

**Approche quantitative pour la prédiction et la cartographie
du drainage des sols par l'intégration de données ASTER et RADARSAT-1**

M.A.Z. Niang¹ et M. Bernier¹

¹ Institut National de la Recherche Scientifique

Centre Eau, Terre et Environnement

490 de la Couronne, Québec (Québec), G1K 9A9

avec la collaboration scientifique de

M.C. Nolin², O. Ka² et J. Liu³

² Agriculture et Agroalimentaire Canada

Laboratoires de pédologie et d'agriculture de précision

979 ave de Bourgogne, local #140, Québec (Québec), G1W 2L4

³ Agriculture et Agroalimentaire Canada

960 Carling Ave., Ottawa (Ontario), K1A 0C6

Rapport de recherche No R-905

Décembre 2006

ISBN : 2-89146-530-X

TABLE DE MATIÈRES

TABLE DE MATIÈRES	iii
LISTE DES TABLEAUX	v
LISTE DES FIGURES	ix
RÉSUMÉ	xi
1 INTRODUCTION.....	1
2 MATÉRIELS ET MÉTHODES	5
2.1 <i>La base de données d'apprentissage</i>	5
2.1.1 <i>Territoire à l'étude</i>	5
2.1.2 <i>Base de données de sols</i>	6
2.1.3 <i>Les données RADARSAT-1</i>	9
2.1.4 <i>Les données de suivi de l'humidité du sol lors de l'acquisition des images RADARSAT-1</i>	11
2.1.5 <i>Les données ASTER</i>	13
3 MÉTHODOLOGIE.....	15
3.1 <i>Analyse statistique des données de sol</i>	15
3.1.1 <i>Étude de corrélation et codage des données</i>	15
3.1.2 <i>Analyse des correspondances multiples (ACM)</i>	16
3.1.3 <i>Comparaison entre la classification experte et différentes classifications automatiques du drainage des sols</i>	16
3.2 <i>Analyse discriminante générale</i>	16
3.2.1 <i>Étude de la normalité avec l'ensemble des données</i>	17
3.2.2 <i>Le problème de multicolinéarité</i>	18
3.2.3 <i>Homogénéité des matrices de variances-covariances</i>	19
4 RÉSULTATS ET DISCUSSIONS.....	21
4.1 <i>Évaluation de la cohérence du modèle de classification du drainage des sols par des experts et proposition d'un modèle de classification automatique</i>	21
4.2 <i>Résultats de la classification</i>	26

4.3	<i>Résultats d'analyse des données RADARSAT-1</i>	29
4.3.1	<i>Les données de suivi</i>	29
4.3.2	<i>Rapport de bandes</i>	32
4.3.3	<i>L'ensemble des données radar</i>	34
4.4	<i>Résultats d'analyse des données ASTER</i>	35
4.4.1	<i>Statistiques générales</i>	35
4.4.2	<i>Analyse en composantes principales (ACP)</i>	37
4.4.3	<i>ACP en fonction des classes de drainage et du type d'occupation du sol</i>	39
4.5	<i>Résultats de l'étude de la normalité des données</i>	41
4.5.1	<i>Test avec la droite de Henry et la statistique (W) de Shapiro-Wilk</i>	41
4.5.2	<i>Transformation logistique (Logit)</i>	43
4.6	<i>Résultats de l'analyse discriminante générale</i>	46
4.6.1	<i>Résultats de comparaison entre transformation logit sans transformation</i>	47
4.6.2	<i>Résultats détaillés en fonction de l'occupation du sol</i>	49
4.6.3	<i>Résultats détaillés des données d'OT pour KMEANS4 et KMEANS3</i>	50
5	CONCLUSION GÉNÉRALE.....	53
	RÉFÉRENCES	55

LISTE DES TABLEAUX

Tableau 1 : Les propriétés des sols choisies pour cette étude	7
Tableau 2. Effectif et pourcentage cumulé de chaque classe de drainage (classification expert).....	8
Tableau 3 : Effectif et pourcentage cumulé de chaque type d'occupation du sol de l'année 2005	9
Tableau 4 : Effectif et pourcentage cumulé selon deux classes générales d'occupation du sol de l'année 2005.	9
Tableau 5 : Date et mode d'acquisition des 6 images RADARSAT-1 ainsi que la quantité de précipitation le jour précédent ainsi que la journée du passage du satellite.	10
Tableau 6 : Occupation du sol et état de surface des parcelles au moment du passage du satellite RADARSAT-1 le 3 mai et le 27 mai (mode S2a).....	12
Tableau 7 : Classification des sols et des pédo-paysages présents dans les différents sites choisis pour mesurer l'humidité du sol lors de l'acquisition des images RADARSAT-1	12
Tableau 8 : Spécifications techniques de l'image Aster du 22 juin 2005	13
Tableau 9 : Mesures d'association non paramétriques entre la classe de drainage des sols et quelques propriétés morphologiques de sol.	21
Tableau 10 : Résultats de l'analyse de correspondances multiples sur les propriétés de sol les plus corrélées avec le drainage.....	22
Tableau 11 : Codes des classes utilisées dans la figure 5 et leurs descriptions	25
Tableau 12 : Matrice des observations classées différemment par la classification automatique (A3 à A7) et par la classification experte (C3-C7).....	26
Tableau 13 : Étude de la concordance (diagonale) entre la classification experte (C3 à C7) et celle issue de la méthode automatique (A3 à A7).....	27
Tableau 14 : Étude de la concordance (diagonale) entre la classification experte et celle issue de la méthode K-MEANS à 4 classes au moyen d'une table de contingence χ^2	27

Tableau 15 : Étude de la concordance (diagonale) entre la classification experte et celle issue de la méthode K-MEANS à 3 classes au moyen d'une table de contingence χ^2	28
Tableau 16 : Classification manuelle des classes de drainage en fonction de la profondeur du gley (PGLE)	28
Tableau 17 : Croisement de la classification experte (DRAI) avec celle de PGLE à 4 classes (GL4 à GL7) au moyen d'une table de contingence χ^2	29
Tableau 18 : Effectif et somme des rangs pour les différentes classes de drainage.....	33
Tableau 19 : Statistiques générales décrivant les 9 bandes spectrales de l'image ASTER acquise en juin 2005.....	36
Tableau 20 : Coefficient de corrélation de Pearson mesuré entre les 9 bandes spectrales de l'image ASTER acquise en juin 2005.....	36
Tableau 21 : Poids factoriels des bandes spectrales d'ASTER pour les deux premiers facteurs....	38
Tableau 22 : Significativité statistique des données d'OT dans la régression multiple avec liaison logit.....	44
Tableau 23 : Qualité d'ajustement de la transformation logit pour les différentes classifications	45
Tableau 24 : Probabilité à priori associées aux différents types de classification	47
Tableau 25 : Étude de la significativité statistique des données avec le test F de Fisher	48
Tableau 26 : Résultats de comparaison de l'analyse discriminante entre les données transformées par logit et celles non transformées.	48
Tableau 27 : Résultats de l'analyse discriminante en fonction de l'occupation du sol pour la classification DRAI.....	49
Tableau 28 : Résultats de l'analyse discriminante en fonction de l'occupation du sol pour la classification PGLE.....	49
Tableau 29 : Résultats de l'analyse discriminante en fonction de l'occupation du sol pour la classification KMEANS4.....	50

Tableau 30 : Résultats de l'analyse discriminante en fonction de l'occupation du sol pour la classification KMEANS3	50
Tableau 31 : Résultats de classification du drainage (modèle de classification = KMEANS4) en condition de sols nus suite à une analyse discriminante appliquée à diverses combinaisons de données d'OT	51
Tableau 32 : Résultats de classification du drainage (modèle de classification = KMEANS4) sous couvert de végétation suite à une analyse discriminante appliquée à diverses combinaisons de données d'OT	51
Tableau 33 : Résultats de l'analyse discriminante en fonction des données d'OT pour les sols nus avec la classification KMEANS3	52
Tableau 34 : Résultats de l'analyse discriminante en fonction des données d'OT pour la végétation avec la classification KMEANS3	52

LISTE DES FIGURES

Figure 1 :	Localisation du territoire à l'étude; le bassin versant du Bras d'Henri, QC.	5
Figure 2 :	Classification physiographique du territoire à l'étude.	6
Figure 3 :	Précipitations et températures minimales et maximales du mois de mai 2005.	10
Figure 4 :	Représentation des observations dans le plan des deux axes principaux (facteurs) définis à partir de l'ACM.	23
Figure 5 :	Représentation des observations dans le plan des deux premiers axes les plus discriminants. Les codes des classes utilisées sont décrits dans le tableau 11 suivant.	24
Figure 6 :	Rétrodiffusion radar des parcelles tests extraites des images RADARSAT-1.....	30
Figure 7 :	a) Simulation de la rétrodiffusion radar avec le modèle IEM avec $s=1\text{cm}$ et $L=8\text{ cm}$ pour l'image du 3 mai et; b) avec $s=0.9\text{ cm}$ et $L=10\text{ cm}$ pour l'image du 27 mai.	31
Figure 8 :	Médiane, quartile et étendue des différentes classes de drainage en fonction du rapport de bandes.....	33
Figure 9 :	Moyenne des rétrodiffusions radar pour les différentes images RADARSAT-1 en fonction des classes de drainage des profils de sol en condition de sol nus.....	34
Figure 10 :	Moyenne des rétrodiffusions radar pour les différentes images RADARSAT-1 en fonction des classes de drainage des profils de sol sous couverture de végétation	35
Figure 11 :	a) Tracé des valeurs propres (a) et pourcentage de la variance originale (b) associé aux neuf composantes extraites de l'ACP à partir des bandes spectrales de l'image d'ASTER	37

Figure 12 : Projection des variables sur le plan factoriel réduit défini à partir des deux premières composantes extraites de l'ACP38

Figure 13 : Classification du drainage du sol dans l'espace réduit défini par les facteurs 1 et 2 extraits de l'ACP a) pour les sols nus et b) pour les sols sous couvert de végétation40

Figure 14 : Médiane, quartile et étendue des différentes classes de drainage en fonction a) du facteur 1 et b) du facteur 2 extraits de l'ACP.....40

Figure 15 : Classification du type d'occupation du sol dans l'espace réduit défini par les facteurs 1 et 2 extraits de l'ACP. LANDUSE2005 : Occupation du sol au printemps 2005. N = sols nus. N*= sols nus . H = sols humides. C = céréales. F = forêts. P = prairies.41

Figure 16 : Test de normalité avec la droite de Henry et la statistique (W) de Shapiro-Wilk pour les six images RADARSAT-1 acquises au printemps 2005.....42

Figure 17 : Test de normalité avec le droite de Henry et la statistique (W) de Shapiro-Wilk pour les deux premiers facteurs extraits de l'ACP à partir des bandes spectrales de l'image ASTER.43

Figure 18 : Statistiques du χ^2 de Pearson en fonction des valeurs théoriques prévues par la régression multiple a) pour la classe 1 de DRAI et b) pour la classe 6 de DRAI46

RÉSUMÉ

Dans cette étude, une méthode quantitative pour la prédiction et la cartographie des classes de drainage utilisant les données d'Observation de la Terre (OT) RADARSAT-1 et ASTER comme données auxiliaires est présentée. L'approche consiste en l'application de l'analyse discriminante sur une base de données de caractérisation du drainage du bassin versant du Bras-d'Henri (Québec) mise en place par le laboratoire de pédologie et d'agriculture de précision (LPAP) d'Agriculture et Agroalimentaire Canada. Une étude préalable de corrélation entre les classes de drainage fournies par le LPAP et certaines propriétés du sol a été faite pour dégager celles qui sont directement en relation avec le drainage. Par la suite, l'application d'une analyse des correspondances multiples après un codage disjonctif de ces propriétés a permis d'étudier la cohérence entre la classification proposée par les experts du LPAP et celle de la méthode automatique des KMEANS. D'une manière générale, la classification proposée correspond relativement bien à la classification automatique obtenue par arbre de décision. Étant donné que près de 87% des observations correspondaient aux classes mal drainées, cette étude est axée sur ces observations à cause de leur représentativité statistique. Une classification KMEANS à 4 classes puis à 3 classes sont alors analysées ainsi que celle de la propriété du sol la plus corrélée au drainage qui est la profondeur du gley. Une analyse discriminante respectant les principales hypothèses théoriques ainsi qu'une standardisation des échelles de mesure requises pour son application a été par la suite effectuée sur les données OT avec les clusters issus de ces classifications. Les résultats obtenus montrent que les classes associées au mauvais drainage des sols sont bien identifiées par les données d'OT. La classification KMEANS à 4 classes fournit 88,06 % et 63,75% de bonnes classifications respectivement en condition de sols nus et sous couvert de végétation tandis que la classification KMEANS à 3 classes fournit elle 72,27% et 74,70% de bonnes classifications respectivement en condition de sols nus et sous couvert de végétation.

1 Introduction

Le drainage est une propriété importante du sol affectant considérablement la croissance des plantes et influençant plusieurs qualités de sol ainsi que la vulnérabilité des sols à la dégradation. Une cartographie précise de cette propriété pouvant être effectuée au moindre coût possible est d'un intérêt majeur aussi bien pour une gestion efficace des périmètres agricoles que celle de l'environnement en général.

La classification des sols en sept classes de drainage (Day et McMenamin, 1982), est établie selon un ensemble de critères morphologiques tels que la couleur de la matrice, la présence ou l'absence de marbrures, la profondeur au gley, la texture, le contenu en matière organique, la capacité de rétention en eau du sol ainsi que la vitesse de retrait de l'eau du sol (longueur de la période de l'année en état de saturation). L'acquisition de ces informations nécessite souvent des moyens humains et matériels considérables surtout pour des zones étendues et éloignées. De plus, si les données sont de faible résolution et non représentatives, la cartographie du drainage pourrait être affectée par des erreurs liées aux interpolations.

Ces dernières années, on assiste plutôt à des recherches de plus en plus orientées vers des méthodes d'intégration de diverses données pour faciliter la classification et la prédiction du drainage. Ces méthodes utilisent des informations auxiliaires plus accessibles et moins coûteuses. C'est dans ce sens que s'inscrit cette étude. L'utilisation des données de télédétection semble l'option la plus prometteuse car plusieurs techniques éprouvées aussi bien en optique que dans le domaine des microondes ont démontré des relations entre les observations satellitaires et certaines propriétés du sol sur lesquelles sont basées la classification du drainage. En outre, avec sans cesse des lancements de nouveaux capteurs satellitaires, ayant des résolutions spectrales et spatiales de plus en plus fines, on gagne à la fois une couverture continue à l'échelle régionale et une bonne répétitivité temporelle, deux attributs nécessaires pour le suivi de la qualité des sols et des diverses productions végétales.

La télédétection dans le domaine des microondes est de prime à bord l'avenue la plus attrayante car la constante diélectrique que mesure ces capteurs est directement fonction de l'humidité du

sol; propriété qui caractérise le mieux le drainage. De plus, les mesures radar ne sont pas affectées par les conditions atmosphériques. Dans ce domaine, on trouve plusieurs travaux dans la littérature (Ulaby et al., 1996; Njoku et Entekhabi, 1996; Niang, 2001).

Dans le domaine optique, l'humidité du sol influence de façon significative la réflectance spectrale dans le visible et le proche infrarouge ainsi que dans le moyen infrarouge (Browers and Hanks, 1965; Skidmore et al., 1975). La couleur du sol a par ailleurs, un lien significatif sur la réflectance spectrale dans le visible et le proche infrarouge (Escadafal, et al., 1989; Escadafal, et al., 1990; Gerbermann and Neher, 1979). Il est à noter que cette propriété peut fournir une information très valable sur l'état du drainage d'un sol.

D'une manière générale, les données d'Observations de la Terre (OT) radar et optique constituent une source intéressante de données pouvant aider à la cartographie du drainage. Cependant, la quantification des certaines propriétés comme l'humidité ou la couleur du sol reste toujours difficile à cause des variabilités spatiales des autres propriétés biophysiques (rugosité et végétation) et chimiques (contenu en matière organique, minéralogie) du même sol. Par ailleurs, du point de vue spectral, l'information d'une seule source d'image satellitaire ne permet pas d'appréhender toute la problématique du drainage à cause de la limitation spectrale pour une détection globale des propriétés morphologiques impliquées dans la classification du drainage (couleur, humidité, texture, matière organique, oxyde de fer, ect...), d'où la nécessité d'utiliser une approche intégrant différentes sources de données.

Comme dans tout problème d'intégration de données multi-sources, à savoir ici les données des capteurs RADARSAT-1 et ASTER, le problème principal réside dans la formulation des règles de décision. Une approche quantitative axée sur l'analyse discriminante a été utilisée pour la formulation des règles de classification à partir de ces données auxiliaires.

L'utilisation de l'analyse discriminante comme méthode pour la prédiction et la classification du drainage des sols n'est pas nouvelle. On peut citer les travaux de Bell et al. (1992, 1994) qui utilisent comme données auxiliaires la topographie et le matériau originel et plus récemment, Kravchenko et al. (2002) qui utilisent en plus des données topographiques, des données de conductivité électrique apparente des sols. Ces études ont montré des résultats très intéressants d'autant plus que les données auxiliaires sont facilement accessibles et mesurées très densément

si ce n'est de façon continue comme dans le cas de l'imagerie. La seule limitation de ces études demeure la taille des territoires étudiés qui est relativement plus petite (14 000 ha dans les travaux de Bell et. al. (1992, 1994) et une zone d'étude de 20 ha au milieu d'une surface agricole de 259 ha dans l'étude de Kravchenko et.al. 2002) comparativement à la présente étude qui cible un bassin versant de 15 000 ha. Par ailleurs, dans leurs essais de classification du drainage des sols, les classes issues de leur expertise n'ont pas fait l'objet de discussion ou de comparaison avec d'autres méthodes de classification.

L'objectif principal de la présente étude est d'évaluer l'apport des données d'OT dans l'identification des classes de drainage du sol. L'originalité de l'approche consiste à analyser la classification experte fournie par les spécialistes de l'organisme responsable de l'inventaire des sols du secteur à l'étude (laboratoires de pédologie et d'agriculture de précision (LPAP) d'Agriculture et Agroalimentaire Canada (AAC)) et à proposer une nouvelle façon d'intégrer les différentes données morphologiques enregistrées lors de la description des profils de sols et prises en compte dans le diagnostic de classification du drainage du sol au champ. La démarche méthodologique retenue commence par une comparaison de la classification experte avec diverses classifications automatiques réalisées à partir de méthodes quantitatives appliquées sur les propriétés du sol les plus pertinentes. Par la suite, une classification du drainage des sols à partir des données radar et optiques sera réalisée d'abord individuellement puis de façon intégrée. La performance de discrimination de ces différentes approches en fonction du type d'occupation du sol (sols nus, céréales, prairies et pâturages, forêt et terres humides) sera également étudiée. La technique d'analyse discriminante sera utilisée à cet effet. Une partie importante de l'étude va être axée sur la qualité d'ajustement des données afin de s'assurer de la fiabilité des résultats.

2 Matériels et Méthodes

2.1 La base de données d'apprentissage

2.1.1 Territoire à l'étude

Le territoire à l'étude, le bassin versant (BV) du Bras d'Henri présente une superficie de 167 km². Il est situé au sud de la ville de Québec, dans la région administrative de Chaudière-Appalaches. (Figure 1). La rivière Bras d'Henri est un tributaire de la rivière Beauvillage, laquelle se jette dans la rivière Chaudière. Le BV du Bras d'Henri couvre deux régions physiographiques; les Appalaches au sud, région caractérisée par une topographie ondulée (>5 % de pente en moyenne) et les Basses-Terres du Saint-Laurent au nord, présentant généralement un relief plutôt plat (2-5% de pente). Entre ces deux régions, se trouve une zone de transition présentant des caractéristiques communes à chacune de ces deux régions physiographiques (Figure 2). Deux micro-bassins (MB) d'environ 3 km² ont été sélectionnés à l'intérieur du bassin pour y mener une expérimentation visant à comparer l'effet de l'implantation de bonnes pratiques de gestion sur la qualité de l'eau de surface (projet WEB's). Ces deux MB jumeaux, désignés ultérieurement MB d'intervention (MB_I) et MB témoin (MB_T), sont situés dans la région physiographique des Basses-Terres du Saint-Laurent (Figure 2). Ces deux MB ont fait l'objet d'une étude pédologique plus intensive.

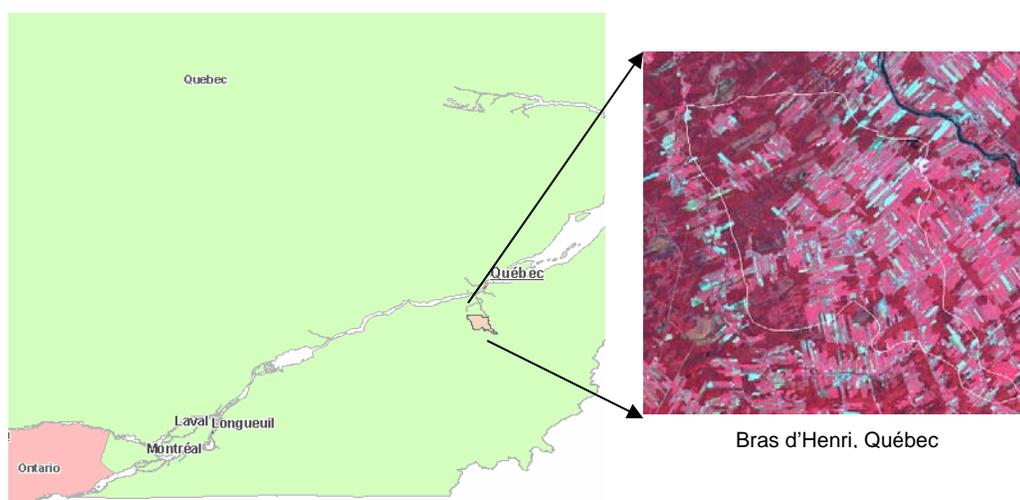


Figure 1 : Localisation du territoire à l'étude; le bassin versant du Bras d'Henri, QC.

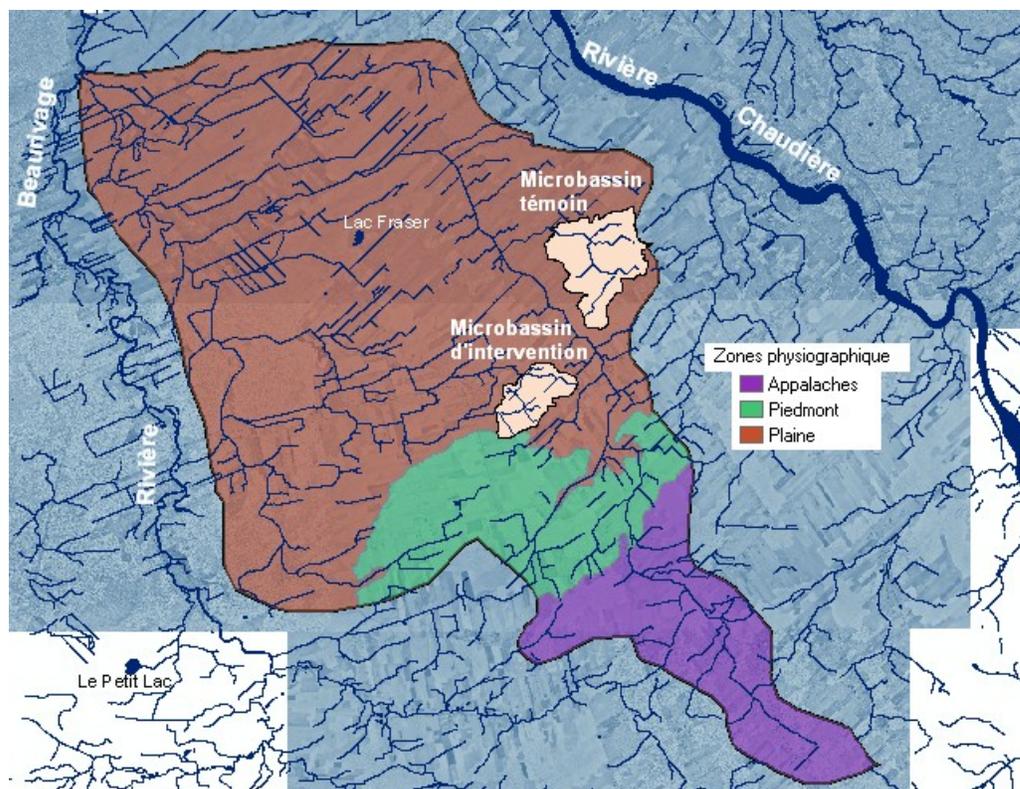


Figure 2 : Classification physiographique du territoire à l'étude.

2.1.2 Base de données de sols

Une base de données sur les sols décrivant les propriétés morphologiques de 1612 profils de sols était disponible pour réaliser cette étude. Cette information a été acquise au cours des étés 2004 et 2005 en suivant la méthode des transects aléatoires stratifiés (Nolin et al., 1994) dans le cadre d'un relevé pédologique semi-détaillé du bassin versant du Bras d'Henri (coordonnées) réalisé à une échelle de 1 :40 000 à une densité de prospection de 1 profil par 16 ha ainsi que lors du relevé détaillé des deux MB expérimentaux qui a été effectué à une échelle de 1 : 15,000 avec une densité de prospection de 1 profil par 2,25 ha (Figure 3). Dans cette base de données, plus 60 propriétés des sols du bassin ont été répertoriées et géoréférencées. Parmi celles-ci, 14 propriétés morphologiques étroitement liées à la classification du drainage du sol ont été sélectionnées (Tableau 1).

Tableau 1 : Les propriétés des sols choisies pour cette étude

PROPRIÉTÉ	DESCRIPTION	NBRE DE CLASSES
DRAI	Classification du drainage	5
PGLE	Profondeur au gley (0-100 cm)	Reclassifiée en 4 ¹
AVAL	Valeur de la couleur de l'horizon A	9
ACHR	Chroma de la couleur de l'horizon A	8
BVAL	Valeur de la couleur de l'horizon B	9
BCHR	Chroma de la couleur de l'horizon B	10
CVAL	Valeur de la couleur de l'horizon C	9
CCHR	Chroma de la couleur de l'horizon C	8
ORDRE	Classification taxonomique des sols (Ordre)	4
AFRAGMENT	Contenu en fragments grossiers de l'horizon A	7
AMO	Teneur en matière organique de l'horizon A	4
ATEXTURE	Texture de l'horizon A	6
BTEXTURE	Texture de l'horizon B	6
CTEXTURE	Texture de l'horizon C	6

Sur les 1612 profils exposés à la pelle, soit la partie supérieure du profil, *i.e.* la couche de surface (horizon A) et le sous-sol (horizon B), ou à la tarière, soit le substratum (horizon C), les analyses faites par les spécialistes du LPAP d'AAC, ont permis d'identifier cinq classes de drainage de sols dans le BV du Bras d'Henri : la classe 3 (bien drainé); la classe 4 (modérément bien drainé); la classe 5 (imparfaitement drainé); la classe 6 (mal drainé) et la classe 7 (très mal drainé). Le tableau 2 donne les effectifs et les pourcentages cumulés de chacune de ces classes.

¹ La profondeur du gley (PGLE) est une donnée semi-quantitative exprimée en classe de 10 cm; elle a été reclassifiée selon quatre classes voir paragraphe 4-2

Tableau 2. Effectif et pourcentage cumulé de chaque classe de drainage (classification expert)

<i>Classes</i>	<i>Effectif</i>	<i>Effectif cumulé</i>	<i>%</i>	<i>% cumulé</i>
3	73	73	4.53	4.53
4	131	204	8.13	12.66
5	457	661	28.35	41.01
6	699	1360	43.36	84.37
7	252	1612	15.63	100.00

Nous constatons que les deux premières classes (3 et 4) sont faiblement représentées (12.66% de l'effectif total). De plus 87.34 % des observations correspondent aux trois autres classes avec une prédominance de la classe 6 (43.36%), ce qui signifie que les sols du bassin versant du Bras d'Henri sont majoritairement mal drainés. Les autres types de classification dont celles dérivées des autres données morphologiques ainsi que celles générées à partir des images satellitaires permettront de confirmer ce constat préliminaire.

Dans cette base de données, un inventaire détaillé et semi-détaillé de l'occupation du sol a également été réalisé pour chacun des points d'observation (Tableaux 3 et 4). Force est de constater que 76.67% des profils étudiés se situaient en zone sous couvert de végétation. Ce constat traduit très bien la mixité de l'agriculture dans le BV du Bras d'Henri, où l'entreprise agricole typique est une ferme laitière avec élevage porcin, donc nécessitant pour la soutenir une production végétale constituée à la fois de plantes fourragères et de grandes cultures (céréales, soya et maïs). Comme la présence d'une couverture végétale peut affecter les modèles développés pour prédire le drainage des sols, une étude des performances de discrimination des données d'OT selon le type d'occupation du sol s'avère donc utile à effectuer pour mesurer l'impact de l'occupation du sol sur les différents modèles de classification.

Tableau 3 : Effectif et pourcentage cumulé de chaque type d'occupation du sol de l'année 2005

<i>Occupation^x</i>	<i>Effectif</i>	<i>Effectif cumulé</i>	<i>%</i>	<i>% cumulé</i>
N	327	327	20.29	20.29
F	483	810	29.96	50.25
H	44	854	2.73	52.98
C	387	1241	24.01	76.99
P	322	1563	19.98	96.96
N*	49	1612	3.04	100.00

^x N : sols nus, F : forêt, H : zone humide, C : céréales, P : prairies, N* : sols nus avec faible couverture végétale.

Tableau 4 : Effectif et pourcentage cumulé selon deux classes générales d'occupation du sol de l'année 2005.

<i>Occupation*</i>	<i>Effectif</i>	<i>Effectif cumulé</i>	<i>%</i>	<i>% cumulé</i>
N	376	376	23.33	23.33
V	1236	1612	76.67	100.00

* Classes d'occupation : sols nus (N) et sous couverture de végétation (V).

2.1.3 Les données RADARSAT-1

Six images RADARSAT-1 en polarisation HH ont été acquises pour cette étude. Le tableau 5 précise la date et le mode d'acquisition de ces images ainsi que la quantité de précipitation observée le jour précédent ainsi que la journée même de l'acquisition de l'image satellitaire. La figure 3 montre les précipitations, les températures minimales et maximales du mois de mai 2005. Les données météo correspondantes n'étaient pas disponibles pour le mois de juin. Ce graphique montre que les conditions d'état de surface sont bien différentes selon les dates d'acquisition. L'impact de ces conditions météorologiques variables sur les données radar sera étudié.

Tableau 5 : Date et mode d'acquisition des 6 images RADARSAT-1 ainsi que la quantité de précipitation le jour précédent ainsi que la journée du passage du satellite.

Date	Mode	Polarisation	Précipitation
03-mai	S2a	HH	0 / 1.8mm
10-mai	S1a	HH	0 / 0
24-mai	S3d	HH	0 / 3.8
27-mai	S2a	HH	0 / 0
10-juin	S2d	HH	NA/NA
17-juin	S3d	HH	NA/NA

NA = Données de précipitation non disponibles.

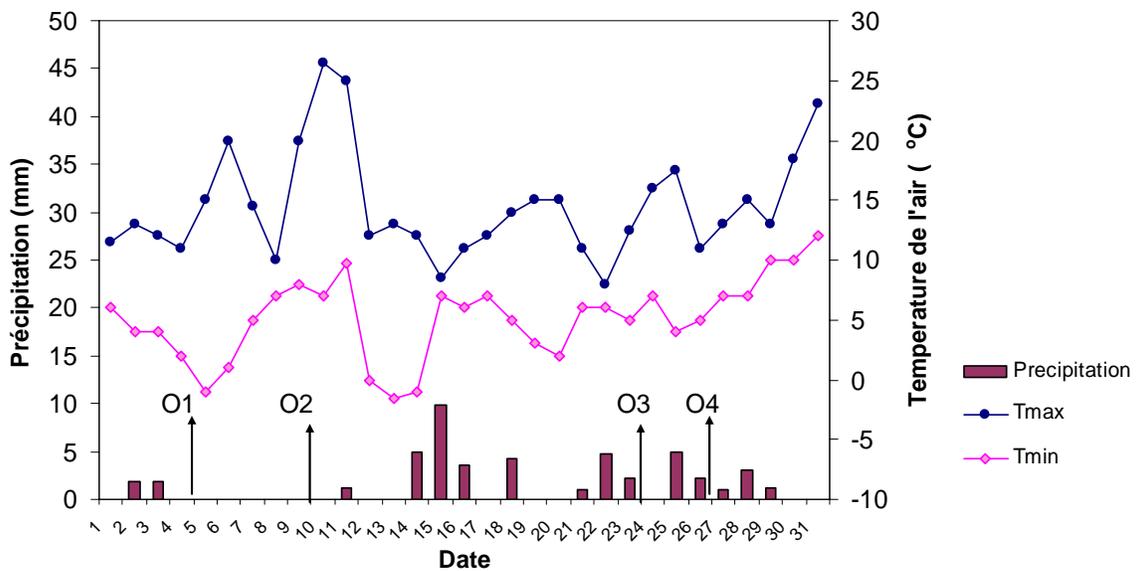


Figure 3 : Précipitations et températures minimales et maximales du mois de mai 2005.

Pour rendre exploitables les données radar, des prétraitements géométriques et radiométriques ont été appliqués sur les images brutes avec le logiciel PCI-Geomatica. Une ortho-rectification géométrique a été effectuée pour rendre les images superposables entre elles ainsi qu'avec les images ASTER et les autres données cartographiques. La projection UTM (WGS84) zone 19N a

été choisie comme projection de référence. Par la suite, un filtre Gamma 3x3 a été appliqué sur les images avant l'extraction du coefficient de rétrodiffusion radar afin de réduire l'effet de chatoiement caractéristique des images radars. Il est à noter que seules les images du 3 mai et du 27 mai ont été acquises selon le même mode (S2a); ainsi il n'y a pas de variation des coefficients de rétrodiffusion liée aux effets d'angle d'incidence entre ces deux images. Un rapport de bandes permettrait de diminuer les effets liés aux variations de la rugosité et de conserver ainsi que l'effet directement lié à la teneur en eau liquide de la couche de surface du sol (Bernier et Fortin, 1998). L'efficacité de ce rapport à la discrimination des classes de drainage du sol sera étudiée, malgré la présence de précipitations entre les deux dates d'acquisition.

2.1.4 Les données de suivi de l'humidité du sol lors de l'acquisition des images RADARSAT-1

Afin de vérifier la relation entre le coefficient de rétrodiffusion radar et la teneur en humidité du sol, des teneurs en eau liquide ont également été mesurées à chaque date d'acquisition des images RADARSAT-1 (sauf lors de l'acquisition du 17 juin où les sols étaient complètement saturés en eau). Six parcelles en condition de sols nus (labour d'automne ou sans travail du sol) représentatives de la diversité des sols (granulométrie, évolution génétique et drainage) du BV du Bras d'Henri ont été sélectionnées à cet effet. À chaque parcelle, ces teneurs en eau volumétriques ont été mesurées en surface (0-15 cm) à l'aide d'un TDR (Time-Domain Reflectometry) ainsi que par la méthode gravimétrique pour différentes profondeurs (0-50 cm). Dans la présente étude, seules les données de teneur en eau mesurées dans les cinq premiers centimètres du sol et correspondant à la profondeur de pénétration du signal radar en bande C ont été retenues. Le type d'occupation du sol ainsi que la classification des sols présents dans les six sites choisis pour mesurer l'humidité du sol lors de l'acquisition des images RADARSAT-1 sont présentés aux tableaux 6 et 7 respectivement.

Tableau 6 : Occupation du sol et état de surface des parcelles au moment du passage du satellite RADARSAT-1 le 3 mai et le 27 mai (mode S2a)

Site	Culture en 2005	Description de l'état de surface (3 mai 2005)	Description de l'état de surface (27 mai 2005)
J02	Orge	Surface labourée, saturée en eau	Surface labourée
J04	Maïs	Surface non travaillée, résidu de surface (85%)	Surface semée en maïs, pas d'émergence
J05	Avoine	Surface labourée, délitescence de la structure, gel présent entre 40-60 cm	Surface semée en avoine, pas d'émergence
J10	Soya	Surface labourée, marques laissées par la circulation de la machinerie.	Surface semée en soya, pas d'émergence
J12	Maïs	Surface labourée et asséchée.	Surface semée en maïs, pas d'émergence
J17	Maïs	Surface labourée, saturée en eau, gel présent entre 40-60 cm.	Surface semée en maïs, pas d'émergence

Tableau 7 : Classification des sols et des pédo-paysages présents dans les différents sites choisis pour mesurer l'humidité du sol lors de l'acquisition des images RADARSAT-1

Site	Série de sols	Drainage	Dépôt	Taxonomie	Pente	Écorégion*
J02	Mawcook	C6	Till remanié	Gleysol Humique orthique	0-2%	132
J04	Le Bras	C6	Fluvio-lacustre / lacustre	Gleysol Humique orthique	0-2 %	132
J05	Neubois	C5	Fluvio-lacustre / lacustre	Podzol humo-ferrique gleyifié	2-5 %	132
J10	Woodbridge	C5	Till appalachien	Brunisol dystrique gleyifié	5-9 %	117
J12	Beaurivage	C3	Fluvio-glaciaire	Podzol humo-ferrique orthique	2-5 %	132
J17	Organique	C7	Organique	Humisol terrique	0-3 %	132

* Écorégion : 132 = Basses-Terres du St-Laurent; 117 = Appalaches.

2.1.5 Les données ASTER

Une image Aster acquise le 22 juin 2005 (la plus proche des dates d'acquisition des images radar) a été également sélectionnée pour l'étude. Les spécifications techniques du capteur sont présentées au tableau 8. Les valeurs des comptes numériques bruts ont été d'abord converties en luminance avant d'être transformées en réflectance bidirectionnelle par le programme de correction atmosphérique 6S (Vermote et al., 1997).

Tableau 8 : Spécifications techniques de l'image Aster du 22 juin 2005

Domaine*	Band e	Longueur d'onde (μm)	Résolutio n spatiale	Codage radiométriqu e
VNIR	1	0.52-0.60	15-m	8 bits
	2	0.63-0.69		
	3N,B	0.78-0.86		
SWIR	4	1.6-1.7	30-m	8 bits
	5	2.145-2.185		
	6	2.185-2.225		
	7	2.235-2.285		
	8	2.295-2.365		
	9	2.360-2.430		
TIR	10	8.125-8.475	60-m	12 bits
	11	8.475-8.825		
	12	8.925-9.275		
	13	10.25-10.95		
	14	10.95-11.65		

* VNIR : visible et proche infrarouge; SWIR : moyen infrarouge; TIR : infrarouge thermique.

3 Méthodologie

3.1 Analyse statistique des données de sol

L'analyse statistique des données de sols comporte trois étapes : l'étude de corrélation et le codage des données, l'analyse en correspondances multiples et la comparaison entre la classification experte et les différents modèles de classification automatique du drainage.

3.1.1 Étude de corrélation et codage des données

L'analyse de la corrélation entre le drainage et les différentes variables de sol retenues permet de sélectionner celles qui sont les plus pertinentes. Étant donné la nature catégorielle de ces variables, des tests non paramétriques ont été utilisés pour guider le choix, soit le coefficient R_S de Spearman, le coefficient Gamma et le Tau de Kendall. Par la suite, un codage quantitatif a été effectué afin de pouvoir appliquer les méthodes générales d'analyse linéaire. Un codage disjonctif complet a donc été appliqué. Ce dernier consiste à transformer la variable qualitative à (c) modalités en (c) variables binaires indicatrices de chaque modalité (Goovaerts, 1977, Khattree and Naik, 2000, Hengl et al., 2004) :

$$q_c(s) = \begin{cases} 1 & \text{si } q_c(s) = \text{classe } (c) \\ 0 & \text{ailleurs} \end{cases} \quad (1)$$

Où $q_c(s)$ est l'indicateur binaire pour chaque modalité.

Cette approche conduit ainsi à 35 indicatrices binaires. Pour réduire la dimension de l'espace, une analyse des correspondances multiples (ACM) a ensuite été appliquée.

3.1.2 Analyse des correspondances multiples (ACM)

L'avantage de l'application d'une ACM sur les données est d'obtenir un espace continu. On peut alors utiliser des distances métriques telles que la distance euclidienne simple sur ces données transformées; car les axes factoriels sont par construction indépendants lorsque considérés deux à deux. Enfin, en ne sélectionnant que les premières composantes extraites par l'ACM, i.e. celles qui présentent la plus grande variance, le bruit lié aux diverses transformations des données se trouve réduit.

3.1.3 Comparaison entre la classification experte et différentes classifications automatiques du drainage des sols

La correspondance entre différentes classifications automatiques du drainage générées à partir des composantes issues de l'ACM et dérivées des propriétés de sol les plus pertinentes avec la typologie du drainage proposée par les spécialistes d'AAC sera d'abord étudiée. Pour ce, une méthode de classification basée sur les arbres de décision (algorithmes 'QUEST' = *Quick, Unbiased, Efficient Statistical Trees*) de Loh et Shih, (1997) sera utilisée. Cette méthode est basée sur des segmentations univariées des variables prédictives en utilisant l'analyse discriminante quadratique.

L'étape suivante consistera à vérifier si la classification proposée, laquelle intègre quand même une contrainte forte; i.e. la construction d'indicateurs logiques pour représenter les classes, concorde avec les résultats produits par les méthodes classiques telles que les K-MEANS (4 et 3 classes). Enfin, les performances de discrimination des classes de drainage générées à partir de trois méthodes de groupement appliquées sur différents jeux de données de sols et d'OT (DRAI : classification proposée par AAC; K-MEANS à 3 classes et K-MEANS à 4 classes).

3.2 Analyse discriminante générale

L'analyse discriminante utilise les formules standards pour calculer les fonctions discriminantes, pour effectuer l'analyse canonique, et pour calculer des statistiques de classification. D'une manière générale, trois hypothèses doivent être souvent bien testées afin d'effectuer correctement

cette analyse : l'hypothèse de la normalité des données, celle de la multicollinéarité ainsi que celle de l'homogénéité des matrices de variances-covariances des classes.

3.2.1 Étude de la normalité avec l'ensemble des données

Il est important de vérifier d'abord cette hypothèse afin de profiter du large éventail des méthodes d'analyse et de classification paramétrique bayésienne. Si cette hypothèse s'avère non applicable, il convient alors de déterminer la transformation à appliquer aux données pour s'en rapprocher.

Il existe plusieurs méthodes pour tester la normalité des données. La droite de Henry sur laquelle est rapportée la statistique (W) de Shapiro-Wilk a été retenue car c'est le test de normalité le plus utilisé et le plus puissant comparativement aux nombreux tests alternatifs (Shapiro et al., 1968). Une extension du test décrit par Royston (1982), permet de l'appliquer à des échantillons de grande taille (jusqu'à 2,000 observations); ce qui convient bien à l'étude en cours comportant 1612 observations.

Si les deux tests effectués sur les données montrent que l'hypothèse gaussienne doit être rejetée, la transformation logistique (*logit*) sera appliquée. Parmi la large gamme de transformations possibles, la transformation logistique (*logit*) est celle qui est la plus utilisée (Hengl et al., 2004). Avant d'effectuer cette transformation, les données seront d'abord centrées et réduites car les données de réflectance et les coefficients de rétrodiffusion ne sont pas mesurées sur une même échelle. Cette transformation rendra toutes les valeurs (indifféremment de leurs distributions et unités de mesures originales) compatibles avec une moyenne 0 et un écart-type de 1. La transformation d'une variable z en variable centrée et réduite est exprimée :

$$z_{cr} = \frac{z - z_{\min}}{z_{\max} - z_{\min}} \quad (1)$$

et la transformation de z_{cr} en *logit* est exprimée par :

$$\log it(z_{cr}) = \ln\left(\frac{z_{cr}}{1 - z_{cr}}\right) \quad (2)$$

L'avantage de cette transformation est qu'elle peut être considérée linéaire en fonction du nombre de variables explicatives ou variables indépendantes centrées et réduites (x_1, \dots, x_n) :

$$\log it(z_{cr}) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n \quad (3)$$

Dans cette étude, les données d'OT sont ces variables explicatives et le coefficient β_0 représente l'ordonnée à l'origine et les valeurs β_i sont les coefficients de régression (des variables 1 à n) calculés à partir des données radar.

Comme la variable dépendante (le drainage) est de nature catégorielle (c'est-à-dire qu'elle contient des données décrivant l'appartenance des observations respectives à une catégorie particulière de drainage), un modèle linéaire multinomial sera utilisé. La qualité d'ajustement pour chaque donnée va être analysée à partir du ratio entre les statistiques et les degrés de liberté.

3.2.2 Le problème de multicollinéarité

Une autre hypothèse de l'analyse discriminante stipule que les variables utilisées pour faire la discrimination entre les groupes ne sont pas complètement redondantes. Cela est dû au fait que dans les calculs de l'analyse discriminante la matrice de variance-covariance des variables du modèle doit être inversée. Si l'une des variables est complètement redondante avec les autres variables, la matrice est dite mal conditionnée ou singulière (déterminant = 0), et ne pourra être inversée. Une étude de multicollinéarité sera faite sur les bandes spectrales de l'image d'ASTER étant donné la proximité de ces intervalles spectraux et le fait que les données soient prises dans les mêmes conditions. Nous utiliserons une analyse en composantes principales pour réduire la dimension de la base de données et éliminer le problème de multicollinéarité. Les données RADARSAT-1 acquises à différentes dates et sous différentes configurations sont supposées être complètement décorréliées contrairement aux bandes spectrales de l'image ASTER qui sont corrélées.

3.2.3 Homogénéité des matrices de variances-covariances

Les matrices de variance-covariance des variables sont supposées homogènes entre les groupes, ce qui n'est pas évident pour les variances-covariances intra-classes. Dépendamment de ces matrices, on calcule la distance $D_j^2(x)$, dite la distance de Mahalanobis soit selon la formule :

$$D_j^2(x) = (x - \mu_j)' \Sigma^{-1} (x - \mu_j) - 2 \ln(z_{crj}) \quad (4)$$

pour réaliser une *discrimination linéaire* si toutes les classes de drainage ont la même matrice de variance-covariance, soit selon la formule :

$$D_j^2(x) = (x - \mu_j)' \Sigma^{-1} (x - \mu_j) + \ln(\Sigma_j) - 2 \ln(z_{crj}) \quad (5)$$

au cas où les matrices de variance-covariance sont différentes pour réaliser une *discrimination quadratique*.

Dans les équations précédentes (4,5) μ_j est le vecteur des valeurs moyennes de la classe de drainage G_j , Σ est la matrice de variance-covariance commune (Eq. 4) et Σ_j est la matrice de variance-covariance de la classe G_j (Eq. 5).

L'observation x sera affectée à la classe de drainage G_j la plus proche au sens de $D_j^2(x)$. z_j étant la probabilité à priori que l'observation x appartienne à la classe de drainage G_j . Dans cette étude, la probabilité à priori considérée proportionnelle à la taille de la classe de drainage.

4 Résultats et discussions

4.1 Évaluation de la cohérence du modèle de classification du drainage des sols par des experts et proposition d'un modèle de classification automatique

L'analyse de la corrélation entre le drainage et les variables de sol retenues pour cette étude permet de sélectionner celles qui sont les plus pertinentes. Le tableau 9 révèle que les variables les plus corrélées avec le drainage sont par ordre d'importance : PGLE, ORDRE, BCHR, ATEXTURE et BTEXTURE. Comme ces propriétés ont été codées de façon qualitative (en classes), l'équation 1 a été appliquée sur celles-ci pour les rendre quantitatives. Cette approche conduit à 25 indicatrices binaires.

Tableau 9 : Mesures d'association non paramétriques entre la classe de drainage des sols et quelques propriétés morphologiques de sol.

<i>Couples de Variables</i>	<i>R de Spearman</i>	<i>Gamma</i>	<i>Tau de Kendall</i>
DRAI & PGLE	-0.83	-0.90	-0.75
DRAI & AVAL	-0.19	-0.21	-0.16
DRAI & ACHR	-0.29	-0.36	-0.25
DRAI & BVAL	0.06	0.09	0.07
DRAI & BCHR	-0.69	-0.78	-0.60
DRAI & CVAL	-0.20	-0.24	-0.17
DRAI & ORDRE	0.82	0.91	0.75
DRAI & CCHR	-0.38	-0.47	-0.33
DRAI & ATEXTURE	0.57	0.68	0.51
DRAI & AMO	0.46	0.62	0.41
DRAI & AFRAGMENT	-0.13	-0.23	-0.12
DRAI & BTEXTURE	0.53	0.64	0.47
DRAI & CTEXTURE	0.37	0.44	0.32

Pour réduire le nombre de variables et pour être dans un espace continu, une analyse des correspondances multiples (ACM) a été appliquée. En ne sélectionnant que les 13 premiers composantes (ou axes) (soit à peu près la moitié du nombre de variables initiales), environ 74.47% de l'information originale est couverte ce qui est acceptable et permet en plus d'éliminer une bonne partie de l'information « bruitée » spécifique aux transformations (Tableau 10). Ce choix est fondé sur le critère de Kaiser (Kaiser, 1960), selon lequel on peut retenir les facteurs avec une valeur propre supérieure ou égale à 1 pour réduire le bruit.

Tableau 10 : Résultats de l'analyse de correspondances multiples sur les propriétés de sol les plus corrélées avec le drainage

Facteur No.	Valeur propre	% variance originale	% variance cumulée
1	4.32	14.42	14.42
2	3.52	11.75	26.17
3	2.27	7.58	33.74
4	1.64	5.47	39.21
5	1.49	4.96	44.17
6	1.35	4.51	48.67
7	1.23	4.09	52.77
8	1.21	4.05	56.81
9	1.17	3.91	60.72
10	1.09	3.62	64.34
11	1.03	3.42	67.76
12	1.01	3.37	71.13
13	1.00	3.34	74.47
14	0.99	3.30	77.77
15	0.97	3.24	81.01
16	0.92	3.06	84.06
17	0.86	2.88	86.95
18	0.82	2.73	89.68
19	0.75	2.49	92.17
20	0.68	2.26	94.43
21	0.55	1.84	96.27
22	0.45	1.49	97.76
23	0.42	1.39	99.15
24	0.23	0.78	99.93
25	0.02	0.07	100.00

Le premier axe représentant 14.41 % de l'inertie totale et le second 11.74 %. Le plan défini par l'ensemble des deux premiers axes représente donc 26.16 % de la variance originale. Ce nombre peut paraître peu important mais doit être mis en regard du grand nombre de variables dans l'espace de représentation.

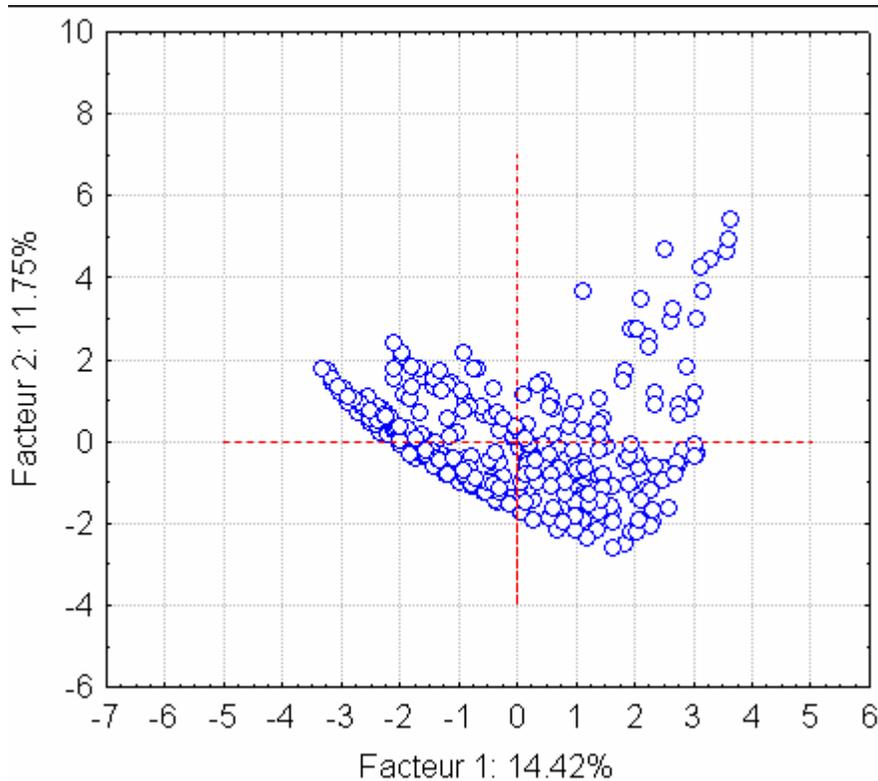


Figure 4 : Représentation des observations dans le plan des deux axes principaux (facteurs) définis à partir de l'ACM.

Néanmoins, à la figure 4, deux régions peuvent être distinguées dans le plan principal formé par les deux premiers facteurs extraits de l'ACM. Dans la première région, les individus sont très rapprochés; dans la seconde au contraire, ils sont plus distants. Bien que cette approche favorise une vision globale des données, elle ne permet pas d'identifier des groupes clairement définis. Pour y arriver, une analyse factorielle discriminante, à l'instar de l'analyse en composantes principales, permet de réorganiser les axes et ainsi d'optimiser la séparation entre les classes. Les

axes discriminants correspondent aux directions du nuage de points telles que les projections des points d'observations soient les mieux différenciées selon leurs classes d'appartenance. En d'autres termes, on cherche, suivant ces axes que la séparation des projections des classes soit maximale et que chacune des classes soit, en projection, bien regroupée.

La figure 5 montre une solution avec les deux premiers axes discriminants. Malgré la très importante réduction de dimensions, la solution avec ces axes (qui ne représente qu'environ 26 % d'inertie totale) montre 4 regroupements de propriétés bien concentrés et un regroupement plus dispersé en bas de la figure.

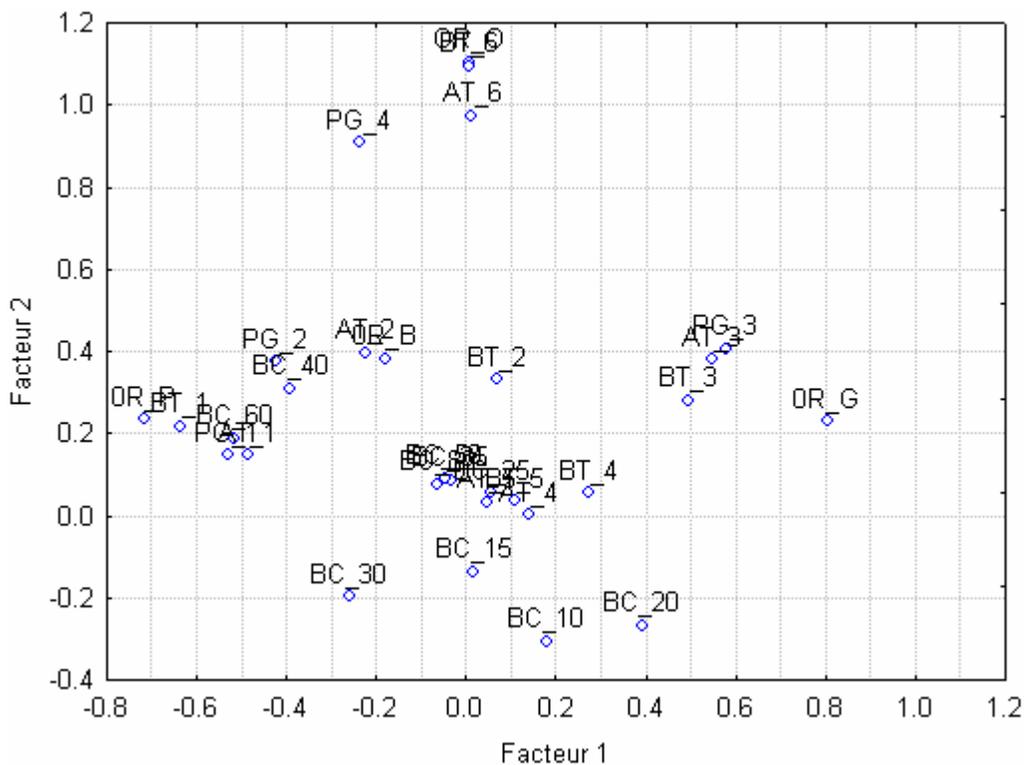


Figure 5 : Représentation des observations dans le plan des deux premiers axes les plus discriminants. Les codes des classes utilisées sont décrits au tableau 11.

Tableau 11 : Codes des classes utilisées à la figure 5 et leur description

Code	Description
PG_1	PGLEY classe 1 ²
PG_2	PGLEY classe 2
PG_3	PGLEY classe 3
PG_4	PGLEY classe 4
OR_O	ORDRE classe O
OR_G	ORDRE classe G
OR_P	ORDRE classe P
OR_B	ORDRE classe B
BC_10	Chroma de la couleur de l'horizon B de valeur 10
BC_15	Chroma de la couleur de l'horizon B de valeur 15
BC_20	Chroma de la couleur de l'horizon B de valeur 20
BC_25	Chroma de la couleur de l'horizon B de valeur 25
BC_30	Chroma de la couleur de l'horizon B de valeur 30
BC_35	Chroma de la couleur de l'horizon B de valeur 35
BC_40	Chroma de la couleur de l'horizon B de valeur 40
BC_50	Chroma de la couleur de l'horizon B de valeur 50
BC_60	Chroma de la couleur de l'horizon B de valeur 60
BC_80	Chroma de la couleur de l'horizon B de valeur 80
AT_1	Code de texture de l'horizon A de valeur 1
AT_2	Code de texture de l'horizon A de valeur 2
AT_3	Code de texture de l'horizon A de valeur 3
AT_4	Code de texture de l'horizon A de valeur 4
AT_5	Code de texture de l'horizon A de valeur 5
AT_6	Code de texture de l'horizon A de valeur 6
BT_1	Code de texture de l'horizon B de valeur 1
BT_2	Code de texture de l'horizon B de valeur 2
BT_3	Code de texture de l'horizon B de valeur 3
BT_4	Code de texture de l'horizon B de valeur 4
BT_5	Code de texture de l'horizon B de valeur 5
BT_6	Code de texture de l'horizon B de valeur 6

Pour voir de façon plus globale si la classification experte de drainage fournie par les spécialistes est conforme à la classification automatique faite à partir des 13 premières composantes les plus pertinentes, une analyse à l'aide d'un arbre de classification a été effectuée.

² Voir tableau 15

4.2 Résultats de la classification

Les arbres de décision permettent de prévoir l'affectation d'observations ou d'objets à des classes d'une variable dépendante catégorielle à partir de leurs mesures sur une ou plusieurs variables prédictives. Cette méthode est assez proche des méthodes plus traditionnelles comme l'analyse discriminante, des méthodes de classification ou des tests non paramétriques puisque son but est de prévoir ou d'expliquer les réponses d'une variable dépendante catégorielle. Cependant, sa flexibilité en fait une option analytique très attrayante si les hypothèses très strictes des méthodes traditionnelles (concernant la distribution) ne sont pas respectées. Dans ce cas, les arbres de décision sont un recours sans égal. La méthode QUEST a été choisie avec des segmentations des combinaisons linéaires basées sur une analyse discriminante. Le tableau 12 montre que pour l'arbre sélectionné, 25 des 1569 observations (sans 0), soit seulement 1.6% étaient mal classées. Il montre également que c'est dans les 3 dernières classes (C5 à C7) que l'on retrouve le plus grand nombre de classifications discordantes. Ceci démontre que la classification experte est généralement en accord avec la classification automatique faite par la méthode automatique basée sur l'ACM.

Tableau 12 : Matrice des observations classées différemment par la classification automatique (A3 à A7) et par la classification experte (C3-C7).

CLASSE	CLASSE A3	CLASSE A4	CLASSE A5	CLASSE A6	CLASSE A7
C3	---	0	1	0	0
C4	0	---	0	1	0
C5	1	1	---	8	0
C6	0	0	4	---	7
C7	0	0	0	2	---

Tableau 13 : Étude de la concordance (diagonale) entre la classification experte (C3 à C7) et celle issue de la méthode automatique (A3 à A7)

CLASSE	CLASSE A3	CLASSE A4	CLASSE A5	CLASSE A6	CLASSE A7
C3	52	0	1	0	0
C4	0	123	0	1	0
C5	1	1	449	8	0
C6	0	0	4	676	7
C7	0	0	0	2	244

Le tableau 14 présente les résultats de l'étude de la concordance (diagonale) entre la classification experte (C4 à C7) et celle issue de la méthode K-MEANS à 4 classes (KM41 à KM44), les profils classés C3 par les experts ayant été retirés du jeu de données. En croisant les deux classifications avec une table de contingence χ^2 , le tableau 14 montre des classes relativement pures, plus particulièrement pour la classe C4 (groupe KM41) et à l'exception des sols très mal drainés (C7) où 75 profils parmi les 251 étudiés appartenant à cette classe se retrouvent plutôt associés aux sols mal drainés (C6) au sein du groupe KM42. En retirant également les sols classés C4 du jeu de données, la classification K-MEANS à 3 classes (Tableau 15) regroupe 1392 profils de sol (87.35 %), lesquels appartiennent, selon la classification experte, aux 3 dernières classes de drainage, les sols imparfaitement à très mal drainés (C5 à C7).

Tableau 14 : Étude de la concordance (diagonale) entre la classification experte et celle issue de la méthode K-MEANS à 4 classes au moyen d'une table de contingence χ^2

CLASSE DRAI	KM41	KM42	KM43	KM44	TOTAL LIGNE
C4	0	0	0	124	124
C5	0	3	429	22	454
C6	0	665	16	6	687
C7	175	75	1	0	251
TOTAL COLONNE	175	743	446	152	1516

Tableau 15 : Étude de la concordance (diagonale) entre la classification experte et celle issue de la méthode K-MEANS à 3 classes au moyen d'une table de contingence χ^2

CLASSE DRAI	KM31	KM32	KM33	TOTAL LIGNE
C5	451	3	0	454
C6	17	670	0	687
C7	1	75	175	251
TOTAL COLONNE	469	748	175	1392

La classification K-MEANS à 3 classes montre également dans le croisement KM32 et C7 la discordance de classification pour 75 profils. Les analyses suivantes permettront de savoir si c'est un problème lié à la procédure de classification automatique ou à l'incohérence de la classification experte.

Comme la profondeur du gley (PGLE) est la propriété la plus corrélée au drainage, la classification experte a été examinée en regard de ce critère. Ainsi au tableau 16, les classes 7 et 6 de la classification experte ont été groupées car elles sont à la même profondeur PGLE de 0 cm. De 0 à 4 cm, on voit bien qu'il y a plus d'observations appartenant à la classe 6, on a donc décidé d'attribuer toutes les observations de cet intervalle de profondeur à la classe 6 et ainsi de suite pour les autres classes.

Tableau 16 : Classification manuelle des classes de drainage en fonction de la profondeur du gley (PGLE)

PGLE	DRAI 3	DRAI 4	DRAI 5	DRAI 6	DRAI 7	Row Totals	Classification
0	0	1	8	144	171	324	Classe 7
1	0	0	0	25	13	38	
2	0	0	1	135	29	165	
3	0	0	3	201	18	222	
4	0	0	14	115	9	138	Classe 6
5	0	1	175	43	9	228	
6	1	0	143	14	2	160	
7	0	1	83	3	0	87	Classe 5
8	0	41	21	3	0	65	
9	0	38	3	3	0	44	
10	0	27	1	0	0	28	
11	0	11	1	0	0	12	
12	1	0	0	0	0	1	
13	1	0	0	0	0	1	
15	1	0	0	0	0	1	
99	50	4	0	1	0	55	Classe 4
Ts Grps	54	124	453	687	251	1569	

Le tableau 17 montre que chaque classe de la classification PGLE est une combinaison de deux classes pures de la classification experte DRAI.

Tableau 17 : Croisement de la classification experte (DRAI) avec celle de PGLE à 4 classes (GL4 à GL7) au moyen d'une table de contingence χ^2

DRAI	GL4	GL5	GL6	GL7	TOTAL
C3	52	1	0	0	53
C4	121	2	0	1	124
C5	28	400	18	8	454
C6	7	60	476	144	687
C7	0	11	69	171	251
TOTAL	208	474	563	324	1569

4.3 Résultats d'analyse des données RADARSAT-1

4.3.1 Les données de suivi

Dans un premier temps, les données d'humidité acquises au moment du passage du satellite ont été analysées. Les données du 3 et du 27 mai 2005 ont été retenues pour l'analyse car ce sont celles qui ont été acquises dans la même configuration (mode S 2 ascendant). La figure 6 montre les coefficients de rétrodiffusion des différentes parcelles. La différence la plus importante observée entre les données du 3 mai et du 27 mai se rapporte aux parcelles J02 et J04.

La figure 7a et 7b montre la simulation de la rétrodiffusion radar en polarisation HH obtenue avec un angle d'incidence de 23.5 degré (angle d'incidence moyen sur les sites) en utilisant le modèle électromagnétique IEM (Integral Equation Method) de Fung (1996). Sur cette figure, sont présentés les coefficients de rétrodiffusion extraits des six champs et leurs teneurs en eau volumétrique correspondantes à la méthode gravimétrique. Toutes les autres propriétés d'entrée du modèle IEM (composition texturale et salinité) sont disponibles. La seule propriété qui n'a pas été acquise est la rugosité de surface caractérisée par l'écart type des hauteurs (s) et la longueur de corrélation (L). En ajustant le modèle avec un s = 1 cm et un L = 8 cm pour l'image du 3 mai et s = 0.9 cm et L = 10 cm pour l'image du 27 mai, on trouve la meilleure correspondance entre

les données radar extraites des images et celles simulées par le modèle IEM. La faible différence de rugosité pourrait s'expliquer par les faibles précipitations enregistrées entre les deux dates (Figure 2) pouvant avoir entraîné des événements de ruissellement ayant résulté en une diminution légère de la rugosité de surface.

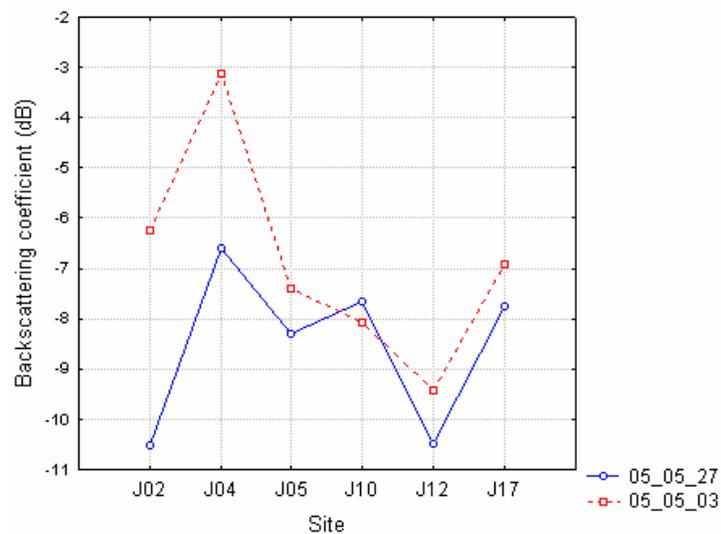


Figure 6 : Rétrodiffusion radar des parcelles tests extraites des images RADARSAT-1

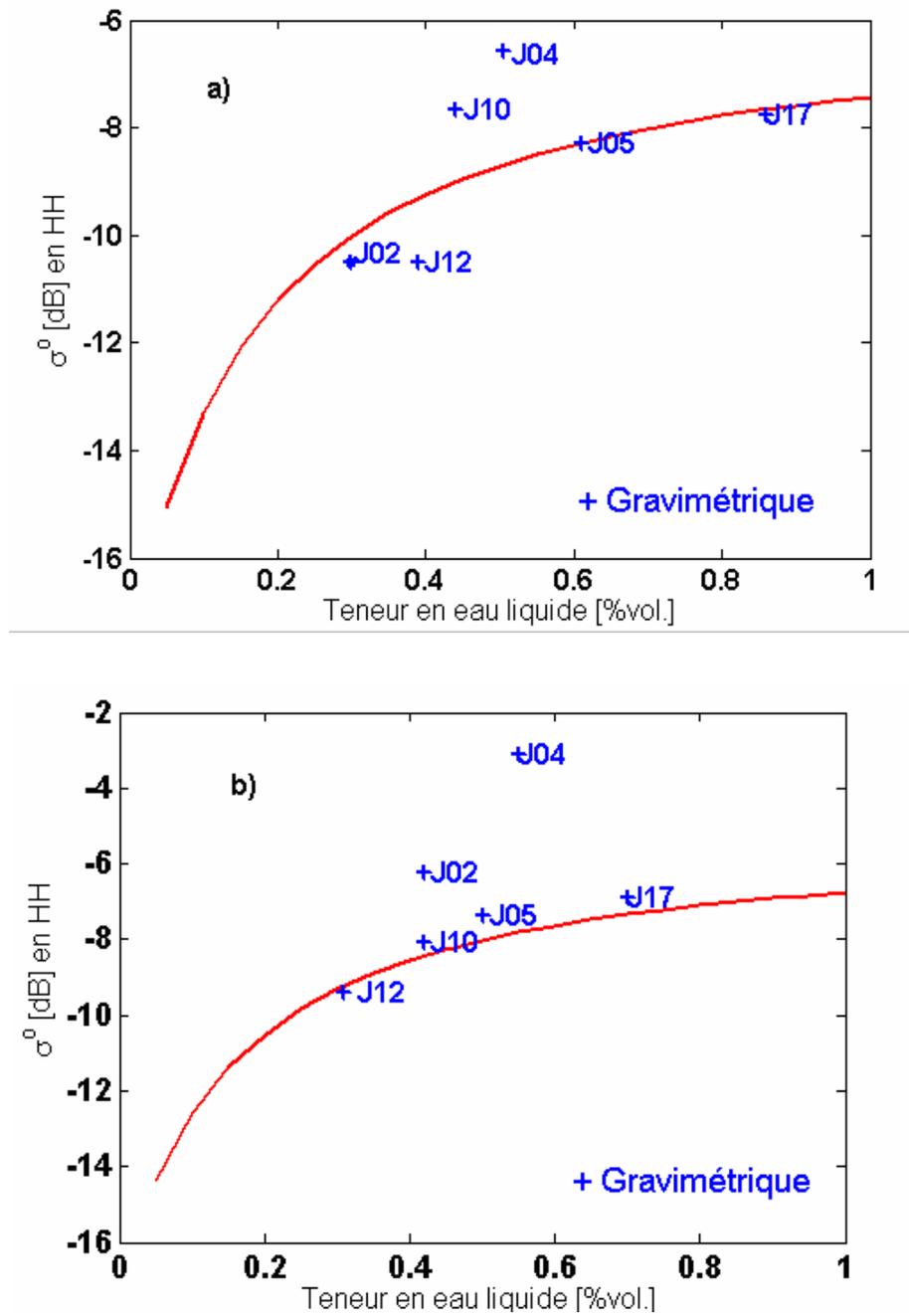


Figure 7 : a) Simulation de la rétrodiffusion radar avec le modèle IEM avec $s=1\text{cm}$ et $L=8\text{ cm}$ pour l'image du 3 mai et; b) avec $s=0.9\text{ cm}$ et $L=10\text{ cm}$ pour l'image du 27 mai.

Dans cette figure on trouve que les coefficients de rétrodiffusion extraits des champs sont relativement bien en accord avec celles simulés par le modèle IEM. Cependant, on constate que c'est le champ J04 qui est le plus éloigné de la rétrodiffusion théorique. Le site J04 est le seul site qui n'a pas été labouré; il comportait beaucoup de résidus au sol ce qui semble avoir influencé la pénétration du gel du sol ainsi que le dégel et le ressuyage des sols au printemps.

4.3.2 Rapport de bandes

En absence de connaissance de la rugosité de surface, propriété complémentaire dans l'étude de l'humidité de surface par analyse de la réponse radar en absence de végétation, un rapport des coefficients de rétrodiffusion extraits des images acquises aux deux dates (03 et 27 mai) pour les sols nus a été utilisé afin de réduire l'effet des variations de rugosité (Bernier et Fortin, 1998). Les variations du signal peuvent alors être attribuées aux seules variations de l'humidité de la couche de surface des sols. Le test ANOVA de Kruskal-Wallis a été retenu pour étudier le pouvoir de discrimination des différentes classes sur le rapport des deux bandes. Le test ANOVA de Kruskal-Wallis par rangs se fonde sur l'hypothèse que les différents échantillons à comparer suivent la même distribution ou qu'ils ont des distributions avec la même médiane. C'est pourquoi, l'interprétation du test de Kruskal-Wallis est très voisine de l'ANOVA paramétrique à un facteur, sauf qu'il se base sur des rangs et non des moyennes.

Le Test de Kruskal-Wallis est hautement significatif ($p = 0,0005$) ce qui permet de conclure que les performances des différentes classes ont été significativement différentes d'une classe à l'autre. La somme des rangs la plus élevée (plus mauvaise performance) apparaît dans les classes 5 et 6 et celle qui est la plus faible (meilleure performance) apparaît dans les classes 3, 4 et 7 (Tableau 18). Ces résultats sont cohérents.

Tableau 18 : Effectif et somme des rangs pour les différentes classes de drainage

CLASSE	EFFECTIF	SOMME DES RANGS
C3	13	2424.50
C4	39	8993.50
C5	119	23854.00
C6	169	30611.00
C7	34	4242.00

À nouveau, le tracé des médianes (Figure 8) montre que c'est la classe 3 qui affiche la meilleure performance. En conclusion de cette analyse réalisée en condition de sols nus, le rapport des coefficients de rétrodiffusion des polarisations HH d'images RADARSAT-1 acquises dans la même configuration (ici mode S 2 ascendant) à 21 jours d'intervalle au printemps (ici 3 mai et 27 mai 2005), permet une bonne discrimination des classes de drainage des sols.

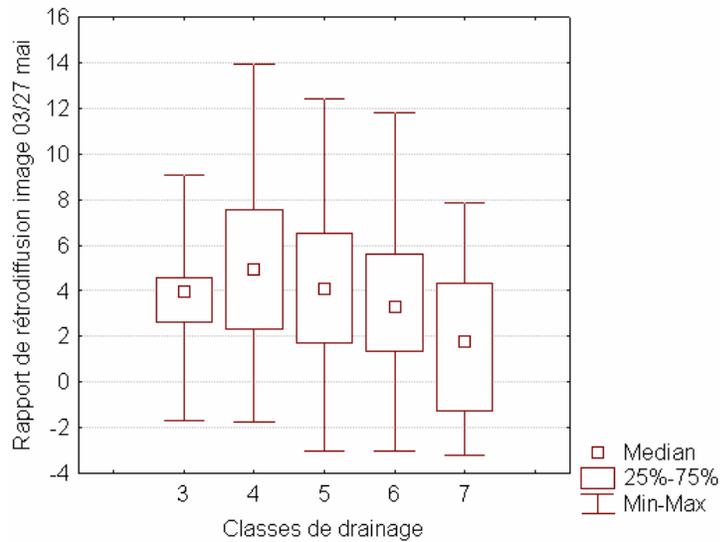


Figure 8 : Médiane, quartile et étendue des différentes classes de drainage en fonction du rapport de bandes

4.3.3 L'ensemble des données radar

En condition de sols nus, en considérant la moyenne des rétrodiffusions de toutes les images RADARSAT-1 acquises au printemps 2005 (Figure 9), les données du 3 mai et celles du 27 mai sont celles qui différencient le mieux les classes de drainage des sols. Pour l'image du 3 mai les classes 7, 6 et 5 donnent la même réponse radar tandis que la classe 4 et la classe 3 ont une différence d'environ 2 dB. Pour l'image du 27 mai, les classes 6 et 5 donnent la même réponse, tandis que toutes les autres sont différents d'un écart moyen de 1.5 dB.

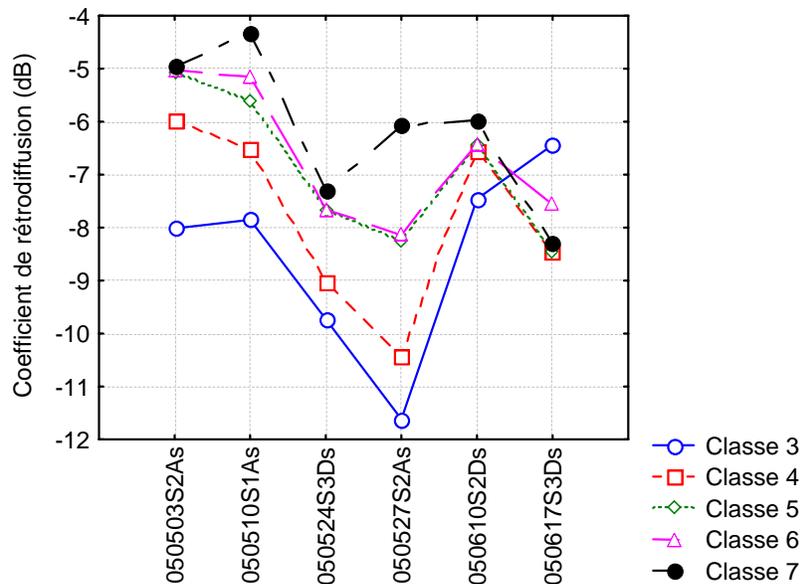


Figure 9 : Moyenne des rétrodiffusions radar pour les différentes images RADARSAT-1 en fonction des classes de drainage des profils de sol en condition de sol nus

Pour les profils sous couvert de végétation, la discrimination entre les classes de drainage est moins évidente (Figure 10), sauf pour les deux classes les plus éloignées (C3 et C7).

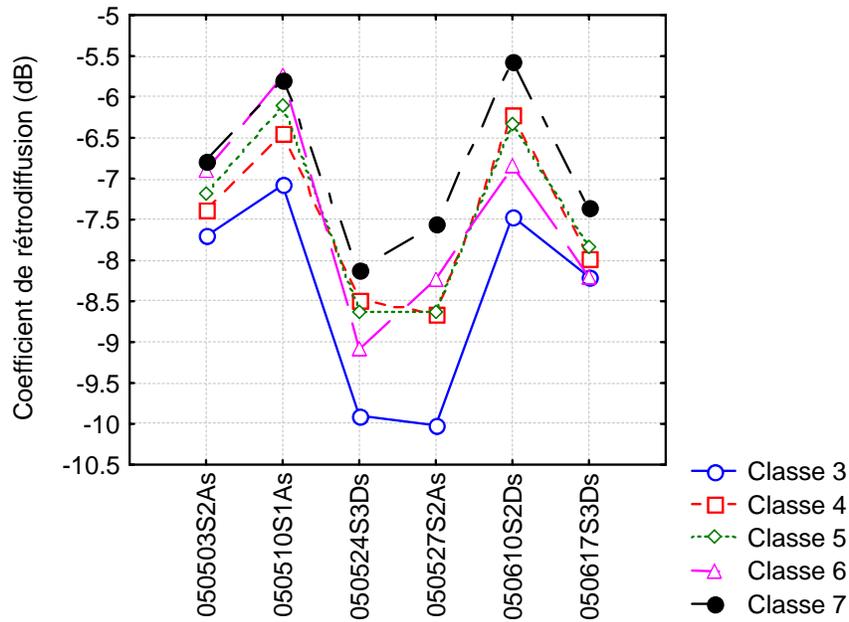


Figure 10 : Moyenne des rétrodiffusions radar pour les différentes images RADARSAT-1 en fonction des classes de drainage des profils de sol sous couverture de végétation

4.4 Résultats d'analyse des données ASTER

4.4.1 Statistiques générales

Les statistiques descriptives (Tableau 19) montrent que parmi les 9 bandes spectrales de l'image ASTER du 22 juin 2005, c'est la bande proche infrarouge (ASTER3) qui affiche les valeurs les plus élevées. De plus, cette bande est la moins corrélée aux autres bandes (Tableau 20). En effet, une forte multicollinéarité est observée entre toutes les autres bandes. Pour contrer ce problème de multicollinéarité, signe de la redondance de l'information spectrale du capteur ASTER, une analyse en composantes principales (ACP) a été appliquée sur les bandes pour créer de nouveaux descripteurs indépendants.

Tableau 19 : Statistiques générales décrivant les 9 bandes spectrales de l'image ASTER acquise en juin 2005

Bande	Effectif	Moyenne	Minimum	Maximum	Écart-type
ASTER_1	1569	22.79605	10.00000	58.0000	9.00045
ASTER_2	1569	20.26577	6.00000	60.0000	12.90122
ASTER_3	1569	79.85214	39.00000	131.0000	15.09535
ASTER_4	1569	59.37859	30.00000	99.0000	13.22194
ASTER_5	1569	37.57425	20.00000	71.0000	12.33120
ASTER_6	1569	38.46144	17.00000	74.0000	13.93338
ASTER_7	1569	36.92097	16.00000	78.0000	14.33096
ASTER_8	1569	33.41300	13.00000	74.0000	14.76902
ASTER_9	1569	32.25813	17.00000	63.0000	11.42752

Tableau 20 : Coefficient de corrélation de Pearson mesuré entre les 9 bandes spectrales de l'image ASTER acquise en juin 2005

Band e	01	02	03	04	05	06	07	08	09
01	1.00	0.95	-0.28	0.80	0.82	0.86	0.87	0.89	0.84
02	0.95	1.00	-0.37	0.79	0.87	0.90	0.90	0.89	0.89
03	-0.28	-0.37	1.00	-0.03	-0.25	-0.28	-0.27	-0.33	-0.31
04	0.80	0.79	-0.03	1.00	0.88	0.92	0.91	0.89	0.87
05	0.82	0.87	-0.25	0.88	1.00	0.98	0.97	0.92	0.98
06	0.86	0.90	-0.28	0.92	0.98	1.00	0.99	0.96	0.98
07	0.87	0.90	-0.27	0.91	0.97	0.99	1.00	0.97	0.98
08	0.89	0.89	-0.33	0.89	0.92	0.96	0.97	1.00	0.94
09	0.84	0.89	-0.31	0.87	0.98	0.98	0.98	0.94	1.00

4.4.2 Analyse en composantes principales (ACP)

La figure 11 montre que les deux premiers facteurs représentent près de 94% de la variance totale. Ainsi, en ne sélectionnant que les deux premiers facteurs, on conserve une proportion très satisfaisante de l'information 'originale' des 9 bandes tout en laissant de côté l'information dite 'bruitée' liée au signal.

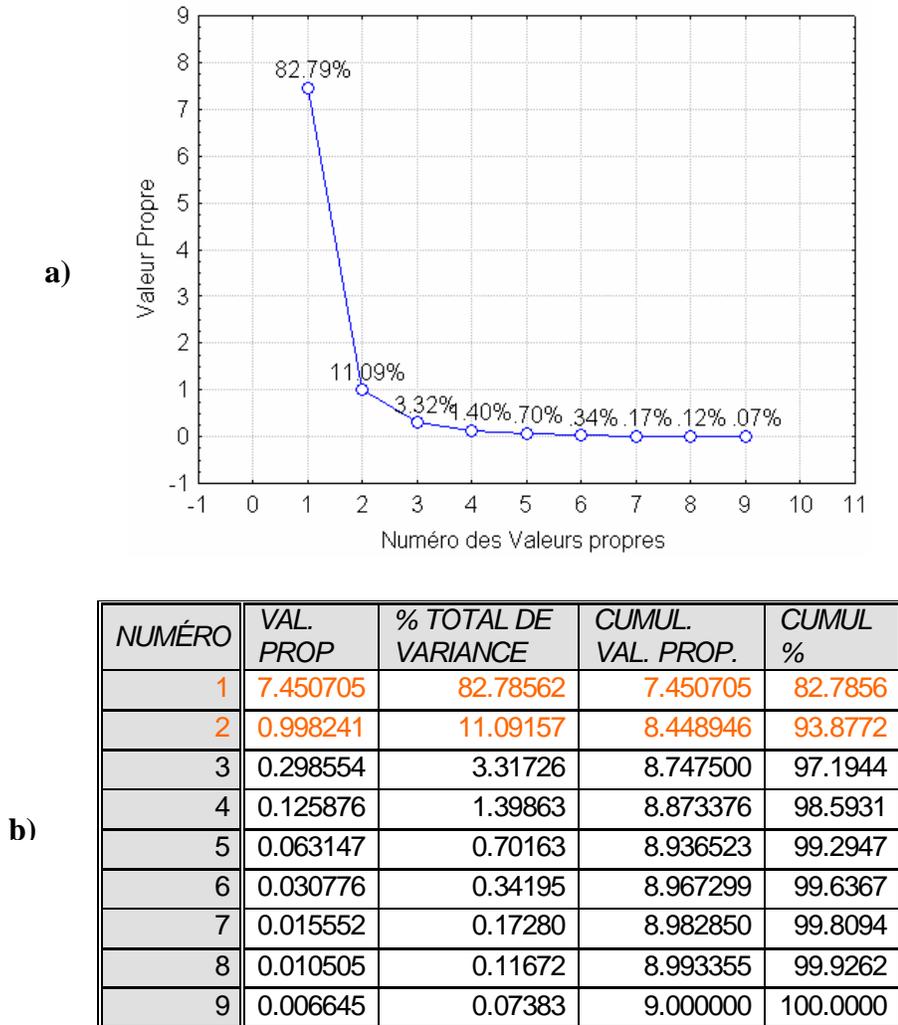


Figure 11 : a) Tracé des valeurs propres (a) et pourcentage de la variance originale (b) associé aux neuf composantes extraites de l'ACP à partir des bandes spectrales de l'image d'ASTER

Sur la base de l'importance des coordonnées factorielles (corrélations entre les variables et les facteurs) (Tableau 20 et figure 12), le premier axe, associé à la première valeur propre 7.45, apparaît fortement corrélé à toutes les bandes spectrales sauf la bande ASTER_3, tandis que le deuxième axe est fortement corrélé à la bande ASTER_3 (coefficients de corrélation élevés et négatifs). La deuxième composante pourrait ainsi être appelée « bande proche infrarouge » et la première « Autres bandes spectrales ASTER ».

Tableau 21 : Poids factoriels des bandes spectrales d'ASTER pour les deux premiers facteurs

BANDE	FACTEUR 1	FACTEUR 2
ASTER_1	-0.916696	0.018142
ASTER_2	-0.940699	0.104526
ASTER_3	0.315492	-0.941404
ASTER_4	-0.909598	-0.304541
ASTER_5	-0.967173	-0.063023
ASTER_6	-0.988834	-0.040405
ASTER_7	-0.989931	-0.045723
ASTER_8	-0.974489	0.016678
ASTER_9	-0.977623	-0.005150

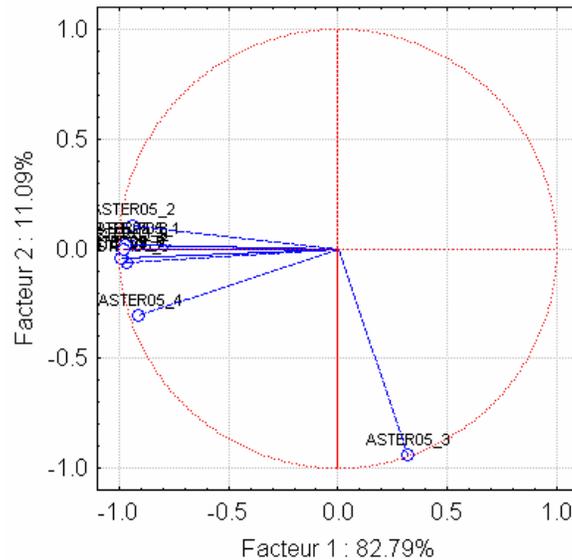


Figure 12 : Projection des variables sur le plan factoriel réduit défini à partir des deux premières composantes extraites de l'ACP

Dans la même optique que l'analyse des données radar, les données extraites de l'ACP réalisée à partir des bandes ASTER seront étudiées quant à leur efficacité à discriminer les classes de drainage des sols globalement et en fonction de l'occupation du sol. L'indice NDVI sera également évalué à cette fin.

4.4.3 ACP en fonction des classes de drainage et du type d'occupation du sol

Les nuages de points (Figure 13) tant pour les sols nus que pour les sols sous couvert de végétation ne permettent pas de discriminer visuellement les classes de drainage. Elles renseignent cependant sur la nature de la distribution; les données ne sont pas distribuées normalement. Le test de la médiane indique que seule la classe 7, laquelle présente la médiane plus élevée, exhibe de faibles performances de discrimination des classes de drainage (Figure 14).

Par ailleurs, le nuage des deux facteurs principaux caractérise relativement bien les types d'occupation du sol. Dans la figure 15, les sols nus (N et N*) et sous couvert de végétation (F et P) sont bien différenciés d'un côté tandis que les céréales (pouvant être une classe mixte) est entre les deux. Les sols nus forment une pente positive entre les deux facteurs tandis que la végétation est quasiment perpendiculaire au facteur 1 mais parallèle au facteur 2 lequel est plus spécifique au proche infrarouge ainsi qu'à la détection de la végétation. Cet aspect des données pourrait être exploité dans d'autres études. Dans cette étude, c'est le drainage qui en est l'objet.

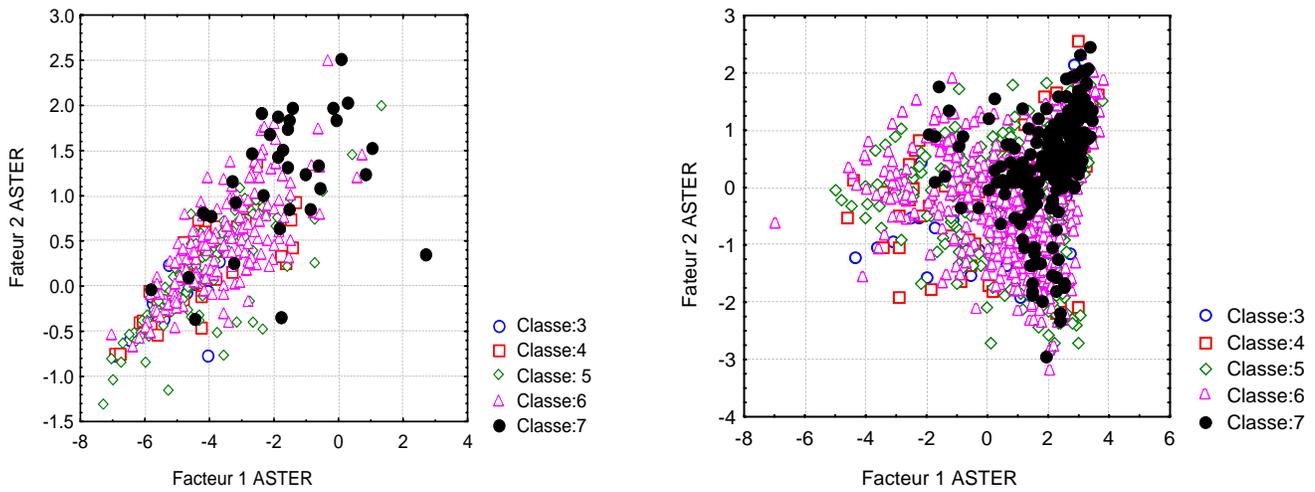


Figure 13 : Classification du drainage du sol dans l'espace réduit défini par les facteurs 1 et 2 extraits de l'ACP a) pour les sols nus et b) pour les sols sous couvert de végétation

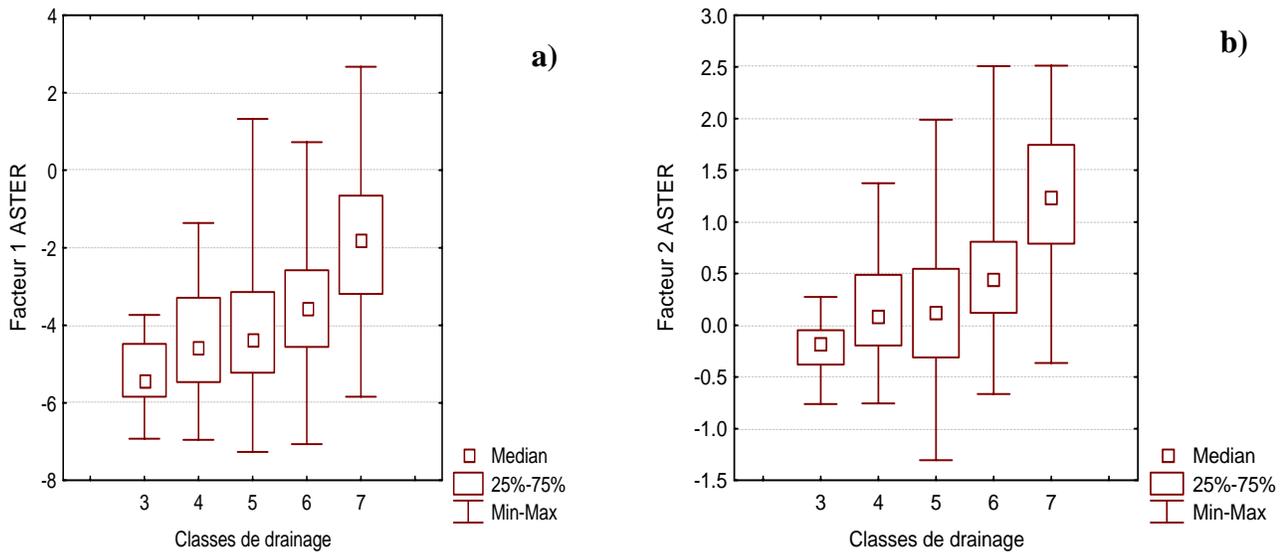


Figure 14 : Médiante, quartile et étendue des différentes classes de drainage en fonction a) du facteur 1 et b) du facteur 2 extraits de l'ACP.

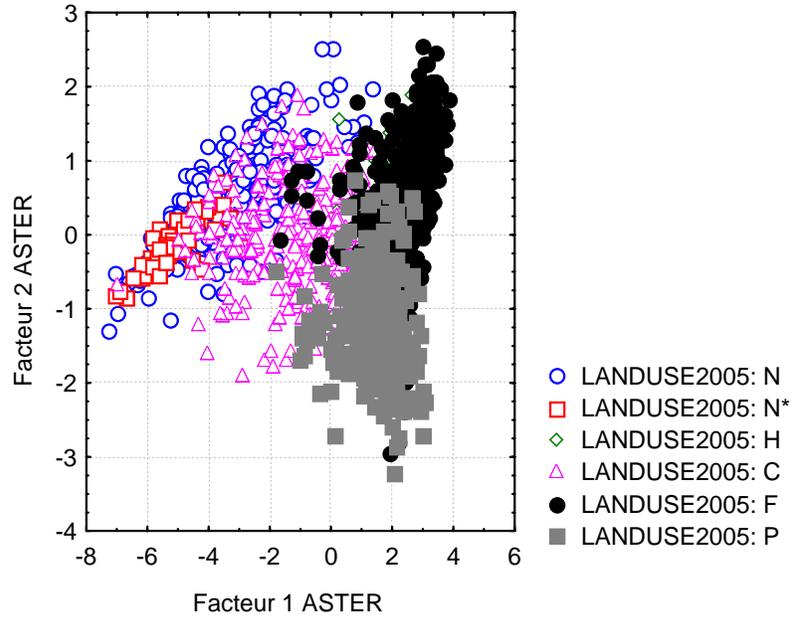


Figure 15 : Classification du type d'occupation du sol dans l'espace réduit défini par les facteurs 1 et 2 extraits de l'ACP. LANDUSE2005 : Occupation du sol au printemps 2005. N = sols nus. N*= sols nus. H = sols humides. C = céréales. F = forêts. P = prairies.

4.5 Résultats de l'étude de la normalité des données

4.5.1 Test avec la droite de Henry et la statistique (W) de Shapiro-Wilk

Pour ce qui est de la droite de Henry, les écarts à la moyenne (résidus) sont classés en rangs. À partir de ces rangs, on calcule les valeurs z (c'est-à-dire les valeurs standardisées (centrées-réduites) de la distribution normale) sous l'hypothèse que les données suivent la loi normale. Ces valeurs z sont tracées selon l'axe Y dans le tracé. Si les résidus observés (tracés selon l'axe X) sont distribués normalement, toutes les valeurs doivent être alignées sur une droite. Ceci n'est pas le cas pour les 6 images RADARSAT-1 acquises au printemps 2005 (Figure 16).

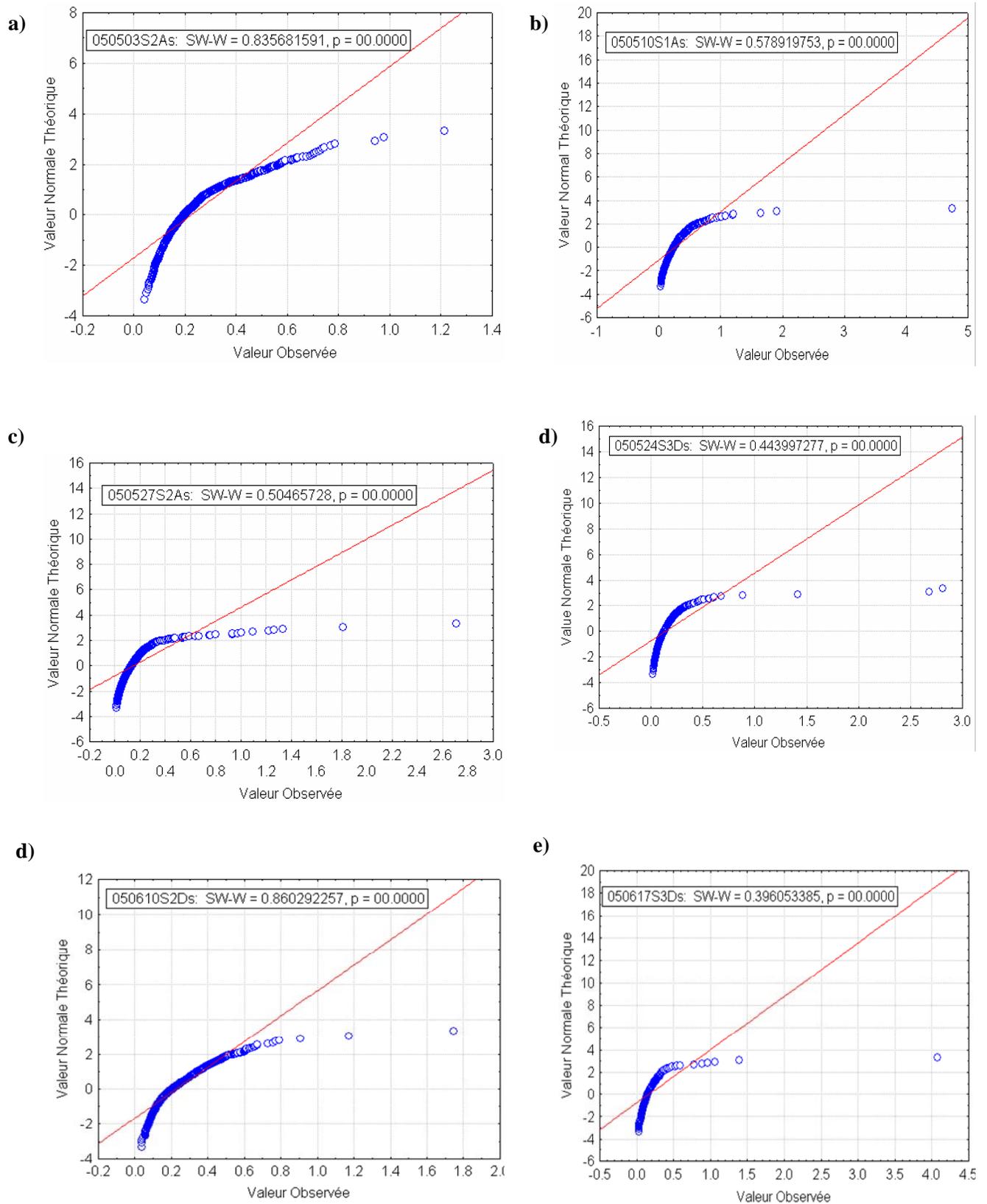


Figure 16 : Test de normalité avec la droite de Henry et la statistique (W) de Shapiro-Wilk pour les six images RADARSAT-1 acquises au printemps 2005.

De plus, les valeurs de la statistique (W) de Shapiro-Wilk sont éloignées de 1 (0.83, 0.58, 0.44, 0.50, 0.86, 0.40); ce qui confirme que l'hypothèse de la normalité des données doit être rejetée. Pour les données ASTER, le facteur 1 est à presque 10% de la normalité tandis que le second facteur est seulement à moins de 2% de la normalité

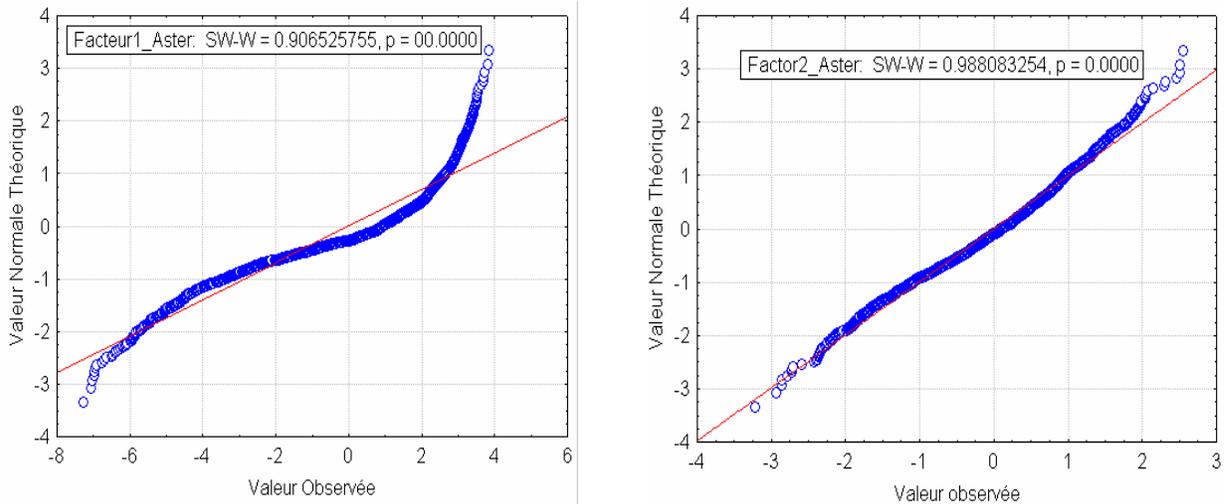


Figure 17 : Test de normalité avec le droite de Henry et la statistique (W) de Shapiro-Wilk pour les deux premiers facteurs extraits de l'ACP à partir des bandes spectrales de l'image ASTER

D'une manière générale, les points atypiques (c'est-à-dire les points qui ne suivent pas la distribution caractéristique du reste des données) mis en évidence dans tous les tracés, semblent former une structure (une forme en S) autour de la droite. Il est alors nécessaire de transformer les données. La transformation logistique (*logit*) semble à priori être mieux adaptée pour redresser les extrémités des distributions.

4.5.2 Transformation logistique (*Logit*)

Comme le drainage est une variable catégorielle à cinq classes donc multimodale, un modèle linéaire multinomial avec une liaison généralisée *logit* a été ajusté. Les données ont d'abord été centrées et réduites avant d'appliquer la transformation. Le modèle a été développé à l'aide d'une régression multiple ascendante pas-à-pas, permettant alors de visualiser les variables entrant et sortant du modèle. Pour chaque effet du modèle, la statistique de *Wald* ainsi que son niveau de

significativité statistique (p) est rapporté afin d'évaluer la contribution de chacune des données d'OT.

Le tableau 22 montre que ce sont deux données RADARSAT-1 qui ont les contributions les plus faibles (10 mai et du 17 juin 2005) sur l'ensemble des classifications. Pour la classification KMEANS à 3 classes, seulement 3 images ont des contributions importantes. Il est à noter également que c'est la classification qui a l'ordonnée à l'origine la plus faible qui devrait donner le meilleur ajustement.

La qualité d'ajustement du modèle *logit* est étudiée par les ratios entre les statistiques (des écarts, des écarts normés, du χ^2 de Pearson et du χ^2 de Pearson normé) et les degrés de liberté. Si ce rapport est très proche de 1, l'ajustement est excellent, sinon on peut déduire qu'il y a une forte dispersion des données (McCullagh et Nelder, 1989). Le tableau 23 montre que ce sont les classifications PGLE et KMEANS à 3 classes qui donnent les meilleures qualités d'ajustement.

Tableau 22 : Significativité statistique des données d'OT dans la régression multiple avec liaison logit

Effets	DRAI		PGLE		KMEANS4		KMEANS3	
	Wald	p	Wald	P	Wald	p	Wald	p
Intercept	538.38	0.00	191.03	0.00	175.07	0.00	94.41	0.00
050503S2As	15.79	0.00	42.95	0.00	21.23	0.00	10.46	0.01
050510S1As	5.30	0.26	1.01	0.80	7.11	0.07	3.89	0.14
050524S3Ds	4.37	0.36	4.74	0.19	10.74	0.01	2.77	0.25
050527S2As	5.64	0.23	0.67	0.88	13.99	0.00	3.00	0.22
050610S2Ds	17.36	0.00	14.00	0.00	25.47	0.00	7.04	0.03
050617S3Ds	7.63	0.11	1.60	0.66	7.26	0.06	4.70	0.10
Rapport03_27	20.02	0.00	106.29	0.00	4.64	0.20	2.42	0.30
Facteur1	31.93	0.00	30.27	0.00	10.63	0.01	8.97	0.01
Facteur2	54.52	0.00	32.45	0.00	56.66	0.00	1.58	0.45
NDVI	28.58	0.00	16.44	0.00	12.74	0.01	4.46	0.11

Le manque d'ajustement de certains modèles de classification du drainage à partir des données d'OT, notamment celui de la classification experte (DRAI), pourrait s'expliquer par la présence d'observations aberrantes dans le jeu de données expliquant les valeurs de χ^2 élevées et des écarts éloignés de la valeur 1.0. On voit par exemple pour DRAI (Figure 18 a et b) que les statistiques du χ^2 de Pearson sont dispersées et importantes pour la classe 3 que pour la classe 6.

Tableau 23 : Qualité d'ajustement de la transformation logit pour les différentes classifications

<i>Statistique</i>	<i>DRAI</i>	<i>PGLE</i>	<i>KMEANS4</i>	<i>KMEANS3</i>
Écart	0.61	0.82	0.75	0.99
Écarts Normés	0.61	0.82	0.75	0.99
Chi ² Pearson	1.17	1.07	0.99	1.01
Chi ² Pearson Normé	1.17	1.07	0.99	1.01

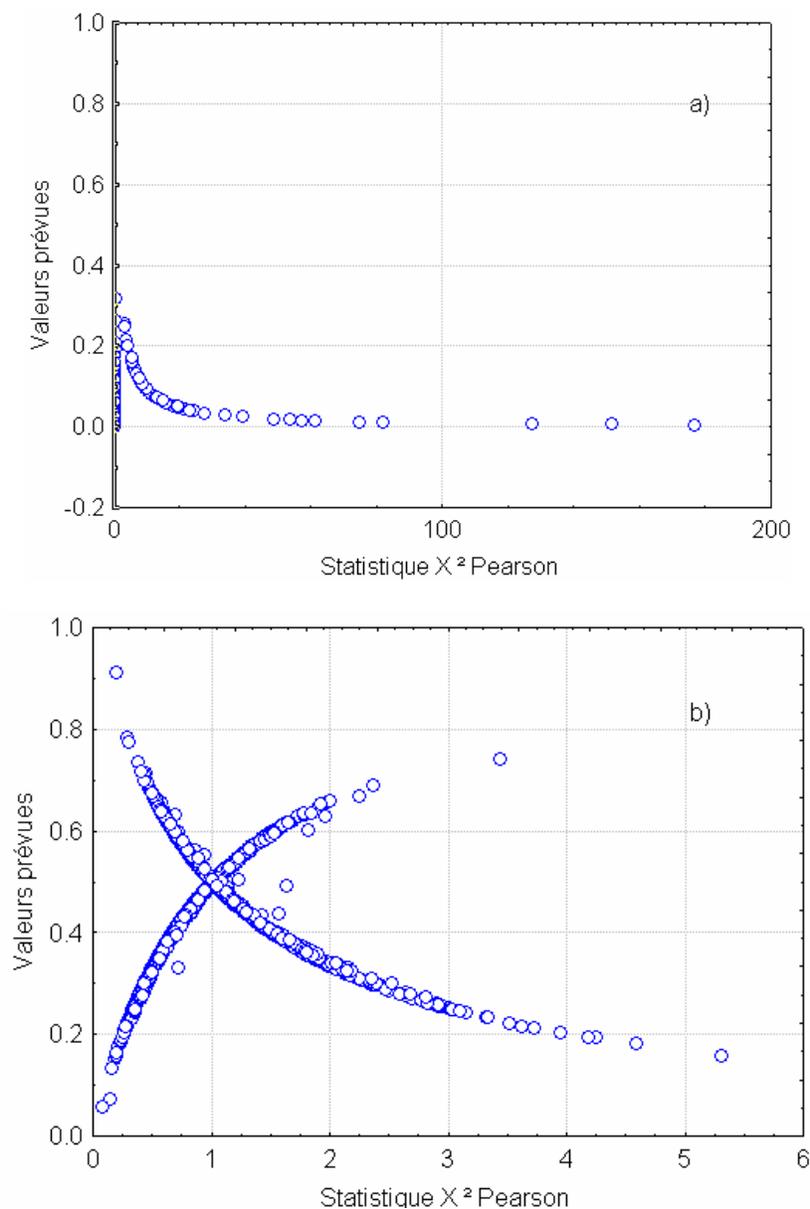


Figure 18 : Statistiques du χ^2 de Pearson en fonction des valeurs théoriques prévues par la régression multiple a) pour la classe 1 de DRAI et b) pour la classe 6 de DRAI

4.6 Résultats de l'analyse discriminante générale

Dans cette étude, les probabilités à priori prises sont celles proportionnelles à la taille des observations dans les classes de drainage. Le tableau 24 montre ces probabilités pour les différents types de classification.

Tableau 24 : Probabilité à priori associées aux différents types de classification

<i>Classes DRAI</i>	<i>Prob. à priori</i>	<i>Classes PGLE</i>	<i>Prob. à priori</i>	<i>Classes KMEANS4</i>	<i>Prob. à priori</i>	<i>Classes KMEANS3</i>	<i>Prob. à priori</i>
C3	0.03	GL4	0.13	KM41	0.21	KM31	0.42
C4	0.08	GL5	0.30	KM42	0.11	KM32	0.20
C5	0.29	GL6	0.36	KM43	0.30	KM33	0.37
C6	0.44	GL7	0.21	KM44	0.38		
C7	0.16						

4.6.1 Résultats de comparaison entre transformation logit et sans transformation

L'ensemble des données d'OT, n'ayant subies aucune transformation autre que les rendre centrées et réduites sont comparées avec les mêmes données ayant subies la transformation *logit*. L'intérêt de cette démarche est d'évaluer l'apport de la transformation *logit* dans la discrimination des classes de drainage. Pour les mêmes raisons que la régression *logit*, le modèle choisi est le pas-à-pas ascendant. Le tableau 25 présente les résultats du test de Fisher pour l'étude de la significativité statistique des données utilisées.

Le résultat est presque le même pour la statistique de Wald en ce qui a trait aux variables auxiliaires contribuant le plus à la discrimination des classes de drainage.

La régression *logit* n'améliore que légèrement le taux de classification globale mais tend à améliorer les performances de discrimination entre les différents types de classe de drainage (tableau 26). Compte tenu de ces résultats et du fait qu'elle respecte entièrement les hypothèses sous-jacentes à l'analyse discriminante, le modèle avec transformation *logit* sera retenu. Il est cependant à noter que les meilleurs succès de classification ont été obtenus avec les deux classifications automatiques dérivés de la technique KMeans (KMEANS4 et KMEANS3).

Tableau 25 : Étude de la significativité statistique des données avec le test F de Fisher

Test F	DRAI	PGLE	KMEANS4	KMEANS3
050503S2As	5.107	18.218	8.789	5.499
050510S1As	1.166	0.313	2.586	2.192
050524S3Ds	1.460	1.660	2.580	1.187
050527S2As	0.573	1.017	5.131	1.821
050610S2Ds	4.574	4.978	10.089	3.984
050617S3Ds	1.670	0.625	1.439	1.603
Rapp05_27	3.748	45.463	2.303	1.395
Factor1_ACP	9.180	11.020	4.279	4.683
Factor2_ACP	23.918	12.644	34.350	0.575
NDVI	9.930	5.322	8.038	2.265

Tableau 26 : Résultats de comparaison de l'analyse discriminante entre les données transformées par logit et celles non transformées.

DRAI	% bonne classification sans logit	% bonne classification avec logit
C3	0.00	24.53
C4	2.42	10.48
C5	12.11	11.23
C6	84.43	80.49
C7	37.45	40.64
Total	46.65	46.65
PGLE	% bonne classification sans logit	% bonne classification avec logit
GL4	18.75	25.96
GL5	34.18	29.75
GL6	68.38	67.32
GL7	22.84	26.23
Total	42.07	42.00
KMEANS4	% bonne classification sans logit	% bonne classification avec logit
KM41	18.15	24.52
KM42	97.67	97.67
KM43	65.70	61.69
KM44	86.92	86.92
Total	67.61	67.74
KMEANS3	% bonne classification sans logit	% bonne classification avec logit
KM31	85.71	89.63
KM32	24.38	16.96
KM33	83.49	83.49
Total	72.41	72.56

4.6.2 Résultats détaillés en fonction de l'occupation du sol

D'une manière générale, le pouvoir de discrimination des données d'OT doit être étudié en fonction de l'occupation du sol pour examiner dans quelle condition et avec quelle efficacité les capteurs satellitaires peuvent prédire un état de drainage. Une synthèse des résultats des quatre classifications est présentée aux tableaux 27 à 30. La classification KMEANS4 (tableau 29), bien que moins performante que celle de KMEANS3 (tableau 30), pour ce qui est du succès global de la classification semble donner une meilleure discrimination entre les classes de drainage.

Tableau 27 : Résultats de l'analyse discriminante en fonction de l'occupation du sol pour la classification DRAI

DRAI	N	N*	C	P	F
C3	0.00	50.00	6.67	0.00	0.00
C4	3.70	58.33	6.25	12.50	0.00
C5	40.20	70.59	13.00	23.16	45.71
C5	78.06	71.43	94.06	91.62	46.26
C7	40.63	100.00	4.00	13.33	33.09
Total	54.15	66.67	55.35	58.31	37.58

Tableau 28 : Résultats de l'analyse discriminante en fonction de l'occupation du sol pour la classification PGLE

PGLE	N	N*	C	P	F
GL4	32.500	66.67	23.08	8.82	13.64
GL5	51.79	73.68	34.55	41.05	35.04
GL6	61.54	57.15	74.82	86.26	81.07
GL7	14.29	85.71	2.74	6.38	26.26
Total	46.46	70.83	41.71	51.47	46.71

Tableau 29 : Résultats de l'analyse discriminante en fonction de l'occupation du sol pour la classification KMEANS4

KMEANS4	N+V	N+N*	V	C	F
KM41	75.52	0.00	68.44	11.11	99.06
KM42	58.51	98.43	16.22	26.61	50.00
KM43	5.88	4.17	1.87	1.92	2.50
KM44	76.99	18.18	84.58	90.48	1.69
Total	60.22	87.78	57.44	56.27	69.56

Tableau 30 : Résultats de l'analyse discriminante en fonction de l'occupation du sol pour la classification KMEANS3

KMEANS3	V+N	N+N*	V	C	P	F
KM31	89.63	58.65	95.25	95.24	94.53	68.26
KM32	16.96	58.82	3.54	5.13	0.00	71.43
KM33	83.49	87.12	82.26	92.16	88.89	74.19
Total	72.56	70.40	73.58	72.78	75.45	71.39

Il est à noter que dans les tableaux 29 et 30, il n'y avait pas de variance pour N*, c'est pourquoi N et N* ont été regroupés. De même, il n'y avait pas de variance pour la classe d'utilisation P dans la classification KMEANS4.

4.6.3 Résultats détaillés des données d'OT pour KMEANS4 et KMEANS3

Cette section traite de l'utilité de chaque type de données d'OT à la classification du drainage des sols selon les modèles KMEANS3 et KMEANS4. En examinant le tableau 31, on constate que les données RADARSAT-1 permettent à elles seules en condition de sols nus (N+N*) de bien discriminer la classe 2, laquelle correspond à la classe 5 (sol imparfaitement drainé) de la classification experte (DRAI). Les autres données contribuent mais faiblement à la discrimination des classes 3 et 4 de KMEANS4. La classe 1, correspondant à la classe 4, n'est pas discriminable pour les sols nus. Les données optiques n'augmentent pas la précision de la classification globale.

Tableau 31 : Résultats de classification du drainage (modèle de classification = KMEANS4) en condition de sols nus suite à une analyse discriminante appliquée à diverses combinaisons de données d’OT

<i>Données d’OT incluses dans le modèle de classification KMEANS4</i>	<i>KM41</i>	<i>KM42</i>	<i>KM43</i>	<i>KM44</i>	<i>Total</i>
RADARSAT-1	0.00	98.43	4.17	18.18	87.78
RADARSAT1 + Rapport de bandes	0.00	98.43	8.33	18.18	88.06
RADARSAT-1+ Rapport de bandes + 2 Facteurs ACP ASTER	0.00	99.06	8.33	18.18	88.61
RADARSAT-1+Rapport de bandes + 2 Facteurs ACP ASTER + NDVI	0.00	98.74	4.17	18.18	88.06

Par contre au tableau 32, ce sont les classes 1 et 4 correspondantes aux classes 4 et 7 de la classification DRAI qui sont les mieux discriminées en condition de végétation (C + P + F). Les données optiques (ASTER et NDVI) apportent près de 6% d’augmentation de la classification globale.

Tableau 32 : Résultats de classification du drainage (modèle de classification = KMEANS4) sous couvert de végétation suite à une analyse discriminante appliquée à diverses combinaisons de données d’OT

<i>Données d’OT incluses dans le modèle de classification KMEANS4</i>	<i>KM41</i>	<i>KM42</i>	<i>KM43</i>	<i>KM44</i>	<i>Total</i>
RADARSAT-1	68.44	16.22	1.87	84.58	57.44
RADARSAT1 + Rapport de bandes	69.50	17.12	1.87	84.36	57.79
RADARSAT-1+ Rapport de bandes + 2 Facteurs ACP ASTER	81.70	37.84	2.80	83.04	63.41
RADARSAT-1+Rapport de bandes + 2 Facteurs ACP ASTER + NDVI	80.90	38.74	3.27	84.14	63.75

Tableau 33 : Résultats de l'analyse discriminante en fonction des données d'OT pour les sols nus avec la classification KMEANS3

<i>Données d'OT incluses dans le modèle de classification KMEANS3</i>	<i>KM31</i>	<i>KM32</i>	<i>KM33</i>	<i>Total</i>
RADARSAT-1	54.81	62.35	87.12	70.09
RADARSAT-1 + Rapport de bandes	56.73	58.82	87.12	69.78
RADARSAT-1+ Rapport de bandes + 2 Facteurs ACP ASTER	63.46	60.00	87.12	72.27
RADARSAT-1+Rapport de bandes + 2 Facteurs ACP ASTER + NDVI	63.46	60.00	87.12	72.27

La classification KMEANS3 (Tableau 33) donne des résultats relativement bons pour les classes C5, C6 et C7 correspondant aux trois dernières classes de la classification experte, la classe 3 (7 du DRAI) étant la meilleure discriminée. Le succès global de la classification KMEANS3 n'a été augmenté que légèrement (2%) par l'ajout des données optiques.

Comme pour la classification KMEANS4 sous couverture de végétation, la classification KMEANS3 discrimine bien les classes 1 et 3 en utilisant l'ensemble des données d'OT. Le pourcentage global de classification ne varie presque pas. L'apport des données optiques est négligeable.

Tableau 34 : Résultats de l'analyse discriminante en fonction des données d'OT pour la végétation avec la classification KMEANS3

<i>Données d'OT incluses dans le modèle de classification KMEANS3</i>	<i>KM31</i>	<i>KM32</i>	<i>KM33</i>	<i>Total</i>
RADARSAT-1	96.69	4.55	82.26	74.42
RADARSAT-1 + Rapport de bandes	96.90	3.03	82.26	74.23
RADARSAT-1+ Rapport de bandes + 2 Facteurs ACP ASTER	95.66	7.07	82.26	74.42
RADARSAT-1+Rapport de bandes + 2 Facteurs ACP ASTER + NDVI	92.98	15.15	82.26	74.70

5 Conclusion générale

L'objectif principal de la présente étude est d'évaluer l'apport des données d'OT dans l'identification des classes de drainage du sol. Les résultats obtenus montrent bien le potentiel des images RADARSAT-1 et ASTER, individuellement et combinées, dans la reconnaissance des états de drainage du sol. Cependant, pour arriver à des résultats intéressants, il faut adopter une méthodologie adéquate aux traitements à la fois des données de sol et des données d'OT. L'étude de corrélation entre les propriétés physiques des sols nous a permis de sélectionner celles qui sont étroitement reliées au drainage. Avec la transformation de ces propriétés qualitatives en données continues et l'application de l'ACM, la classification du drainage proposée par les experts d'AAC correspondait assez bien à la classification automatique obtenue par arbre de décision 'QUEST'. Étant donné que près de 87% des observations correspondaient aux classes mal drainées, la première étape a consisté à évaluer la cohérence de la classification experte du drainage avec la propriété morphologique avec laquelle il était la plus corrélé (PGLE). Par la suite, d'autres modèles de classification générés à partir de 14 propriétés morphologiques du profil étroitement associées au drainage et d'un algorithme de classification automatique, la procédure KMEANS ont été examinés. Une analyse discriminante a été effectuée sur les résultats de ces classifications en respectant au préalable les hypothèses de normalité, d'homogénéité des matrices variances-covariances, de multicollinéarité et de standardisation des échelles de mesure requises pour son application. Les résultats obtenus montrent que les classes associées au mauvais drainage des sols (C5 à C7) sont bien identifiées par les données d'OT lorsqu'on utilise la classification automatique KMEANS à 4 classes en condition de sols nus, avec 98.7 % de bonnes classifications pour l'ensemble des données utilisées. Sous couvert de végétation, on a obtenu 80.9 % de bonnes classifications pour la classe KM41 et 84.1 % pour la classe KM44. Pour la classification KMEANS à 3 classes, en condition de sols nus, toutes les trois classes ont été bien discriminées avec 63.5, 60.0 et 87.1 % de bonnes classifications pour les classes 1, 2 et 3 respectivement. Cependant, sous couvert de végétation, ce sont uniquement les classes 1 et 3 qui ont été bien identifiées avec respectivement 93% et 82% de bonnes classifications.

RÉFÉRENCES

- Bell, J.C, R.L. Cunningham and M.W. Havens (1994). *Soil drainage probability mapping using a soil-landscape model*. Soil Sci. Soc. Am. J. 58:464-470.
- Bell, J.C, R.L. Cunningham and M.W. Havens (1992). *Calibration and validation of a soil-landscape model for predicting soil drainage class*. Soil Sci. Soc. Am. J. 56:1860-1866.
- Browers, S.A. and R.J. Hanks (1965). *Reflection of radiant energy from soils*. Soil Sci. 2:130-138.
- Bernier, M. and J.P. Fortin (1998). *The Potential of Time Series of C-band SAR Data to Monitor Dry and Shallow Snow Cover*. IEEE Transactions on Geoscience and Remote sensing, 36(1):226-241.
- Day, J.H. et J. McMenamin (1982). *Système d'information des sols au Canada* .Comité d'experts sur la prospection pédologique. Agriculture Canada. 196 p.
- Escadafal, R., M.D. Girard and D. Courault (1989). *Modeling the relationship between Munsell soil color and reflectance in the visible spectral bands of Land-sat MSS and TM data*. Remote Sens. Environ. 27:37-46.
- Escadafal, R., A. Heute and D. Post (1990). *Estimating soil spectral properties (visible and NIR) from color and roughness field data*. In the Proceeding of twenty-third International Symposium on Remote Sensing of Environment. Bangkok, Thailand, 18-25 April. ERIM, Ann Arbor, MI: 1263-1272.
- Fung, A.K. (1994). *Microwaver Scattering and Emission and Their Models*. Artech House, 592 p.
- Gerbermann, A.H. and D.D. Neher (1979). *Reflectance of varying mixture of clay and sand*. Photogramm. Eng. Remote Sens. 45:1145-1151.
- Goovaerts, P. (1977). *Geostatistics for natural resources evaluation*. Oxford University New, New York, 483 p.

- Hengl, T., B.M. Heuvelink Gerard and A. Stein (2004). *A generic framework for spatial prediction of soil variables base don regression-kriging*. *Geoderma*, 120(2004):75-93.
- Kaiser, H. F. (1960). *The application of electronic computers to factor analysis*. *Eucational and Psychological Measurement*, 20:141-151.
- Khattree, R. and D.N. Naik (2000). *Multivariate data reduction and discrimination with SAS software*. SAS Institute, Cary, NC, 574 p.
- Kravchenko, A.N, G.A. Bollero, R.A. Omonodo and D.G. Bullock (2002). *Quantitative mapping of soil drainage classes using topographical data and soil electrical conductivity*. *Soil Sci. Soc. Am. J.* 66:235-243.
- Loh, W.Y. and Y.S. Shih (1997). *Split selection methods for classification trees*. *Statistica Sinica* 7:815-840.
- Niang, M. (2001). *Caractérisation des états de surface d'un sol agricole par observations de terrain et par imagerie radar ERS-2*. Thèse de l'Université Paris-7. 278 p.
- Njoku, E.G. and D. Entekhabi (1996). *Passive microwave remote sensing of soil moisture*. *J. Hydrol.* 76:173-180.
- Nolin, M.C., L. Lamontagne et J.C. Dubé (1994). *Cadre méthodologique d'une étude détaillée des sols et son application en terrain plat*. Agriculture et Agroalimentaire Canada. Bulletin technique 1994-4F. Québec. 78 p.
- Royston, J.P. (1982). *An extension of Shapiro and Wilk's W test for normality to large samples*. *Applied Statistics*, 31:115-124.
- Shapiro, S.S., Wilk, M.B. and H.J. Chen (1968). *A comparative study of various tests of normality*. *Journal of the American Statistical Association*, 63:1343-1372.
- Skidmore, E.L., J.D. Dikerson and H. Shimmelpfennig (1975). *Evaluating surface-soil water content by measuring reflectance*. *Soil Sci. Soc. Am. Proc.* 39:238-242.

Ulaby, F.T., P.C. Dubois and J.v. Zyl. 1. (1996). *Radar mapping of surface soil moisture*. J. Hydrol. 184:57-84.

Vermote, E.F., D. Tanré, J.L. Deuzé, M. Herman and J.J. Morcette (1997). *Second simulation of the satellite signal in the solar spectrum, 6S: An overview*. IEEE Transactions on Geoscience and Remote Sensing, 35(3):675-686.