président	Université de Biskra
Dr. YUCEF Leila	Grade	Professeur	Examinatrice	Université de Biskra
Dr. REMINI Boualem	Grade	Professeur	Examineur	Université de Blida
Dr. SAAD HAMOUDI Abdelamir	Grade	Professeur	Examineur	Université de Chelf
Dr. KELLIL Amara	Grade	Professeur	Rapporteur	Université de Blida

A mes parents
A ma petite famille

REMERCIEMENTS

Cette thèse de doctorat a été réalisée sous la direction de Monsieur *AMARA KELLIL*, Professeur à l'université Saad Dahlab à Blida. Ce moment est pour moi, l'occasion de lui témoigner ma plus grande estime et ma reconnaissance.

Monsieur *KELLIL*, je vous remercie de m'a avoir accordé votre confiance tout au long de ce travail.

- Je tiens particulièrement à remercier Monsieur *Mahmoud DEBABECHE*, Professeur à l'université de Biskra pour l'intérêt qu'il a porté à mes travaux. Je lui suis reconnaissant d'avoir accepté de juger ce travail et de présider mon jury de thèse.
- Je suis très reconnaissant de l'honneur que me fait Madame *Leïla YUCEF*, Professeur à l'université de Biskra en acceptant la tâche d'évaluer en qualité d'examineur les travaux présentés dans cette thèse.
- Je remercie vivement monsieur *Abd El-Amir SAAD HAMOUDI*, Professeur à l'université de Chlef pour l'intérêt qu'il a bien voulu porter à ce travail et d'avoir accepté de juger ce travail en qualité d'examineur.
- Je tiens également à remercier Monsieur *Boualem REMINI*, Professeur à l'université de Blida, il m'a fait le plaisir d'accepter de prendre part au jury et d'examiner mon travail, j'apprécie grandement l'intérêt qu'il a porté à mes travaux de thèse.

Une partie des travaux de recherche de cette thèse a été effectuée au laboratoire d'épuration et de valorisation des eaux de l'université Saad Dahlab de Blida, au centre de recherche et de développement CRD et au laboratoire SAIDAI. Je tiens à exprimer mon respect et mes remerciements aux responsables pour les efforts qu'ils déploient afin de réunir les meilleures conditions pour la recherche scientifique et de créer un environnement favorable à son développement.

Mes remerciements s'adressent aussi à mes collègues pour leur précieuse aide. Sans oublier ceux qui ont participé à ce travail, de près ou de loin, et qui ont contribué à le rendre des plus agréable.

Enfin, je tiens à exprimer mes remerciements à mes parents et à ma petite famille pour leur soutien et leur compréhension tout le long de ce travail.

Nom : SEKIOU

Prénom : Fateh

Modélisation de l'efficacité de la floculation : Effets des caractéristiques de la phase solide et des paramètres opératoires

Thèse de doctorat en science, Spécialité Sciences hydrauliques.

RESUME :

La thèse s'inscrit dans le cadre de la modélisation de la floculation d'une suspension aqueuse. Des approches de régression multiple, de régression régularisée, de régression neuronale simple et de régression neuronale basée sur les composantes principales sont proposées afin de développer des modèles statistiques de prédiction de la dose optimale de floculation, des constantes cinétiques de floculation et des paramètres de finalité de traitement "la turbidité finale et l'aluminium résiduel".

L'originalité dans ce travail réside principalement dans la modélisation des constantes cinétiques de floculation et leur intégration dans les modèles de régression.

La démarche adoptée dans cette étude porte sur deux volets :

Le premier volet est expérimental, il a un double objectif : d'une part la création d'une base de données relative à la floculation de six types de particules solides dispersées et d'autre part, l'analyse des effets de la phase solide et des conditions opératoires sur la floculation.

Le deuxième volet du travail a pour objectif la modélisation des paramètres de floculation; il consiste d'une part à utiliser "l'Analyse en Composantes Principales" et "l'Analyse de Classification Hiérarchique" dans le choix des variables d'entrée et d'autre part la formulation des modèles de régression.

Les résultats expérimentaux confirment l'effet des caractéristiques de la phase solide, des paramètres de floculation et de leurs interactions sur le mécanisme, la cinétique et la performance de floculation.

Quant aux résultats de la modélisation, il s'est avéré que les modèles de régression multiple sont simples et satisfaisants mais encourent un grand risque d'instabilité de l'estimation, alors que les modèles de régression régularisée permettent de remédier au problème d'instabilité de l'estimation, ces modèles sont moins complexes et plus crédibles que les modèles de régression multiple. Il ressort aussi de cette étude que la régression neuronale simple et celle basée sur les composantes principales s'adaptent bien à la non linéarité et à la variabilité du procédé. Leurs performances de prédiction sont très élevées, cependant la régression neuronale basée sur les composantes principales réduit grandement la complexité du modèle et permet des modèles parcimonieux.

MOTS CLES

Floculation, Constantes Cinétiques, Matière Organique, Modélisation Analyse en Composantes Principales, Classification Hiérarchique, Régression Multiple, Réseaux de Neurones Artificiels.

ABSTRACT:

The thesis is part of the framework of the modeling of flocculation of an aqueous suspension. Approaches of multiple regression, regularized regression, simple neural regression and neural regression based on principal components are proposed to develop statistical models for predicting the optimal dose of flocculation, the kinetic constants of flocculation and the parameters of treatment finality "final turbidity and residual aluminium".

The originality in this work lies mainly in the modeling of kinetic constants of flocculation and their integration into the regression models.

The approach taken in this study is twofold:

The first is experimental, it has two objectives: firstly creating a database on the flocculation of six types of dispersed solid particles and secondly analyzing the effects of the solid phase and of the operating conditions on the flocculation.

The second part of the work highlights the modeling of flocculation parameters, it consists, on one hand, of the use of "Principal Component Analysis" and "Hierarchical Cluster Analysis" in the choice of input variables and, On the other hand, of the formulation of regression models.

Experimental results confirm the effect of the solid phase characteristics, flocculation parameters, and their interactions on the mechanism, kinetic, and flocculation performance.

As for the modeling results, it turned out that the multiple regression models are simple and satisfying, but incur a risk of instability of the estimation, however the regularized regression models overcomes the problem of estimation instability, these models are less complex and more credible than the multiple regression models. It also appears from this study that the simple neural regression and that based on principal components adapt well to the nonlinearity and process variability. Their performance prediction is very high; however, the neural regression based on principal components greatly reduces the complexity of the model and allows parsimonious models.

KEYWORDS

Flocculation, Kinetic Constants, Organic Matter, Modeling, Principal Component Analysis, Hierarchical Cluster, Multiple Regression, Artificial Neural Networks.

نمذجة كفاءة الترويب : أثر خصائص القسم الصلب و العناصر العملية

ملخص :

هذه الأطروحة تهتم بنمذجة ترويب معلق مائي، وذلك من خلال تطوير نماذج إحصائية قادرة على التنبؤ بالجرعة المثالية للترويب، و بالتوايت الحركية للترويب و كذا بعناصر نهاية المعالجة "العكر النهائي والألمنيوم المتبقي". طرق عديدة مقترحة في هذا الصدد منها : الانحدار المتعدد، الانحدار المتعدد النظامي، الانحدار العنصري البسيط و الانحدار العنصري القائم على المركبات الرئيسية.

الجديد في هذا العمل يكمن أساسا في وضع نماذج للنوابت الحركية للترويب وإدماج هذه الثوابت في نماذج الانحدار.

النهج المتبع في هذه الدراسة ينقسم إلى شقين: الشق الأول تجريبي، يتمحور حول إنشاء قاعدة بيانات عن ترويب ستة أنماط من الجزيئات الصلبة العالقة و أيضا دراسة أثر القسم الصلب و الشروط العملية على الترويب، الشق الثاني من العمل يهدف إلى إعداد النماذج الإحصائية بدءا باستخدام "التحليل وفق المركبات الرئيسية" و "تحليل التصنيف السلمي" في اختيار متغيرات المداخل و مرورا إلى صياغة نماذج الانحدار.

النتائج التجريبية تؤكد تأثير خصائص القسم الصلب و عناصر الترويب و التفاعلات فيما بينها على حركية، وألية و كفاءة الترويب. أما بالنسبة لنتائج النمذجة، اتضح أن نماذج الانحدار المتعدد بسيطة و مرضية، ولكن تنطوي على خطر عدم استقرار التقدير، في حين أن نماذج الانحدار النظامي تغلب على مشكلة العلاقات الخطية المتداخلة و عدم الاستقرار، هذه النماذج هي أقل تعقيدا و أكثر مصداقية من نماذج الانحدار المتعدد.

ومما اتضح أيضا من هذه الدراسة أن الانحدار العنصري البسيط و الانحدار العنصري القائم على المركبات الرئيسية يتكيف بشكل جيد مع لا خطية ومع بعيرات عملية الترويب، ويسمحان بكفاءة تنبؤ عالية جدا، إلا أن الانحدار العنصري القائم على المركبات الرئيسية يقال بدرجة أكبر من تعقيد النموذج و يسمح بنماذج جد شحيحة.

الكلمات المفتاحية

الترويب ، الثوابت الحركية، مادة عضوية، النمذجة، التحليل وفق المركبات الرئيسية، التصنيف السلمي، الانحدار المتعدد، الشبكات العنصرية الاصطناعية.

TABLE DES MATIERES

DEDICACES	i
REMERCIEMENTS	ii
RESUME	iii
ABSTRACT	iv
ملخص	v
LISTE DES FIGURES	xii
LISTE DES TABLEAUX	xx
PRINCIPALES NOTATIONS	xxiv
INTRODUCTION GENERALE	1
Chapitre 1 : COAGULATION/FLOCCULATION DES SUSPENSIONS COLLOIDALES : THEORIE, PRACTIQUE ET MODELISATION	6
I. SYSTEMES COLLOIDAUX	7
I-1. Définitions générales	7
I-2. Propriétés des suspensions colloïdales	8
I-3. Stabilité des suspensions colloïdales	9
I-4. Coagulation des suspensions colloïdales	11
I-4.1. Compression de la double couche	11
I-4.2. Adsorption et neutralisation de charges par ions métalliques polyvalents	12
I-4.3. Adsorption de macromolécules et pontage	12
I-4.4. Emprisonnement des particules dans un précipité	13
I-4.5. Aspect cinétique de la coagulation	13
I-5. Flocculation des suspensions colloïdales	13
I-5.1. Flocculation péricinétique	13
I-5.2. Flocculation orthocinétique	14
II. PRACTIQUE DE LA COAGULATION/FLOCCULATION DES SUSPENSIONS COLLOIDALES	17
II-1. Formation et cassage de floc	17
II-2. Influence de pH sur la flocculation	26
II-3. Effet de la température sur la flocculation	26
II-4. Effet de la nature et charges des colloïdales sur la flocculation	27
II-5. Effet de la minéralisation sur la flocculation	29
II-6. Influence du coagulant	31
II-7. Influence des conditions d'agitation sur la flocculation	35
II-8. Conclusion	37

III. MODELISATION DE LA FLOCCULATION	38
III-1. Appareils et méthodes de détermination de la dose de coagulant.	38
III-1.1. Jar test	38
III-1.2. Détecteur de courant cinétique	38
III-1.3. Mesure de potentiel zêta	39
III-1.4. Titration colloïdale	40
III-1.5. Mesure de la turbidité par l'autofloc.	40
III-1.6. Comptage des particules	40
III-2. Analyse exploratoire multidimensionnelle de la floculation	41
III-2.1. Introduction	41
III-2.2. Analyse exploratoire multivariée de la floculation	42
III-2.3. Conclusion	46
III-3. Modélisation de la floculation	46
III-3.1. Modélisation mathématique de la floculation	46
III-3.2. Modélisation statistique par régression multiple de la floculation	46
III-3.3. modélisation statistique par régression neuronale de la floculation	49
IV. CONCLUSION DU CHAPITRE 1	57
Chapitre 2 : ANALYSE ET MODELISATION DES DONNEES	58
I- INTRODUCTION	59
II-PRINCIPE GENERAL DE L'ANALYSE MULTIDIMENSIONNELLE	60
II-1. Introduction	60
II-2. Représentation géométrique élémentaire d'un tableau de données	61
III-ANALYSE EN COMPOSANTES PRINCIPALES	61
III-1. Introduction	61
III-2. Principes de l'analyse en composantes principales.	62
III-3. Démarche d'interprétation d'une ACP	64
IV- ANALYSE DE CLASSIFICATION	65
IV-1. Classification non hiérarchique	65
IV-2. Classification ascendante hiérarchique	66
IV-3. Choix d'une distance entre individus statistiques	66
IV-4. Choix d'un critère d'agrégation	66
IV-5. Complémentarité entre analyse factorielle et analyse de classification	67

V- MODELISATION	68
V-1. Introduction	68
V-2. Modèles mathématiques	68
V-2.1. Modèles basées sur des principes fondamentaux scientifiques	68
V-2.2. Modèles de balance massique	68
V-2.3. Modèles empiriques (statistiques)	69
V-3. Modélisation paramétrique.	71
V-3.1. Principe de la régression linéaire	71
V-3.2. Sélection du meilleur modèle	75
V-3.2.1. Sélection des données en pratique	75
V-3.2.2. Procédure de choix de modèle	76
V-3.2.3. Critères de choix de modèle	77
V-3.2.4. Qualité de prédiction du modèle	79
V-3.2.5. Stabilité des coefficients de régression	81
V-3.3. Modèle de régression sur composantes principales (régression régularisée ou orthogonalisée)	82
V-3.4. Récapitulatif	
V-4. Modélisation non paramétrique par réseaux de neurones	84
V-4.1. Introduction	84
V-4.2. Définition de neurones artificiels	84
V-4.3. Architecture d'un réseau de neurones	86
V-4.4. Modélisation à l'aide de réseau de neurones	87
V-4.5. Mise en œuvre de la régression neuronale	88
V-4.5.1. Détermination des entrées/sorties du réseau de neurones	88
V-4.5.2. Choix et préparation des échantillons	89
V-4.5.3. Optimisation de la structure du réseau	90
V-4.5.4. Apprentissage du réseau de neurones	90
V-4.5.5. Validation du modèle	93
V-4.5.5. Contribution des variables dans les modèles de réseau de neurones	93
V-4.6. Modélisation à l'aide de réseaux de neurones basés sur les composantes principales	94
V-4.7 Avantages et inconvénients des réseaux de neurones	95
V-4.8. Récapitulatif	96
VI. CONCLUSION DU CHAPITRE 2	98

Chapitre 3 : ETUDE EXPERIMENTALE DE LA FLOCCULATION DES SUSPENSIONS SYNTHETIQUES	99
I. OBJECTIF DE L'ETUDE EXPERIMENTALE	100
II. PRESENTATION DES PARTICULES SOLIDES	100
III. CARACTERISATION DES PARTICULES SOLIDES	101
III-1. Analyse par diffraction aux rayons x	101
III-2. Analyse par spectroscopie infrarouge	103
III-3. Analyse par spectroscopie d'absorption atomique	105
III-4. Analyse de carbone organique total, carbone minéral et perte au feu (à 550 °c)	105
III-5. Analyse granulométrique au laser	106
IV. ESSAIS DE FLOCCULATION DES SUSPENSIONS SYNTHETIQUES	109
IV-1 Suspensions synthétiques	
IV-2 Description des essais de flocculation	110
IV-3 Résultats et discussion	
IV-3.1. Evolution de la performance de la flocculation	110
IV-3.2. Interactions entre la puissance d'agitation, la dose du coagulant et la nature du solide dispersé	114
IV-3.3. Etude de la cinétique de la flocculation	117
IV-3.4. Interaction entre la matière minérale et la matière organique	124
IV-3.4.1. Effet de l'interaction matière minérale/matière organique sur la performance de flocculation.	125
IV-3.4-2. Effet de l'interaction matière minérale/matière organique sur la relation entre la puissance d'agitation et la dose de coagulant.	126
IV-3.4-3. Effet de l'interaction matière minérale/matière organique sur la cinétique de flocculation.	127
V. CONCLUSION DU CHAPITRE 3	129
Chapitre 4 : ANALYSE EXPLORATOIRE DES DONNEES EXPERIMENTALES.	133
I. INTRODUCTION	134
II. DESCRIPTION DES DONNEES	134

III. ANALYSE DES DONNEES	136
III-1. Analyse univariée	136
III-2. Analyse bivariée	139
III-3. Analyse multidimensionnelle.	145
III-3.1. Analyse en composantes principales	145
III-3.2. Analyse de classification hiérarchique	155
IV. CHOIX DES VARIABLES POUR LA MISE EN ŒUVRE DES MODELES	158
IV-1. Recherche bibliographique	158
IV-2. Analyse en composantes principales	161
IV-3. Analyse de classification	161
IV-4. Analyse multidimensionnelle des variables retenues.	163
V. CONCLUSION DU CHAPITRE 4	164
Chapitre 5 : MODELISATION STATISTIQUE	166
I. INTRODUCTION	167
II. MODELISATION PARAMETRIQUE DE LA DOSE OPTIMALE	168
II-1. Choix et préparation des données	168
II-2. Modèles de régression multiple	170
II-2.1. Premier modèle de régression multiple de la dose	170
II-2.2. Deuxième modèle de régression multiple de la dose	177
II-2.3. Troisième modèle de régression multiple de la dose	182
II-2.4. Modèle de régression régularisée	188
III. MODELISATION NON PARAMETRIQUE DE LA DOSE OPTIMALE	195
III.1. Modèle neuronal	195
III.1.1. Choix des paramètres d'entrée et découpage de la base de données	195
III.1.2. Développement du modèle neuronal	197
III-2. Modèle neuronal base sur les composantes principales	203
IV. CONCLUSION	207
V. MODELISATION PARAMETRIQUE DES CONSTANTES CINETIQUES K_A ET K_B	209
V.1. Choix et préparation des données	209
V.2. Modèles de régression multiple	210
V.2.1. Modèle de régression de la constante de cassage K_B	210

V.2.2. Modèle de régression régularisée de la constante de cassage K_B	213
V.2.3. Modèle de régression de la constante de formation K_A	216
V.2.4. Modèle de régression régularisée de la constante de formation K_A	218
VI. MODELISATION NON PARAMETRIQUE DES CONSTANTES CINETIQUES K_A ET K_B	220
VI.1. Modèle neuronal.	221
VI.2. Modèle neuronal basé sur les composantes principales	224
VII CONCLUSION	227
VIII. MODELISATION DES PARAMETRES DE FINALITE DE TRAITEMENT : LA TURBIDITE FINALE "T_F" ET L'ALUMINIUM RESIDUEL "Al_r" .	228
VIII.1. Choix et préparation des données	228
VIII.2. Modèle neuronal	230
VIII.3. Modèle neuronal base sur les composantes principales	236
IX. CONCLUSION	241
X. CONCLUSION DU CHAPITRE 5	242
CONCLUSION GENERALE	248
REFERENCES BIBLIOGRAPHIQUES	255
ANNEXES	273
Annexe A : Table de Durbin watson (test unilatérale au seuil de 5%) Valeurs de d_L et d_U	274
Annexe B : Table des fractiles $f(v_1, v_2)$ pour une loi $F(v_1, v_2)$ (Fisher) (5%)	275
Annexe C : Table des fractiles $t_v(p)$ pour une loi T (Student)	276
Annexe D : Valeurs significatives du coefficient de corrélation	277
Annexe E : Procédure de recherche de l'influence des variables en modèle de régression linéaire	278
Annexe F : Procédure de calcul de l'importance des variables par approche de poids de connexion.	279

LISTE DES FIGURES

Figure 1.1	Diagramme d'action du sulfate d'aluminium en fonction du pH FRANCESCHI (1991).	32
Figure 2.1	Démarche d'analyse et de modélisation suivie dans notre cas d'étude	59
Figure 2.2	Principe de présentation géométrique d'après LEBART (1997)	61
Figure 2.3	Prévision linéaire LEBART (1997).	72
Figure 2.4	Schématisation du modèle linéaire LEBART (1997)	72
Figure 2.5	Echelle de Turkey d'après CONFAIS (2007)	74
Figure 2.6	Modèle de neurone artificiel	85
Figure 2.7	Exemple d'architecture d'un PMC à x_n variables d'entrée, une seule couche cachée de fonction d'activation sigmoïde et une sortie de fonction d'activation identité.	87
Figure 2.8	Architecture du modèle neuronal basé sur les composantes principales	94
Figure 3.1	Spectre DRX du kaolin	102
Figure 3.2	Spectre DRX du la vase	102
Figure 3.3	Spectre DRX du la Boue	102
Figure 3.4	Composition minéralogiques des solides.	103
Figure 3.5	Spectre IR du kaolin	103
Figure 3.6	Spectre IR de la vase	104
Figure 3.7	Spectre IR de la Boue	104
Figure 3.8	Spectre IR de la Tourbe	104
Figure 3.9	Résultats d'analyse d'absorption atomique	105
Figure 3.10	Résultats d'analyse de C.O.T ; C.M et perte au feu (MO).	106
Figure 3.11	Distributions granulométriques des différents solides	106
Figure 3.12	Pourcentage de fractions colloïdales des particules solides	107
Figure 3.13	Valeurs de surface spécifique, uniformité et "span" des particules solides	107
Figure 3.14	Diamètres moyen en volume et en surface des particules solides	108
Figure 3.15	Valeurs de "percentiles" D(0,1) D(0,5) et D(0,9) des particules solides.	108
Figure 3.16	Courbes d'étalonnage (Turbidité- Concentration des particules solides	109
Figure 3.17	Courbe d'étalonnage (Absorbance à 535 nm -Concentration de l'aluminium)	110
Figure 3.18	Evolution de la turbidité finale en fonction de la dose du coagulant	111
Figure 3.19	Evolution de la dose optimale et du rapport R_{Al}/ R_T en fonction du taux de matière organique.	111
Figure 3.20	Evolution du rendement R_T et du taux R_{Al} en fonction du taux de matière organique.	111
Figure 3.21	Evolution du rendement R_T en fonction de la surface spécifique.	113

Figure 3.22	Evolution du rendement R_T en fonction du diamètre moyen de particules.	113
Figure 3.23	Evolution du rendement R_T en fonction du "span"	114
Figure 3.24	Evolution du rendement R_T en fonction de pourcentage de " $\text{SiO}_2 + \text{Al}_2\text{O}_3$ "	114
Figure 3.25	Evolution de la turbidité finale en fonction de l'énergie dissipée pour les suspensions du kaolin, de la vase et de la boue	114
Figure 3.26	Evolution de la turbidité finale en fonction de l'énergie dissipée pour les suspensions de mélange I, de mélange II et de la tourbe	115
Figure 3.27	Evolution de la puissance optimale d'agitation lente en fonction de la nature du solide dispersé et de la dose du coagulant injecté	116
Figure 3.28	Cinétique de floculation des suspensions des particules solides (kaolin, vase, boue et tourbe)	118
Figure 3.29	Evolution des constantes de cinétiques K_A , K_E et du rapport K_E/K_A en fonction de la nature du solide dispersé	119
Figure 3.30	Evolution du K_E/K_A en fonction du taux de matière organique	120
Figure 3.31	Evolution du K_E/K_A en fonction du diamètre moyen des particules primaires.	120
Figure 3.32	Evolution du K_E/K_A en fonction de la surface spécifique des particules primaires.	121
Figure 3.33	Evolution du K_E/K_A en fonction du taux de " $\text{SiO}_2 + \text{Al}_2\text{O}_3$ "	121
Figure 3.34	Evolution du K_E/K_A en fonction de la largeur de distribution "span"	122
Figure 3.35	Evolution du temps d'équilibre en fonction de la nature du solide	123
Figure 3.36	Evolution du rendement d'équilibre en fonction de la nature du solide	123
Figure 3.37	Rendements expérimentaux et théoriques de floculation des suspensions de mélanges I et II	125
Figure 3.38	Evolution de la puissance optimale d'agitation lente pour les suspensions de la vase, des deux mélanges I et II et de la tourbe en fonction de la dose du coagulant injecté.	126
Figure 3.39	Cinétique de floculation des suspensions de la vase, des deux mélanges I et II et de la tourbe.	127
Figure 3.40	Evolution K_E/K_A en fonction de la nature du solide (tourbe, mélange du rapport I, mélange II et vase)	128
Figure 3.41	Evolution de la constante K_E en fonction de la nature du solide (tourbe, mélange I, mélange II et vase)	128
Figure 3.42	Evolution du temps d'équilibre pour les suspensions de la tourbe, du mélange I, du mélange II et de la vase)	129
Figure 4.1	Présentation box plot «boîtes à moustache» des dispersions inter-individuelle des différents types du solide dispersé.	138
Figure 4.2	Nuages de points de la dose optimale en fonction des paramètres de floculation	142
Figure 4.3	Nuages de points de la constante de formation K_A en fonction des paramètres de floculation	143

Figure 4.4	Nuages de points de la constante de formation K_B en fonction des paramètres de floculation	144
Figure 4.5	Evolution des composantes principales	146
Figure 4.6	Projection des individus sur le plan principal 1-2	147
Figure 4.7	Projection des individus sur le plan principal 3-4	148
Figure 4.8	Projection des variables sur le plan principal 5-6	149
Figure 4.9	Projection des individus sur le plan principal 1-2	150
Figure 4.10	Projection simultanée des variables et des solides sur le plan principale 1-2	152
Figure 4.11	Projection des variables sur le plan principal 1-3	153
Figure 4.12	Projection des variables sur le plan principal 1-4	153
Figure 4.13	Projection des variables sur le plan principal 1-5	154
Figure 4.14	Projection des variables sur le plan principal 1-6	154
Figure 4.15	Dendrogramme de classification hiérarchique par la méthode de Ward.	156
Figure 4.16	Dendrogramme de classification hiérarchique par la méthode de la moyenne pondérée.	157
Figure 4.17	Dendrogramme de classification hiérarchique par la méthode de Diamètre.	157
Figure 4.18	Répartition dans le plan principal 1-2 des variables retenues pour la modélisation.	163
Figure 4.19	Répartition dans le plan principal 1-3 des variables retenues pour la modélisation.	164
Figure 4.20	Dendrogramme de classification hiérarchique par la méthode de Ward, des variables retenues pour la modélisation	164
Figure 5.1	Présentation box plot «boites à moustache» des données d'apprentissage et de validation, Modélisation paramétrique de la dose	169
Figure 5.2	Schéma de régression multiple de la dose optimale	170
Figure 5.3	Ajustement des résidus et résidus supprimés, 1 ^{er} modèle de régression multiple de la dose (set d'apprentissage, IC à 95 %)	173
Figure 5.4	Ajustement à une loi Normale des résidus, 1 ^{er} modèle de régression multiple de la dose (set d'apprentissage, IC à 95 %)	173
Figure 5.5	Vérification de l'homoscédasticité par le nuage de points des résidus, 1 ^{er} modèle de régression multiple de la dose (set de validation)	175
Figure 5.6	Ajustement à une loi Normale des résidus, 1 ^{er} modèle de régression multiple de la dose (set de validation, IC à 95 %)	175
Figure 5.7	Ajustement des valeurs observées et des valeurs prévues, 1 ^{er} modèle de régression multiple de la dose [A] set d'apprentissage, IC à 95% [B] set de validation, IC à 95%.	175
Figure 5.8	Influence des paramètres sur l'équation de la régression, 1 ^{er} modèle de régression multiple de la dose.	176
Figure 5.9	Ajustement des résidus et résidus supprimés, 2 ^{ème} modèle de régression multiple de la dose (set d'apprentissage, IC à 95 %)	179

Figure 5.10	Ajustement à une loi Normale des résidus, 2 ^{ème} modèle de régression multiple de la dose (set d'apprentissage, IC à 95 %)	179
Figure 5.11	Vérification de l'homoscédasticité par le nuage de points des résidus, 2 ^{ème} modèle de régression multiple de la dose (set de validation)	180
Figure 5.12	Ajustement à une loi Normale des résidus, 2 ^{ème} modèle de régression multiple de la dose (set de validation, IC à 95 %)	180
Figure 5.13	Ajustement des valeurs observées et des valeurs prévues, 2 ^{ème} modèle de régression multiple de la dose A set d'apprentissage, IC à 95% B set de validation IC à 95%.	181
Figure 5.14	Influence des paramètres sur l'équation de régression, 2 ^{ème} modèle de régression multiple de la dose.	181
Figure 5.15	Ajustement des résidus et des résidus supprimés, 3 ^{ème} modèle de régression multiple de la dose, set d'apprentissage, I.C. à 95 %	184
Figure 5.16	Ajustement à une loi Normale des résidus, 3 ^{ème} modèle de régression multiple de la dose, set d'apprentissage, I.C. à 95 %	184
Figure 5.17	Vérification de l'homoscédasticité par le nuage de points des résidus, 3 ^{ème} modèle de régression multiple de la dose (set de validation)	185
Figure 5.18	Ajustement à une loi Normale des résidus, 3 ^{ème} modèle de régression multiple de la dose (set de validation, IC à 95 %)	185
Figure 5.19	Ajustement des valeurs observées et des valeurs prévues, 3 ^{ème} modèle de régression multiple de la dose A set d'apprentissage, IC à 95% B set de validation IC à 95%.	186
Figure 5.20	Influence des paramètres sur l'équation de la régression, 3 ^{ème} modèle de régression multiple de la dose.	187
Figure 5.21	R ² en fonction du nombre de composantes principales, modèle de régression régularisée de la dose. A les composantes sont prises selon l'importance de variance, B les composantes sont retenues selon la procédure de régression "Forward Regression"	189
Figure 5.22	Ajustement des résidus et résidus supprimés, modèle de régression régularisée de la dose (set d'apprentissage, IC à 95 %)	191
Figure 5.23	Ajustement à une loi Normale des résidus, modèle de régression régularisée de la dose (set d'apprentissage, IC à 95 %)	191
Figure 5.24	Ajustement à une loi Normale des résidus, modèle de régression régularisée de la dose (set de validation, IC à 95 %)	193
Figure 5.25	Vérification de l'homoscédasticité par le nuage de points des résidus, modèle de régression régularisée de la dose (set de validation)	193
Figure 5.26	Ajustement des valeurs observées et des valeurs prévues, modèle de régression régularisée de la dose A set d'apprentissage, IC à 95% B set de validation IC à 95%.	193

Figure 5.27	Présentation box plot « boîtes à moustache » des données d'apprentissage et de validation (test), Modélisation non paramétrique de la dose	196
Figure 5.28	REQM du jeu de validation en fonction du nombre de neurones dans la couche cachée calculé avec [A] la fonction logistique [B], la fonction tangente hyperbolique, modèle de régression neuronale de la dose.	199
Figure 5.29	Architecture du perceptron multicouche 11/8/1 pour la prédiction de la dose	200
Figure 5.30	Ajustement à une loi Normale des résidus, modèle de régression neuronale de la dose [A] set d'apprentissage, IC à 95 % [B] set de validation, IC à 95 %)	200
Figure 5.31	Vérification de l'homoscédasticité par le nuage de points des résidus, modèle de régression neuronale de la dose (set de validation)	201
Figure 5.32	Ajustement des valeurs observées et des valeurs prévues, modèle de régression neuronale de la dose. [A] set d'apprentissage, IC à 95%, [B] set de validation, IC à 95%,	202
Figure 5.33	Importance relative des variables d'entrée, modèle de régression neuronale de la dose.	202
Figure 5.34	R^2 en fonction du nombre de composantes principales dans la couche d'entrée, modèle de régression neuronale de la dose.	203
Figure 5.35	REQM du jeu de validation en fonction du nombre de neurones dans la couche cachée calculée avec [A] la fonction logistique [B], la fonction tangente hyperbolique, modèle de régression neuronale sur composantes principales de la dose.	204
Figure 5.36	Architecture du perceptron multicouche 6/6/1 pour la prédiction de la dose.	205
Figure 5.37	Ajustement à une loi Normale des résidus, modèle de régression neuronale sur composantes principales de la dose, [A] set d'apprentissage, IC à 95 % [B] set de validation, IC à 95 %)	206
Figure 5.38	Vérification de l'homoscédasticité par le nuage de points des résidus, modèle de régression neuronale sur composantes principales (set de validation)	206
Figure 5.39	Ajustement des valeurs observées et des valeurs prévues, modèle de régression neuronale sur composantes principales de la dose. [A] set d'apprentissage, IC à 95%, [B] set de validation, IC à 95%.	207
Figure 5.40	Variations de [A] La racine carrée de l'erreur quadratique moyenne REQM, [B] Le coefficient de détermination R^2 , pour les différents modèles de régression de la dose optimale.	208
Figure 5.41	Schéma de régression multiple des constantes cinétiques K_A et K_E	210
Figure 5.42	Ajustement à une loi Normale des résidus, modèle de régression multiple de la constante de cassage K_E , IC à 95 %.	212

Figure 5.43	Ajustement des valeurs observées et des valeurs prévues, modèle de régression multiple de la constante de cassage K_E , IC à 95 %.	212
Figure 5.44	Influence des paramètres sur l'équation de régression multiple de la constante de cassage K_E .	212
Figure 5.45	Ajustement à une loi Normale des résidus, modèle de régression régularisée de la constante de cassage K_E , IC à 95 %	215
Figure 5.46	Ajustement des valeurs observées et des valeurs prévues, modèle de régression régularisée de la constante de cassage K_E [s], IC à 95 %	215
Figure 5.47	Ajustement à une loi Normale des résidus, modèle de régression multiple de la constante de formation de K_A , IC à 95 %	217
Figure 5.48	Ajustement des valeurs observées et des valeurs prévues, modèle de régression multiple de la constante de formation K_A , IC à 95%	217
Figure 5.49	Influence des paramètres sur l'équation du modèle de régression de la constante de formation K_A A COT < 5.55 B COT ≥ 0.55	218
Figure 5.50	Ajustement à une loi Normale des résidus, modèle de régression régularisée de la constante de formation K_A , IC à 95 %	219
Figure 5.51	Ajustement des valeurs observées et des valeurs prévues, modèle de régression régularisée de la constante de formation K_A , IC à 95%	219
Figure 5.52	R^2 en fonction du nombre de neurones dans la couche cachée, modèle de régression neuronale des constantes cinétiques K_A et K_E	221
Figure 5.53	Architecture du perceptron multicouche 3/3/2 modèle de régression neuronal des constantes cinétiques K_A et K_E	221
Figure 5.54	Ajustement à une loi Normale des résidus, modèle de régression neuronale des constantes cinétiques A Constante de formation K_A , IC à 95 % B Constante de cassage K_E (IC à 95 %)	222
Figure 5.55	Ajustement des valeurs observées et des valeurs prévues A de la constante de formation K_A B de la constante de cassage K_E , modèle de régression neuronale (IC à 95%)	223
Figure 5.56	Importance relative des variables d'entrée, A sur la constante de formation K_A B sur la constante de cassage K_E , modèle de régression neuronale des constantes cinétiques.	223
Figure 5.57	R^2 en fonction du nombre de composantes principales dans la couche d'entrée, modèle de régression neuronale sur composantes principales des constantes cinétiques K_A et K_E	224
Figure 5.58	Architecture du perceptron multicouche 3/2/2 modèle de régression neuronale sur composantes principales des constantes cinétiques K_A et K_E	225

Figure 5.59	Ajustement à une loi Normale des résidus, modèle de régression neuronale sur composantes principales des constantes cinétiques [A] Constante de formation K_A , IC à 95 % [B] Constante de cassage K_E (IC à 95 %)	226
Figure 5.60	Ajustement des valeurs observées et des valeurs prévues [A] de la constante de formation K_A [B] de la constante de cassage K_E , modèle de régression neuronale sur composantes principales (IC à 95%)	226
Figure 5.61	Schéma de régression multiple des paramètres de finalité de traitement Alr et Tf	228
Figure 5.62	Présentation box plot « boîtes à moustache » des données d'apprentissage et de validation (test), Modélisation des paramètres Tf et Alr	229
Figure 5.63	R^2 en fonction du nombre de neurones dans la couche cachée, modèle de régression neuronale des paramètres Alr et Tf (set de validation)	230
Figure 5.64	Architecture du perceptron multicouche 14/17/2, modèle de régression neuronale des paramètres Tf et Alr	231
Figure 5.65	Ajustement à une loi Normale des résidus, modèle de régression neuronale des paramètres Tf et Alr [A] Turbidité finale Tf (IC à 95 %), [B] Aluminium résiduel Alr (IC à 95 %)	232
Figure 5.66	Vérification de l'homoscédasticité, [A] Turbidité finale Tf, [B] Aluminium résiduel Alr, modèle de régression neuronale des paramètres Tf et Alr (set de validation)	233
Figure 5.67	Ajustement des valeurs observées et des valeurs prévues de la turbidité finale Tf, modèle de régression neuronale des paramètres Tf et Alr [A] set d'apprentissage, [B] set de validation, IC à 95%	234
Figure 5.68	Ajustement des valeurs observées et des valeurs prévues de l'aluminium résiduel, modèle de régression neuronale des paramètres Tf et Alr [A] set d'apprentissage, [B] set de validation, IC à 95%	235
Figure 5.69	Importance relative des variables d'entrée, [A] sur la turbidité finale Tf [B] sur l'aluminium résiduel Alr , modèle de régression neuronale des paramètres Tf et Alr	235
Figure 5.70	R^2 en fonction du nombre de composantes principales dans la couche d'entrée, modèle de régression neuronale sur composantes principales des paramètres Tf et Alr	237
Figure 5.71	Architecture du perceptron multicouche 5/10/2, modèle de régression neuronale sur composantes principales des paramètres Tf et Alr	238
Figure 5.72	Ajustement à une loi Normale des résidus, modèle de régression neuronale sur composantes principales des paramètres Tf et Alr [A] Turbidité finale Tf (IC à 95 %), [B] Aluminium résiduel Alr (IC à 95 %)	239

Figure 5.73	Vérification de l'homoscédasticité, A Turbidité finale Tf, B Aluminium résiduel Alr, modèle de régression neuronale sur composantes principales des paramètres Tf et Alr (set de validation)	240
Figure 5.74	Ajustement des valeurs observées et des valeurs prévues de la turbidité finale Tf, modèle de régression neuronale sur composantes principales des paramètres Tf et Alr A set d'apprentissage, B set de validation, IC à 95%	240
Figure 5.75	Ajustement des valeurs observées et des valeurs prévues de l'aluminium résiduel, modèle de régression neuronale des paramètres Tf et Alr A set d'apprentissage, B set de validation, IC à 95%	241

LISTE DES TABLEAUX

Tableau 1.1	Modification de l'équation de Smoluchowski selon différents auteurs KELLIL (1989).	17
Tableau 1.2	Valeurs numériques des constantes cinétiques selon différents auteurs KELLIL (1989).	19
Tableau 1.3	Tendance de force et de résistance de floc JARVIS (2005).	23
Tableau 1.4	Paramètres d'entrée pour la modélisation BRODARD (1986)	43
Tableau 2.1	Choix faits concernant la construction du modèle de régression multiple.	83
Tableau 2.2	Choix faits concernant la construction de modèle de régression neuronale	97
Tableau 3.1	Modèles cinétiques d'Argaman Kauffman de floculation du kaolin, de la vase, de la boue et de la tourbe.	118
Tableau 3.2	Caractéristiques du kaolin utilisé et de la kaolinite utilisée par MIETTA (2009 b)	123
Tableau 3.3	Caractéristiques de la vase utilisée et de celle utilisée par MIETTA (2009 a)	123
Tableau 3.4	Modèles cinétiques de floculation des mélanges I et II.	127
Tableau 4.1	Statistique descriptive des variables.	137
Tableau 4.2	Résultats du calcul du coefficient de corrélation "R"	140
Tableau 4.3	Résultats de détermination des valeurs propres et des variances expliquées	145
Tableau 4.4	Composition du 1 ^{er} et du 2 ^{ème} groupe suite à l'analyse par ACP	150
Tableau 4.5	Composition du 3 ^{ème} , 4 ^{ème} et 5 ^{ème} groupe suite à l'analyse par ACP	151
Tableau 4.6	Répartition en groupes selon la méthode de Ward	156
Tableau 4.7	Paramètres utilisés pour la modélisation du taux de coagulant selon les auteurs [selon FRANCESCHI M. 1991]	159
Tableau 4.8	Paramètres utilisés pour la modélisation du taux de coagulant par régression multiple selon différents auteurs	159
Tableau 4.9	Paramètres utilisés pour la modélisation du taux de coagulant par régression neuronale selon différents auteurs	160
Tableau 4.10	Répartition des variables retenues et celles retirées à la base de l'analyse en composantes principales	161
Tableau 4.11	Résumé des distances d'agrégation et des variables à retirer à la base de la classification hiérarchique.	162
Tableau 4.12	Répartition des variables retenues à la base de la classification hiérarchique	162
Tableau 4.13	Résultats de détermination des valeurs propres et des variances expliquées après réduction des variables.	163
Tableau 4.14	Description des variables retenues pour la modélisation	165
Tableau 5.1	Paramètres d'entrée du 1 ^{er} modèle de régression multiple de la dose.	170

Tableau 5.2	Résultats de la régression, 1 ^{er} modèle de régression multiple de la dose.	171
Tableau 5.3	Synthèse des résidus de la régression de la dose, 1 ^{er} modèle de régression multiple de la dose [mgAl/l] (set d'apprentissage)	172
Tableau 5.4	Synthèse des résidus de la régression, 1 ^{er} modèle de régression multiple de la dose [mgAl/l] (set de validation)	174
Tableau 5.5	Paramètres d'entrée du 2 ^{ème} modèle de régression multiple de la dose.	177
Tableau 5.6	Résultats de la régression, 2 ^{ème} modèle de régression multiple de la dose.	177
Tableau 5.7	Synthèse des résidus de la régression, 2 ^{ème} modèle de régression multiple de la dose [mgAl/l] (set d'apprentissage)	178
Tableau 5.8	Synthèse des résidus de la régression, 2 ^{ème} modèle de régression multiple de la dose [mgAl/l] (set de validation)	179
Tableau 5.9	Paramètres d'entrée du 3 ^{ème} modèle de régression multiple de la dose.	182
Tableau 5.10	Résultats de la régression, 3 ^{ème} modèle de régression multiple de la dose.	182
Tableau 5.11	Synthèse des résidus de la régression, 3 ^{ème} modèle de régression multiple de la dose [mgAl/l] (set d'apprentissage)	183
Tableau 5.12	Synthèse des résidus de la régression, 3 ^{ème} modèle de régression multiple de la dose, (set de validation)	184
Tableau 5.13	Résultats de détermination des valeurs propres et des variances expliquées, modèle de régression régularisée de la dose.	188
Tableau 5.14	Résultats de la régression régularisée de la dose.	190
Tableau 5.15	Synthèse des résidus de la régression régularisée de la dose [mgAl/l] (set d'apprentissage)	191
Tableau 5.16	Synthèse des résidus de la régression régularisée de la dose [mgAl/l](set de validation)	192
Tableau 5.17	Influence des variables explicatives sur l'équation de la régression, modèle de régression régularisée de la dose.	194
Tableau 5.18	Paramètres d'entrée du modèle de régression neuronale de la dose	195
Tableau 5.19	Résultats de la régression neuronale de la dose [mgAl/l]	199
Tableau 5.20	Synthèse des résidus de la régression neuronale de la dose [mgAl/l] (set d'apprentissage)	200
Tableau 5.21	Synthèse des résidus de la régression neuronale de la dose [mgAl/l] (set de validation)	201
Tableau 5.22	Modèles neuronaux sur Composantes Principales, pour la prédiction de la dose.	204
Tableau 5.23	Synthèse des résidus de la régression neuronale sur composantes principales de la dose [mgAl/l] (set d'apprentissage)	205
Tableau 5.24	Synthèse des résidus de la régression neuronale sur composantes principales de la dose [mgAl/l] (set de validation)	205

Tableau 5.25	Paramètres retenues pour la modélisation des constantes de formation et de cassage de floc	209
Tableau 5.26	Paramètres d'entrée du modèle de régression de la constante de cassage K_E	210
Tableau 5.27	Résultats de la régression multiple de la constante de cassage K_E .	210
Tableau 5.28	Synthèse des résidus de la régression de la constante de cassage K_E [s]	211
Tableau 5.29	Résultats de détermination des valeurs propres et des variances expliquées, modèle de régression des constantes cinétiques	213
Tableau 5.30	Résultats de la régression régularisée de la constante de cassage K_E .	214
Tableau 5.31	Synthèse des résidus de la régression régularisée de la constante de cassage K_E [s]	214
Tableau 5.32	Influence des variables explicatives sur l'équation de la régression, modèle de régression régularisée de la constante K_E .	215
Tableau 5.33	Paramètres d'entrée du modèle de régression multiple de la constante de formation K_A	216
Tableau 5.34	Synthèse des résidus de la régression multiple de la constante de formation K_A [-]	217
Tableau 5.35	Résultats de la régression régularisée de la constante de formation K_A	218
Tableau 5.36	Synthèse des résidus de la régression régularisée de la constante de formation K_A [-]	219
Tableau 5.37	Influence des variables explicatives sur l'équation de la régression, modèle de régression régularisée de la constante K_A	220
Tableau 5.38	Paramètres d'entrée du modèle de régression neuronale des constantes cinétiques K_A et K_E	220
Tableau 5.39	Résultats de la régression neuronale des constantes cinétiques K_A et K_E	221
Tableau 5.40	Synthèse des résidus de la régression neuronale des constantes cinétiques K_A et K_E	222
Tableau 5.41	Résultats de la régression neuronale basée sur les composantes principales des constantes cinétiques K_A et K_E	224
Tableau 5.42	Synthèse des résidus de la régression neuronale basée sur les composantes principales des constantes cinétiques K_A et K_E	225
Tableau 5.43	Paramètres d'entrée du modèle de régression neuronale des paramètres Alr et Tf	228
Tableau 5.44	Résultats de la régression neuronale des paramètres Alr et Tf	230
Tableau 5.45	Synthèse des résidus de la régression neuronale des paramètres Alr et Tf (set d'apprentissage)	232
Tableau 5.46	Synthèse des résidus de la régression neuronale des paramètres Alr et Tf (set de validation)	233

Tableau 5.47	Résultats de détermination des valeurs propres et des variances expliquées, modèle de régression neuronale des paramètres Tf et Alr	236
Tableau 5.48	Résultats de la régression neuronale sur composantes principales des paramètres Alr et Tf	237
Tableau 5.49	Synthèse des résidus de la régression neuronale basée sur les composantes principales des paramètres Alr et Tf (set d'apprentissage)	238
Tableau 5.50	Synthèse des résidus de la régression neuronale basée sur les composantes principales des paramètres Alr et Tf (set de validation)	239

PRINCIPALES NOTATIONS

Lettre	Signification
ψ	Potentiel [mV]
ψ_0	Potentiel de surface [mV]
ψ_S	Potentiel de STERN [mV]
δ	Epaisseur de la double couche ionique [μm]
W_A	Energie d'attraction de Van Der Waals [J]
W_R	Energie de répulsion électrique [J]
G	Valeur moyenne du gradient de vitesse [s^{-1}]
GT	Critère de Camp
K_A	Constante cinétique de formation [-]
K_B	Constante cinétique cassage [s]
ζ	Potentiel zêta [mV]
Dose	Dose du coagulant injecté [mgAl/l]
D_c	Dose consommée [mgAl/l]
Pui	Puissance d'agitation lente /unité de volume [Wm^{-3}]
Pr	Puissance d'agitation rapide /unité de volume [Wm^{-3}]
C_p	Concentration en particules solide [mg/l]
TO	Turbidité initiale [NTU]
T_{eli}	Turbidité éliminée [NTU]
T_f	Turbidité finale [NTU]
Al_r	Aluminium résiduel [mg Al/l]
R_T	Rendement d'élimination de la turbidité [%]
R_{Al}	Taux de consommation de coagulant [%]
CM	Carbone minérale [%]
COT	Carbone organique total [%]
MO	Matière organique [%]
D_{ms}	Diamètre moyen en surface
D_{mv}	Diamètre moyen en volume
$D(0,9)$:	Taille pour laquelle 90% de l'échantillon se trouve en dessous de cette dimension [μm]
$D(0,5)$:	Taille pour laquelle 50% de l'échantillon se trouve en dessous de cette dimension [μm]
$D(0,1)$:	Taille pour laquelle 10% de l'échantillon se trouve en dessous de cette dimension [μm]
Span	Largeur de la distribution $span = \frac{D(0,9) - D(0,1)}{D(0,5)}$
S_{sp}	Surface spécifique [m^2/g]
Unif	Uniformité : écart absolu à partir de la médiane $D(0,5)$
Arg	Taux d'argile dans le solide [%]
Kao	Taux de kaolinite dans le solide [%]
illi	Taux d'Illite dans le solide [%]
Mte	Taux de montmorillonite dans le solide [%]

Acronyme	Signification
S.A.	Sulfate d'aluminium
PAC	Polychlorure d'aluminium
PFC	Polychlorure de fer
PAHS	Poly-hydroxy-sulfate d'aluminium
HPAC	Polychlorure d'aluminium composite
PASiC	Polychloro-silicate d'aluminium
PAFSiC	Polychloro-silicate de fer et d'aluminium
WAC	Polychlorosulfate basique d'aluminium
THMFP	Trihalométanes formation potentiel
THM	Trihalométanes
C.O.D.	Carbone organique dissous
CAG	Charbon Activé en Granulats
TAC	Titre alcalimétrique complet
TH	Titre hydrotimétrique
Alc	Alcalinité
CEC	Capacité d'Echange Cationique
DBP	Disinfection by-Products
M.O.N.	Matière organique naturelle
ACP	Analyse en Composantes Principales,
AF	Analyse Factorielle
RNA	Réseaux de neurones artificiels
PMC	Perceptron multicouche
REQM	Racine carrée de l'Erreur Quadratique Moyenne

INTRODUCTION GENERALE

L'eau est la plus précieuse ressource de notre planète et la principale composante de notre corps, elle est à l'origine de la vie. Une eau de surface transporte toutes sortes de substances inertes, vivantes, stables et complexes dont certaines peuvent être nocives pour l'organisme humain. Ces substances proviennent soit du milieu physique dans lequel l'eau a évolué, soit des rejets de certaines activités humaines auxquelles l'eau est devenue le réceptacle. L'eau est ainsi le vecteur de transmission privilégié de nombreuses maladies, pour qu'elle soit consommée sans danger, cette eau doit donc être traitée.

Aujourd'hui, l'eau potable représente l'un des produits alimentaires les plus surveillés dans le monde, faisant du traitement de l'eau une industrie de pointe. Cette industrie est sous une pression croissante pour produire une eau potable d'une plus grande qualité au plus faible coût, la qualité de l'eau traitée est garantie par l'optimisation, le contrôle et la surveillance permanente du procédé de traitement.

L'unité de coagulation/floculation est une étape clé dans la chaîne de traitement des eaux de surface, elle sert à déstabiliser et modifier les propriétés des particules colloïdales et l'agglomération des fines particules en floes de plus grandes tailles, ce qui facilite leur séparation. L'optimisation de cette unité est essentielle pour atteindre deux objectifs : la maîtrise de la qualité de l'eau traitée et le contrôle du coagulant résiduel.

La dose de coagulant à injecter est la variable principale utilisée pour conduire cette unité, la recherche de cette dose repose généralement sur l'essai "Jar Test", cette technique est lourde à mettre en œuvre et souffre d'un temps de retard relativement long, en effet, on ne modifie la dose de coagulant qu'une fois l'événement apparu.

A cet effet, plusieurs tentatives de modélisation, tant mathématique que statistique, des paramètres de floculation et en particulier la dose optimale ont

été élaborées. Cependant, la complexité du phénomène fait qu'il est difficile de trouver un modèle simple et générale permettant de décrire le comportement de la floculation.

Le présent travail s'inscrit dans ce courant de recherche, il s'agit de la modélisation de l'efficacité de la floculation des suspensions aqueuses en tenant compte des caractéristiques de la phase solide et des paramètres opératoires. Il porte sur deux volets :

- Le premier est expérimental, son objectif est double : d'une part la création d'une base de données expérimentales relative à la floculation de six types de particules solides dispersées et d'autre part, la contribution à la compréhension des effets des caractéristiques de la phase solide et des paramètres opératoires sur le mécanisme, la cinétique et la performance de floculation. En effet, de nombreuses études ont amplement discuté ce sujet, néanmoins, la question d'interactions entre les caractéristiques du solide dispersé et les paramètres de floculation semble être moins étudiée, nous accentuons notre recherche expérimentale sur l'incidence de l'interaction entre la matière minérale, la matière organique et les conditions d'agitation sur le comportement de la floculation.

- Le deuxième vise la formulation de modèles mathématiques statistiques permettant la prédiction de la performance de la floculation. Cette modélisation se fait généralement suivant deux approches :

- La modélisation du procédé en utilisant des paramètres descripteurs de la qualité de l'eau brute et des conditions opératoires.
- La régulation en aval basée sur la qualité de l'eau traitée après l'étape de floculation.

Les diverses études de modélisation mathématique statistique de la phase de floculation, montre l'utilisation large des caractéristiques du solide dispersé, des caractéristiques physico-chimiques du milieu aqueux et des conditions hydrodynamiques en tant que paramètres descripteurs des modèles de régression, cependant il s'avère que les constantes cinétiques de formation K_A et de cassage K_B semblent être occultées et non intégrées comme variables explicatives, en outre, à notre connaissance, ces deux constantes n'ont pas fait d'objet de travaux de modélisation statistique empirique.

A cet effet l'innovation dans ce travail réside principalement dans :

- L'intégration des constantes cinétiques de floculation ; la constante de formation K_A et la constante de cassage K_B dans les modèles de régression de la dose optimale et des paramètres de finalité de traitement.

- La modélisation de l'efficacité de la floculation en suivant une troisième approche ; par régression statistique des constantes cinétiques de formation et de cassage, en effet la connaissance de ces deux constantes permet, par

application des modèles cinétiques de floculation, la détermination de l'efficacité de floculation.

Dans la littérature, les approches de modélisation de la floculation par des modèles classiques de régression multiple linéaire semblent présenter des performances non élevées, cela revient principalement à la grande non linéarité du procédé, et à la complexité et la multicollinéarité des données.

Les recherches conduites dans le domaine montrent qu'une amélioration considérable de performance de prédiction peut être réalisée en employant des modèles non linéaires à base de réseau de neurones artificiels, et aussi en combinant les méthodes de régression et les techniques d'analyse multivariées

A cet égard dans cette thèse, l'emphase est mise sur :

- L'utilisation des techniques d'analyse multivariée, en particulier, l'Analyse en Composantes Principales et l'Analyse de Classification Hiérarchique dans la préparation des données et le choix des variables pertinentes à intégrer dans les modèles de régression.

- L'identification des approches statistiques les plus prometteuses pouvant être adaptées à la modélisation de l'efficacité de la floculation par l'établissement des modèles de régression multiple, de régression régularisée, de régression neuronale et de régression neuronale basée sur les composantes principales.

Ce manuscrit est structuré en cinq chapitres :

Le premier chapitre est consacré à une revue bibliographique sur les propriétés des suspensions colloïdales, leur stabilité et leur coagulation/floculation, suivi d'une synthèse des résultats pratiques d'étude de la formation et de cassage de floc et des effets des différents paramètres tel que le pH, la température, la nature et la charge des colloïdes, la minéralisation et les conditions d'agitation sur le procédé. Au cours de ce chapitre, nous présentons un aperçu sur les différents appareils et méthodes de détermination de la dose de coagulant et une description de l'utilisation des techniques multivariées dans l'analyse des données relatives à la floculation. La dernière partie de ce chapitre concerne la synthèse bibliographique des travaux de recherche se rapportant à la modélisation par régression multiple et par régression neuronale des paramètres de floculation.

Le deuxième chapitre consiste à un survol du fondement théorique de l'analyse multidimensionnelle ; principalement l'Analyse en Composantes Principales qui permet, suivant un calcul formalisé, l'obtention de représentation compacte des données, et l'Analyse de Classification Hiérarchique qui sert, suivant une démarche algorithmique, à produire des groupements de variables ou individus.

Ce chapitre aborde aussi les principes de la modélisation statistique paramétrique par régression multiple et non paramétrique par réseaux de neurones, ces derniers s'adaptent bien à la non linéarité de processus, Ils ont de ce fait connu une popularité grandissante. Les modèles combinant la régression multiple et l'analyse des composantes principales et les modèles de régression neuronale basée sur les composantes principales feront l'objet d'une section dans ce chapitre. Ces modèles permettent de réduire la colinéarité et la complexité des modèles.

Le procédé de coagulation et de floculation des suspensions synthétiques est complexe, fortement non linéaire et son fonctionnement reste difficilement modélisable. Dans le but de la modélisation de leur floculation, Nous sommes contraints de créer une base de données et d'étudier l'effet des caractéristiques de ces suspensions et des paramètres opératoires sur le procédé. A cet effet, le troisième chapitre est consacré à l'étude expérimentale de l'effet des caractéristiques du solide dispersé et des conditions opératoires et de leurs interactions sur la performance, la cinétique et le mécanisme de floculation. Il est étalé sur deux points :

- Caractérisation des suspensions et présentation du protocole expérimental de floculation,
- Discussion et interprétation des résultats expérimentaux sur la base des nouvelles recherches, dans le domaine, qui ne cessent d'être de plus en plus importantes et pointues.

Le quatrième chapitre vise à établir l'analyse exploratoire des données expérimentales, étape capitale avant toute étude de modélisation, elle a pour but l'analyse, la préparation des données et le choix des variables explicatives. Cette analyse exploratoire comportera trois étapes ;

1. L'analyse monovariée qui consiste à détecter les valeurs atypiques, et identifier les valeurs manquantes, erronées, aberrantes ou extrêmes.
2. L'analyse bivariée qui sert à identifier les incohérences et liaisons entre les variables prises deux à deux par l'étude des corrélations, par comparaison des dispersions et distributions et par l'observation des nuages de points.
3. L'analyse multidimensionnelle par Analyse en Composantes Principales et par Analyse de Classification Hiérarchique, qui permettent d'identifier les liaisons entre les paramètres prises simultanément et de synthétiser l'information des jeux de données de grandes dimensions en vue d'obtenir des représentations compactes.

Le cinquième chapitre a pour objectif l'établissement des modèles de prédiction de la dose optimale, des constantes cinétiques de formation K_A et de cassage K_B , et des paramètres de finalité de traitement ; l'aluminium résiduel "ALr" et la turbidité finale "Tf".

Ce chapitre est consacré à la régression multiple et aux améliorations apportées par la régression multiple basée sur les composantes principale, communément appelée la régression régularisée, cette régression est considérée comme un modèle combiné entre la régression multiple et l'analyse des composantes principales, elle est utilisée lorsque les variables explicatives sont nombreuses ou fortement corrélées entre elles. La recherche de modèle de régression multiple est présentée étape par étape ; le choix des variables et préparation des données, l'élaboration de modèle sur données d'apprentissage, les tests de signification, l'analyse des résidus, et à la fin la validation du modèle sur un deuxième échantillon selon différents critères d'estimation et de prédiction.

Le chapitre aborde aussi la régression non linéaire par réseaux de neurones, ces réseaux sont utilisés avec succès dans la modélisation des paramètres de floculation, ils offrent l'avantage d'être un outil non linéaire, multivariable, paramétrable par apprentissage sur des données expérimentales. Ce chapitre présente aussi les résultats de la régression neuronale basée sur les composantes principales, à la différence du modèle neuronal simple, les entrées du modèle neuronal basé sur les composantes principales sont les composantes principales des variables initiales.

Ce type de modélisation permet de réduire la colinéarité des données et la complexité du modèle et améliore la performance de prédiction. La mise en œuvre de la régression neuronale consiste à la collecte et la préparation des données, l'apprentissage du modèle, la détermination de l'architecture optimale et en fin l'évaluation et le test du modèle sur un autre échantillon.

Enfin, nous synthétisons dans la conclusion générale les principaux résultats obtenus au cours de ce travail, tout en précisant les avantages et les limites des approches établies, les perspectives de travail et les axes de recherche restant à développer.

Chapitre 1

COAGULATION / FLOCCULATION DES SUSPENSIONS COLLOÏDALES : THEORIE, PRATIQUE ET MODELISATION

I. SYSTEMES COLLOIDAUX

I-1. DEFINITIONS GENERALES

Les particules dont le diamètre varie entre 1 à 1000 nanomètres sont considérées comme étant colloïdales ; en pratique, plusieurs auteurs considèrent les colloïdes dans l'eau, toutes celles dont la taille est inférieure à 10 μ m TAMBO (1991).

Ces particules colloïdales peuvent être constituées d'agrégats de molécules (micelles), de molécules géantes (macromolécules) ou de particules solides argileuses ou biologiques (bactéries, virus) HENRY (1992)

On classe généralement, les colloïdes en deux catégories suivant leur comportement vis à vis du milieu aqueux : les hydrophiles et les hydrophobes.

- Colloïdes hydrophiles

Ils sont, en général, à l'origine de la couleur des eaux de surface, proviennent de la dégradation de l'humus et des substances organiques diverses. Ces colloïdes sont entourés par une enveloppe de molécules d'eau par des liaisons physico-chimiques, avec des groupements -OH⁻, -NH₂, et -COOH⁻, totalement ou partiellement ionisés. Cette enveloppe empêche tout contact ultérieur entre les particules DESJARDINS (1988).

Ces colloïdes sont, donc, stables et cette stabilité est influencée par le pH de la solution, et par la nature, la solvation et le degré d'ionisation des groupements fonctionnels SAWER (1994).

La majorité des colloïdes hydrophiles servent comme couche protectrice des colloïdes hydrophobes et les empêchent alors de réagir avec les électrolytes SAWER (1994).

- Colloïdes hydrophobes

Comme leur nom l'indique, ils n'ont aucune affinité pour l'eau, ce sont en général des particules inorganiques (silice, argile). Ils sont chargés électriquement ; le signe et la valeur de leurs charges dépendent des caractéristiques du colloïde, du pH et des propriétés ioniques de l'eau SAWER (1994). Aussi, ces particules colloïdales hydrophobes, lorsqu'elles sont déshydratées ne se dispersent pas spontanément, contrairement aux particules colloïdales hydrophiles qui réagissent spontanément avec l'eau pour former une suspension qui peut être hydratée et déshydratée à plusieurs reprises. En fait, peu de particules sont exclusivement hydrophobes ou hydrophiles. On trouve plutôt des particules hydratées à différents degrés MASSCHELEIN (1996).

Origines des charges des colloïdes

Une des propriétés de ces colloïdes c'est qu'elles portent généralement des charges électriques, si bien qu'elles présentent certaines propriétés analogues à celle des ions. Leur petite taille leur confère une surface spécifique élevée ce qui rend les phénomènes physico-chimiques à l'interface

solide - liquide complexes, en particulier, la modification de l'environnement ionique autour de la particule.

- L'origine la plus importante de la charge électrique des particules colloïdales est l'adsorption et selon, qu'elles adsorbent plutôt les anions ou cations, elles se chargent négativement ou positivement HENRY (1992).
- Les particules colloïdales peuvent contenir des groupements ionisables ($-NH_2$, $-COOH^-$, $-SO_4H^-$), qui peuvent être répartis régulièrement dans la macromolécule BECHAC (1983).
- Remplacement d'éléments de la maille cristalline par des éléments de charge différente peut aussi être à l'origine des charges des colloïdes MASSCHELEIN (1996).
- MIETTA (2009b) indique que la charge des particules de la vase ou des particules d'argile n'est pas uniforme et dépend des propriétés des suspensions et du pH ; à pH 4 les arêtes du kaolinite et de la montmorillonite sont positivement chargées ce qui permet une rapide agrégation et la formation de grands floccs contrairement au pH 8 où les arêtes et les faces sont négativement chargées.

1-2. PROPRIETES DES SUSPENSIONS COLLOIDALES

Optiques

Parce que les particules colloïdales ont des dimensions proches de la longueur d'onde de la lumière blanche, elles interfèrent avec le passage de la lumière ; cette dernière se reflète. Ce phénomène est appelé l'effet Tyndall SAWER (1994). D'ailleurs, c'est sur cet effet que repose le principe de la mesure de la turbidité pour évaluer la charge en particules d'une eau.

Cinétiques

• Mouvement brownien

Les particules colloïdales sont bombardées par des molécules du milieu de dispersion et à cause de leurs faibles masses, elles se meuvent sous l'effet de ces bombardements d'une manière aléatoire. Ce phénomène est d'autant plus intense que la température de milieu est élevée.

• Vitesse de sédimentation

Les particules colloïdales ne sont pas sensibles à l'action de la pesanteur, elles ne sont pas sédimentables, et cela même, dans les centrifugeuses ordinaires, il faut utiliser des ultracentrifugeuses.

Electriques

• Adsorption

A cause de leur petite taille qui leur confère une surface spécifique élevée, et les charges électriques qu'ils portent, les colloïdes se caractérisent aussi par

leur pouvoir adsorbant important. Cette adsorption est le résultat des attractions électrostatiques qui s'exercent entre les diverses particules et les espèces chimiques en solution. Parmi les facteurs qui influencent l'adsorption : la nature et le rapport surface / masse de l'adsorbant, la température, la pression et la concentration HENRY (1992).

- Électrophorèse

Etant chargées électriquement, les particules colloïdales dispersées dans un milieu liquide se déplacent dans un champ électrique. Si on place, par exemple, deux électrodes reliées à une source de tension électrique, on observe une migration des particules : c'est l'électrophorèse, dont le principe est le suivant :

Les particules sont soumises d'une part à une force d'attraction électrique et d'autre part à une force de freinage due à la viscosité du milieu et du fait de leur charge effective, elles atteignent une vitesse limite. Par ailleurs, les suspensions colloïdales peuvent être purifiées par simple dialyse ou électrodialyse. Leurs tailles permettent leur rétention sur des membranes ordinaires semi-perméables SAWER (1994).

I-3. STABILITE DES SUSPENSIONS COLLOÏDALES

Dans une eau, les particules colloïdales acquièrent une charge électrique effective différente de leur charge réelle HENRY (1992) à cause des ions voisins de signe opposé qui adhèrent plus ou moins fortement à ces particules. Ces ions forment une couche autour des colloïdes, ce qui empêche toute agglomération. Diverses théories ont été proposées pour expliquer ce phénomène FIESSINGER (1976).

- Théorie d'HELMHOLTZ

Elle est basée sur l'hypothèse de l'existence d'une couche d'ions antagonistes fortement liée aux charges portées par les particules ; cette couche recouvre intégralement la surface et assure la neutralité de l'ensemble, le potentiel électrique décroît linéairement à partir de la surface de la particule.

$$\psi = \psi_0 - A x \quad (I-1)$$

ψ : potentiel à une distance x de la surface [mV]

ψ_0 : potentiel à la surface [mV]

x : distance à partir de la surface de la particule [μm]

A : constante qui dépend de la force ionique du milieu et de la température

Toutefois, cette théorie est peu réaliste du fait que les ions sont en mouvement dans l'eau et par conséquent, ils ne peuvent rester fixés continuellement à la particule à cause du mouvement brownien ; il y a un échange d'ions entre la surface de particule et le reste de la suspension, ce qui est à l'origine du développement de la deuxième théorie.

- Théorie de GOUY et CHAPMAN

Basée sur l'idée qu'entre la surface de la particule et le sein de la suspension existe un libre échange d'ions créant une décroissance exponentielle du potentiel électrique à partir de la surface de la particule

$$\psi = \psi_0 e^{-kx} \quad (I-2)$$

ψ : potentiel à une distance x de la surface [mV]

ψ_0 : potentiel à la surface de la particule [mV]

x : distance à partir de la surface de la particule [μm]

k : constante qui dépend de la force ionique du milieu et de température

La couche d'ions positifs est inégalement répartie autour du colloïde ; la neutralité est obtenue à plus grande distance (couche diffuse) ; cette couche est déformable et mobile EDELINE (1996).

- Théorie de STERN

L'originalité de cette théorie réside dans la synthèse des deux précédentes et considère la formation d'une double couche. Elle serait la plus représentative de la réalité ionique autour de la particule. La première couche, appelée "couche liée ou couche de STERN" est fixée au colloïde, d'épaisseur variable en fonction de la quantité et de la nature d'ions absorbés; elle se déplace avec la particule, abaisse sa charge réelle et forme autour d'elle une gaine, un écran protecteur. Dans la "couche liée" le potentiel décroît rapidement de façon linéaire à partir de la surface du colloïde.

$$\psi = \psi_0 - A x \quad (I-3)$$

ψ : potentiel à une distance x de la surface [mV]

ψ_0 : potentiel à la surface [mV]

x : distance à partir de la surface de la particule [μm]

A : constante qui dépend de la force ionique du milieu et de la température

La seconde couche plus étalée, est appelée "couche diffuse ou couche de GOUY", elle freine le mouvement électrophorétique HENRY (1992). Cette couche n'adhère pas à la particule, mais forme un nuage ionique autour d'elle, celui-ci s'estompe et devient nul à une certaine distance de la particule, dans la couche diffuse le potentiel décroît exponentiellement avec la distance selon la loi

$$\psi = \psi_s e^{-kx} \quad (I-4)$$

ψ : potentiel à une distance x dans la couche diffuse [mV]

ψ_s : potentiel de STERN [mV]

x : distance à partir de la surface de la couche de STERN [μm]

k : constante qui dépend de la force ionique du milieu et de la température

Bilan de forces, théorie DELVO (Deryagin-Landau-Verwey-Overbeek)

La stabilité d'une suspension colloïdale dépend du bilan des forces d'attraction et de répulsion SAWER (1994).

- Les forces d'attraction de Van Der Walls qui correspondent aux forces des molécules de chaque colloïde ; ces forces, dont l'intensité est inhérente à la nature du colloïde, ne sont importantes qu'à très courtes distance (de 5 à 10 nm).
- Les forces de répulsion électrique qui sont significatives lorsque deux colloïdes sont proches ; leurs doubles couches de même charge superficielle se repoussent.

L'énergie totale d'interaction W_T tient compte de l'énergie d'attraction de Van Der Walls " W_A " et de l'énergie de répulsion électrique " W_R " et est traduite par l'équation ci - après :

$$W_T = W_A + W_R$$
$$W_T = \frac{-A \cdot r}{12 \cdot d} + \frac{\epsilon r^2 \zeta^2 e^{-\left(\frac{d}{\delta}\right)}}{2r + d} \quad (I-5)$$

A : constant de Hamaker $A=10^{-21}$ [j], $[ML^2T^{-2}]$.

d : distance entre deux particules [L]

ϵ : constante diélectrique $[M^{-1}L^{-3}T^4 A^2]$.

r : rayon de deux particules supposées identiques [L]

δ : épaisseur de la double couche ionique [L]

ζ : potentiel zéta [mV], $[ML^2T^{-3}A^{-1}]$ (Voir définition § III.1.3 p. 39)

L'énergie d'attraction W_A est liée au forces de Van Der Walls, elle ne dépend pas du milieu, alors que l'énergie de répulsion W_R varie en fonction de la force ionique du milieu de dispersion.

I-4. COAGULATION DES SUSPENSIONS COLLOIDALES

La coagulation est la première étape de déstabilisation des suspensions colloïdales ; elle résulte de l'addition de réactif chimique à cations polyvalents (coagulant) dans les dispersions aqueuses afin de permettre l'agrégation des particules dispersées. Plusieurs mécanismes interviennent simultanément lors du processus de coagulation ; parmi ces mécanismes, on citera :

I-4.1. Compression de la double couche

Les particules colloïdales présentes dans l'eau, possèdent des charges électriques réparties sur leur surface et, sont soumises à des forces de répulsion électriques et à des forces d'attraction de Van Der Walls.

On peut agglomérer ces particules en réduisant le potentiel de répulsion entre les particules par l'augmentation de la force ionique de la solution,

ceci grâce à l'ajout d'ions supplémentaires tant positifs que négatifs et de préférence doublement ou triplement chargés de façon à contribuer fortement à l'augmentation de la force ionique globale, selon la règle de Schulzer Hardy SAWER (1994), ainsi on comprime l'épaisseur de la couche diffuse vers la particule.

Le rapport entre les efficacités des ions mono, di et trivalents pour la même suspension est 1 : 8 : 600 MASSCHELEIN (1996).

GÜVEN, (1992) note que pour une même efficacité de coagulation, les concentrations $[C]$ des ions mono, di et trivalent varient selon le potentiel à la surface de la particule :

- Faible potentiel $[C^{+1}] : [C^{+2}] : [C^{+3}] \cong \frac{1}{1^2} : \frac{1}{2^2} : \frac{1}{3^2}$
- Potentiel élevé $[C^{+1}] : [C^{+2}] : [C^{+3}] \cong \frac{1}{1^6} : \frac{1}{2^6} : \frac{1}{3^6}$

Il précise aussi, qu'en cas d'ions de même valence l'efficacité de floculation dépend de la taille et de la polarisation des ions.

Ce mode de coagulation est caractérisé par l'absence de spécificité, toutes les couches diffuses sont contractées et, la dose du coagulant n'obéit à aucune relation stœchiométrique.

I-4.2. Adsorption et neutralisation de charges par ions métalliques polyvalents

Les particules colloïdales peuvent être déstabilisées par adsorption (ou dépôt) à leur surface, des ions provenant du colloïde lui-même ou de ces produits de réactions avec l'eau. Il peut en résulter ainsi une neutralisation de la charge colloïdale et de la précipitation du colloïde.

Du fait de la non spécificité de l'adsorption, plus de charges que possible peuvent être adsorbées afin de neutraliser la charge de surface, mais dans ce cas, la dose active du coagulant est étroite et, elle obéit à une relation stœchiométrique, tout excès du réactif mène à la restabilisation du colloïde par inversion du signe de la charge électrique.

I-4.3. Adsorption de macromolécules et pontage

L'utilisation des polymères synthétiques (polyélectrolytes) comme agents déstabilisants a montré que l'on obtient des résultats encourageants ; ce phénomène, parfois inexplicable, par la théorie de double couche, peut être expliqué par l'adsorption et le pontage. Cette coagulation peut être obtenue même si le potentiel zéta n'est pas amené précisément à la valeur zéro, ceci s'explique par le modèle de pontage développé par Stumm et O'mellia (1968). Le phénomène de restabilisation peut survenir lorsqu'il n'y a plus de sites actifs disponibles pour la formation des "ponts" polymère - particule, soit par manque local des colloïdes, soit par surdosage du coagulant.

I-4.4. Emprisonnement des particules dans un précipité

La déstabilisation des particules colloïdales se fait aussi en les emprisonnant dans une particule de floc en formation constituée de molécules $Al(OH)_3$ ou, $Fe(OH)_3$; Les particules colloïdales jouent un rôle de noyaux DESJARDINS (1988), moins il y a de particules colloïdales en suspension à floculer, plus la dose doit être forte ; c'est la raison pour laquelle on renforce parfois artificiellement la suspension par des colloïdes à grande surface spécifique appelés adjuvants (bentonite, silice activé).

I-4.5. Aspect cinétique de la coagulation

La déstabilisation des particules est un processus relativement complexe impliquant plusieurs phénomènes qui évoluent dans le temps. En 1968 Stumm et O'mellia ont proposé plusieurs étapes de réactions :

1. Hydrolyse de l'ion métallique multivalent et polymérisation subséquente en des espèces polynucléaires hydrolysées.
2. Adsorption des espèces hydrolysées sur l'interface solide - liquide pour déstabiliser le colloïde.
3. Agrégation des particules déstabilisées.
4. Vieillessement des floccs formés et qui s'accompagne par des changements chimiques dans la structure.
5. Précipitation de l'hydroxyde métallique.

I-5. FLOCCULATION DES SUSPENSIONS COLLOÏDALES

La floculation est l'agrégation des particules, déjà déstabilisées, grâce à leurs collisions mutuelles, résultant de leur déplacement relatif. Ces collisions conduisent à la croissance de la taille des particules flocculées et à la diminution du nombre des particules primaires en suspension. Il s'agit donc, d'augmenter la probabilité des contacts entre les particules, celle-ci peut être provoquée par divers mécanismes THOMAS (1999) ;

- Mouvement brownien (floculation péricinétique) : ce mécanisme est prédominant dans le cas des particules de petite taille ($<10 \mu m$)
- Ecoulement du fluide qui donne naissance à un gradient de vitesse (floculation orthocinétique), cette floculation concerne les particules de taille plus importante.
- Mouvement relatif des particules, dû à la différence de leurs masses volumiques respectives.

I-5.1. Floculation péricinétique

La floculation péricinétique est un phénomène extrêmement lent ARFANDI (1979), GREGORY (1989); il favorise l'agrégation des particules par le processus aléatoire naturel de l'agitation thermique. La vitesse de floculation

définie comme la variation du nombre de particules primaires par unité de temps est donnée par :

$$-\frac{dn}{dt} = 4\alpha K_b T n^2 / 3\mu \quad (I-6)$$

n : nombre de particules par unité de volume [m^{-3}]

α : fraction du nombre de chocs efficaces productifs de floes ($\alpha < 1$)

K_b : constante de Boltzman ($K_b = 1,380622 \cdot 10^{-23} J \cdot K^{-1}$)

μ : viscosité dynamique de suspension [$Kgm^{-1}s^{-1}$]

T : température absolue [$^{\circ}K$]

Posant $k_p = \frac{4\alpha K_b T}{3\mu}$ L'équation (I-6) devient : $\frac{dn}{dt} = -k_p n^2$

Donc, (après intégration) $n_t = \frac{n_0}{1 + n_0 k_p t}$ (I-7)

n_0 : nombre de particules par unité de volume à l'instant $t=t_0$

Cette équation montre que le nombre de particules n_t qui restent à l'instant t dépend de n_0 , de la température, de la viscosité du milieu, et du coefficient des collisions efficaces α , cependant il est indépendant de la taille des particules colloïdales. La disparition des particules primaires par ce processus est extrêmement lente et qui s'atténue rapidement dès que la taille devient supérieure à $10\mu m$. Pour accélérer l'agglomération des particules déstabilisées il faut procéder par floculation orthocinétique.

On note que plusieurs chercheurs ont tenté d'améliorer l'équation Smoluchowski, à titre d'exemple FUKASAWA (2006) a étudié l'effet de la structure de floc sur la floculation brownienne et propose ce qui suit :

$$-\frac{dn}{dt} = \frac{4\alpha K_b T n^2}{3\mu} \left(\frac{1}{1 - V_0} \right)^2 \quad (I-8)$$

Avec : V_0 : somme des volumes de floes à l'instant t

I-5.2. Floculation orthocinétique

La floculation orthocinétique est provoquée par le mouvement de l'eau donnant naissance à des gradients locaux de vitesse au sein du liquide. Ces gradients facilitent l'agrégation des particules en augmentant la probabilité des collisions entre ces particules en contact permanent.

En 1917, Smoluchowski a établi un modèle mathématique donnant le nombre de contacts H_{ij} entre toutes les particules i et les particules j par unité de temps et par unité de volume, dans un écoulement laminaire FAIR (1964), IVES (1981).

$$H_{ij} = \frac{4}{3} n_i n_j R_{ij}^3 \frac{du}{dz} \quad (I-9)$$

n_i : est le nombre de particules i par unité de volume [m^{-3}]

n_j : est le nombre de particules j par unité de volume [m^{-3}]

R_{ij} : est la somme de deux rayons R_i et R_j des particules i et j [m]

$\frac{du}{dz}$: est le gradient de vitesse constant dans la direction z [s^{-1}]

L'application de cette équation aux cas réels est évidemment limitée par l'hypothèse d'un écoulement laminaire, en effet, les flocculateurs industriels opèrent en régime turbulent. Dans le cas d'un écoulement turbulent, le gradient de vitesse n'est pas constant, en remplaçant ce gradient constant $\frac{du}{dz}$ par une valeur moyenne mesurable "G" donnée par la formule de Camp et Stein BRATBY(1977).

$$G = \left[\frac{P}{\mu v} \right]^{\frac{1}{2}} \quad [s^{-1}] \quad (I-10)$$

P : puissance totale dissipée dans le liquide [$kg.m^2.s^{-3}$]

μ : viscosité dynamique de liquide [$kg.m^{-1}.s^{-1}$]

v : volume du liquide agité [m^3]

L'équation (I-9) dans un écoulement turbulent devient :

$$H_{ij} = \frac{4}{3} n_i n_j R_{ij}^3 G \quad (I-11)$$

Camp et Stein (1943) ont étendu l'équation de Smoluchowski pour définir le taux de variation de concentration des particules, l'expression donnée par ces deux auteurs est adoptée par Hudson (1964) qui suppose $R_j \gg R_i$

$$\frac{dn_i}{dt} = -\frac{4}{3} n_i n_j R_j^3 G \quad (I-12)$$

Où : n_i : nombre de particules primaires par unité de volume [m^{-3}]

n_j : nombre de flocs par unité de volume [m^{-3}]

R_j : rayon des flocs [m]

En tenant compte du fait que toutes les collisions ne donnent pas lieu à une agrégation, ils introduisent un coefficient dit de collisions efficaces α . On obtient l'expression de la variation de concentration de particules primaires. ARGAMAN (1970).

$$\frac{dn_i}{dt} = \left(-\frac{\alpha}{6} n_i d_j^3 \right) n_j G \quad (I-13)$$

α : taux de collisions efficaces

d_j : diamètre des flocs [m]

En admettant que G est constant dans le temps et qu'à l'équilibre n_j et d_j sont indépendants de n_i et de t , la cinétique de floculation devient du premier ordre par rapport à n_i .

$$\frac{dn_i}{dt} = \left(-\frac{\alpha_0}{6} n_i d_j^3 G \right) n_i \quad (I-14)$$

En faisant intervenir le paramètre adimensionnel ϕ (fraction volumique occupée par les floes) on aboutit à :

$$\phi = \frac{1}{6} \Pi n_i d_j^3 \quad (I-15)$$

Et par conséquent n_t / n_0 s'écrit :

$$\frac{n_t}{n_0} = \exp - \left(\frac{\alpha_0 \phi G}{\Pi} \right) t \quad (I-16)$$

Cette équation fondamentale dans l'étude cinétique de la floculation, présente un inconvénient majeur, c'est qu'elle suppose une croissance continue de la taille du floc avec une augmentation de G et t , ce qui est en contradiction avec la réalité.

Expérimentalement, on constate en effet, l'existence d'une valeur optimale du gradient de vitesse au delà de laquelle l'efficacité de la floculation diminue rapidement (augmentation rapide de la turbidité résiduelle).

Les valeurs de gradient de vitesse G et du critère de CAMP GT recommandées par CAMP et STEIN pour un bon fonctionnement de floculateur doivent respecter les valeurs suivantes

- Paramètre adimensionnel GT (critère de Camp) $10^4 \leq GT \leq 10^6$
- Valeurs optimales de G $20 \text{ s}^{-1} \leq G \leq 40 \text{ s}^{-1}$

Conclusion

L'analyse bibliographique des systèmes colloïdaux nous a permis de constater leur complexité. Celle - ci est liée essentiellement à l'interaction colloïdales, et aux phénomènes à l'interface solide liquide et leur conséquence sur la stabilité; l'environnement ionique au voisinage du colloïde se trouve complètement modifié, ces modifications sont dues aux propriétés du colloïde et mécanismes à l'interface; adsorption, échange ionique, pontage, phénomènes électrophorétiques...etc.

II. PRATIQUE DE LA COAGULATION FLOCCULATION DES SUSPENSIONS COLLOÏDALES

Les mécanismes mis en jeu lors du processus de coagulation/floculation sont complexes et du jour en jour de nouvelles révélations viennent d'expliquer, d'éclaircir et de confirmer des résultats d'étude du mécanisme. Nous voulons à travers ce paragraphe passer en revue des principaux résultats pratiques de coagulation floculation des suspensions colloïdales.

II-1. FORMATION ET CASSAGE DE FLOC

La conception la plus réaliste du processus de floculation consiste à le considérer comme état d'équilibre entre deux phénomènes antagonistes simultanés GALLINARI (1992).

- D'une part l'agrégation de particules primaires et de petits floccs pour en former de plus gros.
- D'autre part l'érosion ou la fragmentation des floccs.

Les principales équations proposées sont présentées dans le tableau 1.1 suivant :

Auteur	Expression proposée
ARGAMAN (1971)	$\frac{dn_f}{dt} = -K_A G n_f + K_B G^2 n_f$
PARKER (1972)	$\frac{dn_f}{dt} = -K'_A G n_f n_f + K'_B G^2 n_f$
BRATBY (1977)	$\frac{dn_f}{dt} = -K''_A G n_f + K''_B G^2 (n_f)_0$
TAMBO (1979)	Expression beaucoup plus complexe

Tableau 1.1 : Modification de l'équation de Smoluchowski selon différents auteurs KELLIL (1989).

Où : K_A, K'_A, K''_A : constantes cinétiques de formation des floccs
 K_B, K'_B, K''_B : constantes cinétiques de cassage des floccs

Ces équations sont en accord sur deux points :

- la vitesse de cassage des floccs est proportionnelle au carré du gradient de vitesse.
- la constante de cassage K_B est inférieure d'environ deux ordres de grandeur à la constante de formation K_A .

L'intégration du modèle d'ARGAMAN pour un réacteur en batch donne :

$$\frac{n_1}{n_0} = \frac{K_B}{K_A} G + \left(1 - \frac{K_B}{K_A} G\right) e^{-K_A T G} \quad (I-17)$$

- n_1 : concentration en nombre des particules restant après période de floculation [m^{-3}]
- n_0 : concentration en nombre des particules avant la période de floculation [m^{-3}]
- K_A : constante de formation (agrégation) [-]
- K_B : constante de cassage [s]
- G : valeur moyenne du gradient de vitesse [s^{-1}]
- T : période de floculation [s]

Pour une floculation dans n réacteurs statiques consécutifs, le modèle est généralisé comme suit :

$$\frac{n_0}{n} = \frac{(K_A T G)^n}{1 + K_B T G^2 \sum_{i=0}^{n-1} (1 + K_A T G)^i} \quad (I-18)$$

La cinétique de floculation a fait l'objet également de modélisation en fonction de la taille de floc, nous citons le modèle lagrangien de Winterwerp. MIETTA (2009b) :

$$L_c = L_p + \frac{k_a C}{k_b G^2} \quad (I-19)$$

- L_p : taille caractéristique des particules primaire [m]
- L_c : taille de floc à l'équilibre [m]
- k_a : paramètre de formation [m^2/kg]
- k_b : paramètre de cassage [$m^{-2}/s^{1/2}$]
- C : concentration massique de sédiment [kg/m^3]
- G : gradient de vitesse [s^{-1}]

Dans ces modèles les constantes K_A et K_B , qui caractérisent respectivement la cinétique de formation et de cassage, ont pu être déterminées théoriquement ou expérimentalement. Pour la détermination expérimentale, HAARHOFF (1996) note que 12 tests expérimentaux sont adéquats pour une bonne estimation, les premières minutes de floculation sont les plus importantes pour l'estimation de la constante de formation K_A vu que les particules primaires disparaissent rapidement ; en revanche, pour l'estimation de la constante de cassage K_B les derniers moments de floculation sont les plus importants parce que le cassage ne peut se faire qu'après formation des floes.

Les constantes cinétiques K_A et K_B sont révélatrices de l'état de la structure du floc ; plus K_A est élevée plus le floc se forme rapidement, plus K_B est faible plus le floc est résistant.

Le tableau 1.2 ci-dessous résume les valeurs numériques de ces constantes proposées par différents auteurs :

Auteur	K_A [-]	K_B [s]
ARGAMAN (1971)	$5,1 \cdot 10^{-5}$	$1,1 \cdot 10^{-7}$
PARKER (1972)	$5,14 \cdot 10^{-5}$	$1,08 \cdot 10^{-7}$
BRATBY (1977)	$2,5 \cdot 10^{-4}$	$4,5 \cdot 10^{-7}$
REITER (1979)	$4,5 \cdot 10^{-5}$	10^{-7}

Tableau 1.2 : Valeurs numériques des constantes cinétiques selon différents auteurs KELLIL (1989).

Les phénomènes de formation et de cassage sont encore mal connus et continuent de faire l'objet de nombreuses recherches tant expérimentales que théoriques. Dans ce qui suit nous présentons quelques résultats pratiques échelonnés dans le temps relatifs à la formation et au cassage de floc.

REBHUN (1983) montre que les floccs à base d'alun se forment plus rapidement que ceux à base de polymère ; mais ils sont moins résistants à la fragmentation et l'érosion.

FRANÇOIS (1985 et 1987a) met en évidence que la structure du floc n'est pas homogène, elle est composée de quatre couches ; particule primaire, "flocculi", floc et agrégation de floc, la liaison entre les particules de chaque niveau est élastique. Il note aussi que le cassage de floc est indépendant de la géométrie du flocculateur et dépend en plus de l'énergie d'agitation, de la dimension, de la structure et de l'âge du floc.

En étudiant le vieillissement du floc d'hydroxyde d'aluminium, FRANÇOIS (1987b) met en exergue que le diamètre, la densité et la résistance de floc ne sont pas constants et que la texture et la structure de floc passe par trois mécanismes :

- Condensation et polymérisation
- Agrégation et cimentation
- Cristallisation

La cimentation est influencée par la quantité de suspension solide et le mélange rapide, alors que le degré de cristallisation dépend des énergies de mélange et du dosage en coagulant.

KUO (1988a) indique que la formation de floccs est influencée par le pH, la turbidité, la dose de l'alun, la préozonation, le temps de floculation et la matière organique dissoute.

KELLIL (1989) confirme l'influence directe des conditions de floculation sur la taille moyenne des floccs, il note que :

Il existe un maximum pour la taille du flocc de l'ordre de 500 μm qui correspond à une valeur de GT de 10^4 , ce qui est en accord avec les valeurs recommandées par CAMP :

- Pour les faibles valeurs de GT , le mécanisme de croissance du flocc prédomine.
- Pour $GT > 5 \cdot 10^4$ le mécanisme de cassage du flocc est plus important, plus la valeur de G est grande et plus le cassage est important.

REBHUN (1990) étudie la formation et le cassage de floccs des suspensions de montmorillonite et de "vase naturelle de rivière+acide humique" avec différents flocculants (alun, polymère, "alun+polymère cationique" et "alun+polymère non-ionique"), il en résulte que :

- Les floccs de l'alun présentent les constantes de formation et de cassage les plus élevées, ces floccs se forment rapidement et sont les moins résistants à la fragmentation et à l'érosion.
- L'utilisation du polymère augmente la résistance des floccs au cassage, le rapport de constante de cassage (K_B alun/ K_B polymère) est de 7-8, cependant la combinaison "polymère+alun" augmente davantage la résistance de flocc.
- Les floccs de "l'alun+polymère cationique" se forment plus rapidement que les floccs de "l'alun + polymère non-ionique" et sont les plus résistants au cassage.

HAARHOFF (1997) étudie la stabilité des constantes cinétiques de floculation des particules de kaolinite mises en suspension dans l'eau de robinet en fonction des paramètres suivants :

- 1) Concentration des particules primaires (exprimé en turbidité)
- 2) Dosage du coagulant $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$
- 3) Temps de décantation

Il ressort de cette étude que pour les deux dosages testés 20 et 30 mg/l:

- L'effet de l'augmentation de la turbidité initiale de 20 NTU à 100 NTU est non significatif pour la constante de formation K_A , et au contraire consistant pour la constante de cassage K_B .
- L'effet de la dose du coagulant sur ces constantes est non significatif.
- Le temps de décantation affecte les valeurs des constantes de formation et de cassage ; une augmentation du temps (de 15 à 60 min) correspond à un léger accroissement de la constante de formation et une diminution de celle de cassage.

BACHE (1997 et 1999) indique que pour des eaux de couleur et de matière organique élevées, les floccs formés sont considérés fragiles, cette fragilité est due au mécanisme de coagulation par neutralisation de charge, majeur mécanisme dans l'élimination de la matière organique naturelle. BACHE note aussi que les floccs "alumino-humique" sont fragiles et que leurs tailles sont proportionnelles à l'échelle de Kolmogorov.

SEKIOU (2001) met en exergue que la floculation des suspensions de particule de kaolin, de vase de barrage, de la tourbe et de la boue de station d'épuration suit le modèle d'Argaman Kaufman, néanmoins les suspensions dont le taux de matière organique est élevé ne répondent pas de la même manière que les suspensions minérales aux changements de pH, de la puissance d'agitation, et du dosage en coagulant, en effet pour les suspensions organiques on constate :

- La plage de pH optimal est rétrécit.
- Le rendement d'élimination de la turbidité est faible.
- Le taux de consommation de coagulant est élevé.
- La puissance optimale d'agitation est faible.
- Les floccs formés sont fragiles.

Duan (2003) résume dans son article les résultats suivants :

- La dynamique de balance entre la formation et le cassage conduit à un état d'équilibre de distribution de la taille de flocc. la taille limite dépend du taux d'agitation appliquée.
- Lorsque l'intensité d'agitation augmente, les floccs préformés subissent un cassage lié étroitement à la taille de flocc et à l'échelle de turbulence.
- La différence constatée en coagulation par le sulfate d'aluminium à 20 et 5°C est liée à la résistance du flocc est non pas aux caractéristiques du écoulement.

TURCHIULI (2004), dans son étude de floculation par le sulfate d'aluminium et le chlorure ferrique de huit types de floccs argileux synthétiques préparés à partir de quatre suspensions argileuses de concentrations en solide allant de 0,5 jusqu'à 5 g/l, montre que :

- Les solides de concentration élevée forment de petits floccs et des boues contenant moins d'eau liée, ces boues se déshydratent plus rapidement que celles produites des solides de faible concentration.
- Les floccs d'alun sont plus grands et plus compactes que les floccs ferriques, et décantent rapidement produisant des boues contenant 20% de moins d'eau liée.

YUKSELENA (2004) note que :

- Pour une agitation élevée le cassage des floccs formés à base de suspension de kaolin et de sulfate d'aluminium est irréversible. Ceci revient à la spécificité du cas de la sweep flocculation ; où Il apparait probable que le cassage de flocc d'hydroxyde métallique entraîne la rupture des liaisons chimiques inaptes à se reformer.
- Les floccs formés à base de polyélectrolyte cationique et qui résultent de déstabilisation des particules d'argile par neutralisation et pontage montrent une reformation complète après cassage.

YOUN (2004) étudie la flocculation et la reflocculation des suspensions d'argiles par différents polymères, et conclut qu'une bonne flocculation initiale est obtenue par pontage en utilisant des polymères de grand poids moléculaire. Cependant, en cas d'ajout des polymères de haute densité de charge, la reflocculation des suspensions d'argile après le cassage initial de flocc est favorisée par attraction électrostatique élevée entre particules.

McCURDY (2004) note que le taux de réagrégation de floccs après le cassage est tributaire du type de coagulant. Le traitement par PAC/alun (polychlorure d'aluminium/sulfate d'aluminium) permet une reformation importante (due à l'augmentation de l'efficacité de collision) par rapport au traitement par l'alun ou le PAC seules.

KOBAYASHI (2004) compare la flocculation de polystyrene latex avec du KCl et la flocculation classique avec des polymères et des sels à base d'aluminium et de fer, il indique que la force de flocc à base de KCl, et qui vaut 2 nN, est de 1 à 2 fois moins que celles des floccs formés à base de polymères et d'hydroxyde métallique. Il mentionne que cette force est indépendante de la taille de flocc.

KOBAYASHI (2005) réétudie la flocculation de particules de terre naturelle, réalisées pour trois cas ; avec 0,5 mole de solution de NaCl (pH = 5,4 à 5,5 et pH = 6,6), avec 0,1 mole de solution de CaCl₂ (pH = 6,4 à 6,9) et avec de l'eau distillée acidifiée par du HCl diluée (pH= 5,6), les résultats montrent que la force de flocc est respectivement 0,3 nN ; 0,7 nN et 4 nN pour les ajouts de NaCl, CaCl₂ et de HCl, ces valeurs de force sont de l'ordre de 1-3 fois moins que ceux des floccs formés par les flocculants polymères et l'hydroxyde d'aluminium ou de fer.

JARVIS (2005) résume les différentes tendances de résistance de floc comme indiqué dans le tableau 1.3 suivant :

Caractéristiques de floc	Observation
Taille de floc	La taille de floc augmente \longrightarrow la résistance de floc diminue
Dose de coagulant	La dose optimale correspond à un floc résistant
Addition de polymère	Pour un floc biologique Addition de polymère \longrightarrow la résistance de floc diminue
	Pour un floc chimique Addition de polymère \longrightarrow la résistance de floc augmente
Type de floc	Floc de pontage de particule > floc de neutralisation de charge de particule > floc de complexation Le signe > signifie plus résistant que

Tableau 1.3 : Tendance de résistance de floc JARVIS (2005).

Jarvis (2006) conclut son étude de coagulation d'une eau de surface de faible alcalinité contenant un taux élevé de carbone organique dissous par :
L'addition du polymère réduit le taux de dégradation de floc mais n'améliore pas réellement l'aptitude de reconstruction de floc après son cassage, cette capacité de reconstruction reste très faible avec ou sans addition de polymère.

KILANDER (2006 et 2007) indique qu'il existe une grande variation spatiale de la taille moyenne de floc et de la forme de sa distribution granulométrique à l'intérieure de la cuve de mélange, et que cette différence diminue lorsque le volume de la cuve augmente de 5 à 28 litres, l'augmentation du volume jusqu'à 560 litres accroît cette variation spatiale.

Il note aussi que pour une configuration de cuve de coagulation et une énergie d'agitation données, il existe un volume optimal de ce bac de mélange permettant de minimiser la variation spatiale de distribution de la taille de floc. Cela permet d'extrapoler de manière significative la cuve de jar test à grande échelle.

COUFORT (2007) montre que dans le cas de floculation des suspensions de bentonite par du sulfate d'aluminium, et quelle que soit la forme de l'agitateur, l'état stationnaire caractérisé par la balance entre les phénomènes d'agglomération et de rupture est rapidement atteint à moins de 15 à 30 min. Le cassage de floc est lié au taux de dissipation visqueux de l'énergie cinétique turbulente.

MIETTA (2009b) montre que le temps d'équilibre dépend du gradient de vitesse, ce temps s'accroît suite à l'élévation du degré d'agitation. MIETTA (2009a) note aussi que les floccs formés des particules de vase naturelle sont de tailles plus grandes que celles des floccs de vase traitée avec du peroxyde d'hydrogène (0% M.O) et plus grandes aussi que celles des floccs de kaolinite (0% M.O).

WANG (2009a) montre que les floccs générés par divers coagulants à base de sels d'aluminium exhibent des différences dans la taille, la force et la structure. WANG note aussi que la formation, le cassage et la reformation des floccs d'acide humique sont tributaires des espèces d'aluminium présentes ; Les floccs formés des espèces monomères et dimères sont larges et dures et sont caractérisés par un temps de croissance élevé, alors que les floccs formés par $Al_3(OH)_4^{3+}$ sont petits et fragiles, caractérisés par un temps de croissance réduit.

XIAO (2009) montre l'effet des basses températures sur la cinétique de la coagulation des particules de la kaolinite ; les constantes de formation et de cassage ont une tendance décroissante lorsque la température diminue de 22°C à 2°C.

SEKIOU (2009) étudie l'évolution des constantes cinétiques de floculation de six types de particules solides dispersées coagulées par le sulfate d'aluminium, il en résulte que :

- La constante de cassage K_B augmente avec l'augmentation du taux de matière organique dans le solide dispersé, ce qui signifie que les floccs formés des suspensions organiques sont moins résistants à l'érosion et à la fragmentation que ceux formés de matière minérale.
- L'interaction entre la matière minérale et la matière organique est prouvée, il apparait que l'addition de la matière minérale renforce, solidifie et réduit le taux de dégradation de flocc.
- La matière organique quelle que soit son origine (tourbe végétale ou boue de station d'épuration,) affecte négativement la formation et la rigidité du flocc, pour des suspensions de plus de 50 % de matière organique les floccs sont fragiles alors que les floccs formés de suspension de moins de 25 % de matière organique sont résistants.

XU(2010) étudie le cassage et la reformation de deux types de floccs ; ceux formés à base du polymère $Al_3(OH)_4^{3+}$ et ceux formés à base du PAC (polychlorure d'aluminium) et montre que :

- Pour des conditions optimales de coagulation, les floccs formés de polymère $Al_3(OH)_4^{3+}$ sont fragiles par rapport aux floccs formés du PAC, ces floccs de polymère $Al_3(OH)_4^{3+}$ présentent une meilleure reformation.

- Aussi bien pour le polymère $Al_3O_4(OH)_7^+$ que pour le PAC, les floccs formés en solution acide d'acide humique sont dures et mieux aptes à la reformation que ceux formés en solution neutre ou alcaline, néanmoins l'effet de la variation de pH est plus faible sur le cassage et la reformation de floccs de $Al_3O_4(OH)_7^+$ que sur ceux des floccs de PAC.

ZHANG (2010) évalue la résistance de flocc, et sa reformation par le polychlorure de fer (PFC) et note que la dose dudit coagulant affecte les paramètres du flocc ; pour une dose de 22 mg/l, des floccs de grandes dimensions se forment rapidement, résistent à l'élévation de la turbulence mais montrent une faible reformation, alors que pour une dose de 3 mg/l les floccs sont petits et faiblement influencés par la turbulence et montrent une réversibilité totale et une meilleure reformation.

Par ailleurs YU (2011) montre que l'aptitude à la reformation des floccs de kaolin cassés dépend de la nature des espèces du PAC (polychlorure d'aluminium) ajouté comme un second coagulant ; il indique que le précipité d'hydroxyde amorphe formé des espèces monomères améliore l'adhésion entre les floccs cassés alors que pour les espèces polymères, la nature du précipité est différente et n'a aucun effet sur la reformation du flocc.

En conclusion, il apparait clairement que le phénomène complexe de formation et de cassage de flocc est tributaire à l'interaction de plusieurs paramètres, nous citons :

- La taille, le type et la structure de flocc.
- La nature et la concentration des particules primaires.
- Le type et le dosage du coagulant.
- Les espèces d'aluminium présentes.
- Le type, le dosage et le poids moléculaire de polymère.
- Le temps de décantation et le temps de cassage de flocc.
- Le taux de matière organique dissoute et particulaire.
- La configuration et la taille de la cuve de coagulation.
- L'alcalinité et les adjuvants minéraux.
- Le mécanisme de floculation.
- La température et le pH de la suspension.
- L'échelle de turbulence et le gradient d'agitation.

II-2. INFLUENCE DU pH SUR LA FLOCCULATION

Le rôle du pH est particulièrement important dans la mesure où il détermine à la fois la charge des colloïdes, la nature des espèces qui naissent de sels minéraux d'Al et de Fe GREGOR (1997), la formation et la décomposition des espèces polymérique d'aluminium ZHAO (2009), et la dissociation des groupements fonctionnels existant dans la structure des composés organiques en fonction de leurs pKa.

DESJARDINS (1996) note que le pH optimal pour l'élimination du COT se situe entre 5 et 6,5 pour les coagulants à base d'aluminium. MIETTA (2009a) lie la taille des floccs à la valeur de pH ; la diminution du pH provoque une élévation de la taille moyenne des particules, cela revient à la floculation rapide promu par les interactions chimiques de protonation-déprotonation des surfaces d'argile.

Xu (2010) note que le mécanisme de coagulation d'acide humique est tributaire au pH ; pour un pH de 5 le mécanisme dominant est "complexation/neutralisation de charge", par contre pour un pH de 7 et un pH de 9 il semble que l'adsorption joue un rôle dominant dans la coagulation de l'acide humique par l'alun.

DESJARDINS (1988), GAO (2007) mentionnent que pour chaque suspension, il existe une plage de pH pour laquelle la coagulation a lieu rapidement, plage qui est fonction du coagulant utilisé, de sa concentration et de la composition de l'eau à traiter.

II-3. EFFET DE LA TEMPERATURE SUR LA FLOCCULATION

Les études réalisées sur l'influence de la température sur le processus de coagulation/floculation, sont nombreuses, nous passons en revue des résultats de quelques travaux :

DENTEL (1988), montrent que l'abaissement de température entraîne :

- Une augmentation de la concentration en sulfate d'aluminium à l'optimum de floculation.
- Une décroissance de la taille des floccs.
- Un déplacement du pH optimum de floculation vers les pH basiques.
- Une inhibition des mécanismes de floculation pour des températures inférieurs à 10°C, en particulier, pour les eaux peu turbides.

BENSCHOTEN (1990) trouve que le pH pour lequel $Al(OH)_3$ précipite s'élève de 4,6 à 25 °C jusqu'à 5,5 à 4,8 °C et que le point isoélectrique se déplace de pH 7 à 25 °C à pH 9 pour 4,8 °C.

DESJARDINS (1996) montre que la température affecte de manière significative l'abattement de la matière organique par le sulfate d'aluminium ; quand la température passe de 22C° à 4C°, on observe une réduction du pourcentage d'élimination du COT ne dépassant pas 15 %. Il note aussi que pour une eau froide la dose optimale pour l'enlèvement de la matière organique est supérieure à celle requise pour enlever la turbidité.

EXALL (2003) note que les proportions des espèces polynucléaires et les complexes Al-SO₄ diminuent en eau froide. Alors que DUAN (2003) note que l'influence de la variation de la température sur la coagulation par les sels métalliques est liée à son effet sur les réactions d'hydrolyse, la précipitation et la solubilité de l'hydroxyde métallique.

XIAO (2009) étudie l'effet de la température sur la cinétique de la coagulation de la kaolinite et sur la morphologie des surfaces des floccs.

Il conclut que :

- les faibles températures ralentissent le processus de coagulation par suite de diminution du taux d'agrégation.
- Les floccs formés à de faibles températures sont moins compacts et présentent des structures irrégulières.

II-4. EFFET DE LA NATURE ET DE LA CHARGE DES COLLOÏDES SUR LA FLOCCULATION

Les études réalisées sur la coagulation/floculation des suspensions colloïdales montrent que :

- La structure minéralogique et les propriétés surfaciques des particules affectent la dose du coagulant LURIE (1983)
- DENTEL (1988), note que dans le cas de déstabilisation par neutralisation de charge, la dose du coagulant est liée à :
 - La solubilité de l'hydroxyde de l'aluminium.
 - La densité de charge des particules primaires.
 - La surface spécifique des particules primaires.
- KUO (1988b) indique que les conditions effectives d'élimination de la matière organique dissoute sont généralement similaires aux conditions permettant la formation et l'accroissement de flocc.

Au cours de notre étude bibliographique nous avons pu constater la mise en évidence de lois stœchiométriques entre les caractéristiques de la suspension et la dose en coagulant ; SPENGLER (1983) indique l'existence de faibles lois stœchiométriques entre l'élimination de la couleur et la dose de l'alun. BAZER-BACHI (1990) stipule que dans le cas de la coagulation par adsorption-déstabilisation, on a observé une relation stœchiométrique entre la surface spécifique et la dose du

coagulant. FRANCESHI (1991) mentionne qu'il existe une relation stoechiométrique entre la C.E.C (capacité d'échange cationique) et le taux de sulfate d'aluminium ajouté. DESJARDINS (1988), ZOUGRANA (1993) et HUANG (1996) notent l'existence de relations stoechiométriques entre la quantité de colloïdes organiques et la dose du coagulant consommé, et montrent que la nature de la suspension influe sur la turbidité résiduelle et sur le taux de sulfate d'aluminium à l'optimum de floculation. ZOUGRANA(1993) ajoute que cette stoechiométrie dépend du pH, de l'alcalinité et du type de coagulant. ACHOUR (2000) précise que les coefficients des lois stoechiométriques entre les concentrations en substances humiques et la dose optimale de coagulant en milieu minéralisé sont tributaires de la composition minérale des eaux.

- L'ajout d'une faible quantité d'argile permet d'améliorer le rendement d'élimination des substances humiques BAZER-BACHI (1990). GREGOR (1977) et DESJARDINS (1996) notent que le temps d'injection de ces adjuvants est aussi important ; un mauvais choix du moment d'injection affaiblit la performance d'élimination de la matière organique.
- L'élimination de la matière organique est liée directement à son poids moléculaire, au degré d'ionisation et au nombre de groupements fonctionnels TAMBO (1991), LEFEBVRE (1993) et HUANG (1996), alors que COLLINS (1986) ajoute l'acidité carbonyle et le taux de substance humique. JULIEN (1994) et PURNENDU (1998) indiquent qu'une meilleure élimination de la matière organique est réalisée par adsorption sur floc d'hydroxyde d'aluminium préformés.

Le mode d'élimination de la matière organique naturelle (M.O.N) par les sels d'aluminium a fait l'objet de plusieurs travaux, EDZWALD (1993) résume cela en plusieurs réactions :

- Réaction de complexation de l'aluminium avec la matière organique naturelle, particulièrement ce sont les substances humiques qui forment de fortes liaisons.
- Réaction d'hydrolyse de l'aluminium où les liaisons inorganiques OH^- sont en compétition avec les liaisons organiques
- Précipitation directe des particules "Al-M.O.N." (ou bien "Al-humate") après complexation et neutralisation de charge
- Adsorption de M.O.N. ou/et des espèces "Al-M.O.N." avec $\text{Al}(\text{OH})_3$ amorphe qui précipite après addition du coagulant.

EXALL (2003) étudie l'effet de la matière organique, les sulfates et la température sur la spéciation de l'alun et la spéciation de deux coagulants préhydrolysés ; le polychlorure d'aluminium (PAC) et le poly-hydroxy-sulfate d'aluminium (PAHS). Les résultats montrent :

- En cas d'addition de l'acide tannique à la solution du coagulant, aucune formation soluble du complexe "matière organique-aluminium" n'est observée.
- En cas de l'alun dilué, la présence de la matière organique affecte légèrement la spéciation de l'aluminium, réduit grandement la proportion des espèces tridecamères Al_{13} observées en solutions préhydrolysés, cette réduction est élevée en cas de PAC par rapport au cas du PAHS, ceci est dû probablement à l'effet d'obstacle du sulfate.

LAGALY (2003) trouve que la floculation des dispersions de montmorillonite est dictée par la forme anisométrique et la distribution de charge des faces et des arêtes des particules. La nature (organique ou inorganique) des contre-ions semble jouer un rôle important dans l'adsorption et la coagulation ; à titre d'exemple à pH 6, la coagulation s'accomplit entre les arêtes positifs et les faces négatives.

II-5. EFFET DE LA MINERALISATION SUR LA FLOCCULATION

L'influence de la minéralisation totale ou de la teneur en certains cations ou anions a été abordée par différents auteurs. Dans ce qui suit nous présentons les résultats de quelques travaux :

FRANCESHI (1991) note que les anions et les cations ont une double action :

- Sur l'environnement des colloïdes, l'augmentation de la teneur en ions entraîne une diminution des forces de répulsion interparticulaires par compression de la double couche.
- Sur la formation des produits d'hydrolyse des sels de fer ou d'aluminium.

DESJARDINS (1988) indique que l'addition des ions sulfates (SO_4^{2-}) élargissait la plage de pH optimal vers l'acidité et favorise la formation de composés neutres ou chargés négativement, alors que l'addition d'ions phosphates (PO_4^{3-}) déplaçait sans élargir la plage de pH optimal. Les ions bicarbonates affectent la précipitation des hydroxydes d'aluminium.

ZOUGRANA (1993) montre que l'alcalinité de l'eau brute joue un rôle important dans le processus de coagulation / floculation, et particulièrement dans l'enlèvement du carbone organique total (COT).

- Avec le sulfate d'aluminium, la reminéralisation améliore sensiblement l'élimination de la turbidité et conduit à de faibles teneurs en aluminium résiduel par rapport au chlorure ferrique.

- Le rendement d'élimination de la matière organique n'excède pas 54,5 % sur l'eau brute naturelle, alors qu'il atteint 64,4 % sur les eaux minéralisées.

PENG (1994) a mis en évidence que les sels de calcium et ceux d'aluminium affectent la floculation des suspensions de kaolin par le polymère anionique.

ACHOUR (2000) démontre que la composante minérale influe grandement sur l'efficacité d'élimination des substances humiques, dopées en échantillons d'eaux souterraines, et celles présentes dans les eaux de surface des barrages algériens (Chiffa, Hamiz).

Dans son étude VELASCO (2007) évalue l'efficacité de l'alun et du PAC pour le traitement des eaux de surface à faible C.O.D. (2,1 à 2,4 mg/l) et à faible turbidité (<1 NTU) pour deux niveaux d'alcalinité (<60 mg CaCO₃/l et >120mg CaCO₃/l).

- Il a été trouvé que le PAC est d'une efficacité supérieure en cas de faibles et fortes alcalinités, quant à l'alun, il semble être plus lié aux conditions opératoires et en particulier au pH.
- La grande différence entre les deux coagulants est marquée par l'élimination finale de THMFP (trihalométanes formation potentiel) cette différence est maximale lorsque le pH est poussé à plus de 8.
- Au contraire de l'alun, lorsque le PAC est utilisé, une relation stœchiométrique est établie entre l'élimination de l'absorbance UV 254 et l'élimination des THMFP.
- Il en résulte aussi que L'optimisation de l'élimination du C.O.D. et de l'élimination de l'absorbance UV 254 n'assure pas l'élimination optimale des précurseurs de THM (trihalométanes).
- L'effet promoteur ou inhibiteur de l'alcalinité apparait étroitement lié au type du coagulant utilisé.

MIETTA (2009a) note que les diamètres moyens des floes de kaolinite et de vase répondent de la même manière aux changements de pH et de concentration des sels ajoutés; pour les suspensions salines à pH 8, la taille moyenne des floes augmente lorsque la concentration des sels et le potentiel zéta augmentent.

WANG (2009b) conclut que dans le cas de coagulation par les sels d'aluminium, la dureté totale affecte l'agrégation des floes ; la taille maximale des floes formés à base de AlCl₃, PAC-1 et PAC-2 est obtenue lorsque la dureté atteint respectivement 8, 4, 1 mmole/l. Pour les différentes duretés testées, WANG note que l'espèce monomère Al montre une rapide réaction et le plus grand floe formé, cependant l'espèce Al₁₃ montre à l'inverse la

réaction la plus lente, la taille de floc la plus petite, et la distribution de tailles la plus rétrécie.

WANG (2010) indique que l'efficacité de coagulation de l'acide humique seul par les sels d'aluminium diminue lorsque la force ionique et la dureté augmentent. Aussi quand la dureté passe de 10 mg/l de CaCO_3 à 160 mg/l de CaCO_3 , les propriétés des floccs formés par le PAC restent constantes, au moment où les floccs formés par AlCl_3 deviennent moins compacts et les floccs formés par le Nano-PAC $[\text{AlO}_4\text{Al}_{12}(\text{OH})_{24}(\text{H}_2\text{O})_{12}]^{7+}$, qui est constitué principalement de Al_{13} , apparaissent les plus petits et les plus denses.

Par ailleurs, l'efficacité de coagulation de "l'acide humique plus particules de Kaolin" augmente avec l'augmentation de la dureté totale et de la force ionique. Lorsque la dureté totale passe de 10mg/l de CaCO_3 à 160mg/l de CaCO_3 , ou bien la force ionique augmente de 0.05 mole/l à 0.25 mole/l, le AlCl_3 et le PAC donnent des floccs moins compacts, les floccs formés par Nano-PAC sont aussi moins compacts et larges.

En règle générale, la minéralisation de l'eau affecte :

- Plage de pH optimal.
- Temps requis pour la floculation (vitesse de réaction).
- Mécanisme et efficacité de floculation.
- Spéciation et hydrolyse de sels.
- Dosage optimal du coagulant
- Résiduel de coagulant dans l'effluent traité.
- Taille, structure et compacité des floccs

II-6. INFLUENCE DU COAGULANT

Le coagulant est l'agent chimique ajouté à la suspension en vue de la déstabiliser. Son rôle consiste à diminuer les forces de répulsion électrique afin de compresser la double couche ionique et baisser la barrière d'énergie permettant aux particules se rapprochant suffisamment d'entrer en contact.

Le coagulant prend très souvent la forme d'un sel de cation trivalent, selon la règle de Schulzer Hardy, Le rapport entre les efficacités des ions mono, di et trivalents pour la même suspension est 1 : 8 : 600.

Le sulfate d'aluminium est le coagulant les plus usuel, ses réactions d'hydrolyse sont extrêmement rapides et se passent dans les microsecondes, si les espèces polymérisées ne se sont pas formées et dans une seconde si celles-ci sont formées. La formation du précipité solide d'hydroxyde d'aluminium est lente et a lieu dans un intervalle de 1 à 7 secondes AMIRTHARAJAH, (1982) et SPENGLER (1983).

De nombreux diagrammes de stabilité des différentes espèces d'aluminium en fonction du pH ont été successivement proposés par Pakham(1963), Stumm et O'mellia (1968), Mc-coot et West (1978), Rubin et Block (1979),

Amirtharajah et Mills (1982). Ces diagrammes permettent de localiser dans différentes zones du graphe le mode d'action du coagulant voir figure 1.1

AMIRTHARAJAH (1982) distingue trois zones d'action pour le sulfate d'aluminium :

- Une zone de coagulation par précipitation ou zone de coagulation par balayage.
- Une zone de coagulation par neutralisation de charge.
- Une zone où les deux mécanismes coexistent.

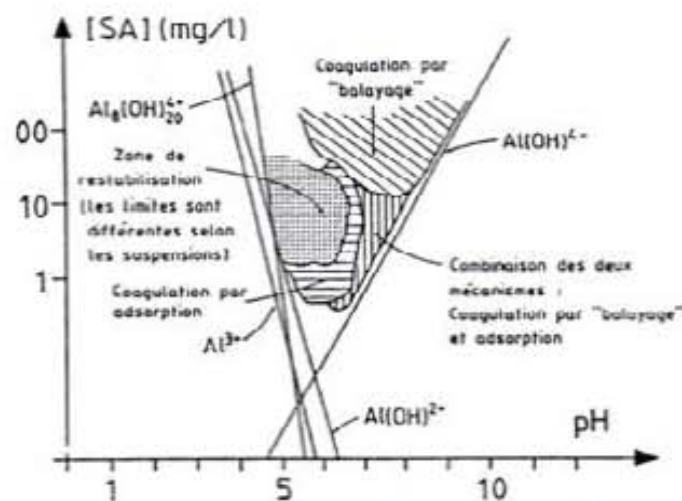


Figure 1.1: Diagramme d'action du sulfate d'aluminium en fonction du pH, d'après FRANCESCHI (1991).

GREGORY (2001) indique que pour des dosages équivalents, le PAC (polychlorure d'aluminium) permet une floculation rapide et donne lieu à des floccs solides comparés aux floccs de l'alun.

MING-QUAN (2007) note que le type du coagulant affecte profondément la performance de la floculation, le HPAC (polychlorure d'aluminium composite) est meilleur que le FeCl₃ (chlorure ferrique) dans l'élimination du carbone organique dissous (C.O.D.) et des trihalométhanes. En fait, le HPAC est montré plus efficace que le FeCl₃ dans l'élimination des particules solides et du C.O.D. aussi bien par flottation que par décantation. MING-QUAN mentionne que sous de faibles températures, le FeCl₃ est plus performant que l'alun, il permet une meilleure élimination de la turbidité et de la couleur, cela revient essentiellement à la formation de grands floccs rapidement décantables.

Duan (2003) résume les résultats suivants :

- La coagulation des suspensions de kaolin par le sulfate ferrique permet de meilleures efficacités de floculation que l'alun, en effet, il apparait que le floc ferrique est plus solide que le floc d'alun pour des températures allant de 5 et de 20°C
- D'un autre coté, pour les faibles températures la floculation par le PAC (polychlorure d'aluminium) montre de meilleurs rendements que la floculation par l'alun, par ailleurs il en découle que le volume de boue produite suite à l'utilisation du PAC est faible.

Wu (2007) note que le mécanisme principal de floculation de l'alun est "neutralisation de charges/sweep-floculation" alors que pour le PAC ce sont les mécanismes de coagulation par adhésion électrostatique et agrégation par pontage en plus de la neutralisation de charges. Il a été montré aussi que le mécanisme d'adsorption n'est pas le même pour l'alun que pour le PAC, pour le premier ; l'adsorption se passe essentiellement par précipitation chimique, par contre, la force prédominante dans le processus d'adsorption du PAC est d'origine électrostatique.

CHENGA (2007) montre que l'élimination effective de la turbidité est liée, non seulement au coagulant utilisé, mais aussi à la nature de la matière organique présente. Par ailleurs, il note que les coagulants : PASiC (polychloro-silicate d'aluminium) et PAFSiC (polychloro-silicate de fer et d'aluminium) sont meilleurs que le PAC (polychlorure d'aluminium) dans le traitement de l'eau de faible turbidité.

GAO (2002 et 2007) compare la floculation par le PASiC et par le PAC des eaux par l'investigation de l'effet des rapports $[OH]/[Al]$, $[Al]/[Si]$, de l'effet du pH et aussi de l'effet du dosage et de la technique de préparation du coagulant. Les résultats de l'étude montrent que la performance de coagulation par le PASiC dépend de la procédure de préparation, et des rapports $[OH]/[Al]$ et $[Al]/[Si]$. Il note aussi que l'efficacité du PASiC augmente avec l'augmentation du rapport $[OH]/[Al]$, et que la performance du PAC est inférieure à celle du PASiC dans l'élimination de la turbidité, des algues, du C.O.D, et des huiles. La marge du pH optimal est plus large pour le PASiC (6.0–8.5) que pour le PAC (6.0–8.0).

YAN(2008) conclut son étude sur les mécanismes d'élimination de la matière organique avec le $AlCl_3$ par :

- L'espèce monomère Al_a montre une grande aptitude à la complexation avec la matière organique, et la majorité des flocs y formés sont petits et décantent difficilement.

- L'espèce d'aluminium moyennement polymérisé Al_6 déstabilise la matière organique plus efficacement mais les floccs formés sont petits et ne précipitent pas facilement comparativement aux floccs formés par "les espèces polymérisés + solide"
- L'espèce polymérisé Al_3 (espèce trimère et tridécamère) adsorbe et élimine efficacement la matière organique naturelle.

YANG (2009) compare la floculation des suspensions de kaolin par deux coagulants, le PAC et le MBFGA1 (floculant composite). Les résultats montrent que le PAC présente plus d'avantages dans le changement du potentiel du colloïde, le MBFGA1 est bon en adsorption et en pontage alors que le coagulant mixte (PAC+MBFGA1) donne les meilleurs rendements.

Wang (2010) constate que l'efficacité d'un coagulant dépend de la dureté totale, de la force ionique et aussi de la qualité des eaux ; Pour l'élimination de l'acide humique seul, l'efficacité du PAC s'avère la plus élevée par rapport aux $AlCl_3$ et Nano-PAC, alors que pour l'élimination de "l'acide humique plus particules de Kaolin" le $AlCl_3$ permet les meilleurs rendements.

Par ailleurs, au cours de notre recherche bibliographique nous avons pu constater que l'utilisation des polyélectrolytes (simple, hybride, ou modifié, avec ou sans adjuvant), montre des résultats très prometteurs, dans la formation de floccs résistants DUSSOUILLEZ (2010), dans la réduction jusqu'à 99% de la turbidité des suspensions de kaolin LEE (2011), dans l'amélioration de la performance de floculation WANG (2011), dans la maximisation de l'élimination de carbone organique dissous "COD" et la minimisation de la turbidité résiduelle ZHU(2011), et dans la formation et l'aptitude à la reformation de flocc ZHAO (2012).

En résumé, nous pouvons mentionner que le choix du coagulant est dicté par la nature de la matière organique présente dans l'eau, la température, l'alcalinité, la dureté, la force ionique, la complexation, la stabilité et le temps de réaction du coagulant, l'efficacité d'élimination de la turbidité et de la matière organique, la résistance et la taille du flocc désiré, la teneur en résiduel du coagulant, la technique de préparation de coagulant, des rapports $[Al]/[Si]$ et $[OH]/[Al]$, et de la présence des particules solides dans l'eau à flocculer.

II-7. INFLUENCE DES CONDITIONS D'AGITATION SUR LA FLOCULATION

Le processus de coagulation floculation se déroule généralement en deux étapes :

- Une étape de mise en contact entre l'eau à traiter et le coagulant. Cette étape se déroule généralement sous forte agitation, son but est à la fois de permettre la neutralisation de charges des particules et d'amorcer le processus de floculation par la formation de microflocs.
- Une étape d'agitation lente qui assure la mise en contact et la croissance des floccs, Camp et Stein (1943) recommandent que le critère GT (dit critère de Camp) soit retenu pour le bon contrôle du bon fonctionnement d'un flocculateur $10^4 < GT < 10^6$ avec $20 < G < 40 \text{ s}^{-1}$, MLYANAMI (1982) confirme ce résultat dans le cas d'agitation par rotation et par vibration.

LETTERMAN (1973) montre que la turbidité résiduelle dépend du temps du mélange rapide, et que pour toute combinaison de gradient de vitesse et dose du coagulant injecté, il existe un temps de mélange rapide associé au minimum de turbidité. FRANCESHI (1991) indique qu'il existe une valeur optimale de gradient de vitesse lors de la phase d'agitation rapide pour laquelle la turbidité résiduelle est minimale. LAI (1975) a mis en évidence que les mélangeurs qui développent la même surface projetée produisent un même gradient de vitesse.

LETTERMAN (1973) et LEENTVAAR, (1980) notent que le gradient d'agitation rapide ne donne pas une caractérisation complète de mélange dans le flocculateur, en effet pour un gradient d'agitation donné, la turbidité résiduelle est influencée par la forme du mélangeur. En utilisant le vélocimètre d'image de particule CHENG (1997) constate que dans le flocculateur le mélange est hétérogène et l'énergie dissipée n'est pas uniforme.

REBHUN (1990) montre que le gradient d'agitation critique G (gradient optimal max) est tributaire de la nature du floc formé ;

- Le floc de "alun+argile pure (montmorillonite)" est sensible aux grandes agitations, son gradient d'agitation critique est de 40 s^{-1} , alors que le floc de "polymère-montmorillonite" est plus résistant, son G critique est de 110 s^{-1}
- Le floc de "alun + vase naturelle + acide humique" montre un comportement légèrement meilleure que le floc "alun+montmorillonite" son G critique est de $G=12 \text{ s}^{-1}$
- La floculation des suspensions de "vase naturelle+acide humique" avec de "l'alun + polymère non-ionique" augmente la résistance de floc au cassage, son G critique s'élève à 85 s^{-1}

- Pour un G de 140 s^{-1} (supérieur au G optimal) la performance d'élimination des particules solides des suspensions organo-argile par l'alun seul est nulle, alors qu'elle est de 73 % pour les suspensions floculées par "l'alun + polymère non-ionique" et atteint à 87,5% pour les floccs "d'alun+polymère cationique".

ELMALEH (1992) a mis en évidence que pour des concentrations en particules supérieures à 2g/l , l'énergie de floculation dépend de la concentration et de la vitesse d'agitation, alors qu'elle est très faible pour des concentrations inférieures à 2g/l .

HOPKINS (2003) indique que lorsque G augmente de 40 à 90 s^{-1} , la formation de floc s'accroît au moment où la moyenne et la variance de distribution de la taille de floc diminuent, Il mentionne aussi qu'à de faibles gradients de vitesse la formation et la taille de floc ne sont pas homogènes et varient spatialement dans le réacteur.

COUFORT (2007) montre que :

- Dans le cas de mélangeur radial "turbine Rushton" la distribution de la taille des floccs est monomodale et peut s'ajuster par une loi log normal, alors que dans le cas du mélangeur axial "Lightnin A310" la distribution est plutôt bi-modale et peut être ajusté par une combinaison pondérée de deux lois log normale.
- La taille caractéristique de floc est liée à l'échelle de Kolmogorov et ainsi à la puissance dissipée.

MIETTA (2009a) étudie la relation entre quelques quantités mesurables (potentiel zéta, taux de matière organique, le degré d'agitation) et la performance de floculation de la vase, il conclut que, pour différents pH et concentrations de sels, la taille moyenne de floc et la micro échelle de Kolmogorov varient de la même manière en fonction du degré d'agitation.

ZHAN (2011) étudie, pour des eaux de surface de faible taux de matière organique, l'effet de gradient de vitesse sur les caractéristiques des floccs, par l'utilisation de deux coagulants pré-hydrolysés ; le polychlorure d'aluminium (PAC) et le polychlorure de fer (PFC), il en résulte que :

- Pour les faibles valeurs de gradient de vitesse G ($G < 7,6 \text{ s}^{-1}$), le diamètre moyen des floccs augmente avec l'augmentation de G , ce qui indique que l'agrégation des floccs est importante par rapport au cassage, les floccs sont moins compactes, de faible capacité de décantation et de grands facteurs de force (résistance) et de

- Pour les valeurs moyens de gradient de vitesse G ($7,6 \text{ s}^{-1} < G < 15,1 \text{ s}^{-1}$ pour le PFC et $7,6 \text{ s}^{-1} < G < 11,2 \text{ s}^{-1}$ pour le PAC) la floculation est très bonne, donnant de grands compacts floccs, caractérisés par de faibles facteurs de force et de reconstruction.
- Pour les valeurs élevées de gradient de vitesse G ($G > 15,1 \text{ s}^{-1}$ pour le PFC et $G > 11,2 \text{ s}^{-1}$ pour le PAC), le cassage des floccs est prépondérant, la turbidité est élevée, et les facteurs de force et de reconstruction augmentent avec l'augmentation de G jusqu'à la valeur de $G=28,3\text{s}^{-1}$.

II-8. CONCLUSION :

L'énumération, non exhaustive, présentée ci-dessus de l'état des connaissances sur les différents résultats pratiques montre la complexité des mécanismes de coagulation et de floculation des suspensions colloïdales.

La performance de la coagulation floculation est liée à deux catégories de paramètres :

- Ceux liés à la qualité de l'eau brute (pH, température, minéralisation, et nature, caractéristiques, concentration et charges des substances présentes).
- et d'autres dépendent du mode de traitement (nature et dose du coagulant, conditions d'agitation, point et mode d'injection des coagulants).

Nous pensons à la suite de cette revue bibliographique que la performance et la cinétique de la floculation et les différentes interactions entre ces paramètres méritent d'être revus à la lumière de nouvelles révélations et résultats de recherches qui ne cessent d'être de plus en plus pointus et importantes.

D'où l'objectif du troisième chapitre qui consiste à l'étude expérimentale de floculation de six différents types de suspensions de particules solides dispersées, à la revue de l'influence des caractéristiques des particules solides primaires sur la performance et la cinétique de floculation, et à l'étude des interactions entre les paramètres de floculation.

III. MODELISATION DE LA FLOCCULATION

III-1. APPAREILS ET METHODES DE DETERMINATION DE LA DOSE DE COAGULANT.

III-1.1. Jar test

La dose optimale du coagulant est, traditionnellement, déterminée à l'aide d'un essai expérimental appelé "Jar-test". On utilise, à cet effet, un flocculateur de laboratoire, appareil permettant d'agiter simultanément à une vitesse bien déterminée, le liquide contenu dans une série de béciers. L'essai consiste à mettre dans ces récipients, contenant la même eau brute, des doses croissantes de coagulant et de faire l'essai de coagulation. Après quelques instants, on procède sur l'eau décantée à toutes les mesures utiles de qualité de l'eau (turbidité, matière organique, pH...etc.). La dose optimale est déterminée en fonction de la qualité des différentes eaux comparées.

L'inconvénient de cette technique est qu'elle nécessite une intervention de l'opérateur. Ce type d'approche a également le désavantage d'avoir un temps de retard relativement long. En effet, on ne modifie la dose de coagulant dans une station de traitement qu'une fois l'événement apparu (augmentation de la turbidité de l'eau brute). De plus, elle ne permet pas de suivre finement l'évolution de la qualité de l'eau brute.

III-1.2. Détecteur de courant cinétique

Le détecteur de courant cinétique, mis au point par Gerdes en 1966, a été utilisé par certains chercheurs pour déterminer la qualité de la coagulation/floculation en mesurant la charge résiduelle des colloïdes et des particules dans l'eau. DENTEL (1989) note que la nature de la relation entre la dose de l'alun et la lecture du courant cinétique permet le contrôle de la dose par la mesure du courant cinétique. GAO (2002) a utilisé cette technique de mesure du courant cinétique et le suivi du diamètre moyen de floc dans l'évaluation de la floculation des eaux par le PAC (polychlorure d'aluminium) et par le PASiC (polychloro- silicate d'aluminium).

A grande échelle la technique de réajustement de la dose de l'alun en traitement des eaux à été pratiqué à la station de purification de Victriaville (Canda). Pour les faibles variations de débit, le dosage de l'alun n'est pas relié au débit d'eau brute, il est asservi à la seule lecture du courant cinétique. Cependant si les variations sont brusques ou/et importantes, il est avantageux de contrôler le dosage du coagulant à partir du débit et de la mesure du courant cinétique (double asservissement). De plus, le fonctionnement en continu de cet appareil de détection de courant cinétique demande le réajustement du point de référence assez fréquemment (environ une fois par mois) ce qui nécessite de faire un essai Jar-test, ou essai de filtrabilité CYR (1990).

L'application de la technique de détection de courant cinétique en automatisation des stations de traitement doit tenir compte des résultats des recherches dans ce domaine ; DENTEL (1989) montre que la mobilité électrophorétique pour une addition de l'alun se corrèle bien avec le courant cinétique, ELICKER (1992) ajoute que ce courant est proportionnelle au potentiel zêta et à la vitesse de piston (jusqu'à 360 tr/mn).

IL a été montré également par DENTEL (1989) que l'addition des sels diminue le courant cinétique et que l'efficacité de contrôle par courant cinétique était limitée pour certains types de qualité d'eau brute ; en particulier lorsque le pH de l'eau brute est élevé (pH > 8).

ELICKER (1992) et DENTEL (1995) notent que tout changement des caractéristiques électriques induit par le changement du pH ou par l'ajout de coagulant se répercute sur la détection de courant cinétique.

III-1.3. Mesure de potentiel zêta

Par définition, on appelle potentiel électrocinétique ou potentiel zêta, la différence de potentiel entre le cœur de la suspension où la neutralité est établie et le plan de cisaillement qui sépare les ions qui se déplacent avec la particule et de ceux qui restent au sein du liquide quand la particule est soumise à un champ électrique. Ce potentiel mesure la charge électrique superficielle et détermine la grandeur des forces répulsives. La mesure du potentiel zêta est une technique bien connue pour évaluer la neutralisation des charges des colloïdes, la turbidité et la taille des floes. NARONG (2006) montre que lorsque le potentiel zêta est faible la turbidité finale est aussi faible et la taille moyenne des floes est élevée, et pour des potentiels élevés l'inverse est observé.

Théoriquement, il est possible de déterminer cette grandeur à partir de n'importe quel phénomène électrocinétique (électrophorèse, électro-osmose, potentiel de sédimentation et potentiel d'écoulement) ARFANDI (1979). Pratiquement, le plus souvent, c'est la technique électrophorétique qui est utilisée ; le potentiel zêta est déterminé expérimentalement à partir de la mobilité des particules soumises à un champ électrique constant par application de la formule classique de SMOLUCHOWSKI

$$\zeta = \frac{4 \pi \cdot \mu \cdot U}{\varepsilon \cdot E} \quad (I-20)$$

ζ : Potentiel zêta [mV]

4π : facteur de forme correspondant aux particules sphériques []

μ : viscosité dynamique de la suspension [$\text{kgm}^{-1}\text{s}^{-1}$]

ε : constante diélectrique [$\text{CV}^{-1}\text{m}^{-1}$]

U : mobilité des particules (vitesse de déplacement dans un champ électrique) [m/s]

E : potentiel électrique par unité de longueur [V/cm]

Le potentiel zêta dépend essentiellement de la force ionique de la solution, de la température du milieu et de la répulsion électrostatique entre les particules colloïdales. Il est considéré par de nombreux auteurs DESJARDINS (1988) ; MASSCHELEIN, (1996) et EDELINE, (1996), comme une base de référence bien utile pour apprécier l'état de coagulation/floculation des particules colloïdales. Cependant, sa mise en œuvre de façon automatique et en continu est encore difficile.

III-1.4. Titration colloïdale

La titration colloïdale, dont le principe est la neutralisation d'un colloïde positif par un colloïde négatif, semble être un paramètre utile pour optimiser la phase de coagulation. Les travaux de Kawmura (1966) et de Verma en (1978) montrent l'intérêt de la titration colloïdale et présentent des modèles reliant le taux de coagulant à la titration colloïdale BAZER-BACHI (1985).

Suite à ses recherches, BAZER-BACHI (1985) montre que parmi les différentes techniques de titration testées, la méthode par indicateur coloré ne peut être appliquée qu'aux eaux fortement chargées alors que la méthode néphélométrique s'avère tout à fait recommandée pour analyser les eaux faiblement chargées. BAZER-BACHI montre également que la mesure de la capacité d'échange cationique CEC au bleu de méthylène est identique à la titration colloïdale en cas des suspensions d'argile. La titration colloïdale permet de déterminer la fraction colloïdale minérale et organique.

III-1.5. Mesure de la turbidité par l'Autofloc.

L'Autofloc est un appareil qui a été développé à partir des résultats obtenus lors des essais au jar test. Le principe de cet appareil repose sur une mesure continue de la turbidité lors des essais de floculation. On a ainsi démontré qu'il existe en batch, une corrélation entre l'accroissement de la taille des floes lors de la phase de formation des floes et l'augmentation de la turbidité, cet appareil a été testé avec succès sur des stations de traitement des eaux.

III-1.6. Comptage des particules

Certains auteurs recommandent l'utilisation du comptage des particules comme critère d'évaluation, celui-ci s'exprime par le nombre de particules par unité de volume de la suspension ; à cet effet, on utilise des compteurs électroniques de particules types COULTER.

KUO (1988a) note que le diamètre moyen en volume et le nombre total de particules sont plus précis que la turbidité dans le contrôle de la floculation, MASSCHELEIN (1996) mentionne que les techniques de comptage et de mesure des particules conviennent assez peu pour l'analyse de suspensions floculées du fait de la nécessité de calibrer l'orifice de passage.

III-2. ANALYSE EXPLORATOIRE MULTIDIMENSIONNELLE DE LA FLOCCULATION

III-2.1. Introduction

L'analyse exploratoire multidimensionnelle et en particulier l'analyse factorielle, l'analyse en composantes principales et l'analyse de classification ont été largement utilisées dans de nombreux domaines se rapportant à l'environnement, pour mettre en évidence les associations entre les individus et/ou les variables, ce sont également de bons outils de réduction de la dimensionnalité d'un ensemble de variables, initialement toutes plus au moins corrélées entre elles.

L'analyse en composantes principales a été utilisée par de nombreux chercheurs afin de mettre en évidence les changements spatial et temporel de la qualité des eaux BENGRAINE (2003), SINGH (2004) et PARINET (2005) et pour simplifier la complexité des relations entre les différentes variables de qualité de l'eau. L'analyse en composante principale a été utilisée pour la prédiction de chlorophylle dans les eaux de barrage ÇAMDEVYREN (2005) et pour la modélisation des concentrations des trihalométhanes PLATIKANOVA (2007).

Cette technique s'est avérée également être un outil tout à fait adapté au développement des indices multivariés. C'est ce qu'a montré PRIMPAS (2009), l'auteur a déterminé l'indice de l'eutrophisation qui est une combinaison linéaire des cinq variables bien corrélées avec le premier axe factoriel d'une ACP. Cet indice est trouvé très efficace dans la classification dudit phénomène en plusieurs niveaux.

Par ailleurs KOKLU (2010) illustre l'importance de l'analyse en composantes principales "ACP", l'analyse factorielle "AF", l'analyse de régression multiple, et l'analyse de classification dans l'évaluation de la qualité des eaux de la rivière Melen (Turquie); L'ACP et l'AF identifient les paramètres importants dans l'estimation des variations de débit alors que l'analyse de régression permet la détermination des relations caractéristiques, cependant l'application de l'analyse de classification donne de meilleurs résultats pour l'étude spatiale et temporelle du phénomène.

L'utilisation combinée de l'analyse en composantes principales et de l'analyse de classification dans les études exploratoires multi-variables des données a fait l'objet de plusieurs travaux; AYOKO (2007) a utilisé l'analyse en composantes principales et l'analyse de classification pour la modélisation exploratoire multivariée et la prédiction des propriétés physico-chimiques des eaux de surface et des eaux souterraines. YIDANA (2008) s'est basé sur la puissance de l'utilisation combinée des techniques multivariées de l'analyse en composantes principales et l'analyse de

classification à la classification des échantillons d'eau du bassin ANKOBRA (GHANA). KAZI (2009) s'est servi de la complémentarité de l'analyse en composantes principales et de l'analyse de classification à l'évaluation de la qualité des eaux du lac Manchar (Pakistan).

III-2.2. Analyse exploratoire multivariée de la floculation

Les techniques statistiques multi-variables comme l'analyse en composantes principales (ACP), et l'analyse factorielle ont été utilisées avec succès dans le domaine du diagnostic et d'analyse des résultats d'essai de floculation. Ce sont des outils capables de compresser les données et de réduire leur dimensionnalité de sorte que l'information essentielle soit conservée et plus facile à analyser que dans l'ensemble original des données, cela permet de décrire correctement les tendances principales du processus.

Nous présentons dans cette section quelques travaux ayant utilisés les techniques multivariées dans l'étude de la floculation, cela montre l'importance de ces utilisations dans l'interprétation, l'analyse, et la prédiction des paramètres.

COLIN (1986) a appliqué l'analyse factorielle des correspondances à des résultats d'essais de traitement des eaux de surface, l'analyse a porté sur une période de trois mois et a traité un total de 387 échantillons, sur chacun, vingt deux paramètres ont été mesurés afin d'apprécier l'efficacité du couplage pré-ozonation –coagulation.

Un objectif modeste en apparence conduit en fait à une alternative difficile : ou bien on restreint le nombre de paramètres interprétés et on risque alors d'en omettre ceux qui présentent un intérêt majeur, ou on cherche à élargir au maximum le champ d'investigation, et on se heurte alors à des difficultés d'interprétation liées à l'excès de données. L'analyse multidimensionnelle des données permet dans le second terme de l'alternative de réduire à un nombre raisonnable les familles des paramètres sur lesquels doit porter l'analyse.

BRODARD (1986) a présenté une étude de modélisation stochastique d'une usine de traitement d'eau potable. L'analyse factorielle de correspondance a permis l'examen d'un nombre important de données (955.585), et d'en réduire les 31 variables représentant la qualité de l'eau et les conditions atmosphériques en 12 paramètres puis en six descripteurs principaux (la transparence (l'inverse de la turbidité), le débit d'eau, la concentration du solide en suspension, la turbidité, les nitrates et la température).

A la base de l'analyse factorielle de correspondance sont déterminés les meilleurs paramètres descripteurs de chacun des cinq modèles élaborés, le tableau 1.4 ci-dessous résume les résultats de détermination des paramètres

d'entrée pour la prédiction de la dose du floculant (sulfate d'aluminium), la concentration des silicates, la concentration de la chaux, la concentration du charbon activé et enfin la dose du chlore.

Paramètre à modéliser	Descripteurs			
Sulfate d'aluminium	Transparence	Débit d'eau	Solide en suspension	Turbidité
Silicates	Débit d'eau	Température		
Chaux	Sulfate d'aluminium	Débit d'eau		
Charbon activé en poudre	Nitrate	Oxydabilité au KMnO ₄		
Dose du Chlore	Demande en chlore	Nitrate		

Tableau 1.4 : Paramètres d'entrée pour la modélisation BRODARD (1986)

JARRET (1986) Caractérise par l'analyse statistique multidimensionnelle descriptive en composantes principales la matière organique dissoute dans les eaux au cours d'affinage en relation avec différentes conditions opératoires de traitement. En effet l'auteur a relié les déterminations physico-chimiques de la qualité des eaux à l'entrée et à la sortie des filtres CAG (Charbon Activé en Granulats) à des paramètres intrinsèques comprenant :

- Le débit, la température et pH de l'eau.
- La dose de coagulant.
- Les taux de préozonation, d'ozonation et d'ozone consommé.
- Le taux d'oxygène dans les eaux d'entrée.
- Le taux d'oxygène consommé en filtration.

Les principaux résultats montrent que la qualité des eaux alimentant les filtres à charbon activé est surtout liée (47% de la variation observée) au taux d'ozonation et à la température. Alors que la qualité des eaux filtrées sur CAG dépend pour 19 % de la température (l'effet de l'ozonation est faible).

Il en résulte aussi que les eaux filtrées sur sable, moyennement ozonées peuvent être caractérisées par trois classes dont le poids moléculaire des composés organiques subsistants après ozonation est inférieur à 5000 Daltons. Alors que les eaux fortement ozonées sont arrangées en deux classes dont le poids moléculaire des composés organiques restants (non détruits par ozonation) est inférieur à 1000 Daltons. JARRET indique, au total, que l'évolution de la matière organique lors de la filtration sur CAG est dicté par deux phénomènes l'adsorption et la biodégradation.

BAZER-BACHI (1990) a utilisé les résultats d'analyse en composantes principales (sur 140 échantillons caractérisés par 13 variables) pour dégager les variables représentatives de la variation de la qualité de l'eau de la Garonne, les paramètres retenus sont : la turbidité (ou MES) la résistivité (ou TAC, TH), la température, la matière organique, la capacité d'échange cationique et le pH.

BAZER-BACHI note aussi, suite à l'étude hydrologique, que les matériaux transportés par la Garonne en période d'étiage sont les illites alors qu'en période de crue ce sont les smectites, ou/et illite et les chlorites.

La méthode systématique d'expérience « dite au coup par coup » de recherche des conditions optimales de floculation et l'évaluation des interactions entre facteurs à été dans plusieurs études larguée au profits des méthodologies de recherche expérimentale en cas ou le nombre de paramètres à étudier est élevé, ces méthodes permettent, avec un nombre d'expériences limités, d'estimer l'importance des paramètres et de leurs interactions GOUPY (1996). Nous présentons dans ce qui suit quelques travaux illustrant l'utilisation des méthodes de plans d'expériences et méthodes de surface de réponses dans l'étude expérimentale de la floculation.

DESJARDINS (2002) a identifié les principaux paramètres de contrôle de la floculation ballastique (basée sur l'addition de micro-sable et polymère durant la coagulation/floculation) à travers un nombre réduit d'expérience, (17 essais jar test) conçus par la méthode de plan d'expérience. Il en résulte que :

- Le dosage de coagulant et le pH présentent l'effet prépondérant dans la variation de la qualité des eaux décantées
- La filtrabilité est trouvée principalement affectée par le rapport "dosage micro-sable/dosage floculant", elle s'améliore pour des rapports élevés.

FRANCESCHI (2002) conclut l'étude d'optimisation de processus de coagulation/floculation de l'eau brute par l'utilisation de plans factoriels fractionnaires et du plan du réseau uniforme de Doehlert (32 essais jar test) par les points suivants :

- Les eaux caractérisées par de faible CEC (capacité d'échange cationique) et un pH autour de 6,3 présentent certaines difficultés en traitement, la dégradation en efficacité de floculation est spécifique au mécanisme de floculation par le sulfate d'aluminium
- L'optimisation de la turbidité résiduelle nécessite moins de paramètres que l'optimisation de l'ajout du coagulant. cependant l'optimisation simultanée apparait impossible à cause d'effet inverse des différents paramètres sur ces deux sorties.

SONMEZ (2007) a appliqué la méthode dite méthode statistique de conception de Box Wilson (méthode de surface de réponse) pour évaluer les effets des variables, comme par exemple la quantité de Na-Oleate ($C_{17}H_{33}COONa$) employée comme substance anionique, la concentration du sel $MgCl_2$ et la vitesse d'agitation, lors de l'étude expérimentale de la floculation de Celestite (le celestite est la principale source du strontium minérale).

L'étude a été réalisée avec un nombre réduit d'expérience (15 expériences seulement), Des modèles polynomiales par régression multiple ont été élaborés pour prédire la qualité de la floculation et le potentiel zéta. Les résultats apparaissent satisfaisants ; les prédictions sont en bonne concordance avec les résultats expérimentaux.

BHATIA (2007) a utilisé le plan expérimental complet de facteurs composites (avec 30 essais jar test) et la méthodologie des surfaces de réponses, pour l'obtention des valeurs optima des paramètres de coagulation/floculation de traitement de POME (palm oil mill effluent), en utilisant l'extrait des graines de *Moringa Oleifera*.

Par le biais de la régression multiple, des modèles polynomiales d'ordre 2, ont été élaborés pour corréliser le solide total et le coefficient de boue avec les quatre variables d'entrée qui sont : le pH, le temps de décantation, le dosage du coagulant (*M. oleifera*), et le dosage du floculant.

YANG (2009) a employé la méthodologie des surfaces de réponses pour l'optimisation de la floculation des suspensions de kaolin avec deux coagulants le PAC (polychlorure d'aluminium) et le coagulant MBFGA1. Deux modèles quadratiques d'optimisation ont été établis pour prédire la taille de floc et l'efficacité de la floculation.

Les résultats montrent que le PAC présente plus d'avantages dans le changement du potentiel du colloïde, le MBFGA1 est bon en adsorption et en pontage alors que le coagulant mixte (PAC+MBFGA1) donne les meilleurs rendements.

Par ailleurs, Plusieurs études WANG (2007), AHMAD (2007), GHAFARI (2009) et ANOUZLA (2009) ont montré l'intérêt de l'utilisation des méthodes de recherche expérimentale à l'optimisation du processus de coagulation/floculation des eaux usées, en effet la méthode des surfaces de réponses permet de dévoiler les interactions entre les paramètres de floculation avec un nombre réduit des expériences. D'un autre coté, l'analyse statistique ANOVA (analyse of variance) a été dûment utilisée par AHMAD (2007), GUIDA (2007) et GHAFARI (2009) à la détermination des facteurs à effets prépondérants sur le processus de floculation.

III-2.3. Conclusion

Cette section bibliographique montre l'utilisation large des techniques multivariées dans l'optimisation de l'expérimentation et l'analyse des données en particulier les résultats d'essais de floculation, ces techniques permettent entre autre de :

- Mettre en évidence les associations entre les paramètres de floculation, compresser les données et de réduire leur dimensionnalité de sorte que l'information essentielle soit conservée.
- Faciliter le diagnostic et l'analyse des résultats.

III-3. MODELISATION DE LA FLOCCULATION

III-3.1. Modélisation mathématique de la floculation

La modélisation mathématique de la floculation a progressé significativement. Au début, c'était l'initiale modélisation de la cinétique d'agrégation et de fragmentation par de simples modèles d'équilibre de population de Smoluchowski, cependant ces dernières décennies plusieurs recherches ont été allouées à l'amélioration de ces modèles en y incorporant d'autres paramètres tenant compte du : le couplage entre les phénomènes physico-chimiques et hydrodynamiques (BOUYER 2002), la structure irrégulière et la perméabilité du floc ainsi que la surface et la chimie du colloïde SOMASUNDARAN (2003). Les propriétés microstructurales de la couche adsorbée et particulièrement la conformation de polymère SOMASUNDARAN (2005). D'autres auteurs se sont attelés à l'étude des modèles de floculation en tenant compte des forces non DLVO, par des approches moléculaires TABOADA-SERRANO (2005) et en faisant incorporer l'attraction par pontage et les forces de répulsion stérique RUNKANA (2006)...etc.

La modélisation mathématique de la floculation ne sera plus explicitée car elle ne sera plus prise en compte dans la suite de l'étude, à contrario dans ce qui suit on détaillera la modélisation statistique de la floculation.

III-3.2. Modélisation statistique par régression multiple de la floculation

L'utilisation de la modélisation statistique pour la détermination du dosage de coagulant en traitement de potabilisation des eaux a fait l'objet de nombreuses études ;

- Gravel (1987) a élaboré un modèle statistique permettant de prédire la dose optimale de l'alun pour le traitement des eaux. L'étude a été menée sur une base de 365 échantillons, le modèle élaboré est de la forme.

$$y = \frac{1313}{C} + 3,66 \times 10 Tu^2 + 1,09 \times 10^{-3} T^2 + \frac{465}{pH} - 46,5 \quad (I-21)$$

Avec

- C : Conductivité électrique [$\mu\text{S}/\text{cm}$]
- Tu : Turbidité [NTU]
- T : Température de l'eau °C
- Y : Dosage de l'alun [mg/l]

Le coefficient de corrélation entre les doses prévues par le modèle et les doses observées et le niveau de confiance sont respectivement $R=0,83$ et 95%

Le modèle de CRITCHLEY en 1990 [DALY 2007] utilise les mesures de la couleur, la turbidité, et le résiduel en coagulant, le modèle se présente sous la forme suite :

$$D = K_1 \times C_{init} \times K_2 \times Tu + K_3 \quad (I-22)$$

D : Dose de coagulant [mg/l]

C_{init} : Couleur colloïdale en unité Hazen

Tu : Turbidité initiale [NTU]

K_1, K_2, K_3 , constantes

K_3 : représente l'influence du pH, de la conductivité, et de la température.

BAZER- BACHI (1990) présente deux modèles mathématiques reliant la dose optimale du coagulant (le sulfate d'aluminium) à la qualité de l'eau brute caractérisée par cinq descripteurs déterminées par analyse en composantes principales. (La turbidité, la résistivité, la température, la teneur en matière organique et la capacité d'échange cationique).

- Le premier modèle, appelé CALME, applicable aux eaux dont la turbidité est inférieure à 20 NTU. Ce modèle, élaboré à partir de 113 échantillons, tient compte de quatre variables caractéristiques : la turbidité, la résistivité, la température et la teneur en matière organique.

$$\begin{aligned} T_{aux} = & 15 + 3,6Tu - 0,4\rho - 4,8T + 4,2MO \\ & + 3,2Tu^2 + 1,9\rho^2 + 4,6T^2 + 3,4MO^2 \\ & - 3,2Tu.\rho - 0,5Tu.T - 4,4Tu.MO \\ & - 0,01\rho.T + 3,1\rho.MO - 0,05T.MO \end{aligned} \quad (I-23)$$

- Le second modèle, appelé CRUE, applicable durant les périodes de crues, (élaboré à partir de 98 échantillons) intègre un cinquième descripteur qui est la nature de la suspension minérale Δ .

$$\begin{aligned}
 Taux = & 66,6 + 40,1Tu + 0,5\rho - 17,2T - 1,2MO + 4,7\Delta \\
 & - 3,5Tu^2 + 1,2\rho^2 + 0,9T^2 - 8,4MO^2 + 3,7\Delta^2 \\
 & + 5,3Tu.\rho - 8,6Tu.T + 1,0Tu.MO + 1,9Tu\Delta \\
 & - 2,0\rho.T - 9,1\rho.MO + 5,7\rho.\Delta \\
 & - 6,2T.MO + 1,2T.\Delta + 6,0MO.\Delta
 \end{aligned}
 \tag{I-24}$$

Avec : Taux : Dose de coagulant [g/m³]

Tu : Turbidité [NTU]

T : température [°C]

ρ : Résistivité (Ω cm)

MO : Teneur en matière organique [m⁻¹]

(mesure UV à 254 nm ramenée à 1m de longueur de cellule)

Δ : Facteur, caractérise la nature de la suspension (Rapport entre la titration colloïdale et la turbidité $\Delta = Tc/Tu$ [10⁻⁴.meq/l/NTU])

- FRANCESCHI (1991 et 1993) a proposé des modèles mathématiques à base de réseau uniforme de Doehrlert en vue d'optimiser le taux de sulfate d'aluminium nécessaire à la floculation d'une eau synthétique.

Les paramètres opératoires retenus sont :

- La vitesse d'agitation (Na) [tours/mn]
- Temps d'agitation (Ta) [mn]
- Temps de décantation (Td) [mn]

Les caractéristiques de l'eau mises en jeu sont :

- Concentration de l'argile (bentonite) (Arg) [NTU]
- Concentration en Ca⁺⁺ introduit sous forme de CaCl₂ [mg/l]

Le taux de sulfate d'aluminium "SA" a été estimé par le modèle polynomiale d'ordre 2 suivant :

$$\begin{aligned}
 SA = & 9,49 - 2,74Na + 4,84Ta - 11,39Ca \\
 & + 6,48Arg + 19,11Na^2 + 12,55Ta^2 + 11,26Ca^2 \\
 & - 7,85Arg^2 - 11,95Na.Ta + 10,39Na.Ca \\
 & - 1,20Na.Arg + 0,49Ta.Ca - 5,19Ta.Arg - 21,71Ca.Arg
 \end{aligned}
 \tag{I-25}$$

Ce modèle a été testé sur trois eaux synthétiques et a permis de déterminer, plusieurs couples de valeurs de vitesse et de temps de décantation (Na, Td) conduisant à l'optimisation de la dose du coagulant.

- MASSCHELEIN (1996) indique qu'en pratique la relation liant la dose aux effets d'un coagulant correspond souvent à l'équation suivante :

$$D = K_1 [Alc] + K_2 C^n \tag{I-26}$$

Où D : Dose de coagulant [g/m³]

[Alc] : Alcalinité initiale de l'eau [g de CaCO₃/m³]

C : Charge du colloïde [10⁴.eq/m³]

K₁, K₂, n : constantes (par ex. K₁=1, K₂=0,33, n=1,06)

HART (1997) présente dans son article un modèle statistique pour décrire l'effet du pH et du dosage du coagulant sur l'élimination de la couleur. Le modèle est à la base des résultats d'essais jar test.

$$C_{\text{finale}} = \frac{C_{\text{brute}}}{1 + k(D - D_{\text{min}})} \quad (I-27)$$

$$k = a + b \times pH + c \times pH^2$$

$$D_{\text{min}} = C_{\text{brute}}(\alpha \times pH - \beta)$$

Où : α, β, a, b, c = constantes [-]

C_{finale} : Couleur finale en unité de Hazen

C_{brute} : Couleur de l'eau brute en unité de Hazen

D : Dose du coagulant [mg/l]

D_{min} : Dose minimale de coagulant nécessaire [mg/l]
à l'élimination de la couleur

k : Facteur de coagulation [-]

LEEUVEN (2005) développe un modèle reliant la dose de coagulant, le type et la concentration de la matière organique dissoute. Ce modèle permet, pour un pH donné, la prédiction de la dose de l'alun afin de maximiser l'élimination de la matière organique et atteindre un taux d'élimination du carbone organique dissous de $\frac{\Delta COD}{\Delta Alun} = 0,015$ (c.à.d. réduction de 0,15 mg/l de COD pour une augmentation de la dose de l'alun de 10 mg).

Le modèle élaboré est sous la forme suivante :

$$Alun = A + B [1 - \exp(-C \times UV \times \log(Col \times 10))] \quad (I-28)$$

Alun : Dose de coagulant en mg/l

UV : Absorbance à 254 nm.

Col : Couleur en unité de Hazen.

A, B, C : Constantes [-]

III-3.3. Modélisation statistique par régression neuronale de la floculation

III-3.3.1 Introduction

En régression neuronale, les prédictions sont à la base de l'établissement des connections entre le ou les sorties (à prédire) et les entrées (facteurs influents). Cette approche ne permet pas la visualisation de relation liant les entrées aux sorties ; néanmoins, il reste capitale la recherche et l'identification des entrées significatives. Plus de détails sur l'aspect théorique de la modélisation par régression neuronale est illustré dans le chapitre 2.

L'importance des modèles neuronaux a été largement illustrée dans de nombreuses applications environnementales nous citons ;

- La prédiction de la performance de traitement des eaux usées GAMAL EL-DIN (2002), ZENGA (2003), HONG (2003), HAMED (2004), SALARI (2005) et ZHENG (2011)
- En traitement des eaux par filtration CONLIN (1997), TEODOSIU (2000) et SAHOO (2005),
- Dans l'analyse de salinité des eaux HUANG (2002), et la prédiction de la qualité des eaux de rivières PARK (2002), BELANGER (2005) et SINGH (2009), des eaux souterraines KUO (2004) et des eaux de barrages CHAVES (2007) et ADELOYE (2009).
- Au développement des algues et à l'eutrophisation KARUL (2000), LEE (2003), ADJEL (2007) et KUO (2007),
- Dans l'étude de la désinfection et la formation des THM, RODRIGUEZ, (1996 et 1999), SERODES (2001) et HAAS (2004),
- A la prédiction de l'élimination de contaminants SHETTY (2003) et de l'élimination de la demande biochimique d'oxygène "DBO" et du carbone organique dissous "COD" AKRATOS (2008)
- A la prédiction de débit et performances des techniques membranaires HILAL (2008) et en étude d'optimisation des systèmes de dessalement TANVIR (2006), GAO (2007) et I.F.F. (2009)

A la lumière de ces travaux, il en ressort que la performance de modèle neuronal dépend étroitement des paramètres suivants :

- le choix du type de réseau,
- la fonction d'activation
- l'algorithme d'apprentissage,
- le nombre de couches cachées,
- la complexité du phénomène et de l'échantillon,
- le nombre de paramètres descripteurs,
- le nombre des itérations.

III-3.3.2 Modélisation par réseaux de neurones des paramètres de floculation.

Nous présentons dans cette section, dans un ordre chronologique, les différents travaux relatifs à la modélisation mathématique de la dose de coagulant par régression neuronale.

GAGNON (1997) développe des modèles neuronaux pour la prédiction de la dose de coagulant, en utilisant pour l'apprentissage la méthode Quasi-Newton avec un arrêt avancée de l'approche pour éviter tout surapprentissage. Les variables descriptives mises en jeu sont la température, le pH, la turbidité et la conductivité des eaux brutes et traitées, avec en plus, le dosage du coagulant actuel.

Cette base de données collectées chaque 5 minutes sur plusieurs années a été la source des échantillons d'apprentissage et de test et a servi à l'élaboration des modèles saisonniers et annuels de prédiction de la dose du coagulant. Le modèle développé présente une erreur de 10-50 mg/l.

VALENTIN (2000) présente dans son travail de thèse la construction d'un réseau neuronal à perceptron multicouche PMC pour la détermination en ligne de la dose optimale du coagulant en fonction de différentes caractéristiques de la qualité de l'eau brute à savoir la turbidité, le pH, la conductivité, la température, l'oxygène dissous, et absorption UV.

Une grande importance a été allouée à la validation et la reconstruction de données manquantes et invalides par l'utilisation des "Cartes auto organisatrice de Kohonen", les données utilisées proviennent de diverses sources.

- des mesures en continu issues de capteurs.
- des analyses faites en laboratoire.
- et des résultats d'essais jar-test.

Ces données en nombre de 260000 échantillons ont subi un prétraitement ce qui a conduit à un ensemble de 3190 vecteurs complets de mesure en supprimant les données invalides et manquantes. Un total de 2230 échantillons (environ 70 %) a été exploité pour l'apprentissage de modèle de prévision de la dose de coagulant à un temps t , le reste (960 échantillons) a été utilisé comme "ensemble de test". Cependant parmi les données d'apprentissage, environ 30 % (670 échantillons) ont été utilisées comme "ensemble de validation" pour optimiser l'architecture.

L'architecture du modèle neuronal élaboré s'écrit sous la forme 6-16-1, six neurones dans la couche d'entrée présentant les descripteurs de l'eau brute, seize neurones dans la seule couche cachée et une sortie qui est la dose du coagulant à prévoir.

Dans le but de faire une comparaison, un modèle de régression linéaire a été également développé. Les valeurs du coefficient de détermination R^2 sont :

- Pour l'ensemble d'apprentissage, $R^2=0,84$ pour le modèle en régression multiple, et $R^2 = 0,985$ pour le modèle de régression neuronale,
- Pour l'ensemble de test ces coefficients valent respectivement $R^2=0,692$ et $R^2=0,974$.

Les valeurs des coefficients de détermination calculées sur l'échantillon d'apprentissage et sur l'échantillon de test montrent que la qualité de prédiction par régression multiple linéaire est inférieure à celle de la régression neuronale, ce qui indique la forte non-linéarité du procédé.

BAXTER (2001), dans son article intitulé « traitement et qualité de l'eau potable, l'utilisation de réseaux de neurones artificiels » met en lumière l'utilité des réseaux de neurones artificiels pour la modélisation de la qualité de l'eau de même que pour la modélisation et le contrôle des procédés de traitement de l'eau potable, ce modèle a été appliqué dans plusieurs usines de production d'eau pour prédire la turbidité et la couleur de l'eau traitée.

MAIER (2004) a développé des modèles de régression neuronale permettant :

1. La prédiction des paramètres de qualité de l'eau traitée : la turbidité, la couleur, et l'absorbance à 254 nm, à partir des caractéristiques des eaux brutes et de la dose de coagulant utilisé. Les performances de ces modèles sont très bonnes et les coefficients de détermination varient de 0,90 jusqu'à 0,98.
2. Le contrôle automatique et la prédiction (par modèle inverse) de la dose optimale de l'alun à la base des caractéristiques des eaux brutes et des paramètres de qualité de l'eau. Le modèle présente un coefficient de détermination de 0,94.

MIRSEPASSI (2004) a appliqué le système d'intelligence artificielle (modèle de réseau de neurones) à l'opération de coagulation/floculation de particules présentes dans l'eau par l'ajout de l'alun et d'un polymère non-ionique, deux tiers de l'ensemble de données (1820 jours de données) ont été allouées à l'apprentissage et 1/6 des données pour le test.

MIRSEPASSI propose également d'utiliser l'historique des différents paramètres influençant la performance de floculation, c'est-à-dire de prendre en compte les valeurs des paramètres aux instants $t-1$, $t-2$, ..., $t-6$ (t représente le jour courant) ce qui a donné 47 descripteurs, seulement 9 entre eux ont été retenus après prétraitement des données. La performance du processus a été nettement améliorée, les coefficients de corrélation entre les valeurs des dosages observées et celles prévues sont pour les deux réactifs (alun et polymère) de l'ordre de 0,97. L'erreur moyenne absolue est respectivement 4,09 % pour l'alun et 8,76 % pour le polymère. L'auteur mentionne que cette bonne performance revient à la qualité et à l'intégrité des données fournies pour l'apprentissage.

WU (2005) propose l'application du modèle de réseau de neurones pour la détermination, en temps réel, de la dose du coagulant à injecter, en cas de situation inhabituelle (fortes pluies engendrant une élévation rapide de la turbidité), dans une telle situation il est souvent difficile à corriger, à répondre et à s'adapter à la bonne posologie en temps opportun par un mode de fonctionnement classique.

Les paramètres choisis dans ce programme de réseau de neurones sont : la turbidité de l'eau brute et de l'eau après clarification, la couleur de l'eau brute et de l'eau après clarification, ainsi que le dosage de la coagulation précédente. Pour la prédiction de la dose du lendemain, l'auteur utilise un modèle d'auto prédiction. Les coefficients de détermination pour le premier et le deuxième modèle sont respectivement $R^2=0,91$ et $R^2=0,94$.

LAMRINI (2005) élabore un modèle neuronal pour la prédiction en ligne de la dose du coagulant en traitant 2511 échantillons (traduisant les données expérimentales de quatre années d'échantillonnage). 3/4 de l'échantillon a été utilisé pour l'apprentissage et le 1/4 restant est employé pour la validation du modèle.

Parmi les différents descripteurs de la qualité de l'eau brute étudiés, figurent la température, le pH, la turbidité, le solide total, l'oxydabilité, l'oxygène dissous et la dose du coagulant (paramètre passive). L'analyse multidimensionnelle par composantes principales a permis d'extraire seulement cinq variables significatives et qui sont : le pH, la turbidité, le solide total, l'oxygène dissous, et la conductivité.

L'architecture du réseau adopté consiste en cinq variables d'entrée, une seule couche cachée avec 18 neurones sigmoïdaux et un seul neurone linéaire formant la couche de sortie. Le modèle présente de bonne performance, le coefficient de corrélation entre les valeurs prévues par le modèle et les valeurs observées (échantillon de validation) est de l'ordre de 0,94.

HECTOR (2006) dans son travail de thèse sur la supervision et le diagnostic des procédés de production de l'eau potable a utilisé la méthodologie de classification LAMDA (Learning Algorithm for Multivariate Data Analysis) pour la construction du modèle à partir des données historiques, ce modèle est caractérisé par un ensemble de classes permettant d'identifier un ensemble de situations correspondant aux états fonctionnels de la station. La modélisation de la dose du coagulant à passer par les étapes suivantes :

- Présentation des données

L'ensemble des données collectées couvrant 2 années de fonctionnement est constitué de 725 échantillons, chacun de ces échantillons possède 09 descripteurs de la qualité de l'eau brute : la température, la couleur, la turbidité, le solide total, la matière organique, le pH, les bicarbonates, le chlorure, et la dureté totale.

- Sélection des variables d'entrée du modèle en utilisant l'ACP

L'interprétation de l'histogramme des valeurs propres et du cercle de corrélation a permis de restreindre le nombre des variables d'entrée à celles les plus pertinentes (les 09 descripteurs en 04 seulement) ; la turbidité, la dureté totale, la température et le pH

- Prédiction de la dose du coagulant.

L'ensemble des échantillons a été divisé en deux, un total de 363 échantillons pour l'apprentissage, le reste, soit 362 échantillons, a été utilisé comme ensemble de test. Il en ressort que l'architecture du perceptron multicouche offrant la meilleure simulation est : 04 entrées, 20 neurones dans la couche cachée et une sortie. Par comparaison au modèle de régression linéaire, le modèle de réseau de neurones artificiels (RNA) présente des coefficients de détermination nettement supérieures à ceux donnés par régression linéaire, (0,96 à 0,97 pour la régression neurale contre 0,61 à 0,72 pour la régression linéaire).

LAMRINI (2007) propose la méthodologie de classification par apprentissage (LAMDA) pour la supervision et le diagnostic d'une unité de coagulation, l'application de cet algorithme de classification donne 10 classes issues de l'auto-apprentissage et cinq états de fonctionnement qui permettent d'identifier, sur l'unité de coagulation, les principales dérives de fonctionnement. L'étude par "analyse en composantes principales (ACP)" a permis de déterminer cinq paramètres principaux à prendre en considération lors du développement du capteur logiciel, et qui sont : la température, le pH, les MES, la conductivité et l'oxygène dissous LAMIRINI (2005).

En fait, dans cette étude un sixième descripteur a été ajouté ; la dose du coagulant. Trois régressions neuronales ont été développées pour prédire la dose de coagulant, les entrées communes des modèles sont le pH, la température, l'oxygène dissous et la conductivité ; pour le réseau 1 le cinquième descripteur est les MES; pour le second, on admet la turbidité (si turbidité est inférieure à 1000 NTU) et pour le réseau 3 on utilise la turbidité par interpolation avec MES (si turbidité est supérieur à 1000 NTU).

Ce travail montre la performance de la technique de classification LAMDA associée à l'analyse en composantes principales et aux réseaux de neurones dans la gestion supervisée d'une unité de coagulation en traitement des eaux de surface.

Dans le but de minimiser la formation des DBP (Disinfection by-Products) et maximiser le résiduel à la désinfection pour la maintenance des réseaux de distribution, DALY (2007) a cité des modèles de prédiction des doses de l'alun et du chlorure ferrique permettant le maximum d'élimination de la

matière organique. Ces modèles mathématiques utilisent comme entrées les descripteurs usuels de l'eau brute : l'absorbance UV à 254nm, la couleur à 456 nm, la turbidité et l'alcalinité.

Ces modèles permettent, en plus, la détermination des doses du coagulant nécessaires pour atteindre des taux fixes d'élimination du carbone organique dissous.

WU (2008) propose deux modélisations afin de surmonter les limites d'utilisation de jar test (la consommation de temps, et l'inadaptation au changement en temps réel de la qualité de l'eau brute), la première est à la base du réseau de neurones artificiels "RNA" alors que la deuxième est à la base du réseau adaptatif du système de déduction floue "ANFIS", la base de données de 819 échantillons a été divisée en deux ; 699 échantillons pour l'apprentissage et 120 pour le test.

La sortie du modèle est le dosage du PAC (polychlorure d'aluminium) en temps réel alors que les entrées du modèle sont : les dosages précédents (d'hier et d'avant hier) PAC (t-1) et PAC(t-2), la température, la turbidité, la couleur, et le pH de l'eau brute, de l'eau floculée, de l'eau décantée, et de l'eau traitée.

Pour les deux réseaux ANN et ANFIS, 17 modèles, rangés dans quatre groupes, ont été testés par combinaison des entrées suivantes ; les dosages précédents PAC(t-1) et PAC (t-2), la turbidité et la couleur de l'eau brute et la turbidité et la couleur de l'eau floculée. Il en résulte que le meilleur modèle de prédiction du dosage optimal de coagulant est obtenu là où les entrées du modèle sont : la turbidité de l'eau brute et le dosage PAC (t-1). Il apparaît aussi que le modèle du réseau de neurones ANN est meilleur que le réseau ANFIS en prédiction de dosage optimal en temps réel lorsque la turbidité est élevée.

HEDDAM (2008) élabore un modèle de régression neuronale pour la prédiction de la dose de l'alun à injecter pour le traitement des eaux de surface du barrage de Keddara.

La base de données utilisées, recueillies pendant deux ans (soit 726 enregistrements), est scindée en deux parties : 2/3 pour l'apprentissage et 1/3 pour la validation. Les variables d'entrée étudiées sont : la température, le pH, la conductivité, l'absorbance à UV, l'oxygène dissous et la turbidité. Les résultats de la modélisation montrent qu'il n'y a pas de descripteur prépondérant et que l'architecture optimale du réseau de neurones est obtenue en faisant intégrer tous ces six descripteurs de l'eau brute.

Le modèle retenu est sous la forme 6-13-1 ; six variables formant la couche d'entrée, treize neurones de l'unique couche cachée et une sortie représentant la dose de coagulant à estimer, le coefficient de corrélation

entre les valeurs calculées et celles mesurées vaut respectivement pour l'ensemble d'apprentissage et l'ensemble de test $R^2=0,80$, et $R^2=0,75$.

AL-ABRI (2008) utilise un modèle de réseau de neurones à rétro-propagation pour la modélisation de la coagulation des substances humiques et la modélisation de l'agglomération des substances humiques et métaux lourds. Les paramètres d'entrée sont : le type de la membrane, niveau de salinité, concentration des substances humiques et métaux lourds, la pression transmembranaire et le type et la concentration du poly-électrolyte. Les données d'apprentissage ont été aléatoirement déterminées et l'analyse par composantes principales a permis la réduction des entrées du réseau de neurones.

AL-ABRI mentionne que les meilleurs résultats sont obtenus avec 2 à 10 neurones en agglomération des substances humiques et métaux lourds et de 5 à 15 neurones pour la coagulation des substances humiques. Les résultats de cette simulation montrent aussi qu'il existe un nombre optimal de données d'apprentissage, un nombre inférieur mène au sous-apprentissage et un nombre supérieur conduit au sur-apprentissage.

ROBENSON (2009) montre que le choix et la préparation des données permet d'élaborer un modèle inverse de réseau de neurones pour la prédiction du dosage optimum du coagulant, permettant de remplacer la technique conventionnelle du jar test.

L'architecture dudit modèle est 11-27-9-1; onze entrées, deux couches cachées comprenant respectivement 27 neurones et 9 neurones et une sortie unique. Ce modèle construit présente pour l'échantillon d'apprentissage un coefficient de corrélation de 0,95, l'erreur quadratique moyenne est de 0,0019 et l'erreur absolue moyenne 0,024 mg/l.

AL-ABRI (2010) reprend la simulation, par réseau de neurones artificiels, de la coagulation des substances humiques et métaux lourds, les entrées du modèle sont : la salinité, la concentration des substances humiques, la concentration des métaux lourds, et la concentration et types de polyélectrolyte. Dans cette étude l'auteur teste cinq métaux lourds (Cu, Ni, Co, Zn et Cd) et deux types de polymère. Plusieurs architectures pour différentes données d'apprentissage et différents nombres de neurones dans la couche cachée ont été testées.

La simulation montre l'existence d'un nombre optimal de données d'apprentissage, et un nombre optimal de neurones dans la couche cachée ; pour des nombres inférieures aux nombres optimaux c'est "le sous-apprentissage" et pour des nombres supérieures, le réseau mémorise les données d'entrée, c'est "le sur-apprentissage".

Les différents modèles de régression multiple et de régression neuronale cités plus haut, laisse présager que les paramètres descripteurs de la régression, la performance, la qualité d'élaboration et de prédiction des modèles sont tributaires de la disponibilité, de l'importance, de la complexité et de l'exactitude des données.

IV. CONCLUSION DU CHAPITRE

Ce chapitre aura permis de présenter dans un premier temps, un état des connaissances sur les suspensions colloïdales, de constater la nature et les caractéristiques très diverses de ces composés et de rappeler les principaux mécanismes de la coagulation et de la floculation.

Les nouvelles révélations de la pratique de la coagulation et de la floculation des suspensions colloïdales ont fait l'objet de la deuxième partie de ce chapitre, en effet plusieurs aspects ont retenu l'attention des chercheurs : la formation et le cassage de floc et les effets de la nature et des charges des colloïdes, des conditions de mélange, du pH, de la température, de la minéralisation et des différents réactifs chimiques ajoutés à la suspension sur la performance de la floculation. La plupart de ces travaux s'accordent sur :

- La complexité des mécanismes mis en jeu lors de la coagulation et de floculation
- L'optimisation des paramètres de coagulation et de floculation peut conduire à une floculation performante.

La modélisation de la phase de floculation a fait l'objet de la troisième partie de cette recherche bibliographique, en fait un bref aperçu sur les appareils et méthodes utilisés pour la détermination de la dose optimale de coagulant a été exposé.

Ce chapitre a permis de mettre aussi l'accent sur l'importance de l'application de l'analyse exploratoire multidimensionnelle à l'identification des corrélations et liaisons entre les variables et le résumé de l'information dans un nombre restreint de données, Il s'est avéré que la régression multiple et la régression neuronale permettent la mise au point des modèles multivariés de prédiction des paramètres de qualité des eaux et du dosage de coagulant, en outre la combinaison de l'analyse exploratoire et des méthodes de régression permet une meilleure optimisation du processus de coagulation/floculation.

L'aspect théorique de l'analyse multidimensionnelle et de la modélisation des données fera l'objet du deuxième chapitre de ce manuscrit.

Chapitre 2

ANALYSE ET MODELISATION DES DONNEES

I. INTRODUCTION

Toute étude de modélisation mathématique des données doit être commencée par une première phase d'analyse exploratoire destinée à faire des représentations, des réductions et de complémentarité de l'information, contenue dans des tableaux de données numériques, en vue d'identifier les problèmes liés à la distribution des variables destinées à être intégrées dans les équations de régression et d'identifier aussi les éventuelles liaisons existantes entre ces mêmes variables. La deuxième phase du travail sera, à partir des résultats obtenus lors de l'analyse exploratoire, de déterminer suivant des procédures, des critères de choix et de qualité d'estimation, les "meilleurs" modèles de prédiction. La démarche d'analyse et de modélisation des données à suivre dans notre cas d'étude est présentée à la figure 2.1

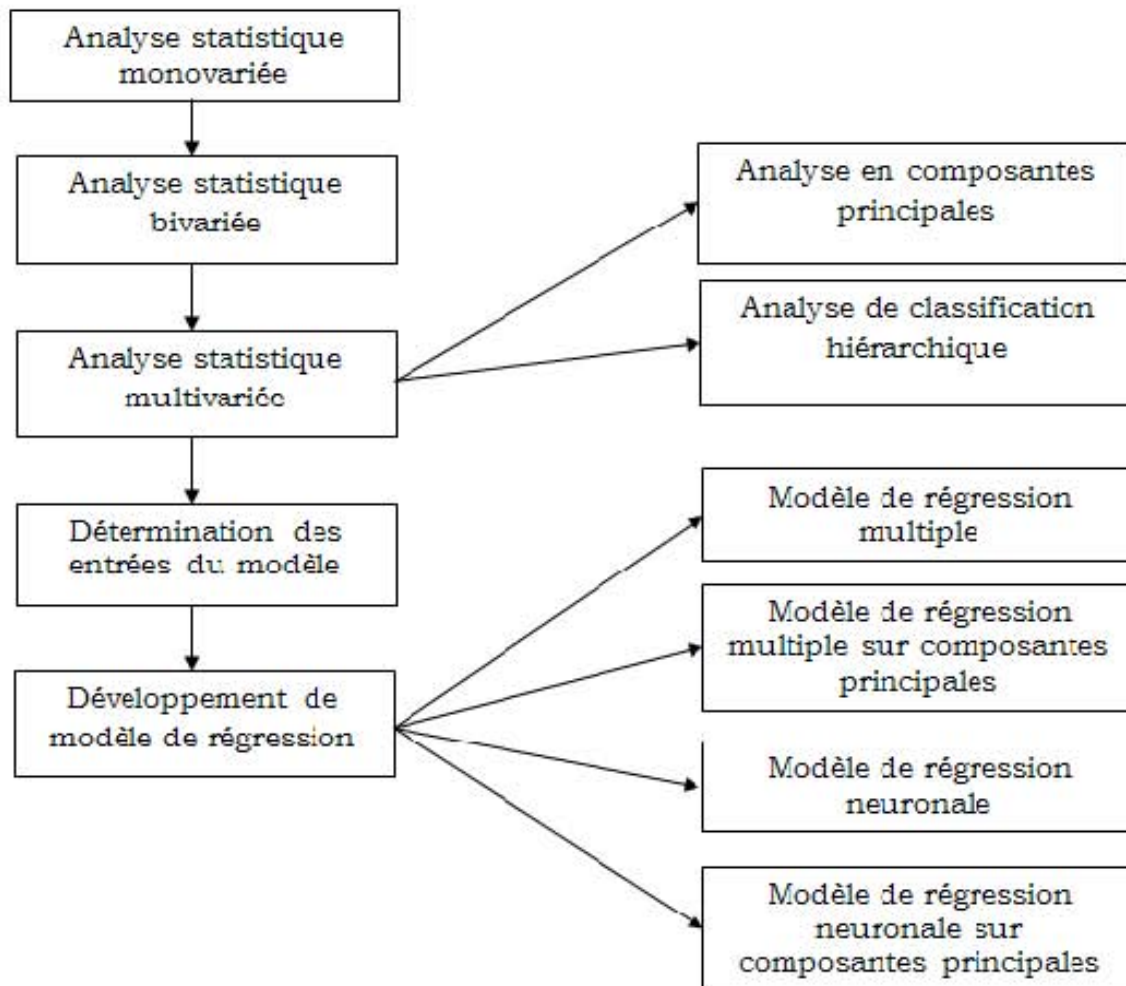


Figure 2.1 : Démarche d'analyse et de modélisation suivie dans notre cas d'étude

La population de données pour les modèles d'inférence est le résultat des essais jar test de floculation des suspensions de particules solides, les caractéristiques de ces particules solides, le protocole expérimental réalisé ainsi que la présentation de l'évolution, de l'influence et de l'interaction des

paramètres et de la cinétique de floculation fera l'objet de présentation et de discussion au chapitre trois. Les sections qui suivent présentent les principes des méthodes statistiques utilisées.

II. PRINCIPE GENERAL DE L'ANALYSE MULTIDIMENSIONNELLE

II-1. INTRODUCTION

Lorsqu'on recueille des informations sur des individus ou unités statistiques, on aboutit à la constitution d'un tableau individus-variables. Pour décrire ces données, si elles sont nombreuses, le statisticien traitera d'abord les variables une par une (traitement univarié), puis il s'intéressera aux éventuelles interactions entre deux variables (traitement bivarié) voire plus (traitement multivarié), et en fin la recherche des prétraitements des données afin de les rendre conformes aux techniques de modélisation.

- transformation : logarithme, puissance.
- Réduction de dimension, classification et premier choix de variables,

Analyse monovariée

Dans un premier temps, la préparation et l'exploration des données pour la modélisation statistique consiste à identifier les valeurs manquantes, erronées ou atypiques, modalités trop rares, distributions "anormales" (dissymétrie, multi modalité), incohérences, liaisons non linéaires, et mettre en forme ces données ; les données incorrectes, (les valeurs manquantes, erronées ou atypiques, modalités trop rares, distributions "anormales" (dissymétrie, multi modalité), incohérences, liaisons non linéaires, et mettre en forme ces données ; les données incorrectes, (les valeurs manquantes, aberrantes ou extrêmes) sont ainsi remplacées ou même supprimées, dans le but d'avoir le moins d'influence "parasite" possible.

Analyse bivariée

Dans un deuxième temps, l'étude des corrélations par les coefficients de Bravais-Pearson "R" et par l'observation des nuages de points permet d'étudier les liaisons entre les variables prises deux à deux. Ces méthodes ne sont pas suffisantes pour la détection de liaisons éventuelles entre les variables considérées simultanément. Pour pallier à ce manque, nous procédons à l'analyse multivariée.

Analyse multivariée

Les techniques d'analyse exploratoire multidimensionnelle sont nombreuses, le choix de la méthode appropriée dépend principalement du type de problème et de la nature des données. Ces techniques se subdivisent en deux grandes familles qui sont les méthodes factorielles et de classification. En effet, les analyses statistiques exploratoires consistent à transformer les données originales par une transformation optimisant un critère géométrique ou statistique. Les méthodes factorielles les plus utilisées sont celles utilisant des statistiques de second ordre, comme l'Analyse en Composantes Principales (ACP).

II-2. REPRESENTATION GEOMETRIQUE ELEMENTAIRE D'UN TABLEAU DE DONNEES

Le tableau de données dispose la masse d'information sous forme rectangulaire, les lignes peuvent représenter les individus ou observations, appelés plus généralement unités statistiques, les colonnes sont alors les p variables qui peuvent être des mesures numériques ou des attributs ou caractères observés sur les individus (cas des variables nominales).

Afin de comprendre le principe des méthodes de la statistique exploratoire multidimensionnelle, il est utile de représenter géométriquement les n lignes et les p colonnes du tableau de données par des points dont les coordonnées sont précisément les éléments de ce tableau (voir figure 2.2) et de représenter en outre, les nuages de points formés par les lignes et les colonnes dans leurs espaces respectifs, cela permet de relever, à partir de la forme de ces nuages, des structures dans l'information.

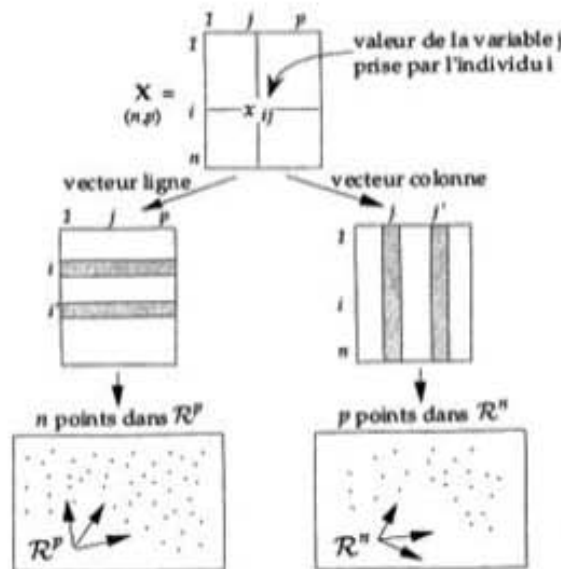


Figure 2.2 : Principe de présentation géométrique d'après LEBART (1997).

Ces représentations géométriques du tableau de données nous conduisent naturellement à utiliser les notions d'espaces vectoriels, de nuages de points, de métrique, mais aussi de masses affectées aux points si l'on ne leur accorde pas la même importance dans le nuage LEBART (1997).

III. ANALYSE EN COMPOSANTES PRINCIPALES

III-1. INTRODUCTION

L'objectif de L'ACP (traitement multivarié) est de revenir à un espace de dimensions réduites en déformant le moins possible la réalité. Il s'agit donc d'obtenir le résumé le plus pertinent possible des données initiales. C'est la matrice des variances-covariances (ou celle des corrélations) qui va permettre de réaliser ce résumé pertinent, parce qu'on analyse essentiellement la dispersion des données considérées.

De cette matrice, on va extraire, par un procédé mathématique adéquat, les facteurs que l'on recherche, en petit nombre. Ces derniers vont permettre de réaliser les graphiques désirés dans cet espace de petites dimensions (le nombre de facteurs retenus), en déformant le moins possible la configuration globale des individus et variables initiales.

C'est l'interprétation de ces graphiques qui permettra de comprendre la structure des données analysées. Cette interprétation sera guidée par un certain nombre d'indicateurs numériques et graphiques, appelés "aides à l'interprétation".

III-2. PRINCIPES DE L'ANALYSE EN COMPOSANTES PRINCIPALES.

Pour observer sous un angle plus favorable les données contenues dans le tableau des données X (figure 2.2), on remplace les anciens axes (donc les anciennes variables) par de nouveaux axes (donc par des variables nouvelles).

Ces nouvelles variables sont appelées composantes principales; elles s'expriment comme des combinaisons linéaires des anciennes variables.

Les nouveaux axes, appelés axes factoriels, sont choisis de la façon suivante :

1^{er} axe factoriel, ou axe principal d'inertie, est la direction de "plus grand allongement" ou "de plus grande dispersion" ou "de plus grande inertie".

Lorsque on projette les points P_i du nuage sur cet axe, leurs projections H_i sont plus dispersées qu'elles ne le seraient sur n'importe quel autre axe. L'axe factoriel F_1 est donc l'axe selon lequel est préservé, par projection, le maximum de la dispersion initiale des points du nuage.

Le fait que le nuage soit allongé précisément dans cette direction doit trouver une explication. La nouvelle variable C_1 (la composante principale n°1) est le caractère selon lequel les individus se différencient le plus.

Le 2^{ème} axe factoriel est la 2^{ème} direction d'allongement du nuage, c'est-à-dire celle qui explique, après le 1^{er} axe, le maximum de l'inertie résiduelle. De plus le 2^{ème} axe est choisi orthogonal au 1^{er}, ce qui traduit le fait que la 2^{ème} composante principale est non corrélée à la première. Comme précédemment, on cherchera à donner un sens à cette 2^{ème} composante principale, et ainsi de suite, jusqu'à avoir remplacé les m anciens axes par n nouveaux axes (les axes factoriels).

Mathématiquement, la détermination des axes factoriels se fait par diagonalisation de la matrice de variances-covariances, d'où le vocabulaire utilisé (valeurs propres, vecteurs propres).

1) Inertie expliquée par les axes factoriels

Le pourcentage d'inertie (ou "variance" du nuage ou "dispersion") expliquée par un axe factoriel permet d'évaluer en quelque sorte la quantité d'information recueillie par cet axe. Notons que l'inertie expliquée par un axe est égale à la valeur propre correspondante et que l'inertie totale (somme des valeurs propres) est égale au nombre de variables de départ dans le cas d'une analyse sur données centrées-réduites.

La qualité de la représentation des données par un plan factoriel s'évalue en ajoutant les pourcentages d'inertie expliquée par les 2 axes.

Il faut en général plusieurs facteurs pour expliquer une part significative de la dispersion. S'il n'y avait pas de direction privilégiée d'allongement du nuage, chaque axe factoriel porterait une part identique de la dispersion: 100 % divisé par le nombre p de variables. Le cas le plus intéressant est évidemment celui où avec un petit nombre d'axes on explique une part importante de l'inertie.

2) Nombre de composantes principales à retenir

Il est difficile de donner une réponse générale à la question : à partir de quel pourcentage peut-on négliger les composantes principales restantes. La qualité des estimations auxquelles conduit l'ACP dépend, de façon évidente, du choix du nombre de composantes à retenir pour reconstituer les données, ou encore de la dimension du sous-espace de représentation BESSE (2005a). Le choix de ce sous espace est une étape crucial pour l'interprétation des résultats, ceci peut conduire à une perte d'information (sous estimation) ou à l'introduction du bruit (surestimation) DRAY (2008). De nombreux critères de choix ont été proposés dans la littérature. Nous présentons, dans ce qui suit, les plus courants BESSE (2005 a).

Part d'inertie

La "qualité globale" des représentations est mesurée par la part d'inertie expliquée : Le nombre de composantes est choisi de sorte que cette part d'inertie expliquée soit supérieure à une valeur seuil, fixée à priori par l'utilisateur.

Règle de Kaiser

On ne considère que, les composantes dont les valeurs propres associées sont plus grandes que 1. Ce critère a tendance à surestimer le nombre de composantes pertinentes.

Eboulis des valeurs propres

Le principe consiste à rechercher, s'il existe, un "coude" dans le graphe présentant la décroissance des valeurs propres, et de ne conserver que les valeurs propres jusqu'à ce coude.

Boîtes à moustaches des composantes principales

Le graphique présente, en parallèle, les boîtes à moustaches des composantes principales qui illustrent bien leurs qualités :

Stabilité : lorsqu'une grande boîte est associée à de petites moustaches,

Instabilité : en présence d'une petite boîte, de grandes moustaches et de points isolés. (Voir exemple de présentation par boîte à moustache au chapitre 04 à la figure 4.1).

III-3. DEMARCHE D'INTERPRETATION D'UNE ACP

Les composantes principales s'expriment comme combinaisons linéaires des anciennes variables, les coefficients de ces combinaisons définissent les nouveaux axes où ils permettent de calculer les nouvelles coordonnées d'un point-individu à partir des anciennes et ils permettent également de voir le poids d'une ancienne variable dans la définition d'un facteur. Pour donner une signification aux axes et aux composantes principales retenues, on examine le nuage des points -variables et le nuage des points-individus BOUROCHE (1980) et BESSE (2005 a).

Les points-variables

Le résultat fondamental concernant les variables est les corrélations linéaires variables-facteurs (composantes principales), ce sont ces corrélations qui vont permettre d'interpréter les composantes principales par l'examen du cercle des corrélations ;

A chaque point-variable, on associe un point dont la coordonnée sur un axe factoriel est une mesure de la corrélation entre cette variable et le facteur, les points-variables s'inscrivent dans un cercle de rayon 1 (le cercle des corrélations) et sont d'autant plus proche du bord du cercle que le point variable est bien représenté par le plan factoriel.

- L'angle entre 2 point-variables, mesuré par son cosinus est égal au coefficient de corrélation linéaire entre les 2 variables:
- si les points sont très proches (α peu différent de 0) : $\cos \alpha = r(X1, X2) = 1$ donc X1 et X2 sont très fortement corrélés positivement
- si α est égal à 90° , $\cos \alpha = r(X1, X2) = 0$ alors pas de corrélation linéaire entre X1 et X2
- si les points sont opposés, α vaut 180° , $\cos \alpha = r(X1, X2) = -1$: X1 et X2 sont très fortement corrélés négativement

Le cercle des corrélations permet de voir, parmi les anciennes variables, les groupes de variables très corrélées entre elles (Voir exemple de projection des variables dans le cercle des corrélations au chapitre 04 à la figure 4.9).

Les points-individus

La qualité de la représentation d'un point M par un axe U dépend de sa distance à l'axe dans le nuage, mesurée par l'angle (OM, U), ou plus exactement par son cosinus ou son \cos^2 . (S'il est proche de 1 le point est bien représenté).

La qualité de la représentation d'un point M par un plan factoriel constitué de 02 axes est mesurée par la somme des \cos^2 avec 2 axes (Pythagore!). La position d'un point-individu par rapport à un axe factoriel, ainsi que les proximités entre les individus, peuvent être interprétées dès lors que ces points sont bien représentés par le plan factoriel observé. Certains individus seront bien représentés par le plan 1-2 (les "très forts" ou "très faibles" en facteur 1 et 2 surtout), d'autres par le plan 1-3 s'ils sont mieux décrits par l'axe 3...etc.

IV. ANALYSE DE CLASSIFICATION

Les méthodes de classification ont pour but de regrouper les individus en un nombre restreint de classes homogènes, il s'agit donc de décrire les données en procédant à une réduction du nombre des individus, les classes seront obtenues au moyen d'algorithme formalisé et non par des méthodes subjectives ou visuelles faisant appel à l'initiative du praticien.

On distingue deux grands types de méthodes de classification :

- Les méthodes non hiérarchiques qui produisent directement une partition en un nombre fixé de classes ;
- Les méthodes hiérarchiques qui produisent une suite de partition en classes de plus en plus vaste à l'image des célèbres classifications des zoologistes en espèces, genres, familles ordre...etc. BOUROCHE (1980).

IV-1. CLASSIFICATION NON HIERARCHIQUE

Il s'agit de regrouper n individus en k classes de telle sorte que les individus d'une même classe soient le plus semblables possible et que les classes soient bien séparées. Ceci suppose la définition d'un critère global mesurant la proximité des individus d'une même classe et donc la qualité d'une répartition. Si on dispose d'un tel critère on pourrait imaginer d'examiner toutes les partitions possibles et de choisir la meilleure. Cette tâche devient impossible en réalité, même avec les plus gros ordinateurs dès que le nombre des individus dépasse quelques dizaines (pour 14 individus il ya plus de 10 millions de partitions possibles en quatre classes) BOUROCHE (1980).

Il est donc à peu près exclu de trouver la meilleure partition possible et il faudra se contenter des algorithmes, comme par exemple l'algorithme de

l'agrégation autour de centres mobiles ou l'algorithme des nuées dynamiques, aboutissant à des solutions approchées minimisant l'inertie intraclasse et maximisant l'inertie interclasse BOUROCHE (1980).

IV-2. CLASSIFICATION ASCENDANTE HIERARCHIQUE

Il s'agit de regrouper itérativement les individus, en commençant par le bas (les deux plus proches) et en construisant progressivement un arbre ou un dendrogramme, regroupant finalement tous les individus en une seule classe, à la racine. On parle de classification hiérarchique car chaque classe d'une partition est incluse dans une classe de la partition suivante. Ceci suppose de savoir calculer, à chaque étape ou regroupement, la distance entre un individu et un groupe ou la distance entre deux groupes. Ceci nécessite donc, pour l'utilisateur de cette méthode, de faire un choix supplémentaire : comment définir la distance entre deux groupes et quel est le critère d'agrégation à retenir. BOUROCHE (1980).

IV-3. CHOIX D'UNE DISTANCE ENTRE INDIVIDUS STATISTIQUES

Ce choix est fonction de la nature des variables et des intentions de l'utilisateur DUMOLARD (2007):

- distance du χ^2 sur des tableaux de contingence
- distance euclidienne multivariée pour des variables quantitatives
- distance euclidienne au carré (pour « éloigner » les individus à valeurs très différentes),
- distance rectilinéaire (« de Manhattan ») pour, les « rapprocher »,
- autre combinaison de différences partielles

IV-4. CHOIX D'UN CRITERE D'AGREGATION

L'utilisateur devra choisir une des variantes suivantes DUMOLARD (2007):

- **plus proche voisin** : la distance (multivariée) entre classes C_i et C_j est la plus petite distance séparant un individu C_i d'un individu C_j (les groupes créés avec cette stratégie sont de longues chaînes, peu homogènes),
- **diamètre maximum** : la distance (multivariée) entre C_i et C_j est la plus grande distance séparant un individu C_i d'un individu C_j (stratégie fonctionnant bien si le nuage de points est constitué de groupes bien distincts et fonctionnant mal si ces groupes sont « allongés »),
- **distance moyenne entre tous les individus de C_i et de C_j** (méthode efficace si les groupes sont bien distincts dans le nuage de points) ; cette distance moyenne entre groupes peut être pondérée (par leur taille) ou non.
- **distance au centre de gravité des groupes** (pondérée ou non par leur taille), méthode bien adaptée aux cas de groupes de taille très différente,
- **méthode de Ward** : elle cherche à minimiser la variance intraclasse (donc à maximiser la variance interclasses) pour créer des groupes homogènes les plus différents possibles les uns des autres.

Dans notre présente étude nous nous limiterons à la classification hiérarchique, nous procéderons à plusieurs classifications selon les différents critères d'inertie, d'agrégation et de mesure de distance entre individus. Le nombre de classes est déterminé à posteriori, à l'allure du dendrogramme (voir exemple de dendrogramme au chapitre 04 à la figure 4.15).

Si les hiérarchies complètes sont en générale différentes, il ne doit pas y avoir de trop grandes variations lorsque l'on regarde uniquement le haut de l'arbre, c'est-à-dire les partitions à faible nombre de classes. Si nous constatons de grosses différences c'est peut être que les variables se prêtent mal à toute classification.

IV-5. COMPLEMENTARITE ENTRE ANALYSE FACTORIELLE ET ANALYSE DE CLASSIFICATION

Les méthodes factorielles sont particulièrement bien adaptées à l'exploration de grands tableaux de données mais elles ne suffisent pas toujours à fournir une vue satisfaisante de l'ensemble des données, les techniques de classification peuvent compléter et nuancer les résultats des analyses factorielles, la complémentarité entre analyse factorielle et analyse de classification concerne la compréhension de la structure des données dans la phase d'interprétation des résultats. La complémentarité peut être marquée dans les cas suivants LEBART (1997) :

1) Difficulté d'interprétation et compression excessive des données

Les visualisations en analyse factorielle sont généralement limitées à très peu de dimensions alors que le nombre d'axes significatifs peut être bien supérieur, cette compression peut entraîner des distorsions et des déformations, aussi, Il est toujours difficile d'interpréter les axes ou les plans factoriels au delà du plan principal. On complète alors l'analyse factorielle pour une classification réalisée sur l'espace tout entier ou de sous espace défini par les premiers facteurs les plus significatifs. Les classes prennent en compte la dimension réelle du nuage de points, elles corrigent donc certaines déformations dues à l'opération de projection.

2) Manque de robustesse

Les visualisations en cercle des corrélations peuvent manquer de robustesse, un point profil aberrant peut notablement influencer le premier facteur et par là toutes les dimensions suivantes alors que les algorithmes des techniques de classification sont robustes et sont indépendants des éventuels points marginaux isolés.

3) Allègement des graphiques factoriels

Les visualisations peuvent concerner des centaines de points et donner lieu à des graphiques chargés ou illisibles il paraît donc utile de procéder à des regroupements d'individus en familles homogènes, les classes peuvent être utilisées pour clarifier les présentations des résultats et aider l'interprétation des plans factoriels en identifiant des zones bien décrites.

V. MODELISATION

V-1. INTRODUCTION

Par définition le modèle est une représentation simplifiée, relativement abstraite, d'un processus ou d'un système, en vue de le décrire, de l'expliquer ou de le prévoir. Le modèle peut être aussi défini comme étant une formalisation mathématique supposée reproduire de manière approchée la réalité d'un phénomène dans le but d'en reproduire le fonctionnement pour permettre de comprendre, de prédire et/ou d'agir.

Le modèle mathématique peut être de nature déterministe ; s'appuyant sur des équations phénoménologiques et des principes scientifiques fondamentaux ou s'appuyant sur des schémas de fonctionnement (Modèle conceptuel, modèle de balance massique). Ou bien il peut être de nature probabiliste (Modèle empirique statistique).

Les modèles mathématiques seront transitoires ou permanents selon que l'on prendra ou non en compte la variation au cours du temps. Ils seront distribués ou agrégés selon que l'on introduira ou non une variabilité spatiale de leurs paramètres.

V-2. MODELES MATHEMATIQUES

V-2.1 Modèles basés sur des principes fondamentaux scientifiques

Ces modèles peuvent être décrits par des réactions physiques ou chimiques et n'exigent pas de calibration. Ils requièrent souvent un nombre élevé de variables.

Un exemple de modèle basé sur des principes fondamentaux scientifiques est le modèle décrivant la précipitation du calcaire dans l'eau potable, ce modèle se fonde sur l'équilibre entre le carbonate de calcium et le dioxyde de carbone. Ce modèle chimique de carbonate peut être utilisé, entre autre, pour la prédiction de l'effet des dosages chimiques sur le pH, la prédiction de la dose nécessaire pour atteindre le pH voulu ou bien le calcul de la solubilité des ions métalliques qui peuvent être utilisées pour la prédiction de l'effet de la température et du pH sur le résiduel en coagulant HART (1997).

V-2.2 Modèles de balance massique

Les procédures de séparation, telle que la clarification, la filtration et l'adsorption sur charbon actif sont caractérisées par des modèles de balance massique qui décrivent le débit massique du matériau à l'intérieur et à l'extérieur de processus et de l'accumulation du matériau.

Par exemple la performance du processus de clarification peut être décrite par l'équation de balance massique suivante :

$$\frac{dX}{dt} = \frac{Q}{V}(X_i - X) - v \frac{A}{V} X \quad (2-1)$$

A : section du décanteur [m²]

Q : débit d'eau [m³/s]

t : temps [s]

V : volume du décanteur [m³]

v : vitesse de décantation [m/s]

X : concentration du solide dans le décanteur [mg/l]

X_i : concentration du solide à l'entrée du décanteur [mg/l]

Cette équation peut être utilisée pour prévoir la variation de la qualité de l'eau et du débit en fonction du temps. En application en traitement des eaux cette équation nécessite d'être calibrer vue que les valeurs de vitesses de décantation des suspensions solides sont indispensables.

Ce type de modèle peut être aussi utilisé pour l'estimation de la vitesse de décantation qui est fonction des propriétés des floes et de la concentration du solide dans le décanteur.

V-2.3 Modèles empiriques (statistiques)

Le développement des modèles empiriques pour décrire la performance des processus est nécessaire quand des réactions chimiques complexes ont lieu ou quand il est difficile de développer un modèle mathématique fondamentale, ces modèles empiriques sont aussi utilisés quand le processus est contrôlé par plusieurs variables ou lorsque le mécanisme du processus n'est pas totalement acquis.

Les modèles empiriques sont développés à partir des données de laboratoire, de pilote ou des installations d'expérimentation et sont présentés par de simples équations reliant les paramètres d'entrée de la qualité de l'eau et celles de sortie en incluant parfois les conditions opératoires du processus.

A titre d'exemple, des modèles empiriques ont été développés pour montrer l'effet de la qualité de l'eau et des conditions de chloration sur la formation des trihalométhanes, d'autres modèles ont été utilisés pour décrire la relation entre la dose optimale de coagulant et la qualité de l'eau, et d'autres également ont été établis pour calculer l'effet de la dose du coagulant et du pH sur l'élimination de la matière organique naturelle (plus d'exemples sont illustrés au chapitre 01).

Les modèles empiriques sont un bon outil de prédiction, cependant ils présentent plusieurs insuffisance, LEK (2007) note que les insuffisances du modèle causent des déviations entre les résultats du modèle et les valeurs observées, ces insuffisances sont dues aux :

- Données imparfaites ; les données peuvent contenir des erreurs résultant de l'échantillonnage, la mesure, et/ou mauvaises estimations.
- L'analyse des données est souvent incomplète et dépourvue de mesure des erreurs.
- Dans la plupart des cas la dépense sur la collecte des données limite le ramassage des données en bon nombre.
- Sensibilité du modèle (variabilité du modèle et des paramètres)
- Le modèle est imparfait parce qu'il peut contenir des erreurs en hypothèse, en formulation, et/ou paramétrisation.
- Les modèles ne sont pas complètement validés parce que les données de validation sont soit indisponibles ou bien les techniques de validation ne sont pas performante.

Bien qu'ils exigent de longues séries de données, les modèles statistiques ont l'avantage de nécessiter que peu de données et un effort de développement souvent moins important. L'analyse des incertitudes de ces modèles suggère l'identification d'erreurs, d'imperfections et du non fiabilité. Généralement, deux familles de modèles sont incluses dans la catégorie statistique :

A) Les modèles paramétriques

Ce sont des modèles prédéfinis qui nécessitent l'ajustement de paramètres dont les valeurs sont déterminées par la calibration desdits modèles. Cette approche statistique peut regrouper des méthodes telles :

- La régression linéaire,
- La régression non-linéaire,
- Les modèles autorégressives (ARMA)
- Les modèles périodiques autorégressives (PARMA).

B) Les modèles non paramétriques

Les modèles non-paramétriques n'ont pas de structure statistique prédéfinie, mais leurs paramètres d'ajustement sont plutôt inférés à partir des données disponibles sans hypothèses sur la distribution de la population. Ce sont des outils efficaces lorsque les systèmes à modéliser sont non-linéaires. Il existe plusieurs modèles non-paramétriques, entre autres, les réseaux de neurones artificiels (RNA) et les k-voisins les plus proches (VPP).

L'avantage des modèles non-paramétriques est leur souplesse relative et le fait qu'il n'y ait pas de paramètres à estimer. Cependant, leur inconvénient est qu'il est difficile d'avoir une fonction mathématique conventionnelle, et donc il sera difficile de les extrapoler.

V-3. MODELISATION PARAMETRIQUE.

D'une façon un peu schématique, on peut associer la modélisation statistique paramétrique à trois objectifs qui peuvent éventuellement être suivis en complémentarité BESSE(2009) :

- **Descriptif** : Il vise à rechercher de façon exploratoire les liaisons entre Y et d'autres variables, potentiellement explicatives, X_j qui peuvent être nombreuses afin, par exemple d'en sélectionner un sous-ensemble.
- **Explicatif** : Le deuxième objectif est sous-entendu par une connaissance à priori du domaine concerné et dont des résultats théoriques peuvent vouloir être confirmés, infirmés ou précisés par l'estimation des paramètres. Dans ce cas, les résultats inférentiels précédents permettent de construire le bon test conduisant à la prise de décision recherchée.
- **Prédictif** : Dans le troisième cas, l'accent est mis sur la qualité des estimateurs et des prédicteurs qui doivent, par exemple, minimiser une erreur quadratique moyenne. C'est la situation rencontrée en apprentissage. Ceci conduit à rechercher des modèles parcimonieux c'est-à-dire avec un nombre volontairement restreint de variables explicatives. Le "meilleur" modèle ainsi obtenu peut donner des estimateurs légèrement biaisés au profit d'un compromis pour une variance plus faible. Un bon modèle n'est donc plus celui qui explique le mieux les données au sens d'une déviance (SSE ; somme des carrés des écarts) minimale (ou d'un R^2 max) au prix d'un nombre important de variables pouvant introduire des colinéarités. Le bon modèle est celui qui conduit aux prévisions les plus fiables.

La construction du modèle statistique peut se décomposer en quatre étapes:

1. Apprentissage du modèle.
2. Tests pour vérifier les modèles obtenus en apprentissage sur un autre échantillon d'individus.
3. Etude des diagnostics pour justifier de la pertinence et de la robustesse des modèles, et du choix du "meilleur" modèle.
4. Validation facultative sur un troisième échantillon pour mesurer les performances du "meilleur" modèle.

V-3.1. Principe de la régression linéaire

La régression linéaire se classe parmi les méthodes d'analyse multivariée qui traitent des données quantitatives. C'est une méthode d'investigation des données d'observations, ou d'expérimentation, où l'objectif principal est de rechercher une liaison linéaire entre une variable Y quantitative et une ou

plusieurs variables X également quantitatives. C'est la méthode la plus utilisée pour deux raisons majeures :

- c'est une méthode ancienne,
- c'est l'outil de base de la plupart des modélisations plus sophistiquées comme la régression logistique, le modèle linéaire généralisé, les méthodes de traitement des séries temporelles,...etc. CONFAIS (2007)

Le modèle linéaire

On dispose d'un ensemble de n observations sur lesquelles ont été effectuées p+1 mesures des variables y, x₁, x₂,..., x_p. On veut expliquer ou prévoir y à l'aide des variables explicatives ou prédicteurs x₁, x₂,..., x_p lesquels sont supposées connues sans erreur. On cherche, à approcher y par une combinaison linéaire des variables explicatives x₁, x₂,..., x_p, pour cela on pose le modèle (voir figure 2.3).

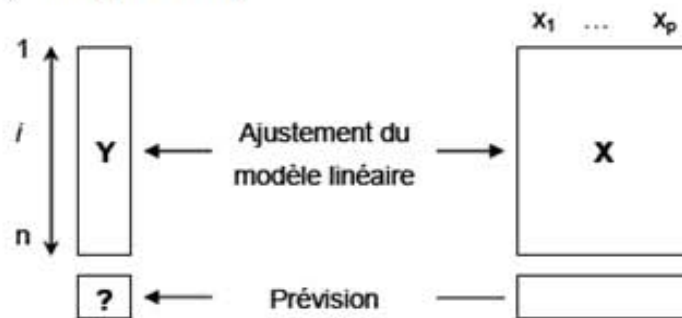


Figure 2.3 : Préviation linéaire LEBART (1997).

$$y_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \dots + \alpha_p x_{ip} + \varepsilon_i \quad \text{avec } i=1, \dots, n \quad (2.2)$$

Où : $\alpha_0, \alpha_1, \alpha_2 \dots \alpha_p$ sont les coefficients inconnus du modèle, et le terme ε_i est le résidu représentant l'écart entre la valeur observée y_i et la partie expliquée de l'observation.

$$\text{Partie expliquée} = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \dots + \alpha_p x_{ip} \quad (2.3)$$

Ce modèle s'exprime sous forme matricielle, la figure 2.4 est la schématisation du modèle

$$\begin{array}{c} Y \\ \begin{array}{|c|} \hline y_i \\ \hline \end{array} \end{array} = \begin{array}{c} X \\ \begin{array}{|c|} \hline 1 \dots 1 \\ \hline \end{array} \begin{array}{|c|} \hline x_{i1} \dots x_{ip} \\ \hline \end{array} \end{array} \times \begin{array}{c} \alpha \\ \begin{array}{|c|} \hline \alpha_0 \\ \alpha_1 \\ \dots \\ \alpha_p \\ \hline \end{array} \end{array} + \begin{array}{c} \varepsilon \\ \begin{array}{|c|} \hline \varepsilon_i \\ \hline \end{array} \end{array} \quad (2.4)$$

Figure 2.4 : Schématisation du modèle linéaire. LEBART (1997)

On recherche des estimateurs a_0, a_1, \dots, a_p de $\alpha_0, \alpha_1, \dots, \alpha_p$ qui minimisent l'erreur

$$e_i = y_i - (a_0 + a_1 x_{i1} + a_2 x_{i2} + \dots + a_p x_{ip}) \quad (2.4)$$

Parmi les critères de minimisation possibles est celui des moindres carrés.

Ajustement par la méthode des Moindres Carrés

L'ajustement du modèle correspond à toute solution du système d'équation

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + \dots + a_p x_{ip} + e_i \quad i=1, \dots, n \quad (2.5)$$

Ce qui correspond sous forme matricielle à :

$$Y = X a + e \quad (2.6)$$

$\begin{matrix} (n,1) & (n,p+1) & (p+1,1) & (n,1) \end{matrix}$

Pour l*i*ème observation, La valeur prédite par le modèle sera

$$\hat{y}_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + \dots + a_p x_{ip} \quad (2.7)$$

Le résidu du modèle vaut

$$e_i = y_i - \hat{y}_i \quad (2.8)$$

On cherchera donc à trouver \hat{y}_i le plus proche à y

L'ajustement par la méthode des moindres carrés fournit les coefficients $a_0, a_1, a_2, \dots, a_p$ conduisant au minimum de la somme des carrés des écarts.

$$\min(\sum e_i^2) \quad \text{Avec} \quad \sum e_i^2 = \left\| y - \hat{y} \right\|^2 \quad (2.9)$$

Lorsque sur un graphique, la liaison entre X et Y n'apparaît pas linéaire (le modèle linéaire n'est pas validé), on peut essayer de transformer les données, pour tenter de linéariser la liaison. La transformation a un double effet, d'une part elle symétrise la distribution et d'autre part elle linéarise la liaison, pour choisir une transformation appropriée, Turkey a proposé ce qu'on appelle maintenant l'échelle de transformation de Turkey CONFAIS (2007) présenté à la figure 2.5

Selon la forme des courbes définies par les points (X_i, Y_i) on pourra soit monter l'échelle, c'est-à-dire transformer X ou Y en ses puissances (Y^2, Y^3 etc.) soit descendre l'échelle, en prenant $\sqrt{Y}, \text{Log}(Y), -1/\sqrt{Y}, -1/Y...$ etc.

L'astuce de lecture proposée par J. Vanpoucke et E. Horber CONFAIS (2007) facilite la transformation, si la forme des courbes du tableau de la figure ci-dessous dessine un arc, la flèche indiquera selon son orientation s'il faut monter ou descendre l'échelle.





Forme de Courbes	Description de l'action	Transformation sur X	Transformation sur Y		
	Descendre l'échelle en X ou Monter l'échelle en Y	$\sqrt{X}, \text{Log}(X), -1/\sqrt{X}$ etc.	Y^2, Y^3 etc.	4	Y^4
	Descendre l'échelle en X ou Descendre l'échelle en Y	$\sqrt{X}, \text{Log}(X), -1/\sqrt{X}$ etc.	$\sqrt{Y}, \text{Log}(Y), -1/\sqrt{Y}$ etc.	3	Y^3
	Monter l'échelle en X ou Descendre l'échelle en Y	X^2, X^3 etc.	$\sqrt{Y}, \text{Log}(Y), -1/\sqrt{Y}$ etc.	2	Y^2
	Monter l'échelle en X ou Monter l'échelle en Y	X^2, X^3 etc.	Y^2, Y^3 etc.	1	Y^1
				1/2	\sqrt{Y}
				0	$\text{Log}(Y)$
				-1/2	$-1/\sqrt{Y}$
				-1	$-1/Y$
				-2	$-1/Y^2$

Figure 2.5 : Echelle de Turkey d'après CONFAIS (2007)

De manière générale, dans toute approche de modélisation statistique, on commence par la régression linéaire dont on vérifie la légitimité. Il arrive cependant que les hypothèses de linéarité du modèle ne soient pas vérifiées. En cas d'invalidité des hypothèses de linéarité, la modélisation non linéaire peut être mise en jeu. La technique dépendra d'une part de la nature des données à modéliser, et d'autre part de la forme de dispersion de ces mêmes données.

CONFAIS (2007) note que les méthodes de régression sont des méthodes très puissantes, mais qui doivent être utilisées avec beaucoup de discernement et de prudence, et qu'il ne faut pas se contenter d'un seul modèle et d'une seule procédure de régression, il faut en tester plusieurs. Dans notre cas d'étude les régressions testées sont :

Régression Simple

Régression simple 1^{er} degré $Y = b_0 + b_1P$ (2.10)

Régression simple de 2^{ème} degré $Y = b_0 + b_1P^2$ (2.11)

Régression Multiple

Régression multiple 1^{er} degré $Y = b_0 + b_1P + b_2Q + b_3R$ (2.12)

Régression multiple 2^{ème} degré
 (Régression polynomiale) $Y = b_0 + b_1P + b_2P^2 + b_3Q + b_4Q^2$ (2.13)

Régression Factorielle (Deux variables ou plus avec interaction)

2^{ème} degré $Y = b_0 + b_1P + b_2Q + b_3P.Q$ (2.14)

3^{ème} degré $Y = b_0 + b_1P + b_2Q + b_3P.Q.R + b_4P.Q + b_5P.R + b_6Q.R$ (2.15)

Régression quadratique de surface de réponse

La régression quadratique de surfaces de réponse est une conception hybride entre les caractéristiques de la régression polynomiale et la régression factorielle, elle contient les mêmes effets de la régression

polynomiale du deuxième degré et les effets d'interaction des variables prédictives. L'équation de régression quadratique de surface de réponse pour trois variables continues P, Q, et R est de la forme :

$$Y = b_0 + b_1P + b_2P^2 + b_3Q + b_4Q^2 + b_5R + b_6R^2 + b_7P.Q + b_8P.R + b_9Q.R \quad (2.16)$$

Régression de surface mixte

La régression de surface mixte est identique à la régression factorielle de degré 2 à l'exception de l'omission de l'intercepte. L'équation de régression de surface mixte pour trois variables continues P, Q et R est de la forme :

$$Y = b_1P + b_2Q + b_3R + b_4P.Q + b_5P.R + b_6Q.R \quad (2.17)$$

V-3.2. Sélection du meilleur modèle

La performance du modèle développé sera évaluée par sa capacité de prédiction ou de généralisation. L'objectif a priori est de trouver le modèle qui ajusterait au mieux les données réelles. Cependant, le meilleur modèle n'est pas nécessairement celui qui possède cette caractéristique. Notre choix sera ainsi basé sur des critères de qualité d'ajustement et de prédiction, tout en recherchant la parcimonie. On désignera donc le meilleur modèle par un compromis entre précision et qualité du modèle BESSE (2005 b).

V-3.2.1. Sélection des données en pratique

Nous travaillons généralement avec deux échantillons ; un échantillon d'apprentissage et un échantillon de validation. Sur l'échantillon d'apprentissage, on choisit des modèles en utilisant les critères et les algorithmes de sélection que nous allons présenter dans les sections suivantes. On aura un ou plusieurs modèles candidats. Parmi ce nombre de modèles candidats, il suffit alors d'utiliser l'échantillon de validation pour choisir le modèle qu'on va conserver et étudier, il nous faut alors suffisamment de données pour pouvoir bien estimer dans le jeu d'apprentissage et suffisamment dans le jeu de validation pour avoir une bonne idée du comportement du modèle.

Souvent, nous avons un seul jeu de données. Une possibilité consiste alors à séparer le jeu initial en deux parties, l'une réservée à l'apprentissage, l'autre à la validation. Cela nécessite donc beaucoup d'observations et évidemment il n'est pas possible de donner de règle quant à la taille minimum n requise. De même, pour les tailles respectives d'apprentissage et de validation sont souvent énoncées les proportions 3/4, 1/4 ou 1/2, 1/2 sans aucune véritable justification BESSE (2009).

MOURAD (2005) résume l'étude de sensibilité des données expérimentales sur la modélisation de la qualité des rejets urbains par :

- Le partage traditionnel à parts égales des données expérimentales entre l'apprentissage et validation est sous-optimal pour tous les cas

étudiés, plus de données doivent être allouée au calage, de l'ordre de 60% au minimum.

- Une vingtaine d'événement est le minimum pour pouvoir utiliser correctement les modèles de régression.

Ces résultats sont, évidemment, relatifs aux données et aux modèles établis par MOURAD (2007), néanmoins, vu le manque de règle de partage, nous essayerons dans la présente étude de respecter ces résultats et de partager la base de données comme suit 2/3 de l'ensemble pour l'apprentissage et le 1/3 restant pour la validation.

Pour le choix des variables d'entrée, plusieurs travaux portant sur la modélisation des paramètres de floculation LAMRINI (2005 et 2007) et HECTOR (2006) montrent l'intérêt d'utiliser l'analyse en composantes principales afin de déterminer les variables explicatives qui influent le plus sur la variable à expliquer.

KETTANEH (2005) note qu'il faut mettre plus d'attention et de soins en analysant de larges ensembles de données, il mentionne qu'avec de grandes tailles, l'interprétation de l'analyse est souvent difficile et l'efficacité des modèles résultant est à discuter, il propose alors afin de contourner les différentes difficultés rencontrées telles que : large base de données, points aberrants, non linéarité, bruit, données manquantes, de procéder à :

- La réduction des données par les méthodes de plan d'expérience, l'analyse en composantes principales et l'analyse par PLS (partial least squares),
- La normalisation, la transformation, la classification et la discrimination des données,
- L'élimination des points aberrants et leviers et comblement des lacunes.

Dans notre cas d'étude, la sélection des entrées du modèle sera effectuée en utilisant la technique statistique par analyse en composantes principales en complémentarité avec l'analyse de classification hiérarchique permettant d'éliminer les informations redondantes et de restreindre le nombre des variables d'entrée à celles les plus pertinentes.

V 3.2.2. Procédure de choix de modèle

La sélection de modèle peut être vue comme la recherche du modèle optimal, au sens d'un critère choisi, parmi toutes les possibilités. Cela peut donc être vu comme une optimisation d'une fonction objectif (le critère). Pour cela, et à l'image des possibilités en optimisation, on peut soit faire une recherche exhaustive car le nombre de modèles possibles est fini, soit partir d'un point de départ et utiliser une méthode d'optimisation de la fonction objectif (recherche pas à pas).

Recherche exhaustive

Lorsque tous les modèles avec p variables sont possibles, il y a 2^p possibilités et donc cette méthode n'est pas envisageable si p est grand. Des techniques algorithmiques permettent cependant de minimiser le nombre de calculs à effectuer et permettent d'envisager cette possibilité dans des cas de taille modérée.

Recherche pas à pas

Il existe des algorithmes permettant de sélectionner de manière optimisée le modèle de notre choix, parmi lesquels:

La méthode descendante "Backward Selection" A la première étape toutes les variables sont intégrées au modèle. Ensuite la méthode consiste à éliminer les variables, moins significatives, dont les t de Student sont en dessous du seuil critique. Cependant l'inconvénient de cette méthode est lorsque le modèle comporte un nombre important de variables explicatives, le risque de multicolinéarité est élevé.

La méthode ascendante "Forward Regression" On introduit les variables une par une, on commence par un modèle à une variable, on ajoute à chaque étape une variable, cela consiste à sélectionner la variable explicative pour laquelle le coefficient de corrélation est le plus élevé. L'algorithme calcule ensuite les corrélations partielles et retient la variable présentant le coefficient de corrélation le plus élevé, et ainsi de suite, la sélection s'arrête lorsque les t de Student des variables sont inférieurs au seuil critique.

La méthode progressive "Stepwise selection" Même principe que pour la méthode ascendante, sauf que l'on peut éliminer des variables déjà introduites. En effet, il peut arriver que des variables introduites en début ne soient plus significatives après introduction de nouvelles variables.

V-3.2.3. Critères de choix de modèle

De nombreux critères de choix de modèle sont présentés dans la littérature, nous citons le R^2 , le $R^2_{\text{ajusté}}$, le C_p , le REQ (Racine carrée de l'Erreur Quadratique Moyenne), l'AIC (Akaike Information Criterion), le BIC (Bayesian Information Criterion). En pratique, les plus utilisés sont les suivants. BESSE (2009).

R^2 et $R^2_{\text{ajusté}}$

Le coefficient de détermination, directement lié à la déviance (SSE), est aussi un indice de qualité mais qui a la propriété d'être monotone croissant en fonction du nombre de variables. Il ne peut donc servir qu'à comparer deux modèles de même niveau c'est-à-dire avec le même nombre de variables. En revanche, le " $R^2_{\text{ajusté}}$ " permet la comparaison entre des modèles de niveaux différents :

$$R^2 = 1 - \frac{SSE}{SST} \quad , \quad R^2_{ajusté} = 1 - \frac{SSE(n-1)}{SST(n-p-1)} \quad (2.18)$$

n : Taille de l'échantillon (nombre d'individus)

P : nombre de variables explicatives

SSE : la somme des carrés des écarts $SSE = \sum (y_i - \hat{y}_i)^2$

SST : la somme totale des carrés $SST = \sum y_i^2$

La statistique de Student.

L'apport de chaque variable explicative dans l'équation de régression est évalué par l'utilisation d'un test d'hypothèse sur le coefficient de régression a_k . L'hypothèse nulle H_0 selon laquelle la variable X_k n'a pas d'influence réelle sur la prédiction se traduit par: $a_k = 0$.

La statistique de Student est alors appliquée:

$$t = \frac{a_k}{S_{a_k}} \quad (2.19)$$

avec S_{a_k} l'estimation de l'écart-type de a_k

Le seuil de confiance pour ce test, pour lequel on rejettera l'hypothèse nulle, est pris généralement $p < 0,05$ le coefficient de régression sera alors significativement (à 95%) différent de 0.

La statistique F de Fisher

Ce critère, justifié dans le cas explicatif car basé sur une qualité d'ajustement est aussi utilisé à titre indicatif pour comparer des séquences de modèles emboîtés. La statistique partielle de Fisher est

$$\frac{(SSR - SSR_q) / q}{SSE / (n - p - 1)} = \frac{(R^2 - R_q^2) n - p - 1}{1 - R^2} \quad (2.20)$$

n : Taille de l'échantillon (nombre d'individus)

P : nombre de variables explicatives

SSE : la somme des carrés des écarts $SSE = \sum (y_i - \hat{y}_i)^2$

SSR : la somme totale des carrés (valeurs estimées) $SSR = \sum \hat{y}_i^2$

Dans laquelle l'indice q désigne les expressions concernant le modèle réduit avec (p-q) variables explicatives.

Cp de Mallows

Cet indicateur est une estimation de l'erreur quadratique moyenne de prévision il s'écrit pour un modèle à j variables explicatives comme suit :

$$C_p = (n - q - 1) \frac{MSE_j}{MSE} - [n - 2(q + 1)] \quad (2.21)$$

n : Taille de l'échantillon (nombre d'individus)

MSE et MSE_j désignant respectivement l'erreur quadratique moyenne du modèle complet et celle d'un modèle à j variables explicatives. Il est alors d'usage de rechercher un modèle qui minimise le C_p .

La racine carrée de l'erreur quadratique moyenne REQM

La racine carrée de l'erreur quadratique moyenne commise sur la base de données est définie comme suit :

$$REQM = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (2.22)$$

Avec O_i et P_i les valeurs respectivement mesurées et estimées par le modèle et n est la taille de l'échantillon.

V-3.2.4. Qualité de prédiction du modèle

Par le biais de l'analyse des résidus on cherchera à valider le modèle de régression postulé. L'étude des résidus, également appelée étude des diagnostics, associée à la régression linéaire multiple nous permet de détecter les violations d'hypothèses ou encore les points influents dans le contexte multidimensionnel. En effet si les résidus se répartissent de façon aléatoire et sont indépendant et identiquement distribuée, le modèle est pertinent. En effet pour un meilleur ajustement du modèle les résidus doivent être selon VITTINGHOFF (2005) et CONFAIS (2007).

- De moyenne nulle
- De variance constante
- Suivent une distribution normale
- Indépendants

Ajustement individuel au modèle, Valeur aberrante, Point levier

Valeur aberrante

Une valeur aberrante est une observation qui est mal expliquée par le modèle et admet un résidu élevé; Pour analyser la qualité de l'ajustement d'une observation, il suffit de regarder le résidu correspondant à cette observation. Si ce résidu est anormalement élevé, alors l'individu i est appelé point aberrant. Il convient alors d'essayer d'en comprendre la raison (erreur de mesure, individu provenant d'une sous-population) et éventuellement d'éliminer ce point car il peut modifier les estimations.

Point levier

Un point est dit "point levier" si les valeurs h_{ii} (Le « poids » de l'observation i sur sa propre estimation vaut h_{ii} .) de la matrice de projection sont élevées (voir CORNILLON 2007). Les points leviers sont donc des points atypiques. Le point levier peut être non aberrant s'il se situe dans le prolongement de la droite de régression et donc son résidu sera faible. Là encore il est bon de repérer les points leviers et de les noter, puis de comprendre pourquoi ces points sont différents : erreur de mesure, erreur d'enregistrement, ou appartenance à une autre population. Le point levier s'il n'est pas aberrant pourra être éliminé ou conservé. Dans le premier cas aucun risque n'est pris au bord du domaine, quitte à sacrifier quelques points. Dans le second cas, le modèle est étendu de manière implicite jusqu'à ces points.

Analyse de la normalité

La normalité des distributions des variables quant à elle, est examinée de plusieurs façons selon un seuil prédéfini ($p < 0,05$) SAPORTA (1990):

Par ajustement graphique ; cela consiste à comparer les quantiles des résidus estimés à l'espérance des mêmes quantiles sous l'hypothèse de normalité. Le graphique (droite de Henry) est alors obtenu en dessinant les observations contre leurs espérances théoriques respectives sous hypothèse de normalité. Si cette hypothèse est respectée, le graphique obtenu sera proche de la première bissectrice (courbe théorique).

Les tests statistiques : par le biais des tests du χ^2 , de Kolmogorov-Smirnov ou de Cramer-von Mises, basé sur l'écart entre la distribution des fréquences observées (donc de l'échantillon) et celle d'une distribution théorique. Pour valider l'ajustement à la valeur théorique, on le compare à une valeur théorique dans une table de distribution. L'étude de la normalité des données est effectuée après traitement des valeurs extrêmes. Dans notre cas d'étude nous nous contenterons de l'ajustement graphique pour l'analyse de la normalité.

Analyse de l'homoscédasticité

La vérification de l'homoscédasticité consiste à rechercher si dans les résidus il n'existe pas une structure organisée ou un contenu informationnel qui prouverait que le modèle postulé se différencie significativement du modèle correct, elle peut être vérifiée par plusieurs tests RACKAUSKAS (2007), dans notre cas d'étude elle sera évaluée par le nuage de points des résidus, il est recommandé de tracer les résidus en fonction des valeurs ajustées y_i . Si une structure apparaît (cône, vagues), l'hypothèse d'homoscédasticité risque fort de ne pas être vérifiée.

La non autocorrélation

La non autocorrélation des résidus empêchera les valeurs moyennes d'être surestimées, et les valeurs extrêmes sous-estimées. Elle peut être détectée grâce au nuage de points des résidus, ou à la statistique de Durbin-Watson en cas des données apparentées CONFAIS (2007).

La statistique d de Durbin-Watson vaut entre 0 et 4; Elle a été tabulée par Durbin et Watson (Annexe A) pour différentes tailles d'échantillon n et de nombre de vraies variables explicatives k (sans compter la constante). La règle de décision est résumée de la manière suivante :

- Non auto corrélation d'ordre 1 si $d_U < d$
- Auto corrélation d'ordre 1 si $d < d_L$
- Incertitude si $d_L < d < d_U$

La distance de Cook D

Comme toute méthode quadratique, l'estimation des paramètres est sensible à la présence de points extrêmes dont la présence peut perturber les résultats. La mesure de la distance de Cook permet de diagnostiquer la présence de ces points influents, Une observation influente est donc une observation qui, enlevée, conduit à une grande variation dans l'estimation des coefficients, c'est-à-dire à une distance de Cook élevée.

Les points présentant des distances de Cook élevées, généralement supérieure à 1, seront des points aberrants, ou leviers, ou les deux, et influenceront l'estimation. A l'image des points aberrants et leviers, nous recommandons de supprimer les observations présentant une forte distance de Cook. Si l'on souhaite toutefois absolument garder ces points, il sera très important de vérifier que les coefficients estimés et les interprétations tirées du modèle ne varient pas trop avec ou sans ces observations influentes.

V-3.2.5. Stabilité des coefficients de régression

La source principale d'instabilité dans l'estimation des coefficients de régression est la multicolinéarité, différents symptômes sont révélateurs des problèmes de colinéarité CONFAIS (2007)

- De grandes corrélations entre les régresseurs
- Un grand changement dans les coefficients quand on ajoute ou on enlève un régresseur
- Des coefficients non significatifs alors que le test global d'analyse de variance sur tous les coefficients est significatif ;
- La non significativité et/ou une très grande variance pour le coefficient d'un régresseur théoriquement important dans le modèle.
- Un coefficient de signe opposé à celui auquel on s'attendait.

Dans le cas de colinéarité, les paramètres du modèle seront estimés avec imprécision et les prédictions pourront être entachées d'erreurs considérables même si R^2 a une valeur élevée, Il est donc important de mesurer l'effet de la colinéarité entre les prédicteurs cela s'effectue au moyen du facteur d'inflation de la variance (VIF) ou de tolérance (tolérance = $1/VIF$), une valeur de VIF plus grande que 10 révèle un problème SAPORTA (1990) et CONFAIS (2007).

En cas de multi-colinéarité plusieurs remèdes sont présentés :

- Sélectionner les régresseurs et retirer ceux "responsables" de la colinéarité
- Transformer les régresseurs par des ratios si on identifie le facteur commun de liaison, ou bien augmenter la taille n de l'échantillon.
- Travailler sur les composantes principales issues des régresseurs (régression régularisée), ou bien faire une régression PLS ou Ridge.

**V-3.3. Modèle de régression sur composantes principales
(régression régularisée ou orthogonalisée)**

L'Analyse en Composantes Principales consiste à la recherche de p variables dites principales qui sont des combinaisons linéaires des variables initiales de variance maximale sous une contrainte d'orthogonalité. Alors que la régression orthogonalisée est considérée comme un mélange de l'analyse des composantes principales et de la régression multiple, son principe est en effet de rechercher l'équation de régression multiple relative à une variable dépendante y (variable à expliquer) à p variables explicatives x_1, x_2, \dots, x_p par l'intermédiaire des composantes principales de ces variables explicatives.

En fait, il est géométriquement équivalent de régresser Y sur les composantes principales que sur les variables initiales LEBART (1997). DAGNELIE (1982) Note que lorsque l'objectif de la régression est beaucoup plus d'expliquer ou d'interpréter un phénomène plutôt que d'essayer de le prévoir il est préférable que toutes les variables observées soient prises en considération alors il peut être utile de procéder à une régression orthogonalisée (régression sur composantes principales).

LEBART (1997) mentionne que L'analyse de régression sur composantes principales est recommandée lorsque les variables explicatives sont nombreuses ou fortement corrélées entre elles, l'analyse factorielle dans ce cas joue donc un double rôle d'exploration et de régularisation.

On élimine alors le problème de la quasi- colinéarité en supprimant les vecteurs correspondant à des valeurs propres nulles ou très faibles. Autrement dit on ne retient que les premières composantes principales de variance non négligeables LEBART (1997).

Par ailleurs SAPORTA (1990) note que les composantes principales de forte variance ne sont pas nécessairement les plus explicatives et qu'il vaut mieux les ordonner en fonction de leurs corrélations avec y (la variable à expliquer) et que les composantes principales de variance proche de zéro fournissent les relations linéaires approchées existantes entre les prédicteurs.

La réalisation de l'analyse de régression orthogonalisée (régularisée) peut être scindée en trois étapes DAGNELIE (1982):

1. La détermination des composantes principales.
2. La recherche des coefficients de corrélation et de régression.
3. Le retour aux variables initiales.

V-3.4. Récapitulatif

Le tableau 2.1 résume les choix faits dans cette étude, en terme de méthodologie suivie en régression multiple ; Détermination des entrées et partage des données, les modèles testés, la procédure et les critères de choix et de qualité de prédiction de modèle.

Rubrique	Critère retenu
Détermination des Paramètres d'entrée	Analyse en composantes principales Analyse de classification hiérarchique
Découpage des sets de données	Apprentissage : 2/3 des données pour élaborer le modèle Validation : 1/3 des données pour tester le modèle
Modèles de régression testés	Régression Multiple 1 ^{er} degré Régression polynomiale Régression Factorielle Régression quadratique de surface de réponse Régression de surface mixte Régression régularisée
Procédure de choix de modèle	La méthode ascendante "Forward Regression"
Critères de choix de modèle	R^2 et R^2 ajusté Statistique du F de Fisher Statistique de Student REQM
Qualité de prédiction du modèle	Ajustement individuel au modèle, Valeur aberrante, Point levier Analyse de la normalité Analyse de l'homoscédasticité La non autocorrélation des résidus La distance de Cook D
Stabilité de modèle	La tolérance

Tableau 2.1: Choix faits concernant la construction du modèle de régression multiple.

V-4. MODELISATION NON PARAMETRIQUE PAR RESEAUX NEURONES

V-4.1. Introduction

Les réseaux de neurones constituent aujourd'hui une technique de traitement de données qui permet d'extraire le maximum d'informations pertinentes des données : commande de procédés, prédiction de propriétés, modélisation de fonctions, reconnaissance de formes...etc. Ils possèdent d'indéniables qualités lorsque l'absence de linéarité et/ou le nombre de variables explicatives rendent les modèles statistiques traditionnelles inutilisables. Leur flexibilité alliée à une procédure d'apprentissage intégrant la pondération des variables et leurs interactions peuvent les rendre très efficaces.

HIM-MELBLAU (2000) note que, c'est à cause de sa structure, le réseau de neurones artificiels fournit de bons modèles empiriques des processus complexes non linéaires ce qui rend son application très utile dans divers buts et domaines et en particulier dans l'engineering chimique.

Les réseaux de neurones sont typiquement applicables dans les domaines d'investigation suivants ZORRIASSATINE (1998) :

- La classification et le diagnostic
- L'approximation de fonction et la prédiction
- L'optimisation

De façon formelle, un réseau de neurones artificiels est une structure composée d'entités (les neurones) capables du calcul et interagissant entre eux.

V-4.2. Définition de neurones artificiels

Les neurones artificiels sont souvent utilisés sous forme de réseaux qui diffèrent selon le type de connections entre les neurones, une cinquantaine de types peut être dénombrée. A titre d'exemples nous citons : le perceptron multicouche (PMC), le perceptron de Rosenblatt, les réseaux de Hopfield ...etc.

Un neurone artificiel est une modélisation simplifiée du neurone biologique, c'est un processeur élémentaire qui reçoit un nombre d'entrées en provenance des autres neurones ou de l'extérieur (voir figure 2.6). A chacune de ces entrées est associé un poids représentant la force de la connexion. La sortie du neurone est une somme pondérée de ses entrées plus un biais, multipliée par une fonction d'activation qui peut être selon BESSE (2009).

- linéaire $f(x) = x$
- sigmoïde : logistique $f(x) = \frac{1}{1 + e^{-x}}$, tangente hyperbolique $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
- seuil $f(x) = 0$ si $x \leq 0$, $f(x) = 1$ si $x > 0$

- stochastique $f(x)=1$ avec la probabilité $1/(1+e^{-x/H})$, 0 si non

Les modèles linéaires et sigmoïdaux sont bien adaptés aux algorithmes d'apprentissage comme celui de rétropropagation du gradient car leurs fonctions de transition est différentiable, ces modèles sont les plus utilisés.

Le modèle à seuil est sans doute plus conforme à la "réalité" biologique mais pose des problèmes d'apprentissage. Alors que le modèle stochastique est utilisé pour des problèmes d'optimisation globale de fonctions perturbées ou encore pour les analogies BRESSE (2009).

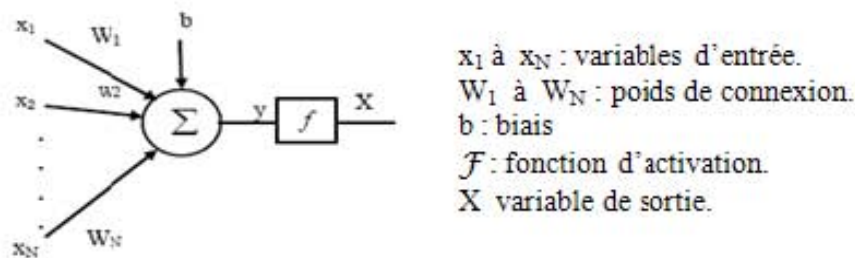


Figure 2.6 : Modèle du neurone artificiel

Choix de la fonction d'activation

Le choix d'une fonction d'activation se révèle être un élément constitutif important des modèles des réseaux neuronaux, il est basé sur la puissance de la fonction dans la simulation de la nature réelle du problème. La fonction d'activation souhaitable pour les unités cachées a pour but d'introduire la non linéarité dans le modèle ce qui donne une grande puissance de simulation de la relation entre les entrées et la sortie. Pour la sortie la sélection de la fonction d'activation est basée sur la distribution de la valeur à estimer. L'utilisation des fonctions sigmoïdes ; (fonctions en forme de « s » symétriques par rapport à l'origine), permet la construction de modèle non linéaire par compression des entrées dans la marge rétrécis de 0-1 (fonction logistique sigmoïde) ou -1 à 1 (fonction tangente hyperbolique). Elles sont particulièrement utilisées quand les valeurs du paramètre à estimé sont bornées.

En bibliographie, on note l'utilisation large des fonctions sigmoïdes logistique et tangente hyperbolique comme fonction d'activation des neurones de la couche cachée; à titre d'exemple, la fonction logistique sigmoïde a été utilisé en modélisation des installations d'ultrafiltration CABASSUD (2002), la variation de la salinité des eaux de rivières HUANG (2002), la contamination fécale des eaux ERION (2002), la modélisation de développement algale LEE (2003), la modélisation des paramètres morphodynamiques des rivières GEVREY (2003 et 2006), la simulation du flux de carbone MELESSE (2005), la détermination des concentrations de l'azote, du phosphore, et du carbone et leurs effets sur les changement climatique HOLMBERG (2006), la prédiction des concentrations de l'ozone SOUSA (2007), la prédiction de la

qualité de l'air IBBRA (2008) et l'optimisation de l'opération de dessalement des eaux de mer GAO (2007) et LEE (2009).

Par ailleurs, HAAS (2004) utilise la fonction sigmoïde tangente hyperbolique en modélisation de la chloration. RUDE (2008) utilise la fonction tangente hyperbolique en description de la pollution par le dioxyde d'azote HAMED (2004) et DELLANA (2009) l'utilisent également en prédiction de la performance de traitement des eaux usées.

Pour les neurones de la couche de sortie, la fonction d'activation est généralement choisie linéaire ; la fonction linéaire s'adapte aux valeurs continues du paramètre à estimer, par contre IBBRA (2008) et DELLANA (2009) utilisent respectivement la fonction logistique et la fonction tangente hyperbolique comme fonction d'activation des neurones de la couche de sortie.

V-4.3. Architecture d'un réseau de neurones

Principe

Un réseau de neurones artificiels (RNA) est formé d'un ensemble de neurones formels interconnectés. Deux grands types d'architectures peuvent être distingués :

- les réseaux de neurones non bouclés "FEED-FORWARD". Ce sont des réseaux dans lesquels l'information se propage de couche en couche sans retour en arrière possible et réalisent une (ou plusieurs) fonction(s) algébrique(s) des entrées, par composition des fonctions réalisées par chaque neurone. Les réseaux de neurones non bouclés sont des outils statiques, utilisés principalement pour effectuer des tâches d'approximation de fonctions non linéaires, de modélisation de processus statiques non linéaires ZORRIASSATINE (1998).
- Les réseaux de neurones bouclés "FEED-BACK". Ce sont des réseaux dans lesquels il y a retour en arrière de l'information, ils constituent un système dynamique "à temps discret", régi par une (ou plusieurs) équation(s) aux différences non linéaires, résultant de la composition des fonctions réalisées par chacun des neurones et des retards associés à chacune des connexions. Ils sont utilisés pour effectuer des tâches de modélisation et d'adaptation de systèmes dynamiques, de commande de processus, ou de filtrage.

Dans notre étude, on a choisi le modèle le plus couramment utilisé le Perceptron Multicouche (PMC) (réseau de neurones non bouclés), illustré par la figure 2.7 et dont les principes sont exposés dans ce qui suit.

Perceptron multicouche P.M.C.

Le P.M.C. est un réseau de neurones non bouclé caractérisé par une structure bien particulière, La figure 2.7 illustre la structure de ce type de réseau. Ses neurones sont organisés en couches successives où les informations circulent dans un seul sens, de la couche d'entrée vers la

couche de sortie. Les neurones d'une même couche ne sont pas interconnectés. Un neurone ne peut envoyer son résultat qu'à un neurone situé dans une couche postérieure à la sienne. Le réseau est formé de plusieurs couches, qui s'appellent, par convention :

- couche d'entrée : cette couche représente toujours une couche associée aux entrées du système, composée des cellules d'entrée qui correspondent aux n variables d'entrée $x_1, x_2 \dots x_n$.
- Couche cachée : chaque neurone de cette couche reçoit l'information de plusieurs couches précédentes, effectue la sommation pondérée par les poids, puis la transforme selon sa fonction d'activation. Par la suite, il envoie cette réponse aux neurones de la couche suivante.
- Couche de sortie : elle joue le même rôle que les couches cachées, la seule différence entre ces deux types de couches est que la sortie des neurones de la couche de sortie n'est liée à aucun autre neurone.

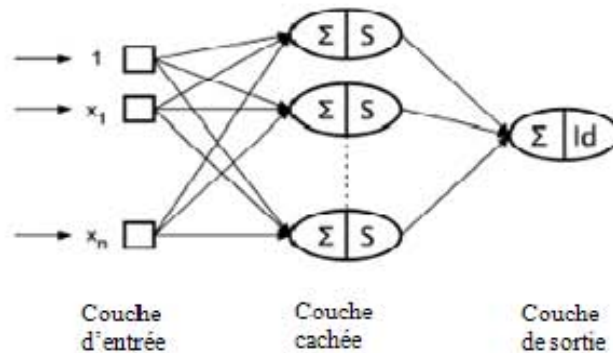


Figure 2.7 : Exemple d'architecture d'un PMC à x_n variables d'entrée, une seule couche cachée de fonction d'activation sigmoïde et une sortie de fonction d'activation identité.

Les neurones sont reliés entre eux par des connexions pondérées. Ce sont les poids de ces connexions qui gouvernent le fonctionnement du réseau et programment une application de l'espace des entrées vers l'espace des sorties, à l'aide de la transformation non linéaire.

V 1.1. Modélisation à l'aide de réseau de neurones

La stratégie de modélisation par réseaux de neurones et communément appelée approche par boîte noire, le processus entier est représenté avec un réseau neuronal, réalisé uniquement à partir des données expérimentales ou d'observations.

La fonction de régression non linéaire neuronale, aura pour expression algébrique

$$y = w_{0,s} + \sum_{j=1}^N \left[w_j \times f \left(w_{0,j} + \sum_{i=1}^n w_{j,i} x_i \right) \right] \quad (2.23)$$

Avec :

- n : nombre de variables d'entrée.
- N : nombre de neurones dans la couche cachée.
- w_i : poids relatifs à la connexion entre variables d'entrée et neurone.
- w_j : poids relatifs à la connexion allant du neurone j vers la sortie.
- $w_{0,j}$: poids relatifs à la connexion entre biais et neurone.
- $w_{0,s}$: poids relatifs à la connexion entre biais et sortie.
- y : variable de sortie.
- x_i : variable d'entrée.

V-4.5. Mise en œuvre de la régression neuronale

Les réseaux de neurones réalisent des fonctions non linéaires paramétrées. Leurs mises en œuvre nécessitent selon KAUNDA (2009) :

- Choix et préparation des entrées et des sorties pertinentes à partir des données brutes, c'est à dire les grandeurs qui ont une influence significative sur le phénomène que l'on cherche à modéliser.
- La collecte des données nécessaires à l'apprentissage et à l'évaluation des performances du réseau de neurones.
- La détermination de l'architecture nécessaire pour obtenir une approximation satisfaisante.
- La réalisation de l'apprentissage (simulation) jusqu'à ce que les erreurs seront minimisées.
- L'évaluation et le test des performances du modèle du réseau de neurones sur de nouvelles données (données de validation) à l'issue de l'apprentissage.
- Raffiner le modèle si nécessaire.

V 4.5.1. Détermination des entrées/sorties du réseau de neurones

Pour toute conception de modèle, la sélection des entrées doit prendre en compte deux points essentiels :

- Premièrement, la dimension intrinsèque du vecteur des entrées doit être aussi petite que possible, en d'autre terme, la représentation des entrées doit être la plus compacte possible, tout en conservant pour l'essentiel la même quantité d'information, et en gardant à l'esprit que les différentes entrées doivent être indépendantes.
- En second lieu, toutes les informations présentées dans les entrées doivent être pertinentes pour la grandeur que l'on cherche à modéliser, les entrées doivent donc avoir une influence réelle sur la valeur de la sortie.

ABDUL-WAHABA (2005) et SOUSA (2007) montrent que l'utilisation des composantes principales permet l'amélioration des données d'entrée du modèle de prédiction, cela par la réduction de leur complexité et l'élimination de la multicolinéarité existante.

HASANI (2008) et AL-ALAWI (2008) montrent que l'utilisation combinée de l'analyse en composantes principales et de la régression par réseau de neurones simplifie l'apprentissage en modélisation neuronale, en effet, travailler avec les composantes principales significatives diminue grandement l'effet des erreurs expérimentales et autres facteurs minoritaires intrus.

Pour AL-ABRI (2008), l'application de l'analyse en composantes principales permet d'élever la vitesse de simulation du réseau de neurones artificiels associée à une légère réduction en efficacité.

Dans la présente étude, nous procéderons au choix des variables descriptives d'entrée du modèle neuronal à la base de l'analyse multidimensionnelle par composantes principales et par classification hiérarchique soutenue par la recherche bibliographique sur les différents modèles établis de régression neuronale des paramètres de floculation.

V-4.5.2. Choix et préparation des échantillons

Le processus d'élaboration d'un réseau de neurones commence toujours par le choix et la préparation des échantillons de données, cette préparation permet selon JOO (2000) et BELICZYNSKI (2005) d'améliorer le taux d'apprentissage et la qualité de prédiction.

En effet, il est nécessaire de disposer de deux bases de données, une pour effectuer l'apprentissage et l'autre pour tester le réseau obtenu et déterminer ses performances. Notons qu'il n'y a pas de règle pour déterminer ce partage d'une manière quantitative, néanmoins chaque base doit satisfaire aux contraintes de représentativité de chaque classe de données et doit généralement refléter la distribution réelle.

On considère, en pratique, que $2/3$ de l'échantillon global doit être utilisé pour l'apprentissage et que le $1/3$ restant doit être pris pour le test. Dans notre cas d'étude nous suivons le même partage mais avec un échantillonnage aléatoire des deux jeux de données d'apprentissage et de test conformément aux travaux de AL-ABRI (2008).

V-4.5.3 Optimisation de la structure du réseau

Les performances d'un réseau de neurones sont étroitement liées à son architecture. Cependant, le choix de l'architecture est un problème difficile qu'il faut résoudre. En effet avec un nombre limité de neurones, le réseau ne sera pas performant sur l'apprentissage, avec un nombre trop important, il sera surentraîné et s'adaptera au bruit généré par les variables qui constituent la base de données.

Comme nous l'avons mentionné, nous avons retenu le perceptron multicouche comme base du modèle. Nous structurerons ce réseau en précisant le nombre de couches et de neurones cachés pour que le réseau soit en mesure de reproduire ce qui est déterministe dans les données.

Nombre de couches cachées : Mis à part les couches d'entrée et de sortie, il n'y a aucune règle pour la détermination du nombre de couches cachées, sans couche cachée, le réseau n'offre que de faibles possibilités d'adaptation, il faut alors déterminer ce nombre par essais empiriques (essai/erreur) selon ZORRIASSATINE (1998).

Par ailleurs, il a été démontré qu'un Perceptron Multicouche avec seulement une couche cachée pourvue d'un nombre suffisant de neurones, peut approximer n'importe quelle fonction continue et bornée avec la précision souhaitée.

Nombre de neurones cachés : Chaque réseau de neurones peut prendre en compte des profils spécifiques de neurones cachés. Un nombre important de neurones cachés permet donc de mieux se coller aux données présentées mais diminue la capacité de généralisation du réseau. Il faut alors trouver le nombre adéquat de neurones cachés nécessaires pour obtenir une approximation satisfaisante. Cette détermination du nombre de neurones cachés est généralement le fruit d'essais empiriques. La technique d'usage consiste à augmenter progressivement la complexité du modèle, en y ajoutant un neurone, et à chaque nouvelle addition d'un neurone, on évalue l'efficacité de généralisation.

Par ailleurs, selon HECHT- NIELSEN (1987), le nombre maximum de neurones dans la couche cachée est de " $2I+1$ " où I est le nombre des entrées du modèle WU (2008).

V-4.5.4. Apprentissage du réseau de neurones

La particularité des réseaux de neurones comme outil mathématique repose sur leur capacité d'apprentissage. Il s'agit d'une procédure adaptative par laquelle les connexions des neurones sont ajustées face à une source d'information.

L'apprentissage est la modification des poids du réseau dans l'optique d'accorder la réponse du réseau aux exemples et à l'expérience. Les poids sont initialisés avec des valeurs aléatoires. Puis des exemples expérimentaux représentatifs du fonctionnement du procédé dans un domaine donné, sont

présentés au réseau de neurones. Ces exemples sont constitués de couples expérimentaux de vecteurs d'entrée et de sortie. Une méthode d'optimisation modifie les poids au fur et à mesure des itérations pendant lesquelles on présente la totalité des exemples, afin de minimiser l'écart entre les sorties calculées et les sorties expérimentales.

Afin d'éviter les problèmes de surapprentissage, la base d'exemples est divisée en deux parties : la base d'apprentissage et la base de test (validation). L'optimisation des poids se fait sur la base d'apprentissage, mais les poids retenus sont ceux pour lesquels l'erreur obtenue sur la base de test est la plus faible. AMMAR (2007).

1) Types d'apprentissage

Il existe de nombreux types de règles d'apprentissage qui peuvent être regroupées en trois catégories : les règles d'apprentissage supervisé, non supervisé, et renforcé. Mais l'objectif fondamental de l'apprentissage reste le même : soit la classification, l'approximation de fonction ou encore la prévision.

Apprentissage supervisé

L'apprentissage est dit supervisé ou apprentissage par des exemples, lorsque le réseau est forcé à converger vers un état final précis, ce qui nécessite la connaissance à priori de la réponse désirée ou les solutions souhaitées. La technique consiste à présenter des exemples au réseau, calculer sa sortie, ajuster les poids de façon à réduire l'écart entre cette sortie et la réponse désirée pour satisfaire un certain critère de performance. Cet apprentissage n'est possible que si un large jeu de données est disponible et si les solutions sont connues pour les exemples de la base d'apprentissage.

Apprentissage renforcé

L'apprentissage renforcé est une technique similaire à l'apprentissage supervisé à la différence qu'au lieu de fournir des résultats désirés au réseau, on lui accorde plutôt un grade (ou score) qui est une mesure du degré de performance du réseau après quelques itérations. Les algorithmes utilisant la procédure d'apprentissage renforcé sont surtout utilisés dans le domaine des systèmes de contrôle.

Apprentissage non supervisé

Dans l'apprentissage non supervisé, aucun résultat désiré n'est fourni au réseau, la procédure d'apprentissage est basée uniquement sur les valeurs d'entrée. Le réseau apprend par lui-même à détecter les similarités et les différences dans l'ensemble d'apprentissage et à classer ces entrées en trouvant des points communs aux stimuli appliqués. L'avantage de ce type d'apprentissage réside dans sa grande capacité d'adaptation reconnue comme une autoorganisation, "self-organizing". L'apprentissage non supervisé est surtout utilisé pour le traitement du signal.

2) Algorithme d'apprentissage

L'algorithme d'apprentissage est la méthode mathématique qui va modifier les poids de connexions afin de converger vers une solution qui permettra au réseau d'accomplir la tâche désirée. Plusieurs algorithmes itératifs peuvent être mis en œuvre, parmi lesquels on note : l'algorithme de rétropropagation, la Méthode Quasi-Newton et Algorithme de BFGS, AMMAR (2007).

- Algorithme de rétropropagation ARP

La technique de rétropropagation du gradient (Backpropagation) est une méthode qui permet de calculer le gradient de l'erreur pour chaque neurone du réseau, de la dernière couche vers la première. Il est important de noter que l'ARP souffre des limitations inhérentes à la technique du gradient à cause du risque d'être piégé dans un minimum local. Il suffit que les gradients ou leurs dérivées soient nuls pour que le réseau se retrouve bloqué dans un minimum local. Ajoutons à cela la lenteur de convergence surtout lorsqu'on traite des réseaux de grande taille.

- Méthode Quasi-Newton

Cet algorithme nécessite la connaissance des valeurs de la fonction objectif à minimiser et du vecteur des dérivées premières au point solution courant. La méthode "quasi-newtonienne", n'est efficace que si elle est appliquée au voisinage d'un minimum.

- Algorithme de BFGS

L'algorithme de BFGS (du nom de ses inventeurs : Broyden, Fletcher, Goldfarb et Shanno) prend implicitement en compte les dérivées secondes et s'avère donc nettement plus performant que la méthode de rétropropagation. Le nombre d'itérations est nettement plus faible et le temps de calcul est réduit d'autant.

3) Problèmes d'apprentissage

Plusieurs problèmes peuvent intervenir au cours de l'apprentissage,

- Insuffisance de la règle d'apprentissage car rien ne garantit que la règle d'apprentissage soit capable de tirer profits du potentiel du réseau.
- Minima locaux, car la technique de descente du gradient utilisée par la majorité des apprentissages supervisés peut conduire à des solutions sous optimales.
- Un mauvais choix de paramètres tels que le nombre de neurones cachés par exemple, peut compromettre l'apprentissage.
- Le surapprentissage (over-fitting) peut se produire quand l'apprentissage du réseau est prolongé, ses poids reflèteront de trop près les particularités des exemples appris.

- Un mauvais échantillonnage du corpus d'apprentissage peut engendrer une mauvaise généralisation.
- Oublis ou interférences, quand le corpus est grand ou que le comportement à apprendre présente beaucoup de cas exceptionnels.

V-4.5.5. Validation du modèle

Une fois les poids de chaque connexion du réseau sont optimisés, il faut évaluer les performances du réseau, en d'autres termes la qualité de la régression, cette qualité garantit les capacités de généralisation du modèle. La validation du modèle doit être réalisée sur un ensemble de données des couples entrée-sortie n'appartenant pas à la base d'apprentissage. On procède à ce que l'on appelle une validation en aveugle. Elle permet à la fois d'apprécier les performances du système neuronal et de détecter le type de données qui pose problème. Les performances seront estimées à l'aide d'un certain nombre d'indicateurs pouvant être utilisés pour évaluer la qualité des modèles. Ce groupe d'indicateurs représente l'accord global entre valeurs observées et estimées.

Nous citons :

- R^2 : coefficient de détermination.
- Résidu moyen.
- Résidus studentisés
- Racine carrée de l'erreur quadratique moyenne (REQM)

V-4.5.5. Contribution des variables dans les modèles de réseau de neurones

L'analyse de contribution des variables d'entrée est extrêmement difficile vue que la fonction implémentée par les réseaux de neurones est complexe. En effet, à partir des variables d'entrée, le réseau est capable de prédire le paramètre de sortie, mais les mécanismes intérieurs de réseau sont complètement ignorés. L'interprétation d'un RNA n'est donc pas aussi directe que pour celle des modèles simples de régression linéaire. Ainsi, diverses méthodes ont été mise en œuvre pour caractériser l'importance des variables d'entrée.

GEVREY (2003 et 2006) résume ces méthodes en 5 techniques :

1. Méthode de dérivation partielle "PAD" qui consiste au calcul de la dérivation partielle de la sortie en fonction des variables d'entrée.
2. Méthode de poids, le calcul consiste à l'utilisation des poids de connexions.
3. Méthode par perturbation, qui consiste à observer l'effet de petites perturbations des variables d'entrée sur la sortie.
4. Méthode de Profil qui est une succession de variation d'une seule variable d'entrée et les autres variables restent constantes.

5. Méthode de Stepwise, se résume à l'observation du changement dans la valeur de l'erreur lorsqu'on ajoute ou on supprime une variable d'entrée.

GEVREY (2003 et 2006) note que la méthode de dérivation partielle PAD s'avère la plus utilisée et donne des résultats les plus complets, en deuxième position vient la méthode du profil, cependant, les méthodes de perturbation et de poids permettent une bonne classification des paramètres d'entrée et finalement la méthode stepwise classique donne les plus mauvais résultats. Dans notre cas d'étude nous utiliserons pour l'étude de contribution des variables, la méthode de poids de connexion, celle-ci permet une bonne classification des paramètres. Un exemple d'application de la méthode est illustré à l'annexe E.

V-4.6. Modélisation à l'aide de réseaux de neurones basés sur les composantes principales

L'application de composantes principales en modélisation par réseau de neurones a pour but de réduire la colinéarité des données, qui peut conduire à de fausses prédictions, cette application permet aussi la détermination des variables indépendantes de prédiction SOUSA (2007)

L'architecture de l'approche de réseau de neurones basée sur les composantes principales est montrée à la figure 2.8

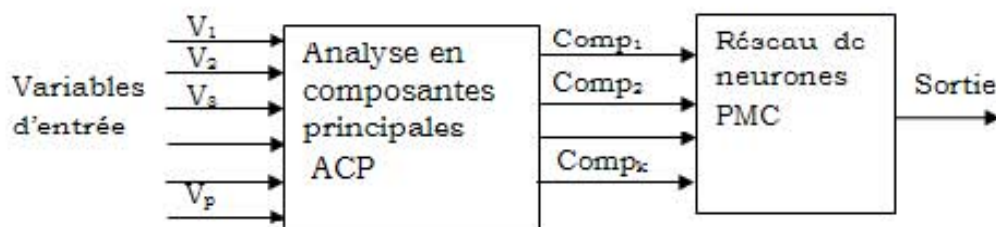


Figure 2.8 : Architecture du modèle neuronal basé sur les composantes principales

A La différence du modèle neuronal simple les entrées du modèle neuronal basé sur les composantes principales sont les composantes principales, par conséquent l'architecture du réseau est moins complexe.

L'utilisation du réseau de neurones basé sur les composantes principales permet selon HIMM-ELBLAU (2000) et AL-ABRI (2008) :

- La réduction des entrées et par conséquent la réduction de la taille et la structure du réseau.
- La réduction du temps d'apprentissage.
- Elimination des problèmes de surapprentissage
- Diminution des erreurs moyennes quadratiques en apprentissage et en validation.

AL-ALAWI (2008) mentionne que le modèle de prédiction combiné entre l'analyse en composante principale et le réseau de neurones est performant et présente un coefficient de corrélation élevé par rapport au modèle de régression neuronale simple, il trouve, à titre d'exemple, que le coefficient de corrélation pour une régression sur composantes principales est de 0,965 et pour une régression neuronale, il vaut 0,986, cependant il s'élève à 0,995 pour le modèle combiné, le modèle combiné permet de minimiser la racine carrée des erreurs.

V-4.13. Avantages et inconvénients des réseaux de neurones

L'intérêt porté au réseau de neurones tient sa justification des propriétés intéressantes qu'il possède, néanmoins ils présentent aussi un certain nombre de limitations ou d'inconvénients DAVALO (1993), OSORIO (1998), BOUTARFA (2006)

[1] Avantages de l'approche neuronale

- L'exploitation de connaissances empiriques : l'apprentissage à partir d'exemple (méthode d'apprentissage empirique) se fait d'une façon assez simple et permet d'obtenir de bons résultats.
- La robustesse
Dans les réseaux de neurones, la mémoire est distribuée, elle correspond à une carte d'activation de neurones, cette carte est en quelque sorte un codage des faits mémorisés ce qui attribue à ces réseaux l'avantage de résister aux bruits (pannes).
- La dégradation progressive
Les réseaux, de par leur nature continue, ne fonctionnent pas en tout ou rien et leurs performances ont plutôt tendance à diminuer progressivement en cas de problème (bruit, panne, entrée inconnue...) les réseaux de neurones sont moins vulnérables aux données approximatives et à la présence de données incorrectes dans la base d'apprentissage
- Le parallélisme massif
Bien que la plupart des réseaux connexionnistes soient réalisées sur des simulateurs séquentiels. Ces réseaux sont composés d'un ensemble d'unités de traitement de l'information qui peuvent opérer en parallèle.
- La prise en compte du non linéarité et du temps
Les réseaux de neurones artificiels présentent l'intérêt de pouvoir prendre en compte la non linéarité (les fonctions d'activation sont en général non linéaire). Certains réseaux peuvent aussi prendre en compte les aspects temporels (cas des réseaux récurrent)

2) Inconvénients de l'approche neuronale

- La difficulté de choix de l'architecture et des paramètres
Il n'existe pas de méthode automatique pour la meilleure architecture possible, pour un problème donné il est assez difficile de trouver la bonne topologie du réseau ainsi que les bons paramètres de réglages de l'algorithme.
- Le problème d'initiation et de codage
Les algorithmes d'apprentissage connexionnistes sont en général très dépendants de l'état initial du réseau (initiation aléatoire des poids) et de la configuration de la base d'apprentissage. Un mauvais choix des poids employés pour initier le réseau, de la méthode de codage des données ou même de l'ordre des données peut bloquer l'apprentissage ou poser des problèmes pour la convergence du réseau vers une bonne solution.
- Le manque d'explicabilité
Les réseaux connexionnistes sont en général des boîtes noires où les connaissances y restent enfermées et sont inintelligibles pour l'utilisateur. Un réseau ne peut pas expliquer le raisonnement qui l'a amené à une solution spécifique
- Le manque d'exploitation de connaissances théoriques
Les réseaux de neurones simples ne permettent pas de profiter des connaissances théoriques disponibles sur le domaine du problème traité. Ils sont dédiés à la manipulation de connaissances empiriques. Une façon simpliste de profiter des connaissances théoriques consiste à convertir des règles en exemples (prototype) cependant cette méthode ne garantit pas que ces exemples vont être bien représentés dans les connaissances du réseau à la fin de l'apprentissage.

V-4.8. Récapitulatif

Le tableau 2.2 suivant résume les choix faits dans cette étude, en terme d'architecture du réseau de neurones. Le terme architecture comprend le nombre de couches cachées, le nombre de neurones appartenant à ces couches, le type de fonction d'activation, l'algorithme d'optimisation retenu, et plus généralement des critères relatifs à l'apprentissage et à l'estimation des performances du modèle.

Rubrique	Critère retenu
Détermination de Paramètres d'entrée	Analyse en composantes principales. Analyse de classification.
Découpage de sets de données	Apprentissage : 2/3 de données pour optimiser les poids des connexions (échantillonnage aléatoire) Validation : 1/3 pour évaluer les performances du modèle (échantillonnage aléatoire)
Architecture du PMC	Neurones d'entrée = nombre de variable d'entrée "I" Neurones de sortie = nombre de variable de sortie Couche cachée = Une seule couche cachée Neurones cachées = le plus petit nombre produisant La plus faible erreur de prédiction sur le set de test (varie de 1 jusqu'à 2I+1)
Fonction d'activation	Pour la couche cachée : Tangente hyperbolique Logistique Pour la couche de sortie : linéaire
Algorithme d'apprentissage	Algorithme de BFGS
Nombre d'essais par configuration (initialisation de poids)	1000 - 5000
Nombre d'itération dans la phase de rétropropagation	200
Performance du modèle	R ² , Résidus moyen, Résidus studentisés, Racine carrée de l'erreur quadratique moyenne REQM
Qualité de prédiction du modèle	Ajustement individuel au modèle, Analyse de la normalité Analyse de l'homoscédasticité
Contribution des variables	la méthode de poids

Tableau 2.2 : Choix faits concernant la construction de modèle de régression neuronale

VI. CONCLUSION DU CHAPITRE

Les méthodes statistiques présentées dans ce chapitre sont le fondement théorique de notre étude de modélisation des paramètres de floculation.

La première partie de ce chapitre a été louée à la présentation de l'analyse multidimensionnelle, en particulier l'analyse en composantes principales et l'analyse de classification et à leur complémentarité.

L'analyse en composantes principales permet de repérer les groupes de variables fortement corrélées entre elles et éventuellement de détecter des caractères complexes sous-jacents à ces groupes, alors que les méthodes de classification ont pour but de regrouper ces variables en un nombre restreint de classes homogènes, au moyen d'algorithmes formalisés et non par des méthodes subjectives ou visuelles, les résultats obtenus permettent de faire un choix sur les variables explicatives les plus discriminantes pour la construction de modèles de régression.

La deuxième partie de ce chapitre a consisté à la présentation de la modélisation mathématique statistique, tout d'abord les différents types de modèles mathématiques ainsi que le principe général de la régression linéaire simple (l'outil de base de la plupart des modélisations).

Nous avons présenté ensuite les concepts fondamentaux de la modélisation par régression multiple; la sélection des données, la procédure et critères du choix du modèle, la qualité de prédiction et enfin la stabilité des modèles. Nous avons aussi introduit les définitions essentielles relatives aux réseaux neuronaux, notamment l'architecture d'un réseau de neurones à perceptron multicouche "PMC", et la mise en œuvre de la modélisation neuronale à savoir : la sélection des données, l'optimisation de la structure et la validation du modèle et enfin l'étude de contribution des variables dans les modèles neuronaux.

Par ailleurs, nous avons mis l'accent sur l'utilisation combinée de la régression multiple et de l'analyse en composantes principales (régression régularisée) et sur la régression neuronale basée sur les composantes principales, ces utilisations mixtes permettent, entre autre, de réduire la complexité des modèles et la colinéarité entre les variables descriptives et permettent des modèles parcimonieux.

Enfin des méthodologies complètes de conception et de mise en œuvre des modèles de régression multiple et de régression neuronale ont été récapitulées.

Chapitre 3

ETUDE EXPERIMENTALE DE LA FLOCCULATION DES SUSPENSIONS SYNTHETIQUES

I. OBJECTIF DE L'ETUDE EXPERIMENTALE

L'objectif de l'étude expérimentale est double : d'une part créer les bases de données expérimentales relatives à la floculation de suspensions synthétiques et d'autre part, apporter une contribution à la connaissance des mécanismes, la cinétique et la performance du procédé de floculation et en particulier l'incidence des interactions entre les caractéristiques du solide dispersé et les conditions opératoires sur la floculation. A cet effet, les recherches réalisées ont été conduites sur des suspensions synthétiques préparées à partir de différentes particules solides à savoir :

- Particules de kaolin.
- Particules de vase de barrage.
- Particules de tourbe végétale
- Particules de boue de station d'épuration.
- Mélanges de particules de vase et de tourbe.
 - Mélange I (72.8% Vase + 27.2% Tourbe)
 - Mélange II (37.2 % Vase + 62.8 % Tourbe)

Les suspensions de kaolin sont souvent utilisées comme une présentation synthétique des particules colloïdales présentes dans les eaux de surface, alors que les particules de vase utilisées par plusieurs auteurs, semblent a priori la simulation la plus proche des particules dispersées dans les eaux de surface. L'utilisation des particules de boue a pour objectif de mettre l'accent sur l'effet du caractère mixte (organo-minéral) sur la floculation. En revanche les particules de tourbe, vû sa déficience en matière minérale et richesse en matière organique et surtout les substances humiques, font d'elle un bon simulateur de colloïdes organiques naturelles.

II. PRESENTATION DES PARTICULES SOLIDES

a) Kaolin : Nom d'origine chinois, Kao Ling signifie lieu où l'on extrayait cette argile de Kao "élevée" et Ling "Colline", le kaolin est une roche argileuse formée essentiellement de la kaolinite et qui est blanche et friable. Elle résulte principalement de l'altération superficielle ou hydrothermale des roches magmatiques acides, riches en feldspaths potassiques et pauvres en minéraux ferromagnésiens FOUCAULT (1984).

b) Vase : Nom d'origine Hollandais "wase", sédiment meuble, gorgé d'eau, à éléments détritiques très fins (sablons, limons, argiles) à colloïdes en grande partie d'origine biologique (bactérie, diatomées, pollens, humus) et à hydroxydes et sulfures de fer. Les vases sont souvent sableuses, mais généralement pauvres en calcaire, la couleur de la vase varie selon sa richesse ; en matière organique (vase noire), en oxyde de fer (vase rouge), en limon (vase jaune), en glauconie parfois (vase verte). Elles sont thixotropes, lacustres et, plus souvent marines, elles sont en particulier abondantes dans les estuaires FOUCAULT (1984). La vase que nous avons utilisée provient du

barrage de Foum El Gherza ; de couleur jaunâtre, elle est utilisée après avoir été séchée, broyée et tamisée.

c) Boue : Elle provient de la station d'épuration des eaux usées urbaines de la ville de Tizi-Ouzou. De façon générale, la composition d'une boue est liée à la nature des eaux résiduaires, au procédé d'épuration et au mode de son traitement. La formule générale proposée pour une boue biologique est « $C_{60} H_{87} O_{23} N_{12} P$ » Mc CARTY (1970) auxquelles s'ajoutent les métaux adsorbés. La boue que nous avons utilisée, est produite par un traitement biologique, à faible charge, d'eau usée urbaine. Elle a été séchée, broyée et tamisée.

d) Tourbe : Selon le dictionnaire de science de sol, la tourbe est définie comme étant des dépôts de matières organiques dont la décomposition et l'humidification sont très lentes, suite à des conditions de milieu mal aéré et saturé en eau de façon permanente ou presque, en toute saison. La tourbe s'accumule en couches épaisses constamment imbibées d'eau et constituées de débris végétaux LOZET (1990).

Les principaux constituants de la tourbe sont la lignine, la cellulose, les substances humiques, les cires, les résines, les stéroïdes et les terpènes.

La tourbe possède des groupes chargés négativement (acides humiques et fulviques) qui se traduisent par une forte adsorption pour les cations et inversement pour les anions. elle permet la fixation de plusieurs métaux dont le cuivre, le cadmium, le fer, le mercure, le plomb, l'uranium et le zinc.

III. CARACTERISATION DES PARTICULES SOLIDES

Les particules solides présentées ci dessus serviront à la préparation des suspensions synthétiques. Ces matières solides ont fait l'objet d'une caractérisation assez complète SEKIOU (2001), dans ce qui suit nous illustrons l'essentielle des résultats de l'analyse de caractérisation.

III-1. Analyse par diffraction aux rayons x

Chaque espèce cristalline donne un diagramme de diffraction dont la position et l'intensité sont relatifs à sa structure. Le composé diffracte plus ou moins bien les rayons x suivant le degré de perfection de son réseau cristallin. Dans un mélange de substances cristallisées, il y a superposition des divers diagrammes BARON (1971).

Un diagramme de diffraction aux rayons x caractérise les points suivants :

- ♦ Position et forme des pics désignés soit par leurs raies (angle de diffraction 2θ) ou par leurs distances réticulaires (basales).
- ♦ Intensité des pics qui renseignent sur la concentration de l'élément présent.

Dans le cas des minéraux inter-stratifiés, il est nécessaire de faire subir à l'échantillon des traitements spéciaux (chauffage à $550^{\circ}C$, traitement à l'éthyle glycol ou traitement au KCl) permettant l'identification des minéraux

inter-stratifiés VATAN (1967). Nous présentons d'une façon très abrégée les résultats de la DRX.

a) Kaolin :

Le diagramme de diffraction aux rayons x du kaolin illustré par la figure 3.1, ci contre, montre l'existence de la Kaolinite, de l'illite, du quartz. Les traitements spéciaux du kaolin ont mis en exergue l'existence de minéraux inter-stratifiés "Illite- Montmorillonite". La composition en pourcentage est reportée sur la figure 3.4

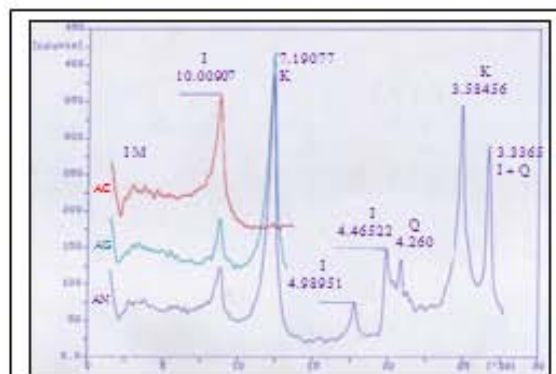


Figure 3.1 : Spectre DRX du kaolin

b) Vase :

Le diagramme DRX de la vase présenté dans la figure 3.2, ci-contre, met en évidence la multiplicité des éléments qui composent la vase à savoir la Kaolinite, l'illite et le Quartz. Les traitements spéciaux montrent également l'existence de La Montmorillonite, la chlorite et de minéraux inter-stratifiés "Illite-Montmorillonite".

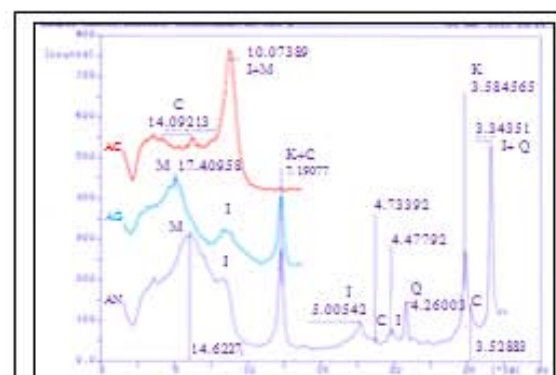


Figure 3.2 : Spectre DRX du la vase

c) Boue :

Le résultat de la DRX de la boue est rapporté sur la figure 3.3. La première observation que l'on peut faire c'est l'intensité modérée des pics comparativement aux deux premiers composés. Nous notons l'existence de la Kaolinite, de l'illite, de la Chlorite et du Quartz, la composition en pourcentage est reportée sur la figure 3.4

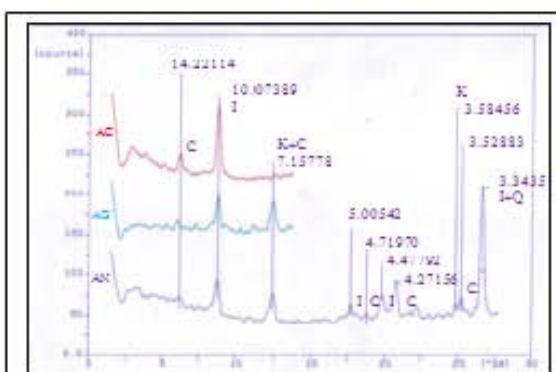


Figure 3.3 : Spectre DRX du la Boue

d) Tourbe :

L'étude par DRX de la tourbe est sans succès, la matière de la tourbe est amorphe (non cristallisée) et ne diffracte pas les rayons x. La composition en pourcentage de différentes particules solides est reportée sur le graphique suivant à la figure 3.4

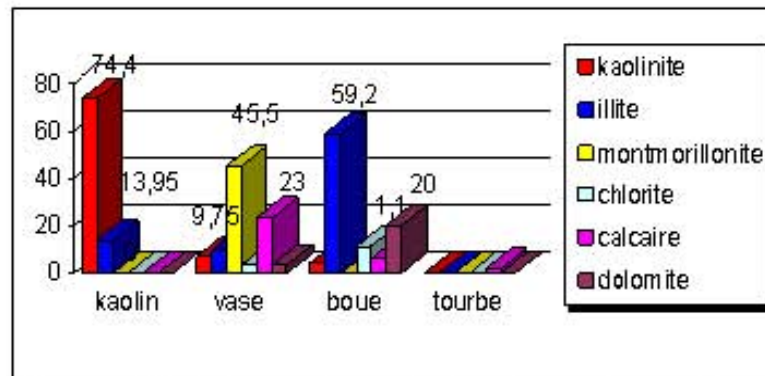


Figure 3.4 : Composition minéralogique des solides.

Conclusion

Les résultats d'analyse par diffraction aux rayons x des particules solides montrent que la kaolinite est le constituant essentiel de la fraction minérale du kaolin avec un taux de 74,4%, alors que la montmorillonite est le constituant principal de la fraction minérale de la vase (45,4%), cependant l'illite présente 59,2% de la fraction minérale de la boue, il en découle aussi que la vase et la boue sont plus carbonatées que le kaolin et la tourbe.

III-2. Analyse par spectroscopie infrarouge

Les échantillons analysés par spectroscopie Infrarouge sont traversés par un faisceau infrarouge dont les spectres sont réalisés entre les fréquences de 4000 à 500 cm^{-1} , les maximums d'absorption sont dus à deux types de vibrations atomiques : les vibrations d'élongation et les vibrations de déformation angulaire, l'analyse par spectroscopie infrarouge permet d'identifier les espèces minérales et groupements fonctionnels, d'étudier les hydrates et de suivre l'évolution de la structure des composés lors d'un traitement.

Un spectre IR se caractérise par trois points essentiels :

- ◆ Le nombre de bandes renseigne sur le nombre d'atomes constituant l'élément analysé (cas d'élément simple).
- ◆ La position de bande indique le mode de vibration des atomes et en particulier, le type de liaisons.
- ◆ L'intensité de bande renseigne sur la concentration de l'élément présent.

a) Kaolin :

L'analyse montre l'existence de bandes spécifiques de la kaolinite due aux vibrations de "Si-O" à 1032 cm^{-1} et de "Al-O" à 913,25 cm^{-1} (PENG 1994). Aussi le spectre révèle la présence de faible taux de matière organique indiqué par le petit pic à 1636,17 cm^{-1} . (voir figure 3.5)

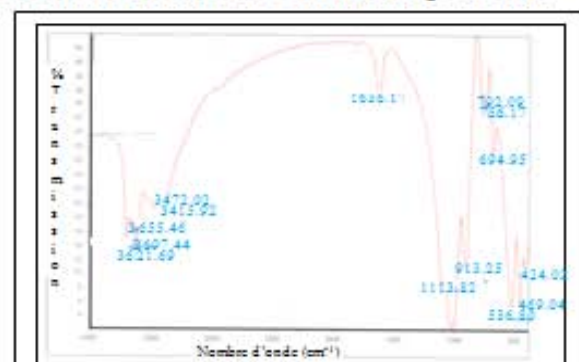


Figure 3.5 : Spectre IR du kaolin

b) Vase :

Le spectre fait apparaître la présence des groupes "C-P et C-N" à $2359,79 \text{ cm}^{-1}$ et la présence des groupes organiques marqués par les bandes $1435,67 \text{ cm}^{-1}$ et $1798,07 \text{ cm}^{-1}$. On note l'intensité remarquable du pic ($1632,10 \text{ cm}^{-1}$) et le faible pic représentatif du groupe "Al-O" ($917,62 \text{ cm}^{-1}$) comparativement aux pics du spectre du kaolin. (voir figure 3.6)

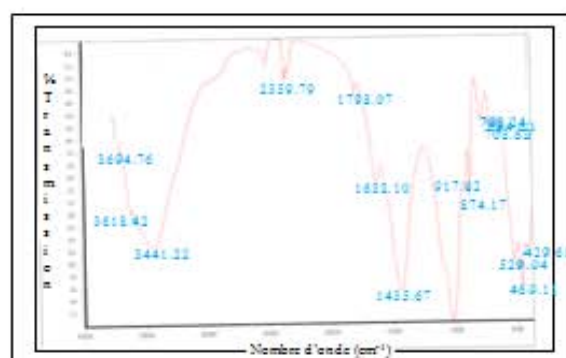


Figure 3.6 : Spectre IR de la vase

c) Boue :

Le spectre IR de la boue ne diffère pas beaucoup de celui de la vase, quoique, il présente plusieurs pics dus à la présence des groupes organiques et un faible pic dû aux vibrations du groupe "Si-O", comparé à ceux de kaolin et de la vase. (voir figure 3.7),

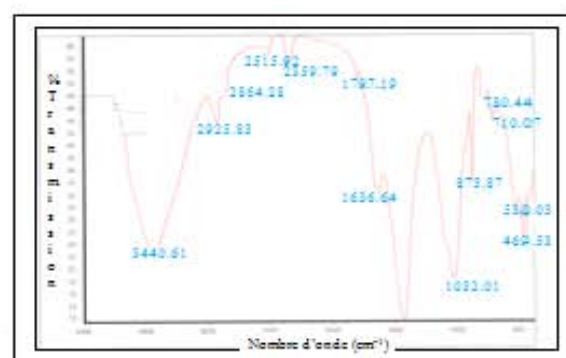


Figure 3.7 : Spectre IR de la Boue

d) Tourbe :

Le résultat d'analyse IR met en exergue la présence des groupements fonctionnels suivants :

- CH_2 et CH_3 à $2921,37 \text{ cm}^{-1}$
- C-P à $2360,08 \text{ cm}^{-1}$
- plusieurs groupes à $1621,87 \text{ cm}^{-1}$.
- CH_2 et C-N à $1386,14 \text{ cm}^{-1}$
- C-H aromatiques et C-O acide carboxylique ou éther à $1059,92 \text{ cm}^{-1}$ caractérisant l'acide humique (CHEMAT (1999) (voir figure 3.8)

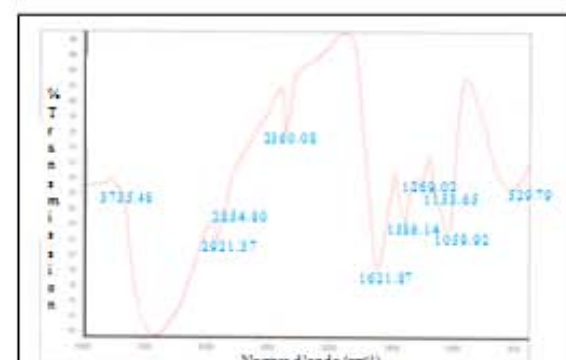


Figure 3.8 : Spectre IR de la Tourbe

Conclusion

Etant donné la diversité des constituants qui entrent dans la composition de chaque composé, il n'est pas possible d'assigner chaque bande à une liaison donnée et c'est plutôt l'ensemble de bandes qui est caractéristique du composé étudié.

- Le spectre du kaolin et de la vase se distinguent par un maximum d'absorption à faible énergie du probablement aux vibrations des liaisons "Si -O" et "Al -O".

- En revanche, le spectre de la tourbe indique plusieurs bandes d'absorption de moyenne énergie dues aux vibrations de groupements fonctionnels organiques, et spécialement les substances humiques.
- Quant à la boue, son spectre indique la présence modérée de matières organique et minérale.

III-3. Analyse par spectroscopie d'absorption atomique

En spectroscopie par absorption atomique, une population d'atomes est générée à l'état fondamental dans un atomiseur (cellule d'absorption) et éclairée par un rayonnement lumineux de longueur d'onde λ_0 et d'intensité I_0 . Lors du passage de ce rayonnement à travers le nuage atomique, les atomes au niveau fondamental (E_0) peuvent absorber de la lumière de telle sorte que, à la sortie, l'intensité lumineuse transmise est égale à I et l'absorbance A est liée à la concentration du produit selon une loi de proportionnalité du type BEER-LAMBERT.

Les résultats d'analyse comme l'illustre la figure 3.9 montrent que le kaolin est composé essentiellement de "SiO₂ et Al₂O₃", en revanche, la vase et la boue présentent des pourcentages moyens de "SiO₂ et Al₂O₃". Quant à la tourbe, la teneur des différents oxydes métalliques est négligeable.

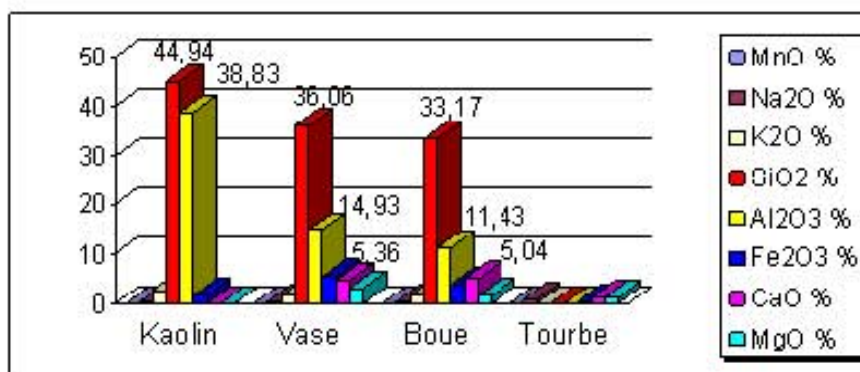


Figure 3.9 : Résultats d'analyse d'absorption atomique

II-4. Analyse de carbone organique total, carbone minéral et perte au feu (à 550 °C)

Les résultats d'analyse de COT (carbone organique total), CM (carbone minéral) et la perte au feu à 550°C présentés à la figure 3.10 ci-après montre que la tourbe est riche en matière organique suivie par la boue, puis la vase et en dernier lieu, le kaolin. La vase et la boue présentent des taux élevés de carbone minéral comparés à ceux du kaolin et de la tourbe.

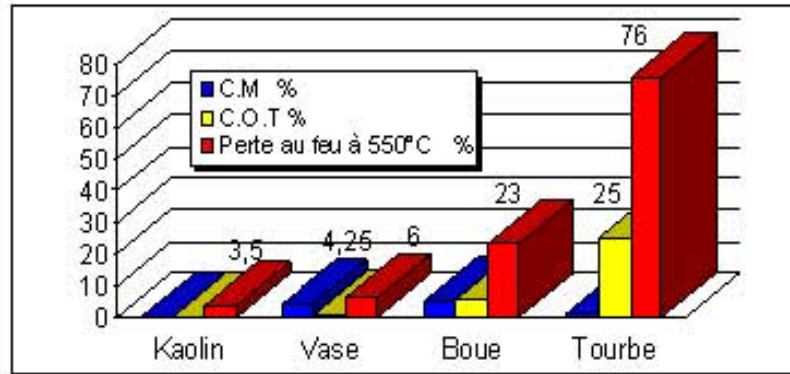


Figure 3.10 : Résultats d'analyse de C.O.T ; C.M et perte au feu (MO).

III-5. Analyse granulométrique au laser

L'analyse granulométrique est une ancienne technique qui a pour objet la caractérisation et la classification du sol par la détermination de la grosseur de ses grains. Le granulomètre au laser permet, entre autres, la détermination de la courbe granulométrique, le diamètre moyen en (volume et surface), la surface spécifique et l'uniformité. Son principe de fonctionnement est basé sur l'interaction d'un faisceau parallèle de lumière (rayons laser) avec les particules de dimensions supérieures à la longueur d'onde du rayonnement, cette interaction cause des diffractions. L'analyse de l'énergie diffractée sur une surface perpendiculaire au faisceau incident entre deux limites S_1 et S_2 donne la surface projetée des particules de même classe granulométrique CASELLAS-SALHA (1981). Les résultats de l'analyse granulométrique en fraction et en cumul des différentes particules solides sont illustrés dans la figure 3.11.

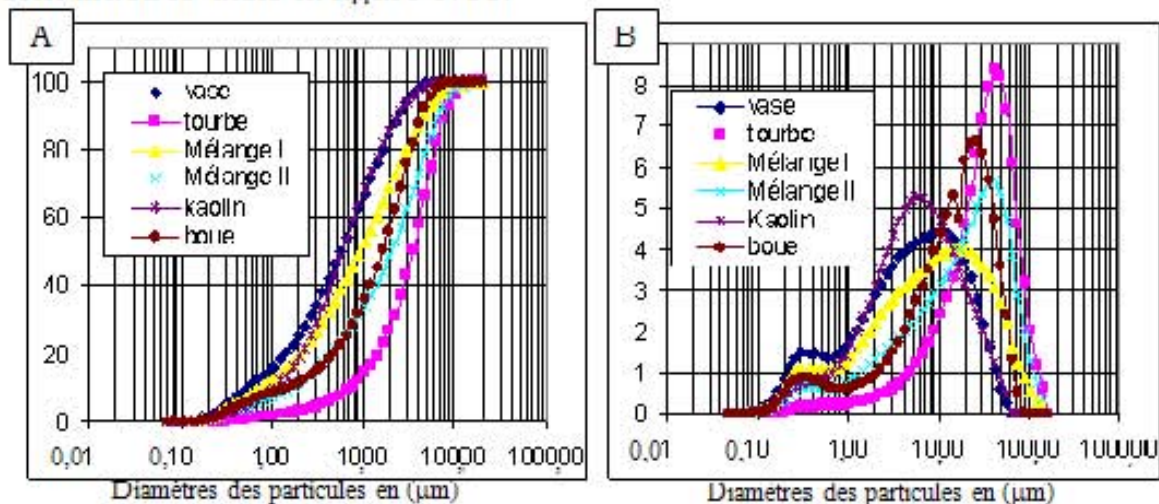


Figure 3.11 : Distributions granulométriques des différents solides (%)

[A] : En cumul, [B] : En fraction

Les résultats montrent que les particules primaires des solides dispersés présentent des fractions colloïdales très différentes ; la fraction colloïdale du kaolin présente plus de 67 % alors qu'elle n'atteint pas 35% en cas de particules de boue et 15 % pour les particules de tourbe (voir figure 3.12).

Le kaolin, la vase et le mélange I semblent présenter des distributions de particules primaires larges et uniformes indiquées par les valeurs élevées de l'uniformité et de la valeur de "Span" alors que la tourbe et le mélange II se distinguent par leurs distributions serrées de faibles uniformités. (Voir figures 3.11 et 3.13).

Uniformité: écart absolu à partir de la médiane $D(0,5)$

Span (largeur de la distribution) $span = \frac{D(0,9) - D(0,1)}{D(0,5)}$.

$D(0,9)$: la taille pour laquelle 90% de l'échantillon se trouve en dessous de cette dimension [μm], $D(0,5)$: la taille pour laquelle 50% de l'échantillon se trouve en dessous de cette dimension [μm], $D(0,1)$: la taille pour laquelle 10% de l'échantillon se trouve en dessous de cette dimension [μm]

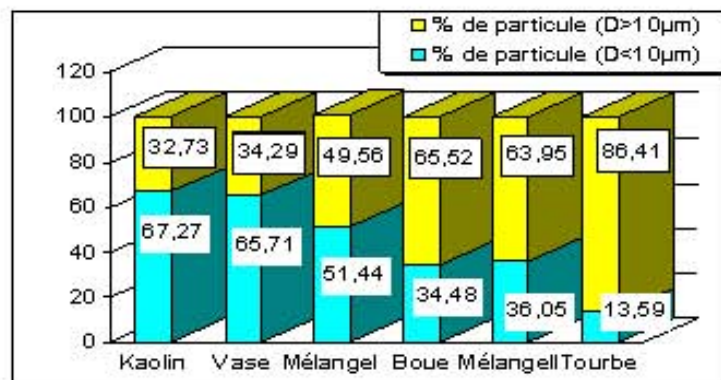


Figure 3.12 : Pourcentages des fractions colloïdales des particules solides

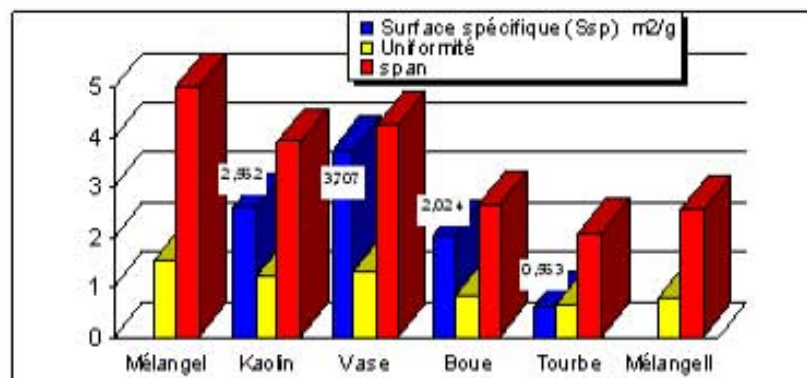


Figure 3.13 : Valeurs de surface spécifique, uniformité et "span" des particules solides

A l'exception de la tourbe dont le diamètre moyen en surface est de $10,66 \mu\text{m}$, les autres composés ont des diamètres inférieurs à $3 \mu\text{m}$ (voir figure 3.14), ceci revient à l'effet d'agglutination des particules solides hydrophiles, aspect confirmé par la valeur élevée de la fraction volumique des particules de tourbe ($4,71 \cdot 10^{-2}$) suivie par celle de la boue ($2,07 \cdot 10^{-2}$) puis celles du kaolin et de la vase ($1,11 \cdot 10^{-2}$; $0,84 \cdot 10^{-2}$).

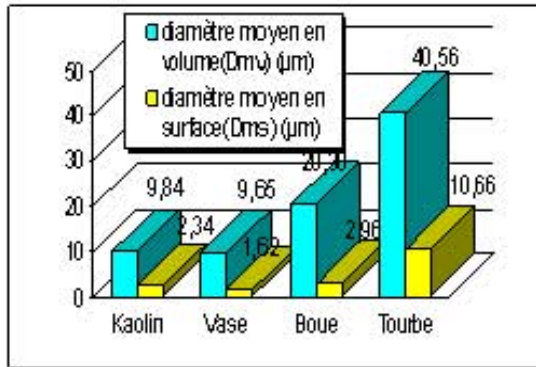


Figure 3.14 : Diamètres moyen en volume et en surface des particules solides

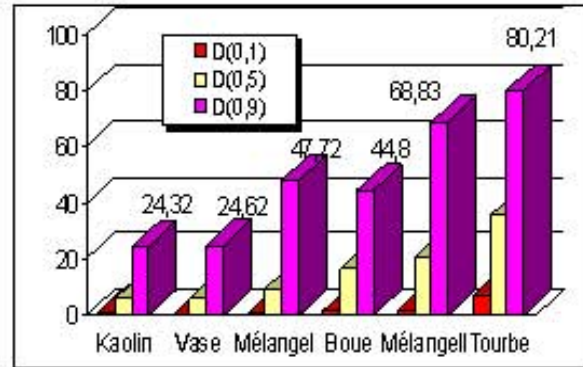


Figure 3.15 : Valeurs de "percentiles" D(0,1), D(0,5) et D(0,9) des particules solides.

Plus les particules s'agglutinent, plus la morphologie devient irrégulière. C'est ce que confirment les valeurs élevées des paramètres D(0,1), D(0,5), D(0,9) observées dans le cas de la tourbe et de la boue. (Voir figure 3.15).

Il en résulte aussi (voir figure 3.13), qu'à part la tourbe qui a une surface spécifique relativement faible ($S_{sp} = 0,5 \text{ m}^2/\text{g}$), les trois autres composés ont des surfaces spécifiques assez proches, la boue ($2,02 \text{ m}^2/\text{g}$), le kaolin ($S_{sp} = 2,56 \text{ m}^2/\text{g}$) et la vase ($S_{sp} = 3,7 \text{ m}^2/\text{g}$), ces dernières valeurs sont en bonne conformité avec celles indiquées par DENTEL (1988) qui note que la surface spécifique des silicates est de l'ordre de (1,35 à $4,52 \text{ m}^2/\text{g}$).

Conclusion

Ces différentes analyses montrent que les solides dispersés présentent des caractéristiques très distinctes et laissent présager que leurs aptitudes à la floculation seront éventuellement différentes, il est donc intéressant d'étudier dans ce qui suit :

1. L'influence des caractéristiques du solide en particulier, le taux de matière organique, la distribution granulométrique, le diamètre et la surface spécifique sur :
 - La performance et mécanisme de floculation.
 - La cinétique de la floculation.
2. Les interactions entre la matière minérale, la matière organique et les paramètres de floculation et leurs effets sur la cinétique et la performance de la floculation.

IV. ESSAIS DE FLOCCULATION DES SUSPENSIONS SYNTHETIQUES

IV-1 SUSPENSIONS SYNTHETIQUES

Préparation

La préparation des suspensions synthétiques est opérée en suivant les étapes ci-après :

- 1) Séchage à 105°C des particules solides pendant deux heures.
- 2) Préparation de la solution mère à 1 g/l avec de l'eau distillée.
- 3) Agitation de la solution préparée durant vingt quatre heures.
- 4) A partir de cette solution mère et avec de l'eau de robinet, on procède à des dilutions pour préparer les suspensions synthétiques.

Courbes d'étalonnages :

1) Turbidité (NTU)-Concentration des particules solides (mg/l)

Les courbes d'étalonnage rapportées sur la figure 3.16 pour les différentes suspensions préparées, (du kaolin, de la vase, de la boue, de la tourbe, du mélange I et du mélange II) montrent une parfaite linéarité entre la concentration en particules et la turbidité. Toutefois, nous observons un grand écart de la valeur de la pente entre les droites correspondantes aux suspensions à caractère minéral et celles à caractère organique, ceci est dû essentiellement, comme l'affirme DESJARDINS (1996) et MASSEHLEIN (1996) à la nature "hydrophile ou hydrophobe" du colloïde. À titre d'exemple pour une concentration de 50 mg/l de particules solides, les turbidités correspondantes des suspensions du kaolin, de la vase, de la boue, de la tourbe et des deux mélanges I et II sont respectivement 19 ; 25 ; 19 ; 4,8 ; 19 et 10 NTU.

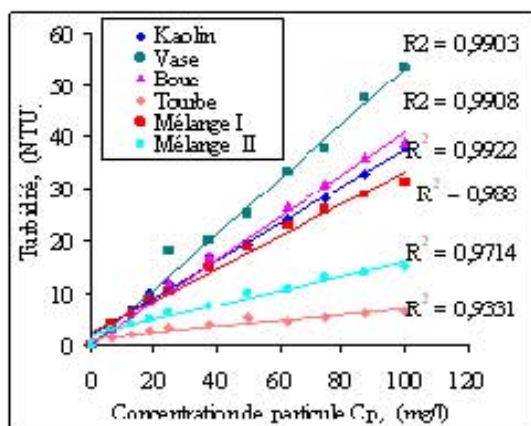


Figure 3.16 : Courbes d'étalonnage (Turbidité - Concentration des particules)

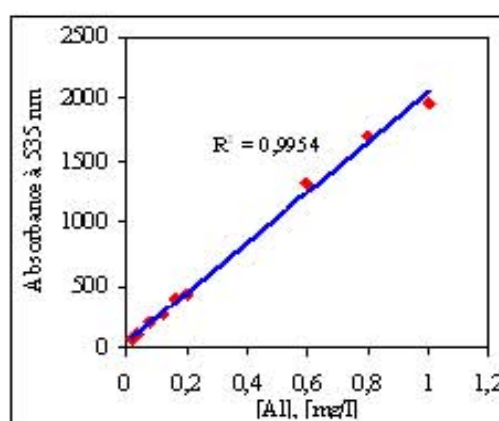


Figure 3.17 : Courbe d'étalonnage (Absorbance à 535 nm - Concentration de l'aluminium)

2) Absorbance à 535 nm - concentration de l'aluminium

La figure 3.17 illustre la courbe d'étalonnage Absorbance = f (concentration de l'aluminium), à la base de cette courbe, la quantification de l'aluminium résiduel est déduite de la mesure de l'absorbance à 535 nm d'un complexe avec l'ériochrome cyanine AUBERT (1978).

VI-2 DESCRIPTION DES ESSAIS DE FLOCCULATION

Les essais sont réalisés selon le protocole du "jar test" ; Chacune des suspensions synthétiques est versée dans quatre béciers de 01 litre et mélangée avec différentes doses de coagulant, le sulfate d'aluminium ($Al_2(SO_4)_3 \cdot 16H_2O$), préparé à chaque essai jar test en vue d'éviter toute évolution d'espèces présentes et de dégradation de capacité coagulante selon PARTHASARATHY (1985), BEECROFT (1995) et PA HO HSU (1997).

Le coagulant est introduit en un seul jet dans la zone active du mélange à l'aide d'une seringue, conformément aux travaux de ZOUGRANA (1993) et CARBONNIER (1997). Un mélange rapide est réalisé afin d'optimiser la phase de déstabilisation des particules avec une puissance d'agitation importante ($Pr = 6.724 \text{ w/m}^3$ puissance maximale du jar test utilisé). Au terme de deux minutes d'agitation rapide (suite aux travaux de LAHOUSSINE (1992), et DESJARDINS (1998)) la vitesse de rotation de l'agitateur est baissée subitement afin de rendre possible la floculation (mélange lent), cette phase est assurée pendant 10 minutes (T_f) conformément aux travaux de DESJARDINS (1996) et ADDACHI (1997). Ensuite, les mélangeurs sont arrêtés et succède alors, une période de décantation (T_d) qui dure 30 minutes. Le surnageant est siphonné à 1 cm en dessous de la surface d'eau selon ZOUGRANA (1993) et CARBONNIER (1997) avec un tuyau de diamètre 3 mm à une faible vitesse pour éviter toute turbulence et tout risque de cassage des floccs formés GIBBS (1982).

La performance de la floculation est évaluée par deux critères :

- Mesure de turbidité finale du surnageant
- Mesure de la teneur résiduelle en aluminium

VI-3 RESULTATS ET DISCUSSION

VI-3.1. Evolution de la performance de la floculation

La recherche du dosage optimal pour les différentes suspensions étudiées (du kaolin, de la vase, de la boue, de la tourbe, du mélange I et du mélange II), montrent l'existence de deux zones distinctes. (Voir figure 3.18)

- Une première zone décroissante, où la turbidité finale décroît avec l'augmentation de la dose du coagulant.
- Une deuxième zone où la turbidité finale commence à croître à partir d'une dose limite (dose optimale) indiquant un surdosage menant à une restabilisation des suspensions.

Cet aspect de restabilisation laisse présager que pour toutes les suspensions minérales et organiques, l'adsorption et la neutralisation de charge interviennent dans leurs mécanismes de floculation ;

Pour les suspensions minérales ce mécanisme est majoritaire, par contre la faible restabilisation dans le cas des suspensions organiques de tourbe et de mélange II comme le montre la figure 3.18 s'explique par la présence de plusieurs mécanismes compétitifs entre les espèces hydrolysées de l'aluminium et la matière organique.

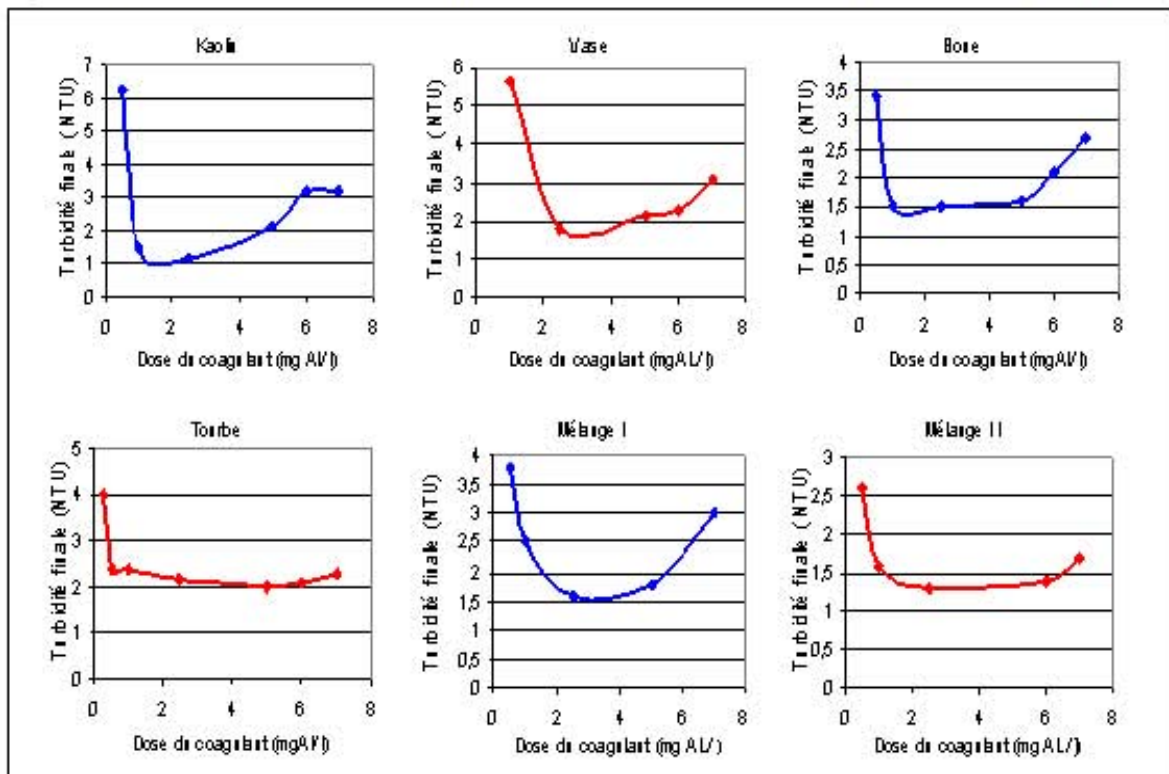


Figure 3.18 : Evolution de la turbidité finale en fonction de la dose du coagulant avec : Pui (puissance d'agitation lente) = 0,9 w/m³, pH= 6,5, Tf (temps de floculation) =10 min, Pr (puissance d'agitation rapide) = 6,724 w/m³, Td (temps de décantation)= 30 min, Cp (concentration de particules solides) = 50 mg/l,

L'évolution de la dose optimale en fonction du taux de matière organique dans le solide dispersé comme l'illustre la figure 3.19 est un résultat qui mérite beaucoup d'attention ; On note une consommation en coagulant des suspensions organiques, nettement supérieure aux suspensions minérales, à titre d'exemple : dans les mêmes conditions opératoires de pH, de concentration des particules Cp, de temps de floculation tf et de temps de décantation td, les suspensions de tourbe à 50 mg/l nécessitent une dose de 5 mgAl/l alors que celle-ci n'est que 2,5 mgAl/l pour les suspensions minérales.

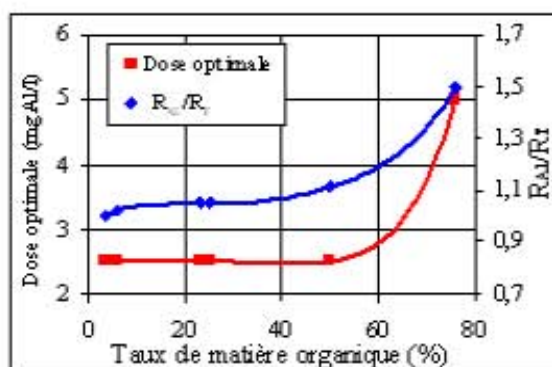


Figure 3.19 : Evolution de la dose optimale et du rapport R_{RAL}/R_T en fonction du taux de matière organique.

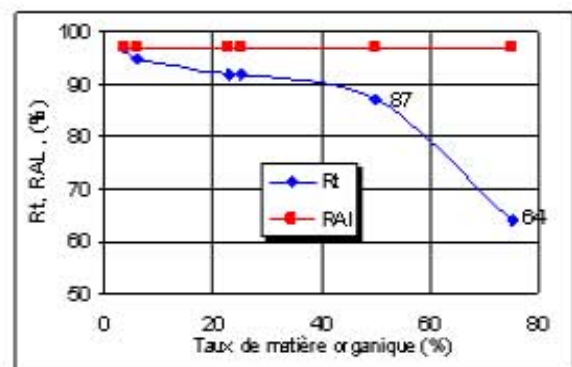


Figure 3.20 : Evolution du rendement R_T et du taux R_{RAL} en fonction du taux de matière organique.

L'examen des évolutions du rendement d'élimination de la turbidité R_T , du taux de consommation du coagulant R_{Al} et du rapport R_{Al}/R_T reportées sur les deux figures 3.19 et 3.20 indique qu'au moment où l'efficacité d'élimination de la turbidité dépasse les 95 % pour les suspensions minérales (kaolin et vase) celle-ci ne dépasse guère 65% pour la tourbe. Il en découle que la présence, à faible concentration, de la matière organique dans le solide dispersé est d'une influence minime sur le rendement d'élimination de la turbidité, alors que l'effet négatif inhibiteur est marqué pour des taux de matière organique qui dépassent 25%.

Compte tenu de l'étude bibliographique sur la floculation des suspensions colloïdales, il ressort que le rendement d'élimination de la turbidité des suspensions organiques dépend essentiellement, selon HUANG (1996), de la quantité de colloïdes organiques et des caractéristiques de la matière organique, (la masse molaire, le nombre, la nature et le degré d'ionisation des groupements fonctionnels) et de la présence de substances humiques et l'acidité carbonyle selon COLLINS (1986) et JULIEN (1994).

Ce faible rendement revient aussi à l'hydrophobicité, la taille et la densité de charge des colloïdes selon PURNENDU (1998) et à la structure minéralogique, les propriétés surfaciques et surfaces de complexation selon LURIE (1996) et PURENDU (1998) et aussi à la nature et au nombre de fonctions acides et nombre d'atome de carbone selon LEFEBVRE (1993).

La figure 3.19 montre aussi que le rapport R_{Al}/R_T augmente avec l'accroissement du taux de matière organique ; il est de 1/1 pour les suspensions à caractère minéral (kaolin et vase) et il aborde 1,5/1 en cas des suspensions organiques de tourbe, cette consommation excessive du coagulant en cas des particules de tourbe revient à priori, à la double consommation de l'aluminium ; à la formation du $Al(OH)_3$, et à la complexation du cation Al^{3+} sur la matière organique et en particulier humiques conformément aux notations de JENKINS (1980), qui mentionne que la matière organique humique complexe les cations métalliques de Fe et Al, Il se forme ainsi de l'humâtes d'aluminium insoluble grâce à l'interaction chimique entre les espèces hydrolysées cationiques de l'aluminium et les groupements fonctionnels organiques.

Les résultats montrent aussi que les suspensions de boue présentent une même aptitude à la floculation que celles du kaolin et de la vase ; le rendement d'élimination de la turbidité et le taux de consommation du coagulant atteints sont respectivement 92% et 97%, ce résultat est tout à fait cohérent à cause de la minéralisation de la boue et de certaines caractéristiques physico - chimiques notamment la surface spécifique et le caractère argileux (% SiO_2 et Al_2O_3 élevé) proches de celles du kaolin et de la vase. Néanmoins, suite aux travaux de WU (2002) sur la structure des floccs

de boues activées, il vient que la floculation des particules de boue peut être gouvernée par le mécanisme de pontage que de neutralisation de charge.

Pour étayer davantage certains aspects liés aux caractéristiques du solide dispersé et leurs effets sur la performance de la floculation, nous présentons ci-après sur les figures 3.21 à 3.24 l'évolution du rendement d'élimination de la turbidité en fonction de la surface spécifique, le diamètre moyen des particules solides, la largeur de distribution granulométrique "span" et le taux de "SiO₂ et Al₂O₃"

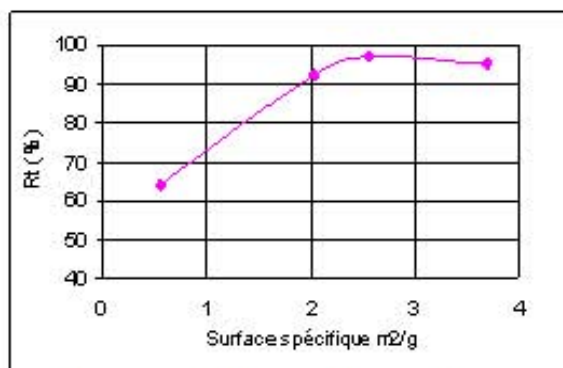


Figure 3.21 : Evolution du rendement R_T en fonction de la surface spécifique.

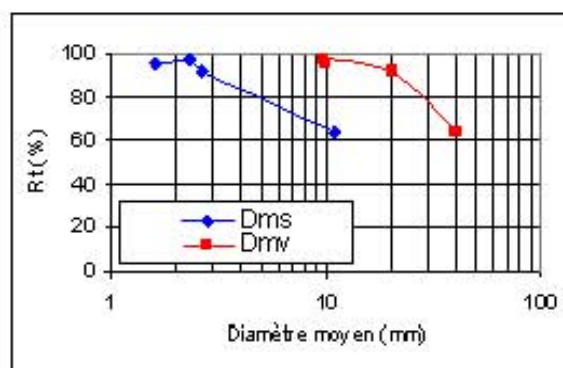


Figure 3.22 : Evolution du rendement R_T en fonction du diamètre moyen de particules.

Le rendement d'élimination de la turbidité est tributaire des caractéristiques des particules solides dispersées, la figure 3.21 montre que le rendement d'élimination de la turbidité s'élève suite à l'accroissement de la surface spécifique. En effet, la faible valeur de la surface spécifique de la tourbe fait en sorte que le rendement de séparation liquide - solide est faible. En effet, pour des surfaces spécifiques élevées la surface d'adsorption est grande permettant une meilleure déstabilisation, adsorption et neutralisation de charge.

A travers la figure 3.22 présentative de l'évolution de rendement d'élimination de la turbidité en fonction du diamètre primaire, nous constatons que pour les grands diamètres, l'efficacité de la floculation est faible, cela revient a priori à la fragilité des floes formés de colloïdes de grands diamètres. Ce point va être plus explicité dans les sections qui suivent.

Peu de recherches ont été consacrées à l'effet de la distribution granulométrique des particules primaires sur la floculation, la figure 3.23 ne semble pas expliquer une évolution significative de la performance de floculation en fonction de la largeur de la distribution "span", néanmoins nous constatons que pour les faibles valeurs de "span" la performance de floculation est minime. Quant au taux de "SiO₂ et Al₂O₃", la figure 3.24 montre que l'accroissement de ce taux dans le solide dispersé améliore l'efficacité d'élimination de la turbidité, en effet la matière minérale argileuse durci et renforce la structure du floe.

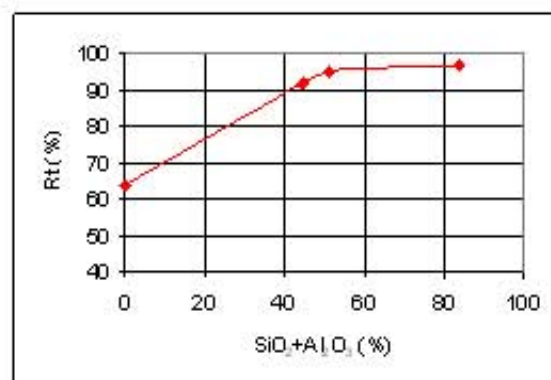
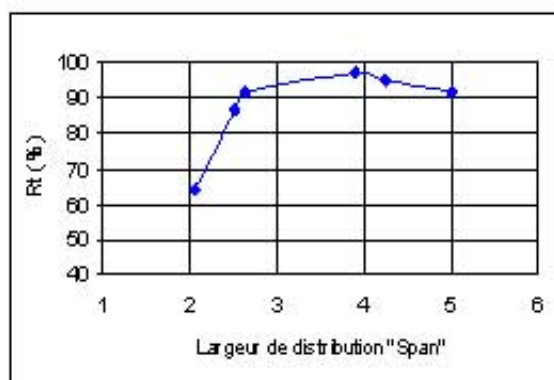


Figure 3.23 : Evolution du rendement R_T en fonction du "span" $span = \frac{D(99) - D(1)}{D(5)}$

Figure 3.24 : Evolution du rendement R_T en fonction de pourcentage de "SiO₂ + Al₂O₃"

VI-3.2. Interactions entre la puissance d'agitation, la dose du coagulant et la nature du solide dispersé

Le gradient d'agitation lente affecte profondément la turbidité finale, des travaux antérieurs montraient que le gradient optimal dépend de la nature du floc formé, sans aucune allusion à l'influence du dosage du coagulant (voir plus de détail au chapitre 1), à cet effet, nous voulons, à travers cette section, montrer l'interaction entre ces trois paramètres qui sont : la puissance optimale d'agitation lente "Pui", la dose du coagulant, et la nature du floc formé. L'expérimentation faite pour toutes les suspensions, (voir figures 3.25 et 3.26), montre l'existence d'une puissance optimale d'agitation permettant d'obtenir une meilleure élimination de la turbidité.

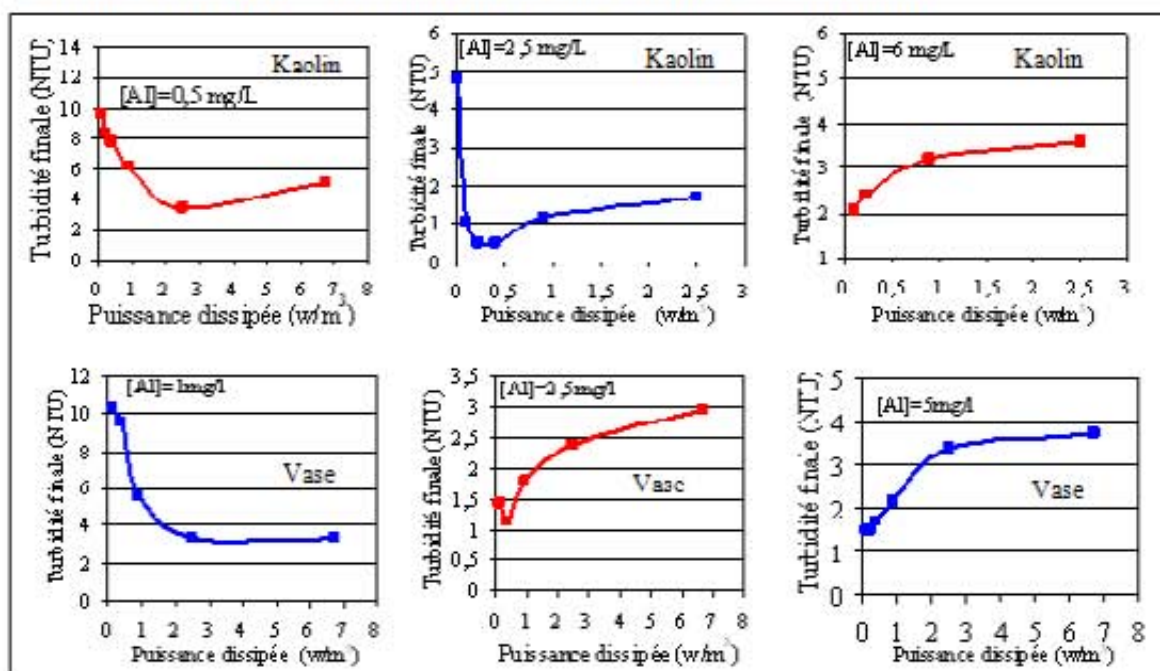


Figure 3.25 : Evolution de la turbidité finale en fonction de la puissance dissipée Pour différentes doses du coagulant et pour les suspensions du kaolin de la vase avec : Pr (puissance d'agitation rapide) = 6,724 w/m³ pH= 6,5, Tf (temps de floculation) =10 min, Cp (concentration de particules solide) = 50 mg/l, Td (temps de décantation)= 30 min

Une puissance inférieure à la valeur optimale est insuffisante pour assurer les contacts et la formation des floccs, une puissance plus grande provoque un cassage des floccs formés et par conséquent, une défloculation des amas formés et augmentation de la turbidité.

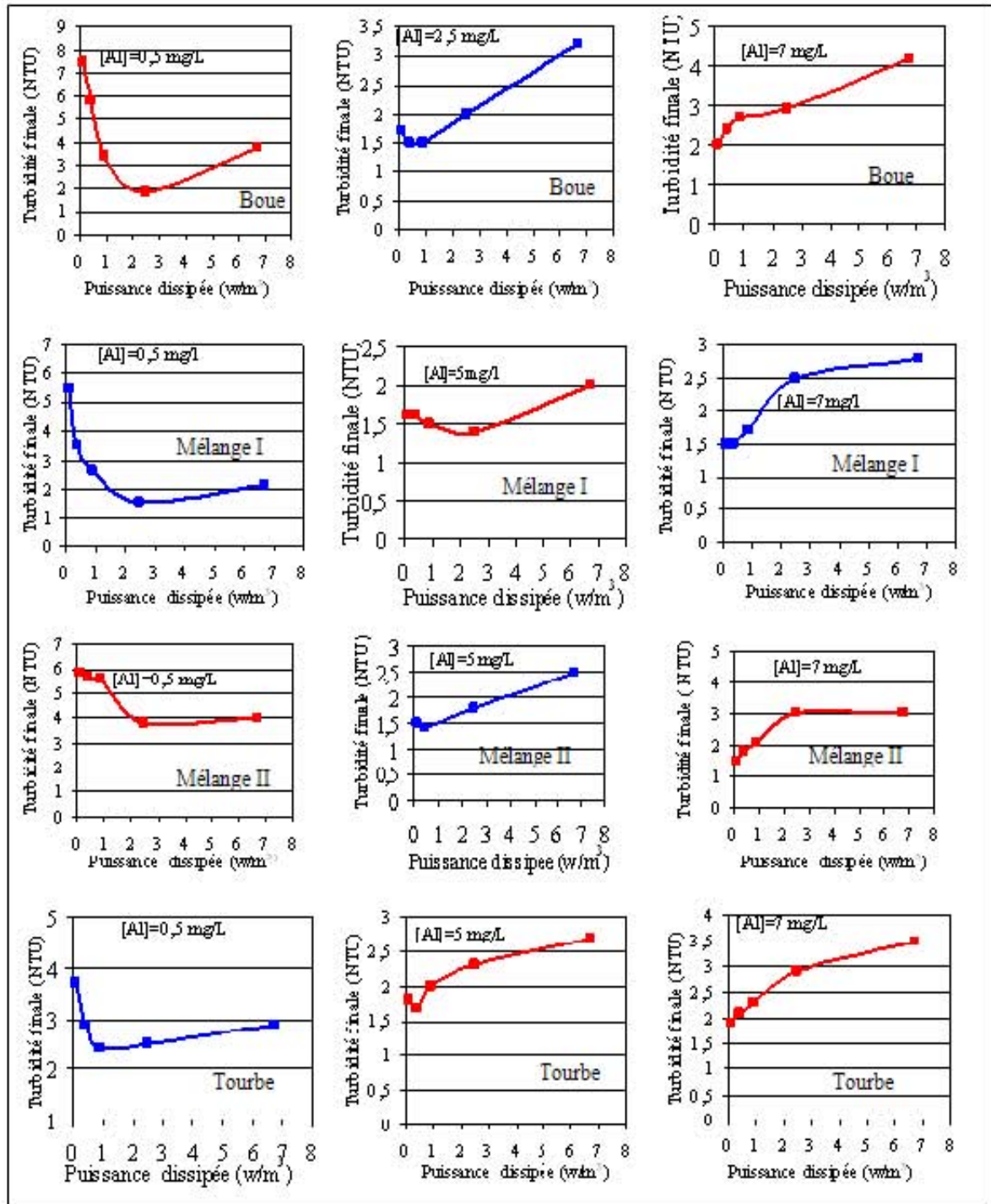


Figure 3.26 : Evolution de la turbidité finale en fonction de la puissance dissipée Pour différentes doses du coagulant et pour les suspensions de la boue, du mélange I, du mélange II et de la tourbe, avec : Pr (puissance d'agitation rapide) = 6,724 w/m³, pH= 6,5, Tf (temps de floculation) =10 min, Cp (concentration de particules solide) = 50 mg/l, Td (temps de décantation)= 30 min.

La marge de la puissance d'agitation lente optimale trouvée varie de 0,1 à 2,5 w/m^3 ce qui correspond à un gradient de vitesse "G" variant de 10 s^{-1} à 50 s^{-1} et GT (critère de camp) variant de 6.10^3 à 3.10^4

- Pour les faibles valeurs de puissance, le mécanisme de croissance de floc prédomine.
- Pour des puissances supérieures à 2,5 w/m^3 le mécanisme du cassage du floc est important.

L'observation minutieuse des figures 3.25 et 3.26 révèle que la puissance optimale d'agitation lente est attachée, à la fois, à la dose du coagulant injecté et à la nature du solide dispersé comme le montre la figure 3.27 suivante :

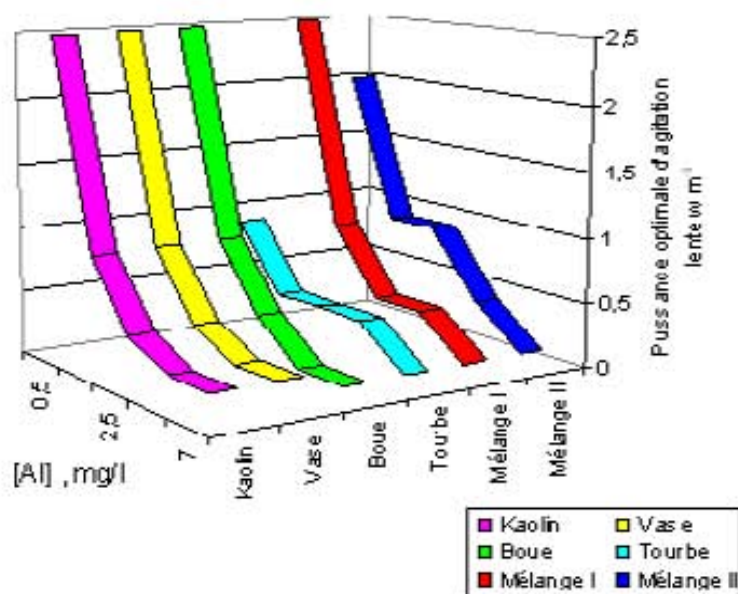


Figure 3.27 : Evolution de la puissance optimale d'agitation lente en fonction de la nature du solide dispersé et de la dose du coagulant injecté.

A première vue, on observe que la puissance optimale d'agitation lente décroît fortement quand la dose du coagulant augmente. Il apparaît aussi que pour les faibles dosages en coagulant, les puissances optimales sont différentes ; pour une concentration de coagulant de 0,5 mg Al/l, les suspensions du kaolin, du mélange I et de la boue exhibent une puissance d'agitation de 2,5 w/m^3 alors que la puissance atteint respectivement pour le mélange II et la tourbe : 2 w/m^3 et 0,7 w/m^3 .

Il en découle alors, que la puissance optimale d'agitation lente, dépend à la fois du dosage en coagulant et de la nature du solide dispersé. Effectivement, en présence de la matière organique et à faible dosage de coagulant c'est la complexation par des espèces monomères qui est majoritaire et les floccs formés sont alors fragiles, comme c'est illustré dans ce qui suit :

D'après JARVIS (2005), les floccs formés par complexation sont fragiles par rapport aux floccs formés par neutralisation ou adsorption à partir des espèces dimères, trimères ou tridécamères (polymérisées et moyennement polymérisées).

Et suite à son étude de l'élimination des substances humiques, SPENGLER (1983), affirme qu'aux faibles dosages en coagulant l'agglomération de particules organiques est réalisée mais la dose est insuffisante pour promouvoir une bonne floculation et formation de floccs séparables par décantation.

Ainsi selon MIETTA (2009a), les floccs organiques ont des diamètres supérieurs aux diamètres des floccs minéraux donc à priori sensibles aux fortes agitations et leur résistance au cassage est faible.

Alors l'interaction établie revient principalement au taux de matière organique dans le solide dispersé et au dosage du coagulant, ces deux facteurs dictent les espèces d'aluminium présentes et le mécanisme de floculation et par conséquent influent sur la sensibilité des floccs formés aux agitations.

Au vu de ces résultats, les sections suivantes auront pour objectif l'étude de la cinétique de floculation des suspensions colloïdales et des effets de l'interaction entre la matière organique et la matière minérale sur la performance de la floculation.

VI-3.3. Etude de la Cinétique de la floculation

Le modèle d'Argaman Kauffman BRATBY (1977) décrivant la cinétique de la floculation dans un flocculateur parfaitement agité, s'exprime :

$$\frac{n_1}{n_0} = \frac{K_A}{K_B} G + \left(1 - \frac{K_A}{K_B} G\right) e^{-K_B G T} \quad (\text{IV-1})$$

n_1 : concentration en nombre de particules primaires non flocculées (particules/m³).

n_0 : concentration en nombre de particules primaires (particules /m³).

K_A : constante de formation (agrégation)

K_B : constante de cassage [s].

G : valeur moyenne du gradient de vitesse [s⁻¹].

T : période de floculation [s]

L'application de ce modèle aux résultats expérimentaux (voir figure 3.28), a permis de déterminer les constantes cinétiques K_A et K_B , le temps et le rendement d'équilibre.

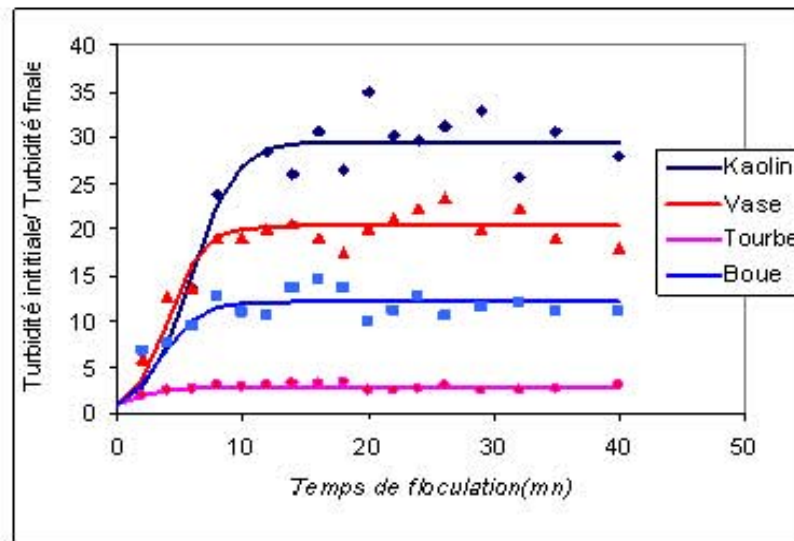


Figure 3.28 : Cinétique de floculation des suspensions des particules solides (kaolin, vase, boue et tourbe)

La cinétique de floculation des différentes suspensions aux conditions optimales de pH, du dosage en coagulant, de la puissance d'agitation lente et pour des concentrations de particules primaire de 50 mg/l, s'écrivent alors comme suit (tableau 3.1) :

Kaolin	$\frac{n_1}{n_0} = 1,7010 \cdot 10^{-3} G + (1 - 1,7010 \cdot 10^{-3} G) e^{-1,70 \cdot 10^{-4} GT}$
Vase	$\frac{n_1}{n_0} = 2,4610 \cdot 10^{-3} G + (1 - 2,4610 \cdot 10^{-3} G) e^{-4,0410 \cdot 10^{-4} GT}$
Boue	$\frac{n_1}{n_0} = 4,16 \cdot 10^{-3} G (1 - 4,16 \cdot 10^{-3}) e^{-5,45 \cdot 10^{-4} GT}$
Tourbe	$\frac{n_1}{n_0} = 1,7910 \cdot 10^{-3} G + (1 - 1,7910 \cdot 10^{-3} G) e^{-3,81 \cdot 10^{-4} GT}$

Tableau 3.1 : Modèles cinétiques de floculation du kaolin, de la vase, de la boue et de la tourbe.

La figure 3.29 ci-après montrent l'évolution de la constante de cassage K_B , de la constante de formation K_A et du rapport K_B/K_A , pour les suspensions du kaolin, de la vase, de la boue et de la tourbe. La connaissance de ces valeurs sont utiles dans la mesure où, elles nous renseignent sur la structure du floc ; plus un floc est résistant plus K_B est faible et inversement, plus K_A est élevé plus le floc se forme rapidement.

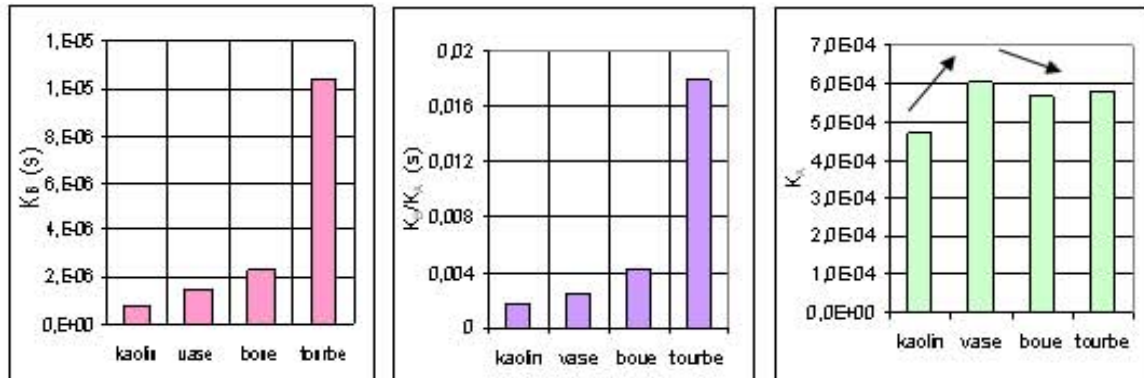


Figure 3.29 : Evolution des constantes de cinétiques K_A , K_B et du rapport K_B/K_A en fonction de la nature du solide dispersé

Nous mentionnons des valeurs croissantes monotones de la constante de cassage K_B , et du rapport K_B/K_A en fonction du type de solide dispersé, En effet on note la valeur élevée de la constante de cassage des particules organiques de la tourbe suivi par celle de la boue, puis de la vase et en dernier lieu la constante de cassage du kaolin. Cela indique que les floccs de particules organiques de la tourbe sont fragiles et de faible résistance alors que les floccs de particules minérales du kaolin et de la vase sont résistants.

La constante de formation K_A apparaît aussi affectée par la nature du solide dispersé. Les floccs de particules de vase se forment rapidement par rapport aux floccs des autres particules solides dispersées ; sur la figure 3.29 on constate que pour les particules dont le taux de matière organique est faible ($MO < 6\%$) la formation de flocc est en accroissement (K_A augmente) et pour des valeurs élevées du taux de matière organique ($MO > 6\%$) la formation de flocc régresse.

Nous mentionnons donc que la présence de la matière organique à de faibles concentrations améliore la formation du flocc, à priori, par complexation et pontage alors que cette formation est inhibée pour des taux élevés de matière organique.

Cela peut confirmer les résultats des travaux de MIETTA (2009a) qui indique que le taux de matière organique influe sur l'efficacité de collision et la résistance de flocc, en fait MIETTA (2009a) indique que pour un pH de 8, les particules de vase traitée (0% MO) et de la kaolinite (0% MO) mises en suspension dans l'eau de robinet sont stables et présentent une mauvaise aptitude à la floculation par rapport aux particules de vase naturelle (6% de MO).

Nous procéderons, dans la section modélisation de la constante de formation K_A du chapitre 05, à la détermination de la valeur seuil du taux de MO pour le quel la formation de flocc se dégrade.

La présentation de l'évolution des constantes cinétiques en fonction des caractéristiques du solide dispersé permet une meilleure visualisation de l'effet des caractéristiques du solide dispersé sur la cinétique de la floculation. La figure 3.30 explicite clairement que la fragilité du floc est tributaire aux diamètres des particules primaires ; le rapport K_B/K_A illustrant le degré de cassage par rapport à la formation, est faible pour les particules de faible diamètre moyen, et quand ce dernier est grand le cassage est prépondérant (le floc est fragile).

Sur la figure 3.31, on note que tant que le taux de matière organique est inférieur à 25 %, (particules du kaolin, de la vase et de la boue) la structure du floc n'est pas trop affecter (K_B/K_A est faible), en revanche, des que le taux de matière organique dépasse cette valeur, la résistance du floc décroît fortement. (K_B/K_A est élevé).

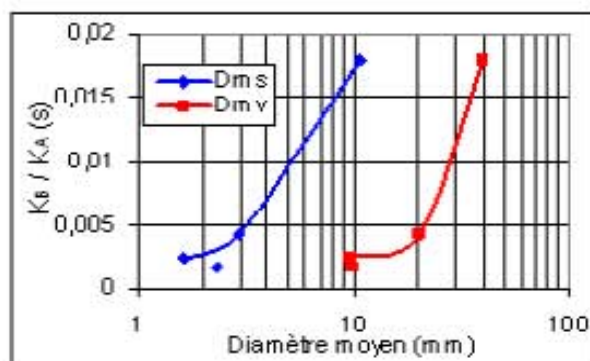


Figure 3.30 : Evolution du K_B/K_A en fonction du diamètre moyen des particules primaires.

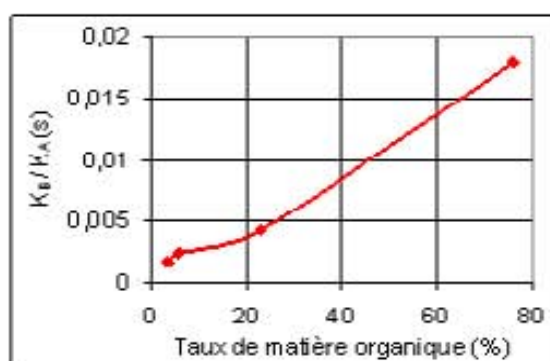


Figure 3.31 : Evolution du K_B/K_A en fonction du taux de matière organique

Il vient alors que la formation et le cassage de floc est influencée par la matière organique présente, celle-ci affecte la spéciation de l'aluminium et dicte le mécanisme de floculation comme l'indique les travaux suivants ;

Selon EXALL (2003), la matière organique affecte légèrement la spéciation de l'aluminium en cas de l'alun dilué, par contre elle réduit grandement la proportion des espèces tridécamer Al_{13} observées dans les solutions préhydrolysées. JARVIS (2005) note que les floccs formés par complexation par des espèces monomères sont fragiles par rapport au floc formés par neutralisation ou adsorption à partir des espèces dimères, trimères ou tridécamères (polymérisées et moyennement polymérisées).

De son côté YAN (2008) confirme que les espèces Ala (monomères), Alb (moyennement polymérisées) et Alc (polymérisées) suivent respectivement les mécanismes de complexation, neutralisation, et adsorption à l'élimination de

la matière organique naturelle, et que l'espèce monomère Ala, montre une grande aptitude à la complexation avec la matière organique, les floccs formés par complexation décantent difficilement.

Par ailleurs WANG (2009b) mentionne que les floccs d'acide humique formés des espèces monomères et dimères sont larges par rapport aux floccs formés à partir de $Al_3O_4(OH)_4^{7+}$ et qui sont petits. A son tour MIETTA (2009a) montre, à la base de son étude de floculation des particules de kaolinite, de la vase naturelle et de la vase traitée par le peroxyde d'hydrogène, que si les conditions d'agitation, de pH et de concentration en sels sont constantes, le diamètre moyen du flocc augmente suite à l'augmentation du taux de matière organique dans le solide dispersé.

Suite à ces résultats nous concluons que les floccs organiques sont formés essentiellement par complexation des espèces monomères, présentent de grands diamètres et montrent une grande difficulté à se décantent. Ce qui confirme les résultats de FRANÇOIS (1987a) qui montre que la force du flocc d'hydroxyde métallique dépend de la dimension et de la structure du flocc et les notations de Duan (2003) ; le cassage des floccs est étroitement lié à la taille de flocc.

En ce qui concerne l'évolution de rapport K_B/K_A en fonction de la surface spécifique, la figure 3.32 montre que pour les faibles surfaces spécifiques le flocc est fragile, cependant la résistance de flocc est meilleure pour les particules dont la surface spécifique est élevée, en fait pour les surfaces spécifiques élevés, la surface d'adsorption est grande permettant une meilleure déstabilisation, adsorption et neutralisation de charge. La figure 3.33 met en évidence que les floccs des particules solides primaires dont le pourcentage en "SiO₂ et Al₂O₃" est élevé sont des floccs résistants et au contraire, ils sont fragiles si ce taux est faible.

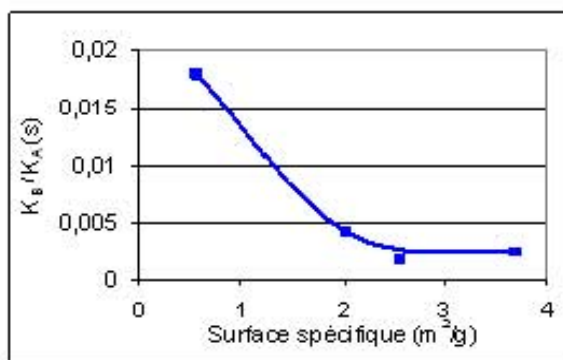


Figure 3.32 : Evolution du K_B/K_A en fonction de la surface spécifique des particules primaires.

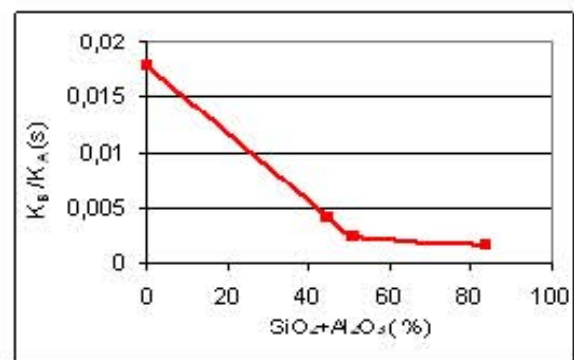


Figure 3.33 : Evolution du K_B/K_A en fonction du taux de "SiO₂ + Al₂O₃"

Quant à la figure 3.34 représentative de l'évolution du rapport K_B/K_A en fonction de la largeur de distribution de tailles de particules primaires "span", elle fait apparaître que la largeur de distribution de taille de particules primaires affecte la résistance et la formation du floc formé ;

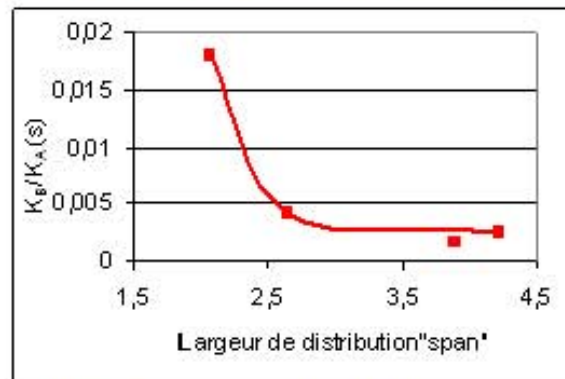


Figure 3.34 : Evolution du K_B/K_A en fonction en fonction de la largeur de distribution "span"

$$span = \frac{D(0,9) - D(0,1)}{D(0,5)}$$

D(0,9): la taille pour laquelle 90% de l'échantillon se trouve en dessous de cette dimension [µm], D(0,5): la taille pour laquelle 50% de l'échantillon se trouve en dessous de cette dimension [µm], D(0,1): la taille pour laquelle 10% de l'échantillon se trouve en dessous de cette dimension [µm]

Pour les distributions serrées (faibles valeurs de span), le floc est fragile, et pour des distributions larges, le floc est résistant. Ce résultat peut être lié à l'état stationnaire de la floculation comme c'est expliqué ci dessous;

La cinétique de floculation est définie par un état stationnaire caractérisé par un équilibre entre deux phénomènes; l'agrégation des "focullis" pour former de gros floc et le cassage et la fragmentation des floccs formés, cette position stationnaire est aboutie en temps limite appelé temps d'équilibre.

Les particules primaires de distributions de tailles de particules primaires serrées mènent à priori à de distributions de tailles de floccs aussi serrées, conduisant en un temps très réduit, la cinétique de floculation à l'état stationnaire, où la performance de séparation ne s'améliore plus. Selon COUFORT (2008), l'état stationnaire est caractérisé par un nombre constant de floccs et un diamètre moyen aussi constant, la distribution de tailles des floccs dépend de l'hydrodynamique et de la répartition initiale de la taille des particules primaires.

MIETTA (2009a) montre que le temps d'équilibre dépend du gradient de vitesse ; en effet il diminue suite à l'augmentation de gradient de vitesse. Les figures ci-après 3.35 et 3.36 montrent les évolutions du rendement d'équilibre et du temps d'équilibre en fonction de la nature du solide dispersé.

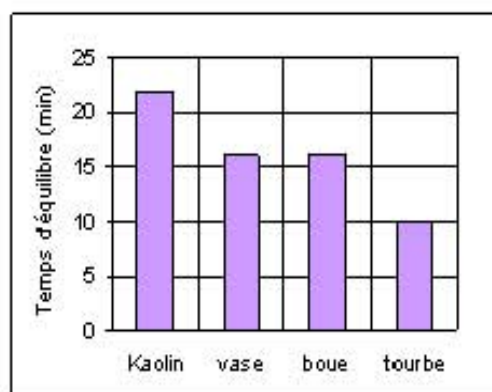


Figure 3.35 : Evolution du temps en fonction de la nature du solide

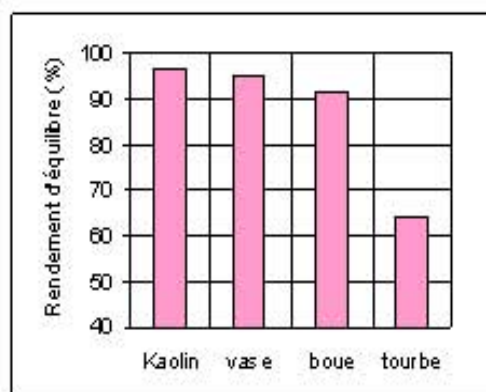


Figure 3.36 : Evolution du rendement d'équilibre en fonction de la nature du solide

Pour les différentes suspensions étudiées, le temps d'équilibre est inférieure à 30 minutes, ce qui est en accord avec les résultats de COUFORT (2007) qui montre que dans le cas de floculation des suspensions de bentonite par le sulfate d'aluminium, et quelle que soit la forme de l'agitateur, l'état stationnaire caractérisé par la balance entre les phénomènes d'agglomération et de rupture est rapidement atteint à moins de 15 à 30 minutes.

Les résultats, rapportés sur la figure 3.35, montrent que pour les particules organiques de tourbe le temps d'équilibre est faible. On constate aussi que le temps d'équilibre de la vase est plus faible que celui du kaolin, ce qui conforme les résultats de MIETTA (2009 a et b) qui utilise des suspensions de vase et de kaolinite de caractéristiques proches à celles de la vase et du kaolin que nous avons utilisé (voir tableau 3.2 et 3.3).

	Al ₂ O ₃	SiO ₂	K ₂ O	Fe ₂ O ₃ %	CaO %	MO	D(0,06) µm	D(0,5) µm	D(1) µm
Kaolinite MIETTA (2009 b)	34,7	48,49	3,32	<1	< 0,3	0	2	8	60
Kaolin (utilisé)	38,83	44,94	2,88	1,65	0	3,5	0,65	5,97	66

Tableau 3.2 : Caractéristiques du kaolin utilisé et de la kaolinite utilisée par MIETTA (2009 b)

	Illite	Montmorillonite	Kaolinite	M.O.%	D(0,02) µm	D(0,93) µm
Vase (MIETTA 2009 a)	53 %	21 %	26 %	6,6	2 mm	6,3
Vase (utilisée)	15	70	10	6%	0,23	29

Tableau 3.3 : Caractéristiques de la vase utilisée et de celle utilisée par MIETTA (2009 a)

Au cours de la floculation, les floccs organiques caractérisés, par de grands diamètres et une grande aptitude au cassage, atteignent en un temps très court (le temps d'équilibre) le volume critique V_c , et se cassent instantanément par la suite, c'est l'état stationnaire. Conformément aux résultats de COUFORT (2008) qui mentionne que lorsque le volume du flocc est inférieur au volume critique " V_c ", il n'y a pas de cassage intense, toutefois lorsque ce volume dépasse le volume critique, le flocc se fragmente instantanément. COUFORT (2008) note aussi que le nombre de floccs par unité de volume de dimension supérieure au volume critique qui se brisent est strictement égale au nombre de floccs de taille V formés par l'agrégation de floccs de tailles inférieures.

Par ailleurs, le temps d'équilibre peut être relié à la densité du solide dispersé, pour les suspensions minérales dont la densité est élevée par rapport à celle de milieu, un pourcentage de particules décantent au cours de la floculation (pour de faible gradient d'agitation), la probabilité de contact diminue alors, et l'état stationnaire prend plus de temps à s'accomplir, au moment où les particules organiques de densité proche de celle du milieu restent en suspension en phase de floculation, la probabilité de contact est élevée et l'état stationnaire est rapidement atteint.

Tout ceci rejoint les résultats obtenus par MIETTA (2009 b) qui montrent l'existence d'un gradient de vitesse limite G^* , tel que si le gradient d'agitation G est inférieure à G^* , les suspensions de particules de densité élevée par rapport à celle du milieu, n'atteignent pas l'équilibre (le temps de séjours de particules dans le jar test est limité par la décantation), et lorsque G est supérieur à G^* l'équilibre est atteint. Toutefois, si on utilise des particules de densité égale à celle de milieu, l'équilibre est atteint pour n'importe quel G (pour des temps de floculation non limités).

Les faibles rendements d'équilibre des particules organiques par rapport aux particules minérales présentés dans la figure 3.36, est le résultat de l'interaction de l'ensemble des caractéristiques des particules primaires, des mécanismes de floculation, de la formation et du cassage du flocc, de la distribution de taille des floccs, des dosages de coagulant injecté et des conditions d'agitation appliquées et aussi cela revient au taux de déshydratation des floccs formés signalé par WU (2002) qui montre que le surdosage en boue dans les mélanges boue/kaolin influe sur la déshydratation des floccs et affecte la floculation du mélange.

VI-3.4 Interaction entre la matière minérale et la matière organique

Pour compléter cette étude, il nous semble opportun de nous situer dans des conditions intermédiaires et de dévoiler les interactions entre la matière minérale et la matière organique et leurs effets sur la performance de floculation, sur la relation entre la puissance d'agitation et la dose de coagulant et sur la cinétique de la floculation,

Afin d'illustrer cet aspect, nous présentons, ci après, les résultats d'étude de la performance et la cinétique de la floculation des suspensions des deux mélanges I et II.

VI-3.4.1. Effet de l'interaction matière minérale/matière organique sur la performance de floculation.

Au vu des résultats de l'évolution des efficacités d'élimination de la turbidité déterminées expérimentalement et théoriquement des deux mélanges I et mélange II de la figure 3.37, Il apparait clairement que pour les suspensions de mélange I (25% MO), le rendement atteint est de 92% dans le meilleur des cas; il est intermédiaire entre ceux des suspensions de la vase et de la tourbe. Quant aux suspensions de mélange II, leur floculation est marquée par un faible rendement d'élimination de turbidité (87%) mais, supérieur à celui des suspensions de tourbe (65%). les suspensions minérales sont caractérisées par de petits diamètres, des surfaces spécifiques élevées, de densité élevée, alors que les suspensions organiques montrent de grands diamètres de particules primaires et effet d'agglutination élevé.

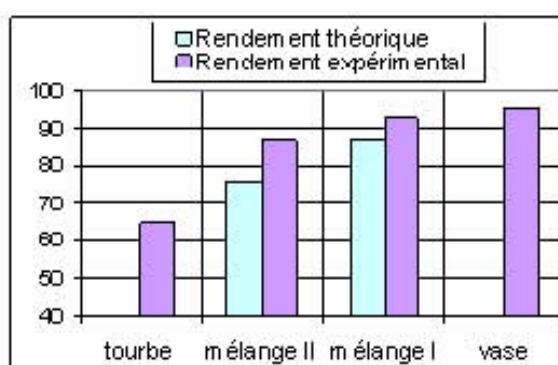


Figure 3.37 : Rendements expérimentaux et théoriques de floculation des suspensions de mélanges I et II

La figure 3.37 montre que les rendements expérimentaux sont nettement supérieures aux rendements théoriques (calculés en considérant que chaque composant de mélange flocule séparément).

Il semble clair que des interactions se produisent davantage entre les particules minérales et les particules organiques au cours de la floculation, ce qui rejoint les résultats de YATES (2008) qui montre que les interactions entre les hétéro-particules sont plus marquées que celles entre les particules de même taille.

Ce résultat remarquable peut être expliqué par le fait que, la floculation ne se fait pas pour chaque composé de mélange à part et qu'il y'aurait des interactions de plus entre les particules du mélange ce qui stimulerait les mécanismes de complexation de déstabilisation et d'agglomération.

Cette amélioration en performance pourrait être reliée au rôle de la matière minérale qui stimule et accélère la floculation conformément aux notations de MAZET (1990) qui indique, une meilleure élimination de la matière organique en présence de particules argileuses de kaolinite.

VI-3.4.2. Effet de l'interaction matière minérale/matière organique sur la relation entre la puissance d'agitation et la dose de coagulant

Pour étayer encore l'interaction entre la matière minérale et la matière organique et son effet sur la puissance optimale d'agitation et la dose en coagulant, nous avons étudié la floculation des suspensions des deux mélanges I et II. Les résultats d'études sont illustrés dans le graphique 3.38 suivant :

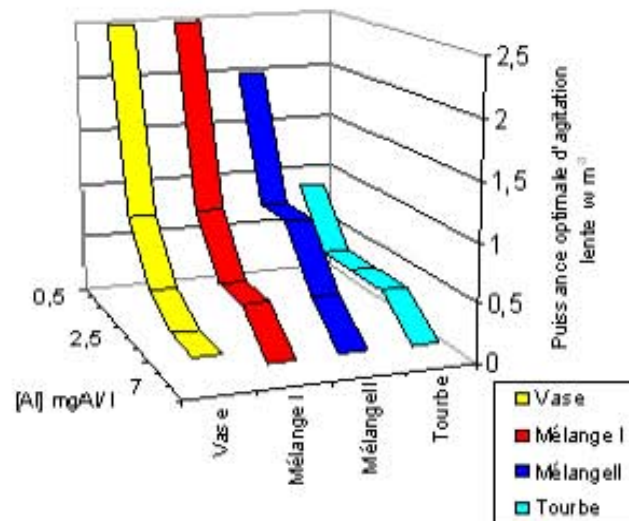


Figure 3.38 : Evolution de la puissance optimale d'agitation lente pour les suspensions de la vase, des deux mélanges I et II et de la tourbe en fonction de la dose du coagulant injecté.

Il apparaît que l'ajout de proportions de matière minérale (la vase) aux particules de la tourbe modifie les valeurs des optimums de puissance d'agitation. Pour les suspensions de tourbe et à faible dosage en coagulant, la puissance optimale d'agitation est faible ($< 1 w/m^3$) cela signifie que le floc

est fragile et sa structure ne permet pas de grandes agitations, l'ajout des suspensions de la vase en proportion de 37 % (mélange II) permet de renforcer la structure de floc et d'élever sa résistance au cassage, permettant ainsi d'augmenter le seuil d'agitation optimale à $2w/m^3$. Alors que pour le mélange I où la vase présente un taux de 73%, la puissance optimale d'agitation s'élève à $2,5 w/m^3$, Il en résulte alors que l'ajout de la matière minérale (la vase) renforce et solidifie le floc et permet d'élever le seuil d'agitation optimale.

VI-4.3. Effet de l'interaction matière minérale/matière organique sur la cinétique de floculation

Les résultats de la cinétique de floculation des suspensions de mélange I et II sont présentés à la figure 3.39, la cinétique de floculation s'écrit sous la forme suivante (Tableau 3.4) :

Mélange I	$\frac{n_1}{n_0} = 3,3310^{-3} G + (1 - 3,3310^{-3} G)e^{-5,3810^{-4} GT}$
Mélange II	$\frac{n_1}{n_0} = 5,62 \cdot 10^{-3} G + (1 - 5,62 \cdot 10^{-3} G)e^{-5,9310^{-4} GT}$

Tableau 3.4 : Modèles cinétiques de floculation des mélanges I et II.

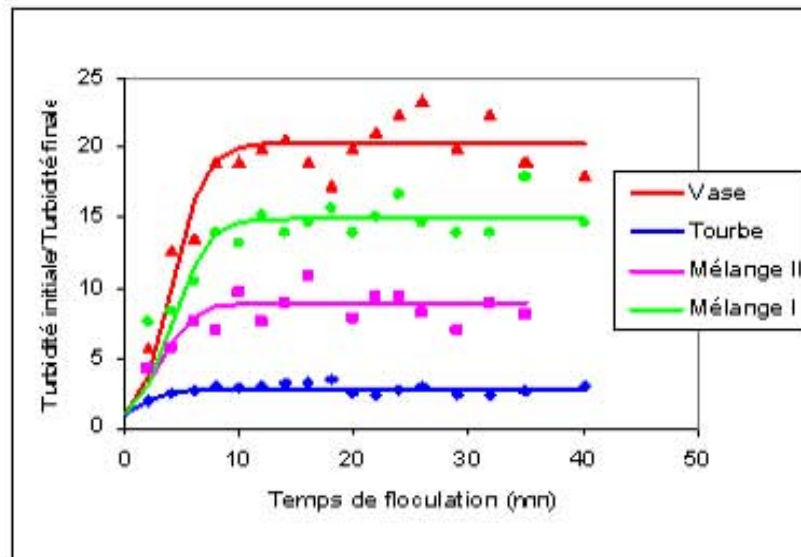


Figure 3.39 : Cinétique de floculation des suspensions de la vase, des deux mélanges I et II et de la tourbe.

Les résultats mis en évidence que les floccs des suspensions de mélange II se forment plus rapidement que les floccs des suspensions de mélange I ($K_A \text{ mélange II} = 5,93 \cdot 10^{-4} > K_A \text{ mélange I} = 5,38 \cdot 10^{-4}$), et au contraire ils sont moins résistants à l'érosion et à la fragmentation ($K_B \text{ mélange II} = 3,33 \cdot 10^{-6} > K_B \text{ mélange I} = 1,79 \cdot 10^{-6}$).

Les figures 3.40 et 3.41 montrent les valeurs de la constante cinétique de cassage K_B et du rapport K_B/K_A pour les deux mélanges I et II comparativement aux valeurs relatives aux suspensions de la tourbe et de la vase.

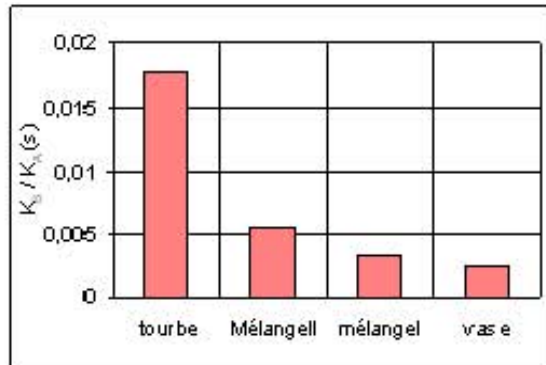


Figure 3.40 : Evolution de K_B/K_A en fonction de la nature du solide (tourbe, mélange I, mélange II et vase)

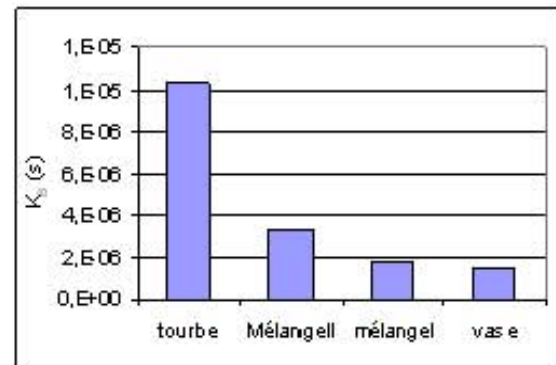


Figure 3.41 : Evolution de K_B en fonction du nature du solide (tourbe, mélange I, mélange II et vase)

Ces deux graphiques montrent que l'ajout de la matière minérale (particules de vase) aux particules de la tourbe diminue la constante de cassage K_B et diminue aussi le rapport entre la constante de cassage et celle de formation (K_B/K_A).

Ce résultat revient principalement aux interactions qui se produisent entre les suspensions minérales et les suspensions organiques et en particulier aux diamètres des floes formés ; les diamètres des floes du mélange II "tourbe/vase" (avec 37,2 % de vase) sont plus grands que les diamètres des floes de mélange I "tourbe/vase" (avec 72,8 % vase), conformément aux résultats des travaux de WU (2002) et MIETTA (2010).

WU (2002) mentionne que les floes de mélange "boue/kaolin" (8% de kaolin) sont plus grands que ceux de mélange "boue/kaolin" (17 % de kaolin), WU (2002) précise encore que le surdosage en boue dans les mélanges boue/kaolin influe sur la déshydratation des floes et affecte la floculation.

Alors que MIETTA (2010) conclut que la taille moyenne d'équilibre du floe varie linéairement avec le rapport entre les paramètres d'agrégation et de cassage, ce ratio peut être alors estimé analytiquement par l'observation de la taille d'équilibre de floe.

Compte tenu des résultats relatifs à la floculation des mélanges I et II, nous confirmons à la figure 3.42 que le temps d'équilibre entre le cassage et la formation n'est pas constant, l'ajout de la matière minérale augmente ce temps d'équilibre permettant ainsi, de temps en plus à la formation et la croissance des floes.

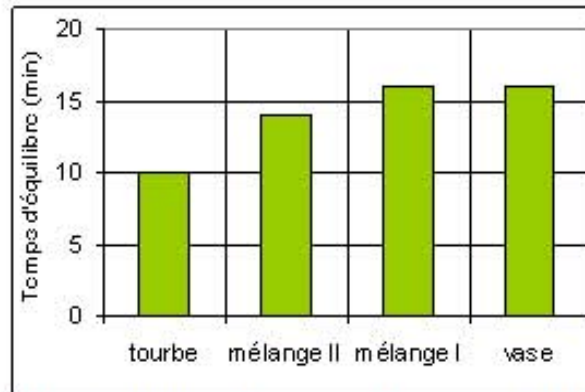


Figure 3.42 : Evolution du temps d'équilibre pour les suspensions de la tourbe, du mélange I, du mélange II et de la vase)

En effet à 0% d'ajout de particules de la vase, le temps était de 10 min, à 37% d'ajout, le temps s'améliore, il saute à 14min, et pour 73 % d'ajout de la vase, le temps s'élève à 16 min.

Nous concluons alors que l'ajout de la matière minérale argileuse retarde l'état stationnaire et permet de promouvoir davantage la floculation.

IV. CONCLUSION DU CHAPITRE

L'objectif du travail expérimental exposé au cours de ce chapitre était d'étudier la floculation de différentes suspensions synthétiques de particules solides du kaolin, de la vase, de la boue, de la tourbe et de deux mélanges vase/tourbe, et principalement la mise en évidence des interactions entre les paramètres opératoires et les caractéristiques du solide dispersé et leurs influences sur le processus de coagulation et de floculation.

Dans la première section de ce chapitre une présentation et une caractérisation assez complète des particules solides ont été exposées et qui ont permis d'affirmer les caractéristiques différentes des solides dispersés ; le kaolin et la vase ont une composition essentiellement minérale, le premier est composé, en majorité, de la kaolinite alors que le major constituant de la vase est la montmorillonite, leurs pourcentages en matière organique est respectivement 3,5 et 6 %, leurs distributions granulométriques sont caractérisées par de faibles diamètres et par des valeurs élevées de surfaces spécifiques.

Cependant la tourbe se caractérise par une teneur élevée en matière organique (soit 76 %), caractérisée par la présence des substances humiques. La tourbe se distingue aussi par la distribution serrée de tailles de ses particules solides, les diamètres moyens en volume et en surface sont élevés alors que la surface spécifique est faible.

Quant à la boue, elle est minéralisée à 77 %, l'illite présente la fraction majoritaire minérale de la boue. La matière organique de la boue est de nature différente de celle de la tourbe. La surface spécifique et le caractère argileux (% SiO_2 et Al_2O_3 élevé) de la boue sont proches de ceux du kaolin et de la vase.

Les résultats d'étude de la floculation montrent que pour les suspensions minérales (kaolin et vase), l'efficacité d'élimination de la turbidité atteint 97%, alors qu'elle ne dépasse guère 65% dans le cas des suspensions organiques de la tourbe. Il en découle que la présence de la matière organique dans le solide dispersé à faible concentration est d'une influence minime, alors que l'effet négatif inhibiteur est marqué pour des taux de matière organique qui dépassent 25 %. Le faible rendement d'élimination de la turbidité des suspensions organiques est tributaire de :

- L'hydrophobicité, la taille et la densité de charge et la quantité de colloïdes organiques
- Des caractéristiques de la matière organique, (la masse molaire, le nombre, la nature et le degré d'ionisation des groupements fonctionnels et le nombre d'atome de carbone et l'acidité carbonyle)
- A la structure minéralogique, les propriétés surfaciques et les surfaces de complexation.

La consommation du coagulant est faible pour les suspensions dont le taux de matière organique est inférieur à 25% alors qu'elle est excessive pour les suspensions dont le taux de matière organique dépassant 50%. Cette consommation élevée en cas des particules organiques revient, à la double consommation de l'aluminium ; à la formation du $Al(OH)_3$, et à la complexation du cation Al^{3+} sur la matière organique ; Il se forme ainsi de l'humates d'aluminium insoluble grâce à l'interaction chimique entre les espèces hydrolysées cationiques de l'aluminium et les groupements fonctionnels organiques.

Le mécanisme de floculation des suspensions minérales est "l'adsorption-neutralisation de charges", alors que pour les suspensions organiques plusieurs mécanismes compétitifs entre les espèces hydrolysées de l'aluminium et de la matière organique se passent ; le mécanisme de floculation, est alors "complexation-adsorption-neutralisation de charges" ; à faible dosage en coagulant c'est la complexation qui est majoritaire, cependant pour des dosages élevés en coagulant c'est l'adsorption et la neutralisation de charges qui dominent. Les floccs formés par complexation sont fragiles par rapport au flocc formés par neutralisation ou adsorption.

Les résultats montrent également que le rendement d'élimination de la turbidité s'élève suite à l'accroissement de la surface spécifique, alors que l'augmentation des diamètres des particules primaires diminue l'efficacité de floculation. Cette diminution est due à la fragilité des floccs ; les floccs organiques ont des diamètres supérieures aux diamètres des floccs minéraux donc sensibles aux fortes agitations et leurs résistances au cassage est faible.

Les résultats montrent aussi que la performance de la floculation des particules de distribution de tailles serrée (faible valeurs de "span") est

minime alors qu'elle est meilleure pour les particules de distribution large. La variation de la performance de floculation en fonction du taux de "SiO₂ et Al₂O₃", indique que l'accroissement de ce taux dans le solide dispersé améliore l'efficacité d'élimination de la turbidité, en effet la matière minérale argileuse durci et renforce la structure du floc.

L'étude des interactions entre la puissance d'agitation lente, la dose du coagulant et la nature du solide dispersé montre que la puissance optimale d'agitation lente est à la fois reliée à la dose du coagulant injecté et à la nature du solide dispersé ; Il apparaît qu'aux faibles dosages en coagulant, les floes organiques sont sensibles aux grandes agitations. Cette sensibilité revient principalement à la présence de la matière organique et des espèces d'aluminium, qui dictent le mécanisme de floculation, cela revient aussi aux diamètres et distributions des tailles des particules primaires et floes, à la composition minérale et organique du solide et aux propriétés surfaciques des particules.

L'étude de la cinétique de la floculation des suspensions synthétiques montre la valeur élevée de la constante de cassage K_B des particules organiques de la tourbe par rapport à celles des autres composés, révélatrice de floc fragile et de faible résistance, suivi par celle de la boue, de la vase et en dernier lieu, la constante de cassage du kaolin.

La constante de formation K_A apparaît aussi affectée par la nature du solide dispersé, les floes des particules de vase se forment plus rapidement par rapport aux floes des autres particules solides. Ceci revient au rôle de la matière organique ; pour des faibles taux de matière organique, cette dernière améliore la formation et la croissance du floc par complexation et par pontage, alors que cette formation est inhibée pour des taux élevés.

Les floes formés par complexation, présentent de grands diamètres et montre une grande difficulté à se décanter interprétée par de faible rendement d'élimination de la turbidité.

L'évolution du rapport K_B/K_A montre que, principalement, pour des taux de matière organique inférieurs à 25 %, la structure du floc n'est pas trop affecter (K_B/K_A est faible) en revanche, des que ce taux dépasse 25 %, la résistance du floc décroît fortement (K_B/K_A est élevé).

Il en résulte aussi de cette étude que la résistance de floc est meilleure tant que la surface spécifique est élevée et le pourcentage en "SiO₂ et Al₂O₃" est aussi élevé et quand la distribution de tailles de particules primaires est large.

Il en découle alors que la cinétique de formation et de cassage de floc est liée aux paramètres suivants :

- Caractéristiques des particules solides primaires à savoir :
 - Le taux de matière organique
 - La densité et la surface spécifique
 - Le diamètre et la largeur de distribution granulométrique
- Caractéristiques des particules de floes à savoir :
 - La nature, la taille et la largeur de distribution granulométrique
- Le dosage en coagulant
- Les espèces d'aluminium présentes.
- Le mécanisme de floculation.
- Les conditions d'agitation.

La recherche des effets de l'interaction entre la matière minérale et la matière organique sur :

- La performance de floculation,
- La relation entre la puissance d'agitation et la dose du coagulant
- La cinétique de floculation

a été traitée dans la dernière section de ce chapitre, il en résulte que l'ajout de la matière minérale argileuse permet :

- L'augmentation des sites d'adsorption où les particules minérales jouent le rôle de noyau, et stimulent les mécanismes de complexation, de déstabilisation et d'agglomération.
- La formation de petits floes, le renforcement de la structure et augmentation de la résistance de floc au cassage ce qui permet d'élever le seuil optimal d'agitation.
- La progression du temps d'équilibre de l'état stationnaire, ce qui permet de stimuler davantage la formation et la croissance des floes et améliore le rendement de floculation.

Chapitre 4

ANALYSE EXPLORATOIRE DES DONNEES EXPERIMENTALES.

V. INTRODUCTION

Les essais expérimentaux d'étude de la performance et de la cinétique de floculation des différents types de solides dispersés ainsi que l'étude des interactions entre les paramètres de floculation présentés dans le chapitre précédent, nous ont permis de construire un tableau de données expérimentales.

L'objectif de ce chapitre est l'analyse de ces données expérimentales en vue de déterminer les variables descriptives qui peuvent être intégrées dans les modèles de prédiction de la dose optimale de floculation, des paramètres de finalité de traitement "l'aluminium résiduel et la turbidité finale" et des constantes cinétiques de formation et de cassage de floc.

Pour ce faire, une analyse exploratoire globale des individus et des variables de ce tableau est ici développée ; elle consiste à l'analyse des variables une par une (analyse univariée), puis l'analyse des corrélations et liaisons entre variables prises deux à deux (analyse bivariée) et ensuite le traitement multivarié qui consiste en la présentation des proximités entre individus ou/et variables sur un graphique (analyse en composantes principales) ou le regroupement de ces mêmes données en catégories homogènes (analyse de classification).

Ces différentes analyses et représentations graphiques suggestives permettent de mettre en évidence des ressemblances ou des différences entre individus ou variables, et permettent aussi d'identifier les problèmes liés aux distributions des données et les éventuelles liaisons entre les variables destinées à être intégrées dans les modèles de régression.

Ce chapitre présente alors l'analyse exploratoire des résultats expérimentaux. Dans un premier temps, après avoir rappelé la provenance des données expérimentales et leur signification, nous procédons à l'analyse exploratoire des données statistiques, qui s'articule en quatre points : la description des données, l'analyse univariée, l'analyse bivariée puis l'analyse multivariée. La deuxième partie expose le choix des variables descriptives des modèles de régression.

VI. DESCRIPTION DES DONNEES

Les données à traiter sont les résultats de l'étude expérimentale de la floculation de six types de suspensions préparées à partir des particules solide du kaolin, de la vase, de la boue, de la tourbe, du mélange I et du mélange II. Ces données répertoriées dans un tableau de données à n individus (lignes) et p variables (colonne) forment une matrice de données de : 40×470 . A chaque ligne de ce tableau est affectée un point expérience, les colonnes, qui représentent les paramètres retenus lors du protocole expérimental, sont scindées en trois groupes de variables : paramètres

opérationnels en nombre de quatre (04), paramètres indiquant le degré de traitement en nombre de huit (08) et les caractéristiques des particules solides mises en suspension en nombre de vingt-huit (28), ces dernières sont le résultat d'analyse de caractérisation des solides dispersés (voir chapitre 3), les variables, selon le modèle proposé, seront divisés en variables d'entrée et variables de sortie. La description de ces variables est listée comme suit :

1. Paramètres opérationnels en nombre de quatre : (4)

Dose	: Dose du coagulant injecté [mgAl/l]
T ₀	: Turbidité initiale [NTU]
P _{ui}	: Puissance d'agitation lente [w/m ³]
C _p	: Concentration en particules solide en [mg/l]

2. Paramètres de performance de floculation (8)

D _c	: Dose consommée [mg/l]
T _{eli}	: Turbidité éliminée [NTU]
D _c /T _{eli}	: Dose consommée/Turbidité éliminée [mg/l.NTU]
T _f	: Turbidité finale [NTU]
Al _r	: Aluminium résiduel [mg Al/l]
R _T	: Rendement d'élimination de la turbidité [%]
R _{Al}	: Taux de consommation de coagulant [%]
R _{Al} /R _T	: Rendement d'élimination de la turbidité/Taux de consommation de coagulant [-]

□

3. caractéristiques des particules solides mises en suspension (28)

K _A	: Constante cinétique de formation [-]
K _E	: Constante cinétique cassage [s]
K _E /K _A	: Constante de cassage /constante de formation[s]
CM	: Carbone minérale [%]
COT	: Carbone organique total [%]
MO	: Teneur en matière organique [%]
Arg	: Pourcentage d'argile dans le solide dispersé [%]
D _{ms}	: Diamètre moyen en surface [µm]
D _{mv}	: Diamètre moyen en volume [µm]
D(0,5):	: La taille pour laquelle 50% de l'échantillon se trouve en dessous de cette dimension [µm]
S _{sp}	: Surface spécifique [m ² /g]
Unif	: Uniformité : écart absolu à partir de la médiane D(0,5)
Si+Al	: Pourcentage de SiO ₂ +Al ₂ O ₃ dans le solide dispersé [%]
Kao	: Taux de kaolinite dans le solide [%]
ILL	: Taux d'Illite dans le solide [%]
Mte	: Taux de montmorillonite dans le solide [%]
Clote	: Taux de chlorite dans le solide [%]
Qtz	: Taux de quartz dans le solide [%]

Calte	: Taux de Calcite dans le solide [%]
Dolo	: Taux de dolomite dans le solide [%]
AL ₂ O ₃	: Taux de AL ₂ O ₃ [%]
SiO ₂	: Taux de SiO ₂ dans le solide [%]
Mg	: Taux Mg dans le solide [%]
CaO	: Taux de chaux dans le solide [%]
Fe ₂ O ₃	: Taux de Fe ₂ O ₃ dans le solide [%]
K ₂ O	: Taux de K ₂ O dans le solide [%]
Na ₂ O	: Taux de Na ₂ O dans le solide [%]
MnO	: Taux de MnO dans le solide [%]

L'imprégnation de tous ces différents paramètres a pour objectif de repérer des éventuelles relations entre ces différents paramètres ou groupes de paramètres.

VII. ANALYSE DES DONNEES

Avant toute tentative de modélisation complexe, il est impératif d'avoir conduit des analyses uni, bi et multidimensionnelles afin d'identifier :

- Les problèmes liés aux distributions des variables : dissymétrie, valeurs atypiques (outliers).
- Les liaisons entre les variables prises deux par deux : non-linéarité.
- Les violations d'hypothèses (homoscédasticité) ou des points influents
- les groupes de variables fortement corrélées entre elles, et éventuellement de détecter des caractères complexes sous-jacents à ces groupes.
- Examiner les distances entre les individus et les regroupant en un nombre restreint de classes.

III-1. ANALYSE UNIVARIEE

Une analyse descriptive univariée est effectuée pour chacune des variables étudiées, pour ce faire, On utilisera, en plus des statistiques descriptives, les boîtes de dispersion communément appelées boîtes à moustaches (box plot). Le tableau 4.1, résume les statistiques de dispersion des variables; les informations présentées ci-dessous concernent la tendance centrale (moyenne et médiane), l'amplitude (minimum et maximum), ainsi que la dispersion autour de la tendance centrale (écart-type).

Le tableau montre clairement que les paramètres caractéristiques suivantes présentent de grandes dispersions: "MO, Arg, Si+Al, Kao, ILL, Mte, COT, SiO₂, Calte, Al₂O₃, Dmv, D(0,5)", alors que pour les paramètres opérationnelles et de performance, les dispersions des valeurs de Pui, R_{AI} et T_{eli} sont élevées.

	N Actifs	Moyenne	Médiane	Minimum	Maximum	Ecart-type
Dose	61	3,01	2,50	0,50	7,00	2,04
Dc	61	2,94	2,45	0,43	6,92	2,02
Pui	61	18,43	9,00	0,00	67,27	23,58
T ₀	61	19,90	19,00	4,80	39,00	9,29
Teli	61	17,96	17,18	1,90	37,70	9,35
Tf	61	1,94	1,56	0,50	4,80	0,89
Alr	61	0,07	0,06	0,01	0,27	0,04
Dc/Teli	61	0,23	0,15	0,02	2,34	0,32
R _T	61	86,41	91,19	39,58	97,37	12,62
R _{Al}	61	96,33	97,72	84,90	99,10	3,44
R _{Al} /R _T	61	1,15	1,07	1,01	2,26	0,22
K _A	61	0,00	0,00	0,00	0,00	0,00
K _B	61	0,00	0,00	0,00	0,00	0,00
K _B /K _A	61	0,00	0,00	0,00	0,02	0,00
C _p	61	53,23	50,00	25,00	100,00	19,98
CM	61	2,49	2,48	0,20	4,80	1,85
COT	61	5,45	0,77	0,01	25,00	8,07
MO	61	19,71	6,00	3,50	76,00	23,40
Arg	61	62,95	65,00	0,00	93,00	29,51
Dms	61	3,64	2,34	1,62	10,66	2,83
Dmv	61	16,44	9,84	9,65	40,60	10,14
D(0,5)	61	12,28	5,97	5,69	35,40	9,74
Ssp	61	2,60	2,56	0,56	3,71	0,94
Unif	61	1,10	1,19	0,64	1,31	0,22
Si+Al	61	51,01	51,00	0,10	84,00	27,12
Kao	61	30,26	10,00	0,00	80,00	34,73
ILL	61	18,53	15,00	0,00	80,00	20,87
Mfe	61	27,77	0,00	0,00	70,00	31,38
Clote	61	3,44	1,85	0,00	15,00	4,38
Qtz	61	6,59	7,00	1,00	9,00	2,36
Calte	61	11,06	8,51	0,00	23,00	10,29
Dolo	61	1,37	1,00	0,00	3,00	1,31
MnO	61	0,04	0,03	0,01	0,07	0,03
Na ₂ O	61	0,39	0,38	0,23	0,60	0,13
K ₂ O	61	1,73	1,56	0,15	2,88	0,91
SiO ₂	61	32,03	36,06	0,11	44,94	14,04
Al ₂ O ₃	61	19,56	14,93	0,00	38,83	14,11
Fe ₂ O ₃	61	3,09	2,13	0,23	5,36	1,80
CaO	61	2,51	2,50	0,00	5,04	2,03
Mg	61	1,69	1,49	0,17	3,15	1,23

Tableau 4.1 : Statistique descriptive des variables.

Analyse de boîte de dispersion « boîte à moustache »

La boîte de dispersion est une représentation graphique, simple et synthétique de la dispersion des variables. Elle consiste à placer la valeur de certains quartiles (ici le premier et le troisième quartile), les percentiles (1% et 99%) sur une échelle et à visualiser instantanément certaines caractéristiques de tendance centrale, de dispersion et de forme de la variable considérée. Ainsi, sur chaque graphe nous pouvons observer la médiane, la distance inter-quartile et la distance inter-percentile. Ce diagramme est facile à interpréter, riche en informations. La figure 4.1 montre la dispersion des différents solides mis en suspension.

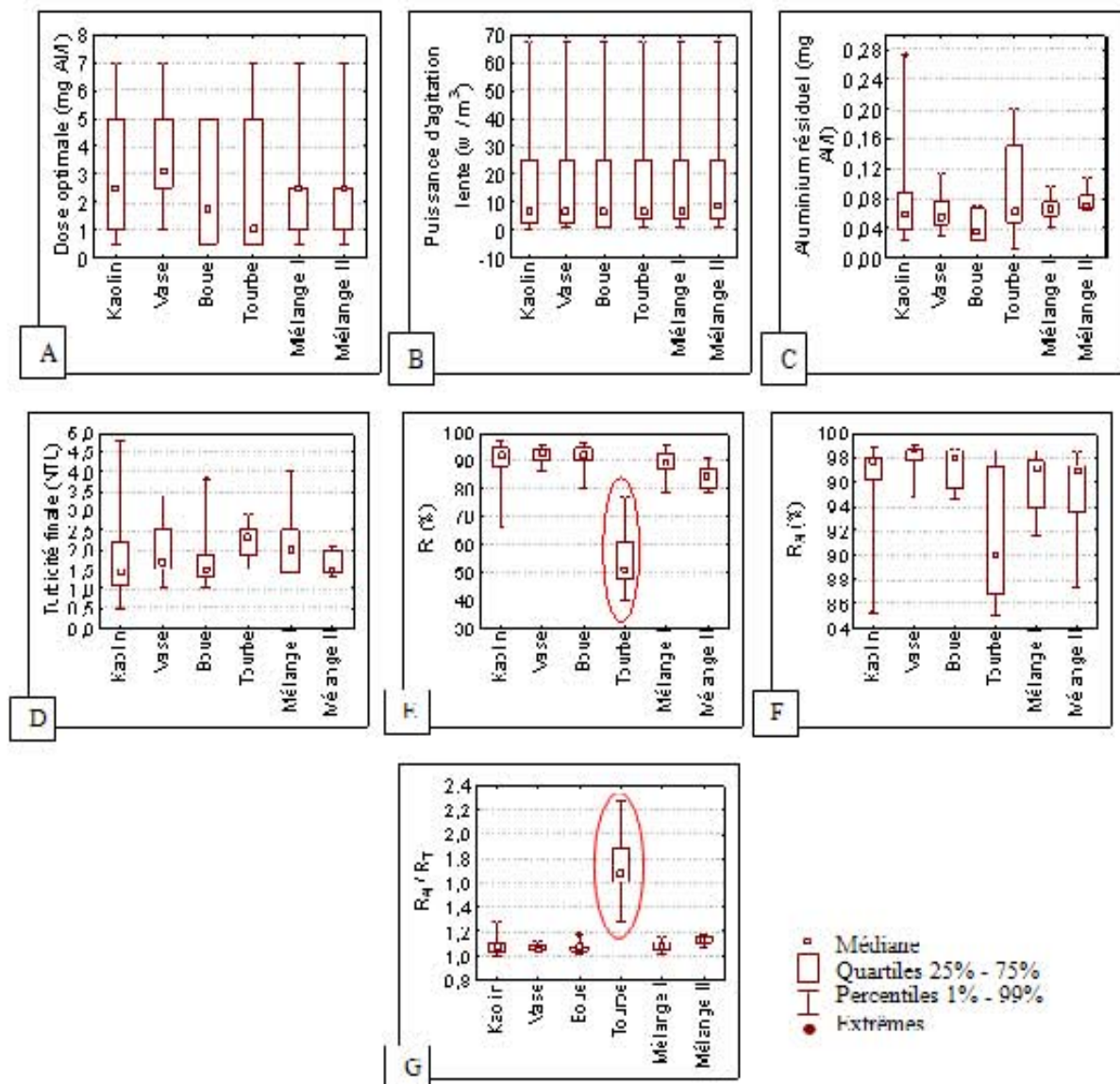


Figure 4.1 : Présentation box plot «boîtes à moustache» des dispersions inter-individuelle des différents types du solide dispersé [A] Dose optimale, [B] Puissance d'agitation lente, [C] Aluminium résiduel, [D] Turbidité finale, [E] R_t (Rendement d'élimination de la turbidité), [F] Taux de consommation du coagulant R_a , [G] R_a / R_t

A la base de l'analyse des boîtes de dispersion nous notons ce qui suit : La disparité de la dose, de la puissance et de la turbidité finale pour les différentes suspensions, montrée à la figure 4.1 [A], [B], [D], semble sans grande différence, cependant la disparité de la l'aluminium résiduel [C] et du taux de consommation du coagulant [F] pour les différentes suspensions montre que les suspensions de la tourbe présentent la plus grande étendue inter-quartiles.

La dispersion inter-suspension présentée également à la figure 4.1 [E] et [G] respectivement pour le rendement d'élimination de la turbidité R_T et le rapport R_{Al}/R_T montre que les suspensions de la tourbe se démarque nettement des autres suspensions ; les suspensions de la tourbe présentent des valeurs de R_T plus faibles que la moyenne, et des valeurs de R_{Al}/R_T plus élevées que la moyenne.

L'analyse de boîtes à moustaches révèle la présence de quelques points extrêmes, ces valeurs atypiques sont souvent dues à la variabilité du processus, et dans ce cas les éliminer de la population d'apprentissage appauvrirait le modèle; dans une autre mesure, les conserver amènerait à considérer une certaine marge d'erreur. Le seul examen des boîtes de dispersion ne nous suffit pas pour décider si une donnée doit être retirée ou non de l'étude, on procède alors pour palier à ce problème à l'analyse multidimensionnelle.

Cette analyse permettra d'observer de plus près les valeurs extrêmes et de déterminer lesquelles des variables sont responsables de l'éloignement anormal afin de pouvoir les corriger, les conserver ou les retirer de la population d'apprentissage LEFEBVRE (1988).

III-2. ANALYSE BIVARIEE

Les statistiques bivariées sont utilisées dans le but de détecter tout d'abord les incohérences entre les variables, mais aussi les liaisons entre elles, prises deux à deux. Les liaisons linéaires, si elles existent, sont obtenues par une étude numérique, le coefficient de corrélation linéaire de Dravais Pearson R , lequel donne une idée sur le réseau d'interrelations et de dépendances linéaires existant entre les différentes variables du tableau de données.

Le coefficient de Bravais Pearson est compris entre -1 et 1, il indique une présomption de liaison linéaire entre les deux séries d'autant qu'il sera proche de 1 en valeur absolue. Le coefficient de corrélation entre x_i et y_i est aussi le cosinus de l'angle θ entre ces deux vecteurs, l'interprétation d'un coefficient de corrélation comme un cosinus est une propriété importante.

Chapitre 4 : Analyse exploratoire des données expérimentales

Les coefficients de corrélation entre les différentes variables d'entrée et de sortie figurent dans le tableau 4.2, les corrélations sont calculées ici pour l'échantillon optimal à savoir toutes les expériences de floculation des six types de particules solides dispersées ayant aboutit aux minimums de turbidité finale et d'aluminium résiduel.

	Dose	Dc	Pui	T ₀	Teli	Tf	Alr	Dc/Teli	R _T	R _{Al}	R _{Al} /R _T	K _A	K _B	K _B /K _A
Dose	1	0,99	-0,64	0,21	0,25	-0,44	0,61	0,52	0,36	0,59	-0,26	0,11	-0,1	-0,1
Dc	0,99	1	-0,64	0,22	0,26	-0,44	0,6	0,51	0,36	0,59	-0,26	0,11	-0,1	-0,11
Pui	-0,64	-0,64	1	-0,05	-0,12	0,77	-0,34	-0,31	-0,35	-0,5	0,24	-0,01	0	0
T ₀	0,21	0,22	-0,05	1	0,99	-0,02	-0,26	-0,4	0,66	0,53	-0,57	0,08	-0,56	-0,57
Teli	0,25	0,26	-0,12	0,99	1	-0,11	-0,25	-0,38	0,7	0,58	-0,6	0,08	-0,57	-0,58
Tf	-0,44	-0,44	0,77	-0,02	-0,11	1	-0,07	-0,17	-0,47	-0,56	0,33	0,04	0,1	0,1
Alr	0,61	0,6	-0,34	-0,26	-0,25	-0,07	1	0,59	-0,13	0	0,1	-0,05	0,15	0,16
Dc/Teli	0,52	0,51	-0,31	-0,4	-0,38	-0,17	0,59	1	-0,31	0,09	0,33	0,11	0,53	0,54
R _T	0,36	0,36	-0,35	0,66	0,7	-0,47	-0,13	-0,31	1	0,81	-0,95	-0,11	-0,83	-0,83
R _{Al}	0,59	0,59	-0,5	0,53	0,58	-0,56	0	0,09	0,81	1	-0,65	0,03	-0,5	-0,5
R _{Al} /R _T	-0,26	-0,26	0,24	-0,57	-0,6	0,33	0,1	0,33	-0,95	-0,65	1	0,15	0,85	0,86
K _A	0,11	0,11	-0,01	0,08	0,08	0,04	-0,05	0,11	-0,11	0,03	0,15	1	0,35	0,32
K _B	-0,1	-0,1	0	-0,56	-0,57	0,1	0,15	0,53	-0,83	-0,5	0,85	0,35	1	0,99
K _B /K _A	-0,1	-0,11	0	-0,57	-0,58	0,1	0,16	0,54	-0,83	-0,5	0,86	0,32	0,99	1
Cp	0,11	0,11	-0,1	0,53	0,53	-0,06	-0,16	-0,03	0,18	0,22	-0,11	0,06	0,12	0,12
CM	0,1	0,11	-0,01	0,37	0,36	0,03	-0,19	-0,15	0,24	0,24	-0,21	0,83	-0,04	-0,07
COT	-0,13	-0,14	0	-0,63	-0,63	0,06	0,15	0,48	-0,75	-0,5	0,73	0,36	0,92	0,92
MO	-0,14	-0,14	0	-0,62	-0,62	0,06	0,14	0,47	-0,74	-0,5	0,73	0,37	0,92	0,92
Arg	0,06	0,06	0	0,47	0,47	-0,08	-0,13	-0,43	0,64	0,4	-0,64	-0,63	-0,85	-0,83
Dms	-0,14	-0,15	0,01	-0,68	-0,68	0,05	0,18	0,5	-0,78	-0,53	0,76	0,22	0,91	0,91
Dmv	-0,15	-0,15	0	-0,63	-0,63	0,05	0,13	0,46	-0,74	-0,5	0,72	0,33	0,91	0,91
D(0,5)	-0,15	-0,16	0	-0,63	-0,63	0,05	0,13	0,46	-0,73	-0,5	0,72	0,32	0,91	0,91
Ssp	0,21	0,21	-0,01	0,68	0,68	-0,01	-0,14	-0,42	0,69	0,51	-0,67	0,15	-0,77	-0,78
Unif	0,19	0,2	0	0,6	0,6	-0,02	-0,07	-0,39	0,65	0,46	-0,63	-0,13	-0,8	-0,81
Si+Al	0,06	0,06	0	0,38	0,39	-0,07	-0,06	-0,37	0,56	0,33	-0,56	-0,73	-0,8	-0,78
Kao	-0,03	-0,03	0,01	0,04	0,04	-0,07	0,05	-0,15	0,22	0,08	-0,24	-0,92	-0,48	-0,45
ILL	-0,06	-0,06	-0,01	0,23	0,24	-0,05	-0,23	-0,23	0,29	0,2	-0,28	0,02	-0,22	-0,22
Mte	0,2	0,2	-0,01	0,44	0,44	0,05	-0,08	-0,17	0,32	0,29	-0,29	0,7	-0,23	-0,26
Clote	0	0,01	-0,01	0,32	0,32	-0,02	-0,25	-0,21	0,28	0,24	-0,26	0,46	-0,13	-0,14
Qtz	0,19	0,19	-0,01	0,7	0,7	-0,03	-0,16	-0,47	0,75	0,54	-0,73	0	-0,85	-0,86
Calte	0,14	0,15	-0,01	0,5	0,49	0,02	-0,19	-0,24	0,39	0,34	-0,35	0,76	-0,24	-0,27
Dolo	0,17	0,18	-0,01	0,29	0,28	0,06	-0,03	-0,05	0,13	0,17	-0,11	0,79	-0,01	-0,04
MnO	0,13	0,14	-0,01	0,49	0,48	0,02	-0,21	-0,24	0,39	0,34	-0,35	0,74	-0,23	-0,26
Na ₂ O	-0,08	-0,08	-0,01	-0,3	-0,31	0,05	-0,02	0,28	-0,45	-0,26	0,46	0,7	0,71	0,7
K ₂ O	0,03	0,03	0	0,32	0,32	-0,08	-0,05	-0,33	0,5	0,29	-0,51	-0,79	-0,75	-0,73
SiO ₂	0,09	0,1	0	0,54	0,54	-0,07	-0,13	-0,45	0,69	0,44	-0,68	-0,54	-0,89	-0,88
Al ₂ O ₃	0,01	0,01	0,01	0,2	0,21	-0,07	0	-0,26	0,39	0,2	-0,4	-0,86	-0,65	-0,62
Fe ₂ O ₃	0,18	0,18	-0,01	0,61	0,61	0,01	-0,2	-0,33	0,34	0,44	-0,5	0,6	-0,46	-0,48
CaO	0,1	0,11	-0,01	0,36	0,36	0,03	-0,19	-0,15	0,24	0,24	-0,2	0,83	-0,04	-0,07
Mg	0,15	0,15	-0,01	0,31	0,3	0,06	-0,09	-0,06	0,15	0,18	-0,11	0,89	0,03	0

Tableau 4.2 : Résultats du calcul du coefficient de corrélation "R"

Il en ressort comme principales observations que :

- La dose optimale se corrèle moyennement négativement avec la puissance d'agitation lente "Pui" ($R = -0.64$) et moyennement positivement avec Alr ($R = 0.61$), R_{Al} ($R = 0.59$), et le rapport $Dc/Teli$ ($R = -0.52$), et parfaitement avec Dc "la dose consommée" ($R = 0.99$)
- La dose optimale présente de faibles corrélations linéaires avec tous les autres paramètres ($|R| < 0.45$)
- La puissance d'agitation lente "Pui" présente une corrélation satisfaisante avec la turbidité finale Tf ($R = 0.77$) et des corrélations faibles avec les autres paramètres ($|R| < 0.5$)
- Le taux de consommation du coagulant " R_{Al} " se corrèle bien avec le rendement d'élimination de la turbidité " R_T " ($R = 0.81$) et moyennement ($R > 0.50$) avec la constante cinétique de passage K_B , le rapport K_B/K_A , COT , MO , Dms , Dmv , et Ssp .
- Le rendement d'élimination de la turbidité R_T et le rapport R_{Al}/R_T se corrént bien avec la constante cinétique de passage K_B et le rapport K_B/K_A ($R > 0.80$) et présentent une corrélation satisfaisante avec COT , MO , Dms , Dmv et Ssp ($R > 0.70$)
- Quant à l'aluminium résiduel Alr , il présente des corrélations assez faibles avec les différents variables.
- La constante cinétique K_A se corrèle bien avec le CM , Al_2O_3 , Mg et le pourcentage de la kaolinite "kao" ($R > 0.8$), et aussi avec : la calcite "Calte", la dolomite "dolo", la montmorillonite "Mte", Na_2O , MnO , CaO , K_2O et $Si+Al$ ($R \geq 0.7$) alors que La constante cinétique K_B et le rapport, K_B/K_A se corrént bien avec COT , MO , Dmv , Dms , $D(0,5)$ ($R > 0.9$), taux d'Argile "Arg", $Unif$, Qtz , SiO_2 ($R > 0.8$), Na_2O , K_2O , $Si+Al$, et SiO_2 ($R > 0.7$).
- Le rendement d'élimination de la turbidité et le taux de consommation de coagulant semblent se corrént mieux que la turbidité finale et l'aluminium résiduel avec les différents paramètres d'entrée.

En vu de déterminer la vraie nature des relations entre les variables à expliquer (dose optimale, paramètres de finalité de traitement et constantes de formation et de passage) et les variables explicatives, le seul calcul du coefficient de corrélation ne nous suffit pas à expliciter les liaisons entre ces variables. De fait, qu'il ne permet de déterminer que des corrélations linéaires. (Deux variables en totale dépendance non linéaire peuvent avoir un coefficient de corrélation nul), alors les nuages de points sont construits. Les graphiques de la figure 4.2 suivante représentent les nuages de points de la dose optimale en fonction des différents paramètres.

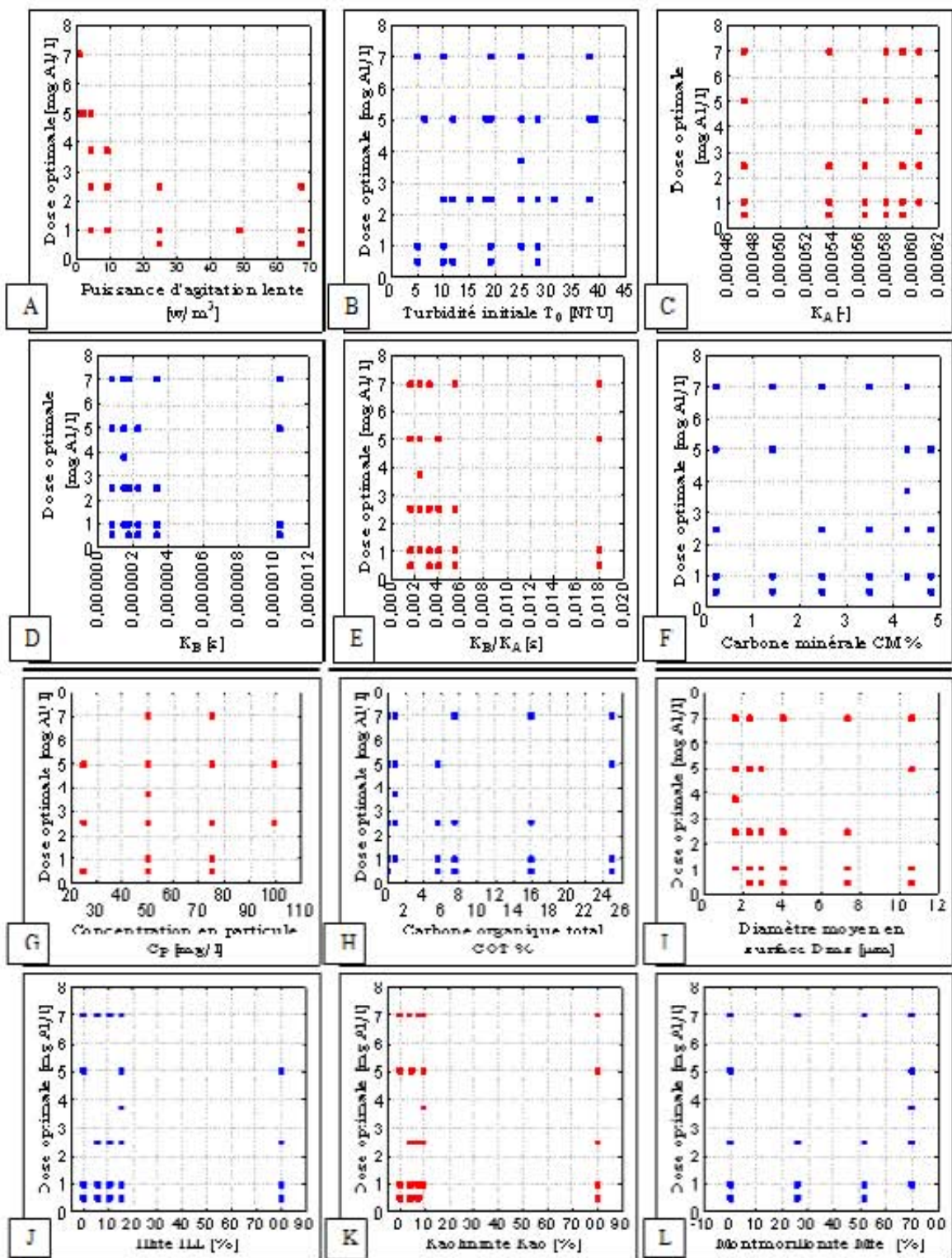


Figure 4.2 : Nuages de points de la dose optimale en fonction de : **A**) puissance d'agitation, **B**) turbidité initiale T_0 , **C**) constante de formation K_A , **D**) Constante de cassage K_B , **E**) K_B/K_A , **F**) carbone minérale CM, **G**) concentration en particule C_p , **H**) carbone organique total COT, **I**) Diamètre moyen en surface D_{ms} , **J**) Illite, **K**) Kaolinite, **L**) Montmorillonite.

L'observation de ces nuages permet de constater qu'il n'existe pas de liaisons linéaires entre la dose optimale et les paramètres : puissance d'agitation lente "Pui", la turbidité initiale "T₀", la concentration en particules solides "Cp", les paramètres "COT, CM, Dms, KB, KA et KB/KA" et les taux de kaolin, de l'illite et de la montmorillonite. Au contraire, il apparaît une liaison exponentielle entre la dose optimale et la puissance d'agitation lente. A fin de rendre cette variable "Pui" conforme aux techniques de modélisation linéaire nous avons procédé, à la base de l'échelle de Turkey (voir figure 2.5 du chapitre 2 à la page 74), à la transformation puissance de cette variable.

Les graphiques de la figure 4.3 suivante représentent les diagrammes de dispersion de la constante cinétique de formation de floc en fonction des différents paramètres, ces graphiques donnent une image des liaisons existantes.

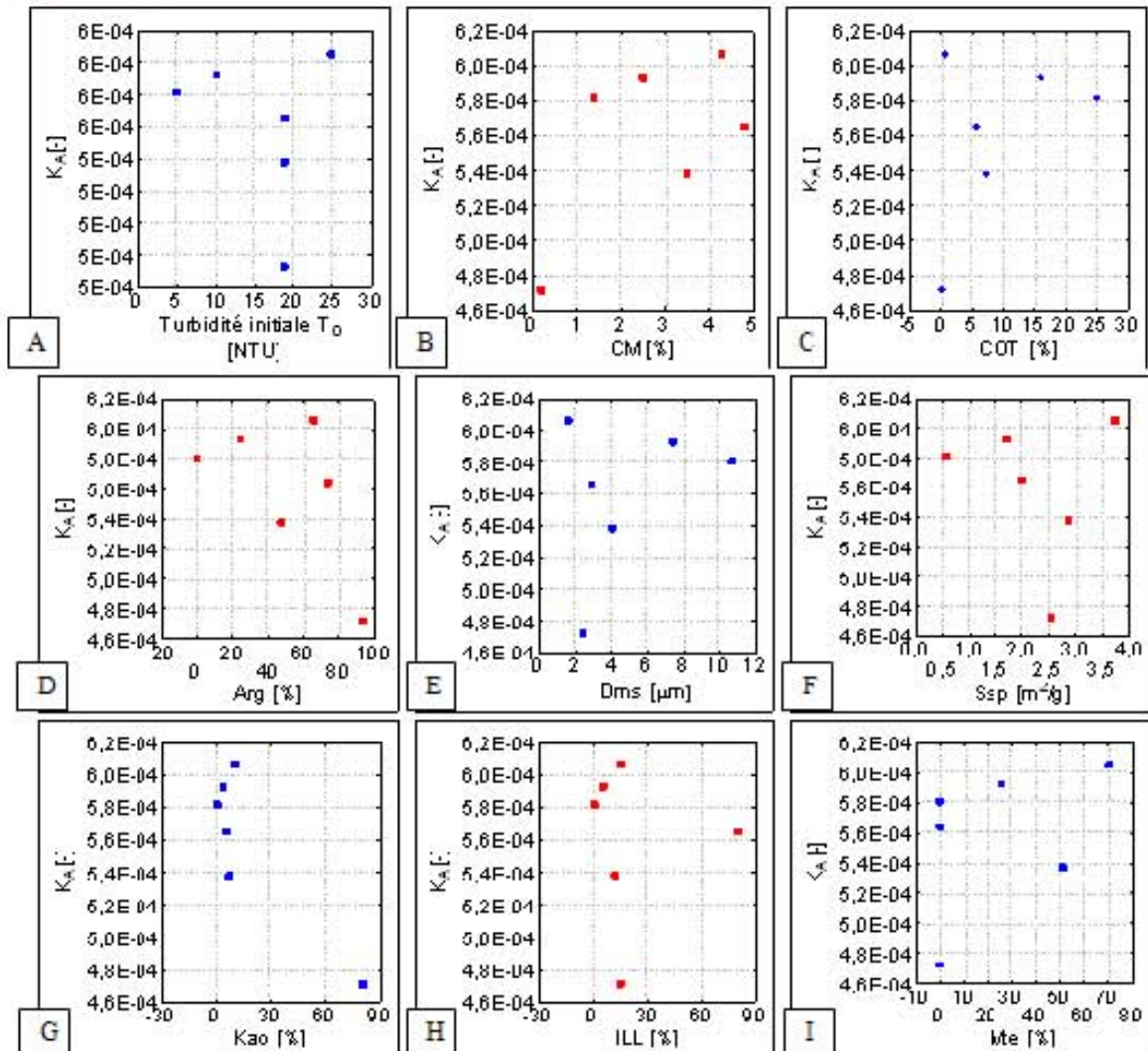


Figure 4.3 : Nuages de points de la constante de formation K_A en fonction de : **A**) turbidité initiale T_0 , **B**) carbone minérale CM, **C**) carbone organique COT, **D**) Argile Arg, **E**) Diamètre moyen en surface Dms, **F**) Surface spécifique Ssp, **G**) Kaolinite Kao, **H**) Illite ILL, **I**) Montmorillonite Mont.

L'observation de la matrice de nuage de points permet de constater qu'aucune liaison linéaire ou autre n'est visualisée entre la constante de formation K_A et les paramètres "T₀, CM, COT, Arg, Dms, Ssp, Kao, ILL et Mont". Quant aux liaisons de la constante de cassage K_B et les mêmes paramètres sont illustrées à la figure 4.4 comme suit :
 On voit que des liaisons non linéaires exponentielles sont constatées entre la constante K_B et les paramètres "Arg, T₀, Ssp, COT et Dms". Mais il faut s'attendre à une grande colinéarité entre ces paramètres.

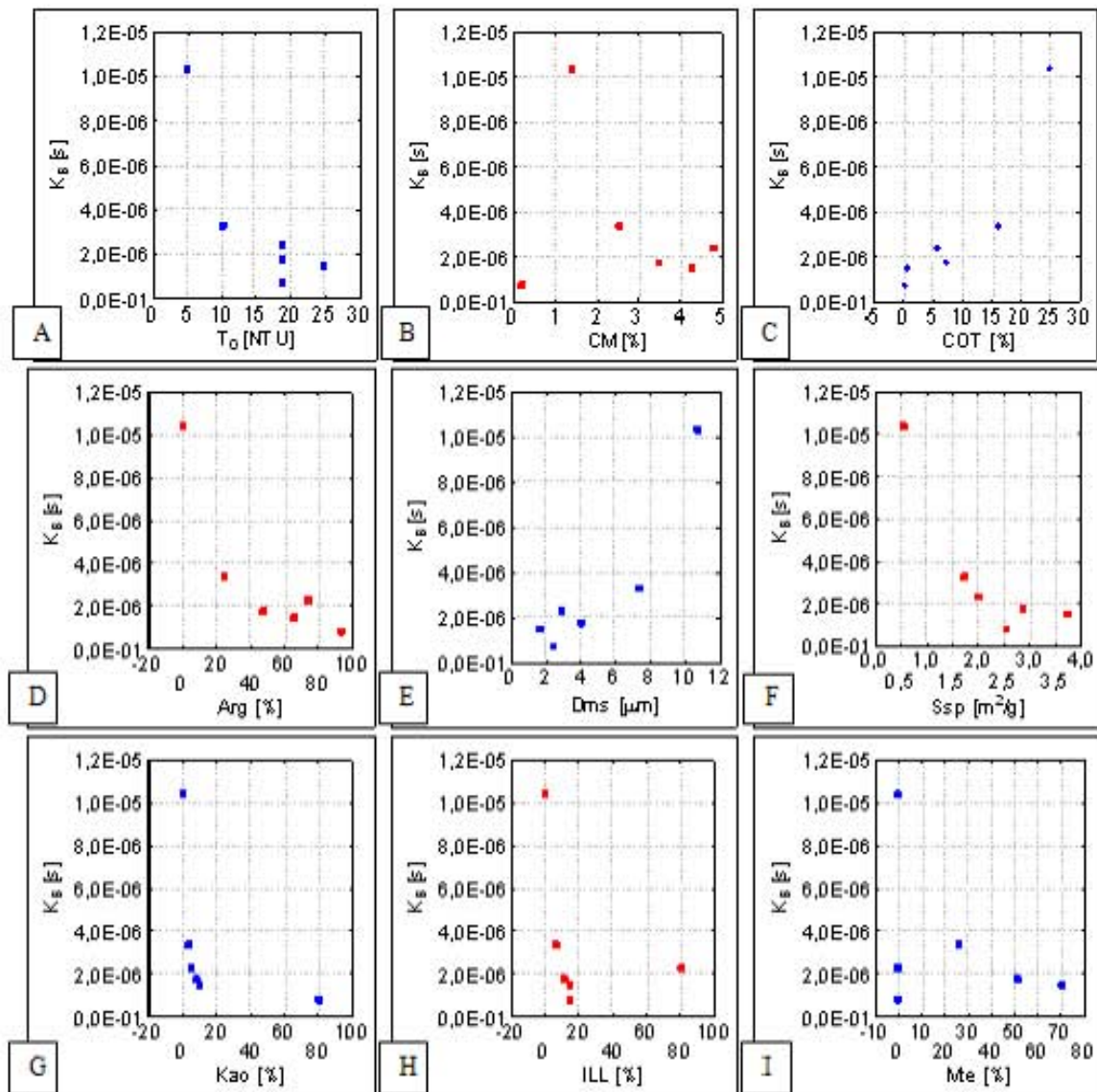


Figure 4.4 : Nuages de points de la constante de formation K_B en fonction de : **A**) turbidité initiale T₀, **B**) carbone minérale CM, **C**) carbone organique COT, **D**) Argile Arg, **E**) Diamètre moyen en surface Dms, **F**) Surface spécifique Ssp, **G**) Kaolinite Kao, **H**) Illite ILL, **I**) Montmorillonite Mont.

Ce calcul bivarié présente un aperçu du réseau d'interrelations existant entre les variables prises deux à deux. La multicolinéarité concerne également les liaisons qui peuvent exister entre les variables explicatives pas seulement prises deux à deux; (il peut exister une relation entre trois variables sans qu'il existe de relation entre deux quelconques de ces trois variables). Or, les éléments de la statistique descriptive univariée et bivariée ne donnent aucune information lorsque les variables sont considérées simultanément. L'étude simultanée des variables est précisément le but de l'analyse multidimensionnelle. Cette méthode nous permet ainsi d'étudier l'apport de plusieurs variables sur une variable.

III-3. ANALYSE MULTIDIMENSIONNELLE.

III-3.1. Analyse en composantes principales

L'Analyse en Composantes Principales (ACP) sert à décrire des tableaux "individus-variables quantitatives" de grande dimension, en examinant les projections du nuage de points sur des plans qui en conserve le mieux la forme. Elle permet également de repérer les groupes de variables fortement corrélées entre elles, et éventuellement de détecter des caractères complexes sous-jacents à ces groupes. Les résultats de l'analyse en composantes principales des données des résultats optimales sont présentés comme suit :

Valeurs propres, et variance expliquée

Dans ce tableau 4.3, sont listées les valeurs propres, et le pourcentage de variance de chaque axe.

	Valeur propre	Variance Totale %	Valeur propre cumulée	Variance Cumulée %
1	16,97	42,42	16,97	42,42
2	11,21	28,02	28,18	70,44
3	4,13	10,32	32,30	80,76
4	2,92	7,29	35,22	88,05
5	1,82	4,55	37,04	92,60
6	1,21	3,02	38,25	95,63
7	0,58	1,44	38,83	97,07
8	0,46	1,16	39,29	98,22
9	0,22	0,56	39,51	98,78
10	0,17	0,41	39,68	99,20
11	0,14	0,36	39,82	99,55
12	0,08	0,21	39,91	99,76
13	0,06	0,16	39,97	99,92
14	0,02	0,07	39,99	99,99
15	0,01	0,01	40,00	100,00

Tableau 4.3 : Résultats de détermination des valeurs propres et des variances expliquées

Il apparait du tableau que le 1^{er} axe, à lui seul revient 42.42 % de variance. Le 1^{er} et le 2^{ème} axe représentent environ plus de 70 % du phénomène, les 6 premiers axes expliquent plus de 95,5 % du phénomène.

Nombre de composantes principales à retenir

Il est difficile de donner une réponse générale à la question : à partir de quel pourcentage peut-on négliger les composantes principales restantes. La qualité des estimations auxquelles conduit l'ACP dépend, de façon évidente, du choix du nombre de composantes à retenir pour reconstituer les données, ou encore de la dimension du sous-espace de représentation. De nombreux critères de choix ont été proposés dans la littérature (Voir chapitre 2).

Pour notre cas d'étude, le choix du nombre de composantes principales est fait à la base de la règle de Kaiser ; On retient les composantes principales correspondantes à des valeurs propres supérieures à 1, il en résulte alors de retenir les six premières composantes principales de variance cumulée de plus de 95%. La figure 4.5 suivante montre la variation de ces différentes composantes.



Figure 4.5 : Evolution des composantes principales

Il apparait clairement qu'à partir de la septième composante, la variation est minimale alors que les six premières composantes sont porteuses d'informations.

Présentation des individus dans le plan principal 1-2

Les composantes C1 et C2 donnent les coordonnées des individus sur le plan principal 1-2, la configuration obtenue est indiquée sur la figure 4.6.

On voit immédiatement apparaître cinq groupes séparés:

- Groupe 1 : Réunit les individus relatifs aux particules du kaolin
- Groupe 2 : Réunit les individus relatifs aux particules de la vase
- Groupe 3 : Réunit les individus relatifs aux particules de la boue et du mélange I
- Groupe 4 : Réunit les individus relatifs aux particules du mélange II
- Groupe 5 : Réunit les individus relatifs aux particules de la tourbe

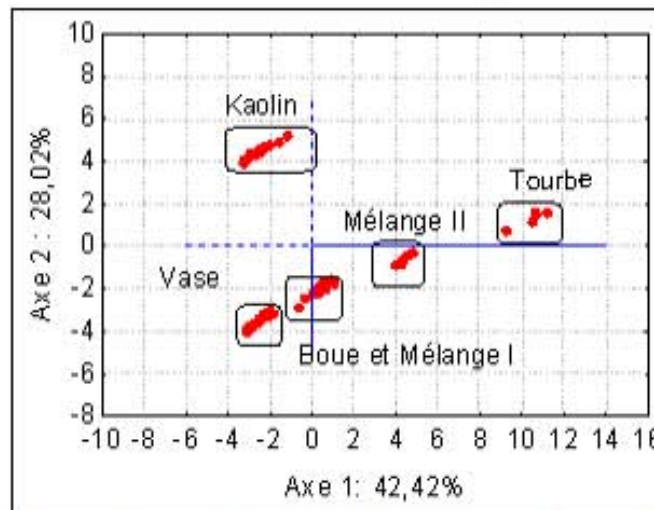


Figure 4.6 : Projection des individus sur le plan principal 1-2

La figure 4.6 fait apparaître que les axes principaux un et deux classifient les individus comme suit :

Axe 1 :

L'axe1 classifie les points individus à priori selon le taux de matière organique dans le solide dispersé

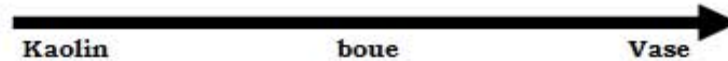


Axe 2

La vase contient de proportion importante en montmorillonite et de faible proportion en kaolinite, par contre le kaolin contient de proportion importante en kaolinite et de faible proportion en montmorillonite, cependant la tourbe présente des valeurs nulles en ces deux éléments.

On peut dire que l'axe 2 fait distinguer la nature de l'argile contenue dans le solide dispersé. Mais si on regarde de près les valeurs de capacité d'échange

cationique de chaque principal composant de ces particules solides, nous pouvons dire que l'axe 2 classe les différents points individuels selon la capacité d'échange cationique (CEC) de la composante argileuse du solide dispersé.



La montmorillonite (majeur constituant de la vase) présente une CEC variant de 80 à 120 meq/100g, l'illite (majeur constituant de la boue) présente une CEC inférieure à celle de la montmorillonite (10-40 meq/100g), alors que la kaolinite (majeure constituant du kaolin) présente les plus faibles valeurs de CEC de 3 à 15 meq/100g CALVET(1977).

LEFEBVRE (1988) note que très souvent la 3^{ème} composante, même si le pourcentage de variance expliqué par la valeur propre correspondante est faible, complète très utilement la description des positions relatives des observations et ceci dans un espace qui reste accessible à nos sens.

La représentation des individus dans le plan 3-4 présenté dans la figure 4.7 ci-dessous permet la classification des individus en deux groupes :

- Groupe 1 : présente les particules de la boue
- Groupe 2 : présente toutes les autres particules

Cela montre la particularité des particules de boue par rapport aux autres particules. La boue englobe les concentrations les plus élevées en Illite.

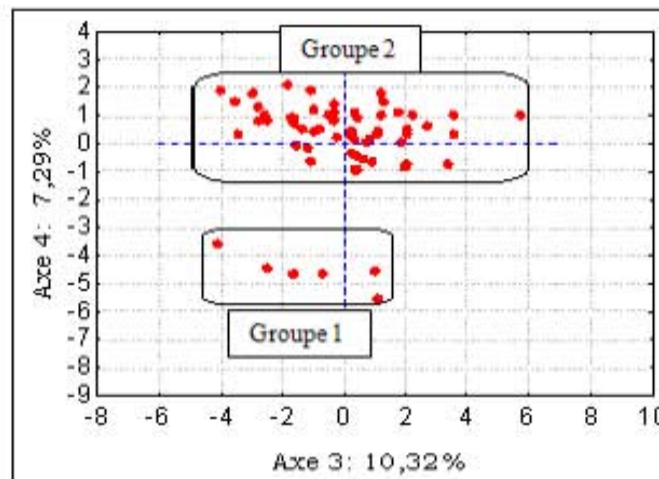


Figure 4.7 : projection des individus sur le plan principal 3-4

La figure 4.8 suivante relative à la représentation des individus dans le plan 5-6 montre un autre partage des individus en trois groupes.

- Groupe 1 : particules solides de concentration élevée (75 mg/l)
- Groupe 2 : particules solides de concentration moyenne (50 mg/l)
- Groupe 3 : particules solides de concentration faible (25 mg/l)

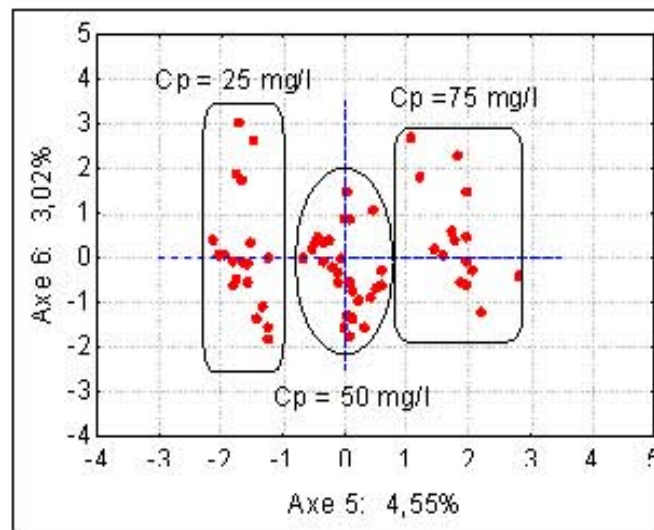


Figure 4.8 : projection des individus sur le plan principal 5-6

Conclusion

Les particules solides mises en suspension présentent des ressemblances et des différences leurs présentations dans les plans principaux a permis de mieux voir leurs caractéristiques ; L'axe 1 montre qu'à 42 %, les particules peuvent être classées selon la teneur en matière organique, l'axe 2 classifie, à 28 % de variance, les individus selon la capacité d'échange cationique de la composante argileuse principale.

Les plans principaux 3-4 et 5-6 permettent respectivement de regrouper, mais à faible variance (inférieure à 15 %), les individus selon la concentration initiale en particules solides et le pourcentage en Illite dans le solide dispersé.

Interprétation des composantes principales et axes principaux

D'une façon générale l'intérêt de l'analyse des composantes est double ; d'une part, cette analyse permet de remplacer les variables, généralement corrélées, par des variables non corrélées, plus faciles à traiter, et d'autre part, l'analyse des composantes permet souvent de réduire dans une certaine mesure le nombre de variables à prendre en considération. En plus l'analyse des composantes permet d'atteindre dans certains cas, selon DAGNELIE (1982, un troisième objectif, en fournissant une interprétation complémentaire des observations.

Examen de la projection des variables sur le plan principal 1-2

La figure 4.9 montre la répartition homogène des variables dans la surface de projection 1-2 et l'absence de l'effet taille, cela indique la bonne présentation de l'ensemble de paramètres du processus floculation.

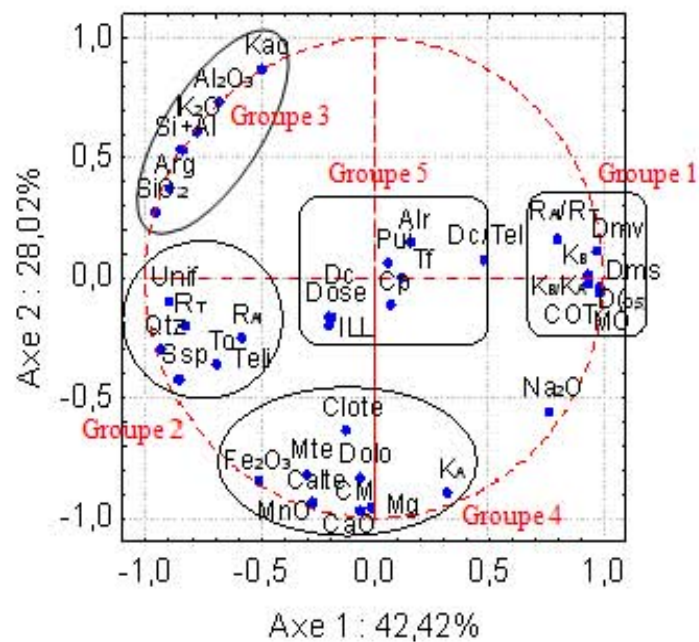


Figure 4.9 : Projection des variables sur le plan principal 1-2

La première composante principale dont l'importance est plus de 42 %, traduit essentiellement l'opposition entre le premier et le deuxième groupe (voir figure 4.9), elle est positivement corrélée avec les paramètres formant le premier groupe et négativement avec les paramètres du deuxième groupe, le tableau 4.4 illustre ces deux groupes.

1 ^{er} groupe	2 ^{ième} groupe
COT	Unif
M.O.	R _T
Dmv	R _A
Dms	Qtz
D(0,5)	Ssp
K _B	Teli
K _B /K _A	To
R _A /R _T	

Tableau 4.4 : Composition du 1^{er} et du 2^{ième} groupe suite à l'analyse par ACP

Il apparaît que la constante de cassage K_B et le rapport entre les constantes cinétiques "K_B/K_A" appartiennent au premier groupe, ils se corrént linéairement positivement avec les paramètres du premier groupe : "COT, MO, Dms, Dmv, R_A/R_T" et négativement avec les paramètres du deuxième groupe : "Unif, R_T, R_A, Qtz, Ssp, Teli, To".

La deuxième composante principale dont l'importance est près de 1.5 fois moindre traduit l'opposition entre le troisième et le quatrième groupe. Le troisième groupe est formé des paramètres suivants "Kao, Arg, SiO₂+Al₂O₃, Al₂O₃ et SiO₂", cependant le quatrième groupe englobe le caractère minéral (CM, dolomite, calcite), les argiles à forte CEC (montmorillonite et chlorite) et la constante cinétique de formation K_A (voir tableau 4.5)

3 ^{ème} groupe	4 ^{ème} groupe	5 ^{ème} groupe
Kao	K _A	Pui
Arg	Calte	Dose
SiO ₂ + Al ₂ O ₃	Dolo	Alr
Al ₂ O ₃	Clote	Tf
SiO ₂	Mte	ILL
K ₂ O	CaO	Cp
	MgO	Dc / Teli
	Fe ₂ O ₃	
	MnO	
	C.M	

Tableau 4.5 : Composition du 3^{ème}, 4^{ème} et 5^{ème} groupe suite à l'analyse par ACP

La projection des variables dans le plan principal 1-2 (Voir figure 4.9) fait apparaître un cinquième groupe composé principalement des paramètres opératoires "Pui, Dose et Cp" et de finalité de traitement "Alr, Tf, Dc/Teli" (voir tableau 4.5), ce groupe prend une position moyenne dans le plan principal 1-2, cette position en milieu de la surface de projection est le trait dominant ; les paramètres de ce groupe sont alors à égale influence des différents groupes de variables, et présentent de faibles corrélations linéaires avec la première et la deuxième composante principale.

La projection sur le plan principal 1-2 montre que la constante de cassage et celle de formation sont interprétées par deux axes différents, la constante de cassage K_B est expliquée par la première composante principale et peut se corrélérer positivement avec les paramètres du premier groupe et négativement avec les paramètres du deuxième groupe, alors que la constante de formation est expliquée par la deuxième composante principale, elle peut se corrélérer positivement avec les paramètres du quatrième groupe.

Cela mène à conclure que la formation et le cassage de floc ne sont pas deux phénomènes antagonistes ; la cinétique de floculation de particules solides peut être simultanément décrite par une valeur élevée de la constante de formation et une valeur aussi élevée de la constante de cassage ou simultanément par de faibles valeurs de ces deux constantes.

Les particules de tourbe en sont un exemple ; la constante de cassage est élevée et la constante de formation est aussi élevée ; Le cassage de floc ne traduit pas nécessairement la faible formation. Ce résultat permet d'éclaircir la mauvaise corrélation linéaire entre la constante de formation et celle de cassage ($R=0,35$).

L'examen de la projection mixte des variables et des solides dispersés sur le plan principal 1-2, illustrée à la figure 4.10, montre clairement que les particules de kaolin sont plus attirées vers le troisième groupe par contre les particules de la vase se trouvent dans le quatrième groupe caractérisé par les argiles de forte CEC.

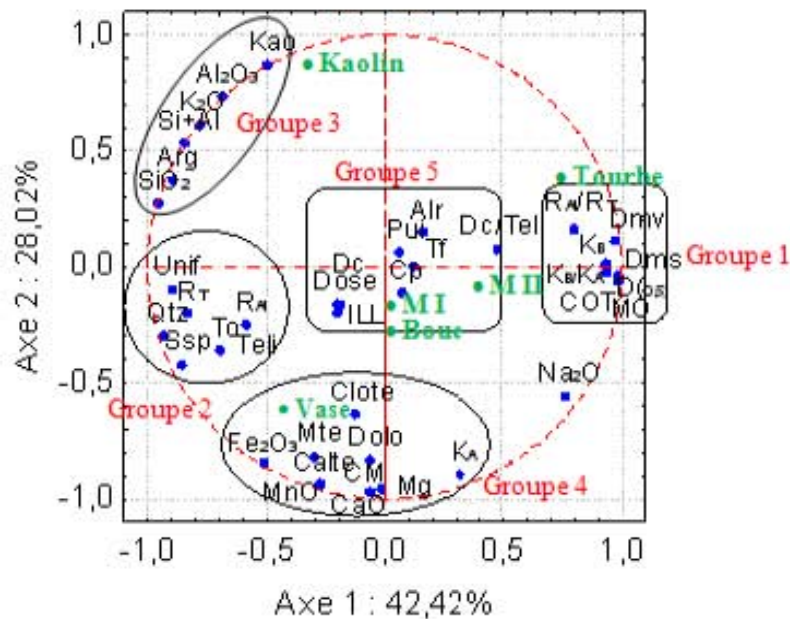


Figure 4.10 : Projection simultanée des variables et des solides sur le plan principal 1-2

Le graphique montre aussi que les particules de tourbe sont du côté du premier groupe caractérisé par les diamètres, la matière organique et la constante de cassage K_B , tandis que les particules organo-minérales de boue, du mélange I et du mélange II prennent des positions moyennes dans le plan de projection.

Examen de la projection des variables sur les plans principaux 1-3 et 1-4

L'axe 03 de la figure 4.11, à lui seul revient 10% de variance, il fait apparaître l'importance des paramètres opératoires et de quelques caractéristiques de particules solides mise en suspension ; Il montre essentiellement l'effet inverse de la dose optimale, et de l'aluminium résiduel contre la puissance d'agitation lente et la turbidité finale.

L'axe 04 de la figure 4.12, expliquant 7% du phénomène, fait montrer l'effet opposant de l'illite "ILL", du chlorite "Clote" et de l'uniformité "Unif" contre la dolomite "Dolo" et la montmorillonite "Mont".

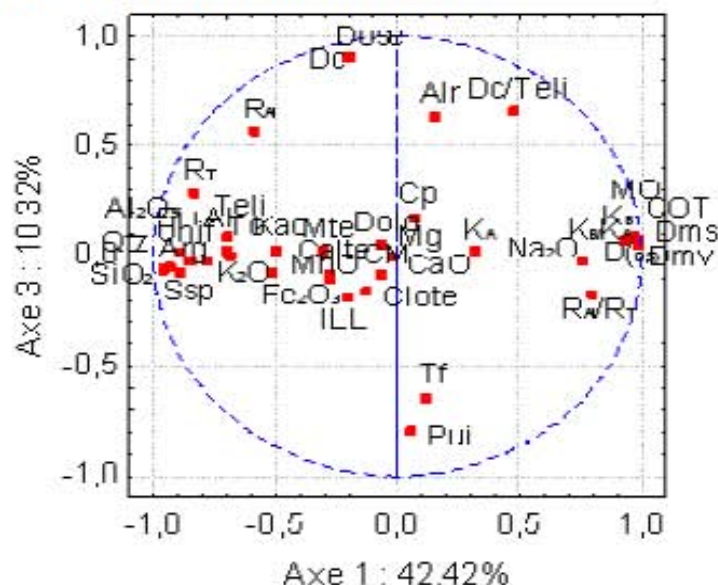


Figure 4.11 : Projection des variables sur le plan principal 1-3

000

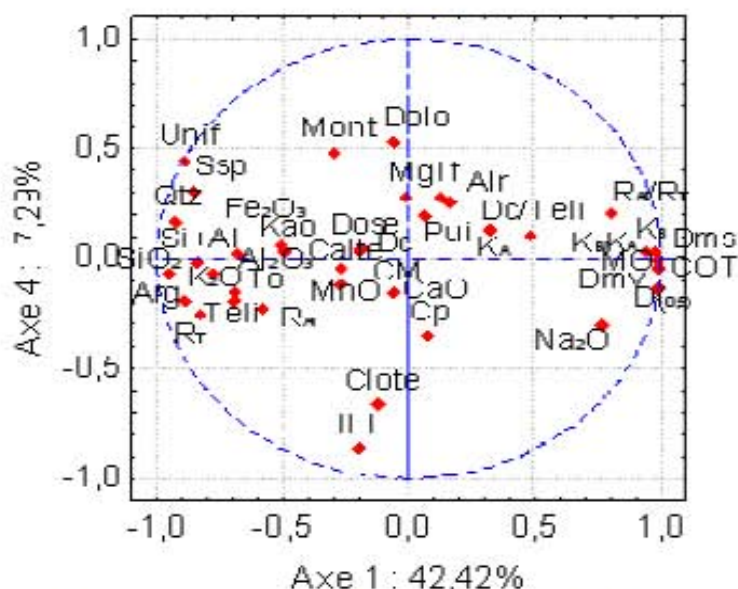


Figure 4.12 : Projection des variables sur le plan principal 1-4

Examen de la projection des variables sur les plans 1-5 et 1-6

L'axe 5, de la figure 4.13, explique 4,55% du phénomène et montre l'effet remarquable de la concentration des particules dans la marge étudiée (25-75 mg/l) sur la floculation alors que l'axe 6 de la figure 4.14 montre l'effet (3 % de variance) de la turbidité finale comme un paramètre d'évaluation de la performance de la floculation.

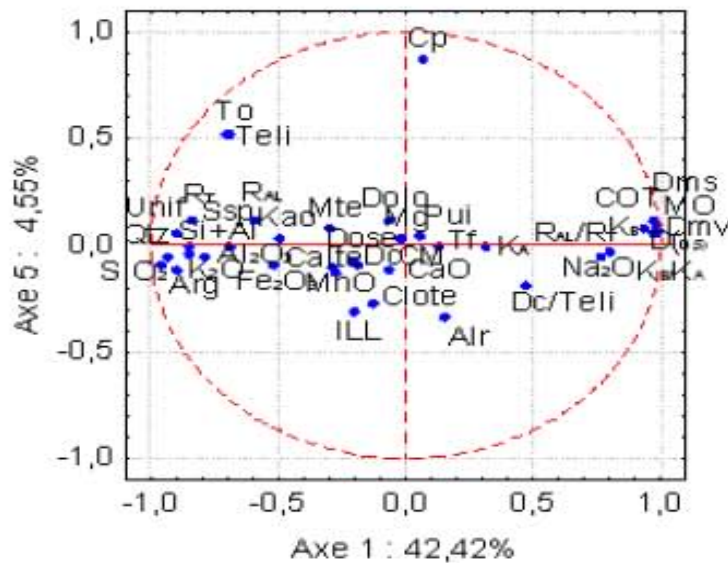


Figure 4.13 : Projection des variables sur le plan principal 1-5

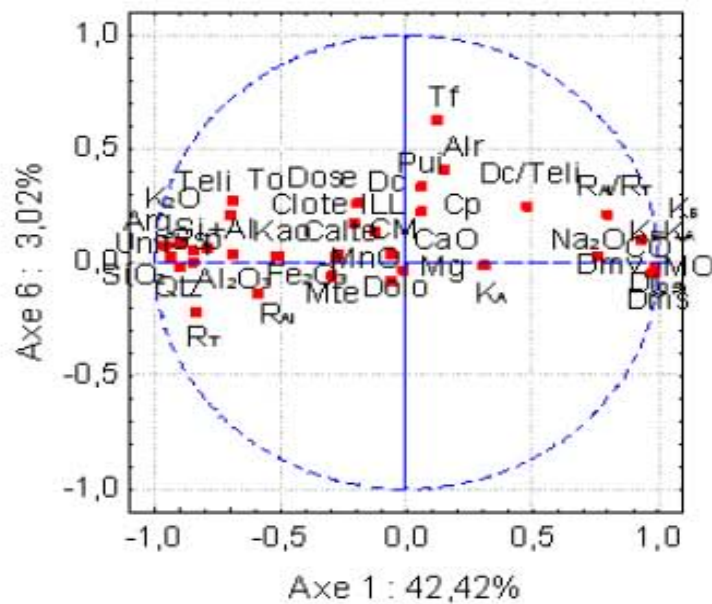


Figure 4.14 : Projection des variables sur le plan principal 1-6

Conclusion

L'analyse en composantes principales a permis de définir deux indices principaux des caractéristiques des solides dispersés.

- Le premier indice : La matière organique et la distribution granulométrique.
- Le deuxième indice : La nature de la matière minérale argileuse.

Cette analyse a permis aussi d'arranger principalement les paramètres de floculation et les caractéristiques des solides dispersés en cinq groupes

- Groupe 1 : Diamètres et matière organique.
- Groupe 2 : Surface spécifique, turbidité initiale et uniformité de distribution granulométrique.
- Groupe 3 : Argile de faible CEC (kaolinite).
- Groupe 4 : Carbone minéral (dolomite et calcaire) et argile de grande CEC (montmorillonite et chlorite).
- Groupe 5 : Paramètres opératoires et de finalité de traitement.

Il est très informative d'examiner les résultats de l'analyse en composantes principales en combinaison avec le dendrogramme de classification hiérarchiques, cela permet de clarifier les présentations des résultats. Dans ce qui suit nous présentons les résultats d'analyse de classification.

III-3.2. Analyse de classification hiérarchique

La méthode de classification hiérarchique est un excellent outil pour l'analyse préliminaire des données, elle examine les distances entre les individus et les regroupant en un nombre restreint de classes, cette information est représentée en graphe à deux dimensions appelé dendrogramme.

Il convient de prendre plus de précaution en cas de classification de variables car la notion de distance entre deux variables pose souvent de délicats problèmes dus à la nature des variables SAPORTA (1990). Dans notre cas d'étude toutes les variables étudiées sont de nature numérique.

Suivant les différents critères d'agrégation et les distances entre individus (voir chapitre 2) nous avons procédé à plusieurs classification hiérarchiques, les figures suivantes 4.15, 4.16, 4.17 illustrent les dendrogrammes de classification correspondantes aux critères de Ward, de moyenne pondérée et de diamètre, pour les autres modes d'agrégation, les variables se prêtent mal à toute classification.

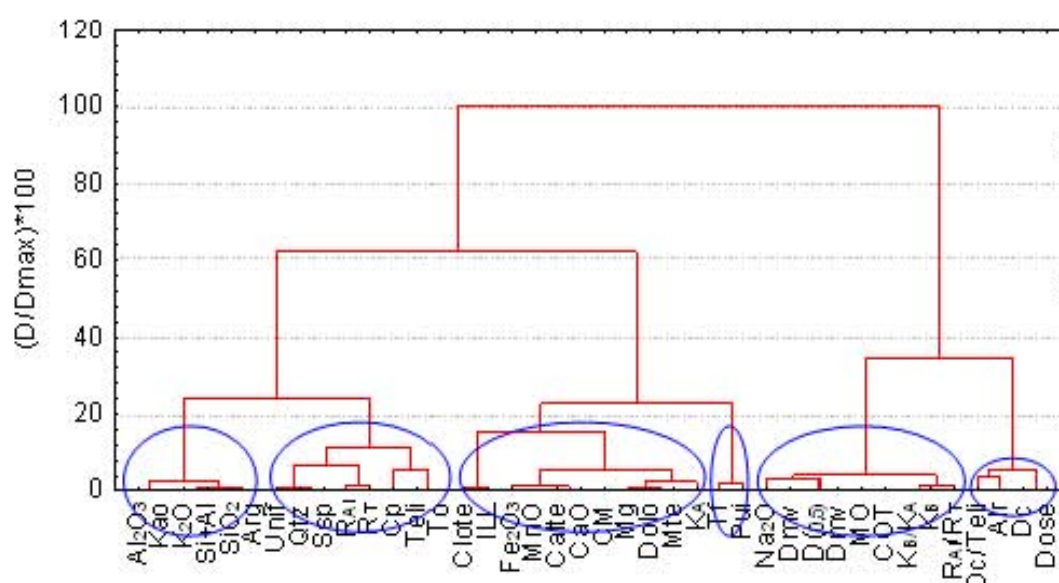


Figure 4.15 : Dendrogramme de classification hiérarchique par la méthode de Ward.

Si on coupe l'arbre de la figure 4.15 à un niveau de 20% (D/Dmax), la classification selon la méthode de Ward permet le classement indiqué dans le tableau 4.6 suivant :

1 ^{er} groupe	2 ^{ème} groupe	3 ^{ème} groupe	4 ^{ème} groupe	5 ^{ème} groupe	6 ^{ème} groupe
COT	Clote	Dose	T0	Arg	Pui
MO	ILL	Dc	Teli	SiO ₂	Tf
Dmv	Fe ₂ O ₃	Alr	Cp	SiO ₂ + Al ₂ O ₃	
Dms	Calte	Dc/ Teli	R ₇	K ₂ O	
D(0,5)	CaO		R _{4a}	K ₂ O	
K _B	CM		Ssp	Al ₂ O ₃	
K _B /K _A	Mg		Qtz		
R _{4a} /R ₇	Dulu		Unif		
Na ₂ O	Mtc				
	K _A				

Tableau 4.6 : Répartition en groupes selon la méthode de Ward

A un niveau de coupure de 40 % de la figure 4.16, la classification selon le critère d'agrégation "moyenne pondérée", ajoute par rapport à la classification précédente un septième groupe composé de l'illite et du chlorite.

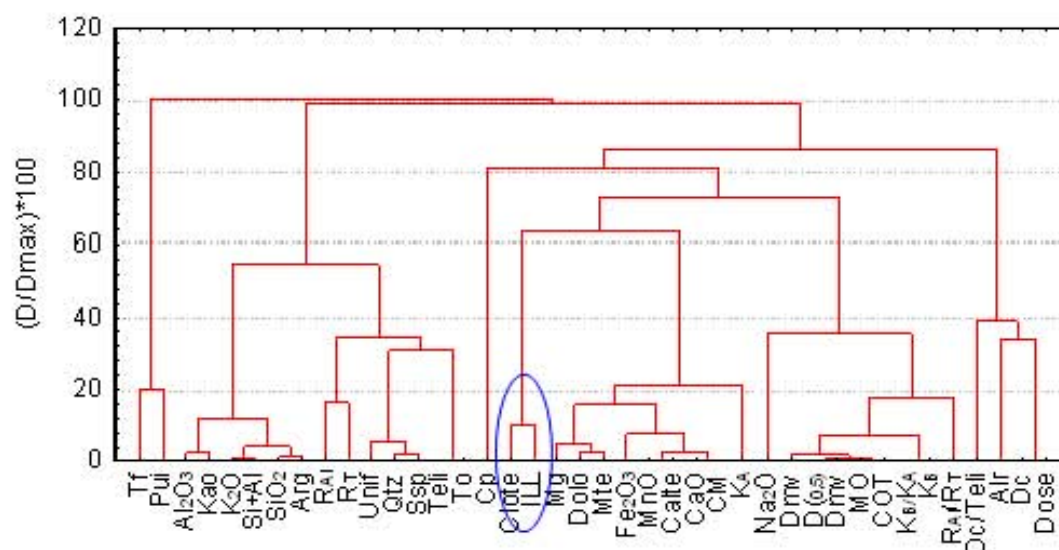


Figure 4.16 : Dendrogramme de classification hiérarchique par la méthode de la moyenne pondérée.

La méthode de classification selon le diamètre pour un niveau de coupure de 24 % illustrée à la figure 4.17 ajoute par rapport à la classification par la méthode "moyenne pondérée" un huitième groupe composé des paramètres K_A et Na_2O .

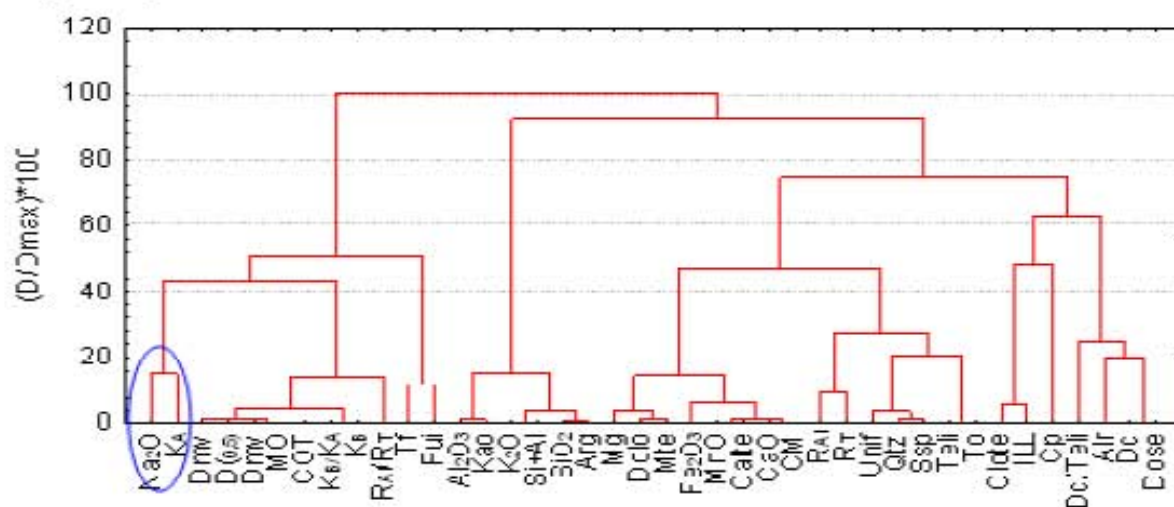


Figure 4.17 : Dendrogramme de classification hiérarchique par la méthode de Diamètre

Conclusion

L'examen de ces dendrogrammes montre l'existence de six groupes selon la classification par la méthode de Ward (voir tableau 4.6) et de sept groupes selon la méthode de moyenne pondérée et de huit groupes selon la classification selon le diamètre.

Ces différentes classifications donnent des arrangements très semblables, nous préférons la classification par la méthode Ward pour :

- Le regroupement est obtenu pour des distances d'agrégation les plus faibles par rapport aux autres méthodes.
- La méthode est la plus utilisée.
- La méthode utilise la notion de variance.

Nous trouvons ici d'après la classification de Ward presque la même typologie retrouvée par l'analyse en composantes principales.

VIII. CHOIX DES VARIABLES POUR LA MISE EN ŒUVRE DES MODELES

La sélection de l'ensemble de variables pour la mise en œuvre de modèle n'est pas claire, le problème vient souvent de la confusion de deux objectifs :

- Garder des variables très liées, donc redondantes, et dont la multiplicité vient renforcer artificiellement l'influence de certains phénomènes.
- Retirer des variables afin de simplifier l'interprétation des axes tout en conservant au mieux les représentations graphiques.

Le premier objectif ne modifie donc pas les représentations en visant à être plus proche de la "réalité", tandis que, le deuxième objectif recherche le sous-ensemble restreint de variables susceptibles d'engendrer le même sous-espace de représentation.

Pour notre cas d'étude on s'est basé pour le choix des variables qui peuvent être intégrer dans les modèles de régression sur :

- La recherche bibliographique sur les différents paramètres utilisés en modélisation statistique de la floculation
- Résultats de l'analyse en composantes principales
- Résultats de l'analyse de la classification hiérarchique.

IV-1. RECHERCHE BIBLIOGRAPHIQUE

GREGORY (1981) a inventorié les différents paramètres et caractéristiques de l'eau brute utilisés pour la modélisation du taux de coagulant selon différents auteurs, le tableau 4.7 récapitule ces paramètres :

Auteurs	ANNÉE	Coagulant	Turbidité	couleur	température	pH	Alcalinité	résistivité	oxydabilité	UV à 254nm	Tc (titration colloïdale)
MAUMARM	1958	S.A.	*				*				
LEWIS	1968	S.A.	*	*		*	*				
ROJANSCHI	1971	S.A.	*		*	*	*				
MANING	1973	S.A.	*		*	*	*				
KAWAMURA	1973	S.A.					*				*
MELBOURNE	1975	S.A.	*	*							
CHIRVAC	1976	S.A.	*		*	*		*			
CAIRO	1976	FeCl ₃	*				*				
BRUECK	1978	S.A.	*		*	*	*				
MUIRA	1978	S.A. et PAC	*		*	*	*				
OVERATH	1980	Fe ₂ (SO ₄) ₃	*	*		*	*		*		
BAZER-BACHI	1986	S.A. et WAC	*		*			*		*	*

S.A. : Sulfate d'aluminium, WAC : Polychlorosulfate basique d'aluminium

PAC : Polychlorure d'aluminium

Tableau 4.7 : Paramètres utilisés pour la modélisation du taux de coagulant selon différents auteurs [inspiré de FRANCESCHI (1991)]

Dans la section modélisation de la floculation du premier chapitre on s'est étalé sur les différents modèles mathématiques statistiques de prédiction du taux de coagulant, les tableaux 4.8 et 4.9 ci-dessous résume respectivement les différents paramètres d'entrée des modèles statistiques de régression multiple et de régression neuronale de la dose du coagulant.

Auteurs	ANNÉE	Coagulant	Turbidité	couleur	MO	MES	[Acide]	Température	pH	Nature de l'argile	Ca++	SO42-	Alcalinité (PAC)	Temps d'agitation	résistivité	UV (254nm)	Vitesse d'égouttement	Debit: m³/s
BRODARD	1986	S.A.	*			*		*										*
GRAVEL	1987	S.A.	+					+	+						+			
CRITCHLEY	1990	-	*	*				*	*						*			
BAZER-BACH	1990	S.A.	*		*			*		*				*	*			
		S.A.	*		*			*		*			*	*	*			
FRANCESCHI	1991	S.A.	*					*	*	*		*	*					
FRANCESCHI	1993	S.A.					*				*		*	*				*
MASCHELEIN	1996	-											*					
HART	1997	-		*					*									
LEEUWEN	2005	S.A.		*												*		

Tableau 4.8 : Paramètres utilisés pour la modélisation du taux de coagulant par régression multiple selon différents auteurs

Auteurs	ANNÉE	Dosages en coagulant ou polymère précédents	Turbidité	couleur	COD.	Température	pH	oxydabilité	Dureté total	Alcalinité	Solide total,	MES	Conductivité	UV (254nm)	Oxygène dissous
GAGNON	1997		*			*	*						*		
VALENTIN	2000		*			*	*						*	*	*
MAIER	2004		*	*	*		*			*				*	
MIRSEPASSI	2004	*	*	*		*	*								
WU	2005	*	*	*											
LAMRINI,	2005		*			*	*	*			*				*
HECTOR	2006		*			*	*		*						
LAMRINI	2007					*	*					*	*		*
DALY	2007		*	*						*				*	
WU	2008														
HEDDAM	2008		*			*	*						*	*	*

Tableau 4.9 : Paramètres utilisés pour la modélisation du taux de coagulant par régression neuronale selon différents auteurs

Conclusion

A la base de cette recherche bibliographique, il vient que les paramètres d'entrée utilisés pour la modélisation de la dose de coagulant peuvent être classés comme suit :

- 1 Paramètres caractérisant le solide dispersé, exprimés en : turbidité, MES, solide total où la turbidité semble être le paramètre incontournable.
- 2 Paramètres caractérisant la matière organique, représentés par : la couleur, l'oxydabilité, MO (perte au feu à 550°C), UV₂₅₄ et C.C.D, il apparaît de la bibliographie que la couleur est le paramètre le plus utilisé.
- 3 La minéralisation de l'eau caractérisée par : la résistivité, la conductivité, la dureté, l'alcalinité (TAC : titre alcalimétrique complet) et les concentrations des ions Ca⁺² et SO₄⁻²
- 4 pH et température du milieu.
- 5 autres : Titration colloïdale, nature de l'argile, vitesse d'agitation, concentration d'acide injecté.

IV-2. ANALYSE EN COMPOSANTES PRINCIPALES

L'ACP est une méthode factorielle conduisant à des représentations graphiques permettant le rangement des variables en groupes. Pour la réduction des variables nous essayons de garder la même répartition (le même regroupement) et nous éliminons de chaque groupe les variables qui expriment la même entité que d'autres ou qui évoluent dans le même sens que d'autres variables, de telle façon que ces suppressions ne modifient que de peu la variance totale expliquée. Le tableau 4.12 résume les paramètres à retenir et ceux à retirer.

	2 ^{ème} groupe	3 ^{ème} groupe	4 ^{ème} groupe	5 ^{ème} groupe
1 ^{er} groupe				
A retenir	A retenir	A retenir	A retenir	A retenir
COT	T _o	Kao	K _A	Pui
Dms	Ssp	Arg	Mte	Dose
K _z	R _r		CM	Alr
K _z /K _A	R _A			Tf
R _A /R _r				ILL
				Cp
A retirer	A retirer	A retirer	A retirer	A retirer
M.O.	Unif	Al ₂ O ₃	Clote	Dc / Teli
D _{IIIv}	Teli	SiO ₂	Calte	
D(0,5)	Qtz	K ₂ O	CaO	
		SiO ₂ + Al ₂ O ₃	MgO	
			Fe ₂ O ₃	
			MnO	

Tableau 4.10 : Répartition des variables retenues et celles retirées à la base de l'analyse en composante principales.

IV-3. ANALYSE DE CLASSIFICATION

En analyse de classification hiérarchique les classes sont obtenues au moyen d'algorithmes formalisés et non par des méthodes subjectives ou visuelles faisant appel à l'initiative du praticien.

Deux variables dont la distance d'agrégation est faible signifient :

- > soit qu'elles évoluent dans le même sens (comportement similaire)
- > soit qu'elles expriment la même entité.

Nous procédons au choix des variables descriptives par élimination (parmi les variables "les plus proches" de distance d'agrégation faible) des variables les moins retrouvées dans la bibliographie. Le tableau 4.11 récapitule les distances d'agrégation et les variables à retirer, alors que le tableau 4.12 résume la répartition finale des variables. A titre d'exemple le CM et le CaO appartenait au même groupe et la distance qui les sépare est très minime, nous retenons le paramètre CM et nous écartons le paramètre CaO, de la même façon nous retenons le Dmv et nous abandonnons D(0,5) ces deux derniers paramètres évoluent dans le même sens.

Variable	Variable	Variable	Distance	Variable à
Obj. N°1	Obj. N° 2	Obj. N° 3	d'agrégation	retirer
CM	CaO		,0000056	CaO
Dmv	D 0,5		,0000859	D 0,5
Dose	Dc		,0001357	Dc
COT	MO		,0008995	MO
Calte	MnO		,0032814	MnO
T ₀	Teli		,0045334	Teli
Si+Al	K2O		,0068251	K2O
COT	Dmv		,0089161	Dmv
Arg	SiO2		,0145632	SiO2
Ssp	Qtz		,0229763	Qtz
KaO	Al ₂ O ₃		,0272184	Al ₂ O ₃
Arg	Si+Al		,0307354	Si+Al
CM	Calte		,0307846	Calte
Mont	Dolo		,0309112	Dolo
Ssp	Unif		,0643303	Unif
Mte	Mg		,0719449	Mg
CM	Calte	Fe ₂ O ₃	,1034292	Fe2O3
ILL	Clote		,1198013	Clote

Tableau 4.11 : Résumé des distances d'agrégation et des variables à retirer à la base de la classification hiérarchique.

1 ^{er} groupe	2 ^{ème} groupe	3 ^{ème} groupe	4 ^{ème} groupe	5 ^{ème} groupe	6 ^{ème} groupe
C.O.T.	ILL	Dose	T ₀	Arg	Pui
Dms	CM	Alr	Cp	Kao	Tf
K _B	Mont		R _T		
R _{Al} /R _T	K _A		R _{Al}		
Na ₂ O			Ssp		

Tableau 4.12 : Répartition des variables retenues à la base de la classification hiérarchique.

Conclusion

Les variables retenues pour la suite du travail de modélisation sont listées comme suit :

- 1 Paramètres caractérisant le solide dispersé : "T₀" et "Cp"
- 2 Paramètres caractérisant la matière organique "COT".
- 3 pH et température sont constants.
pH = 6,5 (suite aux essais préliminaires de recherche de pH optimal).
- 4 Minéralisation est constante (tous les essais ont été conduits avec de l'eau de robinet dont les caractéristiques sont stables).
- 5 Taux d'argile dans le solide dispersé "Arg" et le carbone minéral "CM".
- 6 Type d'argile (Kaolinite, illite, Montmorillonite).
- 7 Diamètre moyen en surface "Dms" et la surface spécifique des particules "Ssp".
- 8 Puissance d'agitation lente "Pui", dose de coagulant "dose".
- 9 Paramètres de finalité de traitement : "Tf", "Alr", "R_{Al}", "R_T", "R_{Al}/ R_T".
- 10 Constantes cinétiques K_A, K_B, K_B/K_A.

IV-4. ANALYSE MULTIDIMENSIONNELLE DES VARIABLES RETENUES.

Cette section est dévolue à la vérification de la stabilité des axes et groupes après réduction des variables, la variance expliquée et la nouvelle répartition des variables sont à revoir après ce choix de variables.

L'analyse en composantes principales est pratiquée pour les 20 variables retenues. La variance totale cumulée jusqu'à la sixième composante principale, comme le montre le tableau 4.13, est de 93,92 % alors qu'elle était de 95.63% pour l'ensemble des 40 variables initialement prises en compte.

	Avant réduction (40 variables)		Après réduction (20 variables)	
	Valeur propre	Variance Cumulée %	Valeur propre	Variance Cumulée %
1	16,97	42,42	8,77	41,76
2	11,21	70,44	4,02	60,91
3	4,13	80,76	2,85	74,47
4	2,92	88,05	1,79	82,98
5	1,82	92,60	1,36	89,45
6	1,21	95,63	0,94	93,92
7	0,58	97,07	0,50	96,27
8	0,46	98,22	0,32	97,78

Tableau 4.13 : Résultats de détermination des valeurs propres et des variances expliquées après réduction des variables.

L'examen de la projection des variables retenues sur les plans principaux 1-2 et 1-3 est illustré sur les figures 4.18 et 4.19 suivantes.

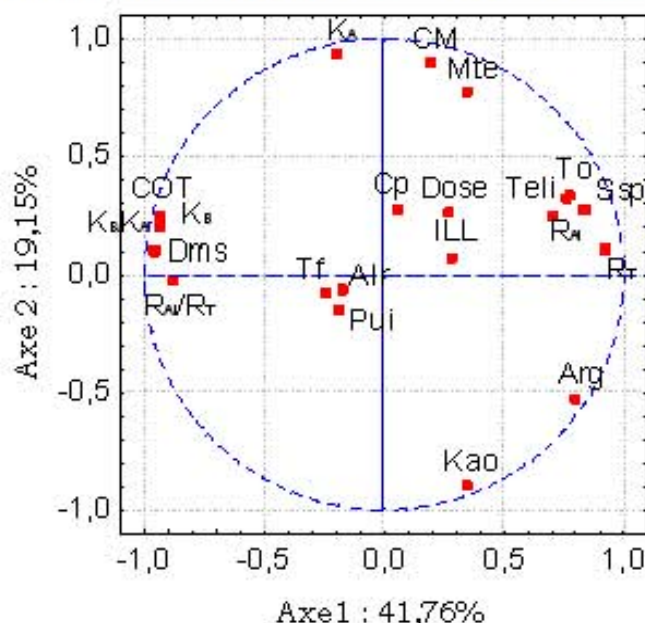


Figure 4.18 : Répartition dans le plan principal 1-2 des variables retenues pour la modélisation

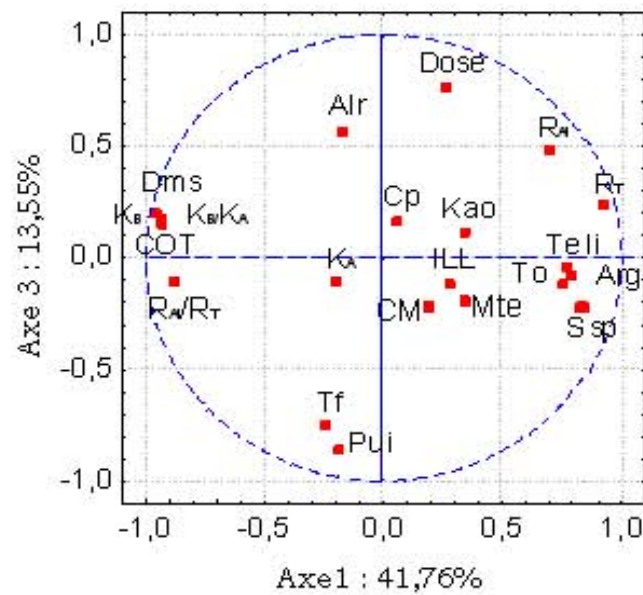


Figure 4.19 : Répartition dans le plan principal 1-3 des variables retenues pour la modélisation

Les figures 4.18 et 4.19 montre la bonne dispersion des variables, le regroupement des variables retenues montre l'importance du caractère organique, de la dimension et de la surface spécifique des particules mises en suspension ainsi que la concentration des particules solides.

Le dendrogramme de la classification hiérarchique, par la méthode de Ward, présenté ci-dessous à la figure 4.20 réaffirme la classification par composantes principales. En effet si on coupe l'arbre de la figure 4.20 à 20%, il en résulte 6 groupes similaires aux groupes trouvés avant la réduction des variables mais avec moins de paramètres dans chacun d'eux.

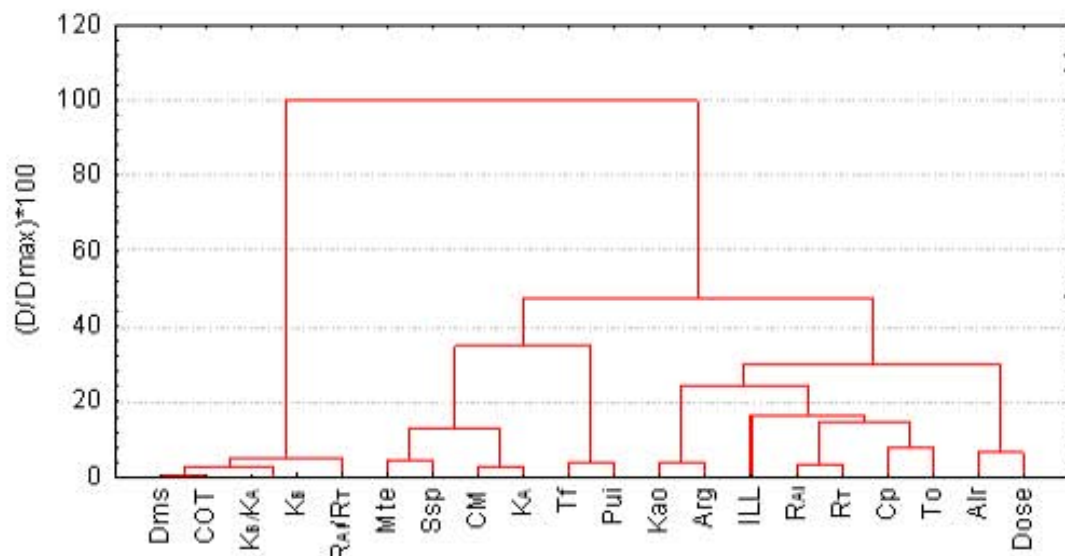


Figure 4.20 : Dendrogramme de classification hiérarchique par la méthode de Ward, des variables retenues pour la modélisation

V. CONCLUSION DU CHAPITRE

Ce chapitre avait pour but l'analyse exploratoire des données de résultats d'étude de la floculation de six types de particules solides dispersées en vue de déterminer les variables descriptives qui peuvent être intégrées dans les modèles de régression. Les données expérimentales, étant de type quantitatif, rassemblent les paramètres opérationnels, les caractéristiques de la suspension, et les paramètres de finalité de traitement.

Dans un premier temps, Les données ont fait l'objet d'une analyse univariée qui a permis d'analyser les variables par l'examen de leurs tendances centrales, distributions et dispersions et l'analyse des diagrammes de dispersion "les box-plot" qui ont servi à décrire les similarités de comportement des différents solides dispersés. Par la suite, l'analyse bivariée a mis en évidence les corrélations entre les variables, alors que les graphiques de nuage de points ont été construits afin de détecter les liaisons existantes entre les différentes variables.

L'analyse des dendrogrammes de l'analyse de classification hiérarchique et les projections des individus et variables sur les plans principaux de l'analyse en composantes principales ont permis de regrouper l'ensemble des caractéristiques des solides dispersés, des conditions opératoires et des paramètres de finalité de traitement principalement en cinq groupes et de montrer deux indices principaux de caractérisation du solide dispersé, le premier indice concerne la matière organique et les distributions granulométriques et le deuxième indice est fonction de la nature de la matière minérale argileuse. Enfin l'analyse exploratoire a été couronnée par le choix des variables qui peuvent être intégrées dans les modèles de régression, comme le montre le tableau 4.14 ci-dessous :

n	Variable	Description
1	Dose	Dose [mgAl/l]
2	Pui	Puissance d'agitation lente [w/m ³]
3	Alr	Aluminium résiduel [mgAl/l]
4	Tf	Turbidité finale [NTU]
5	To	Turbidité initiale [NTU]
6	R _T	Rendement d'élimination de la turbidité [%]
7	R _{Al}	Taux de consommation du coagulant [%]
8	R _{Al} /R _T ,	Taux de consommation du coagulant /Rendement d'élimination de la turbidité [-]
9	Cp	Concentration en particule [mg/l]
10	CM	Carbone minéral [%]
11	COT	Carbone organique total [%]
12	Arg	Taux d'argile [%]
13	Kao	Kaolinite [%]
14	ILL	Illite [%]
15	Mte	Montmorillonite [%]
16	Dms	Diamètre moyen en surface [µm]
17	Ssp	Surface spécifique [m ² /g]
18	K _A	Constante de formation [-]
19	K _E	Constante de cassage [s]
20	K _E /K _A	Constante de cassage/Constante de formation [s]

Tableau 4.14 : Description des variables retenues pour la modélisation

Chapitre 5

MODELISATION STATISTIQUE

I. INTRODUCTION

La coagulation/floculation est l'étape principale dans le traitement physico-chimique des eaux. La modélisation de ce procédé se fait généralement suivant deux approches :

- La modélisation du procédé en utilisant des paramètres descripteurs de la qualité d'eau brute et des conditions opératoires, en effet, cette technique impose de trouver un modèle reliant la dose optimale de coagulant à ces différents paramètres.
- La régulation en aval basée sur la qualité de l'eau traitée après l'étape de floculation, en effet l'efficacité de floculation peut être déterminée en fonction du résultat de l'adjonction du coagulant à l'eau brute.

C'est dans ce cadre que s'insère l'objectif de ce chapitre, il a pour but la modélisation de l'efficacité de la floculation par la recherche de modèles de régression permettant la prédiction de :

- La dose optimale du coagulant,
- Les paramètres de finalité de traitement : l'aluminium résiduel "Alr" et la turbidité finale "Tf",
- Les constantes cinétiques de floculation : la constante de formation K_A et la constante de cassage K_B , la connaissance de ces deux constantes permet, par application des modèles cinétiques de floculation, la détermination de l'efficacité de floculation.

Les premières études de modélisation statistique des paramètres de coagulation et de floculation sont basées sur des modèles linéaires reliant les paramètres explicatives à la variable passive. Ces études ont montré l'intérêt de cette approche mais également les limites de la modélisation linéaire. Le procédé de coagulation et de floculation est complexe et fortement non-linéaire. A cet effet nous proposons dans ce travail d'étudier les modèles suivants :

- a. Modèle de régression multiple
- b. Modèle de régression régularisée.
- c. Modèle de régression neuronale
- d. Modèle de régression neuronale basé sur les composantes principales

En effet, nous exposons, dans ce chapitre, la modélisation statistique de la dose optimale, des constantes cinétiques de floculation et des paramètres de finalité de traitement ; l'aluminium résiduel "Alr", et la turbidité finale "Tf". Les meilleurs modèles d'inférence seront choisis en fonction des critères de qualité d'ajustement et de prédiction et enfin une discussion sur les modèles obtenus, leurs pertinences, stabilités et influences des paramètres régresseurs sera aussi faite. La recherche de ces modèles passera par les étapes suivantes :

- 1 Choix et préparation des données
- 2 Elaboration du modèle
- 3 Validation du modèle

II. MODELISATION PARAMETRIQUE DE LA DOSE OPTIMALE

II-1. CHOIX ET PREPARATION DES DONNEES

Les données utilisées dans cette section sont tirées du fichier brut, il s'agit de résultats optimaux de floculation des six types de particules solides dispersées. La population composée de 61 individus statistiques et 40 variables est d'abord fait l'objet d'une analyse exploratoire multidimensionnelle (voir chapitre 4), cela a permis le choix des variables explicatives qui peuvent être intégrer dans les modèles de prédiction, les variables retenues pour la modélisation sont listées dans le tableau 4.14 du chapitre 04 à la page 165.

En suite la base de données est scindée en deux échantillons ; pour l'identification des coefficients et pour leur validation, respectivement 2/3 de l'échantillon pour l'apprentissage et 1/3 pour la validation, avec au minimum une vingtaine d'événements nécessaires pour pouvoir utiliser correctement les modèles de régression.

Cette sélection doit être réalisée de telle manière que :

- les jeux de données destinées à l'apprentissage soient statistiquement représentatifs de ceux destinés à la validation,
- respecter une certaine homogénéité des caractéristiques des deux échantillons
- le set d'apprentissage doit si possible contenir l'amplitude complète des valeurs des paramètres mesurés.

Alors comme données d'apprentissage nous prenons les résultats de floculation des particules du kaolin, de la boue, du mélange I, et de la tourbe, et pour les données de validation nous prenons les résultats de la floculation des particules de la vase et du mélange II.

Les graphiques "box-plots" communément connus par "boîtes à moustaches" permettent de vérifier la représentativité statistique et l'homogénéité des deux sets d'apprentissage et de validation (Test), la figure 5.1 est une illustration. Sur cette figure le fond, le point central et le plafond de chacune des boîtes représentent respectivement, "la moyenne-2Erreur type", "la moyenne" et "la moyenne+2Erreur type" de la distribution des données, les barres verticales représentent quant à elles les minima et les maxima.

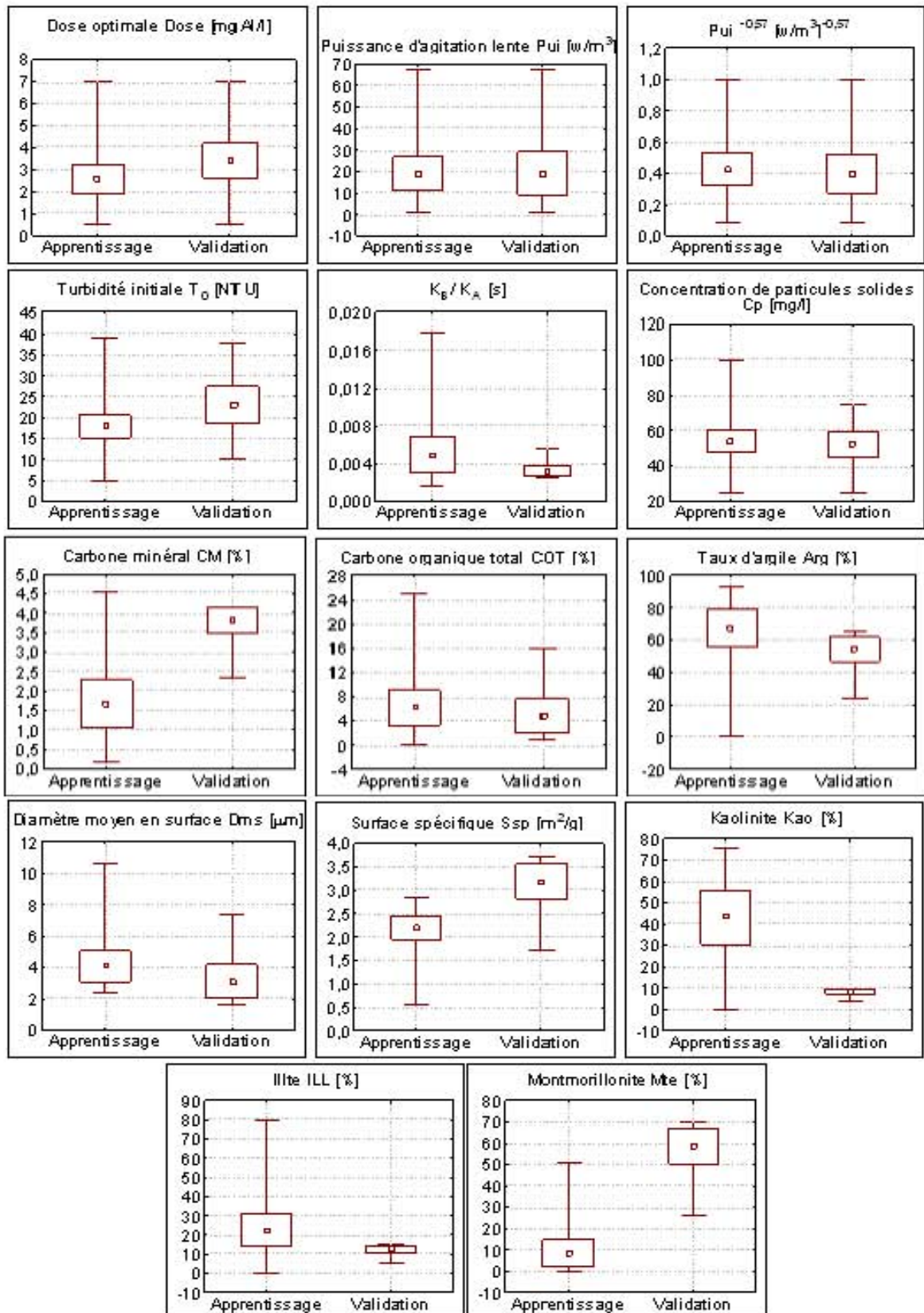


Figure 5.1 : Présentation box-plot «boîtes à moustache» des données d'apprentissage et de validation, Modélisation paramétrique de la dose (Echantillonnage systématique) \square Moyenne \square Moyenne ± 2 Erreur type \top Min-Max

La dispersion des échantillons d'apprentissage est très comparable à celle des échantillons de validation. Trois paramètres présentent des étendus réduits pour le jeu de validation : l'illite, la kaolinite et le rapport K_B/K_A , mais, toujours le jeu d'apprentissage contient l'amplitude complète des valeurs des paramètres mesurés. Plus les boîtes sont identiques, plus les échantillons ont la même répartition statistique. On voit ici une très grande similitude entre ces schémas qui est le signe d'une bonne homogénéité entre les jeux de données.

II-2. MODELES DE REGRESSION MULTIPLE.

Ces modèles consistent à la détermination de la dose optimale en faisant entrer les caractéristiques des suspensions et les conditions opératoires, ils seront élaborés à la base des données des résultats optimaux suivant le schéma de la figure 5.2 ci-dessous.

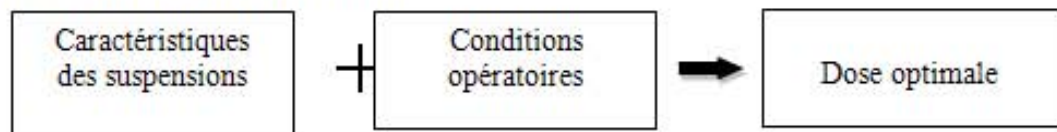


Figure 5.2 : Schéma de régression multiple de la dose optimale

Les choix faits dans en terme de méthodologie suivie en régression multiple ; la détermination des entrées, le partage des données, les modèles testés, la procédure, les critères de choix et la qualité de prédiction du modèle sont résumés dans le tableau 2.1 (chapitre 2).

La recherche pas à pas par la méthode ascendante "Forward Régression" nous à permis de déterminer trois "meilleurs" modèles de régression multiple de la dose optimale. Nous présentons dans ce qui suit le détail de cette modélisation.

II-2.1. Premier modèle de régression multiple de la dose

Les variables exogènes retenues pour le premier modèle statistique de prédiction de la dose optimale sont listées dans le tableau 5.1 ci-après :

N°	Paramètre	Description
1	Pui	Puissance d'agitation lente [w/m³]
2	T _o	Turbidité initiale [NTU]
3	C _p	Concentration en particule [mg/l]
4	COT	Carbone organique total [%]
5	CM	Carbone minéral [%]
6	D _{ms}	Diamètre moyen en surface [µm]

Tableau 5.1 : Paramètres d'entrée du 1^{er} modèle de régression multiple de la dose.

L'équation de prédiction est la suivante :

$$Dose = -22,2468 + 5,9080 \times Pui^{-0,57} - 0,2780 \times T_0 + 0,0913 \times Cp + 2,8220 \times CM - 3,5612 \times COT + 9,7407 \times Dms$$

Les résultats de la régression sont donnés dans le tableau 5.2, dans ce tableau figure :

- Le coefficient de corrélation, le coefficient de détermination et le coefficient de détermination ajusté
- le test de Fisher qui justifie l'apport des variables explicatives au modèle et son niveau de significativité p
- L'erreur standard de l'estimation.
- La valeur de l'ordonnée et des coefficients de régression du modèle.
- Le test de Student qui justifie l'apport de la variable explicative au modèle, et son niveau de significativité p.

+					
R= 0,94 R²= 0,89 R²aj = 0,87					
F(6,30)=39,67 p<,000					
Erreur type de l'estimation 0,74					
	C	Ecart- type de C	I.C _{95%} de C	t(30)	p-niveau
Intercepte	-22,2468	9,246	-40,37 à -4,13	-2,406	0,023
Pui ^{-0,57}	5,9080	0,387	5,15 à 6,67	15,251	0,000
T ₀	0,2780	0,093	0,160 à 0,096	2,995	0,005
Cp	0,0913	0,030	0,033 à 0,149	3,076	0,004
CM	2,8220	1,104	0,66 à 4,99	2,556	0,016
COT	3,5612	1,377	6,26 à 0,86	2,587	0,015
Dms	9,7407	3,906	2,09 à 17, 40	2,494	0,018

Tableau 5.2 : Résultats de la régression, 1^{er} modèle de régression multiple de la dose.

Le modèle présente pour l'échantillon d'apprentissage, des coefficients élevés de corrélation (R= 0,94), de détermination (R²= 0,89) et de détermination ajusté (R²aj = 0,87). Plus ces coefficient se rapprochent de la valeur 1, meilleure est l'adéquation du modèle aux données.

Les résultats des tests F de Fisher et de t de Student sur les coefficients de régression montrent que la valeur de F est grande (F> k= 2,42, k valeur tabulaire, voir annexe B), alors l'hypothèse H₀ de nullité simultanée des coefficients de régression est rejetée par conséquent les variables ont une influence simultanée sur la variable à expliquer. Quant au test de Student pour un niveau de significativité α = 0,05, la variable t de Student est grande (|t| > j_{1-α/2} = 2,04, j valeur tabulaire, voir annexe C) alors l'hypothèse nulle est rejetée ; les coefficients de régression ne sont pas nulles ce qui se traduit

par l'influence réelle des paramètres (Pui, T_0 , CM, COT, Dms) sur la variable à expliquer (la dose optimale),

La valeur "0" n'est pas dans l'intervalle de confiance de l'intercepte, ni dans les intervalles de confiance des coefficients régresseurs, ce qui réaffirme que l'intercepte est différents de 0 et que les régresseurs apportent un pouvoir explicatif réel à la dose optimale.

La comparaison entre le comportement réel et le comportement attendu donné par le modèle fournit une quantité appelée résidu. L'analyse des résidus permet, quant à elle, de confirmer l'hypothèse d'homoscédasticité, de normalité et d'indépendance des résidus.

Le tableau 5.3, ci-dessous, donne pour les données d'apprentissage la synthèse de la régression : les valeurs observées, les valeurs prédites, les résidus et les résidus studentisés, la distance de Cook pour déterminer la présence ou non de points influents.

	Valeur observée	Valeur prévue	Résidus	Résidus studentisés	Distance de Cook	REQM
Minimum	0,50	0,18	-1,42	-1,92	0,00	0,66
Maximum	7,00	6,27	1,35	1,83	0,91	
Moyenne	2,59	2,59	0,00	0,00	0,06	
Médiane	2,50	2,04	-0,09	-0,12	0,01	

Tableau 5.3 : Synthèse des résidus de la régression de la dose, 1^{er} modèle de régression multiple de la dose [mgAl/l] (set d'apprentissage)

La valeur moyenne des résidus est nulle, la valeur de la racine carrée de l'erreur quadratique moyenne REQM est faible (0,66 mgAl/l) cela montre que le modèle est bien ajusté. Les valeurs des résidus studentisés sont bornés dans l'intervalle [-2 : 2] indiquant que, pas de points aberrants et que la divergence entre les doses prévues et observées est systématiquement faible, n'atteigne pas 2 écarts types (2s).

Les valeurs obtenues pour la distance de Cook montrent qu'il n'y a pas de points qui influent considérablement l'ajustement (distance de Cook < 1).

La figure 5.3 des résidus supprimés en fonction des résidus corrobore graphiquement cette observation, le point isolé apparu dans le graphe est relatif à un point expérimental de concentration de particule de 75 mg/l. La vérification de la base de données montre que ce point n'est pas aberrant, alors la suppression de ce point n'est pas nécessaire.

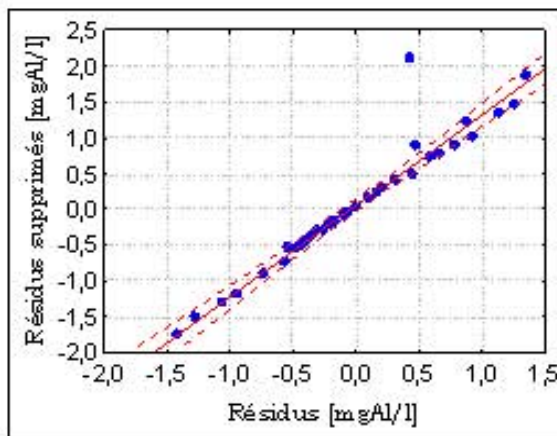


Figure 5.3 : Ajustement des résidus et résidus supprimés, 1^{er} modèle de régression multiple de la dose (set d'apprentissage, IC à 95 %)

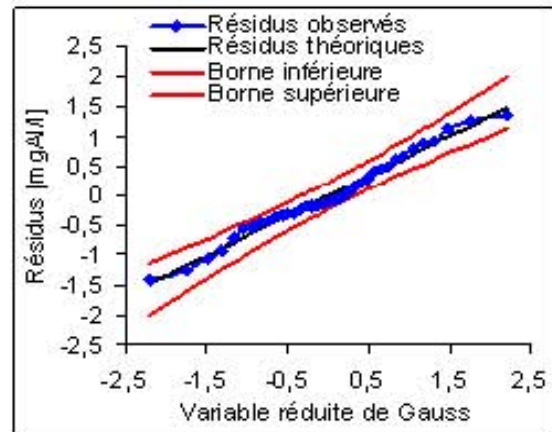


Figure 5.4 : Ajustement à une loi Normale des résidus, 1^{er} modèle de régression multiple de la dose (set d'apprentissage, IC à 95 %)

A) La normalité des résidus

La normalité des résidus de l'échantillon d'apprentissage est évaluée par ajustement graphique à la figure 5.4, les valeurs expérimentales sont proches des valeurs théoriques et sont incluses dans l'enveloppe de l'intervalle de confiance à 95 %. Cette vérification de la normalité des résidus justifie l'utilisation de la statistique F de Fischer et t de Student à l'évaluation de l'apport des variables explicatives sur la variable expliquée.

B) La non autocorrélation des résidus

Le test de Durbin-Watson permet de vérifier si le résidu en i est non-corrélé au résidu en $(i+1)$, la valeur calculé de ce test est de $d=2,64$,

Pour $k=6$ (six variables explicatives) et $n = 37$ (taille de l'échantillon)

$dU = 1,87$ (d'après annexe A)

$d = 2,64 > dU = 1,87$ cela signifie que pas d'autocorrélation des résidus d'ordre 1.

C) Validation du modèle

La validation du modèle statistique établi est réalisée sur le deuxième échantillon (échantillon de validation) composé de données de résultats de floculation des particules de la vase et de mélange II. L'équation de la régression est appliquée sur les données expérimentales de validation, des valeurs de doses optimales sont ainsi obtenues et comparées avec celles expérimentales, alors la qualité de l'ajustement est vérifiée par le calcul des résidus, des résidus studentisés et le REQm, le tableau 5.4 montre les résultats.

		$R = 0,775$ $R^2 = 0,60$			
	Valeur observée	Valeur prévue	Résidus	Résidus studentisés	REQM
Minimum	0,50	-0,26	-1,49	-0,80	1,45
Maximum	7,00	6,77	3,126	1,67	
Moyenne	3,48	2,75	0,73	0,39	
Médiane	2,50	2,38	0,68	0,36	

Tableau 5.4 : Synthèse des résidus de la régression, 1^{er} modèle de régression multiple de la dose [mgAl/l] (set de validation)

Le coefficient de détermination $R^2=0,60$ indique que 60% de la variabilité de la dose optimale de l'échantillon de validation est expliquée par le modèle. Les valeurs moyennes de résidus (0,73 mgAl/l) est proche de zéro, la valeur de REQM est faible, les résidus studentisés sont bornés dans l'intervalle [-1 à 2] ce qui signifie l'absence des points aberrants, et que les fluctuations de la dose optimale estimée est faible ne dépassant pas en valeur absolue 2 écarts types (2s).

D) L'homoscédasticité (égalité des variances)

Le graphique de la figure 5.5, présentant le nuage de points des résidus calculés sur l'échantillon de validation permet de vérifier l'ajustement global, et de repérer les points aberrants, ainsi que de vérifier l'hypothèse d'égalité de variance du vecteur ε (résidu). Aucune structure (cône ou vague) n'est observée, l'homoscédasticité est alors vérifiée. Quant à la normalité des résidus, elle est vérifiée par ajustement graphique à la figure 5.6.

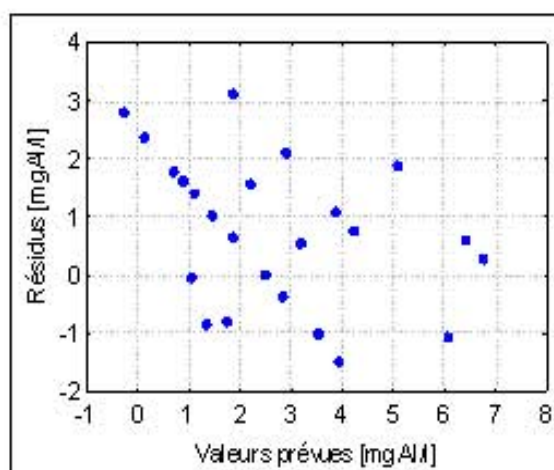


Figure 5.5 : Vérification de l'homoscédasticité par le nuage de points des résidus, 1^{er} modèle de régression multiple de la dose (set de validation)

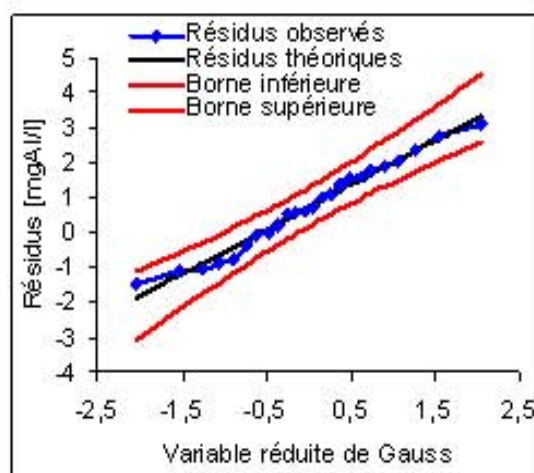


Figure 5.6 : Ajustement à une loi Normale des résidus, 1^{er} modèle de régression multiple de la dose (set de validation, IC à 95 %)

La figure 5.7, ci-dessous, illustre l'ajustement des valeurs prédites et les valeurs observées cas des échantillons d'apprentissage et de validation. L'ajustement est d'autant meilleur que les valeurs se rapprochent de la première bissectrice.

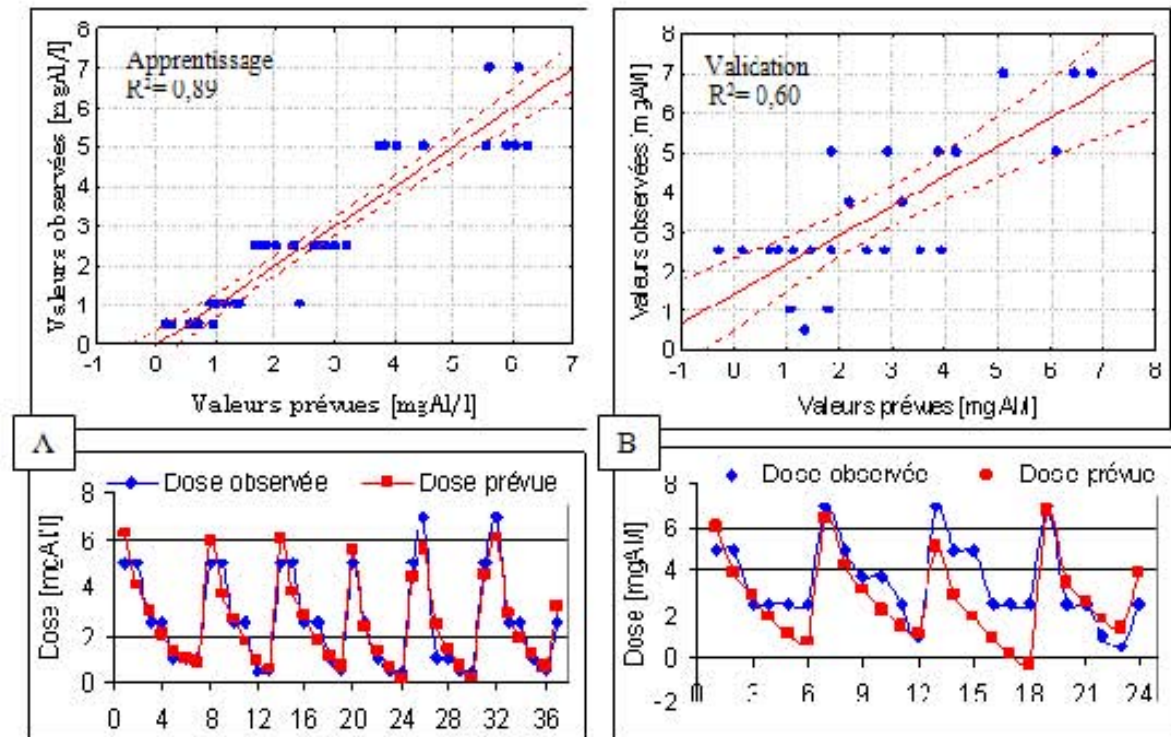


Figure 5.7 : Ajustement des valeurs observées et des valeurs prévues, 1^{er} modèle de régression multiple de la dose **A** : set d'apprentissage, IC à 95% **B** : set de validation, IC à 95%.

E) Stabilité des coefficients de régression et influence des paramètres sur l'équation du modèle.

La grande corrélation entre les variables de régression "T₀, C_p, COT, et D_{ms}" comme le montre le tableau des corrélations 4.2 (chapitre 4) et les faibles valeurs de "tolérance" calculées pour ces régresseurs (< 0,1), montre la multicollinéarité, ce qui risque l'instabilité des coefficients de régression et le changement de leurs signes à ceux auxquels on s'attendait (par exemple le signe de paramètre régresseur COT).

Alors plus d'attention doit être prise à la recherche des effets des variables explicatives sur la variable à expliquer. La méthode de recherche d'influence consiste à pondérer chacun des coefficients du modèle par la valeur moyenne associée de la variable. Nous obtenons ainsi six coefficients pondérés comme suit :

$$C_j = C_j \cdot \bar{x}_j$$

$j = 1 \dots p$ les paramètres de modèle : Pui^{-0,57}, T₀, Cp, COT, CM et Dms

$i = 1 \dots n$ le nombre d'unités statistiques (n= 37)

C_j = les coefficients du modèle

\bar{x}_j = la moyenne du paramètre

L'influence Ix_j (%) de chaque variable de l'équation est ensuite calculée de la

manière suivante :

$$Ix_j (\%) = \frac{|c_j|}{\sum_{j=1}^p |c_j|} \cdot 100$$

La figure 5.8 montre l'influence en pourcentage des différents paramètres de l'équation de régression

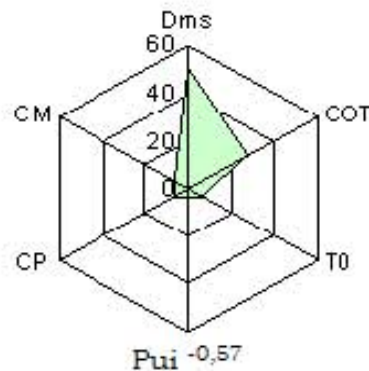


Figure 5.8 : Influence des paramètres sur l'équation de la régression, 1^{er} modèle de régression multiple de la dose.

Dans cette équation le diamètre moyen en surface "Dms" présente la plus grande influence sur la dose (50,52 %), suivi par le COT qui présente 27,95% d'influence, cependant l'influence des autres paramètres n'excède pas 25%, avec un minimum d'influence (3,22%) relatif au paramètre "Pui^{-0,57}", 6,22 % d'influence de la turbidité initiale "T₀", 6,06% d'influence de la concentration de particule "Cp" et 6,04 % d'influence du carbone minéral "CM".

II-2.2. Deuxième modèle de régression multiple de la dose

Les variables exogènes retenues pour le deuxième modèle statistique de prédiction de la dose optimale sont présentées dans le tableau 5.5 suivant :

N°	Paramètre	Description
1	Pui	Puissance d'agitation lente [w/m³]
2	T _o	Turbidité initiale [NTU]
3	C _p	Concentration en particule [mg/l]
4	COT	Carbone organique total [%]
5	Arg	Taux d'argile [%]

Tableau 5.5 : Paramètres d'entrée du 2^{ième} modèle de régression multiple de la dose.

Le meilleur ajustement est mixte. Le modèle ci-dessous prédit la dose optimale en fonction de la puissance d'agitation lente, la turbidité initiale, le carbone organique total et le taux d'argile dans le solide dispersé.

$$Dose = 5,5866 + 5,8966 \times Pui^{-0,57} + 0,0747 \times T_o - 0,2230 \times C_p - 0,3361 \times COT - 0,0506 \times Arg$$

Les résultats de la régression sont donnés dans le tableau 5.6 suivant :

R=0,94 R ² = 0,89 R ² aj = 0,87 F(5,31)=47,999 p < 0 000 Erreur type de l'estimation 0,74					
	C	Ecart type de C	I.C _{95%} de C	t(31)	p-niveau
Intercepte	5,5866	2,067	1,54 à 9,64	2,700	0,011
Pui ^{-0,57}	5,8966	0,385	5,14 à 6,65	15,315	0,000
T _o	0,0747	0,021	0,033 à 0,117	3,489	0,0015
C _p	-0,2230	0,063	-0,347 à -0,099	-3,522	0,0014
COT	-0,3361	0,104	-0,540 à -0,133	-3,239	0,0029
Arg	-0,0506	0,020	-0,090 à -0,011	-2,488	0,018

Tableau 5.6 : Résultats de la régression, 2^{ième} modèle de régression multiple de la dose.

Le coefficient de détermination R²= 0,89 ce qui indique que l'adéquation du modèle aux données est acceptable.

En ce qui concerne la signification globale de la régression, la statistique F de Fischer permet de connaître l'apport global de l'ensemble des variables "T_o, C_p, Pui^{-0,57}, COT et Arg" à la détermination de la dose optimale, la valeur de F est grande (F > k= 2,523, k : valeur tabulaire, voir annexe B) donc on rejette l'hypothèse nulle H₀ qui est l'éventuelle non influence simultanée des variables explicatives.

Le test des espérances de Student est aussi vérifié ; pour un niveau de significativité choisi $\alpha < 0,05$, la variable de Student t est grande ($|t| > j_{1-\alpha/2} = 2,0395$; valeur tabulaire, voir annexe C), les coefficients de régression C sont significativement différent de zéro au niveau α , cela indique que l'apport de chacune des variables explicatives (Pui^{0,57}, T₀, Cp, COT et Arg) est réelle. La valeur 0 n'est pas dans l'intervalle de confiance de l'intercepte ni dans les intervalles de confiance des coefficients de régression ce qui réaffirme que l'intercepte est différents de 0 et que les régresseurs apportent un pouvoir explicatif réel à la dose optimale.

La qualité de l'ajustement est testée par l'analyse des résidus, les résidus doivent être indépendants et identiquement distribués selon une loi normale, le tableau 5.7 ci-dessous donne pour les données d'apprentissage la synthèse des résidus : les valeurs observées, les valeurs prédites, les résidus, les résidus studentisés, le REQM et la distance de Cook ; elle renseigne sur les variations des estimateurs des paramètres dues à la suppression d'une observation et donc à la robustesse de ces estimations.

	Valeur observée	Valeur prévue	Résidus	Résidus studentisés	Distance de Cook	REQM
Minimum	0,50	0,00	-1,52	-0,75	0,00	0,67
Maximum	7,00	6,27	1,26	0,62	0,47	
Moyenne	2,59	2,59	0,00	0,00	0,05	
Médiane	2,50	2,05	-0,14	-0,07	0,01	

Tableau 5.7 : Synthèse des résidus de la régression, 2^{ième} modèle de régression multiple de la dose [mgAl/l] (set d'apprentissage)

La valeur moyenne des résidus est nulle et la valeur de REQM est faible (0,67 mgAl/l), Les résidus studentisés sont bornés dans l'intervalle [-1 à 1] cela montre l'absence de points aberrants. Les valeurs obtenues pour la distance de Cook montrent que le changement en retirant l'observation i sur les estimations de l'ensemble des coefficients est faible (distance de Cook < 1). La figure 5.9 des résidus supprimés en fonction des résidus confirme graphiquement cette observation. Le modèle a priori ajuste bien les données d'apprentissage.

La normalité des résidus

La normalité des résidus est, quant à elle, évaluée par ajustement graphique ; la figure 5.10 montre l'ajustement normal des résidus du jeu d'apprentissage ; les valeurs expérimentales sont proches des valeurs théoriques et sont incluses dans l'enveloppe de l'intervalle de confiance de 95%.

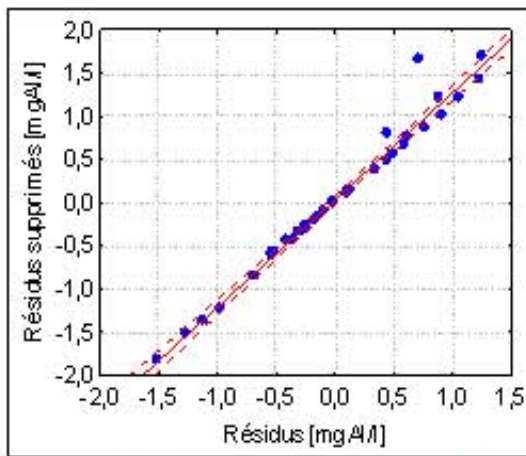


Figure 5.9 : Ajustement des résidus et résidus supprimés, 2^{ème} modèle de régression multiple de la dose (set d'apprentissage, IC à 95 %)

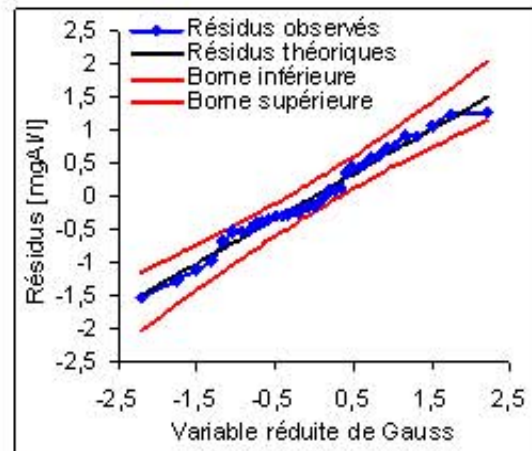


Figure 5.10 : Ajustement à une loi Normale des résidus, 2^{ème} modèle de régression multiple de la dose (set d'apprentissage, IC à 95 %)

La non autocorrélation

Les causes de l'auto-corrélation sont une mauvaise spécification du modèle, ou l'absence d'une variable importante, le test de Durbin-Watson sur les résidus permet d'apprécier le degré de l'autocorrélation. La valeur de ce test est de 2,58, cette valeur est supérieure à la valeur de "dU" (dU= 1,795) Pour k =5 (cinq variables explicatives) et n= 37 (taille de l'échantillon) $dU = 1,795$ (d'après l'annexe A) $d=2,58 > dU = 1,795$ alors pas d'autocorrélation d'ordre 1 des résidus.

Validation du modèle

La validation du modèle statistique est réalisée sur le deuxième échantillon (échantillon de validation) composé de données de résultats de floculation des particules de la vase et de mélange II. L'équation de la régression est appliquée sur les données expérimentales de validation ; des valeurs de doses optimales sont ainsi obtenues et comparées avec celles expérimentales, le tableau 5.8 résume la synthèse des résidus sur le set de validation.

<i>R= 0,81 R²= 0,66</i>					
	Valeur observée	Valeur prévue	Résidus	Résidus studentisés	REQM
Minimum	0,50	-0,31	-0,86	-0,46	1,52
Maximum	7,00	6,38	3,17	1,69	
Moyenne	3,48	2,47	1,01	0,54	
Médiane	2,50	2,02	1,02	0,51	

Tableau 5.8 : Synthèse des résidus de la régression, 2^{ème} modèle de régression multiple de la dose [mgAl/l] (set de validation)

Les valeurs de résidus studentisés sont bornées dans l'intervalle $[-0,5 \text{ à } 2]$, les valeurs moyennes des résidus et le REQM sont faibles, $R^2=0,66$ ce qui signifie que 66% de la variabilité de la dose optimale est validée par le modèle.

L'homoscédasticité et normalité des résidus

L'homoscédasticité est vérifiée graphiquement par traçage d'un graphique avec en ordonnées les résidus et en abscisses la valeur prévue (voir figure 5.11) ; les résidus se répartissent régulièrement en haut et en bas du zéro et aucune forme de cône ou vague n'est observée, l'homoscédasticité est vérifiée et l'ajustement est considéré satisfaisant

Quant à la normalité des résidus, elle est vérifiée par ajustement graphique, les résidus expérimentaux sont alignés autour des valeurs de résidus théoriques (voir figure 5.12).

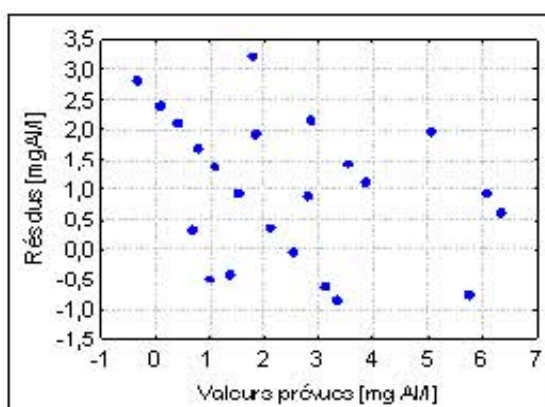


Figure 5.11 : Vérification de l'homoscédasticité par le nuage de points des résidus, 2^{ème} modèle de régression multiple de la dose (set de validation)

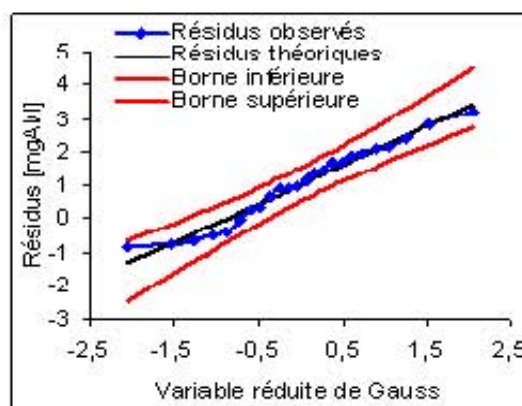


Figure 5.12 : Ajustement à une loi Normale des résidus, 2^{ème} modèle de régression multiple la dose (set de validation, IC à 95 %)

La figure 5.13, ci-dessous, croise pour les sets d'apprentissage et de validation, les valeurs observées avec les valeurs prédites, l'ajustement est d'autant meilleur que les valeurs se rapprochent de la première bissectrice. Le coefficient de détermination R^2 entre les valeurs observés et prévues est de 0,89 pour le jeu d'apprentissage est de 0,66 pour le jeu de validation, l'ajustement est acceptable.

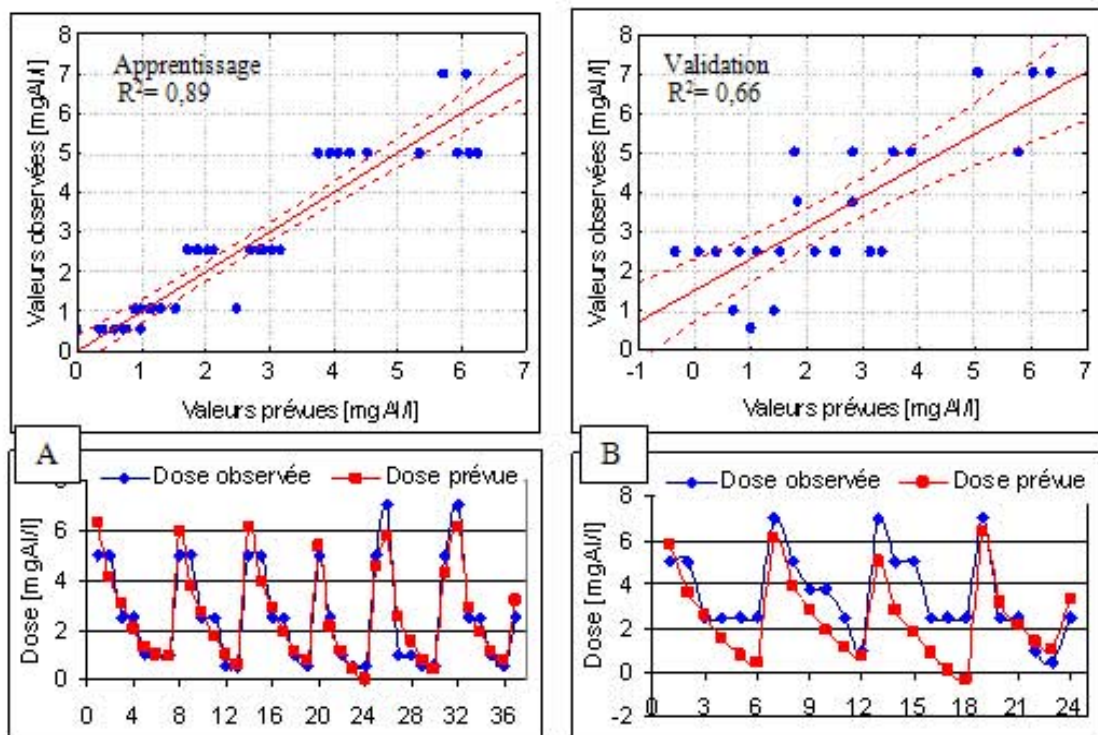


Figure 5.13 : Ajustement des valeurs observées et des valeurs prévues, $2^{ème}$ modèle de régression multiple de la dose [A] : set d'apprentissage, IC à 95% [B] : set de validation IC à 95%.

Stabilité des coefficients de régression et influence des paramètres sur l'équation du modèle.

Le calcul des valeurs de tolérance donne pour le paramètre "Pui^{-0.57}" la valeur de 0,97 et pour le "Cp" la valeur de 0,13, pour les autres paramètres "To, COI" et "Arg" la moyenne des tolérances est inférieure à 0,1 ce qui le signe de multicollinéarité et qui risque l'instabilité des coefficients de régression.

L'influence des variables explicatives est montrée à la figure 5.14 (un exemple de calcul de l'influence des variables est illustré à l'annexe E).

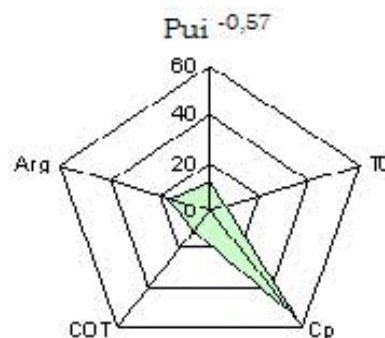


Figure 5.14 : Influence des paramètres sur l'équation de régression, $2^{ème}$ modèle de régression multiple de la dose.

Selon le modèle de prédiction établi, la concentration des particules présente la plus grande influence sur la dose optimale (55,5%) suivi par le taux d'argile (16,3%), la puissance d'agitation lente (12%), le COT (9,9 %) et en dernier lieu la turbidité initiale (6.3%).

II-2.3. Troisième modèle de régression multiple de la dose

Les variables exogènes retenues pour le troisième modèle statistique sont listées dans le tableau 5.9 :

N°	Paramètre	Description
1	P _{ui}	Puissance d'agitation lente [w/m ²]
2	T ₀	Turbidité initiale [NTU]
3	C _p	Concentration en particule [mg/l]
4	K _B /K _A	Constante de cassage/ Constante de formation [s]

Tableau 5.9 : Paramètres d'entrée du 3^{ème} modèle de régression multiple de la dose.

Le meilleur ajustement est mixte. Le modèle de prédiction est comme suit :

$$Dose = 1,5692 + 5,89 \times P_{ui}^{-0,17} - 0,2739 \times T_0 - 280,254 \times [K_B / K_A] + 0,0906 \times C_p$$

Les coefficients de corrélation, de détermination et de détermination ajusté, ainsi que les résultats des tests de Fisher et de Student et leurs niveaux de significativité P et aussi l'erreur de l'estimation sont résumés dans le tableau 5.10 suivant.

	C	Ecart type de C	I.C _{95%} de C	t(32)	p-niveau
Intercepte	1,5692	0,625	0,34 à 2,79	2,512	0,017
P _{ui} ^{-0,17}	5,8900	0,380	5,15 à 6,63	15,499	0,000
T ₀	-0,2741	0,075	-0,42 à -0,13	-3,641	0,000
K _B /K _A	-280,254	78,562	-434,26 à -126,27	-3,567	0,001
C _p	0,0906	0,025	0,042 à 0,139	3,65324	0,000

Tableau 5.10 : Résultats de la régression, 3^{ème} modèle de régression multiple de la dose.

La valeur du coefficient de détermination est de $R^2 = 0,88$, signifiant que 88% de la variabilité de la dose est expliquée par le modèle.

Les résultats des tests de Fischer et de Student justifient l'apport des variables explicatives ; les valeurs de F et t sont élevées ($F > k = 2,668$, k : valeur tabulaire voir annexe B, $|t| > j_{1-\alpha/2} = 2,0369$, j : valeur tabulaire, voir annexe C). Alors l'hypothèse alternative H_1 est acceptée ; les coefficients de régression ne sont pas nuls, par conséquent les variables explicatives : "Pui, T_0 , K_B/K_A , et C_p " présentent une influence réelle simultanée et individuelle sur la variable passive (dose optimale).

La valeur "0" n'est pas, dans les intervalles de confiance de l'intercepte et des coefficients de régression alors l'ordonnée à l'origine n'est pas nulle et les régresseurs apportent un pouvoir explicatif réel à la variable expliquée.

Le tableau 5.11 donne la synthèse des résidus : les valeurs observées, les valeurs prédites, les résidus, les résidus studentisés et la distance de Cook.

	Valeur observée	Valeur prévue	Résidus	Résidus studentisés	Distance de Cook	REQM
Minimum	0,50	0,26	-1,43	-1,97	0,00	0,68
Maximum	7,00	6,31	1,34	1,85	0,41	
Moyenne	2,59	2,59	0,00	0,00	0,05	
Médiane	2,50	2,01	-0,10	-0,14	0,01	

Tableau 5.11 : Synthèse des résidus de la régression, 3^{ème} modèle de régression multiple de la dose [mgAl/l] (sect d'apprentissage)

La valeur moyenne des résidus est nulle, la valeur de REQM est faible (0,68mgAl/l), les résidus sont compris entre -2 et 2, il n'y a pas alors d'observations aberrantes qui peuvent influencer les estimations.

Les valeurs obtenues pour la distance de Cook sont inférieure à 1 ce qui montre que : pas de grande influence de la i^{ème} observation sur l'ajustement.

Le graphe des résidus supprimés en fonction des résidus confirme graphiquement cette observation (voir figure 5.15).

La normalité des résidus est, quant à elle, vérifiée par ajustement des données des résidus à la loi normale (voir figure 5.16).

En ce qui concerne la vérification de la non autocorrélation, il vient que : Pour quatre régresseurs $k=4$, et $n= 37$ (taille de l'échantillon) $dU = 1,723$ (Valeur tabulée voir annexe A). La valeur de test de Durain-Waston $d=2,70 > dU$ ce qui signifie que pas d'autocorrélation d'ordre 1 des résidus.

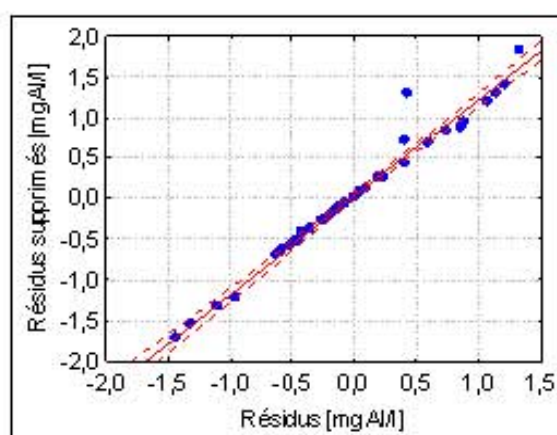


Figure 5.15 : Ajustement des résidus et résidus supprimés, 3^{ème} modèle de régression multiple de la dose, set d'apprentissage, I.C. à 95 %

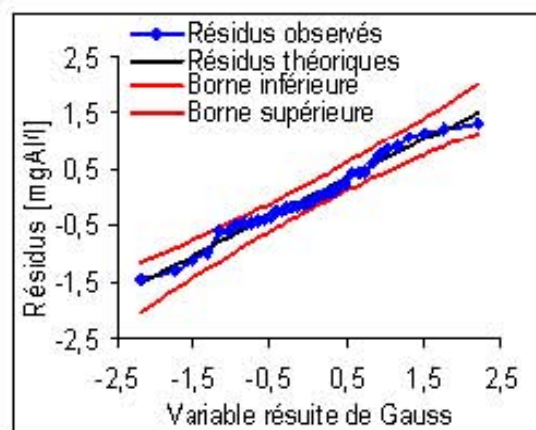


Figure 5.16 : Ajustement à une loi Normale des résidus, 3^{ème} modèle de régression multiple de la dose, set d'apprentissage, I.C. à 95 %

Validation du modèle

Le modèle est validé par un deuxième échantillon (échantillon de validation) le tableau 5.12 résume la synthèse des résidus.

$R = 0,47$ $R^2 = 0,22$					
	Valeur Observée	Valeur Prévue	Résidus	Résidus studentisés	REQM
Minimum	0,5	-2,20	-2,40	-1,29	3,00
Maximum	7	7,68	5,07	2,71	
Moyenne	3,48	1,50	1,98	1,06	
Médiane	2,5	1,10	2,67	1,43	

Tableau 5.12 : Synthèse des résidus de la régression, 3^{ème} modèle de régression multiple de la dose, (set de validation)

La moyenne des résidus est élevée (≈ 2 mgAl/l), la racine carrée de l'erreur quadratique moyenne REQM est aussi élevée (3 mgAl/l), seulement 22% de variance est expliquée par le modèle, cette divergence entre la variance expliquée en apprentissage et en validation revient principalement à la faible similarité des valeurs de K_B/K_A en échantillon d'apprentissage et de validation comme le montre la figure 5.1.

Le résidu studentisé max dépasse la valeur de 2 montrant que quelques observations sont mal expliquées par le modèle à 4 variables.

Test de signification du coefficient de corrélation

Le coefficient de corrélation entre les valeurs observées et les valeurs prévues du jeu de données de validation est faible ($R=0,47$), il est très judicieux de tester sa conformité et sa signification à l'aide du critère statistique t.

$$t = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$$

Ce critère statistique suit une distribution de Student avec n-2 degrés de libertés JAYARAMAN (1999)

Pour n : taille de l'échantillon de validation (n=24)

R : coefficient de corrélation (R=0,47)

La valeur du critère statistique calculée t est de 2,50.

Dans la table de l'annexe C, pour 22 degrés de liberté, au seuil de signification $\alpha = 0,05$ la valeur critique de t est 2,074,

Comme la valeur calculée de t est supérieure à la valeur critique, on conclut que le coefficient de corrélation est significativement différent de 0, indiquant que les valeurs observées de l'échantillon de validation et les valeurs prévues par le modèle sont corrélées de manière significative.

L'homoscédasticité et normalité des résidus

La variabilité des résidus n'augmente pas avec la valeur de la dose prévue comme le montre la figure 5.17 (pas de cône de variance croissante), alors l'homoscédasticité est vérifiée, néanmoins le graphique montre une structure linéaire décroissante se dégageant des résidus, il serait alors bon d'ajouter, dans des cas pareils selon CORNILLON (2007), une variable descriptive au modèle de régression.

La normalité des résidus testée pour l'échantillon de validation par application de l'équation du modèle est montrée vérifiée graphiquement (voir figure 5.18).

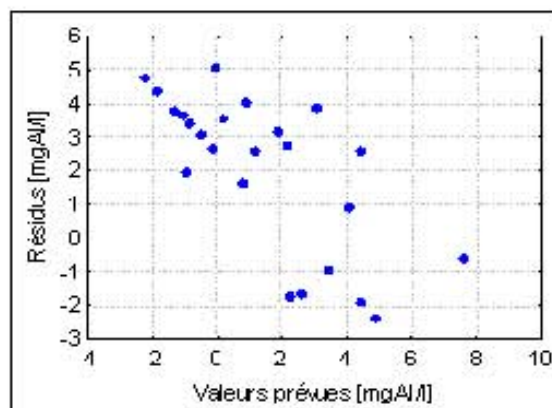


Figure 5.17 : Vérification de l'homoscédasticité par le nuage de points des résidus. 3^{ème} modèle de régression multiple de la dose (set de validation)

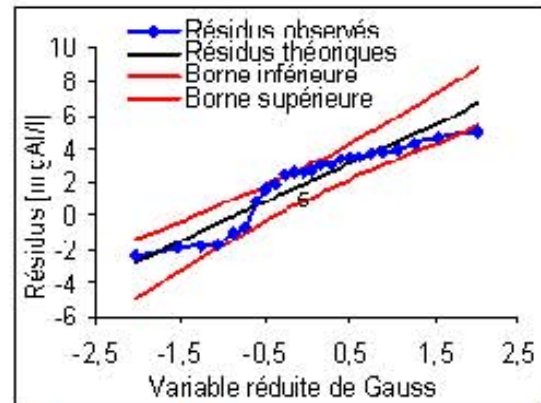


Figure 5.18 : Ajustement à une loi Normale des résidus, 3^{ème} modèle de régression multiple de la dose (set de validation, IC à 95 %)

L'ajustement des valeurs prédites sur les valeurs observées, cas des données d'apprentissage et de validation est illustré à la figure 5.19. L'ajustement est d'autant meilleur que les valeurs se rapprochent de la première bissectrice. Pour le set de validation les points sont dispersés et éloignés de la première bissectrice.

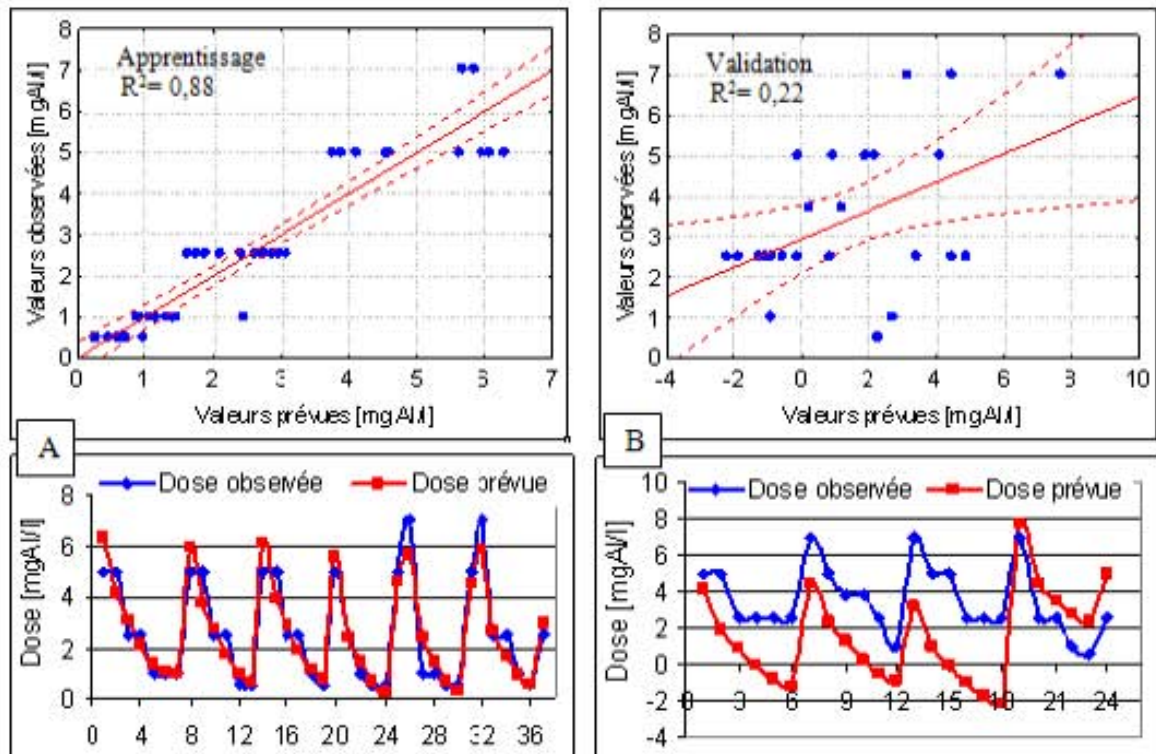


Figure 5.19 : Ajustement des valeurs observées et des valeurs prévues, 3^{ème} modèle de régression multiple de la dose **A** : set d'apprentissage, IC à 95% **B** : set de validation IC à 95%.

Stabilité des coefficients de régression et influence des paramètres sur l'équation du modèle.

Le calcul des valeurs de tolérance donne pour le paramètre "pui^{-0.57}" la valeur 0,97 et pour le "Cp" la valeur 0,1, pour les autres paramètres "T₀" et "K_E/K_A", la tolérance moyenne est inférieure à 0,1 ce qui le signe des colinéarités et ce qui risque l'instabilité des coefficients de régression.

L'influence des variables de régression est montrée à la figure 5.20 (un exemple de calcul de l'influence de variables est illustré à l'annexe E), il apparait selon l'équation du modèle que la concentration de particules "Cp" et la turbidité initiale "T₀" présentent le plus grand effet sur la dose optimale suivi par la puissance "pui^{-0.57}" et en dernière position le rapport "K_E/K_A".

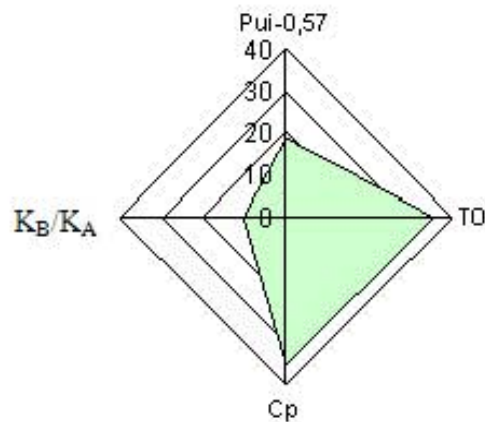


Figure 5.20 : Influence des paramètres sur l'équation de la régression, 3^{ème} modèle de régression multiple de la dose.

Conclusion

Le premier et le deuxième modèle reliant la dose optimale du coagulant aux différents paramètres semblent de qualité d'ajustement acceptable ($R^2=0,89$ en apprentissage et $R^2>0,60$ en validation), alors que le troisième modèle apparaît le moins judicieux en validation ($R^2 =0,88$ en apprentissage et $R^2=0,22$ en validation) cela est due essentiellement à la mauvaise similarité entre les deux jeux de données d'apprentissage et de validation des données du rapport K_B/K_A comme le montre la figure 5.1.

La qualité de prédiction (normalité et non auto-corrélation des résidus, homoscédasticité...etc) apparaît très satisfaisante cependant la grande multicollinéarité existante entre les paramètres régresseurs fait en sorte que les modèles de régression établis encourent un grand risque d'instabilité de l'estimation.

Par ailleurs, les variables retrouvées les plus influents sur la dose optimale selon les trois modèles sont : le diamètre moyen en surface "Dms", le carbone organique total "COT", la concentration des particules "Cp" et la turbidité initiale "T₀".

Afin de remédier au problème d'instabilité de l'estimation et d'améliorer encore la qualité d'ajustement et de prédiction de la dose optimale, nous travaillons dans ce qui suit sur les composantes principales issues des régresseurs, c'est la régression sur composantes principales communément connue par régression régularisée ou orthogonalisée, les problèmes de colinéarité sont alors résolus en supprimant les variables principales de plus faibles variances.

II-2.4. Modèle de régression régularisée

L'analyse en composantes principales permet d'apprécier l'existence de colinéarité entre les variables explicatives, de repérer les individus occupant des positions aberrantes ou simplement suspectés et peut également fournir des variables artificielles orthogonales qui sont les coordonnées des points-observations sur les nouveaux axes comme nouveaux prédicteurs LEBART (1997).

Le principe de la régression sur composantes principales (régression régularisée) revient à remplacer les variables explicatives par ces variables artificielles orthogonales. Cette technique est recommandée lorsque les variables explicatives sont nombreuses ou fortement corrélées entre elles, CONFAIS (2007).

Les problèmes de colinéarité sont alors résolus en supprimant les variables principales de plus faibles variances c'est-à-dire associées aux plus petites valeurs propres. Ou encore en exécutant un algorithme de choix de modèle sur les composantes principales.

L'analyse en composantes principales des résultats optimaux (sans la variable passive "dose") est illustrée dans le tableau 5.13 suivant, il vient qu'à six composantes la variance expliquée dépasse 96%.

	<i>Valeur propre</i>	<i>Variance totale %</i>	<i>Valeur propre cumulée</i>	<i>Variance cumulée %</i>
1	17,09	43,83	17,09	43,83
2	11,02	28,26	28,11	72,09
3	3,79	9,69	31,9	81,78
4	2,9	7,43	34,8	89,21
5	1,84	4,72	36,64	93,93
6	1,07	2,74	37,71	96,67
7	0,4	1,04	38,11	97,71
8	0,35	0,89	38,46	98,6
9	0,2	0,5	38,66	99,1
10	0,15	0,39	38,81	99,49
11	0,08	0,2	38,89	99,69

Tableau 5.13 : Résultats de détermination des valeurs propres et des variances expliquées, modèle de régression régularisée de la dose.

La figure 5.21 montre l'importance du choix des composantes principales sur la performance du modèle ; à la figure 5.21 **A**, les composantes sont prises selon l'importance de leurs variances et à la figure 5.21 **B**, les composantes sont retenues selon la procédure de régression "Forward Regression". Il apparaît clairement que les composantes principales de forte variance ne sont pas nécessairement les plus explicatives comme l'indique SAPORTA (1990) et il apparaît aussi qu'à avec un grand nombre de composantes principales le R^2 est élevé.

Par ailleurs, le calcul de la régression avec les composantes ordonnées en fonction de leurs corrélations avec la variable à expliquer (dose optimale) selon les recommandations de SAPORTA (1990) n'a pas permis de bonne performance.

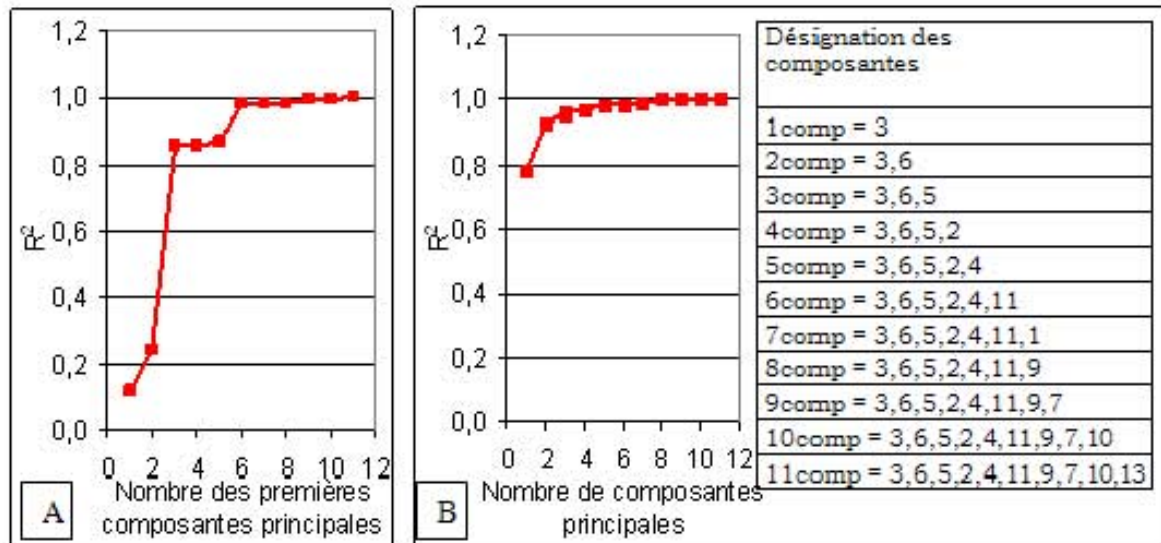


Figure 5.21 : R² en fonction du nombre de composantes principales, modèle de régression régularisée de la dose. **A** : les composantes sont prises selon l'importance de variance, **B** : les composantes sont retenues selon la procédure de régression "Forward Regression"

Il apparaît clairement qu'avec six composantes principales le modèle est performant, les coefficients de détermination R² valent 0,98. Alors nous retenons pour la prédiction de la dose optimale les six premières composantes principales (la colinéarité est faible et le R² est élevé). Le meilleur ajustement est linéaire. L'équation de prédiction est la suivante:

$$Dose = 3,1349 - 0,0416 \times comp_1 - 0,077 \times comp_2 + 1,0214 \times comp_3 + 0,0927 \times comp_4 - 0,2516 \times comp_5 + 0,7277 \times comp_6$$

Les résultats de la régression sont donnés dans le tableau 5.14, dans ce tableau figure :

- Le coefficient de corrélation, le coefficient de détermination et le coefficient de détermination ajusté.
- L'erreur standard de l'estimation.
- La valeur de l'ordonnée à l'origine et les coefficients de régression du modèle avec leurs intervalles de confiance.
- Les résultats des tests statistiques de Fischer et de Student.

$$R = 0,991 \quad R^2 = 0,982 \quad R^2_{aj} = 0,979$$

$$F(6,30) = 304,28 \quad p < 0,00$$

Erreur type de l'estimation 0,29

	C	Ecart type de C	I.C _{95%} de C	t(30)	p-niveau
Intercepte	3,1349	0,048	3,04 à 3,33	65,565	0,000
Comp ₁	-0,0416	0,012	-0,065 à -0,018	-3,446	0,002
Comp ₂	-0,0770	0,015	-0,107 à -0,047	-5,055	0,000
Comp ₃	1,0214	0,028	0,967 à 1,075	37,104	0,000
Comp ₄	0,0906	0,025	0,042 à 0,139	3,653	0,000
Comp ₅	-0,2516	0,035	-0,321 à -0,182	-7,118	0,000
Comp ₆	0,7277	0,050	0,62 à 0,827	14,318	0,000

Tableau 5.14 : Résultats de la régression régularisée de la dose.

La corrélation linéaire entre les valeurs expérimentales et prévues est très bonne. La valeur du coefficient de détermination $R^2=0,982$ cela signifie que 98,2 % de la variabilité de la dose est expliquée par le modèle.

Pour le test de Fisher, la valeur de F est grande ($F > k=2,421$, k : valeur tabulaire, voir annexe B) alors l'effet des variables explicatives est important et on devra rejeter donc l'hypothèse nulle H_0 qui est l'éventuelle nullité simultanée des coefficients de régression.

Pour le test de Student la probabilité est jugée très faible la valeur de t est grande ($|t| > j_{1-\alpha/2} = 2,0423$, j : valeur tabulaire, voir annexe C) alors on rejette l'hypothèse nulle H_0 qui est l'éventuelle non influence de la variable explicative, alors les composantes principales ont chacune une influence réelle sur la variable à expliquer (dose optimale).

Le calcul de l'intervalle de confiance confirme ce résultat ; la valeur "0" n'est pas dans les intervalles de confiance, alors les variables explicatives apportent un pouvoir explicatif réel à la dose optimale.

L'analyse des résidus permet de détecter les différences significatives entre les valeurs observées et les valeurs prédites ce qui aide donc à connaître les points mal prédits et les faiblesses du modèle en terme de moyenne ou de variance.

Cependant, cela ne nous renseigne pas sur les variations des estimateurs des paramètres dues à la suppression d'une observation et donc à la robustesse du modèle, pour cela le calcul de la distance de Cook est alors nécessaire à la détection des points influents.

Le tableau 5.15 ci-dessous donne pour les données d'apprentissage la synthèse de la régression : les valeurs observées, les valeurs prédites, les résidus, les résidus studentisés, la valeur de REQM et la distance de Cook.

	Valeur observée	Valeur prévue	Résidus	Résidus studentisés	Distance de Cook	REQM
Minimum	0,50	-0,13	-0,56	-1,93	0,00	0,00
Maximum	7,00	7,14	0,66	2,26	0,43	
Moyenne	2,83	2,83	0,00	0,00	0,03	
Médiane	2,50	2,61	-0,01	-0,04	0,01	

Tableau 5.15 : Synthèse des résidus de la régression régularisée de la dose [mgAl/l] (set d'apprentissage)

La valeur moyenne des résidus, et la valeur de REQM sont nulles, les résidus studentisés sont faibles et sont bornés dans l'intervalle [-2 à 2,3] cela montre qu'il y a peu d'incertitude pour l'estimation des paramètres. Les valeurs de distances de Cook sont inférieures à 1 ce qui indique l'absence de points trop influents.

La figure 5.22 des résidus supprimés en fonction des résidus confirme graphiquement cette observation. Le modèle à priori ajuste bien les données.

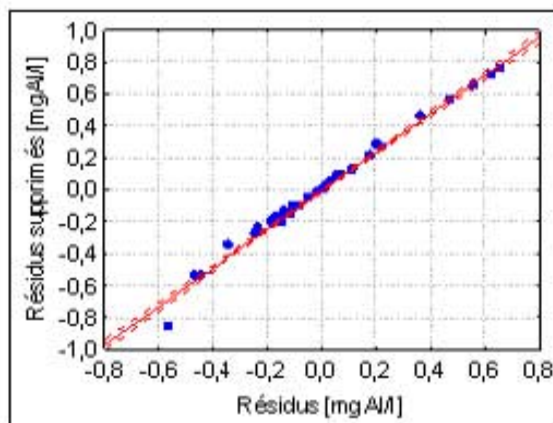


Figure 5.22 : Ajustement des résidus et résidus supprimés, modèle de régression régularisée de la dose (set d'apprentissage, IC à 95 %)

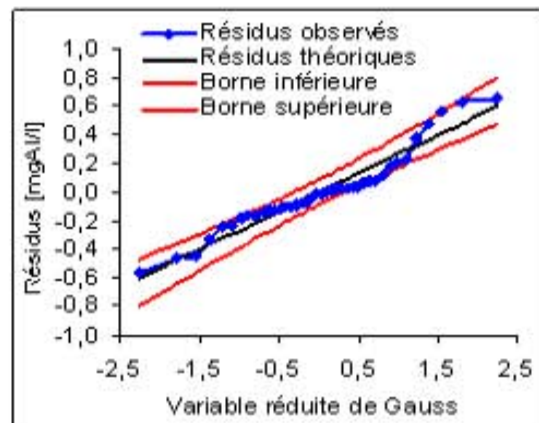


Figure 5.23 : Ajustement à une loi Normale des résidus, modèle de régression régularisée de la dose (set d'apprentissage, IC à 95 %)

La normalité des résidus

La normalité des résidus de l'échantillon d'apprentissage est, quant à elle, évaluée graphiquement à la figure 5.23, les valeurs expérimentales (observées) sont proches des valeurs théoriques et sont incluses dans l'enveloppe de l'intervalle de confiance à 95 %.

La non autocorrélation

La statistique de Durbin-Watson est utile pour évaluer la présence ou l'absence d'autocorrélation entre les résidus (c'est à dire, si les résidus

d'observations adjacentes sont ou non corrélés, indiquant que les observations du fichier de données ne sont pas indépendantes).

La valeur calculée de la statistique de Durbin-Watson est de $d = 2,27$ cette valeur est supérieure à la valeur de "dU" $dU = 1,87$ cela signifie que pas d'autocorrélation d'ordre 1 des résidus.

(Voir annexe A, pour 6 descripteurs et $n=37$ taille de l'échantillon $dU = 1,87$).

Validation du modèle

La validation du modèle statistique développé est réalisée sur l'échantillon de validation composé de données de résultats de floculation des particules de la vase et de mélange II. L'équation de la régression est appliquée sur ces données expérimentales de validation ; des valeurs de doses optimales sont ainsi obtenues et comparées avec celles expérimentales. L'ajustement est vérifié par le calcul des résidus, des résidus studentisés et de REQM, le tableau 5.16 résume les résultats.

$R = 0,975 \quad R^2 = 0,95$					
	Valeur Observée	Valeur Prévue	Résidus	Résidus studentisés	REQM
Minimum	1,00	0,67	-1,52	-3,715	0,41
Maximum	7,0	6,92	0,37	0,91	
Moyenne	2,96	3,05	-0,09	-0,21	
Médiane	2,50	2,49	-0,03	-0,07	

Tableau 5.16 : Synthèse des résidus de la régression régularisée de la dose [mgAl/l](set de validation)

Les valeurs moyennes de résidus, des résidus studentisés et de REQM sont très faibles ce qui signifie que l'ajustement est bon. La valeur minimale des résidus studentisés est inférieure à "-2" ce qui indique la présence de point aberrant mal prédit par le modèle, minimisant la qualité de l'ajustement graphique des résidus à la loi normale (voir figure 5.24).

L'homoscédasticité

Le nuage de points des résidus calculés sur l'échantillon de validation, illustré à la figure 5.25 montre que les résidus se répartissent régulièrement en haut et en bas du zéro. L'homoscédasticité est vérifiée. On note l'existence d'un point extrême en bas du graphique, ce qui pourra réduire la qualité de validation du modèle.

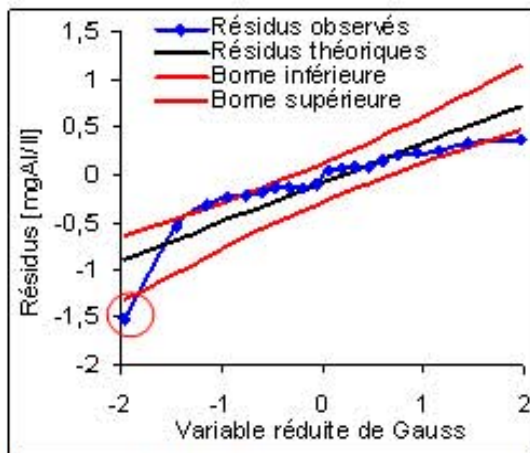


Figure 5.24 : Ajustement à une loi Normale des résidus, modèle de régression régularisée de la dose (set de validation, IC à 95 %)

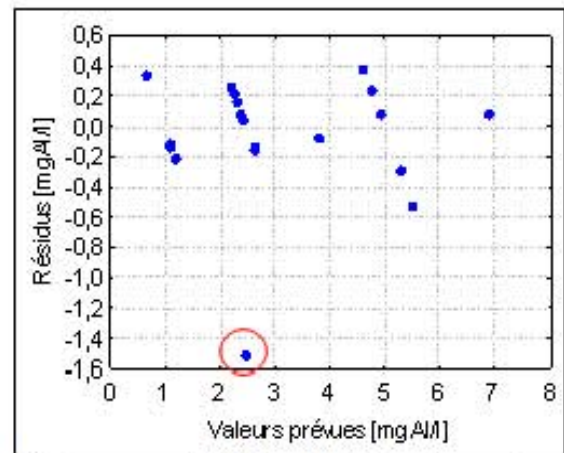


Figure 5.25 : Vérification de l'homoscédasticité par le nuage de points des résidus, modèle de régression régularisée de la dose (set de validation)

Ajustement des données

L'ajustement des valeurs de doses observées et prévues par le modèle est présenté sur la figure 5.26. L'ajustement est d'autant meilleur que les valeurs se rapprochent de la première bissectrice. La valeur du coefficient de détermination pour le jeu de données d'apprentissage est de $R^2 = 0.98$, alors que pour le jeu de validation le point extrême (se trouvant éloigné de la première bissectrice) a réduit sensiblement la performance de validation du modèle et a permis un $R^2 = 0.95$.

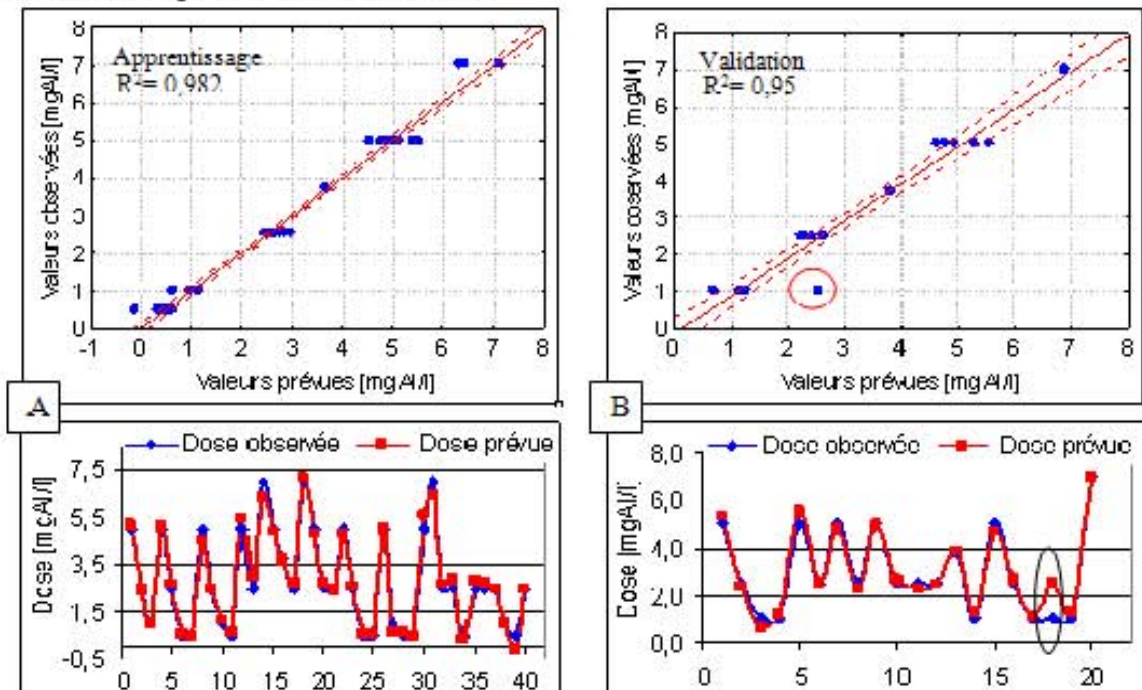


Figure 5.26 : Ajustement des valeurs observées et des valeurs prévues, modèle de régression régularisée de la dose **A** : set d'apprentissage, IC à 95% **B** : set de validation IC à 95%.

Stabilité des coefficients de régression et influence des paramètres sur l'équation du modèle.

Les variables de régression et qui sont les composantes principales sont orthogonales et présentent de faibles corrélations linéaires $R < 0,08$ (il s'agit du cas idéal). Le calcul des valeurs de tolérance donne pour tous les régresseurs des valeurs variant de 0,77 à 0,96, ces valeurs sont largement supérieures à 0,1 donc pas de grande colinéarité entre les régresseurs et aucun risque d'instabilité de l'estimation n'est encouru.

La solution obtenue présente de meilleures qualités prédictives mais, les coefficients de la régression s'appliquent aux composantes principales, un calcul complémentaire nécessaire est fait afin d'évaluer et d'interpréter les effets de chacune des variables initiales.

Chaque composante principale dans le modèle est alors réécrite en fonction des variables initiales, tout calcul fait de l'influence des paramètres initiaux est résumé dans le tableau 5.17 suivant :

<i>Paramètre</i>	<i>Importance</i>	<i>Paramètre</i>	<i>Importance</i>	<i>Paramètre</i>	<i>Importance</i>
Illite	40,47	MO	1,82	Mg	0,28
Cp	10,12	Dmv	1,50	Dolomite	0,20
Puissance	7,30	AL ₂ O ₃	1,49	Ssp	0,18
Chlorite	6,09	T ₀	1,18	K ₂ O	0,14
Argile	5,53	D(0,5)	1,12	Uniformité	0,09
SiO ₂ +AL ₂ O ₃	4,50	Quartz	0,526	Na ₂ O	0,03
Montmorillonite	4,21	Fe ₂ O ₃	0,50	Al _r	0,02
Kaolinite	3,65	COT	0,50	MnO	0,0067
SiO ₂	2,91	CAO	0,42	K _B /K _A	0,0004
Tf	2,58	CM	0,42	K _A	7,5.10 ⁻⁶
Calcite	1,91	Dms	0,32	K _B	2,1.10 ⁻⁷

Tableau 5.17 : Influence des variables explicatives sur l'équation de la régression, modèle de régression régularisée de la dose.

Selon cette méthode de calcul, illustrée à l'annexe E, Il apparait bien que le paramètre "ILL" est le plus influant, suivi par le "Cp" ensuite la puissance d'agitation lente "Pui". La composante argileuse prend une position avancée dans les effets des caractéristiques du solide dispersée sur la dose optimale, les diamètres, la matière organique et la matière minérale prennent une position moyenne, alors que les paramètres K_A et K_B montrent à leurs tours les plus faibles influences sur la dose optimale.

Conclusion

L'intérêt de la régression orthogonalisée n'est évidemment pas d'obtenir ainsi, par une voie détournée, des résultats identiques à ceux que donnent, en général sans difficultés, les méthodes classiques de la régression multiple. Mais son intérêt apparait lorsque les variables explicatives possèdent une matrice de corrélation singulière ou quasi singulière. En régression régularisée le problème de colinéarité et d'instabilité de l'estimation est alors résolu. Le résultat de la régression régularisée montre que le modèle retenu est robuste et présente de meilleures qualités prédictives.

III. MODELISATION NON PARAMETRIQUE DE LA DOSE OPTIMALE

III.1. MODELE NEURONAL

Un réseau de neurones peut être considéré comme un modèle mathématique de traitement composé de plusieurs éléments de calcul non linéaire (neurone) opérant en parallèle et connectés entre eux par des poids. Il réalise simplement une somme pondérée des entrées, ajoute un seuil à cette somme et fait passer le résultat par une fonction de transfert pour obtenir sa sortie. HIMM-ELBLAU 2000 résume les différents applications des réseaux de neurones en engineering chimiques en :

- Détection des défauts
- Prédiction de qualité
- Rectification de la base de données
- Modélisation et control de processus

III.1.1. Choix des paramètres d'entrée et découpage de la base de données

Les variables exogènes retenues comme variables d'entrée pour le modèle statistique neuronal sont le résultat d'analyse en composantes principales et d'analyse de classification hiérarchique, c'est ce que à procéder LAMIRINI (2005) et HECTOR (2006) à la détermination des paramètres d'entrée des modèles de régression neuronale. Les variables d'entrée retenues pour la régression neuronale sont listées dans le tableau 5.18 ci-dessous.

N°	Paramètre	Description
1	Pui	Puissance d'agitation lente [w/m³]
2	T _o	Turbidité initiale [NTU]
3	Cp	Concentration des particules [mg/l]
4	COT	Carbone organique total [%]
5	CM	Carbone minéral [%]
6	Dms	Diamètre moyen en surface [µm]
7	Ssp	Surface spécifique [m²/g]
8	Kao	Kaolinite [%]
9	ILL	Illite [%]
10	Mte	Montmorillonte [%]
11	Arg	Argile [%]

Tableau 5.18 : Paramètres d'entrée du modèle de régression neuronale de la dose

La mise en œuvre de réseau de neurones comporte deux phases : une phase d'apprentissage et une phase de validation. La phase d'apprentissage n'est autre que l'identification des poids du réseau telle que les sorties calculées par le réseau soient le plus proche possible de celles observées. (Valeurs cibles). Nous attribuons par échantillonnage aléatoire, 2/3 de la base de données à l'apprentissage et 1/3 à la validation. Les choix faits en termes de construction, d'architecture de réseau de neurones et de mode d'évaluation du modèle sont récapitulés dans le tableau 2.2 du chapitre 2.

La figure 5.27 illustre les box-plots des données d'apprentissage et celui des données de validation, cette figure permet d'examiner la représentativité et l'homogénéité des partitions. Sur cette figure le fond, la point central et le plafond de chacune des boites représentent respectivement "la moyenne-2Erreur type", "la moyenne" et "la moyenne+2Erreur type" de la distribution des données, les barres verticales représentent quant à elles les minima et les maxima.

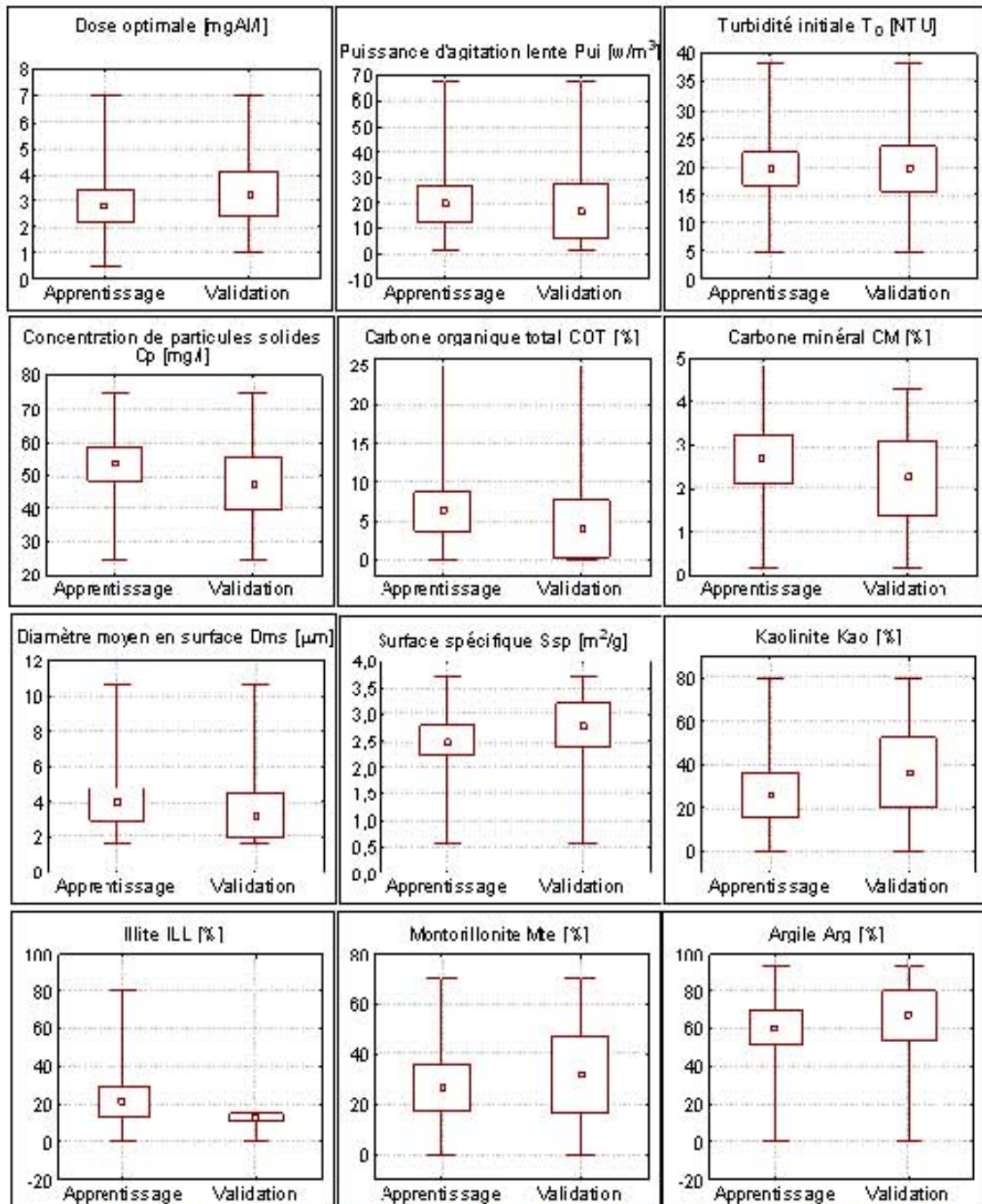


Figure 5.27 : Présentation box-plot « boîtes à moustache » des données d'apprentissage et de validation (test), Modélisation non paramétrique de la dose (Echantillonnage aléatoire) □ Moyenne □ Moyenne ± 2Erreur type T Min-Max

Cette figure montre qu'effectivement les box-plots des données d'apprentissage et ceux des données de validation sont très comparables, la disparité est similaire pour les deux jeux de données alors les données destinées à l'apprentissage sont statistiquement représentatives de celles destinées à la validation.

III.1.2. Développement du modèle neuronal

L'utilisation d'une régression neuronale passe par le choix des paramètres d'entrée mais aussi par l'optimisation du réseau de neurones lui-même.

Il n'existe aucune méthode optimale pour déterminer à priori la structure interne d'un réseau de neurones. Elle ne peut être déterminée que de façon heuristique : on s'efforce en général à avoir beaucoup moins de paramètres à estimer que d'exemples dans la base d'apprentissage. Il a été montré que le perceptron multicouche composé d'une seule couche cachée avec pour fonction d'activation une fonction sigmoïde et d'une couche de sortie linéaire était un approximateur universel.

L'optimisation à la fois du type de fonction non linéaire et de la géométrie du réseau de neurones (nombre de neurones dans la couche cachée) consiste à tester plusieurs réseaux ayant des architectures différentes jusqu'à atteindre la meilleure performance.

En bibliographie nous mentionnons que pour le contrôle automatique du procédé de coagulation en traitement d'eau potable VALENTIN (2000) utilise un réseau de six entrées, seize neurones non linéaires dans la couche cachée et une sortie linéaire, l'architecture est alors sous la forme 6/16/1. Alors que HECTOR (2006) utilise, pour la supervision et le diagnostic des procédés de production d'eau potable, un modèle neuronal de configuration 4/20/1 ; quatre neurones d'entrée, vingt neurones dans la seule couche cachée et un neurone dans la couche de sortie.

L'architecture du réseau adopté par LAMIRINI (2005) en vue de prédire le dosage de coagulant est 5/18/1 ; elle consiste en cinq variables d'entrée, une seule couche cachée avec 18 neurones sigmoïdaux et un seul neurone linéaire formant la couche de sortie.

Les travaux de MAIER (2004) portant sur la prédiction de la dose optimale de l'alun et des paramètres de qualité des eaux traitées, ont été couronnés par trois modèles dont les architectures sont comme suit :

- 1^{er} modèle pour la prédiction des paramètres de qualité de l'eau traitée ; son architecture est comme suit : sept neurones d'entrées, une couche cachée de huit neurones et trois paramètres de sorties (la turbidité, la couleur, et l'absorbance à 254 nm, 7/8/0/3.

- 2^{ème} modèle pour la prédiction du pH et de l'aluminium résiduel, il consiste en dix entrées, une seule couche cachée de huit neurones et de 2 paramètres de sortie 10/8/0/2.
- 3^{ème} modèle pour la prédiction de la dose optimale de l'alun ; son architecture est comme suit : neuf entrées, deux couches cachées de huit neurones dans chacune d'elle et une seule sortie (qui est la dose optimale) 9/8/8/1.

Parmi les différentes configurations de réseau de neurones testées HADDAM (2008) a retenu, pour la modélisation de la dose du coagulant, le modèle 6/13/1 ; six variables formant la couche d'entrée, treize neurones de l'unique couche cachée et une sortie représentant la dose de coagulant.

L'architecture du réseau pour la prédiction du dosage optimum du coagulant selon ROBENSON (2009) est 11/27/9/1 ; onze paramètres d'entrées, deux couches cachées respectivement de vingt-sept et neuf neurones et une sortie.

MIRSEPASSI (2004) teste un premier modèle de 17 entrées, dix neurones dans la couche cachée et deux paramètres de sortie (dosages de l'alun et du polymère) 17/10/2, l'analyse de performance a permis d'optimiser l'architecture du modèle, en neuf variables d'entrées deux couches cachées respectivement de vingt-cinq et cinq neurones et deux sorties, 9/25/5/2.

Dans notre cas d'étude, pour déterminer l'architecture du modèle, on a augmenté progressivement le nombre de neurones dans la couche cachée de 1 à 23,

Chaque réseau construit a été entraîné avec les fonctions sigmoïdes tangente hyperbolique et logistique, à noter que pour chaque configuration neuronale 5000 valeurs de jeu du poids initiaux ont été utilisées et c'est le modèle qui présente le meilleur résultat qu'on sélectionne. Enfin parmi les différentes configurations testées, nous avons retenu celles produisant la plus faible racine carrée de l'erreur quadratique moyenne REQM sur le set de validation en aveugle.

La figure 5.28 montre la variation de REQM en fonction du nombre de neurones dans la couche cachée sur le jeu de validation respectivement avec les deux fonctions d'activation sigmoïdes tangente hyperbolique et logistique. On constate de ces représentations que c'est la structure à 8 neurones pour la fonction logistique et à 12 neurones pour celle de tangente hyperbolique qui permettent d'obtenir la valeur minimale de REQM par la suite sur des structures neuronales plus complexes la REQM ne s'améliore plus.

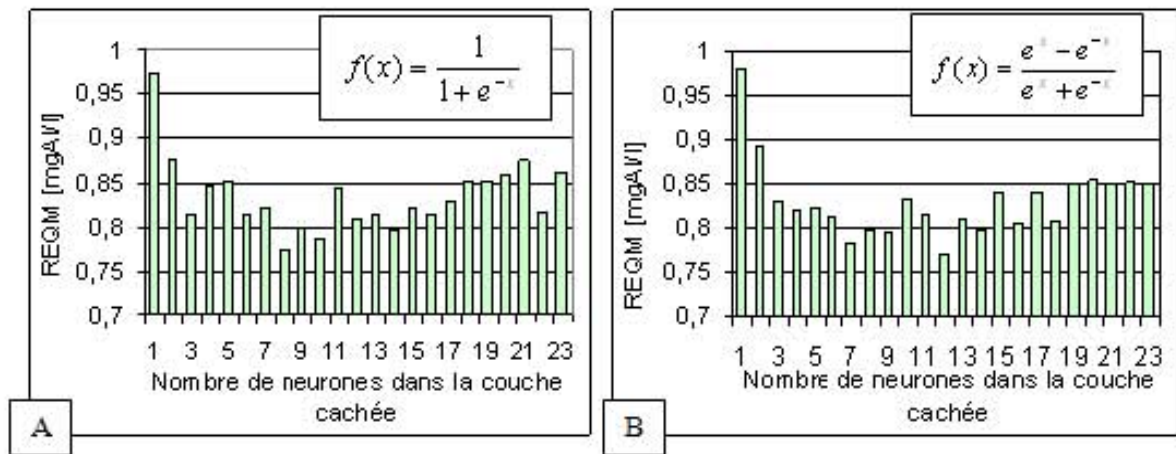


Figure 5.28: REQM du jeu de validation en fonction du nombre de neurones dans la couche cachée calculé avec : **A** la fonction logistique **B**, la fonction tangente hyperbolique, modèle de régression neuronale de la dose.

Les caractéristiques des meilleurs modèles obtenus pour les deux fonctions d'activation sigmoïdes logistique et tangente hyperbolique sont récapitulées dans le tableau 5.19

Modèle PMC	Fonction de transfert	Apprentissage			Validation		
		Résidus	REQM	R ²	Résidus	REQM	R ²
11/8/1	logistique	0,006	0,560	0,922	0,122	0,773	0,852
11/12/1	Tangent hyp.	0,037	0,530	0,929	0,110	0,773	0,850

Tableau 5.19 : Résultats de la régression neuronale de la dose [mgAl/l]

Le modèle retenu est le modèle le moins complexe et qui présente les meilleures performances. Il vient alors de prendre le modèle 11-8-1 muni de la fonction sigmoïde logistique comme fonction d'activation et de transfert expliquant respectivement 92 % et 85 % la corrélation de la dose observée à la dose prévue pour l'échantillon d'identification et celui de validation, l'architecture dudit modèle est présentée à la figure 5.29

Les valeurs moyennes des résidus et la valeur de REQM, listées dans le tableau 5.20, sont très faibles cela montre le bon ajustement des données d'apprentissage par le modèle. Les valeurs des résidus studentisés sont bornées dans l'intervalle [-2,1 à 2,65] montrant la présence de quelques points mal prédits par le modèle mais sans grande influence sur l'ajustement.

Nous observons sur le graphique de la figure 5.30 que la normalité semble bien respectée, tous les points étant autour de la première bissectrice et sont incluses dans l'intervalle de confiance de 95%.

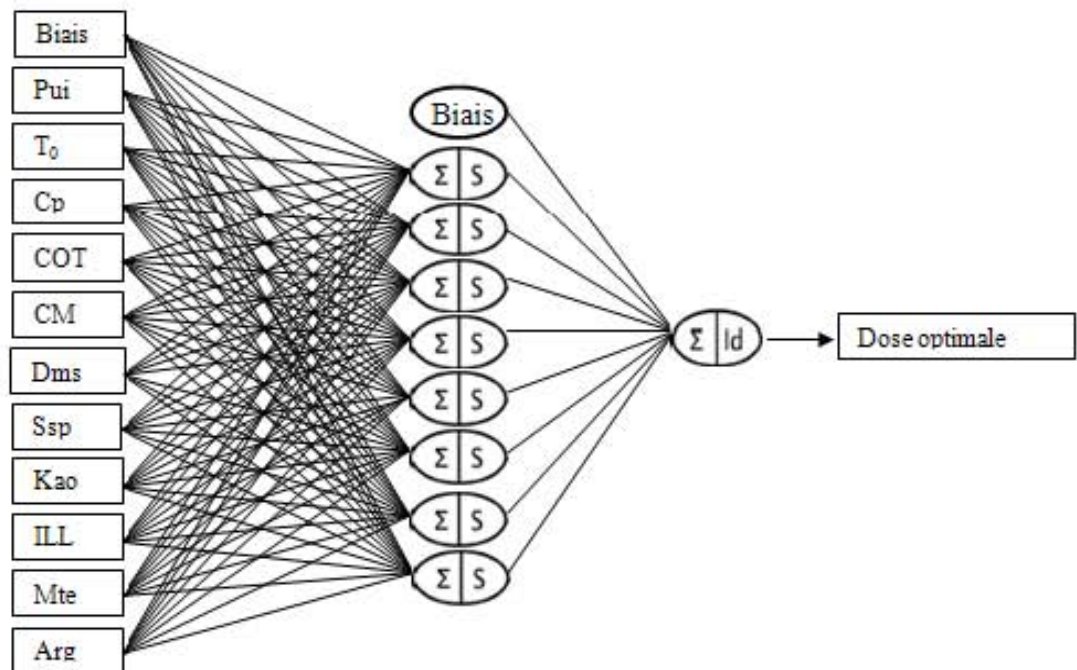


Figure 5.29 : Architecture du perceptron multicouche 11/8/1 pour la prédiction de la dose

Valeur	Minimale	Maximale	Moyenne	Médiane
Observée	0,5	7	2,83	2,5
Prévue	-0,021	6,59	2,84	2,44
Résidus	-1,16	1,46	0,006	-0,063
Résidus studentisés	-2,09	2,62	0,01	-0,11
REQM	0,560			

Tableau 5.20 : Synthèse des résidus de la régression neuronale de la dose [mgAl/l] (set d'apprentissage)

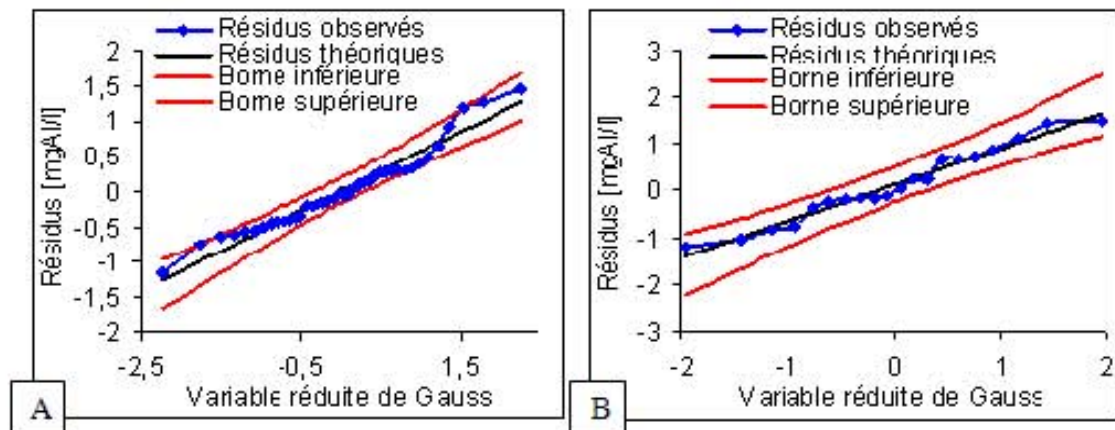


Figure 5.30 : Ajustement à une loi Normale des résidus, modèle de régression neuronale de la dose [A] : set d'apprentissage, IC à 95 % [B] : set de validation, IC à 95 %)

La synthèse des résidus de la régression sur le jeu de validation est résumée dans le tableau 5.21, les valeurs moyennes et médianes des résidus, et la valeur de REQM sont faibles, les résidus studentisés sont bornées dans l'intervalle [-2,2 à 2,66], ce qui est l'indice d'une validation justifiée du modèle.

Valeur	Minimale	Maximale	Moyenne	Médiane
Observée	1	7	3,19	2,5
prévue	0,82	6,83	3,31	2,96
Résidus	-1,21	1,47	0,122	-0,03
Résidus studentisés	-2,17	2,65	0,22	-0,06
REQM	0,773			

Tableau 5.21 : Synthèse des résidus de la régression neuronale de la dose [mgAl/l] (set de validation)

L'homoscédasticité est vérifiée par le nuage de points des résidus calculés sur l'échantillon de validation, illustrée dans la figure 5.31 ; Les résidus sont régulièrement répartis autour de la valeur de "0".

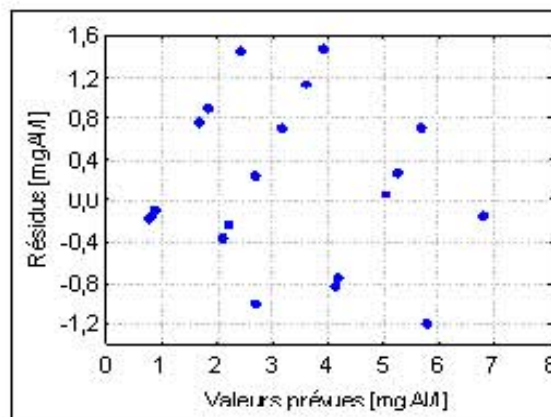


Figure 5.31 : Vérification de l'homoscédasticité par le nuage de points des résidus, modèle de régression neuronale de la dose (set de validation)

L'ajustement des valeurs observées et prévues par le modèle est présenté à la figure 5.32 ; pour l'échantillon d'apprentissage 92 % de la variabilité de la dose est expliquée par le modèle, alors que pour l'échantillon de validation 85 % de la variabilité de la dose est validée par le modèle.

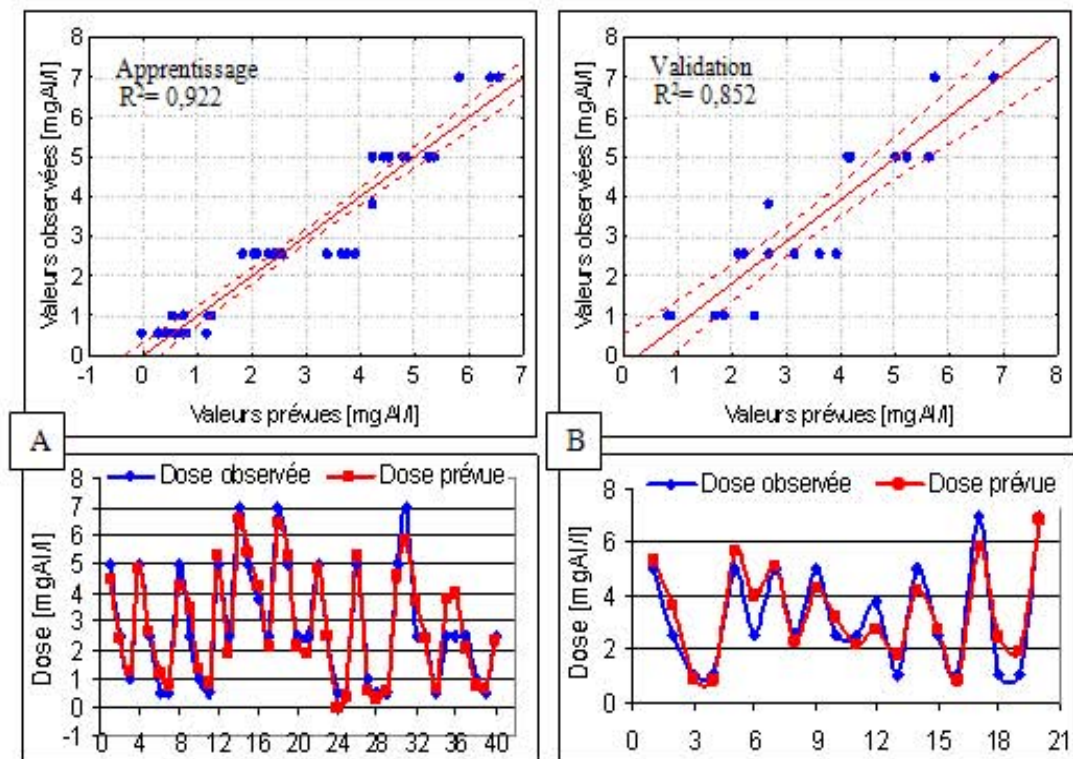


Figure 5.32 : Ajustement des valeurs observées et des valeurs prévues, modèle de régression neuronale de la dose. **A** : set d'apprentissage, IC à 95%, **B** : set de validation, IC à 95%,

Analyse de l'importance relative des variables d'entrée

L'importance de chaque variable d'entrée doit être déterminée de manière à confirmer l'influence des paramètres sélectionnés en entrée sur l'évolution de la dose optimale. La méthodologie de détermination de l'importance relative des variables d'entrée est présentée à l'annexe F, et le résultat est illustré par la figure 5.33 ci dessous.

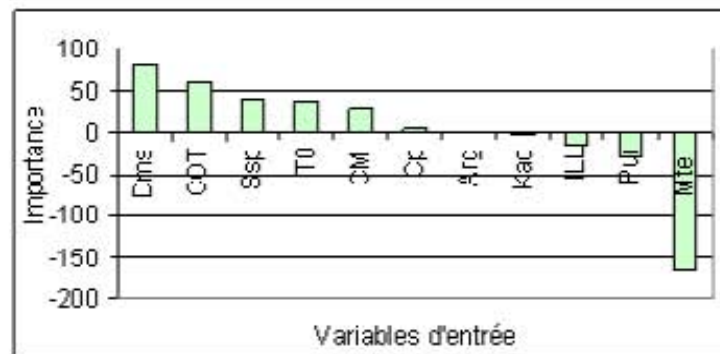


Figure 5.33 : Importance relative des variables d'entrée, modèle de régression neuronale de la dose.

La figure met en évidence trois groupes de variables distinctes, le premier groupe est constitué des paramètres : "Mte, Pui et ILL "dont l'influence sur la sortie, selon cette méthode, est négative avec une influence majeure de la montmorillonite; la dose optimale est faible pour des valeurs élevées des paramètres de ce groupe.

Le deuxième groupe contient les variables dont l'effet est positif ; une augmentation des valeurs de ces paramètres "Dms, COT, Ssp, T₀ et CM" nécessite un accroissement de la dose optimale. Le troisième groupe est composé de "Cp, Arg et Kao" dont l'importance est très faible.

III-2. MODELE NEURONAL BASE SUR LES COMPOSANTES PRINCIPALES

Pour la modélisation par régression neuronale basée sur les composantes principales les entrées du modèle neuronal simple sont remplacées par les leurs composantes principales.

Un essai préliminaire permet de visualiser l'importance du nombre des entrées sur la performance du modèle, la figure 5.34 montre qu'à partir de la troisième composante principale (les composantes sont prises selon l'importance de la variance expliquée), les performances du modèle sont très bonnes (le coefficient de détermination est élevé) ce qui montre que le choix du nombre de composantes principales est déterminant.

Nous retenons, pour raison de comparaison avec les autres modèles élaborés, les composantes principales dont les valeurs propres sont supérieures à 1 (voir tableau 5.14), il vient alors de prendre 6 composantes principales, la variance expliquée est plus de 96 %.

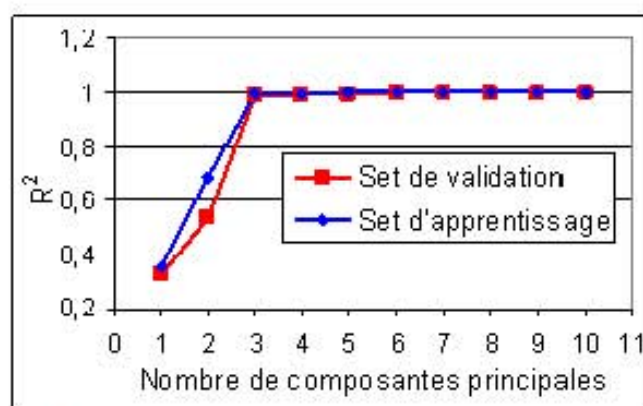


Figure 5.34 : R² en fonction du nombre de composantes principales dans la couche d'entrée, modèle de régression neuronale de la dose.

Pour déterminer le nombre de neurones de la couche cachée, on a augmenté progressivement le nombre de neurones dans cette couche de 1 à 13, chaque réseau construit a été entraîné avec la fonction tangente hyperbolique et la

fonction logistique, à noter que pour chaque configuration neuronale 5000 valeurs de jeu du poids initiaux ont été utilisées et c'est le modèle qui présente le meilleur résultat qui est sélectionné.

Enfin parmi les différentes configurations testées, nous avons retenu celles produisant la plus faible REQM sur le set de validation en aveugle.

La figure 5.35 montre la variation de REQM en fonction du nombre de neurones dans la couche cachée sur le jeu de validation et cela pour les deux fonctions sigmoïdes testées.

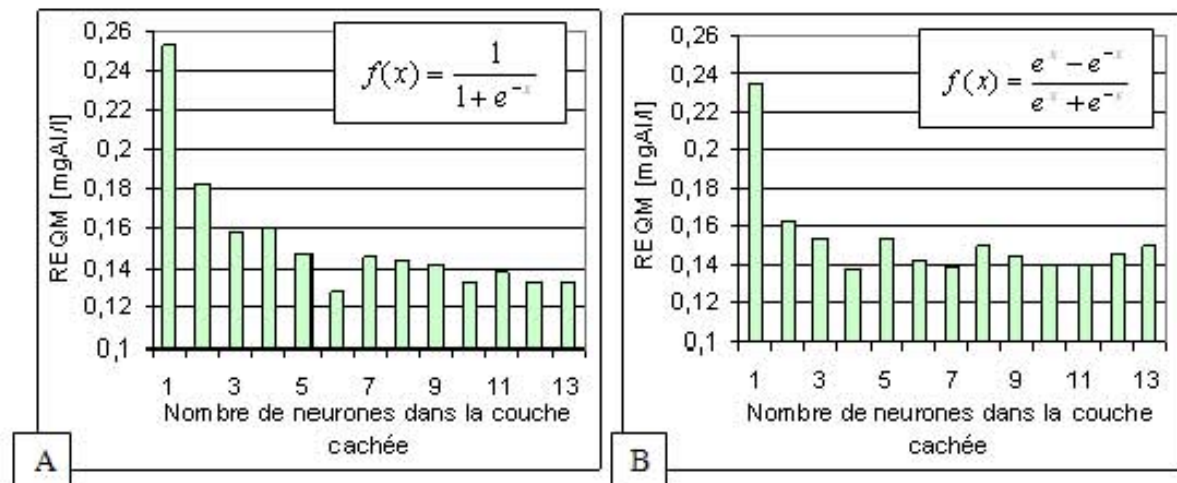


Figure 5.35: REQM du jeu de validation en fonction du nombre de neurones dans la couche cachée calculé avec : **A** la fonction logistique **B**, la fonction tangente hyperbolique, modèle de régression neuronale sur composantes principales de la dose.

On constate que c'est la structure à 1 neurone qui permet d'obtenir une REQM minimale, en utilisant la fonction d'activation tangente hyperbolique et à 6 neurones pour la fonction d'activation logistique. Par la suite sur des structures neuronales plus complexes la valeur de REQM ne s'améliore plus. Les performances de ces deux configurations neuronales sont résumées dans le tableau 5.22.

Modèle PMC	Fonction de transfert	Apprentissage			Validation		
		Résidus	REQM	R ²	Résidus	REQM	R ²
6/6/1	logistique	-0,002	0,110	0,997	0,006	0,128	0,995
6/4/1	Tangent hyp.	-0,001	0,105	0,995	-0,009	0,138	0,994

Tableau 5.22 : Modèles neuronaux sur Composantes Principales, pour la prédiction de la dose.

La configuration retenue est 6/6/1 ; six entrées qui sont les composantes principales, six neurones dans la couche cachée et une sortie (la dose optimale), l'architecture du modèle est présentée ci-dessous à la figure 5.36.

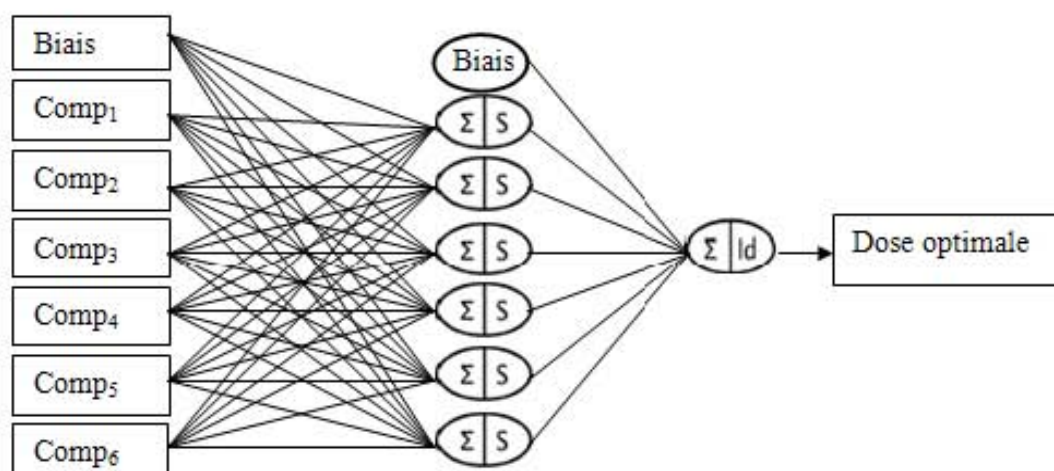


Figure 5.36 : Architecture du perceptron multicouche 6/6/1 pour la prédiction de la dose

La synthèse de régression pour le l'échantillon d'apprentissage est résumée dans le tableau 5.23 et celle de l'échantillon de validation dans le tableau 5.24.

<i>Valeur</i>	<i>Minimale</i>	<i>Maximale</i>	<i>Moyenne</i>	<i>Médiane</i>
Observée	0,5	7	2,84	2,5
prévue	0,32	7	2,84	2,49
Résidus	-0,25	0,26	-0,002	0,00
Résidus studentisés	-2,22	2,31	0,02	-0,02
REQM	0,110			

Tableau 5.23: Synthèse des résidus de la régression neuronale sur composantes principales de la dose [mgAl/l] (set d'apprentissage)

<i>Valeur</i>	<i>Minimale</i>	<i>Maximale</i>	<i>Moyenne</i>	<i>Médiane</i>
Observée	1	7	3,22	2,5
prévue	0,99	7,06	3,22	2,57
Résidus	-0,34	0,18	0,006	0,00
Résidus studentisés	-3,00	1,59	0,07	0,00
REQM	0,128			

Tableau 5.24: Synthèse des résidus de la régression neuronale sur composantes principales de la dose [mgAl/l] (set de validation)

Il apparaît nettement que la valeur de REQM et les valeurs moyennes et médianes des résidus sont très faibles. Les résidus studentisés sont bornés dans l'intervalle [-3 à 2,31] indiquant la présence de quelques points mal prédis par le modèle mais sans grande influence sur la qualité d'ajustement et de prédiction, en effet, les coefficients de détermination R^2 pour le jeu de données d'apprentissage et le jeu de données de validation sont respectivement $R^2 = 0.997$ et $R^2=0.995$, ces grandes valeurs montrent le pouvoir explicatif élevé du modèle neuronal basé sur les composantes principales.

La normalité des résidus est vérifiée graphiquement par le traçage de la droite de Henry, les valeurs théoriques sont proches des valeurs observées comme le montre la figure 5.37 et sont incluses dans l'enveloppe de l'intervalle de confiance de 95%.

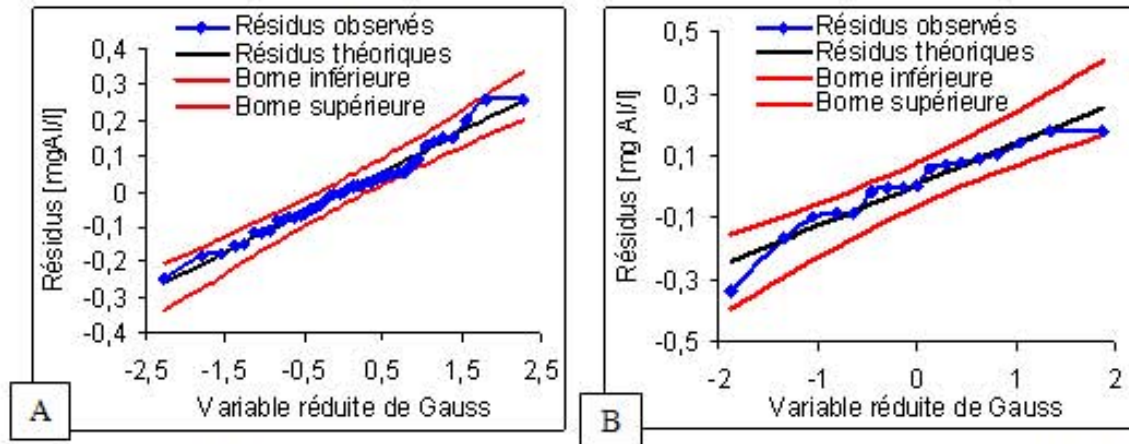


Figure 5.37 : Ajustement à une loi Normale des résidus, modèle de régression neuronale sur composantes principales de la dose, **A** set d'apprentissage, IC à 95 % **B** set de validation, IC à 95 %)

L'homoscédasticité

La représentation graphique à la figure 5.38 des résidus suivant les doses prédites par le modèle montrent que les points sont repartis de façon égale en haut et en bas du zéro et ne suivent pas une tendance particulière. On peut donc dire alors que les résidus sont indépendants, l'homoscédasticité est vérifiée et que la prédiction est de bonne qualité.

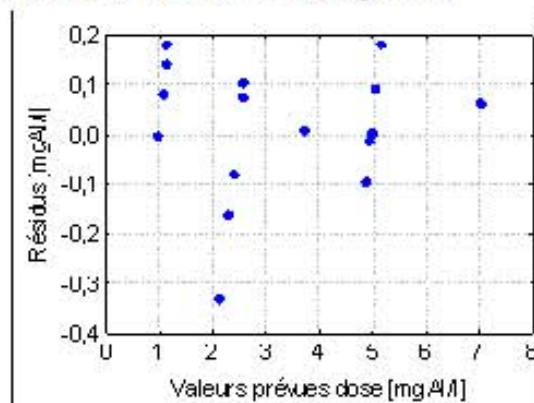


Figure 5.38 : Vérification de l'homoscédasticité par le nuage de points des résidus, modèle de régression neuronale sur composantes principales (set de validation)

La figure 5.39 croise les valeurs observées avec les valeurs prévues. Les points sont alignés avec la première bissectrice, l'ajustement est très bon. Il en ressort que le modèle de régression neuronale basé sur les composantes principales de la dose optimale est moins complexe que le modèle neuronal simple et présente de meilleures performances de prédiction.

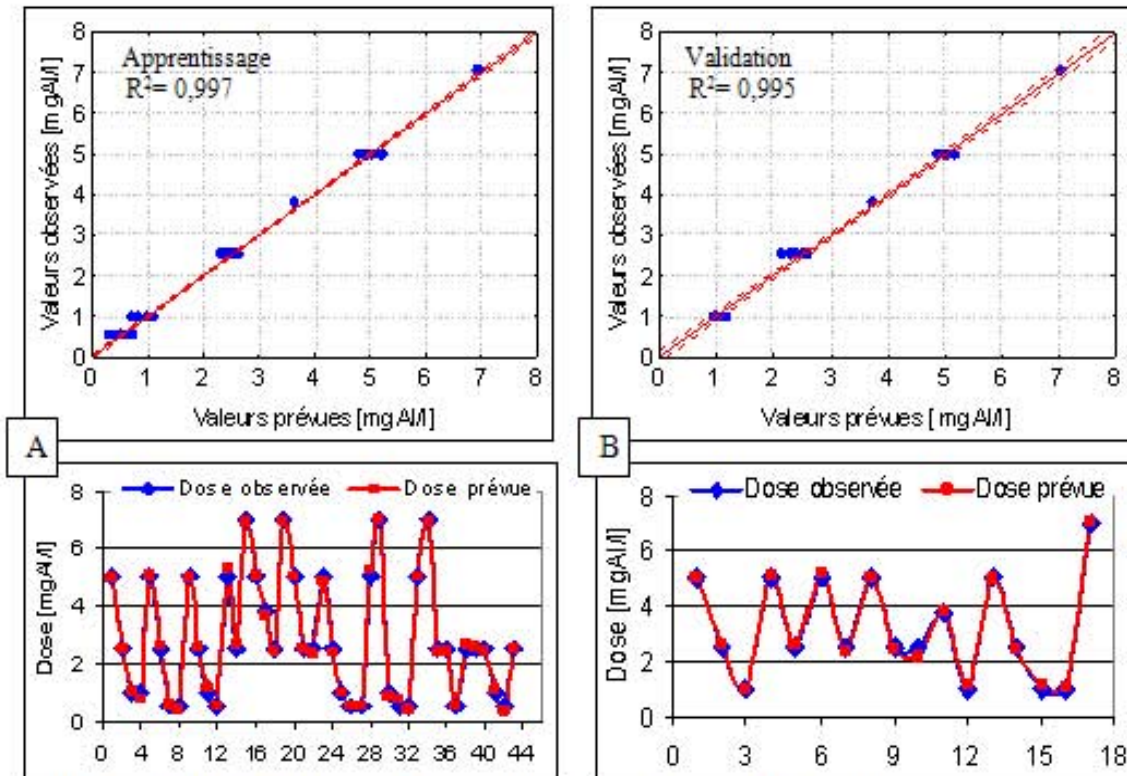


Figure 5.39 : Ajustement des valeurs observées et des valeurs prévues, modèle de régression neuronale sur composantes principales de la dose.

A : set d'apprentissage, IC à 95%, **B** : set de validation, IC à 95%.

IV. CONCLUSION

Les modèles élaborés de régression multiple de la dose optimale présentent des performances acceptables ($R^2=0,89$) en apprentissage pour les trois modèles et plus de 0,60 en validation seulement pour le premier et le deuxième modèle, Alors que pour le troisième modèle où a été insérer le rapport K_B/K_A comme paramètre régresseur, la validation de ce modèle n'a pas été vraiment judicieuse (R^2 est faible) cela est du essentiellement à la mauvaise similarité des valeurs du rapport K_D/K_A entre le jeu de données d'apprentissage et celui de validation.

Un grand risque d'instabilité de l'estimation de ces trois modèles de régression multiple est relevé, cela revient à la grande multi colinéarité entre les variables explicatives. La régression sur composantes principales (régression régularisée) a permis de remédier à ce problème de multi-colinéarité en travaillant avec des variables orthogonales qui résument l'information de toutes les variables initiales en un nombre réduit de descripteurs (les composantes principales), le coefficient de détermination obtenu est de $R^2=0,98$ en apprentissage et $R^2=0,95$ en validation.

La régression non paramétrique par réseau de neurones présente des performances très intéressantes mais un peu faible par rapport à la régression régularisée ($R^2 = 0,92$ en apprentissage et $R^2= 0,85$ en validation), cela explique la sensibilité de choix des variables descripteurs de la couche d'entrée du modèle PMC.

Cependant la régression neuronale basée sur les composantes principales rassemble les avantages des deux techniques ; les variables descriptives de la couche d'entrée sont bien résumés en un nombre limité de variable grâce à l'analyse en composantes principales et l'adaptation du réseau de neurones à la non linéarité des relations entre la variable à expliquer et les variables explicatives.

On se rapportant à l'annexe D qui donne les valeurs au-delà desquelles un coefficient de corrélation peut être déclaré significatif (pour un nombre donné d'observations au seuil de signification voulu), il s'avère que pour les différents modèles établis, les coefficients de corrélation entre les valeurs observées (expérimentales) et les valeurs prévues par les modèles sont tous significatifs. La figure 5.40 suivante récapitule les valeurs des coefficients de détermination R^2 et des valeurs de la racine carrée de l'erreur quadratique moyenne REQM pour les différents modèles de régression de la dose optimale.

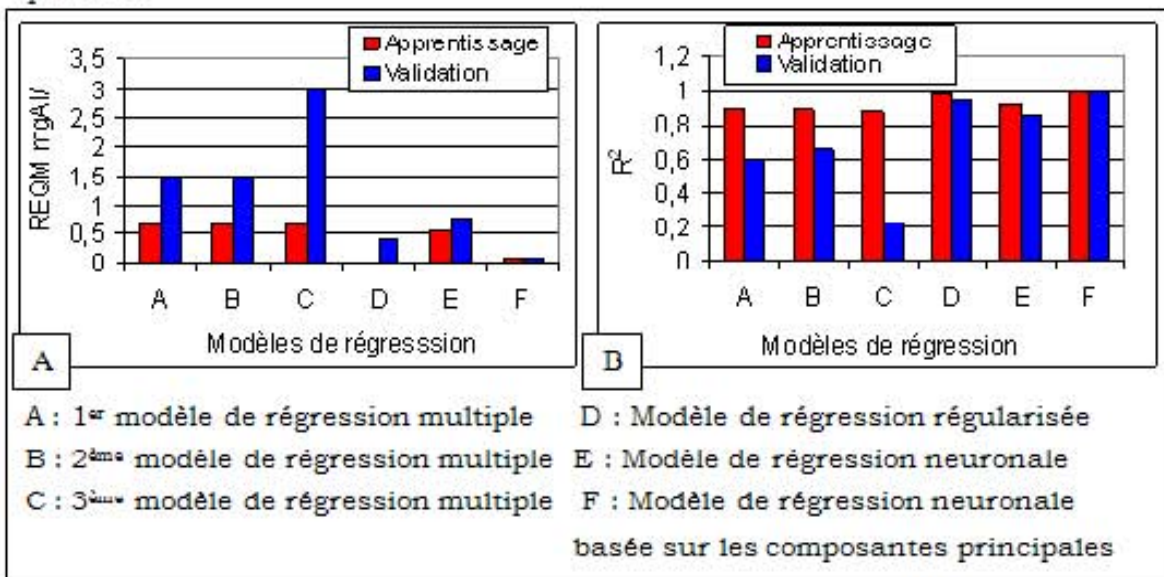


Figure 5.40 : Variations de : **A** La racine carrée de l'erreur quadratique moyenne REQM, **B** Le coefficient de détermination R^2 , pour les différents modèles de régression de la dose optimale.

Il apparait clairement que le modèle neuronal basé sur les composantes principales est le plus performant, il présente les valeurs les plus élevées du coefficient de détermination R^2 et les valeurs les plus faibles de REQM, il est suivi par le modèle de régression régularisée, puis le modèle neuronal et en dernier lieu les modèles de régression multiple.

Les différences relevées du calcul des influences des paramètres descripteurs sur l'estimation de la dose optimale, pour les différents modèle testés : régression multiple, régression régularisée et régression neuronale sont tributaires de la complexité des modèles et de la méthode du calcul des influences et peuvent être à priori synonyme d'absence des influents prépondérants.

V. MODELISATION PARAMETRIQUE DES CONSTANTES CINETIQUES

K_A ET K_B

Le processus de floculation est considéré comme un état d'équilibre entre deux réactions simultanées, d'une part l'agrégation de particules primaires et de petits floccs pour en former de plus gros, et d'autre part l'érosion ou la fragmentation des floccs.

Plusieurs modèles mathématiques (voir chapitre I § II-1) décrivent la cinétique de la floculation en faisant entrer deux constantes, la première de formation K_A et la deuxième de cassage K_B . Ces constantes cinétiques K_A et K_B sont révélatrices de l'état de la structure du flocc ; plus K_A est élevée plus le flocc se forme rapidement, plus K_B est faible plus le flocc est résistant. La connaissance de ces constantes permet, par application des modèles cinétiques de floculation, la détermination de l'efficacité de floculation.

L'objectif de cette section est l'établissement de modèles statistiques de prédiction des valeurs de ces constantes cinétiques K_A et K_B à partir des caractéristiques de solides dispersés.

V.1. Choix et préparation des données

Les données utilisées dans cette section sont tirées du fichier brut, il s'agit des caractéristiques du solide dispersé et des valeurs des constantes de formation et de cassage des suspensions de kaolin, de la vase, de la tourbe, de la boue et des deux mélange I et II. La matrice de données est composée de 6 individus statistiques et 26 variables.

L'analyse exploratoire multidimensionnelle (voir chapitre 4) a permis le choix des variables explicatives qui peuvent être injectées dans les modèles de prédiction des constantes K_A et K_B . Les variables retenues pour la modélisation sont listées dans le tableau 5.25 ci-dessous.

N°	Paramètre	Description
1	T _o	Turbidité initiale [NTU]
2	COT	Carbone organique total [%]
3	Dms	Diamètre moyen en surface [µm]
4	Arg	Taux d'argile [%]
5	Ssp	Surface spécifique [m ² /g]
6	CM	Carbone minéral [%]
7	Mte	Montmorillonite [%]
8	Kao	Kaolinite [%]
9	ILL	Illite [%]

Tableau 5.25 : Paramètres retenues pour la modélisation des constantes de formation et de cassage de flocc

V.2. Modèles de régression multiple

Les modèles recherchés vont permettre de prévoir les valeurs des constantes cinétiques de floculation, la constante de formation K_A et la constante de cassage K_B , en fonction des caractéristiques des suspensions, en conditions optimales de floculation selon le schéma de la figure 5.41 suivante :

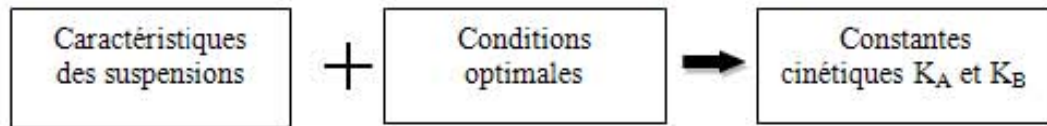


Figure 5.41 : Schéma de régression multiple des constantes cinétiques K_A et K_B

V.2.1. Modèle de régression de la constante de cassage K_B

Les variables exogènes retenues par recherche pas à pas, par la méthode ascendante "Forward Régression" sont listées dans le tableau 5.26 suivant :

N°	Paramètre	Description
1	T_0	Turbidité initiale [NTU]
2	COT	Carbone organique total [%]
3	Dms	Diamètre moyen en surface [μm]
4	Arg	Taux d'argile [%]

Tableau 5.26 : Paramètres d'entrée du modèle de régression de la constante de cassage K_B

Le meilleur ajustement est mixte. L'équation du modèle est la suivante:

$$K_B = -1,70 \cdot 10^{-5} + 1,14 \cdot 10^{-6} \times e^{0,2286Dms} + 4,57 \cdot 10^{-7} \times T_0 + 7,70 \cdot 10^{-8} \times COT + 4,88 \cdot 10^{-7} \times Arg$$

La synthèse de la régression est résumée dans le tableau 5.27. Ce modèle permet un R^2 très élevé (0,9995) c'est à dire que la variation de la constante de cassage K_B est expliquée à 99.95 % par les paramètres régresseurs : diamètre moyen en surface "Dms", la turbidité initiale " T_0 ", le carbone organique total "COT" et le taux d'argile dans le solide dispersé "Arg". C'est donc un très bon ajustement des données.

$R^2=0,9995$ $R^2_{aj} = 0,997$ $F(4,1) = 9109,1$ $p < 0,00786$ Erreur type de l'estimation $< 0,01$						
	C	Ecart-type de C	IC _{95%} de C	t(1)	p-niveau	
Intercepte	$-1,70 \cdot 10^{-5}$	$4,63 \cdot 10^{-7}$	$-2,29 \cdot 10^{-5}$ à $-1,11 \cdot 10^{-5}$	-36,7	0,017	
$e^{0,2286 \times Dms}$	$1,14 \cdot 10^{-6}$	$1,68 \cdot 10^{-8}$	$9,27 \cdot 10^{-7}$ à $1,35 \cdot 10^{-6}$	67,6	0,009	
T_0	$4,57 \cdot 10^{-7}$	$1,23 \cdot 10^{-8}$	$3,01 \cdot 10^{-7}$ à $6,13 \cdot 10^{-7}$	37,2	0,017	
COT	$7,70 \cdot 10^{-8}$	$2,67 \cdot 10^{-9}$	$4,31 \cdot 10^{-8}$ à $1,11 \cdot 10^{-7}$	28,8	0,022	
Arg	$4,88 \cdot 10^{-7}$	$1,82 \cdot 10^{-8}$	$2,57 \cdot 10^{-7}$ à $7,19 \cdot 10^{-7}$	26,8	0,024	

Tableau 5.27 : Résultats de la régression multiple de la constante de cassage K_B .

La valeur F de test de Fischer et la probabilité p induite sont utilisées à tester la relation entre la variable dépendante K_B et l'ensemble des variables indépendantes explicatives : T_0 , $e^{0,2286 \times Dms}$, COT et Arg , le résultat de ce test montre que la valeur de F est grande ($F > k=224.58$, k valeur tabulaire, voir annexe B), la probabilité est jugée faible, alors l'ensemble des variables ont une influence simultanée réelle sur la valeur de K_B .

La valeur t de test de Student et la probabilité p induite sont utilisées pour tester l'hypothèse d'une ordonnée à l'origine nulle et la nullité des autres coefficients de régression. Le résultat du test montre que la valeur de la variable t est grande ($|t| > k=12,7$, k valeur tabulaire, voir annexe C), la probabilité (p -niveau) est jugée faible alors les variables T_0 , $e^{0,2286 \times Dms}$, COT et Arg ont, chacune seule, une influence réelle sur la constante cinétique K_B . Le calcul des intervalles de confiance confirme l'apport explicatif des régresseurs ; la valeur "0" n'est pas dans les intervalles de confiance de l'intercepte et des coefficients de régression C .

Le tableau 5.28, ci-dessous, donne la synthèse de la régression : la valeur moyenne des résidus est très faible $5,684.10^{-14}$. Les résidus studentisés sont faibles et sont bornés dans l'intervalle $[-1 \text{ à } 1]$ ce qui montre l'absence de points aberrants, la valeur du REQM est faible $1,7.10^{-8}$, l'ajustement est alors très bon.

Le tableau 5.28, ci-dessous, donne la synthèse de la régression : la valeur moyenne des résidus est très faible $5,684.10^{-14}$. Les résidus studentisés sont faibles et sont bornés dans l'intervalle $[-1 \text{ à } 1]$ ce qui montre l'absence de points aberrants, la valeur du REQM est faible $1,7.10^{-8}$, l'ajustement est alors très bon.

	Valeur observée	Valeur prévue	Résidus	Résidus studentisés	REQM
Minimum	8.10^{-7}	$8,06.10^{-7}$	$-0,32.10^{-7}$	-0,78	$1,7.10^{-8}$
Maximum	$1,04.10^{-6}$	$1,04.10^{-6}$	$0,002.10^{-6}$	0,42	
Moyenne	$3,36.10^{-6}$	$3,36.10^{-6}$	$5,684.10^{-14}$	0,00	
Médiane	$2,07.10^{-6}$	$2,08.10^{-6}$	$0,001.10^{-6}$	0,04	

Tableau 5.28 : Synthèse des résidus de la régression de la constante de cassage K_B [s]

La normalité des résidus

La normalité des résidus est vérifiée graphiquement à la figure 5.42 ; pour un seuil de signification $\alpha=0,05$, les points observés sont très proches des points théoriques et sont incluses dans l'enveloppe de l'intervalle de confiance de 95%, L'ajustement des valeurs observées et valeurs prévues est illustré sur la figure 5.43, les points sont bien ajustés sur la première bissectrice.

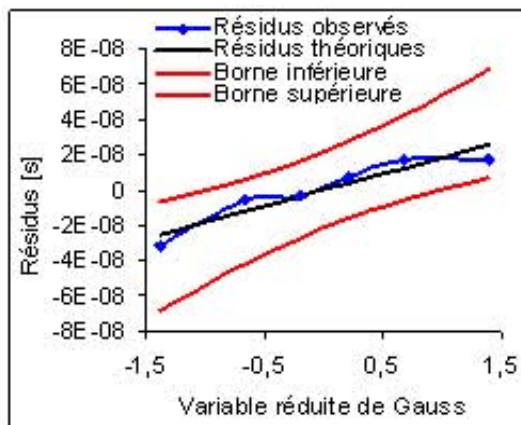


Figure 5.42 : Ajustement à une loi Normale des résidus, modèle de régression multiple de la constante de cassage K_B , IC à 95 %.

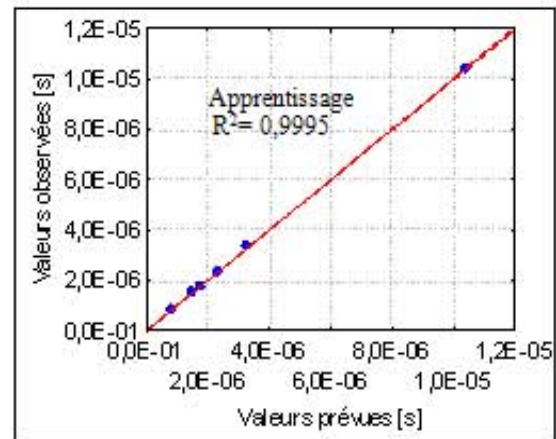


Figure 5.43 : Ajustement des valeurs observées et des valeurs prévues, modèle de régression multiple de la constante de cassage K_B , IC à 95 %.

Stabilité des coefficients et influence des paramètres sur l'équation de régression multiple de la constante de cassage K_B .

Les coefficients de corrélation entre les paramètres descripteurs du modèle sont élevés $R > 0,80$, les valeurs de tolérance calculées sont inférieure à 0,1. Cela est le signe de multi-colinéarité entre les régresseurs, il en ressort que les coefficients de régression du modèle encourent un grand risque d'instabilité.

Les résultats de l'étude de l'influence des paramètres sur l'équation du modèle sont présentés à la figure 5.44, la méthodologie de calcul est illustrée à l'annexe E, il en ressort que la turbidité initiale "T0" présente la plus grande influence (36,27 %) sur la constante de cassage K_B , les autres paramètres "COT, Arg et " montrent des influences presque de même importance respectivement 21,88% 19,14% et 22,72%.

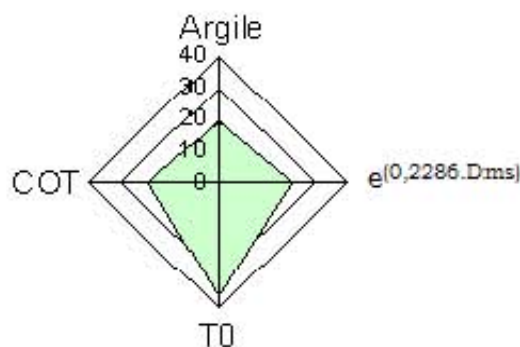


Figure 5.44 : Influence des paramètres sur l'équation de régression multiple de la constante de cassage K_B .

V.2.2. Modèle de régression régularisée de la constante de cassage K_B

La régression régularisée est peut être considérée comme un mélange de régression multiple et d'analyse en composantes principales. Son principe est en effet de rechercher l'équation de régression multiple relative à une variable dépendante y et à p variables explicatives x_1, \dots, X_p par l'intermédiaire des composantes principales de ces variables explicatives. L'analyse des composantes principales des caractéristiques de solide dispersé en nombre de 26 a donné les résultats résumés dans le tableau 5.29 ci-dessous :

	<i>Valeur propre</i>	<i>Variance Totale %</i>	<i>Valeur propre cumulée</i>	<i>Variance Cumulée %</i>
1	14,76751	56,79812	14,76751	56,7981
2	7,69729	29,60497	22,46480	86,4031
3	3,52335	13,55136	25,98816	99,9545
4	0,01182	0,04548	25,99998	99,9999
5	0,00002	0,00006	26,00000	100,0000

Tableau 5.29 : Résultats de détermination des valeurs propres et des variances expliquées, modèle de régression des constantes cinétiques

Il apparait que les 26 variables (valeur propre cumulée) peuvent être remplacées par 5 variables indépendantes orthogonales (composantes principales) liées linéairement aux variables initiales.

La régression multiple permet un ajustement linéaire de la constante K_B , L'équation de régression est la suivante:

$$K_B = 3,36.10^{-4} + 8,22.10^{-7} \times comp_1 - 1,70.10^{-4} \times comp_4 - 2,86.10^{-4} \times comp_5$$

Ce modèle prédit la constante de formation K_B en fonction de la première, la quatrième et la cinquième composante principale, l'absence des composantes 2 et 3 de l'équation du modèle confirme les notations de SAPORTA (2007) ; les composantes de grandes variances ne sont pas nécessairement les plus explicatives.

Les résultats de la régression sont donnés dans le tableau 5.30 dans ce tableau figure

- Le coefficient de corrélation, de détermination et le coefficient de détermination ajusté
- Les tests de Fisher et de Student et leurs niveaux de significativité P
- L'erreur standard de l'estimation.
- La valeur de l'ordonnée et des coefficients de régression du modèle.

R= 0,9985 R²= 0,9969 R²aj = 0,9923
 F(3,2)=215,55 p < 0.00462
 Erreur de l'estimation 3,12.10⁻⁷

	Coefficients C	Ecart-type de C	IC _{95%} de C	t(2)	p-niveau
Intercepte	3,36.10 ⁻⁶	1,27.10 ⁻⁷	2,81.10 ⁻⁶ à 3,91.10 ⁻⁶	26,4	0,001
Comp ₁	8,22.10 ⁻⁷	3,50.10 ⁻⁸	6,71.10 ⁻⁷ à 9,73.10 ⁻⁷	23,5	0,002
Comp ₄	-1,70.10 ⁻⁶	2,97.10 ⁻⁷	-2,98.10 ⁻⁶ à -4,22.10 ⁻⁷	-5,73	0,030
Comp ₅	-2,86.10 ⁻⁶	3,66.10 ⁻⁷	-4,43.10 ⁻⁶ à -1,29.10 ⁻⁶	-7,81	0,016

Tableau 5.30 : Résultats de la régression régularisée de la constante de cassage K_B.

R²=0,9969, ce qui signifie que l'adéquation du modèle aux données d'apprentissage est excellente. Toutes les p-niveau associées aux estimateurs des paramètres sont <0.05. Les valeurs de t du test de Student (|t|>j_{1α/2} = 4,30) et les valeurs de la statistique F de Fisher (F >19,16) sont grandes, on rejette alors l'hypothèse de nullité des coefficients de régression.

Il en résulte cependant que l'apport des variables explicatives (composantes principales) au modèle de prédiction de la constante de cassage K_B est justifié. Les intervalles de confiance calculées confirment cet apport explicatif des composantes principales. (La valeur "zéro" n'est pas dans les intervalles de confiance).

Le tableau 5.31, ci-dessous donne la synthèse de la régression : les valeurs observées, les valeurs prédites, les résidus, les résidus studentisés et le REQM.

	Valeur Observée	Valeur Prévue	Résidus	Résidus studentisés	REQM
Minimum	8,00.10 ⁻⁷	4,4210 ⁻⁷	-1,6810 ⁻⁷	-0,54	0,18.10 ⁻⁶
Maximum	1,0410 ⁻⁵	1,0310 ⁻⁵	3,5810 ⁻⁷	1,15	
Moyenne	3,3610 ⁻⁶	3,3610 ⁻⁶	0,00	-0,00	
Médiane	2,0710 ⁻⁶	2,1910 ⁻⁶	-5,03.10 ⁻⁸	-0,16	

Tableau 5.31 : Synthèse des résidus de la régression régularisée de la constante de cassage K_B [s]

On vérifie aisément que la valeur de REQM est faible, les résidus et les résidus studentisés sont de moyenne nulle, le résidu studentisé max ne dépasse pas en valeur absolue 2, signe d'absence de points extrêmes qui appauvriraient la qualité d'ajustement du modèle, alors l'ajustement des données d'apprentissage par le modèle est très bon.

Quant à la normalité des résidus, elle est évaluée par ajustement graphique à la figure 5.45 : Les valeurs expérimentales (observées) sont proches des valeurs théoriques et sont incluses dans l'enveloppe de l'intervalle de confiance à 95 %.

De plus, le graphique de la figure 5.46 présentant les valeurs prédites en fonction des valeurs observées montre que les points sont alignés avec la première bissectrice ce qui permet de valider ce modèle.

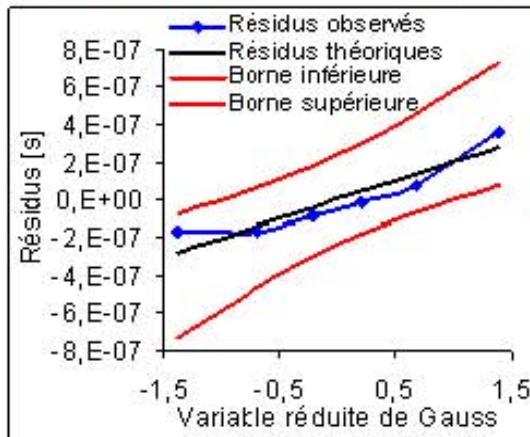


Figure 5.45 : Ajustement à une loi Normale des résidus, modèle de régression régularisée de la constante de cassage K_B , IC à 95 %

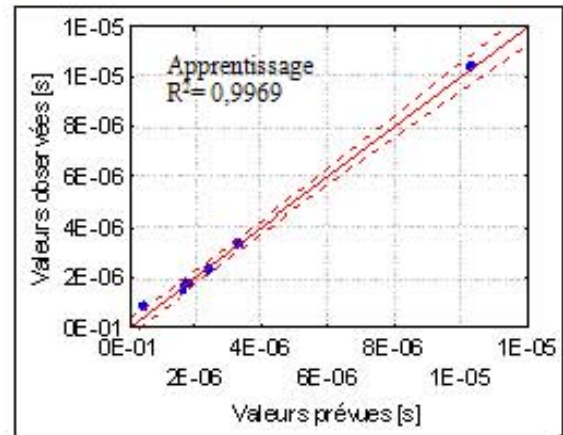


Figure 5.46 : Ajustement des valeurs observées et des valeurs prévues, modèle de régression régularisée de la constante de cassage K_B [s], IC à 95 %

Stabilité des coefficients de régression et influence des paramètres sur l'équation du modèle.

Les variables explicatives sont orthogonales, de coefficients de corrélation très faibles (il s'agit du cas idéal), les valeurs calculées de tolérance sont supérieure à 0,98, alors la multi-colinéarité entre les paramètres explicatifs est très faible. Il vient alors que les coefficients de régression n'encourent aucun risque ; ils sont stables.

Le retour aux variables initiales et le calcul des influences des paramètres initiales est présenté dans le tableau 5.32. Il apparait que la composante argileuse est d'une influence majeure sur la constante K_B , il vient ensuite la turbidité initiale "T₀", la matière organique "MO" et le diamètre "D_{mv}", suivi par l'influence du type d'argile "kaolinite, montmorillonite et illite" et parmi les dernières on note : la surface spécifique "Ssp", le carbone minérale "CM", et la Dolomite "Dolo" qui semblent de plus faibles influences.



Paramètre	Influence	Paramètre	Influence	Paramètre	Influence
Argile	19,91	D(0,5)	4,35	Chlorite	0,41
SiO ₂ +Al ₂ O ₃	15,67	Quartz	2,30	Uniformité	0,34
SiO ₂	10,77	Calcite	2,29	CaO	0,30
T ₀	8,57	Illite	2,23	CM	0,30
MO	7,04	COT	1,96	Mg	0,11
D _{mv}	5,79	D _{ms}	1,30	Na ₂ O	0,10
Kaolinite	4,77	Ssp	0,85	Dolomite	0,07
Al ₂ O ₃	4,74	Fe ₂ O ₃	0,84	MnO	0,008
Montmorillonite	4,52	K ₂ O	0,49		

Tableau 5.32 : Influence des variables explicatives sur l'équation de la régression, modèle de régression régularisée de la constante K_B .

V.2.3. Modèle de régression de la constante de formation K_A

Par observation des nuages de points présentés dans le chapitre 4 (figure 4.3), il ne semble pas y avoir de liaison linéaire ou autre entre la constante de formation K_A et les variables "T₀, CM, Arg, COT, Ssp, Dms, ILL et Mte".

Les tentatives de recherche d'un modèle de régression continu ou les variables explicatives présentent des influences réelles sur la variable à expliquer ont été vouées à l'échec. Dans des cas pareils, il est recommandé de réaliser une régression pondérée ou une régression linéaire par morceaux (modèle de régression discontinu).

En se référant au résultat, signalé au chapitre 3 (paragraphe VI-3.3), qui montre que la présence de la matière organique à de faibles concentrations améliore l'aptitude de la floculation et qu'à fortes concentrations l'effet inverse est marqué, nous proposons alors un modèle de régression multiple linéaire discontinu de la constante de formation K_A de la forme suivante :

$$Y = a_1 + a_2 \times X_1 + a_3 \times X_2 + \dots + a_n \times X_n \quad (\text{Si } COT < COT_{rupture})$$

$$Y = b_1 + b_2 \times X_1 + b_3 \times X_2 + \dots + b_n \times X_n \quad (\text{Si } COT \geq COT_{rupture})$$

X_1, X_2, \dots, X_n : Variables explicatives
 a_1, a_2, \dots, a_n et b_1, b_2, \dots, b_n : Coefficients de régression
 Y : Variable dépendante à expliquer

Les paramètres retenus pour le modèle sont listés dans le tableau 5.33 suivant :

N°	Paramètre	Description
1	T ₀	Turbidité initiale [NTU]
2	COT	Carbone organique total [%]
3	Arg	Taux d'argile [%]

Tableau 5.33 : Paramètres d'entrée du modèle de régression multiple de la constante de formation K_A

Le modèle de prédiction de la constante de formation K_A est sous la forme :

Pour **COT ≤ 5.55** (correspond à MO ≤ 25 %)

$$K_A = 0,272423.10^{-4} + 0,216870.10^{-4} \times T_0 + 0,179864.10^{-4} \times COT + 0,349716.10^{-4} \times Arg$$

Pour **COT ≥ 5.55**

$$K_A = 0,910832.10^{-3} - 0,16541.10^{-4} \times T_0 - 0,10017.10^{-4} \times COT + 0,324891.10^{-4} \times Arg$$

Le tableau 5.34 ci-dessous donne la synthèse de la régression : Le coefficient de détermination, les valeurs observées, les valeurs prédites et les résidus.

$R^2 = 0,9985$			
	Valeur Observée	Valeur Prévue	Résidus
Minimum	$4,72.10^{-4}$	$4,72.10^{-4}$	$-1,02.10^{-8}$
Maximum	$6,06.10^{-4}$	$6,06.10^{-4}$	$-6,52.10^{-10}$
Moyenne	$5,59.10^{-4}$	$5,59.10^{-4}$	$-3,28.10^{-9}$
Médiane	$5,73.10^{-4}$	$5,73.10^{-4}$	$-1,13.10^{-9}$

Tableau 5.34 : Synthèse des résidus de la régression multiple de la constante de formation K_A [-]

IL apparaît bien que la moyenne des résidus est très faible, elle est de l'ordre de $-3,28.10^{-9}$. Par ailleurs l'ajustement à la loi normale des résidus ainsi que l'ajustement des valeurs prévues et les valeurs observées sont montrés très satisfaisants respectivement aux figures 5.47 et 5.48

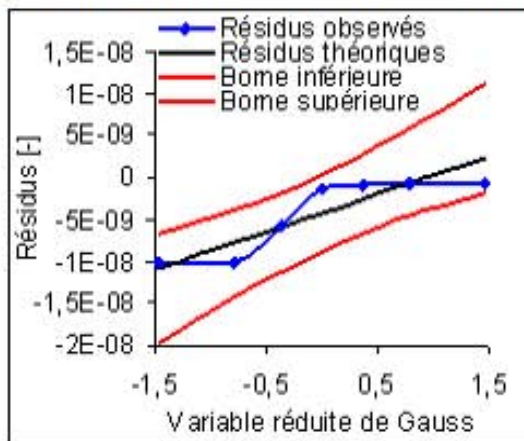


Figure 5.47 : Ajustement à une loi Normale des résidus, modèle de régression multiple de la constante de formation de K_A IC à 95 %

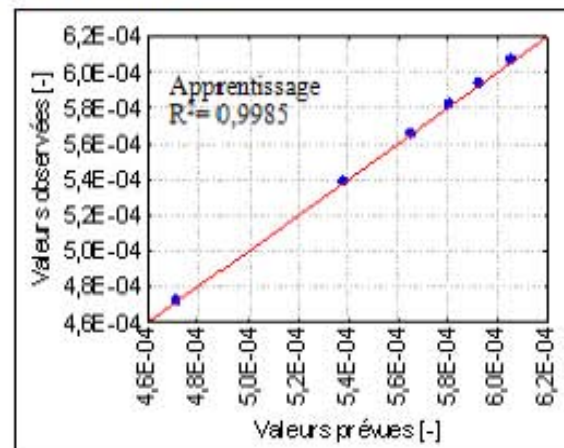


Figure 5.48 : Ajustement des valeurs observées et des valeurs prévues, modèle de régression multiple de la constante de formation K_A , IC à 95%

Influence des paramètres sur l'équation du modèle.

Les résultats de l'étude de l'influence des paramètres sur l'équation du modèle sont présentés sur la figure 5.49, la méthodologie de calcul est illustrée à l'annexe F.

Il apparaît de la figure que la turbidité présente l'influence majeure sur la constante de formation K_A suivi par le COT et en dernier lieu le taux d'argile dans le solide dispersé cela pour les deux cas ($COT \leq 5,55$ et $COT \geq 5,55$).

Selon l'équation du modèle il apparaît que jusqu'à 5,55 % de COT, la matière organique améliore la formation du floc (K_A augmente) et pour des COT supérieures, la matière organique inhibe cette formation.

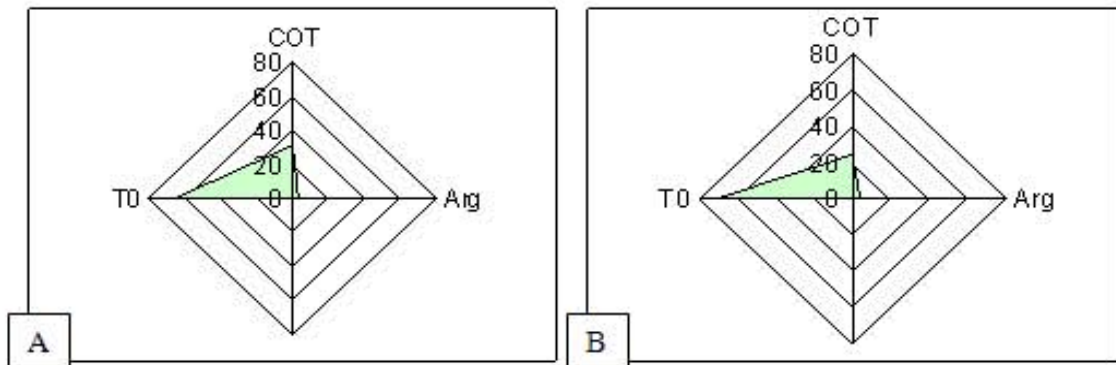


Figure 5.49 : Influence des paramètres sur l'équation du modèle de régression de la constante de formation K_A **A** : $COT < 5,55$ **B** : $COT \geq 5,55$

Il vient aussi que pour de faibles valeurs de COT ($COT \leq 5,55$), l'augmentation de la turbidité augmente la constante de formation K_A , et pour des valeurs élevées de COT ($COT \geq 5,55$) l'augmentation de la turbidité engendre une diminution de la constante de formation K_A .

Cet effet variable de la turbidité initiale sur la constante formation K_A a été signalée par HARHOOF (1996) qui note que pour des faibles dosages en coagulant l'augmentation de la turbidité initiale du kaolin augmente la constante de formation et pour des dosages élevés en coagulant, l'augmentation de la turbidité diminue la constante de formation.

V.2.4. Modèle de régression régularisée de la constante de formation K_A

Le meilleur ajustement de la régression régularisée de la constante de formation K_A est linéaire, les composantes principales retenues sont en nombre de trois ($Comp_1$, $Comp_2$, $Comp_4$). L'équation du modèle est la suivante:

$$K_A = 0,559167.10^{-3} + 0,500453.10^{-5} \times comp_1 + 0,128158.10^{-4} \times comp_2 - 0,39996.10^{-4} \times comp_4$$

Les résultats de la régression sont donnés dans le tableau 5.35, le modèle présente un R^2 égale à 0.95 cela montre que 95% de la variabilité de la constante de formation K_A est expliquée par ce modèle.

	C	Ecart-type de C	IC _{95%} de C	t(2)	p-niveau
R=0,975 R ² = 0,950 R ² aj= 0,875 F(3,2)= 12,676 p < 0,07400 Erreur type de l'estimation ,00002					
Intercepte	5,59167.10 ⁻⁴	7,04.10 ⁻⁶	5,29.10 ⁻⁴ à 5,89.10 ⁻⁴	79,5	0,000
Comp ₁	5,00453.10 ⁻⁶	1,93.10 ⁻⁶	-	2,59	0,120
Comp ₂	1,28158.10 ⁻⁵	2,54.10 ⁻⁶	1,87.10 ⁻⁶ à 2,38.10 ⁻⁵	5,04	0,037
Comp ₄	-3,99996.10 ⁻⁵	1,64.10 ⁻⁶	-	-2,44	0,130

Tableau 5.35 : Résultats de la régression régularisée de la constante de formation K_A .

Les tests F de Fischer et de t de Student sont un moyen d'apprécier la régression dans son ensemble, pour les probabilités indiquées dans le tableau, les tests de Student et de F de Fisher justifient l'apport simultané et individuel des variables explicatives au modèle.

Le tableau 5.36 ci-dessous, donne la synthèse de la régression : les valeurs observées, les valeurs prédites, les résidus, les résidus studentisés et le REQM.

Il en résulte que les résidus studentisés sont bornés dans l'intervalle [-1 à 1,5], ce qui signifie que pas de points extrêmes et les fluctuations des résidus sont faibles, ne dépassant pas 1,5 écart-type. La moyenne des résidus est nulle ; la qualité de l'ajustement est considérée alors très élevée.

	Valeur Observée	Valeur Prévue	Résidu	Résidu Studentisé	REQM
Minimum	4,72E-04	4,72E-04	-1,17E-05	-0,68	0,1.10 ⁻⁴
Maximum	6,06E-04	6,03E-04	1,93E-05	1,12	
Moyenne	5,59E-04	5,59E-04	0,00	0,00	
Médiane	5,73E-04	5,751E-04	-1,30E-06	-0,08	

Tableau 5.36 : Synthèse des résidus de la régression régularisée de la constante de formation K_a [-]

La normalité des résidus est évaluée par ajustement graphique à la figure 5.50; Les valeurs expérimentales (observées) sont proches des valeurs théoriques et sont incluses dans l'enveloppe de l'intervalle de confiance de 95%.

L'ajustement des valeurs prédites et observées est présenté sur la figure 5.50 ; Les points sont alignés autour de la première bissectrice et sont à plus de 80% incluses dans l'enveloppe de l'intervalle de confiance de 95%.

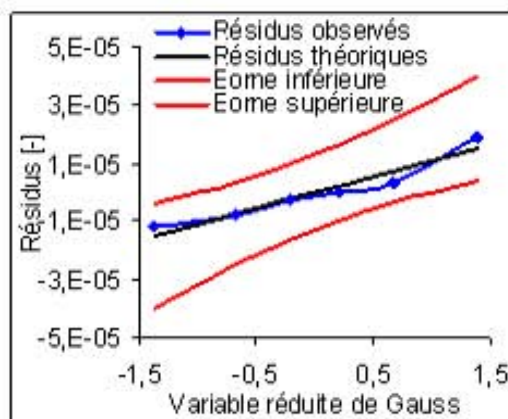


Figure 5.49 : Ajustement à une loi Normale des résidus, modèle de régression régularisée de la constante de formation K_a , IC à 95 %

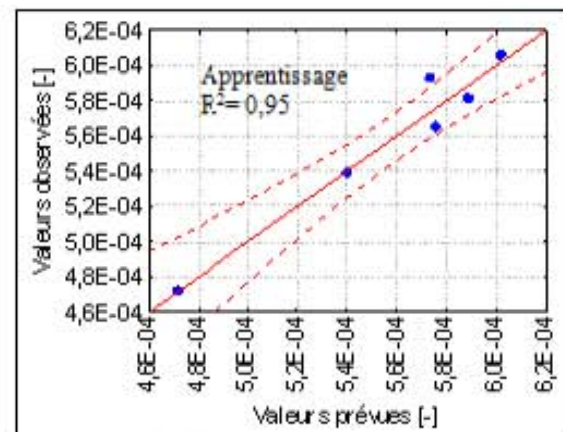


Figure 5.50 : Ajustement des valeurs observées et des valeurs prévues, modèle de régression régularisée de la constante de formation K_a , IC à 95%

Stabilité des coefficients de régression et influence des paramètres sur l'équation du modèle.

Les composantes principales descriptives présentent des coefficients de corrélation très faibles, les valeurs de tolérance calculées sont supérieures à 0,98 alors la multi-colinéarité entre les composantes principales explicatives est très faible. Il vient alors que les coefficients de régression du modèle de prédiction de la constante de formation n'encourent aucun risque, ils sont stables.

Ces composantes principales sont liées linéairement aux variables initiales, le remplacement de ces composantes et le retour aux variables initiales permet de calculer les influences de ces variables initiales sur la constante de formation K_A . Le tableau 5.37 récapitule le résumé de ce calcul.

<i>Paramètre</i>	<i>Importance</i>	<i>Paramètre</i>	<i>Importance</i>	<i>Paramètre</i>	<i>Importance</i>
SiO ₂ +AL ₂ O ₃	18,44	Calcite	3,07	Mg	0,62
Argile	17,68	D(0,5)	2,73	Dms	0,59
Kaolinite	14,05	COT	1,18	Fe ₂ O ₃	0,50
SiO ₂	8,61	Illite	1,11	Dolomite	0,35
AL ₂ O ₃	8,47	CaO	0,90	Uniformité	0,20
Montmorillonite	5,32	CM	0,89	Ssp	0,17
T ₀	4,83	Quartz	0,73	Na ₂ O	0,15
MO	4,43	Chlorite	0,66	MnO	0,01
Dmv	3,62	K ₂ O	0,66		

Tableau 5.37 : Influence des variables explicatives sur l'équation de la régression, modèle de régression régularisée de la constante K_A

(La méthodologie de calcul de l'influence des paramètres descripteurs est montrée à l'annexe E). Nous concluons de ce tableau que les paramètres les plus influents sont : la composante argileuse, puis la turbidité initiale "T₀", la matière organique "MO" et les diamètres "Dmv, D(0,5)" et parmi les moins influents, nous notons la surface spécifique "Ssp" et l'uniformité de distribution granulométrique des particules solides mises en suspension.

VI. MODELISATION NON PARAMETRIQUE DES CONSTANTES CINÉTIQUES K_A ET K_B

VI.1. Modèle neuronal.

Les variables retenues pour la régression non paramétriques par réseau de neurones sont listées dans le tableau 5.38 suivant :

<i>N°</i>	<i>Paramètre</i>	<i>Description</i>
1	T ₀	Turbidité initiale [NTU]
2	COT	Carbone organique total [%]
3	Arg	Taux d'argile [%]

Tableau 5.38 : Paramètres d'entrée du modèle de régression neuronale des constantes cinétiques K_A et K_B

Pour déterminer le nombre de neurones dans la couche cachée, on a augmenté progressivement le nombre de neurones dans cette couche de 1 à 5. Chaque réseau construit a été ainsi entraîné avec la fonction sigmoïde logistique. La figure 5.52 montre l'évolution du R^2 en fonction du nombre de neurones dans la couche cachée.

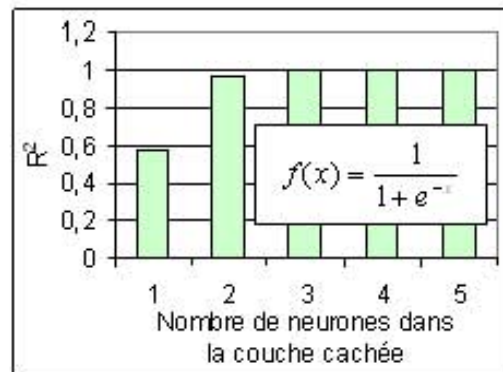


Figure 5.52 : R^2 en fonction du nombre de neurones dans la couche cachée, modèle de régression neuronale des constantes cinétiques K_A et K_B

Le modèle retenu est le modèle 3-3-2 ; trois neurones dans la couche d'entrée, trois neurones dans la couche cachée et deux sorties K_A et K_B , le PMC retenu est non complexe et présente un R^2 très satisfaisant (voir tableau 5.39), l'architecture de ce modèle est illustrée par la figure 5.53

Modèle	R^2 sur le jeu d'apprentissage		
	K_A	K_B	Moyen
PMC 3-3-2	0,999981	0,999955	0,999968

Tableau 5.39 : Résultats de la régression neuronale des constantes cinétiques K_A et K_B

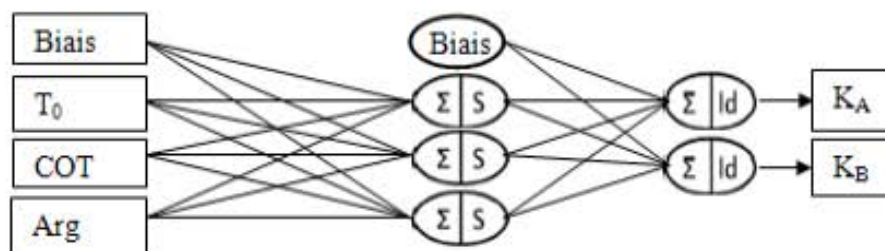


Figure 5.53 : Architecture du perceptron multicouche 3/3/2 modèle de régression neuronal des constantes cinétiques K_A et K_B

Le tableau 5.40, ci-dessous donne la synthèse de la régression : il apparaît que la moyenne et la médiane des résidus et la valeur de REQM sont très faibles, les résidus studentisés sont bornés dans l'intervalle $[-1,5 \text{ à } 2]$, ce qui montre l'absence des points extrêmes qui peuvent influencer l'estimation du modèle.

		Minimum	Maximum	Moyenne	Médiane
K_a [-]	Observée	$4,72 \cdot 10^{-4}$	$6,06 \cdot 10^{-4}$	$5,59 \cdot 10^{-4}$	$5,73 \cdot 10^{-4}$
	prévue	$4,72 \cdot 10^{-4}$	$6,06 \cdot 10^{-4}$	$5,59 \cdot 10^{-4}$	$5,73 \cdot 10^{-4}$
	Résidus	$-2,52 \cdot 10^{-7}$	$1,62 \cdot 10^{-7}$	$1,03 \cdot 10^{-9}$	$8,22 \cdot 10^{-9}$
	Résidus studentisés	-0,444	0,285	$1,81 \cdot 10^{-3}$	0,0145
	REQM	$0,001 \cdot 10^{-4}$			
K_E [s]	Observée	$8,0 \cdot 10^{-7}$	$1,04 \cdot 10^{-6}$	$3,36 \cdot 10^{-6}$	$2,07 \cdot 10^{-6}$
	prévue	$7,74 \cdot 10^{-7}$	$1,04 \cdot 10^{-6}$	$3,36 \cdot 10^{-6}$	$2,08 \cdot 10^{-6}$
	Résidus	$-5,00 \cdot 10^{-8}$	$7,88 \cdot 10^{-8}$	$9,16 \cdot 10^{-11}$	$-1,84 \cdot 10^{-9}$
	Résidus studentisés	-1,23	1,94	$2,25 \cdot 10^{-3}$	-0,451
	REQM	$0,039 \cdot 10^{-6}$			

Tableau 5.40: Synthèse des résidus de la régression neuronale des constantes cinétiques K_a et K_E

La normalité des résidus est vérifiée graphiquement, à la figure 5.54. Les valeurs expérimentales (observées) sont proches des valeurs théoriques pour un seuil de signifiante $\alpha = 0,05$ et sont incluses dans l'enveloppe de l'intervalle de confiance de 95%.

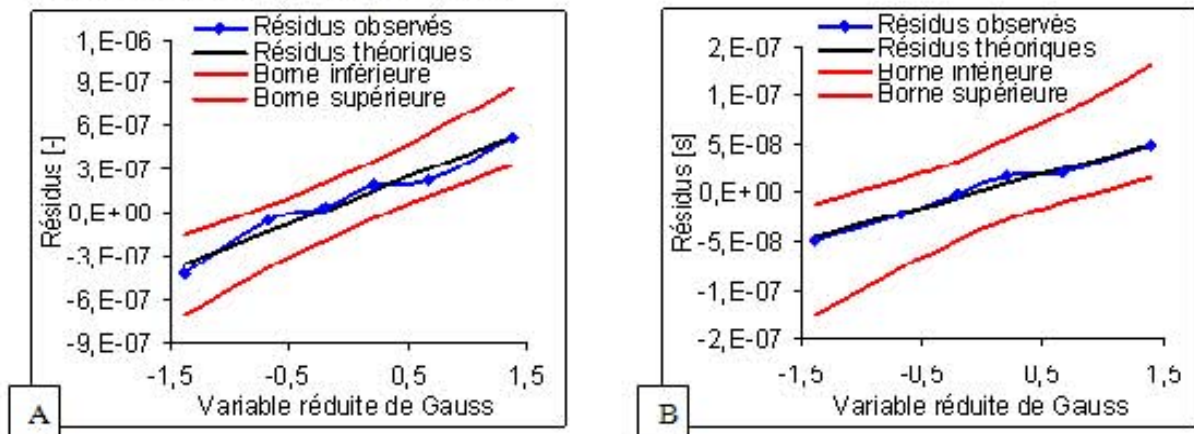


Figure 5.54 : Ajustement à une loi Normale des résidus, modèle de régression neuronale des constantes cinétiques **A** : Constante de formation K_a , IC à 95 % **B** : Constante de passage K_E (IC à 95 %)

L'ajustement des valeurs prévues sur les valeurs observées est présenté à la figure 5.55 suivante ; les points sont ordonnés sur la première bissectrice, l'ajustement est excellent.

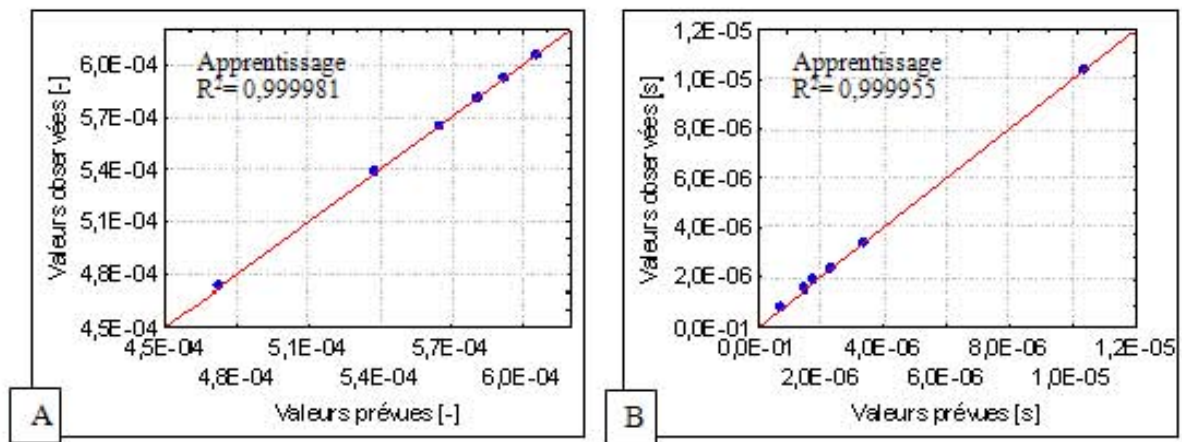


Figure 5.54 : Ajustement des valeurs observées et des valeurs prévues **A** : de la constante de formation K_A **B** : de la constante de cassage K_B , modèle de régression neuronale (IC à 95%)

Importance relative des variables d'entrée

Suivant la méthode des poids de connexion illustrée dans l'annexe F, il apparaît comme le montre la figure 5.56 que la constante de formation K_A est influencée positivement par la turbidité initiale "T₀" et négativement par le COT. Par contre, la constante de cassage K_B est influencée positivement par le COT et négativement par la turbidité initiale "T₀", le taux d'argile présente pour les deux constantes une faible influence.

L'augmentation de la turbidité initiale améliore la constante de formation K_A et diminue la constante de cassage K_B , ce qui est en accord avec les résultats de HARHOOF (1996) qui note que quelle que soit la dose de coagulant injecté, l'accroissement de la turbidité initiale de 20 à 100 NTU diminue la constante de cassage K_B .

Par ailleurs l'augmentation de taux de COT dans le solide dispersé diminue la constante de formation K_A (effet négatif) et à l'inverse augmente la constante de cassage K_B (effet positif)

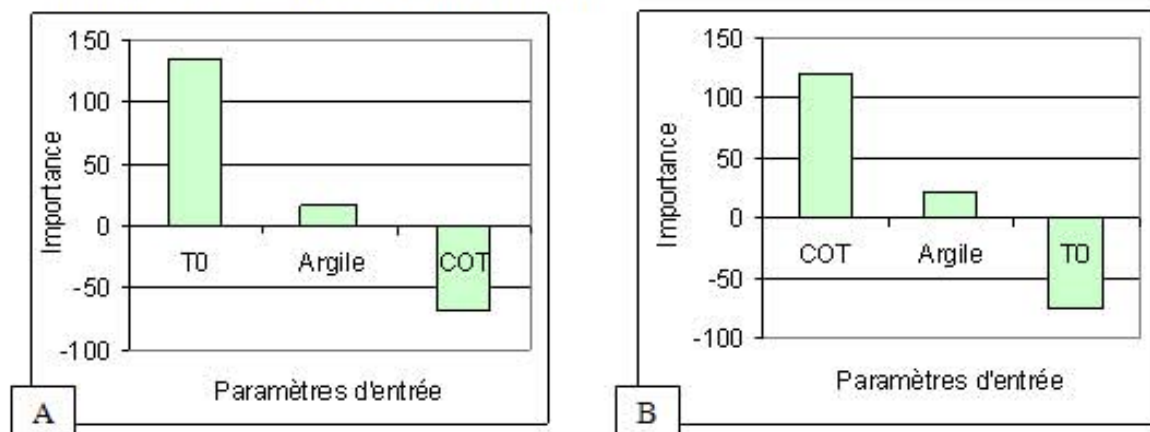


Figure 5.56 : Importance relative des variables d'entrée, **A** : sur la constante de formation K_A **B** : sur la constante de cassage K_B , modèle de régression neuronale des constantes cinétiques.

VI.2. Modèle neuronal basé sur les composantes principales

Dans la régression neuronale basée sur les composantes principales, les variables d'entrées sont remplacées par les composantes principales.

Pour déterminer l'architecture optimale du modèle neuronal on a augmenté progressivement le nombre des entrées (nombre de composantes) "I" de 1 à 3 (ce qui correspondait à des valeurs propres supérieures à 1 et une variance expliquée qui dépasse, pour trois composantes, 99 %, voir tableau 5.29) et on a varié le nombre de neurones de la couche cachée, de 1 à "2I+1" jusqu'à atteindre la meilleure précision. Chaque réseau construit a été ainsi entraîné avec la fonction sigmoïde logistique.

La figure 5.57 illustre les meilleures configurations du réseau de neurones, il apparaît qu'à avec une seule composante (une seule entrée) le coefficient de détermination R^2 est faible alors que pour deux et trois composantes le R^2 est élevé.

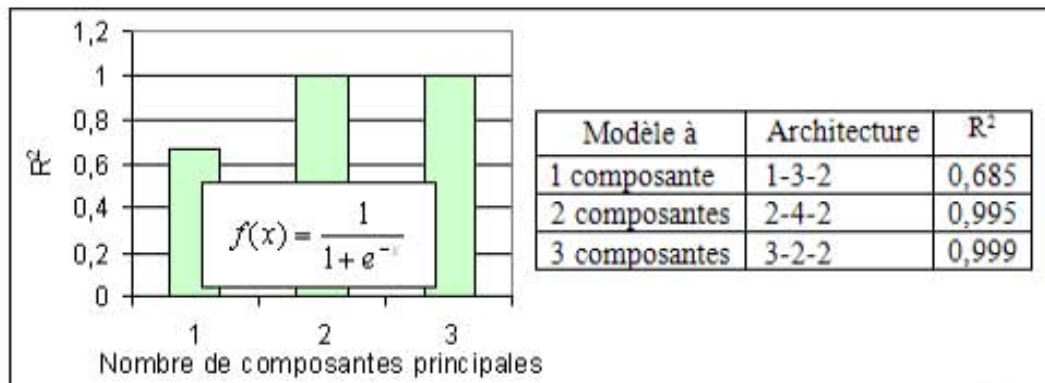


Figure 5.57 : R^2 en fonction du nombre de composantes principales dans la couche d'entrée, modèle de régression neuronale sur composantes principales des constantes cinétiques K_A et K_B

Parmi ces trois configurations le modèle retenu est d'architecture suivante : 3/2/2 ; trois composantes principales comme neurones d'entrée, deux neurones dans la couche cachée et deux neurones de sortie K_A et K_B , ce modèle est le plus performant et le moins complexe. Le tableau 5.41 résume la performance du modèle et la figure 5.58 illustre son architecture

Modèle	R^2 sur le jeu d'apprentissage		
	K_A	K_B	Moyen
PMC 3/2/2	0,9999	0,9995	> 0,999

Tableau 5.41 : Résultats de la régression neuronale basée sur les composantes principales des constantes cinétiques K_A et K_B

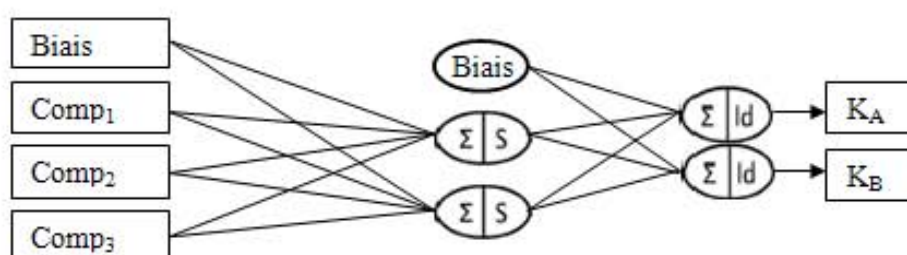


Figure 5.58 : Architecture du perceptron multicouche 3/2/2, Modèle de régression neuronale sur composantes principales des constantes cinétiques K_A et K_B

Le tableau 5.42, ci-dessous donne la synthèse de la régression : les valeurs observées, les valeurs prédites, les résidus, les résidus studentisés et la valeur de REQM. Il apparaît que la moyenne, la médiane des résidus et la REQM sont très faibles, les résidus studentisés sont bornés dans l'intervalle [-2 à 2], ce qui est le signe du bon ajustement du modèle neuronal.

La normalité des résidus est vérifiée graphiquement (figure 5.59). Les valeurs expérimentales (observées) sont proches des valeurs théoriques pour un seuil de signifiante $\alpha = 0,05$ et sont incluses dans l'enveloppe de l'intervalle de confiance de 95%.

		<i>Minimum</i>	<i>Maximum</i>	<i>Moyenne</i>	<i>Médiane</i>
K_A [-]	Observée	$4,72 \cdot 10^{-4}$	$6,06 \cdot 10^{-4}$	$5,59 \cdot 10^{-4}$	$5,73 \cdot 10^{-4}$
	Prévue	$4,72 \cdot 10^{-4}$	$6,06 \cdot 10^{-4}$	$5,59 \cdot 10^{-4}$	$5,73 \cdot 10^{-4}$
	Résidus	$-6,08 \cdot 10^{-8}$	$1,40 \cdot 10^{-7}$	$2,28 \cdot 10^{-9}$	$-4,88 \cdot 10^{-9}$
	Résidus studentisés	-0,0438	0,101	$1,64 \cdot 10^{-8}$	$-3,51 \cdot 10^{-8}$
	REQM	$0,0007 \cdot 10^{-4}$			
K_B [s]	Observée	$8,0 \cdot 10^{-7}$	$1,04 \cdot 10^{-6}$	$3,36 \cdot 10^{-6}$	$2,07 \cdot 10^{-6}$
	prévue	$8,98 \cdot 10^{-7}$	$1,04 \cdot 10^{-6}$	$3,37 \cdot 10^{-6}$	$1,98 \cdot 10^{-6}$
	Résidus	$-2,05 \cdot 10^{-7}$	$9,76 \cdot 10^{-8}$	$-5,70 \cdot 10^{-10}$	$1,84 \cdot 10^{-8}$
	Résidus studentisés	-2,00	0,981	$-5,73 \cdot 10^{-8}$	$1,85 \cdot 10^{-1}$
	REQM	$0,099 \cdot 10^{-6}$			

Tableau 5.42 : Synthèse des résidus de la régression neuronale basée sur les composantes principales des constantes cinétiques K_A et K_B

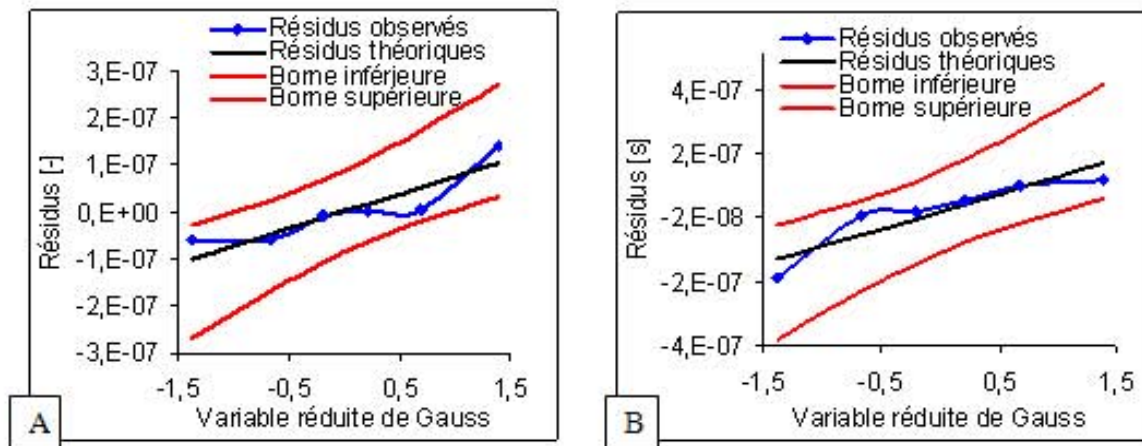


Figure 5.58 : Ajustement à une loi Normale des résidus, modèle de régression neuronale sur composantes principales des constantes cinétiques **A** : Constante de formation K_A (IC à 95 %) **B** : Constante de cassage K_B (IC à 95 %)

L'ajustement des valeurs prévues sur les valeurs observées est présenté à la figure 5.60 suivante ; les points sont alignés avec la première bissectrice l'ajustement est excellent.

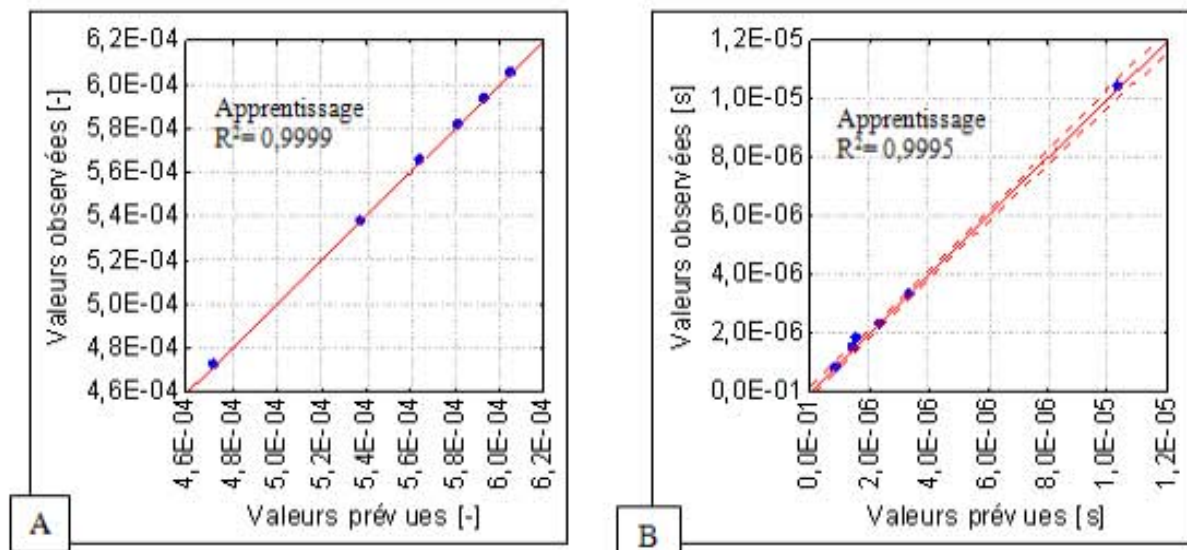


Figure 5.60 : Ajustement des valeurs observées et des valeurs prévues **A** : de la constante de formation K_A **B** : de la constante de cassage K_B , modèle de régression neuronale sur composantes principales (IC à 95%)

VII CONCLUSION

Dans cette section nous avons établi quatre types de modèles de régression des constantes cinétiques de formation K_A et de cassage K_B :

- Modèles de régression multiple
- Modèle de régression régularisée
- Modèles de régression neuronale
- Modèle par régression neuronale basée sur les composantes principales

La régression multiple discontinue de la constante de formation K_A et celle continue de la constante de cassage K_B présentent des performances très élevées ($R^2 > 0,998$), mais la grande colinéarité entre les descripteurs, traduite par les faibles valeurs de tolérance, risque l'instabilité de l'estimation et appauvrit la robustesse du modèle.

Par ailleurs la régression régularisée de ces deux constantes apparaît légèrement inférieure à celle de la régression multiple ($0,95 < R^2 < 0,997$), mais elle présente l'avantage d'être plus stable.

Le modèle neuronal simple et celui basé sur les composantes principales ajustent pareillement bien la constante de formation K_A et la constante de cassage K_B . Ces modèles ont permis des performances de prédiction excellentes, démontrés par la valeur très élevée du coefficient de détermination moyen ($R^2 > 0,9995$) et des faibles valeurs des résidus moyens (les résidus sont inférieurs à 10^{-8}).

Les différences relevées du calcul des influences des paramètres descripteurs sur l'estimation de la constante de formation K_A et de la constante de cassage K_B , pour les trois régressions établies multiple, régularisée et neuronale sont tributaires de la complexité du modèle et de la méthode du calcul des influences et peuvent être à priori synonyme d'absence des influents prépondérants.

VIII. MODELISATION DES PARAMETRES DE FINALITE DE TRAITEMENT : LA TURBIDITE FINALE "T_F" ET L'ALUMINIUM RESIDUEL "AL_r"

Dans cette partie nous avons repris la démarche de modélisation par réseau de neurones développée dans les sections précédentes pour l'appliquer à la prédiction de la qualité des eaux après floculation, par l'estimation des paramètres de finalité de traitement ; l'aluminium résiduel "Al_r" et la turbidité finale "T_f". Ces modèles sont élaborés à la base des résultats bruts d'essai jar test, suivant le schéma de la figure 5.61, ci-dessous.

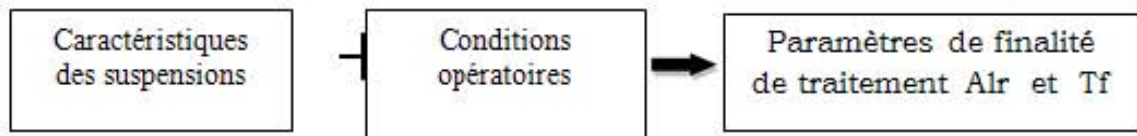


Figure 5.61 : Schéma de régression neuronale des paramètres de finalité de traitement Al_r et T_f

VIII.1. CHOIX ET PREPARATION DES DONNEES

La population est composée de 319 individus statistiques et de 14 variables descriptives récapitulées dans le tableau 5.43 ci-dessous. La base de données est divisée aléatoirement en deux jeux de données : respectivement 2/3 pour l'apprentissage et 1/3 pour la validation.

N°	Paramètre	Description
1	Dose	Dose du coagulant injectée [mgAl/l]
2	Pui	Puissance d'agitation lente [w/m³]
3	T _o	Turbidité initiale [NTU]
4	C _p	Concentration des particules [mg/l]
5	COT	Carbone organique total [%]
6	CM	Carbone minéral [%]
7	D _{ms}	Diamètre moyen en surface [µm]
8	S _{sp}	Surface spécifique [m²/g]
9	Kao	Kaolinite [%]
10	ILL	Illite [%]
11	Mte	Montmorillonte [%]
12	Arg	Argile [%]
13	K _a	Constante de formation [-]
14	K _E	Constante de cassage [s]

Tableau 5.43 : Paramètres d'entrée du modèle de régression neuronale des paramètres Al_r et T_f

Les graphiques "box-plots" de la figure 5.62 suivante montrent la dispersion des échantillons d'apprentissage et celle de validation. Sur cette figure le fond, le point central et le plafond de chacune des boîtes représentent respectivement : "la moyenne-2Erreur type", "la moyenne" et "la moyenne+2Erreur type", de la distribution des données, les barres verticales représentent quant à elles les minima et les maxima. Il apparaît clairement

que la disparité est similaire pour les deux jeux de données, cela montre l'homogénéité des caractéristiques des deux échantillons.

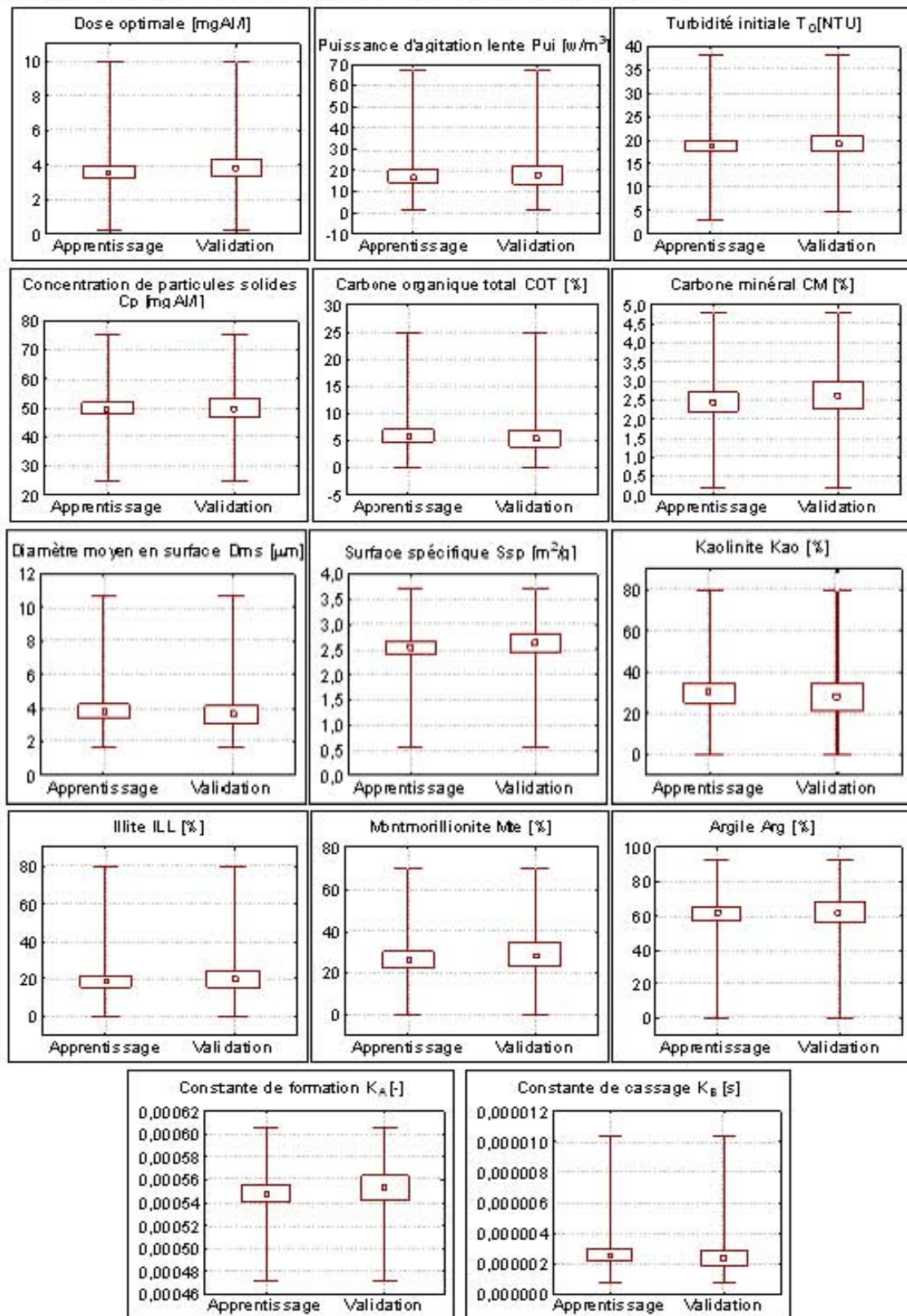


Figure 5.62 : Présentation box-plot « boîtes à moustache » des données d'apprentissage et de validation (test), Modélisation des paramètres Tf et Alr (Echantillonnage aléatoire) □ Moyenne □ Moyenne ± 2Erreur type T Min-Max

VIII.2. MODELE NEURONAL

Le modèle proposé est un modèle neuronal PMC (perceptron multicouche), à double sorties : l'aluminium résiduel "Alr" et la turbidité finale "Tf".

Les choix faits en termes de construction, d'architecture de réseau de neurones et de mode d'évaluation du modèle sont récapitulés dans le tableau 2.2 (Chapitre 2).

Développement du modèle

Quatorze variables sont retenues pour la prédiction des paramètres de turbidité finale et d'aluminium résiduel. Pour déterminer le nombre de neurones de la couche cachée, on a augmenté progressivement le nombre de neurones dans cette couche de 1 à 29. Chaque réseau construit a été ainsi entraîné avec la fonction sigmoïde logistique, à noter que pour chaque configuration neuronale, 1000 valeurs de jeu de poids initiaux ont été testées et c'est le modèle qui présente le meilleur résultat qui est sélectionné. Parmi les différentes configurations vérifiées, nous avons retenu celles produisant le plus élevé coefficient de détermination R^2 moyen sur le jeu de validation, la figure 5.63 présente la variation de R^2 pour les différentes configurations testées.

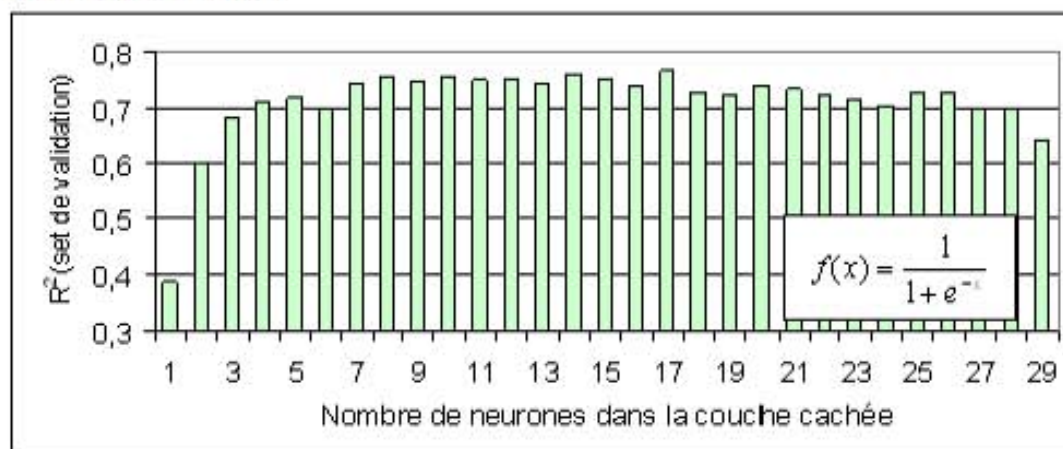


Figure 5.63 : R^2 en fonction du nombre de neurones dans la couche cachée, modèle de régression neuronale des paramètres Alr et Tf (set de validation)

Le modèle retenu est celui qui donne la meilleure performance (R^2 élevé) il vient alors de choisir le réseau d'architecture suivante 14/17/2 ; 14 neurones de la couche d'entrée (14 paramètres d'entrée), une seule couche cachée composée de 17 neurones et deux neurones de sortie, pour des configurations neuronales plus complexes le R^2 ne s'améliore plus.

La performance du modèle neuronale est résumée dans le tableau 5.44 suivant :

Modèle	R^2 sur le jeu d'apprentissage			R^2 sur le jeu de validation		
	Tf	Alr	Moyen	Tf	Alr	Moyen
PMC 14-17-2	0.90	0.77	0,84	0.78	0.75	0.765

Tableau 5.44 : Résultats de la régression neuronale des paramètres Alr et Tf

Pour le jeu d'apprentissage le R^2 moyen est de 0,84 alors que pour le jeu de validation le R^2 moyen est de 0,765 signifiant respectivement que 84% de la variabilité du jeu d'identification est expliquée par le modèle et que 76,5% de la variabilité du jeu de validation est validée par le modèle, le pouvoir explicatif et prédictif du modèle est bon.

La figure 5.64 illustre l'architecture de ce modèle neuronal.

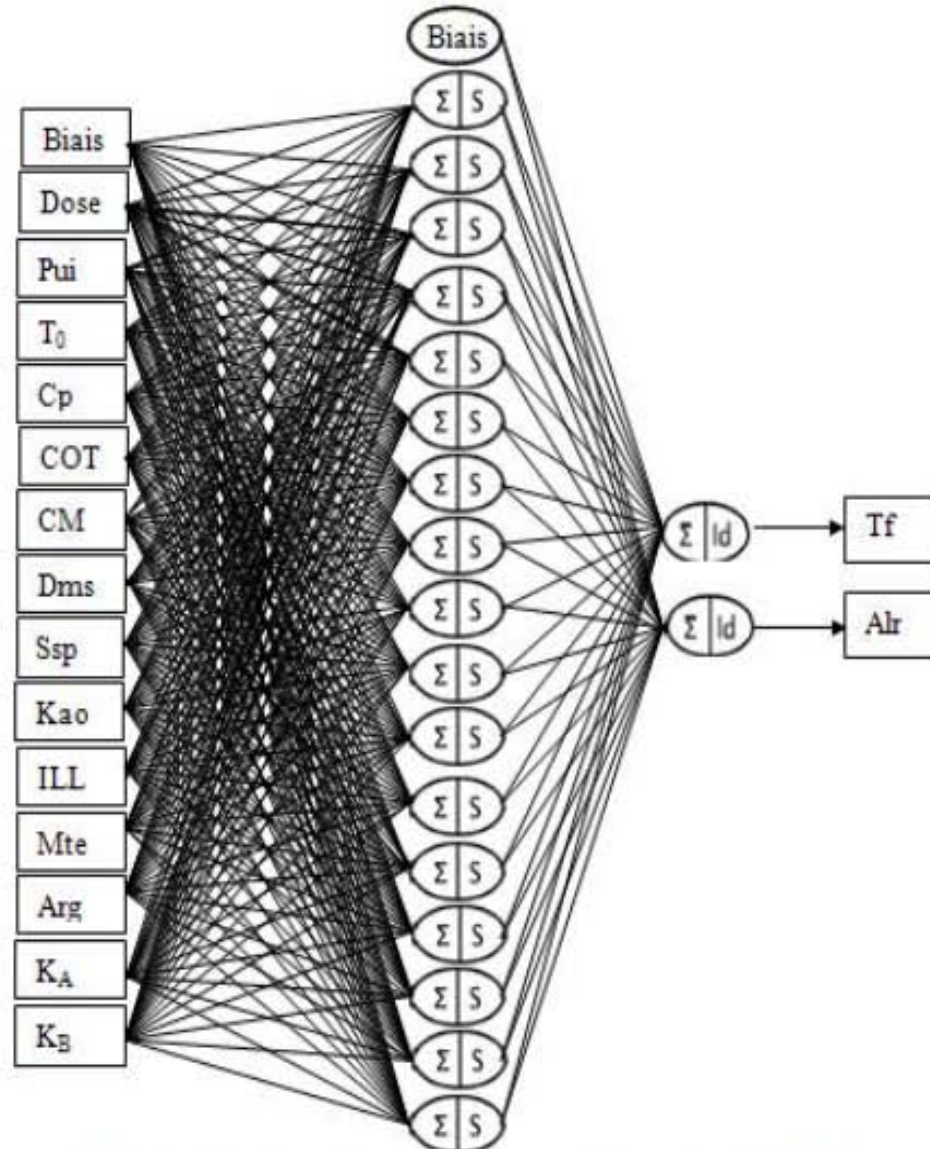


Figure 5.64: Architecture du perceptron multicouche 14/17/2, Modèle de régression neuronale des paramètres : Tf et Alr

Le tableau 5.45, ci-dessous donne la synthèse de la régression pour l'échantillon d'apprentissage : les valeurs observées, les valeurs prédites, les résidus, les résidus studentisés, et la valeur de REQM.

La moyenne des résidus est nulle, les valeurs de REQM sont faibles (0,75 NTU pour la turbidité et 0,03 mgAl/l pour l'aluminium résiduel) ce qui est le signe de bon ajustement du modèle neuronal.

Les résidus studentisés sont, pour la turbidité finale, bornés dans l'intervalle [-2,27 à 2,48], cependant pour l'aluminium résiduel l'intervalle est un peu large [-3,9 à 2,6], indiquant la présence de quelques points extrêmes, qui peuvent appauvrir la robustesse du modèle.

		<i>Minimum</i>	<i>Maximum</i>	<i>Moyenne</i>	<i>Médiane</i>
Tf [NTU]	Observée	0,5	17	3,09	2,40
	Prévue	0,49	15,37	3,09	2,53
	Résidus	-2,98	3,26	0,00	0,07
	Résidus studentisés	-2,27	2,48	0,00	0,05
	REQM	0,75			
Alr [mgAl/l]	Observée	0,011	0,549	0,101	0,080
	prévue	0,014	0,505	0,101	0,085
	Résidus	-0,167	0,112	0,000	0,002
	Résidus studentisés	-3,906	2,605	0,010	0,054
	REQM	0,03			

Tableau 5.45 : Synthèse des résidus de la régression neuronale des paramètres Alr et Tf (set d'apprentissage)

La normalité des résidus est vérifiée graphiquement (voir figure 5.65). Les valeurs expérimentales (observées) sont proches des valeurs théoriques pour un seuil de signifiacnce $\alpha = 0,05$ et sont incluses, sauf quelques points extrêmes, dans l'enveloppe de l'intervalle de confiance de 95%.

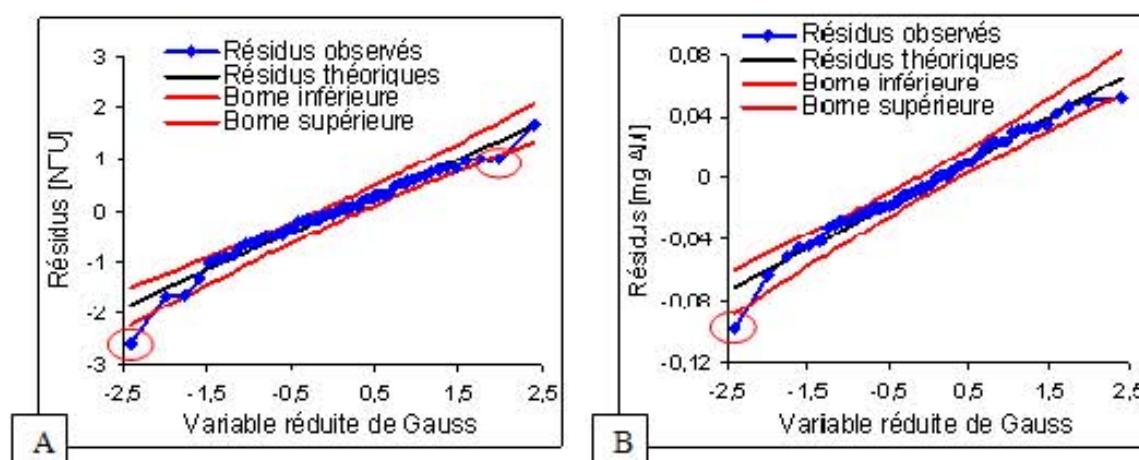


Figure 5.65 : Ajustement à une loi Normale des résidus, modèle de régression neuronale des paramètres Tf et Alr **A** : Turbidité finale Tf (IC à 95 %), **B** : Aluminium résiduel Alr (IC à 95 %)

Validation du modèle

La validation est faite sur le deuxième échantillon, le tableau 5.46, ci-dessous donne la synthèse : les valeurs observées, les valeurs prédites, les résidus, les résidus studentisés, le REQM ; il apparaît que la moyenne des résidus est très faible (-0,07 NTU pour la turbidité et -0,003 mgAl/l pour l'aluminium résiduel), les valeurs des résidus studentisés ne dépasse pas en valeur absolue 2,3 alors que les valeurs de REQM sont faibles : 0,71 NTU pour la turbidité et 0,03 mgAl/l de l'aluminium résiduel ce qui est le signe de bon validation du modèle neuronal.

		<i>Minimum</i>	<i>Maximum</i>	<i>Moyenne</i>	<i>Médiane</i>
Tf [NTU]	Observée	0,50	9,60	2,68	2,35
	Prévue	0,56	7,94	2,61	2,44
	Résidus	-2,59	1,66	-0,07	-0,04
	Résidus studentisés	-1,97	1,26	-0,05	-0,03
	REQM	0,71			
Alr [mgAl/l]	Observée	0,026	0,302	0,099	0,084
	prévue	0,000	0,322	0,096	0,086
	Résidus	-0,098	0,052	-0,003	-0,005
	Résidus studentisés	-2,297	1,214	-0,070	-0,118
	REQM	0,03			

Tableau 5.46 : Synthèse des résidus de la régression neuronale des paramètres Alr et Tf (set de validation)

La vérification de l'homoscédasticité consiste à rechercher si dans les résidus, il n'existe pas une structure organisée ou un contenu informationnel (cône, vague) qui prouverait que pas d'égalité de variance des résidus. Le nuage de points des résidus calculés sur l'échantillon de validation, illustré dans la figure 5.66 n'a indique aucune hétéroscédasticité.

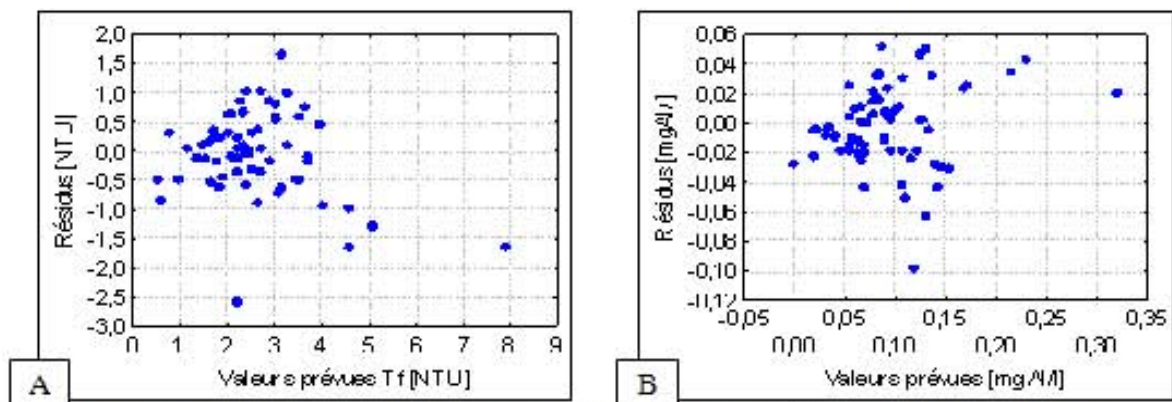


Figure 5.66 : Vérification de l'homoscédasticité, **A** : Turbidité finale Tf, **B** : Aluminium résiduel Alr, modèle de régression neuronale des paramètres Tf et Alr (set de validation)

L'ajustement des valeurs de turbidité finale observées et prévues pour les échantillons d'apprentissage et de validation est présenté à la figure 5.67. L'ajustement est d'autant meilleur que les valeurs se rapprochent de la première bissectrice.

Il résulte de cette modélisation neuronale que respectivement 90% de la variabilité du jeu d'apprentissage de la turbidité finale est expliquée par le modèle et que seulement 78% de la variabilité du jeu de validation est validée par le modèle.

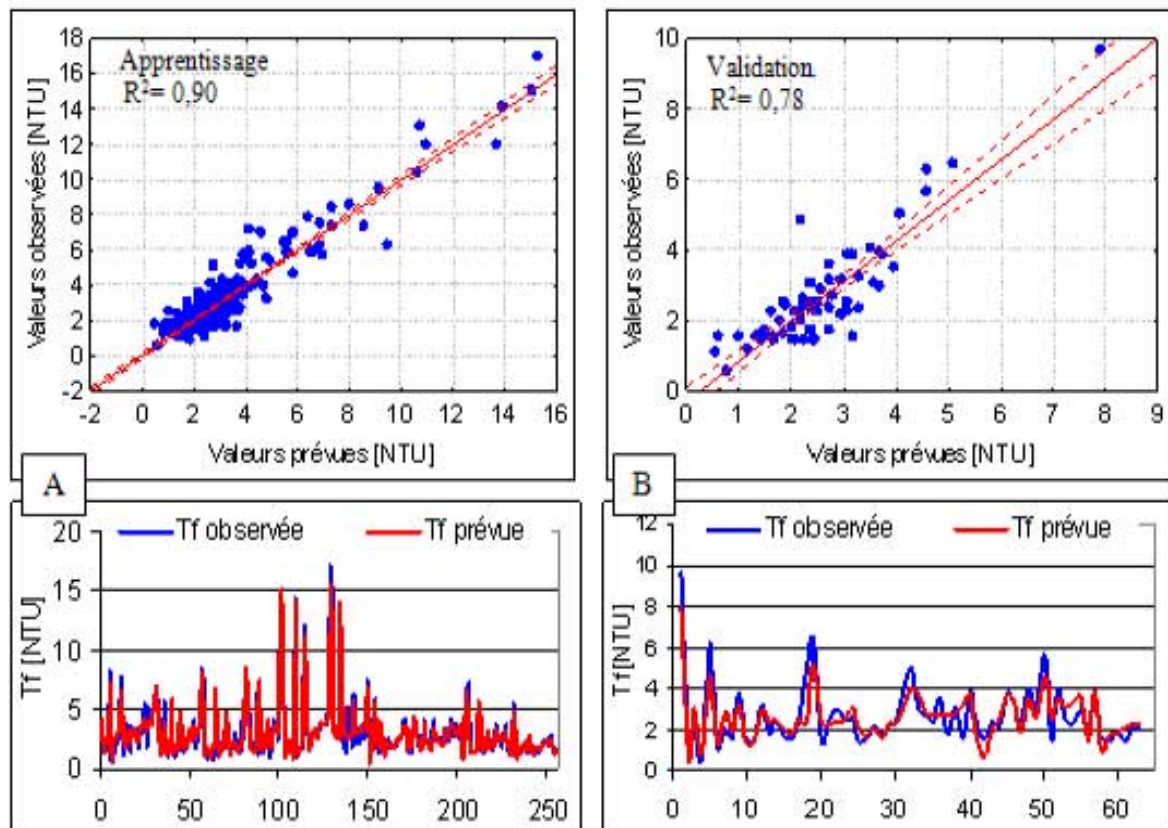


Figure 5.67 : Ajustement des valeurs observées et des valeurs prévues de la turbidité finale Tf, modèle de régression neuronale des paramètres Tf et Alr

A : set d'apprentissage, **B** : set de validation, IC à 95%

Pour l'aluminium résiduel, l'ajustement des valeurs observées (expérimentales) et prévues par le modèle pour les deux échantillons d'apprentissage et de validation est présenté à la figure 5.68. L'ajustement est d'autant meilleur que les valeurs s'alignent avec la première bissectrice. Pour l'échantillon d'apprentissage 77% de la variance est expliquée par le modèle cependant pour l'échantillon de validation 75% de la variance est validée. Le pouvoir explicatif du modèle est bon.

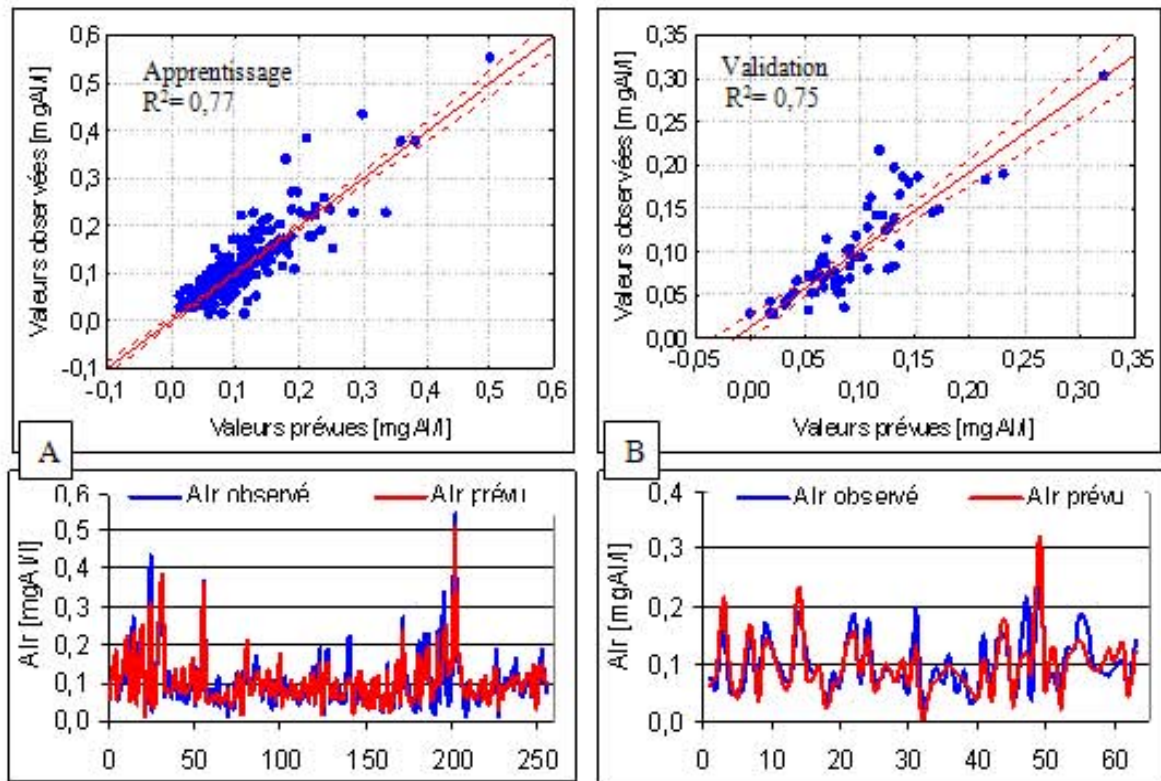


Figure 5.68 : Ajustement des valeurs observées et des valeurs prévues de l'aluminium résiduel, modèle de régression neuronale des paramètres Tf et Alr **A** : set d'apprentissage, **B** : set de validation, IC à 95%

Analyse de l'importance relative des variables d'entrée

Le résultat d'analyse de l'importance des variables explicatives de la régression neuronale des paramètres ALr et Tf selon la méthode de poids de connexion est montré sur la figure 5.69

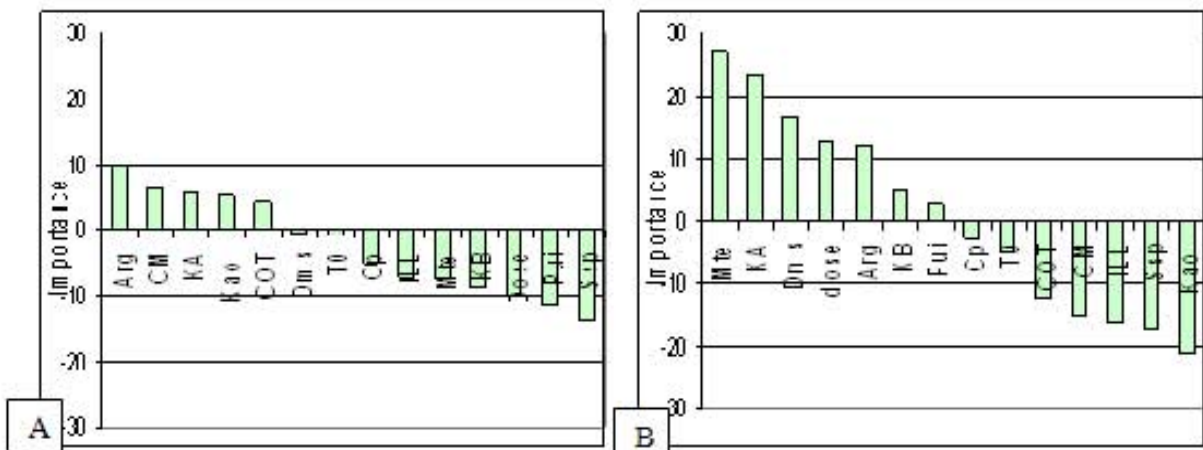


Figure 5.69 : Importance relative des variables d'entrée, **A** sur la turbidité finale Tf **B** sur l'aluminium résiduel Alr, modèle de régression neuronale des paramètres Tf et Alr

Il en ressort que les influences des paramètres descripteurs sur la turbidité finale sont en majorité faibles (<10), les paramètres "Ssp, Pui, Dose, K_B , Mte, ILL, et Cp" influent négativement sur la turbidité finale alors que les paramètres "Arg, CM, K_A , COT" présentent des effets positifs, cependant les deux paramètres "T₀, Dms" leurs importances est très faible.

Pour l'aluminium résiduel les importances des paramètres sont plus marquées ; les paramètres "Mte, K_A , Dms, Dose, Arg, K_B et Pui" influent positivement sur l'aluminium résiduel alors que les paramètres "Kao, Ssp, ILL, CM, COT, T₀ et Cp" influent négativement sur l'aluminium résiduel. Il n'est pas possible d'aller au delà de cette interprétation.

VIII.3. MODELE NEURONAL BASE SUR LES COMPOSANTES PRINCIPALES

Dans ce tableau 5.47 sont listées les valeurs propres et le pourcentage de variance de chaque axe de l'analyse en composantes principales de la matrice des données du fichier brut. On retient, pour la modélisation neuronale les cinq premières composantes principales dont les valeurs propres sont supérieures à 1, ce qui correspond à une variance expliquée de plus de 96%.

	Valeur propre	Variance totale %	Valeur propre cumulée	Variance Cumulée
1	16,33	48,04	16,33	48,04
2	11,27	33,15	27,60	81,19
3	2,78	8,18	30,38	89,37
4	1,40	4,13	31,78	93,50
5	1,01	2,90	32,78	96,40
6	0,84	2,50	33,62	98,90
7	0,27	0,80	33,89	99,70
8	0,08	0,24	33,97	99,94
9	0,03	0,06	34,00	100

Tableau 5.47 : Résultats de détermination des valeurs propres et des variances expliquées, modèle de régression neuronale des paramètres Tf et Air

Développement du modèle

Pour déterminer l'architecture du réseau de neurones (le nombre de composantes d'entrée "I" et le nombre de neurones de la couche cachée) on a augmenté le premier "I" de 1 à 5 et le deuxième de 1 à 2×I+1, chaque réseau construit a été ainsi entraîné avec la fonction sigmoïde logistique. Enfin parmi les différentes architectures testées, nous avons retenu celles produisant le plus élevé coefficient de détermination R² sur le jeu de validation.

La figure 5.70 montre la variation de R² moyen en fonction du nombre de composantes principales sur le jeu de validation.

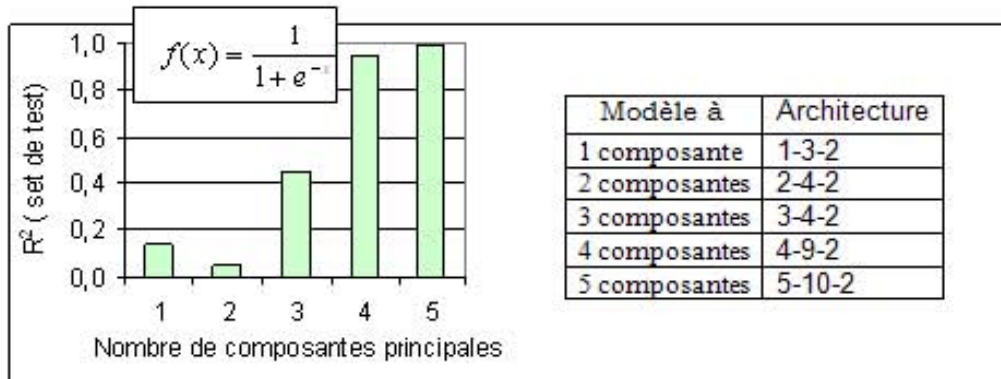


Figure 5.70 : R^2 en fonction du nombre de composantes principales dans la couche d'entrée, modèle de régression neuronale sur composantes principales des paramètres Tf et Alr.

On note que c'est la structure à 5 composantes qui permet d'obtenir la meilleure performance, le tableau 5.48 résume les caractéristiques du modèle retenu.

Modèle	R^2 sur le jeu d'apprentissage			R^2 sur le jeu de validation		
	Tf	Alr	Moyen	Tf	Alr	Moyen
PMC 5-10-2	0,997	0,994	0,9955	0,994	0,991	0,9925

Tableau 5.48 : Résultats de la régression neuronale sur composantes principales des paramètres Alr et Tf

Les valeurs de coefficient de détermination moyen pour les jeux d'identification des poids et de validation de modèle sont très élevées ; Pour le jeu d'apprentissage le R^2 moyen est de 0,9955 alors que pour le jeu de validation le R^2 moyen est de 0,9925 signifiant respectivement que 99,55% de la variabilité du jeu d'identification est expliquée par le modèle et que 99,25% de la variabilité du jeu de validation est validée par le modèle, le pouvoir explicatif et prédictif du modèle est excellent.

Le tableau 5.49, ci-après donne la synthèse de la régression sur l'échantillon d'apprentissage: les valeurs observées, les valeurs prédites, les résidus, les résidus studentisés, et les valeurs de REQM.

Il apparait clairement que les valeurs moyennes, les médianes des résidus et les valeurs de REQM sont très faibles ce qui est le signe de bon ajustement du modèle neuronal.

		Minimum	Maximum	Moyenne	Médiane
Tf [NTU]	Observée	0,5	17	3,09	2,40
	Prévue	0,26	16,56	3,09	2,41
	Résidus	-0,49	0,62	0,00	0,00
	Résidus studentisés	-2,28	2,88	0,00	0,01
	REQM	0,14			
Alr [mgAl/l]	Observée	0,011	0,549	0,101	0,080
	prévue	0,009	0,540	0,101	0,079
	Résidus	-0,028	0,021	-0,000	-0,000
	Résidus studentisés	-4,040	2,931	-0,028	-0,018
	REQM	0,00			

Tableau 5.49 : Synthèse des résidus de la régression neuronale basée sur les composantes principales des paramètres Alr et Tf (set d'apprentissage)

Les résidus studentisés de régression de Tf sont bornés dans l'intervalle [2,3 à 2,9], alors que pour l'aluminium résiduel cette intervalle est plus large [-4,1 à 3], cela est du à la présence de points extrêmes. Ces points, après vérification, ne seront pas supprimés des échantillons d'apprentissage et de validation (leurs résidus sont faibles), ils contribuent par conséquent à la variabilité du phénomène.

L'architecture du modèle retenu est illustrée à la figure 5.71 suivante :

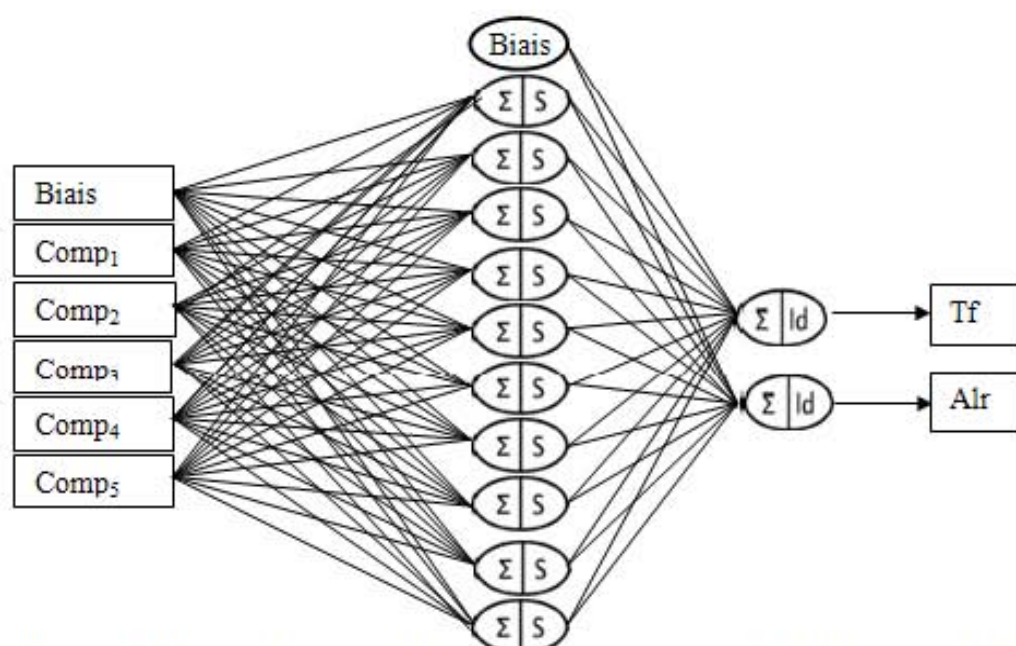


Figure 5.71 : Architecture du perceptron multicouche 5/10/2, modèle de régression neuronale sur composantes principales des paramètres Tf et Alr

En ce qui concerne la normalité des résidus, elle est vérifiée graphiquement, à la figure 5.72 ; les valeurs expérimentales (observées) sont proches des valeurs théoriques pour un seuil de signifiacance $\alpha= 0,05$, quelques points de résidus (points extrêmes), semblent être non incluses dans l'intervalle de confiance de 95% cela explique les valeurs élevées des résidus studentisés.

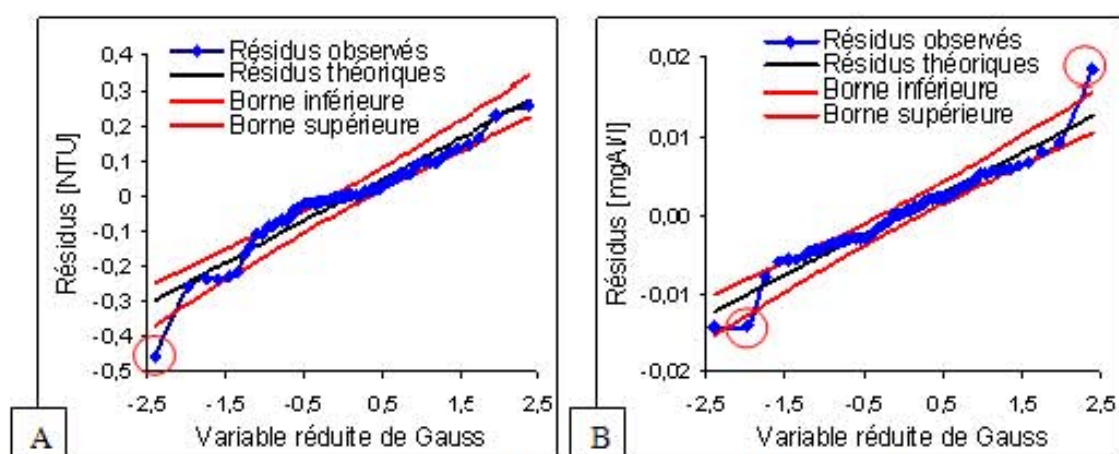


Figure 5.72 : Ajustement à une loi Normale des résidus, modèle de régression neuronale sur composantes principales des paramètres Tf et Alr **A** : Turbidité finale Tf (IC à 95 %, **B** : Aluminium résiduel Alr (IC à 95 %)

Validation du modèle

La validation est faite sur l'échantillon de validation, le tableau 5.50, ci-dessous donne la synthèse de cette validation : les valeurs observées, les valeurs prédites, les résidus, les résidus studentisés et les valeurs de REQM ; il apparaît bien que les valeurs moyennes et les médianes des résidus et les valeurs de REQM sont très faibles ce qui est le signe de bon ajustement du modèle neuronal. Les valeurs des résidus studentisés ne sont pas élevées, elles sont bornées dans l'intervalle [-2,2 à 2,7] comme le montre le tableau 5.50 ci-dessous.

		<i>Minimum</i>	<i>Maximum</i>	<i>Moyenne</i>	<i>Médiane</i>
Tf [NTU]	Observée	0,50	9,60	2,68	2,35
	Prévue	0,26	9,49	2,67	2,36
	Résidus	-0,45	0,26	-0,01	-0,00
	Résidus studentisés	-2,10	1,20	-0,04	-0,00
	REQM	0,1			
Alr [mgAl/l]	Observée	0,026	0,302	0,099	0,084
	prévue	0,024	0,303	0,099	0,083
	Résidus	-0,014	0,019	0,000	0,000
	Résidus studentisés	-2,023	2,645	0,031	0,059
	REQM	0,00			

Tableau 5.50 : Synthèse des résidus de la régression neuronale basée sur les composantes principales des paramètres Alr et Tf (set de validation)

Par ailleurs l'homoscédasticité est vérifiée par le nuage de points des résidus calculés sur l'échantillon de validation, illustrée dans la figure 5.73, quelques points se démarquent du nuage mais restent de faibles résidus.

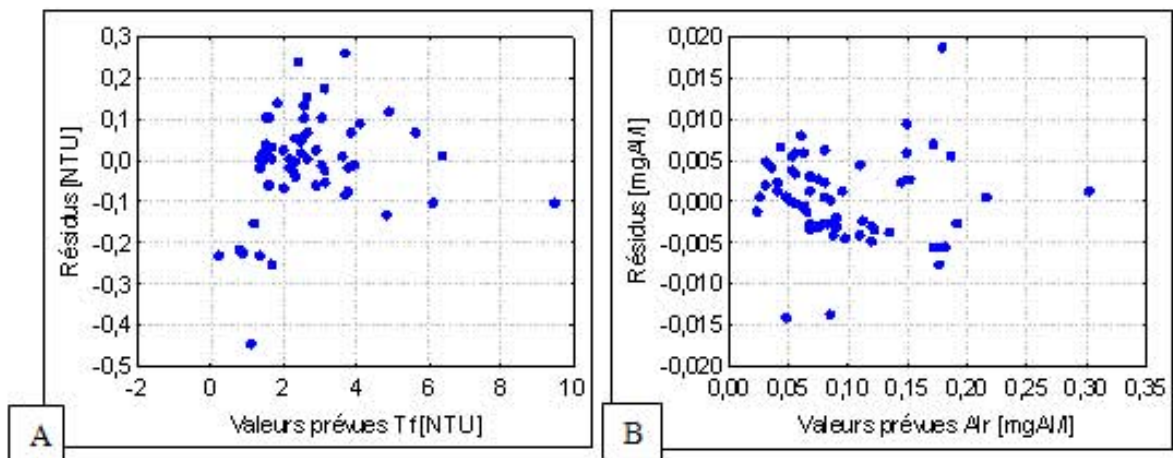


Figure 5.73 : Vérification de l'homoscédasticité, **A** : Turbidité finale Tf, **B** : Aluminium résiduel Alr, modèle de régression neuronale sur composantes principales des paramètres Tf et Alr (set de validation)

Les prédictions de la turbidité finale effectuées sont très bonnes, les valeurs prévues doivent être normalement égales aux valeurs observées. La figure 5.74 croise les valeurs observées avec les valeurs prévues, il apparaît dans cette figure que tous les points sont alignés sur la première bissectrice, cela correspond à une variance expliquée de 99,7% en apprentissage et de 99,4% en validation. L'ajustement est excellent.

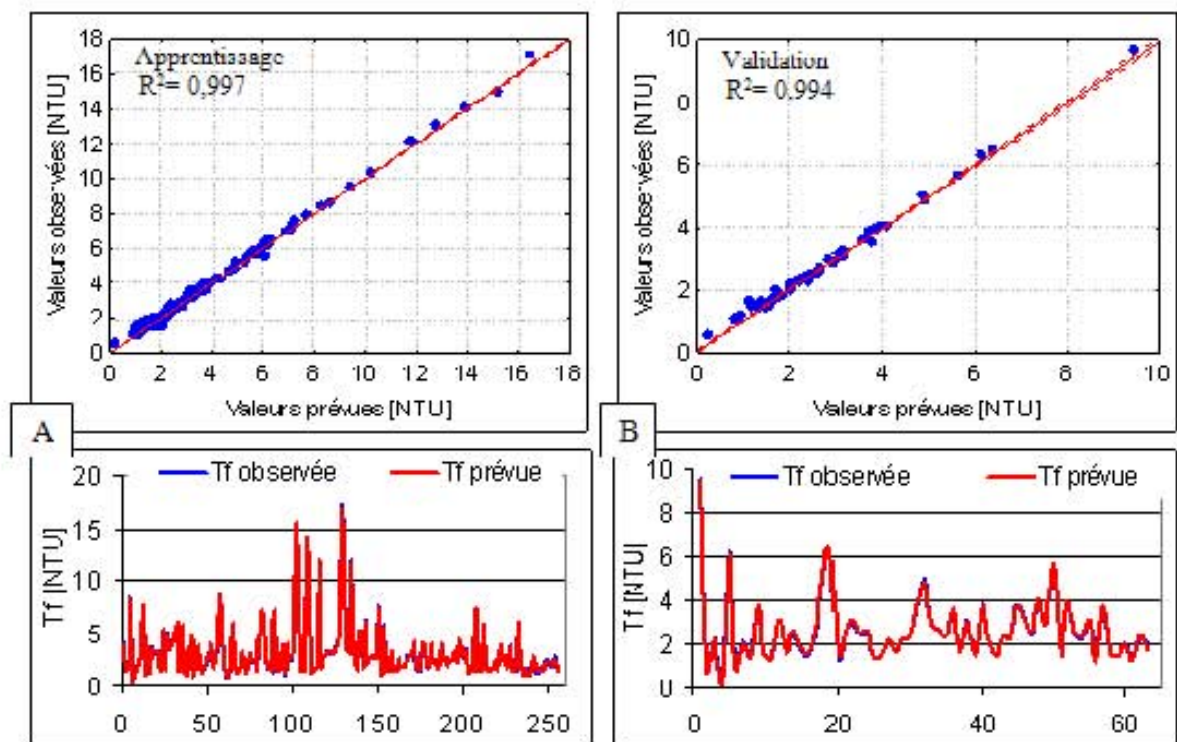


Figure 5.74 : Ajustement des valeurs observées et des valeurs prévues de la turbidité finale Tf, modèle de régression neuronale sur composantes principales des paramètres Tf et Alr **A** : set d'apprentissage, **B** : set de validation, IC à 95%

De même pour la prédiction de l'aluminium résiduel, l'ajustement des valeurs observées (expérimentales) et prévues par le modèle pour les échantillons d'apprentissage et de validation est présenté sur la figure 5.75, l'ajustement est excellent, les points s'alignent sur la première bissectrice. Cela correspond à une variance expliquée de 99,4% en apprentissage et de 99,1%. L'ajustement est excellent.

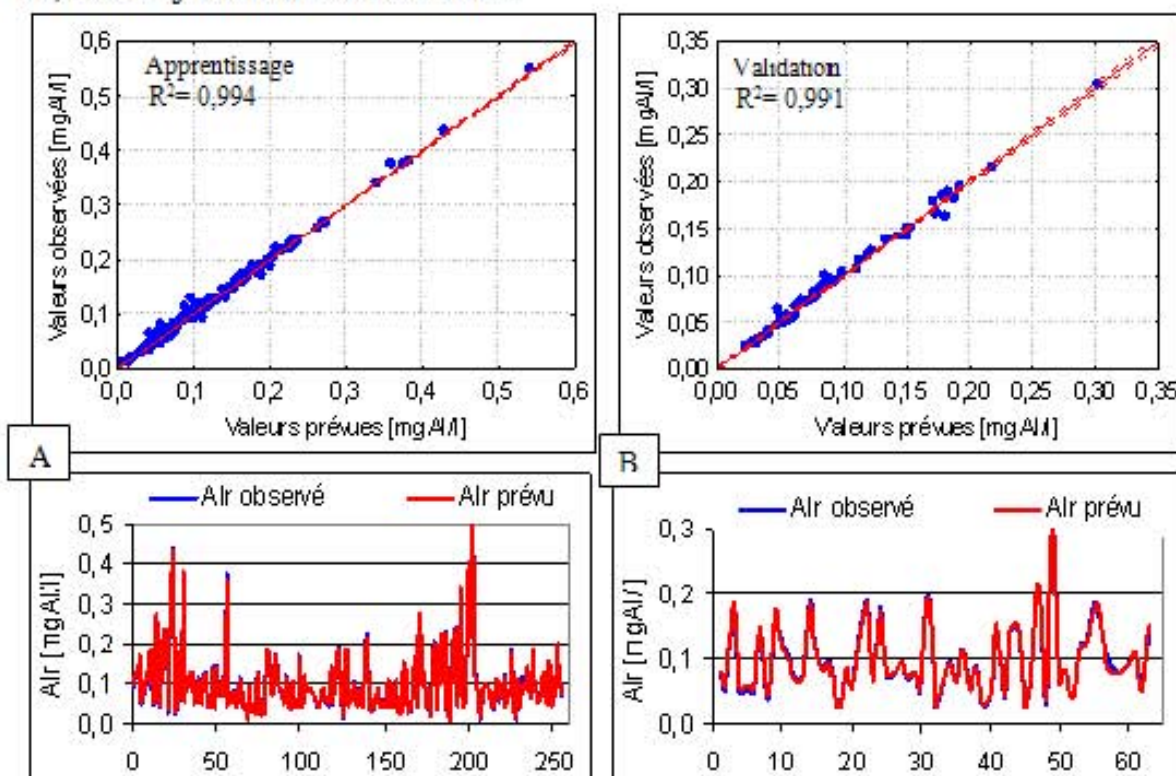


Figure 5.74 : Ajustement des valeurs observées et des valeurs prévues de l'aluminium résiduel, modèle de régression neuronale des paramètres Tf et Alr
A : set d'apprentissage, **B** : set de validation, IC à 95%

IX. CONCLUSION

La modélisation des paramètres de finalité de traitement (l'aluminium résiduel "Alr" et la turbidité finale "Tf") par régression neuronale nous a permis de dégager la conclusion suivante :

Les bonnes performances des modèles montrent la non linéarité de l'évolution de ces deux paramètres. Le modèle neuronal basé sur les composantes principales présente des performances élevées, son pouvoir explicatif et de prédiction est supérieur au modèle neuronal simple cela revient à ce que le modèle neuronal basé sur les composantes principales bénéficie des avantages des composantes principales dans le résumé de l'information en un nombre réduit de variables descriptives et bénéficie aussi de la puissance de la régression neuronale en traitement de la non linéarité. La variance de validation expliquée en cas de régression neuronale basée sur les composantes principales dépasse 99% alors qu'elle atteint juste 76,5% en cas de régression neuronale simple. Ces deux modèles neuronaux à double sorties ajustent pareillement l'aluminium résiduel et la turbidité finale.

X. CONCLUSION DU CHAPITRE

Ce chapitre a été consacré à la modélisation de l'efficacité de la floculation par l'élaboration de modèles de régression de la dose optimale, des constantes cinétiques de floculation et des paramètres de finalité de traitement qui sont la turbidité finale et l'aluminium résiduel.

Plusieurs modèles de régression ont été retenus dans ce travail :

- a. La régression multiple
- b. La régression régularisée
- c. La régression neuronale
- d. La régression neuronale basée sur les composantes principales

Au commencement, la base de données utilisée est issue des résultats de floculation de six types de suspensions préparées à la base des particules solides dispersées du kaolin, de la vase, de la tourbe, de la boue, et de deux mélanges de la vase et de la tourbe. Cette matrice de données a fait l'objet dans le quatrième chapitre d'une analyse multidimensionnelle ce qui a permis à faire le choix des variables explicatives, les plus discriminantes pour la construction des modèles de régression. Les différentes modélisations réalisées sont résumées comme suit :

Modélisation de la dose optimale de la floculation

Pour la régression multiple de la dose optimale, la recherche, pas à pas, par la méthode ascendante nous a permis suivant plusieurs critères de choix, et critères de qualité de prédiction de retenir trois "meilleurs" modèles de régression multiple de la dose optimale.

Les coefficients de régression pour chacune des variables explicatives ont été estimés par la méthode des moindres carrés et la significativité de l'apport des descripteurs a été établie par les tests de Fisher et de Student.

Le premier et le deuxième modèle présentent des performances semblables, la variance expliquée respectivement pour le jeu de données d'apprentissage et de validation du premier modèle sont : 89% pour l'apprentissage et 60% pour la validation.

Le REQM d'apprentissage est égale à 0,66 mgAl/l, tandis que il s'élève à 1,45mgAl/l pour le set de validation, le paramètre le plus influent dans l'équation du premier modèle est le diamètre moyen en surface "Dms", il présente 50,52 % d'influence.

Cependant pour le deuxième modèle, les performances en apprentissage et en validation sont respectivement 89 % pour l'identification du modèle et 66% pour la validation. La valeur de REQM est de 0,67mgAl/l en apprentissage et de 1,52 mgAl/l en validation, dans ce modèle l'influence majeure revient au paramètre : concentration des particules solides "Cp".

Pour le troisième modèle où a été intégré le rapport de la constante de cassage par la constante de formation " K_B/K_A ", 88% de la variabilité du jeu d'apprentissage est expliquée par le modèle alors que seulement 22% de la variabilité de l'échantillon de validation est expliquée.

Ce faible taux (22%) revient principalement à la similarité médiocre entre les deux échantillons d'identification et de validation du paramètre K_B/K_A .

La racine carrée de l'erreur quadratique moyenne en apprentissage est de $REQM = 0,68$ mgAl/l alors qu'en validation le REQM atteint à 3 mgAl/l.

Le calcul de la tolérance qui est l'inverse du facteur d'inflation de variance montre que ces trois modèles de régression multiple encourent un grand risque d'instabilité des coefficients régresseurs par conséquent, la recherche des effets des variables explicatives sur la variable à expliquer et les prédictions doivent être réalisées avec beaucoup d'attention même si R^2 a une valeur élevée.

La régression régularisée de la dose optimale permet des modèles plus stables, Il s'agit de remplacer toutes les variables initiales par un nombre réduit de variables orthogonales qui résument toute l'information. La régression multiple sur composantes principales de la dose optimale a permis, selon différents critères de choix et de qualité de prédiction, un modèle linéaire à six descripteurs qui sont les six premières composantes principales. Il apparaît clairement que le choix du nombre de composantes principales est déterminant.

Le modèle de régression régularisée a permis un coefficient de détermination $R^2 = 0,982$ pour l'échantillon d'apprentissage et $R^2 = 0,95$ pour l'échantillon de validation correspondants à un REQM de 0,00 mgAl/l en apprentissage et de 0,41 mgAl/l en validation.

Le retour aux variables initiales et le calcul des influences des paramètres descripteurs montre que la turbidité initiale et la concentration de particules solides partagent l'influence majeure sur la dose optimale.

Par régression neuronale de la dose optimale nous avons retenu le perceptron multicouche PMC d'architecture suivante 11-8-1 onze neurones de la couche d'entrée, une seule couche cachée de huit neurones et une sortie (la dose optimale), muni de la fonction sigmoïde logistique comme fonction d'activation et de transfert expliquant respectivement 92,2% et 85,2% la variabilité de la dose observée pour l'échantillon d'identification et celui de validation. La racine carrée de l'erreur quadratique moyenne est de 0,56 mgAl/l pour le set d'apprentissage et de 0,773 mgAL/l pour l'échantillon de validation.

Par ailleurs, la modélisation de la dose optimale par régression neuronale basée sur les composantes principales apparaît la plus performante.

Le modèle retenu est moins complexe et présente les meilleurs performances, son architecture est 6-6-1 (six neurones de la couche d'entrée, six neurones de la seule couche cachée et une sortie), muni de la fonction sigmoïde logistique comme fonction d'activation et de transfert permettant un coefficient de détermination $R^2= 0,997$ en apprentissage et $R^2= 0,995$ en validation. Les valeurs de REQM pour l'apprentissage et pour la validation sont très faibles ($REQM < 0,111$ mgAl/l) ce qui est synonyme d'excellent ajustement des données.

Modélisation des constantes cinétiques

La connaissance des constantes de formation K_A et de cassage K_B permet, par application des modèles cinétiques de floculation, la détermination de l'efficacité de floculation. La recherche de modèles de régression permettant de prédire ces constantes cinétiques a fait l'objet d'une section importante dans ce chapitre, il en ressort que :

Le modèle de régression multiple retenu par recherche pas à pas par la méthode ascendante de la constante de cassage K_B est très performant le coefficient de régression est plus de 0,999, le résidu moyen est très faible ($5,684.10^{-14}$ s) et la valeur de REQM est aussi très faible ($1,7.10^{-8}$ s)

La statistique F de Fisher et le test de student montre l'apport réelle des paramètres descripteurs sur la constante de cassage K_B . Cependant le calcul de l'influence des régresseurs sur la constante de cassage K_B montre que la turbidité " T_0 " présente la plus grande influence (36,27 %).

L'inconvénient de ce modèle et la grande colinéarité existante entre les variables explicatives d'où le risque d'instabilité de l'estimation. La régression régularisée permet de remédier à ce problème de colinéarité, les composantes principales descriptives sont orthogonales (pas de multicollinéarité).

Pour ce modèle 99,69% de la variabilité de la constante de cassage est expliquée, le résidu moyen est nul et la valeur de REQM est de l'ordre de $0,18.10^{-6}$ s. Le retour aux variables initiales et le calcul des effets des variables montre que la composante argileuse est d'une influence majeure sur la constante K_B .

La régression multiple de la constante de formation K_A a permis de retenir un modèle linéaire discontinu dont le coefficient de détermination est de $R^2=0,999$. Le calcul de l'influence des paramètres descripteurs montre que la turbidité initiale présente l'influence majeure sur la constante de formation K_A suivi par le COT et en dernier lieu le taux d'argile dans le solide dispersé.

La régression régularisée par un modèle continu de la constante de formation n'a pas été vraiment judicieuse : l'apport réel des variables descripteurs est confirmé mais cela pour des niveaux de significances qui

dépasse parfois 0,05. Le retour aux variables initiales et le calcul des effets des régresseurs montre la grande influence de la composante argileuse sur la variabilité de la constante de formation K_A .

La régression neuronale par un modèle à double sorties des constantes cinétiques de formation et de cassage a permis de retenir le modèle à perceptron multicouche PMC 3-3-2 : trois neurones de la couche d'entrée, trois neurones de l'unique couche cachée et deux sorties, 99,99% de la variabilité des deux constantes cinétiques est expliquée par le modèle.

Le résidu moyen et la racine carrée de l'erreur quadratique moyenne de l'estimation sont très faibles ; pour la constante de formation et celle de cassage le résidu moyen est respectivement de $1,03 \cdot 10^{-9}$ et de $9,16 \cdot 10^{-11}$ s, alors que la valeur REQM est de $0,001 \cdot 10^{-4}$ pour la constante de formation K_A et est égale à $0,039 \cdot 10^{-6}$ s pour la constante de cassage K_B .

Le calcul de l'importance relative des variables d'entrée par la méthode des poids montre que la constante de formation K_A est influencée positivement par la turbidité initiale " T_0 " est négativement par le "COT", cependant la constante de cassage K_B est influencée positivement par le "COT" et négativement par la turbidité initiale " T_0 ", Le taux d'argile présente, pour les deux constantes, une faible influence.

Le modèle de régression neuronale basée sur les composantes principales des constantes cinétiques de formation et de cassage a permis de retenir le perceptron multicouche PMC d'architecture suivante 3/2/2, performant et moins complexe que le modèle neuronal simple, tels que démontrés par le coefficient de détermination moyen R^2 qui est supérieure à 0,999 et par les faibles valeurs du résidu moyen et des valeurs de la racine carrée des erreurs quadratiques moyennes REQM

Résidus moyen sur $K_A = 2,28 \cdot 10^{-9}$, REQM sur $K_A = 0,0007 \cdot 10^{-4}$

Résidus moyen sur $K_B = -5,70 \cdot 10^{-10}$ s, REQM sur $K_B = 0,099 \cdot 10^{-6}$ s

Ces différents modèles de régression des constantes cinétiques K_A et K_B ont montré une grande performance d'ajustement, mais faute de faible déterminations expérimentales, la validation de ces régressions n'a pas eu lieu, alors pour une meilleur utilisation de ces modèles, il est nécessaire de valider ses modèles sur d'autres données.

Modélisation de l'aluminium résiduel "Alr" et de la turbidité finale "Tf"

La modélisation des paramètres de finalité de traitement, l'aluminium résiduel et la turbidité finale par régression neuronale nous a conduit à retenir le perceptron multicouche 14/17/2 : 14 neurones de la couche d'entrée et qui sont les principaux caractéristiques du solide dispersé plus la

puissance d'agitation lente et la dose du coagulant injectée, une seule couche cachée munie 17 neurones sigmoïdaux et deux neurones (Alr et Tf) formant la couche de sortie.

Le coefficient de détermination R^2 en apprentissage est de 0,84 signifiant que 84% de la variabilité des deux paramètres est expliquée par le modèle. Les résidus moyens sont nulles, la racine carrée de l'erreur quadratique moyenne de la turbidité finale est de $REQM = 0,75$ NTU, alors que pour l'aluminium résiduel le $REQM$ est égale à 0,03 mgAl/l.

La validation du modèle sur un deuxième échantillon permet un coefficient de détermination de 0,765 interprété par des résidus moyens très faibles (-0,07 NTU pour la turbidité et -0,003 mgAl/l pour l'aluminium résiduel). Les valeurs de $REQM$ sont aussi très faibles : 0,71 NTU pour la turbidité finale et 0,03 mgAl/l pour l'aluminium résiduel, ce qui est le signe de très bonne validation du modèle neuronal.

La modélisation par réseau de neurones basé sur les composantes principales de l'aluminium résiduel et de la turbidité finale a permis de retenir le modèle PMC 5/10/2 : Cinq neurones de la couche d'entrée et qui sont les cinq premières composantes principales retenues suite à l'analyse en composantes principales, dix neurones de l'unique couche cachée et deux neurones de la couche de sortie, ce modèle est moins complexe et plus performant que le modèle de régression neuronale simple.

99,55% de la variabilité des deux paramètres en échantillon d'apprentissage est expliquée par le modèle, les résidus moyens sont nulles, le $REQM$ de la turbidité finale est de 0,14 NTU alors que le $REQM$ de l'aluminium résiduel est nulle.

Le modèle a été validé sur un deuxième échantillon, 99,25% de la variabilité du jeu de données de validation est expliquée par le modèle. Les résidus moyens sont très faibles (-0,01 NTU pour la turbidité finale et -0,0009 mgAl/l pour l'aluminium résiduel), les valeurs de $REQM$ sont aussi très faibles, 0,10 NTU pour la turbidité finale et 0,009 mgAl/l pour l'aluminium résiduel, ce qui est le signe d'excellent ajustement des données de validation.

A travers ces différentes modélisations réalisées ci-dessus il en ressort que :

- Les modèles paramétriques de régression multiple sont simples et présentent des performances très satisfaisantes mais la grande colinéarité relevée entre les variables descriptives risque l'instabilité de ces modèles.
- L'utilisation des modèles de régression régularisée permet d'augmenter la performance de prédiction et la robustesse des modèles par l'amélioration de la qualité des données d'entrée (réduction de la

complexité et de la multicolinéarité). Cependant la régression régularisée nécessite le retour aux variables initiales pour déterminer l'influence des variables explicatives.

- Les modèles non paramétriques de régression neuronale à perceptron multicouche sont plus performants que les modèles de régression multiple, ces modèles s'adaptent bien à la non linéarité du processus, et permettent de minimiser le résidu moyen et la racine carrée de l'erreur quadratique moyenne.
- Les modèles neuronaux basés sur les composantes principales permettent de réduire la colinéarité des données, de minimiser la taille et la structure du perceptron multicouche et permettent d'améliorer davantage la qualité d'ajustement et de prédiction.

CONCLUSION GENERALE

Bien que plusieurs études ont amplement discuté la modélisation des paramètres de floculation, la problématique de remplacer le jar test, la recherche de la dose optimale et la prédiction de la qualité des eaux traitées, la modélisation de la floculation continue d'être de plus en plus étudiée et raffinée surtout avec l'émergence des techniques numériques de modélisation en aide au diagnostic, à l'automatisation, et à la supervision des stations de production des eaux.

Le travail établi avait pour objectif la modélisation de l'efficacité de la floculation, il a été consacré à l'élaboration de modèles de régression de la dose optimale, de modèles de régression des constantes cinétiques de floculation et de modèles de régression des paramètres de finalité de traitement : "l'aluminium résiduel et la turbidité finale". Pour cela, l'étude consistait : en premier temps aux réalisations expérimentales permettant la collecte des données et résultats sur le procédé de coagulation/floculation des suspensions de six types de particules solides dispersées. En second temps l'utilisation des techniques statistiques exploratoires dans la préparation des données et le choix des variables explicatives des modèles et l'établissement des modèles de régression multiple, des modèles de régression neuronale, et des modèles de régression combinés : la régression régularisée et la régression neuronale basée sur les composantes principales.

Dans le premier chapitre, une synthèse de la théorie, la pratique et la modélisation de la floculation des suspensions colloïdales a été réalisée. Dans un premier résultat, cette synthèse a montré la complexité des mécanismes de coagulation et de floculation des suspensions colloïdales et nous a permis de faire un bilan de l'état actuel de la pratique des connaissances scientifiques dans ce domaine.

Par ailleurs, dans l'optique de créer des modèles d'inférence statistiques, nous avons fait le point sur les techniques, les appareils et les méthodes utilisées pour le contrôle du procédé de coagulation et de floculation.

La recherche bibliographique a été aussi consacrée, d'une part à montrer l'importance de l'application de l'analyse exploratoire dans l'analyse, la prédiction et l'interprétation des paramètres de floculation, et d'autre part à la présentation des modèles statistiques de prédiction de la dose optimale, et des paramètres de floculation par régression multiple et par régression non paramétrique par réseau de neurones artificiels.

Les différents modèles établis s'accordent tous sur l'importance de la disponibilité des données en bon nombre et en bonne qualité, du choix des paramètres descripteurs et du type de régression dans l'apprentissage et la validation des modèles.

Le deuxième chapitre nous a permis de faire le tour d'horizon des principes théoriques nécessaires à la compréhension de l'analyse exploratoire et de la modélisation mathématique statistique des données ;

Après une brève introduction où nous rappelons l'importance de l'analyse exploratoire multidimensionnelle comme première phase de la modélisation, nous avons présenté le principe de l'Analyse en Composantes Principales "ACP" et de l'Analyse de Classification Hiérarchique. L'ACP permet de réduire la dimension d'une matrice de données, elle transforme un premier jeu de données en un second jeu de plus petite dimension composé de nouvelles variables qui sont des combinaisons linéaires des variables originelles et qui décrivent correctement les tendances principales. Alors que l'Analyse de Classification Hiérarchique permet de regrouper les variables en classes de tel sorte que les individus d'une même classe soient le plus semblables possible et que ces classes soient bien séparées, cette technique de classification peut compléter et nuancer les résultats de l'Analyse en Composantes Principales.

La grande partie dans ce chapitre a concerné la définition des différents modèles mathématiques, en particulier le modèle de régression multiple et le modèle de régression neuronale. Nous avons exposé la démarche de sélection du meilleur modèle de régression multiple, les procédures et critères de choix, la qualité de prédiction, l'influence des paramètres descripteurs et enfin la robustesse de modèle.

La régression non paramétrique par réseau de neurones a fait l'objet de la dernière partie de cette revue bibliographique, elle a porté sur les définitions essentielles relatives aux réseaux de neurones, la présentation de la structure du perceptron multicouche et de sa capacité de modélisation de fonctions non linéaires. Nous avons décrit la méthodologie utilisée pour la construction du modèle neuronal et les techniques de détermination de la contribution des entrées.

Dans ce 2^{ième} chapitre nous avons mis l'accent aussi sur "la régression sur composantes principales" communément appelée régression régularisée et sur la régression neuronale basée sur les composantes principales, ces utilisations mixtes permettent de réduire la complexité des modèles, et la colinéarité entre les variables descriptives et permettent des modèles parcimonieux performants

Le troisième chapitre avait deux objectifs, le 1^{er} objectif est de créer une base de données relative à la floculation des suspensions synthétiques alors que le 2^{ième} objectif est de contribuer à la compréhension des effets des conditions opératoires et des caractéristiques du solide dispersé en particulier, sa nature, sa distribution granulométrique, son diamètre et sa surface spécifique et leurs interactions sur la cinétique, le mécanisme et la performance de la floculation, tous cela pour un meilleur discernement de la modélisation des paramètres de floculation.

A cet effet, une campagne d'essais expérimentaux a été conduite, Il s'agit des essais de floculation de suspensions synthétiques de six types de particules solides du kaolin, de la vase de barrage, de la tourbe végétale, de la boue de station d'épuration et de deux mélanges de la vase et de la tourbe. Il ressort comme principales de ce chapitre que :

- La présence de la matière organique dans le solide dispersé à des taux inférieure à 25 % est d'une influence minime, alors que l'effet négatif inhibiteur sur la performance de la floculation est marqué pour des taux de matière organique plus élevés (>25 %), pour ces suspensions organiques l'aluminium est doublement consommé ; à la formation du $\text{Al}(\text{OH})_3$, et à la complexation du cation Al^{3+} sur la matière organique.
- Le mécanisme de coagulation des suspensions minérales est l'adsorption et la neutralisation de charge alors que pour les suspensions organiques c'est la complexation, l'adsorption et la neutralisation de charge, il s'avère aussi que pour les faibles dosages en coagulant c'est la complexation qui est majoritaire alors que pour des dosages élevés en coagulant c'est l'adsorption et la neutralisation de la charge qui dominent.
- L'ajout de la matière minérale argileuse stimule les mécanismes de déstabilisation, augmente les sites d'adsorption, retarde l'état stationnaire permettant de promouvoir davantage la floculation, elle renforce la structure de floc et augmente sa résistance au cassage ce qui accroît le seuil optimal d'agitation.

- La formation, le cassage et la résistance de floc sont liés à l'état stationnaire de la floculation et sont tributaires de l'interaction entre plusieurs paramètres ; la matière organique, la matière minérale, le dosage en coagulant, les espèces d'aluminium présentes, le mécanisme de floculation, les diamètres, la densité, la distribution de taille des particules primaires et des floes et des propriétés surfaciques des particules.
- L'étude de la cinétique de la floculation montre que l'état stationnaire est influencée par la nature et les caractéristiques du solide dispersé, il apparaît que la constante de cassage K_B est élevée pour les suspensions organiques ce qui met en évidence que les floes formés des colloïdes organiques sont moins résistants à l'érosion et à la fragmentation que ceux formés des colloïdes minéraux. La constante de formation montre deux évolutions différentes en fonction du taux de matière organique, pour de faibles taux, la constante de formation K_A augmente avec l'augmentation de la matière organique et pour des taux élevés, la constante de formation diminue, ce qui signifie que la présence de la matière organique à de faibles concentrations améliore la formation du floc, par complexation et pontage, alors que cette formation est inhibée pour des taux élevés de matière organique.

L'objectif du quatrième chapitre, était l'analyse exploratoire des résultats expérimentaux en vue de la détermination des variables pertinentes qui peuvent être intégrées dans les modèles de régression. A cet effet une attention particulière a été donnée à la nature de nos données expérimentales et aux relations qui lient entre elles. Trois phases d'analyse statistique complémentaire ont été mises en œuvre.

Une première phase exploratoire monovariée, des statistiques descriptives et des boîtes de dispersion "box plot" des données, qui a permis de constater la présence de quelques points extrêmes et le comportement démarqué des particules de la tourbe par rapport aux autres solides.

Une deuxième phase d'analyse bivariée est réalisée, où la nature des relations liant les différentes variables a été étudiée par le biais du coefficient de corrélation de Bravais Pearson, puis l'observation des nuages de points, ce qui a permis de constater l'absence de liaisons linéaires entre la dose optimale, la constante de formation et la constante de cassage avec les différentes caractéristiques du solide dispersé et les conditions opératoires.

Les éléments de la statistique descriptive univariée et bivariée ne donnent aucune information lorsque les variables sont considérées simultanément, alors une troisième phase d'analyse multidimensionnelle par composantes principales et par classification hiérarchique a été faite, l'examen des dendrogrammes de classification hiérarchiques et l'analyse des contributions

aux axes et des représentations graphiques des variables et individus montrent que les caractéristiques du solide dispersé peuvent s'ordonner selon deux indices principaux ; le premier concerne la matière organique et la distribution granulométrique et le deuxième est fonction de la nature de la matière minérale argileuse.

Les résultats de l'analyse multidimensionnelle ont permis de retenir parmi toutes les variables descriptives du solide dispersé et des paramètres de floculation, celles les plus pertinentes et qui peuvent être intégrées dans les modèles de régression.

Nous avons ensuite, dans le chapitre cinq, procédé à la phase d'inférence statistique de modélisation de la dose optimale, des constantes cinétiques de formation " K_A " et de cassage " K_B " et des paramètres de finalité de traitement ; l'aluminium résiduel " AL_r " et la turbidité finale " T_f "

Quatre types de modèles ont été élaborés

- Modèle de régression multiple
- Modèle de régression régularisée
- Modèle de régression neuronale |
- Modèle de régression neuronale basée sur les composantes principales

Pour la dose optimale, trois régressions multiples ont été établies. La population a été scindée en deux échantillons, échantillon d'apprentissage et échantillon de validation, la procédure ascendente pas à pas a permis la sélection des variables d'entrée (les régresseurs). Les critères de qualité appliqués à ces modèles montrent le bon ajustement, démontré par les valeurs suffisamment élevées des coefficients de détermination et les faibles valeurs des résidus moyens et des racines carrées des erreurs quadratiques moyennes REQM. L'intégration des constantes cinétique dans le modèle de prédiction de la dose optimale est une application très intéressante et qui a donné des résultats acceptables, cependant un modèle performant nécessite une bonne base de données.

Ces modèles de régression multiple sont classiques et simples et permettent des ajustements satisfaisants, cependant la grande colinéarité entre les variables explicatives risque l'instabilité des coefficients de régression et par conséquent affecte la robustesse des modèles. Le calcul des influences des régresseurs montre que le diamètre moyen en surface " D_{ms} ", le carbone organique total " COT ", la concentration de particules solides " C_p ", et la turbidité initiale " T_0 " sont les paramètres les plus influents sur la prédiction de la dose optimale.

Le modèle de régression régularisée de la dose optimale est plus stable, moins complexe et plus performant que les trois modèles de régression multiple, la multicollinéarité est réduite, le coefficient de détermination R^2 est élevé, les résidus moyens et les valeurs de REQM sont très faibles. Il

apparaît clairement que le choix du nombre de composantes principales est déterminant. Par ailleurs le retour aux variables initiales montre que la composante argileuse influe grandement sur la qualité de prédiction de la dose optimale.

La régression neuronale de la dose optimale a révélé des performances très élevées ce qui montre la non linéarité du processus. L'avantage de ce modèle non-paramétrique est sa souplesse du fait qu'il n'y ait pas de paramètres à estimer, cependant, son inconvénient est qu'il est la difficulté d'avoir une fonction mathématique conventionnelle.

La régression neuronale basée sur les composantes principales de la dose optimale rassemble les avantages des deux techniques ; les variables descriptives de la couche d'entrée sont bien résumées en un nombre réduit grâce à l'analyse en composantes principales, et le perceptron multicouche s'adapte bien à la non linéarité des relations entre la variable à expliquer et les variables explicatives.

Ce modèle combiné apparaît moins complexe que le modèle neuronal simple et présente de meilleures performances de prédiction, les coefficients de détermination sont élevés, les résidus moyens sont nuls et les racines carrées des erreurs quadratiques moyennes REQM sont très faibles.

Les constantes cinétiques de formation et de cassage ont fait l'objet de modélisation dans ce chapitre. Suivant la méthode ascendante de recherche pas à pas de modèle de régression multiple, il apparaît que la constante de cassage peut être modélisée par régression multiple linéaire continue, et la constante de formation par un modèle de régression multiple linéaire discontinue, ces modèles, et à cause de la grande colinéarité entre les régresseurs, encourent un grand risque d'instabilité de l'estimation.

La régression régularisée des constantes cinétiques présente l'avantage de remédier aux problèmes de colinéarité et permet alors des modèles plus stables. Le modèle de régression neuronale à double sorties muni de la fonction d'activation sigmoïde logistique permet un très bon ajustement des constantes cinétiques de formation et de cassage. Par ailleurs la combinaison des composantes principales et le modèle à perceptron multicouche offre le modèle le moins complexe et le plus performant.

Les résultats de l'étude des effets des variables descriptives montrent que la turbidité initiale, la composante argileuse, le taux de matière organique et les diamètres des particules primaires influent grandement sur les estimations des constantes cinétiques.

Les modèles établis de prédiction des constantes cinétiques montrent des performances d'apprentissage très élevées, il est recommandé alors de les valider sur d'autres échantillons de données.

La modélisation neuronale à double sorties ajustent pareillement les deux paramètres de finalité de traitement : l'aluminium résiduel "Alr" et la turbidité finale "Tf" et montre la grande performance du modèle neuronal basé sur les composantes principales par rapport au modèle neuronal simple. Les effets des paramètres descripteurs sur la turbidité finale sont faibles et sans grande différence, alors que pour l'aluminium résiduel, l'importance des paramètres est plus marquée.

En résumé et au vu de tous ces résultats, la comparaison entre les différentes régressions montre que :

- les modèles de régression multiple sont simples et donnent des résultats satisfaisants, mais encourent un grand risque d'instabilité à cause de la colinéarité existante entre les variables explicatives, un modèle de régression multiple performant nécessite des données en bon nombre, en bonne qualité et en bonne représentativité.
- Dans les modèles de régression régularisée, les variables d'entrée sont remplacées par les composantes principales, la colinéarité est alors réduite et le modèle devient moins complexe et plus stable et les prévisions donc plus crédibles.
- Les modèles non paramétriques à la base de perceptron multicouche sont très prometteurs à la prédiction des paramètres de floculation, cela est dû principalement à leur capacité de simulation des relations non linéaires complexes entre la variable passive et les variables explicatives. La modélisation paramétrique par réseau de neurones ne nécessite pas la connaissance explicite sur le procédé à modéliser, elle est basée seulement sur des données expérimentales, donc la qualité des données décide de l'exactitude et la robustesse du modèle. Cependant les modèles de régression neuronale basée sur les composantes principales apparaissent moins complexes et plus performants que les modèles neuronaux simples.

Les résultats obtenus dans le cadre de cette thèse nous conduisent à proposer les perspectives suivantes :

- Les modèles de régression établis sont inhérents aux données de base d'élaboration, il est alors recommandé de les recalculer sur d'autres données.
- Optimiser les résultats expérimentaux en utilisant les techniques des plans d'expériences et des surfaces de réponses.
- Appliquer d'autres techniques en modélisation tel que : Réseaux de Neurones Hybrides, Réseaux neuroflous, k-voisins les plus proches (VPP), La méthodologie de classification LAMDA (Learning Algorithm for Multivariate Data Analysis), L'analyse en composantes principales non linéaires ...etc

REFERENCES BIBLIOGRAPHIQUES

- ABDUL-WAHAB A.S., BAKHEIT C. S., AL-ALAWI S. M.(2005)** "Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations" Environmental Modelling & Software, 20, pp.1263-1271
- ADACHI Y., TANAKA Y. (1997)** "Settling velocity of an aluminium kaolinite floc water" Water Research, vol. 31, N°3, pp. 449-454
- ADELOYE A.(2009)** "The relative utility of regression and artificial neural networks models for rapidly predicting the capacity of water supply reservoirs" Environmental Modelling & Software, 24, pp.1233-1240
- ACHOUR S., (2000)** Incidence des procédés de chloration, de floculation, et d'adsorption sur l'évolution de composés organiques et minéraux des eaux naturelles, Thèse de doctorat d'état Univ. Mouloud MAMMERI- TIZI-OUZOU
- AHMAD A.L., WONG S.S., TENG T.T., ZUHAIRI A. (2007)** "Optimization of coagulation-flocculation process for pulp and paper mill effluent by response surface methodological analysis" Journal of Hazardous Materials, 145, pp.62-168
- AKRATOS C., PAPASPYROS J., N.,E., TSIHRINTZIS V., A. (2008)** "An artificial neural network model and design equations for BOD and COD removal prediction in horizontal subsurface flow constructed wetlands" Chemical Engineering Journal 143, pp.96-110
- AL-ABRI M., HILAL N. (2008)** "Artificial neural network simulation of combined humic substance coagulation and membrane filtration" Chemical Engineering Journal 141, pp. 27-34
- AL-ABRI M., AL ANEZI K., DAKHEEL A., HILAL N. (2010)** "Humic substance coagulation: Artificial neural network simulation" Desalination, 253, pp.153-157
- AL-ALAWI S. M., ABDUL-WAHAB S. A, BAKHEIT C. S. (2008)** "Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone Environmental" Modelling & Software 23, pp.396-403.
- AMMAR M. Y. (2007)** Mise en œuvre de réseaux de neurones pour la modélisation de cinétiques réactionnelles en vue de la transposition batch/continu. Thèse de Doctorat de l'I.N.P., Toulouse 177p
- AMIRTHARAJAH A., MILLS K. M. (1982)** "Rapid-mix design for mechanisms of alum coagulation" J. Am. water works Assoc., 74, 4, pp. 210-216
- ANOUZLA A., ABROUKI Y., SOUABI S., SAFI M., RHBAL H. (2009)** "Colour and COD removal of disperse dye solution by a novel coagulant: Application of statistical design for the optimization and regression analysis" Journal of Hazardous Materials, 166, pp.1302-1306
- ARFANDI M. (1979)** Mise au point d'une méthode de contrôle en continue des procédés de floculation décantation. Thèse de docteur ingénieur USTL Montpellier 175p

Références bibliographiques

- ARGAMAN Y., KAUFFMAN W.J. (1970)** "Turbulence and flocculation" Journal of the sanitary Engineering Division, Vol. 92, N° SA2, pp. 223 – 241
- AUBERT GUY (1978)** Méthodes d'analyses des sols CNDP, CRDP Marseille 190 p
- AYOKO G. A., SINGH K., BALEREA S., KOKOT S. (2007)** "Exploratory multivariate modeling and prediction of the physico-chemical properties of surface water and groundwater" Journal of Hydrology, 336, pp.115– 124
- BACHE D.H., JOHNSON C., MCGILLIGAN J.F., RASOOL E., (1997)** "A conceptual view of floc structure in the sweep floc domain" Water Science Technology, Vol.36, 4, pp. 49–56.
- BACHE D.H., RASOOL E., MOFFAT D., MCGILLIGAN F.J. (1999)** "On the strength and character of alumino-humic flocs" Water Science Technology, Vol.40, 9, pp. 81-88
- BARON A., BERNARD A. (1971)** L'analyse minéralogique des sols argileux Editions Eyrolles 93 p
- BAXTER, C.W., ZHANG, Q., STANLEY, S.J., SHARIFF, R., TUPAS, R.-R.T., STARK, H.L. (2001)** "Drinking water quality and treatment: the use of artificial neural networks". Canadian Journal of Civil Engineering, 28 (Suppl. 1), pp.26–35.
- BAZER - BACHI A., PUECH - COSTE E., BEN AIM R. (1990)** "Modélisation mathématique du taux de coagulant dans une station de traitement d'eau" Revue des sciences de l'eau, 3 pp. 377 – 397
- BECHAC J.P., BOUTIN P., MERCIER B. (1983)** Traitement des eaux usées Editions Eyrolles, 281 p
- BEECROFT J.R.D., KOETHER M.C., VANLOON G.W. (1995)** "The chemical nature of precipitates formed in solutions of partially neutralised aluminium sulphate" Water Research, Vol. 29, N°6, pp. 1461 – 1464
- BELANGER M., EJABI N., CAISSIE D., ASHKAR F., RIBI J. M. (2005)** "Estimation de la température de l'eau en rivière en utilisant les réseaux de neurones et la régression linéaire multiple" Revue des sciences de l'eau, 18 (3) pp.403-421
- BELICZYNSKI B. (2005)** "Certain comments on data preparation for neural networks based modeling" Proc.ICANN'05 Coimbra, Springer, pp. 5-8.
- BENGRABINE K., MARHABA T. F. (2003)** "Using principal component analysis to monitor spatial and temporal changes in water quality" Journal of Hazardous Materials, B100 pp.179–195
- BENSCHOTEN V.J.E., EDZWALD J.K. (1990)** "Chemical aspects of coagulation using aluminum salts—II. Coagulation of fulvic acid using alum and polyaluminum chloride" Water Research, Vol. 24 (12), pp. 1527-1535
- BESSE P., BACCINI A. (2005a)** Data mining I Exploration Statistique Publication du laboratoire de statistique et Probabilités, Université Paul Sabatier Toulouse, 142p

Références bibliographiques

- BESSE P. (2005b)** Data Mining. II. Modélisation statistique et apprentissage. Publications du laboratoire de statistique et probabilités. Université Paul Sabatier Toulouse 103 p
- BESSE P. (2009)** Apprentissage Statistique & Data mining Publication du Laboratoire de Statistique et Probabilités, Institut de Mathématiques de Toulouse, Institut National des Sciences Appliquées Toulouse. 124 p
- BHATIA S., OTHMAN Z., AHMAD A. (2007)** "Coagulation–flocculation process for POME treatment using Moringa oleifera seeds extract: Optimization studies" Chemical Engineering Journal, 133, pp.205–212
- BOUROCHE J., SAPORTA G. (1980)** l'analyse des données. Presses Universitaires de France 127p.
- BOUTARFA A. (2006)** Reconnaissance de formes 3D par approche neuronale associant la transformée de Hough en robotique mobile application à la robotique. Thèse de doctorat Es-science, Univ. de Batna. 140p
- BOUYER D. (2002)** Analyse expérimentale de la floculation : influence de l'hydrodynamique sur les phénomènes d'agglomération et de rupture Thèse de doctorat INSA 368p.
- BRATBY J., MILLER M. W., MARAIS G. R. (1977)** "Design of flocculation from batch test data" Water S.A., Vol. 3, N°4, pp 173 – 182
- BRION G. M., NEELAKANTAN T.R., LINGIREDDY S. (2002)** "A neural-network-based classification scheme for sorting sources and ages of fecal contamination in water" Water Research, 36, pp. 3765–3774
- BRODARD E., LEPRINCE A., FIESSINGER F. (1986)** "Modélisation stochastique d'une usine de traitement d'eau potable" Eau et informatique, pp.34- 42
- CABASSUD M., DELGRANGE V.N., CABASSUD C., DURAND B. L., LAINE J.M. (2002)** "Neural networks: a tool to improve UF plant productivity" Desalination, 145, pp.223-231
- CALVET R., TERCE M. (1977)** "Adsorption des herbicides par les constituants minéraux du sol" Colloque national Protection des eaux souterraines captées pour l'alimentation humaine, pp.75-88, ORLEANS-LA-SOURCE-France.
- ÇAMDEVYREN H. , DEMYR N., KANIK A., KESKYN S. (2005)** "Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs" Ecological Modelling, 181, pp. 581–589
- CARBONNIER F, NAULEAU F, GOURITEN Y. (1997)** "Coagulation mixte" T.S.M., N°7-8, pp. 3-8
- CHAVES P., KOJIRI T. (2007)** "Deriving reservoir operational strategies considering water quantity and quality objectives by stochastic fuzzy neural networks" Advances in Water Resources, 30, pp.1329–1341
- CHEMAT S. (1999)** La matière organique naturelle dissoute des eaux de surface, fractionnement, caractérisation et dégradation. Thèse de magister USTB

Références bibliographiques

- CHEN-Y.C, JOSEPH F.A, MARCUS I.B, (1997)** "Direct measurement of structures in mixing jar using piv" Journal of Environmental Engineering, Vol.123, N°2, pp.115-124
- CHENG W. P., CHI F. H., LI C. C., YU R. F., (2008)** "A study on the removal of organic substances from low-turbidity and low-alkalinity water with metal-polysilicate coagulants" Colloids and Surfaces A: Physicochem. Eng. Aspects, Vol.312, Issues 2-3, 15, pp.238-244
- COLIN J.L, DUTOT A. L., BABLON G., VIE LE SAGE R. (1986)** "Application de l'analyse des correspondances à des résultats d'essais de traitement d'eaux de surface" Water Research, Vol.20, N°6, pp.675-684
- COLLINS M.R, AMY G.L., STEELINK C. (1986)** "Molecular weight distribution, carboxylic acidity, and humid substances content of aquatic organic mater: Implications for removal water during treatment" Environmental sciences technology, Vol. 20 N°10, pp 1028 - 1032
- CONLIN J, PEEL C, MONTAGUE GA. (1997)** "Modelling pressure drop in water treatment". Artif. Intell Eng. 11, pp.393-400
- CONFAIS J., LE GUEN M. (2007)** Premiers pas en régression linéaire avec SAS, Documents de Travail du Centre d'Economie de la Sorbonne, CNRS, Université paris I, 145 p
- CORNILLON P.A, MATZNER L.É. (2007)** Régression Théorie et applications, les presses de l'Imprimerie BARNÉOUD, Springer-Verlag France, Paris, p303
- COUFORT C, BOUYER D., LINÉ A., HAUT B. (2007)** "Modelling of flocculation using a population balance equation" Chemical Engineering and Processing, 46(12), pp.1264 - 1273
- COUFORT C., DUMAS C., BOUYER D., LINÉ A. (2008)** "Analysis of floc size distributions in a mixing tank" Chemical Engineering and Processing, 47(3), pp. 287-294
- CRITCHLEY RF, SMITH E., PETTIT P. (1990)** "Automatic coagulation control at water treatment plants in the North-West region of England" Water and Environment Journal, 4(6), pp.535-543.
- CYR S., CHABOT R., SIMARD H. (1990)** "Centrale de contrôle de la coagulation (ccc) utilisant un détecteur de courant électrique" sciences et techniques de l'eau, Vol. 23, N°1, pp.75-80
- DAGNELIE P. (1982)** Analyse statistique à plusieurs variables les presses agronomiques de Gembloux ASBL 362p.
- DALY R., LEEUWEN J. V., HOLMES M.(2007)** Modelling Coagulation to Maximise Removal of Organic Matter - A Pilot Plant and Laboratory Based Study Chemical Dose Prediction Research Report N° 36 177 p Cooperative Research Centre for Water Quality and Treatment, AUSTRALIA.
- DAVALO E., NAIM P. (1993)** Des réseaux de neurones Eyrolles, 140p.

Références bibliographiques

- DELLANA S. A, WEST D. (2009)** "Predictive modeling for wastewater applications: Linear and nonlinear approaches" Environmental Modelling & Software, 24 pp.96-106
- DENTEL S. K. (1988)** Application of precipitation - charge neutralisation model of coagulation" Environmental science technology Vol. 2 N°7, pp 825 - 832
- DENTEL S. K., THOMAS A. V., KINGERY K. M.(1989)** "Evaluation of the streaming current detector—I. Use in jar tests" Water Research, Vol.23, 4, pp.413-421
- DENTEL S.K. (1995)** "Use of the streaming current detector in coagulation monitoring and control. Journal of Water Science Research and Technology - Aqua 44(2) pp.70-79.
- DESJARDINS R. (1988)** Le traitement des eaux Editions de l'école polytechnique de Montréal 366 p
- DESJARDINS R., LENGU K., DIONGUE S. (1996)** "Etude comparative de la performance de divers coagulants préhydrolysés sur l'enlèvement de la matière organique dissoute" Vecteur Environnement, vol. 29, N°5, pp. 37-45
- DESJARDINS C., KOUDJONOU B., DESJARDINS R. (2002)** "Laboratory study of ballasted flocculation" Water Research, 36 pp.744-754
- DRAY S. (2008)** "On the number of principal components: A test of dimensionality based on measurements of similarity between matrices" Computational Statistics & Data Analysis, Vol.52, 4, pp.2228-2237.
- DUAN J., GREGORY J. (2003)** "Coagulation by hydrolysing metal salts" Advances in Colloid and Interface Science, 100 -102, pp 475-502
- DUCHAUFOUR P. (1965)** Précis de pédologie 2ème édition
- DUCHAUFOUR P. (1997)** Abrégé de pédologie Sol, végétation, environnement 5ème édition Masson
- DUMOLARD P. (2007)** Analyse multivariée des données géographiques. Disponible sur http://iga.ujf-grenoble.fr/Cours-en-Ligne./bk_multiv.pdf 144p.
- DUSSOUILLEZ E. B., BOTTERO J.Y., MOULIN P, (2010)** "Coagulation of bentonite suspension by polyelectrolytes or ferric chloride: Floc breakage and reformation" Chemical Engineering Journal 156, pp.83-91
- EDELIN F. (1996)** L'épuration physico chimique des eaux, 4^{ème} édition Lavoisier 288 p
- EDZWELD J. K. (1993)** "Coagulation in drinking water treatment: particles, organics, and coagulants" Water Science Technology, Vol. 27, N°11, pp21-35
- ELMALEH S., JABBOURI A. (1991)** "Flocculation energy requirement" Water Research, Vol.25, N°8, pp 939 - 943
- ELICKER M.L., RESTA J.J., HUNT J.W.,DENTEL S.K (1992)** "Fundamental considerations in use of streaming current detector for chemical dose control" in

Références bibliographiques

Chemical Water and Wastewater Treatment II , R. Klute and H. Hahn, eds., Springer Verlag, Berlin(1992).

EXALL K. N., VANLOON G. W. (2003) "Effects of raw water conditions on solution-state aluminium speciation during coagulant dilution" Water Research, 37, pp.3341-3350

FAIRE G. M., GEMMELL R. S. (1964) "A mathematical model of coagulation" Journal of colloid science 19, pp 360 - 372

FIESSINGER F. (1976) "La coagulation : Errements anciens et connaissances nouvelles" T.S.M avril pp 147 - 162

FOUCAULT A., RAOULT J. F. (1984) Dictionnaire de géologie, 2^{ème} édition 347p

FRANCESCHI M. (1991)Contribution à l'étude des mécanismes de coagulation floculation, modélisation de la phase de floculation. Étude de la morphologie des agrégats formés Thèse de doctorat physique et chimie de l'environnement Université de Paul Sabatier - Toulouse.

FRANCESCHI M., PUECH-COSTE E., MAURETTE M. T., GIROU A. (1993) "Première étape dans l'élaboration d'un modèle mathématique générale permettant de remplacer le jar test" Environmental Tecnology, Vol.14, pp.341-350

FRANCESCHI M., GIROU A., CARRO-DIAZ A.M., MAURETTE M.T.,PUECH-COSTES E.(2002) "Optimisation of the coagulation-floculation process of raw water by optimal design method" Water Research, 36 pp.3561-3572

FRANCOIS R.J. HAUTE V., A.A.(1985) "Structure of hydroxide flocs" Water Research, Vol. 19, N°10, pp.1249-1254

FRANÇOIS R.J. (1987a) "Strength of aluminium hydroxide flocs" Water Research, Vol. 21, N°9, pp. 1023-1030

FRANCOIS R.J. (1987b) "Ageing of aluminium hydroxide flocs" Water Research, Vol. 21, N°5, pp. 523-531

FUKASAWA T., ADACHI Y.(2006) "Effect of floc structure on the rate of Brownian coagulation" Journal of Colloid and Interface Science, 304, pp. 115-118

GAGNON C., GRANDJEAN B. P. A., THIBAUT J. (1997) "Modelling of coagulant dosage in a water treatment plant" Artificial Intelligence in Engineering, V.11, Issue 4, pp.401-404

GALLINARI F.P (1992) Etude comparative de la décantation et de la floculation à l'air dissous; application au traitement des eaux. Thèse de docteur d'état sciences physiques USTHB

GAMAL E.A., SMITH D. W. (2002) "A neural network model to predict the wastewater inflow incorporating rainfall events" Water Research, 36, pp.1115-1126

GAO B.Y., HAHN H.H., HOFFMANN E. (2002) "Evaluation of aluminum-silicate polymer composite as a coagulant for water treatment" Water Research, 36, pp.3573-3581

Références bibliographiques

- GAO B., YUE Q., WANG Y.,(2007a)** "Coagulation performance of polyaluminum silicate chloride (PASiC) for water and wastewater treatment" Separation and Purification Technology, 56, pp.225-230
- GAO P., ZHANG L., CHENG K., ZHANG H. (2007b)** "A new approach to performance analysis of a seawater desalination system by an artificial neural network" Desalination, 205, pp.147-155
- GEVREY M., DIMOPOULOS I., LEK S. (2003)** "Review and comparison of methods to study the contribution of variables in artificial neural network models" Ecological Modelling, 160, pp. 249-264
- GEVREY M. DIMOPOULOS I., LEK S. (2006)** "Two-way interaction of input variables in the sensitivity analysis of neural network models" Ecological modelling 195, pp.43-50
- GHAFARI S., HAMIDI A. , HASNAIN M. I., ZINATIZADEHD A .A. (2009)** "Application of response surface methodology (RSM) to optimize coagulation-flocculation treatment of leachate using poly-aluminum chloride (PAC) and alum" Journal of Hazardous Materials, 163, pp. 650-656
- GIBBS R.J., KONWAR L.N. (1982)** "Effect of pipetting on mineral flocs" Environmental science technology Vol. 16, N°2, pp. 119-121
- GOUPY J. (1996)** la méthode des plans d'expériences optimisation du choix des essais et de l'interprétation des résultats édition DUNOD 303p.
- GRAVEL S. (1987)** "modélisation mathématique du dosage de produits chimiques pour le traitement de l'eau potable à l'usine de Sainte -Foy" Science et technique de l'eau, pp.175-178
- GREGOR J.E., NOKES C. J., FENTON E. (1997)** "Optimising natural organic matter removal from low turbidity waters by controlled pH adjustment of aluminium" Water Research, Vol.31, N°12, pp. 2949-2958
- GREGORY J. (1989)** "Fundamentals of flocculation" critical reviews in Environmental control, pp. 185-231
- GREGORY J., DUPONT V. (2001)** "Properties of flocs produced by water treatment coagulants" Water Science Technology, 44(10), pp.231-236
- GUIDA M., MATTEI M., ROCCA C.D., MELLUSO G., MERIÇ S. (2007)** "Optimization of alum-coagulation/flocculation for COD and TSS removal from five municipal wastewater" Desalination, 211, pp.113-127
- GÜVEN N., (1992)** "Rheological aspect of aqueous smectite suspension" pp. 81-156, in C.M.S workshop lectures vol. 4, clay -water interface and its rheological implications, the clay mineral society, Colorado
- HAARHOFF J, BEEK JC. V., H.J. VAN ZYL (1996)** "Practical application of the Argaman - Kauffman flocculation model" Proceeding of the 1996 Biennial conference of water institute of Southern Africa, held in port Elizabeth South Africa.
- HAARHOFF J., HANNES J. (1997)** "Determination of aggregation and break-up constants during flocculation" Water Science Technology, 36 (4), pp. 33-40

Références bibliographiques

- HAAS C. N. (2004)** "Neural networks provide superior description of Giardia lamblia inactivation by free chlorine" Water Research, 38, pp.3449-3457
- HAMED M.M., KHALAFALLAH M.G., EZZAT A. H. (2004)** "Prediction of wastewater treatment plant performance using artificial neural networks" Environmental Modelling & Software, 19, pp. 919-928
- HASANI M., MOLOUDI M. (2008)** "Application of principal component-artificial neural network models for simultaneous determination of phenolic compounds by a kinetic spectrophotometric method" Journal of Hazardous Materials, 157, pp161-169
- HART J. (1997)** "Application of process simulation in water treatment" TSM, N°7-8, juillet -aout, pp.30-36
- HECHT-NIELSEN, R. (1987)** "Kolmogorov's mapping neural network existence theorem" First IEEE International Joint Conference on Neural Networks, San Diego, California, pp. 11-14.
- HECTOR R. H. D. L. (2006)** Supervision et diagnostic des procédés de production d'eau potable Thèse de doctorat, laboratoire d'analyse et d'architecture des systèmes CNRS
- HEDDAM S., DECHEMI N. (2008)** "Modeling of Optimal Coagulant Dose Using Artificial Neural Network, Application to Water Treatment Plant of Boudouaou (Algeria)" European Journal of Scientific Research Vol.20, N°3, pp.652-663
- HENRY M.T, (1989)** "Evolution des dérivés de l'aluminium utilisés comme agents coagulants" Science et Technique de l'eau, 22, N°4, pp. 297-304.
- HENRY M,T, (1992)** Chimie des eaux 2ème édition le griffon d'argile 537 p
- HIMM-ELBLAU D. M. (2000)** "Applications of Artificial Neural Networks in Chemical Engineering" Korean J Chem. Eng., 17(4), 373-392
- HILAL N., OGUNBIYI O., AL-ABRI M. (2008)** "Neural network modeling for separation of bentonite in tubular ceramic membranes" Desalination, 228, pp.175-182
- HOLMBERG M. FORSIUS M. STARR M. HUTTUNEN M. (2006)** "An application of artificial neural networks to carbon, nitrogen and phosphorus concentrations in three boreal streams and impacts of climate change" Ecological modelling, 195, pp.51-60
- HONG Y.T., ROSEN M. R., BHAMIDIMARRI R. (2003)** "Analysis of a municipal wastewater treatment plant using a neural network-based pattern analysis" Water Research, 37, pp.1608-1618
- HOPKINS D., DUCOSTE C. J. (2003)** "Characterizing flocculation under heterogeneous turbulence" Journal of Colloid and Interface Science, 264, pp.184-194
- HU Y., QIU G., MILLER J.D. (2003)** "Hydrodynamic interactions between particles in aggregation and flotation", Int. J. Miner. Process., 70, pp.157- 170

Références bibliographiques

- HUANG C., SHIU H. (1996)** "Interaction between alum and organics in coagulation" Colloids and surfaces A)Physico-chemical and Engineering Aspect 113 pp 155 – 163
- HUANG W., FOO S. (2002)** "Neural network modeling of salinity variation in Apalachicola River" Water Research, 36, pp.356–362
- IBARRA B G., ELIAS A., BARONA A., SAENZ J. EZCURRA A., ARGANDONA J., D. (2008)** "From diagnosis to prognosis for forecasting air pollution using neural networks: Air pollution monitoring in Bilbao" Environmental Modelling & Software, 23, pp.622-637
- IVES K.J. (1981)** "Coagulation flocculation part II orthokinitic flocculation" dans 'solid-liquide séparation' Ed. L SVAROVSKY, second edition, Butter.
- JARRET M., CAVELIER C., DUCAUZE C. (1986)** "Caractérisation de la matière organique présente dans les eaux en cours d'affinage : influence des traitements appliqués et du climat" Water Research, Vol. 20, N°12, pp. 1477-1488
- JARVIS P., JEFFERSON B., GREGORY J., PARSONS S.A. (2005)** "A review of floc strength and breakage" Water Research, 39, pp.3121–3137
- JARVIS P., JEFFERSON B., PARSONS S. A. (2006)** "Floc structural characteristics using conventional coagulation for a high doc, low alkalinity surface water source" Water Research, 40, pp.2727–2737
- JAYARAMAN K. (1999)** Manuel de statistique pour la recherche forestière
Organisation des nations unies pour l'alimentation et l'agriculture
FO : GCP/ RAS/163/NET, FO : GCP/INT/679/EC, 239p.
- JENKINS D., SNOEYINK V.L. (1980)**, water chemistry copyright by John Wiley et sons, Inc., 363 p.
- JOO D.S., CHOI D.J., PARK H. (2000)** "The effects of data preprocessing in the determination of coagulant dosing rate" Water Research, Vol.34, 13, pp.3295-3302
- JULIEN F., GUEROUX B., MAZET (1994)** "Comparaison de l'élimination de molécules organiques par coagulation flocculation et par adsorption sur floes d'hydroxyde métallique préformés" Water Research, Vol. 28, N°12, pp. 2567-2574.
- KARUL C., SOYUPAK S., ÇILESIZ A. F., AKBAY N., GERMEN E.(2000)** "Case studies on the use of neural networks in eutrophication modeling" Ecological Modelling, V.134, 2-3, pp.145-152
- KAUNDA R. B., CHASE R. B., KEHEW A. E., KAUGARS K., SELEGEAN J. P (2009)** "Neural network modeling applications in active slope stability problems" Environ Earth Sci, DOI 10.1007/s12665-009-0290-3 Springer
- KAZI T.G., ARAIN M.B., JAMALI M.K., JALBANI N., AFRIDI H.I., SARFRAZ R.A., BAIG J.A., SHAH A. Q. (2009)** "Assessment of water quality of polluted lake using multivariate statistical techniques: A case study" Ecotoxicology and Environmental Safety, 72, pp.301–309
- KELLIL A., (1989)**, Caractéristiques morphologiques des particules floculées : influence des conditions hydrodynamiques de formation thèse docteur d'état, sciences physique I.N.P.T 310 p.

Références bibliographiques

- KETTANEH N., BERGLUND W. S. A., (2005)** "PCA and PLS with very large data sets" Computational Statistics & Data Analysis, 48, pp.69-85
- KILANDER J., BLOMSTRÖM S., RASMUSON A. (2006)** "Spatial and temporal evolution of floc size distribution in a stirred square tank investigated using PIV and image analysis" Chemical Engineering Science, 61, pp.7651-7667
- KILANDER J., BLOMSTRÖM S., RASMUSON A.(2007)** "Scale-up behaviour in stirred square flocculation tanks" Chemical Engineering Science, 62, pp.1606-1618
- KOBAYASHI M. (2004)** "Breakup and strength of polystyrene latex flocs subjected to a converging flow" Colloids and Surfaces A: Physicochem. Eng. Aspects 235, pp.73-78
- KOBAYASHI M. (2005)** "Strength of natural soil flocs" Water Research, 39, pp.3273-3278
- KOKLU R., SENGORUR B., TOPAL B. (2010)** "Water Quality Assessment Using Multivariate Statistical Methods —A Case Study: Melen River System (Turkey)" Water Resour. Manage. 24, pp.959-978
- KUO C.J., AMY G.L, BRYANT C.W. (1988a)** "Factors affecting coagulation with aluminium sulfate I: Particle Formation and growth" Water Research, Vol. 22, N°7, pp 853 - 862.
- KUO C.J., AMY G.L, (1988b)** "Factors affecting coagulation with aluminium sulfate II: Dissolved organic matter removal" Water Research, Vol. 22, N°7, pp. 863-872
- KUO Y.M., LIU C.W., LIN K.H. (2004)** "Evaluation of the ability of an artificial neural network model to assess the variation of groundwater quality in an area of black foot disease in Taiwan" Water Research, 38, pp.148-158
- KUO J. T., HSIEH M.H., LUNG W.S., SHE N. (2007)** "Using artificial neural Network for reservoir eutrophication prediction" Ecological modelling 200, pp.171-177
- LAGALY G., ZIESMER S. (2003)** "Colloid chemistry of clay minerals: the coagulation of montmorillonite dispersions" Advances in Colloid and Interface Science, 100 -102, pp. 105-128
- LAHOUSSINE T.V., WIESNER M., BOTTERO J.Y. et al (1992)** "Coagulation flocculation à l'aide de sels d'aluminium: influence sur la filtration par des membranes micro - poreuses" Water Research, Vol. 26, N°5 pp 695 - 702.
- LAI R J, HUDSON H.E.J., SINGLEY J E. (1975)** "Velocity gradient calibration of jar-test equipment" Journal of AWWA, pp. 553-557
- LAMRINI B., BENHAMMOU A., LANN M.V. L., KARAMA A. (2005)** "A neural software sensor for on-line prediction of coagulant dosage: Application to a drinking water treatment plant", Trans. Inst. Meas. Control (TIMC), 27, pp.195-213.
- LAMRINI B., LANN M.V.L., LAKHAL E. BENHAMMOU A. (2007)** "Gestion supervisée d'une unité de coagulation pour la potabilisation des eaux à partir d'une méthodologie d'apprentissage et d'expertise" Revue des sciences de l'eau, vol.20, N°4, pp.325-338

Références bibliographiques

- LEBART L., MARINEAU A., PIRON M. (1997)** Statistique exploratoire multidimensionnelle, 2^{ème} édition Dunod 437p
- LEE J. H.W., HUANG Y., DICKMAN M., JAYAWARDENA A.W. (2003)** "Neural network modelling of coastal algal blooms" Ecological Modelling, 159, pp.179-201
- LEE L., HELSEL D. (2007)** "Statistical analysis of water-quality data containing multiple detection limits II: S-language software for nonparametric distribution modeling and hypothesis testing" Computers & Geosciences, 33, pp.696-704
- LEE Y. G., LEE Y. S., JEON J. J., LEE S., YANG D. R., KIMA I. S., KIMA J.H., (2009)** "Artificial neural network model for optimizing operation of a seawater reverse osmosis desalination plant" Desalination, 247, 11, pp.80-89
- LEE K., E., MORAD N., POH B., T., TENG T., T., (2011)** "Comparative study on the effectiveness of hydrophobically modified cationic polyacrylamide groups in the flocculation of kaolin" Desalination 270, pp. 206-213
- LEENTVAAR J., YWEMAT. S. J. (1980)** "Some dimensionless parameters of impeller power in coagulation flocculation processes" Water Research vol.14, pp. 135 -140.
- LEEUWEN J.V., DALY R., HOLMES M. (2005)** "Modeling the treatment of drinking water to maximize dissolved organic matter removal and minimize disinfection by-product formation" Desalination, 176, pp. 81-89
- LEFEBVRE J. (1988)** Introduction aux analyses statistiques multidimensionnelles avec exemples d'applications, 3^{ème} édition, 273p.
- LEFBVRE E, LEGUBE B. (1993)** " Coagulation-floculation par le chlorure ferrique de quelques acides organiques et phénols en solution aqueuse" Water Research, Vol. 27, N°3, pp 433 - 447.
- LEK S. (2007)** "Uncertainty in ecological models" Ecological modeling, 207, pp.1-2
- LETTERMAN R.D. (1973)** "Influence of rapid mix parameters on flocculation" Journal of AWWA, pp.716-722
- LOZET J., MATHIEU C. (1990)** Dictionnaire de science de sol. 2^{ème} édition Lavoisier Paris 395 p
- LURIE M., REBHUN M.,** "Effect of properties of Polyélectrolytes on their interaction with particulates organics", pp. 84 - 91
- MAIER H. R., MORGAN N., CHOWB C. W.K. (2004)** "Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters Environmental" Modelling & Software 19, pp.485-494.
- MASSCHELEIN W.J. (1996)** Processus Unitaire de traitement de l'eau potable. Trad. de Unit process in drinking water treatment par CHEVOLET H. édition CEBE Doc.sprL leige 693 p
- McCURDY K., CARLSON K., GREGORY D. (2004)** "Floc morphology and cyclic shearing recovery: comparison of alum and polyaluminum chloride coagulants" Water Research, 38, pp.486-494

Références bibliographiques

- McCARTY P.L. (1970)** "Phosphorus and nitrogen removal by biological systems", proceeding wastewater reclamation and reuse workshop, Tahae, Calif., 226, pp25-27
- MELESSE A.M., HANLEYB R. S. (2005)** "Artificial neural network application for multi-ecosystem carbon flux simulation" Ecological Modelling, 189, pp.305-314
- MIETTA F., CHASSAGNE C., WINTERWERP J. C., MANNING A. J. (2009a)** "Influence of shear rate, organic matter content, pH and salinity on mud flocculation" Ocean Dynamics, 59, pp.751-763
- MIETTA F., CHASSAGNE C., WINTERWERP J. C. (2009b)** "Shear-induced flocculation of a suspension of kaolinite as function of pH and salt concentration" Journal of Colloid and Interface Science 336, pp. 134-141
- MIETTA F., CHASSAGNE C., VERNEY R., WINTERWERP J. C. (2010)** On the behavior of mud floc size distribution: model calibration and model behavior Ocean Dynamics,
- MING-QUAN Y., DONG-SHENG W., BAO-YOU S., QUN-SHAN W., JIU-HUI Q., HONG-XIAO T. (2007)** "Transformations of particles, metal elements and natural organic matter in different water treatment processes" Journal of Environmental Sciences, 19, pp.271-277
- MIRSEPASSI A (2004)** "Application of Intelligent System for Water Treatment Plant Operation" Iranian Journal of Environmental Health Science & Engineering, Vol.1, N. 2, pp.51-57
- MLYANAMI K., TOJO K., YOKOTA M. et al (1982)** "Effect of mixing on flocculation" Ind. Eng. chem. Fund. , 21, pp 132-135
- MOURAD M. (2005)** Modélisation de la qualité des rejets urbains de temps de pluie: sensibilité aux données expérimentales et adéquation au besoin opérationnel, Thèse de doctorat, Institut National des Sciences Appliquées Lyon 296 p.
- NARONG P., JAMES A.E. (2006)** "Effect of pH on the ζ -potential and turbidity of yeast suspensions" Colloids and Surfaces A: Physicochem. Eng. Aspects, 274, pp.130-137
- OSORIO F. S., (1998)** INSS: Un système hybride neuro-symbolique pour l'apprentissage constructif automatique, Thèse de doctorat, L'Institut National Polytechnique de Grenoble - I.N.P.G. Laboratoire LEIBNIZ - IMAG 305 p.
- PA. IIO. IISU (1997)** "Effect of temperature on the degradation of Al_{13} complex" Clays and clay minerals, Vol. 45, N°2, pp. 286-289
- PARINET B., BOUKARI S.O.B., ADIMA A. (2005)** "Evaluation de l'état d'une eau de rivière par analyse multidimensionnelle en utilisant certains paramètres caractéristiques de la matière organique dissoute" Revue des sciences de l'eau, 18/2, pp.133-163
- PARK S. S., LEE Y. S. (2002)** "A water quality modeling study of the Nakdong River, Korea" Ecological Modelling, 152, pp.55-75
- PARTHASARATHY N., BUFFLE J. (1985)** "Study of polymeric aluminium (III) hydroxide solutions for application in wastewater treatment. Properties of the polymer and optimal conditions of preparation" Water Research, vol.19, N°1, pp. 25-36

- PENG F.F., PINGKUAN DI (1994)** "Effect of multivalent salts - calcium on the flocculation of kaolin suspension with anionic polyacrylamide" Journal of colloid and Interface science 164 pp. 229-237
- PLATIKANOV S., PUIG X., MARTIN J., TAULER R. (2007)** "Chemometric modeling and prediction of trihalomethane formation in Barcelona's water works plant" Water Research, 41, pp.3394-3406
- PRIMPAS I., TSIRTSIS G., KARYDIS M., KOKKORIS G. D. (2010)** "Principal component analysis: Development of a multivariate index for assessing eutrophication according to the European water framework directive" Ecological Indicators, Vol. 10, Issue 2, pp.178-183.
- PURNENDU B, RECKHOW D.A. (1998)** "Adsorption of natural organic matter on performed aluminium hydroxide flocs" Journal of Environmental Engineering, Vol. 124, N°9, pp. 803 - 811
- RACKAUSKAS A., ZUOKAS D.(2007)** "New tests of heteroskedasticity in linear regression model" Lithuanian Mathematical Journal, Vol. 47, N° 3
- RAKOTOMALALA R. (2009)** Pratique de la Régression Linéaire Multiple, Diagnostic et sélection de variables, Version du 20 sept. 2009 Université Lumière Lyon 2 disponible sur : <http://dis.univ-lyon2.fr/>.
- REBHUN M., MSPERBER H. (1983)** "Somme kinetic studies of flocculation with polymers in continuous flow reactors" International Symposium "Advances in Solid-Liquid Separation", London.
- REBHUN M. (1990)** "Floc formation and break up in continuous flow flocculation and in contact filtration", H. H. Hahn and R Klute (Eds) Chemical Water and Wastewater Treatment, Spring- verlag berline heidelberg
- RIETVELD L., DUDLEY J. (2006)** Models for Drinking Water Treatment Review State-of-the-Art, Techneau, 06, 22p
- ROBENSON A., ABDSHUKOR S.R., AZIZ N.(2009)** "Development of Process Inverse Neural Network Model to Determine the Required Alum Dosage at Segama Water Treatment Plant Sabah, Malaysia" Computer Aided Chemical Engineering, Vol.27, pp.525-530
- RODIER J. (1996)** L'analyse de l'eau 8ème édition Dunod 1384 p
- RODRIGUEZ M. J., SERODES J.B. (1996)** "Neural network-based modelling of the adequate chlorine dosage for drinking water disinfection" Can. J. Civ. Eng., 23(3), pp.621-631
- RODRIGUEZ MJ, SERODES J.B. (1999)** "Assessing empirical linear and non-linear modelling of residual chlorine in urban drinking water systems", Environ Modelling Software, 14, pp.93-102
- RUNKANA V., SOMASUNDARAN P., KAPUR P.C.(2006)** "Apopulation balance model for flocculation of colloidal suspensions by polymer bridging" Chemical Engineering Science, 61, pp.182-191

Références bibliographiques

- RUDE J. (2008)** Développement d'un modèle statistique neuronal pour la description fine de la pollution par le dioxyde d'azote : application à la région parisienne, Thèse de Doctorat Université Paris XII - Val de Marne, p140
- SAHOO G. B., RAY C., WANG J. Z., HUBBS S. A., SONG R., JASPERSE J., SALARI D., DANESHVAR N., AGHAZADEH F., KHATAEE A.R. (2005)** "Application of artificial neural networks for modeling of the treatment of wastewater contaminated with methyl tert-butyl ether (MTBE) by UV/H₂O₂ process" Journal of Hazardous Materials, B125, pp.205-210
- SAPORTA G. (1990)** Probabilité, analyses des données et statistique éditions TECHNIP 476p
- SAWER C.N., PERRY L.C., GERE F.P. (1994)** Chemistry of Environmental Engineering 4ème édition Mc Graw - Hill 658 p
- SERODES J.B., RODRIGUEZ M.J., PONTON A. CHLORCAST C. (2001)** "a methodology for developing decision-making tools for chlorine disinfection control", Environ Modelling Software, 16, pp.53-62
- SEYMOUR D. (2005)** "Use of artificial neural networks to evaluate the effectiveness of riverbank filtration" Water Research, 39, pp.2505-2516
- SHETTY G.,R. MALKI H., CHELLAM S. (2003)** "Predicting contaminant removal during municipal drinking water nanofiltration using artificial neural networks" Journal of Membrane Science 212, pp. 99-112
- SINGH K. P., MALIK A., MOHAN D., SINHA S. (2004)** " Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study" Water Research, 38, pp.3980-3992
- SINGH K. P., BASANT A., MALIK A., JAIN G. (2009)** "Artificial neural network modeling of the river water quality—A case study" Ecological Modelling, 220, pp.888-895
- SEKIOU F. (2001)** Effets de la nature des particules solides dispersées et des conditions de formation de floc sur l'efficacité de floculation, Mémoire de Magister, E.N.S.H. Blida 153p.
- SEKIOU F., KELLIL A. (2009)** "Effect of organic and mineral matters on kinetic and performance of flocculation" Desalination, 249, pp.891-894
- SOMASUNDARAN P., RUNKANA V. V. (2003)** "Modeling flocculation of colloidal mineral suspensions using population balances" Int. J. Miner. Process., 72, pp. 33-55
- SOMASUNDARAN P., RUNKANA V. V. (2005)** "Investigation of the flocculation of colloidal suspensions by controlling adsorbed layer microstructure and population balance modelling" Chemical Engineering Research and Design, 83(A7), pp.905-914
- SONMEZ I. (2007)** "Application of a statistical design method to the shear flocculation of celestite with Na-Oleate" Colloids and Surfaces A: Physicochem. Eng. Aspects, 302, pp.330-336

Références bibliographiques

- SOUSA S.I.V., MARTINS F.G., ALVIM-FERRAZ M.C.M., PEREIRA M.C. (2007)** "Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations" Environmental Modelling & Software, 22, pp.97-103
- SPENGLER S., BENEFIELD L., JENKINS S.R. (1983)** "Coagulation for removal of humic materials from groundwater" Separation science and technology, 18(2), pp.135-163
- TABOADA-SERRANO P., CHIN C., YIACOUMI S., TSOURIS C. (2005)** "Modeling aggregation of colloidal particles" Current Opinion in Colloid & Interface Science, 10, pp.123-32
- TAMBO N. (1991)** "Basic concepts and innovative turn of coagulation/ flocculation" Water supply, Vol.9, pp. 1-10
- TANVIR M.S., MUJTABA I.M. (2006)** "Neural network based correlations for estimating temperature elevation for seawater in MSF desalination process" Desalination, 195, pp.251-272
- TEODOSIU C, PASTRAVANU O, MACOVEANU M. (2000)** "Neural network models for ultrafiltration and backwashing" Water Research, 34(18), pp.4371-4380
- THOMAS D.N., JUDDS J., FAWCETT N., (1999)** "Flocculation modelling a review" Water Research, Vol. 33, N°7, pp 1579-1592
- TURCHIULI C., FARGUES C., (2004)** "Influence of structural properties of alum and ferric flocs on sludge dewaterability" Chemical Engineering Journal, 103, pp.123-131
- VALENTIN N. (2000)** "Construction d'un capteur logiciel pour le contrôle automatique du procédé de coagulation en traitement d'eau potable thèse de doctorat de l'UTC, CIRSEE, Centre International de Recherche sur l'Eau et l'Environnement, 168 p.
- VATAN A. (1967)** Manuel de sédimentologie Editions technique 39/ p
- VELASCO U. I., URIARTE J. I., VELASCO A. J. R. G. (2007)** "Enhanced coagulation under changing alkalinity-hardness conditions and its implications on trihalomethane precursors removal and relationship with UV absorbance" Separation and Purification Technology, 55, pp.368-380
- VITTINGHOFF E., SHIBOSKI C. S. C., GLIDDEN D. V., MCCULLOCH C. E. (2005)** Regression Methods in Biostatistics Linear, Logistic, Survival and Repeated Measures Models, Library of Congress Cataloging-in-Publication Data, Springer Science+Business Media, Inc. 340 p
- WANG J., CHIEN Y., GE X., YU H. (2007)** "Optimization of coagulation-flocculation process for a paper-recycling wastewater treatment using response surface methodology" Colloids and Surfaces A: Physicochem. Eng. Aspects, 302, pp.204-210
- WANG Y., GAO B., XU X., XU W., XU G. (2009a)** "Characterization of floc size, strength and structure in various aluminum coagulants treatment" Journal of Colloid and Interface Science, 332, pp.354-359

Références bibliographiques

- WANG Y., ZHOU W., GAO B., XU X., XU G. (2009b)** "The effect of total hardness on the coagulation performance of aluminium salts with different Al species" Separation and Purification Technology, 66, pp.457-462
- WANG Y., GAO B., XU X., , XU W., (2010)** "The effect of total hardness and ionic strength on the coagulation performance and kinetics of aluminum salts to remove humic acid" Chemical Engineering Journal, 160, pp.150-156
- WANG H., CUI J., JIANG W. (2011)** "Synthesis, characterization and flocculation activity of novel Fe(OH)₃-polyacrylamide hybrid polymer" Materials Chemistry and Physics 130, pp.993- 999
- WU R. M., LEE D. J., WAITE T. D., GUAN J. (2002)** "Multilevel Structure of Sludge Flocs" Journal of Colloid and Interface Science, 252, pp.383-392
- WU G D.(2005)** Une approche de réseau de neurones pour la coagulation en stations d'épuration Thèse de maitrise, Université nationale de technologie de Taipei chine, 137 p
- WU G.D., LO S. (2008)** "Predicting real-time coagulant dosage in water treatment by artificial neural networks and adaptive network-based fuzzy inference system" Engineering Applications of Artificial Intelligence, 21, pp.1189-1195
- WU X., GE X., WANG D., TANG H. (2007)** "Distinct coagulation mechanism and model between alum and high Al₁₃-PACl" Colloids and Surfaces A: Physicochem. Eng. Aspects, 305, pp.89-96
- XIAO F., HUANG J.H., ZHANG B., CUI C. (2009)** "Effects of low temperature on coagulation kinetics and floc surface morphology using alum" Desalination 237, pp. 201 213
- XU W., GAO B., YUE Q., WANG Y. (2010)** "Effect of shear force and solution pH on flocs breakage and re-growth formed by nano-Al₁₃ polymer" Water Research, 44, pp.893 -899
- YAN M., WANG D., NI J., QU J., CHOW C.W.K., LIU H. (2008)** "Mechanism of natural organic matter removal by polyaluminum chloride: Effect of coagulant particle size and hydrolysis kinetics" Water Research, 42, pp.3361-3370
- YANG Z., HUANG J., ZENG G., RUAN M., ZHOU C., LI L., RONG Z. (2009)** "Optimization of flocculation conditions for kaolin suspension using the composite flocculant of MBFGA1 and PAC by response surface methodology" Bioresource Technology, 100, pp.4233-4239
- YATES P., D., FRANKS G., V., JAMESON G., J. (2008)** "Orthokinetic heteroaggregation with nanoparticles: Effect of particle size ratio on aggregate properties" Colloids and Surfaces A: Physicochem. Eng. Aspects 326, pp. 83-91
- YIDANA S. M., OPHORI D., BANOENG-YAKUBO B. (2008)** "A multivariate statistical analysis of surface water chemistry data—The Ankobra Basin, Ghana" Journal of Environmental Management, 86, pp.80-87
- YOON Y., DENG Y. (2004)** "Flocculation and reflocculation of clay suspension by different polymer systems under turbulent conditions " Journal of Colloid and Interface Science, 278, pp.139-145

Références bibliographiques

- YU W., GREGORY J., CAMPOS L., C. (2011)** "Breakage and re-growth of flocs: Effect of additional doses of coagulant species" water re search 45, pp. 6718-6724
- YUANG J., MURRAY H. (1997)** "The importance of crystal morphology on the viscosity concentrated suspensions of kaolin" Applied clay suspensions 12 pp209 – 219
- YUKSELEN M., A., JOHN GREGORY J. (2004)** "The reversibility of floc breakage" Int. J. Miner. Process., 73, pp. 251-259
- ZHAN X., GAO B., YUE Q., WANG Y., CAO B. (2010)** "Coagulation behavior of polyferric chloride for removing NOM from surface water with low concentration of organic matter and its effect on chlorine decay model" Separation and Purification Technology, 75, pp. 61-68
- ZHAN X., GAO B., WANG Y., YUE Q. (2011)** "Influence of velocity gradient on aluminum and iron floc property for NOM removal from low organic matter surfacewater by coagulation" Chemical Engineering Journal, 166, pp.116-121
- ZHAO H.A.B., LIUA H., QUA J., (2009)** "Effect of pH on the aluminum salts hydrolysis during coagulation process: Formation and decomposition of polymeric aluminum species" Journal of Colloid and Interface Science, 330, pp.105-112
- ZHAO Y.X., GAO B.Y., WANG Y. SHON H.K. BO X.W. YUE Q.Y. (2012)** "Coagulation performance and floc characteristics with polyaluminum chloride using sodium alginate as coagulant aid: A preliminary assessment" Chemical Engineering Journal 183, pp.387- 394
- ZENG G.M., QIN X.S., HE L., HUANG G.H., LIU H.L., LIN Y.P. (2003)** "A neural network predictive control system for paper mill wastewater treatment" Engineering Applications of Artificial Intelligence, 16, pp.121 129
- ZHENG H., ZHU G., TSHUKUDU T., XIANG X., ZHANG P., HE Q. (2011)** "Investigations of coagulation-flocculation process by performance optimization, model prediction and fractal structure of flocs" Desalination 269, pp.148-156
- ZHU G., ZHENG H., ZHANG Z., TSHUKUDU T., ZHANG P., XIANG X. (2011)** "Characterization and coagulation-flocculation behavior of polymeric aluminum ferric sulfate (PAFS) Chemical Engineering Journal 178, pp.50-59
- ZORRIASSATINE F., TANNOCK J. D. T. (1998)** "A review of neural networks for statistical process control" Journal of Intelligent Manufacturing, 9, pp.209-224
- ZOUGRANA C.J., RAYMOND D., PREVOST M., (1993)** "Effets de la reminéralisation sur l'élimination de la matière organique par coagulation-flocculation" Sciences et techniques de l'eau, Vol. 26, N°4, Nov. pp. 256 - 263

ANNEXES

- Annexe A :** Table de Durbin watson (test unilatérale au seuil de 5%)
Valeurs de dL et dU
- Annexe B :** Table des fractiles $f_{(v_1, v_2)}$ pour une loi $F_{(v_1, v_2)}$ (Fisher) (5%)
- Annexe C :** Table des fractiles $t_v(p)$ pour une loi T (Student)
- Annexe D :** Valeurs significatives du coefficient de corrélation
- Annexe E :** Procédure de recherche de l'influence des variables en modèle de régression linéaire
- Annexe F :** Procédure de calcul de l'importance des variables par approche de poids de connexion.

Annexe A

Table de Durbin watson (test unilatérale au seuil de 5%)
Valeurs de dL et dU

n	k*=1		k*=2		k*=3		k*=4		k*=5		k*=6		k*=7		k*=8		k*=9		k*=10		
	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	
6	0.610	1.400	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
7	0.700	1.356	0.467	1.896	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
8	0.763	1.332	0.559	1.777	0.367	2.287	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588	----	----	----	----	----	----	----	----	----	----	----	----	----
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822	----	----	----	----	----	----	----	----	----	----	----
11	0.927	1.324	0.758	1.604	0.595	1.928	0.444	2.283	0.315	2.645	0.203	3.004	----	----	----	----	----	----	----	----	----
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.380	2.506	0.268	2.832	0.171	3.149	----	----	----	----	----	----	----
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.444	2.390	0.328	2.692	0.230	2.985	0.147	3.266	----	----	----	----	----
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360	----	----	----
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.471	0.343	2.727	0.251	2.979	0.175	3.216	0.111	3.438	----
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304	----
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975	0.198	3.184	----
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.258	0.502	2.461	0.407	2.668	0.321	2.873	0.244	3.073	----
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783	0.290	2.974	----
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.691	2.162	0.595	2.339	0.502	2.521	0.416	2.704	0.336	2.885	----
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.731	2.124	0.637	2.290	0.546	2.461	0.461	2.633	0.380	2.806	----
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571	0.424	2.735	----
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514	0.465	2.670	----
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.750	2.174	0.666	2.318	0.584	2.464	0.506	2.613	----
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.013	0.784	2.144	0.702	2.280	0.621	2.419	0.544	2.560	----
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379	0.581	2.513	----
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342	0.616	2.470	----
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.959	0.874	2.071	0.798	2.188	0.723	2.309	0.649	2.431	----
29	1.341	1.483	1.270	1.563	1.198	1.650	1.124	1.743	1.050	1.841	0.975	1.944	0.900	2.052	0.826	2.164	0.753	2.278	0.681	2.396	----
30	1.352	1.489	1.284	1.567	1.214	1.650	1.143	1.739	1.071	1.833	0.998	1.931	0.926	2.034	0.854	2.141	0.782	2.251	0.712	2.363	----
31	1.363	1.496	1.297	1.570	1.229	1.650	1.160	1.735	1.090	1.825	1.020	1.920	0.950	2.018	0.879	2.120	0.810	2.226	0.741	2.333	----
32	1.373	1.502	1.309	1.574	1.244	1.650	1.177	1.732	1.109	1.819	1.041	1.909	0.972	2.004	0.904	2.100	0.836	2.203	0.769	2.306	----
33	1.383	1.508	1.321	1.577	1.258	1.651	1.193	1.730	1.127	1.813	1.061	1.900	0.994	1.991	0.927	2.085	0.861	2.181	0.796	2.281	----
34	1.393	1.514	1.333	1.580	1.271	1.652	1.208	1.728	1.144	1.808	1.079	1.891	1.015	1.978	0.950	2.069	0.885	2.162	0.821	2.257	----
35	1.402	1.519	1.343	1.584	1.283	1.653	1.222	1.726	1.160	1.803	1.097	1.884	1.034	1.967	0.971	2.054	0.908	2.144	0.845	2.236	----
36	1.411	1.525	1.354	1.587	1.295	1.654	1.236	1.724	1.175	1.799	1.114	1.876	1.053	1.957	0.991	2.041	0.930	2.127	0.868	2.216	----
37	1.419	1.530	1.364	1.590	1.307	1.655	1.249	1.723	1.190	1.795	1.131	1.870	1.071	1.948	1.011	2.029	0.951	2.112	0.891	2.197	----
38	1.427	1.535	1.373	1.594	1.318	1.656	1.261	1.722	1.204	1.792	1.146	1.864	1.088	1.939	1.029	2.017	0.970	2.098	0.912	2.180	----
39	1.435	1.540	1.382	1.597	1.328	1.658	1.273	1.722	1.218	1.789	1.161	1.859	1.104	1.932	1.047	2.007	0.990	2.085	0.932	2.164	----
40	1.442	1.544	1.391	1.600	1.338	1.659	1.285	1.721	1.230	1.786	1.175	1.854	1.120	1.924	1.064	1.997	1.008	2.072	0.952	2.148	----
45	1.475	1.566	1.430	1.613	1.383	1.666	1.336	1.720	1.287	1.776	1.238	1.835	1.189	1.895	1.139	1.958	1.089	2.022	1.038	2.088	----
50	1.505	1.585	1.462	1.628	1.421	1.674	1.378	1.721	1.335	1.771	1.291	1.822	1.246	1.875	1.201	1.930	1.136	1.996	1.110	2.044	----
55	1.528	1.601	1.490	1.641	1.452	1.681	1.414	1.724	1.374	1.768	1.334	1.814	1.294	1.861	1.253	1.909	1.212	1.959	1.170	2.010	----
60	1.549	1.616	1.514	1.652	1.480	1.689	1.444	1.727	1.408	1.767	1.372	1.808	1.335	1.850	1.298	1.894	1.260	1.939	1.222	1.984	----
65	1.567	1.629	1.536	1.662	1.503	1.696	1.471	1.731	1.438	1.767	1.404	1.805	1.370	1.843	1.336	1.882	1.301	1.923	1.266	1.964	----
70	1.583	1.641	1.554	1.672	1.525	1.703	1.494	1.735	1.464	1.768	1.433	1.802	1.401	1.838	1.369	1.874	1.337	1.910	1.305	1.948	----
75	1.598	1.652	1.571	1.680	1.543	1.709	1.515	1.739	1.487	1.770	1.458	1.801	1.428	1.834	1.399	1.867	1.369	1.901	1.339	1.935	----
80	1.611	1.662	1.586	1.688	1.560	1.715	1.534	1.743	1.507	1.772	1.480	1.801	1.453	1.831	1.425	1.861	1.397	1.893	1.369	1.925	----
85	1.624	1.671	1.600	1.696	1.575	1.721	1.550	1.747	1.525	1.774	1.500	1.801	1.474	1.829	1.448	1.857	1.422	1.886	1.396	1.916	----
90	1.635	1.679	1.612	1.703	1.589	1.726	1.566	1.751	1.542	1.776	1.518	1.801	1.494	1.827	1.469	1.854	1.445	1.881	1.420	1.909	----
95	1.645	1.687	1.623	1.709	1.602	1.732	1.579	1.755	1.557	1.778	1.535	1.802	1.512	1.827	1.489	1.852	1.465	1.877	1.442	1.903	----
100	1.654	1.694	1.634	1.715	1.613	1.736	1.592	1.758	1.571	1.780	1.550	1.803	1.528	1.826	1.506	1.850	1.484	1.874	1.462	1.898	----
150	1.720	1.747	1.706	1.760	1.693	1.774	1.679	1.788	1.665	1.802	1.651	1.817	1.637	1.832	1.622	1.846	1.608	1.862	1.593	1.877	----
200	1.758	1.779	1.748	1.789	1.738	1.799	1.738	1.809	1.718	1.820	1.707	1.831	1.697	1.841	1.686	1.852	1.675	1.863	1.665	1.874	----

K^* est le nombre des régresseurs (sans compter l'intercepte)

Annexe BTable des fractiles $f_{[\nu_1, \nu_2]}$ pour une loi $F_{[\nu_1, \nu_2]}$ (Fisher) (5%)

Cette table donne les valeurs pour lesquelles le pourcentage de la distribution de F dans le titre est supérieur à la valeur tabulaire de F pour ν_1 (degrés de liberté du numérateur) et ν_2 (degrés de liberté du dénominateur) associés au rapport F .

Degré de liberté (ν_1)	Degré de liberté (ν_2)										
	1	2	3	4	5	6	7	8	10	12	24
1	161	199	216	225	230	234	237	239	242	244	249
2	18.5	19.0	19.2	19.2	9.3	19.3	19.4	19.4	19.4	19.4	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.79	8.74	8.64
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.96	5.91	5.77
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.74	4.68	4.53
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.06	4.00	3.84
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.64	3.57	3.41
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.35	3.28	3.12
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.14	3.07	2.90
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.98	2.91	2.74
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.85	2.79	2.61
12	4.75	3.88	3.49	3.26	3.11	3.00	2.91	2.85	2.75	2.69	2.51
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.67	2.60	2.42
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.60	2.53	2.35
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.54	2.48	2.29
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.49	2.42	2.24
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.45	2.38	2.19
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.41	2.34	2.15
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.38	2.31	2.11
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.35	2.28	2.08
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.30	2.23	2.03
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.25	2.18	1.98
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.22	2.15	1.95
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.19	2.12	1.91
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.16	2.09	1.89
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.11	2.04	1.83
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.08	2.00	1.79
45	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.05	1.97	1.76
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.03	1.95	1.74
55	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.01	1.93	1.72

Annexe CTable des fractiles $t_{\nu}(p)$ pour une loi T (Student)

Cette table donne les points de pourcentage de la distribution de t avec ν degrés de liberté. Ce sont les valeurs de t pour les quelles un pourcentage donné, P , de la distribution de t est situé en dehors de la fourchette allant de $-t$ à $+t$. Au fur et à mesure que le nombre de degrés de liberté augmente, la distribution se rapproche de la distribution normale standard.

Degré de liberté (ν)	Test unilatéral		Test bilatéral	
	Pourcentage (P)			
	5%	1%	5%	1%
1	6.31	31.8	12.7	63.7
2	2.92	6.96	4.30	9.92
3	2.35	4.54	3.18	5.84
4	2.13	3.75	2.78	4.60
5	2.02	3.36	2.57	4.03
6	1.94	3.14	2.45	3.71
7	1.89	3.00	2.36	3.50
8	1.86	2.90	2.31	3.36
9	1.83	2.82	2.26	3.25
10	1.81	2.76	2.23	3.17
11	1.80	2.72	2.20	3.11
12	1.78	2.68	2.18	3.05
13	1.77	2.65	2.16	3.01
14	1.76	2.62	2.14	2.98
15	1.75	2.60	2.13	2.95
16	1.75	2.58	2.12	2.92
17	1.74	2.57	2.11	2.90
18	1.73	2.55	2.10	2.88
19	1.73	2.44	2.09	2.86
20	1.72	2.53	2.09	2.85
22	1.72	2.51	2.07	2.82
24	1.72	2.49	2.06	2.80
26	1.71	2.48	2.06	2.78
28	1.70	2.47	2.05	2.76
30	1.70	2.46	2.04	2.75
35	1.69	2.44	2.03	2.72
40	1.68	2.42	2.02	2.70
45	1.68	2.41	2.01	2.69
50	1.68	2.40	2.01	2.68
55	1.67	2.40	2.00	2.67
60	1.67	2.39	2.00	2.66
∞	1.64	2.33	1.96	2.58

Annexe DValeurs significatives du coefficient de corrélation

Cette table présente les valeurs au-delà desquelles le coefficient de corrélation est déclaré significatif, pour un seuil de signification déterminé et un nombre donné de paires d'observations de x et y

<i>n</i>	.1	.05	.02	.01	.001
1	.9877	.9969	.9995	.9999	.9999
2	.9000	.9500	.9800	.9900	.9990
3	.8054	.8783	.9343	.9587	.9912
4	.7293	.8114	.8822	.9172	.9741
5	.6694	.7545	.8329	.8745	.9507
6	.6215	.7067	.7887	.8343	.9249
7	.5822	.6664	.7498	.7977	.8982
8	.5494	.6319	.7155	.7646	.8721
9	.5214	.6021	.6851	.7348	.8471
10	.4973	.5760	.6581	.7079	.8233
11	.4762	.5529	.6339	.6835	.8010
12	.4575	.5324	.6120	.6614	.7800
13	.4409	.5139	.5923	.6411	.7603
14	.4259	.4973	.5742	.6226	.7420
15	.4124	.4821	.5577	.6055	.7246
16	.4000	.4683	.5425	.5897	.7084
17	.3887	.4555	.5286	.5751	.6932
18	.3783	.4438	.5155	.5614	.6787
19	.3687	.4329	.5034	.5487	.6652
20	.3598	.4227	.4921	.5368	.6524
25	.3233	.3809	.4451	.4869	.5974
30	.2960	.3494	.4093	.4487	.5541
35	.2746	.3246	.3810	.4182	.5189
40	.2573	.3044	.3578	.3932	.4896
45	.2428	.2875	.3384	.3721	.4648
50	.2306	.2732	.3218	.3541	.4433
60	.2108	.2500	.2948	.3248	.4078
70	.1954	.2319	.2737	.3017	.3799
80	.1829	.2172	.2565	.2830	.3568
90	.1726	.2050	.2422	.2673	.3375
100	.1638	.1946	.2301	.2540	.3211

Annexe EProcédure de recherche de l'influence des variables
en modèle de régression linéaire

La méthode de recherche d'influence des variables en modèle de régression linéaire consiste à pondérer chacun des coefficients du modèle par la valeur moyenne associée de la variable. Nous obtenons ainsi six coefficients pondérés comme suit :

$$C_j = C_j \cdot \bar{x}_j$$

$j = 1 \dots p$ les paramètres de modèle Pui^{-0,57}, T₀, C_p, COT, CM et Dms

$i = 1 \dots n$ le nombre d'unités statistiques ($n = 37$)

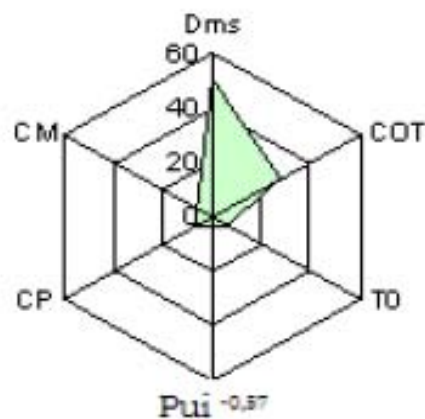
C_j = les coefficients du modèle

\bar{x}_j = la moyenne du paramètre

L'influence $I_{x_j}(\%)$ de chaque variable de l'équation est ensuite calculée de la

manière suivante :
$$I_{x_j}(\%) = \frac{|c_j|}{\sum_{j=1}^p |c_j|} \cdot 100$$

La figure 5.7 montre l'influence en pourcentage des différents paramètres de l'équation de régression



Annexe FProcédure de calcul de l'importance des variables
par approche de poids de connexion.

La procédure de calcul consiste à multiplier la valeur du poids de connexion des neurones cachée-sortie, pour chaque neurone caché (tableau 1), par les valeurs des poids de connexion de la couche entrée-caché (tableau 2). En effectuant cette démarche pour chaque neurone d'entrée, nous identifions sa contribution à la sortie. Le produit effectué donne les résultats présentés dans le tableau 3.

Tableau 1 : Matrice des poids de connexion neurones cachés-sortie.

	N1	N2	N3	N4	N5	N6	N7	N8
sortie	0,6843	-0,4727	-0,8284	-9,0024	0,2741	8,8541	0,5860	1,7602

Tableau 2 : Matrice des poids de connexion entrée-neurones cachés.

	N1	N2	N3	N4	N5	N6	N7	N8
Energie	-2,5460	1,8893	0,0644	0,8517	-1,9928	-2,0998	1,4324	-0,5068
T0	-0,4498	2,3462	0,1819	-2,7695	1,3357	1,9834	1,1252	-2,6985
Cp	-2,8879	1,4904	-1,6986	-0,9473	3,3761	-0,6641	-4,9773	2,9211
CM	0,5039	2,0939	-3,1134	-3,3326	2,4514	-1,1407	0,1564	3,5089
COT	0,7516	-3,5191	0,3691	-6,0042	0,8838	0,4441	0,2584	-1,2082
DMS	-3,0133	-0,4746	0,8395	-2,5803	-11,8988	6,1760	0,2626	4,1138
SSP	-3,6942	-4,0120	4,3907	-2,0074	-0,0119	6,4955	1,1792	-19,5264
KAOLIN	-1,1324	0,8167	-0,0483	-0,5254	-0,6161	-0,5541	-0,9444	-0,0363
ILLite	-0,0532	-0,8673	-5,2539	2,8720	-0,3219	1,1960	-2,4173	-2,6230
Montmoëillo	2,2745	-0,8763	-1,0882	3,7742	0,1155	-14,8916	1,2301	-1,1770
Arg	-0,5968	-2,3724	-2,3895	0,3749	0,1221	-0,1854	0,4620	0,1165

Tableau 3 : Matrice résultante du produit des poids de connexions

	N1	N2	N3	N4	N5	N6	N7	N8
Energie	-1,7422	-0,8931	-0,0534	-7,6671	-0,5462	-18,5921	0,8394	-0,8920
T0	0,3078	-1,1090	-0,1507	24,9319	0,3661	17,5612	0,6593	-4,7499
Cp	-1,9762	-0,7045	1,4071	8,5276	0,9254	-5,8801	-2,9167	5,1417
CM	0,3448	-0,9898	2,5792	30,0012	0,6719	-10,1000	0,0917	6,1764
COT	0,5143	1,6635	-0,3058	54,0527	0,2422	3,9323	0,1514	-2,1266
DMS	-2,0620	0,2244	-0,6954	23,2287	-3,2615	54,6831	0,1539	7,2410
SSP	-2,5279	1,8965	-3,6373	18,0715	-0,0033	57,5119	0,6910	-34,3704
KAOLIN	-0,7749	-0,3860	0,0400	4,7299	-0,1689	-4,9058	-0,5534	-0,0639
ILLite	-0,0364	0,4100	4,3523	-25,8547	-0,0882	10,5897	-1,4166	-4,6170
Montmoëillo	1,5564	0,4142	0,9015	-33,9770	0,0317	-131,8516	0,7208	-2,0717
Arg	-0,4084	1,1215	1,9795	-3,3754	0,0335	-1,6416	0,2707	0,2051

On effectue ensuite la somme des produits précédents pour chaque neurone d'entrée suivant

On peut alors quantifier l'importance ou la distribution de tous les poids de sortie attribués aux variables d'entrée données. Les résultats obtenus, pour chaque variable étudiée sont présentés dans le tableau 4

Matrice résultante du poids des connexions

Tableau 4 : Matrice résultante de la somme du produit des poids de connexions

Dms	79,51
COT	58,12
Ssp	37,63
T0	37,20
CM	28,78
Cp	4,52
Arg	-1,82
Kaolinite	-2,08
illite	-16,67
Puissance	-29,55
Mont.	-164,28

