



HAL
open science

Stockage de grands volumes de données géospatiales floues dans les bases de données NoSQL

B Khalfi, C. de Runz

► **To cite this version:**

B Khalfi, C. de Runz. Stockage de grands volumes de données géospatiales floues dans les bases de données NoSQL. LFA 2017 - RENCONTRES FRANCOPHONES SUR LA LOGIQUE FLOUE ET SES APPLICATIONS, 2017, Amiens, France. hal-02117097

HAL Id: hal-02117097

<https://hal.science/hal-02117097>

Submitted on 2 May 2019

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

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

Stockage de grands volumes de données géospatiales floues dans les bases de données NoSQL

Une approche guidée par le schéma pour la cohérence de données

B. Khalfi¹

C. de Runz^{1,2}

¹ Laboratoire LIASD, Université de Paris 8

² Laboratoire CReSTIC, Université de Reims Champagne-Ardenne

2 rue de la liberté 93526 Saint -Denis cedex, France, {khalfi, derunz}@ai.univ-paris8.fr
IUT de Reims, Chemin des Rouliers, CS30012 51687 REIMS CEDEX 2, cyril.de-runz@univ-reims.fr

Résumé :

De nos jours, de nombreuses applications exploitent les systèmes NoSQL pour le stockage des données géospatiales. Cependant, la question de la véracité des données dans ces systèmes se pose toujours. Si les données sont géospatiales et plus encore imprécises, des complexités supplémentaires apparaissent liées à leur syntaxe et leurs caractéristiques sémantiques. Notre proposition aborde certains problèmes de véracité dans les systèmes NoSQL document. Elle présente une approche guidée par le schéma permettant l'identification des critères à respecter pour valider un stockage cohérent d'entités géospatiales imprécises.

Mots-clés :

Données géospatiales floues, imprécision, cohérence, JSON, Fuzzy GeoJSON, systèmes NoSQL.

Abstract:

Nowadays, many applications use NoSQL systems for geospatial data storage. However, the question of the veracity of data in NoSQL systems still arises. If the data is geospatial and even more imprecise, additional complexities arise associated to their syntax and semantic characteristics. Our proposal addresses some veracity issues in NoSQL document systems. It presents a schema-guided approach allowing the identification of the criteria to validate a coherent storage of inaccurate geospatial entities.

Keywords:

Fuzzy geospatial data, imprecision, consistency, JSON, Fuzzy GeoJSON, NoSQL systems.

1 Introduction

Afin d'assurer une meilleure qualité des données géospatiales, il est important de considérer leur nature imparfaite, de l'étudier et donc de l'intégrer dans le processus d'analyse [4]. Les données géospatiales sont, par nature, imprécises [2]. Basées sur la théorie des ensembles flous [9] utilisée pour représenter les

connaissances imprécises, de nombreuses approches ont été proposées afin de modéliser ce type d'imperfection dans les bases de données [8].

Aujourd'hui, la croissance significative de données a conduit à l'apparition de nouveaux systèmes non relationnels appelés NoSQL. Ces systèmes s'attaquent aux limites de scalabilité et de performance des systèmes de gestion de bases de données relationnelles (SGBDR). Ils sont distribués dépassant les SGBDR dans la capacité de stockage, la montée en charge et l'hétérogénéité des données [1].

Désormais, les systèmes NoSQL peuvent être utilisés pour les données massives et géospatiales. En effet, plusieurs systèmes NoSQL ont été étendus pour pouvoir stocker de grandes masses de données spatiales, tels que les systèmes Cassandra et MongoDB.

Cependant, les systèmes NoSQL ne fournissent pas de cohérence stricte [7] : ils n'ont pas de restrictions sur le modèle de données et ne nécessitent pas de contraintes ACID (atomicité, cohérence, isolation et durabilité). Même si le concept *sans-schéma* est un atout, les systèmes NoSQL ont besoin de faire quelques suppositions sur la structure de données et d'ajouter plus de code pour les manipuler, ce qui implique un fort impact sur la qualité du traitement des données et questionne sur la qualité des données.

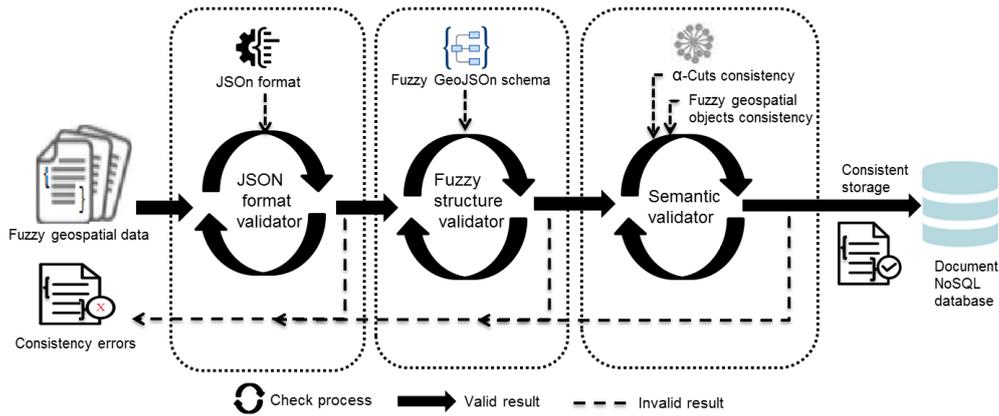


Figure 1 – Modèle du système proposé

Les données géospatiales imprécises, modélisées à l'aide d'ensembles flous, ont une structure complexe et une sémantique spécifique à l'instar des approches proposées pour les gérer dans les modèles relationnels [5, 10]. Nous cherchons à pouvoir les gérer dans des systèmes NoSQL, particulièrement ceux orientés document tout en assurant leur cohérence structurelle et sémantique.

Nous proposons un système de contrôle automatique afin de fournir un moyen efficace de stockage cohérent de données géospatiales floues dans les bases de données Document. La Figure 1 montre l'architecture du processus exploitant des validateurs du format, de la structure et de la sémantique et qui sont utilisés pour identifier des données géospatiales floues conformes et non conformes. Nous proposons un schéma logique dédié pour valider la structure et des modèles supplémentaires basés sur des contraintes sémantiques pour valider les caractéristiques sémantiques.

Dans cet article, nous n'étudions pas la cohérence des systèmes mais plutôt la cohérence des structures pour la représentation des données géospatiales floues. La démarche proposée permet d'améliorer la modélisation de données spatiales imprécises et contribue davantage à la qualité des données dans les systèmes NoSQL.

La plan de l'article est le suivant. Nous discuterons, dans la section 2, le cadre formel de représentation de données géospatiales imprécises. Nous présentons le schéma logique *Fuzzy GeoJSON* dans la section 3 et les contraintes sémantiques dans la section 4. Les résultats expérimentaux sont introduits dans la section 5. Enfin, les conclusions et perspectives sont exposées dans la section 6.

2 Données géospatiales imprécises : concepts et définitions

2.1 Définition et exemples

Les observations de l'espace révèlent la complexité et la variation de la réalité géographique. Contrairement aux espaces géographiques bien définis ayant des limites précises, nettes et linéaires, beaucoup d'autres situations dégagent des objets géospatiales avec des limites plus ou moins nettes et continues [2]. Ces caractéristiques présentent des objets géospatiales qui ont des limites larges, dont leur qualification et quantification ne passe pas nécessairement par des lignes nettes mais par des zones de transitions [3].

Un premier exemple de distribution spatiale floue est la densité du couvert d'arbres dans une forêt. Il s'agit de la difficulté à déterminer où commence une forêt. Entre un point d'intérieur

et un point de l'extérieur de la forêt, le territoire change de densité, cette dernière diminue jusqu'à ce qu'on ne retrouve plus d'arbres. Pour la représentation de ce territoire, des interrogations se posent : comment déterminer le seuil des arbres à considérer pour définir la limite de la forêt, ou encore comment représenter les arbres périphériques qui appartiennent partiellement à la forêt. Une limite nette n'évoque pas vraiment la transition de densité de la forêt avec son environnement extérieur. Elle modélise cette transition d'une manière drastique. Par conséquent, la valeur réelle des frontières peut être située à diverses positions. La forêt est un objet spatial à limites mal définies. En effet, pour tout modèle approximatif donné de la forêt, diverses limites peuvent être représentées.

Un deuxième exemple d'imprécision peut se présenter dans le cas d'une transition d'une zone urbaine à une zone rurale. Dans certains cas, cette transition est continue, de sorte qu'il n'y a pas de limite nette entre les deux zones. Ainsi, nous ne pouvons pas avoir une seule limite, mais plusieurs.

2.2 Théorie des ensembles flous pour les données géospatiales imprécises

Un objet géospatial imprécis présente un cœur où tous les éléments en son sein appartiennent pleinement au dit objet et des bordures qui lui appartiennent plus ou moins. Par conséquent, la question « Comment caractériser un territoire imprécis ? » est une question liée aux outils de représentation en termes de modèles mathématiques et de modèles informatiques.

La théorie des ensembles flous est une des solutions mathématiques définies pour représenter l'imprécision. Grâce à elle, il est possible d'exprimer une appartenance partielle d'une valeur à un ensemble. Alors, si E est un ensemble flou et e est un élément de E , la proposition « e est un membre de E » n'est pas nécessairement soit vraie soit fausse. Elle peut être vraie que dans une certaine mesure. E est caractérisée par une fonction d'appartenance μ_E prenant ses valeurs

dans $[0, 1]$.

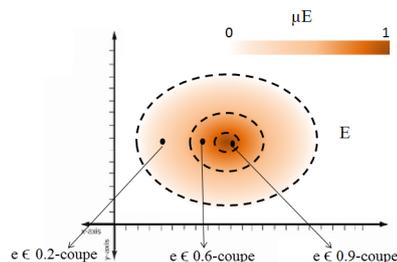


Figure 2 – Un ensemble flou E avec trois α -coupes

Pour stocker un ensemble flou, [10] propose de le discrétiser sur les valeurs de sa fonction d'appartenance. L'ensemble flou est ainsi présenté par un nombre fini d' α -coupes (voir Figure 2). Un objet géospatial flou est un objet imprécis défini par la théorie des ensembles flous comme suit :

$$O : \{ O_\alpha, \alpha \in [0, 1] \} \text{ et } \text{Card}(O) = n$$

L'objet géospatial flou est représenté par n α -coupes géographiques. Nous considérons O_α l'ensemble des valeurs du domaine ayant un degré d'appartenance supérieur ou égal à α ($\mu_E(e) \geq \alpha$); E_1 est le noyau de l'objet flou avec $\alpha = 1$ et E_0 - support de l'objet flou - est l'ensemble des éléments ayant un degré d'appartenance strictement supérieur à 0.

Projetant les principes de l'arithmétique floue sur les objets géospatiaux 2D flous, un objet géospatial flou cohérent doit être connexe et normalisé. Cela signifie que :

- les α -coupes géographiques sont des formes géométriques connexes, donc, $\forall \alpha$ avec α ($0 < \alpha \leq 1$), $\{O_\alpha\}$ est connexe ;
- le degré d'appartenance maximum doit être égal à 1, ce qui signifie que $\text{Max}_{O_\alpha \in O} (\mu(O_\alpha)) = 1$;
- quelle que soit l' α -coupe géographique O_1 de degré α_1 , toutes les α -coupes géographiques ayant un degré supérieur à α_1 doivent être incluses dans O_1 .

Pour la modélisation de l'espace flou, nous distinguons deux grandes catégories d'objets : simple (ou primitive) et composite. Les formes simples peuvent être de type *Point flou*, de type *Ligne floue* ou de type *Polygone flou*. La définition des α -coupes géographiques de ces objets suit le formalisme suivant :

$$G(O_\alpha) = \begin{cases} Point & \text{si } \alpha=1 \text{ et } O \text{ est un point flou} \\ Ligne & \text{si } \alpha=1 \text{ et } O \text{ est une ligne floue} \\ Polygone & \text{sinon} \end{cases}$$

Un objet point flou possède une α -coupe noyau de géométrie point et de degré =1 englobé d' α -coupes de géométrie polygone. Un objet ligne floue possède une α -coupe noyau de géométrie ligne et de degré =1 et le reste des α -coupes sont de géométrie polygone. Un objet polygone flou possède des α -coupes de géométrie polygone.

Chaque type de géométrie primitive est excellent pour représenter des espaces homogènes, mais souvent le monde réel est composé des espaces beaucoup plus complexes basés sur des formes composites [6]. Nous considérons, pour cette catégorie, deux familles : (1) les objets composites à géométries homogènes (une collection de forme unidimensionnelle) – *multi-Points flous*, *multi-Polygones flous* et *multi-Lignes floues* – et (2) les objets composites avec des structures hétérogènes telles que les géométries *alternatives*, *facultatifs* et *multiples*.

Nous proposons le schéma logique *Fuzzy GeoJSON* qui fournit la structure des objets géospatiaux flous à géométrie simple, à géométrie composite et la structure d'une collection d'objets géospatiaux flous.

3 Schéma Fuzzy GeoJSON

Le schéma Fuzzy GeoJSON est une extension du format GeoJSON. Il introduit des descriptions syntaxiques et des contraintes pour les données géospatiales floues.

La Figure 3 illustre la structure générale du schéma Fuzzy GeoJSON. La section *definitions* précise la syntaxe des géométries floues en se basant sur les définitions de structure des géométries primitives présentées dans la section *geometries*.

```

{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "title": "Fuzzy Geo JSON object",
  "description": "Schema for a fuzzy Geo JSON data",
  "type": "object",
  "oneOf": [
    { "$ref": "#/definitions/simpleFuzzyFeature" },
    { "$ref": "#/definitions/compositeFuzzyFeature" },
    { "$ref": "#/definitions/fuzzyFeatureCollection" }
  ],
  "definitions": {
    "geometries": {
      "type": "array",
      "items": {
        "$ref": "#/definitions/geometry"
      }
    }
  }
}

```

Figure 3 – Schéma Fuzzy GeoJSON

Le schéma Fuzzy GeoJSON fournit la classe *simpleFuzzyFeature* qui est la super-classe des objets géospatiaux flous à géométrie simple, la classe *compositeFuzzyFeature* qui est la super-classe des objets géospatiaux flous à géométrie composite et la classe *fuzzyFeatureCollection* qui désigne une collection d'objets géospatiaux flous.

Généralement, la structure d'un objet géospatial flou doit définir le type de l'objet, sa géométrie (forme et coordonnées) et la structure de l'ensemble de ses α -coupes géographiques. Dans ce qui suit, nous présentons la structure des trois types simples flous et nous nous limitons à deux types composites flous.

3.1 Point flou

La structure Fuzzy GeoJSON d'un point flou est illustrée par la Figure 4. Il s'agit d'un objet JSON avec deux propriétés : (i) un *type* qui doit prendre la valeur *FuzzyPoint* ("*type*" : "*enum*" : [*FuzzyPoint*]) et (ii) un ensemble d' α -coupes (*cuts*) qui est un tableau d'au minimum deux α -coupes (*minItems* = 2). Chaque α -coupe est de type *pointCut*. Le schéma considère les propriétés *type* et *cuts* comme obligatoires et ne supporte pas d'autres propriétés

supplémentaires pour décrire l'objet flou.

```

"fuzzyPoint": {
  "properties": {
    "type": { "enum": [ "FuzzyPoint" ] },
    "cuts": { "type": "array", "minItems": 2,
              "items": { "$ref": "#/definitions/pointCut" } } }
},
"pointCut": {
  "required": [ "type", "geometry", "properties" ],
  "properties": {
    "type": { "enum": [ "Feature" ] },
    "geometry": { "oneOf": [ { "$ref": "#/geometries/Polygon" },
                             { "$ref": "#/geometries/point" } ]
    },
    "properties": {
      "type": [ "object", "null" ],
      "required": [ "degree" ],
      "degree": { "type": "number", "minimum": 0, "maximum": 1,
                  "exclusiveMaximum": false, "exclusiveMinimum": false
    } } }
},

```

Figure 4 – Schéma d'un objet point flou

Un *pointCut* présente la structure d'une α -coupe appartenant à un objet de type point flou. Il s'agit d'un objet JSON qui requiert un type *Feature*, une géométrie (*geometry*) et un ensemble d'attributs. Particulièrement, chaque α -coupe doit avoir un degré d'appartenance (*degree*) à l'objet point flou. Selon la valeur du degré d'appartenance, la géométrie de l' α -coupe doit être de type point si le degré est égal à 1 sinon la géométrie doit être de type polygone.

3.2 Ligne flou

La structure d'un objet ligne floue doit avoir un type *FuzzyLineString* et un ensemble d' α -

```

"fuzzyLineString": {
  "required": [ "type", "cuts" ],
  "properties": {
    "type": { "enum": [ "FuzzyLine" ] },
    "cuts": {
      "type": "array", "minItems": 2,
      "items": { "$ref": "#/definitions/LineCut" }
    }
  }
},
"lineCut": {
  "required": [ "type", "geometry", "properties" ],
  "properties": {
    "type": { "enum": [ "Feature" ] },
    "geometry": {
      "oneOf": [
        { "$ref": "#/geometries/LineString" },
        { "$ref": "#/geometries/Polygon" }
      ]
    },
    "properties": {
      "required": [ "degree" ],
      "degree": {
        "type": "number", "minimum": 0, "maximum": 1,
        "exclusiveMaximum": false, "exclusiveMinimum": false
      }
    }
  }
},

```

Figure 5 – Schéma d'un objet ligne floue

coupes de type *lineCut*.

Selon la valeur du degré d'appartenance, la géométrie de l' α -coupe doit être de type ligne (*LineString*) si le degré est égal à 1 sinon la géométrie doit être de type polygone (cf. Figure 5).

3.3 Polygone flou

La structure d'un objet polygone flou doit avoir un type *FuzzyPolygon* et un ensemble d' α -coupes de type *polygonCut*. La géométrie des α -coupes est toujours de type *Polygone* (cf. Figure 6).

```

"FuzzyPolygon": {
  "required": [ "type", "cuts" ],
  "properties": {
    "type": { "enum": [ "FuzzyPolygon" ] },
    "cuts": {
      "type": "array", "minItems": 2,
      "items": { "$ref": "#/definitions/polygonCut" }
    }
  }
},
"polygonCut": {
  "required": [ "type", "geometry", "properties" ],
  "properties": {
    "type": { "enum": [ "Feature" ] },
    "geometry": { "$ref": "#/geometries/Polygon" },
    "properties": {
      "type": [ "object", "null" ],
      "required": [ "degree" ],
      "degree": { "type": "number", "minimum": 0, "maximum": 1,
                  "exclusiveMaximum": false, "exclusiveMinimum": false
    } } }
},

```

Figure 6 – Schéma d'un objet polygone flou

3.4 Géométrie alternative floue

La géométrie alternative floue introduit des classes d'objets qui peuvent être représentées par plusieurs représentations spatiales possibles, mais à chaque objet ne correspond qu'une seule géométrie floue. Par exemple, une zone verte peut être un parc urbain numérisé en tant que point ou peut être un parc régional numérisé par un polygone.

La Figure 7 illustre le schéma d'un objet alternatif flou ayant les trois dimensions. Il est défini par le type *FuzzyAlternativeGeometry*, un ensemble d' α -coupes et un ensemble de propriétés. Les α -coupes peuvent être de type *pointCuts*, *polygonCuts* ou *lineCuts*. La structure de chaque objet stocké dans la base de données doit valider un seul sous-schéma parmi

```

"fuzzyAlternativeGeometry":{
  "title": "FuzzyAlternativeGeometry",
  "description": "A alternative fuzzy Geo JSON object",
  "properties": {
    "type": { "enum": [ "FuzzyAlternativeGeometry" ] },
    "cuts": {
      "oneOf": [
        { "$ref": "#/definitions/pointCuts" },
        { "$ref": "#/definitions/polygonCuts" },
        { "$ref": "#/definitions/lineStringCuts" }
      ]
    }
  },
  "required": [ "type", "cuts" ]
},

```

Figure 7 – Schéma de la géométrie alternative floue

ces trois. Le mot-clé *oneOf* est utilisé pour valider les valeurs des α -coupes (*cuts*) contre l'un des sous-schémas donnés.

3.5 Géométrie agrégée floue

La géométrie agrégée décrit un objet spatial qui est une agrégation de géométries similaires. Par exemple, une route est composée de plusieurs segments de route. Le schéma Fuzzy GeoJSON distingue trois catégories d'agrégation : (1) *MultiFuzzyPoint* : une collection de géométries de type point flou, (2) *MultiFuzzyPolygon* : une collection de géométries de type polygone flou et (3) *MultiFuzzyLineString* : une collection de géométries de type ligne floue. Nous présentons, ci-après, l'exemple de *MultiFuzzyPoint* illustré par la Figure 8.

```

"multiFuzzyPoint": {
  "title": "MultiFuzzyPoint",
  "description": "A collection of fuzzy point",
  "required": [ "features" ],
  "properties": {
    "type": { "enum": [ "MultiFuzzyPoint" ] },
    "features": {
      "type": "array",
      "minItems": 2,
      "items": { "$ref": "#/definitions/fuzzyPoint" }
    }
  }
},

```

Figure 8 – Schéma de la géométrie MultiFuzzyPoint

Un objet *MultiFuzzyPoint* nécessite l'élément *Features* qui est un tableau d'objets de type *FuzzyPoint*. La contrainte ("minItems" : 2) indique

que chaque collection est composée d'au moins deux entités floues (deux points flous).

4 Les contraintes sémantiques

Le schéma Fuzzy GeoJSON définit les contraintes de structure au niveau logique mais il ne peut pas considérer les contraintes liées à la sémantique floue des objets. Étant donné un ensemble d'objets géospatiaux flous $O^E = \{O_1, \dots, O_n\}$ où chaque objet flou O possède une géométrie $G_O \in \{FuzzyPoint, FuzzyLineString, FuzzyPolygon\}$ et est composé d'un ensemble fini d' α -coupes $C^O = \{C_1, \dots, C_m\}$. Chaque α -coupe C est définie par le couple (G_C, α_C) où $G_C \in \{Point, Line, Polygon\}$ est la géométrie de l' α -coupe et $\alpha_C \in [0, 1]$ est le degré d'appartenance de l' α -coupe à l'objet flou O .

Dans cette section, nous donnons le formalisme utilisé pour valider les contraintes sémantiques relatives à la cohérence des α -coupes et à la cohérence des objets flous.

4.1 α -coupes géographiques cohérentes

Une α -coupe cohérente est mesurée en fonction de son degré d'appartenance et de la géométrie de l'objet flou auquel elle appartient. Une α -coupe cohérente $C \in C^O$, est définie par $(O, consistentCut^C)$ (cf. Algorithme 1).

Algorithme 1 Vérifier la cohérence des coupes

Requis : Initialiser avec O

$G_O \leftarrow$ la géométrie de chaque objet O

pour chaque C_i dans C^O **faire**

$G_{C_i} \leftarrow$ la géométrie de C_i

$\alpha_{C_i} \leftarrow$ le degré d'appartenance de C_i

si $consistentCut(G_O, G_{C_i}, \alpha_{C_i})$ est

FAUX alors

retourner message d'erreur et SORTIR

fin si

fin pour

retourner VRAI

O est l'objet flou associé et $consistentCut^C$ la

fonction qui vérifie que (1) pour un degré d'appartenance $\alpha_C = 1$, la géométrie de l' α -coupe (G_C) doit être la même que celle de l'objet O , et (2) pour un degré d'appartenance $\alpha_C \neq 1$, la géométrie de l' α -coupe G_C doit être de type *Polygon*.

4.2 Objets géospatiaux flous cohérents

Un objet flou cohérent doit avoir un ensemble connexe et normalisé d' α -coupes. Un objet flou connexe et normalisé $O \in O^E$, est défini par $(C^O, Norm^O, Inclus^O)$ (cf. Algorithme 2) où $C^O = \{C_1, \dots, C_m\}$ présente l'ensemble d' α -coupes cohérentes de l'objet flou O . La fonction $Norm^O$ vérifie que $\forall C \in C^O$ avec le degré d'appartenance α_c , $Max(\alpha_c) = 1$. Ainsi, l'objet flou O est de hauteur = 1 et est donc normalisé.

Algorithme 2 vérifier la cohérence de chaque objet

Requis : Initialiser avec O

si $norm(O) \neq 1$ **alors**
 retourner message d'erreur et SORTIR

sinon

pour chaque C_i dans C^O **faire**
 $G_{C_i} \leftarrow$ la géométrie de la coupe
 $\alpha_{C_i} \leftarrow$ le degré d'appartenance de la coupe
 $RC \leftarrow$ le reste des coupes de l'objet C^O

pour chaque C_k dans RC **faire**
 $G_{C_k} \leftarrow$ la géométrie de la coupe
 $\alpha_{C_k} \leftarrow$ le degré d'appartenance de la coupe
 si $Inclus((G_{C_i}, \alpha_{C_i}), (G_{C_k}, \alpha_{C_k}))$
 est FAUX **alors**
 retourner message d'erreur et
 SORTIR

fin si

fin pour

fin pour

fin si
 retourner VRAI

Pour chaque couple $\{C_i, C_k\}$ avec respective-

ment leurs types de géométrie et leurs degrés d'appartenance $\{(G_{C_i}, G_{C_k}), (\alpha_{C_i}, \alpha_{C_k})\}$, la fonction $Inclus^O$ vérifie la relation d'inclusion entre les α -coupes formant l'objet flou. La fonction teste si $(\alpha_{C_i} > \alpha_{C_k})$ alors $(G_{C_i} \subset G_{C_k})$. Si l'ensemble des α -coupes n'est pas normalisé ou si la contrainte d'inclusion est violée, un message d'erreur est signalé à l'utilisateur.

5 Résultats expérimentaux

Les principes de la méthodologie de validation sont implémentés à travers un prototype développé en Java. Dans la suite, des résultats d'évaluation sont donnés.

Log session start time Tue Nov 29 16:11:25 CET 2016

Time	Level	Category	Log Messages
0	INFO	com.validation.CheckSchema	-----Checking JSON Structure and Fuzzy Schema Check -----
6343	ERROR	com.validation.CheckSchema	Schema error: wrong structure or incomplete definition
6344	ERROR	com.validation.CheckSchema	Schema error: wrong structure or incomplete definition
6344	ERROR	com.validation.CheckSchema	#/features/0/cuts/0/geometry: # 0 subschemas matched instead of one

Figure 9 – Détection d'erreur de géométrie d' α -coupe géographique

Un premier exemple concerne la validation de la structure Fuzzy GeoJSON. On suppose qu'un polygone flou doit avoir des α -coupes géographiques forcément avec des géométries de type Polygone. Le fichier de journalisation (cf. Figure 9) montre une erreur liée à une incohérence de structure d'une α -coupe géographique.

Un deuxième exemple concerne la validation de la sémantique d'objets géospatiaux flous. Le fichier de journalisation, illustré par la Figure 10, affiche une erreur liée à la violation de la contrainte de normalisation.

Time	Level	Category	Log Messages
0	INFO	com.validation.CheckSchema	-----Checking JSON Structure and Fuzzy Schema Check -----
6276	INFO	com.validation.CheckSchema	-> The document 'semanticError(normalization) geojson' is conform to the JSON structure and to the 'FuzzyPolygonSchema.json' schema
6277	INFO	com.validation.CheckSchema	-----Checking Semantic Properties of Fuzzy Objects -----
6277	INFO	com.validation.CheckSchema	1-Checking Cut Consistency
6480	INFO	com.validation.CheckSchema	-> Each cut is consistent: membership degree and geometry type are valid
6482	INFO	com.validation.CheckSchema	2-Checking Fuzzy Objects Consistency
6638	ERROR	com.validation.CheckSchema	Normalization Violation for feature n° 18

Figure 10 – Détection d'erreur de normaliza-

6 Conclusion

L'approche et l'outil proposés contribuent à la qualité des données géospatiales floues dans un contexte de Big data. Nous avons présenté dans ce papier un moyen de modélisation floue qui considère à la fois les données géospatiales et leur imprécision ce qui n'est pas possible avec les approches classiques non floues. L'approche proposée assure un stockage cohérent des données géospatiales floues et le système développé reste scalable.

Cette méthodologie de validation est une première étape dans l'objectif de fournir un cadre complet de gestion du Big Geo Data flou en offrant d'autres possibilités d'interrogation, d'exploration et de visualisation.

Références

- [1] Vinayak R Borkar, Michael J Carey, and Chen Li. Big data platforms : what's next ? *ACM*, 19(1) :44–49, 2012.
- [2] Éric Desjardin, Bastien Lefebvre, and Cyril de Runz. Intégration de l'imperfection de l'information dans les dynamiques spatiales. définitions, outils et exemples. *Revue Internationale de Géomatique/International Journal of Geomatics and Spatial Analysis*, 25(3) :437.
- [3] Areti Dilo. Representation of and reasoning with vagueness in spatial information : A system for handling vague objects. *PhD thesis*, 2006.
- [4] Michael F Goodchild. Beyond metadata : Towards user-centric description of data quality. In *International Symposium on Spatial Data Quality (ISSDQ)*, pages 13–15, 2007.
- [5] Bisma Khalfi, Cyril de Runz, and Herman Akdag. Enhanced F-Perceptory approach for dealing with geographic data imprecision from the conceptual modeling to the fuzzy geographical database building. In Sami Faiz and khaoula Mahmoudi, editors, *Handbook of Research on Geographic Information Systems Applications and Advancements*, pages 238–267. IGI-Global, 2017.
- [6] Bisma Khalfi, Cyril de Runz, Sami Faiz, and Herman Akdag. A new fuzzy composite objects modeling with F-Perceptory for GIS under Modelio. *Journal of Transactions in GIS (TGIS)*, in press, 2017.
- [7] Songnian Li, Suzana Dragicevic, Francesc Anton Castro, Monika Sester, Stephan Winter, Arzu Coltekin, Christopher Pettit, Bin Jiang, James Haworth, Alfred Stein, and Tao Cheng. Geospatial big data handling theory and methods : A review and research challenges. *Journal of Photogrammetry and Remote Sensing (JPRS)*, 115 :119 – 133, 2016.
- [8] Markus Schneider. Fuzzy spatial data types for spatial uncertainty management in databases. In *Handbook of Research on Fuzzy Information Processing in Databases*, pages 490–515. Information Science Reference, 2008.
- [9] Lotfi A Zadeh. Information and control. *Fuzzy sets*, 8(3) :338–353, 1965.
- [10] Asma Zoghalmi, Cyril de Runz, and Herman Akdag. F-perceptory : an approach for handling fuzziness of spatiotemporal data in geographical databases. *International Journal of Spatial, Temporal and Multimedia Information Systems (IJST-MIS)*, 1(1) :30.