



**HAL**  
open science

## Génération de classes de transition vectorielles par recouvrement de classes floues

Grandchamp Enguerran, Sébastien Regis, Rousteau Alain

► **To cite this version:**

Grandchamp Enguerran, Sébastien Regis, Rousteau Alain. Génération de classes de transition vectorielles par recouvrement de classes floues. RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle), Jan 2012, Lyon, France. pp.978-2-9539515-2-3. hal-00656502

**HAL Id: hal-00656502**

**<https://hal.science/hal-00656502>**

Submitted on 17 Jan 2012

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

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

# Génération de classes de transition vectorielles par recouvrement de classes floues

Enguerran Grandchamp  
LAMIA

Sébastien Régis  
LAMIA

Alain Rousteau  
DYNECAR

Université des Antilles et de la Guyane  
97157 Pointe-à-Pitre  
{egradch,sregis,aroustea}@univ-ag.fr

## Résumé

*Nous présentons ici une méthode permettant de créer et de représenter sous forme vectorielle des classes de transition à partir d'une classification supervisée basée sur des arbres de décision flous. Cette méthode est particulièrement utile pour la classification stricte de données dont les frontières sont naturellement diffuses ainsi que pour leur représentation au sein d'un Système d'Information Géographique. C'est le cas par exemple des formations forestières entre lesquelles les transitions sont généralement étalées sur plusieurs centaines de mètres avec des gradients dépendant de conditions environnementales locales ne permettant pas de définir une transition identique dans toutes les directions. La méthode présentée permet d'ajouter des classes lorsque les zones de transition sont trop étendues plutôt que de fixer une frontière arbitrairement.*

## Mots Clef

SIG, classification, arbre de décision, flou.

## Abstract

*We present in this article a method to create transition classes from a supervised classification based on fuzzy decision trees. This method is useful to classify in a strict way data having diffuse borders and to represent the data within a Geographical Information System. This is the case of forest area where transitions between different kinds of forest are often spread over several hundred meters. Moreover, the transitions depend on local environmental conditions and the gradients are not the same for every transitions and directions. The proposed approach allows adding transition classes when transition areas are too much spread instead of fixing an arbitrary limit.*

## Keywords

GIS, classification, decision tree, fuzzy.

## 1. Introduction

Les systèmes d'information géo-référencées (SIG) ainsi que les applications les manipulant ont connu une croissance spectaculaire au cours des dernières années. Les données modélisées à l'aide de ces systèmes vont de simples positions géographiques représentant des points, à des données complexes telles que des villes, des phénomènes naturels (cyclones, pluies, etc.), des réseaux, etc.

Tous les outils présents dans les SIG permettent de manipuler des données sous forme de vecteurs ou d'images (raster). Les données dites vecteurs sont plus souples que les données raster et permettent l'ajout de données contextuelles [16].

Dès 1997, dans les conclusions du numéro spécial *Spatial Data Types for Database Systems* [23] les auteurs soulignent l'importance des structures de données dans les SIG et leur manque d'adaptation aux données floues. Même si plusieurs pistes semblent se dégager à l'époque, la difficulté d'implémentation, de modélisation et la complexité des calculs ont conduit, plus de 10 ans après à une avancée très minime sur le sujet.

Les premiers travaux sur la modélisation floue en 2D ont été menés par Peter Burrough en 1986 [1]. L'essentiel des travaux menés depuis sur la représentation et la manipulation de données 2D floues utilisent une modélisation de type raster ([17], [20], [19], [25], [26]). Cette modélisation réalise donc une partition de l'espace à l'aide d'un maillage donc chaque maille représente une surface régulière carrée donc la taille dépend de l'échelle d'analyse. Les nombreux inconvénients de cette approche ont été maintes fois soulignés dans la littérature ([1], [7], [8], [15], [22]) et sont principalement liés au manque de souplesse de ces modèles (pas de données contextuelles associées), à la lourdeur de manipulation des données (grand volume de données généré pour une analyse fine) et à la dégradation de la précision (approximation réalisée par le maillage).

Une représentation vectorielle consiste à représenter les entités présentes dans les scènes observées par des modèles géométriques construits à partir d'objets simples

tels que les points, les lignes ou les polygones. Cette représentation est adaptée à l'ajout de données contextuelles et à l'interaction avec d'autres sources d'information ainsi qu'à leur manipulation au sein d'un Système d'Information Géographique (SIG). L'inconvénient majeur de cette représentation vectorielle pour nos données est que les structures de données utilisées sont exclusivement basées sur des représentations strictes. Seules quelques rares études ont introduit cette dimension vectorielle aux Geo-traitements flous ([2], [24]) en proposant de modéliser les ensembles flous par une série de zones tampons (polygones) régulières et concentriques. La difficulté réside dans le découpage vectoriel de la fonction d'appartenance.

Nous proposons dans un premier temps un modèle flou vectoriel basé sur des ensembles imbriqués non réguliers permettant une représentation plus fiable des données. Dans un second temps, nous proposons une représentation stricte d'un ensemble de classes vectorielles floues en créant des classes de transition plutôt que de fixer une frontière arbitraire entre deux ou plusieurs classes.

Nous présentons en effet une méthode permettant de créer des classes de transition à partir d'une classification supervisée basée sur des arbres de décision flous. Cette méthode est particulièrement utile pour la classification stricte de données dont les frontières sont naturellement diffuses. C'est le cas par exemple des formations forestières entre lesquelles les transitions sont généralement étalées sur plusieurs centaines de mètres avec des gradients dépendant de conditions environnementales locales.

La section 2 présente la modélisation par des classes floues en appliquant des arbres de décision puis en donnant des représentations brutes et simplifiées des classes. La section 3 présente l'algorithme permettant de générer des classes de transition. La section 4 illustre les deux premières sections en définissant les données de référence puis en appliquant la classification floue et la génération des classes de transition. La section 5 présente quelques perspectives à ces travaux et la section 6 la conclusion.

## 2. Modélisation par des classes floues vectorielles

Une classe vectorielle est une classe dont les instances sont des objets définis à l'aide de points, de lignes et de polygones. Chaque instance est donc localisée en fonction des coordonnées des objets la définissant et est renseignée par un certain nombre d'attributs.

Pour que les classes obtenues soient vectorielles, nous pouvons soit classer directement des objets vectoriels, soit

vectoriser les résultats d'une classification raster. Etant données les couches d'information dont nous disposons (qui sont majoritairement vectorielles) et la perte de précision liée à leur rasterisation, nous avons opté pour la première approche.

Les objets à classer sont appelés Unités Vectorielles ( $UV = \{UV_i, i \in \{1 \dots M\}\}$ ), ils sont issus de l'union de différentes couches d'information ( $L = \{L_1, \dots, L_N\}$ ) représentant chacune une partition de l'espace en objets vectoriels. L'union de ces couches conduit à découper les objets provenant des celles-ci par des intersections géométriques. Ces UV représentent des zones uniformes vis-à-vis de l'information apportée par chaque couche. Chaque  $UV_i$  est donc associée à un vecteur d'attributs  $F_i = \{F_{i1}, F_{i2}, \dots, F_{iN}\}$ . La Figure 1, illustre la construction des UV à partir de deux couches.

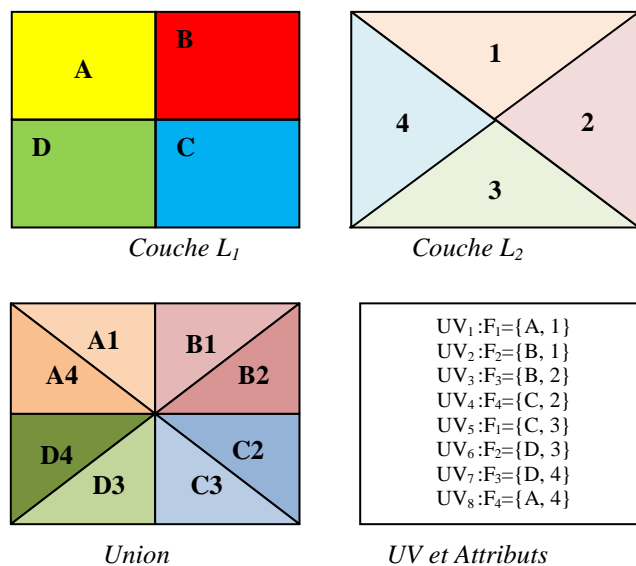


Figure 1. Principe d'union de couches

Les classes floues sont obtenues par des méthodes de classification supervisées floues. Nous utilisons pour cela des arbres de décision tels que les arbres fonctionnels (FT, [7]) ou les arbres de type C4.5 [18].

On considère un ensemble de  $n$  classes ( $C_k, k \in \{1 \dots n\}$ ) et un ensemble des  $A \subset UV$  labélisé et servant pour l'apprentissage et la validation (validation croisée).

Après apprentissage, les arbres de décision utilisés fournissent pour chaque  $UV_i$  à classer un vecteur  $\mu_i = \{\mu_{i1}, \dots, \mu_{in}\}$  indiquant le degré de confiance dans l'affectation de  $UV_i$  à chacune des classes ( $\mu_{ik}$  = degré de confiance dans l'assertion :  $UV_i$  appartient à la classe  $k$ ). Les degrés de confiance sont calculés par une régression logistique [9] donnant la probabilité d'appartenance d'un échantillon à chaque classe calculée à l'aide d'une fonction de régression définie comme une combinaison linéaire des attributs durant la phase d'apprentissage.

Le cœur de la classe  $C_k$  est défini par

$$H_k = \{UV_i \mid i \in \{1 \dots M\}, \mu_{ik} = 1\}$$

L'intérieur de la classe  $C_k$  est défini par

$$E_k = \{UV_i \mid i \in \{1 \dots M\}, \mu_{ik} > 0\}$$

L'extérieur de la classe  $C_k$  est défini par

$$\bar{E}_k = \{UV_i \mid i \in \{1 \dots M\}, \mu_{ik} = 0\}$$

On peut donc représenter pour chaque classe  $C_k$  une carte vectorielle floue basée sur les valeurs des degrés de confiance des  $UV_i$  vis-à-vis de la classe  $C_k$  :  $m_k = \{\mu_{1k}, \dots, \mu_{Mk}\}$ . L'affichage de ce type de carte est classique le niveau de gris étant directement la valeur du degré de confiance (0 pour le noir, 1 pour le blanc).

Cette représentation est vectorielle puisque chaque  $UV_i$  est un objet vectoriel. Mais elle représente un volume de données très important et n'est pas toujours très lisible, donc peu exploitable par un utilisateur voulant par exemple avoir une idée de la répartition de la classe sur un territoire. Ceci tient principalement au fait que l'on ne peut pas regrouper des  $UV_i$  adjacents pour créer des ensembles de plus grande taille car les  $UV_i$  ne sont pas affectés à une unique classe et le degré de confiance est une valeur continue.

Pour résoudre le problème, nous proposons différentes vues simplifiées d'une classe floue  $C_k$  basées sur des regroupements de valeurs du degré de confiance. Les paramètres définissant une vue  $v$  sont le nombre de groupements  $g_v$  et les bornes des intervalles correspondants sur la valeur du degré de confiance  $I^v = \{I_0^v, I_1^v, \dots, I_{g_v}^v\}$  avec  $I_0^v = 0$ ,  $I_{g_v}^v = 1$  et  $I_p^v < I_{p+1}^v$ .

Le  $p^{\text{ème}}$  groupement ( $p \in \{1 \dots g_v\}$ ) de la classe  $C_k$  ( $k \in \{1 \dots n\}$ ) selon la vue  $v$  est donc défini par

$$G_{kp}^v = \{UV_i \mid i \in \{1 \dots M\}, I_{p-1}^v < \mu_{ik} \leq I_p^v\}$$

Les UV adjacents spatialement et appartenant à un même groupement peuvent être regroupés pour ne former qu'une seule et même UV au sein d'une vue.

Nous pouvons noter que

$$\forall v, E_k = \bigcup_{p=1}^{g_v} G_{kp}^v \quad \text{et} \quad H_k \subseteq G_{k1}^v$$

Le choix des bornes des intervalles permet de mettre en relief les classes de manière différente. Les trois principales manières de fixer les intervalles sont les suivants

1. Partition régulière de  $[0,1]$ . Dans ce cas,  $I_p^v = ]\frac{p-1}{g_v}, \frac{p}{g_v}]$
2. Intervalles centrés sur les valeurs du degré de confiance les plus représentées.  $I_p^v$  est fixé de manière à ce que  $|G_{kp}^v| \approx \frac{M}{g_v}$  (/. renvoyant le cardinal et  $M$  étant le nombre d'UV,  $M =$

$\sum_{p=1}^{g_v} |G_{kp}^v|$ ). Pour cela on choisit les intervalles de manière à minimiser  $\sum_{p=1}^{g_v} (|G_{kp}^v| - \frac{M}{g_v})^2$ .

3. En définissant des intervalles de faible amplitude autour des valeurs élevées du degré de confiance, on représente plus en détail le centre des classes. Dans ce cadre, la règle à respecter pour fixer les valeurs des bornes est :  $I_{p+1}^v - I_p^v < I_p^v - I_{p-1}^v$ . Toute fonction  $f$  strictement croissante et concave de  $[0,1]$  dans  $[0,1]$  (avec  $f(0) = 0$  et  $f(1) = 1$ ) permettra d'obtenir une telle partition à partir d'une partition régulière telle qu'exprimée en 1 :  $I_p^v = ]f(\frac{p-1}{g_v}), f(\frac{p}{g_v})]$

### 3. Génération de classes de transition

Nous nous intéressons maintenant à la recherche de zones de transition étendues de manière à en déduire des classes de transition. Ces zones sont localisées en se basant sur un seuillage des degrés de confiance de la manière suivante :

1. On fixe un seuil  $S$
2. Pour chaque  $UV_i$  on trie les classes par ordre décroissant de degrés de confiance :  $\{\mu_{ij_1} \geq \mu_{ij_2} \geq \dots \mu_{ij_n}\} \rightarrow \{C_{j_1}^i, C_{j_2}^i, \dots, C_{j_n}^i\}$
3. On attribue à chaque  $UV_i$  la liste minimale des classes définie par  $E_i(S) = \{C_{j_1}^i, C_{j_2}^i, \dots, C_{j_p}^i \mid \sum_{k=1}^p \mu_{ij_k} > S\}$

Pour tout  $UV_i$ ,  $E_i(0) = \{C_{j_1}^i\}$ , nous obtenons donc une classe unique pour chaque UV ce qui revient à produire une classification stricte à partir d'un ensemble de  $n$  classes floues en choisissant la classe ayant le degré de confiance le plus élevé et de manière générale, si  $E_i(S)$  est un singleton, c'est qu'il n'y a pas d'ambiguïté sur la classe.

Dans le cas contraire, il s'agit d'une zone de transition. En fonction de la valeur du seuil, les zones de transition sont plus ou moins nombreuses et étendues mais elles ne sont ni constantes, ni régulières dans l'espace.

En faisant varier  $S$  de 0 à 1 nous pouvons hiérarchiser l'ordre d'apparition des transitions. Plus une transition apparaît pour une valeur faible de  $S$ , plus elle témoigne de l'incertitude sur l'affectation des classes dans la zone concernée. Plus elle est étendue et plus la transition entre les classes concernées est lente. Le choix d'une valeur de  $S$  permet de figer les transitions à prendre en compte.

Théoriquement toutes les transitions entre classes sont possibles et la hiérarchie pourrait faire apparaître jusqu'à  $2^n$  transitions mais dans notre cas, la cohérence spatiale des classes que nous étudions limite le nombre de transitions possibles.

## 4. Application

Nous allons maintenant appliquer notre approche à un cas concret de classification d'ensembles ayant des frontières diffuses. C'est le cas des formations forestières entre lesquelles les transitions dépendent des conditions environnementales locales. Ces transitions sont généralement étalées sur plusieurs centaines de mètres. En effet, la classification des forêts est réalisée en tenant compte de la composition floristique de celles-ci (espèces présentes, proportion de chaque espèce, etc.) qui est un paramètre variant de manière continue dans l'espace avec un gradient plus ou moins élevé. Même si la granularité des classes peut être discutée, l'aspect transitoire des formations forestières est attesté par les experts. Nous nous intéressons ici plus particulièrement aux forêts couvrant l'île de la Basse-Terre en Guadeloupe.

### 4.1. La carte de référence

Afin de comparer nos résultats, nous nous basons sur une carte écologique de référence obtenue par des biologistes [20] en 1996. Quatorze classes  $\{C_1, \dots, C_{14}\}$  ont été définies à partir de l'analyse des données floristiques de 47 placettes d'observation (Figure 2. à gauche). Les classes présentes au sein du Parc National et présentées dans les différentes figures sont nommées dans le tableau 1.

Num.	Nom
3	Horizon de formations non forestières
4	Horizon des forêts altimontaines - Etage G4
5	Horizon des forêts altimontaines - Etage G5
6	Forêts ombrophiles, montagnardes et submontagnardes au vent – Etage G6
7	Forêts ombrophiles, montagnardes et submontagnardes au vent – Etage G7
8	Forêts ombrophiles, montagnardes et submontagnardes au vent – Etage G8
9	Forêts ombrophiles, montagnardes et submontagnardes au vent – Etage G9
10	Forêts ombrophiles, montagnardes et submontagnardes sous le vent – Etage G10
11	Forêts ombrophiles, montagnardes et submontagnardes sous le vent – Etage G11
12	Forêts sempervirentes saisonnières

Tableau 1. Classes des forêts de la Guadeloupe (extrait)

La carte a ensuite été créée manuellement par les experts en tenant compte de la répartition géographique des classes et des conditions environnementales.

La partie droite de la figure 2 présente un extrait de la cartographie correspondant au Parc National de la Guadeloupe.

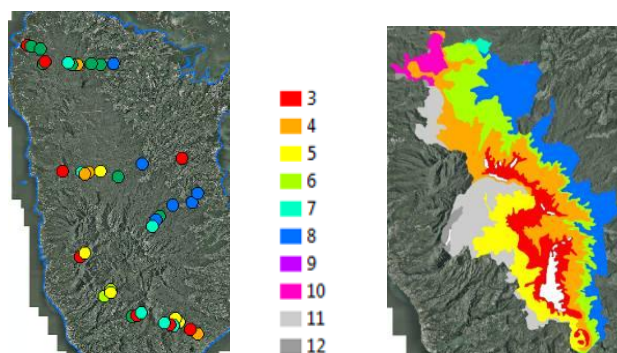


Figure 2. Vérité terrain (gauche) et extrait de la carte écologique (droite)

### 4.2 Les données utilisées pour la classification supervisée

Les couches utilisées pour l'obtention des UV sont les 4 couches d'information topographiques sur la pente, l'altitude, l'exposition et le versant ( $L = \{L_1, L_2, L_3, L_4\}$ ). Il y a au total 417 273 UV ( $M = 417\,273$ ).

Ces couches fournissent les attributs pour les UV qui seront utilisés pour l'apprentissage et la classification :

$$F_i = \{pente(UV_i), altitude(UV_i), exposition(UV_i), versant(UV_i)\}$$

Les classes utilisées sont les classes définies initialement par les experts (tableau 1).

L'ensemble d'apprentissage (A) est obtenu en sélectionnant les UV contenant les 47 placettes labellisées et en leur attribuant le label de la placette correspondant.

### 4.3 La classification stricte (S=0)

Afin de comparer notre approche avec la classification existante nous avons appliqué notre algorithme en fixant  $S$  à 0, ce qui permet d'obtenir une classification stricte.

La figure 3 montre deux extraits (Parc National de la Guadeloupe) de la classification stricte obtenue à partir des arbres de décisions FT et C4.5. Une première analyse visuelle permet de voir une certaine cohérence des résultats entre eux et par rapport à la carte de référence.

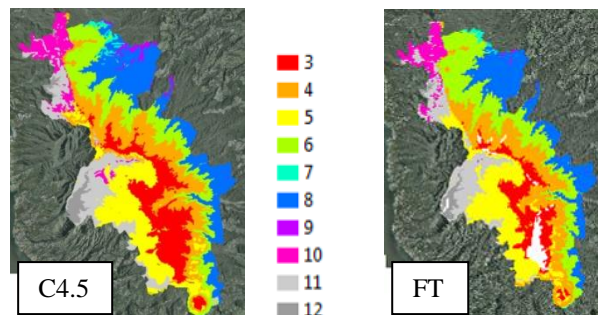


Figure 3. Extrait de la classification automatique

La figure 4 présente la carte des différences (en noir) entre la carte de référence et la classification obtenue en appliquant FT. On constate que globalement les deux



classifications se recouvrent. Les zones de conflits sont irrégulières et localisées aux frontières entre classes. Certaines frontières de la classification automatique coïncident presque parfaitement avec celles de la carte écologique. D'autres diffèrent sensiblement et l'espace de conflit s'élargit.

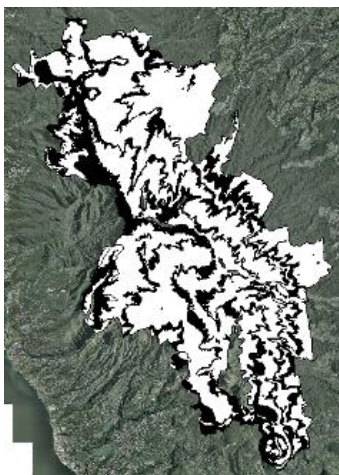


Figure 4. Différences de classification entre FT et la classification de référence

Le tableau 1 montre la matrice de confusion pour les classes concernées par l'extrait présenté. On constate que la diagonale est dominante avec des taux de bonnes classifications allant de 68% à 95.5% soit en moyenne 82.3% de bonne classification.

En tenant compte du caractère manuel et arbitraire de certaines frontières de la carte de référence, nous pouvons valider la classification obtenue à l'aide des algorithmes FT et C4.5.

	Classification de référence									
FT	3	4	5	6	7	8	10	11	12	
3	<b>77</b>	10.5	12.5							
4	4.7	<b>71.9</b>	4.3	9.1			2.2	7.8		
5	1.4	2.8	<b>95.3</b>				0.2	0.3		
6		3.3		<b>93.7</b>		3				
7				19.7	<b>78.7</b>	0.7	0.9			
8				10.2	1	<b>88</b>	0.8			
10				0.2			<b>95.5</b>	3.8	0.5	
11		1.1	10.4				7.6	<b>72.4</b>	8.5	
12							10.8	21.2	<b>68</b>	

Tableau 1. Matrice de confusion entre FT et la classification de référence (% de surface commune)

#### 4.4 Classification et représentation floue

Nous allons maintenant nous intéresser à la représentation floue des classes. C'est-à-dire à la visualisation pour chaque classe  $C_k$  de la carte des degrés de confiance  $m_k = \{\mu_{1k}, \dots, \mu_{mk}\}$ .

La figure 5. montre un affichage des cartes  $m_3$  et  $m_9$  brutes. Plus la valeur du degré de confiance est élevée, plus l'UV correspondante est claire. On remarque que la classe de gauche est beaucoup plus étalée sur le territoire que la classe de droite. Cette représentation brute est fiable mais peu exploitable en raison du grand nombre d'UV à représenter et à manipuler (environ 210 000 dans l'exemple présenté).

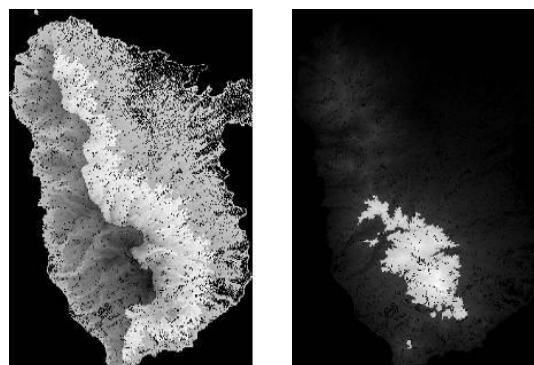


Figure 5. Cartes floues vectorielles brutes

Nous allons donc donner une représentation simplifiée de ces classes de manière à faire ressortir l'aspect général de la classe et l'évolution spatiale du degré de confiance. Etant donnée la continuité spatiale des classes et des degrés de confiance, la simplification des cartes par des regroupements de valeur de degré de confiance conduit à l'apparition de vastes zones homogènes imbriquées.

La figure 6 donne deux représentations simplifiées de la carte  $m_9$  basées sur 5 (gauche) et 10 (droite) zones imbriquées ( $g_v = 5, 10$ ). Les différences entre les deux représentations sont également liées à la manière de centrer les intervalles, à gauche sur les valeurs les plus nombreuses ( $|G_{kp}^v| \approx \frac{M}{5}$ ).  $I = \{0, 0.03, 0.07, 0.13, 0.22, 1\}$ , à droite sur les plus hautes valeurs  $I = \{0, 0.14, 0.5, 0.65, 0.73, 0.80, 0.86, 0.91, 0.95, 0.98, 1\}$ . La représentation de droite est celle qui semble la plus intéressante puisqu'elle donne plus de finesse dans la représentation des zones à fort degré de confiance.

L'aspect irrégulier des différentes zones permet de mieux représenter les valeurs réelles du degré de confiance contrairement aux études faisant intervenir des zones tampon concentriques régulières.

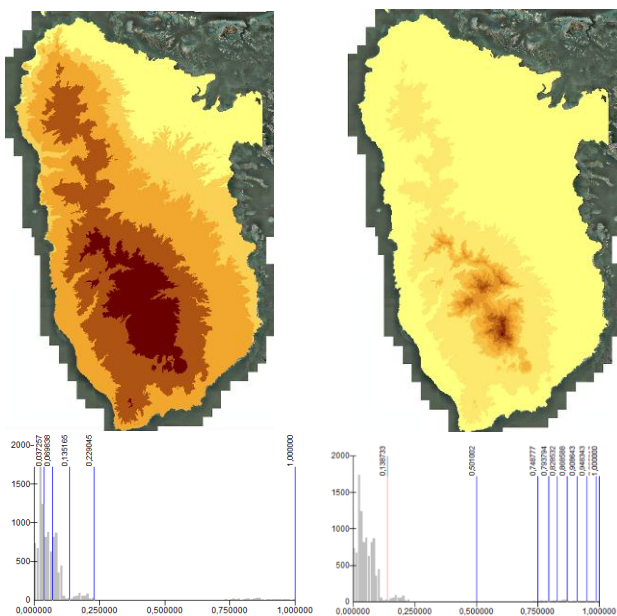


Figure 6. Cartes floues vectorielles simplifiées de  $m_9$

#### 4.5 Recherche des transitions

Nous nous intéressons maintenant à la recherche de classes de transition.

La figure 7 montre (en noir) la localisation des transitions pour deux valeurs de seuil ( $S=0.8$  et  $0.9$ ). On remarque que les zones de transition sont bien localisées à la frontière entre deux ou plusieurs classes et non au sein même d'une classe ce qui montre une certaine cohérence des attributs topographiques et environnementaux choisis. Il faut noter également que certaines transitions sont très rapides comme entre les classe 3 (en rouge) et 5 (en jaune). Ce phénomène coïncide avec un fort gradient altitudinal et environnemental.

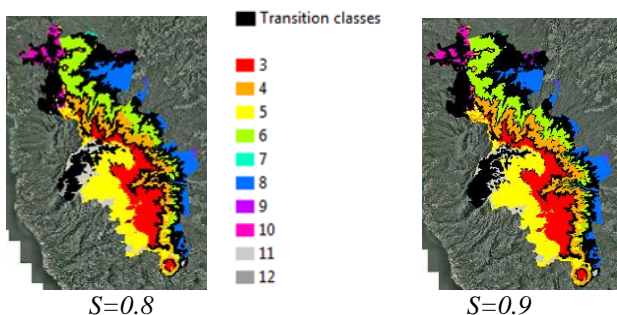


Figure 7. Classes de transition

Le tableau 2 présente la hiérarchie des transitions obtenues en faisant varier  $S$  de 0 à 1. Pour chaque valeur de  $S$  l'ensemble des transitions est indiqué puisque certaines peuvent apparaître tandis que d'autres sont englobées progressivement dans des transitions

préexistantes (par exemple la transition 10-12 présente pour  $S=0.9$  a été remplacée par la transition 10-11-12 pour  $S=1$ ).

En fixant une valeur pour  $S$  on obtient donc entre 3 et 13 classes de transitions.

S	Transitions	Superficie
0	$\phi$	0
0.1	<b>6-7, 6-8, 7-8</b>	1.2 %
0.3	6-7, 6-8, 7-8, <b>10-12, 11-12</b>	2.1 %
0.5	6-7, 6-8, <b>6-11, 7-8, 8-11, 10-11, 10-12, 11-12</b>	4.4 %
0.7	<b>3-4, 4-6, 6-7, 6-8, 6-11, 7-8, 8-11, 10-11, 10-12, 11-12</b>	21.4 %
0.9	3-4, 4-6, <b>4-11, 6-7, 6-8, 6-11, 7-8, 8-11, 10-11, 10-12, 11-12</b>	25.1 %
1	3-4, 4-6, 4-11, 6-7, 6-8, 6-11, 7-8, 8-11, 10-11, 11-12, <b>6-7-8, 6-8-11, 10-11-12,</b>	27.2 %

Tableau 2. Hiérarchie des transitions

Les premières transitions à apparaître correspondent aux classes 6, 7 et 8. La transition 6-8 s'étend également rapidement avec l'augmentation de  $S$ . C'est également le cas de la transition 10-11 qui apparaît pour  $S=0.5$  et qui s'étend rapidement. Etant donnée leur apparition pour des valeurs faibles de  $S$  et leur étendue, ces classes de transition présentent un intérêt puisqu'elles mettent en évidence de vastes zones d'incertitude.

#### 5 Perspectives

Une perspective prometteuse est l'intégration d'un algorithme de type OKM (Overlap K-Means) ([4], [5], [6]) permettant de faire émerger des classes par une approche non supervisée. En utilisant une distance dans l'espace des attributs l'algorithme permet d'affecter un échantillon à classer à l'union de deux (ou  $n$ ) classes plutôt qu'à une seule classe si celui-ci se trouve plus proche du milieu des deux centres (ou du barycentre des  $n$  centres) que du centre de l'une des classes.

Cette étape serait appliquée sur les zones de transition, après les avoir localisées tel que décrit dans notre approche. L'algorithme OKM permettrait ainsi d'affiner les zones de transition.

Une étude en cours a également pour objectif de déterminer automatiquement les classes de transition à prendre en considération en tenant compte à la fois de l'importance de la transition (valeur de  $S$  à l'apparition de la transition) et son étendue.

Enfin, la localisation des zones de transition peut guider le choix des bornes des intervalles de  $I$  de manière à les faire apparaître dans les modèles simplifiés.

#### 6 Conclusion

Nous avons présenté dans un premier temps un modèle flou vectoriel adaptatif pour des données spatiales. Ce

modèle est particulièrement utile pour des données dont les frontières sont diffuses et que l'on souhaite représenter dans un SIG avec des structures de données basées sur des ensembles stricts. L'avantage de ce modèle par rapport à d'autres approches est qu'il permet une représentation vectorielle et non raster et qu'il autorise des évolutions irrégulières du degré de confiance dans l'espace.

Dans un second temps, nous avons proposé un algorithme permettant, à partir des modèles flous, de faire apparaître des classes de transition. L'objectif consiste à identifier des zones transitoires lorsque les zones d'incertitude sont trop étendues.

Les résultats obtenus pour la classification et les classes de transition sont cohérents visuellement et numériquement avec les classifications de référence ce qui a permis de valider les approches.

## Bibliographie

- [1] Altman, D., 1994, Fuzzy set theoretic approaches for handling imprecision in spatial analysis, *Internat. J. Geographical Inform. Systems* 8 (3) 271–289.
- [2] Benz, U. C., and al., 2004, Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information, *ISPRS Journal of Photogrammetry & Remote Sensing* 5839– 258
- [3] Burrough P. A. and Frank. A. U., 1987, *Geographic Objects with Indeterminate Boundaries*. London, Taylor & Francis. chapter 12: 171-187,
- [4] Cleuziou, G., 2007, OKM : une extension des k-moyennes pour la recherche de classes recouvrantes, in 7èmes journées d'Extraction et de Gestion des Connaissances (EGC'2007). Pp. 691-702.
- [5] Cleuziou, G., 2008, An extended version of the k-means method for overlapping clustering, in 19th International Conference on Pattern Recognition (ICPR'2008). Pp. 1-4.
- [6] Cleuziou, G., 2010, Two variants of the OKM for Overlapping Clustering, in *Advances in Knowledge Discovery and Management*, Springer.
- [7] Cross, V.V., 2001, Fuzzy extensions for relationships in a generalized object model, *International Journal on Intelligent Systems* 16, 843–861.
- [8] Fisher, P., 2000, Sorites paradox and vague geographies, *Fuzzy Sets and Systems* 113 7–18.
- [9] Gama, J, 2005, *Functional Trees*. Niels Landwehr, Mark Hall, Eibe Frank. *Logistic Model Trees*.
- [10] Goodchild M., and al., 2009, Discriminant Models of Uncertainty in Nominal Fields, *Transactions in GIS*
- [11] Grandchamp, E., 2009, GIS information layer selection directed by remote sensing for ecological unit delineation, *IGARSS*
- [12] Grandchamp, E., 2010, Raster-vector cooperation algorithm for GIS, *GeoProcessing*
- [13] Grandchamp, E., 2011, Specification for a Shared Conceptual Layer in GIS, *GeoProcessing*
- [14] Kainz, W., *Fuzzy Logic and GIS, book chapter 1*
- [15] Kainz, 2011, Introduction to Fuzzy Logic and Applications in GIS – Example,
- [16] Lewinski, S., Zaremski. K., 2004, Examples of Object Oriented Classification Performed On High Resolution Satellite Images. *Miscellanea Geographica*. Volume 11.
- [17] Mukhopadhyay, B., Integrating exploration dataset in GIS using fuzzy inference modeling, *GISdevelopment*
- [18] Quinlan, R., 1993, C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA.
- [19] Raines, G. L. and al., 2010, New fuzzy logic tools in ArcGIS 10, , ESRI Communication.
- [20] Rousteau, A., 1996, Carte écologique de la Guadeloupe. 3 feuilles au 1/75.000ème et notice (36 p.). Conseil Général de la Guadeloupe, Office National des Forêts et Parc National de la Guadeloupe.
- [21] Ruiz, M.C., and al., 2007, The Development of a New Methodology Based on GIS and Fuzzy Logic to Locate Sustainable Industrial Areas 10th AGILE International Conference on Geographic Information Science
- [22] Sawatzky, D. Raines G. L., , and Bonham-Carter, G., 2008, *Spatial Data Modeller*, Technical Report,
- [23] Schneider, M., 1997, *Spatial Data Types for Database Systems Finite Resolution Geometry for Geographic Information Systems Series: Lecture Notes in Computer Science*, ISBN 978-3-540-63454-6, Vol. 1288, 275 p
- [24] Schneider, M., 1999, Uncertainty management for spatial data in databases: fuzzy spatial data types, in: *Advances in Spatial Databases, Lecture Notes in Computer Science*, vol. 1651, Springer, Berlin, pp. 330–351.
- [25] Sunila, R., and Horttanainen, P., 2009, Fuzzy Model of Soil Polygons for Managing the Imprecision Interfacing GeoStatistics and GIS.
- [26] Zhu, A. X., and al., 2001, Soil Mapping Using GIS, Expert Knowledge, and Fuzzy Logic. *Simonson, Soil Sci. Soc. Am. J.* 65:1463–1472