



République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de  
la Recherche Scientifique  
Université Ibn Khaldoun – Tiaret –



Faculté des Sciences et de la Technologie et Sciences de la Matière

Département des Sciences et de la Technologie

MEMOIRE EN VUE DE L'OBTENTION DU DIPLOME DE MAGISTER

SPECIALITE: Sciences et Technologies de l'information et Communication (STIC)

OPTION : Système d'informations et connaissances (SIC)

Présenté par

Mr. BOUALEM Adda

SUJET DU MEMOIRE :

*Théorie de l'évidence et de la logique floue :  
Application à l'appariement de données géographiques*

SOUTENU LE .....2013 Devant Le Jury Composé de :

Mr. Amar BALLA	Professeur	(ESI) -Alger -	Président
Mr. Youcef DAHMANI	Maitre de conférences A	(UIK) -Tiaret-	Rapporteur
Mr. Omar NOUALI	Directeur de recherche	CERIST -Alger-	Examineur
Mr. Rachid CHALAL	Maitre de conférences A	(ESI) -Alger -	Examineur

Année Universitaire : 2012/2013

---

## **Remerciements**

*Je tiens à remercier en premier lieu **Dieu** tout puissant de nous avoir donné la foi et courage afin de réaliser ce travail.*

*J'exprime mon profonde gratitude à mon promoteur **Monsieur DAHMANI Youcef, maitre de conférences A (UIK)-Tiaret-** le directeur de la thèse pour la qualité du sujet qu'il a proposé et la confiance qu'il m'a accordé, et accepté d'être mon tuteur.*

*Je voudrais remercier également **Mr. Amar BALLA, Professeur (ESI)-Alger-** pour d'avoir accepté d'être **président** de jury d'examen de notre mémoire.*

*Je voudrais remercier également **Mr. Omar NOUALI, Directeur de recherche CERIST-Alger-** pour d'avoir accepté d'être **examineur** de jury d'examen de notre mémoire.*

*Je voudrais remercier également **Mr. Rachid CHALAL , Maitre de conférences A (ESI)-Alger-** pour d'avoir accepté d'être **examineur** de jury d'examen de notre mémoire.*

*Je voudrais remercier également **Mr. Abdelkader ALEM, Maitre-Assistant B (UIK)-Tiaret-** pour d'avoir accepté d'être **invité** de jury d'examen de notre mémoire.*

*Ainsi tous les enseignants de l'école supérieure d'informatique qu'ils ont contribué à notre formation « école doctorat en informatique », tous les enseignants, tous les administrateurs de l'université de tiaret qu'ils ont facilité le procédure d'accès au concoure de post graduation.*

*Je tiens à remercier ma mère, qui m'a aidé par leur prière, et ma femme qui m'encouragé de continuer la recherche.*

*Je tiens à remercier vivement, **GHALEM Taher, YUCEFI Sadek, LAZRI Ahmed, Hamache AEK** qu'il ont m'a aidé à corriger ce travail.*

*J'exprime mon reconnaissance à toute personne ayant contribué de près ou de loin à l'élaboration de ce travail.*

Résumé:

L'appariement est le processus consistant à établir les correspondances entre les objets géographiques des différentes bases qui représentent le même phénomène du monde réel. Il est utilisé dans de nombreuses applications manipulant l'information géographique : regroupement de bases de données juxtaposées [31], propagation des mises à jour dans une base de données client, recalage de données sur un référentiel [32.33], intégration de bases de données géographiques (BDG) (Gouvernement du Québec), contrôle qualité [18], superposition de couches pour fusionner les géométries [42].

Pour intégrer des BDG, une clause d'appariement géométrique des données (AGD) qui spécifie le prédicat d'appariement entre les instances peut être définie. Cependant, elle peut rarement être employée du fait de la difficulté à identifier les objets homologues. Une alternative consiste à remplacer l'ensemble des clauses AGD par un processus d'appariement global. Ce processus est exécuté parallèlement à l'intégration des schémas des BDG.

Dans la littérature spécialisée, plusieurs outils peuvent être utilisés pour apparier des objets géographiques. Ils s'appuient sur :

- l'utilisation de distances entre leur géométrie
- la ressemblance de leur forme
- les correspondances entre leurs relations topologiques.
- la comparaison de la sémantique des objets,

Cependant, aucun outil pris individuellement n'est pas suffisant à cause d'imperfection dans les connaissances (l'imprécision, l'incertitude, l'incomplétude). Effectivement, les objets géographiques des différentes bases représentant le même phénomène du monde réel, présentent des différences importantes (conflits de données). Afin d'obtenir un résultat fiable, il faut donc combiner un grand nombre d'outils et s'appuyer sur les résultats d'appariements déjà réalisés pour en déterminer de nouveaux.

Les outils d'appariement sélectionnés vont renvoyer, en plus des objets à apparier, à des objets dits parasites (objets répondant aux critères définis par l'outil d'appariement mais ne devant pas être

appariés). Il faudra donc compléter les outils d'appariement par des outils de filtrage permettant de restreindre le résultat aux éléments devant être appariés.

La théorie des fonctions de croyance de Dempster [20] et Shafer [43] qui porte aussi leurs noms, contient la théorie des possibilités et des probabilités dans le cas discret. De plus, la problématique du conflit reste central dans la théorie des fonctions de croyance. C'est donc naturellement que nous cherchions à bien modéliser, gérer et utiliser la notion de conflit dans la théorie des fonctions de croyance, d'où, une des premières raisons de l'utilisation de la théorie des fonctions de croyance est certainement sa richesse en termes de modélisation des imperfections de l'information.

La logique floue née en 1965, consiste à accorder aux affirmations décrivant la réalité, un certain degré de vérité rappelant la façon de raisonner des ordinateurs à celle des humains. Le raisonnement sera donc basé sur des données approximatives pour extraire des solutions précises.

Le concept de la logique floue est la généralisation de la logique classique, qui tient compte uniquement de deux événements distincts. En effet, cette théorie considère un certain degré d'appartenance à une classe admettant des situations intermédiaires entre le «vrai » et le « faux ». En conséquence des concepts de sous-ensembles flous ont été adoptés. Ces concepts ont été introduits pour éviter les passages brusques d'une classe à une autre et autoriser une appartenance partielle à chacune des classes.

L'appariement est un processus complexe. Comme thème de recherche, le développement d'un processus d'appariement s'appuyant sur la logique floue pour définir les paramètres d'appariements, puis, le caractère d'évidence dans le raisonnement pour définir les candidats potentiels à l'appariement et pour but final initialiser les fonctions de masse pour permettre de les utiliser dans le processus d'appariement des données géographique, et une décision est prise.

Notre approche est composée de neuf étapes : une étape de définition des variables floues pour chaque critère d'appariement pris, suivie par une étape de fuzzification (la deuxième étape), qui comporte la valorisation des variables floues, selon les différentes valeurs de ces variables floues. La troisième étape est la construction d'une table de décision comportant les différentes décisions possibles. La quatrième étape est la sélection des candidats. La cinquième est l'initialisation de masses. La sixième est la définition des facteurs de fiabilités, puis la combinaison des sources

d'information. La septième est la combinaison des candidats. La huitième est la combinaison des critères. La neuvième est la prise de décision.

---

# Contenu

Résumé.....	3
Introduction générale .....	8

## Chap I : Etat de l'art

1. Les Systèmes d'informations géographiques.....	13
1.1. Généralités.....	13
1.2. Les fonctionnalités d'un SIG .....	14
1.3. Types des données géographique (les modes de représentation).....	16
1.3.1. La représentation vectorielle .....	16
1.3.2. La représentation matricielle (raster) .....	17
2. Les bases de données géographiques .....	18
2.1. La qualité des données géographique .....	18
2.1.1. Introduction .....	18
2.1.2. L'imperfection.....	19
2.2. Présentation formelle de l'information géographique par les théories de l'incertain. ....	25
2.3. Le Processus d'appariements des données géographiques .....	26
3. Conclusion .....	29

## Chapitre II: La théorie de l'évidence

1. Introduction :.....	31
2. Les fondamentaux de la Théorie de Dempster-Shafer.....	32
2.1. L'univers (cadre de discernement).....	32
2.2. La fonction de masse.....	33
2.3. La fonction de croyance et la fonction de plausibilité .....	34
2.3.1. La fonction de croyance $Bel(A)$ :.....	35
2.3.2. La fonction de plausibilité $Pl(A)$ :.....	35
2.4. Initialisation des masses:.....	36
2.5. Combinaison d'information avec la théorie de l'évidence .....	37
2.6. La prise de décision Par TDS.....	40
3. Conclusion.....	42

## Chapitre III: La théorie de la logique floue

1. Introduction .....	45
2. Les Sous-ensembles flous .....	45

2.1.	Définition d'un sous-ensemble flou .....	46
2.2.	Caractéristiques d'un sous-ensemble flou.....	46
3.	Fondamentaux de la logique floue.....	48
3.1.	Introduction .....	48
3.2.	Les opérations en logique floue .....	49
3.3.	Le vocabulaire flou : .....	50
3.4.	Règle floue: .....	51
3.4.1.	Prémisse .....	51
3.4.2.	Conséquence .....	51
3.4.3.	Implications floues .....	51
4.	Raisonnement par la logique floue.....	52
4.1.	Système d'inférence floue.....	52
4.2.	Fuzzification.....	52
4.3.	Inférence.....	53
4.4.	Defuzzification.....	54
4.5.	Exemple illustratif:.....	56
5.	Conclusion : .....	58
<b>Chapitre IV: Méthodologie et implémentation</b>		
1.	Justification du choix .....	60
2.	Approche et implémentation .....	60
2.1.	Initialisation des fonctions des masses.....	61
2.2.	Critère de la distance toponymique.....	73
3.	Résultats et comparaisons .....	75
4.	Discussion .....	78
5.	Conclusion.....	79
<b>Conclusion et perspectives :</b> .....		80
Table des figures et tableaux.....		80
Bibliographie.....		82
Annexe: Dictionnaire de domaine routier.....		85

## Introduction générale :

---

De nos jours l'information géographique constitue une ressource incontournable pour les domaines qui prennent la décision pertinente, joue un rôle très important, et pour les organisations qui considèrent l'information géographique comme support et aide à la décision, soit pour planifier, analyser (données numériques, cartes papiers, photographies aériennes), ou encore exécuter leurs projets.

Les systèmes d'informations géographiques, leurs buts situés dans la collecte, l'analyse, le stockage dans des bases de données géographiques selon les différents modes de représentation (mode vecteur, mode raster), et selon les différents points de vue d'utilisateurs et de producteurs, et cela signifie que l'information géographique qui représente les phénomènes du monde réel dans des bases de données géographiques BDG est toujours entachée des erreurs à cause de la non-unification de la représentation... , et cela met en cause l'information géographique entachée de l'imparfait (que ce soit leur type : incertaine, incomplète ou imprécise).

Le but derrière l'étude des systèmes d'informations géographiques (SIG) est la décision pertinente prise pour un problème étudié. Cependant, ces informations géographiques sont entachées par l'imperfection dans leurs représentations . Pour prendre une décision, sur des données géographiques en se basant sur la qualité des données qui représentent la différence entre la réalité et les données recueillies , il ne faut pas simplifier ces données, c'est-à-dire en tenant compte des imprécisions, des incertitudes, et des incomplétudes dans les données représentant le phénomène géographique, pour ne pas perdre l'information.

En raison de tous ces besoins liés à la mise à jour des données, à la détection des incohérences, au recalage des données, à l'étude de zones adjacentes ou à l'analyse multi-niveaux, de nombreux travaux de recherche ont été menés et ont permis d'améliorer le processus de mise en correspondance des différentes bases de données. Plus concrètement, définir les relations entre les différentes représentations consiste à expliciter les relations entre les objets homologues des bases de données géographiques existantes. Ce processus est connu sous le nom d'appariement de données. L'appariement de données est donc un outil qui permet de mettre en correspondance des objets homologues, c'est-à-dire des objets qui représentent la même réalité [48].



Le problème d'appariement des données géographiques est parmi les problèmes étudiés pour prendre une décision soit concernant l'intégration, la mise à jour des données géographiques ou le contrôle de qualité, ou pour produire des cartes, et des plans terrestres.

La modélisation des SIGs : C'est réaliser des modèles ou prototypes pour automatiser les différents types d'appariement basés nécessairement sur les théories mathématiques de l'incertain (ou de l'imparfait). Ces théories traitent les différents types d'imparfaits trouvés dans les données géographiques (BDG).

Selon la littérature spécialisée dans le domaine des (SIGs), la modélisation par ces théories montre une efficacité sur la décision finale, mais l'utilisation d'une seule théorie pour modéliser les données géographique (appariement), ne suffit pas pour une décision plus améliorante, performante, car les études faites sur ces théories montrent que l'utilisation d'une seule théorie est moins pertinente que l'utilisation d'un compromis de théories:

*Devozele* (utilise les concepts de SGBD traditionnelles sur les BDG), *Ana Maria* (utilise la théorie de (croyance) l'évidence pour l'appariement), *Anne DROMIGNY- BADIN* (Fusion d'images par la théorie de l'évidence en vue d'applications médicales et industrielles), *PLOIX M.A., GARNIER V., MOYSAN J.* (Théorie des possibilités appliquée à l'end du béton), *Essa Reem, Diab Youssef, Morand Denis* (L'application de la théorie des possibilités dans l'évaluation de la qualité environnementale des bâtiments), *Fouad ALOUANI* (Commande par Logique Floue Appliquée aux pendules inversées et au Simulateur de Vol d'Hélicoptère : Simulation et Expérimentation). De plus, ces théories presque similaires, peuvent donner une précision plus efficace pour le traitement des problèmes de l'appariement des données géographiques et leur efficacité surtout dans des domaines stratégiques comme le domaine militaire (identification des ennemis, leurs armes ), urbanisation des grandes villes, les sources naturelles (eaux, mines,...).

La contribution de cette thèse s'inscrit dans le cadre de l'appariement des données géographiques dans le domaine de la gestion de la mise à jour du fait de la complexité et l'hétérogénéité des données.

Notre objectif final est de proposer des solutions permettant l'intégration cohérente des données géographiques autant que possible d'une manière automatiques et génériques.

L'originalité de cette approche réside dans l'application de la logique floue couplée par la théorie de l'évidence pour l'initialisation des fonctions de masses pour fournir des améliorations sur les processus

décisionnels de l'appariement des données géographiques et une amélioration du critère sémantique notamment distance toponymique basé sur la distance de Levenshtein.

Le reste de ce manuscrit est organisé comme suit :

**Le chapitre I:** dans ce chapitre, un état de l'art sur les systèmes d'informations géographiques en matière de gestion des données géographiques est représenté. Nous définissons premièrement les systèmes d'informations géographiques, leurs avantages, les données géographiques, la structure de l'information géographique, leur type de représentation. Nous nous intéressant dans notre étude aux données géographiques représentées selon le mode vectoriel, les différents types d'imperfection qui peuvent entacher l'information géographique.

En suite nous avons présenté la qualité et la représentation formelle de l'information géographique par les différentes théories de l'imparfait, dans laquelle on a cité les différentes théories mathématiques de l'incertain et les types d'information géographique qui permettent de les modéliser. A la fin, le processus d'appariement des données géographique est décrit dans sa globalité.

**Le chapitre II:** aborde la première théorie de notre choix « la théorie de l'évidence », dans lequel on a présenté les fondements de la théorie de l'évidence, ses domaines d'utilisation, suivi d'une conclusion sur ce chapitre.

**Le chapitre III:** aborde la deuxième théorie choisie « la théorie de la logique floue », dans lequel on a défini le vocabulaire, les opérations de la logique floue, et présenté les étapes de cette dernière, et terminé par une conclusion sur les avantages d'utilisation (les points forts), et citer quelques domaines d'application de cette théorie.

**Le chapitre IV:** intitulé « Méthodologie & Implémentation », composé de trois parties, *la première* aborde les raisons nécessaires et suffisantes pour notre choix « théorie de l'évidence et de la logique floue », à l'appariement des données géographiques, *la deuxième* partie s'inscrit dans la stratégie de trouver une approche théorique basée sur ce compromis des théories pour améliorer; le calcul des fonctions de masses, et la prise des décisions pour l'appariement des données géographiques, et la définition d'un autre algorithme pour calculer la distance toponymique, *la troisième* partie s'inscrit dans la stratégie de projection de ces approches sur le terrain, dans lequel on a développé deux applications sous *java netbeans 7.0.1* et l'utilisation d'une base de données géographiques du réseau routier des « *ressources naturelles et faune Québec*»[6], et on a montré des exemples de comparaison

concrets par les deux approches d'appariement (Travaux d'Ana Maria [8], et la notre), les résultats trouvés, les comparaisons faites. Après, nous avons réalisé une étude de cas utilisant un exemple illustratif complet traité par (l'approche d'Anna Maria [8] utilisant l'algorithme de samal [40] pour le calcul de la distance toponymique, et notre approche. Nous avons terminé ce travail par une conclusion et perspectives pour d'éventuels travaux futures.

---

A decorative frame with a light blue border and four ornate floral corner pieces in shades of green, yellow, and red. The text is centered within this frame.

Chapitre I: « Etat de l'art »  
Les systèmes d'informations  
géographiques

---

## 1. Les Systèmes d'informations géographiques

### 1.1. Généralités

Un SIG est un système informatique permettant, à partir de diverses sources, de rassembler et d'organiser, de gérer, d'analyser, de combiner, d'élaborer et de présenter des informations localisées géographiquement et contribuant notamment à la gestion de l'espace [47] (Figure 1).

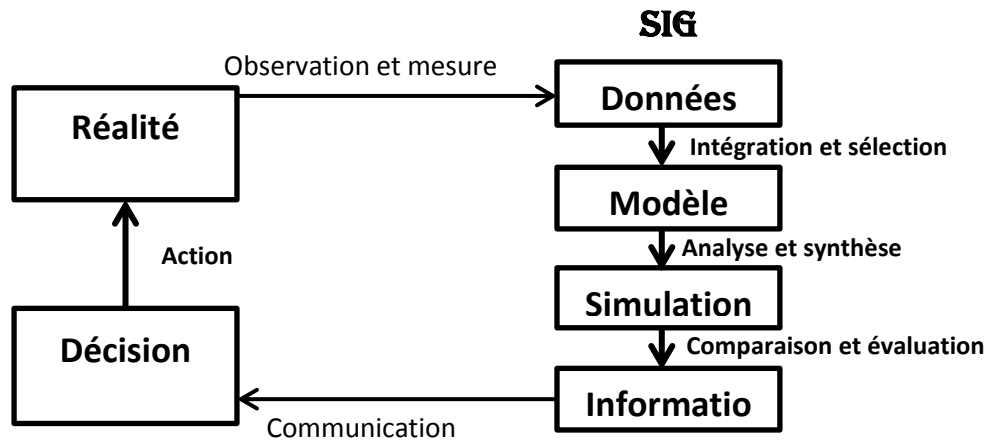


Figure 1 : Définition d'un SIG [47]

Un SIG est l'ensemble de matériels, de logiciels, de données, des personnes et de compétences mises en place pour analyser un territoire [1] (Figure 2).

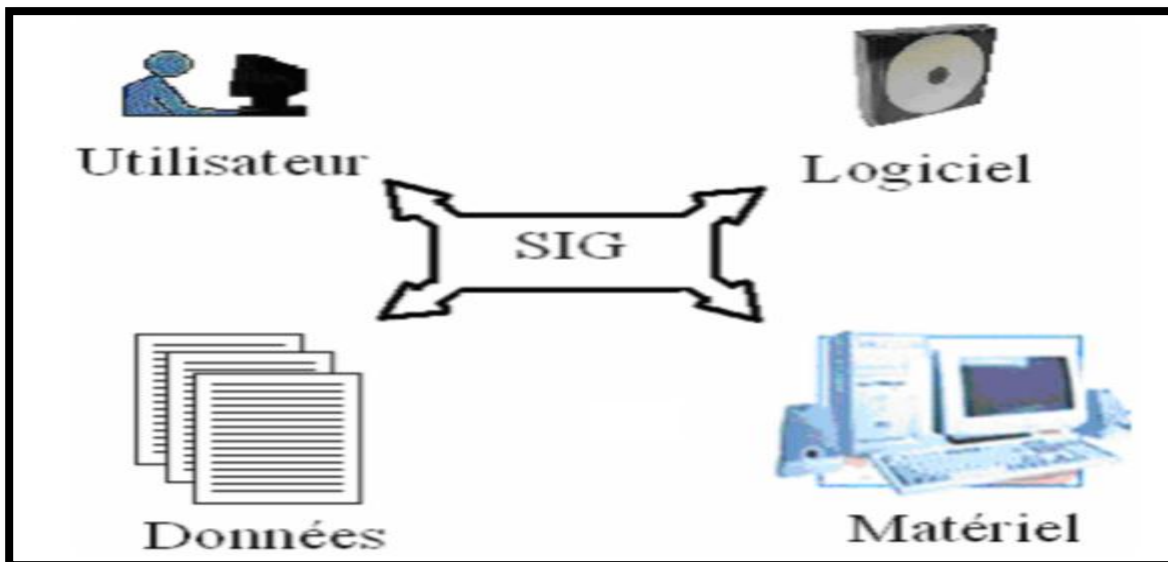


Figure 2 : Les constituants d'un SIG [3]

Les Systèmes d'Informations Géographiques sont le résultat de la combinaison et de la fusion de multiples disciplines:

- **La géographie** : qui fournit les méthodes d'analyses et d'organisation spatiales.
- **La cartographie** : qui donne les méthodes de représentation de l'espace géographique.
- **La télédétection** : qui offre les techniques d'acquisition et de traitement de données.
- **La photogrammétrie** : qui explique des photographies aériennes.
- **La géodésie** : qui fait des localisations très précises.
- **La mathématique** : qui met à la disposition des spécialistes en systèmes d'informations géographiques, les différentes méthodes de calcul matriciel, de calcul différentiel et intégrale, de trigonométrie, de géométrie, d'interpolation et d'extrapolation des courbes.
- **La statistique** : qui offre de la possibilité d'analyse et d'interprétation des mesures.
- **L'informatique** : qui permet de faire l'infographie et la conception et de réaliser des bases de données.
- **Le génie civil** : qui est lié aux applications en transport et en génie urbain [47].

## 1.2. Les fonctionnalités d'un SIG

Un SIG offre la possibilité de saisir des données géolocalisées, les stocker et les réorganiser, les modéliser, les analyser et aussi les représenter sous forme alphanumérique ou graphique [3]. Entre autres, un SIG assure comme fonctions :

- La conversion d'un système de coordonnées à un autre.
- La disposition des objets dans un système géoréférencé.
- La possibilité de superposition des cartes de sources différentes.
- L'extraction de tous les objets géographiques situés à une distance bien déterminée d'un objet spécifique.
- La fusion des objets ayant une caractéristique commune.
- La détermination du plus court itinéraire pour se rendre à une destination donnée [47].

Les Systèmes d'Informations Géographiques apportent un ensemble de bienfaits, à savoir :

- Le stockage des informations est clair et définitif.
- La compréhension des phénomènes et la prévention des risques.
- L'établissement des cartographies rapides.
- La localisation dans l'espace et dans le temps.

- La réaction rapide face aux catastrophes.
- La fourniture des itinéraires.
- L'exploitation d'une grande quantité d'informations.
- Le tri par la superposition sur des couches distinctes.
- La possibilité d'ajouter des étiquettes de texte.
- La lisibilité de l'affichage [1].

Les SIGs manipulent des données géographiques. On entend par données géographiques une représentation ou une reformulation de la réalité localisée dans l'espace et le temps. C'est en fait la combinaison des données sémantiques en gérant des données alphanumériques spatialement localisées et des données géométriques. L'information géographique est représentée habituellement sous forme cartographique. Elle est composée de deux composantes :

- *Une composante graphique* : il s'agit de la carte décrivant la forme et les caractéristiques d'un objet en déterminant ses coordonnées cartographiques.
- *Une composante attributaire* : il s'agit de la légende identifiant les objets représentés en donnant leurs significations [1].

En se basant sur des primitives graphiques comme les points, les lignes, les arcs et les polygones, les Systèmes d'Informations Géographiques permettent de réaliser une représentation réaliste de l'environnement concret. Enfin, l'information géographique est de qualités diverses. Les données stockées dans les SIG ne donnent pas une vision exacte et parfaite sur la réalité de phénomène géographique étudié. En effet, l'étude de la qualité des données est nécessaire pour l'obtention de degré de confiance dans les SIG's, la précision et la fiabilité des résultats requis à la fiabilité des instruments de mesure sur les sources d'informations et selon les modes d'expertises utilisés.

L'information géographique est formée par deux types d'informations : les informations spatiales et les informations alphanumériques. Les deux types jouent deux rôles indispensables à la représentation des informations géographiques, mais chacun à sa manière. En effet, les données spatiales sont organisées en couches ; par contre, les données alphanumériques sont structurées en base de données.

L'ensemble forme une base de données géographique qui peut être définie comme étant un ensemble de couches superposées accompagnées par des étiquettes de texte qui y fournissent de la sémantique [47], (Figure I.3).

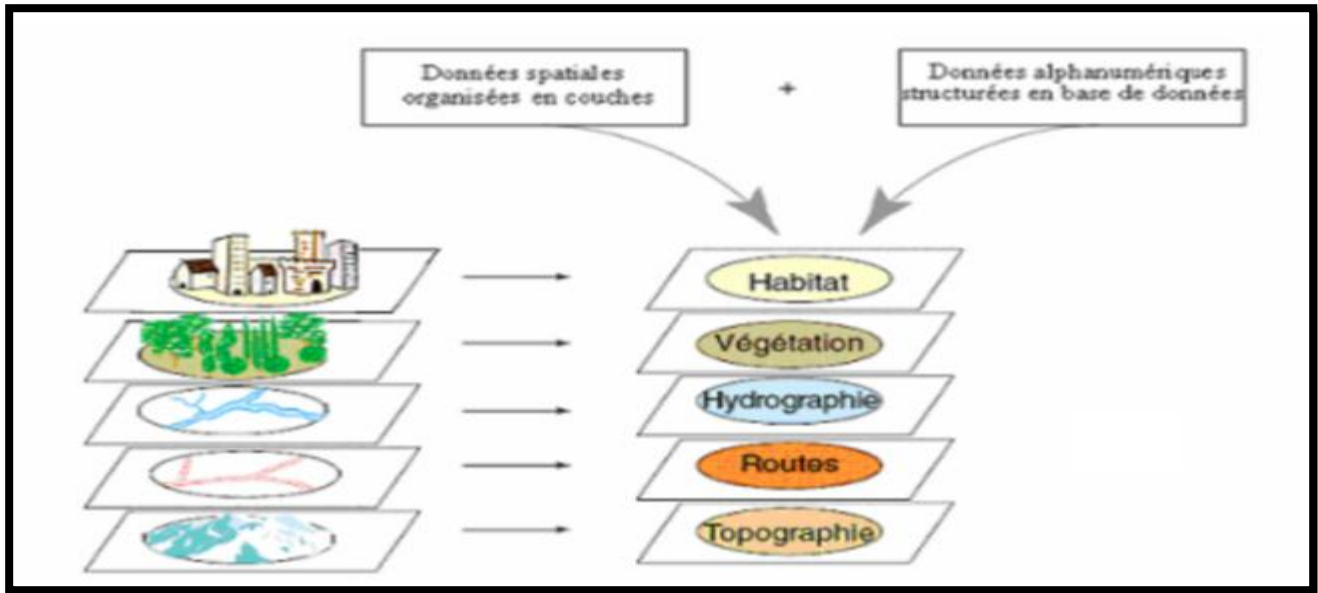


Figure 3 : Structure de l'information géographique [47].

### 1.3. Types des données géographiques (les modes de représentation)

Les logiciels SIGs, sont capables de gérer les données géographiques sous deux grandes représentations.

#### 1.3.1. La représentation vectorielle

Un objet géographique possédant une position, une forme, et des attributs descriptifs, classés :

- Sous une grandeur géométrique qui désigne la position par les coordonnées spatiales comme les abscisses  $x_i$ , les ordonnées  $y_i$ , les hauteurs  $z_i$ , la forme désignée par les trois primitives : les points, les lignes et les polygones.
- Sous une grandeur sémantique qui désigne ses attributs descriptifs comme les noms, le nombre des voix, types (routes principales, autoroutes,...) pour les objets géographiques routes.
- Sous une grandeur topologique qui désigne les différentes relations de voisinage entre les objets géographiques, comme « l'objet A est proche de la rivière », « l'objet A limité par le nord par la forêt », « inclus .... », et d'autres relations de voisinage.

Parmi les avantages d'utilisation de la représentation vectorielle :

- Donne une représentation très conforme à la réalité
- La localisation et les dimensions des objets sont calculées avec précision



- On peut individualiser les objets, donc leur attacher des attributs.
- La taille du fichier est réduite.

### 1.3.2. La représentation matricielle (raster)

Appelée aussi la représentation raster, l'information géographique est représentée aussi par une matrice, constituant les éléments qui forment un pavage (les pixels dans une image). Chaque élément de la matrice porte un ou plusieurs attribut. Il s'agit d'une image, d'un plan ou d'une photo numérisée et affichée dans le SIG en tant qu'image.

Ce mode correspond à une division régulière de l'espace sous forme de cellules ou mailles généralement carrées appelées pixels, qui définissent la précision minimale de la structure. Les pixels sont par exemple, les centaines de milliers de points lumineux et colorés qui composent votre écran d'ordinateur.

Le mode raster s'applique aux traitements d'images (satellites, photos aériennes). On appelle résolution la taille du pixel : (un pixel équivaut à x mètres sur le terrain).

L'insertion d'une couche raster dans le SIG se fait par scannérisation. La précision du document scanné (carte ou image) dépend de son échelle et de la résolution du scanner, exprimée en dpi (dots per inches) ou ppp (pixels par pouce), sachant que 1 pouce = 2,54 cm.

Parmi les avantages d'utilisation de la représentation raster :

- Facilité d'utilisation : données sont sous forme de tableau. Par rapport au mode vecteur, la dimension thématique est donnée par des valeurs numériques de la grille et la dimension spatiale est déduite par la position relative du pixel dans la grille.
- Le croisement des données est facile à réaliser : toutes les grandeurs sont ramenées à la même unité de base (le pixel).
- Elle se prête bien à certains types de traitements numériques car chaque pixel contient une valeur numérique (ex : classification supervisée). Cette valeur est stockée dans un canal (raster monocanal) ou plusieurs canaux (raster multi-canal), chacun d'eux représentant une information distincte. Le canal est au raster ce que l'attribut est au vecteur.

Il est possible de convertir les rasters en vecteurs, après avoir scanné une image, et l'inverse est aussi possible et s'appelle la rasterisation.

Parmi les inconvénients d'utilisation de la représentation raster :

- fichier lourd en mémoire
- manque de précision
- qualité médiocre des documents à l'impression (phénomène d'aliasing ou marches d'escalier)
- pas d'individualisation des objets.

## 2. Les bases de données géographiques

Une base de données géographique est un composant indispensable pour la mise en œuvre de l'outil envisagé. Elle va servir à stocker les données d'une façon structurée et avec le moins de redondances possibles. De plus, elle va correspondre à la nature spatiale des données manipulées.

Dans notre cas, la base de données géographiques en question est largement rattachée à l'aspect réseau. Elle joue le rôle d'une entité centrale, implémentée dans un serveur et gérée par un Système de Gestion de Bases de Données (SGBD). Ce dernier met en commun la totalité des données du problème. Le serveur répartit et permet l'accès aux données entre plusieurs utilisateurs. Il offre la possibilité d'accès aussi bien simultané du point de vue temporel que distant dans le sens spatial. Il permet également de mettre les données à la disposition des utilisateurs pour une consultation ou même une mise à jour selon les droits d'accès accordés à ces derniers.

Ces données vont être, par la suite, utilisées par des applications spécifiques, installées au niveau des postes utilisateurs, pour répondre à des besoins supérieurs. On parle alors d'un SIG puisqu'il s'agit d'une structure regroupant les moyens permettant de partager des données de nature spatiale entre différents utilisateurs. La fiabilité est une propriété importante pour n'importe quel SGBD. Elle se résume par le fait d'enregistrer les modifications induites par les transactions, bien entendu après une commutation, sur un support de stockage non volatile pour garantir la permanence et la conservation des données même en absence de l'alimentation électrique, l'endommagement du matériel ou bien la défaillance du système d'exploitation, sans considérer les problèmes pouvant affecter la mémoire non volatile.

### 2.1. La qualité des données géographiques

#### 2.1.1. Introduction

La qualité des données géographiques est très complexe, car les données géographiques localisées doivent être correctement référencées sur la terre, leurs caractéristiques doivent être

correctement décrites, et leur agencement les unes par rapport aux autres doit également être conforme à la réalité, donc la problématique de la qualité des informations géographiques est associée directement à la gestion de leurs imperfections, c.-à-d. l'imprécision, l'incertitude, l'incomplétude, et l'ambiguïté, et d'autres formes d'imperfection.

### 2.1.2. L'imperfection

Les systèmes d'information géographique (SIGs) dédiés à la gestion de l'information géographique permettent la gestion de grands volumes d'information sur l'environnement naturel présents et passés. Ces connaissances sont sujettes à différentes formes d'imperfection. Si cette imperfection est minorée dans la représentation des données, la validité des résultats des processus de généralisation et de mise en relation des objets géographiques peut être remise en question. Ainsi, l'imperfection de l'information impacte sur la qualité des données et des décisions prises pour les différents domaines (sécurité des forêts contre les feux, l'aménagement, l'agriculture, les cadastres, le domaine militaire...).

Il est donc essentiel que la qualité de données spatiales soit étudiée et intégrée au processus d'analyse en fonction de la nature de l'information disponible.

Dans ce cadre, la compréhension des différentes formes d'imperfection est fondamentale [37].

Dans ce qui suit, on va examiner la nature conceptuelle des différentes sortes d'imperfection pouvant être attachées à l'information géographique dans ses caractéristiques spatiales, sémantiques, et topologiques.

En se référant à [28], on peut distinguer, en étudiant les données, si les classes d'objets ou les objets suivants sont bien ou mal définis. Dans le cas où l'objet et la classe sont bien définis, on sera soumis à l'incertitude. Dans les autres cas, l'objet ou la classe sont mal définis et l'imperfection des données sera due à l'imprécision, à l'ambiguïté et/ou à l'incomplétude.

- **Incertitude**

L'incertitude est liée à la validité d'une connaissance, c'est-à-dire ne pas avoir à répondre à une question dans un contexte donné : comme par exemple :

- On ne sait pas si une proposition est vraie ou fausse
- On ne sait pas si un événement va se produire

Il peut y avoir plusieurs causes à l'incertitude :

- La validité d'un phénomène empêche de prévoir le résultat de la prochaine expérience
- Le manque d'information
- La présence d'informations contradictoires.

Traditionnellement, toutes ces incertitudes sont traitées par la théorie de probabilités.

Pour l'information géographique [29], propose l'ensemble, présenté dans le *tableau 1*, des types et causes classiques d'erreurs qui sont sources d'incertitudes dans les données. En suivant la présentation de Fisher, l'incertitude à laquelle est sujette l'information peut être issue d'erreurs de mesure, d'enregistrement, de classement, de regroupement de classes, de généralisation spatiale, temporelle et de traitement. Ces erreurs sont pour certaines constituantes de l'information contenue dans le système (mesure, enregistrement, temporelle, classement) et pour les autres de l'exploitation du système (regroupement, analyse, traitement). L'incertitude est alors attachée à l'information se dégageant de l'utilisation du système (données + résultats d'analyses) [36].

*Tableau 1 -Types et causes d'erreurs qui sont sources d'incertitudes dans une base de données géographiques - d'après [29].*

Types d'erreurs	Causes d'erreurs
De mesure	La mesure d'un objet est erronée (localisation)
De classement	L'objet est affecté à une mauvaise classe d'objets à cause d'une erreur dans l'expertise
De regroupement de classes d'objets	L'objet est regroupé avec des objets aux propriétés différentes
De généralisation spatiale	Généralisation de la représentation cartographique de l'objet avant la numérisation (déplacement, simplification, échelles d'approches).
D'enregistrement	Mauvais codage en entrée au SIG
Temporelle	L'objet à changer de caractéristique entre la date d'enregistrement dans la base et celle de son utilisation
De traitement	Une erreur est causée pendant le traitement par un arrondi ou une erreur algorithmique

#### ▪ **Imprécision**

L'imprécision de l'information est due au caractère vague ou approximatif de la sémantique utilisée. Le caractère vague de l'information est la conséquence d'une insuffisance des instruments d'observation et de l'utilisation de connaissances subjectives ou flexibles.

L'approximatif est issu de catégories aux limites mal définies. Contrairement à l'incertitude qui porte sur la validité de l'information et que l'on peut voir d'un point de vue statistique ou probabiliste, l'imprécision est un des défis du domaine de la logique et de la cognition, en particulier :

- **Vague**

L'information est vague si elle est définie à l'aide de connaissances subjectives ou flexibles, ou si elle est la conséquence d'une insuffisance des instruments d'observation.

- **Approximation**

Le caractère approximatif de l'information provient de catégories aux limites mal définies, c'est-à-dire l'objet dont la représentation ne peut être délimitée clairement.

À l'imprécision et l'incertitude, on peut ajouter l'ambiguïté et l'incomplétude aux formes possibles de l'imperfection de l'information. L'ambiguïté est à considérer dans le cas de choix entre de multiples identifications possibles d'un même objet.

- **Ambiguïté**

L'ambiguïté survient lorsqu'il y a un doute sur la manière de définir un objet ou un phénomène, c'est-à-dire quand un élément peut appartenir à plusieurs catégories disjointes ou d'échelles différentes, ou encore quand la description de l'élément peut donner lieu à plusieurs sens. Peter Fisher, dans [28], distingue les deux cas d'ambiguïté suivants : le désaccord ou conflit, le manque de spécificité ou non-spécificité.

- **Conflit**

Il y a conflit si au minimum deux classifications contradictoires pour un unique objet sont possibles.

En archéologie, le cas d'un objet, auquel on pourrait attribuer deux représentations temporelles, spatiales ou fonctionnelles différentes et non incluses, illustre une situation conflictuelle. Par exemple, si deux expertises décrivent un même objet comme étant pour l'une un mur, pour l'autre un pavage de rues, la description de l'objet de manière unique amènera à choisir l'une ou l'autre.

La résolution du conflit ou désaccord revient à choisir une approche (à gérer l'incertitude) sur des objets (ou relations) imprécis (ayant plusieurs représentations disjointes possibles). Cependant les sémantiques peuvent ne pas être disjointes. C'est le cas d'informations non-spécifiques.

### ○ Non-spécificité

Lorsqu'une définition d'une relation ou d'un objet peut amener à plusieurs sens, ou lorsque l'échelle de l'analyse est susceptible d'amener à de multiples interprétations, on parle alors de non-spécificité.

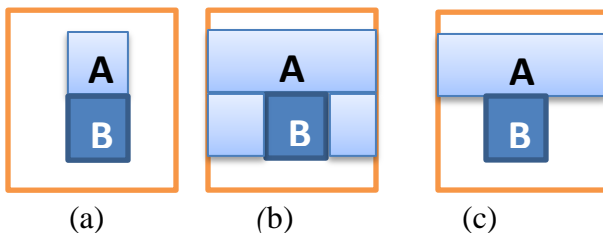
Pour l'exemple extrait de [29] illustrant la non-spécificité spatiale par la relation « au nord de ». La relation A est « au nord de » B, cette relation peut être interprétée par :

- que les objets sont sur la même longitude mais que A est plus au nord que B sur cette ligne (voir figure 4.a) ou
- que A est quelque part au nord du parallèle qui passe par B (voir figure 4.b) ou
- que A se situe entre le nord-nord-est et le nord-nord-ouest de B (voir figure 4.c).

Les deux premières définitions sont précises et valides, la troisième est vague.

Cependant, l'incertitude, l'imprécision et l'ambiguïté ne permettent pas de décrire toutes les formes d'imperfections. En archéologie, comme en géographie, et dans toutes les bases de données en général, l'incomplétude de l'information est très répandue.

Figure 4 : Non-spécificité de la relation A « au nord de » B



### ▪ Incomplétude

Les incomplétudes sont des absences de connaissances ou des connaissances lacunaires sur des informations du système.

- **Absence**

L'absence est le phénomène qui survient quand dans une base de données, des valeurs manquent à la description de certains objets. On parle d'absence quand, pour un objet particulier, il manque une information nécessaire à sa complète définition. En géographie, par exemple, si l'on souhaite stocker dans un SIG un pays dont on a la représentation spatiale mais pas le nom, la description du pays est incomplète du fait de l'absence de l'information descriptive.

- **Lacune**

La lacune est le fait qu'un ou plusieurs objets de la base de données ne décrivent que de manière partielle une structure les englobant. Les connaissances expertes sur la structure étudiée permettent de mettre en évidence l'aspect fragmentaire de la description du phénomène obtenue selon les données contenues dans la BDG associée au SIG.

On remarque bien qu'une échelle de description archéologique donnée (comme par exemple l'échelle du fait), les objets peuvent être bien définis (des tronçons de rue peuvent théoriquement être bien définis), par le changement d'échelle (le passage à l'échelle de la structure), les objets à la nouvelle échelle (les rues) sont mal définis car l'information qu'ils comportent apparaît alors comme lacunaire (les rues ne sont que partiellement définies).

On Remarque que l'absence et la lacune de l'information nécessitent généralement des traitements spéciaux des données en présence afin de compléter cette information. Ce traitement peut, par exemple, être une analyse spatiotemporelle ou une généralisation.

[30] proposent la classification des incertitudes présentée dans la figure 5. À partir de cette structuration, plusieurs approches spécifiques à l'information géographique ont été proposées, comme celle proposée dans [28] et dans [29].

Dans cette partie, la typologie de l'imperfection présentée, qui se base sur celle proposée par Fisher, considère l'imperfection et ses formes. Elle est illustrée dans la *figure 6*.

Elle se différencie de l'approche de [30] qui observe l'incertitude. Par l'utilisation de cette typologie, l'étude de la nature des imperfections auxquelles sont sujettes les données géographiques est facilitée. L'information peut alors être analysée en tenant compte de ses imperfections, et ainsi de sa qualité et par la même d'une partie de sa complexité.

En décrivant, selon leur qualité, les données stockées dans la base de données associée à un système d'information à références spatiales dédié à un usage géographique, nous guidons le choix de la théorie de représentation pour la modélisation de l'information, ce choix est nécessaire à la préparation des données en vue de les analyser.

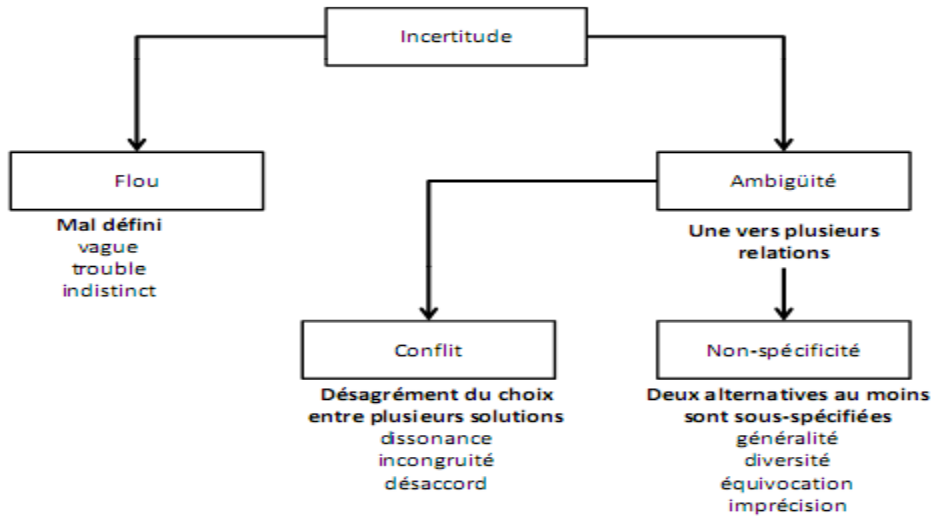


Figure 5 : Topologie de l'incertitude selon Klir et Yuan, [30]

Le choix de la théorie dépend de la nature de l'imperfection. Pour faire ce choix, on dispose, entre autres, des différentes théories citées dans (section 2.2 «Présentation formelle de l'information géographique par les théories de l'incertain »).

Ces théories ont toutes pour objectif de permettre la modélisation de l'information selon leurs types d'imperfection. Elles sont complémentaires et permettent la représentation des données imparfaites, qu'elles soient de nature incertaine, imprécise ou ambiguë. L'incomplétude doit-elle être comblée par des traitements de données.



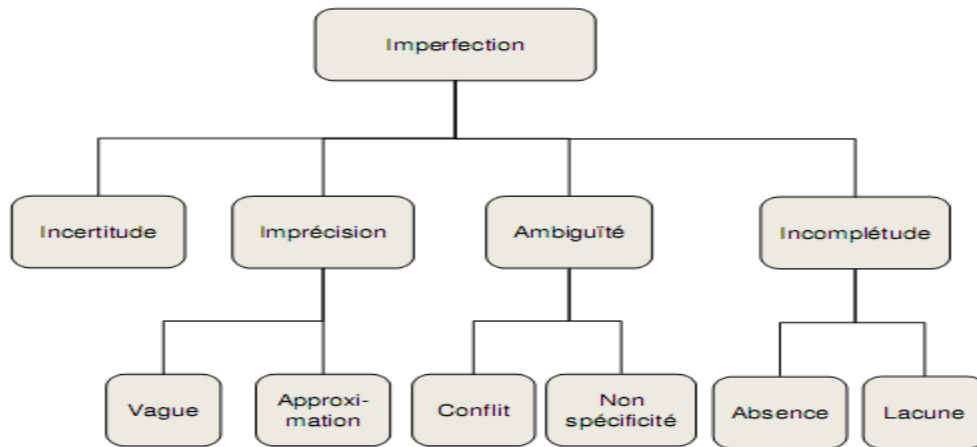


Figure 6 : Typologie de l'imperfection des données géographie selon Fisher [29]

## 2.2. Présentation formelle de l'information géographique par les théories de l'incertain.

Pour modéliser l'information imparfaite, plusieurs approches et méthodes ont été présentées dans la littérature. Ces approches présentent les différents formalismes pour formaliser l'information imparfaite.

Fisher, propose de relier chaque type d'imperfection au choix d'une théorie de représentation pour les informations géographiques [29], il propose pour toute topologie de [30], dans laquelle il a proposé pour chaque taxonomie une théorie de représentation.

L'objectif de cette section est de présenter une panoplie des théories mathématiques, les théories de l'incertain sont appelées aussi les théories de l'imparfait.

L'avènement et l'émergence de l'intelligence artificielle, et la théorie de la décision font apparaitre les théories de présentation de l'information imparfaite (Figure 7).

Ces deux grands domaines ont abordé la modélisation (la formalisation) de l'imparfait de deux façons différentes. L'intelligence artificielle cherche à formaliser les modes de raisonnement basé sur les approches *qualitatives*. Par contre, la théorie de la décision cherche à formaliser l'observation des choix individuels basés sur les approches *probabilistes*.

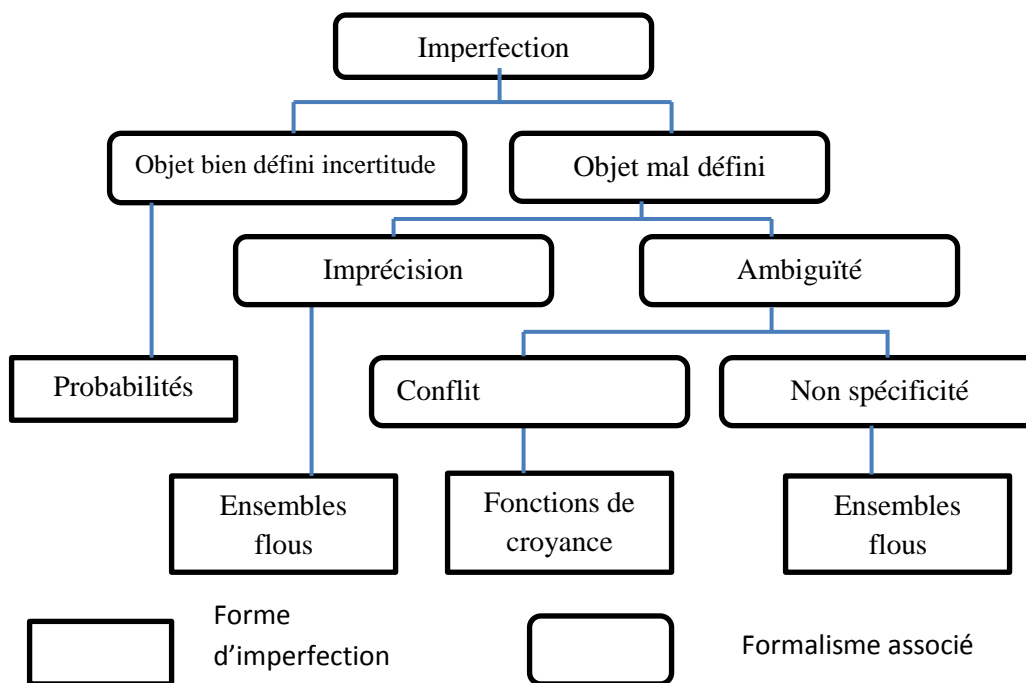


Figure 7 : Taxonomie de l'imperfection de l'information géographique adaptée au Fisher [28]

Au vu de nombreuses formes d'informations imparfaites, diverses théories ont été développées, généralement :

- La théorie des probabilités : considérée comme une théorie de représentation de *l'incertitude*.
- La théorie de la logique floue : s'intéresse à la représentation de *l'imprécis* ou des *vagues*.
- La théorie de possibilités : est une approche pour formaliser *l'imprécision et l'incertitude*.
- La théorie de l'évidence (croyance), ainsi s'intéresse à la formalisation de *l'imprécision et l'incertitude*.
- L'approche mixte (possibilité, évidence) pour formaliser *l'imprécision et l'incertitude*.

Dans le cadre de notre travail, nous avons tenté d'intégrer deux approches (évidence, logique floue), afin de formaliser un certain type d'imperfection (l'incertain et l'imprécis).

### 2.3. Le processus d'appariement des données géographiques

L'appariement de données géographiques est une technique complexe consistant à identifier, dans des jeux de données différents, les entités représentant les mêmes phénomènes du monde réel [8].

Donc, l'appariement est un processus qui définit les liens entre les objets de différentes bases de données qui représentent le même phénomène réel.

Le processus a été décomposé en deux niveaux :

- un niveau global constitué par l'analyse préalable qui détermine la stratégie globale que le processus doit adopter pour appairer les bases de données géographiques. Il est réalisé pour exprimer le contexte et le but de cet appariement sous forme de connaissances et utilise une base de connaissances,
- un niveau local qui réalise effectivement les liens d'appariement et fait appel à de nombreux algorithmes.

Plusieurs outils pouvant être utilisés pour appairer des objets géographiques, ont été présentés. Ils s'appuient sur :

- La comparaison de la sémantique des objets
- L'utilisation de distances entre leur géométrie
- La ressemblance de leur forme
- Les correspondances entre leurs relations topologiques

Cependant aucun outil pris individuellement n'est suffisant. Devogele décrit un module générique d'appariement qui regroupe les étapes communes aux processus d'appariement. Ce squelette permettra de construire des processus d'appariement complexes sans développer uniquement sur les choix des outils d'appariement pour appairer ses données.

Le processus générique d'appariement s'appuie sur une boîte à outils ; composée de 6 étapes :

- L'enrichissement
- La sélection des objets candidats à l'appariement
- Le calcul des mesures d'appariement
- Le filtrage ou la prolongation
- Le regroupement d'appariement
- L'analyse du résultat

Thomas Devogele [22]. distingue trois types d'appariement de données géographiques possible:

- ***L'appariement géométrique*** qui se base sur des notions de distance, de ressemblances et de forme pour comparer la géométrie des objets et retrouver les correspondants.

- **L'appariement sémantique** qui consiste à mettre en correspondance les objets grâce à la valeur de leurs attributs sémantiques. Cette technique est dérivée des bases de données classiques, et basée sur les relations entre les concepts, aussi bien au niveau des données attributaires, qu'au niveau toponymique.
- **L'appariement topologique** qui utilise les relations de composition et les relations topologiques pour obtenir un ensemble d'objets candidats et limiter les recherches géométriques. C'est un complément des deux autres et ne peut être utilisé seul.

L'imprécision qui entache les données géographiques est une des principales sources d'erreurs que l'on peut rencontrer lorsqu'on utilise un processus d'appariement. En effet, il arrive que le processus détecte un lien entre deux objets qui ne correspondent pas en réalité au même phénomène dans le monde réel. Mais la principale limite du processus d'appariement reste la difficulté à le rendre automatisable.

[34] pense qu'un appariement complètement automatique est certainement impossible car certains cas conduisent souvent à des résultats erronés (traitement de nombreux carrefours proches ou d'échangeurs complexes par exemple). En effet dans l'idéal, il faudrait définir un processus informatique capable de retranscrire l'appariement visuel de l'homme (notion intuitive de ressemblance) tout en évitant les conflits. Il faut par conséquent concevoir des outils pour évaluer et corriger interactivement les appariements réalisés automatiquement.

*Thierry Badard* [11.21], a défini un processus d'appariement des données géographiques ayant des actualités différentes. Ce processus a été conçu avec un comportement ascendant et est quasi-automatique. Il procède d'abord à l'appariement des primitives géométriques des objets, puis des objets simples pour terminer par les objets complexes (groupes d'objets). Il interprète ensuite les résultats afin d'établir une typologie des évolutions spatio-temporelles, qui permet de traduire les évolutions subies entre deux actualités d'une même base de données. L'information obtenue est détaillée et proche des évolutions intervenant dans les bases de données géographiques, ce qui facilite ensuite l'intégration des mises à jour dans les bases de données de l'utilisateur.

Parmi les difficultés d'implémentation du processus d'appariement qui sont cités dans [22], souligne cependant que les processus d'appariement sont complexes car ils utilisent plusieurs outils pour comparer un grand nombre d'informations. Leur conception est donc difficile mais peut être facilitée par la spécification d'un processus générique et par une interaction avec l'utilisateur dans le choix et le paramétrage des outils.

L'appariement géométrique reste celui le plus souvent utilisé pour les données géographiques, mais il est également le plus complexe à mettre en œuvre du fait de plusieurs facteurs :

- Il est difficile de détecter que deux objets occupent la même position dans l'espace sachant que chaque point de l'objet est entaché d'erreurs de localisation plus au moins aléatoires.
- Etant donné qu'un objet d'un jeu de données peut être apparié avec plusieurs objets de l'autre jeu de données ou même aucun objet, la notion d'objet le plus proche n'est pas suffisante.
- L'appariement doit tenir compte de l'information contextuelle, en analysant la répartition des objets les uns par rapport aux autres dans l'espace géographique.

### 3. Conclusion

Au vu des énormes potentialités offertes par les SIG en termes de collecte, d'analyse et de diffusion des données, L'utilisation des SIG est indispensable à la mise en œuvre efficace de programmes tels que les recensements de la population et de l'habitat, les recensements agricoles ainsi que les autres enquêtes telles que les enquêtes sur la santé et la démographie.

De nos jours, Les systèmes d'information géographique (SIGs) sont utilisés dans plusieurs domaines incluant celui de la production et de l'utilisation des données géographiques, l'évaluation de la qualité des données géographiques, l'intégration des données de diverses sources et la visualisation en relations causales.

Ces connaissances sont sujettes à différentes formes d'imperfection. Si cette imperfection est minorée dans la représentation des données, la validité des résultats des processus d'appariement des données géographiques et de mise en relation des objets géographiques peut être remise en question. Ainsi, l'imperfection de l'information impacte sur la qualité des données et des décisions.

Parmi les avantages offerts par les SIGs, contribuent à l'amélioration de la qualité des données géographiques, de la qualité de l'analyse et la compréhension des ensembles des données et d'autres phénomènes complexes.

La prise de décision est ainsi rendue plus facile et plus précise utilisant le processus de l'appariement des données géographiques ou d'autres processus d'analyse des données géographiques.

---



## Chapitre II: La théorie de l'évidence

## II. La théorie de l'évidence

---

### 1. Introduction :

Un système d'information que ce soit un système d'information de gestion ou d'information géographique est une abstraction du monde réel. L'utilisation de ce genre de systèmes est leurs portés sur la précision et la pertinence de cette représentation, mais nos connaissances et notre vue au monde réel sont plus souvent subjectives et non objectives.

Nos connaissances sont imparfaites, ainsi que de nombreuses théories, telles que la théorie des probabilités, la théorie des ensembles flous [14.51], la théorie des possibilités [25.52], la théorie des fonctions de croyance [20.43]. Elle possèdent des outils pour modéliser, présenter et manipuler les imperfections.

L'enjeu est de représenter explicitement les connaissances que nous avons sur des phénomènes réels et de raisonner sur ces connaissances imparfaites afin de les utiliser pour prendre une décision finale. D'une manière générale, certaines connaissances se présentent sous la forme de règles tandis que d'autres sont parfois difficiles à modéliser explicitement (difficile à formaliser mathématiquement).

Par conséquent le besoin d'exprimer le doute ou l'ignorance nous semble nécessaire dans le processus d'appariement.

Les phénomènes naturels sont souvent entachés par des connaissances incertaines, vagues, imprécises, et plutôt incomplètes à cause des instruments de mesure. Le degré de croyance sur ces phénomènes reste toujours difficile, et soulève des interrogations qui sont résolues de nos jours par des modèles mathématiques de croyances. Parmi eux, nous pouvons citer :

- Le modèle Bayesian (modèle de bayes)
- Modèle de probabilités inférieures et supérieures (Model Upper and Lower Probabilities UPL) de Dempster
- Modèle de valeur d'évidence (Evidentiary Value Model EVM) de Dempster,
- La Probabilité des Propositions modales et autre.

De nos jours l'utilisation d'un modèle ou d'un autre dépend de plusieurs domaines d'application et suivant un mode d'application.

Ainsi pour une distribution de probabilité avec valeurs connues on utilisera le Bayésien. Pour distribution avec quelques valeurs connues, on utilisera l'ULP. Pour une distribution qui n'est pas connue, on utilise le TBM (modèle des croyances transférables (Transferable Belief Model)).

Cette théorie, basée sur les travaux de Dempster (1966-1968), formalisée, et complétée par les travaux de Shafer en 1974, énormément contribué au développement de cette théorie grâce à son modèle des croyances transférables TBM, dérivée de celui de Shafer, a donné naissance à un modèle mathématique de croyance connue sur le nom de la théorie de l'évidence. Cette théorie est un modèle d'inférence statistique qui généralise l'inférence Bayésienne et qui connaît des applications diverses telles que:

- Modélisation des incertitudes dans les systèmes experts.
- Fusion d'informations (télé-détection, identification de cibles, imagerie, médicale, ...).

Cette théorie permet de représenter explicitement, à partir d'outils mathématiques, l'incertitude liée aux connaissances (tels que l'avis des experts sur le degré de plausibilité ou de croyance que la tempête (l'orage) sera dans 15 jours). Par cette méthode, on peut modéliser le degré de croyance, notion « de croyance » qui s'avère particulièrement efficace lors de la combinaison de différents points de vue. Connue pour son efficacité dans de multiples applications telles que la classification d'images, prise de décision, nous donnons comme objectif de présenter la théorie de *Dempster-Shafer* à travers ses principaux termes fondamentaux, afin de permettre d'en saisir le sens.

## 2. Les fondamentaux de la Théorie de Dempster-Shafer

Dans cette théorie probabiliste introduite par Dempster, à travers ses fonctions de probabilité inférieure et probabilité supérieure auxquelles Shafer attribue respectivement la fonction de croyance et la fonction de plausibilité, l'on prend en compte l'extension de la théorie des probabilités subjectives. Les applications ne concernent que des ensembles de définition discrets au niveau crédible et pignistique. Cette théorie est caractérisée par les fondements suivants

### 2.1. L'univers (cadre de discernement)

La théorie des évidences décrit un « cadre de discernement »  $\Omega$  « frame of discernement » ou ensemble « d'hypothèses » élémentaires envisageables pour la solution du problème posé [23]. Ce cadre contient toutes les hypothèses ( $H_i$ ) nécessaires à la description complète d'un problème posé :

$$\Omega = \{H_1, H_2, H_3, \dots, H_N\}$$



Le nombre  $N$  indique le nombre total d'hypothèses ;  $\Omega$  doit être exhaustif et l'ensemble des hypothèses qui le composent exclusives. Dans la théorie de Dempster-Shafer, le raisonnement porte sur l'ensemble des parties  $2^\Omega$  appelée également « power-set » ou « l'ensemble des partitions », qui est l'ensemble de tous les sous-ensembles de  $\Omega$ . Tous les éléments de ce « power-set » correspondent à une proposition disjonctive (union de « singletons ») du cadre de discernement.

Le nombre de sous-ensembles est dépendant du nombre d'hypothèses contenues dans  $\Omega$ . Pour trois hypothèses  $A, B$  et  $C \subseteq \Omega$ , on obtient ainsi sept sous-ensembles:

$$2^\Omega = \{\{A\}, \{B\}, \{C\}, \{A,B\}, \{A,C\}, \{B,C\}, \{A,B,C\}, \emptyset\}$$

L'union de deux hypothèses est considérée par Shafer comme une « proposition » exprimant en cela l'incertitude.  $A, B$  sont nommés « singletons ». On obtient pour cet exemple :

$$2^\Omega = \{A, B, C, A \cup B, A \cup C, B \cup C, A \cup B \cup C, \emptyset\}$$

La théorie de Dempster-Shafer comprend trois types d'opérations [41]:

- la définition des fonctions de masses, de crédibilité et de plausibilité.
- la fusion des informations par la règle de Dempster
- la prise de décision.

Soit  $\Omega$  est l'ensemble des propositions, c'est-à-dire l'espace du monde possible de cardinal fini (un référentiel fini), elle associe des degrés de croyance et de plausibilité à des parties de  $A$  (groupe de propositions ou d'informations) de  $\Omega$ .

## 2.2. La fonction de masse :

La théorie TDS, basée sur le pouvoir d'associer à un élément quelconque une probabilité inférieure et une probabilité supérieure qui constitue en principe la généralisation de la théorie de probabilité bayésienne.

L'idée de base de cette théorie est qu'à partir de l'ensemble des classes  $\Omega$ , appelé cadre de discernement. On définit sur leur ensemble des parties une fonction  $m$ , dite fonction de masse, ayant ses valeurs dans  $[0,1]$ , et vérifiant les conditions suivantes :

$$\begin{cases} m(\emptyset) = 0 \\ \sum_{A \subseteq \Omega} m(A) = 1 \end{cases}$$

- La quantité  $m(A)$  s'interprète comme la part de croyance placée strictement sur  $A$ .
- $m(A)$  se différencie d'une probabilité par le fait que la totalité de la croyance est répartie non seulement sur les hypothèses Singletons mais sur tous les ensembles des hypothèses.

### 2.3. La fonction de croyance et la fonction de plausibilité

La fonction de croyance ou de plausibilité (appelée aussi fonction de masse) est une fonction qui, pour une information ou proposition émise ou déduite d'une analyse, permet d'attribuer des degrés de croyance ou de plausibilité comprise entre 0 et 1.

Soit  $\Omega$  un ensemble d'hypothèses  $\{H_1, H_2, \dots, H_n\}$  mutuellement exclusives, appelé cadre de discernement. L'ensemble des parties  $A$  de  $\Omega$  est noté  $2^\Omega$ .

La construction de la fonction de croyance dépend de l'application concernée [24].

Formellement :

$$\begin{cases} m : 2^\Omega \rightarrow [0, 1] \\ A \rightarrow m(A) \end{cases}$$

Si on définit la fonction  $m : 2^\Omega \rightarrow [0, 1]$  une distribution d'évidence sur les parties de  $\Omega$  (ou la partition de  $\Omega$ ), alors  $A \rightarrow m(A)$  représente la distribution d'évidence sur la partie  $A$  (représente la masse d'évidence associée à une source d'information sur un certain nombre d'hypothèses ou de propositions du cadre de discernement.).

- $m(A)$  est la confiance portée strictement dans  $A$  sans que celle-ci puisse être répartie sur les hypothèses qui la composent.
- $A$  est un élément focal si  $m(A) \neq 0$ , (c.-à-d.  $m(A) > 0$  ).

Dans le cas où pour  $m : 2^\Omega \rightarrow [0, 1]$  les deux conditions suivantes sont remplies :  $m(\emptyset) = 0$  et

$$\sum_{A_i \subset \Omega} m(A_i) = 1, \text{ elle est appelée fonction de masse sur } \Omega.$$

Par ailleurs, il faut noter que si l'on considère deux sources d'information ou évidences  $S_1$  et  $S_2$ , contiennent respectivement  $H_1$ ,  $H_2$  comme hypothèses.  $m(H_1)$  et  $m(H_2)$  peuvent être toutes les deux nulles sans pour autant que l'on obtienne  $m(H_1 \cup H_2) = 0$ . Par conséquent, un jeu de masse quelconque  $m(\cdot)$  s'avère différent d'une mesure de probabilité.

### 2.3.1. La fonction de croyance $Bel(A)$ :

Elle est définie par la somme des probabilités de tous les éléments focaux qui impliquent forcément la proposition d'une solution au problème posé.

La fonction de croyance ( $Cr$ ) (ou crédibilité), « Belief » est dérivée des fonctions de masse. Ainsi,  $Cr(A)$  totalise toutes les raisons de soutenir l'hypothèse  $A$ . La mesure de la croyance ( $Cr$ ) s'effectue sur le cadre de discernement  $\Omega$ , mais il est possible de la représenter par rapport à une fonction de masse comme suit :

$$Cr(A) = Bel(A) = \sum_{B_i \subset A} m(B_i)$$

Pour calculer la croyance de l'hypothèse  $A$ , il faut donc effectuer la somme de toutes les propositions  $B$  qui sont strictement incluses dans  $A$ . La fonction de croyance peut donc être interprétée comme la borne inférieure d'une mesure de vraisemblance [17].

Une fonction de croyance nulle représente ainsi l'ignorance totale sur le problème considéré. Dans le cas d'une fonction de crédibilité particulière où :

$$\begin{cases} Cr(\Omega) = 1 \\ Cr(A \cup B) = Cr(A) + Cr(B) \end{cases} \text{ quand } A, B \subset \Omega \text{ et } A \cap B = \emptyset$$

On a alors affaire à une fonction de crédibilité bayésienne qui correspond à une mesure classique de probabilité. C'est à ce titre que la TDS est généralement considérée comme une extension de la théorie des probabilités.

### 2.3.2. La fonction de plausibilité $Pl(A)$ :

Elle est la somme des probabilités de tous les éléments focaux qui ne contredisent pas nécessairement la proposition d'une solution au problème posé.

Donc la plausibilité que la vérité soit dans  $A$  est la somme de masses des propositions dont l'intersection avec  $A$  n'est pas nulle.

Elle est définie formellement par :

$$Pl(A) = \sum_{A \cap B_i \neq \emptyset} m(B_i)$$

La plausibilité (Pl) est également dérivée des fonctions de masse. La fonction de plausibilité Pl (A) exprime l'intensité avec laquelle on devrait soutenir l'hypothèse A si tous les faits non connus allaient dans le sens de cette hypothèse. Elle peut être définie de deux manières différentes [16] :

- Soit par la fonction de croyance de l'évènement contraire à A, notée  $A^c$  :

$$Pl = 1 - Cr(A^c)$$

- Soit par rapport à la fonction de masse qui correspond à la définition originale de Dempster :

$$Pl(A) = \sum_{A \cap B_i \neq \emptyset} m(B_i)$$

Ainsi, la plausibilité représente le niveau de vraisemblance maximum d'une hypothèse donnée.

D'après Dempster, le couple (Cr, Pl), peut interpréter comme la mesure de probabilité respectivement basse et haute.

Cr(A) et Pl(A) ont une valeur comprise dans l'intervalle [0,1]. A partir de ces deux valeurs, on peut déterminer une probabilité inconnue nommée « intervalle de croyance (Ic) » ou « intervalle de confiance » qui correspond à :

$$Cr(A) \leq Ic(A) \leq Pl(A)$$

#### 2.4. Initialisation des masses :

La principale difficulté dans la théorie des fonctions de croyance (TDS) réside dans l'étape d'initialisation des masses de croyance. Elle consiste à déterminer les masses pour chaque hypothèse supposée comme une solution pour un problème donné afin de prendre une décision finale pertinente et optimale. Par contre, la mauvaise initialisation guide à une mauvaise décision ou à un conflit.

Plusieurs méthodes peuvent être utilisées pour choisir les masses, comme celle des fonctions de vraisemblance construites à partir d'un ensemble de données d'apprentissage [9.45.19] utilise les gaussiennes, [21] utilise les mesures de distance connues sous le nom d'algorithme des K plus proches voisins, d'où des grandeurs physiques ont été choisies pour l'initialisation des masses, telle que la distance (euclidienne, toponymique, sémantique,...), l'angle, les représentations floues [26.38.39]. L'initialisation de jeu de masses nécessite l'expérience d'un expert pour le domaine modélisé et pour lequel sont définies les fonctions de croyance qui sont basées sur une base de connaissance.

## 2.5. Combinaison d'information avec la théorie de l'évidence

Une fois le jeu de masse défini pour chacun des sources d'information, utilisant l'opérateur de conjonction de Dempster (opérateur de combinaison de Dempster).

Le principe de la règle de combinaison est relativement simple. Dans ses travaux, G. Shafer (1976) a proposé la règle de combinaison de Dempster, symboliser par  $\oplus$ , pour combiner ou fusionner deux sources d'informations distinctes. Ces dernières sont couramment appelées « sources d'évidence » ( $A_1$  et  $A_2$ ). Présumées équitablement fiables, elles expriment des degrés de confiance sur les propositions d'un cadre de discernement. Pour reprendre une notation courante, on notera  $m_1(\cdot)$  et  $m_2(\cdot)$ ,  $Cr_1$  et  $Cr_2$ , les jeux de masses et les fonctions de croyance relatives à  $A_1$  et  $A_2$ .

La fonction de croyance globale est obtenue, dans le cadre de la TDS, à partir de la combinaison des jeux de masse  $m_1(\cdot)$  et  $m_2(\cdot)$ . Elle s'exprime de la façon suivante :

Soient  $A_1, A_2$  des parties de  $\Omega$  soutenues respectivement par les sources  $S_1, S_2$ , la masse associée à leur jonction est  $A = A_1 \cap A_2$  est  $m(A) = m_1(A_1) \times m_2(A_2)$ , cependant, si  $A_1 \cap A_2 = \emptyset$ , ( $A_1, A_2$  sont disjointes), alors la masse associée à leurs intersection est vide  $m(A_1 \cap A_2) = 0$  et cela peut être normaliser par :

$$m_1 \oplus m_2(A) = \begin{cases} 0 & \text{si } A = \emptyset \\ \frac{1}{1 - \lambda_C} \sum_{A_1 \cap A_2 = A} m_1(A_1) \times m_2(A_2) & \text{si } A \neq \emptyset \end{cases}$$

Où  $\lambda_C = \sum_{A_1 \cap A_2 = \emptyset} m_1(A_1) \times m_2(A_2)$  représente le conflit entre les sources, appelé aussi l'ignorance (utilisé comme facteur de normalisation).

Cette règle de combinaison a été critiquée dans plusieurs travaux dont [50].

Il y a d'autres opérateurs de combinaison comme celle de [44], « opérateurs non normalisés », opérateurs disjonctifs [25], généralisation d'opérateur.

La règle de fusion de Dempster-Shafer peut être illustrée par un exemple appliqué à l'appariement des objets géographique de deux BDG :

L'expert sélectionne des deux objets **A**, **B** comme des candidats considérés comme homologues potentiels de l'objet **O**.

Soit le cadre de discernement  $\Omega$  :

$\Omega = \{A \text{ homologue de } O (\mathbf{H}_{A,O}), B \text{ homologue de } O (\mathbf{H}_{B,O}), (A \cup B) \text{ homologue de } O (\mathbf{H}_{A \cup B,O})\}$ , c'est-à-dire  $\Omega = \{(\mathbf{H}_{A,O}), (\mathbf{H}_{B,O}), (\mathbf{H}_{A \cup B,O})\}$ , décrivant les hypothèses supposés comme solutions potentielles de l'appariement des données (objets géographique).

Pour chaque hypothèse, deux experts donnent les mesures (émettent l'avis)  $m_1$  et  $m_2$  suivant :

La source  $S_1$ , est la distance euclidienne :

$$\begin{cases} m_1(H_{A,O}) = 0,85 \\ m_1(H_{B,O}) = 0,05 \\ m_1(H_{A \cup B,O}) = 0,10 \end{cases}$$

La source  $S_2$ , est la distance toponymique (la distance entre les chaînes de caractères) :

$$\begin{cases} m_2(H_{A,O}) = 0,70 \\ m_2(H_{B,O}) = 0,02 \\ m_2(H_{A \cup B,O}) = 0,28 \end{cases}$$

Dans ce cas, l'application de la règle de fusion de TDS pour l'hypothèse «A homologue O ( $\mathbf{H}_{X,O}$ ) » est la suivante :

$$\begin{aligned} m(\mathbf{H}_{A,O}) &= \frac{m_1(\mathbf{H}_{A,O}) \times m_2(\mathbf{H}_{A,O}) + m_2(\mathbf{H}_{A,O}) \times m_1(\mathbf{H}_{A \cup B,O}) + m_1(\mathbf{H}_{A,O}) \times m_2(\mathbf{H}_{A \cup B,O})}{1 - [m_1(\mathbf{H}_{A,O}) \times m_2(\mathbf{H}_{B,O}) + m_1(\mathbf{H}_{B,O}) \times m_2(\mathbf{H}_{A,O})]} \\ &= \frac{0,595 + 0,07 + 0,238}{1 - [0,017 + 0,035]} \\ m(\mathbf{H}_{A,O}) &= 0,9525 \end{aligned}$$

$$\begin{aligned} m(\mathbf{H}_{B,O}) &= \frac{m_1(\mathbf{H}_{B,O}) \times m_2(\mathbf{H}_{B,O}) + m_2(\mathbf{H}_{B,O}) \times m_1(\mathbf{H}_{A \cup B,O}) + m_1(\mathbf{H}_{B,O}) \times m_2(\mathbf{H}_{A \cup B,O})}{1 - [m_1(\mathbf{H}_{A,O}) \times m_2(\mathbf{H}_{B,O}) + m_1(\mathbf{H}_{B,O}) \times m_2(\mathbf{H}_{A,O})]} \\ &= \frac{0,001 + 0,002 + 0,014}{1 - [0,017 + 0,035]} \end{aligned}$$

$$m(\mathbf{H}_{B,O}) = 0,0179$$

$$m(\mathbf{H}_{A \cup B,O}) = \frac{m_1(\mathbf{H}_{A \cup B,O}) \times m_2(\mathbf{H}_{A \cup B,O})}{1 - [m_1(\mathbf{H}_{A,O}) \times m_2(\mathbf{H}_{B,O}) + m_1(\mathbf{H}_{B,O}) \times m_2(\mathbf{H}_{A,O})]}$$

$$= \frac{0,028}{1 - [0,017 + 0,035]}$$

$$m(\mathbf{H}_{A \cup B,O}) = 0,0295$$

$$\text{Donc } \begin{cases} m(\mathbf{H}_{A,O}) = 0,9525 \\ m(\mathbf{H}_{B,O}) = 0,0179 \\ m(\mathbf{H}_{A \cup B,O}) = 0,0295 \end{cases}$$

Vérifiant la condition:  $m(\mathbf{H}_{A,O}) + m(\mathbf{H}_{B,O}) + m(\mathbf{H}_{A \cup B,O}) = 1$ , c'est-à-dire normalisés, Dans ce cas, le niveau de conflit entre les sources est très faible puisque  $(1 - \lambda_c = 0,948)$  est très proche de 1. La prise de décision est aisée, la croyance, affectée à l'hypothèse «A homologue O ( $\mathbf{H}_{A,O}$ ) », étant très élevée. On constate à travers cet exemple que la règle de fusion de TDS est commutative et associative, c'est à dire que l'ordre de la combinaison n'a pas d'importance.

Par ailleurs, la définition d'une fonction de masse représentant l'ignorance totale, par exemple  $m(A) = 0$  ne modifierait pas les informations acquises au préalable.

La théorie de Dempster-Shafer permet de combiner des informations issues de sources différentes, car d'après Shafer, si deux fonctions de croyance sont définies sur un même cadre de discernement alors ce cadre de discernement distingue l'interaction pertinente entre deux informations. De ce fait les informations doivent être définies sur un même cadre de discernement de façon que l'on puisse en déduire leur somme orthogonale suivant la règle de combinaison de Dempster. Cette somme est toujours une fonction de croyance et prend en compte l'influence de toutes les autres. Cette règle peut s'énoncer comme suit :

Soit  $m_1$  et  $m_2$  deux jeux de masses associant aux fonctions de croyance  $Cr_1$  et  $Cr_2$  respectivement, sur le même cadre de discernement  $\Omega$ . On note  $Cr = Cr_1 * Cr_2$  la somme orthogonale de  $Cr_1$  et  $Cr_2$ , et  $A_1, A_2, \dots, A_n$  les éléments focaux de  $Cr_1$  ainsi que  $B_1, B_2, \dots, B_n$  ceux de  $Cr_2$ .

De ce fait :

Si  $\sum A_i \cap B_j = \Omega^{m_1(A_i)m_2(B_j)} < 1$  : alors la fonction  $Cr : 2^\Omega \rightarrow [0, 1]$  est définie par :

$$Cr(\emptyset) = 0 \text{ et } Cr(C) = \frac{\sum_{A_i \cap B_j = C} m_1(A_i)m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i)m_2(B_j)}$$

Il est à noter que cette règle de combinaison donne lieu à des propriétés de symétrie, d'associativité, d'élément neutre. Elle permet aussi de combiner des fonctions bayésiennes pour créer d'autres semblables et donne aussi lieu à la règle de conditionnement.

## 2.6. La prise de décision Par TDS

La théorie de fonctions de croyance propose plusieurs critères de décisions, donc la prise de décision dans la TDS est une étape difficile qui a fait et fait toujours l'objet de nombreux débats. De multiples critères de décision ont ainsi été proposés [17. 46.49], contrairement à la théorie bayésienne, où le maximum de vraisemblance est le critère le plus souvent retenu. Parmi les règles de décision retenues se trouvent : le maximum de crédibilité ; le maximum de plausibilité, les règles basées sur l'intervalle de confiance, la probabilité pignistique et la décision par maximum de vraisemblance. La première peut être considérée comme totalement non risquée, mais sous-estime de fait l'appartenance à une classe donnée, dans la mesure où elle ne retient que les cas pour lesquels il n'y a pas d'incertitude. A l'inverse, la seconde peut être considérée comme trop risquée, puisqu'elle retient la limite supérieure de l'intervalle de vraisemblance, incluant tous les cas incertains. Dans la mesure où la réalité se situe quelque part dans l'intervalle [Crédibilité , Plausibilité], il apparaît ainsi préférable d'exclure ces deux règles. Dès lors, la question de la détermination du seuil sur cet intervalle se pose avec acuité. Deux voies peuvent être considérées : une voie supervisée, la décision se fondant empiriquement sur des connaissances à priori ; une voie non supervisée, la décision étant alors prise sur des critères statistiques. La deuxième voie doit être privilégiée pour deux raisons : éviter d'introduire des erreurs de jugement, et rendre la méthode reproductible.

Ainsi, la probabilité pignistique est une mesure de probabilité subjective qui se trouverait entre la mesure de crédibilité et de plausibilité [44]. Elle est exprimée de la façon suivante :

$$\forall A \in 2^\Omega, \quad P(A) = \sum_{A \subseteq B} \frac{|A \cap B|}{|B|} m(B)$$



$|X|$  est le nombre d'hypothèses qui construit la proposition  $X$ , et  $m(B)$  représente la fonction de masse de l'hypothèse (proposition)  $B$ .

L'intérêt d'utilisation de la mesure de la probabilité *pignistique* réside dans le fait qu'elle offre un compromis intéressant entre la croyance et la plausibilité, défini de façon objective, ainsi traite les hypothèses singletons qui correspondent à l'approche de l'appariement de données géographiques, ainsi que notre objectif est d'apparier un objet d'une  $BDG_1$  à un seul et unique objet dans  $BDG_2$ , dans le cas où il y a plusieurs homologues pour un objet, on regroupe les objets homologues obtenus par TDS appliquée.

Afin de clarifier le sujet, on donne l'exemple suivant qui montre comment prendre la décision par Cr, Pl, probabilité pignistique ?

Soit le cadre de discernement  $\Omega$ , concerne le processus de l'appariement des données géographiques est :

- $\Omega = \{A, B, C\}$ .

- Initialisation de fonctions des masses :

$$\left\{ \begin{array}{l} m(\emptyset) = 0 \\ m(A) = 0,30 \\ m(B) = 0,10 \\ m(C) = 0,20 \\ m(A \cup B) = 0,10 \\ m(A \cup C) = 0,20 \\ m(B \cup C) = 0,05 \\ m(A \cup B \cup C) = 0,05 \end{array} \right.$$

Qui vérifient la condition suivante :  $\sum_{i=1..|x|} m_i = 1$ , tel que  $|X|$  représente la cardinalité de  $\Omega$

- la prise de la décision

- selon les fonctions de croyance :

C'est la décision pessimiste, donc après l'étape 1 :

$$\left\{ \begin{array}{l} Cr(A) = \mathbf{0,30} \\ Cr(B) = 0,10 \\ Cr(C) = 0,20 \end{array} \right.$$

- selon les fonctions de plausibilité :

C'est la décision optimiste, donc après l'étape 1 :

$$\left\{ \begin{array}{l} Pl(A) = m(A) + m(A \cup B) + m(A \cup C) + m(A \cup B \cup C) \\ \quad = 0,30 + 0,10 + 0,20 + 0,05 \\ \quad \quad = \mathbf{0,65} \\ Pl(B) = m(B) + m(A \cup B) + m(B \cup C) + m(A \cup B \cup C) \\ \quad = 0,10 + 0,10 + 0,05 + 0,05 \\ \quad \quad = 0,30 \\ Pl(C) = m(C) + m(A \cup C) + m(B \cup C) + m(A \cup B \cup C) \\ \quad = 0,20 + 0,05 + 0,05 + 0,20 \\ \quad \quad = 0,50 \end{array} \right.$$

c. selon la mesure de probabilité Pignistique

$$\left\{ \begin{array}{l} P(A) = m(A) + \frac{1}{2} m(A \cup B) + \frac{1}{2} m(A \cup C) + \frac{1}{3} m(A \cup B \cup C) \\ \quad = 0,30 + 0,1 \times 0,5 + 0,5 \times 0,20 + 0,33 \times 0,05 \\ \quad \quad = \mathbf{0,466} \\ P(B) = m(B) + \frac{1}{2} m(A \cup B) + \frac{1}{2} m(B \cup C) + \frac{1}{3} m(A \cup B \cup C) \\ \quad = 0,10 + 0,5 \times 0,1 + 0,5 \times 0,05 + 0,33 \times 0,05 \\ \quad \quad = 0,191 \\ P(C) = m(C) + \frac{1}{2} m(A \cup C) + \frac{1}{2} m(B \cup C) + \frac{1}{3} m(A \cup B \cup C) \\ \quad = 0,20 + 0,5 \times 0,20 + 0,5 \times 0,05 + 0,33 \times 0,05 \\ \quad \quad = 0,341 \end{array} \right.$$

Donc selon les fonctions de croyance, la décision est A ( $Cr(A)=0,30$ ), selon les fonctions de plausibilité la décision est aussi A ( $Pl(A)=0,65$ ), mais selon la mesure de probabilité pignistique la décision est aussi A ( $P(A)=0,466$ ).

### 3. Conclusion

La théorie de l'évidence, à partir des mathématiques probabilistes, présente un cadre formel de raisonnement dans l'imparfait, un modèle qui permet de modéliser la connaissance. Car à travers les fonctions de croyance qui sont des outils de mesure de la probabilité subjective, on peut évaluer le degré de vérité d'une affirmation, d'un avis d'expert. Avec l'introduction des masses d'évidence, des coefficients d'affaiblissement de ces masses et aux moyens de la règle de combinaison elle permet de traiter l'information au point d'aboutir à sa fiabilité. Ce qui aide grandement dans la prise de décision, ou dans d'autres domaines (tels que la télédétection).

Toutefois, certains analystes et utilisateurs (tels que F. Voorbraak, H. Hyburg) relèvent des difficultés d'application de cette théorie en raison : « de la sensibilité de la méthode de combinaison sur les petites

valeur de masses de croyance, des grands temps de calcul qu'elle demande en comparaison aux autres méthodes, de l'absence de sémantique claire, de la difficulté à suivre une méthode systématique et quelque peu générique ».

Néanmoins d'autres analystes et utilisateurs voient en elle, une théorie qui permet de raisonner dans l'incertain avec des *degrés de croyance* qui sont suffisamment généraux pour englober les approches *probabilistes classiques* et *possibilistes*, de comparer des vues, de fusionner des critères afin de traiter les informations qui peuvent provenir de différentes sources, de différents domaines.

L'utilisation de cette théorie montre plusieurs avantages et inconvénients.

- Parmi les avantages d'utilisation de cette théorie réside dans son caractère commutatif et associatif de la règle de la fusion de Dempster-shafer, qui rend son implémentation relativement aisée.
- Parmi les inconvénients d'utilisation de cette théorie réside dans la complexité de la mise en œuvre de la règle de fusion de la TDS qui varie exponentiellement avec la cardinalité du cadre de discernement  $\Omega$ . De plus, l'indépendance des sources, prérequis nécessaires dans la règle de fusion de la TDS.

La théorie de l'évidence a été utilisée pour de multiples applications, comme pour fusionner des données de différentes sources, incluant les images de télédétection pour certaines applications...

Tenant compte de toutes les informations, il est conseillé à un utilisateur d'analyser de façon judicieuse la portée de cette théorie dans le domaine d'application dont il veut en faire usage.

---



Chapitre III:  
La théorie de la logique floue

### III. La Théorie de la Logique floue

---

#### 1. Introduction

Dans le fonctionnement de l'esprit humain, les incertitudes sont particulièrement remarquables, par exemple dans les fonctions de reconnaissance et de raisonnement. La capacité d'établir des classes d'éléments de la nature ayant des propriétés analogues est très naturelle chez l'homme. Il sait reconnaître un chien, déterminer l'âge approximatif d'un individu en l'observant. Il sait aussi rendre compte de données vagues « large ouverture », imprécises « de 3 à 5 m », soumises à des erreurs « 100 M à 3% près », mal définies « massif altéré », dont la validité n'est pas absolue « dans 85% des cas », soumises à une certitude « très probable ». Il est tout aussi naturel à l'homme de traiter des données affectées d'incertitude, inhérentes à l'univers ou dues à sa méconnaissance de certains facteurs, que d'utiliser des critères subjectifs, donc imprécis [13].

De nos jours, la logique floue est un axe de recherche important sur lequel se focalise de nombreuses recherches. Des applications de grand public sont déjà disponibles (appareils photos, machines à laver, fours à micro-onde, ...etc.), mais aussi industrielles (commande de processus complexes liés à l'énergie, à la transformation de la matière, à la robotique, etc...).

Les problèmes du monde géographique doivent tenir compte des imprécisions, la logique floue offre cette opportunité, car selon *Zadeh* père de cette discipline, un élément peut plus au moins appartenir à un ensemble donnée (ie : avec un certain degré d'appartenance).

Les imprécisions et les incertitudes peuvent ainsi être modélisées, et le raisonnement acquiert une flexibilité que la logique classique (basé sur l'algèbre de Boole) ne permettait pas.

La logique floue s'applique désormais dans plusieurs domaines : automatisation, instrumentation, traitement d'information soit géographique ou de gestion ...etc, pour extraire des connaissances (prendre des décisions) pertinentes. On ne peut pas aborder la logique floue sans aborder les notions des sous-ensembles flous.

#### 2. Les Sous-ensembles flous

Cette théorie est une extension de la théorie des ensembles classiques pour la prise en compte des ensembles définis d'une manière imprécise. Elle introduit la notion des fonctions d'appartenance pour déterminer le degré d'appartenance d'un élément à un ensemble donné. Cette théorie se réduit à la

théorie des sous-ensembles classiques dans le cas où les fonctions d'appartenance considérées prennent des valeurs 0 ou 1.

La théorie des ensembles flous permet de modéliser les connaissances imprécises, vagues, et traiter des connaissances fournies numériquement et d'autres symboliquement.

### 2.1. Définition d'un sous-ensemble flou

Dans la théorie des ensembles classiques, un élément peut prendre deux situations : soit appartenir ou ne pas appartenir à un sous-ensemble flou. Le professeur L. Zadeh a été tenté de sortir de cette logique classique (Booléenne) en introduisant la notion d'appartenance pondérée, c'est-à-dire introduire des graduations dans l'appartenance d'un élément à un sous-ensemble.

Formellement :

Soit  $X$  un ensemble de référence et soit  $x$  un élément quelconque de  $X$ , un sous-ensemble flou  $A$  de  $X$  est défini comme l'ensemble des couples :

$$A = \{(x, \mu_A(x)), x \in X\}, \text{ avec } \mu : X \longrightarrow [0, 1]$$

### 2.2. Caractéristiques d'un sous-ensemble flou

Un sous-ensemble flou est caractérisé par :

- **Fonction d'appartenance :**

Le sous-ensemble flou  $A$  de  $X$  est caractérisé par une fonction d'appartenance  $\mu_A(x)$ , qui est associée à chaque élément  $x$  de  $X$  un réel dans l'intervalle  $[0, 1]$ ,  $\mu_A(x)$  représente le degré d'appartenance de  $x$  à  $A$ .

Formellement on observe qu'il y a trois cas possibles :

$$\left[ \begin{array}{l} \mu_A(x) = 0 \\ 0 < \mu_A(x) < 1 \\ \mu_A(x) = 1 \end{array} \right.$$

Où  $\mu_A(x) = 0$  signifie  $x$  n'appartient pas à  $A$  et  $\mu_A(x) = 1$  signifie que  $x$  appartient complètement ou entièrement à  $A$ ,  $0 < \mu_A(x) < 1$  signifie que  $x$  appartient partiellement à  $A$ .

On peut remarquer que si  $A$  est un sous ensemble classique, la fonction d'appartenance qui lui est associée ne prend que les valeurs entre 0 et 1

On a donc :

$$\begin{cases} \mu_A(x) = 0 \\ \mu_A(x) = 1 \end{cases}$$

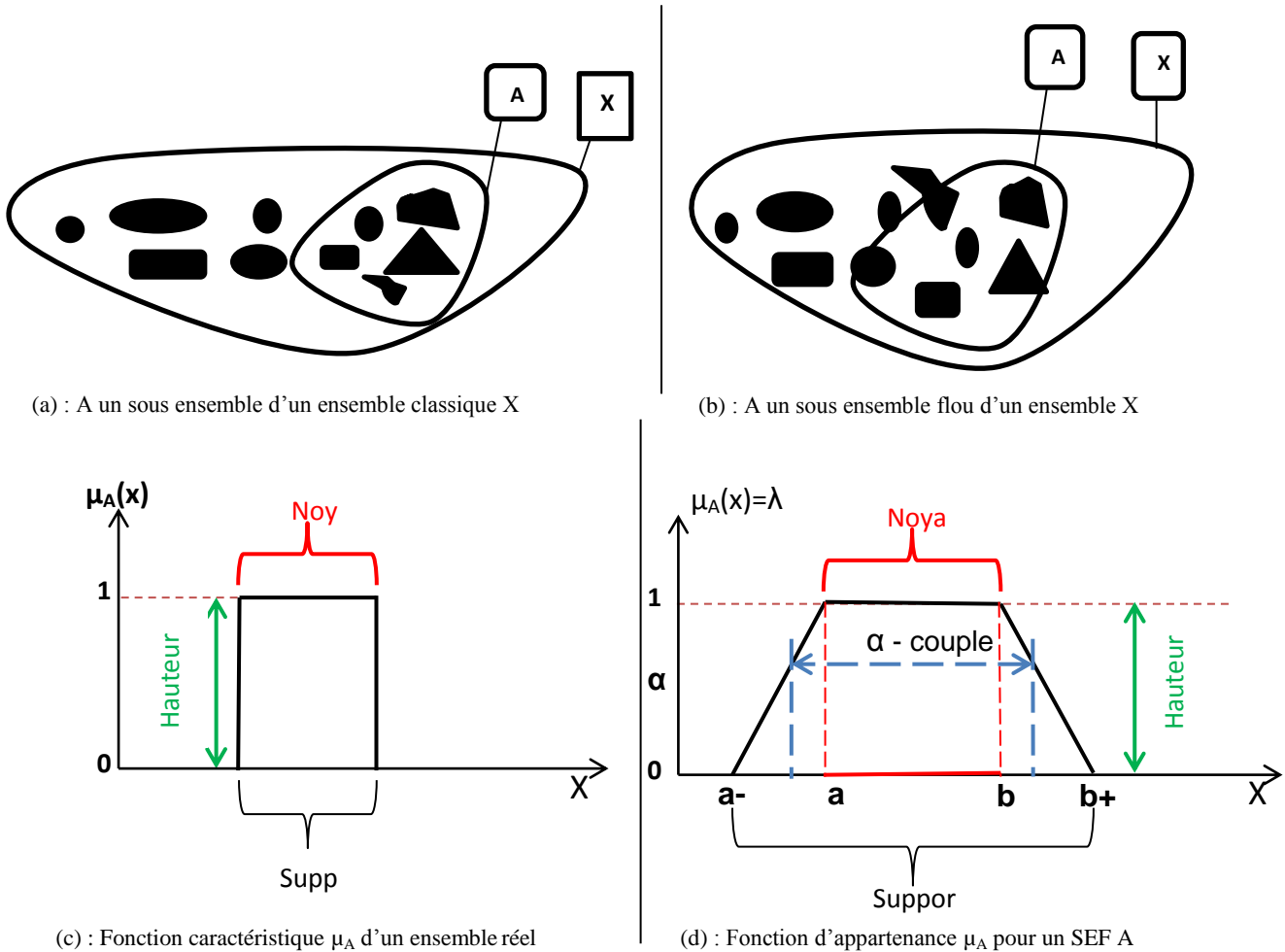


Figure 8 : Fondements de la logique floue

• **Le support et La hauteur :**

Ces deux caractéristiques, pour l'essentiel montrent dans quelle mesure un sous-ensemble flou  $A$  de  $X$ , diffère d'un sous-ensemble classique de  $x$ , la première est *le Support* et la deuxième *la hauteur*, comme le montre la *Figure 8.(d)*.

Formellement :  $Supp(A) = \{x \in X / \mu_A(x) > 0\}$

C'est à dire l'ensemble de tous les éléments qui appartiennent au moins un petit peu (partiellement).

$$H(A) = Sup_{x \in X} \mu_A(x)$$

C'est le plus fort degré avec lequel un élément de  $x$  appartient à  $A$ .

- **Le Noyau :**

Est l'ensemble des éléments dont la hauteur ( $h(A) = 1$ ), c'est-à-dire un ensemble normalisé, simplement de tous les éléments qui lui appartiennent totalement (entièrement).

Formellement :  $Noy(A) = \{x \in X / \mu_A(x) = 1\}$

- **Cardinalité :**

Soit  $A$  est un sous-ensemble flou de  $X$ , la cardinalité de  $A$ , noté  $|A|$  ou bien  $Card(A)$ , est le nombre d'éléments appartenent à  $A$ , pondéré par leur degré d'appartenance.

Formellement :  $|A| = \sum_{x \in X} \mu_A(x)$

- **$\alpha$ -coupes**

Une partie floue  $A$  de  $E$  peut aussi être caractérisée par l'ensemble de ses  $\alpha$ -coupes. Une  $\alpha$ -coupe d'une partie floue  $A$  est le sous-ensemble net (classique) des éléments ayant un degré d'appartenance supérieur ou égal à  $\alpha$ , comme le montre la figure 8 (d).

Formellement :  $\alpha\text{-coupe}(A) = \{x \in E / \mu_A(x) \geq \alpha\}$

### 3. Fondamentaux de la Logique floue

#### 3.1. Introduction

Une approche fut développée à partir 1965 par le professeur L.A.Zedeh, basée sur la notion des ensembles flous, un élément peut plus ou moins appartenir à un certain ensemble. Les imprécisions et les incertitudes peuvent ainsi être modélisées, et les raisonnements acquièrent une flexibilité que la logique classique ne permet pas.

De nombreuses applications sont alors développées dans divers domaines, là où aucun modèle déterministe n'existe ou n'est pratiquement implémentable, ainsi que dans des situations pour lesquelles l'imprécision sur les données rend le contrôle par des méthodes classiques impossible.



### 3.2. Les opérations en logique floue

Parmi les opérations sur les sous-ensembles flous, la logique floue qui permet à une condition d'être en un état autre que *vrai* et *faux*, d'où il y a des degrés dans la vérification d'une condition. Elle permet d'exprimer des connaissances imprécises et déduire des connaissances et prises des décisions pertinentes.

La logique floue s'appuie sur la théorie mathématique des sous-ensembles flous. Quand les entrées et les sorties sont binaires alors la logique floue se réduit à la logique classique, c'est-à-dire que la logique floue inclut la logique classique, ça nous permet de déduire que nous pouvons exprimer toutes les opérations binaires de base en se basant sur la logique floue. En effet, à partir des opérations ET, OU et NON (AND, OR, NOT), qui sont appelés *les opérateurs Min-Max de Zadeh*, on peut présenter les 8 opérations de base. Afin de pouvoir manipuler les sous-ensembles flous [4].

1. **ET (AND)** :  $A \text{ AND } B = \min(A,B)$
2. **OU (OR)** :  $A \text{ OR } B = \max(A,B)$
3. **NON (NOT)** :  $\text{NOT } A = 1 - B$
4. **OU EXCLUSIF (XOR)**:  $A \text{ XOR } B = (A \text{ OR } B) \text{ AND NOT } (A \text{ AND } B)$   

$$A \text{ XOR } B = A + B - 2 \times \min(A,B)$$
5. **NON-OU (NOR)**:  $A \text{ NOR } B = 1 - \max(A,B)$
6. **NON-ET (NAND)** :  $A \text{ NAND } B = 1 - \min(A,B)$
7. **NON-XOR (NXR)** :  $A \text{ NXR } B = 1 + 2 \times \min(A,B) - (A+B)$
8. **SUIVEUR (NOP)** :  $\text{NOP } A = A$ .

La définition des opérateurs sur les ensembles flous est choisie, à l'instar des fonctions d'appartenance. Voici les ensembles d'opérateurs pour le complément (NON), l'intersection (ET) et l'union (OU) utilisés le plus couramment

Il y a plusieurs types d'opérations pour les ET et OU, soit de type min-max, arithmétique, probabiliste ou floue, comme nous le montre le tableau suivant.

Tableau 2 : les différents types d'opérations

Type d'opération	Opérateur NON	Opérateur OU	Opérateur ET
Les opérateurs MIN-MAX (Mamdani)	$\mu_C=1-\mu_A(X)$	$\mu_{A\cup B}=\max(\mu_A(X), \mu_B(X))$	$\mu_{A\cap B}=\min(\mu_A(X), \mu_B(X))$
Les opérateurs arithmétiques(larsen)	$\mu_C=1-\mu_A(X)$	$\mu_{A\cup B}=1/2[\mu_A(X)+ \mu_B(X)]$	$\mu_{A\cap B}=\mu_A(X). \mu_B(X)$
Les opérateurs probabilistes (Sugeno)		$\mu_{A\cup B}=\mu_A(X)+\mu_B(X)-\mu_A(X). \mu_B(X)$	$\mu_{A\cap B}=\mu_A(X). \mu_B(X)$
Les opérateurs ET flou et OU flou		$\mu_{A\cup B}=\beta.\max(\mu_A(X), \mu_B(X))+ [(1- \beta)/2][ \mu_A(X)+ \mu_B(X)]$ Avec $\beta \in[0,1]$	$\mu_{A\cap B}=\beta. \min(\mu_A(X), \mu_B(X)) + [(1- \beta)/2][ \mu_A(X)+ \mu_B(X)]$ Avec $\beta \in[0,1]$
Opérateurs min-max et opérateur $\beta$	$\mu_A(X)=\beta.\min(\mu_A(X), \mu_B(X)) + (1- \beta).\max[ \mu_A(X), \mu_B(X)]$ Avec $\beta \in[0,1]$		

### 3.3. Le vocabulaire flou :

Avant de commencer cette partie, il est indispensable de définir certains vocabulaires, utilisés pour raisonner dans les systèmes flous.

- **Univers de discours :**

Domaine de variation d'une variable considérée, comme par exemple « Bon », « Moyen », « Faible » pour la variable « résultat ».

- **Variable linguistique :**

Dite « variable floue », comme par exemple « âge ».

- **Ensembles flous :**

Les valeurs d'une variable linguistique par exemple « âge » sont : les valeurs « jeune », « entre deux âges », « âgé »

- **Fonctions d'appartenance :**

Par exemple  $\mu_{jeune}(\text{âge})$

- **Inférence :**

Sont les déductions floues (une ou plusieurs règles)

- **Méthode d'inférence :**

Comment prendre en compte toutes les règles simultanément

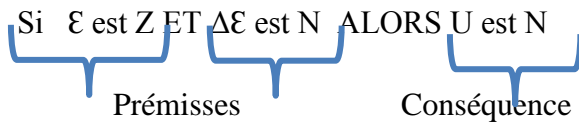
- **Prédicats :**

Appelés aussi classes d'appartenance comme par exemple : grand, petit, moyenne

### 3.4. Règle floue

Appelée aussi, règle d'inférence.

C'est-à-dire une expression subdivisée en deux parties, la première appelée « prémisses », et la deuxième est appelée « conséquence », par exemple



#### 3.4.1. Prémisses :

C'est la première partie d'une règle, peut être composée par une ou plusieurs prémisses, séparées entre eux par des conjonctions ET ou par des disjonctions OU.

#### 3.4.2. Conséquence :

C'est la partie conclusion de la règle.

#### 3.4.3. Les implications floues

Le raisonnement à partir de connaissances imparfaitement définies utilise la logique floue pour pallier les insuffisances de la logique classique. Le raisonnement flou nécessite la définition des notions de variable linguistique, proposition floue et implication floue.

Dans la logique floue, la structure de l'implication est définie par la forme suivante :

**SI X est A, ALORS Y est B.**

#### 4. Raisonnement par la logique floue

il est toujours possible de faire aussi mieux que l'être humain. C'est raisonner à l'aide de plusieurs règles floues composées, et des variables linguistiques expliquant l'état interne d'un système d'inférence floue.

##### 4.1. Système d'inférence floue

La méthodologie générale de la logique floue consiste à estimer l'imprécision d'une moyenne de résultats d'une simulation. Cette méthodologie consiste à définir un système flou composé par trois étapes ; la fuzzification, l'inférence floue, la défuzzification.

L'étape de fuzzification nécessite les sous-ensembles flous, l'inférence nécessite la base de connaissances qui peut être composée par ; base de règles floues, base de faits, l'étape de défuzzification demande un défuzzificateur (une méthode de défuzzification), comme illustré et détaillé dans la *Figure 9*.

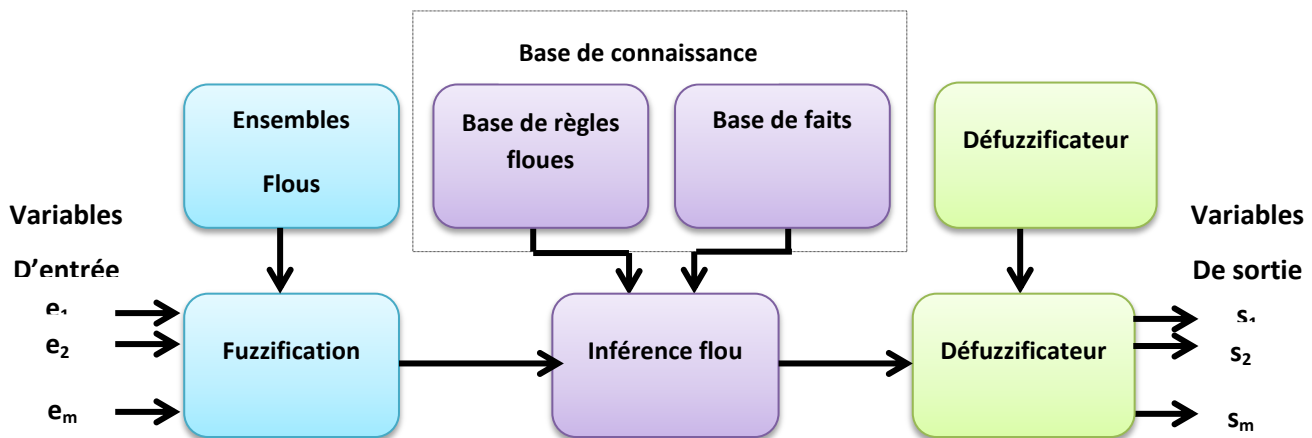


Figure 9 : Implémentation d'un système flou (Méthode d'inférence floue)

La modélisation des systèmes flous passe par trois étapes:

##### 4.2. Fuzzification:

C'est l'étape qui consiste en qualification floue des valeurs réelles d'une variable, c'est-à-dire déterminer les valeurs d'entrées et de sorties nécessaires et /ou suffisantes à la modélisation.

La fuzzification d'un système flou consiste à définir :

- L'univers du discours : la plage de variation possible de l'entrée considérée c'est-à-dire la conversion des données en entrée en degré d'appartenance par l'intermédiaire de fonctions d'appartenances.
- Une partition en classe floue de cet univers.
- Les fonctions d'appartenances : il y a un degré d'appartenance par terme linguistique s'appliquant à l'entrée.

Donc il faut fuzzifier les entrées et les sorties du système flou, c'est-à-dire selon les valeurs des entrées, le système flou indiquera qu'en sortie les valeurs linguistiques « faible », « moyen », « bon », « très bon ».

**Remarque** : la fuzzification est souvent réalisé de manière itérative et requiert l'expérience d'un spécialiste (expert) pour le domaine à modéliser.

### 4.3. Inférence:

C'est raisonner à l'aide de plusieurs règles floues composées, et des variables linguistiques expliquant l'état interne des systèmes flous, cette étape permet l'identification des paramètres des sous-ensembles flous.

La représentation de propositions par des formules affectant des ensembles flous comme valeurs aux variables floues.

Ce principe basé sur la proposition « plus la condition sur les entrées est vraie alors plus l'action préférée sur les sorties doit être conservée (respectée).

La conclusion d'une règle floue par ce principe est « l'appartenance d'une variable de sortie par exemple « chauffer » à une classe floue « fort », c'est-à-dire prendre, réaliser l'action « chauffer fort ».

Le problème de raisonnement approximatif est de trouver la fonction d'appartenance de la conséquence.

Il y a plusieurs méthodes d'inférence dans la littérature, ces méthodes utilisées pour traduire le terme **ALORS** pour les règles (les implications floues), ainsi les opérations pour définir l'interaction entre ces règles.

- **Méthode de Mamdani** (max-min):  $\mu_{\text{Conclusion}}(y) = \min_y(\mu_{\text{prémisse}}(X_0), \mu_{\text{Conclusion}}(y))$

- **Méthode de Larsen:**  $\mu_{\text{Conclusion}}(y) = \mu_{\text{prémisse}}(X_0) \times \mu_{\text{Conclusion}}(y)$

Parmi les autres méthodes, la méthode d'inférence **somme-prod**...

La méthode la plus utilisée est la méthode d'inférence max-min, et pas la méthode somme-prod ou la méthode de max-prod qui demandent beaucoup de calcul analytique de la sortie.

#### 4.4. Defuzzification:

C'est l'approximation linguistique, c'est-à-dire utiliser des variables linguistiques, cette opération faite après un raisonnement, nous sommes obligés de déduire un unique résultat à des fins d'interprétations ou de comparaisons, c'est-à-dire associer à tout résultat flou une étiquette linguistique en utilisant différentes méthodes :

Parmi les méthodes existantes, pour la défuzzification, la méthode de centre de gravité, la méthode Moyenne des minimums et Moyenne des maximums.

La méthode de centre de gravité est la plus utilisée.

- **Centre de gravité**

Cette méthode est la plus souvent utilisée car elle donne généralement de meilleurs résultats, elle consiste à prendre comme solution l'abscisse du centre de gravité des solutions. C'est-à-dire il s'agit de calculer la position du centre de gravité de la fonction d'appartenance résultante, l'abscisse de ce centre de gravité devient la sortie.

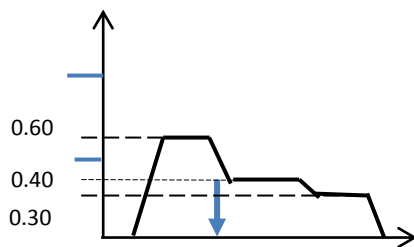


Figure 10: Défuzzification par la méthode du centre de gravité

- **Comment calculer le centre de gravité d'un polygone ? :**

En utilisant la méthode définie dans [5]

- Surface ou aire donnée par la formule suivante :

$$A = \frac{1}{2} \sum_{i=0}^{n-1} (x_i y_{i+1} - x_{i+1} y_i)$$

- Centre de gravité d'un polygone irrégulier

L'abscisse de centre de gravité donné par la formule suivante :

$$G_x = \frac{1}{6A} \sum_{i=0}^{n-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

L'ordonnée du centre de gravité donné par la formule suivante :

$$G_y = \frac{1}{6A} \sum_{i=0}^{n-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

- **Exemple :** Soit les points sommets du polygone : P<sub>1</sub>(5,0), P<sub>2</sub>(10,0.75), P<sub>3</sub>(15,0.75), P<sub>4</sub>(20,0.50), P<sub>5</sub>(25,0.50), P<sub>6</sub>(30,0.25), P<sub>7</sub>(35,0.25), P<sub>8</sub>(40,0).

Tableau 3 : clacul du centre de gravité

I	$(x_i y_{i+1} - x_{i+1} y_i)$	$(x_i y_{i+1} - x_{i+1} y_i)(x_i + x_{i+1})$	$(x_i y_{i+1} - x_{i+1} y_i)(y_i + y_{i+1})$
I = 1	5x0.75-10x0=3.75	3.75 x (5+10) = 56.25	3.75 x (0+0.75) = 2.8125
I = 2	10x0.75-15x0.75=-3.75	-3.75 x (10+15) = -93.75	3.75 x (0.75+0.75) = 5.6250
I = 3	15x0.50-20x0.75=-7.50	-7.50 x (15+20) = -262.50	-7.50 x (0.75+0.50) = -9.3750
I = 4	20x0.50-25x0.50=-2.50	-2.50 x (20+25) = -112.50	-2.50 x (0.50+0.50) = -2.50
I = 5	25x0.25-30x0.50=-8.75	-8.75 x (25+30) = -481.25	-8.75 x (0.50+0.25) = -6.5625
I = 6	30x0.25-35x0.25=-1.25	-1.25 x (30+35) = -81.25	-1.25 x (0.25+0.25) = -0.6250
I = 7	35x0-40x0.25=-10	-10 x (35+40) = -750	-10 x (0.25+0) = -2.5000
Total	-30	-1725	-24.375

Les résultats obtenus sont positifs ou négatifs selon le sens ou on numérote les points. Pour l'aire nous prendrons la valeur absolue du nombre. Par contre, pour le calcul du centre de gravité, il convient de respecter les signes.

Aire :  $A = \frac{-30}{2}$  , soit A = 15

Centre de gravité :  $G_x = \frac{-1725}{6A}$      $G_x = \frac{-1725}{6(-15)}$

Soit  $G_x = 19.17$

$G_y = \frac{-24.375}{6A}$

$G_y = \frac{-24.375}{6(-15)}$

Soit  $G_y = 0.27$

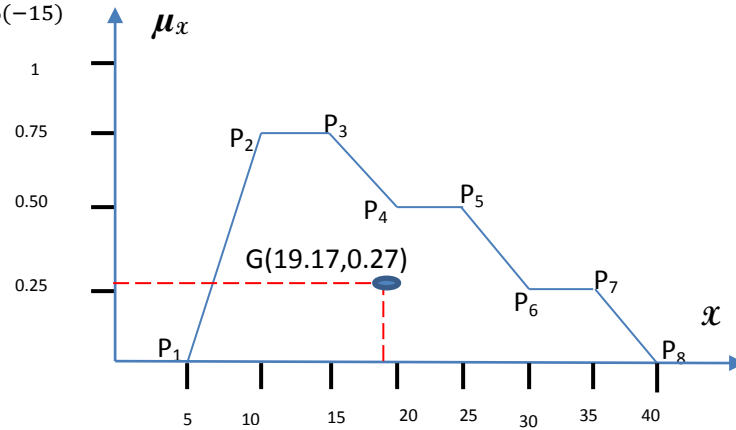


Figure 11 : calcul du centre de gravité (exemple)

**4.5. Exemple illustratif :**

Nous avons une machine de presse à café industrielle reliée par une interface électronique à un ventilateur électrique pour régler le couple (Température, Pression) c’est-à-dire si la pression augmente, le ventilateur se déclenche automatiquement pour diminuer la température, donc baisser la pression. On veut régler leur fonctionnement pour durer longtemps. On décide de réaliser l’action qui va provoquer contre chaque situation, pour tout ça, on a modélisé ce problème en construisant les règles suivantes, en définissant les variables d’entrées, sorties suivantes.

*Variables entrées :* Pression, Température, *variables de sortie :* Action

**Les sous-ensembles flous :**

Pression  $\in$  {Min, Max}, Température  $\in$  {Basse, élevée}, Action  $\in$  {produire plus le café, régler la machine, arrêter la machine}.

SI Pression est Min et Température est Basse ALORS Action est PP

SI Pression est Min et Température est Élevée ALORS Action est R

SI Pression est Max et Température est Basse ALORS Action est R

SI Pression est Max et Température est Élevée ALORS Action est A.

Pour représenter les valeurs floues (base de règles), on va utiliser les abréviations suivantes : Pression **P**, Température **T**, Basse **B**, élevée **E**, produire plus le café **PP**, régler la machine **R**, arrêter la machine **A**



<b>P</b> <b>T</b>	<b>Min</b>	<b>Max</b>
<b>B</b>	<b>PP</b>	<b>R</b>
<b>E</b>	<b>R</b>	<b>A</b>

Tableau 4 : Base de règles floues (Mamdani)

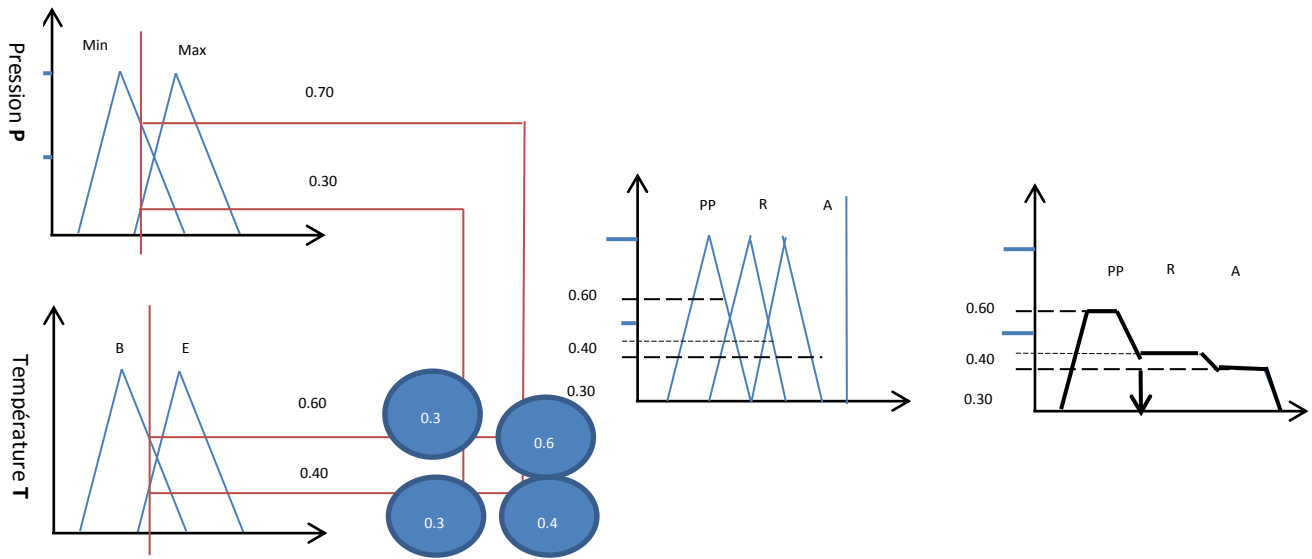


Figure 12 : Raisonnement flou par la méthode d'inférence floue de Min-Max de Mamdani

$$\mu_{PP} = \text{Min}(\mu_{0.70}, \mu_{0.60}) = \mu_{0.60}$$

$$\mu_R = \text{Min}(\mu_{0.70}, \mu_{0.40}) = \mu_{0.40}$$

$$\mu_R = \text{Min}(\mu_{0.30}, \mu_{0.60}) = \mu_{0.30}$$

$$\mu_A = \text{Min}(\mu_{0.30}, \mu_{0.40}) = \mu_{0.30}$$

Donc:

$$\mu_{PP\text{-Final}} = \mu_{0.60}$$

$$\mu_{R\text{-Final}} = \text{Max}(\mu_{0.40}, \mu_{0.30}) = \mu_{0.40}$$

$$\mu_{R\text{-Final}} = \mu_{0.30}$$

soit les sommets du polygone suivant :  $P_1(1,0)$ ,  $P_2(2,0.60)$ ,  $P_3(2.75,0.60)$ ,  $P_4(3,0.40)$ ,  $P_5(3.75,0.40)$ ,  $P_6(4.25,0.30)$ ,  $P_7(4.75,0.30)$ ,  $P_8(5,0)$ .

Après le calcul on trouve  $G_y=0.22$ , d'où l'action est : « produire plus le café **PP** »

## 6. Conclusion :

L'approche de la logique floue montre leur efficacité dans la modélisation de données imprécises, surtout pour prendre une décision pertinente sans perdre les informations.

La modélisation floue peut fortement contribuer à accroître les possibilités des SIG en:

- intégrant dans les bases de données certaines imprécisions inhérentes à l'information géographique.
- permettant l'utilisation de termes linguistiques non formels pour l'interrogation de bases de données, ce qui est confortable pour l'utilisateur moyen.

Bien que la base théorique complète et les outils logiciels pour les données géographiques floues ne soient pas encore bien établis, il est possible de profiter des deux avantages précités, si les précautions suivantes sont prises.

- vérifier que le concept portant l'imprécision est sensible aux paradoxes,
- trouver des fonctions d'appartenances et/ou des méthodes d'agrégation adaptées à la situation.
- vérifier que les  $\alpha$ -coupes des ensembles flous sont significatifs dans le contexte.

---



Chapitre IV :  
Méthodologie & Implémentation

---

# Chapitre IV : Implémentation & Implémentation

## 1. Justification du choix

Pour notre cas d'étude qui est l'appariement des données géographiques utilisant un compromis entre la théorie de la logique floue et la théorie de l'évidence est pris pour plusieurs raisons parmi lesquelles

- Après une étude détaillée dans les chapitres précédents, nous avons constaté que la théorie des possibilités et de probabilités sont comprises dans la théorie de l'évidence.
- La théorie des possibilités est une branche importante de la logique floue, car elle est basée sur les sous-ensembles flous, et toutes les opérations appliquées sur les sous-ensembles flous sont extensibles pour les possibilités
- La méthode la plus utilisée pour la prise de la décision qui est l'étape la plus importante pour la théorie de l'évidence basée sur la probabilité pignistique, et parfois la mesure de croyance ou la plausibilité.
- La méthode la plus utilisée pour la prise de la décision pour la théorie de possibilité est la méthode de defuzzification de la logique floue
- La théorie de l'évidence traite les données imprécises, incertaines, et incomplètes, pour la théorie de possibilité, la théorie des probabilités subjectives, ainsi que pour la théorie de logique floue.
- Le compromis que nous avons choisi comme méthode de modélisation des appariements des données géographiques, a pour but d'améliorer la décision finale.

## 2. Approche et implémentation

Le présent travail a pour objectif de trouver une solution théorique idéale pour résoudre les problèmes de l'appariement de données géographiques basées sur l'application des théories mathématiques (les théories de l'incertain, appelées aussi théories de l'imparfait) : théorie de l'évidence et de la logique floue.

Il s'agit de développer un système géographique SIG, qui permet facilement d'accomplir les tâches suivantes : l'accès aux bases de données géographiques et faire l'appariement entre ces bases basées sur la nouvelle approche dans le but d'intégrer des données: c'est à dire à partir des bases de données

géographiques, une base de données de référence, et l'autre comme une base de données à comparer, et construire une base de données géographique dérivée contenant les objets homologues de ces bases.

Notre participation touche deux grands axes, un axe qui concerne l'approche proposée et la deuxième axe qui concerne la nouvelle méthode pour calculer la distance toponymique entre les candidats qui participent dans l'opération d'appariement.

*Premièrement* : en définissant un cadre théorique basé sur un compromis entre la théorie de l'évidence et la théorie de la logique floue, puis on développe une application c.-à-d. un SIG qui permet d'interroger une ou plusieurs bases de données géographiques pour construire une base de données dérivée contenant les données communes entre ces bases et les enrichir. On se basant sur l'appariement des données géographiques sur les critères géométriques, toponymiques de différentes sources d'informations.

*Deuxièmement* : pour définir les critères toponymiques, nous avons utilisé notre nouvelle méthode basée sur la distance de Levenshtein pour calculer la distance toponymique.

Concernant l'implémentation, nous avons utilisé une base de données géographiques représentant les routes extraite d'une «*Base de données géographiques et administratives à l'échelle 1/1 000 000 (BDGA 1M)*», ressources naturelles et faune du Québec[6], dans laquelle l'information spatiale présentée par deux dimensions :

1. *La dimension géométrique* : la taille et la forme.
2. *La dimension toponymique* : les noms et numéros des routes, les attributs.

Une implémentation de ces deux approches a été réalisée.

## **2.1. Initialisation des fonctions de masses**

L'initialisation des fonctions de masses est une étape importante dans la TDS, qui présente un problème et leurs effets peut influencer la prise de décision.

Notre approche est basée sur le chevauchement des étapes de logique floue et de théorie des fonctions de croyance, pour permettre d'initialiser des fonctions de masse et prendre une décision pertinente, (*Figure 13*).

L'approche est une succession des étapes suivantes:

**1<sup>ère</sup> étape:** définition des variables d'entrée ( $E_1, E_2, \dots, E_n$ ) de différentes sources d'information utilisées, et les variables de sortie ( $S_1, S_2, \dots, S_n$ ).

**2<sup>ème</sup> étape:** la Fuzzification, qui consiste à la qualification floue des valeurs floues d'une variable, c'est-à-dire déterminer les valeurs d'entrées et de sorties nécessaires et/ou suffisantes à la modélisation (*la représentation floue des variables floues selon les sous-ensembles flous*).

**3<sup>ème</sup> étape:** construction de la table de décision, selon un expert dans le domaine

**4<sup>ème</sup> étape:** Sélection des candidats.

Sélection des candidats et définition du cadre de discernement, après la définition des seuils par la définition des sous-ensembles flous.

**5<sup>ème</sup> étape:** l'initialisation des masses de croyance à partir du graphe des valeurs floues selon les étapes qui sont expliquées en dessous.

**6<sup>ème</sup> étape:**

- a. définition d'un *facteur de fiabilité* pour les sources d'informations.
- b. Combinaison des sources d'informations, en multipliant les masses de chaque source par leur facteur de fiabilité  $F$ , qui sont expliquées en dessous.

**7<sup>ème</sup> étape:** fusion des critères c'est-à-dire la combinaison des masses de croyance par l'opérateur *Dempster-Shafer*.

**8<sup>ème</sup> étape:** fusion des candidats c'est-à-dire la normalisation des masses de croyance telle que la somme des masses soit égale à 1.

**9<sup>ème</sup> étape :** prise de décision en utilisant la probabilité pignistique.

La figure 13 récapitule les étapes de notre approche.

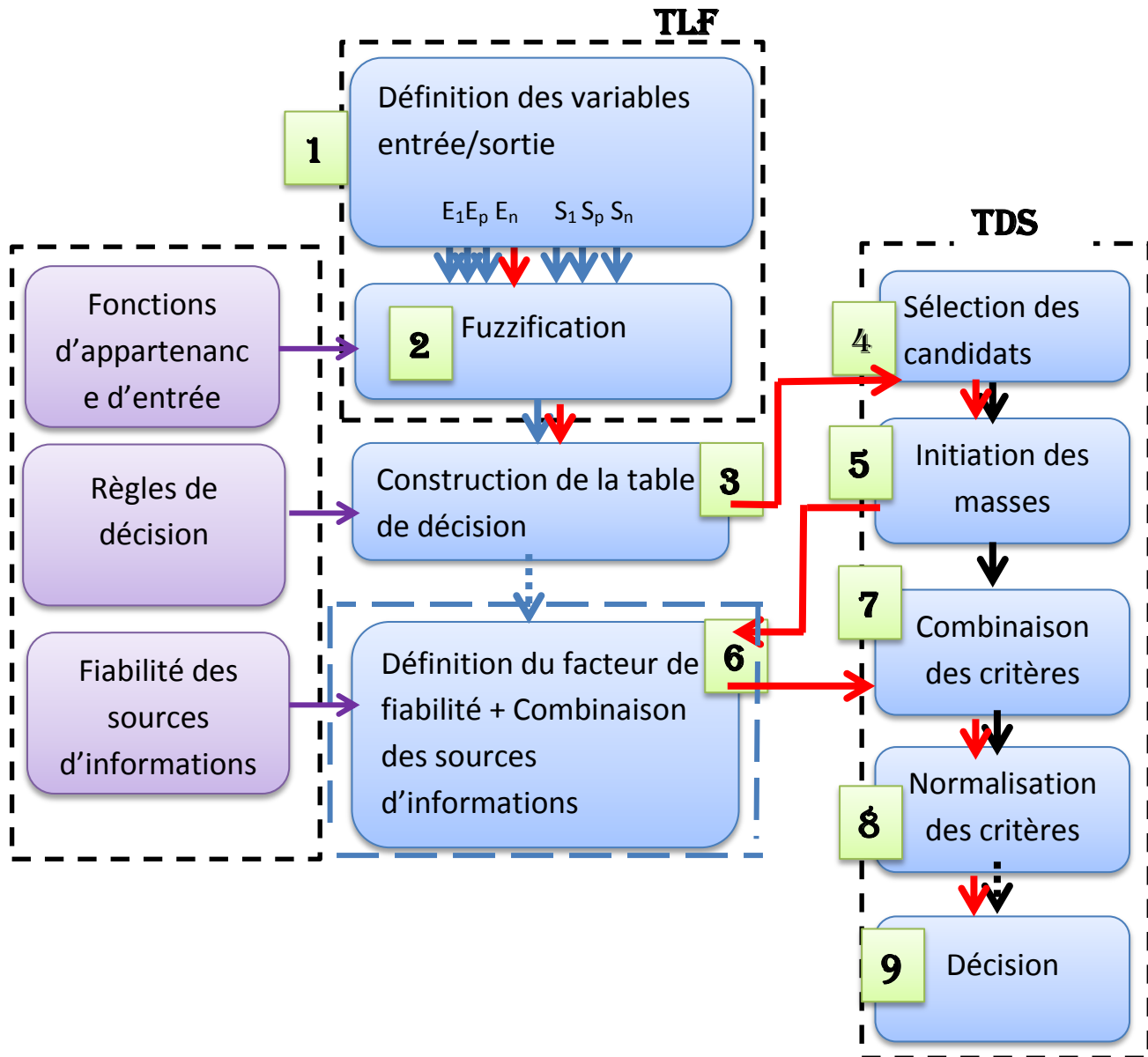


Figure 13: Approche sur théorie de l'évidence & la logique floue : Application à l'appariement des données géographiques

- **Explication de l'approche :**

Pour mieux cerner les principes utilisés par notre méthode, on explique les différentes étapes en détail :

**1<sup>ème</sup> étape : définition des variables floues d'entrée et de sortie**

Dans cette étape on doit définir le sous-ensemble flou nécessaire pour chaque variable floue, leur type, et leur abréviation.

Chaque critère d'appariement lié par une classe comme (classe géométrie, classe sémantique, classe topologique, classe toponymique...). Les critères d'appariement liés chacun par une variable floue, et un sous-ensemble flou.

**Exemple :** nous avons pris deux variables d'entrée, et une variable de sortie (Homologue)

- le variable flou : **Distance Euclidienne ( $D^e$ )**, Leur Type : **Entrée (E)**,  $D^e \in \{\text{Proche (P)}, \text{Moyen (M)}, \text{Loin (L)}\}$
- le variable flou : **Distance Toponymique ( $D^t$ )**, Leur Type : **Entrée (E)**,  $D^t \in \{\text{Proche (P)}, \text{Moyen (M)}, \text{Loin (L)}\}$

### 2<sup>ème</sup> étape : la fuzzification

C'est une étape pour définir chaque sous-ensemble flou (donner la plage des valeurs pour chaque variable).

Dans cette étape on définit l'intervalle (valeur Min et valeur Max) selon le contexte, et on selectionne les sources d'informations prises pour l'opération d'appariement, c'est-à-dire quel critère utiliser pour lancer la procédure d'appariement.

**Exemple :**

Pour le variable  $D^e$  : **Proche(P)** $\in[0,300]$ , **Moyen(M)** $\in[0,600]$ , **Loin(L)** $\in[300,1500]$ .

Nous avons choisi les critères suivants : distance euclidienne, distance toponymique.

### 3<sup>ème</sup> étape : construction de la table de décision

Nous avons ajouté une étape pour définir les décisions possibles (Homologue, ne pas se prononcer, non homologue), selon la distance utilisée, c'est-à-dire pour quelle distance on peut dire que le candidat est homologue, ou non homologue, ou on ne peut pas se prononcer.

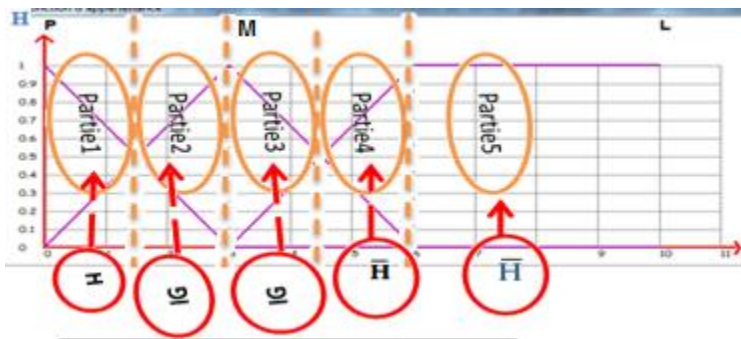


Figure 14: les différentes décisions possibles



<b>D<sup>e</sup></b>	<b>P</b>	<b>M</b>	<b>L</b>
<b>P</b>	<b>H</b>	<b>H</b>	<b>IG</b>
<b>M</b>	<b>H</b>	<b>IG</b>	<b>H̄</b>
<b>L</b>	<b>IG</b>	<b>H̄</b>	<b>H̄</b>

Tableau 5 : table de décision

#### 4<sup>ème</sup> étape : Sélection des candidats

Après la définition des seuils par la définition des sous-ensembles flous de l'étape précédente, on peut sélectionner des candidats et définir le cadre de discernement, c'est-à-dire définir l'ensemble des hypothèses possibles comme solution du problème d'appariement.

Premièrement, on sélectionne les bases de données (BDG de référence, BDG à comparer), puis on définit le seuil expérimental de sélection des candidats potentiels selon la distance euclidienne, et voir quels candidats sont homologues potentiels dans la deuxième base de données géographiques à comparer.

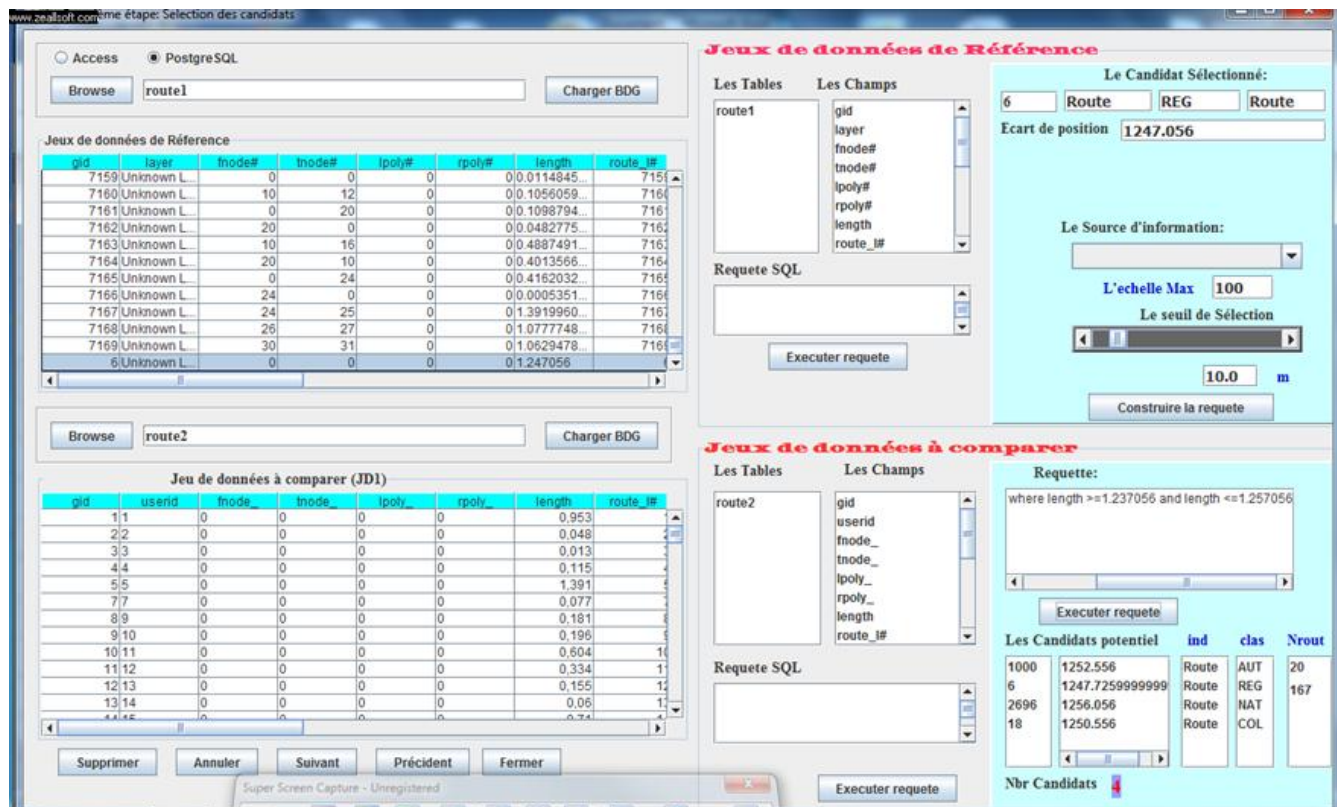


Figure 15 : 4<sup>ème</sup> étape « Sélection des candidats »

Dans notre cas nous avons pris un candidat de référence (gid=6) de la base route1, on a défini un seuil expérimental ( $S=10m$ ), nous avons trouvé les candidats qui peuvent être leur homologues potentiels (gid=1000, 6, 2696, 18) de la base route2, comme illustré sur la Figure 15.

### 5ème étape : Initiation des masses

Cette étape représente une de nos deux contributions, elle permet de spécifier la méthode d'initialisation des fonctions de masses.

D'abord nous choisissons aléatoirement à partir de la base de données de référence (DB1) un objet de référence. Nous définissons un seuil expérimental pour choisir les candidats potentiels selon la distance euclidienne. Nous obtenons un ensemble d'objets qui sont potentiellement des candidats de la base de données à comparer (DB2).

L'initialisation de la masse de base de croyance exige un expert. Dans ce travail, nous avons essayé de donner une méthode conviviale pour la faire.

- **Le critère distance euclidienne**

La distance euclidienne, entre l'objet de référence et son homologue potentiel, est « fuzzifié » comme montré sur la Figure 16 et divisé en 5 parties.

À partir de cette figure, nous avons défini trois termes: proche (N), milieu (M) et loin (F). Pour chaque candidat, nous prenons trois possibilités, le candidat est apparié quand cette distance est proche

(l'homologue H), ou non-homologue si la distance est très lointaine (U) ou nous ne pouvons pas se prononcer si la distance est assez lointaine (ignorance I). Pour chaque possibilité, nous calculons la masse de base de croyance.

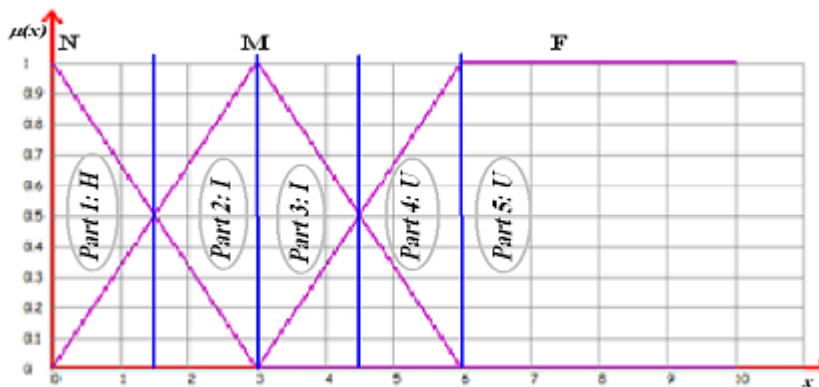


Figure 16 : Le partage de distance en ensembles flous.

Nous avons choisi le calcul du centre de gravité sur les cinq parties du graphe pour l'initialisation des fonctions de masses, car la fonction du centre de gravité est monotone sur chaque partie de ce graphe suite aux points suivants :

- Pour la partie 1 : (P,M)→H, selon la table de décision (Tableau 5)

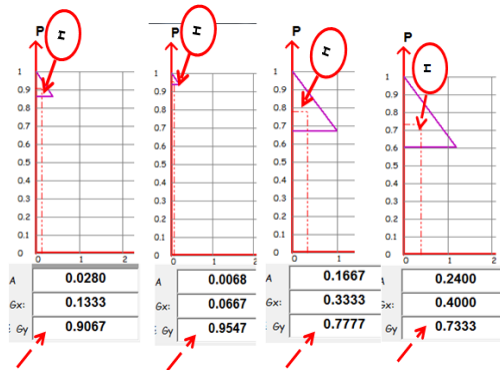


Figure 17: calcul de centre de gravité de la première partie du graphe

Donc :  $m_H(C)=0.9067$ ,  $m_H(C)=0.9547$ ,  $m_H(C)=0.7777$ ,  $m_H(C)=0.7333$

On remarque que chaque fois qu'on s'approche vers la partie 2, qui représente la partie d'ignorance, le centre de gravité diminue, et qu'on s'approche vers le début de la partie 1, le centre de gravité augmente, donc la fonction de centre de gravité est monotone.

- Pour la partie 2 : (M,P)→IG, selon la table de décision (Tableau 5)

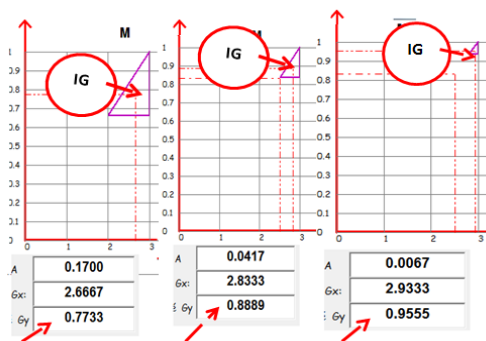


Figure 18: calcul de centre de gravité de la deuxième partie du graphe

Donc :  $m_{IG}(C)=0.7733$ ,  $m_{IG}(C)=0.8889$ ,  $m_{IG}(C)=0.9555$

On remarque que chaque fois qu'on s'approche vers le maximum de  $\mu_m(C)$  (fin de la partie2), qui représente la zone d'ignorance, le centre de gravité augmente, et qu'on s'approche vers le début de la partie2, le centre de gravité diminue, donc la fonction de centre de gravité est monotone.

- Pour la partie3 : (M,L)→IG, selon la table de décision (Tableau 5)

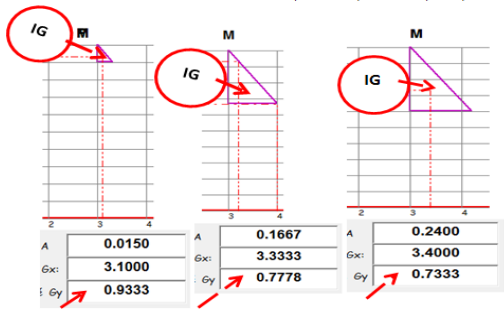


Figure 19: calcul de centre de gravité de la troisième partie du graphe

Donc :  $m_{IG}(C)=0.7778$ ,  $m_{IG}(C)=0.7333$  ,  $m_{IG}(C)=0.9333$

On remarque que chaque fois qu'on s'approche vers la partie 4, qui représente la zone de non homologue, le centre de gravité diminue, et qu'on s'approche vers le début de la partie 3, le centre de gravité augmente, donc la fonction de centre de gravité est monotone.

- Pour la Partie4 : (L ,M)→ $\bar{H}$ , selon la table de décision (Tableau 5)

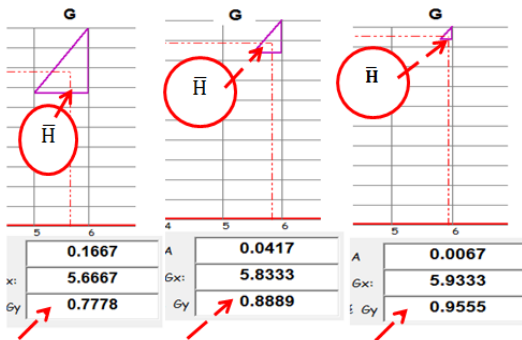


Figure 20: calcul de centre de gravité de la quatrième partie du graphe

Donc :  $m_{\bar{H}}(C) = 0.7778$ ,  $m_{\bar{H}}(C) = 0.8889$ ,  $m_{\bar{H}}(C) = 0.9555$

On remarque que chaque fois qu'on s'approche vers la partie5, qui représente la zone non homologue parfait, le centre de gravité augmente, et qu'on s'approche vers le début de la partie 4, le centre de gravité diminue, donc la fonction de centre de gravité est monotone.

- Pour la Partie5 :  $(L, M) \rightarrow \bar{H}$ , selon la table de décision (Tableau 5)

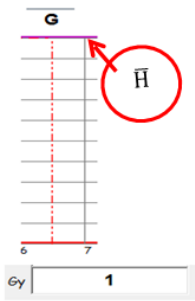


Figure 21: calcul de centre de gravité de la cinquième partie du graphe

Dans cette partie la fonction de centre de gravité est 1, donc :  $m_{\bar{H}}(C) = 1$

**Formellement :**

Afin de donner une formulation pour l'initialisation des masses, nous donnons cette approche sous forme de règle.

Si  $x$  dans *Partie<sub>i</sub>* Alors  $m(H)=f_1(x)$ ,  $m(I)=f_2(x)$ ,  $m(U)=f_3(x)$

Où  $f_i$  est définie comme suit:

Pour chaque partie, nous calculons le centre de gravité du triangle avec la base  $y = \max(\mu_1(x), \mu_2(x))$  comme montré sur la Figure 22.

Soit  $G(G_x, G_y)$ , ce centre de la gravité, les masses sont donnés par:

• **Partie 1:**

$$m(H)=G_y, m(I)=1-m(H), m(U)=0$$

Dans la présente partie, la référence est très proche du candidat, ainsi nous favorisons l'hypothèse : le candidat est homologue à la référence.

• **Parties 2 et 3:**

La référence est assez près du candidat, l'hypothèse est partagée entre l'ignorance et l'homologue avec une légère préférence à l'ignorance.

- **Partie 2 :**  $m(I)=G_y, m(H)=1-m(I), m(U)=0$

- **Partie 3:**  $m(I)=G_y, m(U)=1-m(I), m(H)=0$

• **Parties 4 et 5:**

La référence est loin du candidat alors la préférence est entre les hypothèses non-homologue et l'ignorance.

Partie 4 :  $m(U)=Gy$ ,  $m(I)=1-m(U)$ ,  $m(H)= 0$ .

Partie 5 :  $m(U)=1$ ,  $m(I)=0$ ,  $m(H)= 0$ .

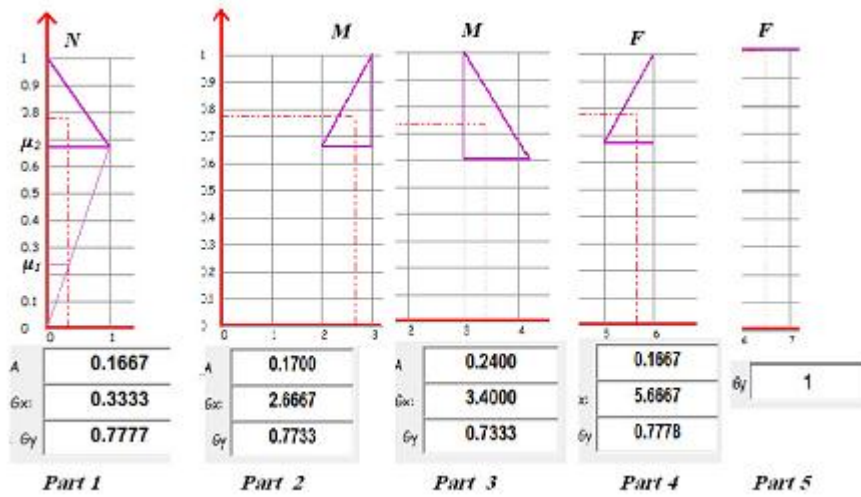


Figure 22. Méthode de calcul des masses(Méthode récapitulative).

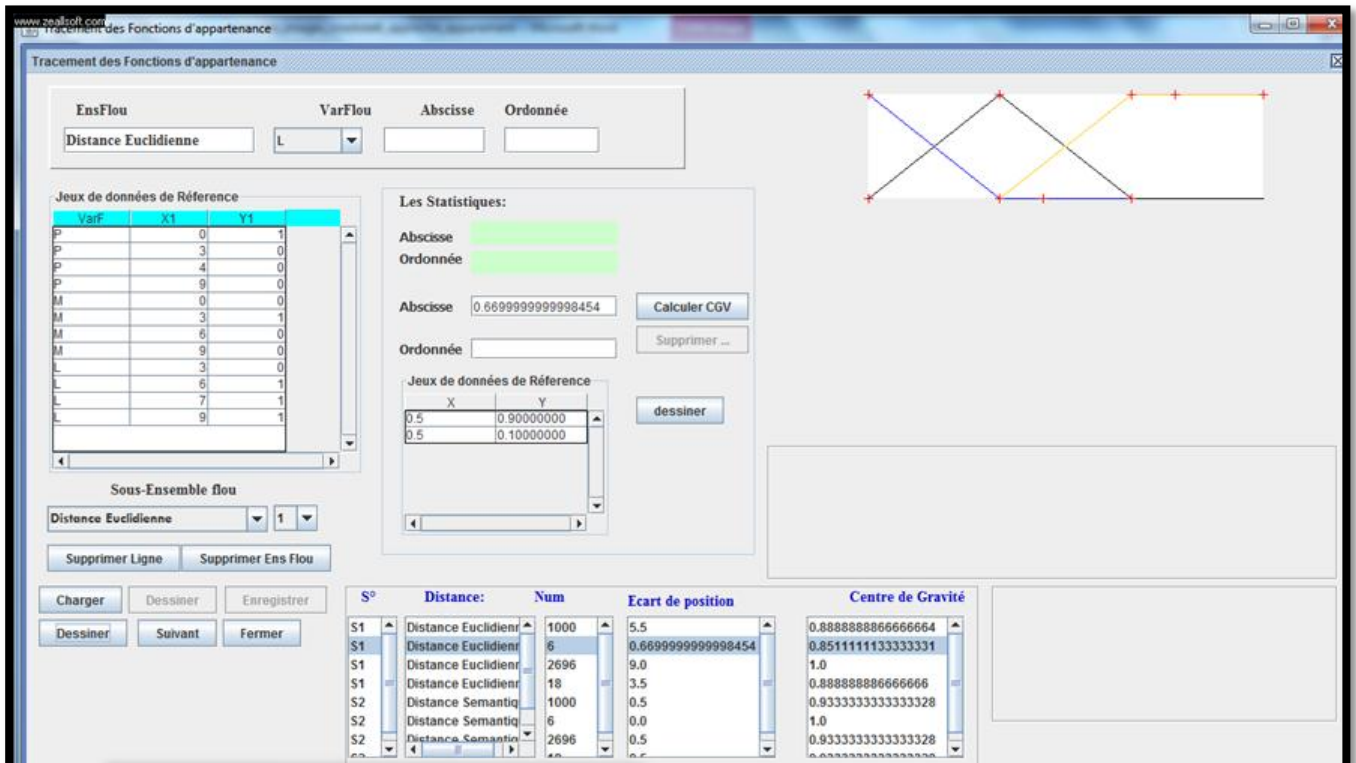


Figure 23: calcul du centre de gravité

En suivant l'exemple précédent, nous avons calculé leurs centres de gravités, selon les deux sources distance euclidienne, distance toponymique ( Figure 23)

Les résultats trouvés après l'utilisation la table de décision :

Source  $S_1$ : fourni Distance euclidienne

$$\begin{cases} m_{A1H}^{S1} = 0.85111 \\ m_{A2H}^{S1} = 0.14889 \\ m_{A3H}^{S1} = 0.00000 \\ m_{A4H}^{S1} = 0.00000 \end{cases}$$

Source  $S_2$ : fournie Distance toponymique

$$\begin{cases} m_{A1H}^{S2} = 1.00000 \\ m_{A2H}^{S2} = 0.76333 \\ m_{A3H}^{S2} = 0.61000 \\ m_{A4H}^{S2} = 0.00000 \end{cases}$$

## 6<sup>ème</sup> étape :Définition du facteur de fiabilité + Combinaison des sources d'informations

### a. Définition du facteur de fiabilité

Cette étape consiste à évaluer les sources d'informations par la définition d'un facteur de fiabilité  $F \in ] 0, 1]$ , déterminé par un expert, selon plusieurs critères comme les circonstances atmosphériques, la technologie du matériel de mesure,...

- Si  $F=0$ , la source est non fiable, dans ce cas-là nous perdons l'information parvenant de la source, ce cas-là ne se produire pas dans la réalité.
- Si  $F=1$ , la source est entièrement fiable.
- Si  $F \in ]0, 1[$ , la source est partiellement fiable.

### b. Combinaison des sources d'informations

La combinaison de la source nécessite la multiplication des masses de chaque source par leur facteur de fiabilité  $F$ , et l'obtention d'une nouvelle distribution des masses.

En suivant l'exemple précédent on trouve les résultats suivants :

L'expert donne :

Pour  $S_1$  :  $F=0.98$  et pour  $S_2$  :  $F=0.67$

La combinaison des sources d'informations :

$$\begin{cases} m_{A1H}^{S1} = 0.83408 \\ m_{A2H}^{S1} = 0.14591 \\ m_{A3H}^{S1} = 0.00000 \\ m_{A4H}^{S1} = 0.00000 \end{cases}$$

Et

$$\begin{cases} m_{A1H}^{S2} = 0.67000 \\ m_{A2H}^{S2} = 0.51143 \\ m_{A3H}^{S2} = 0.64908 \\ m_{A4H}^{S2} = 0.00000 \end{cases}$$

#### 7ème étape: la combinaison des masses :

Est la fusion des critères c'est-à-dire la combinaison des masses de croyance par l'opérateur *Dempster-Shafer*.

En suivant l'exemple précédent, on trouve les résultats suivants :

$$\begin{cases} m_{A1H}^{S12} = 0.55883 \\ m_{A2H}^{S12} = 0.07462 \\ m_{A3H}^{S12} = 0.00000 \\ m_{A4H}^{S12} = 0.00000 \end{cases}$$

#### 8ème étape: fusion des candidats

C'est-à-dire la normalisation des masses de croyance telle que la somme des masses soit égale à 1.

En suivant l'exemple précédent on trouve les résultats suivants :

$$\begin{cases} m_{A1H}^{S12} = \mathbf{0.88220} \\ m_{A2H}^{S12} = 0.11779 \\ m_{A3H}^{S12} = 0.00000 \\ m_{A4H}^{S12} = 0.00000 \end{cases}$$



## 9ème étape : la décision

On utilisant la probabilité pignistique, pour prendre la décision final.

La décision en utilisant la probabilité pignistique, la probabilité de crédibilité et la probabilité de plausibilité donne le même résultat pour les hypothèses singletons.

En suivant l'exemple précédent on trouve le résultat suivant :

$\mathbf{A}_1$  est l'homologue de  $\mathbf{O}_1$ , avec une masse  $m_{A_1H}^{S12} = 0.88220$

### 2.2. Critère de la distance toponymique

La distance toponymique joue un rôle important comme critère d'appariement. Son avantage est que, en calculant la distance toponymique, vous pouvez contrôler et manipuler des erreurs telles que l'omission, la substitution, la transposition, l'abréviation, et la faute d'orthographe.

Cette partie représente notre deuxième contribution, une méthode améliorée de Samal [40] qui est basés sur la distance de Levenshtein pour une décision appropriée de la similitude toponymique entre deux objets, elle s'inscrit dans le contexte d'appariement multi-critères.

Cette méthode est basée sur l'algorithme suivant :

- a. Etape de *Tokenization* : on découpe les deux chaînes en mots (unités lexicales), et éliminer les mots vides (de, des, le, la, ...).
- b. Etape de correction : cette étape consiste à comparer ces mots avec un dictionnaire et corriger les mots erronés, en se basant sur un dictionnaire du domaine étudié (concerné).
- c. Etape de construction d'une matrice dont les lignes, représentent les mots de la chaîne1, et les colonnes représentent les mots de la chaîne 2,
- d. Etape du calcul de la distance de la distance de *Levenshtein*, entre les mots, et les stockées dans la matrice,
- e. Normaliser la distance de levenshtein pour tous les éléments de la matrice par la méthode suivante :

Soit  $a_{ij}$  un élément de la matrice  $\mathbf{A}$ , soit  $\hat{a}_{ij}$ , un élément normalisé de la matrice  $\mathbf{A}$ , tel que  $\hat{a}_{ij} = \frac{dL(chaine1,chaine2)}{\max(l1,l2)}$ , et  $l1$  est la longueur de la chaîne1,  $l2$  est la longueur de la chaîne2

- f. Normaliser les lignes de la matrice présentant la distance de levenshtein normalisée,

$$D_{L1} = \frac{\sum_{i=1}^{i=\text{nbr lignes}} \min(a_{ij})}{\text{nbr des lignes}}$$

C'est-à-dire la somme des éléments minimal de chaque ligne de la matrice sur le nombre des lignes.

*Condition* : on ne pas normaliser si tous les éléments d'une ligne =1, car nous sommes certains qu'il y a non similarité totale.

On définit la distance *toponymique* par la formule suivante :

$$Dt = \frac{\min1 + \max1}{2}$$

Soit  $\min1 = \min(D_{L1}, D_{L2})$ ,  $\max1 = \max(D_{L1}, D_{L2})$ .

Tel que  $D_L$ , est la distance de levenshtein calculée par la formule suivante :

$D_{L1}$  : la distance de levenshein calculé sur la matrice  $(A)_{ij}$ ,

$D_{L2}$  : la distance de levenshein calculé sur la matrice transposée  $(A)_{ji}$ ,

$$D_{L1} = \frac{\sum_{i=1}^{i=\text{nbr lignes}} \min(a_{ij})}{\text{nbr des lignes}}$$

$$D_{L2} = \frac{\sum_{j=1}^{j=\text{nbr lignes}} \min(a_{ji})}{\text{nbr des lignes}}$$

Afin de valider notre algorithme, nous avons développé une autre application, qui permet le calcul de la distance de Levenshein, la distance toponymique par la méthode de Samal, et la distance toponymique par notre méthode. Nous avons testé cette approche sur 4 types (catégories) des exemples, possibles, c'est-à-dire chaque type par un exemple.

- a. Les chaines ch1, ch2, sont identiques
- b. Les chaines ch1, ch2, sont différentes
- c. Les chaines ch1 inclus dans la chaine ch2
- d. Les chaines ch1, ch2 sont partiellement chevauchées (par des mots ou des lettres).

Les différents résultats sont représentés dans le (*Tableau 6*).


Categorie	Chainel	Chaîne2	D_top_Samal	D_top_Nouv_meth
Ch1=ch2	main road	main road	$D_{top\_samal}(ch1, ch2) = 0.0$	$D_{top\_nov\_Moy}(ch1, ch2) = 0.0$
			$D_{top\_samal}(ch2, ch1) = 0.0$	$D_{top\_nov\_Moy}(ch1, ch2) = 0.0$
Ch1≠ch2	Street	Highway	$D_{top\_samal}(ch1, ch2) = 1.0$	$D_{top\_nov\_Moy}(ch1, ch2) = 1.0$
			$D_{top\_samal}(ch2, ch1) = 1.0$	$D_{top\_nov\_Moy}(ch1, ch2) = 1.0$
Ch1 inclus dans ch2	Main principal national propriator road	main road	$D_{top\_samal}(ch1, ch2) = 0.200$	$D_{top\_nov\_Moy}(ch1, ch2) = 0.142$
			$D_{top\_samal}(ch2, ch1) = 0.428$	$D_{top\_nov\_Moy}(ch1, ch2) = 0.142$
Ch1, ch2 sont partiellement chevauchés	main road	Secondary road	$D_{top\_samal}(ch1, ch2) = 0.44$	$D_{top\_nov\_Moy}(ch1, ch2) = 0.234$
			$D_{top\_samal}(ch2, ch1) = 0.388$	$D_{top\_nov\_Moy}(ch1, ch2) = 0.234$

Tableau 6: la validation de la distance toponymique.

### 3. Résultats et comparaisons:

Afin de valider notre approche, nous avons tenté d'implémenté les méthodes utilisées par Ana-Maria [8], avec les notes

Le tableau suivant (une étude de cas) résume les résultats trouvés par l'approche d'Ana-Maria utilisant l'algorithme de Samal pour la distance toponymique, et notre approche utilisant le nouvel algorithme pour le calcul de la distance toponymique.

Etape N° :	Description	Notre approche	L'approche de : anna maria
Etape 1 :	Définition des variables floues d'E/S	<p><u>Var Distance euclidienne</u> <math>D^e</math>, type E, <math>D^e \in \{ \text{Proche P, Moyenne M, Loin L} \}</math></p> <p><u>Var Distance toponymique</u> <math>D^t</math>, type E, <math>D^t \in \{ \text{Proche P, Moyenne M, Loin L} \}</math></p> <p><u>Var Homologue</u> S, type S, <math>S \in \{ \text{Homologue H, Ignorance IG, Non Homologue NH} \}</math></p>	

<b>Etape 2 :</b>	Fuzzification	$\text{Var } D^e : P \in \{0 : 3\}, M \in \{0 : 6\}, L \in \{3 : 9\}$ $\text{Var } D^l : P \in \{0 : 0.5\}, M \in \{0 : 1.0\}, L \in \{0.5 : 1.0\}$	/																
<b>Etape 3 :</b>	Construction de la table de décision	<table border="1" style="width: 100%; text-align: center;"> <tr> <td style="color: red;"><b>D<sup>e</sup></b></td> <td><b>P</b></td> <td><b>M</b></td> <td><b>L</b></td> </tr> <tr> <td><b>P</b></td> <td><b>H</b></td> <td><b>H</b></td> <td><b>IG</b></td> </tr> <tr> <td><b>M</b></td> <td><b>IG</b></td> <td><b>IG</b></td> <td><b>H̄</b></td> </tr> <tr> <td><b>L</b></td> <td><b>IG</b></td> <td><b>H̄</b></td> <td><b>H̄</b></td> </tr> </table>	<b>D<sup>e</sup></b>	<b>P</b>	<b>M</b>	<b>L</b>	<b>P</b>	<b>H</b>	<b>H</b>	<b>IG</b>	<b>M</b>	<b>IG</b>	<b>IG</b>	<b>H̄</b>	<b>L</b>	<b>IG</b>	<b>H̄</b>	<b>H̄</b>	/
<b>D<sup>e</sup></b>	<b>P</b>	<b>M</b>	<b>L</b>																
<b>P</b>	<b>H</b>	<b>H</b>	<b>IG</b>																
<b>M</b>	<b>IG</b>	<b>IG</b>	<b>H̄</b>																
<b>L</b>	<b>IG</b>	<b>H̄</b>	<b>H̄</b>																
<b>Etape 4 :</b>	Sélection des candidats	<p>Définition d'un seuil de sélection : dans notre exemple = <b>10U</b>, et ça nous permet de récolter les candidats potentiels <b>A<sub>1</sub>, A<sub>2</sub>, A<sub>3</sub>, A<sub>4</sub></b> de l'objet de référence <b>O<sub>1</sub></b> :</p> <p><b>Distance euclidienne</b></p> $\begin{cases} De_{O_1, A_1} = 0.67U \\ De_{O_1, A_2} = 3.50U \\ De_{O_1, A_3} = 9.00U \\ De_{O_1, A_4} = 5.50U \end{cases}$ <p><b>Distance toponymique (la notre)</b></p> <p><b>O<sub>1</sub></b>= « main road », <b>A<sub>1</sub></b>= « main road », <b>A<sub>2</sub></b>= « main principal national proprietor road », <b>A<sub>3</sub></b>= « secondary road », <b>A<sub>4</sub></b>= « highway »</p> $\begin{cases} Dt_{O_1, A_1} = 0.000U \\ Dt_{O_1, A_2} = 0.142U \\ Dt_{O_1, A_3} = 0.234U \\ Dt_{O_1, A_4} = 0.928U \end{cases}$	<p>Définition d'un seuil de sélection : dans notre exemple = <b>10U</b>, et ça nous permet de récolter les candidats potentiels <b>A<sub>1</sub>, A<sub>2</sub>, A<sub>3</sub>, A<sub>4</sub></b> de l'objet de référence <b>O<sub>1</sub></b> :</p> <p><b>Distance euclidienne</b></p> $\begin{cases} De_{O_1, A_1} = 0.67U \\ De_{O_1, A_2} = 3.50U \\ De_{O_1, A_3} = 9.00U \\ De_{O_1, A_4} = 5.50U \end{cases}$ <p><b>Distance toponymique (samal)</b></p> <p><b>O<sub>1</sub></b>= « main road », <b>A<sub>1</sub></b>= « main road », <b>A<sub>2</sub></b>= « main principal national proprietor road », <b>A<sub>3</sub></b>= « secondary road », <b>A<sub>4</sub></b>= « highway »</p> $\begin{cases} Dt_{O_1, A_1} = 0.000U \\ Dt_{O_1, A_2} = 0.428U \\ Dt_{O_1, A_3} = 0.440U \\ Dt_{O_1, A_4} = 0.904U \end{cases}$																
<b>Etape 5 :</b>	Initiation des masses	<p>On utilise notre graphe, et le centre de gravité on trouve :</p> <p><b>Source S<sub>1</sub></b>: fourni Distance euclidienne</p> $\begin{cases} m_{A_1 H}^{S_1} = 0.85111 \\ m_{A_2 H}^{S_1} = 0.14889 \\ m_{A_3 H}^{S_1} = 0.00000 \\ m_{A_4 H}^{S_1} = 0.00000 \end{cases}$ <p><b>Source S<sub>2</sub></b>: fournie Distance toponymique</p>	<p>On utilise le graphe de anna maria, on trouve :</p> <p><b>Source S<sub>1</sub></b>: fourni Distance euclidienne</p> $\begin{cases} m_{A_1 H}^{S_1} = 0.88833 \\ m_{A_2 H}^{S_1} = 0.43333 \\ m_{A_3 H}^{S_1} = 0.10000 \\ m_{A_4 H}^{S_1} = 0.16666 \end{cases}$ <p><b>Source S<sub>2</sub></b>: fournie Distance toponymique</p>																

		$\begin{cases} m_{A1H}^{S2} = 1.00000 \\ m_{A2H}^{S2} = 0.76333 \\ m_{A3H}^{S2} = 0.61000 \\ m_{A4H}^{S2} = 0.00000 \end{cases}$	$\begin{cases} m_{A1H}^{S2} = 1.00000 \\ m_{A2H}^{S2} = 0.35800 \\ m_{A3H}^{S2} = 0.34000 \\ m_{A4H}^{S2} = 0.10000 \end{cases}$
<b>Etape6 :</b>	Définition des facteurs de fiabilité + combinaison des sources d'informations	<p><b>L'expert donne :</b>                  Pour la source <math>S_1</math> : <math>F=0.98</math>                  Pour la source <math>S_2</math> : <math>F=0.67</math>  <b>La combinaison des sources d'informations</b></p> $\begin{cases} m_{A1H}^{S1} = 0.85111 * 0.98 \\ m_{A2H}^{S1} = 0.14889 * 0.98 \\ m_{A3H}^{S1} = 0.00000 * 0.98 \\ m_{A4H}^{S1} = 0.00000 * 0.98 \end{cases}$ <p><b>donc</b></p> $\begin{cases} m_{A1H}^{S1} = 0.83408 \\ m_{A2H}^{S1} = 0.14591 \\ m_{A3H}^{S1} = 0.00000 \\ m_{A4H}^{S1} = 0.00000 \end{cases}$ <p><b>Pour la source <math>S_2</math></b></p> $\begin{cases} m_{A1H}^{S2} = 1.00000 * 0.67 \\ m_{A2H}^{S2} = 0.76333 * 0.67 \\ m_{A3H}^{S2} = 0.61000 * 0.67 \\ m_{A4H}^{S2} = 0.00000 * 0.67 \end{cases}$ <p><b>Donc</b></p> $\begin{cases} m_{A1H}^{S2} = 0.67000 \\ m_{A2H}^{S2} = 0.51143 \\ m_{A3H}^{S2} = 0.64908 \\ m_{A4H}^{S2} = 0.00000 \end{cases}$	/
<b>Etape 7 :</b>	Combinaison des critères par l'opérateur dempster-shafer	$\begin{cases} m_{A1H}^{S12} = 0.83408 * 0.67000 \\ m_{A2H}^{S12} = 0.14591 * 0.51143 \\ m_{A3H}^{S12} = 0.00000 * 0.64908 \\ m_{A4H}^{S12} = 0.00000 * 0.00000 \end{cases}$ <p><b>Donc</b></p> $\begin{cases} m_{A1H}^{S12} = 0.55883 \\ m_{A2H}^{S12} = 0.07462 \\ m_{A3H}^{S12} = 0.00000 \\ m_{A4H}^{S12} = 0.00000 \end{cases}$	$\begin{cases} m_{A1H}^{S12} = 0.88833 * 1.00000 \\ m_{A2H}^{S12} = 0.43333 * 0.35800 \\ m_{A3H}^{S12} = 0.10000 * 0.34000 \\ m_{A4H}^{S12} = 0.16666 * 0.10000 \end{cases}$ <p><b>Donc</b></p> $\begin{cases} m_{A1H}^{S12} = 0.88833 \\ m_{A2H}^{S12} = 0.15513 \\ m_{A3H}^{S12} = 0.03400 \\ m_{A4H}^{S12} = 0.01666 \end{cases}$

<b>Etape 8 :</b>	Normalisation des critères	$\begin{cases} m_{A1H}^{S12} = 0.55883/0.63345 \\ m_{A2H}^{S12} = 0.07462/0.63345 \\ m_{A3H}^{S12} = 0.00000/0.63345 \\ m_{A4H}^{S12} = 0.00000/0.63345 \end{cases}$ <p><b>Donc</b></p> $\begin{cases} m_{A1H}^{S12} = \mathbf{0.88220} \\ m_{A2H}^{S12} = 0.11779 \\ m_{A3H}^{S12} = 0.00000 \\ m_{A4H}^{S12} = 0.00000 \end{cases}$	$\begin{cases} m_{A1H}^{S12} = \frac{0.88833}{1.09412} \\ m_{A2H}^{S12} = \frac{0.15513}{1.09412} \\ m_{A3H}^{S12} = \frac{0.03400}{1.09412} \\ m_{A4H}^{S12} = \frac{0.01666}{1.09412} \end{cases}$ <p><b>Donc</b></p> $\begin{cases} m_{A1H}^{S12} = \mathbf{0.81191} \\ m_{A2H}^{S12} = 0.14178 \\ m_{A3H}^{S12} = 0.03107 \\ m_{A4H}^{S12} = 0.01522 \end{cases}$
<b>Etape 9 :</b>	La décision	On utilisant la probabilité pignistique, on trouve : $A_1$ est l'homologue de $O_1$ , avec une masse $m_{A1H}^{S12} = \mathbf{0.88220}$	On utilisant la probabilité pignistique, on trouve : $A_1$ est l'homologue de $O_1$ avec une masse $m_{A1H}^{S12} = \mathbf{0.81191}$

Tableau 7 : étude de cas : (Comparaison entre l'approche d'Ana-Maria & Notre approche)

#### 4. Discussion

Nous avons trouvé les mêmes résultats, en utilisant les deux approches, sauf que la notre est plus simple suite aux points suivants :

- L'utilisation d'un graphe plus simple regroupant les distances (l'écart de position P, M, L), ne demande pas l'expert comme celui d'Ana Maria et subdivisé sur 5 zones (parties), définissant un tableau de décision (H, IG, NH) c'est-à-dire à partir de l'écart de position on peut décider si l'objet est H, IG, ou NH, par contre Ana-Maria utilise directement les notions (H, IG, NH), ainsi l'ajout d'une étape qui consiste à définir et remplir d'une table de décision qui contient les différentes décisions possibles (H, IG, NH) à partir de ce tableau. Pour montrer les points forts de notre approche, nous donnons 4 exemples qui montrent la différence.
  - Le candidat  $A_1$  : l'écart de position est de 0.67U (soit 0.67m ou 0.67km ou autre unité de mesure), donne presque le même résultat avec l'approche d'Ana Maria, mais notre approche plus améliorée que la sienne car elle prend en considération l'ignorance avec 0.11167 comme vous voyez dans la figure précédente: ( $y_1=0.88833$  pour H,  $y_2=0.11167$  pour IG,  $y_3=0.00$  pour Non H), donc dans la réalité il n'y a pas de notion d'ignorance si la distance est plus proche selon la table de décision

- Le candidat  $A_2$  : l'écart de position est de 3.50U (soit 3.50m ou 3.50km ou autre unité de mesure). L'objet situé ni proche ni moyenne (c-à-dire situé dans la demi partie entre (P, M)) , à l'objet à apparier, l'approche de Ana Maria elle donne une masse importante au non homologue (0.33333) même que l'ignorance, mais la réalité montre que l'objet  $A_2$  n'est pas loin de la zone «Proche » et pas plus proche à la zone « Moyenne » : ( $y_1=0.4333$  pour H,  $y_2=0.4334$  pour IG,  $y_3=0.13333$  pour Non H), comme elle montre notre approche (donne 0 au Non ignorance)
- Le candidat  $A_3$  : l'écart de position est de 9U (soit 9m ou 9km ou autre unité de mesure), l'objet  $A_3$  est plus loin au O c'est-à-dire il n'y a pas une notion de « Homologue », Ana Maria donne une masse non nulle à cette hypothèse ( $y_1=0.1000$  pour H,  $y_2=0.1000$  pour IG) et  $y_3=0.8000$  pour Non H
- Si un candidat  $A_4$ , situé à un écart de position =5.5U, L'objet situé ni loin ni à distance moyenne (c-à-dire situé dans la demi partie entre (M, L)) , au l'objet à apparier, l'approche d' Ana Maria donne une masse importante au homologue (0.16666) même que l'ignorance (0.16666), mais la réalité montre que l'objet  $A_4$  n'est pas loin de la zone «Moyenne » et pas plus proche à la zone « Loin » : ( $y_1=0.16666$  pour H,  $y_2=0.16666$  pour IG,  $y_3=0.66667$  pour Non H), comme montrer notre approche (donne 0 au Homologue)
- Pour la distance toponymique, l'ancienne approche utilise l'algorithme de [40] , cette approche calcule un coefficient de similarité  $D_{sim}$  à partir de la matrice mot-mot, puis définit  $D_t$  (distance toponymique), par la formule  $D_t= 1-D_{sim}$ . Par contre, on peut utiliser une autre approche presque similaire de celle de Samal, et définir un nouvel l'algorithme détaillé et illustrée dans (Chapitre IV, section 2.2). Dans notre étude nous utilisons la distance toponymique, définie par notre approche.

## 5. Conclusion

Nous avons tenté de porter dans le processus d'appariement de données géographiques en proposant deux solutions : l'une concerne l'initialisation des fonctions de masses et l'autre le calcul de la distance topologique.

Afin de vérifier notre contribution, nous avons donnée des exemples tout en comparants notre méthode avec les travaux d'Ana-Maria [8].

## Conclusion & Perspectives

---

- **Conclusion & perspectives:**

Dans le cadre de notre thèse, nous avons implémenté une méthodologie qui nous permet d'initialiser les fonctions de masses utilisées en théorie de l'évidence via la notion de fuzzification utilisée en logique floue. Nous avons également proposer un algorithme pour le calcul de la distance toponymique, qui est un critère important dans le processus d'appariement des données géographiques.

Des résultats intéressants ont été trouvés et une comparaison avec des travaux récents ont été faits, ceci dit, le travail est loin d'être achevé, nous projetons ajouter d'autres critères pour l'appariement et implémenter notre approche sur plusieurs bases de données géographiques afin de fusionner et d'intégrer les données géographiques.

Une nouvelle utilisation d'autres opérateurs de combinaison de la théorie de l'évidence nous permette de comparer leurs efficacités, et par conséquent une prise de décision pertinente.



**Table des figures et tableaux**

<b>Figure 1</b> : Définition d'un SIG	<b>P13</b>
<b>Figure 2</b> : Les constituants d'un SIG	<b>P13</b>
<b>Figure 3</b> : Structure de l'information géographique	<b>P16</b>
<b>Figure 4</b> : Non-spécificité de la relation A « au nord de » B	<b>P22</b>
<b>Figure 5</b> : Topologie de l'incertitude selon Klir et Yuan, [30]	<b>P24</b>
<b>Figure 6</b> : Typologie de l'imperfection des données géographie selon Fisher [28]	<b>P25</b>
<b>Figure 7</b> : Taxonomie de l'imperfection de l'information géographique adaptée au Fisher,[28]	<b>P26</b>
<b>Figure 8</b> : Fondements de la logique floue	<b>P47</b>
<b>Figure 9</b> : Implémentation d'un système flou (Méthode d'inférence floue)	<b>P52</b>
<b>Figure 10</b> : Déffuzzification par la méthode du centre de gravité	<b>P54</b>
<b>Figure 11</b> : Calcul du centre de gravité (exemple)	<b>P56</b>
<b>Figure 12</b> : Raisonnement flou par la méthode d'inférence floue de Min-Max de Mamdani	<b>P57</b>
<b>Figure 13</b> : Approche sur théorie de l'évidence & la logique floue : application à l'appariement des données	<b>P63</b>
<b>Figure 14</b> : les différentes décisions possibles	<b>P64</b>
<b>Figure 15</b> : 4 <sup>ème</sup> étape « sélection des candidats »	<b>P65</b>
<b>Figure 16</b> : Le partage de distance en ensembles flous	<b>P66</b>
<b>Figure 17</b> : calcul de centre de gravité de la première partie du graphe	<b>P67</b>
<b>Figure 18</b> : calcul de centre de gravité de la deuxième partie du graphe	<b>P67</b>
<b>Figure 19</b> : calcul de centre de gravité de la troisième partie du graphe	<b>P68</b>
<b>Figure 20</b> : calcul de centre de gravité de la quatrième partie du graphe	<b>P68</b>
<b>Figure 21</b> : calcul de centre de gravité de la cinquième partie du graphe	<b>P69</b>
<b>Figure 22</b> : Méthode de calcul des masses(Méthode récapitulative)	<b>P70</b>
<b>Figure 23</b> : calcul du centre de gravité	<b>P70</b>

<b>Tableau 1</b> : Types et causes d'erreurs qui sont sources d'incertitudes dans une base de données géographiques — d'après [29].	<b>P20</b>
<b>Tableau 2</b> : les différents types d'opérations	<b>P50</b>
<b>Tableau 3</b> : calcul du centre de gravité	<b>P55</b>
<b>Tableau 4</b> : base de règles floues (Mamdani)	<b>P57</b>
<b>Tableau 5</b> : table de décision	<b>P65</b>
<b>Tableau 6</b> : la validation de la distance toponymique	<b>P75</b>
<b>Tableau 7</b> : étude de cas : (Comparaison entre l'approche d'Ana-Maria & Notre approche)	<b>P75</b>

## Bibliographie

- [1]. <http://logiciels-libres-cndp.ac-versailles.fr/>, 18/02/2012.
- [2]. [http://www.notre-planete.info/geographie/outils/sig\\_0.php](http://www.notre-planete.info/geographie/outils/sig_0.php)., 04/12/2011.
- [3]. <http://photogeo.insa-strasbourg.fr/cours/sig12.htm>., 04/12/2012
- [4]. [http://fr.wikipedia.org/w/index.php?title=Logique\\_floue&oldid=10899471](http://fr.wikipedia.org/w/index.php?title=Logique_floue&oldid=10899471)., 18/02/2012
- [5]. [http://fr.wikipedia.org/wiki/Aire\\_et\\_centre\\_de\\_masse\\_d'un\\_polygone.htm](http://fr.wikipedia.org/wiki/Aire_et_centre_de_masse_d'un_polygone.htm)., 18/02/2012
- [6]. <http://www.mrnf.gouv.qc.ca/territoire/portrait/portrait-donnees-mille.jsp>., 15 /08/2012.
- [7]. **ALOUANI Fouad**. *Commande par Logique Floue Appliquée aux pendules inversés et au Simulateur de Vol d'Hélicoptère : Simulation et Expérimentation*. 24/06/2006.
- [8]. **Ana-Maria OLTEANU**. *Fusion de connaissances imparfaites pour l'appariement de données géographiques , « Proposition d'une approche s'appuyant sur la théorie des fonctions de croyance »*. 24/10/2008.
- [9] **Appriau, A**. Multisensor signal processing in the framework of the theory of evidence. Application of Mathematical Signal Processing Techniques to Mission Systems, 1999.
- [10]. **Badard, T**. Extraction des mises a jour dans les BDG : de l'utilisation des méthodes d'appariement. Revue Internationale de Géomatique, 1998.
- [11]. **Badard, T. (2000)**. Propagation des mises à jour dans les bases de données géographiques multi-représentations par analyse des changements géographiques. Thèse de doctorat, Université de Marne la Vallée.
- [12]. **BADIN, Anne DROMIGNY**. *Fusion d'images par la théorie de l'évidence en vue d'applications médicales et industrielles*. 6 mai 1998.
- [13]. **Bouchon-Meunier, B**. «La logique floue», Edition Que sais -je?, Presses universitaire de France, [1993].
- [14]. **Bouchon-Meunier, B**. La logique floue et ses applications. Paris : Addison-Wesley France, 1995.
- [15]. **Bouziani, M**. Définition d'une méthode d'extraction des mises à jour de l'information spatiale dans un réseau routier en milieu urbain. In 2nd FIG Regional Conference. 2003.
- [16]. **Bracker H., Mouchot M.C.**, Classification multisources en télédétection par la théorie de Dempster - Shafer, Séminaire CNRS GdR 134 : Fusion en imagerie aérienne et satellitaire , Nice. 1994.
- [17]. **Bracker H**. Utilisation de la théorie de Dempster/Shafer pour la classification d'images satellitaires à l'aide de données multi-sources et multitemporelles. Thèse de l'université de Rennes I, 1996.
- [18]. **P. Brooker**. Development of a Multiresolutional Integrated Land Cover Database for Scotland and the Application of GIS in Predictive Spatial Modeling. In Proceedings of the ESRI User Conference, 1995.
- [19]. **Colot, O**. Systèmes de perception d'informations incertaines - Application au diagnostic médical, Mémoire de HDR, Université de Rouen, 2000.

- [20]. **Dempster, A.** *Upper and lower probabilities induced by multivalued mapping. Annals of Mathematical Statistics, 1967.*
- [21]. **Denoeux, T.** A k-nearest neighbour classification rule based on Dempster-Shafer theory. IEEE Transactions on Systems, Man and Cybernetics, 1995.
- [22]. **Devogele, T.** Processus d'intégration et d'appariement de bases de données Géographiques, Applications à une base de données routières multi-échelles, Thèse de Doctorat, Université de Versailles, 1997.
- [23]. Fondations pour une nouvelle théorie du raisonnement plausible et paradoxal. Application à la fusion d'informations incertaines et conflictuelles, Rapport Technique, ONERA, 2003.
- [24]. **Dromigny-Badin A. , Zhu Y.M., Magnin I.E., Revel D.** *Fusion of cine magnetic resonance and contrast-enhanced first -pass magnetic resonance data in patients with coronary artery disease : A feasibility study, Investigative Radiology , 1998.*
- [25]. **Dubois, D. et Prade, H.** Representation and combination of uncertainty with belief functions and possibility measures. Computer Intelligence, 1988.
- [26]. **El Najjar, M.E.** Localisation dynamique d'un véhicule sur une carte routière numérique pour l'assistance à la conduite. Thèse de doctorat, Université de Compiègne, 2003.
- [27]. **Essa Reem, Diab Youssef, Morand Denis.** *L'application de la théorie des possibilités dans l'évaluation de la qualité environnementale des bâtiments, 2006.*
- [28]. **Fisher, P.F., Comber, A. et Wadsworth, R.** *Nature de l'incertitude pour les données spatiales. In : Qualité de l'information géographique. Paris : Hermès et Lavoisier : s.n., 2005.*
- [29]. **Peter F.Fisher.** *Models of uncertainty in spatial data. Dans [Longley et al., 2005b], Second Edition. 2005.*
- [30]. **George J. KLIR et Bo YUAN.** *Fuzzy Sets and Fuzzy Logic : Theory and Application.* Prentice Hall PTR . 1995.
- [31]. **R. Laurini.** Raccordement géométrique de bases de données géographiques fédérées, Ingénierie des systèmes d'informations, vol.4, no. 3, 1996.
- [32]. **A. E. Lupien et W. H. Moreland.** *A general approach to map conflation. In Proceedings of the ACSM/ASPRS Annual Convention & Exposition, Technical Papers, AutoCarto 8, Baltimore, Maryland, USA, 1987.*
- [33]. **M. Lynch et A. Salford.** Conflation : Automated map compilation. A video game approach. In Proceedings of Auto Carto 7, Washington D.C, USA, 1985.
- [34]. **Mustière, S., Sheeren, D. et Gesbert, N.** Unications des bases de données géographiques : Recherches au laboratoire cogit de l'ign. Revue Géomatique Expert, 2004.
- [35]. **PLOIX M.A., GARNIER V., MOYSAN J.** *Théorie des possibilités appliquée à l'end du béton, 2010 .*
- [36]. **Xavier RODIER et Laure SALIGNY.** *Modélisation des objets urbains pour l'étude des dynamiques urbaines dans la longue durée. Dans SAGEO'07. Clermont-Ferrand, France : s.n., 2007.*

- [37]. **Christiane ROLLAND-May**. *Evaluation des territoires*. Hermes. 2000.
- [38]. **Rombaut, M.** Decision in multi-obstacle matching process using Dempster-Shafer's theory. In : Proceedings of Advances in Vehicle Control and Safety, Amiens, 1998.
- [39]. **Royère, C.** Contribution à la résolution du conflit dans le cadre de la théorie de l'évidence : Application à la perception et à la localisation des véhicules intelligents. Thèse de Doctorat, Université de Compiègne, 2002.
- [40]. **Samal A., Seth S., and Cueto K.**, "A feature based approach to conflation of geospatial sources". International Journal of Geographic Information Systems, vol. 18, n°5, 2004.
- [41]. **Samuel CORGNE**. Modélisation prédictive de l'occupation des sols en contexte agricole intensif : application à la couverture hivernale des sols en Bretagne, 2004.
- [42]. **Schorter, al**, Building an OOGIS prototype: experiments with Geo2. In Proceedings of ACSM/ASPRS Annual Convention and Exposition (AutoCarto12), Charlotte, North Carolina, USA, Volume 4, 1995.
- [43]. **Shafer, G. A.** *Mathematical Theory of Evidence*. Princeton : Princeton University Press. 1976.
- [44]. **Smets, P.** Constructing the pignistic probability function in a context of uncertainty. in Uncertainty in Artificial Intelligence5, Henrion M., Shachter R.D., Kanal L.N., Lemmer J.F. Eds., North Holland, Amsterdam, 1990
- [45]. **Smets, P.** Belief functions: The disjunctive rule of combination and the generalized bayesian theorem. International Journal of Approximate Reasoning, 1993.
- [46]. **Sudano J.** Inverse pignistic Probability Transforms. Proceedings of the 5th International Conference of Information Fusion, (Fusion 2002), Annapolis, MD, USA, pp 1277-2183, 2002.
- [47]. **Tebourbi Riadh**. « Cours SIG ». École Supérieure des communications de Tunis, Tunisie : s.n., 2007.
- [48]. **Walter, V. et Fritsch. D.** *Matching Spatial Data Sets: Statistical Approach*. International Journal of Geographical Information Science. 1999.
- [49]. **Yager R.** A general approach to decision making with evidential knowledge. In Uncertainty in Artificial Intelligence, Kanal L. et Lemmer J. Eds., Elsevier Science Publishers B.V., pp 317-327, 1986.
- [50]. **Yager, R.** On the Dempster-Shafer framework and new combination rules. Informations Sciences, 1987.
- [51]. **Zadeh, L.A.** *Fuzzy sets*. Information and Control, 1965.
- [52]. **Zadeh, L.A.** *Fuzzy sets as a basis for a theory of possibility*. Fuzzy sets and systems, 1978.

# Annexe1 : Dictionnaire du domaine routier

Selon la classification de « Ressources naturelles et Faune Québec »[6], et après [8] il y a une autre classification des routes, selon DBCARTO et DBMultiNet, nous avons construit un petit dictionnaire du domaine routier, en basant sur ces trois classifications, et on prend l'union de ces concepts. Ce dictionnaire sert à corriger les concepts du réseau routier définies par l'utilisateur, ou le producteur.

<b>Dictionnaire du domaine routier</b>	
<b>Type de la route</b>	<b>Abv</b>
1. autoroute	(AUT)
2. route nationale	(NAT)
3. route régionale	(REG)
4. route collectrice	(COL)
5. route locale	(LOC)
6. route d'accès aux ressources	(ACC)
7. autre chemin	(ACH)
8. Route Départementale	(RDP)
9. Vocation, ce type des routes contient	
a. Type autoroutier	(VTA)
b. Liaison principale	(VLP)
c. Liaison régionale	(VLR)
d. Liaison locale	(VLL)
e. Bretelle	(VBR)
10. route prioritaire	(RPR)
11. Route prioritaire moins importante qu'une autoroute	(RPM)
12. Autre route prioritaire	(ARP)
13. Route secondaire	(RSC)
14. Route locale d'importance élevée	(RLIE)
15. Route locale d'importance moins élevée	(RLME)
16. Autre route	(ATR)