Qualitative topological relationships for objects with possibly vague shapes: implications on the specification of topological integrity constraints in transactional spatial databases and in spatial data warehouses

Lofti Bejaoui

N° d'ordre : D.U. 1924
EDSPIC : 434

Université Blaise Pascal - Clermont-Ferrand II
Ecole Doctorale Sciences Pour l'Ingénieur de Clermont-Ferrand

Université Laval, Québec, Canada
Faculté de Foresterie et de Géomatique

THESE EN COTUTELLE

présentée par

Lotfi BEJAOUI

pour obtenir le grade de

DOCTEUR D'UNIVERSITE

Spécialité : INFORMATIQUE

Qualitative topological relationships for objects with possibly vague shapes: implications on the specification of topological integrity constraints in transactional spatial databases and in spatial data warehouses

Soutenue publiquement le 25 Mai 2009 devant le jury :

Rapporteurs :	M. Robert LAURINI Mme Monica WACHOWICZ
Examinateurs :	M. Mir Abolfazel MOSTAFAVI (Président) M. Yvan BEDARD (Co-Directeur de thèse) M. François PINET (Co-Directeur de thèse) M. Michel SCHNEIDER (Co-Directeur de thèse)

LOTFI BEJAOUI

QUALITATIVE TOPOLOGICAL RELATIONSHIPS FOR OBJECTS WITH POSSIBLY VAGUE SHAPES: IMPLICATIONS ON THE SPECIFICATION OF TOPOLOGICAL INTEGRITY CONSTRAINTS IN TRANSACTIONAL SPATIAL DATABASES AND IN SPATIAL DATA WAREHOUSES

Thèse de doctorat présentée en cotutelle
à la Faculté des études supérieures de l'Université Laval, Québec
dans le cadre du programme de Sciences géomatiques
pour l'obtention du grade de Philosophiae Doctor (Ph.D.)

FACULTÉ DE FORESTERIE ET DE GÉOMATIQUE
UNIVERSITÉ LAVAL
QUÉBEC, CANADA

et

pour l'obtention du grade de Docteur en Informatique

UNIVERSITÉ BLAISE PASCAL
CLERMONT-FERRAND, FRANCE

2009

Résumé Court

Dans les bases de données spatiales actuellement mises en oeuvre, les phénomènes naturels sont généralement représentés par des géométries ayant des frontières bien délimitées. Une telle description de la réalité ignore le vague qui caractérise la forme de certains objets spatiaux (zones d'inondation, lacs, peuplements forestiers, etc.). La qualité des données enregistrées est donc dégradée du fait de ce décalage entre la réalité et sa description.

Cette thèse s'attaque à ce problème en proposant une nouvelle approche pour représenter des objets spatiaux ayant des formes vagues et caractériser leurs relations topologiques. Le modèle proposé, appelé *QMM model* (acronyme de Qualitative Min-Max model), utilise les notions d'extensions minimale et maximale pour représenter la partie incertaine d'un objet. Un ensemble d'adverbes permet d'exprimer la forme vague d'un objet (ex : a region with a *partially* broad boundary), ainsi que l'incertitude des relations topologiques entre deux objets (ex : *weakly* Contains, *fairly* Contains, etc.). Cette approche est moins fine que d'autres approches concurrentes (modélisation par sous-ensembles flous ou modélisation probabiliste). Mais elle ne nécessite pas un processus d'acquisition complexe des données. De plus elle est relativement simple à mettre en œuvre avec les systèmes existants de gestion de bases de données.

Cette approche est ensuite utilisée pour contrôler la qualité des données dans les bases de données spatiales et les entrepôts de données spatiales en spécifiant des contraintes d'intégrité par l'intermédiaire des concepts du modèle QMM. Une extension du langage de contraintes OCL (Object Constraint Language) a été étudiée pour spécifier des contraintes topologiques impliquant des objets ayant des formes vagues. Un logiciel existant (outil OCLtoSQL développé à l'Université de Dresden) a été étendu pour permettre la génération automatique du code SQL d'une contrainte lorsque la base de données est gérée par un système relationnel. Une expérimentation de cet outil a été réalisée avec une base de données utilisée pour la gestion des épandages agricoles. Pour cette application, l'approche et l'outil sont apparus très efficients.

Cette thèse comprend aussi une étude de l'intégration de bases de données spatiales hétérogènes lorsque les objets sont représentés avec le modèle QMM. Des résultats nouveaux ont été produits et des exemples d'application ont été explicités.

Résumé long

Les bases de données spatiales et les systèmes d'information géographique (SIG) sont de plus en plus utilisés pour répondre à des besoins transactionnels liés à la gestion des phénomènes du monde réel. De même, les cubes de données géo-décisionnelles sont devenus des outils incontournables qui permettent au preneur de décisions d'analyser l'extension spatiale d'un phénomène donné. Cette analyse est facilitée par la possibilité d'une navigation cartographique au niveau de la dimension spatiale du phénomène. Un point commun entre ces outils transactionnels et décisionnels consiste à représenter les phénomènes spatiaux en utilisant des géométries bien définies ou considérées comme telles. Une telle description simplifiée de la réalité *ignore* le vague de forme de certains objets comme des zones d'inondation ou des peuplements forestiers. Par exemple, une région *crisp* (ayant des frontières bien définies) ne peut être une représentation correcte d'un lac physiquement entouré par des frontières *partiellement* ou *complètement* larges; les berges du lac dépendent du niveau des précipitations). Il s'agit donc d'un problème de qualité puisque la fiabilité des données est dégradée par ce décalage entre la réalité et sa description.

Cette thèse propose une approche permettant de représenter des objets spatiaux ayant des formes vagues et de caractériser leurs relations topologiques. Plus spécifiquement, nous définissons un modèle qualitatif appelé *QMM model* (acronyme de Qualitative Min-Max model) qui utilise les notions d'extensions minimale et maximale pour représenter la partie incertaine d'un objet. Un ensemble d'adverbes permet alors d'exprimer le vague de forme des objets (ex : a region with a *partially* broad boundary, a line with a *completely* broad interior) ainsi que l'incertitude des relations topologiques (*weakly* Contains, *fairly* Contains, *strongly* Covers, etc.). Cette approche fournit une évaluation de l'incertitude moins fine que d'autres approches concurrentes (modélisation par sous-ensembles flous ou modélisation probabiliste) mais elle ne nécessite pas un processus d'acquisition complexe des données. De plus elle est relativement simple à mettre en œuvre avec les systèmes existants de gestion de bases de données.

Cette approche est ensuite utilisée pour contrôler la qualité des données dans les bases de données spatiales et les entrepôts de données spatiales en spécifiant des contraintes d'intégrité par l'intermédiaire des concepts du modèle QMM. Une extension du langage de contraintes OCL (Object Constraint Language) a été étudiée pour spécifier des contraintes topologiques impliquant des objets ayant des formes vagues. Plus précisément les expressions de

contraintes s'appuient sur une forme adverbiale d'où l'acronyme $AOCL_{OVS}$ (*Adverbial OCL for Objects with Vague Shapes*) pour caractériser cette extension d'OCL. Un logiciel existant (outil OCLtoSQL développé à l'Université de Dresde) a été étendu pour permettre la génération automatique du code SQL d'une contrainte lorsque la base de données est gérée par un système relationnel. Une expérimentation de cet outil a été réalisée avec une base de données utilisée pour la gestion des épandages agricoles. Pour cette application, l'approche et l'outil sont apparus très efficients.

Cette thèse comprend aussi une étude de l'intégration de bases de données spatiales hétérogènes lorsque les objets sont représentés avec le modèle QMM. Des résultats nouveaux ont été produits et des exemples d'application ont été explicités.

Avant Propos

Comme le veut la tradition, je vais tenter de satisfaire au difficile exercice de la rédaction de l'avant propos, une tâche que je considère parmi les plus ardues de ces années de thèse. Non qu'exprimer ma gratitude envers les personnes en qui j'ai trouvé un soutien soit contre ma nature, bien au contraire. La difficulté tient plutôt dans le fait de n'oublier personne. C'est pourquoi, je remercie par avance ceux dont le nom n'apparaît pas sur cette page et qui m'ont aidé d'une manière ou d'une autre. Ils se reconnaîtront.

Mon premier remerciement s'adresse à mes parents : Arbi et Beya qui m'ont permis de réaliser cette thèse grâce à leurs encouragements et leurs sacrifices tout au long de mes études. Je remercie aussi chaleureusement mes frères : Sami, Adel et Lassaad qui m'ont aidé à finir mes études dans de bonnes conditions grâce à leur support moral et financier. J'exprime aussi mes plus vifs remerciements à ma conjointe Kalthoum qui était très courageuse en supportant mon absence lors des mes séjours à Québec et qui a su me redonner de la confiance en moi dans les moments de doute.

Je voudrais aussi adresser un très grand merci à Yvan Bédard, François Pinet et Michel Schneider, mes *chefs*, qui ont su me laisser la liberté nécessaire à l'accomplissement de mes travaux, tout en y gardant un œil critique et avisé. Nos continuelles discussions ont sûrement été la clé de notre travail commun. Je les remercie également pour leur bonne humeur, leurs conseils ainsi que leur façon de diriger un travail de recherche dans un climat de respect mutuel et de bonne ambiance. Merci à vous!

Merci également au Dr. Mir Abolfazel Mostafavi d'avoir accepté être un prélecteur de cette thèse. J'éprouve un profond respect pour son travail et son parcours, ainsi que pour ses qualités humaines. Je remercie également Pr. Robert Laurini et Pr. Monica Wachowicz d'avoir accepté d'être les rapporteurs de cette thèse. Leur présence dans mon jury de thèse m'honore.

Je remercie *en vrac* les membres (passé et présent) du Département des sciences géomatiques de l'université Laval ainsi que ceux du Centre de Recherche en Géomatique. Je remercie particulièrement Carmen Couture pour sa disponibilité, son efficacité et sa bonne humeur. Je remercie également les professionnels de recherches et les post-doctorants (passé et présent) au CRG: Suzie, Marie-Josée, Eveline, Sonia, Martin, Nouri (post-doctorant) et Fred (ex post-doctorant et actuellement professeur au département des sciences géomatiques à l'université Laval). Merci pour votre aide! Et un très gros merci à mes deux amies Tarek et Mehrdad pour les diverses discussions sur différents sujets et les bons moments qu'on a vécu ensemble. Merci à vous!

Merci aux membres de l'équipe COPAIN de l'UMR TETIS du Cemagref de Clermont-Ferrand pour leur accueil, leur soutien et leur confiance. Merci à Jean-Pierre, Vincent Abt, Vincent Soulignac, Stephan, Julie, Eliane, Christiane. Je veux remercier particulièrement mon amie Marie-Angélina Magne pour son soutien, son amitié et ses excellents conseils. Merci à chacun!

Le financement étant un point crucial dans une thèse, je remercie les différents organismes ayant contribués au financement de cette thèse! Merci donc au CRSNG et aux 9 partenaires de la chaire de recherche industrielle en bases de données géospatiales décisionnelles, au CEMAGREF, à la Faculté de foresterie et de géomatique de l'université Laval, au Bureau de l'Aide Financière de l'université Laval ainsi qu'à la Mission Universitaire de la Tunisie en Amérique du Nord.

Enfin, un très gros merci à mes amis Tarek (pour la deuxième fois), Nidhal, Kamel, Anis, Sonia, Ali… pour leurs encouragements ainsi que pour les innombrable agréables moments que nous avons passé ensemble. Merci à tous!

Quatre articles composent les chapitres 3, 4, 5 et 6 de cette thèse dont je suis l'auteur. Pour les quatre articles, j'ai réalisé l'ensemble de la recherche et ai rédigé les manuscrits. De ce fait, j'en suis l'auteur principal. Le Dr Yvan Bédard, le Dr Michel Schneider et le Dr François Pinet ont contribué aux articles par la revue des manuscrits et par l'apport de leurs commentaires judicieux. Le premier article intitulé «*Qualified topological relationships between objects with possibly vague shapes*» va apparaître bientôt, en version papier, dans la revue *International Journal of Geographical information Science* (il est disponible en version électronique sur le site http://www.informaworld.com/smpp/content~content=a902649624~db=all~order=pubdate, depuis Sepembre 2008). Le deuxième article, « *Qualitative Min-Max model for lines with vague shapes and their topological relations* », a été soumis à la revue *Transaction in GIS* (*TGIS*). Le troisième article, «*Reducing the vagueness of topological relationships in spatial data integration*», a été soumis à la revue *International Journal of Geographical Information Science*. Enfin, le dernier article intitulé «*An adverbial approach for the formal specification of topological integrity constraints involving regions with broad boundaries*» a été publié dans une série de *Lecture Notes in Computer Sciences* (Volume 5231/2008, pp. 383-396) qui correspond au *proceeding* de la 27ème édition de la conférence Conceptual Modeling - *ER* 2008. Cet article a été également pré-sélectionné pour être publié dans un numéro spécial de la revue Geoinfromatica. Une version étendue a été alors soumise pour répondre à cette sollicitation. Dans cette version étendue, François Pinet a rédigé des paragraphes dans les sections 3, 4 et 5 ainsi que dans la conclusion.

A mes parents
A mes frères et mon épouse
A mes amis

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

Chapter 1: Introduction

1.1 Research context

Over the last two decades, Geographical Information Systems (GIS) and spatial databases have been increasingly used to meet some transactional and decisional needs in various areas. The rise of GIS and spatial databases has been stimulated by the technological advances and an increasing relevance of multi-source integrated spatial information in the management of phenomena such as forestry, geology, agriculture, disaster control and emergency management, land cover/land use planning, national defence and security, etc. The increasing use of GIS leads to increasing requirements about presenting a reliable description of geographic information. Such a description should always consider the imperfection that is an endemic feature of the geographic information (Goodchild 1995a, Duckham et *al.* 2001). The imperfection can be present, in the description of a spatial object, in different forms including vagueness (e.g. Erwig and Schneider 1997), error (e.g. Heuvelink 1998), imprecision (e.g. Worboys 1998(b)), inconsistency (e.g. Rodriguez 2005), etc.

Dealing with imperfection is generally based on general taxonomies that propose definitions of its different types and causes (Bédard 1987, Smithson 1989, Parsons 1996, Smets 1996, Goodchild and Jeansoulin 1998, Fisher 1999a, Worboys 1998a, Hazarika and Cohn 2001, Devillers and Jeansoulin 2005). The first aim of such taxonomies is in distinguishing the nuances between the imperfection types, rather than accurately characterising the nature of imperfection (Parsons and Hunter 1998). According to Dilo (2006), these taxonomies has led to the development of different formalisms, each intended to capture a particular nuance of imperfection. The definitions of imperfection types and the nuances between them are explained in details in the literature review (Chapter 2).

The inherent imperfection of geographic information leads to deficiencies in spatial data quality (Guptill and Morrison 1995, Goodchild and Jeansoulin 1998, Aalders 2002, Devillers 2004, Devillers and Jeansoulin 2005, Van Oort 2006). The '*quality*' can be defined as "the totality of *features* and *characteristics* of a product or service that *bear* on its ability to satisfy stated or implied needs" (ISO 2002, originally in ISO standard 8402). Spatial data quality is generally described by a set of elements such as the positional accuracy and genealogy called the *elements of spatial data quality* (Guptill and Morrison 1995). The description of such elements is made by the data producer and helps the users to determine if the available data meet their needs. Moreover, the information about spatial data quality is increasingly required by users of transactional spatial databases and spatial data warehouses (Devillers 2004). In the latter case, it became the first criterion needed because the relevance of a decision depends strongly on the quality of data loaded in the data warehouse (Knightbridge Solutions 2006).

Spatial data quality may also be degraded when inappropriate spatial models are used to describe the geographic reality (Dilo 2006). For instance, the *traditional* (this term is used in the remainder of the thesis to refer crisp spatial models) spatial models assume that the geographic reality is certain, crisp, unambiguous and independent of context. (Duckham et *al.* 2001). Then, natural phenomena such as an earthquake or an inundation are represented using crisp spatial objects; although they include inherent shape vagueness (e.g., broad boundaries separate the different disaster areas). This simplification of geographic reality decreases the reliability of its description because a relevant property of spatial objects is lost (Tang 2004) (i.e. their inherent shape vagueness). According to Clementini and Di Felice (1997), this mismatch between the geographic modeling and the complex geographic reality presents a big limitation of traditional spatial models. It entails a gap between the spatial reality and its description in spatial databases and GIS. Consequently, the users cannot have knowledge about the uncertainty of the spatial objects and of their relationships. They may miss-interpret the available data and make wrong decisions.

Furthermore, the traditional spatial models do not always meet the modeling needs in a spatial integration process especially when crisp source geometries are used to represent vague concepts in the source databases. The spatial data integration aims to make heterogeneous geometries compatible with each other in a final database, so that they can be displayed on the same map and their relationships can be analysed (Shepherd 1992, Devogel

1997). Spatial data integration is a complex problem that can be defined, addressed and resolved differently according to different needs. In this context, we are interested in a *vertical* integration (Poulliot 2005) where the same objects are represented by heterogeneous and redundant crisp geometries in different sources with different specifications. Then, the final geometries resulting from the latter integration process (ex. displacement, rubber sheeting, size modification, distortion) may be plagued by increased vagueness and then the traditional spatial models are not able to reliably represent them. For example, a forest stand is a vague concept that may be falsely represented by heterogeneous and redundant crisp geometries in different data sources with different specifications, each intended to represent a different interpretation of an aerial photo that represents the object (De Groeve et *al.* 2000). When such crisp source geometries have similar qualities, a better final geometry is obtained by considering all of them (Devogel 1997). In the example of forest stands, a region with a broad boundary is then generated from the integration. The broad boundary refers to the difference between the union and intersection of crisp source polygons and reflects the disaccord between the experts in the interpretation of aerial photos. If considered in the same context, the final geometry should then more reliable than those representing the same object in the data sources because the shape vagueness is now explicitly represented. Figure 1.1 shows a spatial object A represented by three crisp heterogeneous and redundant polygons P_1, P_2 and P_3 in three different source databases S_1, S_2 and S_3. The final geometry of A is a region with a broad boundary obtained by merging P_1, P_2 and P_3. The intersection of P_1, P_2 and P_3 is the kernel or the certain part (*the black sub-region* in Figure 1.1) of the final geometry R. The broad boundary (*the grey part of R* in Figure 1.1) of R corresponds to the difference between the intersection and union of P_1, P_2 and P_3.

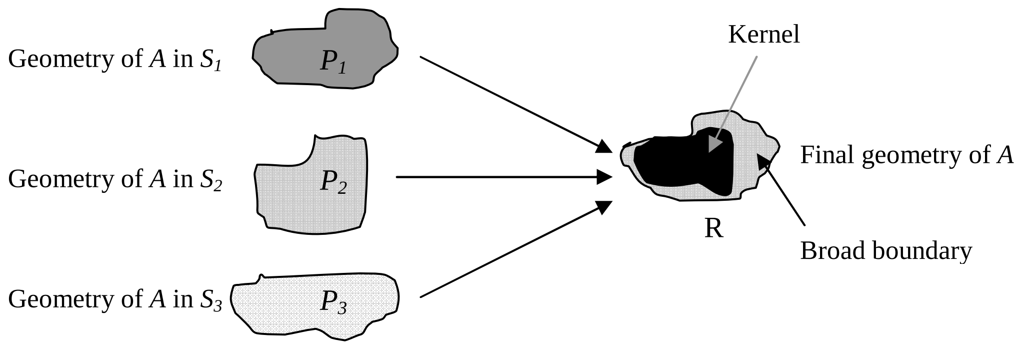


Figure 1.1 Example of a region with a broad boundary resulting from the integration of redundant and heterogeneous source polygons

In spatial modeling, the importance of topological relationships such as *Overlap* or *Contains* is widely recognised (Clementini and Di Felice 1997). These relationships are preserved under continuous geometric transformations (e.g. rotation, scaling, translation).

Several spatial models studied the topological relationships between objects with crisp shapes (Egenhofer and Herring 1990, Egenhofer and Franzosa 1991, Mark and Egenhofer 1994, Cohn et *al.* 1997). In GIS applications, these models (called traditional in the remainder of the thesis) provide the theoretical bases for the spatial reasoning and computation of topological relationships involved in the spatial queries and in topological integrity constraints (Clementini and Di Felice 1997). Nonetheless, the traditional spatial models do not describe the shape vagueness of spatial objects that exist in the geographic reality as well as that resulted from a vertical integration (see above). Existing approaches such as Burrough (1996), Erwig and Schneider (1997), Zhan (1997), Clementini and Di Felice (1997), Worboys (1998b), Roy and Stell (2001), Schneider (2001), Morris (2003), Tang (2004), Pfoser and Tryfona (2005), Pfoser et *al.* 2005, Dilo (2006) and Reis et *al.* (2006) proposed methods to represent spatial objects with vague shapes and to compute their topological relationships. In these proposals, the problem of shape vagueness is generally addressed without studying the possibilities of expressing the topological integrity constraints involving spatial objects with vague shapes. The specification of such topological integrity constraints cannot be based on traditional spatial models and remains unexplored. For example, an integrity constraint controlling a topological relationship between two regions with broad boundaries such as geopolitical conflict zones should consider the case where the relationship is *partially* respected (e.g. *weakly overlap*, *fairly inside*.). Such a specification is not available in traditional approaches. To overcome this limitation, one can suggest reusing existing approaches that deal with objects with vague shapes. However, these approaches have some limits (presented in the next section) that make difficult their use to specify topological constrains involving objects with vague shapes.

1.2 Problem statement

Shape vagueness is a type of imperfection arising when there is an uncertainty to sharply distinguish an object shape from its neighbourhood. This imperfection concerns the presence of broad boundaries for regions (Burrough and Frank 1996), broad endpoints and/or interiors for lines (Clementini 2005) and broad interiors for points (Santos and Moreira 2007). For instance, some spatial objects such as a *lake* or a *forest stand* are delimited in real life by broad boundaries rather than crisp ones. Likewise, when mapping the vegetation, the transition from one class to another may be gradual. It may be difficult to decide whether a location belongs to one vegetation class or another (Dilo 2006). Dealing with spatial objects with vague shapes is also recurrent in decisional applications such as the evaluation of the risk of fire in the Sydney Olympic Park (Zeng et *al.* 2003) or the management of data about the environmental phenomena in the forests of central Africa (FAO 2001). In this same context, Groeve et *al.* (2000) proposed a method to represent a forest stand as a region with a broad boundary by merging its different representations. The shape vagueness of a spatial object can also be caused by the ignorance. For example, one might have a vague idea about the spatial extent of an oil deposit; i.e. additional information could reduce this vagueness (Cohn and Gotts 1996a) but is not available. Thus, the shape vagueness concerns the spatial extents of spatial objects in various geographic applications.

Several approaches investigated the importance and possibility to handle the spatial objects with vague shapes (Burrough 1996, Erwig and Schneider 1997, Zhan 1997, Clementini and Di Felice 1997, Worboys 1998b, Roy and Stell 2001, Schneider 2001, Morris 2003, Tang 2004, Pfoser and Tryfona 2005, Pfoser et *al.* 2005, Dilo 2006, Reis et *al.* 2006). These approaches can be categorized in two main groups: (2) the models based on mathematical theories such as Fuzzy Logic (Zadeh 1965) and (1) the qualitative or exact models. The principles of each model category and their differences are explained in details in chapters 2 and 3. In this section, we just introduce the different categories and enumerate some of their limits in order to justify the problems addressed in the thesis.

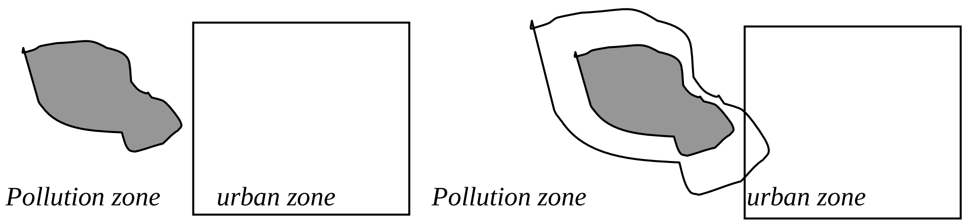
For the first category of models, fuzzy logic is the most often used theory (Dilo 2006). The fuzzy approaches such as (Robinson and Thongs (1986) Altman 1987, Burrough 1989, Brown 1998, Schneider 2001, Tang 2004, Hwang and Thill 2005, Dilo 2006) allow a finite quantification of the vagueness of spatial objects and of their topological relationships. The fuzzy approaches are better adapted to raster data where the vagueness levels are shown by computing the membership degree of each pixel to the object class involved, i.e. these approaches support a field-oriented view of the geographic reality. However, the hardest problem of fuzzy approaches is to define the membership functions intended to compute the shape vagueness inside the geometry of a given object. The definition of such functions is based on quantitative hypotheses that are also difficult to be set (Clementini 2005). It is also problematic to combine different membership functions to compute the shape vagueness inside a same object, where different factors entail the vagueness (Godjjevac 1999). Moreover, the current computational technology does not allow efficient processing to define and manage probabilistic and fuzzy models (other limits of this category of models are presented in the literature review (Chapter 2)). According to Erwig and Schneider (1997), the qualitative approaches refer to a pertinent alternative to represent shape vagueness.

The qualitative or exact[1] approaches such as (Cohn and Gotts 1996, Erwig and Schneider 1997, Clementini and Di Felice 1997, Clementini 2005) represent the spatial objects with vague shapes by extending the traditional spatial models. The advantage of these approaches is that existing definitions, techniques, data structures, algorithms, etc., do not need to be redeveloped but only modified and extended, or simply used (Erwig and Schneider 1997). For example, Cohn and Gotts (1996) proposed the Egg-Yolk model that extends the RCC model (Randell and Cohn 1989). In the Egg-Yolk model, a region is composed by a core (the yolk) that is surrounded by a broad boundary that partially belongs to the region. With regards to fuzzy approaches, Egg-Yolk model does not allow computing of the membership degree of a given point inside the broad boundary. However, such a model provides a representation of vagueness notion while retaining the simplicity of using traditional spatial models. Furthermore, quantitative hypotheses are not required to represent shape vagueness using a qualitative approach. Nevertheless, exiting qualitative approaches do not consider the case of spatial objects with *partial* vague shapes. For example, a region with a partial broad boundary (e.g. a lake with swamp banks on one side and rocky banks on the other side) cannot be represented using existing approaches since a broad boundary is defined as a connected and closed area that surrounds the region's core (this definition is not respected if the lake's boundary is linear in some locations and broad in some others). In the same way, a line can be partially vague when only one endpoint is broad or when the interior is partially broad. Also the latter cases are not supported by existing approaches. Other limits of existing qualitative models are discussed in the literature review.

Based on the limits of existing approaches dealing with shape vagueness, a new qualitative spatial is required to cover the different cases of spatial objects with different levels of shape vagueness. Such a model is necessary to control the topological consistency of spatial databases supporting this type of objects. According to Frank (2001), the consistency of vague data should be controlled through specific constraints which *tolerate a partial satisfaction* of the defined rules. Nonetheless, the principal approaches dealing with the specification of topological integrity constraints are based on traditional spatial models such as the 9-intersection model (Egenhofer and Herring 1991), the CBM approach (Clementini and Di Felice 1995) and the RCC theory (Randell and Cohn 1989, Cohn et *al*. 1997) that

[1] The terms qualitative and exact are used interchangeably along the thesis

ignore the shape vagueness. These approaches (Cockcroft 1997, Normand 1999, Servigne et *al.* 2000, Duboisset 2007) are also based on a binary logic to evaluate whether a topological relationship is respected or not. They do not consider the case where a topological relationship is *partially* (e.g. *weakly*, *fairly*, *strongly*, etc.) verified. Such a case is recurrent in the relationships involving objects with vague shapes. For example, let the topological integrity constraint *TC1*: *"a pollution zone should* ***weakly overlap*** *an urban zone"*. Figure 1.2 shows two representations of the spatial objects involved in the constraint *TC1*. In the first case (Figure 1.2(a)), the pollution zone is represented as a crisp polygon. The spatial objects are disjoint and therefore the first representation does not satisfy the topological integrity constraint presented above. In the second case (Figure 1.2(b)), the pollution zone is represented as a region with a broad boundary that partly overlaps the urban zone. Since the broad boundary is an uncertain part of the pollution zone, it is possible to associate the adverb *weakly* to the overlap relation. If *TC1* is specified using a traditional approach, the expression '*weakly overlap*' should be replaced by '*disjoint or meet*' in order to accept the *crisp* configurations. The first configuration is then accepted while it is not reliable (the broad boundary of the pollution zone is ignored). However, the second configuration is not valid because the vague shapes are not supported by the approach used to define the integrity constraint. More specifically, the partial satisfaction of the overlap relation cannot be tested since the term '*weakly*' is not supported. The term '*weakly*' requires the representation of the broad boundary of the pollution zone (i.e. the existence of such a boundary can be used to justify that the overlap relation is *weak*, otherwise the relation is *true* or *false*).



(a) crisp representation of the pollution zone (b) vague representation of the pollution zone

Figure 1.2 Two different representations of a pollution zone and of the resulting differences regarding its topological relationship with an urban zon

The general problem addressed in this thesis is:

Insufficiencies of existing approaches regarding the specification of topological integrity constraints involving spatial objects with vague shapes and their topological relationships, both in transactional spatial databases and in spatial data warehouses.

for objects with vague shapes in a spatial database environment) and a decisional one (*topological relationships between geometries with vague shape in a spatial dimension of a spatial data warehouse*). These axes are related since the geometries stored in transactional databases may be integrated and loaded into a spatial data warehouse through what is known as an ETL process (Extract-Transform-Load). In this case, shape vagueness may result from spatial data integration when heterogeneous crisp geometries (representing the same object in different data sources) are merged in order to produce a final geometry (with its vagueness) that represents a given spatial object in the data warehouse (see the example in Section 1.1). Existing exact spatial models generally study shape vagueness as an imperfection that characterises some natural objects. In this work, we show that shape vagueness can also result from integration and causes some difficulties in the final databases. Among these difficulties, we only concerned with the specification of topological relationships between geometries with vague shape in the final databases (see the second specific problem).

The general problem presented above is decomposed into three specific ones:

- ***Insufficiencies of existing exact models regarding the representation of spatial objects with different levels of shape vagueness (i.e. partial shape vagueness, complete shape vagueness) and the specification of their topological relationships***

  The literature review presented in section 2.3 shows that most of existing exact models do not model spatial objects with *partially vague* shapes such as a region with a partially broad boundary (i.e. a boundary that is crisp in certain areas and broad in other areas) or a line with one broad endpoint and one crisp endpoint. For example, a lake may be surrounded by crisp rocky banks on one side and swamp banks on the other side. Likewise, the itinerary of a XVth century explorer can be sharply known in some locations and only broadly known in some others. Most of existing works evaluate the shape vagueness through a binary logic that considers an object as *vague* or *not vague* (crisp). However, the geographic reality is more complex and an object may be partially vague, i.e. it may include vagueness and crispness at the same time in different parts of the boundary. Accordingly, there is today no exact approach to evaluate the vagueness of topological relationships that occur between objects with different levels of shape vagueness.

- ***Problem of topological relationships vagueness for geometries with vague shapes resulting from the integration of heterogeneous and redundant crisp geometries of a same object***

  A spatial data warehouse is generally loaded from several data sources that are heterogeneous on several levels. In this work, we are interested in considering the geometrical heterogeneities between geometries representing the same object at the same epoch in different sources in order to better know this object and its vagueness. Accordingly, these geometries should be merged before being loaded in the spatial data warehouse as they represent a same object in the reality. The final geometry may be vague if it is generated from heterogeneous crisp geometries that have a similar quality level. In this case, the integration process requires a method to identify the appropriate topological relationships between the final geometries. These topological relationships should consider the shape vagueness because they cannot be identified to those defined in the data sources. Consequently, there is a problem of *topological relationships vagueness* that we define as the uncertainty about the valid topological relationships for geometries with vague shapes loaded into the final database. In Figure 1.3, an example of a vertical integration of redundant crisp geometries is presented to illustrate the problem of topological relationships vagueness. In this example, two spatial objects O_1 and O_2 are represented using heterogeneous crisp geometries in two different data sources S_1 and S_2. Regions with broad boundaries are then resulted from the integration of available geometries of O_1 and O_2. The broad boundaries refer to the difference between the intersection and union of source geometries of the object involved. In this context, the topological relationship defined in the sources (i.e. *Disjoint* in our example) between geometries of O_1 and O_2 can be just partially respected by final geometries with vague shapes. Even though one chooses to ignore the shape vagueness by crisping (e.g. choose the unions, intersections, union/intersection or intersection/union as crisp geometries of O_1 and O_2 in the final database) the final geometries, the problem remains since other relationships are also possible (e.g. *Meet* is also possible in our example).

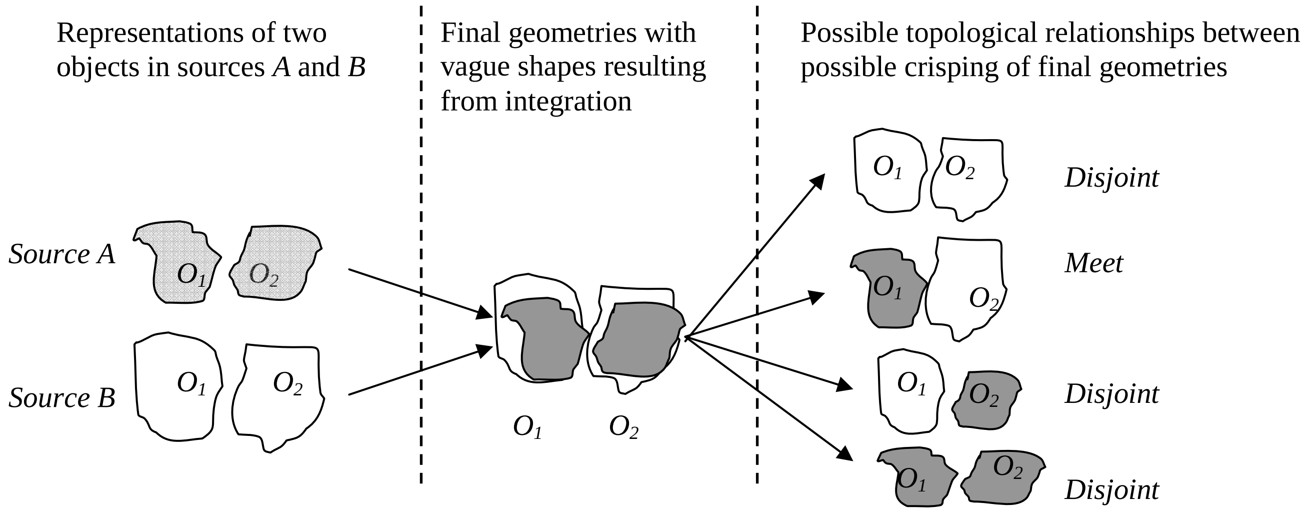


Figure 1.3. Example of topological relationships vagueness in a vertical integration of redundant crisp geometries

The *topological relationships vagueness* concerns the relationships between geometries representing the members of one hierarchy level of a spatial dimension as well as those between the geometries belonging to its different hierarchy levels. For example, let the spatial dimension of a spatial data warehouse (intended to analyze the distribution of taxes) defined by the following hierarchy (*building*, *county*, *state*, *region*, *nation*). If the geographic union of points representing the buildings (commercial, residential and industrial) is not within the spatial extent of their county, every individual building should be analyzed to determine how the required taxes should be distributed between two or more counties[2]. In this thesis, we deal only with the *intra-level* topological relationships vagueness. We are conscious that *inter-levels* topological relationships are also very important since the shape vagueness should be considered to correctly compute the aggregations of fact measures. This latter aspect exceeds the objectives of this thesis and requires additional investigations that will be made in future researches.

- ***Inadequacy of existing approaches regarding the formal specification of integrity constraints involving objects with vague shapes***

Several approaches (see section 2.4) handle the specification of integrity constraints in spatial databases. Generally, the shape vagueness is not considered, neither in the geometric representations of some spatial objects nor during the specification of their topological integrity constraints (see example in figure 1.2).

[2] This example is adapted from another one presented in (Malinowski and Zimányi 2005).

The problem of formally expressing the integrity constraints involving spatial objects with vague shapes remains, to our knowledge, always unexplored. There exists an extension of the *Object Constraint Language* (*OCL* for short) that allows the modeling of topological integrity constraints involving spatial objects represented by crisp shapes (Pinet et *al.* 2007). This method allows generating SQL code from spatial OCL constraints in order to check the consistency of a given spatial database. Nonetheless, it cannot express topological integrity constraints involving spatial objects with vague shapes. Additional syntax elements are required to express the possible partial satisfaction (see above) of topological relations between the objects with vague shape involved.

1.3 Objectives and hypotheses of the research

1.3.1 Objectives

The general objective of this research consists of proposing **an approach to specify topological integrity constraints in both transactional spatial databases and data warehouses** that support **spatial objects with vague shapes and their topological relationships.** Three specific objectives are set:

- ***To propose a spatial model in order to represent spatial objects having different levels of shape vagueness and to identify their topological relationships***
- ***To develop an approach in order to reduce the topological relationships vagueness for geometries with vague shapes resulting from the integration of heterogeneous crisp geometries of a same object.*** This approach reuses the spatial model proposed in the first objective.
- ***To add required syntax to the Object Constraint Language (OCL):***

  - To formally express the topological integrity constraints involving spatial objects with vague shapes and their topological relationships

  - To generate SQL scripts from OCL constraints in order to check the consistency of a given spatial database

1.3.2 Hypotheses

The general hypothesis of this research can be presented as follows: **it is possible to provide an approach that supports the specification of topological integrity constraints involving spatial objects with vague shapes and of their topological relationships, both in transactional spatial databases and in spatial data warehouses**.

Three specific hypotheses have been established for this research:

- It is possible to propose a new qualitative model that supports the description of spatial objects with different levels of shape vagueness. Such a model may be integrated in a general approach intended to express topological integrity constraints for spatial object with vague shapes and their relationships.
- It is possible to deal with topological relationships vagueness in the spatial dimension of a data warehouse using a qualitative spatial model able to describe the shape vagueness. In other words, we assume that it is possible to study the shape vagueness using the same approach independently of the factors causing this vagueness.
- It is possible to enrich the constraints language OCL in order to formally express the integrity constraints involving spatial objects with vague shapes and their topological relationships.

1.4 Methodology

This thesis has been realized in the context of a global research project dealing with the integrity constraints in transactional spatial databases and in data cubes. Two other PhD students participated in this project: Magali Duboisset, a PhD student at Blaise Pascal University in France, and Mehrdad Salehi, a PhD student at Laval University. The research of Magali Duboisset has been supported by *Cemagref* (*Institut de recherche Français pour l'ingénierie de l'agriculture et de l'environnement*). She proposed extensions of OCL in order to express integrity constraints involving topological relationships between spatial objects with well-defined shapes. A part of her work consisted in studying the expressiveness of *OCL* (Duboisset et *al.* 2005). An extension called OCL_{9IM} has been implemented into an existing OCL editor called *OCL2SQL* and developed by the Dresden University (Demuth and Hussmann 1999, Demuth *et al.* 2001). OCL2SQL allows the translation of OCL constraints in

SQL queries or triggers. Then, she implemented a second extension called $OCL_{9IM+adverbe}$ where she used a set of adverbs (e.g., *partially*, *entirely*, *etc*.) to describe the topological relationships. She compared the two extensions and proved that they have the same *expressiveness*.

Mehrdad Salehi proposes a formal model for spatial datacubes where he distinguishes different types of components of a datacube structure with regards to the spatial component of data. Such a formal model is required before proposing a framework for identifying different types of integrity constraints in spatial datacubes. Based on this model, he identifies different types of integrity constraints in spatial datacubes. Examples of these integrity constrains are: summarizability integrity constraints, hyper-cellability integrity constraints, fact integrity constraints and traditional integrity constraints in spatial datacubes. Each one of these categories of integrity constraints are further categorized into several sub-categories. Using these results as well as a formal classification of integrity constraints in spatiotemporal databases, he finally develops a formal integrity constraints specification language (ICSL) for defining various types of integrity constraints in spatial datacubes. This ICSL is developed based on a controlled natural language and a natural hybrid language with pictograms.

In practice, the research projects of Mehrdad Salehi and Magali Duboisset have started one year before the present thesis. Then, the results of these research projects have been reused in this thesis and they accelerated the realization of my objectives. The general objective of our research group is to study different problems related to the specification of spatio-temporal integrity constraints for different types of spatial objects (*objects with well-defined shapes* as well as *objects with vague shapes*) in the context of spatial transactional databases and spatial data cubes.

In this thesis, the methodology followed is composed of four phases:

- ***Phase 1: literature review and formulation of the research problem***

This step began with an in-depth literature review in the following domains: (1) modeling of spatial objects with vague shapes in spatial databases and GIS, (2) the formal specification of integrity constraints for spatial objects and their topological relationships. The literature review is justified by the complexity of spatial vagueness problem which has three dimensions at least: a philosophical dimension in addition to the modeling and technological ones. In this research, we principally contribute in the modeling and implementation of spatial objects with vague shapes. We reviewed several research works such as Smithson (1989), Smets (1996), Worboys (1998a), Fisher (1999a), Hazarika and Cohn (2001) and Smith (2001) that proposed different categorizations of spatial imperfection types and defined the *spatial vagueness* and its different uses. Then, we studied some works on the formal specification of integrity constraints for spatial objects and their topological relationships (Cockcroft 1997, Normand 1999, Elmasri and Navathe 2000, Servigne et *al.* 2000, Borges et *al.* 2002, Pinet et *al.* 2004). We concluded that these approaches do not consider the shape vagueness of spatial objects because they are based on traditional spatial models. For that, we explored some research works such as Robinson and Thongs (1986), Altman (1987), Burrough (1989), Cohn and Gotts (1996a), Clementini and Di Felice (1997), Erwig and Schneider (1997), Tang (2004), Reis *et al.* (2006), Verstraete et *al.* (2007) that proposed different spatial models to represent spatial objects with vague shapes. These approaches are generally categorized into two types of models: the exact models in addition to the models based on quantitative mathematical theories. Finally, we studied the advantages and limitations of existing exact models in order to justify the research questions and the objectives of this thesis.

- ***Phase 2: proposing a spatial model for spatial objects with vague shapes and their topological relationships***

According to the literature review, the existing exact models cannot present spatial objects with partially vague shapes such as a lake with rocky borders on one side and swamp borders on the other side. These models consider this type of objects with vague shapes as invalid. We used the principles of the point-set topology (Egenhofer and Herring 1990) to propose a new exact model. We defined three types of spatial objects with vague shapes: *broad points*, *lines with vague shapes* and *regions with broad boundaries*. Additionally, we propose a general framework to identify the topological relationships between objects with vague shapes. The vagueness of a topological relationship can be qualitatively evaluated using a set of adverbs such as *weakly* or *strongly.* Then, this model is reused to deal with the *topological relationships vagueness* for geometries with vague shapes resulting from the integration of heterogeneous and redundant source geometries loaded in a spatial data warehouse. We studied the topological relationships that are possible between final geometries according to those which can occur between source geometries. We intended to reduce the topological relationships vagueness by preventing the impossible relationships between final geometries loaded in the data warehouse.

- ***Phase 3: Extending Spatial OCL in order to express the integrity constraints involving spatial objects with vague shapes and their topological relationships***

There are different approaches to express the integrity constraints in spatial databases (Cockcroft 1997, Elmasri and Navathe 2000, Servigne et *al.* 2000, Borges et *al.* 2002, Bédard et *al.* 2004, Pinet et *al.* 2004, Rodriguez 2005). The *Object Constraint Language* is based on the object-oriented development principles (Pinet et *al.* 2004). This Language has been extended by Duboisset (2007) in order to formally express the integrity constraints involving topological relationships between objects with crisp shapes. Then, two reasons motivated the selection of Spatial OCL to express topological integrity constraints involving spatial objects with vague shapes. First, Spatial OCL is based on the standard constraint language OCL associated to the UML formalism. It allows a declarative specification of constraints; it has a pertinent expressiveness and has been implemented into an existing constraint editor called *OCL2SQL* (Duboisset 2007). Second, Spatial OCL is an element of context of this research; the motivated choice of this language was mainly initiated during the thesis of Magali Duboisset (2007). Extending the language Spatial OCL includes two stages:

1. *Extending the meta–model of Spatial OCL*: three new objects types have been introduced into the meta-model of OCL. These objects types refer to: *broad point*, *line with a vague shape* and *region with a broad boundary*.

2. *Enriching the syntax of Spatial OCL to support integrity constraints involving spatial objects with vague shapes*: we introduced a method to identify the topological relationships between spatial objects with vague shapes. We enriched Spatial OCL by a set of topological operators where their vagueness may be expressed using a set of specific adverbs (e.g. *weakly contains*, *fairly contains*, *strongly disjoint*, *etc*.). These topological operators have been defined in the proposed spatial model and can be introduced in the expression of a spatial query or an integrity constraint.

- ***Phase 4: Validation of the research results***

In this phase, we tested the validity of the results obtained in the first three phases. This validation phase is composed by four principal stages:

1. *Implementing an architecture to store the objects with vague shapes and their topological relationships*: in *Oracle Spatial*, geometric attributes are managed through a generic type called *SDO_Geometry*. In this research, we reused this data type to define the geometries of objects with vague shapes.

2. *Extending Spatial OCL2SQL editor by introducing topological operators adapted to objects with vague shapes*: *OCL2SQL* has been extended in order to express topological integrity constraints involving objects with vague shapes. This extension is based on the spatial model proposed in the phase 2 and the extension of Spatial OCL made in phase 3. The topological operators for objects with vague shapes have been implemented as Oracle functions that reuse the method *SDO_Relate* of *Oracle Spatial*. Each defined function refers to a Java method which realizes necessary controls before executing the operator on the database and displaying a final result.

3. *Testing the application on a real spatial database storing objects with vague shapes:* In this step, we tested our approach using the extension of *OCL2SQL* in order to express some integrity constraints in an agricultural database. This database stores vector data describing the parcels that received organic fertilizers produced by wastewater plants and the agro-food industry in France. In the database, these parcels generally have vague shapes because they have been drawn approximately by users with a GIS-based interface; there is usually a difference between the drawn parcel and the real parcel. We define the agricultural parcels as regions with broad boundaries using an extension of *Oracle Spatial*. Then, we define the integrity constraints using *OCL2SQL* before generating a SQL script that can be executed in the database. The objective of this step is to prove that the extension of Spatial OCL is operational. However, we did not aim at testing the execution performances of the implementation of proposed topological operators.

- ***Phase 5: Analyzing the results obtained in the different phases***

This phase is composed by three main steps:

1. *Reviewing the contributions of the thesis:* the results obtained in the phases 2-4 are reviewed according to the objectives set in the beginning of the thesis. This revision aims at showing the validity of hypotheses presented above.

2. *Comparing the results obtained in the thesis to those of existing approaches:* this step aims at showing the similarities and differences between the results of this thesis and those of existing approaches. It also discusses the advantages and limits of our contributions with regards to other approaches.

3. *Drawing the possible perspectives of this work:* based on the limits discussed in the previous step, some future researches are proposed. The future researches aims at achieving the objectives that cannot be reached in this thesis.

The next activity diagram describes the methodology followed in this thesis:

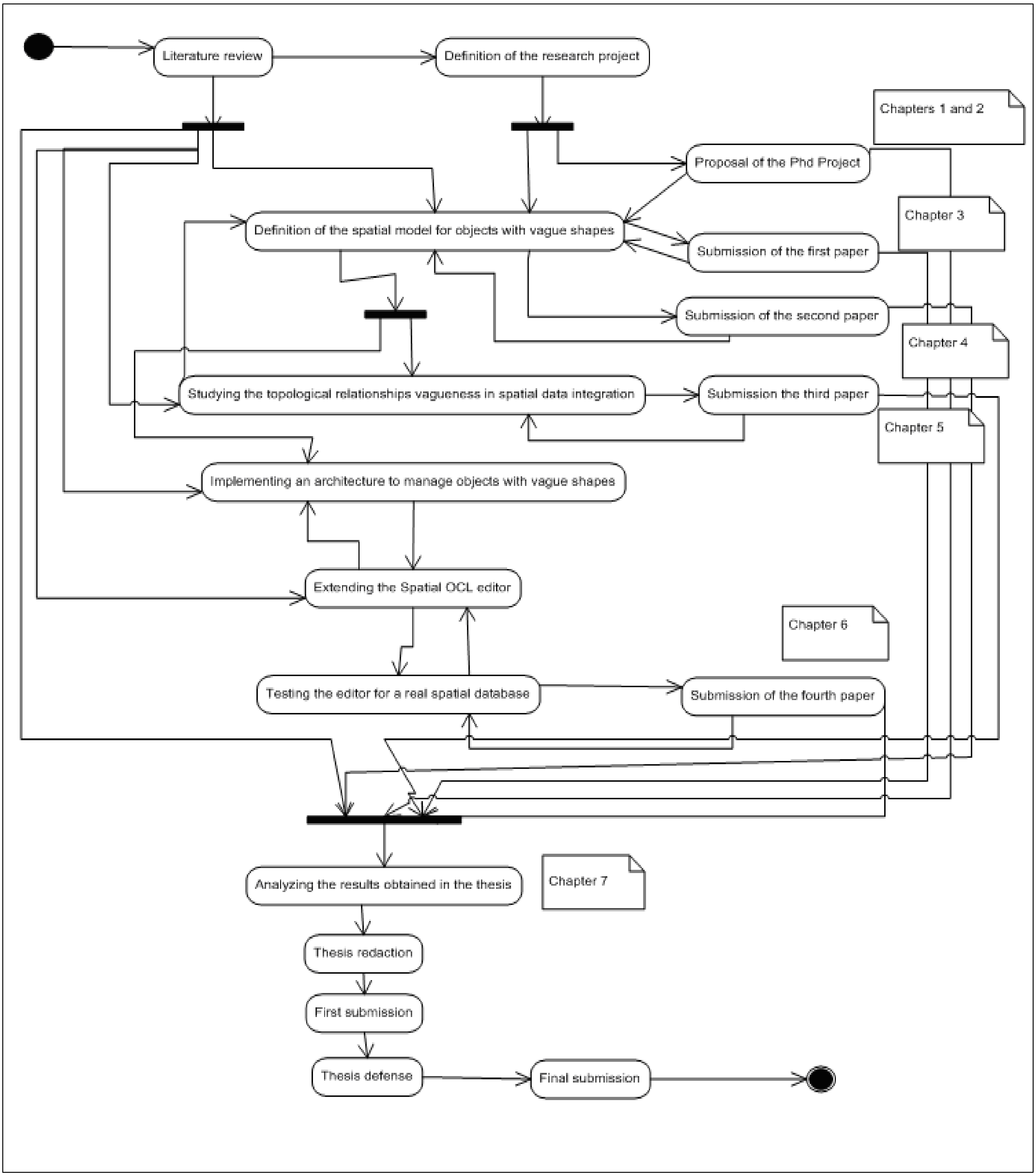


Figure 1.3 UML activity diagram of the research methodology

1.5 Structure of the thesis

The results of this research are presented in seven chapters. Chapter 2 presents a literature review which sets the background of this research and justifies the research questions. This chapter reviews: (1) taxonomies of spatial data imperfections, (2) modeling of spatial objects with vague shapes and of their topological relationships, and (3) the specification of integrity constraints in spatial databases. Chapters 3, 4, 5 and 6 present the contributions of this research and refer to four papers realized during the thesis. These papers have not been substantially modified after being integrated in the thesis. Therefore, the content of some chapters may look redundant. This redundancy is generally required to set the context of our research and to help the journals reviewers to understand the background of our contributions.

Chapter 3 explains the terminology used in this thesis. It also presents a qualitative (or exact) model to represent spatial objects with vague shapes and to identify their topological relationships. We call this approach the *Qualitative Min-Max* (*QMM for short*)[3] *model*. In Chapter 3, we mainly focus on the identification of topological relationships involving regions with broad boundaries. In Chapter 4, we are interested in the identification of topological relationships involving lines with vague shapes. Chapter 5 reuses the principles of the *QMM* model to deal with the topological relationships vagueness for final geometries with vague shapes resulted from the spatial data integration. Chapter 6 presents the extension of Spatial OCL to express the topological integrity constraints involving regions with broad boundaries and their topological relationships. Chapter 7 draws the conclusions and perspectives of this research.

[3] This term has introduced in our second paper (Chapter 4) in order to reference our spatial model. Nonetheless, it is important to denote that we speak about the same spatial model in the remainder of the thesis.

CHAPTER 2: Literature Review

2.1 Introduction

This chapter describes the researches related to the present thesis work. The discussion is organized in three parts. Section 2.2 presents some categorizations of spatial data imperfections as well as the definitions of principal terms used to express its different types. This section is also interested in: (1) the management of the spatial imperfection in spatial databases and spatial data warehouses, and (2) the relationships between the spatial data quality and spatial data imperfections. Sections 2.3 and 2.4 respectively review related works in two domains: (1) the modeling of spatial objects with vague shapes, and (2) the formal specification of spatial integrity constraints.

2.2 Spatial data Imperfections

Two types of data are generally used to describe a spatial phenomenon: (1) qualitative data and (2) quantitative data. These data may be *vague*, *imprecise*, *incomplete, contradictory, etc.* (Dutta 1991). Works such as Smithson (1989), Fisher (1999a) and Mowrer (1999) proposed categorizations of the spatial objects as well as definitions and taxonomies of the spatial imperfection types. Other works such as Burrough (1996), Cohn and Gotts (1996a), Clementini and Di Felice (1997), Erwig and Schneider (1997), Tang (2004), Dilo (2006) and Reis et *al.* (2006) studied the possibilities of modeling the spatial objects with vague shapes and of computing their topological relationships. Finally, some researches such as Pfoser and Jensen (1999), Pfoser and Tryfona (2001) and Pfoser et *al.* (2005) were interested in modeling the imperfection types in spatio-temporal phenomena.

Section 2.2.1 presents the principal taxonomies of spatial imperfection types. Section 2.2.2 focuses on the definition of principal terms used in the literature to express the various types of spatial data imperfections. Sections 2.2.3 and 2.2.4 present the levels of spatial data imperfections and principal strategies to manage it, respectively. Section 2.2.5 relates the spatial data imperfection questions to the transactional spatial databases. In the same way, Section 2.2.6 studied the forms of imperfections in spatial data warehouses. Section 2.2.7 is interested in the relation between the spatial data quality and spatial data imperfections.

2.2.1 Taxonomies of spatial imperfections

The definition of spatial imperfection types is a very complex question where different disciplines such as philosophy, sciences and technology can overlap each other. The objective of this section is to show the divergence of taxonomies of spatial data imperfections proposed in GIS and the spatial databases domain. These taxonomies refer to the background of any framework aiming at modeling a spatial imperfection type (Dilo 2006). Generally, the taxonomies organize spatial imperfection types by using *generalization/specialization* relationships. Devillers (2005) reviewed the principal taxonomies in this domain (Smithson 1989, Smets 1996, Worboys 1998a, Fisher 1999a, Hazarika and Cohn 2001, Smith 2001).

Smithson (1989) considers the *ignorance* concept as the origin of any other type of spatial data imperfection (figure 2.1). Such a philosophical point of view finds its roots in the works of *Socrate* who limited the perfect knowledge to only one certainty: *the ignorance*. Using the reflexivity property, he considers the ignorance of this basic knowledge as a *double ignorance*. This idea was also reused by (Bédard 1987) who introduced the notion of "*meta-uncertainty*": the uncertainty about uncertainty (cf. Section 2.2.3).

Fisher (1999a) focuses, in his taxonomy, on the notion of uncertainty that appears differently for the well-defined objects and ill-defined ones. Two types of objects have been also distinguished by Smith (2001): *bona fide* (well-defined) objects and *fiat* (ill-defined) objects (see section 2.3.1). For the well-defined objects, the uncertainty is often modeled through the probabilities theory such as a confusion matrix which determines whether an object is ill-classified or not (Fisher 1999b). For the ill-defined objects, uncertainty refers to the ambiguity of the object definition as well as of thematic and/or spatial attributes. The latter case relates to a qualitative imperfection which occurs at the conceptual level.

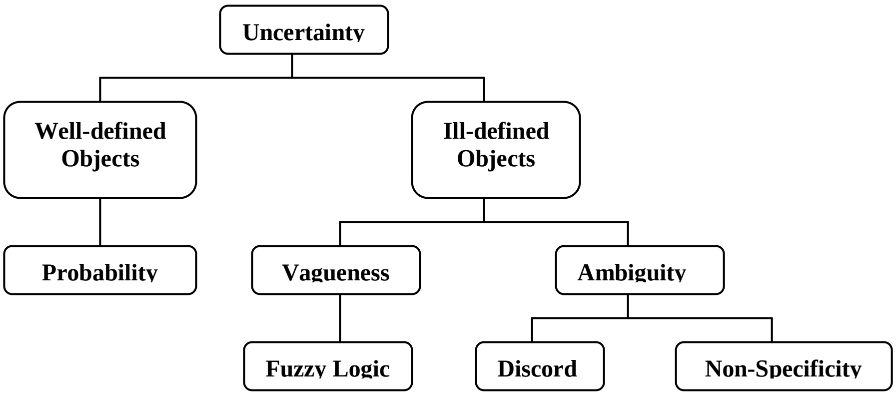


Figure 2.1 Taxonomy of spatial data imperfections (Fisher 1999(b))

According to Worboys (1998b), the spatial data imperfections refer to the factors causing a deficiency in the spatial data quality. These factors relates to the *error* component (a deviation of the data from one value considered as true), *incompleteness* (a lack of relevant information to describe a spatial phenomenon), *inconsistency* (conflicts between data stored in the same structure), *inaccuracy* (a coarse level of granularity or resolution at which the measurement is made or the data is represented), and *vagueness* defined as a lack of precision in the definition of the concepts used to describe the geographic information.

Smets (1996) distinguishes three types of imperfections: *inaccuracy*, *inconsistency* and *uncertainty*. The inaccuracy and inconsistency are two imperfections that can characterise the data whereas the uncertainty relates to the knowledge state about the world (the relationship or *distance* between the available information and the geographic reality).

Couclelis (1996) proposed a first attempt to consider the spatial vagueness in the classification of spatial objects. She proposed to examine the spatial vagueness according to three aspects: (1) the empirical nature of the objects, (2) the observation mode of spatial objects and (3) the user's needs. Hazarika and Cohn (2001) are also interested in the notion of *spatial vagueness*. This notion is considered as the root of their spatial imperfection taxonomy. In (Hazarika and Cohn 2001), the spatial vagueness notion exceeds the simple difficulty of drawing a linear boundary around a given region. It can also occur for objects with well-defined boundaries where there is an uncertainty about their locations.

2.2.2 Terminology related to spatial data imperfections

In the literature, several terms have been used to express the different types of spatial data imperfection. In this section, we review the definitions of these terms.

• **Uncertainty:** it can characterize the knowledge state about a given assertion (Smets 1996).

It refers to the difficulty to determine whether a data is true or false. Uncertainty is considered as a root of different categorizations of spatial data imperfections (Smets 1996, Worboys 1998b, Fisher 1999a). It is presented as a generic imperfection that can be specialized into different forms such as the *imprecision* for quantitative data and the *fuzziness* for qualitative data (Bédard 1987, Erwig and Schneider 1997). According to Bédard (1987), the uncertainty can result from the intrinsic limitations of the modeling process (*omission of details, omission of compatibility between cognitive and physical level, etc.)*). It can also result from the gap between the geographic reality and its description. For example, this gap occurs when fiat spatial objects such as air pollution zones (i.e., regions with broad boundaries in the reality) are presented using crisp polygons. Uncertainty can appear at various levels and in different forms during the development process of a spatial database (see section 2.3.1). Then, the terms '*imperfection'* and '*uncertainty*' can be used interchangeably since the uncertainty includes different types of spatial imperfections. In section 2.3.1, we use the term '*uncertainty'* in order to respect the contributions of Bédard (1987). However, in the remainder of this thesis, the term '*imperfection*' is generally preferred.

• **Error**: it refers to the difference between the available value and another one considered as

true (Goodchild 1995a, David and Fasquel 1997). The error can result from an inadequate calibration of the measurement device, an inadequate use of this device or an erroneous application of the procedures using these measurements as input data. Then, erroneous measurements of the spatial phenomena are introduced as true values to be stored in the database. The error is also related to the concept of *reliability*. The reliability expresses the closeness of collected data to the reality observed (Azouzi 1999).

• **Imprecision:** it refers to limitations on the granularity or resolution at which the observation

is made, or the information is represented (Worboys 1998b). A data value is imprecise when it corresponds to an *interval* (*e.g.*, the age of a person is between 35 and 45), a *disjunction of values* (*e.g.* the age of Jean can be is 35 or 36) or a negation of a given assertion (*e.g.* John do not have 35 years old) (Motro 1995). In the context of spatial data, the *precision* can be *statistical* when it refers to the dispersion around an average value (Mowrer 1999). It can be also *numerical* when it corresponds to the number of significant decimals given by a measurement device (Goodchild 1995a, Mowrer 1999). Statistical precision is generally computed through a probabilistic method using available measurements. It can also be given by computing an ellipse of error (Chrisman 1991). The error and imprecision are orthogonal concepts since the level of the first does not affect that of the second (Mowrer 1999, Duckham et *al*. 2001). For example, the observation "*Quebec is in the north of America*" is more accurate and, at the same time, less precise than the statement "*Quebec is in the United States* ». The second statement is simply inaccurate.

• **Vagueness**: according to Fisher (1999a), the vagueness is an inherent imperfection that characterizes the definitions of some concepts called *vague* (e.g. *young person, bald person, large surface, North, South, etc.*). The membership degree to a given vague concept cannot be computed using a binary logic (i.e., 0 or 1) because its definition is partially respected by elements involved in most cases. The vague concepts can be modeled using Fuzzy Logic (Zadeh 1965). Then, a membership degree is expressed as a value (i.e., belonging to the interval [0,1]) computed using a membership function that defines the vague concept. In the spatial domain, the vagueness is an inherent property of geometries of fiat spatial objects such as *valleys*, or *oceans*. It relates to the difficulty of distinguishing an object shape from its neighborhood. For example, an air pollution zone is a region with a vague shape because it is surrounded by a broad boundary rather the sharp one. Navratil and Frank (2006) consider that the vagueness of concepts entail ambiguous classification of spatial objects. Spatial vagueness can also characterise bona fide objects when there is an uncertainty about their locations. In this case, Hazarika and Cohn (2001) speak about '*location vagueness*'. Nonetheless, an object with a vague shape can be also vaguely located.

Hazarika and Cohn (2001) do not correlate the *shape vagueness* to the difficulty of drawing a linear boundary for a given region (e.g. a *lake*). They consider the temporal data dimension that may affect certainty about the shapes of spatial objects. Accordingly, it is important to denote that *shape vagueness* is a more general notion than *fuzziness*. *Fuzziness* is generally associated to the problem of drawing linear boundaries for regions (Hazarika and Cohn 2001). However, the shape vagueness can also refer to the broadness of a line interior and/or boundary (Reis et *al.* 2006). In the same way, shape vagueness may occur for composed geometries that may contain uncertain parts in addition to certain ones (Schneider 1999). In this work, we are interested only in the shape vagueness for simple fiat objects without considering the temporal dimension. We use the term "*shape vagueness*" because it is more exhaustive than *fuzziness* to describe the shape imperfection of some geographic objects. Moreover, *fuzziness* is often correlated to the use of Fuzzy Logic (Zadeh 1965) to model the boundary broadness. Using this term can be falsely interpreted by assuming that we use Fuzzy Logic to realize the objectives of this thesis (which is not the case as explained later).

Figure 2.2(a) shows an example of a region with a broad boundary. Figure 2.2(b) presents an example of a line where the interior is broad whereas its endpoints remain well-defined. Figure 2.2(c) shows an example of a composed vague region (white polygons for uncertain sub-regions and grey polygons for certain ones).

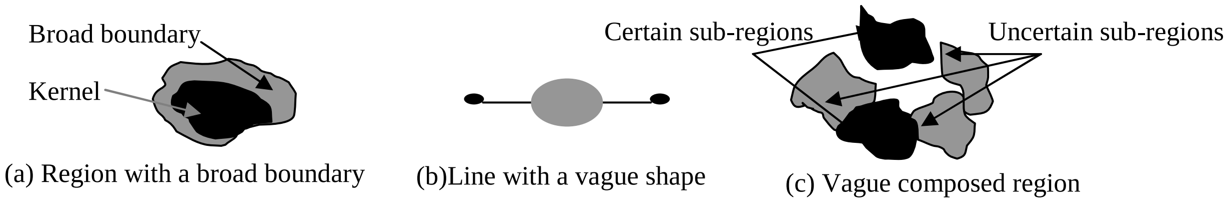


Figure 2.2 Examples of spatial objects with vague shapes

• **Ambiguity**: it appears when different results are obtained using different classification methods for the same set of elements. In this context, broad boundaries can be considered as the result of an ambiguity to affect a set of spatial points to different object classes. Nonetheless, it is important to denote that ambiguity results from the classification process and not from an inherent property of the classes. It corresponds to an imperfection type occurred at the conceptual level defined in (Bédard 1987). Ambiguity can affect the identification (*being or not being such an entity?*) or the categorization (*Being an entity of type A or type B?*) of a given object.

• **Discord**: it appears when different conceptual schemas are proposed by different designers of a same geographic phenomenon. According to Van Oort (2006), each designer uses his proper terminology to define the spatial concepts in the database dictionary. He defines his specific "*product ontology*". The existence of different product ontologies is a first discord type. In the same way, the database users have their specific terminologies and definitions (i.e. *their own problem ontologies*). Then, the heterogeneities between the product and problem ontologies present a second type of discord.

• **Indeterminacy**: it occurs when a spatial object is ill-classified because its definition is ambiguous or coarsely described (Roy and Stell 2001). Indeterminacy is a reflexive, symmetric and transitive relation and is generally modeled through the theory of Rough Sets (Pawlak 1994).

• **Incompleteness**: it refers to a lack of some relevant values and/or occurrences of spatial objects involved. It is generally defined as a partial description of a spatial phenomenon.

• **Inconsistency**: it relates to the existence of logical contradictions in the same database (Worboys and Duckham 2004). For example, an implicit inconsistency can be deduced from the following premises:

Dijon contains 300000 inhabitants
A city of less 500000 inhabitants is not a big city
Dijon is a big city

Inconsistencies are generally managed through integrity constraints (Kainz 1995, Motro 1995, Cockcroft 1997, Normand 1999, Servigne et al. 2000, Pinet et *al.* 2004). Inconsistencies arise when integrity constraints are violated. According to Rodriguez (2005), inconsistency is related to what are called *primary* or *secondary* forms of error. The primary form of error corresponds to a wrong description of location or characteristics/qualities of spatial objects. For example, if an integrity constraint that states that a given object have only one location, there is an inconsistency derived from a primary type of error if there is more than one location for the involved object. This type of inconsistency occurs because there are differences in data accuracy or precision, but also because many observations of spatial phenomena are essentially vague. For example, the boundaries of forests, mountains, lakes, and oceans cannot be determined with precision; i.e. two observers may draw two different shapes/locations for the same object.

A spatial inconsistency related to a secondary error refers to a contradiction between stored data and constraints associated with definitions of geometric primitives. For example, a polygon must be bounded by closed and non self-intersecting polylines that represents its boundary. Inconsistency may also be related to semantic contradictions, such as when a road overlaps a building. These types of inconsistency depend on the spatial domain, and they are captured by rules that should be expressed within the data model.

2.2.3 Levels of uncertainty

Bédard (1987) considers four levels of uncertainty:

- *Conceptual level:* the uncertainty refers to the fuzziness in the identification of an observed reality. For example, a house can be defined as "*a surface greater than 100 m^2 intended for a residential exploitation* ». The definition of the '*house*' concept presented above is fuzzy. It brings to raise the following questions: When is-it possible to consider that a building is principally used for residential exploitation? Moreover, the conceptual imperfection may also refer to the categorization fuzziness. For example, it is possible to have uncertainty to consider a given building as a *house* or a *commerce* (assuming they are two different classes of objects with different sets of properties) if it is exploited simultaneously for these two finalities (commercial at ground level, residential at first level).

- *Descriptive level:* it concerns the uncertainty in the attribute values of an observed reality. At this level, the uncertainty can relate to the fuzziness in the qualitative values and the imprecision in the quantitative values. For example a thematic attribute describing the *vulnerability* of a forest stand can have the following fuzzy values: *"weak"*, "*fair*" or "*strong*".

- *Spatio-temporal level:* a spatial object is generally described by a geometry and a temporality. These data are managed in the database likewise the thematic attributes. For an object geometry, the uncertainty refers to the shape vagueness where there is an inherent difficulty to distinguish the object partially or completely from its neighborhood (e.g. *a zone of pollution*). In the same way, it relates to inaccuracy or imprecision of an object location or other spatial data such as its area or perimeter. For temporal data, the uncertainty relates to the vagueness when there is an inherent difficulty to distinguish an event extension on the time axis (e.g., the birthday of one historic person). It can also correspond to the imprecision or inaccuracy about an event location on the time axis.

- *Meta-uncertainty level*: it refers to the uncertainty about the uncertainties occurred in the first three levels (ex. *95%* certainty of a point to fit within its error ellipse in geodetic adjustment; a population survey about voting preferences that claims a precision of ± 3% *19 times out of 20*). Bédard (1987) spoke about the "*uncertainty of uncertainty"*.

2.2.4 Management of uncertainty

Bédard (1987) distinguishes two approaches to manage the uncertainty in spatial databases:

- ***Reduction***: uncertainty reduction refers to a rigorous definition of modeling rules (i.e. *defining the contents of a model, what to observe and how)* and communication rules (i.e. *defining the model form, the modeling language to use*). From a technical point of view, the uncertainty reduction is realized by using specific tools: mathematical procedures to improve the data precision (e.g. statistics with overabundant measurements), Fuzzy Logic to reduce the qualitative uncertainty, inclusion of lineage in digital maps, the use standard specifications and symbols (e.g. ISO standards), etc. (Bédard 1987, Hunter 1998).

- ***Absorption***: uncertainty absorption refers to the risk related to the uncertainty that remains after all reduction means have been used. For example, it may refer to the guarantees made by a database producer in order to compensate the users damaged by poor data. In the same way, the user can *absorb* the imperfection when he accepts to use non-guaranteed databases. Absorption can also take place when a professional guarantees data (then his professional liability insurances absorb the risk). Bédard (1987) defined the uncertainty absorption as the level of monetary risk in providing or using of a given database. When damages occur, the uncertainty is absorbed by the ones who pay for these damages. This solution is often perceived as a protection against the potential liability claims whether the database entail damages for the users (Hunter 1998).

Finally, the reduction and absorption are substantially different. The r*eduction* is ensured through technical tools and methods whereas the absorption is guaranteed through institutional and legal tools. In practice, the imperfection is managed by combining these two approaches.

2.2.5 Spatial imperfections in spatial databases

2.2.5.1 Introduction

A spatial database is a data collection describing the thematic and spatial properties of real world phenomena (temporal properties are also possible) (Bédard 1999). According to Kemp (2008), spatial databases can be implemented using various technologies, the most common being the relational technology. They can have various structure architectures according to their intended purpose. There are two categories of spatial databases: transactional and analytical. Transactional spatial databases are the most frequent ones; they are often used to facilitate collection, storage, integrity checking, manipulation and display of the characteristics of spatial phenomena. For example, data about precipitations or temperature variations can be stored in a transactional spatial database. The geometry of a spatial object refers to a geometrical primitive (i.e., *a point*, *a line* or *a polygon*) or a collection of these primitives. Analytical spatial databases are more recent and they are very useful in business intelligence applications. This type of databases includes data warehouses and data marts used to meet strategic analytical needs. They can comprise multidimensional structures termed datacubes or hypercubes. When spatial data are involved, the datacubes become spatial datacubes.

The spatial databases are managed through specific software tools called Spatial Database Management Systems (Spatial DBMS). *" A Spatial DBMS is a DBMS whose the meta-model allows the definition and implementation of spatial data types, proposes a query language for spatial data and provides definitions of spatial indexes and algorithms for spatial joins"* (Guting 1994). According to Vauglin (1997), a spatial DBMS supports the management of geometries and the execution of spatial queries (e.g., *finding rivers crossing a forest*) in addition to the functionalities available for *non-spatial* databases. Several DBMS such as *Oracle Spatial* contains specific libraries to store and manage geographical data (Gregan 2004). The spatial DBMS provide additional functionalities in the *Data Definition Language* (*DDL*) and the *Data Manipulation Language* (*DML*). For example, *Oracle Spatial* proposes a specific data structure called *SDO_Geometry* in order to store geometries of spatial objects. In the same way, the function *SDO_Relate* executes spatial queries where the conditions concern topological relationships between spatial objects (i.e., *finding spatial objects that meet a river*). A spatial indexing method is also integrated into *Oracle Spatial*.

2.2.5.2 Imperfection aspects in spatial databases modeling

Bédard (1999) proposed a pictogram-based language in order to help the database designer to describe the geometry properties of a given spatial object. Temporal pictograms are also provided to represent the temporal *existence* and geometric *evolution* of a given spatial object. These pictograms are available through a design editor for spatial databases called *Perceptory* (Bédard et *al.* 2004). In (Miralles 2006), *Perceptory* has been extended to support the description of the spatial extensions of spatial objects with vague shapes. In this same way, Parent et *al.* (1997) provided syntactic tools to build class diagrams of spatio-temporal applications. This approach has been extended to model the *random imperfections* (*measurement problems*) and the *vagueness* of spatial concepts (Shu et *al.* 2003). Likewise, several works (Duckham et *al.* 2001, Yazici et *al.* 2001, Fonseca et *al.* 2003, Shu et *al.* 2003) enriched the meta-models of some design methods in order to support the spatial vagueness. For example, Yazici et *al.* (2001) proposed an extension of UML (*Unified Modeling Language*) by adding two constructors: *U* used to represent *inaccuracy and imprecision (that can characterize an object location)* and *F* used to describe the shape vagueness. They applied this extension to describe an environmental information system for a pollution phenomenon.

2.2.5.3 Management of imperfections in spatial databases

A spatial database is a formal *description* of the geographic reality where two types of operations can be done: *transformations* and *modifications* (Motro 1995). A *description* of database refers to its structure and its contents. However, the operations of *transformation* and *modification* consist in the update of the contents and the structure of the database, respectively. In this thesis, we are interested in the *description* component because we are focused on the modeling of spatial databases. Accordingly, a modeling process generally aims at producing a database description that respects two principles properties: the *soundness* and *completeness*. On the one hand, a description is *sound*, if it includes only necessary data to describe the reality. On the other hand, a description is *complete* if it includes all of data that describe the reality. At the conceptual level of a database description, the vagueness results from a simplification of the complex reality and/or an ambiguous definition of the spatial objects (Yazici et *al.* 2001). At the physical level, several solutions can be implemented to deal with different aspects of imperfections in relational databases (Motro 1995):

1. "Null" values: a "*null*" value denotes that no information is available. It can be also used to denote the inapplicability; i.e., that a specific attribute is inapplicable to a given object.

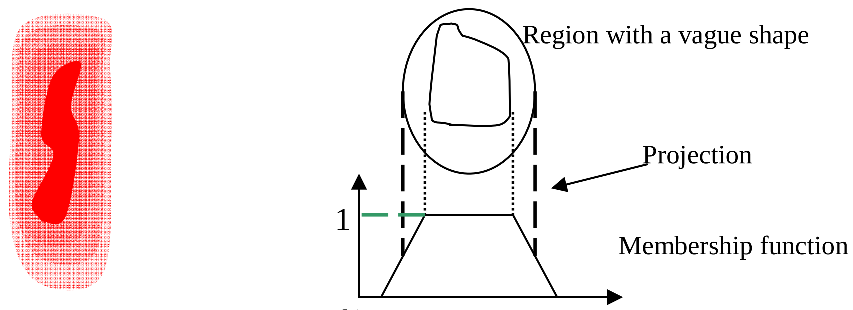
2. Disjunctive value: it is a set of values that necessarily include the true one (but we don't know precisely which one is true). A disjunctive value occurs when there is an uncertainty to assign one value to a given attribute. Then, a set of values (separated by *OR* operator) are assigned to the attribute.

3. Confidence factors: they denote the *confidences* that one can have about the description elements (Motro 1995). For example, confidence factors have been used into retrieval systems to indicate the confidence that a specific word describes in a given document.

4. Probabilistic databases: in these databases, data are represented through variables where each is related to a probability distribution function. The data are stored in the database with a probability that present their truth degrees. Examples: *P (age (Jean) = 32) = 0,6; P (age (Jean) = 33) = 0,4.*

5. Possibilistic or fuzzy databases: in these databases, concepts are modeled as fuzzy subsets (Zadeh 1965). These concepts are managed by the DBMS through a fuzzy inference system that computes a membership degree for each instance according to a membership function associated to the concept involved.

In the context of spatial databases, a geometry with a vague shape can be represented though a fuzzy membership function defined in raster data where the shape vagueness is shown using a color degradation (figure 2.3(a)). This method has high implementation and management *costs*. It is only possible for a limited surface and consists in computing the membership degree of each pixel to a given class. However, the vector format allows a less expensive representation of geometries with vague shapes (Cohn and Gotts 1996(a), Clementini 2005). Morris (2003) proposes a model to store geometries with vague shapes in a vector format using fuzzy subsets. In this approach, a region is represented as a set of sub-regions. For each one, a membership degree is computed through a membership function defining the global *fuzzy region* (figure 2.3(b)). A membership degree refers to the projection of a sub-region on the membership function.



(a) Raster representation (b) Vector representation

Figure 2.3 Representation of a region with a vague shape

The databases were initially invented to meet transactional needs that consist in managing one or several daily activities of an organization. However, the economic competition encouraged the rise of decisional needs where it is required to analyze time-variant data in order to make the best decisions. Decisional needs are met using specific structures called data warehouses (Malinowski and Zimányi 2007). The management of data warehouses requires large storage capacities in order to store a large amount of data loaded from different source databases. The data warehouses are used to load data cubes intended to meet the needs in analysis and decision-making processes. In the next section, we focus on some spatial imperfections (those related to spatial data integration) in spatial data warehouses.

2.2.6 Spatial imperfections in spatial data warehouses

A data warehouse is a *subject-oriented*, *integrated*, *time-variant* and *non-volatile* collection of data in order to support a decision-making process (Inmon 1992). It can be also defined as a time-variant data collection that is extracted from different transactional databases and files, organised by subject, and stored into one final data structure in order to support a decision-making process (Kimball 1996). The data warehouses are generally represented using a multidimensional model such as the star schema. A star schema is composed by a single fact table connected to a set of dimensions tables. The dimensions refer to the analysis perspectives such as the *time* or *space*. A dimension contains one or several *hierarchies* typically composed by several granularity *levels* such as the *country*, *region*, and *county* for a spatial dimension. According to Malinowski and Zimányi (2005), a *level* refers to a set of instances called *members* that have common characteristics. For example, the level '*region*' of the spatial dimension contains the following members: *East*, *West*, *North*, and *South*. Two consecutive levels of a hierarchy are called *child* and *parent* depending on whether they include more detailed or more general data, respectively. The members of a parent level are obtained by aggregating its child members of the immediately lower level. The fact table stores one or several attributes that represent the analysis such as the sales amount or the number of accident victims. They are generally numerical attributes that are summarized before being analyzed according to the set of dimensions (Rafenelli et *al.* 2003). According to Rivest et *al.* (2003), a fact refers to a combination of dimension members, with the measures value for a particular aggregation level. For example, a fact can correspond to the "*car sales in Quebec city at the first half of 2008*"; i.e. the sum of car sales for the member *Quebec* of the dimension *Space* and for the member *first half of 2008* of the dimension *Time*. The combination of all facts and dimensions refers to a *data cube*. Different data cubes can be obtained from the same data warehouse.

Franklin (1992) estimates that 80% of transactional data have a spatial component. This fact justifies the rise of spatial data warehouses that support the management of significant amounts of time-variant data including a spatial component. The spatial data are captured by their geometries and can be managed in the dimension tables as well as in the fact table. In spatial dimensions, the members of different levels can be related by classical relationships or topological relationships. In this second case, each member of a hierarchy level has a geometry that is normally *within* the geometry of its parent member belonging to the immediately higher level. However, other topological relationships such as *Overlap* or *Covers* are possible but require the use of specific operations to compute measure aggregations. These relationships occur when the hierarchy levels are loaded from heterogeneous source databases. In practice, it is generally difficult to geometrically deduce the topological relationships between objects belonging to different hierarchy levels. These relationships can be managed through semantic links between the geometries involved stored in the data warehouse. In a fact table, the spatial measures can correspond to geometries or quantitative spatial data such as the area or distance.

A spatial data warehouse is generally loaded from different data sources. These data sources are involved in an integration process in order to be adapted to the structural and semantic requirements of the data warehouse. In the spatial data integration, the data sources are generally heterogeneous at different levels such as the database structures heterogeneities, the geometric heterogeneities, etc (Devogel 1997). Then, different forms of imperfections can be observed in a spatial data warehouse. On the one hand, each source database includes its own imperfections that can be propagated in the data warehouse. For example, the hierarchical levels of a spatial dimension are typically extracted from different sources. Then, inconsistencies can be shown during the navigation from one level to another: the navigation from a *county* level to a *municipality* level can be inconsistent whether data are extracted from different data sources; some municipalities are not completely inside their parent county. Moreover, the imperfection in spatial data warehouses can be related to the data aggregations. For example, the aggregated values may be different to the sum (when the aggregation function is SUM) of values stored at the lower level. For example, the inhabitants living in the broad boundary of an urban zone may not be computed in the sum of urban inhabitants the region involved. In some cases, the inconsistencies between the different aggregation levels are managed through *warnings* that inform the users about the possible incoherencies (Levesque et *al.* 2007).

The geometric heterogeneities between source geometries can also entail the shape vagueness. The geometry of each member of a spatial hierarchy level can refer to the final geometry obtained by merging heterogeneous geometries available in the source databases. The principal tool to merge source geometries is the *Overlay* method (Frank 1987, Demirkesen and Schaffrin 1996, Harvey and Vauglin 1996). This method compute the intersection of the different source geometries using a *tolerance value* around the nodes of a source geometric representation taken as a reference in order to merge the others. A source geometry is excluded from the integration process whether it is not inside the tolerance zone. When the quality of source geometries cannot be evaluated, the shape of a final geometry becomes vague if there is a non-empty difference between the union and intersection of source geometries (Shepherd 1992). The topological relationships, between the members of the same spatial hierarchy level as well as those between the child and parent members belonging to different levels, should then consider the shape vagueness of the geometries involved.

Dealing with the spatial data imperfections leads to investigate how they entail deficiencies in the spatial data quality. Moreover, the advances in the information technologies domain gave place to increasingly powerful material solutions at the level of storage capacities and personal use of spatial data. From this perspective, the spatial data quality is increasingly described by the spatial databases producers and required by the users.

2.2.7 Spatial data quality and management of imperfections

2.2.7.1 Notion of spatial data quality

- ***Definitions***

In the standard ISO 19113 (ISO/TC211 2002), the general definition of quality is "*the totality of features and characteristics* of a product *or service that* bear *on its ability to satisfy stated or implied needs*". According to Devillers (2004), various definitions have been associated to the concept of quality in the domain of geographical information systems. Two main groups of definitions can be then identified. The first group associates the quality of a product or a service to the standards and specifications, allowing to reduce the errors in the product. The second group associates quality with the satisfaction of the users' needs, i.e. a product with a good quality level should meet or exceed the users' needs. These two groups of definitions are commonly identified by "internal quality" and "external quality" (Aalders, 2002, Dassonville et *al.* 2002). In GIS, the first group is generally placed from the point of view of the producers of data, compared to the second group which is placed from the point of view of the users (Kahn and Strong, 1998).

The *internal quality* relates to the meeting by the data producer of the requirements defined by the user or by himself. These requirements represent the theoretical specifications or the nominal ground (David and Fasquel 1997) that is used to evaluate the internal quality. Generally, the data producer describes the internal quality of its product using the following elements: (1) *actuality of data,* (2) *geometric and thematic accuracy,* (4) *genealogy,* (5) *logical consistency* and (6) *completeness* (Mostafavi et *al.* 2004). This description generally appears as a quality report associated to the database (Boin and Hunter 2006). The *internal quality* can be evaluated by making the comparison with theoretic specifications of the reality description called the "*nominal ground*" (David and Fasquel 1997). On the other hand, the *external quality* corresponds to the concept *of adequacy to the user's needs* or "*fitness for uses*" (Juran et *al*. 1979). Bédard and Vallière (1995) define *"the external quality as the set of characteristics which make spatial data ready to meet user's needs in a given application"*. The external quality cannot be objectively described by the data producer because a same database can be intended for different uses. Accordingly, Devillers (2004) proposes a fast and intuitive approach to communicate the information about the spatial data quality and to improve the evaluation of the *external quality*.

- ***Elements of spatial data quality***

In (Guptill and Morrison 1995, Azouzi 2000, Aalders 2002, Van Oort 2006), the spatial data quality is described through the following elements:

- *Genealogy (or lineage)*: it refers to the history of a geographic dataset. It describes the source of data as well as the acquisition and derivation methods including all transformations involved in the data production process (Van Oort 2006).
- *Completeness*: it measures the exhaustiveness of the data in terms of the spatial and thematic properties (Brassel et *al.*1995). In the case of absence of data, one speaks about data *omission.* In the case of excess data, one speaks about a data *commission* (Guptill and Morrison 1995, Van Oort 2006).

- *Logical consistency*: it relates to the fidelity of relationships encoded in the data structure of the digital spatial data (Guptill and Morrison 1995, Van Oort 2006). The consistency is composed by: (1) the conceptual consistency (i.e., *the validity of data according to the conceptual schema*), the thematic consistency (i.e., *the validity of data according to the value domains*), the structural consistency (i.e., *the validity of data according to the physical structure of data in the DBMS*) and the topological consistency (i.e., *the validity of geometrical properties of the spatial objects and of their topological relationships*).
- *Positional accuracy*: it relates to the positions exactness of geographic objects. A distinction is generally made between the *relative accuracy* and *absolute accuracy* (Guptill and Morrison 1995). The *absolute accuracy* refers to the relationship between a geographic position on a map (a street corner, for instance) and its real-world position measured on the surface of the earth. The *relative accuracy* is the difference in the distance measured between two points on a map and the true distance between these same two points, which is measured using conventional surveying methods.
- *Attribute accuracy*: it provides an assessment of the accuracy of the identification of entities and assignment of attribute values in a data set. It measures the accuracy of quantitative and qualitative values assigned to the thematic attributes (*the population of an urban area, the city name, etc.*) of the spatial objects involved. The thematic attributes can be measured according to different measurement scales: *cardinal*, *ordinal* and *nominal*. Each type of values requires specific procedures to measure the attribute accuracy (Azouzi 2000).
- *Temporal accuracy*: it refers to the accuracy of the temporal information describing geographic entities and their temporal relationships. It is also called the "*temporal quality*" (Van Oort 2006). It can be subdivided in: (1) the accuracy of temporal measurements, (2) the consistency of temporal topology (i.e., *the relationships between the temporal events*) and (3) the temporal validity (i.e., *the actuality of data and their validity according to the time*).

The elements of spatial data quality can be used to evaluate the spatial data imperfections in a spatial database. These elements cover principally the problems of *inaccuracy*, *incompleteness*, *inconsistency*, *imprecision and vagueness.* Then, improving the internal spatial data quality leads to the reduction of spatial data imperfections. However, reducing spatial data imperfections is not a solution for the spatial objects with vague shapes. This strategy would decrease the reliability of spatial databases because the geographic reality would be *excessively* simplified. The spatial objects with vague shapes and their topological relationships are not always properly represented using the traditional spatial models. In the next section, we review existing approaches that proposed different models to represent objects with vague shapes and to identify their topological relationships.

2.3 Spatial objects with vague shapes and their topological relationships

Section 2.3.1 presents a categorization of spatial objects. Section 2.3.2 is interested in the modeling of spatial objects with vague shapes. Section 2.3.3 is focused on the identification of their topological relationships. Section 2.3.4 reviews the classifications of integrity constraints and existing tools to formally express them.

2.3.1 Fiat objects *vs* bona fide objects

Two categories of spatial objects are distinguished: (1) *fiat* objects and *bona fide* objects (Smith 1994, Smith and Varzi 2000, Brodeur et *al.* 2003). This categorization is based on the distinction between "*fiat boundary*" and "*bona fide boundary".* A fiat boundary cannot be directly observed in the reality (Bittner 2000). For example, the boundaries between the hills of a mountain chain are *fiat*. *Forest stands* and *lakes* are two examples of *fiat* objects. However, a *bona fide boundary* establishes a discontinuity in the space. It refers to a *sharp* line *or* a *physical* demarcation between two objects having qualitative and physical differences (Smith 1994). *Buildings* and *roads* are examples of bona fide objects. Nonetheless, the notion of *fiat* and *bona fide* classification cannot be applied independently to the users' needs and specificities of the studied phenomenon. In other words, a given object cannot be inherently classified as fiat or bona fide. In practice, any object can be in the first or in the second class according to the definition given to this object. For example, it is generally difficult to determine the start and final points of a road. In the latter case, it is more appropriate to consider a road as a fiat object than a bona fide one. Then, it is possible to conclude that the boundary between the fiat class and bona fide class is broad.

The *fiat* objects refer to spatial objects with vague shapes (in our terminology) such as regions with broad boundaries or broad lines. For this type of objects, several researches (Guarino and Wetly 2000, Hwang and Thill 2005) made the distinction between the *identity vagueness* and *unit vagueness*. This distinction reminds the first uncertainty level (i.e., *a conceptual uncertainty*) defined in (Bédard 1987).

Generally, the traditional geometrical models do not allow the representation of vague shapes. They reduce the spatial extensions of the spatial objects to their certain parts (Yazici et *al.* 2001). For example, a lake with a broad boundary is represented as a region with a sharp boundary despite the non-reliability of this representation (in the best cases, metadata are stored in the databases to describe the data imperfection). This approach can be motivated by two reasons: (1) a tendency to eliminate the shape vagueness in the geometric representations and (2) the absence of a technology that allows the storing and management of spatial objects with vague shapes. This modeling approach reduces the reliability of spatial databases. For example, let a database intended for the storage of spatio-temporal data describing some phenomena related to climatic changes. In this example, the climatic zones should be represented as regions with broad boundaries because they have fiat boundaries that cannot be reliably represented as linear demarcations. These zones are modeled as being *bona fide* in order to allow their management using existing technologies. Consequently, an inherent property of these objects is lost. Let a second example of a spatial database that stores data about the moving traffic in a navigation system. In this database, a vehicle coordinates represent only an estimation of its real position at a moment *t*. Moreover, there are generally no data that inform the user about the truth degree of such estimation. For that reasons, there is a necessity to meet new needs by managing spatial objects with vague shapes and computing their topological relationships using a new modeling approach.

In general, we distinguish between at least two categories of models used to represent the spatial vagueness. In the first category, crisp spatial concepts are transferred and extended to formally express the spatial vagueness; we speak about the *exact models* such as Cohn and Gotts (1996b), Clementini and Di Felice (1997), Erwig and Schneider (1997). In the second category, three principal mathematical theories are generally used: (1) the models based on the Fuzzy Logic (Zadeh 1965) (e.g., Altman 1987, Burrough 1989, Brown 1998, Schneider 2001, Tang 2004, Hwang and Thill 2005, Dilo 2006), which can be used to represent continuous phenomena such as temperature, (2) the models based on the Rough Sets theory (e.g., Ahlqvist *et al.* 1998, Worboys 1998b), which represents the objects with vague shapes as a pair of approximations (*the upper approximation and lower approximation*), and (3) the models based on the probability theory (e.g., Burrough and Frank 1996, Pfoser *et al.* 2005), which is principally used to evaluate the errors in positions and attributes. In the next section, we review the principal approaches belonging to these categories of models.

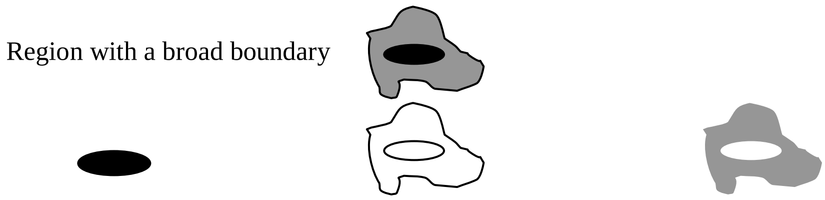
2.3.2 Modeling of spatial objects with vague shapes

2.3.2.1 Definitions based on exact models

The exact or qualitative models reuse the existing definitions in traditional spatial models to represent the spatial objects with vague shapes. The Egg-Yolk theory (Cohn and Gotts 1996a) is an extension of the RCC (*Region Connection Calculus*) model (Randell and Cohn 1989, Cohn et *al.* 1997). This theory has been the first that introduced the concept of *regions with broad boundaries* (Hazarika and Cohn 2001). In this approach, a region with a broad boundary is made up of two crisp sub-regions (surrounded by crisp boundaries). The internal sub-region is called "*Yolk*" (i.e., the certain part of the geometry) which is surrounded by an external sub-region called "*White*" (i.e., the broad boundary or the uncertain part of the geometry). The union of the "*Yolk*" and "*White*" refers to the "*Egg*" (i.e., an Egg-Yolk region is made up of two sub-regions with crisp boundaries). Cohn and Gotts (1996b) consider the "*Yolk*" as a region vaguely localised inside a container sub-region (i.e., the "*Egg*"). Since the points and lines are not considered in the *RCC* model, the *Egg-Yolk* theory does not model the shape vagueness of these two types of objects. In addition, regions with broad boundaries with empty "*yolk*" or empty "*egg*" are not admitted. The crisp regions cannot be represented using the *Egg-Yolk* theory.

Likewise, Clementini and Di Felice (1997) proposed a definition of regions with broad boundaries based on the principles of the general point-set topology (Egenhofer and Herring 1990). A region with a broad boundary is defined as a composition of two sub-regions with crisp boundaries A_1 and A_2, with $A_1 \subseteq A_2$. The broad boundary of A refers to the closure of the difference between A_1 and A_2, $\Delta A = A_1 - A_2$. In this approach, A_1 and A_2 should be topologically valid; i.e. they should be closed, regular and connected (Clementini and Di Felice 1997). For the linear geometries, Clementini and Di Felice (1997) distinguish two types of lines with vague shapes: *completely broad lines* and *lines with broad boundaries* (i.e.

the line endpoints are broad). Tang (2004) proposed an extension of the approach defined in Clementini and Di Felice (1997) by giving a more detailed formal definition for regions with broad boundaries. He distinguishes four mutually disjoint topological invariants: an *interior*, an *interior of the boundary*, a *boundary of the boundary*, and an *exterior* (figure 2.4).



(a) Interior (b) boundary of the boundary (c) Interior of the boundary

Figure 2.4 Topological invariants of a simple region with a broad boundary (Tang 2004)

The condition $A_1 \subseteq A_2$ in Clementini and Di Felice (1997) does not exist in Erwig and Schneider (1997). Erwig and Schneider (1997) are interested in another kind of vagueness, where a *region with a vague shape* is a composed geometry. The geometry components belong to a pair of subsets. First, the *kernel* subset contains the sub-regions that *definitely* belong to the region with a vague shape. Second, the *boundary* subset contains the sub-regions that *possibly* belong to the region with a vague shape. Likewise, the *points with vague shapes* and *lines with vague shapes* are respectively defined as a pair of subsets of points and lines. Crisp spatial objects can be expressed through this model when the *boundary* subset is empty. Figure 2.5 gives an example of a region with a vague shape *A*, in which the *white* sub-regions compose to the *boundary* subset and *gray* ones compose the *kernel* subset.

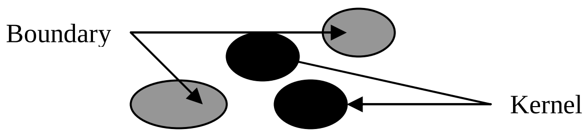


Figure 2.5 Representation of a region with a vague shape according to (Erwig and Schneider 1997)

2.3.2.2 Models based on mathematical approaches

- **Probabilistic approaches**

  - *Principles*

The probabilities theory is a branch of mathematics concerned with random phenomena (Wikipédia 2008). This theory evaluates the uncertainty by computing a value that belongs to an interval bounded by 0 *for impossible events* and 1 *for certain events*. Two types of probabilities are generally distinguished. On the one hand, the *random probability* refers to the realization chance of a future event which depends on some unpredictable physical phenomena (e.g., *obtaining a certain number while turning a chance wheel*). On the other hand, the *epistemic probability* relates to the uncertainty of the assertions when there is a lack of knowledge about the circumstances and causalities. This type of probabilities has got to do with our possession of knowledge, or information (Berglund 1993).

More formally, if X indicates the universe of probable events, it is possible to define a probability distribution P: $X \rightarrow [0, 1]$. The value given by $P(X)$ specifies the probability that an event x occurs. A probability distribution should satisfy the following axioms:

$$P(X) = 1; P(\varnothing) = 0$$
$$P(A \cup B) = P(A) + P(B);\ \text{iff } A \cap B = \varnothing$$

- *Modeling spatial imperfection by using probabilistic approach*

In the case of spatial data, probabilistic methods are quantitative approaches mainly used to deal with the positional inaccuracy and precision by using *probability distributions* (Worboys and Duckham 2004). For example, Shu et *al.* (2003) use this theory to represent the random positions of spatial objects. Accordingly, the probability distributions are intended to two principal uses. A spatial probability distribution can model the random position of a spatial object (Fisher 1999(b), Shu et *al.* 2003, Worboys and Duckham 2004). Other approaches (Bordoloi et *al.* 2004, Pbesma et *al.* 2006) use the same concept to visualize the uncertainty by using raster data. In the latter case, a probability distribution allows to assign a *weight* for each pixel belonging to the spatial extension of an object visualized. In geodesy, the least squares compensation method has been related to the probability theory and is taught in every basic geodesy class as a fundamental approach to model imperfection in position and measurements. This method allows to estimate the non-systematic errors (due to independent factors non-related to a failure in the measurement device) in a dataset when there is a superabundance of measurements.

The probability theory is a quantitative approach which has two principal advantages: (1) an advanced mathematical background (Yao 1998), and (2) a simplicity of application. However, in the spatial domain, it is rarely used for other types of imperfections than the inaccuracy or imprecision. These imperfections can result from the difficulty of observations, the linguistic vagueness, the inherent shape vagueness of some objects, the complexity of human spatial reasoning, etc.

❖ **Fuzzy approaches**

➢ *Basic elements of Fuzzy Logic*

Fuzzy Logic (Zadeh 1965) is based on the notion of Fuzzy Subsets that are generally used to model vague concepts such as "*young person*", "*small*", etc. This theory is an *extension* of the binary logic (i.e., the use of only two values {0, 1} to evaluate the truth of an assertion). The works of Zadeh (1965) represent the beginning of the proposals of modeling approaches based on fuzzy inference systems. His first contributions were the use of the fuzzy logic to represent the natural language.

The binary logic distinguishes *firmly* between the members (i.e. *elements having 1 as a membership degree to the universe*) and non-members (i.e. *elements having 0 as a membership degree to the universe*) of a given universe *X*. The fuzzy logic is a generalization of the binary logic since it establishes the correspondence between the members of the universe *X* with all values belonging to the interval [0, 1]. Then, the elements of *X* do not have a strict membership (i.e., *0 or 1*) but rather a *membership degree* belonging to the interval [0, 1] and computed by using a *membership function*. Godjjevac (1999) defines the notion of membership degree as the compatibility of a given element with the concept represented by the fuzzy subset involved. A membership function can take different forms according to the application: it can be *monotonous*, *triangular*, *trapezoidal*, *bell-shaped*, etc. A membership function is generally expressed as follows:

$$\mu_{\tilde{A}} : X \rightarrow [0,1]$$

According to this function, the non-members of a given subset *A* have a membership degree equal to 0. However, the members which are *certainly* in *A* have a membership degree equal to 1. Other elements which *partially* belong to *A* have a membership degree between 0 and 1.

➢ *Fuzzy modeling of spatial objects with vague shapes*

In the context of spatial databases, several approaches such as (Robinson and Thongs 1986, Altman 1987, Burrough 1989, Zhan 1997, Schneider 1999, Tang 2004, Dilo et *al.* 2005, Hwang and Thill 2005, Verstraete et *al.* 2007) used the theory of fuzzy subsets to model the spatial objects with vague shapes and their topological relationships (Dutta 1991). In these approaches, the spatial objects with vague shapes are called *fuzzy spatial objects.* The term '*fuzzy*' does not express a type of spatial imperfections but rather an indication about the mathematical approach used to model shape vagueness. Figure 2.3 shows examples using fuzzy approaches to represent the spatial objects with vague shapes.

Zhan (1997) and Dilo (2006) interpret a spatial object with a vague shape as a fuzzy subset. In (Zhan 1997), the membership function of a spatial object with a vague shape is made up of n $\alpha - cuts$ (an $\alpha - cut$ is a crisp set containing the elements having membership degrees higher or equal to a value α belonging to the interval [0, 1] (Godjjevac 1999)) in order to facilitate its interpretation. In the same way, Somodevilla and Petry (2003) represented a region with a vague shape by a set of $\alpha - cuts$ organized inside a minimum rectangle including the region. Schmitz and Morris (2006) proposed a fuzzy model (in the sense of Fuzzy Logic) also based on the concept of $\alpha - cuts$ to represent fuzzy regions. They use this concept to describe the internal structure of the broad boundary that surrounds the interior of the region. The use of $\alpha - cuts$ allows to deal with principal limitations of fuzzy approaches related to the interpretation and use of the membership functions defining the fuzzy subsets. Figure 2.6 shows a region where the broad boundary is decomposed into $n \alpha - cuts$. α = 0 in the exterior, α = 1 in the interior and α belongs to the interval]0,1[inside the broad boundary (with $\alpha_1 > \alpha_2 > \alpha_3$).

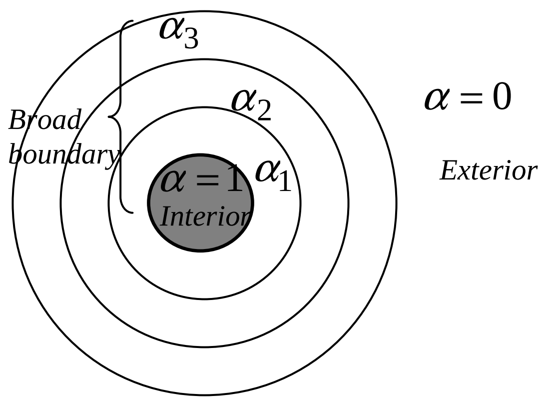


Figure 2.6 A region with multiple $\alpha - cuts$ (Schmitz and Morris 2006)

According to Schmitz and Morris (2006), the definition of fuzzy regions assumes that the boundary is broad everywhere and an $\alpha - cut$ should uniformly surround the interior of the region. This assumption is not realistic because a region can have a partially vague shape; i.e., broad boundaries in some locations and sharp boundaries in some others (e.g., a lake with *rocky banks* on one side and *swamp banks* on the other side). In this case, a $\alpha - cut$ should have more than one definition in order to be always inside the broad boundary. However, an $\alpha - cut$ cannot have more than one definition inside the same fuzzy subset. Consequently, regions with *partially* broad boundaries cannot be represented by using this approach.

In (Tang et *al.* 2003, Tang 2004), a spatial object with a vague shape is defined in two different ways. The first definition is based on the properties of a *crisp topological space* (Tang 2004). In a crisp topological space, the membership of a given point is evaluated by using a binary logic (1 if the point belongs to the object, 0 else). The second definition respects the topological properties of a *fuzzy topological space.* The concept of *fuzzy topological space* is a generalization of crisp topological space, in which the spatial objects are defined as fuzzy subsets (i.e. the membership degree of a point is α, where $0 \le \alpha \le 1$). Bjørke (2004) and Schneider (2001) proposed a method to identify the broad boundary of a region by computing the membership of each point to the interior and boundary, respectively.

❖ Rough Sets

Rough sets theory (Pawlak 1994) is a formal approach to deal with the difficulty to distinguish between the elements belonging to a first set *A* and those contained by a second set *B*. For example, let two data sources *A* and *B* involved in an integration process. *A* and *B* store the same set of forest stands where the geometries are defined with different resolutions and precisions. To distinguish similar forest stands, Rough sets theory can be used to define two approximations for each stand: a minimal approximation and a maximal one. They correspond to the geometric representation having the smallest resolution and that having the highest resolution, respectively (figure 2.7).

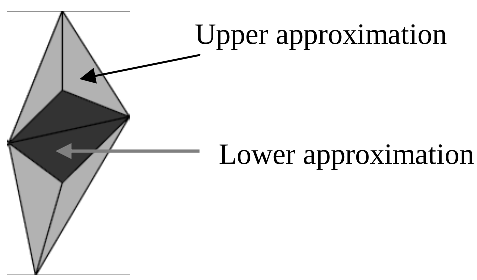


Figure 2.7 Example of an integration of two geometries based on the rough sets (Worboys 1998b)

In the case of spatial data, Rough sets theory has been also used to model spatial objects with vague shapes and their topological relationships (Beaubouef and Petry 2001). Worboys (1998b) used this theory in a context of multi-resolution representations. Ahlqvist et *al.* (1998) introduced the concept of *approximate classification* which corresponds to the set of *rough sets* associated to the data. In this approach, the membership to the maximum approximation reflects the uncertainty of the concerned element.

Roy and Stell (2001) deal with *indeterminacy* defined as a knowledge imperfection that prevents a bivalent evaluation of a given assertion (true or false). They define an indeterminate region by using approximate sets (Pawlak 1994). An indeterminate region is composed of a lower approximation and an upper one. The difference between these approximations refers to the broad boundary of a region. When this difference is empty, the region is crisp because the two approximations are equal (Roy and Stell 2001).

2.3.3 Topological relationships between spatial objects with vague shapes

In the context of objects with crisp shapes, several models (Egenhofer and Herring 1990, Egenhofer and Franzosa 1991, Mark and Egenhofer 1994, Cohn et *al.* 1997) studied the specification of topological relationships in GIS and spatial databases. These models are based on two principal approaches: (1) the *point-set topology* (Egenhofer and Herring 1990) and (2) *mereology*[4]. The principles of mereology have been reminded in (Varzi 2004). First, we review these models used for characterising the topological relationships between crisp objects. Then, we present the extensions of these *traditional* models to deal with topological relationships between vague objects.

❖ **RCC model and 9-intersection model for characterising the topological relationships between crisp objects**

The *RCC* (*Region Connection Calculus*) model is based on the *mereology*. The RCC model has been presented in different papers as a tool to identify the spatial and temporal relationships (Randell and Cohn 1989, Cohn et *al*. 1997, Stell 2000). In the RCC model, the "*region*" is the only geometric primitive used to represent spatial objects; i.e. the *points* and *lines* are not considered. Moreover, a primitive relationship called "*Connection*" noted *C* is used to express a general relationship between two simple regions with crisp shapes: *C (A, B)* (*A* "*is connected*" to *B*).

[4] Region is the only geometric primitive defined (i.e. the points and lignes are not considered). The region is the elementary component of the space

Two versions of RCC model have been proposed:

- RCC-5: this model proposes five relationships between two simple regions: *DR* (*Disjoint*), *PO* (*Partial Overlap*), *PP* (*Proper Part*), *PPi* (*Proper Part inverse*) and *EQ* (*Equal*).
- RCC-8: the relationships proposed in this model can be derived from those defined the RCC-5 model (figure 2.8).

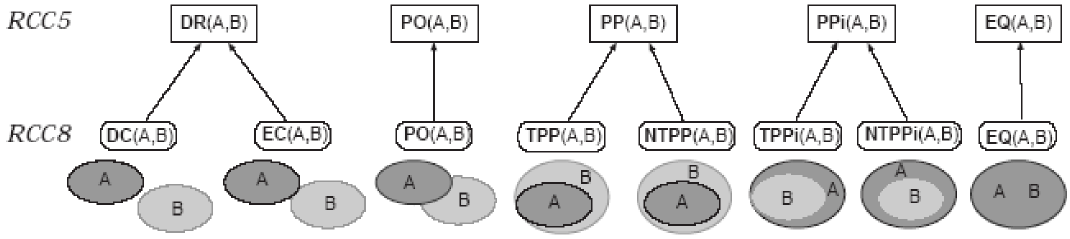


Figure 2.8 Topological relationships according to RCC-5 and RCC-8 models (Dilo 2006)

The 9-Intersection model allows an identification of topological relationships based on the principles of the point-set topology (Egenhofer and Franzosa 1991). This model is typically referenced when one speaks about the topological relationships and it has been integrated in different frameworks to specify these relationships (Chen and Li 1997). In this model, the topological relationships are identified by using 9-Intersection matrices that denote the intersections between the *boundaries*, *interiors* and *exteriors* of the objects involved. The 9-Intersection model distinguishes 8 topological relationships between two simple crisp regions (*Disjoins*, *Equal*, *Overlap*, *Contains*, *Inside*, *Covers*, *Covered by*, *Meet*), 36 relationships between two simple crisp lines, 19 relationships between a simple crisp region and a simple crisp line, 2 relationships between two crisp points, 3 relationships between a crisp point and a simple crisp line, and 3 relationships between a crisp point and a simple crisp region. This model is an extension of the 4-Intersection model (Egenhofer 1989) where only the interior and boundaries of objects are considered to identify the topological relationships. The 9-Intersection model also includes the intersections with exteriors.

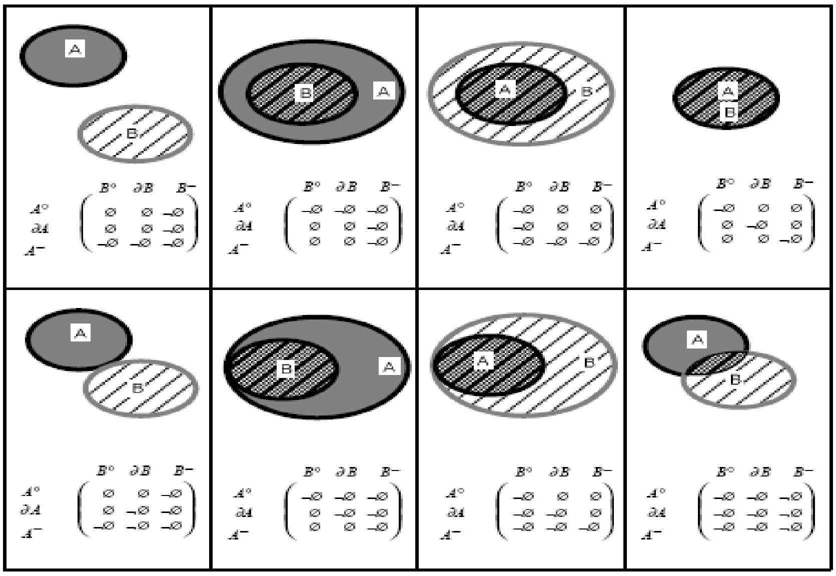


Figure 2.9 Topological relationships between two simple regions with well-defined shapes according to the 9-Intersection model (Egenhofer and Herring 1990)

- **Extensions of traditional models to deal with topological relationships between objects with vague shapes**

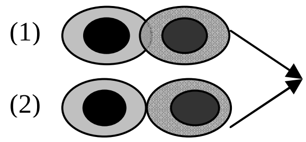
In the context of spatial objects with vague shapes, the topological relationships can be specified by extending the RCC and 9-Intersection models (Cohn and Gotts 1996a, Clementini and Di Felice 1997, Erwig and Schneider 1997, Roy and Stell 2001, Tang 2004). Erwig and Schneider (1997) used a three-valued logic to compute the topological relationships involving objects with vague shapes. Then, an intersection between two topological invariants can be *true*, *false*, or *may be* (i.e. when an uncertain part of the geometry is involved in the intersection).

In Cohn and Gotts (1996b), a topological relationship between two Egg-Yolk regions *A* and *B* is identified using a 4-Intersection matrix which enumerates four sub-relations: $R_1(Egg(A)$ - $Egg(B))$, $R_2(Egg(A)$ - $Yolk(B))$, $R_3(Yolk(A)$ - $Egg(B))$, and $R_4(Yolk(A)$ - $Yolk(B))$ (figure 2.10). These four sub-relations are those defined in RCC-5 model: *Partially Overlap* (*PO*), *Proper part* (*PP*), *Equal* (*E*), *Proper Part inverse* (*PPi*), and *Distinct* (*D*). In (Cohn and Gotts 1996b), only 46 matrices are consistent and refer to 46 topological relationships that can be drawn between two regions with broad boundaries. Figure 2.10 presents the relationship number 15 identified in Cohn and Gotts (1996b). The principal advantage of this approach relates to its simplicity to identify the topological relationships. However, it does not provide a framework to specify topological relationships involving *points*, *lines*, or *regions* with crisp shapes.

$$\begin{bmatrix} PO\ (Egg(A), Egg(B)) & PPI\ (Egg(A), Yolk(B)) \\ PO\ (Yolk(A), Egg(B)) & PO\ (Yolk(A), Yolk(B)) \end{bmatrix}$$

Figure 2.10 Identification of topological relationships in (Cohn and Gotts 1996(b))

Clementini and Di Felice (1997) introduced the concept of *approximate topological relationships* defined as relationships between regions with broad boundaries. They used a formalism based on a 3*3-Intersection matrix where the crisp boundary is replaced by a broad one. This approach considers the rules defined in Clementini and Di Felice (1997) to check the consistency of a matrix (12 rules to eliminate each matrix that cannot be drawn). Then, only 44 matrices are consistent and refer to 44 relationships which can be drawn between two regions with vague shapes. These relationships are grouped into 17 clusters that are organized in a conceptual neighborhood graph. This approach may be very useful when the topological relationships are coarsely described by the user. However, it is not sufficiently expressive when the needs are more specific and the user has a clear idea about the required relationship between regions with broad boundaries involved. For example, figure 2.11 shows an example of two different relationships which belong to the same cluster and identified by the same matrix.

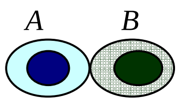


$$\textit{Meet} \longrightarrow \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Figure 2.11 Identification of the topological relationships in (Clementini and Di Felice 1997)

In the same way, Reis et *al.* (2006) reused the model proposed in (Clementini and Di Felice 1997) in order to identify the topological relationships between lines with vague shapes. In this approach, 2 conditions defined in (Clementini and Di Felice 1997) are used to reduce the number of topological relationships. Then, 5 topological relationships are distinguished between two *completely broad lines* and 77 between two *lines with broad boundaries (or endpoints).*

Tang (2004) proposed an extension of the 9-Intersection model where he identifies more topological relationships than (Clementini and Di Felice 1997) by using a 4*4-Intersection matrix. Indeed, this approach distinguishes 152 topological relationships described by 152 matrices (see the example in figure 2.12). In practice, the absence of a classification of these topological relationships reduced the utility of the model because it is very difficult to make an easy and intuitive distinction between them. Moreover, Tang (2004) does not make the distinction between the internal boundary and external one for a region with a broad boundary. Consequently, several topological relationships cannot be identified by a 4*4-Intersection matrix.



$$\begin{array}{c} \\ A° \\ llA \\ lA° \\ A^{--} \end{array} \begin{array}{c} \begin{array}{cccc} B° & llB & lB° & llB^{--} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \end{array} \quad \text{with}$$

$A°$: the interior of A

$lA°$: the boundary 's interior

llA : the boundary's boundary

A^{--} : the exterior

Figure 2.12 Identification of the topological relationships in (Tang 2004)

❖ Using Fuzzy Set theory to deal with topological relationships between objects with vague shapes

Fuzzy Set theory is also used to identify the topological relationships between objects with vague shapes (Zhan 1997, Schneider 2001**,** Somodevilla and Petry 2003, Bjørke 2004, Du et *al*. 2005**,** Dilo 2006). According to Zhan (1997), a topological relationship is called *R* (i.e. *a parameter used to replace the eight relationships identified in the 4-Intersection model* (Egenhofer 1989)). For each pair of $\alpha - cuts$ of regions involved, a sub-relation *r* is identified. Then, the *possibility* of the global relation *R* is deduced from the number of sub-relations arising between the different $\alpha - cuts$. This approach is easy to use in practice, but it presents some complexity when the $\alpha - cuts$ are non-uniformly distributed between 0 and 1. In the same way, Dilo (2006) identifies six possible topological relationships (i.e. *Disjoint*, *Touches*, *Crosses*, *Overlaps*, *Within*, and *Equal*) between two spatial objects with vague shapes. A topological relationship is defined by using fuzzy operators (e.g. *union*, *intersection*, *absolute difference*, and *bounded difference*) applied to the fuzzy subsets that define the objects involved. According to Dilo (2006), many topological relationships may exist at the same time with different *Truth degrees* (e.g. *Overlap*(*A*, *B*) with the *Truth degree* = 0,2; *Meet*(*A*,*B*) with the *Truth degree* = 0,3; *Disjoint*(*A*, *B*) with the *Truth degree* = 0,5). Du et *al*. (2005) proposed an extension of the 9-Intersection model in order to describe the fuzziness of topological relationships. Shi and Liu (2007) consider two stages to model the topological relationships: (1) giving a qualitative definition for each relationship and (2) computing each instance of this relationship by using Fuzzy Logic. In the same way, Bjørke (2004) uses a linguistic variable which gives an association to a crisp relation and a quantifier which indicates the strength of the topological relationship computed by using fuzzy operators.

The fuzzy models allow a description of the internal structure of broad parts of an object with a vague shape. However, some quantitative hypotheses are generally required in order to define the membership functions either for the computation of spatial objects or the evaluation of their topological relationships. This requirement can be considered as a limitation of the fuzzy approaches because the definition of these hypotheses is generally arbitrary; i.e. they are neither based on perception studies nor application evaluations (Bjørke 2004). Additionally, the fuzzy approaches are expensive in the implementation and more adapted to the raster data than to the vector data. In the raster data, the gradual transition of the interior or boundary of a given fuzzy object can be shown through the membership degree computed for each pixel (Clementini 2005).

In this section, we made a bibliographical study on the modeling of topological relationships between spatial objects with vague shapes. These topological relationships present relevant data that should be consistent and reliable in spatial databases. The consistency of topological relationships is generally controlled through a set of rules called the topological integrity constraints. In the next section, we review these constraints in the context of spatial databases.

2.4 Consistency of spatial databases and integrity constraints

2.4.1 Introduction

The specification of integrity constraints is an important design step in a development process of spatial databases (Borges et *al.* 2002). The integrity constraints should be respected when the database is updated in order to preserve its logical consistency (Elmasri and Navathe 2000). The logical consistency requires the specification of different types of constraints which can relate to the spatial object attributes as well as the relationships between spatial entities (*topological*, *metric*, *order*, *temporal*). According to Bédard (1987), a spatial object has *a definition*, *a thematic description* (*thematic attributes*), *a spatial extension* (*a geometry*) and *a temporal description* (*existence and geometric evolution*). All of these aspects can be concerned by integrity constraints.

2.4.2 Classification of integrity constraints

In spatial databases, the terms "*integrity*" and "*consistency*" are used to remind that the data should be *exact*, *correct*, *valid* and *consistent* (Kainz 1995). Accordingly, the integrity constraints are used to define the characteristics of valid data that can be accepted in a given database. Integrity constraints can relate to the properties of relational databases such as the uniqueness of some keys. They can also relate to semantic properties (e.g. *a house is build on 1,1 ground*), to spatial properties and relationships (e.g., *a building should not overlap a road*), or to temporal properties and relationships. The integrity constraints can be defined at the conceptual level through specific tools (Bédard et *al.* 2004).

According to Hendrik et *al.* (1997), the integrity constraints can be *intra-object* when they are defined on the attributes of only one object. In the same way, they can be *inter–objects* when they control the validity of a spatial relationship (*topological*, *metric*, *directional* and *order relationships*) between two objects.

Mehrdad Salehi (2005) proposes a formal classification of integrity constraints based on the distinctive components of spatio-temporal databases that refer to space, time, themes, and their combinations. This classification of integrity constraints is based on a classification of objects in spatio-temporal databases that has been widely used and considered as a base in developing spatio-temporal schema modeling languages such as Perceptory (Bédard *et al*. 2004) and MADS (Parent *et al*. 2006). In spatio-temporal databases, objects are classified based on their spatial, temporal, and thematic (i.e. non-spatial and non-temporal) properties and on the combinations of these properties. Objects that hold geometric attributes are usually called "spatial". Objects for which the existence is managed (e.g., their birth and death dates) and their non-spatial attribute values that evolve through time are called "temporal". "Spatio-temporal" objects are those having a geometry evolving in time. Objects that are not in these previous categories are usually called "thematic". Accordingly, Mehrdad Salehi assumes that an IC is an assertion carrying a number of concepts that are related to space, time, themes, and their combinations. These concepts are in fact used to build an integrity constraint language for spatio-temporal databases called "ICLS concepts". Based on the nature of ICSL concepts appearing in the IC assertion, an integrity constraint is then classified. Following to this, spatial-only integrity constraints and spatial integrity constraints are specialized to primary, topological, metric, ordering, and hybrid integrity constraints. Sub-classes of temporal-only and temporal integrity constraints are primary, topological, and metric integrity constraints. He distinguishes two types of spatio-temporal-only integrity constraints as inherent and hybrid. Finally, three types of spatio-temporal integrity constraints, i.e., inherent, composite, and hybrid are distinguished.

Elmasri and Navathe (2000) distinguish three categories of integrity constraints. Firstly, the *inherent* constraints refer to the rules related to the data model and not to the application. For example, the uniqueness of primary keys is an inherent constraint of the relational databases. Secondly, the *implicit* constraints are defined on the physical schema of the database by using the Data Definition Language (DDL for short). For example, the integrity constraints on the domain values are *implicit*. Thirdly, the *explicit* constraints are defined using application languages at the level of class methods. The business rules can be considered as examples of explicit constraints.

Fahrner et *al.* (1995) proposed a classification based on the *impact* of an integrity constraint on the database states. Then, an integrity constraint can be *static* when it should be checked according to a single state of the database. For example, "*the surface of an administrative region should be higher than each of its municipalities*». Likewise, the *transitional* constraints are used to restrict the number of possible transitions from one state of the database to another. For example, the constraint, "*when the data describing an administrative region are updated, its budget should never be reduced"*, is transitional. Moreover, *dynamic* constraints allow restricting sequences of transitions between possible states of a given database.

Cockcroft (1997) distinguishes three principal categories of integrity constraints in spatial databases:

- *Topological integrity constraints*: they refer to the topological relationships between spatial objects belonging to the same data collection. They can also refer to the geometrical properties of the objects without considering the meaning of geographical features involved. These constraints relate mainly to the *connectedness* and *adjacency* between geometries involved. For example, "*a polygon should be closed*" or "*objects belonging to the same collection should form a connected* *graph*” (*isolated objects are not admitted*). In (Cockcroft 1997), these constraints are inherent to the data model itself and do not need to be specified in the conceptual schema of the database.

• *Semantic integrity constraints*: these constraints are defined according to the meaning of geographical entities (Cockcroft 1997). The semantic constraints result from the combination of the geometric information, spatial relationships, and meaning of spatial objects involved. Then, the semantic constraints may contain topological conditions. For example, “*a road network should be connected”*. This integrity constraint is semantic because the definition of a road network should be considered. Moreover, the network *connectedness* is a topological condition that should be respected by this type of objects.

• *User-defined integrity constraints*: according to Cockcroft (1997), the user-defined constraints express esoteric rules defined by the domain experts. They can express legislative rules, environmental constraints, etc. For example, “*the distance between a military zone and the closest urban area should be greater than 3 km*”.

The classification of Cockcroft (1997) has its specific limitations. Firstly, metric constraints cannot be classified into one of the three categories proposed by Cockcroft (1997). In these constraints, the topological conditions are replaced by metric ones. For example, “*The distance between two polygons or two lines is defined as the minimal distance between all nodes of the objects involved*». Secondly, the semantic constraints can contain metric conditions (e.g. ***the maximum distance*** *between a house and a fire hydrant is lower or equal to* ***20 m***). Likewise, they can be purely semantic (e.g., *a house has only one owner*). Thirdly, it is difficult to distinguish between the semantic *constraints* from *user-defined ones*. Moreover, a semantic component may exist in a topological constraint especially when it verifies a topological relationship between two spatial objects.

In the data warehouses, it is also necessary to control the logical consistency of aggregations. This consistency is managed through specific constraints defined on the aggregation functions such as *min, max, sum, count* and *average* (Ross et *al.* 1998). Aggregative integrity constraints can be integrated into an optimization process which prevents the execution of an expensive computing process whether a set of data cannot be aggregated (Levy and Mumick 1996).

Some simple integrity constraints can be represented in the conceptual schema of a given data warehouse. For example, an aggregation association between two hierarchical levels of a spatial dimension means that the members of an intermediate level has only one *parent* member at the immediately higher level (e.g.,*" Montpellier belongs only to the South_France category"*). In this context, *Perceptory* is a design tool which provides a set of pictograms extending the *Unified Modeling Language* (UML) in order to establish the conceptual schema of a spatial data warehouse (Bédard 2006).

Salehi (2005) aims also at specifying complex integrity constraints in a spatial data warehouse. In this context, intra-level topological relationships (*between spatial objects of the same hierarchical spatial dimension level*) or inter-level ones (*between spatial objects belonging to different hierarchical levels of the same spatial dimension*) should be controlled through specific integrity constraints. These constraints are often difficult to be managed since the geometrical data stored in the different levels result from an integration process involving several heterogeneous data sources. More specifically, topological constraints should consider the uncertainty about the appropriate intra-level and inter-level topological relationships between the integrated geometries that can be vague. Accordingly, Frank (2001) and Rodriguez (2005) proposed the implementation of *tolerant* constraints that considers the shape vagueness of data resulting from an integration process. Figure 2.13 shows the integration of different geometric representations (of two spatial objects *A* and *B*) loaded from two different data sources. According to Rodriguez (2005), the integration result is partially consistent because final geometries have vague shapes. Then, the maximal *consistency* of these geometries is obtained by the intersection. However, the *minimal consistency* is that obtained by the unions of source geometries of *A* and *B*.

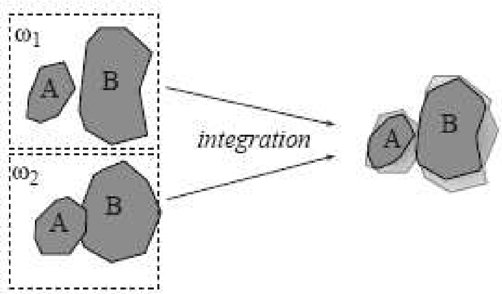


Figure 2.13 Integration of two heterogeneous source geometries (Rodriguez 2005)

The implementation of integrity constraints in a database is preceded by a formal specification done through specific languages or representations that we review in the next section.

2.4.3 Formal specification of spatial integrity constraints

2.4.3.1 First-order logic based languages

A formula of first-order logic can be made up of symbols representing *variables*, *constants*, *predicates*, *functions*, *quantifiers* and *logical connectors* (Dehornoy 2006). First-order logic has been used to specify integrity constraints as in languages used to model the knowledge in artificial intelligence (Reiter 1987). For example, the constraint *TC3:* "*a person gender should correspond to one of the following values: male or female*" is expressed as follows:

$$(\forall x)\, person(x) \supset male(x) \vee female(x)$$

In spatial domain, Hadzilacos and Tryfona (1992) used the first-order logic and the 4-intersection model to specify spatial integrity constraints. For example, the constraint *C3*: "*parcels should not intersect buildings*" is expressed as follows:

DEFINE CONSTRAINT CONSTRAINT
CONSTRAINT_IN_BUILDING_BLOCK
AMONG (LANDPARCEL, BUILDING_BLOCK)
AS r_6 (LANDPARCEL, BUILDING_BLOCK) OR
r_7 (LANDPARCEL, BUILDING_BLOCK)

With r_6 and r_7 refer to the following topological relationships: "*Contains*" and "*Covered by*" defined in the 4-Intersection model, respectively.

However, the first-order logic based languages are generally difficult to be used to express the integrity constraints. Long formulas are required to specify the integrity constraints because their syntaxes are often limited. Benzaken and Doucet (1993) used the object-oriented concepts through a specific programming language called *THEMIS*. The integrity constraints are implemented as methods written using this language.

2.4.3.2 Visual specification of spatial integrity constraints

According to Proulx et *al.* (1995), a visual language requires the use of visual expressions (e.g. icons, diagrams) to formally express a topological integrity constraint. The main advantage of a visual language relates to its facility of use. These languages can be useful to help novice users to express simple topological integrity constraints. For example, CIGALES is a visual language proposed by Calcinelli and Mainguenaud (1994) in order to express simple spatial queries.

Servigne et *al.* (2000) proposed a visual interface to define the topological integrity constraints. In this approach, a topological integrity constraint is defined in three phases using a visual interface. The first phase is to choose the objects involved, before specifying a spatial relationship and setting a specification related to the validity of this relationship. Then, the general form of a spatial integrity constraint is presented as follows:

Constraint = (Class Object 1, Relation, Class Object 2, Specification)

The argument *"specification"* can refer to a *prohibition*, an *authorization, the maximum number of occurrences,* etc. The main advantage of this approach lies in its simplicity and intuitive use. For example, the constraint *C4*: "*a parcel should not be crossed by a road*" can be expressed as follows:

(Parcel, crossed by, Road, forbidden)

In the same way, Erwig and Schneider (2003) proposed a visual language to specify valid topological relationships between spatio-temporal objects. The logical consistency of such relationships is verified through a set of graphs that describe their valid evolutions (Erwig and Schneider 2003).

According to Proulx et *al.* (1995), visual languages have various limitations related principally to a lack of normalization symbols and pictograms used in the interfaces depend on cultural aspects (e.g. some symbols change from one country to another). Moreover, it is generally difficult to specify in the same integrity constraint two topological constraints involving the same objects.

2.4.3.3 Tabular specification

Normand (1999) proposed a tabular approach based on the formalism defined in (Government of Canada 1996) in order to express the spatial integrity constraints. This approach consists in exploring the constraint description given by the expert in order to represent its necessary elements in the cells of a related table. Table 1 shows the tabular specification of the following spatial constraint: "*a stream may be adjacent to a river whether its endpoints are on the boundary of this river. In the other cases, it should be adjacent to two other streams*".

Table 2.1 Tabular specification of integrity constraints

<table>
<tr><th>Operator</th><th>Relations</th><th>Cardinalities</th><th>Objets</th><th>Dimensions of objects involved</th></tr>
<tr><td></td><td>Equality</td><td>0-0</td><td>River</td><td>1</td></tr>
<tr><td rowspan="2">Or</td><td>Disjunction</td><td>-</td><td rowspan="4">Stream</td><td rowspan="2">1</td></tr>
<tr><td>Adjacency</td><td>1-2</td></tr>
<tr><td rowspan="2">Or</td><td>Disjunction</td><td>-</td><td rowspan="2">-</td></tr>
<tr><td>Adjacency</td><td>1-2</td></tr>
</table>

2.4.3.4 Spatial Extension of Object Constraint Language (*OCL*)

OCL (Waremer and Kleppe 1998) is a formal language mainly used for the specification of integrity constraints; this language has been integrated in UML. In OCL, the integrity constraints are expressed through the notion of "*invariants".* An *invariant* refers to a condition which should be always satisfied by each instance of a given class. The constraint *context* refers to the element of the conceptual schema on which the constraint is defined: *a class, an interface* or *a type* defined in the UML class diagram. OCL is based on the principle of "*navigation*". This principle relates to the possibility of defining constraints involving different classes related to the *context* class.

For spatial integrity constraints, Duboisset et *al.* (2005) and Pinet et *al.* (2004) proposed an extension of the meta-model of OCL. A new generic type called *BasicGeoType* has been proposed in order to integrate geographical data types in OCL. Moreover, new functions have been defined in order to introduce topological operators as additional syntax elements of OCL. These operators find their theoretical background in (Egenhofer and Franzosa 1991). In the case of topological relationships between crisp regions, eight operators have been defined where each allows the identification of a topological relationship proposed in the 9-Intersection model. According to this approach, a spatial integrity constraint is an *invariant* defined for a given *context* class. For example, let the constraint *C4*: "*buildings and roads should be disjoint or adjacent*". *C4* can be expressed as follows:

```
Context road inv :
Building.allInstances→forAll(b|Self.geometry→aredisjoint(b) or self.geometry→areAdjacent(b))
```

2.5 Conclusion

The spatial imperfection is an inherent property of spatial data. In Section 2.2, we stressed the diversity of the contributions around the question of spatial data imperfections. Several taxonomies such as (Bédard 1987, Parsons 1996, Smets 1996, Smithson 1989, Fisher 1999a, Worboys 1998, Hazarika and Cohn 2001) have been proposed to classify these imperfections according to various points of view: the origin and nature of imperfection, the nature of objects involved (i.e. *well-defined*, *ill-definite*), the factors causing a deficiency of data quality, *etc*. Bédard (1987) studied the forms and levels of uncertainties in a spatial object description (Bédard 1987). Then, we reviewed the management of uncertainty in spatial databases and data warehouses. We concluded that the data quality of these databases is directly affected by different forms of spatial imperfections.

In this thesis work, we are specifically interested in the logical consistency of spatial objects with vague shapes and of their topological relationships. The logical consistency is controlled through integrity constraints which represent a set of rules specified at the conceptual level and applied to the data in order to prevent the inconsistencies in a given spatial database. The definition and application of integrity constraints can be affected by various forms of imperfection such as concepts ambiguity, shape vagueness, and inaccuracy of the quantitative information checked by these constraints. Among these forms of imperfection, we are interested in the shape vagueness that may characterise the geometry of some spatial objects such as pollution zones. Representing spatial objects with vague shapes requires the use of a specific spatial model which allows a more reliable description of reality. This model presents the background of any approach aiming at the management of integrity constraints for spatial objects with vague shapes and their topological relationships.

In Section 2.3, we studied related works to the problem of modeling spatial objects with vague shapes. These models can be grouped in two principal categories. First, exact models extend concepts and structures of models defined for crisp objects in order to represent spatial objects with vague shapes (Burrough and Frank 1996, Cohn and Gotts 1996(a), Clementini and Di Felice 1997, Roy and Stell 2001). The advantage of these models lies in their *low* development cost. However, the existing exact models do not represent spatial objects with *partially* vague shapes. For example, a lake can be surrounded by a broad boundary (swamp banks) on one side and a linear boundary on the other side (rocky banks) at the same time. The existing exact models consider that a broad boundary should correspond to a closed and connected polygonal zone that surrounds the interior of the region involved. These models consider these regions as invalid. Moreover, the topological relationships between such objects with vague shapes cannot be computed through the existing exact models.

The second category of models (Dilo et *al.* 2005, Schneider 2001, Zhan 1997, Worboys 1998(b), Roy and Stell 2001, Tang 2004, Pfoser and Tryfona 2005) includes approaches based on mathematical theories such as Rough Sets theory (Pawlak 1994) or Fuzzy Logic (Zadeh 1965). Fuzzy Logic has been used in the principal proposals in this category of models. The fuzzy approaches have the advantage of modeling the internal structure of vague parts of a given object. For example, a fuzzy approach generally allows to compute the membership degree of each point inside the broad boundary of a region. However, these approaches are *expensive* in terms of implementation as well as they require the setting of some quantitative assumptions necessary to define the membership functions.

Section 2.4 presents a literature review on the classification of integrity constraints and their formal specification. We concluded the absence of an approach which allows the specification and implementation of topological constraints involving spatial objects with vague shapes and their topological relationships. Existing methods for integrity constraints modelling do not support spatial objects with vague shapes. The fuzzy approaches provide a quantitative evaluation of shape vagueness of this type of objects and of their spatial relationships. Consequently, De Tré et *al.* (2004) proposed an extension of the notion of *generalized constraints* (presented in (Zadeh 1965)) in order to model a partial satisfaction of integrity constraints involving spatial objects with vague shapes. An integrity constraint is partially respected when it is satisfied with a membership degree between 0 and 1. However, the fuzzy approaches present different limitations discussed in section 2.3.

References

AALDERS, H.J.G.L. 2002, The Registration of Quality in a GIS. *Spatial Data Quality* (W. Shi, P. Fisher, et M.F. Goodchild, Eds, Taylor & Francis, p. 186-199.

ABITEBOUL, S., R. HULL ET V. VIANU (traduit par Patrick Cegielski) 2000, Page : 78 Fondements des Bases de Données, Addison Wesley, p. 715, ISBN 2-7117-8645-5.

AHLQVIST, O., J. KEUKELAA AND A. OUKBIR, 1998,. Using Rough classification to Represent Uncertainty in Spatial Data. In Proceedings of the Tenth Annual Colloquium of the Spatial Information Research Cent*re*. P. Firns (ed). 16 - 19 Dec, Dunedin, New Zealand. University of Otago, ISBN 1877139122. pp.1-10.

AZOUZI, M., 2000, Suivi de la qualité des données spatiales au cours de leur acquisition et de leurs traitements, Thèse de doctorat en sciences techniques EPFL, Ecole technique polytechnique fédéral de Lausanne.

AZOUZI, M., 1999, Introducing the concept of reliability in spatial data. In Kim Lowell and Annick Jaton, editors, *Spatial Accuracy Assessment: land information uncertainty in Natural Resources*, pages 139–144. Ann Arbor Press.

BENZAKEN, V. AND A. DOUCET 1993, Thémis: a database programming language with integrity constraints. DBPL 1993: 243-262.

BÉDARD Y., 1987, Uncertainties in Land Information Systems Databases. Proceedings of Eighth International Symposium on Computer-Assisted Cartography. Baltimore, Maryland (USA), 29 Mars - 3 Avril 1987, American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping, p. 175-184.

BEDARD Y. AND D. VALLIERE 1995, Qualité des données à référence spatiale dans un contexte gouvernemental. Rapport de recherche sur la mise en place d'une méthode d'évaluation de la qualité des données à référence spatiale préparé pour le plan géomatique du Gouvernement du Québec, Université Laval, Québec, Canada.

BÉDARD, Y., 1999, Principles of Spatial Database Analysis and Design. In GIS: Principles, Techniques, Applications & Management, Wiley, 2nd Ed., Chap. 29, pp. 413-424.

BÉDARD, Y, S. LARRIVÉE, M.J. PROULX & M. NADEAU, 2004, Modeling Geospatial Databases with Plug-Ins for Visual Languages: A Pragmatic Approach and the Impacts of 16 Years of Research and Experimentations on Perceptory, S. Wang et al. (Eds.): Conceptual Modeling for Geographic Information Systems (COMOGIS) Workshop ER2004, LNCS 3289, pp. 17–30.

BEDARD, Y. 2006, Notions de bases de données avancées SIG, Notes de cours, Département de sciences géomatiques, Université Laval.

BEAUBOUEF, T. AND F. PETRY 2001, Vague regions and spatial relationships: a rough set approach. In Proc. Of the 4th International Conference on Computational Intelligence and Multimedia Applications (ICCIMA ’01), Yokosuka City, Japan, October 30-November 1.

BEJAOUI, L., BÉDARD, Y., PINET, F., SALEHI, M., SCHNEIDER, M., 2007, Logical consistency for vague spatiotemporal objects and relations. In: the 5th International Symposium on Spatial Data Quality (ISSDQ 2007) Enschede, Netherlands, June.

BEJAOUI, L., BÉDARD, Y., PINET, F., SCHNEIDER, M., 2008, Qualified topological relations between objects with possibly vague shape. In: International Journal of Geographical Information Sciences, Francis & Taylor, *to appear*.

BERGLUND, G., 1993, Epistemic probability and epistemic weight, Department of Statistics, University of Stockholm, available at http://www.internetional.se/pgbepist.pdf, last visit (25-06-2008)

BITTNER, T., 2000, A qualitative formalization of built environments. In: Proceedings of DEXA2000, LNCS Vol 1873, Springer-Verlag, Berlin-Heidelberg.

BJØRKE, J. T., 2004, Topological Relations Between Fuzzy Regions: Derivation of Verbal Terms. *Fuzzy sets and systems*, **141**, 449-467.

BOIN, A.T. AND G.J. HUNTER 2006, Do spatial data consumers really understand data quality information?. In. Proc. of 7th international Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. 5 - 7 July 2006, Lisbon, Portugal, pp : 215-235.

BORDOLOI, U., D.L. KAO AND H.W. SHEN 2004, Visualization techniques for spatial probability density function data. Data Science Journal 3: 153-162.

BORGES, K.A.V., J.R. DAVIS AND A.H.F LAENDER 2002, Integrity Constraints in Spatial Databases. In: Door, J, H.; Rivero, L. C.. (Org.). Database Integrity: Challenges and Solutions. Hershey, PA: Idea Group Publishing, 2002, p. 144-171.

BRASSEL, K., F. BUCHER, E.M. STEPHAN AND A. VCKOVSKI 1995, Completness. In: Guptill, S.C. and Morrison, J.L. (eds.), Elements of spatial data quality. Elsevier Science inc., New York, pp. 81-107.

BRODEUR, J., Y. BÉDARD, G. EDWARDS ET B. MOULIN 2003, Revisiting the Concept of Geospatial Data Interoperability within the Scope of Human Communication Processes, Transactions in GIS, Vol. 7, No. 2, pp. 243-265.

BURROUGH, P.A. 1989, Fuzzy mathematical methods for soil survey and land evaluation. *Journal of Soil Science*, 40, 477–492.

BURROUGH, P.A 1996, Natural Objects with Indeterminate Boundaries. In G*eographic Objects with Indeterminate Boundaries*, GISDATA Series, vol. 3, Taylor & Francis, pp. 3-28, 1996.

BURROUGH, P.A. AND A.U. FRANK 1996, *Geographic objects with indeterminate boundaries*. Number 2 in GISDATA. Taylor & Francis, London.

CALCINELLI, D., AND MAINGUENAUD, M., 1994, Cigales, a Visual Query Language for a Geographical Information System: the User Interface. J. Vis. Lang. Comput. 5(2): 113-132.

CLEMENTINI, E., 2005, A model for uncertain lines. *Journal of Visual Languages and Computing* 16(4): 271-288.

CLEMENTINI, E., AND DI FELICE, P., 1995, A Comparison of Methods for Representing Topological Relationships. *Information Sciences* 3: 149-178.

CLEMENTINI, E., AND P., DI FELICE, 1997, Approximate topological relations. International Journal of Approximate Reasoning, 16:173-204.

CHEN, J., AND C., LI 1997, Improving 9-Intersection Model by Replacing the Complement with Voronoi Region. In Proceedings of Inter. Workshop on “Dynamic of Multi-dimensional GISs”, Hong Kong,1997, 36-48.

CHRISMAN, N.R., 1991, The error component in spatial data. In: Maguire, D.J., Goodchild, M.F and Rhind, D.W. (eds.), Geographical Information Systems: Principles and Applications. Volume 1, p. 165-174.

COCKCROFT, S., 1997, A Taxonomy of Spatial Data Integrity Constraints. Geoinformatica **1**(4): 327-343.

COHN, A.G. AND N.M. GOTTS 1996a, Representing spatial vagueness: A mereological approach. In Carlucci Aiello, L., Doyle, J., Shapiro, S., eds.: KR'96: Principles of Knowledge Representation and Reasoning. Morgan Kaufmann, pp. 230—241.

COHN, A.G. AND N.M. GOTTS 1996b, The 'egg-yolk' representation of regions with indeterminate boundaries in: Burrough, P. & Frank, A. M. (editors) Proceedings of the GISDATA Specialist Meeting on Spatial Objects with Undetermined Boundaries, pp. 171-187 Francis Taylor.

COHN, A.G., B. BENNETT, J. GOODAY AND N.M. GOTTS 1997, Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. GeoInformatica, 1(3):275–316.

COUCLELIS, H., 1996, 'Towards an operational typology of geographic entities with illdefined boundaries', in *Geographic Objects with Indeterminate Boundaries*, eds., P. Burrough and A.M. Frank, GISDATA, pp. 45–55. Taylor & Francis.

DASSONVILLE, L., VAUGLIN, F., JAKOBSSON, A., AND LUZET, C., 2002, Quality Management, Data Quality and Users, Metadata for Geographical Information. In *Spatial Data Quality*, edited by W. Shi, P.F. Fisher, and M.F. Goodchild, (Taylor & Francis), pp. 202-215.

DAVID, B. AND P. FASQUEL 1997, Bulletin d'information de l'IGN - Qualité d'une base de données géographique: concepts et terminologie, N. 67, IGN France.

DE GROEVE, T., LOWELL, K., 2000, Improving local forest volume estimates by fusion of multi-temporal forest type maps. Environmental Modelling and Software (ENVSOFT) 15(4):373-385.

DEHORNOY, P., 2006, Logique et théorie des ensembles. Chapitre 7: Logique du premier ordre. *Notes de cours, FIMFA ENS, version 2005-2006.*

DEMIRKESEN A. AND B. SCHAFFRIN, 1996, Map Conflation : Spatial point data merging and transformation, GIS/LIS'96, Denver, pages 393-404.

DEVILLERS, R., 2004, Conception of a Multidimensional Information System for Geospatial Data Quality Management and Communication. Geomatics Sciences. Quebec City, Laval University. Ph.D.Thesis**.**

DEVILLERS, R., Y. BÉDARD AND R. JEANSOULIN 2005, Multidimensional management of geospatial data quality information for its dynamic use within GIS. Photogrammetric Engineering & Remote Sensing, 71(2): 205-215.

DEVILLERS, R. AND R. JEANSOULIN, 2005, "Qualité de l'information géographique: concepts généraux", *in: Lavoisier (Ed.), ``Qualité de l'information géographique", Hermès Sciences traité IGAT, série Géomatique, Paris*, octobre.

DEVOGEL, T., 1997, Processus d'intégration et d'appariement de bases de données géographiques: application à une base de données routière multi-échelle. Thèse de doctorat, soutenu le 12 décembre 1997.

DEMUTH, B., HUSSMANN, H., 1999, Using UML / OCL Constraints for Relational Database Design, Proceedings of the Conference on the Unified Modelling Language (UML'99), Fort Collins, USA, Lecture Notes in Computer Science, vol. 1723, 28-30, October, pp. 598-613.

DEMUTH, B., HUSSMANN, H., LOECHER, S., 2001, OCL as a Specification Language for Business Rules in Database Applications, Proceedings of the Conference on the Unified Modelling Language (UML 2001), Toronto, Canada, Lecture Notes in Computer Science, vol. 2185, October, pp. 104-117.

DILO, A., R.A. BY, ET AL., 2004, Definition and Implementation of Vague Objects. In the Proceeding of the ISSDQ '04, GeoInfo Series, Bruck/Leitha, Austria.

DILO, A., R. A. BY, ET AL. 2005, A Proposal for Spatial Relations between Vague Objects. In the Proceeding of the International Symposium on Spatial Data Quality ISSDQ'05, Beijing, China, pp 50-59, August 25-26.

DILO, A. 2006, "Representation of and reasoning with vagueness in spatial information: A system for handling vague objects", thesis, 187p.

DU, S., QI, Q., WANG, Q. AND LI, B., 2005, Fuzzy description of topological relations I: a unified fuzzy 9-intersection model. *In: L. Wang, K. Chen, Y.S. Ong (Eds.), Advances in Natural Computation*, Lecture Notes in Computer Science, vol. 3612, pp. 1260-1273.

DUBOISSET, M., F. PINET, M.-AH KANG AND M. SCHNEIDER 2005, Precise Modeling and Verification of Topological Integrity Constraints in Spatial Databases: From an Expressive Power Study to Code Generation Principles. ER 2005: 465-482.

DUCKHAM, M., K. MASON, J. STELL AND M. WORBOYS 2001, A formal ontological approach to imperfection in geographic information. Computer, Environment and Urban Systems, 25, 2001, pp. 89-103.

DUTTA, S. 1991, Approximate Spatial Reasoning: Integrating Qualitative and Quantitative Constraints. International Journal of Approximate Reasoning 5, pp. 307-331.

DUTARTE 2006, http://dutarte.club.fr/Sitestat/HISTOIRE%20STATISTIQUE%201.htm, dernière visite (10-10-2006).

DYRESON, C.E. AND R.T. SNODGRASS 1993, Valid-time Indeterminacy. In Proceedings of the Ninth International Conference on Data Engineering (ICDE), Vienna, Austria, April, pp. 335-343. Winner of Best Paper Award.

EGENHOFER, M.J. 1989, A formal definition of binary topological relationships, in W.Litin and H.J.Scheck(eds.), In proceedings of the third international conference on Foundations of Data Organisation and Algorithms (FODO), lecture notes in computer science 367 (springer-verlag NY), pp. 457-472.

EGENHOFER, M. J. AND R.D. FRANZOSA 1991, Point-set Topological Relations. International journal of geographical Information Systems, 5(2), 161-174.

EGENHOFER, M. AND J. HERRING, 1990, A mathematical framework for the definition of topological relationships. Proceedings of the Fourth International Symposium on Spatial Data Handling, Zurich, Switzerland (eds. K. Brassel and H. Kishimoto), 803--813.

ELMASRI, R. AND S. B. NAVATHE 2000, Fundamentals of Database Systems. Addisom-Wesley.

ERWIG, M. AND M. SCHNEIDER 1997, Vague regions. In 5th International Symposium on Advances in Spatial Databases (SSD'97), number 1262 in Lecture Notes in Computer Science, pp. 298--320.

ERWIG, M. AND M. SCHNEIDER 2003, A visual langage for the evolution of spatial relationships and its translation into spatio-temporal calculus. Journal of Visual Languages and Computing, 14, 181-211.

FAHRNER, C., T. MARX AND S. PHILIPPI 1995, Integration of Integrity Constraints into Object Oriented database schema according to ODMG-93 (RR-9-95), University of Koblenz.

FAO, 2001, Détermination des zones prioritaires pour la lutte contre la trypanosomie au moyen de données satellite et de la logique floue. Food and Agricultral Organisation.

FISHER, P.F. 1999a, Models of uncertainty in spatial data. Geographical Information Systems (P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind, Eds), John Wiley & Sons, New-York, p. 191-205.

FISHER P.F. 1999b, "Set Theoritic Considerations in the Conceptualization of Uncertainty in Natural Resource Infomation". *Spatial Accuracy Assessment, Land Information Uncertainty in Natural Ressources* (K. Lowell, and A. Jaton, Eds), Quebec, Ann Arbor Press, p. 147-150.

FONSECA, F.T., C. DAVIS, ET G. CÂMARA 2003, Bridging Ontologies and Conceptual Schemas in Geographic Information Integration. GeoInformatica. v.7, n.4, p.355-378. Kluwer Academic Publishers.

FRANK, A.U. 2001, Tiers of ontology and consistency constraints in geographical information systems. In: *International Journal of Geographical Information Science*, 15(7):667--678.

FRANKLIN, C., 1992, An Introduction to Geographic Information Systems: Linking Maps to databases. Database. pp. 13-21.

GODFREY, P., J. GRANT, ET AL. 1997, Integrity constraints: semantics and applications**:** 1-46.

GODJJEVAC, J. 1999, Idées nettes sur la logique floue. Presses polytechniques et universitaires romandes, Lausanne.

GOODCHILD, M.F. 1995a, Attribute Accuracy. In: Guptill, S.C. and J.L. Morrison (eds.), Elements of spatial data quality, Elsevier Science inc., New York, pp. 59-79.

GOODCHILD, M. 1995b, Sharing imperfect data. In H. Onsrud, & G. Rushton, Sharing geographic information (pp. 413±425). New Jersey: Rutgers.

GOODCHILD, M. AND R. JEANSOULIN, 1998, "Data Quality in Geographic Information: from Error to Uncertainty", *Hermés, Paris*, 192 pages.

Gouvernement du Canada, 1996, Normes et spécifications de la Base Nationale de données topographiques, édition 3.1. Géomatique Canada.

GREGAN, M. 2004, Tutorial : Oracle Spatial, http://www.iict.ch/Tcom/Presentations/EI2004/Oracle_Spatial_Tutorial.pdf

GUARINO, N. AND C. WELTY 2000, Ontological analysis of taxonomic relationships. In: Laender A, Storey V (eds) Proceedings of ER-2000: The International Conference on Conceptual Modeling, Lecture Notes in Computer Science (LNCS) Vol 1920, Springer-Verlag, Berlin, pp. 210-224.

GUTING, R.H. 1994, An Introduction to Spatial Database Systems, VLDB Journal 3, pp. 357-399.

GUPTILL S.C. AND J.L. MORRISON 1995, Spatial data quality. Elements of spatial data quality (S. C. Guptill and J. L. Morrison, Eds), Elsevier Science inc., New York.

FRANK, A. 1987, Overlay processing in spatial information systems, Auto Carto 8, Baltimore.

HADZILACOS, T. AND N. TRYFONA 1992, A model for expressing topological integrity constraints in geographic databases. In Theories and Methods of SpatioTemporal Reasoning in Geographic Space, volume 639. Springer-Verlag, LNCS.

Harvey F and F. Vauglin, 1996, Geometric Match-processing : Applying Multiple Tolerances, Spatial Data Handling (SDH), Delft (Pays-Bas), Kraak et Molenaar (Eds.), pages 155-171.

Hazarika, S.M. et A.G. Cohn (2001). A taxonomy for spatial vagueness, an alternative egg-yolk interpretation. In Proceeding of the COSIT 2001, The fifth international Conference On Spatial Information Theory (ed. Montello, D.R.), Californie, Springer-Verlag Berlin Heidelberg. Lecture Notes in Computer Science, 2205, pp. 92-107.

HEUVELINK, G. B. M. 1998, Error propagation in environmental modelling with GIS. London: Taylor & Francis.

HUNTER, G.J., 1998, Managing Uncertainty in GIS, *NCGIA Core Curriculum in GIScience*, http://www.ncgia.ucsb.edu/giscc/units/u187/u1871.html, posted February 03.

HURTADO, C., C. GUTIERREZ AND A. MENDELZON, 2003, *Capturing summarizability with integrity constraints in OLAP*. In Technical Report, Departamento de Ciencias de la Computacin, Universidad de Chile, TR/DCC-2003-6.

HWANG, S. AND J-C. THILL, 2005, Modeling Localities with Fuzzy Sets and GIS, In: Cobb M, Petry F, and Robinson V (eds) Fuzzy Modeling with Spatial Information for Geographic Problems, Springer-Verlag, pp. 71-104.

INMON, W.H., 1992, Building the Data Warehouse. John Wiley.

ISO/TC211, 2002, ISO 19113:2002 Geographic information — Quality principles, International Organization for Standardization.

JENSEN, C.S AND C.E. DYRESON 1998, A Consensus Glossary of Temporal Database Concepts - February 1998 Version, in O. Etzion, S. Jajodia, and S. Sripada, editors, Temporal Databases: Research and Practice, Lecture Notes in Computer Science 1399, Springer-Verlag, pp. 367-405.

JURAN J.M., F.M.J. GRYNA, AND R.S. BINGHAM 1979, Quality Control Handbook , New-York, McGraw-Hill, 3rd Edition.

KAINZ, W., 1995, Logical consistency. In: Guptill, S.C. and Morrison, J.L. (eds.), Elements of spatial data quality. Elsevier Science inc., New York, pp. 109-137.

KEMP, K.K. 2008, *Encyclopedia* of Geographic Information Science. Sage, Los Angeles, 584p.

KIMBALL, R. 1996, The data warehouse toolkit : practical techniques for building dimensional data warehouses, New York : John Wiley & Sons.

KNIGHTBRIDGE SOLUTIONS 2006, Top 10 Trends in Business Intelligence for 2006. White Paper, accessed on BitStream, Septembre 26th, 2006, 12 p.

LENZ, H.-J. AND A. SHOSHANI 1997 *Summarizability in OLAP and statistical data bases*. In SSDBM '97: Proceedings of the Ninth International Conference on Scientific and Statistical Database Management, pages 132--143. IEEE Computer Society.

LEVESQUE, M.-A., Y. BÉDARD, M. GERVAIS, R. DEVILLERS, 2007, Towards managing the risks of data misuse for spatial datacubes, Proceedings of the 5th International Symposium on Spatial Data Quality, June 13-15, Enschede, Netherlands.

LEVY, A. AND I. MUMICK 1996, Reasoning with aggregation constraints. In the Proceedings of the 5th International Conference on Extending Database Technology: Advances in Database Technology, Avignon, France, March 25-29.

MALINOWSKI, E. AND ZIMÁNYI, E. 2005, Spatial hierarchies and topological relationships in the Spatial MultiDimER model. In *Proc. of the 22nd British Nat. Conf. on Databases*, pages 17–28.

MALINOWSKI, E. AND ZIMÁNYI, E., 2007, Implementing spatial datawarehouse hierarchies in object-relational DBMSs. ICEIS (1) 2007: 186-191.

MIRALLES, A. 2006, Ingénierie des modèles pour les applications environnementales. Thèse de doctorat. Université Montpellier II. Montpellier. 322p.

MORRIS, A. 2003, A Framework for Modeling Uncertainty in Spatial Databases. Transactions in GIS, January 2003, vol. 7, no. 1, Blackwell Publishing Ltd, Oxford., pp. 83-101(19).

MOWRER, H. T. 1999, "Accuracy (Re)assurance: Selling Uncertainty Assessment to the Uncertain". *Spatial Accuracy Assessment, Land Information Uncertainty in Natural Ressources* (K. Lowell, and A. Jaton, Eds), Quebec, Ann Arbor Press, p. 3-10.

MOSTAFAVI, M. A., G. EDWARDS, et *al*. 2004, An Ontology-Based Method for Quality Assessment of Spatial Data Bases. ISSDQ'04, GeoInfo Series, Austria.

MOTRO, A. 1995, Imprecision and uncertainty in database systems. In: Bosc P, Kacprzyk J, editors. Fuzziness in database management systems. Heidelberg: Physica-Verlag; 1995. p 3-22.

NAVRATIL G. ET A. FRANK 2006, Quality of Spatial Data for e-Government from an Ontological View. Vortrag: FIG Workshop on eGovernance, Knowledge Management and eLearning, Budapest, Hungary; 27.04.2006 - 28.04.2006; in: "Workshop on eGovernance, Knowledge Management and eLearning", B. Markus (Hrg.); College of Geoinformatics, 963-229-423-8; S. 106 - 116.

NORMAND, P. 1999, Modélisation des contraintes d'intégrité spatiale, théorie et exemples d'applications. Département des sciences géomatiques. Québec, Université Laval.

PARENT, C., S. SPACCAPIETRA, E. ZIMANYI, P. DONINI, C. PLAZANET, C. VANGENOT, N. ROGNON, AND P.-A. CRAUSAZ 1997, MADS, modèle conceptuel spatio-temporel. Revue Internationale de Géomatique. 7(3-4):317-352, December.

PARSONS, S. 1996, Current approaches to handling imperfect information in data and knowledge bases. *Knowledge and Data Engineering*, 8(3):353–372.

PARSONS, S. AND A. HUNTER 1998, A review of uncertainty handling formalisms. In *Applications of Uncertainty Formalisms*, pages 8–37.

PAWLAK, Z. 1994, Rough sets: present state and further prospects. In Third International Workshop on Rough Set and Soft Computing (RSSC '94), pages 72–76, November.

PBESMA, E.J., D. KARSSEMBERG AND K.D JONG 2006, Dynamic visualisation of spatial and spatio-temporal probability distribution functions. In. Proc. of 7th international Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. 5 - 7 July 2006, Lisbon, Portugal, pp : 825-839.

PINET, F., M.-AH KANG AND F. VIGIER 2004, Spatial Constraint Modelling with a GIS Extension of UML and OCL: Application to Agricultural Information Systems. Metainformatics 2004: 160-178.

Pinet, F., Duboisset, M. and V. Soulignac (2007) : Using UML and OCL to maintain the consistency of spatial data in environmental information systems. In: Environmental modeling & software, 22(8), pp. 1217-1220.

PFOSER, D. AND C.S. JENSEN 1999, Capturing the Uncertainty of Moving-Object Representations. In Proc. Intl. Conf. on Large Spatial Databases (SSD), pages 111–132.

PFOSER, D. AND N. TRYFONA 2001, Capturing fuzziness and uncertainty of spatiotemporal objects, Proceedings of the 5th East European Conference on Advances in Databases and Information Systems, 149Y162.

PFOSER, D., N. TRYFONA AND C.S JENSEN 2005, Indeterminacy and Spatiotemporal Data: Basic Definitions and Case Study. GeoInformatica 9(3): 211-236.

PROULX, M.J., Y. BEDARD & B. MOULIN 1995, Développement d'un nouveau langage d'interrogation de base de données spatio-temporelle, La 7e Conférence internationale sur la géomatique, on CD-Rom, 16p., 11-15 juin.

RAFANELLI, M. 2003, Multidimensional Databases: Problems and Solutions, Idea Group.

RANDELL, D.A. AND A.G. COHN, 1989, Modelling topological and metrical properties of physical processes. In Brachman, R.J., Levesque, H.J. and Reiter, R. editors, Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning (KR'89), pages 357–368. Morgan Kaufmann.

REITER, R., 1987, On integrity constraints. In Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge, pages 97--111, Pacific Grove, CA.

REIS, R., M.J. EGENHOFER AND J. MATOS 2006, Topological relations using two models of uncertainty for lines. In. Proc. of 7th international Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. 5 - 7 July 2006, Lisbon, Portugal, pp : 286-295.

RIVEST, S., Y. BÉDARD, M.J. PROULX AND M. NADEAU 2003, SOLAP: a new type of user interface to support spatio-temporal multidimensional data exploration and analysis, Workshop International Society for Photogrammetry and Remote Sensing (ISPRS), Quebec, Canada, October 2-3.

ROBINSON, V.B. AND D. THONGS 1986, Fuzzy Set Theory Applied to the Mixed Pixel problem of Multi-spectral Land Cover Databases. GIS in Government, 2, 1986. Edited by Opitz, B.K. Washington D.C: A.Deepak Publication.

RODRIGUEZ, A. 2005, Inconsistency Issues in Spatial Databases. Lecture Notes In Computer Science 3300: 237-269.

ROSS, K.A, D. SRIVASTAVA, P.J. STUCKEY AND S. SUDARSHAN 1998, Foundations of Aggregation Constraints. Theor. Comput. Sci. 193(1-2): 149-179.

ROY, A.J. AND J.G. STELL 2001, Spatial relations between indeterminate regions. International Journal of Approximate Reasoning 27 (2001) 205—234.

SALEHI, M. 2005, Categorization and Formal Specification of the Integrity Constraints in Spatial Multidimensional Database. Proposé de projet de thèse de doctorat, Université Laval.

SANTOS, M.Y. E A. MOREIRA 2007, Topological Spatial Relations between a Spatially Extended Point and a Line for Predicting Movement in Space, 10th. AGILE International Conference on Geographic Information Science, Aalborg, Denmark, 8-11 May, 2007. [ISBN 978-87-918-3004-4].

SCHNEIDER, M. 1999, Uncertainty Management for Spatial Data in Database: Fuzzy Spatial Data Types. The 6th Int. Symp. on Advances in Spatial Databases(SSD), LNCS 1651, Springer Verlag, 330-351.

SCHNEIDER, M. 2001, A design of topological predicates for complex crisp and fuzzy regions. In ER '01: Proceedings of the 20th International Conference on Conceptual Modeling, pages 103–116. Springer-Verlag, ISBN 3-540-42866-6.

SERVIGNE, S., U. THIERRY, A. PURICELLI AND R. LAURINI 2000, A Methodology for Spatial Consistency Improvement of Geographic Databases. GeoInformatica 4(1): 7-34.

SHEPHERD, I. 1992, Geographic Information Systems, chapitre 22 : Information Integration and GIS, Maguire, Goodchild, Rhind (Eds.), Longman Scientific & Technical, pages 337-358.

SHOKRI, T., M.J. DELAVAR, M.R. MALEK AND A.U. FRANK 2006, Modelling uncertainty in spatio-temporal objects. In. Proc. of 7th international Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. 5 - 7 July, Lisbon, Portugal, pp : 469-478.

SHU, H., S. SPACCAPIETRA, C. PARENT AND S. QUESADA SEDAS 2003, Uncertainty of Geographic Information and its Support in MADS. Proc. 2nd International Symposium on Spatial Data Quality, Hong Kong.

SMETS, P. 1996, Imperfect information: Imprecision and uncertainty. *Uncertainty Management in Information Systems*, pages 225–254.

SMITH, B. 1994, Fiat Objects. In Proceedings of Workshop on Parts and Wholes: Conceptual Part-Whole Relations and Formal Mereology, 11th European Conference on Artificial Intelligence : 15-23.

SMITH, B. AND A.C. VARZI 2000, Fiat and Bona Fide Boundaries. Philosophy and Phenomenological Research, 60(2): 401-420.

SMITH, B. 2001, Fiat objects, in Topoi 20:2, 131-148.

SMITHSON M., 1989, Ignorance and Uncertainty: Emerging Paradigms, New York, Springer Verlag.

STELL, J.G. 2000, Boolean connection algebras: A new approach to the region-connection calculus. Artificial Intelligence, 122(1–2):111–136, September.

TANG, XI., 2004, Spatial object modeling in fuzzy topological spaces: with applications to land cover change. PhD thesis, University of Twente, ISBN 90-6164-220-5.

VAN OORT, P. 2006, Spatial data quality: from description to application, Publication on Geodesy 60, Netherlands Geodetic Commission, ISBN 90 6132 295 2, Delft, December.

VARZI, A. 2004, Mereology. In Edward N. Zalta, editor, The Stanford Encyclopedia of Philosophy.

VERSTRAETE, J., HALLEZ, A., AND DE TRÉ, G., 2007, Fuzzy Regions: Theory and Applications. In A. Morris, S. Kokhan (Eds.). *Geographic Uncertainty in Environmental Security*, pp. 1-17. Dordrecht, The Netherlands: Springer.

WAREMER, J. B. AND A. G. KLEPPE, 1998, The Object Constraint Language: precise modeling with UML. Addison-Wesley Professional, 1st edition.

WIKIPÉDIA 2006, http://en.wikipedia.org/wiki/Probability_theory, dernière visite (11-10-2008).

WORBOYS, M.F 1998a, Computation with imprecise geospatial data. Computers, Environment and Urban Systems, 22:85-106.

WORBOYS, M.F., 1998b, Imprecision in finite resolution spatial data. GeoInformatica, 2:257-279.

WORBOYS., M.F. AND M. DUCKHAM 2004, Spatial reasoning and uncertainty, In *GIS: A Computing Perspective*, Taylor & Francis.

YAO Y.Y. 1998, A comparative study of fuzzy sets and rough sets. Information Sciences, Vol. 109, No. 1-4, pp. 227-242.

YAZICI A, ZHU Q. AND SUN N. 2001, Semantic data modeling of spatiotemporal database applications. Int. J. Intell. Syst. 2001, p: 881-904.

ZADEH, L.A. 1965, Fuzzy sets. Inform. Control, vol. 8, pp. 338–353.

ZENG, T., J. HUDSON, S. KAY AND E. LOGINESTRA 2003, A fuzzy GIS approach to fire risk assessment: a case study of Sydney Olympic Park, Australia, Spatial Science Conference 2003. 22-27 September 2003, Canberra, Australian Capital Territory

ZHAN, B.F. 1997, Topological relations between fuzzy regions. In Proceedings of the 1997 ACM Symposium on Applied Computing, pages 192–196. ACM Press, 1997. ISBN 0-89791-850-9.

ZHAN, F. B. AND LIN., H., 2003, Overlay of Two Simple Polygons with Indeterminate Boundaries. *Transactions in GIS*, 7(1), pp. 67-81.

Chapter 3: Qualified topological relationships between objects with possibly vague shapes

Lotfi BEJAOUI, François PINET, Yvan BÉDARD, Michel SCHNEIDER

International Journal of Geographical Information Sciences

(In press)

3.1 Résumé de l'article

La notion de *frontière large* est généralement utilisée pour remplacer les frontières linéaires pour des objets ayant des formes vagues. Une frontière large est un invariant topologique qui doit respecter les conditions de fermeture et de connexité. En effet, les régions ayant des frontières partiellement larges sont considérées comme inconsistantes dans les modèles existants (e.g. un lac avec des berges rocheuses et d'autres marécageuses). L'objectif de ce travail est de représenter différent niveaux de vague de forme et de les considérer lors de l'identification des relations topologiques. Ainsi, un objet ayant une forme vague est défini comme étant une composition de deux extensions spatiales: une extension minimale et une autre maximale. Ensuite, les relations topologiques sont identifiées en appliquant le modèle de 9-Intersection pour les sous-relations entre les extensions minimales et maximales des objets impliqués. Quatre sous-relations sont ainsi représentées dans une matrice 4*4 que nous utilisons également pour établir une classification des relations topologiques. Pour les régions ayant des frontières larges, 242 relations sont distinguées et classées dans 40 groupes. Cette

approche permet une expression adverbiale des contraintes d'intégrité et des requêtes spatiales.

3.2 Abstract

A broad boundary is generally used to replace one-dimensional boundary for spatial objects with vague shapes. For regions with broad boundaries, this concept should respect both connectedness and closeness conditions. Therefore, some real configurations, like regions with *partially broad boundaries* (e.g., *lake with rocky and swamp banks*), are considered invalid. This paper aims to represent different levels of shape vagueness and consider them during the identification of topological relationships. Then, an object with a vague shape is composed by two crisp components: a *minimal extent* and a *maximal extent*. Topological relationships are identified by applying the 9-Intersection model for the sub-relations between the minimal and maximal extents of objects involved. Four sub-relations are then represented through a 4-Intersection matrix used to classify the topological relationships. For regions with broad boundaries, 242 relationships are distinguished and classified into 40 clusters. This approach supports an adverbial expression of integrity constraints and spatial queries.

3.3 Introduction

To satisfy the requirements of several categories of users, Geographic Information Systems (GIS) and spatial databases provide tools to store, retrieve, analyze, and display spatial data. Ensuring their usability requires controlling the spatial data quality, which can be degraded by several types of imperfections. Several approaches (Smithson 1989, Fisher 1999(b), Mowrer 1999, Duckham *et al.* 2001) proposed different categorizations of data imperfections that are generally caused by the complexity of reality and limitations of the instruments and processes used in the measurements (Bédard 1987). Moreover, inappropriate spatial data representations can also be another source of data quality degradation (Dilo 2006). Spatial reality is generally forced to be represented by *crisp spatial object types* (i.e., *points*, *lines*, and *regions*), whereas the shapes of many spatial objects are inherently vague (e.g., *forest stand*, *pollution zone*, *valley*, or *lake*). Shape vagueness occurs when it is difficult to distinguish the boundary (*e.g., regions with broad boundaries*) and/or the interior (*e.g., broad points or lines with broad interior*) of an object's geometry from other spatial objects of the neighborhood. Using crisp spatial object types to represent *spatial objects with vague shapes* entails a clear gap between the spatial reality and its formal representation in databases and GIS (Cheng *et al.* 2001, Yazici *et al.* 2001).

Pertinent solutions were found to overcome the "*classical*" sources of spatial data quality degradation (Bédard 1987, Goodchild 1995, Guptill and Morrison 1995, Ubeda and Egenhofer 1997, Frank 2001, Van Oort 2006, Devillers *et al.* 2007, Pinet *et al.* 2007). Several approaches (Burrough and Frank 1996, Cohn and Gotts 1996, Clementini and Di Felice 1997, Erwig and Schneider 1997, Schneider 2001, Tang 2004, Pfoser *et al.* 2005, Dilo 2006) have studied specificities of *objects with vague shapes* to determine their appropriate representations. A review of the literature in this domain (cf. Section 2.3) stresses that current GIS and spatial database systems do not offer the specific structure to formally represent this type of objects (as pointed by Clementini and Di Felice 1997 ten years ago). With regard to this problem, researchers are increasingly more motivated to model *shape vagueness* in order to: (1) reduce the gap between the geographic reality and the spatial models (Cohn and Gotts 1996), (2) provide formal modeling tools to represent shape vagueness (Yazici *et al.* 2001), and (3) specify spatial queries involving *spatial objects with vague shapes* (Erwig and Schneider 1997). In the same way, the spatial data integration requires the extraction of heterogeneous representations of the same objects from different data sources. The main difficulty lies in choosing one of them when no information exists about their quality (Rodriguez 2005). By using a spatial model that supports *shape vagueness*, it becomes possible to merge different representations in such a way that the integration result looks like an object with a vague shape. For example, figure 3.1 shows a spatial object that has a representation *A* in a first source and a representation *B* in a second one. The integration result can correspond to one geometry with a vague shape made up of *A* and *B* (figure 3.1). The intersection of *A* and *B* corresponds to the *certain part* (i.e., the part that exists in both representation *A* and representation *B*) or the *minimal extent* of the spatial object. However, the union is the *maximal extent* that the object can fill; it groups the certain and the uncertain parts (i.e., a geometry part is uncertain when it does not exist in all candidate representations for the integration) of the geometry. Indeed, there are strong and different motivations to present pertinent solutions in order to adequately model the shape vagueness.

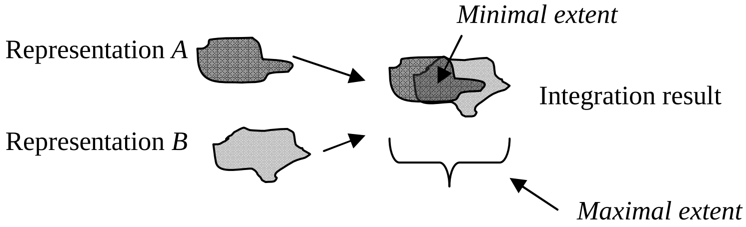


Figure 3.1 Integration of different spatial representations of a same object (e.g., lake)

To model objects with vague shapes, researchers were firstly inspired by the modeling of crisp spatial objects. In point-set topology (Egenhofer and Herring 1990), crisp spatial objects are typically decomposed into three mutually disjoint topological invariants: an *interior*, a *boundary*, and an *exterior*. Several approaches (Clementini and Di Felice 1997, Tang 2004, Reis *et al.* 2006) extend the crisp models by identifying other topological invariants for the objects with vague shapes. For example, Clementini and Di Felice (1997) distinguish three topological invariants for regions with broad boundaries: an *interior*, a *broad boundary* (i.e., *a two–dimensional boundary*), and an *exterior*. In this approach, the shape vagueness is correlated to the *broad boundary,* which should respect the closeness and the connectedness conditions (Clementini and Di Felice 1997, Tang 2004). Thus, any representation that does not verify these conditions is considered invalid. Nonetheless, the shape vagueness can also characterize only some parts of an object's geometry. For example, figure 3.2 shows a *lake* surrounded by crisp rocky banks on one side and swamp ones on the other side at the same time (figure 3.2). We denote this kind of feature as *objects with partially vague shapes* that cannot be represented by existing models. Then, the main questions are: How is it possible to define an exact model where different levels of shape vagueness could be considered? How can we retain this expressivity during the specification of topological relations between such objects?

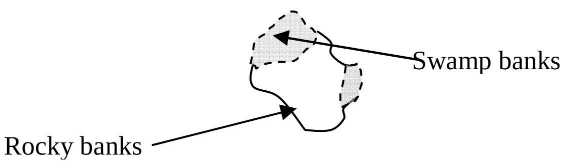


Figure 3.2 A lake with a partially broad boundary

The first objective of this paper is to allow the representation of three levels of shape vagueness: *crispness*, *partial shape vagueness*, and *complete shape vagueness*. Modeling objects with vague shapes requires a framework for identifying topological relations. The second objective is to consider the different levels of vagueness in the identification of topological relations between objects with vague shapes. In several studies (Clementini and Di Felice 1997, Tang 2004, Reis *et al.* 2006), topological relations can be identified by enumerating the intersections between the topological invariants of the objects with vague shapes involved. For each model, the number of relations depends upon the number of topological invariants. In this work, we look for an expressive model in which it is possible to specify the vagueness level of the topological relation instances. We think that it would be pertinent for the user to know whether objects are *weakly* or *strongly* disjoint. Accordingly, the third specific objective of this work is to classify the topological relations according to their vagueness level. We should denote that this model is called Qualitative Min-Max model (*QMM* model) in Chapter 4. This label has been proposed after the acceptance of this paper in order to facilitate using and reference to our approach. In this Chapter, we do not use this label in order to preserve as well as possible the original version of the paper.

The remainder of the paper is organized as follows. In sections 3.4, we present previous works on the modeling of objects with vague shapes and their topological relationships. Section 3.5 addresses the problem of this paper. In section 3.6, we present three basic types of spatial objects with vague shapes: *regions with broad boundaries*, *lines with vague shapes* and *broad points*. Then, section 3.7 gives a proposition based on the 9-Intersection model (Egenhofer and Herring 1990) in order to identify the topological relations among objects with vague shapes. The model is applied to regions with broad boundaries, and their topological relations are studied in detail in the appendix 1. As a result of this approach, 242 relations can be distinguished through a 4-Intersection matrix. Section 3.8 proposes a hierarchical clustering of topological relations between regions with broad boundaries, and section 3.9 explains how to use our approach to express spatial queries and integrity constraints. In section 3.10, our model is compared with existing exact approaches (Cohn and Gotts 1996, Clementini and Di Felice 1997, Tang 2004). Finally, section 3.11 presents our conclusions and discusses future research.

3.4 Previous works

3.4.1 Spatial vagueness

According to (Erwig and Schneider 1997, Hazarika and Cohn 2001, Pfoser *et al.* 2005), spatial vagueness can characterize the position and/or shape of the spatial extent of a given object. From this perspective, the *shape vagueness* refers to the difficulty of distinguishing an object shape from its neighborhood. Shape vagueness is an intrinsic property of an object that certainly has an extent in a known position but cannot or does not have a well-defined shape (Erwig and Schneider 1997). For example, a region has a vague shape when it is surrounded by a broad boundary instead a sharp one. One could normally use the term « fuzziness » to speak about «shape vagueness» since it would correspond to the unclearness of an object shape as it is defined in a general ontology (i.e., to the definitions found in the Oxford and the Cambridge dictionaries). Nevertheless, in order to avoid confusion with the mathematical definition found in the specialized ontology of Fuzzy Set Theory (Zadeh 1965) which is used in several GIS-related papers (e.g., Altman 1987, Burrough 1989, Brown 1998, Schneider 2001), we have decided to use the expression “shape vagueness”. Accordingly, one must not confuse “fuzziness” as defined in Fuzzy Set Theory with the concept of “shape vagueness” as defined in the present paper.

Spatial vagueness can also characterize well-defined (or crisp) objects when there is uncertainty about objects' positions despite their sharp shapes; we refer to this scenario as *positional vagueness*. Positional vagueness is a measurement imperfection related to the accuracy and precision of the instruments and processes used in the measurements (Mowrer 1999). Figure 3.3 shows this categorization of spatial vagueness into "*shape vagueness*" and "*positional vagueness*". In this paper, we only deal with the formal representation of spatial objects with vague shapes and the topological relations between them.

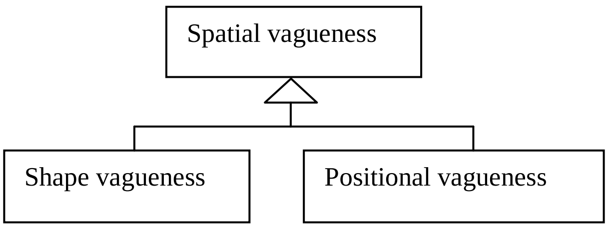


Figure 3.3 Categorization of spatial vagueness

In general, we distinguish between at least two categories of models used to represent spatial vagueness. In the first category, crisp spatial concepts are transferred and extended to formally express spatial vagueness; we speak about *exact models* (Cohn and Gotts 1996, Clementini and Di Felice 1997, Erwig and Schneider 1997) as explained in the next section. In the second category, three principal mathematical theories are generally used: (1) models based on the Fuzzy Logic (Zadeh 1965) (e.g., Altman 1987, Burrough 1989, Brown 1998, Schneider 2001, Tang 2004, Hwang and Thill 2005, Dilo 2006), which can be used to represent continuous phenomena such as temperature, (2) models based on rough sets (e.g., Ahlqvist *et al.* 1998, Worboys 1998), which represent the objects with vague shapes as a pair of approximations (*upper and lower approximations*), and (3) models based on probability theory (e.g., Burrough and Frank 1996, Pfoser *et al.* 2005), which is principally used to model errors of positions and attributes.

3.4.2 Formal definitions of objects with vague shapes

In the original version of paper, this section reviews previous works that formally define objects with vague shapes. In the present manuscript, this review literature has been transferred in Chapter 2 (cf. Section 2.3) in order to reduce redundancies and improve the readability of the thesis.

3.5 Problem statement

The exact models presented earlier (Cohn and Gotts 1996, Clementini and Di Felice 1997, Erwig and Schneider 1997, Tang 2004, Reis *et al.* 2006) have the advantage of explicitly distinguishing the topological invariants of objects involved. Through this discrete viewpoint of space, the specification of topological relations can be improved (Clementini and Di Felice 1997). For these reasons, we propose an exact model in order to achieve objectives. Nevertheless, we think that the existing models do not distinguish between different levels of shape vagueness and are not sufficiently expressive to represent partial shape vagueness. In reality, a region with a broad boundary is not always surrounded by a large boundary everywhere. For example, the boundary of a given lake can be broad in some locations and sharp in some others. This situation cannot be represented by existing exact models, because the connectedness condition is violated. The same problem is present for lines. Only two cases of shape vagueness are distinguished for lines (cf. section 2.3). Nonetheless, a line can have a partially broad interior independently of the boundary. Moreover, the studied models are not sufficiently expressive in terms of topological relations since there is no distinction between the inner and outer boundary for regions with broad boundaries. Some works try to offer more expressivity by increasing the number of topological invariants (e.g., Tang 2004). Nevertheless, the absence of relation clustering limits their practical use. Indeed, the main research questions of our paper are the following:

1- How can we obtain more expressive definitions of the objects with vague shapes through an exact model? How can we represent shape vagueness?

2- What are the topological relations between objects with vague shapes? How is it possible to identify topological relations between objects that have different levels of shape vagueness? How can we formally identify these relations?
3- How can we classify the topological relations between regions with broad boundaries in order to facilitate their use in practice? How could resulting clusters reflect the vagueness level of a topological relation?

3.6 Spatial objects with vague shapes

In general, there is no agreement regarding the appropriate formal definition of spatial objects with vague shapes, because shape vagueness can be interpreted in different ways. It is not the objective of this work to unify these interpretations. We are interested in proposing an expressive and easy definition of spatial objects with vague shapes through an exact model. In our approach, we transfer the Egg-Yolk model into point-set topology context in order to both consider points and lines and permit the representation of *objects with partially vague shapes*. In the literature, many expressions have been used to speak about shape vagueness of spatial objects. For example, Burrough and Frank (1996) used the terms "*objects with indeterminate boundaries*", Dilo (Dilo 2006) used the terms "*vague spatial objects*" and Clementini and Di Felice (1997) used "*objects with large boundaries".* We find these different expressions pertinent but they are not sufficiently expressive to cover the *shape vagueness* for a line's interior or a point (i.e. a point does not have a boundary; it is composed by an interior). In other words, we make distinction between "*broad interior*" and "*broad boundary*" that we consider as specializations of "*shape vagueness*". This distinction is useful especially in the cases of lines and points. From this perspective, we distinguish three basic types of *spatial objects with vague shapes*: *broad points, lines with vague shapes (i.e., lines with broad boundaries, lines with broad interiors or broad lines),* and *regions with broad boundaries*. Figure 3.4 shows our categorization of objects with vague shapes.

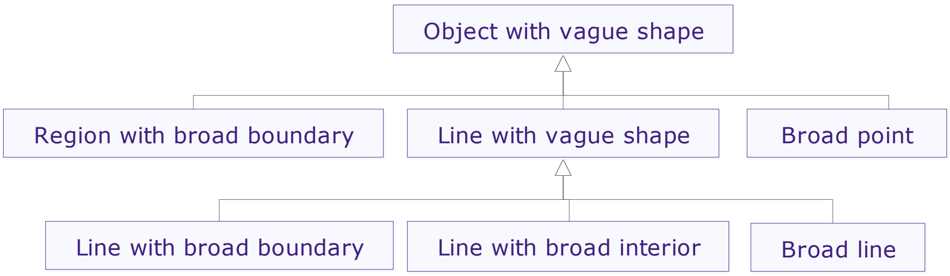


Figure 3.4 Categorization of objects with vague shapes

Each object with a vague shape is composed of *n* crisp object types (i.e., *point*, *line*, and *region*) distributed into a pair of sets called (1) the *minimal extent* and (2) the *maximal extent* (figure 3.5). Figure 3.5 presents an example of broad points, lines with vague shapes, and regions with broad boundaries. A broad point is a zone that we approximate to a crisp region containing all of elementary space portions that the point can possibly fill. The minimal extent of point is equal to its maximal extent because the shape vagueness concerns a unique topological invariant: *the interior* (cf. section 3.6.1 for more details). For a line with a vague shape (cf. section 3.6.2), the minimal extent is the union of the linear parts. However, its maximal extent can contain some broad parts (i.e., presented as broad points in figure 3.5(b)), at which the line can have any shape. For a region with a broad boundary (cf. section 3.6.3), the shape vagueness concerns the boundary. The minimal extent refers to the geometry when the boundary is as close as possible (i.e. it is drawn around the area which *certainly* belongs to the region). The maximal extent is the geometry of the object when the boundary is as far away as possible (i.e. it is drawn around the area, which contains all of points *possibly* belonging to the region).

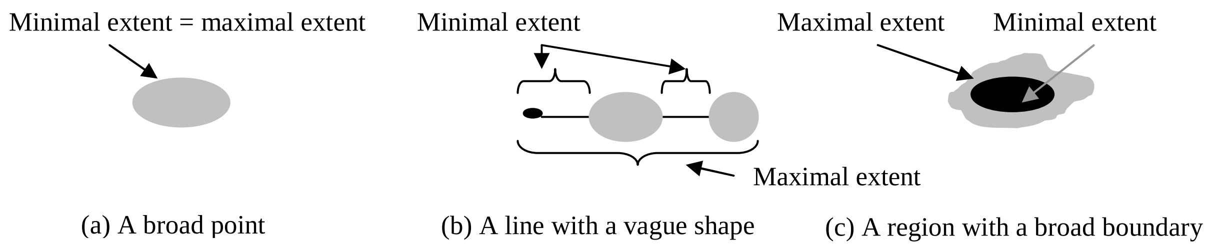


Figure 3.5 Minimal and maximal extents for (a) a broad point, (b) a line with a vague shape and (c) a region with a broad boundary

Generally, the minimal extent refers to the geometry's parts definitely belonging to the spatial object. The maximal extent corresponds to the object's geometry when shape vagueness is taken into account and added to the minimal extent. Outside of the maximal extent, there are no spatial points that can possibly belong to the object. The number *n* of crisp object types composing an object with a vague shape is 1 for *a broad point* (i.e., a zone that we represent as a crisp region composed of the quasi-totality of possible elementary space portions that the point can fill (cf. section 3.6.1)), 2 for a *region with a broad boundary* (i.e., two crisp regions (cf. section 3.6.3)), and *n* for *lines with vague shapes* (i.e., 1 or *n* points of the line are broad (cf. section 3.6.3)). For example, a region with a broad boundary corresponds to a pair of crisp regions that respectively represent the minimal and maximal extents. This general definition of spatial objects with vague shapes is based on the following principles:

1- A spatial object with a vague shape is a generalization of a crisp spatial object.
2- The minimal and the maximal extents are made up of crisp spatial object types. Only the combination of two extents corresponds to the object with a vague shape.
3- For the minimal and the maximal extents, the topological invariants should be mutually disjoint.

The first principle means that the spatial extent of an object with a vague shape is crisp when its minimal extent is equal to its maximal one. The second principle requires that the minimal and maximal extents verify the topological consistency conditions of the crisp spatial object types (e.g., *a simple crisp region should be connected*). Finally, the third principle permits the identification of topological relations based on the intersections between the topological invariants of the minimal and maximal extents of spatial objects with vague shapes involved. In the next sections, we present our definitions of broad points, lines with vague shapes, and regions with broad boundaries.

3.6.1 Broad point

In the crisp context, a point *p* is a 0-dimensional object type which corresponds to an elementary portion of the space. This portion refers to the interior of the point (i.e. the only topological invariant of the point). Because a point does not have a boundary (the dimension of the boundary of an object with a dimension *n* is *n-1*), the shape vagueness can characterize only the interior and thus the point itself. Semantically, a broad point occurs when an intrinsic property of the point or a lack of knowledge does not permit to sharply distinguish the point from its neighborhood. For such a case, the spatial extent of the object is typically replaced by a zone that we represent as a crisp region composed of the family of elementary space portions that the point can fill (see an example of broad point in figure 3.6). The closure[5] of this crisp region represents an infinity of possible elementary space portions for the point. Consequently, a broad point does not have a minimal extent; it only has a maximal extent.

Simple Crisp point	Broad point

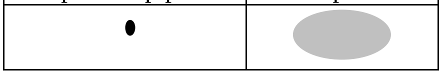


Figure 3.6 Broad point

Since a simple broad point corresponds, in fact, to a simple crisp region, it should verify the following conditions:

1- The closure is a non-empty connected regular closed set.

2- The interior is a non-empty connected regular open set.

3- The boundary and exterior are connected.

To provide an example of a broad point, consider an application to help the fire brigades in their interventions. Generally, a fire fighter cannot precisely localize the fire source. However, he can draw an area in which the fire source should exist. This intervention area corresponds to a broad point and can be represented through our model. It is clear that the size of the region representing the broad point depends on the shape vagueness level (i.e., *a larger region refers to a fuzzier point*).

3.6.2 Line with a vague shape

Shape vagueness for lines has been studied in-depth in another paper that presents Chapter 4 of this thesis. In order to reduce the redundancies, we summarize the original content of this section.

A crisp line is composed by an interior connected by two endpoints that refer to its boundary. We consider that shape vagueness can characterize any point of the line. Indeed, the line boundary can be partially or completely broad while the interior remains well-defined; we then speak about lines with broad boundary. For example, the trajectory of an aircraft (for which the pilot attempted a crash-landing) can be represented as a line with a partially broad boundary (only the final endpoint is ill-known). In the same way, the interior can be partially or completely broad while the endpoints are well-defined; we then speak

[5] The closure, in point set topology, is the union of the interior and the boundary.

about lines with partially and completely broad interior, respectively. The extreme case of line shape vagueness arises when all topological invariants of the line (i.e. the interior and the boundary) are broad. For example, the trajectory of an historical person can be represented as a completely broad line whether few information are available about it. Thus, a completely broad line arises when there is a difficulty to sharply distinguish each point one the line from its neighborhood. However, a completely crisp line is a particular case of lines with vague shapes, for which all of the interior and boundary are well-defined. In Chapter 4, lines with vague shapes are specifically studied. All of these aspects are presented more in detail.

3.6.3 Region with a broad boundary

A crisp region is a two-dimensional spatial object type in which the shape is typically composed of an *interior*, a *boundary,* and an *exterior.* For a region, shape vagueness occurs when there is difficulty in precisely distinguishing between the interior and exterior through a sharp boundary. From this perspective, shape vagueness is generally correlated with the boundary, which can itself be *sharp*, *partially broad*, or *completely broad*. It is possible to draw a *minimal spatial extent* by considering the boundary to be as close as possible (i.e., it is drawn around the area which *certainly* belongs to the region). In the same way, a *maximal spatial extent* can be drawn by considering the boundary to be as far as possible (i.e., it is drawn around the area which contains all of points *possibly* belonging to the region). Figure 3.7 represents an example of a *region with a partially broad boundary*. The spatial extent of a region with a broad boundary is composed of a portion called the *minimal extent* (i.e., *all of the points definitely belonging to the spatial object*) and covered by a *maximal extent* (i.e., *all of the points possibly belonging to the spatial object*).

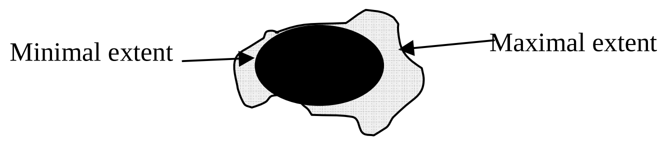


Figure 3.7 Region with partially broad boundary

We consider that a simple region with a broad boundary is made up of two crisp regions: (1) the *maximal extent*, which can *be* "*Equal*", "*Contains*", or "*Covers*" (2) the *minimal extent* (see examples in figure 3.7). When the boundary is completely sharp (i.e. it does not contain any broad point), the region is completely crisp. This is a particular case of *regions with broad boundaries* for which the *maximal extent* is *equal* to the *minimal extent*; we speak about regions with none broad boundary (or crisp regions). In the second case, the region boundary is broad only in some locations. We speak about *regions with partially broad boundaries,* where the maximal extent *covers* the minimal extent. For example, a forest stand or a lake can have sharp boundaries (e.g., *rocky borders for a lake* and *a total cut for a forest stand*) and broad boundaries (e.g., *swamp borders for a lake*) at the same time. The third case represents a typical region with a broad boundary for which the boundary is completely broad. For example, the boundary of a pollution zone is broad everywhere since the pollution decreases from its kernel to the region exterior. In figure 3.8, we present an example of each of these three cases.

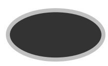
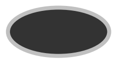
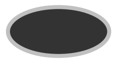
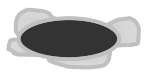
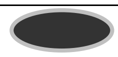
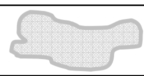
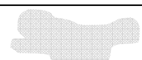
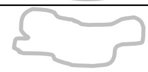
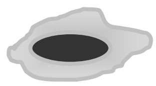
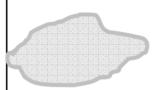
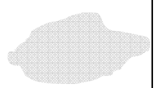
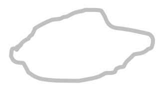
Region with a broad boundary	Representation	Maximal and minimal extents		Topological invariants of minimal and maximal extents: Interior	Topological invariants of minimal and maximal extents: Boundary
Region with none broad boundary (i.e., *crisp region*)		Minimal extent			
		Maximal extent			
Region with partially broad boundary (i.e., *region with partially vague shape*)		Minimal extent			
		Maximal extent			
Region with completely broad boundary (i.e., *region with completely vague shape*)		Minimal extent			
		Maximal extent			

Figure 3.8 Regions with broad boundaries

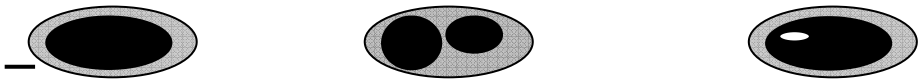
Since the minimal and maximal extents are crisp regions, we distinguish three mutually disjoint topological invariants for each of them: an *interior*, a *boundary*, and an *exterior*. Thus, a region with a broad boundary $\tilde{A}$ is made up of six topological invariants: *the interior of the minimal extent* $\tilde{A}^{\circ}_{\min}$, *the boundary of the minimal extent* $\partial\tilde{A}_{\min}$, *the exterior of the minimal extent* $\tilde{A}^{-}_{\min}$, *the interior of the maximal extent* $\tilde{A}^{\circ}_{\max}$, *the boundary of the maximal extent* $\partial\tilde{A}_{\max}$, and *the exterior of the maximal extent* $\tilde{A}^{-}_{\max}$ (figure 3.8).

- ***Definition 1***: A simple region with a broad boundary $\tilde{A}$ is composed of two simple crisp regions $\tilde{A}_{max}$ and $\tilde{A}_{min}$, *where Equal*[6]*(* $\tilde{A}_{max}$, $\tilde{A}_{min}$ *), Contains(* $\tilde{A}_{max}$, $\tilde{A}_{min}$ *), or Covers(* $\tilde{A}_{max}$, $\tilde{A}_{min}$ *).* $\tilde{A}_{min}$ is the *minimal extent* of $\tilde{A}$, $\partial\tilde{A}_{min}$ is the inner boundary of $\tilde{A}$, $\tilde{A}_{max}$ is the *maximal extent of* $\tilde{A}$*, and* $\partial\tilde{A}_{max}$ is the outer boundary of $\tilde{A}$. $\tilde{A}_{min}$ is the set of points *certainly* belonging to $\tilde{A}$. However, the maximal extent $\tilde{A}_{max}$ is the union of the minimal extent and the set of points *possibly* belonging to the region with a broad boundary.

The following conditions should be respected for any type of regions with broad boundaries:

1- The closures of the maximal and the minimal extents are non-empty regular connected closed subsets.
2- The interiors of the maximal and minimal extents are non-empty regular open sets.
3- The boundaries and exteriors of the maximal and minimal extents are connected.

In this paper, we limit our investigations to simple regions with broad boundaries (i.e., we do not consider vague regions with complex vague shapes such as regions with broad boundaries and holes or regions with broad boundaries and several cores). We adopt this strategy in order to clearly present the bases of our model before improving it**.** Figure 3.9 presents some examples of invalid regions with broad boundaries. In case (a), the region is invalid because its closure is non-regular, i.e. there is an isolated line that belongs to the closure. In case (b), the interior of the region is non-connected because it is composed of two disjoint minimal extents (or cores). Then, this shape cannot be considered as a simple region with a broad boundary and therefore it is invalid according to our model. In the case (c), the exterior does not respect the connectedness condition of the exterior (see condition 3 presented above) since the interior contains a hole. This type of regions is considered as invalid because we only deal with simple regions with broad boundaries and without holes.



(a) Non-regular closed closure (b) Non-connected interior (c) Non-connected exterior

Figure 3.9 Examples of invalid regions with broad boundaries

[6] The spatial relations (i.e., *Equal*, *Contains*, *Covers*) used in this definition are those defined in (Egenhofer and Herring 1990).

This general definition covers the crisp regions occurring when *Equal* $(\tilde{A}_{\max}, \tilde{A}_{\min})$. Accordingly, this property can be used to represent a region with only one extent and without a full membership to the object (i.e., a region without any core; shape vagueness is about all of the region and not only about its boundary). Our model is capable to represent this type of regions but we do not study them in detail in the present paper. Hereafter, we only focus on the typical regions with broad boundaries where *Contains*$(\tilde{A}_{\max}, \tilde{A}_{\min})$ or *Covers*$(\tilde{A}_{\max}, \tilde{A}_{\min})$ and their topological relations.

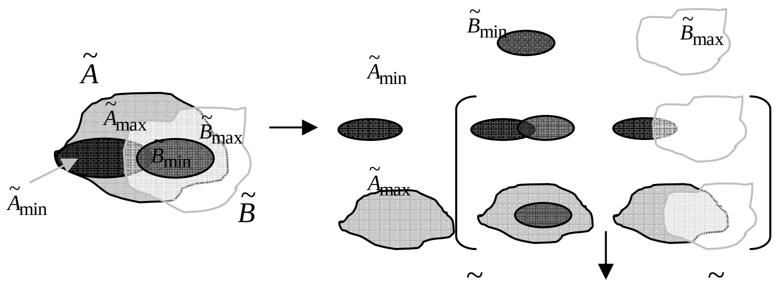
3.7 Topological relations between spatial objects with vague shapes

3.7.1 Principles

To identify the topological relations between two objects with vague shapes, we interpret their maximal and minimal extents as independent crisp geometries. In fact, our methodology consists of identifying four specific topological relations between the minimal and maximal extents of the objects with vague shapes involved. For that, we define a 4-Intersection matrix containing the following four topological sub-relations: $R_1(\tilde{A}_{\min}, \tilde{B}_{\min})$, $R_2(\tilde{A}_{\min}, \tilde{B}_{\max})$, $R_3(\tilde{A}_{\max}, \tilde{B}_{\min})$, and $R_4(\tilde{A}_{\max}, \tilde{B}_{\max})$(see example in figure 3.10). These topological sub-relations assigned to the matrix's cells are those defined in the 9-Intersection model (Egenhofer and Herring 1990). For example, if we study the topological relations between two regions with broad boundaries, each cell receives one of the eight possible topological relations between two simple crisp regions (i.e., *Disjoint*, *Overlap*, *Meet*, *Equal*, *Contains*, *Inside*, *Covers*, *Covered by*). Then, the 4-Intersection matrix corresponds to the following representation:

$$\begin{array}{cc} & \begin{array}{cc} \tilde{B}_{\min} & \tilde{B}_{\max} \end{array} \\ \begin{array}{c} \tilde{A}_{\min} \\ \tilde{A}_{\max} \end{array} & \begin{pmatrix} R_1(\tilde{A}_{\min}, \tilde{B}_{\min}), R_2(\tilde{A}_{\min}, \tilde{B}_{\max}) \\ R_3(\tilde{A}_{\max}, \tilde{B}_{\min}), R_4(\tilde{A}_{\max}, \tilde{B}_{\max}) \end{pmatrix} \end{array}$$

Figure 3.10 shows the content of the matrix that describes the topological relation between a *region with a partially broad boundary* $\tilde{A}$ and a *region with a completely broad* *boundary* $\tilde{B}$. In the matrix (b), the letters *O* and *C* are used to denote the relations *Overlap* and *Contains*, respectively.



(a)

$$\begin{array}{c|c}
 & \begin{array}{cc} \tilde{B}_{min} & \tilde{B}_{max} \end{array} \\
\begin{array}{c} \tilde{A}_{min} \\ \tilde{A}_{max} \end{array} & \begin{pmatrix} O(\tilde{A}_{min}, \tilde{B}_{min}), O(\tilde{A}_{min}, \tilde{B}_{max}) \\ C(\tilde{A}_{max}, \tilde{B}_{min}), O(\tilde{A}_{max}, \tilde{B}_{max}) \end{pmatrix}
\end{array}$$

(b)

Figure 3.10 Description of the topological relation between two regions with broad boundaries: (a) visual content of the matrix, (b) formal identification of the relations between the minimal and maximal extents of the objects involved

The content of a given matrix corresponds to the topological sub-relations relating the minimal and maximal extents. We use the topological sub-relation between the maximal extents $R_4(\tilde{A}_{max}, \tilde{B}_{max})$ to label the global topological relation. For example, if $R_4(\tilde{A}_{max}, \tilde{B}_{max})$ is *Overlap*, we consider that spatial objects with vague shapes globally *Overlap* each other. If $R_4(\tilde{A}_{max}, \tilde{B}_{max})$ is *Contains*, we consider that the global topological relation is *Contains*.

In figure 3.11, we present examples of an identification of topological relations between spatial objects with vague shapes. The first example presents a description of the topological relation between two regions with completely broad boundaries $\tilde{A}$ and $\tilde{B}$. The second example concerns a line with a fairly vague shape $\tilde{L}$ and a region with a completely broad boundary $\tilde{A}$. The third example shows the identification of the topological relation between two lines with fairly vague shapes $\tilde{L}$ and $\tilde{K}$. Finally, the last example concerns the relation between a region with a completely broad boundary $\tilde{A}$ and a broad point $\tilde{P}$.

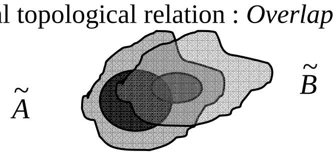
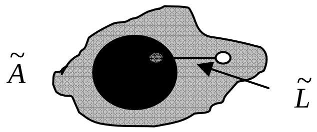
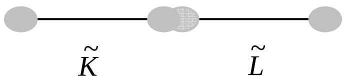
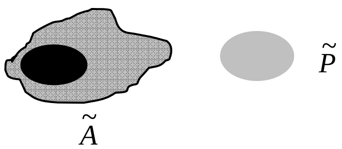
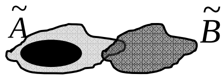
Spatial representation	Correspondent matrix
Global topological relation : *Overlap* 	$\begin{array}{c c} & \begin{array}{cc} \tilde{B}_{\min} & \tilde{B}_{\max} \end{array} \\ \begin{array}{c} \tilde{A}_{\min} \\ \tilde{A}_{\max} \end{array} & \begin{bmatrix} \mathit{Overlap}(\tilde{A}_{\min}, \tilde{B}_{\min}), & \mathit{Overlap}(\tilde{A}_{\min}, \tilde{B}_{\max}) \\ \mathit{Contains}(\tilde{A}_{\max}, \tilde{B}_{\min}), & \mathit{Overlap}(\tilde{A}_{\max}, \tilde{B}_{\max}) \end{bmatrix} \end{array}$
Global topological relation : *Contains* 	$\begin{array}{c c} & \begin{array}{cc} \tilde{L}_{\min} & \tilde{L}_{\max} \end{array} \\ \begin{array}{c} \tilde{A}_{\min} \\ \tilde{A}_{\max} \end{array} & \begin{bmatrix} \mathit{Overlap}(\tilde{A}_{\min}, \tilde{L}_{\min}), & \mathit{Overlap}(\tilde{A}_{\min}, \tilde{L}_{\max}) \\ \mathit{Contains}(\tilde{A}_{\max}, \tilde{L}_{\min}), & \mathit{Contains}(\tilde{A}_{\max}, \tilde{L}_{\max}) \end{bmatrix} \end{array}$
Global topological relation : *Overlap* 	$\begin{array}{c c} & \begin{array}{cc} \tilde{L}_{\min} & \tilde{L}_{\max} \end{array} \\ \begin{array}{c} \tilde{K}_{\min} \\ \tilde{K}_{\max} \end{array} & \begin{bmatrix} \mathit{Disjoint}(\tilde{K}_{\min}, \tilde{L}_{\min}), & \mathit{Disjoint}(\tilde{K}_{\min}, \tilde{L}_{\max}) \\ \mathit{Disjoint}(\tilde{K}_{\max}, \tilde{L}_{\min}), & \mathit{Overlap}(\tilde{K}_{\max}, \tilde{L}_{\max}) \end{bmatrix} \end{array}$
Global topological relation : *Disjoint* 	$\begin{array}{c c} & \tilde{P}_{\max} \\ \begin{array}{c} \tilde{A}_{\min} \\ \tilde{A}_{\max} \end{array} & \begin{bmatrix} \mathit{Disjoint}(\tilde{A}_{\min}, \tilde{P}_{\max}) \\ \mathit{Disjoint}(\tilde{A}_{\max}, \tilde{P}_{\max}) \end{bmatrix} \end{array}$

Figure 3.11 Examples of identification of topological relations through a 4-Intersection matrix

3.7.2 Topological relations between a region with a broad boundary and a crisp one

In our approach, the 4-Intersection matrix highlights the sub-relations that exist between the components of the geometries with vague shape. Indeed, this expressivity is highlighted when the maximal extent of the spatial object with a vague shape is non-empty and different from the minimal extent. In the other cases, some cells in the matrix will have the same values. For example, figure 3.12 shows a region with a completely broad boundary that overlaps a crisp region. The topological relation can be reduced to a 2-Intersection matrix, because the region $\tilde{B}$ is crisp and so its minimal extent equals its maximal one. Hereafter, we do not study topological relations that involve crisp regions. We focus on regions with different non-empty maximal extent and non-empty minimal extent.



$$\begin{array}{ccc} & \tilde{B}_{\min} & \tilde{B}_{\max} \\ \tilde{A}_{\min} & \multicolumn{2}{c}{\multirow{2}{*}{}} \end{array}$$

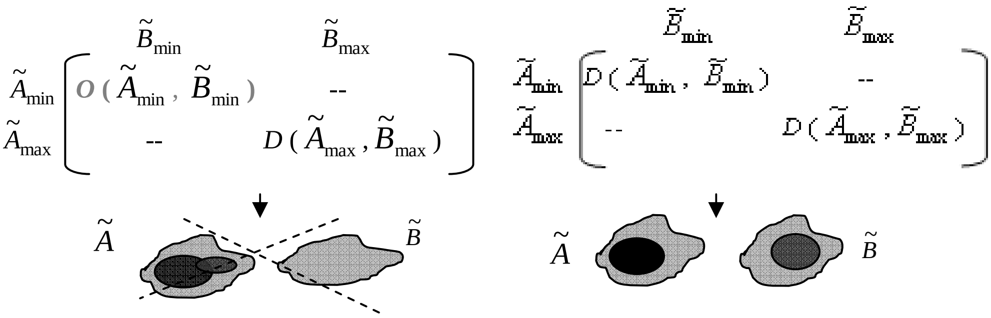
$$\begin{array}{c} \tilde{A}_{\min} \\ \tilde{A}_{\max} \end{array} \begin{bmatrix} \mathit{Overlap}\,(\tilde{A}_{\min}, \tilde{B}_{\min}), \mathit{Overlap}\,(\tilde{A}_{\min}, \tilde{B}_{\max}) \\ \mathit{Overlap}\,(\tilde{A}_{\max}, \tilde{B}_{\min}), \mathit{Overlap}\,(\tilde{A}_{\max}, \tilde{B}_{\max}) \end{bmatrix} \Longleftrightarrow \begin{array}{c} \\ \tilde{A}_{\min} \\ \tilde{A}_{\max} \end{array} \begin{array}{c} \tilde{B}_{\min} \\ \begin{bmatrix} \mathit{Overlap}\,(\tilde{A}_{\min}, \tilde{B}_{\min}) \\ \mathit{Overlap}\,(\tilde{A}_{\max}, \tilde{B}_{\min}) \end{bmatrix} \end{array}$$

Figure 3.12 Example of a topological relation between a region with a broad boundary and a crisp region

The values assigned to the different cells of the matrix should not be arbitrarily chosen. In general, the value of $R\ (\tilde{A}_{\max}, \tilde{B}_{\max})$ enforces the other values. In the next section, we study these aspects specifically for the topological relations between regions with broad boundaries.

3.7.3 Topological relations between regions with broad boundaries

Eight topological relations are possible between two simple crisp regions. By considering these as the possible values in the four cells of the matrix, there are $8^4 = 4096$ possible matrices. However, definition 1 imposes a condition mandating that the extents of a region $\tilde{A}$ with a broad boundary should be related by one of the following relations: *Equal(*$\tilde{A}_{\max}, \tilde{A}_{\min}$*)*, *Contains(*$\tilde{A}_{\max}, \tilde{A}_{\min}$*), or Covers(*$\tilde{A}_{\max}, \tilde{A}_{\min}$*)*. Indeed, a 4-Intersection matrix cannot identify a topological relation between two regions with broad boundaries when this condition is violated. Thus, the contents of the matrix cells are not independent. For example, if the maximal extents are disjoint, it is inconsistent for an *Overlap* to exist between the minimal extents (figure 3.13). In figure 3.13, the sub-relation $O\ (\tilde{A}_{\min}, \tilde{B}_{\min})$ is grey to denote that is not allowed whereas $D\ (\tilde{A}_{\min}, \tilde{B}_{\min})$ is black to show that is permitted. Consequently, several of the 4096 possible matrices are invalid because the dependency between the cells of the matrix involved is not respected.



(a) Invalid *Disjoint*' relation (b) Valid *Disjoint*' relation

Figure 3.13 Controlling the validity of a *Disjoint* relation

In order to enumerate the valid 4-Intersection matrices, we firstly studied possible values in the other three cells for each of the eight possible values of $R(\tilde{A}_{max}, \tilde{B}_{max})$. For example, if *Contains* $(\tilde{A}_{max}, \tilde{B}_{max})$, the only possible relation between $\tilde{A}_{max}$ and $\tilde{B}_{min}$ is *Contains;* otherwise, the expected relation cannot respect the general definition of a region with a broad boundary. Figure 3.13 shows an example of an inconsistent matrix in which *Disjoint* $(\tilde{A}_{max}, \tilde{B}_{max})$ and *Contains* $(\tilde{A}_{min}, \tilde{B}_{min})$. This matrix is inconsistent because $R(\tilde{B}_{max}, \tilde{B}_{min}) \notin$ {*Contains*, *Covers*, *Equal*}. In the second step, we also fix the relation between $\tilde{A}_{min}$ and $\tilde{B}_{min}$ to deduce the possible values of $R(\tilde{A}_{min}, \tilde{B}_{max})$. For example, when *Contains* $(\tilde{A}_{max}, \tilde{B}_{max})$ and *Contains* $(\tilde{A}_{min}, \tilde{B}_{min})$, $R(\tilde{A}_{min}, \tilde{B}_{max})$ should not be *Meet* or *Equal*. In this way, 31 rules (cf. appendix 2) are defined in order to ensure the consistency of matrices and to minimize the number of topological relations. In the premises of rules, we specify either $R(\tilde{A}_{max}, \tilde{B}_{max})$ or ($R(\tilde{A}_{max}, \tilde{B}_{max})$ and $R(\tilde{A}_{min}, \tilde{B}_{min})$). Then, we deduce the possible values in the remaining cells. In figure 3.13, the matrix on the left is not valid because it requires the minimal extent to be disjoint to the minimal extent (i.e., the definition of regions with broad boundaries is not respected, because $R(\tilde{A}_{max}, \tilde{B}_{max})$ should be *Contains*, *Covers*, or *Equal*).

This study proves that only 242 topological relations are possible between two simple regions with broad boundaries (cf. appendix 1). More specifically, only one matrix is valid when *Disjoint* $(\tilde{A}_{max}, \tilde{B}_{max})$, 29 matrices are valid when *Contains* $(\tilde{A}_{max}, \tilde{B}_{max})$, 29 for *Inside* $(\tilde{A}_{max}, \tilde{B}_{max})$, 46 for *Covers* $(\tilde{A}_{max}, \tilde{B}_{max})$, 46 for *Covered by* $(\tilde{A}_{max}, \tilde{B}_{max})$, 65 for *Overlap* $(\tilde{A}_{max}, \tilde{B}_{max})$, 4 for *Meet* $(\tilde{A}_{max}, \tilde{B}_{max})$, and 22 when *Equal* $(\tilde{A}_{max}, \tilde{B}_{max})$. The topological relations are numbered from 1 to 242 according to the relation between $\tilde{A}_{max}$ and $\tilde{B}_{max}$. Table 3.1 shows this numbering (see the appendix 1 to explore these relations).

Table 3.1 Relations' numbers

The relation between ($\tilde{A}_{\max}$, $\tilde{B}_{\max}$)	Correspondent matrices
Disjoint ($\tilde{A}_{\max}$, $\tilde{B}_{\max}$)	1
Contains ($\tilde{A}_{\max}$, $\tilde{B}_{\max}$)	2→30
Equal ($\tilde{A}_{\max}$, $\tilde{B}_{\max}$)	31→52
Covers ($\tilde{A}_{\max}$, $\tilde{B}_{\max}$)	53→98
Covered by ($\tilde{A}_{\max}$, $\tilde{B}_{\max}$)	99→144
Inside ($\tilde{A}_{\max}$, $\tilde{B}_{\max}$)	145→173
Meet ($\tilde{A}_{\max}$, $\tilde{B}_{\max}$)	174→177
Overlap ($\tilde{A}_{\max}$, $\tilde{B}_{\max}$)	178→242

3.8 Clustering of topological relations between regions with broad boundaries

3.8.1 Principles

In our work, the proposed model is expressive in terms of the topological relations distinguished between regions with broad boundaries. In this context, 242 topological relations are enumerated. Consequently, the clustering of relations into larger groups of relations is an important step, because it is very difficult to keep in the mind this high number of relations. It is additionally very difficult to find a name for each one of them, and so the user will have difficulty of choosing the appropriate topological operator in order to express a query or an integrity constraint. Mark and Egenhofer (Mark and Egenhofer 1994) studied the clustering of topological relations between simple crisp regions and simple crisp lines both through a formal basis and by taking into account cognitive aspects. Clementini and Di Felice (1997) defined a *topological distance* to classify the approximate topological relations between regions with completely broad boundary. In this way, they deduced 17 clusters that they represent in a conceptual neighborhood graph.

In our approach, most of the distinguished topological relations are not completely different from each other. For example, two simple regions with broad boundaries can *weakly* or *completely overlap* each other depending on the content of the 4-Intersection matrix involved. In the first case, only the maximal extents overlap. In the second case, however, *Overlap* is the unique value in the matrix cells. Thus, it is possible to deduce the relation vagueness level according to the content of the 4-Intersection matrix. The objective of this section is to group the 242 topological relations into a limited number of clusters based on the content of their respective matrices.

3.8.2 Clustering results

In section 3.7, we showed that the global topological relationship is identified through a 4-Intersection matrix that enumerates four sub-relations. Thus, a topological relation becomes possible if it appears at least once in the matrix. This possibility increases according to the number of similar sub-relations. For example, a *Covers* topological relation in which *Covers* ($\tilde{A}_{max}$, $\tilde{B}_{max}$) and *Covers* ($\tilde{A}_{min}$, $\tilde{B}_{min}$) is stronger than another where only *Covers* ($\tilde{A}_{max}$, $\tilde{B}_{max}$). Because there are eight possible values for the matrix cells, we distinguish eight basic clusters that we call: *DISJOINT*, *CONTAINS*, *INSIDE*, *COVERS*, *COVERED BY*, *EQUAL*, *MEET*, *and OVERLAP*. Each cluster contains all of the topological relations for which at least one of the four sub-relations has the same name. For example, figure 3.14 shows a topological relation that belongs to the following clusters: *DISJOINT*, *CONTAINS*, and *COVERS*. Nevertheless, it belongs to the *DISJOINT cluster* more strongly than to the *CONTAINS* and *COVERS clusters.*

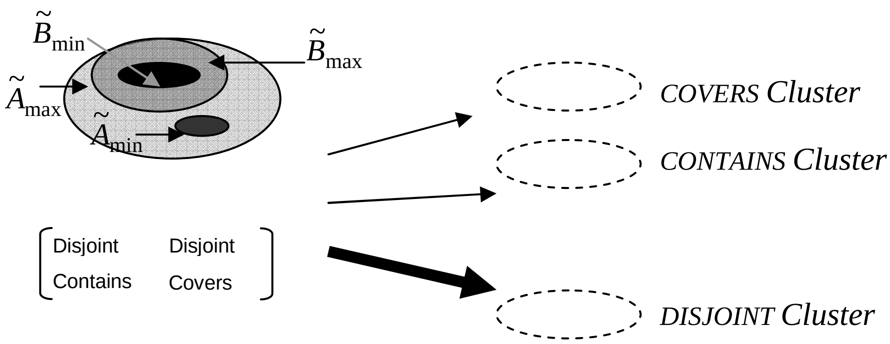


Figure 3.14 Example of clustering of a topological relation

For each one of the eight basic clusters, we identify four levels of relation membership: (1) *completely*, (2) *strongly*, (3) *fairly*, and (4) *weakly* (table 2). A topological relation belongs to the cluster *completely* when the four sub-relations are similar. It belongs to the cluster *strongly* when only three sub-relations have the same name as the cluster. The level labelled *fairly* contains all relations for which two sub-relations have the same name as the cluster. Finally, the level called *weakly* contains the relations for which only one sub-relation has the same name as the cluster. Figure 3.15 presents some relations that belong to different levels of

CONTAINS and *DISJOINT* clusters, respectively, according to the contents of their correspondent matrices.

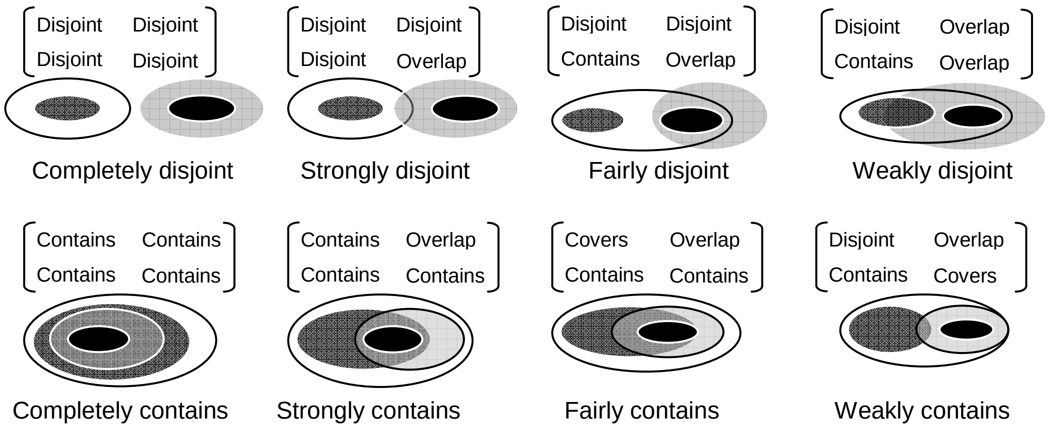


Figure 3.15 Evaluation of a topological relation membership to one of the eight basic clusters

3.8.3 Overlapping clusters

The main result of this clustering process is a hierarchical classification of the topological relations (figure 3.16). The top level is made up of eight basic clusters that each contains typically four levels: *completely*, *strongly*, *fairly*, and *weakly*. The resulting 32 sub-clusters overlap each other because a topological relation typically belongs to different levels of 1, 2, 3, or 4 clusters at the same time. For example, topological relation number 56 (see the appendix 1 and the table 3.2) belongs *fairly* to the *CONTAINS* cluster and *weakly* to the *COVERS* and *INSIDE* clusters. The bottom level of the classification contains the 242 topological relations that appear in different sub-clusters.

Table 3.2 Clustering results

Cluster's name	Vagueness level	Topological relations' numbers (cf. appendix 1)
DISJOINT	*Weakly*	13, 14, 15, 17, 41, 42, 43, 44, 67, 69, 70, 71, 72, 74, 75, 80, 113, 115, 116, 117, 118, 120, 121, 126, 157, 159, 161, 162, 193, 194, 195, 196, 197, 198, 199, 200, 201, 204, 205, 208, 213, 214, 215, 216
	Fairly	16, 73, 76, 119, 122, 158, 175, 176, 202, 203, 206, 207, 209, 210, 211, 212
	Strongly	174, 217
	Completely	1
CONTAINS	*Weakly*	31, 34, 36, 39, 43, 44, 45, 48, 51, 52, 57, 59, 61, 63, 67, 68, 69, 71, 73, 76, 77, 79, 80, 82, 85, 86, 88, 91, 93, 94, 95, 96, 102, 105, 110, 113, 118, 125, 128, 130, 135, 137, 140, 145, 146, 153, 157, 163, 167, 173, 181, 184, 186, 189, 193, 195, 198, 210, 213, 218, 219, 221, 223, 226, 230, 232, 234, 238, 240

	Fairly	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 32, 33, 37, 53, 54, 55, 56, 60, 103, 104, 152, 178, 179, 180
	Strongly	2, 3, 4, 5, 7
	Completely	6
EQUAL	*Weakly*	4, 25, 26, 29, 32, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 55, 77, 78, 79, 84, 91, 101, 123, 124, 125, 131, 138, 145, 148, 170, 171, 230, 231, 232, 233
	Fairly	24, 30, 31, 169
	Strongly	
	Completely	
COVERS	*Weakly*	7, 18, 19, 21, 25, 30, 35, 38, 40, 41, 42, 46, 47, 50, 52, 53, 55, 56, 57, 59, 60, 61, 63, 67, 68, 69, 71, 73, 76, 77, 79, 80, 88, 93, 94, 95, 96, 99, 107, 108, 115, 116, 123, 124, 129, 132, 135, 137, 139, 141, 147, 156, 159, 160, 166, 167, 171, 183, 185, 187, 192, 194, 196, 201, 211, 215, 218, 219, 221, 223, 225, 227, 228, 231, 233, 235, 239, 241
	Fairly	20, 49, 54, 58, 62, 64, 65, 66, 70, 72, 74, 75, 78, 82, 85, 86, 89, 90, 91, 92, 97, 98, 136, 142, 168, 220, 222
	Strongly	83, 84, 87
	Completely	81
COVERED BY	*Weakly*	3, 12, 14, 21, 22, 23, 26, 30, 33, 37, 39, 40, 42, 44, 45, 47, 49, 51, 53, 61, 62, 69, 70, 77, 78, 83, 85, 88, 90, 92, 94, 99, 101, 102, 103, 105, 106, 107, 109, 113, 114, 115, 117, 122, 123, 125, 126, 135, 140, 141, 142, 143, 151, 163, 164, 166, 170, 180, 186, 187, 191, 195, 196, 200, 212, 216, 219, 220, 223, 225, 226, 227, 229, 232, 233, 234, 235, 242
	Fairly	50, 89, 95, 100, 104, 108, 110, 111, 112, 116, 118, 119, 120, 121, 124, 128, 130, 132, 133, 136, 137, 138, 139, 144, 165, 224, 228
	Strongly	129, 131, 134
	Completely	127
INSIDE	*Weakly*	2, 9, 13, 18, 22, 23, 28, 29, 31, 32, 36, 38, 41, 43, 46, 48, 51, 52, 56, 59, 64, 67, 72, 79, 82, 88, 90, 93, 98, 103, 105, 107, 109, 113, 114, 115, 117, 122, 123, 125, 126, 128, 132, 133, 135, 138, 140, 141, 142, 143, 178, 184, 185, 188, 193, 194, 197, 209, 214, 218, 225, 226, 227, 229, 230, 231, 236, 240, 241
	Fairly	8, 34, 35, 57, 58, 99, 100, 101, 102, 106, 145, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 181, 182, 183
	Strongly	146, 147, 148, 149, 151
	Completely	150
MEET	*Weakly*	9, 11, 12, 15, 36, 38, 39, 40, 59, 61, 62, 63, 64, 66, 74, 80, 105, 107, 108, 109, 110, 112, 120, 126, 153, 155, 156, 161, 174, 184, 185, 186, 187, 188, 189, 190, 191, 192, 204, 205, 206, 207, 213, 214, 215, 216
	Fairly	10, 65, 68, 111, 114, 154, 175, 176, 208
	Strongly	

	Completely	177
OVERLAP	*Weakly*	5, 11, 17, 19, 28, 33, 45, 46, 47, 48, 60, 63, 66, 71, 75, 86, 87, 92, 93, 94, 98, 109, 112, 117, 121, 130, 133, 134, 139, 140, 141, 149, 155, 160, 162, 164, 173, 178, 180, 181, 183, 184, 185, 186, 187, 193, 194, 195, 196, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 223, 225, 226, 227, 228, 230, 231, 232, 233
	Fairly	27, 96, 97, 143, 144, 172, 179, 182, 188, 189, 191, 192, 197, 198, 200, 201, 202, 203, 204, 205, 221, 222, 224, 229, 234, 235, 240, 241
	Strongly	190, 199, 236, 238, 239, 242
	Completely	237

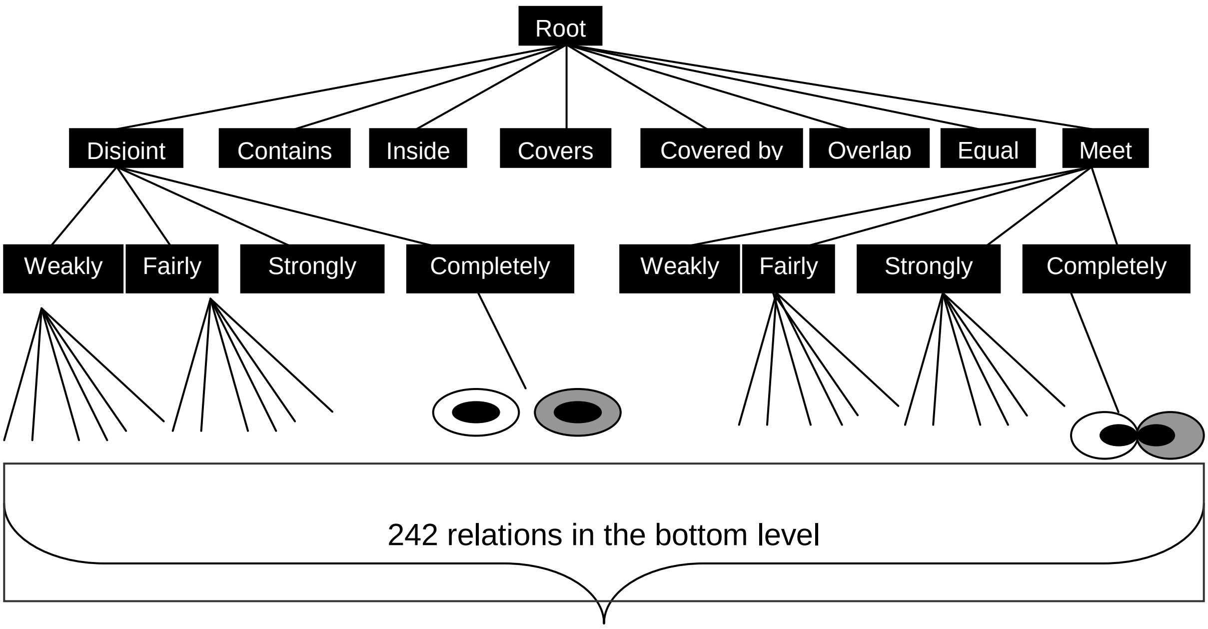


Figure 3.16 Hierarchical classification of the topological relations

3.9 Specification of spatial queries and integrity constraints

In the previous sections, we presented a framework for identifying topological relations between regions with broad boundaries. Because it uses the 9-Intersection model (Egenhofer and Herring 1990), our model can be easily integrated in a spatial database system. Indeed, the SQL language can be extended in order to retrieve regions with broad boundaries based on the qualitative information given by the user regarding their topological relations. In fact, a topological relation between two regions with broad boundaries can be recognized through the combination of four crisp topological operators. For example, relation number 56 corresponds to (*Disjoint*, *Disjoint*, *Contains*, *Covers*). Hereafter, we suppose that we integrated our spatial model in a relational engine in order to give an example of its possible use in spatial queries involving regions with broad boundaries. We suppose that the spatial database stores pollution zones, which are represented as regions with broad boundaries. In the first query example, the user gives a coarse description of the topological relation when he introduces the specification *fairly DISJOINT*. The query results should contain the *pollution zones* related to a user-defined zone *A* by a topological relation belonging to this sub-cluster. In the second example, the query is more specific because the user identifies all topological sub-relations that relate ($\tilde{A}_{\min}$, $\tilde{B}_{\min}$), ($\tilde{A}_{\min}$, $\tilde{B}_{\max}$), ($\tilde{A}_{\max}$, $\tilde{B}_{\min}$), and ($\tilde{A}_{\max}$, $\tilde{B}_{\max}$). The third example shows another use of our model, in which it is possible to display the different strength levels of a relation (e.g., *weakly Overlap* or *strongly Overlap*) that occurs between two regions with broad boundaries (cf. table 3.3). Table 3.3 shows a possible result for the query presented in example 3.

Table 3.3 Result of query 3

P1.id	*P2.id*	*Determine*
11	23	*Weakly overlap*
45	14	---
18	26	*Strongly Overlap*

Example 1: **Select** *Pollution_Zone.geometry* **From** *Pollution_Zone* **Where** **vague_Relate** (*pollution_zone.geometry*, *A.geometry*, *fairly DISJOINT*);

Example 2: **Select** *Pollution_Zone.geometry* **From** *Pollution_Zone* **Where** **vague_Relate** (*Pollution_Zone.geometry, A.geometry, Disjoint, Meet, Contains, Contains*);

Example 3: **Select** *P1.id, P2.id, determine (P1.geometry, P2.geometry, "Overlap")* **From** *Pollution_Zone P1, P2* **Where** *P1.id<>P2.id*;

In the same way, it is possible to use the model to formally express spatial integrity constraints for objects with vague shapes. For example, let the constraint saying that '*two different lakes can be only fairly meet or completely disjoint'*. This constraint can be formally expressed by integrating new spatial operators (e.g., *completely Contains*, *weakly Covers*, etc.) in a formal constraint language like the Object Constraint Language (OCL) (Pinet *et al.* 2007). The database storing the *lakes* is consistent only if the topological relations between the different entities belong to *fairly MEET* or *completely DISJOINT* sub-clusters (see example 4).

Example 4: **Context** Lake **inv**:
Lake.allInstances → **forAll** (a, b| a<>b implies ***fairly MEET***(a,b) **or** ***completely DISJOINT***(a,b));

3.10 Discussion

Clementini and Di Felice (1997) propose an extension of the 9-Intersection model (Egenhofer and Herring 1990) that uses a broad boundary to replace the sharp boundary. In this approach, 44 topological relations are distinguished between two regions with broad boundaries. By considering a topological distance, Clementini and Di Felice (1997) draw a conceptual neighborhood graph that shows similarity degrees between relations classified into 17 clusters. The main advantage of this approach is the ability to support a coarser spatial reasoning involving regions with broad boundaries. When the needs are more specific, it becomes more difficult to use this model. Furthermore, the identification of a broad boundary as a two-dimensional topological invariant requires respecting consistency conditions related to closeness and connectedness. Tang (2004) presents a more expressive model than that defined by Clementini and Di Felice (1997), because he decomposes the broad boundary into the *boundary's interior* and the *boundary's boundary*. Based on this definition, Tang (2004) presents another extension of 9-Intersection model, in which topological relations are identified through a 4*4-Intersection matrix. He distinguishes 152 topological relations presented as variants of the 44 relations proposed by Clementini and Di Felice 1997). Nonetheless, this model does not distinguish between the boundaries of the *minimal* and *maximal extents*. Accordingly, many topological relations cannot be distinguished (see examples in Section 2.3). Moreover, regions with *partially* broad boundaries (see example in figure 3.2) are considered invalid and cannot be presented through existing exact models. In our approach, we resolve this problem by considering a simple region with a broad boundary as a general concept which can be specialized into*: regions with none broad boundary* (or crisp regions), *regions with a partially broad boundary* and *regions with a completely broad boundary*. A region is then defined as a maximal extent and a minimal extent, in which either *Equal (* $\tilde{A}_{\max}, \tilde{A}_{\min}$) or *Contains (* $\tilde{A}_{\max}, \tilde{A}_{\min}$) or *Covers (* $\tilde{A}_{\max}, \tilde{A}_{\min}$). The notion of broad boundary (i.e., in the sense of connected and closed polygonal zone) is not formally defined as a topological invariant in our model. It can be deduced from the difference between the minimal extent and the maximal one. This difference can be non-empty everywhere around the minimal extent (i.e., *region with completely broad boundary*), non-empty in some location and empty in some others (i.e., *region with partially broad boundary*) or empty everywhere around the minimal extent (i.e., *crisp region*). Our main motivations for adopting this framework are (1) to consider regions with partially broad boundaries and (2) to present an expressive model in terms of the identification of topological relations between regions with broad boundaries. With regards to principal exact models (Clementini and Di Felice 1997, Cohn and Gotts 1996, Erwig and Schneider 1997, Tang 2004), our approach allows to make distinction between *partial shape vagueness* and *complete shape vagueness*. This distinction is very important in order to deal with two main problems: an ontological problem and a modeling one. First, the ontological problem means that "*shape vagueness*" is generally considered as a "*binary imperfection*" (i.e., only two possibilities are considered for an object's shape: crisp or vague). Spatial objects can be characterized by different levels of shape vagueness (e.g., how can we classify a region with partially broad boundary? Is - it a crisp or a vague region?). These levels are easily computed in fuzzy models by using a *quantitative* approach. In our submission, we try to categorize two levels by using a *qualitative* approach because we believe that "*shape vagueness*" is a qualitative problem. It is clear that our approach cannot provide a fine computation of shape vagueness as in fuzzy models. However, we believe that our model provides a solution to qualitatively distinguish different levels of shape vagueness in the category of exact models. We do not claim that exact models are better than fuzzy ones, because the needs are not identical and therefore the direct comparison is not appropriate. Second, the modeling problem refers to the lack of expressivity in existing exact models to represent the objects, which include sharpness and broadness in their topological invariants at the same time. To deal with this second problem, our model can formally represent regions with *partially broad boundary* in addition to those with *completely broad boundary*. This distinction is ignored in the most of existing exact models; notably in (Clementini and Di Felice 1997, Cohn and Gotts 1996, Erwig and Schneider 1997, Tang 2004)).

For topological relationships, we propose a 4-Intersection matrix where it is possible to identify respective sub-relations between minimal extents and maximal ones: ($\tilde{A}_{\min}$, $\tilde{B}_{\min}$), ($\tilde{A}_{\min}$, $\tilde{B}_{\max}$), ($\tilde{A}_{\max}$, $\tilde{B}_{\min}$), and ($\tilde{A}_{\max}$, $\tilde{B}_{\max}$). These sub-relations are labelled by using the 9-Intersection model (Egenhofer and Herring 1990). In our paper, 31 rules (or strategies) have been defined in order to minimize the number of topological relations between regions with broad boundaries and to control their consistency. In this context, we would clarify that the seven first strategies defined in (Schmitz and Morris 2006) can be considered as a subset of our 31 rules (see these rules in the appendix 2). More specifically, Strategy 1 (Schmitz and Morris 2006) can correspond to Rule 1 in our model, Strategy 2 <==> Rule 2, Strategy 3 <==> Rule 3, Strategy 4 <==> Rule 3 (this rule is applied for *Inside* and *Contains* relations), Strategy 5 <==> Rule 5, Strategy 6 <==> Rule 6 and Strategy 7 <==> Rule 6 (this rule is applied for *Inside* and *Contains* relations). The 8th strategy presented in (Schmitz and Morris 2006) does not provide any indication about the appropriate topological sub-relations when overlap relations arise between components of regions with broad boundaries involved (i.e., it recommends additional investigations). However, in our paper, we propose eight strategies when an *Overlap* relation occurs between maximal extents of two regions with broad boundaries (Rule 20 – Rule 27). Then, incoherent and redundant topological relations have been removed by using the 31 rules presented in the appendix 2. We distinguish 242 different topological relations that we classify into eight overlapping basic clusters. Each cluster has four membership levels (or sub-clusters): *completely, strongly, fairly,* and *weakly*. This classification of the topological relations is proposed to support an adverbial expression of topological integrity constraints. Nevertheless, our model is not able to quantify the gradual change inside the maximal extent in the same way as the fuzzy approaches do (Zhan 1997, Schneider 2001, Du *et al.* 2005, Dilo 2006, Verstraete *et al.* 2007). Finally, we are convinced that a more detailed comparison of the models' expressivity requires to be thoughtfully investigated in another paper.

The Egg-Yolk model (Cohn and Gotts 1996) was our main inspiration to develop this framework for identifying topological relations. However, there are some fundamental differences between our model and that defined in (Cohn and Gotts 1996). For instance, the topological relations used in (Cohn and Gotts 1996) are those defined in the RCC-5 model (Randell and Cohn 1989, Cohn *et al.* 1997). In contrast, the topological relations used in the cells of our matrix are those defined in the 9-Intersection model (Egenhofer and Herring 1990). It is true that we follow the same methodology to identify topological relations. However, our definitions of *objects with vague shapes* are substantially different. Our model is based on the point-set theory where points and lines are considered as basic crisp spatial object types. In terms of originality, we do not formally redefine the concept '*broad boundary*' as it is done in most of existing exact models. Our approach is based on the distinction between a minimal extent and a maximal one. The broad boundary can be deduced from the difference between these two extents but it is not defined as a topological invariant of the object. In (Cohn and Gotts 1996), a conceptual neighborhood graph was drawn with 44 topological relations are classified into 13 clusters. In our model, we define a hierarchical classification based on the content of the matrices we use to identify the topological relations. This classification is the basis of an adverbial approach that we use to specify topological integrity constraints between regions with broad boundaries.

3.11 Conclusions and future works

Shape vagueness is an inherent property of many spatial objects like *lakes*, *valleys*, and *mountains*. In GIS and spatial databases, it is a general practice to neglect shape vagueness and formally represent spatial objects with vague shapes as crisp geometries. Using such inappropriate representations can provide a source of spatial data quality degradation, because the reliability of spatial data is decreased. With emergence of prediction applications, data integration, and strategic decisional needs, researchers are increasingly motivated to propose different methods for the formal representation of shape vagueness. A review of the literature regarding this topic proves that existing exact models do not permit the representation of *objects with partially vague shape*. For such objects, shape vagueness partially characterizes one or several of its topological invariants. For example, a lake can have *rocky banks* on one side and *swamp banks* on the other side at the same time; the boundary is broad only for the *swamp* part. In this work, we have proposed an exact model in order to represent spatial objects that can have: *crisp shapes*, *partially vague shapes*, or *completely vague shapes*. We have considered this categorization of shape vagueness during the identification of topological relations.

More specifically, this paper contributes in three main ways. Based on point-set topology, we firstly define three basic *types of spatial objects with vague shapes*: *broad point*, *line with a vague shape (i.e., lines with broad boundaries, lines with broad interiors or broad lines),* and *region with a broad boundary*. Each one of them is typically defined as a minimal extent $\tilde{A}_{\min}$ and a maximal extent $\tilde{A}_{\max}$, and these extents must verify some topological conditions in order to be valid. This model permits the representation of spatial objects with partially vague shapes considered as invalid in the existing models of (Clementini and Di Felice 1997, Tang 2004, Reis *et al.* 2006). Then, we identify a topological relation through use of a 4-Intersection matrix that permits the enumeration of four sub-relations: R_1 ($\tilde{A}_{\min}, \tilde{B}_{\min}$), R_2 ($\tilde{A}_{\min}, \tilde{B}_{\max}$), R_3 ($\tilde{A}_{\max}, \tilde{B}_{\min}$), and R_4 ($\tilde{A}_{\max}, \tilde{B}_{\max}$). By using this formalism for simple regions with broad boundaries, 242 relations can be distinguished (cf. appendix 1). In order to retain our propositions useful in practice, we propose the clustering of these topological relations. A topological relation can belong to one or several clusters with various qualitative strengths: *completely*, *strongly*, *fairly,* and *weakly*. The objective of this qualitative clustering is to improve the specification of spatial queries and integrity constraints involving spatial objects with vague shapes.

In this paper, our study is limited to the regions with broad boundaries which are composed by a simple core (or minimal extent). Extending this approach to regions with more complex shapes (e.g., regions with broad boundaries and holes, regions with several cores, regions composed by disjoint uncertain sub-regions, etc.) is one of our future researches. We are conscious that it can be a limitation of our current model but considering this type of regions requires additional investigations which exceed the objectives of this paper. The goal of this paper is to clearly present the basis of our approach before improving it. Another extension consists of using this approach to improve the logical consistency of spatial databases involving spatial objects with vague shapes. More specifically, we are interested in the specification of integrity constraints in spatial databases storing objects with vague shapes. We hope to identify both integrity constraint categories and the requirements for their formal expression. The framework presented earlier can provide a basis for the extension of a formal constraint language like OCL (Pinet *et al.* 2007) to express *tolerant integrity constraints* for objects with vague shapes.

Finally, this approach can be used to deal with geometric heterogeneities between sources databases in decisional applications. These applications require the integration of spatial data from heterogeneous sources before they are stored in a spatial data warehouse (Bédard *et al.* 2007). The main difficulty lies in choosing one of the available geometric representations. We suggest merging the different representations in such way that the result looks like a spatial object with a vague shape. The *tolerant* integrity constraints can be used to increase the logical consistency of such data.

References

AHLQVIST, O., KEUKELAA J., and OUKBIR A., 1998, Using Rough classification to Represent Uncertainty in Spatial Data. In *Proceedings of the Tenth Annual Colloquium of the Spatial Information Research Centre*. P. Firns (editors). 16 - 19 Dec, Dunedin, New Zealand. University of Otago, ISBN 1877139122, pp. 1-10.

ALTMAN, D., 1987, Fuzzy set theoretic approaches for handling imprecision in spatial analysis. *International Journal of Geographical Information Systems*, **8**, pp. 271-289.

BÉDARD, Y., 1987, Uncertainties in Land Information Systems Databases, In *Proceedings of Eighth International Symposium on Computer-Assisted Cartography*, Baltimore, Maryland (USA), 29 Mars - 3 Avril 1987, American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping, pp. 175-184.

BÉDARD, Y., RIVEST, S. and PROULX, M.-J., 2007, Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective. In *Data Warehouses and OLAP: Concepts, Architectures and Solutions*, Robert Wrembel and Christian Koncilia (Ed.) (London: IRM Press (Idea Group)), Chap. 13, UK, pp. 298-319.

BJØRKE, J. T., 2004, Topological Relations Between Fuzzy Regions: Derivation of Verbal Terms. *Fuzzy sets and systems*, **141**, 449-467.

BORDOGNA, G. and CHIESA, S., 2003, A Fuzzy Object-Based Data Model for Imperfect Spatial Information Integrating Exact Objects and Fields. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **11(1)**, pp. 23-42.

BROWN, D.G., 1998, Classification and boundary vagueness in mapping presettlement forest types. *International Journal of Geographical Information Systems*, **12**, pp. 105-129.

BURROUGH, P.A., 1989, Fuzzy mathematical methods for soil survey and land evaluation. *Journal of Soil Science*, **40**, pp. 477–492.

BURROUGH, P.A. and FRANK, A.U., 1996, Geographic Objects with Indeterminate Boundaries (London: Taylor & Francis).

CHENG, T., MOLENAAR, M. and LIN, H., 2001, Formalizing fuzzy objects from uncertain classification results. *International Journal Geographical Information Science*, **15(1)**, pp. 27-42.

CLEMENTINI, E. and DI FELICE, P., 1997, Approximate topological relations. *International Journal of Approximate Reasoning,* **16**, pp. 173-204.

Clementini, E., 2005, A model for uncertain lines. In *J. Vis. Lang. Comput,* **16(4)**, pp. 271-288.

COHN, A.G. and GOTTS N.M., 1996, The 'egg-yolk' representation of regions with indeterminate boundaries. In *Proceedings of the GISDATA Specialist Meeting on Spatial Objects with Undetermined Boundaries*, Burrough, P. & Frank, A. (Ed.) (Taylor & Francis), pp. 171-187.

COHN, A.G., BENNETT, B., GOODAY J. and GOTTS, N.M., 1997, Qualitative Spatial Representation and Reasoning with the Region Connection Calculus, *GeoInformatica*, **1(3)**, pp. 275–316.

CHRISMAN, N.R., 1991, The error component in spatial data. In: Maguire, D.J., Goodchild, M.F and Rhind, D.W. (eds.), *Geographical Information Systems: Principles and Applications,* **Volume 1**, pp. 165-174.

DE TRÉ, G., DE CALUWE, R., VERSTRAETE, J. and HALLEZ, A., 2004, The applicability of generalized constraints in spatio-temporal database modeling and querying. In *Spatio-Temporal Databases: Flexible Querying and Reasoning,* De Caluwe, R., De Tré, G. and Bordogna, G. (Eds.), Heidelberg, Germany: Springer, pp. 127-158.

DEVILLERS, R., BÉDARD, Y., JEANSOULIN, R., and MOULIN, B., 2007, Towards Spatial Data Quality Information Analysis Tools for Experts Assessing the Fitness for Use of Spatial Data, International Journal of Geographical Information Sciences (IJGIS), Vol. 21, **3**, pp. 261-282.

DILO, A., 2006, Representation of and reasoning with vagueness in spatial information: A system for handling vague objects. PhD thesis, ITC, Netherlands, 187p.

DU, S., QIN, Q., WANG, Q. and LI, B., 2005, Fuzzy description of topological relations I: a unified fuzzy 9-intersection model. *In: L. Wang, K. Chen, Y.S. Ong (Eds.), Advances in Natural Computation*, Lecture Notes in Computer Science, vol. **3612**, pp. 1260-1273.

DUCKHAM, M., MASON, K., STELL, J., and WORBOYS M., 2001, A formal ontological approach to imperfection in geographic information. *Computer, Environment and Urban Systems*, **25**, pp. 89-103.

EGENHOFER, M.J., 1989, A formal definition of binary topological relations. In *Poceedings of the third international conference on Foundations of Data Organisation and Algorithms (FODO),* W.Litin and H.J.Scheck (Ed.) (NY: Springer-Verlag), Lecture notes in computer science, **367**, pp. 457-472.

EGENHOFER, M. J. and FRANZOSA, R.D., 1991, Point-set Topological Relations. *International journal of geographical Information Systems*, **5(2)**, pp. 161-174.

EGENHOFER, M. and HERRING J., 1990, A mathematical framework for the definition of topological relations. *In Proceedings of the Fourth International Symposium on Spatial Data Handling,* K. Brassel and H. Kishimoto (Ed.), Zurich, Switzerland , pp. 803-813.

ERWIG, M. and SCHNEIDER, M., 1997, Vague regions. In *5th International Symposium on Advances in Spatial Databases (SSD'97),* Lecture Notes in Computer Science, **1262**, pp. 298-320.

FISHER, P.F., 1999, Models of uncertainty in spatial data. In *Geographical Information Systems*, P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind (Ed.) (New York: John Wiley & Sons), pp. 191-205.

FRANK, A.U., 2001, Tiers of ontology and consistency constraints in geographical information systems. In *International Journal of Geographical Information Science*, **15(7)**, 667--678.

GODJJEVAC, J., 1999, Idées nettes sur la logique floue. Presses polytechniques et universitaires romandes, Lausanne.

GOODCHILD, M.F., 1995, Attribute Accuracy. In *Elements of spatial data quality*, S.C. Guptill and J.L. Morrison (Ed.) (New York: Elsevier Science inc.), pp. 59-79.

GUPTILL, S.C. and MORRISON, J.L., 1995, Spatial data quality. In *Elements of spatial data quality*, S. C. Guptill and J. L. Morrison (Ed.) (New York: Elsevier Science inc.).

HAZARIKA, S.M. and COHN, A.G., 2001, A taxonomy for spatial vagueness, an alternative egg-yolk interpretation. In *Proceeding of the COSIT 2001, The fifth international Conference On Spatial Information Theory*, D.R. Montello (Ed.) (California: Springer-Verlag), Lecture Notes in Computer Science, **2205**, pp. 92-107.

HWANG, S., and THILL, J-C., 2005, Modeling Localities with Fuzzy Sets and GIS, In: Cobb M, Petry F, and Robinson V (eds) *Fuzzy Modeling with Spatial Information for Geographic Problems*, Springer-Verlag, pp. 71-104.

MARK, D. and EGENHOFER, M., 1994, Modeling Spatial Relations Between Lines and Regions: Combining Formal Mathematical Models and Human Subjects Testing, *Cartography and Geographical Information Systems*, **21 (3)**, pp. 195-212.

MOWRER, H.T., 1999, Accuracy (Re)assurance: Selling Uncertainty Assessment to the Uncertain. In *Spatial Accuracy Assessment, Land Information Uncertainty in Natural Ressources*, K. Lowell and Jaton A. (Ed.) (Quebec: Ann Arbor Press), pp. 3-10.

PAWLAK, Z., 1994, Rough sets: present state and further prospects. In *Third International Workshop on Rough Set and Soft Computing (RSSC '94)*, pp. 72–76.

PFOSER, D., TRYFONA, N. and JENSEN, C.S., 2005, Indeterminacy and Spatiotemporal Data: Basic Definitions and Case Study. *GeoInformatica*, **9(3)**, pp. 211-236.

PINET, F., DUBOISSET, M. and SOULIGNAC, V., 2007, Using UML and OCL to maintain the consistency of spatial data in environmental information systems. *Environmental modeling & software*, **22(8)**, pp. 1217-1220.

RANDELL, D.A. and COHN A.G., 1989, Modeling topological and metrical properties of physical processes. In *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning (KR'89)*, R. J. Brachman, H. J. Levesque, and R. Reiter (Ed.) (Morgan Kaufmann), pp. 357–368.

REIS, R., EGENHOFER, M.J. and MATOS, J., 2006, Topological relations using two models of uncertainty for lines. In. *Proceeding of the 7th international Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, 5 - 7 July, Lisbon, Portugal, pp. 286-295.

ROBINSON, V.B. and THONGS, D., 1986, Fuzzy Set Theory Applied to the Mixed Pixel problem of Multi-spectral Land Cover Databases. *GIS in Government*, **2**, B.K. Opitz (Ed.) (Washington D.C: A.Deepak Publication).

RODRIGUEZ, A., 2005, Inconsistency Issues in Spatial Databases. *Lecture Notes In Computer Science*, **3300**, pp. 237-269.

ROY, A.J. and STELL J.G., 2001, Spatial relations between indeterminate regions. *International Journal of Approximate Reasoning*, **27**, pp. 205—234.

SCHMITZ, A. and MORRIS, A., 2006, Modeling and Manipulating Fuzzy Regions: Strategies to Define the Topological Relation between two Fuzzy Regions. *Control and Cybernetics*, May 2006.

SCHNEIDER, M., 2001, A design of topological predicates for complex crisp and fuzzy regions, In ER '01: *Proceedings of the 20th International Conference on Conceptual Modeling*, Springer-Verlag, ISBN 3-540-42866-6 pp. 103–116.

SHI W. and LIU K., 2007, A fuzzy topology for computing the interior, boundary, and exterior of spatial objects quantitatively in GIS. *Computer & Geosciences*, **33(7)**, July 2007, pp. 898-915.

SHU H., SPACCAPIETRA, S., PARENT, C. and QUESADA SEDAS, S., 2003, Uncertainty of Geographic Information and its Support in MADS. In *Proceeding of the 2nd International Symposium on Spatial Data Quality*, Hong Kong.

SMITHSON M., 1989, *Ignorance and Uncertainty: Emerging Paradigms* (New York: Springer – Verlag).

SOMODEVILLA, M. and PETRY, F.E., 2003**,** Approximation of topological relations on fuzzy regions: an approach using minimal bounding rectangles Fuzzy Information Processing Society. *NAFIPS 2003, 22nd International Conference of the North American*, 24-26 July 2003, pp. 371 – 376.

TANG, T., 2004, Spatial object modeling in fuzzy topological spaces: with applications to land cover change. PhD thesis, University of Twente, ISBN 90-6164-220-5.

TANG, X., Kainz W., and Fang Y., 2003, Modeling of fuzzy spatial objects and their topological relations. *In: Proceedings of the 2nd Symposium on Spatial Data Quality* (*ISSDQ*), W.Z. Shi, M.F. Goodchild and P. Fisher (eds), Hong Kong, pp. 34-50.

UBEDA, T. and EGENHOFER M., 1997, Topological Error Correcting in GIS. In *the Proceedings of International Symposium on Large Spatial Databases*, *Lecture Notes in Computer Science*, Vol. **1262**, Springer-Verlag, pp. 283-297.

VAN OORT, P., 2006, Spatial data quality: from description to application. *Publication on Geodesy 60* (Netherlands: Geodetic Commission), ISBN 90 6132 295 2, Delft, December.

VARZI, A., 2004, Mereology. In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (Ed.).

VERSTRAETE, J., HALLEZ, A., and De TRÉ, G., 2007, Fuzzy Regions: Theory and Applications. In A. Morris, S. Kokhan (Eds.). *Geographic Uncertainty in Environmental Security*, pp. 1-17. Dordrecht, The Netherlands: Springer.

WORBOYS, M.F., 1998, Imprecision in finite resolution spatial data. *GeoInformatica*, **2**, pp. 257-279.

WIKIPEDIA, 2008, http://en.wikipedia.org/wiki/Flight_19, last modification on 17 February 2008.

YAZICI A., ZHU, Q. and SUN N., 2001, Semantic data modeling of spatiotemporal database applications. *Int. J. Intell. Syst*, pp. 881-904.

YONGMING, L. and SANJIANG, L., 2004, A Fuzzy Sets Theoretic Approach to Approximate Spatial Reasoning. *IEEE Transaction on Fuzzy Systems*, **12(6)**, pp. 745- 754.

ZADEH, L.A., 1965, Fuzzy sets. *Inform. Control*, vol. **8**, pp. 338–353.

ZHAN, B.F.,1997, Topological relations between fuzzy regions. In *Proceedings of the 1997 ACM Symposium on Applied Computing*, (ACM Press), 192–196., ISBN 0-89791-850-9.

ZHAN, F. B. and LIN., H., 2003, Overlay of Two Simple Polygons with Indeterminate Boundaries. *Transactions in GIS*, **7(1)**, pp. 67-81.

Chapter 4: Qualitative Min-Max model for lines with vague shapes and their topological relations

Lotfi BEJAOUI, Yvan BÉDARD, François PINET, Michel SCHNEIDER

Transactions in GIS

(Submitted)

4.1 Résumé de l’article

Le vague de forme est plus difficile à modéliser pour les lignes que pour les régions (Clementini 2005). Deux types de lignes ayant des formes vagues sont généralement distingués : (1) *les lignes ayant des frontières vagues* et (2) *celles qui sont complètement vagues* (Clementini and Di Felice 1997, Reis et *al.* 2006). Cependant, l'intérieur d'une ligne peut être *partiellement* ou *complètement vague* indépendamment des points finaux. La forme d'une ligne peut être également vague quand seulement une des points finaux est *vague*. En effet, un problème conceptuel caractérise les travaux existants où différents types et niveaux de vague de forme ne sont pas considérés. Ce problème implique le besoin d’une méthode permettant l’identification des relations topologiques entre les lignes avec différentes formes vagues. Cet article propose une approche qualitative appelée le modèle QMM (acronyme de Qualitative Min-Max), où des lignes avec des niveaux différents de vague de forme sont distingués : *aucun*, *vague de forme partiel* et *vague de forme complet*. Nous définissons formellement une ligne avec la forme vague en tant qu’une combinaison d’une extension

minimale et une autre maximale. Les relations topologiques sont alors identifiées en fonctions des sous-relations entre les extensions minimales et maximales respectives des lignes impliquées. Le *poids* d'une relation topologique peut être exprimée qualitativement en employant des adverbes tels que *faiblement* ou *fortement*. Cette approche peut être servir à exprimer des contraintes topologiques et des requêtes spatiales sur des lignes ayant des formes vagues.

4.2 Abstract

Shape vagueness about lines is more complicated to model than about regions (Clementini 2005). Two types of lines with vague shapes are generally distinguished: (1) *lines with broad boundary* and (2) *completely broad lines* (Clementini and Di Felice 1997, Reis et *al.* 2006). However, a line's interior can be *partially* or *completely broad* independently of the endpoints. A line's shape can be also vague when only one of the endpoints is *broad*. Then, there is a conceptual problem, because different types and levels of shape vagueness are ignored in existing works. Overcoming this problem implies studying the identification of topological relations between lines with different vague shapes. This paper proposes a qualitative approach called Qualitative Min-Max model (QMM model for short), where different levels of shape vagueness of lines are distinguished: *none*, *partial vagueness* and *complete vagueness*. We formally define a line with vague shape as having *a minimal extent* and *a maximal one*. The topological relations are then specified according to the sub-relations between respective minimal and maximal extents of lines involved. The strength of a topological relation can be qualitatively expressed by using a set of adverbs such as *weakly* or *fairly*. This approach can be integrated into a framework to express topological integrity constraints and spatial queries.

4.3 Introduction

Topological errors can refer to the anomalies in an object's shape (e.g., *unclosed polygon*) or more often to an invalid topological relation between two objects (e.g., an overlap relation between two buildings). These topological properties and relations change according to the shapes of spatial objects stored in the database (Ubeda and Egenhofer 1997) as well as over time if objects move, enlarge, rotate, etc. Moreover, some researches notably in (Altman 1987, Burrough and Frank 1996, Cohn and Gotts 1996, Hunter and Goodchild 1996, Erwig and Schneider 1997, Couclelis 1996)) proved that spatial objects can have vague shapes (e.g. *regions with broad boundaries like a pollution zone*) and/or uncertain positions. These spatial data imperfections are generally caused by the complexity of reality and limitations of the measurement instruments and processes (Bédard 1987). Shape vagueness occurs when there is a difficulty to distinguish an object shape from its neighborhood and therefore the topological invariants (e.g., *a broad boundary*) could not have the same definitions as in the crisp context (Winter 2000). Using crisp spatial object types to represent spatial objects with vague shapes entails a gap between the knowledge that we have about spatial objects and their formal representation in spatial databases and GIS (Cheng and Lin 2001, Yazici et al. 2001). Then, the topological properties and relations can also change whether the objects manipulated have vague shapes such as regions with broad boundaries (e.g., *a pollution zone*), lines with vague shapes (e.g., *the trajectory of an historic explorer*) or broad points (e.g., *a wreck on the bottom of the sea*).

In the literature, the topological aspects for regions with broad boundaries have been thoughtfully explored (Altman 1987, Burrough and Frank 1996, Cohn and Gotts 1996, Erwig and Schneider 1997, Zhan 1997, Hazarika and Cohn 2001, Roy and Stell 2001, Winter 2000, Morris 2003, Robinson 2003, Zhan and Lin 2003, Tang 2004, Dilo 2006, Bejaoui et *al.* 2008). However, lines with vague shapes have not received the same attention except in few works (Clementini and Di Felice 1997, Clementini 2002, Clementini 2005, Reis et *al.* 2006). These last approaches proposed modeling of lines by using the appropriate shapes (i.e., *using two-dimensional parts which denote the shape vagueness such as broad endpoints*) and emphasizing of lines shape vagueness during the identification of topological relations (e.g., *connection*, *crossing*, etc). Two types of lines with vague shapes are generally distinguished: (1) *lines with broad boundary* and (2) *completely broad lines*. However, the interior of a given line can be *partially* or *completely broad* independently of the boundary (or endpoints). A line's shape can also be considered as vague when only one of the endpoints is *broad* (e.g., *an engine trajectory with only one ill-defined endpoint*). However, existing works (Clementini and Di Felice 1997, Clementini 2002, Clementini 2005, Reis et *al.* 2006) do not explicitly and exhaustively distinguish these different types and levels of shape vagueness for lines. Furthermore, the shape vagueness affects the identification of topological relations, which depend on the objects' shapes. It is the second main problem addressed in this work.

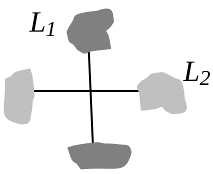
In this paper, we study the different types and levels of shape vagueness which can characterize the topological invariants of a given line (i.e., *boundary* and *interior*). We look for a new geometric model to describe different levels of shape vagueness of the boundary and/or interior. More specifically, we aim to make a distinction between the notions of broad interior and broad boundary, because each can be vague independently of the other. In this paper, this distinction is useful for simple lines with vague shapes and it will be extended for multi-lines and polygons. Additionally, a topological invariant (i.e., the interior or the boundary) of a given line can be characterized by one of the following levels of shape vagueness: *none* (i.e. *the topological invariant is well-defined*), *partial shape vagueness* and *complete shape vagueness.* In the same way, we aim to describe the vagueness of a topological relation by using a qualitative approach. We think that is pertinent for users to know whether two lines with vague shapes are *weakly* or *strongly* connected. For that, we define a line with a vague shape as a *minimal extent* (i.e., *it contains all of the points which certainly belong to the line*) included into a *maximal extent* (i.e., *it contains all of the points which possibly belong to the line*). The difference between these two extents refers to the shape vagueness of the line involved. Therefore, the topological relations between two lines with vague shapes can be qualitatively identified according to sub-relations between their respective extents. These sub-relations are identified through an extension of CBM method (Clementini and Di Felice 1995) which provides a limited number of topological operators that are more expressive than those defined in the 9-Intersection model (Egenhofer and Herring 1990, Clementini and Di Felice 1995, Clementini 2005). Our approach can be then seen as an extension of existing geometric models for objects with well-defined shapes. This model can be simply used to support the specification of topological relations in queries and integrity constraints by using a set of adverbs (e.g., *weakly*, *fairly*, *strongly,* and *completely*), which denote the vagueness of a relation to occur between the crisping of lines with vague shapes. The crisping of a line with a vague shape refers to any line with well-defined endpoints and interior that is strictly inside the spatial extent covered by the line with a vague shape (Bennett 2000, Clementini 2005). We call this approach: the *Qualitative Min-Max* model (*QMM* model for short), because it deals with shape vagueness in a qualitative way by distinguishing different types of lines with vague shapes according to the difference between the minimal extent and the maximal one. This first part of model is called *Qualitative Min-Max Definitions* (QMM$_{Def}$ for short), because it includes the principles of the spatial model to represent the shape vagueness for linear geometries. In addition, QMM model includes a second part called *Qualitative Min-Max Topological relations (*QMM$_{TR}$ for short) used to identify the topological relations between lines with vague shapes by studying the sub-relations between the minimal extents and maximal extents of lines involved. The vagueness of each topological relation can be qualified by using a set of adverbs such as *weakly* or *fairly*. We denote that we speak about the same model presented in the previous chapter and applied to regions with broad boundaries. We recall that the acronym QMM model have been proposed after acceptance of the first paper in order to facilitate reference to our approach.

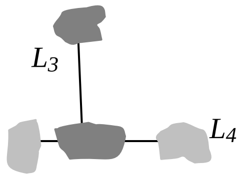
The remainder of this paper is organized as follows. In Section 4.4, we explore some previous works on the definition of lines with vague shapes and their topological relationships. In Section 4.5, we present the QMM_{Def} model for lines with vague shapes, where we thoughtfully underline the different levels of shape vagueness. In Section 4.6, we propose the QMM_{TR} model in order to identify topological relations between lines with vague shapes. For that, we propose an extension of the CBM method in order to identify the topological sub-relations, which occur between minimal and maximal extents of lines involved. After that, we define a 4-Intersection matrix in order to describe these sub-relations and classify topological relations. Section 4.7 proposes an adverbial approach to classify the topological relations by using the similarity between the sub-relations enumerated in their respectives 4-Intersection matrices. In Section 4.8, we show how this *adverbial* approach can be used to express topological integrity constraints and spatial queries involving lines with vague shapes. Section 4.9 draws the conclusions and some perspectives of this work.

4.4 Shape vagueness for lines

Shape vagueness occurs when an intrinsic property of the object or a lack of knowledge does not allow to sharply distinguish this object from its neighborhood (Bejaoui et *al.* 2008). For regions, the shape vagueness is generally correlated to the boundary which should be *broad*. For example, a lake can be considered as a region with a broad boundary, because its limits change according to the level of precipitation. Two types of models are generally used to represent objects with vague shapes. Exact models such as Burrough and Frank (1996), Cohn and Gotts (1996), Clementini and Di Felice (1997), Erwig and Schneider (1997) and Hazarika and Cohn (2001) proposed the extension of the models defined for crisp objects to underline the vagueness of the boundary (e.g., *the one-dimensional boundary is replaced by a broad one*) without any hypothesis about its internal structure. The main advantage of this approach is its simplicity to be integrated in existing spatial database systems (Erwig and Schneider 1997, Clementini and Di Felice 1997). Other approaches (Altman 1987, Brown 1998, Burrough and Frank 1996, Dilo 2006, Robinson and Thongs 1986, Schneider 2001, Tang 2004, Zhan 1997, Morris 2003, Robinson 2003, Zhan and Lin 2003) are based on Fuzzy Sets Theory (Zadeh 1965) in order to precisely describe the structure of broad boundary, or on Rough Sets (Pawlak 1994) (e.g., (Worboys 1998(b))), or (3) on the probability theory (e.g., (Burrough and Frank 1996, Pfoser et *al.* 2005)). For fuzzy models, some quantitative hypotheses should be set in order to define mathematical functions associated to the spatial objects with vague shapes. Furthermore, these approaches are expensive in implementation and they generally require an important effort to be manipulated by users (Clementini 2005).

For lines, the shape vagueness cannot be only correlated to the boundary (i.e. the line's endpoints). In (Clementini and Di Felice 1997, Reis et al. 2006), two categories of lines with vague shapes are generally distinguished: *lines with broad boundary* and *completely broad lines*. Reis et *al.* (Reis et *al.* 2006) distinguish 77 topological relations between lines with broad boundary and 5 between completely broad ones. They apply the 9-Intersection model (Egenhofer and Herring 1990) on lines with vague shapes in order to identify their topological relations. Figure 4.1 shows two examples of topological relations between two lines with vague shapes according to (Hazarika and Cohn 2001). In Clementini (2002), Clementini (2005), Clementini explained that the line's interior can be also *broad* (or *vague*) and therefore it is important to distinguish between the notions of *broad boundary* and *broad interior*. This second approach is more expressive than Clementini and Di Felice (1997), Reis et *al.* (2006) model, because it allows to distinguish the case where only the line's interior is broad and not the boundary. By using the 9-Intersection model, Clementini (2005) distinguishes 146 topological relations between two lines with vague shapes. He considers these lines as complex geometries composed by two-dimensional parts (*for broad parts of the line*) and one-dimensional parts (*for certain parts*). Therefore, the line's interior corresponds to the union of interiors of *two-dimensional* and *one-dimensional* parts (line's boundary is the union of boundary of one-dimensional and two-dimensional parts). Clementini (2005) distinguishes 146 topological relations without any labelling or clustering process. This approach has two main limitations. First, the participation of each one of *two-dimensional* (uncertain parts of the line) and *one-dimensional parts* (certain parts of the lines) of lines in the topological relation is not described. In other words, the lines are defined as complex shapes without a formal distinction between their certain and uncertain parts. Second, this approach does not allow the description of partial shape vagueness and the 146 topological relations are not labelled.





$$\begin{array}{c|ccc|} & IL_2 & BL_2 & E\,L_2 \\ \hline IL_1 & 1 & 0 & 1 \\ B\,L_1 & 0 & 0 & 1 \\ E\,L_1 & 1 & 1 & 1 \\ \hline \end{array} \qquad \begin{array}{c|ccc|} & IL_4 & BL_4 & E\,L_4 \\ \hline IL_3 & 0 & 1 & 1 \\ B\,L_3 & 0 & 0 & 1 \\ E\,L_3 & 1 & 1 & 1 \\ \hline \end{array}$$

Figure 4.1 Identification of topological relations between lines with vague shapes in (Reis *et al.* 2006) (with *IL*, *BL* and *EL* refer respectively to the interior, boundary and exterior of the lines involved)

According to Clementini (2005), we agree about the importance of making the difference between the shape vagueness of an interior and that of a boundary. However, existing approches dealing with lines with vague shapes do not cover the cases where the boundary and/or interior of the line is partially vague. For example, figure 4.2 shows the trajectory of an historic explorer where the final destination is ill-known (i.e., only one of the endpoints is broad). The final destination is presented by a broad point which covers the set of the points, which can be the destination of the explorer. In the same way, only a part of the interior can be broad for an *aircraft* which traversed a turbulence area and that has not be detected by radars during this time period. In this paper, we aim to stress these different types and levels of shape vagueness in a new classification of lines with vague shapes. After that, an exact model is proposed in order to formally represent the lines with vague shapes. This formalization allows to overcome the limitations of existing works in terms of identification of the topological properties and relations between this type of lines.

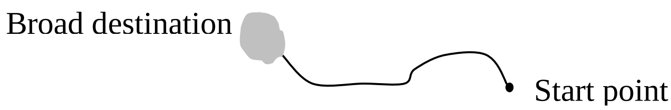


Figure 4.2 An example of a trajectory with vague shape of an historic explorer

4.5 QMM$_{Def}$ model for lines with vague shapes

4.5.1 Evaluation of shape vagueness for linear geometries

A simple crisp line is a one-dimensional object type made up of an interior and a disconnected boundary (i.e. two endpoints). The endpoints represent the boundary of a crisp line, whereas the interior is the set of points connecting them. Shape vagueness can characterize the interior or the boundary of a given line. Consequently, the line's boundary can be *partially* or *completely broad* while the interior remains well-defined; we then speak about lines with *broad boundaries*. In the same way, the interior can be *partially* or *completely broad* while the endpoints are well-defined; we speak about lines with *partially* and *completely broad interior*, respectively (figure 4.4). The *extreme* case of shape vagueness for lines arises when all of the line's topological invariants (i.e. *the interior* and *the boundary*) are *broad* (figure 4.4). Thus, a *completely broad line* arises when it is not possible to sharply distinguish the line from its neighborhood. It is also possible to have a line with completely broad line with broad boundary where there is a vague indication about the endpoints (see examples in lower-right cell of figure 4.4). In our categorization, we also consider a *completely crisp line* as a particular case of lines with vague shapes, for which both the interior and endpoints are well-defined. According to Clementini (2005), shape vagueness of a line *interior* is always present even only endpoints are broad. In other words, a broad endpoint implies that there is a part of space where each point can be: *the endpoint*, in *interior* or in *exterior* of the line. Figure 4.4 presents our general categorization of lines with vague shapes. A line with a vague shape can correspond to one or a combination of three basic object types: *lines with broad boundary*, *lines with broad interior* or *completely broad lines*. In figure 4.3, the specification "*overlapping*" means that different types of shape vagueness can be combined in a same line at the same time. For example, it can have a broad boundary and a broad interior at the same time.

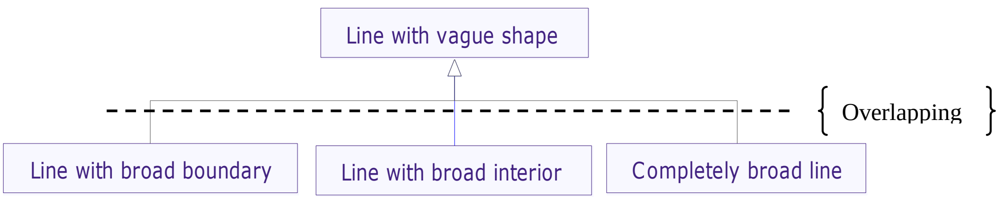


Figure 4.3 Categorization of lines with vague shapes

The different levels of shape vagueness for lines can be combined as presented in figure 4.4. We use one pronoun and four adverbs to underline these levels: (1) *none* (for crisp lines), (2) *weakly*, (3) *fairly*, (4) *strongly*, and (5) *completely*. The term "*weakly*" indicates that one of the topological invariants is *partially broad*. The term "*fairly*" reflects either a complete shape vagueness of one of topological invariants or the case where the interior and boundary are *partially broad* at the same time. The term "*strongly*" specifies complete shape vagueness for one of the topological invariants and partial shape vagueness for the second one. Finally, the term "*completely*" is used to express total shape vagueness of the line's components. Figure 4.4 shows a symmetrical matrix, in which the shape vagueness increases from "*none*" in the upper-left cell to "*completely*" in the lower-right cell through a progression including "*weakly*", "*fairly,*" and "*strongly*".

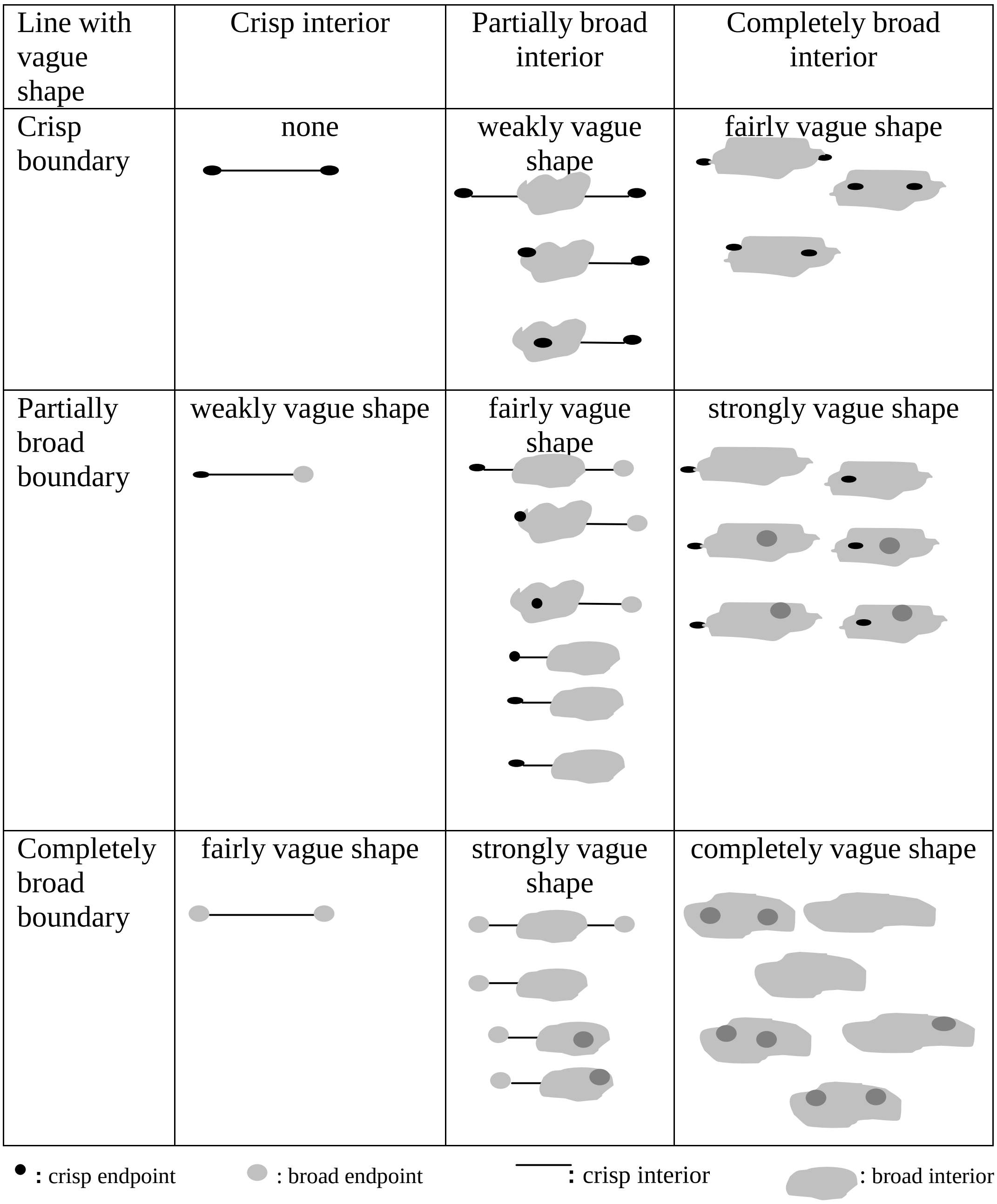


Figure 4.4 Lines with vague shapes

- Example of a *line with a weakly vague shape* (e.g., only one broad endpoint) :

The Bermuda triangle is a region in the Atlantic Ocean where some aircrafts and surface vessels have disappeared. Fight 19 is the designation of five American fighters which disappeared in this triangle on December 9, 1945. The five fighters left Naval Air Station of Lauderdale for a patrol. Their plan is to fly over the south east coast before landing in Florida. However, communication was interrupted when they enter into Bermuda Triangle. Then, only the start point (i.e. Naval Air Station of Lauderdale) and a part of the trajectory's interior are well-known before the communication interruption. The final endpoint is broad because the trajectory can have any shape inside the triangle. This situation can be modeled through a *line with weakly vague shape*.

- Example of a *line with a weakly vague shape* with two crisp endpoints :

We suppose that an aircraft disappeared for some time from radar screens because it traversed a turbulence area. After that, the communication returns to normal and the engine arrives at its destination. In this case, the aircraft trajectory is composed of two crisp endpoints. However, the interior is *partially broad* because the trajectory can take any unpredictable shape inside the turbulence zone. The trajectory of the aircraft can also be represented as a *line with a weakly vague shape*.

This approach is called the QMM_{Def} model, because different levels of shape vagueness can be distinguished by using a set of adverbs (i.e., a qualitative approach). Furthermore, the level of shape vagueness of a given line is deduced from the difference between its minimal extent and its maximal extent. Hereafter, we present the formal definition of a line with a vague shape in the QMM_{Def} model.

4.5.2 Definition of lines with vague shapes

In the QMM_{Def} model, a line with a vague shape is typically composed of two-dimensional parts that correspond to the vague parts of the line and one-dimensional parts that refer to the crisp parts of the line. We define *the maximal extent* of a line with a vague shape as a crisp complex geometry resulting from the union of the one-dimensional and two-dimensional parts. The interior of *maximal extent* corresponds to the union of interiors of one-dimensional parts and those of two-dimensional parts. In the same way, the boundary of the maximal extent is the union of boundaries of one-dimensional parts and those of two-dimensional parts. The maximal extent cannot be empty.

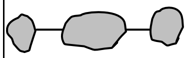
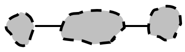
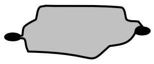
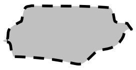
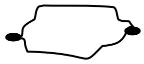
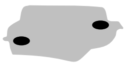
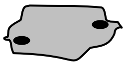
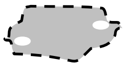
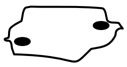
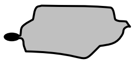
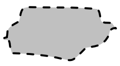
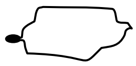
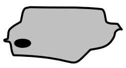
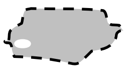
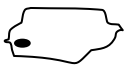
The *minimal extent* corresponds only to the crisp parts of the line involved (i.e., one-dimensional parts and well-defined endpoints). The minimal extent is also a crisp geometry and it is a subset of the maximal extent. It can be empty if the line is completely broad. The minimal and maximal extents are not mutually exclusive; i.e. $L_{min} \subseteq L_{max}$.

A line with a vague shape geometrically (but not semantically) refers to the maximal extent L_{max}. L_{min} and L_{max} are crisp geometries. L_{max} can include two-dimensional parts as well as one-dimensional parts. However, L_{min} includes only one-dimensional parts and well-defined endpoints of the line. The interpretation of shape vagueness of each part of the maximal extent L_{max} is made with regards to the related object represented by the line with a vague shape. Then, L_{max} is semantically different from the line itself; i.e. L_{max} cannot have a definition and a semantic independently of the line involved. The notion of maximal extent is distinguished from the minimal extent in order to distinguish the crisp parts of the line from the broad ones.

The notions of broad boundaries and broad interiors are proper to the line with a vague shape. For a line with a broad boundary, each point inside the broad boundary may be an endpoint, inside the interior or outside the line. The latter property proves that a point of the broad boundary cannot be outside the broad interior. Then, the concept of broad interior includes that of the broad bounadry. A broad interior is always present, even if the shape vagueness concerns only the endpoints (i.e. broad interior and broad boundary are not mutually exclusive). In other words, a point of the broad boundary is also a point of the broad interior at the same time.

For the maximal extent as well as for the minimal extent of a line with a vague shape, the interior can b e disconnected. The boundary can be also disconnected. Figure 4.5 shows different cases of decomposition of topological invariants composing extents of lines with vague shapes. We should denote that these different representations of lines with vague shapes correspond to a set of pictograms. In other words, these representations are not based on a mathematical model that allows to consider the error component of spatial data as in (Chrisman 1991). In Figure 4.5, the semantic difference between a line with a vague shape and its maximal extent is stressed by drawing linear boundaries for broad parts of the maximal extent. Such boundaries show that the maximal extent is a crisp complex geometry where we can distinguish the interiors and boundaries of its subparts as presented in the next figure.

Lines with vague shapes	Extents		Topological invariants	
	Minimal extent		Interior	
			Boundary	
	Maximal extent		Interior	
			Boundary	
	Minimal extent		Interior	
			Boundary	
	Maximal extent		Interior	
			Boundary	
	Minimal extent		Interior	
			Boundary	∅
	Maximal extent		Interior	
			Boundary	
	Minimal extent		Interior	
			Boundary	
	Maximal extent		Interior	
			Boundary	
	Minimal extent		Interior	
			Boundary	
	Maximal extent		Interior	
			Boundary	

	Minimal extent	— —	Interior	— —
			Boundary	∅
	Maximal extent		Interior	
			Boundary	
	Minimal extent		Interior	∅
			Boundary	
	Maximal extent		Interior	
			Boundary	
	Minimal extent		Interior	∅
			Boundary	
	Maximal extent		Interior	
			Boundary	
	Minimal extent		Interior	∅
			Boundary	
	Maximal extent		Interior	
			Boundary	
	Minimal extent		Interior	∅
			Boundary	
	Maximal extent		Interior	
			Boundary	

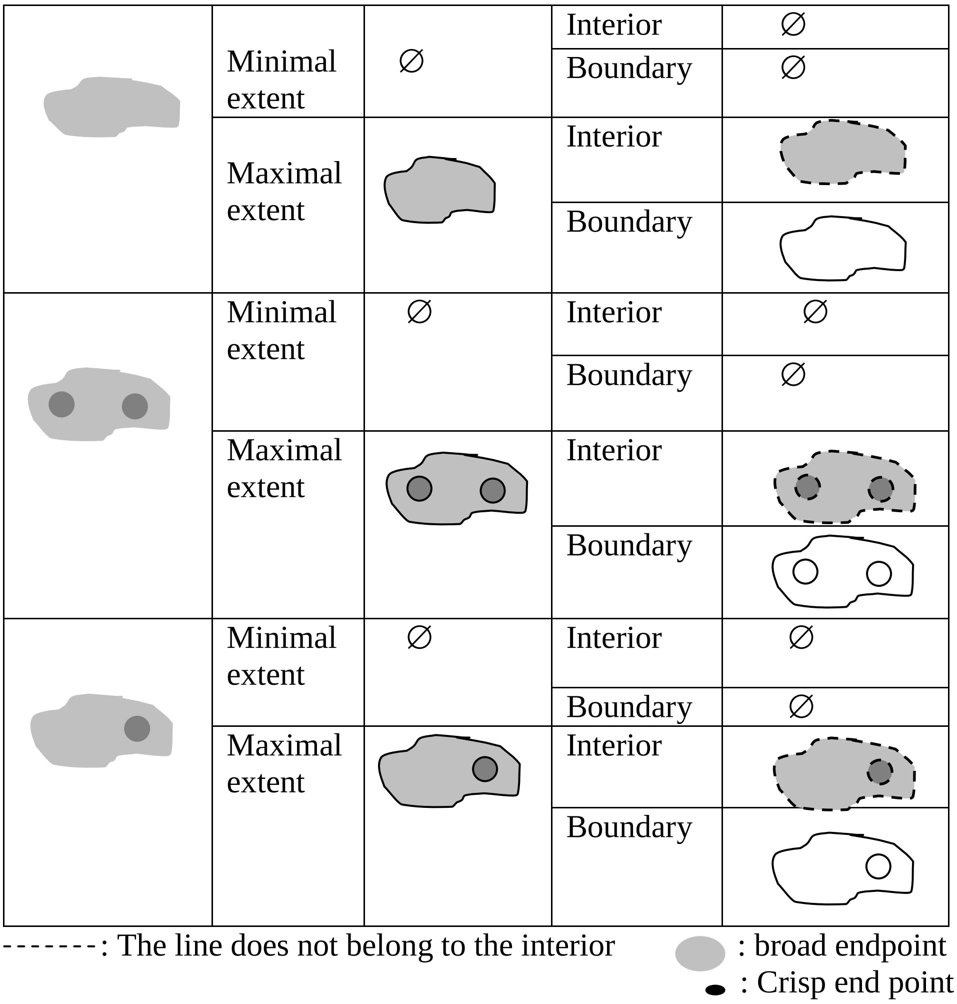


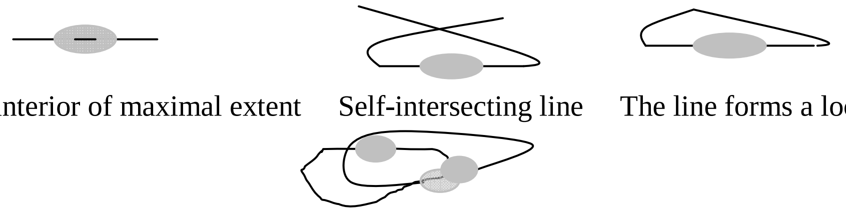
Figure 4.5 Topological invariants according to the line shape vagueness

More formally, a line $\tilde{L}$ with vague shape is composed by a maximal extent $\tilde{L}_{\max}$ and a minimal extent $\tilde{L}_{\min}$. The minimal extent corresponds to the one-dimensional parts and well-defined endpoints of the line. The maximal extent refers to the spatial extent of the line when the shape vagueness is considered. The maximal extent includes the minimal extent and the difference between them corresponds to the shape vagueness of the line. In our approach, we focus on the definition of the topological invariants for the maximal extent $\tilde{L}_{\max}$ and the minimal one $\tilde{L}_{\min}$. For each one, we distinguish an interior and a boundary that can be empty according to the configuration of the line (figure 4.5). From a point-set topology view point, a simple line with a vague shape should verify the following conditions:

1- Each one-dimensional part of the simple line with a vague shape is connected.
2- Each one-dimensional part of the simple line with a vague shape is not self-intersecting.
3- Each one-dimensional part of the simple line with a vague shape does not form a loop.

4- If the endpoints are broad, they do not overlap with each other.

The three first conditions are those defined for a crisp line in the general point-set topology. Then, we apply these conditions to each linear part of the line with a vague shape. The last condition is defined to eliminate any risk of a self-intersection or loop configurations. Figure 4.6 shows some cases of lines that are invalid according to our model.



Non-regular interior of maximal extent   Self-intersecting line   The line forms a loop

The endpoints can be identical

Figure 4.6 Examples of invalid lines

In the next section, we propose a qualitative approach to identify topological relations of between lines with vague shapes. This approach is called the *Qualitative Min-Max* model for *Topological Relations* (QMM$_{TR}$ for short) between lines with vague shapes and it is based on the QMM$_{Def}$ model presented above.

4.6 QMM Topological Relationships between lines with vague shapes

4.6.1 Extending of CBM method

In general, two models have been used for specifying topological relations between lines with vague shapes: the 9-Intersection model (Egenhofer and Herring 1990) and the CBM method (Clementini 1995). In the 9-Intersection model, topological relations between two spatial objects are defined in terms of nine intersections between their topological invariants (interiors, boundaries and exteriors). This approach has been extended to simple regions with broad boundaries in (e.g. Clementini and Di Felice 1997, Tang 2004, Bejaoui et al. 2008) as well as for lines with vague shapes (e.g. Clementini 2005, Reis et *al.* 2006). In the case of lines, the 9-Intersection model generally distinguishes a high number of topological relations either for crisp lines or for lines with vague shapes (Clementini 2005). In absence of any clustering method, the 9-Intersection model becomes useless because users cannot intuitively distinguish all of possible topological relations between lines with vague shapes. For example, 33 relations are possible between two simple crisp lines and 77 between two lines with broad boundary (Reis et *al.* 2006). In (Clementini 2005), 146 topological relations are distinguished computationally by using 9-Intersection matrices.

However, the CBM method (Clementini and Di Felice 1995) proposes five high-level operators (*touch*, *in*, *cross*, *overlap* and *disjoint*) in addition to the interior and boundary operators. Clementini and Di Felice (1995) proved that this approach is more expressive than the 9-Intersection model. Furthermore, each relationship identified by the 9-Intersection model can be classified into one of the five clusters associated to the five high-level operators of CBM. The main advantages of this approach are its expressivity and simplicity in identifying topological relations. CBM method was extended for regions and lines with broad boundaries (Clementini 2002). In this paper, we adapt the CBM method to our model of lines with vague shapes. More specifically, we propose an additional operator called *ext_min* that we use to extract the minimal extent of the line. This operator allows to underline the participation of one-dimensional parts in a topological relation. Furthermore, new topological operators are suggested in order to improve the expressivity of the approach regarding the specification of topological relationships between lines with vague shapes. *cross_min* and *overlap_min* are respective specializations of *Overlap* and *Cross*. These new operators can be applied between minimal extents of lines involved. The extension of CBM method provides the set of topological operators of QMM_{TR} to identify the topological relations between lines with vague shapes (cf., Section 4.6). In the next definitions, the formal definitions of basic and new operators are presented and some examples are given in figure 4.7. We assume that O_1 and O_2 are two lines with vague shapes:

- Definition 1: *touch*

  $(O_1, touch, O_2) \Leftrightarrow (O_1^\circ \cap O_2^\circ = \emptyset) \cap (O_1 \cap O_2 \neq \emptyset)$

- Definition 2: *in*

  $(O_1, in, O_2) \Leftrightarrow (O_1 \cap O_2 = O_1)$

- Definition 3: *Disjoint*

  $(O_1, in, O_2) \Leftrightarrow (O_1 \cap O_2 = \emptyset)$

- Definition 4: *cross_min* (arises between $ext\_\min(O_1)$ and $ext\_\min(O_2)$ *where* $dim(ext\_\min(O_1))=1$ and $dim(ext\_\min(O_2))=1$)

  $(O_1, cross\_\min, O_2) \Leftrightarrow \quad ((ext\_\min(O_1) \cap ext\_\min(O_2) \neq ext\_\min(O_1))$

  $\cap (ext\_\min(O_1) \cap ext\_\min(O_2) \neq ext\_\min(O_1)))$

$\cap (\dim(O_1^\circ \cap O_2^\circ) = 0)$

- Definition 5: *cross* (arises between a line with a vague shape and the minimal extent of another one, example O_1 and *ext_*min*(O₂) where dim(ext_*min*(O₂))=1*)

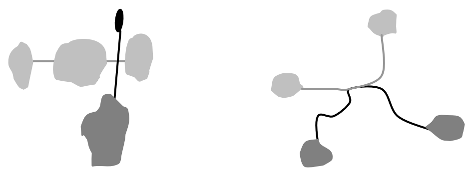
$$(O_1, cross, O_2) \Leftrightarrow (O_1, cross\_\min, O_2)$$
$$\cup (((O_1 \cap ext\_\min(O_2) \neq ext\_\min(O_2))$$
$$\cap (ext\_\min(O_1) \cap O_2 \neq ext\_\min(O_1))$$
$$\cap (\dim(O_1^\circ \cap O_2^\circ) = 1))$$

- Definition 6: *overlap_min* arises between *ext_*min(O_1) and *ext_*min(O_2) *where dim(ext_*min(O_1)*)=1* and *dim(ext_*min(O_2)*)=1*)

$$(O_1, overlap\_\min, O_2) \Leftrightarrow (\dim(O_1^\circ \cap O_2^\circ) = 1)$$
$$\cap (ext\_\min(O_1) \cap ext\_\min(O_2) \neq ext\_\min(O_1))$$
$$\cap (ext\_\min(O_1) \cap ext\_\min(O_2) \neq ext\_\min(O_2))$$

- Definition 7: *overlap*

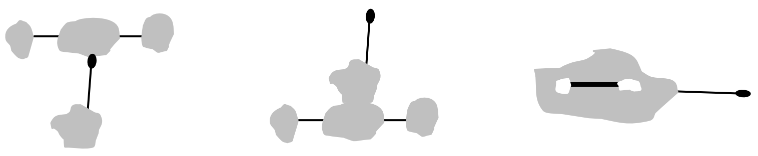
$$(O_1, overlap, O_2) \Leftrightarrow (O_1, overlap\_\min, O_2)$$
$$\cup ((O_1 \cap O_2 \neq O_2)$$
$$\cap (O_1 \cap O_2 \neq O_1)$$
$$\cap (\dim(O_1^\circ \cap O_2^\circ) = 2))$$



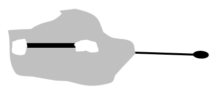
(a) *cross_min relation* (b) *overlap_min relation*

Figure 4.7 Examples of *cross_min* and *overlap_min* relations

Additionally, we look for highlighting the dimension of an intersection resulting from a *touch* relation. In essence, *0-dim_touch* and *1-dim_touch* are specializations of the *touch* operator; they are used to specify whether the dimension of an intersection in a touch relation is a point or a line. In the same way, the CBM method does not explicitly distinguish the *Covered by* relation as in the 9-Intersection model. In this work, we consider it as a specialization of the *in* relation; we call this relation *in_touch*(*b*) (*b* is an operator to extract a line's boundary), because it requires that the boundary of the inner object touches that of outer one. *in_disjoint*(*b*) is another specialization of the *in* relation; it means that boundaries of the inner object and the outer one are disjoint. Figure 4.8 shows examples of these four relations.



(a) *0-dim_touch relation* (b) *1-dim_touch relation* (c) *in-disjoint(b) relation*



(d) *in-touch(b) relation*

Figure 4.8 Examples of (a) 0-dim_touch relation, (b) 1-dim_touch relation, (c) in-disjoint(b) relation and (d) in-touch(b) relation

The relations *0-dim_touch, 1-dim_touch, in_touch(b), in_disjoint(b)* are defined as follows:

- Definition 8: *0-dim_touch*

$$(O_1, 0-\dim\_touch, O_2) \Leftrightarrow (O_1^\circ \cap O_2^\circ = \varnothing) \cap (O_1 \cap O_2 \neq \varnothing)$$
$$\cap (\dim(O_1 \cap O_2) = 0)$$

- Definition 9: *1-dim_touch*

$$(O_1, 1-\dim\_touch, O_2) \Leftrightarrow (O_1^\circ \cap O_2^\circ = \varnothing) \cap (O_1 \cap O_2 \neq \varnothing)$$
$$\cap (\dim(O_1 \cap O_2) = 1)$$

- Definition 10: *in_touch(b)*

$$(O_1, in\_touch(b), O_2) \Leftrightarrow (O_1 \cap O_2 = O_1)$$
$$\cap (b(O_1) \cap b(O_2) \neq \varnothing)$$

- Definition 11: *in_disjoint(b)*

$$(O_1, in\_disjoint(b), O_2) \Leftrightarrow (O_1 \cap O_2 = O_1) \cap (b(O_1) \cap b(O_2) = \varnothing)$$

Figure 4.9 shows generalization/specialization relations between the topological operators in the QMM$_{TR}$ model applied for lines with vague shapes:

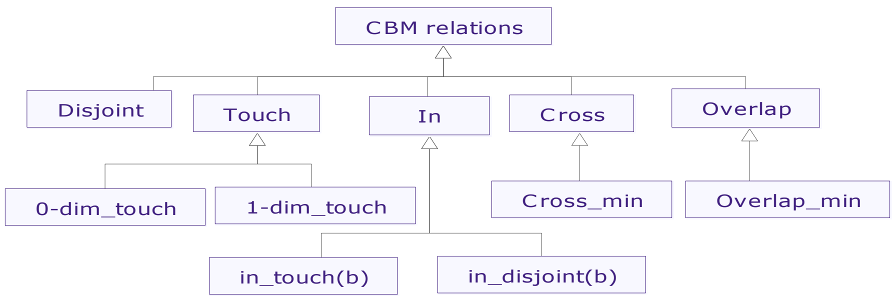


Figure 4.9 Generalization/Specialization links between relations of the QMM$_{TR}$ model

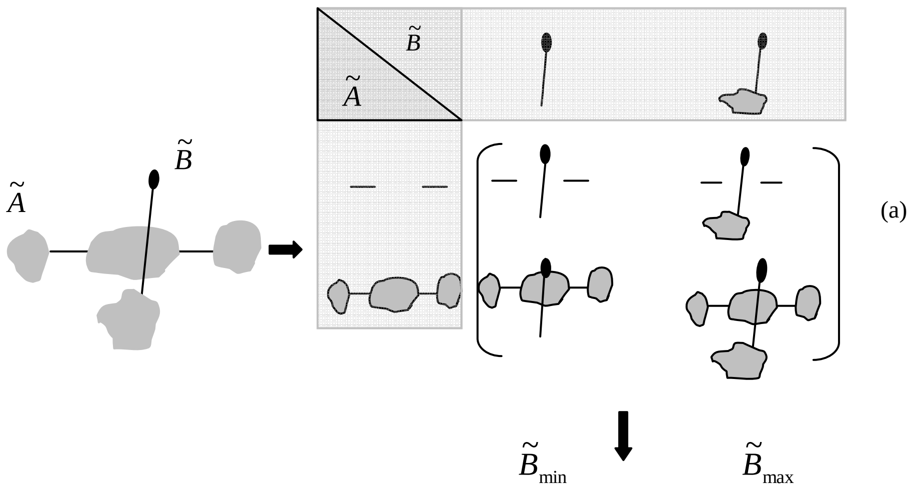
The topological operators of each level of the QMM_{TR} model are mutually exclusive. This property is verified for the first level, which contains the following relations: *Disjoint*, *Touch*, *In*, *Cross* and *Overlap*. In the same way, *Disjoint*, *0-dim_touch*, *1-dim_touch*, *in_touch(b)*, *in_disjoint(b)*, *Cross_min* and *Overlap_min* are also mutually exclusive. In the next section, we explain how we use these operators to identify topological relations between lines with vague shapes.

4.6.2 Principles of identification of topological relations in the QMM_{TR} model

We interpret the maximal extents of lines with vague shapes as composite geometries. It is composed by one-dimensional parts and two-dimensional ones. The minimal extent is a subset of the maximal one (i.e., it corresponds to one-dimensional parts and crisp points of the line). In fact, our methodology consists in identifying four specific topological relations between minimal and maximal extents of lines with vague shapes involved. For this purpose, we define a 4-Intersection matrix containing the following four topological sub-relations: $R_1(\tilde{A}_{\min}, \tilde{B}_{\min})$, $R_2(\tilde{A}_{\min}, \tilde{B}_{\max})$, $R_3(\tilde{A}_{\max}, \tilde{B}_{\min})$, and $R_4(\tilde{A}_{\max}, \tilde{B}_{\max})$ (see example in figure 4.10) (with $\tilde{A}$ and $\tilde{B}$ two lines with vague shapes). According to this idea, we should remind that the the structure of 4-Intersection matrix has been used by (Eegnhofer 1989) to identify topological relationships between crisp regions. In the present work, we propose a model based on the use of 4-Intersection matrices in the specific context of lines with vague shapes. These matrices are just containers; i.e. a formal representation of the topological relationships between lines with vague shapes involved. The method used to fill the matrices' cells is different to that used in (Egenhofer 1989). In the present approach, the basic idea consists in using the extension of CBM method (i.e., the topological operators presented above in the QMM_{TR}) to fill the four cells of the matrix. Then, the 4-Intersection matrix corresponds to the following representation:

$$\begin{array}{c} \\ \tilde{A}_{\min} \\ \tilde{A}_{\max} \end{array} \begin{array}{cc} \tilde{B}_{\min} & \tilde{B}_{\max} \\ \multicolumn{2}{c}{\begin{bmatrix} R_1(\tilde{A}_{\min}, \tilde{B}_{\min}), & R_2(\tilde{A}_{\min}, \tilde{B}_{\max}) \\ R_3(\tilde{A}_{\max}, \tilde{B}_{\min}), & R_4(\tilde{A}_{\max}, \tilde{B}_{\max}) \end{bmatrix}} \end{array}$$

Figure 4.10 shows the content of the matrix that describes a topological relation between two lines with vague shapes $\tilde{A}$ and $\tilde{B}$.



$$\begin{array}{c} \\ \tilde{A}_{\min} \\ \tilde{A}_{\max} \end{array} \begin{array}{cc} \tilde{B}_{\min} & \tilde{B}_{\max} \\ \multicolumn{2}{c}{\left[\begin{array}{cc} \textit{Disjoint}\,(\tilde{A}_{\min}, \tilde{B}_{\min}), & \textit{Disjoint}\,(\tilde{A}_{\min}, \tilde{B}_{\max}) \\ \textit{Cross}\,(\tilde{A}_{\max}, \tilde{B}_{\min}), & \textit{Cross}\,(\tilde{A}_{\max}, \tilde{B}_{\max}) \end{array}\right]} \end{array} \quad \text{(b)}$$

Figure 4.10 Description of a topological relation between two lines with vague shapes: (a) visual content of the matrix, (b) formal identification of the relations between the minimal and maximal extents of the objects involved

The content of the matrix corresponds to the four topological sub-relations between respective minimal and maximal extents of lines involved. Since the maximal extents geometrically (but not semantically) refer to the lines, we use the topological sub-relation between them $R_4(\tilde{A}_{\max}, \tilde{B}_{\max})$ (value of the down-right cell) in order to label the global topological relation. For example, if $R_4(\tilde{A}_{\max}, \tilde{B}_{\max})$ is *Cross*, we consider that one of the lines with vague shapes globally *Crosses* the other. If $R_4(\tilde{A}_{\max}, \tilde{B}_{\max})$ is *Contains*, we consider that the global topological relation is *Contains*. In the example of figure 4.10, $\tilde{A}$ globally *Crosses* $\tilde{B}$.

4.7 Clustering of topological relations between lines with vague shapes

4.7.1 Principles

In this work, topological relationships between lines with vague shapes are specified through the topological operators defined in the QMM$_{TR}$ model (cf., section 4.6) that we apply between minimal and maximal extents of involved lines. Eleven topological operators can be used to specify these sub-relations. These operators allow to describe any topological relation between two lines with vague shapes. However, it is very difficult to enumerate all of possible relations, because the shapes of such composite objects are unpredictable. It is also not realistic to find a name for each one of possible relations, and therefore the user will have difficulty to choose the appropriate operator in order to express a spatial query or a topological integrity constraint. For this purpose, the clustering of topological relations into larger groups may be a pertinent alternative followed by previous works such as (e.g. Clementini and Di Felice 1997, Mark and Egenhofer 1994).

In this paper, we use the content of the proposed 4-Intersection matrix in order to classify the topological relations. Five basic clusters are distinguished: *DISJOINT*, *IN*, *CROSS*, *OVERLAP* and *TOUCH*. Each cluster contains all of the topological relations for which at least one of the four sub-relations has the same name as the cluster. A topological relation becomes possible if it appears at least once in the matrix. This possibility increases according to the number of similar sub-relations. For example, a *Cross* topological relation in which *Cross* $(\tilde{A}_{\max}, \tilde{B}_{\max})$ and *Cross* $(\tilde{A}_{\min}, \tilde{B}_{\min})$ is stronger than another where only *Cross* $(\tilde{A}_{\max}, \tilde{B}_{\max})$. In order to distinguish these different levels of a relation's membership, we use four adverbs to evaluate the vagueness of a topological relation: (1) *completely*, (2) *strongly*, (3) *fairly,* and (4) *weakly.* A topological relation belongs to one cluster *completely* when the four sub-relations are identical. It belongs to one cluster *strongly* when only three sub-relations have the same name as the cluster. The level termed *fairly* contains all relations for which two sub-relations have the same name as the cluster. Finally, the level called *weakly* contains the relations for which only one sub-relation has the same name as the cluster. For example, figure 4.11 shows a topological relation that belongs to the following clusters: *DISJOINT*, *TOUCH,* and *IN*. Nevertheless, it belongs to the *IN cluster* more strongly than to the *DISJOINT* and *COVERS clusters.* By using our adverbial approach, we can conclude that the topological relation is *fairly IN*, *weakly DISJOINT*, and *weakly TOUCH*.

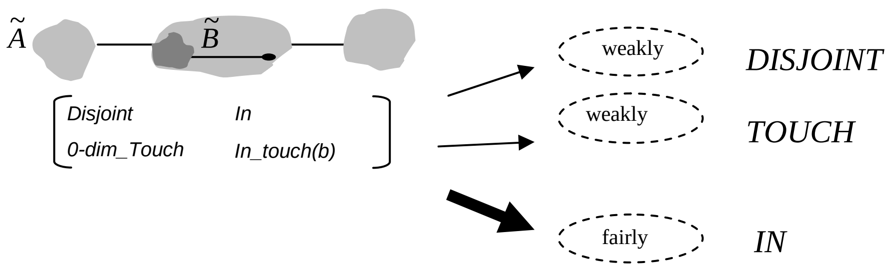


Figure 4.11 Example of clustering of a topological relation

Figure 4.12 presents examples of relations that belong to different levels of *CONTAINS* and *DISJOINT* clusters, respectively, according to the contents of their correspondent matrices.

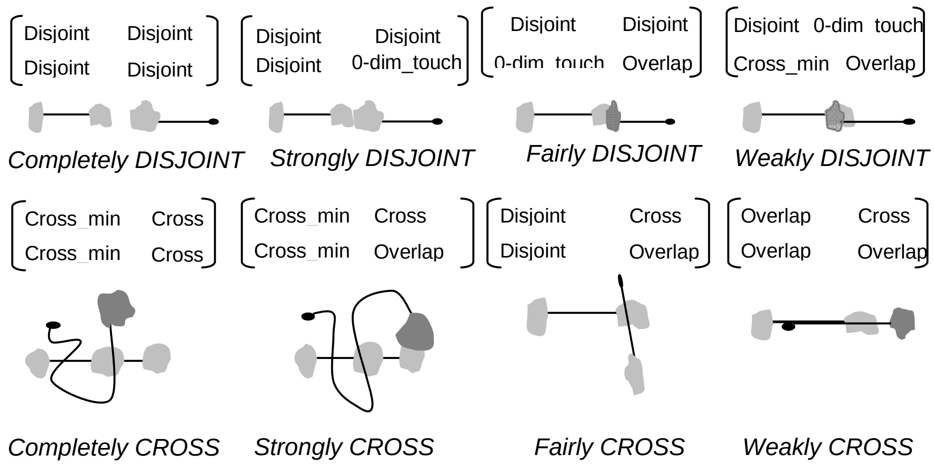


Figure 4.12 Evaluation of topological relationship strength

4.7.2 Overlapping clusters

The main result of this clustering process is a hierarchical classification of the topological relations (figure 4.13). The top level is made up of five basic clusters (*DISJOINT*, *TOUCH*, *IN*, *CROSS*, *OVERLAP*) that each contains typically four levels: *completely*, *strongly*, *fairly*, and *weakly*. The resulting 32 sub-clusters overlap each other because a topological relation can belongs to different levels of 1, 2, 3, or 4 clusters at the same time. Figure 4.13 shows the structure of this hierarchical classification. The bottom level includes all of possible cases that can occur between two lines with vague shapes.

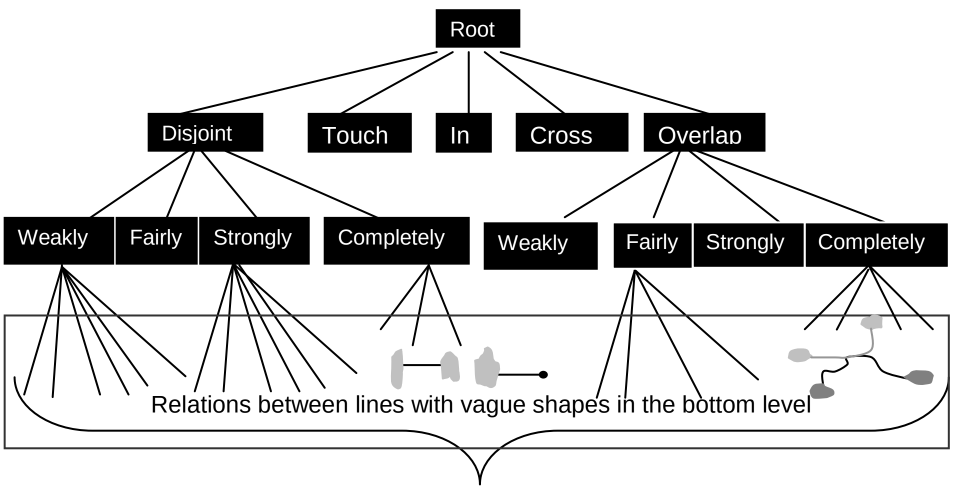


Figure 4.13 A hierarchical classification of the topological relations between lines with vague shapes

4.8 Specification of topological integrity constraints and spatial queries for lines with vague shapes

The simplicity and expressivity of the CBM method (Clementini and Di Felice 1995, Clementini 2002, Clementini 2005) can be inherited by the QMM_{TR} especially with a qualitative classification of topological relations between lines with vague shapes. Then, this approach can provide necessary conceptual tools in order to formally express topological integrity constraints for lines with vague shapes. A topological integrity constraint is a rule that insure that a topological property of an object or a topological relation is not violated. These constraints are used to insure the consistency of a spatial database (Frank 2001). For example, we assume that a spatial database stores the geometries of some protected animals' trajectories and that shape vagueness is considered in this database. A topological integrity constraint can be defined in order to say that '*Different trajectories of one species in one season should not be* ***completely*** or ***strongly Disjoint'***. This constraint can be formally expressed by integrating new spatial operators (e.g., *completely Disjoint*, *weakly Covers*, etc.) in a formal constraint language such as the Object Constraint Language (OCL) (Pinet et al. 2007). The database storing the *trajectories* is consistent only whether the topological relations between the different trajectories do not belong to the following subclusters: ***completely*** or ***strongly Disjoint***. This constraint can be expressed through Spatial OCL as follows:

Example 1 : **Context** Trajectory **inv**:

Trajectory.allInstances → **forAll** (a, b| a<>b implies not

(***strongly DISJOINT***(a,b) **or** ***completely DISJOINT***(a,b)));

In the same way, the QMM model can be integrated in a spatial database system. Indeed, the SQL language can be extended in order to express spatial queries involving lines with vague shapes based on the qualitative information given by the user regarding their topological relation. In this section, we suppose that we integrated our spatial model in a relational engine in order to give an example of its possible use in spatial queries involving lines with vague shapes. In the next query example, the user would select the animals' trajectories that *weakly Overlap* or *weakly Meet* each other. According to our approach, this query can be expressed as follows:

Example 2: **Select** *A.geometry, B. geometry* **From** *Trajectories A, Trajectories B*
Where A.trajectories_id<> B.trajectories_id AND
vague_Relate (*A.geometry, B.geometry, weakly Meet*) OR
vague_Relate (*A.geometry, B.geometry, weakly Overlap*)

4.9 Conclusion

Shape vagueness has been thoughtfully studied for regions (notably in (Burrough and Frank 1996, Cohn and Gotts 1996, Dilo 2006, Erwig and Schneider 1997, Roy and Stell 2001, Tang 2004, Zhan and Lin 2003)). However, shape vagueness of lines has been generally considered more complicated to model than regions. Some approaches (Clementini and Di Felice 1997, Clementini 2002, Clementini 2005, Reis et *al.* 2006) was interested in modeling lines with vague shapes and their topological relations. The main limitation of these approaches is that they do not make the distinction between different types and levels of shape vagueness of lines (i.e. *partial shape vagueness*, *complete shape vagueness*, *partial broad interior*, *and partial broad interior*, *etc*.). In this paper, we proposed a new geometric model called QMM model composed by two sub-models: (1) the QMM_{Def} model and QMM_{TR} model. The QMM_{Def} model proposes an expressive taxonomy of lines with vague shapes and their formal definitions. In the proposed taxonomy, we made the distinction between the shape vagueness of the interior of a given line from that arising in its boundary. The line interior can be *partially* or *completely* broad independently of the boundary, and vice versa. We identified four levels of shape vagueness for lines according to the *crispness*, *partial broadness and complete broadness* of the interior and/or boundary: (1) *weakly*, (2) *fairly*, (3) *strongly* and (4) *completely*. Generally, we defined a line with a vague shape as a minimal extent composed only by one-dimensional parts of the line and a maximal extent that additionally includes the two-dimensional or broad parts. Topological relations between lines with vague shapes are then identified through an extension of the CBM method (Clementini and Di Felice 1995) that we integrate into the QMM_{TR} model and apply for sub-relations between minimal and maximal extents of involved lines. After that, we proposed a 4-Intesersection matrix to describe these four sub-relations and classify topological relations between lines with vague shapes. A topological relation can belong with different strengths (i.e., *weakly*, *fairly*, *strongly*, and *completely*) to one or multiple of the following basic clusters: *DISJOINT*, *IN*, *CROSS*, *OVERLAP* and *TOUCH*. This adverbial approach can provide the basis of an extension of a constraint language to express topological integrity constraints involving lines with vague shapes. Finally, the main perspective of this work is to extend our model to the composite lines with vague shapes.

References

ALTMAN D. 1987, Fuzzy set theoretic approaches for handling imprecision in spatial analysis. *International Journal of Geographical Information Systems* 8: 271-289

BENNETT B. 2000, Application of superevaluation semantics to vaguely defined spatial concepts. In: D.R. Montello (Ed.) *Spatial Information Theory*. Foundations of geographic *Information Science International Conference COSIT 2001* Springer Berlin 2000: 108-123

BÉDARD Y. 1987, Uncertainties in Land Information Systems Databases. In *Proceedings of the Eighth International Symposium on Computer-Assisted Cartography* Baltimore Maryland (USA) 29 Mars - 3 Avril 1987 American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping: 175-184

BEJAOUI L. BÉDARD Y. Pinet F. and Schneider M. 2008, Qualified topological relations between spatial objects with vague shapes. *International Journal of Geographical Information Science. accepted.*

BROWN D.G. 1998, Classification and boundary vagueness in mapping presettlement forest types. *International Journal of Geographical Information Systems* 12: 105-129

BURROUGH P. A. AND FRANK A. U. 1996, *Geographic Objects with Indeterminate Boundaries* (London: Taylor & Francis).

CHENG T. MOLENAAR M. AND LIN H. 2001, Formalizing fuzzy objects from uncertain classification results. *International Journal Geographical Information Science* 15(1): 27-42.

CLEMENTINI E. 2002, A model for lines with broad boundary. In *Proceedings of the Ninth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems* (IPMU 2002), Annecy, France, 2002, pp. 579-1586.

CLEMENTINI E. 2005, A model for uncertain lines. *Journal of Visual Languages and Computing* 16(4): 271-288

CLEMENTINI E. AND DI FELICE P. 1995, A Comparison of Methods for Representing Topological Relationships. *Information Sciences* 3: 149-178

CLEMENTINI E. AND DI FELICE P. 1997, Approximate topological relations. *International Journal of Approximate Reasoning* 16: 173-204

COHN A.G. AND GOTTS N.M. 1996, The 'egg-yolk' representation of regions with indeterminate boundaries. In *Proceedings of the GISDATA Specialist Meeting on Spatial Objects with Undetermined Boundaries*, Burrough, P. & Frank, A. (Ed.) (Taylor & Francis): 171-187

CHRISMAN N.R. 1991, The error component in spatial data. In: Maguire, D.J., Goodchild, M.F and Rhind, D.W. (eds.) *Geographical Information Systems: Principles and Applications* Volume 1: 165-174.

COUCLELIS H. 1996, 'Towards an operational typology of geographic entities with illdefined boundaries', in *Geographic Objects with Indeterminate Boundaries*, eds., P. Burrough and A.M. Frank, GISDATA, pp. 45–55. Taylor & Francis, (1996).

DILO A. 2006, *Representation of and reasoning with vagueness in spatial information: A system for handling vague objects*. PhD thesis ITC Netherlands, 187p.

EGENHOFER M.J., 1989, A formal definition of binary topological relations. In *Poceedings of the third international conference on Foundations of Data Organisation and Algorithms (FODO)*, W.Litin and H.J.Scheck (Ed.) (NY: Springer-Verlag), Lecture notes in computer science, **367**, pp. 457-472.

Egenhofer M. and Herring J. 1990, *Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases*. Technical Report, Department of Surveying Engineering, University of Maine Orono ME.

Erwig M. and Schneider M. 1997, Vague regions In *Proceedings of the 5th International Symposium on Advances in Spatial Databases (SSD'97)* Lecture Notes in Computer Science 1262: 298-320.

Frank A U 2001 Tiers of ontology and consistency constraints in geographical information systems. *International Journal of Geographical Information Science* 15(7): 667-678

HAZARIKA S.M. AND COHN A.G. 2001, A taxonomy for spatial vagueness, an alternative egg-yolk interpretation. In *Proceedings of the COSIT 2001 The fifth international Conference On Spatial Information Theory* D.R. Montello (Ed.) (California: Springer-Verlag) Lecture Notes in Computer Science 2205: 92-107

HUNTER G.J. AND GOODCHILD M.F. 1996, A new model for handling vector data uncertainty in geographical information systems. *URISA Journal* 8: 51-57

MARK D. AND EGENHOFER M. 1994, Modeling Spatial Relations Between Lines and Regions: Combining Formal Mathematical Models and Human Subjects Testing. *Cartography and Geographical Information Systems* 21 (3): 195-212.

MORRIS A. 2003, A framework for modeling uncertainty in spatial databases. *Transactions in GIS* 7: 83-102

PAWLAK Z. 1994, Rough sets: present state and further prospects. In *Proceedings of the third International Workshop on Rough Set and Soft Computing* (RSSC ’94): 72–76

PFOSER D. TRYFONA N. AND JENSEN C.S. 2005, Indeterminacy and Spatiotemporal Data: Basic Definitions and Case Study. *GeoInformatica* 9(3): 211-236

PINET F. DUBOISSET M. AND SOULIGNAC V. 2007, Using UML and OCL to maintain the consistency of spatial data in environmental information systems. *Environmental modeling & software* 22(8): 1217-1220

REIS R. EGENHOFER M.J. AND MATOS J. 2006, Topological relations using two models of uncertainty for lines. In *Proceeding of the 7th international Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, 5 - 7 July, Lisbon, Portugal: 286-295

ROBINSON V.B. AND THONGS D. 1986, Fuzzy Set Theory Applied to the Mixed Pixel problem of Multi-spectral Land Cover Databases. *GIS in Government*, **2**, B.K. Opitz (Ed.) (Washington D.C: A.Deepak Publication)

ROBINSON V.A. 2003, A perspective on the fundamentals of fuzzy sets and their use in Geographic Information Systems, *Transactions in GIS* 7 (1): 3–30

ROY A.J. AND STELL J.G. 2001, Spatial relations between indeterminate regions. *International Journal of Approximate Reasoning* 27: 205—234

SCHNEIDER M. 2001, A design of topological predicates for complex crisp and fuzzy regions. In ER ’01: *Proceedings of the 20th International Conference on Conceptual Modeling* Springer-Verlag ISBN 3-540-42866-6: 103–116.

UBEDA T. AND EGENHOFER M. 1997, Topological Error Correcting in GIS. In *Proceedings of International Symposium on Large Spatial Databases Lecture Notes in Computer Science* 1262 Springer-Verlag: 283-297.

WINTER S. 2000, Uncertain Topological Relations between Imprecise Regions. *International Journal of Geographical Information Science* 14(5): 411-430

WORBOYS M.F. 1998, Imprecision in finite resolution spatial data. *GeoInformatica* **2**: 257-279

TANG XI. 2004, *Spatial object modeling in fuzzy topological spaces: with applications to land cover change.* PhD thesis University of Twente ISBN 90-6164-220-5

YAZICI A. ZHU Q. AND SUN N. 2001, Semantic data modeling of spatiotemporal database applications. *Int. J. Intell. Syst:* 881-904

ZADEH, L.A., 1965, Fuzzy sets. *Inform. Control*, vol. **8**, pp. 338–353.

ZHAN B.F. 1997, Topological relations between fuzzy regions. In *Proceedings of the 1997 ACM Symposium on Applied Computing*, (ACM Press) ISBN 0-89791-850-9: 192-196

ZHAN F.B. AND LIN H. 2003, Overlay of Two Simple Polygons with Indeterminate Boundaries. *Transactions in GIS* 7(1): 67-81

Chapter 5: Reducing the vagueness of topological relationships in spatial data integration

Lotfi BEJAOUI, François PINET, Michel SCHNEIDER, Yvan BÉDARD

International Journal of Geographical Information Science
(Submitted)

5.1 Résumé de l'article

L'intégration des bases de données spatiales peut être basée sur l'analyse de leur qualité interne. Cette analyse justifie la sélection d'une base de données *référence* contenant les *meilleures* (dans le sens de la qualité interne) géométries qui peuvent représenter les objets dans une base de données finale. Toutefois, cette approche n'est pas toujours possible, en particulier lorsque des éléments de qualité sont mal décrits au niveau des bases de données sources. Dans cet article, nous nous sommes intéressés à un cas particulier de l'intégration de bases de données spatiales visant à fusionner (1) des représentations géométriques hétérogènes stockées dans des sources différentes pour lesquelles (2) la qualité interne est mal-décrite. Dans ce cas, une approche commune consiste à supposer que toutes les géométries sources d'un objet contribuent d'une façon égale dans sa géométrie finale. Par conséquent, un objet spatial peut avoir une géométrie finale de *forme vague (par exemple,*

régions ayant des frontières larges) lorsqu’il y a une différence entre l'union et l'intersection de ses géométries sources. Dans cet article, nous adressons le problème du *vague topologique* que nous définissons comme l’incertitude par rapport à la relation topologique appropriée entre les géométries finales. Ces relations topologiques sont généralement différentes de celles définies dans les bases de données sources car le vague de forme doit être pris en compte. L’objectif de cet article est de *réduire* le vague par rapport aux relations topologiques entre les géométries finales. Dans notre approche, nous énumérons les relations topologiques possibles et proposons différentes stratégies pour les vérifier.

5.2 Abstract

The integration of multiple spatial databases takes into account the analysis of their spatial data quality. This comparison leads to select or to generate the *best* geometries to be loaded in the final database. Such a process is a challenge when the elements of spatial data quality are poorly described in the data sources. In this case, a common approach consists of assuming that all the crisp source geometries of each object contribute, in an equal way, to produce the final geometric representation. Then, a spatial object may be represented through a *geometry with a vague shape (e.g. region with a broad boundary)* in the final database. The shape vagueness results from the difference between crisp source geometries. In addition, for a same pair of objects, the topological relationships between their final geometry cannot be deduced from those defined between their former crisp geometries in the original data sources. Therefore, we address the problem of *topological relationships vagueness*, i.e. the uncertainty about the appropriate topological relationships between the final geometries. This paper aims at *reducing* the topological relationships vagueness in a given final database. We analyze which topological relationships are possible, and propose different strategies to manage them.

5.3 Introduction

Spatial data integration is a complex problem that can be defined, addressed and resolved differently according to different needs. In (Shibasaki *et al.* 1994, Ziegler and Dittrich 2004), spatial data integration aims at combining data stored in different sources in order to produce a more complete final database with respect to the areas, epochs or themes to be covered. In (Uitermark *et al.* 2005), the integration of multiple spatial databases consists of establishing the relationships between corresponding instances in different spatial databases representing the same geographic space. It can be also used to (1) load a multi-representation spatial database (Laurini 1996, Megrin 1996), (2) reuse the data in another context (Breunig and Perkhoff 1992), (3) improve the completeness and non-redundancy of an existing database (Nyerges 1989), and so on. It is also possible to distinguish vertical integration (integrating spatial data describing different themes in the same location) from horizontal integration (integrating spatial data describing the same theme but in different locations) (Poulliot 2005). In the context of decision-support systems, spatial data integration is often a necessary process to load spatial data warehouses (Malinowski and Zimányi 2005, Bédard *et al.* 2007, Sboui *et al.* 2007). According to Franklin (1992), 80% of data have a spatial component. When a spatial data warehouse is modelled and implemented with a hypercube structure, this property of data is exploited in order to improve the data analysis by providing the geometric navigation in a spatial dimension. A spatial dimension includes different geometric levels which are organised in a hierarchy. The members of each level of analysis can involve geometries loaded from different source databases selected in an integration process. In this work, we deal with a special case of the vertical integration; i.e. where the same spatial objects are represented with heterogeneous redundant geometries measured at the same epoch but using different specifications for different data sources. Then, we assume that the final geometries resulted from the integration may be loaded into the spatial dimensions of a spatial data warehouse (with a hypercube structure) and provide vague shapes for the members of a hierarchy level.

The internal quality of a spatial database refers to the respect of the specifications defined by the data producer, and generally includes the following elements: (1) data *actuality*, (2) *geometric and thematic accuracies*, (4) *lineage (i.e. genealogy* of data), (5) *logical consistency* (i.e. *thematic*, *geometric*, *temporal*, *topological*, and *structural* coherencies of data; generally controlled with integrity constraints) and (6) *completeness* (Devillers and Jeansoulin 2005, Mostafavi *et al.* 2004). An internal quality analysis involves the comparison of these quality elements to the theoretical specifications or the nominal ground (David and Fasquel 1997). Multiple spatial databases can be integrated based on a comparison of their respective internal quality. According to Devogel (1997), the integration process requires the selection of a source database as a *reference*. The geometries of these sources are used to control the integration of the geometries coming from the other sources. Then, if a source geometry does not occur inside a matching area, it is not considered in the final geometry of the spatial object involved. The *reference* database is selected based on the internal quality analysis. In the final database, the topological relationships between spatial objects are defined according to the *reference* database. This is the most desirable case amongst those we can meet for spatial data integration.

Unfortunately, the internal quality is not always well-described. Therefore, a comparison between the available source databases cannot usually support the selection of one *reference*. Let's assume that a set of spatial objects is given, each having different geometries in different sources, and that the spatial data quality of each object is poorly described, then no clear conclusion can be drawn from such a situation. In this case, one possibility is to consider that the available source geometries of each spatial object contribute in an equal way to its final geometry. Then, the spatial intersection of the source geometries of a given object provides the subpart where a consensus has been found. In addition, the spatial union of these same geometric representations provides the exhaustive area where the object might be found. If the difference between the intersection and union is non-empty, then the object shape can be considered as vague since only a subpart was agreed upon (i.e. the result of the spatial intersection). In other words, by using only the knowledge provided by the data sources, it is not possible to be certain about the object shape; however, it is possible to deduce the complete or partial vagueness of this shape. *Regions with broad boundaries such as forest stands* and *lines with broad interiors such as canoe routes between two piers* are examples among several of objects with vague shapes. A database designer may use this type of geometries to integrate heterogeneous redundant geometries in order to improve the data reliability, especially in a data feeding process. For example, the management of the wood industry in a given forest should consider the broad boundaries of forest stands (i.e. it is an oversimplification of the reality to surround a forest stand by a linear boundary), to decrease the risks of wrong analyses and decisions. In Figure 1, we assume that a spatial object *A* is represented by three heterogeneous geometries in three different databases, respectively. The final geometry resulted from the integration of the source geometries of *A* is a region with a broad boundary. The broad boundary refers to the difference between the union and intersection of source geometries. Then, the decision-maker takes account of broad boundary in order to get the most appropriate decision. For example, if *A* refers to a forest stand, he can adjust the production of wood inside the broad boundary according to the available data.

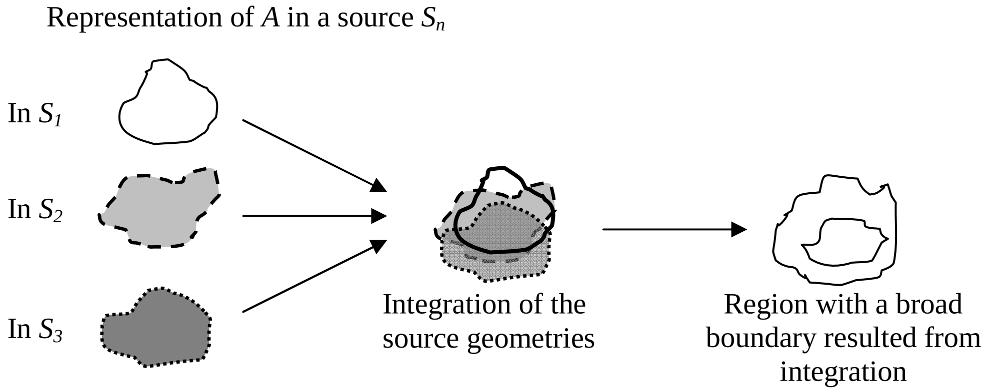


Figure 5.1 Example of the integration of three heterogeneous geometries representing the same object

In the source databases, topological relationships happen between well-defined shapes can be controlled by topological integrity constraints (Pinet *et al.* 2007). Such control allows one to make sure the quality of data is on par with the specifications. Topological (integrity) constraints are an important class of integrity constraints for such spatial databases. They refer to a set of rules defined at the conceptual level in order to reduce the topological inconsistencies in spatial databases (e.g. *roads and buildings should be Disjoint*) (Cockcroft 1997, Normand 1999, Servigne *et al.* 2000). These constraints can be specified by using specific languages such as the Object Constraint Language (OCL) (Waremer and Kleppe 1998, Pinet *et al.* 2007).

The heterogeneity of partly or totally redundant source geometries and the poor description of their internal quality entail a shape vagueness and may produce an uncertainty concerning the topological relationships between objects of the final database. In the final database, the shape vagueness must be taken into account in order to define the topological integrity constraints properly. An adequate model of topological relationships is necessary and an adapted method is needed to characterize them in a given situation. The characterization of such relationships can be also useful for the specification of spatial queries.

In this paper, we address the problem of *topological relationships vagueness*, which we define as the uncertainty about the appropriate topological relationship between possibly vague shapes resulting from the integration of multi-source redundant data. These relationships are generally different from those occurring in the source databases, because they involve shape vagueness. The main objective is to reduce this topological relationships vagueness when specifying the topological integrity constraints for a given final database (e.g. a warehouse). We propose a model to define this vagueness and an approach to characterize the possible topological relationships between the geometries resulting from the integration process. We apply these concepts to the case where the final database is a spatial data warehouse (with. a hypercube structure) and where the final geometries are loaded in one hierarchy level of a spatial dimension (for the details of hypercube structures in spatial data warehousing, we refer the readers to (Bédard and Han 2008)). In this context, we assume that the semantic heterogeneities have been resolved and only the appropriate intra-level topological relationships need to be specified in the final constraints. In the same way, we assume that no integrity constraints are defined between geometries belonging to different hierarchy levels. We do not deal with the topological relationships between child and parent members in a spatial dimension hierarchy.

The paper is organised as follows. In Section 5.4, we refer to some works related to the topic of geometric heterogeneities in spatial data integration and the use of specific spatial models to represent the shape vagueness. In Section 5.5, we explain the problem studied in this paper. Section 5.6 presents the spatial model to merge heterogeneous redundant geometries that represent a given spatial object. Section 5.7 describes our approach to analyze possible topological relationships between geometries with vague shapes resulting from the integration process. We propose two strategies to reduce the topological relationships vagueness: (1) *modifying the final geometries in order to completely respect topological relationships* (i.e. using topological operators for objects with well-defined shapes, such as those defined in the 9-Intersection model (Egenhofer and Herring 1990)) or (2) *using an adverbial approach to partially characterize these relationships*. Section 5.8 presents an example of reducing the vagueness of intra-level topological relationships in a spatial data warehouse. Finally, Section 5.9 presents the conclusions and some perspectives of this work.

5.4 Previous works

5.4.1 Geometric heterogeneities in spatial data integration

In spatial databases, the values of geometric attributes can be observed and measured in different ways (Mowrer 1999). This property of geometric data allows room for more than one value and could entail some difficulty when heterogeneous geometries for a same object need to be integrated (Devogel 1997). Figure 2 presents three examples of spatial objects with heterogeneous redundant representations in source databases. Figure 5.2(a) shows a set of points, each one of them being an heterogeneous redundant representation measured at the same epoch for the same spatial object (e.g. a fire hydrant). In the same way, Figure 5.2(b) and Figure 5.2(c) show the same thing for lines and regions that can represent objects such as a river and a lake, respectively.

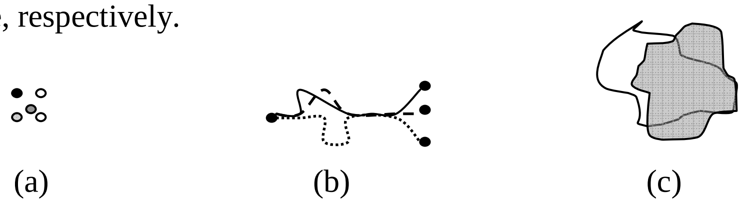


Figure 5.2 Example of redundant heterogeneous representations: (a) 5 representations of the same 0-D object, (b) 3 representations of the same 1-D object, (c) 2 representations of the same 2-D object

The principal function used to merge source crisp geometries is the *overlay* method (Frank 1987, Demirkesen and Schaffrin 1996, Harvey and Vauglin 1996). This approach consists in identifying features in different data sources intended to represent a same world object before merging them into a final geometric representation. The overlay method assumes that one data source (called *reference*) has a higher quality then the other available data sources. The nodes of a geometry belonging to A should remain fixed. A tolerance error termed *tolerance match* is associated to the geometries of A in order to consider the geometries of the other sources within this tolerance in the integration process. In other words, if a feature F_S belonging to a data source S is within the match tolerance of a feature F_A belonging to the reference A, then each node of F_S should be moved to an existing or newly created node of F_A (Ware and Jones 1998).

The overlay approach requires that the internal quality is well described in the data sources in order to select a reference among them. Then, a final geometry with a possibly vague shape may result from the integration of source geometries (Shepherd 1992). Accordingly, some approaches use specific spatial models in order to represent inherent shape vagueness of several spatial objects such as inundation areas or pollution zones (Clementini and Di Felice 1997, Cohn and Gotts 1996, Erwig and Schneider 1997, Tang 2004, Bejaoui et *al.* 2008) (Section 5.4.2). These models can be also used to represent final geometries with vague shapes resulted from the integration of redundant and heterogeneous geometries. Nevertheless, the specification of topological integrity constraints involving these geometries is still an open question since it is required to consider the shape vagueness. Spaccapietra and Parent (1991) suggested choosing the least constrained database as a reference. This approach can be efficient when the least constrained database has also the highest quality. Rodriguez (2005) proposed to disable any constraints when different topological relationships are possible for geometries resulted from the integration process. For example, figure 5.3 shows two spatial objects that are represented differently in two data sources *A* and *B*. Three topological relationships are possible between final geometries: *Overlap*, *Meet*, *Disjoint*. The topological inconsistencies in the final database are increased.

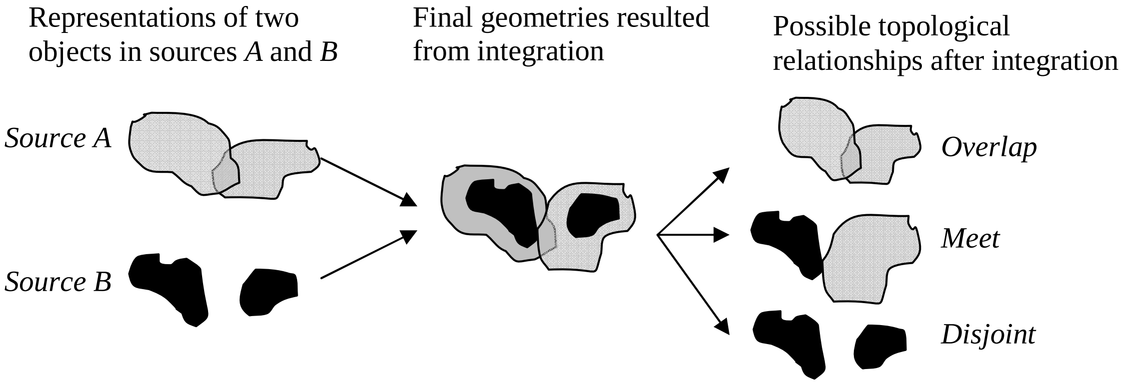


Figure 5.3. Possible topological relationships for final geometries

5.4.2 Formal specification of objects with vague shapes and their topological relationships

Two categories of models are generally used to deal with spatial vagueness. In the first category, crisp spatial concepts are transferred and extended to formally express the spatial vagueness: we speak about *exact models* (Clementini and Di Felice 1997, Cohn and Gotts 1996, Erwig and Schneider 1997, Tang 2004). In the second category, three principal mathematical theories are generally used: (1) the models based on Fuzzy Logic (Zadeh 1965) (e.g. Dilo 2006) which can be used to represent continuous phenomena such as temperature, (2) the models based on Rough Sets (e.g. Worboys 1998) which represent spatial objects with vague shapes as a pair of approximations (*upper approximation*, *lower approximation*) and (3) the models based on probability theory (e.g. Burrough and Frank 1996, Pfoser *et al.* 2005) which are primarily used to model position errors. A literature review on specification of spatial objects with vague shapes and their topological relationships has been realized in Section 2.3.

5.5 Problem Statement

In the integration process, the *topological relationships vagueness* increases when the available source geometries of each spatial object are heterogeneous (*measured using different methods which not give the same results*) and the internal quality is ill described in source databases. In this case, geometric heterogeneities entail shape vagueness for the final geometries whenever the difference between the union and intersection of available source geometries is non-empty. The topological relationships between final geometries should be redefined to take into account their possible shape vagueness.

Let A and B be two spatial objects with heterogeneous geometries (A_1, B_1) and (A_2, B_2) in two source databases S_1 and S_2 respectively. The final geometries of A and B can be represented by two regions with broad boundaries. A broad boundary refers to the difference between the union and intersection of the source geometries (i.e. (U_A, I_A) for A and (U_B, I_B) for B). For example, assume that the geometries of A and B are *Disjoint* in both sources (Figure 5.4). It appears that the *Disjoint* relationship is *partially* respected in the final database because it holds for the intersections (I_A, I_B), whereas the unions (U_A, U_B) meet each other.

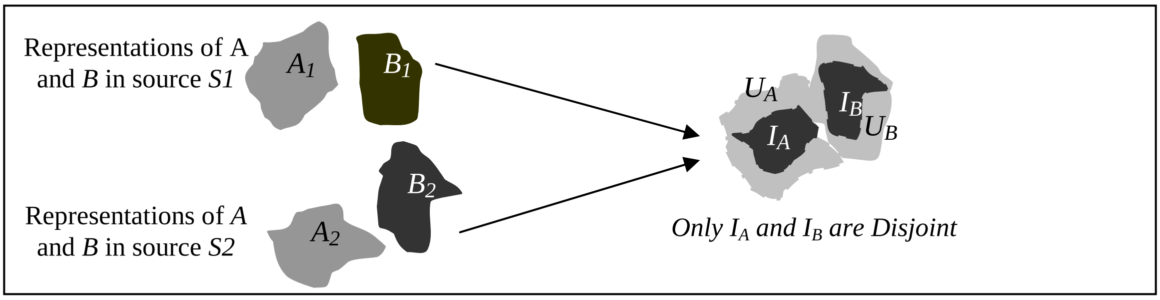


Figure 5.4. Example of topological relationships vagueness

From this perspective, there is a need for a specific spatial model to represent the shape vagueness and to compute the topological relationships between final geometries. Therefore, the primary existing exact models (Cohn and Gotts 1996, Clementini and Di Felice 1997, Erwig and Schneider 1997, Tang 2004, Reis *et al.* 2006) show some limitations. Most of these models cannot formally represent objects with *partially* vague shapes, such as a lake with rocky banks on one side and swamp banks on the other. For example, *regions with partially broad boundaries* are considered invalid because the connectedness condition is violated (i.e. the boundary should be broad everywhere around the region interior). However, it is important to consider this type of regions, as they can result from integration when the difference between the union and intersection of source geometries is non-empty on some locations and empty on other ones. The spatial model proposed by Bejaoui *et al.* (2008) can be used to define regions of this type and their topological relationships (Section 5.6).

In this paper, we address the problem of characterizing the *topological relationships vagueness* for final geometries resulting from the integration process. Our aim is to answer the following questions:

4- How is it possible to represent a region with a broad boundary that results from merging the heterogeneous polygons representing a given spatial object in different source databases?
5- How can we deduce the possible topological relationships between final geometries from the relationships defined in the data sources? The answer can help the specification of topological integrity constraints.
6- Which strategies can be defined to reduce the *topological relationships vagueness*?

5.6 Merging heterogeneous polygons through regions with broad boundaries

The shape vagueness can characterise any geometric primitive (*point*, *line* or *region*). In this paper, we focus our investigation on the regions. For other geometric primitives, we suggest the following references (Clementini 2005, Bejaoui *et al.* 2008). The present section presents the spatial model for regions with broad boundaries defined in (Bejaoui *et al*. 2008). This model is not a contribution of the present paper. However, it is one of the primary elements on which our proposed approach is based. The definition of a region with a broad boundary is adapted to the geometric heterogeneity problem and semantically different from that proposed in (Bejaoui *et al*. 2008) (cf. Section 5.6.1). For that reasons, we present the spatial model in a separate section, in order to facilitate understanding of the remainder of the paper.

5.6.1 Regions with broad boundaries resulting from integration

In spatial data integration, a region with a broad boundary may result from geometric heterogeneities of the sources' geometries of the object involved. It corresponds to the difference between the union and intersection of the source geometries. The intersection refers to the minimal extent of the object, whereas the union refers to its maximal extent. Figure 5.5 represents an example of a region with a broad boundary resulting from the integration of two heterogeneous source geometries of the same spatial object *A* at the same epoch. The final geometry of this object corresponds to a region where the minimal extent is covered by the maximal one. In this example, the boundary of the minimal extent is partially superposed on the boundary of the maximal one. Then, the region has a *partially broad boundary*.

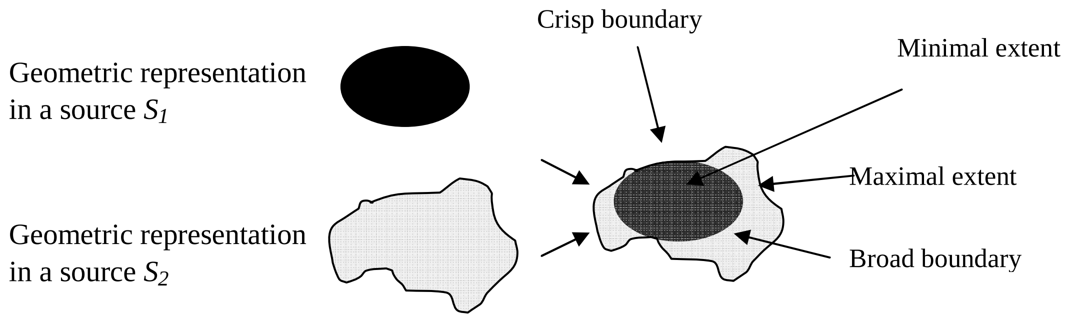


Figure 5.5. Region with a partially broad boundary in spatial data integration

We consider a region with a broad boundary $\tilde{A}$ resulted from the integration of heterogeneous geometries of the same object at the same epoch. It is made up of two crisp sub-regions: *a maximal extent* $\tilde{A}_{\max}$ *(i.e. the union of source representations), and a minimal extent* $\tilde{A}_{\min}$ *(i.e. the intersection of source representations) where Equal*[7]*(*$\tilde{A}_{\max}, \tilde{A}_{\min}$) *or Contains(*$\tilde{A}_{\max}, \tilde{A}_{\min}$) *or Covers(*$\tilde{A}_{\max}, \tilde{A}_{\min}$) (Figure 5.5). In this definition, we assume that the source geometries should intersect each other (i.e. an isolated geometry cannot be considered in the integration process). In other words, we do not deal with the case where there is no intersection between source geometries, because we consider that they do not represent the same object if they represent the geometry for the same time. The boundary of $\tilde{A}$ can be *completely crisp* (or not at all broad) when the difference between the union (*maximal extent*) and the intersection (*minimal extent*) of the source geometries (i.e. *identical geometries in all source databases*) is empty. In another case, the boundary is *partially broad* when the intersection and union are different only in some locations. In this case, the union of source geometries *covers* their intersection. Finally, the third possibility is a region with a *completely broad boundary*. In other words, the union of source geometries *contains* their intersection. In Figure 5.6, we present an example of each of these three cases.

[7] The spatial relations (i.e. *Equal, Contains, Covers*) used in this definition are those defined in (Egenhofer and Herring 1990).

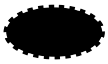
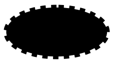
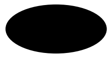
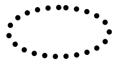
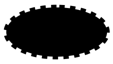
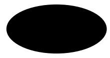
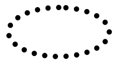
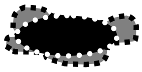
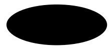
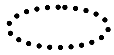
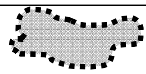
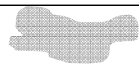
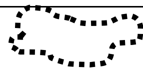
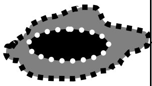
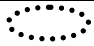
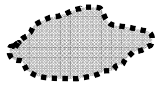
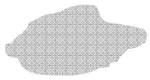
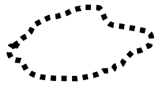
Region with a vague shape	Representation	Maximal and minimal extents		Topological invariants of minimal and maximal extents	
				Interior	Boundary
A region with a crisp broad boundary (i.e. *crisp region*)		Minimal extent			
		Maximal extent			
A region with a *partially* broad boundary (i.e. *partially vague region*)		Minimal extent			
		Maximal extent			
A region with a completely broad boundary (i.e. *completely vague region*)		Minimal extent			
		Maximal extent			

Figure 5.6. Regions with broad boundaries

The consideration of all the available geometries in the integration result increases the shape vagueness of the final geometry. However, it decreases the uncertainty about the possible shape vagueness (or the meta-uncertainty (Bédard 1987)) of the spatial object involved. The user has a reliable idea of the data imperfection despite the fact that he does not have the well-defined shapes that would be obtained by a simplification of the reality nor well described quality information. One can make an analogy between this approach and the computation of an error ellipse in geodesy (Chrisman 1991). An error ellipse provides an area of a probable position around the true position of a point (its uncertainty) along with a percentage of probability for this point to be within this area (meta-uncertainty). This percentage increases according to a parameter C that defines the size and orientation of the ellipse. For example, there is a 38 per cent chance that the true position will fall within a standard error ellipse (i.e. $C=1$). Similarly, there is a 90 per cent chance that the true position fall within an ellipse defined by $C=4.6$. Then, increasing the error ellipse radius improves the spatial accuracy (or exactness) whereas it decreases the data precision, and in both cases the meta-uncertainty is known (38%, 90%). The user manipulates the data with better accuracy despite the more limited precision.

5.6.2 Topological relationships between regions with broad boundaries

In this paper, we defined the result of integration of heterogeneous source polygons representing as a region with a broad boundary. In order to specify their topological relationships, we use the $QMM_{Topological\ Relationships}$ model for regions with broad boundaries defined in (Bejaoui et *al.* 2008) (cf. section 3.7).

5.7 Controlling the validity of topological relationships in spatial data integration

5.7.1 The different situations

Let's consider a spatial object represented by heterogeneous and redundant crisp geometries stored in different data sources with different specifications. The final geometry of this object displays shape vagueness if there is a difference between the intersection and union of the source geometries. The source geometries contribute equally to the final geometry of the object while the comparison of their quality does not allow the selection of a reference geometry. Then, we speak about a *non-distinctive* internal quality. In this case, the topological relationships vagueness can be *reduced* when the source topological relationships are identical. The impossible relationships may be deduced despite the shape vagueness of the final geometries. For example, let's assume that the geometries of two spatial objects *A* and *B* are respectively disjoint in two data sources. Then the intersections of the source geometries of *A* and the source geometries of *B* are necessarily disjoint. *Contains* is an inconsistent relationship between the final geometries of *A* and *B*. In this paper, we consider only the situation where the internal quality is non-distinctive and the topological relationships in the data sources are identical. We also study the problem of topological relationships vagueness only for spatial objects represented by polygons. The same methodology can be used to address such a problem for objects represented by lines and points.

Considering that internal quality of sources can be distinctive or non-distinctive and that topological relationships in sources can be identical or different, there exist four situations that can be studied (Table 5.1). Only the second situation (the grey cell) is explored in the present work.

Table 5.1 potential cases of topological relationship in different sources

	Distinctive internal quality	Non-distinctive internal quality
Identical topological relationships in the sources	Non-Studied	Studied
Different topological relationships in the sources	Non-Studied	Non-Studied

In the following, we assume that the source geometries are topologically consistent in their respective data sources (i.e. the topological relationships specified in the integrity constraints are respected). Then, we explore the topological relationships to be respected by the final geometries resulting from the integration process. Two points of view are considered in the following study: (1) modifying the topological relationships and keeping the shapes of final geometries invariant, (2) keeping the source topological relationships invariant and modifying the final geometries.

5.7.2 Characterizing the possible topological relationships for the final geometries when a same topological relationship is specified in the sources

We assume that the final geometry of a given object is a region with a broad boundary. Likewise, we assume that the same topological relationships are specified for a same set of objects. The main objective of this section is to characterize the possible topological relationship between the final geometries.

In the cases studied below, we suppose that an object A has n heterogeneous source geometries A_1, A_2,.. and A_n in the sources S_1, S_2, .., S_n, respectively. In the same way, an object B has n heterogeneous source geometries B_1, B_2,.., and B_n in S_1, S_2, .., and S_n, respectively. Then, we consider separately the different cases of the topological relationships that can arise in the data sources (i.e. *Disjoint*, *Contains*, *Inside*, *Covers*, *Covered by*, *Meet* and *Overlap*). In Figures 5.10-5.14, we represent only two heterogeneous redundant geometries for A and B in order to improve the paper readability. These examples can be easily extended to any other number of sources.

We term I_A and I_B the intersection of A_1, A_2,.. and A_n and that of B_1, B_2,.. and B_n, respectively. In the same way, we term U_A and U_B the union of A_1, A_2,.. and A_n and that of

B_1, B_2,.. and B_n, respectively. The demonstrations of the possible relationships between the intersections (I_A, I_B) and between the unions (U_A, U_B) are presented in Appendix 3. These demonstrations concern the eight cases of topological relationships that can arise in the data sources.

❖ **Disjoint**

In each source database, the geometries of spatial objects *A* and *B* have the same topological relationship: they are *Disjoint*. Then, I_A and I_B are certainly *Disjoint*. Otherwise, the topological relationship is not respected in one data source at least. For U_A and U_B, one of the following relationships is possible: *Disjoint*, *Meet* or *Overlap* (see the demonstrations in Appendix 3). A union can contain points that do not belong to all of the source geometries. Therefore, whether the unions overlap or meet each other, the *Disjoint* relation is still possible between the objects involved. The *maximal topological consistency* is obtained when *Disjoint* (I_A, I_B) and *Disjoint* (U_A, U_B), because then the source topological relationship is respected despite the shape vagueness. Finally, we conclude that the *final geometries* should conform to the restrictions of the next matrix:

$$\begin{array}{cc} & \begin{array}{cc} \quad I_B \quad & \qquad\qquad U_B \end{array} \\ \begin{array}{c} I_A \\ U_A \end{array} & \left[\begin{array}{cc} \textit{Disjoint}\ (I_A, I_B) & \text{--} \\ \text{--} & \{\textit{Disjoint, Meet, Overlap}\}\ (U_A, U_B) \end{array}\right] \end{array}$$

In the matrix presented above, no restrictions are imposed for U_A/I_B and I_A/U_B; this situation is marked by “--" in the corresponding cells of the matrix. Figure 5.10 shows an example illustrating this case with two heterogeneous geometries of two spatial objects *A* and *B*. In S_1 and S_2, the geometries of *A* and *B* are disjoint. In this example, regions with broad boundaries R_A and R_B are resulted from the integration process to represent *A* and *B* in the final database. In R_A and R_B, the intersections I_A and I_B are *Disjoint*. The unions U_A and U_B are also *Disjoint*. Then, we conclude that these final geometries are *topologically consistent*.

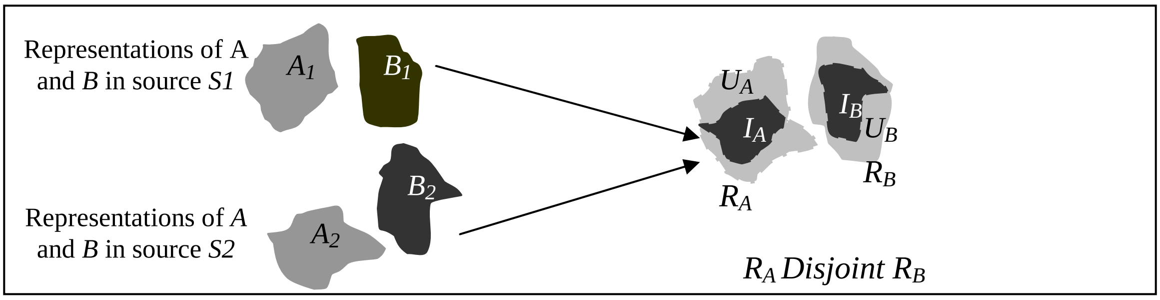


Figure 5.10. Integration example where the topological relationship defined in source databases is *Disjoint*

❖ **Contains/Inside**

In the data sources, the geometry of *A* contains that of *B*. Then, relationship between intersections I_A and I_B is necessarily *Contains*. Likewise, U_A contains U_B. Consequently, the topological relationship (i.e. *Contains/Inside*) holds despite the shape vagueness. The final geometries are consistent while the source geometries involved in the integration respect the topological relationship *Contains/Inside* (see the demonstrations in Appendix 3). Hence, we say that *Contains* and *Inside* are *invariant topological relationships*. They are still invariant despite the heterogeneity of the source geometries. For these relationships, the final geometries should conform to the restrictions of one of the following matrices:

- For *Contains* (*A*, *B*)

$$
\begin{array}{cc}
 & \begin{array}{cc} I_B & \qquad\qquad\qquad U_B \end{array} \\
\begin{array}{c} I_A \\ U_A \end{array} & \left(\begin{array}{cc} \textit{Contains}\ (I_A, I_B) & \text{--} \\ \text{--} & \textit{Contains}\ (U_A, U_B) \end{array}\right)
\end{array}
$$

- For *Inside* (*A*, *B*)

$$
\begin{array}{cc}
 & \begin{array}{cc} I_B & \qquad\qquad\qquad U_B \end{array} \\
\begin{array}{c} I_A \\ U_A \end{array} & \left(\begin{array}{cc} \textit{Inside}\ (I_A, I_B) & \text{--} \\ \text{--} & \textit{Inside}\ (U_A, U_B) \end{array}\right)
\end{array}
$$

Figure 5.11 shows an example of two heterogeneous geometric representations of two spatial objects *A* and *B*, where A_1 contains B_1 and A_2 contains B_2. Then I_A *contains* I_B and U_A *contains* U_B.

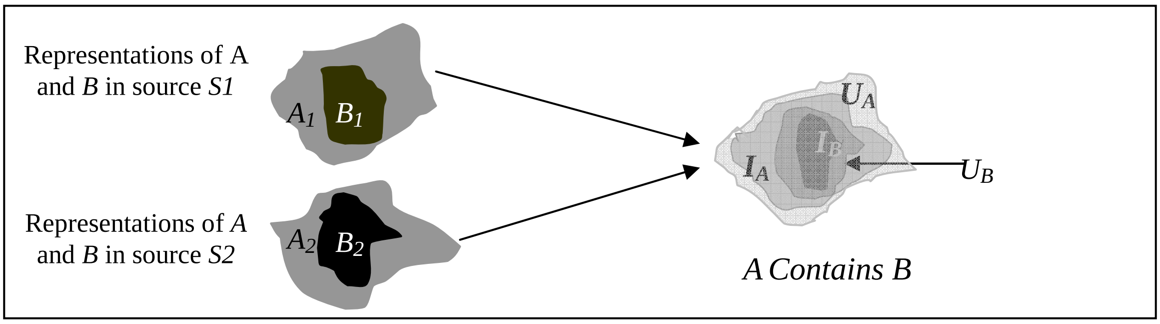


Figure 5.11. Integration example where the topological relationship in the data source is *Contains*

❖ **Covers/Covered by**

In the data sources, we assume that the geometry of *A* covers that of *B*. The intersection of source geometries of *A* covers the intersection of source geometries of *B* (I_A Covers I_B). The relationship between the intersections is also valid when it corresponds to *Contains*. In the latter case, the intersections between the boundaries of source geometries of the objects *A* and *B* do not arise in the same location in the different data sources. The relationship between intersections cannot be different to *Contains* or *Covers* (see the demonstrations in Appendix 3).

The union U_A should always cover U_B. The maximal topological consistency is obtained when *Covers* (I_A, I_B) and *Covers* (U_A, U_B) because the topological relationship is respected despite the shape vagueness. The same conclusions may be made to *Covered by*: {*Covered by or Inside*} between the intersections (I_A, I_B) and *Covered by* between the unions (U_A, U_B). *Covers* and *Covered by* are also two *invariant topological relationships* for the unions. Finally, the final geometries should conform to the restrictions of one of the following matrices:

- *Covers* (*A*, *B*)

$$\begin{array}{ccc} & I_B & U_B \\ \begin{array}{c} I_A \\ U_A \end{array} & \multicolumn{2}{c}{\begin{bmatrix} \{\textit{Covers, Contains}\}\ (I_A, I_B) & \text{--} \\ \text{--} & \textit{Covers}\ (U_A, U_B) \end{bmatrix}} \end{array}$$

- *Covered by* (*A*, *B*)

$$\begin{array}{ccc} & I_B & U_B \\ \begin{array}{c} I_A \\ U_A \end{array} & \multicolumn{2}{c}{\begin{bmatrix} \{\textit{Covered by, Inside}\}\ (I_A, I_B) & \text{--} \\ \text{--} & \textit{Covered by}\ (U_A, U_B) \end{bmatrix}} \end{array}$$

Figure 5.12 shows an example two spatial objects *A* and *B* that are represented in two different data sources *S1* and *S2*. In S1 and S2, the representations of *A* and *B* are respectively related by the following relationships: *Covers* (A_1, B_1) and *Covers* (A_2, B_2). For the final geometries, we have *Covers* (I_A, I_B) and *Covers* (U_A, U_B).

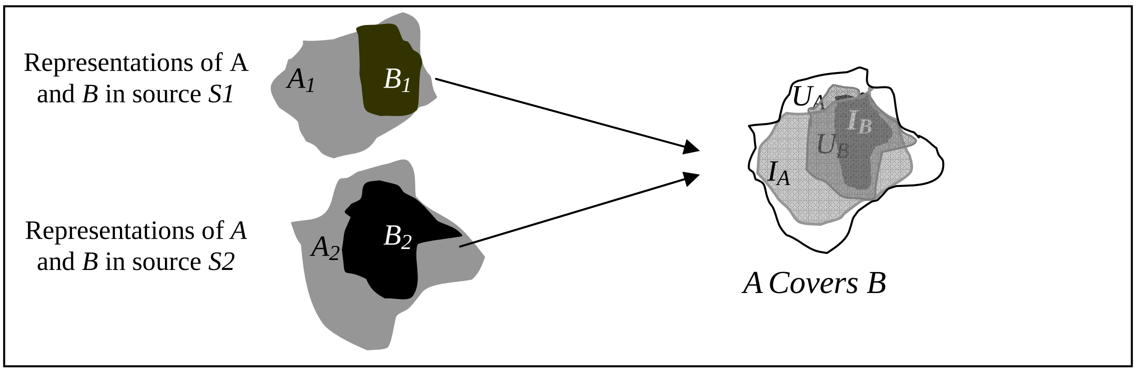


Figure 5.12. Integration example where the topological relationship defined in the sources is *Covers*

❖ Overlap

In the data sources, the geometry of *A* and that of *B* overlap each other. In the final geometries, one of the following relationships may arise between the intersections I_A and I_B: *Disjoint*, *Meet* or *Overlap*. However, the unions U_A and U_B should overlap each other. When the *Overlap* relationship does not occur in the same location for all the source geometries, this part of the interior cannot appear in the intersections. In the latter case, the *Overlap* relationship is still possible between the objects involved while the relationship between the intersections I_A and I_B is *Meet* or *Disjoint* (see the demonstrations in Appendix 3). The maximal topological consistency is obtained when *Overlap* (I_A, I_B) and *Overlap* (U_A, U_B), i.e. the topological relationship is preserved despite the shape vagueness. Finally, the final geometries should conform to the restrictions of the following matrix:

$$\begin{array}{cc} & \begin{array}{cc} I_B & \qquad\qquad U_B \end{array} \\ \begin{array}{c} I_A \\ U_A \end{array} & \begin{pmatrix} \{\mathit{Overlap, Meet, Disjoint}\}\ (I_A, I_B) & \text{--} \\ \text{--} & \mathit{Overlap}\ (U_A, U_B) \end{pmatrix} \end{array}$$

Figure 5.13 shows an example of two heterogeneous geometric representations of two spatial objects *A* and *B*. In S_1 and S_2, the respective geometries of *A* and *B* overlap each other. After the integration, the intersections I_A and I_B overlap each other. In the same way, the relationship between the unions U_A and U_B is *Overlap*.

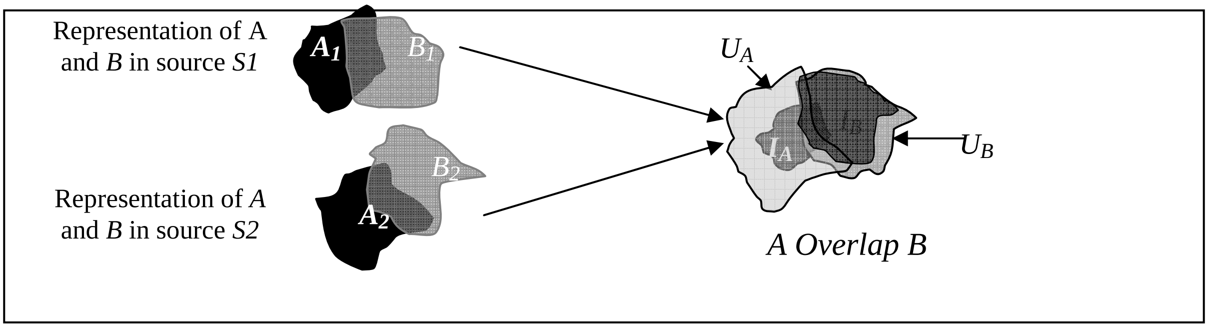


Figure 5.13. Integration example where the topological relationship defined in the sources is *Overlap*

❖ Meet

Let's assume that *Meet* is the relationship between the geometries of *A* and *B* in each one of data sources. In the final database, the topological relationship between the intersections I_A and I_B can be either *Disjoint* or *Meet*. However, the unions U_A and U_B should be connected by one of the following relationships: *Meet*, *Overlap*. The unions U_A and U_B should overlap or meet each other while the intersections I_A and I_B meet each other. Otherwise, the source geometries of *A* and *B* do not respect the topological relationship in one data source at least (see the demonstrations in Appendix 3).

When the intersections I_A and I_B are *Disjoint*, the unions U_A and U_B should overlap or meet each other. The latter case occurs when the source geometries of *A* and *B* do not meet each other in the same locations. A maximal topological consistency is obtained when *Meet* (I_A, I_B) and *Meet* (U_A, U_B), because the topological relationship is preserved despite the shape vagueness. Finally, the final geometries should conform to the restrictions of one of the following matrices:

- *Meet*(I_A, I_B):

$$\begin{array}{cc} & \begin{array}{cc} I_B & \qquad\qquad U_B \end{array} \\ \begin{array}{c} I_A \\ U_A \end{array} & \begin{bmatrix} Meet\ (I_A, I_B) & \text{--} \\ \text{--} & \{Overlap, Meet\}\ (U_A,U_B) \end{bmatrix} \end{array}$$

- *Disjoint*(I_A, I_B)

$$\begin{array}{cc} & \begin{array}{cc} I_B & \qquad\qquad U_B \end{array} \\ \begin{array}{c} I_A \\ U_A \end{array} & \begin{bmatrix} Disjoint\ (I_A, I_B) & \text{--} \\ \text{--} & \{Overlap, Meet\}\ (U_A,U_B) \end{bmatrix} \end{array}$$

Figure 5.14 shows two spatial objects *A* and *B* where the geometry of *A* meets that of *B* in each data source. In this example, the intersections I_A and I_B are *Disjoint* even though the unions U_A and U_B overlap each other. The final geometries with vague shapes are topologically valid because they satisfy the specifications of the second matrix presented above: *Disjoint* (I_A, I_B) and *Overlap* (U_A, U_B).

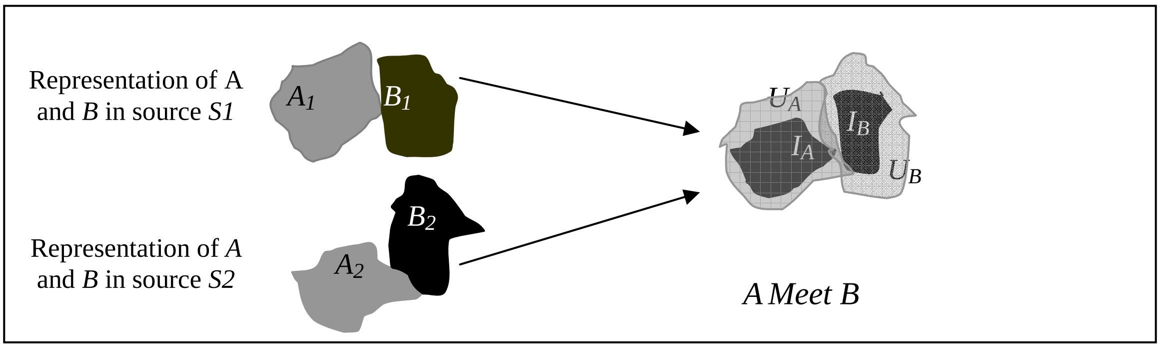


Figure 5.14. Integration example where the topological relationship defined in the sources is *Meet*

5.7.3 Strategies to reduce the vagueness of topological relationships

5.7.3.1 Principles of the strategies

In this paper, we propose two strategies to reduce the vagueness about the topological relationships in a final database:

1) *Choosing the best extents of objects involved and modifying them if they violate the topological relationship defined in the sources. The final geometries are modified to be crisp (or well-defined).*

2) *Using final geometries with vague shapes and apply an adverbial approach to stress the partial respect of the topological relationship.*

The geometric modifications of final geometries aims at *forcing* them to respect the topological relationships defined in the data sources. Such a strategy may be used when the topological relationships are more important to the final users than the objects' shapes involved. For example, it is sometimes required to prevent an overlap relationship between the forest stands in spite of their broad boundaries. The principles of geometric modifications are proposed and discussed in (Ubeda and Egenhofer 1997). For example, it is possible to retain only the relationship between intersections of the source geometric representations even though the unions may violate the source topological relationship, and vice versa. In the case of the *Disjoint* relationship (Section 5.7.2), the intersections of the objects involved are usually *Disjoint*. Consequently, the intersections give rise to consistent final geometries of the objects involved.

The second strategy retains the final geometries and uses topological operators adapted to regions with broad boundaries, as defined in (Bejaoui *et al.* 2008). This strategy considers the intersection as the minimal extent of the object and the union as its maximal extent. Then, a given topological relationship is *partially* respected, because only the impossible relationships (*based on the source geometries and the topological relationships defined in the sources*) are forbidden. For example, if the source geometries are *Disjoint*, then it is impossible to have a *Contains* relationship between the intersections or between the unions (Section 5.7.2).

5.7.3.2 First strategy: modifying the final geometries

The modification of geometries is an important issue that has been thoroughly studied in several works on the spatial data conflation (e.g. Saalfled 1993, Rodriguez 2005, Casado 2006) and the control of spatial databases consistency (e.g. Ubeda and Egenhofer 1997). In this section, pragmatic examples are provided in order to illustrate several possible ways to modify a final geometry resulted from an integration process. The goal of this modification is to force the verification of a topological relationship in a final database and to reduce the vagueness of topological relationship due to geometric heterogeneities.

In (Ubeda and Egenhofer 1997), two types of geometric modifications are proposed in order to correct topological relationship violations: *moving and reshaping*. *Moving* an object *A* consists of translating it in one of five main directions: along the *X* axis, along the *Y* axis, perpendicular to *B* (a second object), parallel to *B*, and along *A*. Moving an object can be used to change its relative position according to another object while preserving its area. *Reshaping* an object *A* consists of deforming its original geometry. According to Ubeda and Engenhofer (1997), reshaping refers to move one or several parts of A’s geometry and leaving another part unchanged. In this context, it is important to note that both the original and reshaped geometries are crisp, simple and connected. Reshaping an object aims to modify its shape in order to force the topological relationship between the two objects involved without changing their relations with other objects of the database.

When the spatial data quality is poorly described, the integration process can produce two geometries for each integrated spatial object: (1) the intersection and (2) union of the source geometries. The goal of this section is to apply geometric modifications on the intersections and/or unions of source geometries in order to force a given topological relationship.

The intersection refers to the parts that exist in all of its source geometries while the unions integrate all of the points that belong to any of the source geometries. It is less risky to choose the intersections if we assume that all of the source geometries have a poor accuracy (Rodriguez 2005). However, the unions can be selected when we assume that all of the source geometries are incomplete (i.e. they do not include all of points that they should). The unions become more reliable geometric representations than the intersections. In other cases, the union may be more appropriate than the intersection for the first object involved, whereas the intersection is better for the second object.

Our approach is to choose the *best extent* among the union and intersection of source geometries of each spatial object. Then, depending on whether the source topological relationship is preserved or not, two principal methods are proposed: (1) preserving the shapes of the *best extents* or (2) changing them (i.e. moving and/or reshaping). The goal of the first method is to leave the shapes of the best extents of the objects unchanged when the topological relationship is respected. The second method aims to modify these extents in order to insure the topological relationship specified in the sources. In Figures 5.15a-5.15e, we consider that two sources are available and we apply the methods presented above on the different cases for the topological relationship (i.e. *Disjoint*, *Contains*, *Inside*, *Covers*, *Covered by*, *Overlap*, or *Meet*). According to Section 5.7.2, we look for the most appropriate method to respect the topological relationship. The goal is to force the *best* extents selected (i.e. *intersections*, *unions* or *intersection – union*) to respect this relationship.

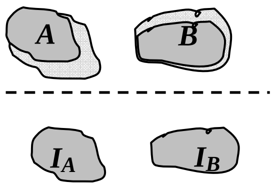
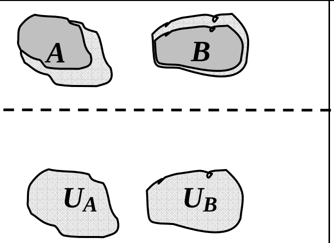
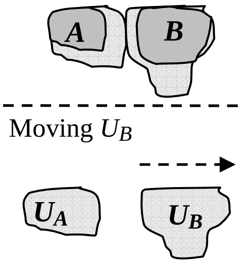
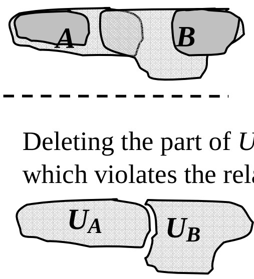
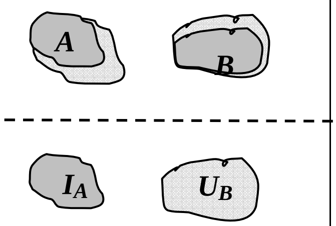
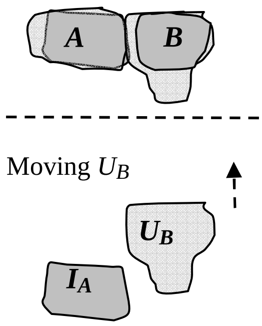
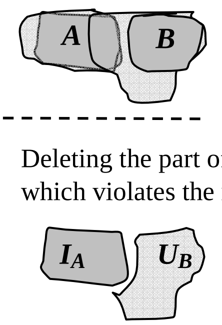
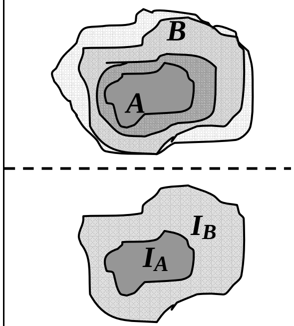
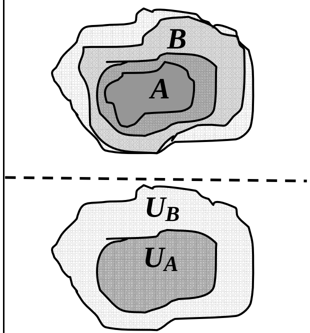
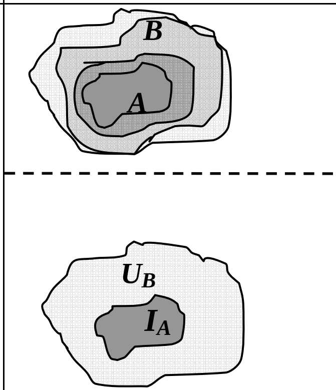
Source topological relationship : *Disjoint*			
Best extents chosen	The topological relationship is respected	The topological relationship is violated	
		Moving one of the best extents	Reshaping one of the best extents
Intersections	 Intersections are disjoint	Not needed because the intersections are necessarily disjoint (cf. Section 5.7.2)	Not needed because the intersections are necessarily disjoint (cf. Section 5.7.2)
Unions	 Occurs when the unions are disjoint		
The intersection for one and the union for the other	 Occurs when the best extents chosen are disjoint		

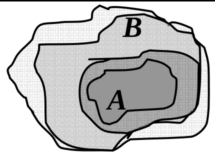
Figure 5.15a. Strategies for forcing a topological relationship in spatial data integration

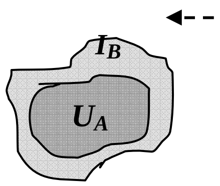
(case of *Disjoint*)

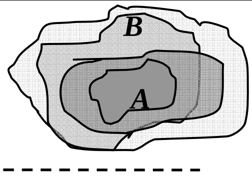
<table>
<tr><th colspan="4">Source topological relationship : Contains/Inside</th></tr>
<tr><th rowspan="2">Best extents chosen</th><th rowspan="2">The topological relationship is respected</th><th colspan="2">The topological relationship is violated</th></tr>
<tr><th>Moving one of the best extents</th><th>Reshaping one of the best extents</th></tr>
<tr><td>Intersections</td><td>


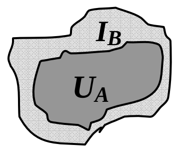
</td><td>Not needed because the relation between the intersections is necessarily Contains/ Inside (cf. Section 5.7.1)</td><td>Not needed because the relation between the intersections is necessarily Contains/ Inside (cf. Section 5.7.2)</td></tr>
<tr><td>Unions</td><td>


</td><td>Not needed because the relation between the unions is necessarily Contains/ Inside (cf. Section 5.7.1)</td><td>Not needed because the relation between the unions is necessarily Contains/ Inside (cf. Section 5.7.2)</td></tr>
<tr><td>The intersection for one and the union for the other</td><td>


</td><td>


Moving U_A


</td><td>


Deleting the part of U_A which violates the relation


</td></tr>
</table>

Figure 5.15b. Strategies for forcing a topological relationship in spatial data integration

(case of *Contains/Inside*)

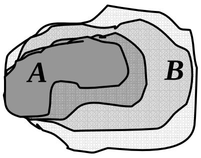
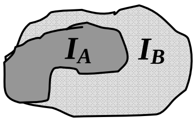
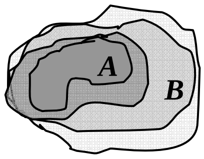
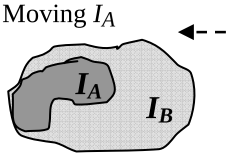
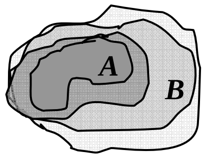
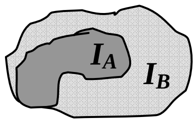
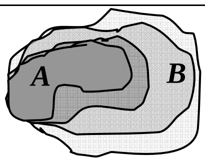
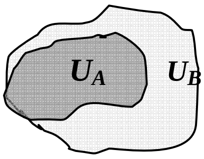
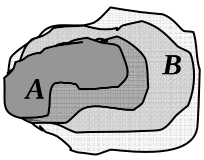
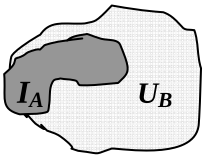
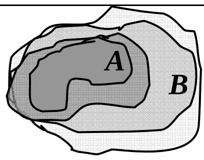
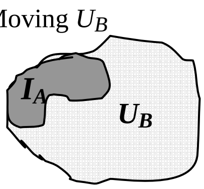
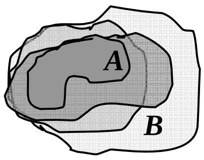
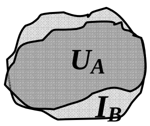
Source topological relationship : *Covers*/*Covered by*			
Best extents chosen	The topological relationship is respected	The topological relationship is violated	
		Moving one of the best extents	Reshaping one of the best extents
Intersections	 	 Moving I_A 	 Expanding I_A 
Unions	 	Not needed because the relation between the unions is necessarily *Covers/ Covered by* (cf. Section 5.7.1)	Not needed because the relation between the unions is necessarily *Covers/ Covered by* (cf. Section 5.7.2)
The intersection for one and the union for the other	 	 Moving U_B 	 Deleting the part of U_A that violates the relation 

Figure 5.15c. Strategies for forcing a topological relationship in spatial data integration

(case of *Covers/Covered by*)

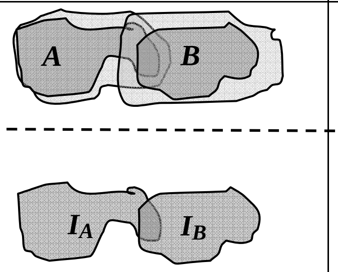
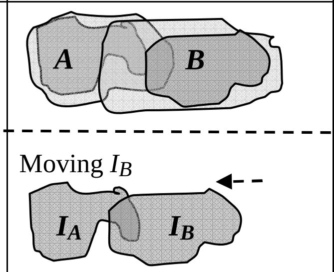
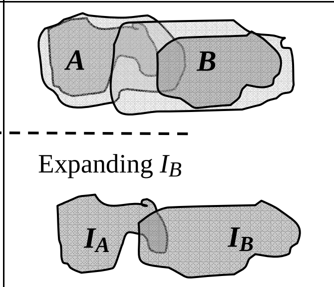
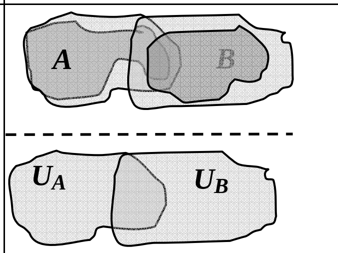
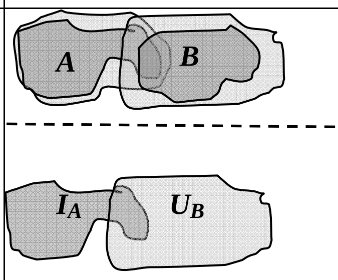
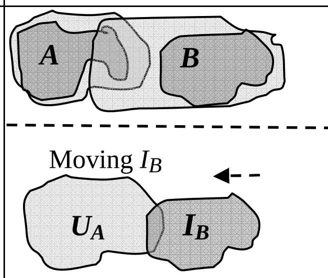
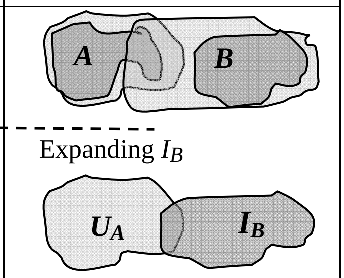
Source topological relationship : *Overlap*			
Best extents chosen	The topological relationship is respected	The topological relationship is violated	
		Moving one of the best extents	Reshaping one of the best extents
Intersections			
Unions		Not needed because the relation between the unions is necessarily *Overlap* (cf. Section 5.7.2)	Not needed because the relation between the unions is necessarily *Overlap* (cf. Section 5.7.2)
The intersection for one and the union for the other			

Figure 5.15d. Strategies for forcing a topological relationship in spatial data integration

(case of *Overlap*)

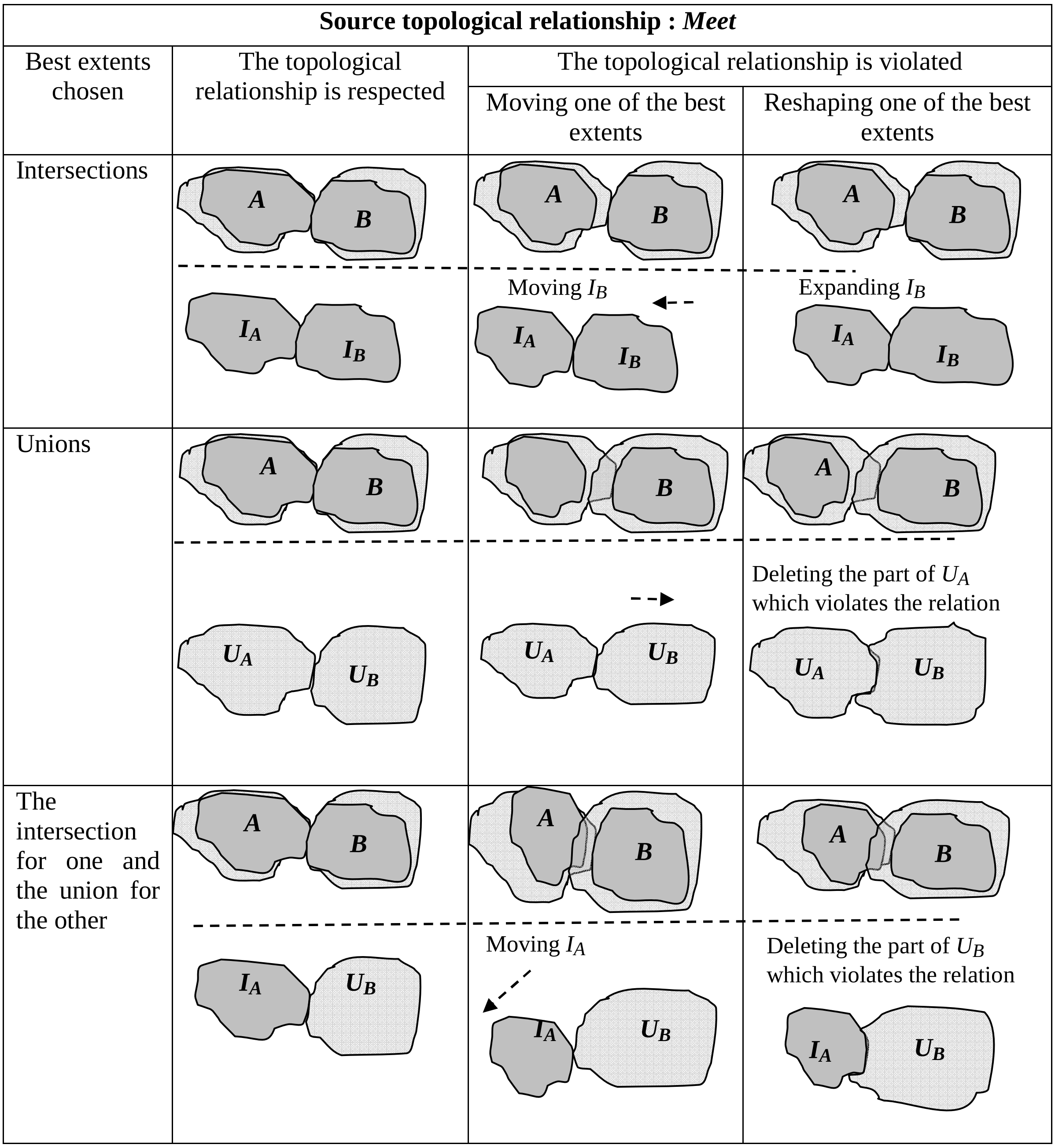


Figure 5.15e. Strategies for forcing a topological relationship in spatial data integration

(case of *Meet*)

5.7.3.3 Second strategy: using an adverbial approach to reduce the vagueness of the topological relationships

The objective of this second strategy is to preserve the final geometries resulting from the integration of heterogeneous source geometries. The shape vagueness of the final geometries implies that the topological relationships cannot be *completely* respected. Our methodology consists of studying the possible topological relationships between the respective unions and intersections of the objects involved. For each topological relationship, the final geometries are valid if the specifications of the related matrices are satisfied (cf. Section 5.7.2). For example, if the source geometries are *Disjoint*, then their respective intersections are necessarily *Disjoint*. In order to express this specification, we propose using the topological relationships introduced in (Bejaoui *et al.* 2008). This adverbial approach reduces the topological relationships vagueness because the topological relationship defined in the source databases is partially respected, and impossible configurations are forbidden. Figure 5.16 shows how the topological relationships are redefined for final geometries with vague shapes.

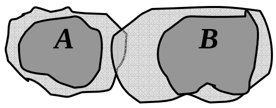
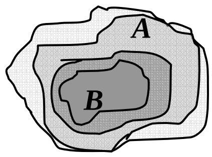
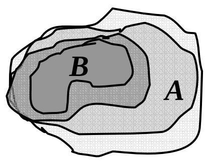
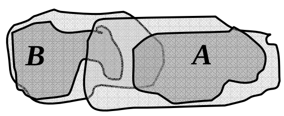
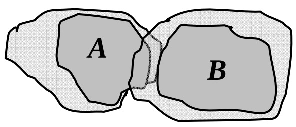
Topological relation displayed in the source databases	Example of geometries resulting from integration	Topological integrity constraint defined for the final geometries
Disjoint		(*A weakly Disjoint B* / *Disjoint*(I_A,I_B) and R(U_A,U_B)={*Overlap*, *Meet*}) or (*A completely Disjoint B* / *Disjoint*(I_A,I_B) and *Disjoint*(U_A, U_B))
Contains/Inside		For Contains: *A fairly Contains B* / *Contains*(I_A,I_B) and *Contains*(U_A, U_B) For Inside: *A fairly Inside B* / *Inside*(I_A,I_B) and *Inside*(U_A, U_B)
Covers/ Covered by		For Covers: (*A weakly Covers B* / *Contains*(I_A,I_B) and *Covers*(U_A,U_B)) or (*A fairly Covers B* / *Covers*(I_A,I_B) and *Covers*(U_A,U_B)) For Covered by: (*A weakly Covered by B* / *Inside*(I_A,I_B) and *Covered by*(U_A,U_B)) or (*A fairly Covered by B* / *Covered by*(I_A,I_B) and *Covered by*(U_A,U_B))
Overlap		(*A weakly Overlap B* / R(I_A,I_B)= {*Disjoint*, *Meet*} and *Overlap*(U_A,U_B)) or (*A fairly Overlap B* / *Overlap*(I_A,I_B) and *Overlap*(U_A,U_B))
Meet		(*A weakly Meet B* / R(I_A,I_B)= {*Disjoint*, *Meet*} and *Overlap*(U_A,U_B)) or (*A weakly Meet B* / *Disjoint*(I_A,I_B) and *Meet*(U_A,U_B)) or (*A fairly Meet B* / *Meet*(I_A,I_B) and *Meet*(U_A,U_B))

Figure 5.16. An adverbial approach to reduce the topological vagueness

5.8 Example of reducing the intra-level topological relationships vagueness in a spatial data warehouse

In the present example, we consider the case of population density transitions from the urban zones to rural ones. The urban-rural classification may be based on the population density, which often decreases progressively from the urban zones to rural ones. Generally, the urban planners are appointed to estimate the boundary of the urban zones. An urban zone can have heterogeneous geometries in different databases when the estimates are made by different experts. Therefore, it is not reliable to surround an urban zone using a linear boundary. Nonetheless, these spatial objects are falsely represented as regions with crisp boundaries that replace the real broad boundaries.

In this example, a spatial data warehouse (with a cube structure) is required to support a decision-making process in the domain of urbanism. Figure 5.17 shows a star schema of the spatial data warehouse. The fact table is connected to three dimensions. The spatial dimension is made up of four hierarchy levels: *Building_group*, *Urban_zone*, *Region*, *Country*. The temporal dimension contains three hierarchy levels: *Year*, *Five_years*, *Twenty_years*. The last dimension is called *Taxe_Category* and it is composed of one hierarchy level that describes the categories of required taxes (e.g. *provincial*, *federal*, *property tax*, *house tax*, *etc*). Only one measure, called *required_taxes*, is considered in this spatial data cube.

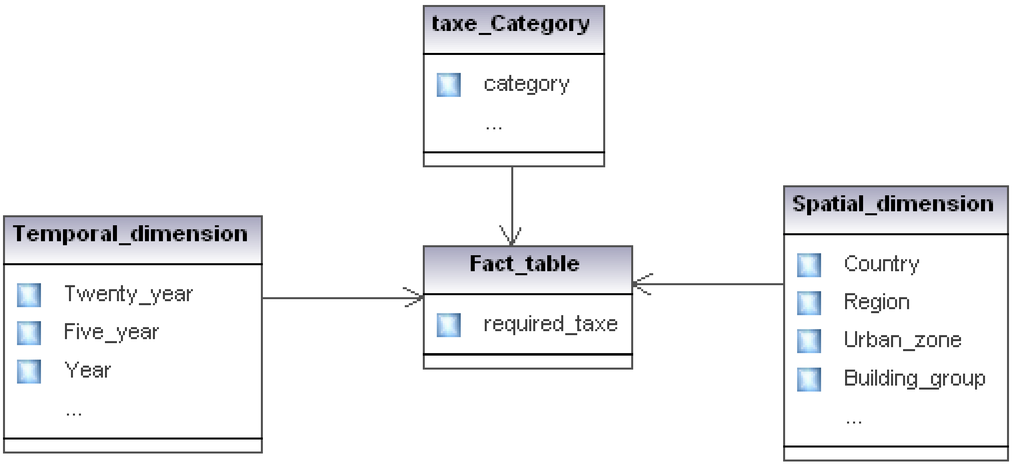


Figure 5.17. The star schema of a spatial data warehouse in the domain of urban planning

In this example, the urban zones are loaded from different data sources. For an urban zone *A*, we assume that the intersection between its source geometries is non-empty. The integration of source geometries of *A* generates a final geometry with a vague shape. We are interested in specifying final topological integrity constraints for geometries stored at the *Urban_zone* level of the spatial dimension.

We remind the hypotheses made in this example. First, the same topological relationships between urban zones are specified in the different data sources. Second, semantic heterogeneities are not considered in our specification of final constraints. Finally, we do not deal with inter-levels topological relationships.

Figure 5.18 shows an example of two urban zones A and B that are disjoint in their respective sources S_1, S_2, and S_3 but are represented by heterogeneous crisp geometries in each one. The same topological relationship is specified in the topological integrity constraints defined in the data sources: "*the geometries of two different urban zones should be disjoint*". The final geometries are regions with broad boundaries that overlap each other (the unions U_A and U_B overlap each other, whereas the intersections I_A and I_B are disjoint).

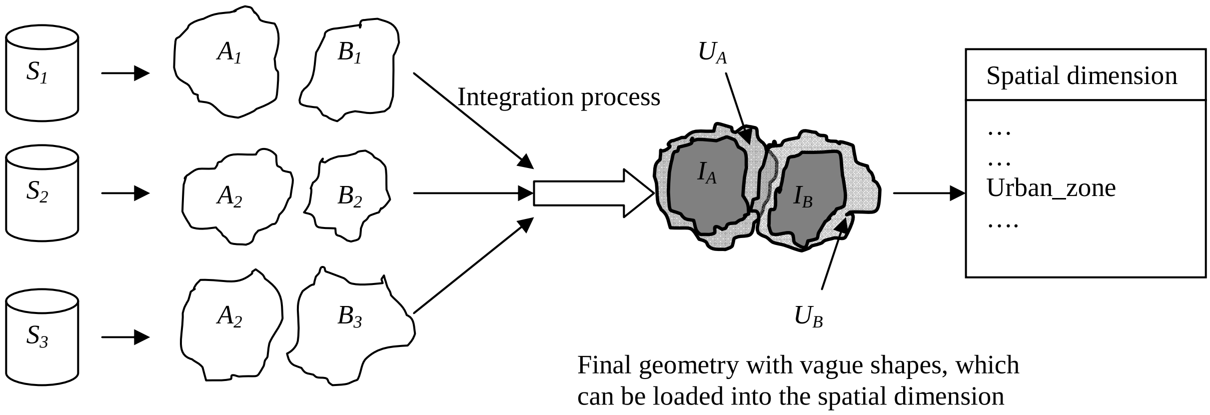


Figure 5.18. Integration of the heterogeneous source geometries of an urban zone

The source topological integrity constraints cannot be completely respected by the final geometries with vague shapes. Therefore, we use the two strategies defined above to reduce the topological relationships vagueness in the level *Urban_zone* of the spatial dimension. The following constraints are expressed using a spatial extension of OCL (Pinet *et al.* 2007, Bejaoui *et al.* 2008).

- Using the first strategy (modifying the final geometries):

In this example, the intersections are certainly disjoint because the source geometries respect this constraint. If we assume that the topological relationship is more important than objects shapes, then it is less risky to choose the intersections as the best spatial extents of the urban zones. In the other cases, the best extensions should be translated (see the first case in Figure 5.15) to verify the constraint before loading them in the *Urban_zone* level. The final topological integrity constraint is expressed as follows:

Context Urban_zone **inv**:

Urban_zone.allInstances → **forAll** (a, b| a<>b implies ***DISJOINT***(a,b));

- Using the second strategy (an adverbial approach to express final topological integrity constraints):

In this strategy, we consider that the shapes of objects are more relevant than the topological relationships in the decision-making process. In other words, decision-makers need to take into account the shape vagueness of urban zones in order to improve the quality of their decisions by considering the uncertainty of input data. Nonetheless, the vagueness of the topological relationships can be reduced by preventing impossible relationships between the final geometries. For this purpose, we use the second strategy.

According to the constraints defined in the data sources, the intersections should be disjoint at least. Then we use the topological operators "*weakly Disjoint*" and "*completely Disjoint*" to express this specification (see the first case in Figure 5.16). The unions should be *Disjoint*, *Overlap* or *Meet* each other. Otherwise, the final geometries are invalid and cannot be accepted in the spatial dimension. The final constraint is expressed as follows:

Context Urban_zone **inv**:

Urban_zone.allInstances → **forAll** (a, b| a<>b implies (*weakly Disjoint(a,b)*/ (*Disjoint*(I_a,I_b) and *R*(U_a,U_b)={*Overlap*, *Meet*})) OR (*completely Disjoint(a,b)*/ (*Disjoint*(I_a,I_b) and *Disjoint*(U_a,U_b))));

5.9 Conclusion

In spatial data integration, the spatial data quality can be used to compare the data sources in order to deal with geometric heterogeneities. When the data quality is distinctive, one data source can be selected as a reference in order to apply an overlay method and generate crisp geometries from the integration process. Otherwise, the spatial data integration consists of merging all of the source geometries that contribute equally in the final geometry of a given spatial object. In the latter case, the final geometries may be geometries with vague shapes (e.g. *regions with broad boundaries*) when there is a difference between the union and intersection of source geometries. The topological relationships between final geometries should be redefined in order to take into account their shape vagueness. In this paper, we studied the problem of topological relationships vagueness, defined as the uncertainty about the appropriate topological relationship between the final geometries. The main objective is to *reduce the topological relationships vagueness* between the final geometries. Table 1 explained the scope of this work. We limited our study to the case where the same topological relationship is specified in the source for the objects involved and where the data quality is not distinctive. For the other cases, additional investigations are required to deal with topological relationships vagueness.

This paper provides three main contributions. First, a model for objects with vague shapes (Bejaoui *et al.* 2008) has been reused to merge all the source geometries of a given object when the internal quality analysis is *non-distinctive*. Second, we have studied the valid topological relationships between the final geometries resulted from an integration process considering the case where the same *topological relationship* is found in the data sources. For each of the eight topological relationships (Egenhofer and Herring 1990) that can arise in the data sources, we studied which topological relationships can occur between the respective unions and intersections of the source geometries of objects involved. We proposed patterns of matrices to verify the validity of relationships between final geometries with vague shapes (cf. Section 5.7.2). Third, we proposed two main strategies to reduce the topological relationships vagueness: (1) *choosing the best extents of objects involved and modifying them if they violate a given topological relationship*, (2) *preserving the final geometries with vague shapes and using an adverbial approach to stress the partial satisfaction of a given topological relationship*. The first strategy can be used when the topological relationships are considered more important than the shapes of objects involved to meet the users' needs. The second strategy aims to preserve the possible vague shapes of final geometries and to partially satisfy the source topological relationships. These strategies were tested in an example of a spatial data warehouse (with a cube structure) in the domain of urbanism.

According to Malinowski and Zimányi (2005), topological relationships between hierarchy levels have been the focus of many works (e.g. Tryfona and Egenhofer 1997). However, neither the shape vagueness of the geometries involved in these relationships nor their implications in computing of measure aggregations were considered (Pedersen and Tryfona 2001, Jensen *et al.* 2004). In the future researches, we aim at studying these problems using the contributions of the present work.

A code generator could be also proposed and implemented. Such a generator could produce triggers or SQL queries from OCL constraints in order to check the validity of data in the data warehouses. The generated code will be used to control if the data comply with the topological conditions of the constraints. In order to reach this goal, a specific extension of the existing code generator OCL2SQL could be considered for data warehouses (Pinet *et al.* 2007).

References

BÉDARD Y. 1987, Uncertainties in Land Information Systems Databases. Proceedings of Eighth International Symposium on Computer-Assisted Cartography. Baltimore, Maryland (USA), 29 Mars - 3 Avril 1987, American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping, p. 175-184.

BÉDARD Y. RIVEST S. AND PROULX M.J. 2007, Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective, Dans: Robert Wrembel & Christian Koncilia (ed(s)), Data Warehouses and OLAP : Concepts, Architectures and Solutions, Chap. 13, IRM Press (Idea Group), London, UK, pp. 298-319.

BÉDARD Y. AND HAN J. 2008. Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery. Chap. 3 in: *Geographic Data Mining and Knowledge Discovery, 2nd edition*, Taylor & Francis, *in press*.

BEJAOUI L. PINET F. BÉDARD Y. SCHNEIDER M. 2008, Qualified topological relations between objects with possibly vague shape. In: International Journal of Geographical Information Sciences, Francis & Taylor, *to appear*.

BREUNIG M. AND A. PERKHOFF, 1992, Data and System Integration for Geoscientific Data, Spatial Data Handling (SDH), Charleston, pages 272-280.

BURROUGH P.A. AND FRANK A.U. 1996, Geographic Objects with Indeterminate Boundaries (London: Taylor & Francis).

CASADO M.L. 2006, Some Basic Mathematical Constraints for the Geometric Conflation Problem. In Proceedings of 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Edited by M. Caetano and M. Painho.

CLEMENTINI, E. 2005, A model for uncertain lines. *Journal of Visual Languages and Computing* 16(4): 271-288.

CLEMENTINI E. AND P. DI FELICE 1997, Approximate topological relations. International Journal of Approximate Reasoning, 16:173-204.

CHRISMAN, N.R. 1991, The error component in spatial data. In: Maguire, D.J., Goodchild, M.F and Rhind, D.W. (eds.), Geographical Information Systems: Principles and Applications. Volume 1, p. 165-174.

COCKCROFT S. 1997, A Taxonomy of Spatial Data Integrity Constraints. Geoinformatica **1**(4): 327-343.

COHN A.G. AND N.M. GOTTS 1996, The 'egg-yolk' representation of regions with indeterminate boundaries in: Burrough, P. & Frank, A. M. (editors) Proceedings of the GISDATA Specialist Meeting on Spatial Objects with Undetermined Boundaries, pp. 171-187 Francis Taylor.

DAVID B. FASQUEL P. 1997, Bulletin d'information de l'IGN - Qualité d'une base de données géographique: concepts et terminologie, N. 67, IGN France.

DEMIRKESEN A. AND B. SCHAFFRIN 1996, Map Conflation : Spatial point data merging and transformation, GIS/LIS'96, Denver, pages 393-404.

DEVILLERS R. AND R. JEANSOULIN 2005, "Qualité de l'information géographique: concepts généraux", *in: Lavoisier (Ed.), ``Qualité de l'information géographique'', Hermès Sciences traité IGAT, série Géomatique, Paris*, octobre.

DEVOGEL T. 1997, Processus d'intégration et d'appariement de bases de données géographiques: application à une base de données routière multi-échelle. Thèse de doctorat, soutenue le 12 décembre 1997.

DILO A. 2006, Representation of and reasoning with vagueness in spatial information: A system for handling vague objects. PhD thesis, ITC, Netherlands, 187p.

EGENHOFER M. AND J. HERRING 1990, A mathematical framework for the definition of topological relationships. Proceedings of the Fourth International Symposium on Spatial Data Handling, Zurich, Switzerland (eds. K. Brassel and H. Kishimoto), 803--813.

ERWIG M. AND M. SCHNEIDER 1997, Vague regions. In 5th International Symposium on Advances in Spatial Databases (SSD'97), number 1262 in Lecture Notes in Computer Science, pp. 298--320.

FRANK A. 1987, Overlay processing in spatial information systems, Auto Carto 8, Baltimore.

FRANKLIN C. 1992, An Introduction to Geographic Information Systems: Linking Maps to databases. Database. pp. 13-21.

HARVEY F AND F. VAUGLIN 1996, Geometric Match-processing : Applying Multiple Tolerances, Spatial Data Handling (SDH), Delft (Pays-Bas), Kraak et Molenaar (Eds.), pages 155-171.

JENSEN, C. KLYGIS A. PEDERSEN T. AND TIMKO I. 2004, Multidimensional data modeling for location-based services. *VLDB Journal*, 13(1):1-21.

LAURINI R. 1996, Raccordement géométrique de bases de données géographiques fédérées, Ingénierie des systèmes d'informations, vol.4, num. 3, pages 361-388.

LEE M.L. AND RAMAKRISHNAN R. 1999, "Integration of Disparate Information Sources: A Short Survey." ACM Multimedia.

MALINOWSKI E. ZIMÁNYI E. 2005, Spatial Hierarchies and Topological Relationships in the Spatial MultiDimER model. *Actes de British National Conf. on Databases, BNCOD22*, LNCS 3567, Springer, pp. 7-28.

MEGRIN 1996, Seamless Administrative Boundaries of Europe (SABE) Technical Overview, http ://www.ign.fr/megrin/sabe/sabedesc.exe.

MOSTAFAVI M.A. EDWARDS G. et *al*. (2004). An Ontology-Based Method for Quality Assessment of Spatial Data Bases. ISSDQ'04, GeoInfo Series, Austria.

MOWRER H.T. 1999, Accuracy (Re)assurance: Selling Uncertainty Assessment to the Uncertain. In *Spatial Accuracy Assessment, Land Information Uncertainty in Natural Ressources*, K. Lowell and Jaton A. (Ed.) (Quebec: Ann Arbor Press), pp. 3-10.

NORMAND P. 1999, Modélisation des contraintes d'intégrité spatiale, théorie et exemples d'applications. Département des sciences géomatiques. Québec, Université Laval.

NYERGES T. 1989, Schema integration analysis for the development of GIS databases, Int. J. Geographical Information Systems, vol.3, num. 2, pages 153-183.

PEDERSEN T. AND TRYFONA N. 2001, Pre-aggregation in spatial data warehouses. In *Proc. of the 7th Int. Symposium on Advances in Spatial and Temporal Databases*, pages 460-478.

PINET F. DUBOISSET M. AND V. SOULIGNAC 2007, Using UML and OCL to maintain the consistency of spatial data in environmental information systems. In: Environmental modeling & software, 22(8), pp. 1217-1220.

PFOSER D. TRYFONA N. AND JENSEN C.S. 2005, Indeterminacy and Spatiotemporal Data: Basic Definitions and Case Study. *GeoInformatica*, 9(3), pp. 211-236.

PULLIOT J., 2005, Intégration des données spatiales, Notes de cours, Département de sciences géomatiques, Université Laval, Automne.

REIS R. M.J. EGENHOFER AND J. MATOS 2006, Topological relations using two models of uncertainty for lines. In. Proc. of 7th international Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. 5 - 7 July 2006, Lisbon, Portugal, pp : 286-295.

RODRIGUEZ A. 2005, Inconsistency Issues in Spatial Databases. Lecture Notes In Computer Science 3300: 237-269.

SAALFELD A. 1993, Automated Map Conflation, Center for Automation Research, CAR-TR-670, (CS-TR-3066), University of Maryland, College Park.

SERVIGNE S. THIERRY U. PURICELLI A. AND LAURINI R. 2000, A Methodology for Spatial Consistency Improvement of Geographic Databases. GeoInformatica 4(1): 7-34.

SBOUI T. BÉDARD Y. BRODEUR J. BADARD T. 2007, A Conceptual Framework to Support Semantic Interoperability of Geospatial Datacubes, LNCS, Vol. 4802, pp. 378-387.

SHIBASAKI R. 1994, Handling Spatio-Temporal Uncertainties of Geo-Objects for Real-Time Update of GIS Databases from Multi-Source Data. In *Advanced Geographic Data Modeling*, Publications on Geodesy, Netherlands Geodetic Commission, Vol. 40:228-243.

SHEPHERD I. 1992, Geographic Information Systems, chapitre 22 : Information Integration and GIS, Maguire, Goodchild, Rhind (Eds.), Longman Scientific & Technical, pages 337-358.

SPACCAPIETRA S. AND C. PARENT 1991, Conflicts and Correspondence Assertions in Interoperable Databases, ACM SIGMOD RECORD, vol.20, num. 4.

TANG XI. 2004, Spatial object modeling in fuzzy topological spaces: with applications to land cover change. PhD thesis, University of Twente. ISBN 90-6164-220-5.

TRYFONA N. AND EGENHOFER M. 1997, Consistency among parts and aggregates: a computational model. *Transactions in GIS*, 4(3):189-206.

UBEDA T. AND M.J. EGENHOFER 1997, Topological Error Correcting in GIS. In the Proceedings of International Symposium on Spatial Databases.

UITERMARK H. T. VAN OOSTEROM P. J. M. MARS N. J. I. MOLENAAR M. 2005, Ontology-based integration of topographic data sets. In International Journal of Applied Earth Observation and Geoinformation, v. 7, iss. 2, p. 97-106.

WARE J.M. AND JONES C.B. 1998, Matching and aligning features in overlayed coverages. In proceedings of the 6th ACM international symposium on Advances in geographic information systems, p.28-33, November 02-07, Washington, D.C., United States.

WAREMER J. B. AND A. G. KLEPPE 1998, The Object Constraint Language: precise modeling with UML. Addison-Wesley Professional, 1st edition.

WORBOYS M.F. 1998, Imprecision in finite resolution spatial data. *GeoInformatica*, 2, pp. 257-279.

ZADEH L.A. 1965, Fuzzy sets. *Inform. Control*, vol. 8, pp. 338–353.

ZIEGLER P. AND DITTRICH K.R. 2004, Three Decades of Data Integration - All Problems Solved (2004). In 18th IFIP World Computer Congress (WCC 2004), Volume 12.

Chapter 6: An adverbial approach for the formal specification of topological integrity constraints involving regions with broad boundaries

Lotfi BEJAOUI, François PINET, Michel SCHNEIDER, Yvan BÉDARD

Accepted in 27th International Conference on Conceptual Modeling (ER 2008)

(Extended version of the paper appeared in the Lecture Notes in Computer Sciences proceeding edited by Springer and submitted to Geoinformatica)

6.1 Résumé de l'article

Dans les bases de données spatiales, les contraintes topologiques d'intégrité contrôlent les propriétés topologiques des objets spatiaux ainsi que la validité de leurs relations topologiques. Ces contraintes peuvent être exprimées en utilisant des langages formels tels que l'extension spatiale d'OCL (acronyme d'*Object Constraint Language*). OCL spatial permet l'expression des contraintes topologiques impliquant les objets spatiaux ayant des formes bien définis. Cependant, ce langage ne fournit pas les éléments de syntaxe requis pour exprimer des contraintes topologiques impliquant les objets spatiaux ayant des formes vagues (*ex. régions ayant des frontières larges*). Le vague de forme requiert des opérateurs

topologiques appropriés (ex. *fortement disjoint, faiblement adjacent*) pour désigner les relations valides entre ce type d'objets. Cet article adresse le problème de manque des éléments de syntaxe pour exprimer des contraintes topologiques impliquant des régions ayant des frontières larges. Nous proposons une extension d'*OCL spatial* basée sur notre modèle géométrique pour des objets ayant des formes vagues et une approche adverbiale pour des relations topologiques entre des régions ayant des frontières larges. Cette extension a été validée sur un exemple d'une base de données stockant des informations sur les épandages agricoles.

6.2 Abstract

Topological integrity constraints control the topological properties of spatial objects and the validity of their topological relationships in spatial databases. These constraints can be specified by using formal languages such as the spatial extension of the *Object Constraint Language* (OCL). Spatial OCL allows the expression of topological integrity constraints involving crisp spatial objects. However, topological integrity constraints involving spatial objects with vague shapes (e.g., *regions with broad boundaries*) are not supported by this language. Shape vagueness requires using appropriate topological operators (e.g., *strongly Disjoint, fairly Meet*) to specify valid relations between these objects; otherwise, the constraints cannot be respected. This paper addresses the problem of the lack of terminology to express topological integrity constraints involving regions with broad boundaries. We propose an extension of *Spatial OCL* based on the QMM model for objects with vague shapes and an adverbial approach for topological relations between regions with broad boundaries. This extension of *Spatial OCL* is then tested on a database storing data related to agricultural spreading.

6.3 Introduction

Integrity constraints are well-know techniques to guarantee the consistency of the data. According to Altman (1994), integrity constraints are rules that are dependent on a problem domain, and they must be held to be true for all meaningful states of information systems. The modeling of integrity constraints in a conceptual data model may be viewed as a representation of a set of business rules for the information system. The satisfaction of these constraints tends to guarantee the consistency and quality of data.

In geographical databases, the spatial integrity constraints are required to control the topological properties of geometries, the semantic aspects (e.g., a house has one floor at least), and topological relations (e.g., cultural parcels should be disjoint or adjacent) in addition to basic constraints (e.g., domain constraints) (Frank 2001, Souris 2006). The formal specification of these integrity constraints requires using an unambiguous language adapted to geographical databases. A spatial database-oriented language should allow the specification of both alphanumeric and spatial constraints (Duboisset et *al.* 2005).

The work of (Demuth and Hussmann 1999, Demuth et *al.* 2001) proposed to make use of the Object Constraint Language (OCL) (Warner and Kleppe 1999) to model alphanumerical database integrity constraints. OCL provides a framework to define constraints on UML class diagrams. This language has several advantages. First, it allows a declarative expression of constraints. Second, it is based on UML which is commonly used in the information system and software engineering domains. Third, it can be interpreted by code engines/compilers to generate integrity checking mechanisms automatically. Some tools allow producing code in different languages (Java, C#, SQL, etc) from specifications of constraints expressed in OCL (Klasse 2005). For instance, different tools can produce SQL code (Demuth 2005, Demuth et *al.* 2004). The produced SQL queries can be used to check if a database verifies the constraints or to forbid inserting data that do not verify a constraint.

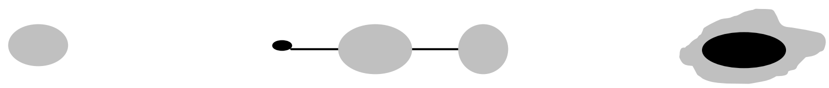
A recent extension of OCL called “Spatial OCL” has been proposed to model complex spatial integrity constraints (Hasenohr and Pinet 2006, Pinet et *al.* 2007). Currently, Spatial OCL cannot be used to define topological constraints involving objects with vague shapes. However, the shapes of many spatial objects are inherently vague (Bejaoui et *al.* 2008, Burrough and Frank 1996, Clementini and Di Felice 1997, Cohn and Gotts 1996, Dilo 2006, Reis et *al.* 2006, Tang 2004). This is the case of regions with broad boundaries (e.g. forest stand, pollution zone, valley or lake). In this paper, we propose a formalism based on Spatial OCL to model integrity constraints involving topological relations in databases storing vague objects. More precisely, we focus on regions with broad boundaries and we integrate the recent *Qualitative Min-Max* model (*QMM*) (Bejaoui et *al.* 2008) into Spatial OCL. In this model, a region with a broad boundary is composed of crisp and vague parts. The advantage of *QMM* is its capacity to represent regions with partially broad boundaries. The *QMM* is also very expressive in terms of topological relationships. 242 topological relationships have been distinguished between two regions with broad boundaries. An intuitive method based on adverbs is proposed in (Bejaoui et *al.* 2008) to term the relationships. This makes *QMM* adapted to query or constraint languages.

The paper is organized as follows. In section 2, we briefly review the concept of objects with vague shapes and we present the *QMM* model. In section 3, we review related works on the specification of spatial integrity constraints involving topological relationships. These constraints are termed topological (integrity) constraints in the present paper. In section 4, we introduce our extension of Spatial OCL. Section 5 presents a case study inspired of a spatial database storing information about agricultural spreading activities. Some spatial objects in this database could be represented by vague shapes. Their topological constraints are expressed using the proposed extension of Spatial OCL. Section 6 focuses on the implementation of the approach and Section 7 presents the conclusions of our work.

6.4 Objects with vague shapes in QMM model

6.4.1 Categorization of spatial objects with vague shapes

According to Erwig and Schneider (1997) and Hazarika et *al.* (2001), *shape vagueness* refers to the difficulty of distinguishing the shape of one object from its neighborhood. It is an intrinsic property of an object that has a spatial extent in a known position but does not have a well-defined shape (e.g., a pollution zone, a lake, a forest stand, etc.). According to the QMM model defined in Bejaoui et *al.* (2008) (cf. Chapter 3), we distinguish three basic types of *spatial objects with vague shapes*: *broad points*, *lines with vague shapes* (i.e., *lines with broad boundaries*, *lines with broad interiors* or *broad lines*), and *regions with broad boundaries*. Figure 6.1 shows an example of each one of these types of objects. A region has a vague shape when it is surrounded by a broad boundary instead of a sharp one (Figure 6.1(c)); we refer to these as *regions with broad boundaries* (e.g., a pollution zone). A line has a vague shape when its boundary (endpoints) and/or its interior are broad (Figure 6.1(b); e.g., the itinerary of an historic explorer). For lines, we make a distinction between *broad interior* and *broad boundary* as we consider them specializations of *linear shape vagueness* (cf. Chapter 4). This distinction is also useful for points because a point does not have a boundary; it is only composed of an interior. The shape of a given point corresponds to the elementary space portion, which refers to its interior (Figure 6.1(a)). A broad point arises when there is a difficulty to distinguish the punctual object from its neighborhood (e.g., a mountain peak). Principles of QMM model can be retrieved in Section 3.6. The original version of this section has been modified to reduce redundancy and improve the readability of the thesis manuscript.

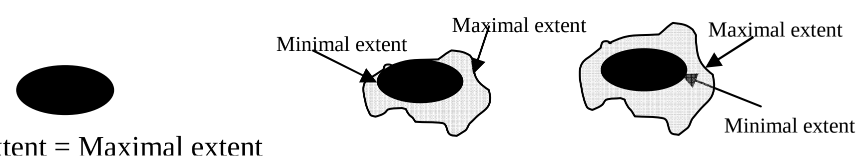


(a) A broad point (b) A line with a vague shape (c) A region with a broad boundary

Figure 6.1 Examples of objects with vague shapes

6.4.2 Regions with broad boundaries and their topological relations

In this paper, we define a region with a broad boundary according to the *QMM* model. A region with a broad boundary is then composed by two crisp sub-regions: (1) *a maximal extent* A_{max} (i.e., the representation of the region when the boundary is considered as far as possible) and (2) *a minimal extent* A_{min} *(*i.e., the representation of the region when the boundary is considered as close as possible). These two extents should are related by one of the following topological relations: *Equal*[8]*(* A_{min}, A_{max}*) or Contains(*A_{min}, A_{max}*) or Covers(*A_{min}, A_{max}) (Figure 6.2). The broad boundary refers to the difference between these two extents. This difference may include area everywhere around the minimal extent (i.e., regions with completely broad boundaries), may include area in some locations but not others around the minimal extent (i.e., regions with partially broad boundaries) or empty everywhere around the minimal extent (i.e., regions with no broad boundaries, or crisp regions). In figure 6.2(b), we present an example of a region with a partially broad boundary. The boundary is partially broad because the difference between the maximal extent and the minimal one is empty in some locations. Figures 6.2(a) and 6.2(c), represent an example of a crisp region and another one of a region with a completely broad boundary, respectively.



Minimal extent = Maximal extent

(a) A crisp region (b) A region with a partially broad boundary (c) A region with a completely broad boundary

Figure 6.2 Regions with broad boundaries

[8] The spatial relations (i.e., *Equal*, *Contains*, *Covers*) used in this definition are those defined in (Egenhofer and Herring 1990).

In this paper, the $QMM_{Topological\ relations}$ for regions with broad boundaries (cf. section 3.7) has been used to identify topological relationships between spatial objects concerned by a topological integrity constraint.

6.5 Specification of topological integrity constraints in spatial databases

Topological (integrity) constraints are defined as rules, which control the validity of topological relations between objects in spatial databases. They may be also viewed as spatio-semantic constraints, in the sense given by Bejaoui et *al.* (2007) and Salehi et *al.* (2007). In this Section, we study the formal expression of topological constraints.

6.5.1 OCL

The Object Constraint Language (OCL) is a subset of the well-known Unified Modeling Language (UML) that allows specifying constraints over entities representing concepts from the application domain (Warner and Kleppe 1999, OMG 2007). OCL constraints are defined on UML diagrams. OCL was first developed by a group of IBM's scientists around 1995 during a business modeling project. It was influenced by Syntropy that is an object-oriented modeling language that makes heavy use of mathematical concepts (Cook and Daniel 1994). OCL is supported by the Object Management Group and its role is important in the Model Driven Architecture approach (Kleppe and Warner 2003). OCL is used to specify invariants, i.e. conditions that "must be true for all instances of a class at any time" (Schmid et *al.* 2002). In the context of databases, an important advantage of OCL is due the fact that constraints are expressed in declarative manner at a conceptual level. OCL integrates notations close to a spoken language to express constraints. It is easier for database users to express the integrity constraints using OCL than SQL.

OCL provides a platform-independent and generic method to model constraints. It can be interpreted by compilers to generate code automatically. Some tools allow producing integrity checking mechanisms in different languages (Java, C#, SQL, etc) from specifications of constraints expressed in OCL (Klasse 2005). For example, OCL2SQL can generate SQL code from OCL constraints (Demuth and Hussmann 1999, Demuth et *al.* 2001, Demuth et *al.* 2004, Demuth 2005). The produced code can be used to check if a database verifies the constraints or to forbid inserting data that do not verify a constraint (Demuth and Hussmann 1999, Demuth et *al.* 2001).

Let us consider a class *Agricultural_Parcel* in a spatial database. The declaration of the class is *Agricultural_Parcel(id: Integer, shape: Region, surface_area: Real)*. Some of these parcels may have no spatial representation stored in the database. In this case, the value of the attribute *shape* is equal to *NULL*. The following OCL constraint models that the surface area of a parcel is greater than 0 if a spatial representation is available for this parcel:

```
context Agricultural_Parcel inv:
self.shape→notEmpty() implies self.surface_area > 0
```

In OCL constraints, self always represents an instance of a class. This class is specified in "context". An OCL constraint defines a condition that must be true for each instance of the class, i.e. for each value of self. Thus the above constraint specifies a condition that must be true for each instance of Agricultural_Parcel; self.shape and self.surface_area are attributes of self. The OCL function *notEmpty*() returns true if self.shape has a value and false otherwise. The operator “implies” corresponds to the logical implication.

6.5.2 Spatial OCL

Some tools and methods have been proposed to model visually spatial integrity constraints (Cockcroft 1997, Cockcroft 1998, Servigne et *al.* 2000, Borges et *al.* 2001, Cockcroft 2001, Cockcroft 2004, Parent et *al.* 2006, Raffaeta et *al.* 2008); their goal is to enable end-users to specify simple constraints thanks to specific GUI and different visual representations. They provide very interesting possibilities to end-users but they cannot be used to model complex constraints (e.g., topological constraints depending on complex conditions (Kang et *al.* 2004)).

In order to define complex spatial integrity constraints, Kang et *al.* (2004), Duboisset et *al.* (2005) and Pinet et *al.* (2007) proposed an extension of OCL meta–model. This extension called Spatial OCL adds geographic basic types (i.e., *point*, *line*, and *region*) into the OCL meta-model - see Figure 6.3. These spatial types are generalized through an abstract type called *BasicGeoType*. Topological constraints between simple regions can be expressed through Spatial OCL; this language integrates spatial functions based on Egenhofer's relationships between simple regions into OCL. The general syntax of these Spatial OCL functions is:

(*A*) .*EgenhoferTopologicalRelation* (*B*) : *Boolean*

Thus, *EgenhoferTopologicalRelation* can be: *disjoint*, *contains*, *inside*, *equal*, *meet*, *covers*, *coveredBy*, *overlap*. *A* and *B* are the parameters of the functions, i.e. the two simple regions to compare. These operations return true or false depending on whether the topological relation between *A* and *B* is true or false. The following example of Spatial OCL constraint illustrates the use of the proposed functions.

Let *Road* and *Building* be two classes; these two classes have a *shape* attribute. The topological constraint « buildings and roads *should not overlap each other* » is specified as follows in Spatial OCL:

```
context Road inv:
Building.allInstances()→forAll( b|
     self.shape .disjoint (b.shape)  or
     self.shape .meet     (b.shape) )
```

In OCL, the function `C.allInstances()` returns a collection that contains all the instances of a class `C`. Consequently, `Building.allInstances()` returns a collection that contains all the instances of the *Building* class. The OCL operation `forAll` corresponds to the universal quantifier. In the constraint, `self` is an instance of *Road* class, i.e. an instance of the context. The semantics of the constraint is "For each instance `self` in *Road* and for each instance *b* in *Building*, the shapes of `b` and `self` must be disconnected or must meet each other."

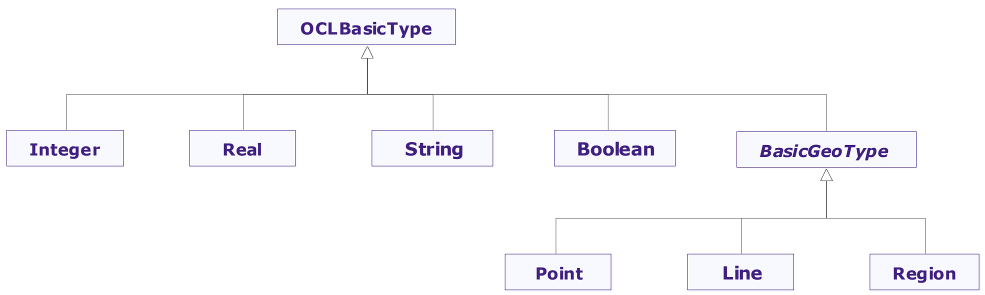


Figure 6.3 Extension of the meta–model of OCL proposed in (Kang et *al.* 2004).

6.6 Adverbial spatial OCL for Objects with vague shapes (AOCL$_{OVS}$)

As seen in previous sections, the shapes of *RBB* are more complex than those of crisp ones and their topological relations should be identified differently. Then, topological constraints cannot be specified in the same way as for crisp regions. Additional OCL extensions are required to deal with topological constraints for *RBB*. For example, how can we express a topological constraint, which specifies that "two zones should be *completely disjoint* or *fairly meet* each other"? We need more *tolerant* topological functions than those currently used in Spatial OCL.

Hereafter, we propose an extension of the Spatial OCL in order to support the formal expression of topological constraints between *RBB*. We call this extension Adverbial spatial OCL for *Objects with vague shapes* (AOCL$_{OVS}$ for short). For that, we integrate the specifications of *QMM* spatial model defined for objects with vague shapes into the meta–model of Spatial OCL. Moreover, we integrate our adverbial approach into a set of new functions of Spatial OCL in order to express the *strength* of topological relations specified in a constraint.

We propose to distinguish two abstract subclasses of geometries generalized by *BasicGeoType* in the meta-model of Spatial OCL: a type for *Objects with vague shapes* (*OVSType*) and another one for *Objects with Crisp Shapes* (*OCSType*). *OVSType* is a generalization of three basic types of objects with vague shapes: *broad point*, *line with a vague shape* and *region with a broad boundary*. These additional geometric basic types are defined according to the *QMM* model. Then, a *RBB* is composed by two crisp polygons (i.e., this relation is expressed through aggregations between the object type *Region with a broad boundary* and the object type *Region*), which respectively represent the minimal extent and the maximal extent of the object. Figure 6.4 shows a general extension of the Spatial OCL meta–model, which covers three basic types of objects with vague shapes. Hereafter, we focus on topological constraints for only regions with broad boundaries.

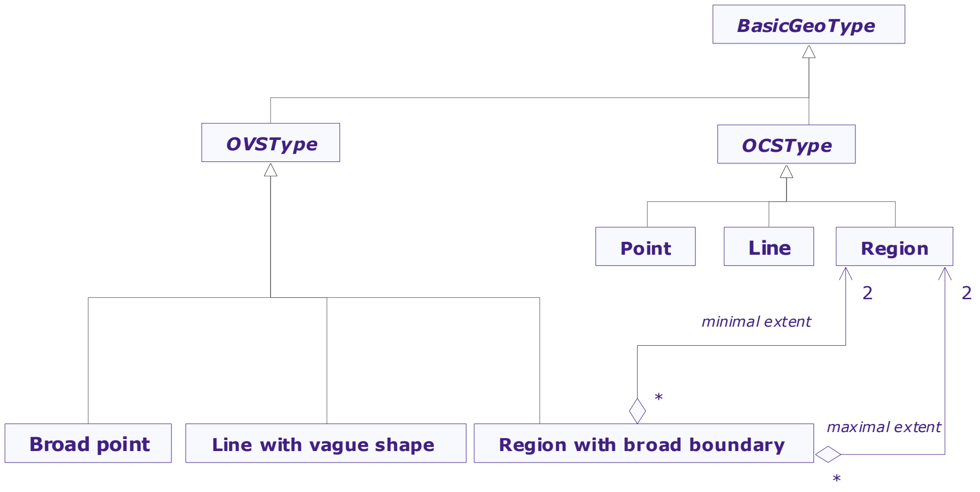


Figure 6.4 Extension of the meta–model of Spatial OCL

As presented in Section 2.2., the qualitative approach of the *QMM* model permits to model a relation between *RBB* by an Egenhofer's relation associated to an adverb (*weakly*, *fairly*, *strongly*, *completely*). The proposed Spatial OCL extension introduces new topological functions adapted to *RBB*. The general syntax of these new Spatial OCL functions is:

(*A*) .*Adverb_EgenhoferTopologicalRelation* (*B*) : *Boolean*

Thus, *EgenhoferTopologicalRelation* can be: *disjoint*, *contains*, *inside*, *equal*, *meet*, *covers*, *coveredBy*, *overlap*. *Adverb* can be *weakly*, *fairly*, *strongly*, *completely*. *A* and *B* are the parameters of the functions, i.e. the two objects having the *Region with broad boundary* type. These functions return true or false depending on whether the topological relation between *A* and *B* is true or false.

Note that an object having the *Region with broad boundary* type is considered valid when it verifies the next conditions:

- Each one of the minimal extent and maximal extent verifies the closeness and connectedness conditions of a simple crisp region.
- The minimal and maximal extents of a region with a broad boundary are related by one of the following topological relations: *Contains* (*max*, *min*), *Covers* (*max*, *min*) or *Equal* (*max*, *min*) (cf. section 2.2).

These last conditions are the *invariants* of the spatial model. We call these invariants *meta-constraints*, which control the validity of *RBB*.

6.7 Example in agricultural spreading activities

To illustrate the practical use of our extension of Spatial OCL, we introduce a case study related to an environmental information system for the traceability of agricultural spreading activities. Agricultural spreading activities consist of putting an organic substance *on* or *into* the soil in order to improve its agricultural productivity. In France, this activity is strictly controlled by public organizations, because the substances used in spreading may be dangerous for ecological systems whether they are not reasonably applied (Pinet et *al.* 2007, Pinet et *al.* 2009). The quantities and types of substances allowed in agricultural spreading activities depend on several criteria such as the parcel emplacement and soil type.

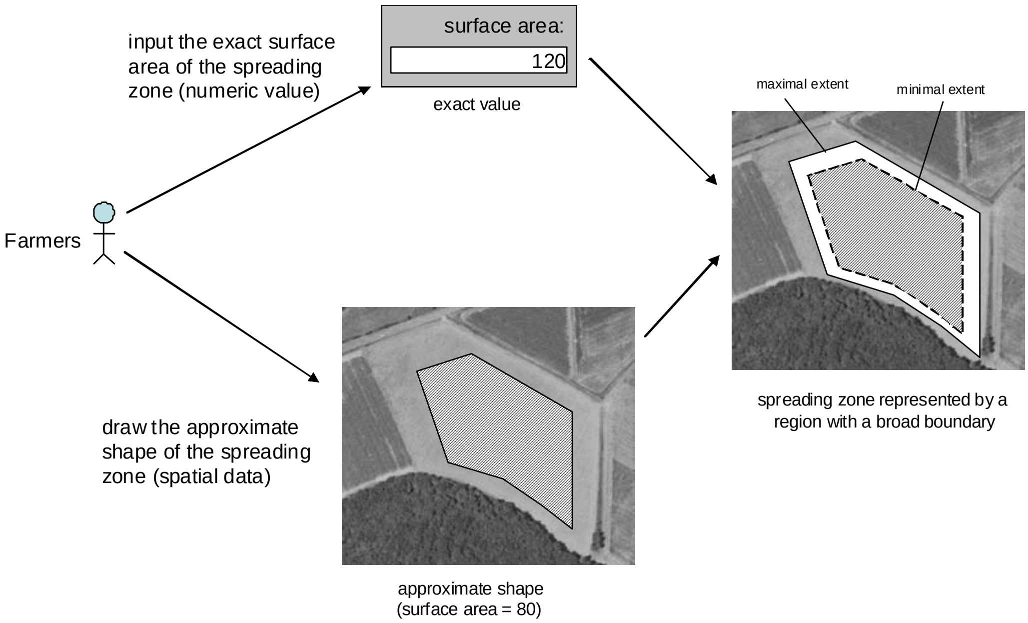


Figure 6.5 Example of *RBB* deduction – in the present case the exact surface area is greater than the surface area of the drawn shape

In France, the farmers should declare the areas to be spread (i.e. the spreading parcels) thanks to a Web-based tool (i.e., they declare an *outline for the geographical area to be spread*). These data are stored into a national spatial database (Pinet et *al.* 2007, Pinet et *al.* 2009). In practice, the farmers use the Web-based tool to input a numeric value indicating the surface areas of parcels before approximately drawing their respective geometries on a map through a GIS-based interface. The surface areas indicated by farmers are generally calculated thanks to expertise of land parcels. While declared surface areas are considered exact, the geometries drawn by farmers only provide approximate information about the location of spreading. The surface areas of drawn geometries are also computed by a GIS-based tool.

They are generally different from those declared by farmers. It could be possible to deduce a spreading parcel with a broad boundary (i.e., a *RBB*) from the drawn geometry and the surface area indicated by farmers. The crisp part of this *RBB* is the zone where spreading is considered as certain, and its vague part is the zone where spreading is uncertain. The surface area of the *RBB* should be equal to the declared surface area. Figure 6.5 provides an example of produced RBB.

A spreading parcel may include several capacity zones, which correspond to subparts of the parcel where the spreading is allowed with conditions (e.g., preserve the soil quality). The approximate geometry of capacity zones is also drawn by farmers thanks to the Web-based application; consequently they can be also represented by *RBB*. A spreading perimeter is an area that includes all the parcels of a farm. Figure 6.6 shows a spreading parcel and a capacity zone both represented by *RBB*.

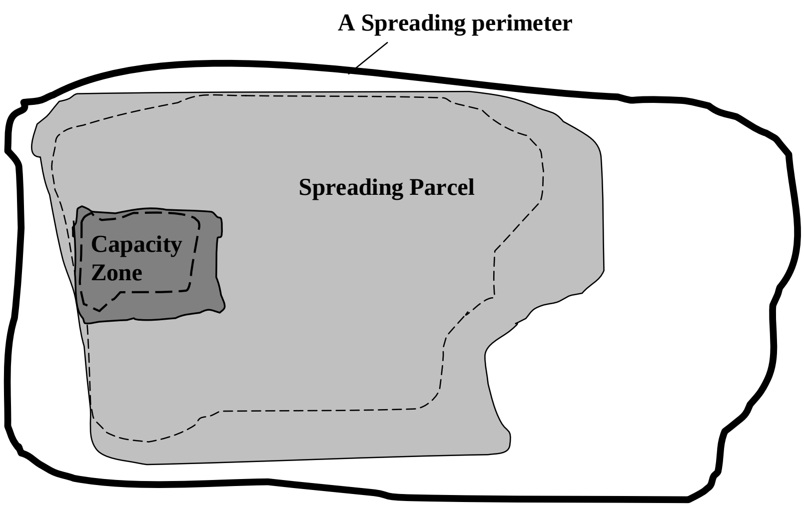


Figure 6.6 Spreading perimeter, spreading parcel, and capacity zone

Figure 6.7 presents the conceptual model of our example. The class *Parcel* refers to spreading parcels. A parcel is described by an identifier, a declared surface area, a surface area computed from the drawn geometry (*Draw_area*) and a *RBB*. Capacity zones are also represented by *RBB*.

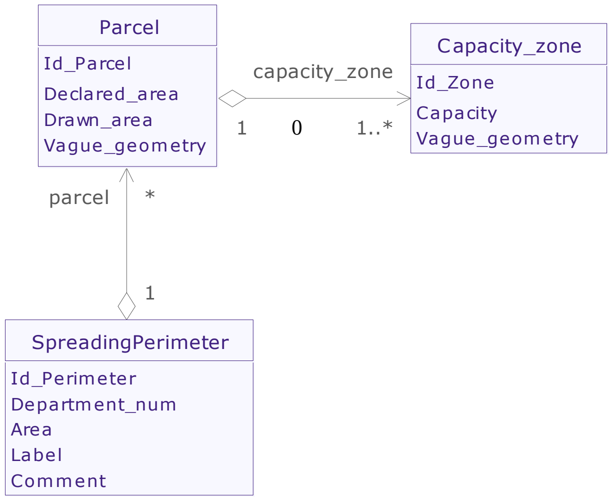


Figure 6.7 Conceptual model

6.7.1 Formal expression of constraints

We present a set of spatial constraints expressed in $AOCL_{OVS}$. They mainly concern the spreading parcels and their capacity zones.

Constraint 1:

The spreading parcels of farmers should be disjoint or meet each other. In the present example, a parcel is represented by an object with a vague shape. The topological relation between two vague parcels is valid, when it belongs to one of the following relations:

- *completely Disjoint* (i.e., occurs when both minimal and maximal extents are disjoint, respectively),
- *completely Meet* (i.e., occurs when both minimal and maximal extents meet each other, respectively),
- *strongly Disjoint* and *weakly Meet* (i.e., occurs when maximal extents meet each other whereas minimal extents are disjoint, respectively), or
- *fairly Disjoint* and *fairly Meet* (i.e., occurs when maximal extents meet each other, minimal extents are disjoint and one of the minimal extents meets one the maximal extents).

The context of this topological constraint is the class *Parcel*. The constraint is formally expressed as follows:

```
Context Parcel inv:
Parcel.allInstances() → forAll (b| self<>b implies
self.vague_geo  .completely_Meet      (b.vague_geo) or
self.vague_geo  .completely_Disjoint (b.vague_geo) or
(self.vague_geo .strongly_Disjoint    (b.vague_geo) and
self.vague_geo  .weakly_Meet          (b.vague_geo)) or
(self.vague_geo .fairly_Disjoint      (b.vague_geo) and
self.vague_geo  .fairly_Meet          (b.vague_geo)))
```

Constraint 2:

A spreading parcel is composed by one or several capacity zones. The geometry of a capacity zone is drawn by the farmer after drawing the parcel's geometry. A capacity zone is then inside, covered by or equal to the drawn geometry of the parcel involved. The relations that should be respected between the respective RBB of a parcel and each one of its capacity zones, are: *completely Contains*, *completely Covers*, (*strongly Contains* and *weakly Covers*), (*strongly Contains* and *weakly Overlap*), (*fairly Contains* and *fairly Covers*), (*fairly Contains* and *weakly Covers* and *weakly Overlap*), (*strongly Covers* and *weakly Contains*), (*fairly Contains* and *fairly Covers*) or (*strongly Covers* and *weakly Overlap*). The constraint can be specified declaratively as follows:

```
Context Parcel inv:
self.vague_geo→ forAll (b| self.capacity_zone → exists(d|
(b.vague_geo .completely_Contains (d.vague_geo)) or
(b.vague_geo .completely_Covers     (d.vague_geo)) or
(b.vague_geo .strongly_Contains     (d.vague_geo) and
b.vague_geo  .weakly_Covers         (d.vague_geo)) or
(b.vague_geo .strongly_Contains     (d.vague_geo) and
b.vague_geo  .weakly_Overlap        (d.vague_geo)) or
(b.vague_geo .fairly_Contains       (d.vague_geo) and
b.vague_geo  .fairly_Covers         (d.vague_geo)) or
(b.vague_geo .fairly_Contains       (d.vague_geo) and
b.vague_geo  .weakly_Covers         (d.vague_geo) and
b.vague_geo  .weakly_Overlap        (d. vague_geo)) or
(b.vague_geo .strongly_Covers       (d.vague_geo)  and
b.vague_geo  .weakly_Contains       (d.vague_geo)) or
(b.vague_geo .fairly_Contains       (d.vague_geo)  and
b.vague_geo  .fairly_Covers         (d.vague_geo)) or
(b.vague_geo .strongly_Covers       (d.vague_geo)  and
b.vague_geo  .weakly_Overlap        (d.vague_geo))))
```

The OCL operation exists expresses the existential quantifier. The subexpression self.capacity_zone returns a collection that contains all the capacity zones associated to the parcel self.

Constraint 3:

Inside a spreading parcel, two different capacity zones should verify one of the following relations: *completely Disjoint*, *completely Meet*, (*strongly Disjoint and weakly Meet*) or (*fairly Disjoint and fairly Meet*). The context of this topological constraint is the class *Capacity_zone*. The constraint is then formally expressed as follows:

```
Context Capacity_zone inv:
Capacity_zone.allInstances()
→ forAll (a | a<>self and a.parcel=self.parcel implies
a.vague_geo  .completely_Meet      (self.vague_geo)  or
a.vague_geo  .completely_Disjoint (self.vague_geo)  or
(a.vague_geo .strongly_Disjoint     (self.vague_geo) and
a.vague_geo  .weakly_Meet           (self.vague_geo)) or
(a.vague_geo .fairly_Disjoint     (self.vague_geo) and
a.vague_geo  .fairly_Meet         (self.vague_geo)))
```

The subexpression self.parcel returns the parcel associated to the capacity zone self.

Constraint 4:

Let *P* be a spreading perimeter composed by *N* spread parcels. The sum of areas of minimal extents of spread parcels is inferior or equal to the area of *P*. However, the sum of areas of maximal extents of spreading parcels is superior or equal to the declared area of *P*. The constraint is expressed as follows:

```
Context SpreadingPerimeter inv:
self.parcel.vague_geo.minimal_extent.area→sum() ≤ self.area          and
self.parcel.vague_geo.maximal_extent.area→sum() ≥ self.area
```

The subexpression self.parcel.vague_geo.minimal_extent.area → sum() provides the sum of areas of minimal extents of parcels belonging to the spreading perimeter involved (i.e., this function makes the same thing for maximal extents of capacity zones in one spread parcel).

6.7.2 Implementation of AOCL$_{OVS}$

We developed a prototype to automatically generate SQL queries from the AOCL$_{OVS}$ expressions. More precisely, we extended OCL2SQL developed by TU Dresden University. This tool has been extended by Duboisset et *al.* (2005) and Pinet et *al.* (2007) to express the topological constraints involving crisp regions. The code generator is a Java application. The constraints are defined on an UML class diagram that is stored in an *xmi* file. They are written using AOCL$_{OVS}$. Our extension of OCL2SQL translates these constraints in Oracle SQL using new topological operators implemented in the database. Each topological operator defined in the *QMM* model is implemented as a SQL spatial operator that refers to a PL-SQL function. The PL-SQL function verifies if the concerned relationship is respected by the geometries of objects involved. For example, the $AOCL_{OVS}$ operation `completely_Disjoint` corresponds to a SQL operation (that has the same name) implemented by a specific PL-SQL functions.

All the *RBB* of a same database are stored in a single table termed *VAGUE_GEO*. The other tables of the database can have an attribute called *vague_geo* that references the primary key of *VAGUE_GEO*. The attribute *geo_max* has the type *MDSYS.SDO_Geometry* and stores the maximal extent of the object. The attribute *geo_min* is used to store the minimal extent. When a topological operator (e.g. '*completely Disjoint*') is executed for two given objects, a PL-SQL function compares their minimal (*geo_min*) and maximal (*geo_max*) extents. The SQL expression below shows the definition of the *VAGUE_GEO* table.

```
Create table VAGUE_GEO
(   PK_VG       NUMBER(10)  primary key
,   GEO_MAX MDSYS.SDO_Geometry
,   GEO_MIN   MDSYS.SDO_Geometry
);
```

To illustrate the generation of SQL code we introduce an example concerning pollution zones. The SQL expression below shows the definition of the *POLLUTION_ZONES* table. The attribute *geometry_pk_vg* is the foreign key that references *VAGUE_GEO.pk_vg*.

```
Create table POLLUTION_ZONES
(   PK_PZ              NUMBER(10)  primary key
,   DESCRIPTION          VARCHAR2
,   GEOMETRY_PK_VG    NUMBER(10)
);
```

The constraint 5 models that two pollution zones should be strongly disjoint.

Constraint 5:

Context Pollution_zones **inv**: Parcel.allInstances()→**forAll** (b| self<>b implies self.vague_geo ***.strongly_Disjoint*** (b.vague_geo)

The SQL query generated by OCL2SQL for this constraint is presented below. This query selects all the rows that violate the $AOCL_{OVS}$ constraint. Thus this SQL query can be executed by the users of a spatial database in order to retrieve possible inconsistencies.

```
Select * from OV_Pollution_Zone SELF
Where not (not exists ( (select PK_ PZ from Pollution_Zone) minus
      Select PK_ PZ from Pollution_Zone SELF2  Where (SELF.PK_PZ = SELF2.PK_PZ) OR
         stronglyDisjoint((select PK_VG from VAGUE_GEO
            Where PK_VG IN (Select GEOMETRY_PK_VG From
                                                Pollution_Zone where PK_PZ = SELF2.PK_PZ)),
                                                (Select PK_VG from VAGUE_GEO Where PK_VG in
                                                (Select GEOMETRY_PK_VG From Pollution_Zone
                                                Where PK_PZ = SELF2.PK_PZ)) , VAGUE_GEO)=0 ));
```

Figure 6.8 schematizes the architecture of the extension of OCL2SQL, which covers topological constraints involving regions with broad boundaries. This Figure is adapted from (Duboisset et al. 2005). Other platforms (MySQL, SQL Server, etc.) could be considered in the future.

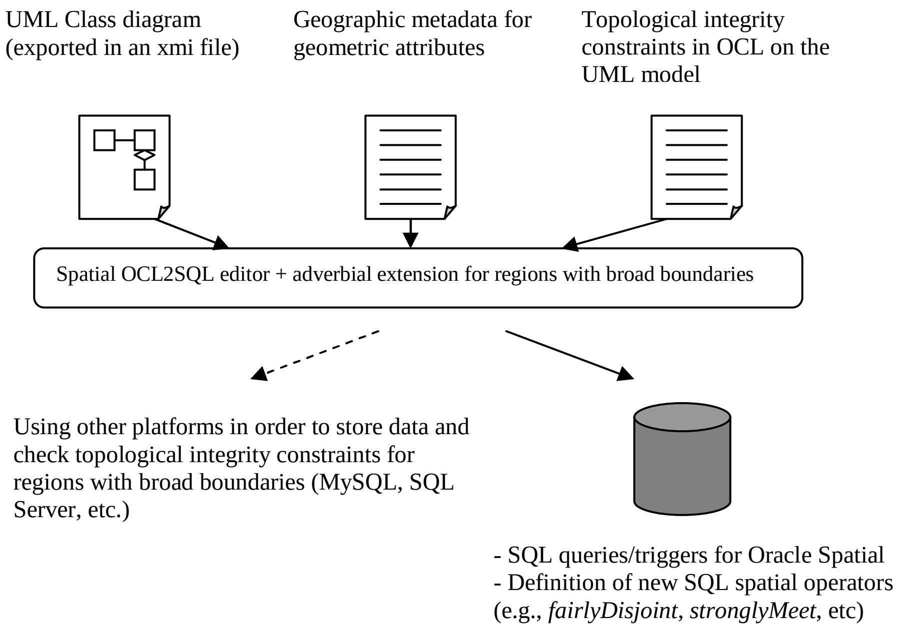


Figure 6.8 Architecture of the application used to check the OCL constraints (this figure is adapted from (Duboisset et *al.* 2005))

6.8 Conclusion

Controlling topological constraints is an important aspect of the spatial data quality. Visual tools and methods proposed in (Cockroft 1997, Cockroft 1998, Servigne et *al.* 2000, Borges et *al.* 2001, Cockroft 2001, Cockroft 2004, Parent et *al.* 2006, Raffaeta et *al.* 2008) enable end-users to easily specify simple constraints but they cannot be used to model complex spatial constraints (e.g., topological constraints depending on complex conditions (Kang et *al.* 2004). As presented in (Duboisset et *al.* 2005, Pinet et *al.* 2007), complex topological constraints can be expressed through Spatial OCL which integrates the Egenhofer's relations. This language provides easiness in the specification of formal constraints in UML class diagrams.

However, Spatial OCL assumes that objects are represented using crisp geometries whereas they can have vague shapes (e.g. a pollution zone, the itinerary of an historic explorer, etc.). Spatial OCL lacks syntactical tools to express the topological constraints for objects with vague shapes. In this paper, we addressed the problem of the formal specification of topological constraints for regions with broad boundaries. It contributes in two main directions.

First, the meta-model of Spatial OCL has been extended in order to consider new object types covering spatial objects with vague shapes. We proposed a new abstract type called *OVSType* (Object with Vague Shape Type), which can be specialized into: *broad point*, *line with a vague shape*, and *region with a broad boundary*. The adverbial approach for topological relations presented in (Bejaoui et *al.* 2008) has been integrated into Spatial OCL; new topological functions are proposed in this language. We called this extension *Adverbial spatial OCL for Objects with Vague Shapes* ($AOCL_{OVS}$).

Second, $AOCL_{OVS}$ has been implemented into OCL2SQL. This extension allows generating Oracle SQL code from $AOCL_{OVS}$ constraints. The generated SQL queries control the consistency of spatial databases. These queries are executed by the database administrators to detect possible inconsistencies. The main objective of this $AOCL_{OVS}$ implementation was to show the feasibility of our approach. Some constraints of the case study presented in Section 5 have been used to experiment our extension. These constraints principally concern spreading parcels and their capacity zones both represented by regions with broad boundaries.

In future, we will generalize our framework in order to specify topological relations between different objects with vague shapes (i.e., broad points, lines with vague shapes, and regions with broad boundaries). We will also study the specification of topological constraints involving complex regions with vague shapes (i.e. regions with several kernels, regions composed by several sub-regions with broad boundaries, etc.).

The syntax of $AOCL_{OVS}$ could be also simplified by grouping the adverbs that concern the same topological relations. For instance, the following constraint:

```
"self.vague_geo .strongly_Disjoint (b.vague_geo) or
self.vague_geo  .weakly_Disjoint   (b.vague_geo)"
```

It could be expressed more directly as follows:

“self.vague_geo .**{strongly|weakly}Disjoint** (b.vague_geo)”

In this case, it is needed to introduce additional OCL operators in order to group adverbs.

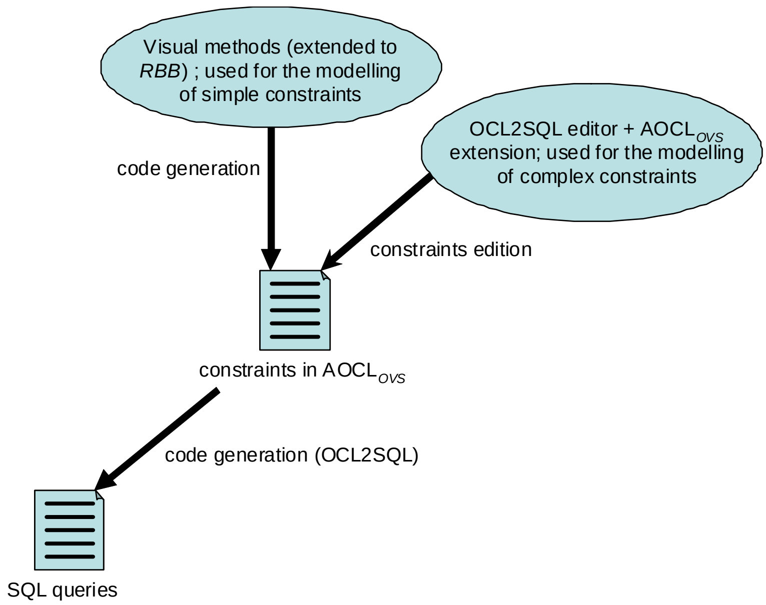


Figure 6.9 Combination of different tools to generate the SQL code

AOCL$_{OVS}$ and the extension of OCL2SQL are intended to computer scientists. This approach can be used jointly with other existing methods to specify the spatial constraints. For example, the simple constraints could be specified with user-oriented methods such as those presented in (Cockroft 1997, Cockroft 1998, Servigne et *al.* 2000, Borges et *al.* 2001, Cockroft 2001, Cockroft 2004, Parent et *al.* 2006, Raffaeta et *al.* 2008) before being translated into AOCL$_{OVS}$ expressions. The user-oriented methods are very efficient to visually and easily model simple constraints. Complex constraints may be directly specified using AOCL$_{OVS.}$ For that purpose, the user-oriented methods should be preliminary extended. They should cover the *RBB* and generate AOCL$_{OVS}$ constraints .Figure 6.9 illustrates this solution.

It could be also possible to generate triggers (with OCL2SQL) that are executed automatically with each update of the databases (Demuth and Hussmann 1999, Demuth et *al.* 2001). Difficulties of performances may be observed in the case of large spatial databases. In our opinion, optimizing the generated code requires an in-depth study.

References

ALTMAN D. 1994, Fuzzy set theoretic approaches for handling imprecision in spatial analysis. International Journal of Geographical Information Systems 8:271-289.

BEJAOUI L. BÉDARD Y. PINET F. SAHELI M. AND SCHNEIDER M. 2007, Logical consistency for vague spatiotemporal objects and relations. In: 5th International symposium on spatial data quality (ISSDQ 2007), Enschede, NLD, 13-15 June, p 8.

BEJAOUI L. PINET F. BÉDARD Y. SCHNEIDER M. 2008, Qualified topological relations between spatial objects with possibly vague shape. to appear in International Journal of Geographical Information Science

BORGES K.A.V. DAVIS C.A. AND LAENDER A.H.F. 2001, OMT-G: An Object-Oriented Data Model for Geographic Applications. GeoInformatica 5:221-260

BROWN D.G. 1998, Classification and boundary vagueness in mapping presettlement forest types. International Journal of Geographical Information Science 12:105-129

BURROUGH P.A. 1989, Fuzzy mathematical methods for soil survey and land evaluation. Journal of Soil Science 40:477-492

BURROUGH P.A. AND FRANK A.U. 1996, Geographic objects with indeterminate boundaries.

CLARAMUNT C. 2000, Extending Ladkin's algebra on non-convex intervals towards an algebra on union-of regions. In: Proceedings of the 8th ACM international symposium on Advances in geographic information systems. ACM, Washington, D.C., United States

CLEMENTINI E. AND DI FELICE P. 1997, Approximate Topological Relations. International Journal of Approximate Reasoning 16:173-204

COCKCROFT S. 1997, A Taxonomy of Spatial Data Integrity Constraints. GeoInformatica 1:327-343

COCKCROFT S. 1998, User Defined Spatial Business Rules: Storage, Management and Implementation – A Pipe Network Case Study. In: 10th Colloquium of the Spatial Information Research Centre, University of Otago, Dunedin, New-Zealand, 16-19 novembre, pp. 73-81

COCKCROFT S. 2001, Modelling Spatial Data Integrity Rules at the Metadata Level. In: 6th International Conference on GeoComputation, Brisbane, Australia, 2001 september 24-26.

COCKCROFT S. 2004, The Design and Implementation of a Repository for the Management of Spatial Data Integrity Constraints. GeoInformatica 8:49-69.

COHN A.G. BENNETT B. GOODAY J. AND GOTTS N.M. 1997, Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. GeoInformatica 1:275-316.

COHN A.G. AND GOTTS N.M. 1996, The 'Egg-Yolk' Representation of Regions with Indeterminate Boundaries. In: GISDATA Specialist Meeting on Spatial Objects with Undetermined Boundaries. Taylor & Francis, p 171-187.

COOK S. AND DANIELS J. 1994, Designing object systems-object oriented modeling with Syntropy. Prentice-Hall

DEMUTH B. 2005, The Dresden OCL Toolkit and the Business Rules Approach. In: European Business Rules Conference (EBRC2005), Amsterdam.

DEMUTH B AND HUSSMANN H. 1999, Using UML/OCL Constraints for Relational Database Design. In: «UML»'99 — The Unified Modeling Language. p 751-751

DEMUTH B. HUSSMANN H. AND LOECHER S. 2001, OCL as a Specification Language for Business Rules in Database Applications. In: «UML» 2001 — The Unified Modeling Language. Modeling Languages, Concepts, and Tools. p 104-117

DEMUTH B. LOECHER S. ZSCHALER S. 2004, Structure of the Dresden OCL Toolkit. In: 2nd International Fujaba Days "MDA with UML and Rule-based Object Manipulation", Darmstadt, Germany, September, pp. 15 - 17

DILO A. 2006, Representation of and reasoning with vagueness in spatial information: A system for handling vague objects. In:Wageningen University and ITC, p 187.

DUBOISSET M. PINET F. KANG M.A. AND SCHNEIDER M. 2005, Precise modeling and verification of topological integrity constraints in spatial databases: from an expressive power study to code generation principles. Lecture Notes in Computer Science 3716:465-482

EGENHOFER M. HERRING J. 1990, A mathematical framework for the definition of topological relations. In: Fourth International Symposium on Spatial Data Handling, Zurich, Switzerland, pp. 803-813

ERWIG M. AND SCHNEIDER M. 1997, Vague regions. In: Advances in Spatial Databases, pp 298-320.

FRANK. A.U. 2001, Tiers of ontology and consistency constraints in geographical information systems. International Journal of Geographical Information Science 15:667-678.

HASENOHR P. AND PINET F. 2006, Modeling of a spatial DSS template in support to the Common agricultural policy. Journal of decision systems 15:181-196.

HAZARIKA S. AND COHN A. 2001, Qualitative Spatio-Temporal Continuity. In: Spatial Information Theory, pp. 92-107.

HWANG S. THILL J-C. 2005, Modeling Localities with Fuzzy Sets and GIS. In: Fuzzy Modeling with Spatial Information for Geographic Problems, pp. 71-104.

KANG M.A. PINET F. SCHNEIDER M. CHANET JP. AND VIGIER F. 2004, How to design geographic database? Specific UML profile and spatial OCL applied to wireless Ad Hoc networks. In: 7th Conference on Geographic Information Science (AGILE'2004), Heraklion, GRC, April 29-May, pp. 289-299.

KLASSE O. 2005, OCL Tools and Services Web Site. In: <http://www.klasse.nl/ocl>.

KLEPPE A. AND WARMER J. 2003, Object Constraint Language, the Getting your Models Ready for MDA. Addison-Wesley

OMG 2007, Unified Modelling Language: OCL, version 2.0. OMG Specification In:

PARENT C. SPACCAPIETRA S. ZIMANYI E. 2006, Conceptual Modeling for Traditional and Spatio-temporal Applications. Springer

PINET F. DUBOISSET M. DEMUTH B. SCHNEIDER M. SOULIGNAC V. BARNABE F. 2009, Constraints modeling in Agricultural Databases. In: Advances in Modeling Agricultural Systems. Springer

PINET F. DUBOISSET M. AND SOULIGNAC V. 2007, Using UML and OCL to maintain the consistency of spatial data in environmental information systems. Environmental modelling & software 22:1217-1220.

RAFFAETA A. CECCARELLI T. CENTENO D. GIANNOTTI F. MASSOLO A. PARENT C. RENSO C. SPACCAPIETRA S. AND TURINI F. 2008, An application of advanced spatio-temporal formalisms to behavioural ecology, GeoInformatica 12:37-72.

RANDELL D.A. AND COHN A.G. 1989, Modelling topological and metrical properties of physical processes. In: 1st International Conference on Principles of Knowledge Representation and Reasoning (KR'89), p 357–368.

REIS R. EGENHOFER M.J. AND MATOS J. 2006, Topological relations using two models of uncertainty for lines. In: 7th international Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 5 - 7 July, Lisbon, Portugal, pp. 286-295.

SALEHI M. BÉDARD Y. MIR A.M. AND BRODEUR J. 2007, Classification of integrity constraints in spatiotemporal databases: toward building an integrity constraint specification language, Report.

SCHMID B. WARMER J. AND CLARK T. 2002, Object Modeling with the OCL: the Relational Behind the Object Constraint Language. Springer.

SCHNEIDER M. 2001, A Design of Topological Predicates for Complex Crisp and Fuzzy Regions. In: Conceptual Modeling — ER 2001, pp. 103-116

SERVIGNE S. UBEDA T. PURICELLI A. AND LAURINI R. 2000, A methodology for spatial consistency improvement of geographic databases, GeoInformatica 4:7-34.

SOURIS M. 2006, Contraintes d'intégrité spatiales. In: Devillers R, Jeansoulin R (eds) Qualité de l'information géographique. Lavoisier, pp. 100-123.

TANG XI. 2004, Spatial object modeling in fuzzy topological spaces: with applications to land cover change. In: University of Twente.

WARMER J. AND KLEPPE A. 1999, The Object Constraint Language Precise Modeling with UML. Addison-Wesley

Chapter 7: Conclusions and discussion

The shape vagueness is considered as an inherent property of some spatial objects such as lakes, pollution zones, forest stands, etc. This type of imperfection can also result from merging heterogeneous and crisp redundant geometries that describe the same spatial object in different source databases.

The representation of spatial objects with vague shapes requires using specific spatial models in order to stress the vagueness of topological invariants such as broad boundaries for regions. This thesis proposes a general approach to represent spatial objects with partially or totally vague shapes and their topological relationships (Chapters 3 and 4). The spatial model is also used to study the topological relationships vagueness that arises between geometries with vague shapes that result from an integration process (Chapter 5). Then, it is integrated into the Spatial Object Constraint Language (Chapter 6) in order to express topological constraints involving regions with broad boundaries.

7.1 Contributions

The main contributions of this research work are presented in four papers which refer to Chapters 3, 4, 5 and 6 of this thesis.

Chapter 3 proposes an exact spatial model to represent spatial objects with vague shapes. Three basic *types of spatial objects with vague shapes* have been defined: *broad point, line with a vague shape (i.e., lines with broad boundaries, lines with broad interiors or broad* *lines),* and *region with a broad boundary*. In the proposed model, the shape vagueness relates to the difference between the certain and uncertain knowledge about the appropriate shape of a given spatial object. From this perspective, an object with a vague shape is defined as a minimal extent $\tilde{A}_{\min}$ (the object geometry including space points that *certainly* belong to the object) and a maximal extent $\tilde{A}_{\max}$ (includes space points that *possibly* belong to the object) that respect some topological conditions. The difference between the minimal extent and maximal one can be empty (*objects with well-defined shapes*), empty in some locations and non-empty in some others (*objects with partially vague shapes*) or non-empty everywhere (*objects with completely vague shapes*). The advantage of this model is that spatial objects with partially vague shapes are represented whereas they are considered as invalid in the existing models of (e.g. Clementini and Di Felice 1997, Tang 2004, Reis *et al.* 2006). Then, the topological relationships between spatial objects with vague shapes are identified using a 4-Intersection matrix that enumerates four sub-relations: R_1 ($\tilde{A}_{\min}, \tilde{B}_{\min}$), R_2 ($\tilde{A}_{\min}, \tilde{B}_{\max}$), R_3 ($\tilde{A}_{\max}, \tilde{B}_{\min}$), and R_4 ($\tilde{A}_{\max}, \tilde{B}_{\max}$). We distinguished 242 relations between regions with broad boundaries (cf. appendix 1). In order to retain our propositions useful in practice, we classify these topological relationships into eight basic clusters using the contents of their respective matrices. We use four adverbs strengths to describe the membership to a given cluster: *completely*, *strongly*, *fairly*, and *weakly*. This model is termed *Qualitative Min-Max* (QMM) model.

Chapter 4 focused on the shape vagueness of lines and the identification of their topological relationships. Then, two components of the QMM model are proposed: (1) the QMM_{Def} model and QMM_{TR} model. The QMM_{Def} model proposes an expressive taxonomy of lines with vague shapes and their formal definitions. In this taxonomy, we make the distinction between the shape vagueness of the interior and boundary of a given line. For each topological invariant, shape vagueness can be partial or total. The line boundary can be *partially* or *completely* broad while the boundary remains well-defined, and vice versa. We identify four levels of shape vagueness for lines according to the *crispness, partial broadness and complete broadness* of the interior and/or boundary: *weakly*, *fairly*, *strongly* and *completely*. In this chapter, lines are defined according to the principles of the QMM model set in Chapter 3. We define a line with a vague shape as a minimal extent composed by only one-dimensional parts and a maximal extent that additionally includes two-dimensional (or broad) parts. The topological relationships between lines with vague shapes are then identified through an extension of the CBM method (Clementini and Di Felice 1995) that we integrate into the QMM_{TR} model and apply to compute the sub-relations between minimal and maximal extents of the lines involved. Then, a 4-Intesersection matrix is proposed to describe these four sub-relations and to classify the topological relationships between lines with vague shapes.

In Chapter 5, we are interested in a vertical integration where heterogeneous and redundant crisp geometries that represent the same object, in different data sources, are intended to be integrated and loaded in a final database. In this study, we assume that the data quality is poorly described in the data sources and can be used neither to choose geometries with best quality nor to identify the appropriate topological relationships in the final database. Geometries with vague shapes can then result from the integration because source geometries are heterogeneous and contribute in an equal way in the final geometry of a given object. Consequently, for a same set of objects, the topological relationships between their final geometries cannot be identified to those defined for crisp geometries in the data sources. Therefore, we address the problem of *topological relationships vagueness*, i.e. the uncertainty about the appropriate topological relationships between the final geometries. Accordingly, we aim at *reducing* the topological relationships vagueness in a given final database. For this purpose, Chapter 5 contributes in two main directions. First, heterogeneous and redundant crisp geometries that represent a given same object, in different source databases, are merged using the QMM model for regions with broad boundaries. The broad boundaries of final regions result from the difference between the union and intersection of the source geometries involved. Second, we propose a method to deduce the valid topological relationships between them. In this method, we assume that the same topological relationship is defined between the objects involved in source databases. This assumption is required to allow the reasoning about topological relationships between the final geometries of the same collection of objects in the final database. For example, let's assume that the geometries of two spatial objects *A* and *B* are respectively disjoint in two data sources. Then the intersections of the source geometries of *A* and the source geometries of *B* are necessarily disjoint. *Contains* is an inconsistent relationship between the final geometries of *A* and *B*. Then, for each topological relationship of the 9-Intersection model (Egenhofer and Herring 1990), we define patterns of matrices that specify the valid relationships between the unions and intersections of the source geometries of objects involved, respectively (section 5.7.1). The patterns matrices are used to reduce the topological relationships vagueness through two main strategies: (1) *choosing the best extents of concerned objects and modifying them whether they violate the recommended topological* *relation*, (2) *preserving the geometries with vague shapes and using an adverbial approach to stress the partial respect of a given topological relation*. The first strategy can be used when the topological relationships are considered as more important than the shapes of objects to meet the users' needs. The second strategy is more appropriate to preserve possible vague shapes of final geometries that partially respect source topological relationships.

The results obtained in Chapter 5 can be very useful to deal with geometric heterogeneities in the context of spatial data warehouses (especially those with a hypercube structure). The spatial dimension of a spatial data warehouse is generally loaded from different sources that have different specifications. Our approach proposes to represent the final geometry of a given spatial object using geometries with vague shapes while the source crisp geometries are heterogeneous and have a same quality level. Such approach allows the decision-maker to distinguish between the certain and uncertain data and to consider the shape vagueness in his decision. An example of a spatial data warehouse in the urban planning domain is presented to illustrate the contributions of chapter 5.

Chapter 6 proposes an extension of Spatial OCL for regions with broad boundaries and their topological relationships. First, we extend the meta-model of Spatial OCL in order to consider new geometric types covering objects with vague shapes. Then, the geometry of an object with a vague shape is defined as a new abstract type called *OVSType* (*Object with a vague shape Type*), which can be specialized into *broad point*, *line with a vague shape*, and *region with a broad boundary*. Second, the topological constraints involving regions with broad boundaries are specified using the QMM model defined in Chapter 3. We integrate forty new topological operators as additional keywords of Spatial OCL. These topological operators refer to the forty clusters distinguished in the QMM model for regions with broad boundaries. We term this extension *Adverbial spatial OCL for Objects with vague shapes* (*AOCL$_{OVS}$* for short). Third, we integrate AOCL$_{OVS}$ in the constraint editor *OCL2SQL* (Duboisset 2007). Then, the SQL query that implements a topological constraint (in the physical level of the database) can be automatically generated from the AOCL$_{OVS}$ expression. An example of agricultural spreading database is presented in order to show the possibilities to express topological constraints involving regions with broad boundaries. This example has been inspired from the existing application called SIGEMO used to control the traceability of agricultural spreading activities in France (Soulignac *et al.* 2005).

7.2 Discussion

This thesis provides a general qualitative approach to deal with spatial objects with vague shapes and their topological relationships. We propose this approach in the context of controlling topological consistency of such objects and of their topological relationships. The general hypothesis made in this work is: **it is possible to provide an approach that supports the specification of topological integrity constraints involving spatial objects with vague shapes and of their topological relationships, both in transactional spatial databases and in spatial data warehouses**. This hypothesis requires a specific spatial model to represent different levels of shape vagueness and evaluate the vagueness of a topological relationship. Therefore, we proposed an adverbial approach to express the topological constraints involving regions with broad boundaries using an extension of Spatial OCL. We think that the general hypothesis has been verified in this thesis work.

The QMM model is principally inspired from the Egg-Yolk model (Cohn and Gotts 1996). However, there are some fundamental differences between our model and that defined in (Cohn and Gotts 1996). First, the sub-relations described in the 4-Intersection matrix of the Egg-Yolk theory (Cohn and Gotts 1996) are those defined in the RCC-5 model (Randell and Cohn 1989, Cohn *et al.* 1997) whereas we use those defined in the 9-Intersection model (Egenhofer and Herring 1990). In addition, the same methodology is used to identify topological relationships between objects with vague shapes. However, our definitions of this type of objects are substantially different because they are based on the point-set topology. Then, points and lines are also considered as basic crisp spatial object types. Moreover, the concept of '*broad boundary*' is not redefined in our model as it is done in most of existing exact models. In our approach, shape vagueness of a given object refers to the difference between its minimal extent and maximal one. Finally, the topological relationships are organised into a hierarchical classification based on the content of their respective matrices. This classification is the basis of an adverbial approach that we use to specify the topological constraints between regions with broad boundaries.

In (Clementini and Di Felice 1997), the notion of broad boundary has been used to replace linear (or well-defined) boundary. According to Clementini and Di Felice (1997), 44 topological relations are distinguished between two regions with broad boundaries using an extension of the 9-Intersection model (Egenhofer and Herring 1990). These relations have classified into 17 clusters and organised into a conceptual neighborhood graph that shows their similarity degrees (Clementini and Di Felice 1997). The main advantage of this approach is the ability to support a coarser spatial reasoning involving regions with broad boundaries. However, it becomes more difficult to use this model when the needs are more specific. Furthermore, the identification of a broad boundary as a two-dimensional topological invariant requires respecting the consistency conditions related to closeness and connectedness. Tang (2004) decomposed the broad boundary into the *boundary's interior* and *boundary's boundary*. He distinguished 152 topological relationships presented as variants of the 44 ones defined in (Clementini and Di Felice 1997). Nonetheless, many topological relationships cannot be identified because there is no distinction between the boundaries of *minimal extent* and those of the *maximal extent*. Moreover, spatial objects with partially vague shapes such as regions with partially broad boundaries cannot be presented in existing exact models. In this thesis, we resolved this problem by considering a simple region with a broad boundary as a general concept that can be specialized into*: regions with none broad boundary* (or crisp regions in fuzzy approaches terminology), *regions with a partially broad boundary* and *regions with a completely broad boundary*.

With regards to the principal exact models (Clementini and Di Felice 1997, Cohn and Gotts 1996, Erwig and Schneider 1997, Tang 2004), we made the distinction between the *partial shape vagueness* and *complete shape vagueness* to deal with two main problems: an ontological problem and a modeling one. The ontological problem occurs because "*shape vagueness*" is generally considered as a "*binary imperfection*" (an object shape can be *well-defined* or *vague*). However, spatial objects can be characterized by different levels of shape vagueness that can be easily computed in fuzzy models by using a *quantitative* approach. In this thesis, the shape vagueness levels are categorized using a *qualitative* approach because we believe that "*shape vagueness*" is a qualitative problem. In this context, we denote that the computation of shape vagueness provide coarse values contrary to evaluation based on fuzzy models. Different levels of shape vagueness are qualitatively distinguished using a set of adverbs (*completely*, *weakly*, *fairly*, etc.). We do not claim that exact models are better than fuzzy ones, because the needs are not identical and therefore the direct comparison is not appropriate.

The modeling problem refers to the difficulty of existing exact models (notably (Clementini and Di Felice 1997, Cohn and Gotts 1996, Erwig and Schneider 1997, Tang 2004)) to represent spatial objects with partially vague shapes and their topological relationships. For example, a region can have well-defined boundaries on one side and broad ones on the other side at the same time. In this work, we made the distinction between the regions with *partially broad boundaries* and those with *completely broad boundary*. Topological relationships between regions with broad boundaries have been classified into eight overlapping basic clusters. This adverbial classification supports the specification of topological constraints involving spatial objects with vague shapes. Nonetheless, it is important to denote that the QMM model is not able to quantify the gradual change inside the maximal extent in the same way as the fuzzy approaches done in (Zhan 1997, Schneider 2001, Du *et al.* 2005, Dilo 2006, Verstraete *et al.* 2007).

7.3 Future researches

This thesis provides a qualitative approach to represent spatial objects with vague shapes and reduce their topological relationship vagueness. This sets a starting point for future research projects that we present in the next paragraphs:

- ***Modeling complex spatial objects with vague shapes and their topological relationships***

  In this thesis, we studied shape vagueness for simple objects types: simple regions, simple lines and simple points. In the practice, complex spatial objects may also have vague shapes such regions with broad boundaries and holes, regions with several cores, regions composed by disjoint uncertain sub-regions, lines with several start broad points, etc. Currently, the QMM model does not cover this type of objects and their topological relationships. Studying this type of objects requires additional investigations that exceed the objectives of this thesis. Extending the present approach to model the complex objects with vague shapes is one of our future researches. Our methodology consists in generalizing the principles of the QMM model for complex objects with vague shapes by verifying appropriate conditions for each component of the object's shape involved.

- ***Considering topological relationships between objects with vague shapes and different dimensions***

  Topological relationships studied in this research are those between spatial objects with vague shapes having the same dimension. We studied relationships between simple regions with broad boundaries as well as those between lines with vague shapes. We also showed that our approach can be applied for objects with different dimensions such as topological relationships between a region with a broad boundary and a line with a partially vague shape. However, additional investigations are required to study specificities of these relationships and to propose a method to classify them.

- ***Studying the temporal vagueness***

  In many applications such as the management of forest stands, the temporal information is generally required in order to follow the existence of spatial objects and their geometric evolution. Temporal data may be vague, difficult to be collected and represented. For example, the birthday of an historic person and the construction period of a monument are often poorly known. Dyreson and Snodgrass (1993) distinguished four sources that affect the perfection about the dimension (i.e. *an interval* or *an instant*) of a time event as well as its location on the time axis: *granularity, dating techniques, future planning and unknown/imprecise time events.* The temporal vagueness has been studied in several works (e.g. (Dreyson and Snodgrass 1993, Pfoser and Tryfona 2001)). One perspective of the present work is to extend the QMM model in order to represent time events with vague temporal dimensions and/or vague locations. We are specifically interested in the partial temporal vagueness. For example, a time period can be bounded by a vague start time point on one side and a well-defined final one on the other side. We are also interested in the identification of topological relationships between vague temporal primitives using the same qualitative approach defined in the context of spatial objects with vague shapes. We look for an adverbial approach that can help to express topological constraints involving spatio-temporal objects with vague shapes and/or vague temporalities.

- ***Considering vagueness in the definition of topological relationships***

  In spatial databases, a topological relationship has a definition given by the spatial model (e.g. the 9-Intersection model) or by the model-maker. A topological relationship has also an extension that refers to its instance for two spatial objects stored in the database. In this thesis, we studied the vagueness of a topological relationship because it depends on the shape vagueness of objects involved. However, the definition can also be vague while the shapes of spatial objects involved remain well-defined. For example, it is possible to define a topological relationship called "*weakly meet*" that can arise between two crisp objects. In this case, objects weakly meet each other if the intersection between their boundaries occurs in three points at most. A vague topological relationship can be also associated to a quantitative function which returns its strength according to the definition and not to the shape vagueness of objects involved. In our future researches, we aims at studying this type of vagueness for topological relationships such as the metric (e.g. *far*, *close*) and directional (e.g. *in the north of*, *in the south of*) relationships.

- ***Coupling quantitative and qualitative approaches***
  Qualitative approaches are generally simple to be used and provide a coarse evaluation of vagueness. These approaches can be the base of an intuitive interface to communicate the vagueness to the users of spatial databases and GIS. However, the quantitative approaches provide a fine computation of vagueness using specific mathematical theories such as Fuzzy Logic (Zadeh 1965) or Rough sets (Pawlak 1994). For example, they can model the gradual changes of shape vagueness inside a broad boundary. We think that it is possible to couple these approaches in only one framework where the qualitative aspects are placed at its high level and quantitative ones in the bottom level. For example, it is possible to implement vagueness adverbs (e.g weakly, fairly) by using fuzzy sets in a lower level. The user may have the choice to use the qualitative approach or to drill-down in the vagueness detail by using the values provided by the membership functions. Such a framework provides the easiness of qualitative approaches and the precision of quantitative ones.

- ***Considering the shape and semantic vagueness in topological relationships between different level of a spatial dimension in a spatial data warehouse***
  In Chapter 5, we studied topological relationships vagueness at the level of final geometries with vague shapes resulted from the integration of heterogeneous and redundant source geometries. In spatial data warehouses, final geometries can be stored in different hierarchy levels (e.g. *country*, *region*, *county*) of a spatial dimension. One perspective of the present work is to consider the topological *inter-levels* relationships vagueness that can arise between the final geometries belonging to different hierarchy levels of the spatial dimension. In this context, the topological relationships vagueness affects the measure aggregations. For example, how to compute the required taxes for a given object with a vague shape that is partially contained in different members belonging to the immediately higher hierarchy level?

According to Malinowski and Zimányi (2005), the topological relationships between hierarchy levels are the focus of several works such as (Tryfona and Egenhofer 1997). However, neither the shape vagueness of geometries which can be involved in these relationships nor their implications on the computation of measure aggregations are considered (Pedersen and Tryfona 2001, Jensen et *al.* 2004).

- ***Extending $AOCL_{OVS}$ to support the specification of other types of spatial objects with vague shapes***

  $AOCL_{OVS}$ provides syntactic tools to express the topological constraints involving regions with broad boundaries. In our future researches, we look for extending this language in order to express the constraints involving lines with vague shapes, spatial objects with vague shapes having different dimensions as well as objects with complex vague shapes. We think that the same adverbial approach can be used to express the strength of topological relationships between these types of objects. However, these relationships will be termed by considering the type of objects involved.

- ***Testing the approach in other domains and for other uses***

  In the future researches, we aim at testing the present approach in other domains such land cover/land use, urbanism, forestry, pollution, climatic changes, erosion of beaches, etc. The same spatial model may be used to represent the shape vagueness of spatial objects in these domains. In the same way, we preview to develop a framework in order to express spatial queries for objects with vague shapes and their topological relationships (see example in section 3.9). This framework can be easily implemented using the existing prototype *OCL2SQL* where the spatial SQL queries are automatically generated.

7.4 General conclusion

According to the general objective set in the beginning of this work, we develop a spatial model that supports different types of objects with different levels of shape vagueness. The vagueness of topological relationships is stressed using a set of adverbs that are integrated in an existing integrity constraint language. This language is Spatial OCL that we have extended to support the specification of topological integrity constraints on objects with vague shapes in a transactional database. The proposed approach is also used to deal with the problem of topological relationships vagueness in the context of a vertical integration with redundant source geometries. In the latter case, we propose two strategies to reduce the uncertainty about the appropriate topological relationships between final geometries resulting from an integration process, both based on the same spatial model proposed in the first phase. Then, the spatial model has been integrated in an existing DBMS and the constraint language is easily implemented in an existing editor of integrity constraints OCL2SQL.

Nevertheless, it is important to denote that the proposed approach is not perfect; i.e. it does not resolve all the problems related to the modeling of spatial objects with vague shapes. The first problem is that the shape vagueness cannot be directly computed through a measurement device. Some computational functions (such as that we applied to deduce the broad boundary for the spreading agricultural parcels) should be applied on the initial data in order to deduce the shape vagueness. The computation of shape vagueness should be preceded by a strong study to build required functions that correctly use the input data to meet the need of computing vagueness. Otherwise, the shape vagueness is wrongly computed and serious risks of a degradation of spatial data quality could appear. Our approach does not provide a solution to this problem since we assume that the appropriate functions to compute vagueness are defined.

The present approach is also developed in the context of a feature-oriented view of spatial phenomena. In other words, the space is coarsely subdivided into three parts: a first one that certainly belongs to the object, a second that may belong to the object and a third that is certainly outside the object. However, an extension (coupling with a quantitative approach) of the approach is required to provide a fine computation of shape vagueness using a field-oriented view of space. In the latter case, the fuzzy and probabilistic models are more advantageous.

Furthermore, the number of topological operators (forty) used to express the topological relationships between regions with broad boundaries, in our approach, is high with regards to the most of existing GIS and spatial DBMS that generally propose eight topological operators at most to express the same relationships between crisp regions. Additional investigations are then required to allow an implementation in existing software intended to meet different needs of users with different skills. In addition, the proposed approach can be used to deal with topological relationships vagueness in a specific case of integration where different hypotheses have been set to identify the possible topological relationships between geometries resulting from integration. Consequently, the problem of topological relationships vagueness remains an open question for other types of integration and should be studied regarding the specificities of each one. Finally, we conclude that the present thesis leads to address many complex problems that require several projects and a real research community to be resolved.

References

CLEMENTINI E. AND DI FELICE P. 1995, A Comparison of Methods for Representing Topological Relationships. *Information Sciences* 3: 149-178.

CLEMENTINI E. AND P. DI FELICE 1997, Approximate topological relations. International Journal of Approximate Reasoning, 16:173-204.

COHN, A.G. AND GOTTS N.M. 1996, The 'egg-yolk' representation of regions with indeterminate boundaries in: Burrough, P. & Frank, A. M. (editors) Proceedings of the GISDATA Specialist Meeting on Spatial Objects with Undetermined Boundaries, pp. 171-187 Francis Taylor.

COHN A.G. BENNETT B. GOODAY J. AND GOTTS N.M. 1997, Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. GeoInformatica, 1(3):275–316.

DUBOISSET M. PINET F. KANG M.A. AND SCHNEIDER M. 2005, Precise Modeling and Verification of Topological Integrity Constraints in Spatial Databases: From an Expressive Power Study to Code Generation Principles. ER 2005: 465-482.

DILO A. 2006, “Representation of and reasoning with vagueness in spatial information: A system for handling vague objects”, thesis, 187p.

DU S. QIN Q. WANG Q. AND LI B. 2005, Fuzzy description of topological relations I: a unified fuzzy 9-intersection model. *In: L. Wang, K. Chen, Y.S. Ong (Eds.), Advances in Natural Computation*, Lecture Notes in Computer Science, vol. 3612, pp. 1260-1273.

EGENHOFER M. AND J. HERRING 1990, A mathematical framework for the definition of topological relationships. Proceedings of the Fourth International Symposium on Spatial Data Handling, Zurich, Switzerland (eds. K. Brassel and H. Kishimoto), 803--813.

ERWIG M. AND M. SCHNEIDER 1997, Vague regions. In 5th International Symposium on Advances in Spatial Databases (SSD'97), number 1262 in Lecture Notes in Computer Science, pp. 298--320.

RANDELL D.A. AND COHN A.G. 1989, Modelling topological and metrical properties of physical processes. In Brachman, R.J., Levesque, H.J. and Reiter, R. editors, Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning (KR’89), pages 357–368. Morgan Kaufmann.

REIS R. EGENHOFER M.J. AND MATOS J. 2006, Topological relations using two models of uncertainty for lines. In. Proc. of 7th international Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. 5 - 7 July, Lisbon, Portugal, pp : 286-295.

SCHNEIDER M. 2001, A design of topological predicates for complex crisp and fuzzy regions. In ER ’01: Proceedings of the 20th International Conference on Conceptual Modeling, pages 103–116. Springer-Verlag, ISBN 3-540-42866-6.

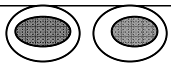
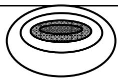
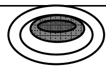
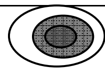
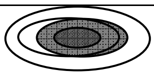
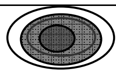
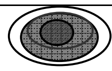
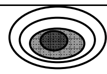
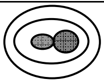
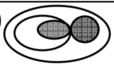
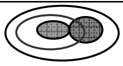
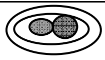
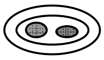
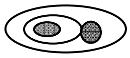
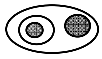
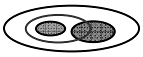
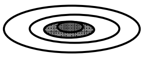
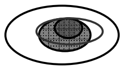
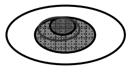
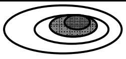
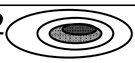
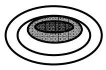
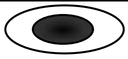
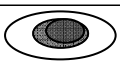
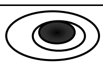
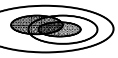
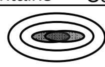
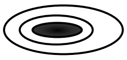
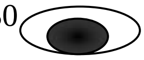
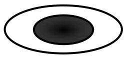
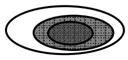
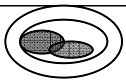
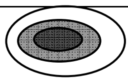
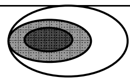
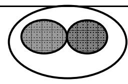
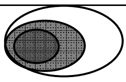
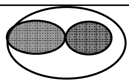
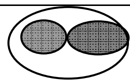
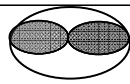
SOULIGNAC V. GIBOLD F. PINET F. AND VIGIER F. 2005, Spreading Matter Management in France within Sigemo. In: the 5th European Conference for Information Technologies in Agriculture (EFITA 2005), Portugal , July.

TANG XI. 2004, Spatial object modeling in fuzzy topological spaces: with applications to land cover change. PhD thesis, University of Twente, ISBN 90-6164-220-5.

VERSTRAETE J. HALLEZ A. AND DE TRÉ G. 2007, Fuzzy Regions: Theory and Applications. In A. Morris, S. Kokhan (Eds.). *Geographic Uncertainty in Environmental Security*, pp. 1-17. Dordrecht, The Netherlands: Springer.

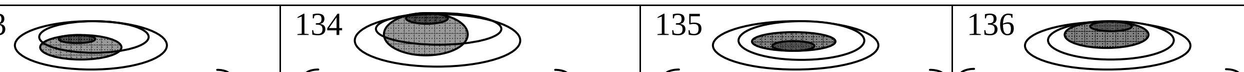
ZHAN B.F. 1997, Topological relations between fuzzy regions. In Proceedings of the 1997 ACM Symposium on Applied Computing, pages 192–196. ACM Press, 1997. ISBN 0-89791-850-9.

Appendix 1: 242 topological relations between regions with broad boundaries and required rules to deduce them

1  Disjoint Disjoint Disjoint Disjoint	2  Contains Inside Contains Contains	3  Contains Covered by Contains Contains	4  Contains Equal Contains Contains
5  Contains Overlap Contains Contains	6  Contains Contains Contains Contains	7  Contains Covers Contains Contains	8  Inside Inside Contains Contains
9  Meet Inside Contains Contains	10  Meet Meet Contains Contains	11  Meet Overlap Contains Contains	12  Meet Covered by Contains Contains
13  Disjoint Inside Contains Contains	14  Disjoint Covered by Contains Contains	15  Disjoint Meet Contains Contains	16  Disjoint Disjoint Contains Contains
17  Disjoint Overlap Contains Contains	18  Covers Inside Contains Contains	19  Covers Overlap Contains Contains	20  Covers Covers Contains Contains
21  Covers Covered by Contains Contains	22  Covers Inside Contains Contains	23  Inside Covered by Contains Contains	24  Equal Equal Contains Contains
25  Covers Equal Contains Contains	26  Equal Covered by Contains Contains	27  Overlap Overlap Contains Contains	28  Overlap Inside Contains Contains
29  Equal Inside Contains Contains	30  Equal Covered by Covers Equal	31  Equal Inside Contains Equal	32  Contains Inside Contains Equal
33  Overlap Covered by Contains Contains	34  Inside Inside Contains Equal	35  Inside Inside Covers Equal	36  Meet Inside Contains Equal
37  Contains Covered by Contains Equal	38  Meet Inside Covers Equal	39  Meet Covered by Contains Equal	40  Meet Covered by Covers Equal

41 Disjoint Inside Covers Equal	42 Disjoint Covered by Covers Equal	43 Disjoint Inside Contains Equal	44 Disjoint Covered by Contains Equal
45 Overlap Covered by Contains Equal	46 Overlap Inside Covers Equal	47 Overlap Covered by Covers Equal	48 Overlap Inside Contains Equal
49 Covers Covered by Covers Equal	50 Covered by Covered by Covers Equal	51 Covered by Inside Contains Inside	52 Covers Inside Contains Equal
53 Contains Covered by Contains Covers	54 Contains Covers Contains Covers	55 Contains Equal Contains Covers	56 Contains Inside Contains Covers
57 Inside Inside Contains Covers	58 Inside Inside Covers Covers	59 Meet Inside Contains Covers	60 Contains Overlap Contains Covers
61 Meet Covered by Contains Covers	62 Meet Covered by Covers Covers	63 Meet Overlap Contains Covers	64 Meet Inside Covers Covers
65 Meet Meet Covers Covers	66 Meet Overlap Covers Covers	67 Disjoint Inside Contains Covers	68 Meet Meet Contains Covers
69 Disjoint Covered by Contains Covers	70 Disjoint Covered by Covers Covers	71 Disjoint Overlap Contains Covers	72 Disjoint Inside Covers Covers
73 Disjoint Disjoint Contains Covers	74 Disjoint Meet Covers Covers	75 Disjoint Overlap Covers Covers	76 Disjoint Disjoint Contains Covers
77 Equal Covered by Contains Covers	78 Equal Covered by Covers Covers	79 Equal Inside Contains Covers	80 Disjoint Meet Contains Covers
81 Covers Covers Covers Covers	82 Covers Inside Contains Covers	83 Covers Covered by Covers Covers	84 Covers Equal Covers Covers

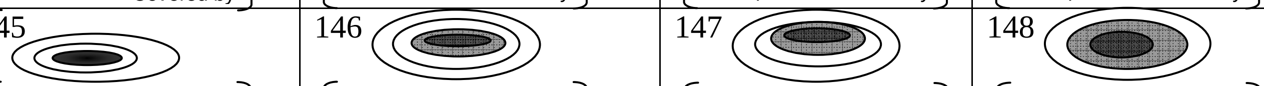
85 Covers Covered by Contains Covers	86 Covers Overlap Contains Covers	87 Covers Overlap Covers Covers	88 Covered by Inside Contains Covers
89 Covered by Covered by Covers Covers	90 Covered by Inside Covers Covers	91 Covers Equal Contains Covers	92 Overlap Covered by Covers Covers
93 Overlap Inside Contains Covers	94 Overlap Covered by Contains Covers	95 Covered by Covered by Contains Covers	96 Overlap Overlap Contains Covers
97 Overlap Overlap Covers Covers	98 Overlap Inside Covers Covers	99 Inside Inside Covers Covered by	100 Inside Inside Covered by Covered by
101 Inside Inside Equal Covered by	102 Inside Inside Contains Covered by	103 Contains Inside Contains Covered by	104 Contains Covered by Contains Covered by
105 Meet Inside Contains Covered by	106 Inside Inside Overlap Covered by	107 Meet Inside Covers Covered by	108 Meet Covered by Covers Covered by
109 Meet Inside Overlap Covered by	110 Meet Covered by Contains Covered by	111 Meet Covered by Meet Covered by	112 Meet Covered by Overlap Covered by
113 Disjoint Inside Contains Covered by	114 Meet Inside Meet Covered by	115 Disjoint Inside Covers Covered by	116 Disjoint Covered by Covers Covered by
117 Disjoint Inside Overlap Covered by	118 Disjoint Covered by Contains Covered by	119 Disjoint Covered by Disjoint Covered by	120 Disjoint Covered by Meet Covered by
121 Disjoint Covered by Overlap Covered by	122 Disjoint Inside Disjoint Covered by	123 Equal Inside Covers Covered by	124 Equal Covered by Covers Covered by
125 Equal Inside Contains Covered by	126 Disjoint Inside Meet Covered by	127 Covered by Covered by Covered by Covered by	128 Covered by Inside Contains Covered by
129 Covered by Covered by Covers Covered by	130 Overlap Covered by Contains Covered by	131 Covered by Covered by Equal Covered by	132 Covered by Inside Covers Covered by



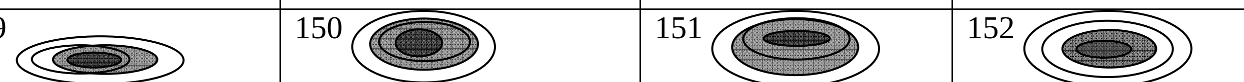
133 Covered by · Inside Overlap · Covered by	134 Covered by · Covered by Overlap · Covered by	135 Covers · Inside Contains · Covered by	136 Covers · Covered by Covers · Covered by

137 Covers · Covered by Contains · Covered by	138 Covered by · Inside Equal · Covered by	139 Overlap · Covered by Covers · Covered by	140 Overlap · Inside Contains · Covered by

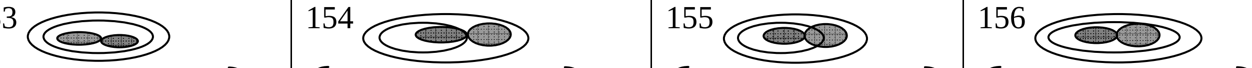
141 Overlap · Inside Covers · Covered by	142 Covers · Inside Covers · Covered by	143 Overlap · Inside Overlap · Covered by	144 Overlap · Covered by Overlap · Covered by



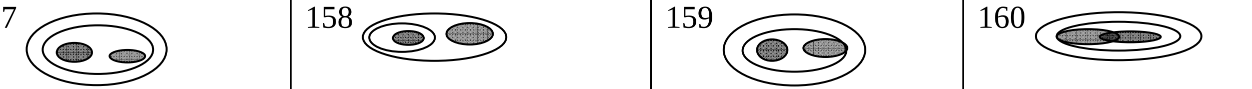
145 Equal · Inside Contains · Inside	146 Inside · Inside Contains · Inside	147 Inside · Inside Covers · Inside	148 Inside · Inside Equal · Inside



149 Inside · Inside Overlap · Inside	150 Inside · Inside Inside · Inside	151 Inside · Inside Covered by · Inside	152 Contains · Inside Contains · Inside



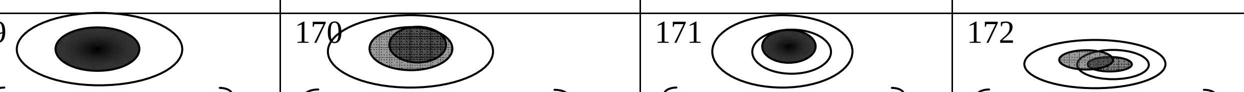
153 Meet · Inside Contains · Inside	154 Meet · Inside Meet · Inside	155 Meet · Inside Overlap · Inside	156 Meet · Inside Covers · Inside



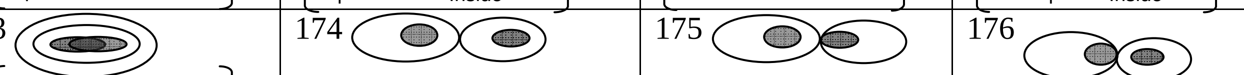
157 Disjoint · Inside Contains · Inside	158 Disjoint · Inside Disjoint · Inside	159 Disjoint · Inside Covers · Inside	160 Overlap · Inside Covers · Inside

161 Disjoint · Inside Meet · Inside	162 Disjoint · Inside Overlap · Inside	163 Covered by · Inside Contains · Inside	164 Covered by · Inside Overlap · Inside

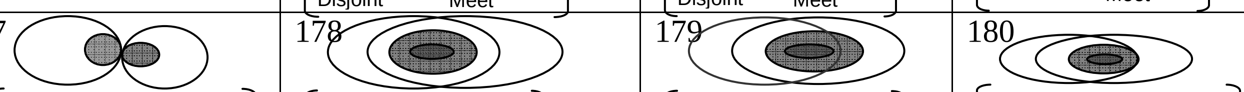
165 Covered by · Inside Covered by · Inside	166 Covered by · Inside Covers · Inside	167 Covers · Inside Contains · Inside	168 Covers · Inside Covers · Inside



169 Equal · Inside Equal · Inside	170 Covered by · Inside Equal · Inside	171 Equal · Inside Covers · Inside	172 Overlap · Inside Overlap · Inside

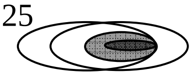
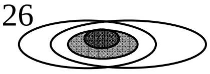
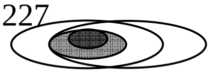
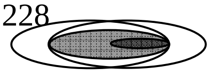
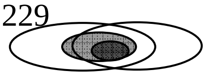
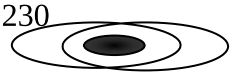
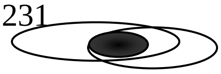
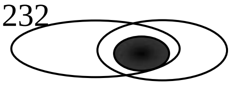
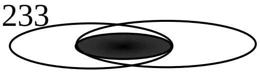
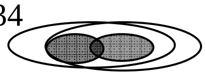
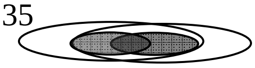
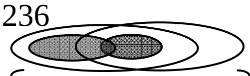
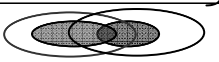
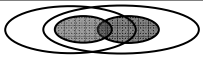
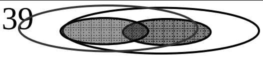
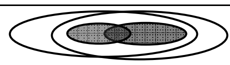
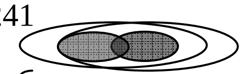
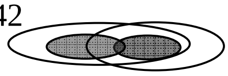


173 Overlap · Inside Contains · Inside	174 Disjoint · Disjoint Disjoint · Meet	175 Disjoint · Meet Disjoint · Meet	176 Disjoint · Disjoint Meet · Meet



177 Meet · Meet Meet · Meet	178 Contains · Inside Contains · Overlap	179 Contains · Overlap Contains · Overlap	180 Contains · Covered by Contains · Overlap

181 Inside Inside Contains Overlap	182 Inside Inside Overlap Overlap	183 Inside Inside Covers Overlap	184 Meet Inside Contains Overlap
185 Meet Inside Covers Overlap	186 Meet Covered by Contains Overlap	187 Meet Covered by Covers Overlap	188 Meet Inside Overlap Overlap
189 Meet Overlap Contains Overlap	190 Meet Overlap Overlap Overlap	191 Meet Covered by Overlap Overlap	192 Meet Overlap Covers Overlap
193 Disjoint Inside Contains Overlap	194 Disjoint Inside Covers Overlap	195 Disjoint Covered by Contains Overlap	196 Disjoint Covered by Covers Overlap
197 Disjoint Inside Overlap Overlap	198 Disjoint Overlap Contains Overlap	199 Disjoint Overlap Overlap Overlap	200 Disjoint Covered by Overlap Overlap
201 Disjoint Overlap Covers Overlap	202 Disjoint Disjoint Overlap Overlap	203 Disjoint Overlap Disjoint Overlap	204 Disjoint Meet Overlap Overlap
205 Disjoint Overlap Meet Overlap	206 Disjoint Meet Disjoint Overlap	207 Disjoint Disjoint Meet Overlap	208 Disjoint Meet Meet Overlap
209 Disjoint Inside Disjoint Overlap	210 Disjoint Disjoint Contains Overlap	211 Disjoint Disjoint Covers Overlap	212 Disjoint Covered by Disjoint Overlap
213 Disjoint Meet Contains Overlap	214 Disjoint Inside Meet Overlap	215 Disjoint Meet Covers Overlap	216 Disjoint Covered by Meet Overlap
217 Disjoint Disjoint Disjoint Overlap	218 Covers Inside Contains Overlap	219 Covers Covered by Contains Overlap	220 Covers Covered by Covers Overlap
221 Covers Overlap Contains Overlap	222 Covers Overlap Covers Overlap	223 Covers Covered by Contains Overlap	224 Covered by Covered by Overlap Overlap

225  $\begin{bmatrix}\text{Covered by} & \text{Inside}\\ \text{Covers} & \text{Overlap}\end{bmatrix}$	226  $\begin{bmatrix}\text{Covered by} & \text{Inside}\\ \text{Contains} & \text{Overlap}\end{bmatrix}$	227  $\begin{bmatrix}\text{Covered by} & \text{Inside}\\ \text{Covers} & \text{Overlap}\end{bmatrix}$	228  $\begin{bmatrix}\text{Covered by} & \text{Covered by}\\ \text{Covers} & \text{Overlap}\end{bmatrix}$
229  $\begin{bmatrix}\text{Covered by} & \text{Inside}\\ \text{Overlap} & \text{Overlap}\end{bmatrix}$	230  $\begin{bmatrix}\text{Equal} & \text{Inside}\\ \text{Contains} & \text{Overlap}\end{bmatrix}$	231  $\begin{bmatrix}\text{Equal} & \text{Inside}\\ \text{Covers} & \text{Overlap}\end{bmatrix}$	232  $\begin{bmatrix}\text{Equal} & \text{Covered by}\\ \text{Contains} & \text{Overlap}\end{bmatrix}$
233  $\begin{bmatrix}\text{Equal} & \text{Covered by}\\ \text{Covers} & \text{Overlap}\end{bmatrix}$	234  $\begin{bmatrix}\text{Overlap} & \text{Covered by}\\ \text{Contains} & \text{Overlap}\end{bmatrix}$	235  $\begin{bmatrix}\text{Overlap} & \text{Covered by}\\ \text{Covers} & \text{Overlap}\end{bmatrix}$	236  $\begin{bmatrix}\text{Overlap} & \text{Inside}\\ \text{Overlap} & \text{Overlap}\end{bmatrix}$
237  $\begin{bmatrix}\text{Overlap} & \text{Overlap}\\ \text{Overlap} & \text{Overlap}\end{bmatrix}$	238  $\begin{bmatrix}\text{Overlap} & \text{Overlap}\\ \text{Contains} & \text{Overlap}\end{bmatrix}$	239  $\begin{bmatrix}\text{Overlap} & \text{Overlap}\\ \text{Covers} & \text{Overlap}\end{bmatrix}$	240  $\begin{bmatrix}\text{Overlap} & \text{Inside}\\ \text{Contains} & \text{Overlap}\end{bmatrix}$
241  $\begin{bmatrix}\text{Overlap} & \text{Inside}\\ \text{Covers} & \text{Overlap}\end{bmatrix}$	242  $\begin{bmatrix}\text{Overlap} & \text{Covered by}\\ \text{Overlap} & \text{Overlap}\end{bmatrix}$		

Appendix 2: Rules of consistency

Table A2.1 Required rules for topological relations between regions with broad boundaries

<table>
<tr><td>Rule 1: Let $\tilde{A}$ and $\tilde{B}$ are two simple regions with broad boundaries, if Disjoint ($\tilde{A}_{max}$, $\tilde{B}_{max}$) then
Disjoint ($\tilde{A}_{min}$, $\tilde{B}_{min}$).</td><td>$$\begin{array}{cc} & \begin{matrix} \tilde{B}_{min} & \tilde{B}_{max} \end{matrix} \\ \begin{matrix} \tilde{A}_{min} \\ \tilde{A}_{max} \end{matrix} & \begin{pmatrix} D\,(\tilde{A}_{min}\,,\,\tilde{B}_{min}) & -- \\ -- & D\,(\tilde{A}_{max}\,,\tilde{B}_{max}) \end{pmatrix} \end{array}$$</td></tr>
<tr><td colspan="2">Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where Disjoint ($\tilde{A}_{max}$, $\tilde{B}_{max}$). Now, we suppose that R ($\tilde{A}_{min}$, $\tilde{B}_{min}$) $\neq$ Disjoint. In this case, the relation between minimal extent $\tilde{A}_{min}$ and maximal extent $\tilde{A}_{max}$ of a region with a broad boundary $\tilde{A}$ or that between $\tilde{B}_{max}$ and $\tilde{B}_{min}$ does not correspond to Contains, Covers, Equal. Thus, there is a contradiction with definition 1.</td></tr>
<tr><td>Rule 2: Let $\tilde{A}$ and $\tilde{B}$ two regions with broad boundaries, if Meet ($\tilde{A}_{max}$, $\tilde{B}_{max}$) then R($\tilde{A}_{min}$, $\tilde{B}_{min}$)$\in$ {D, M}.</td><td>$$\begin{array}{cc} & \begin{matrix} \tilde{B}_{min} & \tilde{B}_{max} \end{matrix} \\ \begin{matrix} \tilde{A}_{min} \\ \tilde{A}_{max} \end{matrix} & \begin{pmatrix} R(\tilde{A}_{min}\,,\tilde{B}_{min}) \in \{D, M\} & -- \\ -- & M\,(\tilde{A}_{max}\,,\tilde{B}_{max}) \end{pmatrix} \end{array}$$</td></tr>
<tr><td colspan="2">Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where Meet ($\tilde{A}_{max}$, $\tilde{B}_{max}$). Now, we suppose that R($\tilde{A}_{min}$, $\tilde{B}_{min}$) $\notin$ {Disjoint, Meet}. In this case, relation between minimal extent $\tilde{A}_{min}$ and maximal extent $\tilde{A}_{max}$, R' ($\tilde{A}_{max}$, $\tilde{A}_{min}$) or that between $\tilde{B}_{max}$ and $\tilde{B}_{min}$, R''($\tilde{B}_{max}$, $\tilde{B}_{min}$) does not correspond to Contains, Covers, Equal. Thus, there is a contradiction with definition 1.</td></tr>
<tr><td>Rule 3: Let $\tilde{A}$ and $\tilde{B}$ two regions with broad boundaries, if Contains ($\tilde{A}_{max}$, $\tilde{B}_{max}$) then Contains ($\tilde{A}_{max}$, $\tilde{B}_{min}$), and vice versa.</td><td>$$\begin{array}{cc} & \begin{matrix} \tilde{B}_{min} & \tilde{B}_{max} \end{matrix} \\ \begin{matrix} \tilde{A}_{min} \\ \tilde{A}_{max} \end{matrix} & \begin{pmatrix} -- & -- \\ C\,(\tilde{A}_{max}\,,\tilde{B}_{min}) & C\,(\tilde{A}_{max}\,,\tilde{B}_{max}) \end{pmatrix} \end{array}$$</td></tr>
<tr><td colspan="2">Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where Contains ($\tilde{A}_{max}$, $\tilde{B}_{max}$). According to definition 1, any region with a broad boundary $\tilde{A}$ should respect the principal following condition: Equal($\tilde{A}_{max}$, $\tilde{A}_{min}$), Contains($\tilde{A}_{max}$, $\tilde{A}_{min}$) or Covers($\tilde{A}_{max}$, $\tilde{A}_{min}$). Moreover, Contains is a transitive topological relation: Contains (A,B) and Contains(B,C)➔Contains(A,C). Then, since Contains ($\tilde{A}_{max}$, $\tilde{B}_{max}$) and R ($\tilde{B}_{max}$, $\tilde{B}_{min}$) = {Contains, Covers, Equal} then Contains ($\tilde{A}_{max}$, $\tilde{B}_{min}$) and vice versa.</td></tr>
<tr><td>Rule 4: Let $\tilde{A}$ and $\tilde{B}$ two regions with broad boundaries, if Covers ($\tilde{A}_{max}$, $\tilde{B}_{max}$) then R ($\tilde{A}_{max}$, $\tilde{B}_{min}$) $\in$ {Contains, Covers}, and vice versa.</td><td>$$\begin{array}{cc} & \begin{matrix} \tilde{B}_{min} & \tilde{B}_{max} \end{matrix} \\ \begin{matrix} \tilde{A}_{min} \\ \tilde{A}_{max} \end{matrix} & \begin{pmatrix} -- & -- \\ R(\tilde{A}_{max}\,,\tilde{B}_{min}) \in \{C, CV\} & C\,(\tilde{A}_{max}\,,\tilde{B}_{max}) \end{pmatrix} \end{array}$$</td></tr>
<tr><td colspan="2">Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where Covers ($\tilde{A}_{max}$, $\tilde{B}_{max}$). According to definition 1, any region with a broad boundary $\tilde{A}$ should respect the principal following condition: Equal($\tilde{A}_{max}$, $\tilde{A}_{min}$), Contains($\tilde{A}_{max}$, $\tilde{A}_{min}$) or Covers($\tilde{A}_{max}$, $\tilde{A}_{min}$). Contains is a transitive topological relation: if Contains(A,B) and Contains(B,C) ➔ Conatins(A,C). Then, if Contains ($\tilde{B}_{max}$, $\tilde{B}_{min}$) then Contains ($\tilde{A}_{max}$, $\tilde{B}_{min}$) else if R ($\tilde{B}_{max}$, $\tilde{B}_{min}$)$\in$ {Covers, Equal} then Covers ($\tilde{A}_{max}$, $\tilde{B}_{min}$) else if Covers($\tilde{B}_{max}$, $\tilde{B}_{min}$) then R ($\tilde{A}_{max}$, $\tilde{B}_{min}$) $\in$</td></tr>
</table>

{*Contains*, *Covers*} and vice versa.	
Rule 5: Let $\tilde{A}$ and $\tilde{B}$ two regions with broad boundaries, if *Equal* ($\tilde{A}_{max}, \tilde{B}_{max}$) then *R* ($\tilde{A}_{max}, \tilde{B}_{min}$) $\in$ {*Contains*, *Covers*} and *R* ($\tilde{B}_{max}, \tilde{A}_{min}$)$\in$ {*Inside*, *Covered by*}, and vice versa.	$\begin{array}{cc} & \begin{matrix} \tilde{B}_{min} & \tilde{B}_{max} \end{matrix} \\ \begin{matrix} \tilde{A}_{min} \\ \tilde{A}_{max} \end{matrix} & \begin{pmatrix} -- & R(\tilde{B}_{max}, \tilde{A}_{min}) \in \{I, CVB\} \\ R(\tilde{A}_{max}, \tilde{B}_{min}) \in \{C, CV\} & E(\tilde{A}_{max}, \tilde{B}_{max}) \end{pmatrix} \end{array}$
Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Equal* ($\tilde{A}_{max}, \tilde{B}_{max}$). According to definition 1, any region with a broad boundary $\tilde{A}$ should respect the principal following condition: *Equal(*$\tilde{A}_{max}, \tilde{A}_{min}$*)*, *Contains(*$\tilde{A}_{max}, \tilde{A}_{min}$*)* *or* *Covers(*$\tilde{A}_{max}, \tilde{A}_{min}$*)*. In this case, we don't consider *Equal(*$\tilde{A}_{max}, \tilde{A}_{min}$*)* because the topological relation becomes between crisp regions thoughtfully studied in other works (e.g, Egenhofer and Herring 1990). *Equal* and *Contains* are transitive topological relations: *Equal(A,B)* and *Equal(B,C)* ➔ *Equal(A,C)*, *Contains(A,B)* and *Contains(B,C)* ➔ *Contains(A,C)*. Then, if *Equal*($\tilde{A}_{max}, \tilde{B}_{max}$) and *Contains*($\tilde{B}_{max}, \tilde{B}_{min}$) then *Contains*($\tilde{A}_{max}, \tilde{B}_{min}$) (1) else if *Covers*($\tilde{B}_{max}, \tilde{B}_{min}$) then *Covers*($\tilde{A}_{max}, \tilde{B}_{min}$) (2). Then, (1) and (2) implies that *R* ($\tilde{A}_{max}, \tilde{B}_{min}$)$\in$ {*Contains*, *Covers*} and *R* ($\tilde{B}_{max}, \tilde{A}_{min}$)$\in$ {*Inside*, *Covered by*}.	
Rule 6: Let $\tilde{A}$ and $\tilde{B}$ two regions with broad boundaries, if *Contains* ($\tilde{A}_{max}, \tilde{B}_{max}$) and *Contains* ($\tilde{A}_{min}, \tilde{B}_{min}$) then *R* ($\tilde{A}_{min}, \tilde{B}_{max}$)$\notin$ {*Meet*, *Equal*}, and vice versa.	$\begin{array}{cc} & \begin{matrix} \tilde{B}_{min} & \tilde{B}_{max} \end{matrix} \\ \begin{matrix} \tilde{A}_{min} \\ \tilde{A}_{max} \end{matrix} & \begin{pmatrix} C(\tilde{A}_{min}, \tilde{B}_{min}) & R(\tilde{A}_{min}, \tilde{B}_{max}) \notin \{D, M\} \\ -- & C(\tilde{A}_{max}, \tilde{B}_{max}) \end{pmatrix} \end{array}$
Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Contains* ($\tilde{A}_{max}, \tilde{B}_{max}$) and *Contains* ($\tilde{A}_{min}$, $\tilde{B}_{min}$). According to definition 1, we have *Equal(*$\tilde{A}_{max}, \tilde{A}_{min}$*)*, *Contains(*$\tilde{A}_{max}, \tilde{A}_{min}$*)* *or Covers(*$\tilde{A}_{max}, \tilde{A}_{min}$*)*. We suppose now that *Disjoint* ($\tilde{A}_{min}, \tilde{B}_{max}$) or *Meet* ($\tilde{A}_{min}, \tilde{B}_{max}$) (1). By considering definition 1 and *Contains* ($\tilde{A}_{max}, \tilde{B}_{max}$)**,** since *R* ($\tilde{B}_{max}, \tilde{B}_{min}$){*Contains*, *Covers*, *Equal*} then *Contains* ($\tilde{A}_{max}, \tilde{B}_{min}$) (2). In addition, since *Contains* ($\tilde{A}_{min}$, $\tilde{B}_{min}$) and (1) then *R* ($\tilde{B}_{max}, \tilde{B}_{min}$)$\notin$ {*Contains*, *Covers*, *Equal*}. Thus, there is a contradiction with definition 1.	
Rule 7: Let $\tilde{A}$ and $\tilde{B}$ two regions with broad boundaries, if *Contains* ($\tilde{A}_{max}, \tilde{B}_{max}$) and *Inside*($\tilde{A}_{min}$, $\tilde{B}_{min}$) then *Inside* ($\tilde{A}_{min}, \tilde{B}_{max}$), and vice versa.	$\begin{array}{cc} & \begin{matrix} \tilde{B}_{min} & \tilde{B}_{max} \end{matrix} \\ \begin{matrix} \tilde{A}_{min} \\ \tilde{A}_{max} \end{matrix} & \begin{pmatrix} I(\tilde{A}_{min}, \tilde{B}_{min}) & I(\tilde{A}_{min}, \tilde{B}_{max}) \\ -- & C(\tilde{A}_{max}, \tilde{B}_{max}) \end{pmatrix} \end{array}$
Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Contains* ($\tilde{A}_{max}, \tilde{B}_{max}$) and *Inside*($\tilde{A}_{min}, \tilde{B}_{min}$). We suppose now that *R* ($\tilde{A}_{min}, \tilde{B}_{max}$) $\notin$ {*Inside*} (1). By considering definition 1 and *Contains* ($\tilde{A}_{max}, \tilde{B}_{max}$), since *R* ($\tilde{B}_{max}, \tilde{B}_{min}$)$\in$ {*Contains*, *Covers*, *Equal*} and *Inside* ($\tilde{A}_{min}, \tilde{B}_{min}$) then *Inside* ($\tilde{A}_{min}, \tilde{B}_{max}$) (2). Thus, there is contradiction among (1) and (2).	
Rule 8: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries, if *Contains* ($\tilde{A}_{max}, \tilde{B}_{max}$) and *Meet*($\tilde{A}_{min}, \tilde{B}_{min}$) then *R* ($\tilde{A}_{min}, \tilde{B}_{max}$)$\notin$ {*Contains*, *Equal*, *Covers*, *Disjoint*}, and vice versa.	$\begin{array}{cc} & \begin{matrix} \tilde{B}_{min} & \tilde{B}_{max} \end{matrix} \\ \begin{matrix} \tilde{A}_{min} \\ \tilde{A}_{max} \end{matrix} & \begin{pmatrix} M(\tilde{A}_{min}, \tilde{B}_{min}) & R(\tilde{A}_{min}, \tilde{B}_{max}) \notin \{C, E, CV, D\} \\ -- & C(\tilde{A}_{max}, \tilde{B}_{max}) \end{pmatrix} \end{array}$
Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Contains* ($\tilde{A}_{max}, \tilde{B}_{max}$) and *Meet* ($\tilde{A}_{min}$, $\tilde{B}_{min}$) (1). We suppose now that *R* ($\tilde{A}_{min}, \tilde{B}_{max}$) $\in$ {*Contains*, *Equal*, *Covers*, *Disjoint*} (2). By considering definition 1 and *Contains* ($\tilde{A}_{max}, \tilde{B}_{max}$), if *Contains* ($\tilde{A}_{min}, \tilde{B}_{max}$) then there is a contradiction because *R* ($\tilde{B}_{max}, \tilde{B}_{min}$)$\in$ {*Contains*,	

Covers, *Equal*} and (1). If *Equal* ($\widetilde{A}_{min}, \widetilde{B}_{max}$) then there is a contradiction because *R* ($\widetilde{B}_{max}, \widetilde{B}_{min}$)$\in$ {*Contains*, *Covers*, *Equal*} and (1). If Covers($\widetilde{A}_{min}, \widetilde{B}_{max}$) then *R* ($\widetilde{B}_{max}, \widetilde{B}_{min}$)$\notin$ {*Contains*, *Covers, Equal*}or (1) is *false*. Finally, if *Disjoint* ($\widetilde{A}_{min}, \widetilde{B}_{max}$) then there is a contradiction because *R* ($\widetilde{B}_{max}, \widetilde{B}_{min}$)$\in$ {*Contains*, *Covers*, *Equal*} and (1). Thus, (2) cannot be true.

Rule 9: Let $\widetilde{A}$ and $\widetilde{B}$ two regions with broad boundaries, if *Contains* ($\widetilde{A}_{max}, \widetilde{B}_{max}$) and *Covers*($\widetilde{A}_{min}, \widetilde{B}_{min}$) then *R* ($\widetilde{A}_{min}, \widetilde{B}_{max}$)$\notin$ {*Meet*, *Disjoint*}, and vice versa.

$$\begin{array}{c} \\ \widetilde{A}_{min} \\ \widetilde{A}_{max} \end{array} \begin{array}{c} \begin{array}{cc} \widetilde{B}_{min} & \widetilde{B}_{max} \end{array} \\ \left(\begin{array}{cc} CV(\widetilde{A}_{min}, \widetilde{B}_{min}) & R(\widetilde{A}_{min}, \widetilde{B}_{max}) \notin \{M, D\} \\ \text{--} & C(\widetilde{A}_{max}, \widetilde{B}_{max}) \end{array} \right) \end{array}$$

Proof: Let $\widetilde{A}$ and $\widetilde{B}$ two simple regions with broad boundaries where *Contains*($\widetilde{A}_{max}, \widetilde{B}_{max}$) and *Covers*($\widetilde{A}_{min}, \widetilde{B}_{min}$). We suppose now that *Disjoint* ($\widetilde{A}_{min}, \widetilde{B}_{max}$) or *Meet* ($\widetilde{A}_{min}, \widetilde{B}_{max}$) (1). By considering definition 1 and *Contains*($\widetilde{A}_{max}, \widetilde{B}_{max}$), since *R* ($\widetilde{B}_{max}, \widetilde{B}_{min}$)$\in$ {*Contains*, *Covers*, *Equal*} then *Contains* ($\widetilde{A}_{max}, \widetilde{B}_{min}$) (2). In addition, since *Covers* ($\widetilde{A}_{min}, \widetilde{B}_{min}$) and (1) then *R* ($\widetilde{B}_{max}, \widetilde{B}_{min}$) $\notin$ {*Contains*, *Covers*, *Equal*}. Thus, there is a contradiction with definition 1.

Rule 10: Let $\widetilde{A}$ and $\widetilde{B}$ two regions with broad boundaries, if *Contains* ($\widetilde{A}_{max}, \widetilde{B}_{max}$) and *Equal* ($\widetilde{A}_{min}, \widetilde{B}_{min}$) then *R* ($\widetilde{A}_{min}, \widetilde{B}_{max}$)$\notin$ {*Contains*, *Covers*, *Disjoint, Meet, Overlap*}, and vice versa.

$$\begin{array}{c} \\ \widetilde{A}_{min} \\ \widetilde{A}_{max} \end{array} \begin{array}{c} \begin{array}{cc} \widetilde{B}_{min} & \widetilde{B}_{max} \end{array} \\ \left(\begin{array}{cc} E(\widetilde{A}_{min}, \widetilde{B}_{min}) & R(\widetilde{A}_{min}, \widetilde{B}_{max}) \notin \{C, CV, D, M, O\} \\ \text{--} & C(\widetilde{A}_{max}, \widetilde{B}_{max}) \end{array} \right) \end{array}$$

Proof: Let $\widetilde{A}$ and $\widetilde{B}$ two simple regions with broad boundaries where *Contains* ($\widetilde{A}_{max}, \widetilde{B}_{max}$) and *Equal* ($\widetilde{A}_{min}, \widetilde{B}_{min}$) (1). We suppose now that *R* ($\widetilde{A}_{min}, \widetilde{B}_{max}$)$\in$ {*Contains*, *Covers*, *Disjoint*, *Meet*, *Overlap*} (2). If (2) then *R* ($\widetilde{B}_{max}, \widetilde{B}_{min}$)$\notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1 and (1), there is a contradiction and (2) cannot be true.

Rule 11: Let $\widetilde{A}$ and $\widetilde{B}$ two regions with broad boundaries, if *Contains* ($\widetilde{A}_{max}, \widetilde{B}_{max}$) and *Covered by* ($\widetilde{A}_{min}, \widetilde{B}_{min}$) then *R* ($\widetilde{A}_{min}, \widetilde{B}_{max}$)$\in$ {*Covered by*, *Inside*}, and vice versa.

$$\begin{array}{c} \\ \widetilde{A}_{min} \\ \widetilde{A}_{max} \end{array} \begin{array}{c} \begin{array}{cc} \widetilde{B}_{min} & \widetilde{B}_{max} \end{array} \\ \left(\begin{array}{cc} CVB(\widetilde{A}_{min}, \widetilde{B}_{min}) & R(\widetilde{A}_{min}, \widetilde{B}_{max}) \in \{CVB, I\} \\ \text{--} & C(\widetilde{A}_{max}, \widetilde{B}_{max}) \end{array} \right) \end{array}$$

Proof: Let $\widetilde{A}$ and $\widetilde{B}$ two simple regions with broad boundaries where *Contains* ($\widetilde{A}_{max}, \widetilde{B}_{max}$) and *Covered by* ($\widetilde{A}_{min}, \widetilde{B}_{min}$) (1). We suppose now that *R* ($\widetilde{A}_{min}, \widetilde{B}_{max}$)$\notin$ {*Covered by*, *Inside*} then *R* ($\widetilde{A}_{min}, \widetilde{B}_{max}$)$\in$ {*Contains*, *Covers*, *Disjoint*, *Meet*, *Overlap*, *Equal*} (2). If (2) then *R* ($\widetilde{B}_{max}, \widetilde{B}_{min}$)$\notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1 and (1), there is a contradiction and (2) cannot be true.

Rule 12: Let $\widetilde{A}$ and $\widetilde{B}$ two regions with broad boundaries, if *Contains*($\widetilde{A}_{max}, \widetilde{B}_{max}$) and *Overlap*($\widetilde{A}_{min}, \widetilde{B}_{min}$) then *R* ($\widetilde{A}_{min}, \widetilde{B}_{max}$)$\in$ {*Covered by*, *Inside*, *Overlap*}, and vice versa.

$$\begin{array}{c} \\ \widetilde{A}_{min} \\ \widetilde{A}_{max} \end{array} \begin{array}{c} \begin{array}{cc} \widetilde{B}_{min} & \widetilde{B}_{max} \end{array} \\ \left(\begin{array}{cc} O(\widetilde{A}_{min}, \widetilde{B}_{min}) & R(\widetilde{A}_{min}, \widetilde{B}_{max}) \in \{CVB, I, O\} \\ \text{--} & C(\widetilde{A}_{max}, \widetilde{B}_{max}) \end{array} \right) \end{array}$$

Proof: Let $\widetilde{A}$ and $\widetilde{B}$ two simple regions with broad boundaries where *Contains*($\widetilde{A}_{max}, \widetilde{B}_{max}$) and *Overlap* ($\widetilde{A}_{min}, \widetilde{B}_{min}$) (1). We suppose now that *R* ($\widetilde{A}_{min}, \widetilde{B}_{max}$)$\in$ {*Contains*, *Covers*, *Disjoint*, *Meet*, *Equal*} (2). If (2) then *R* ($\widetilde{B}_{max}, \widetilde{B}_{min}$)$\notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1 and (1), there is a contradiction and (2) cannot be true.

Rule 13: Let $\tilde{A}$ and $\tilde{B}$ two regions with broad boundaries, if *Covers* ($\tilde{A}_{\max}, \tilde{B}_{\max}$) and *Contains*($\tilde{A}_{\min}$, $\tilde{B}_{\min}$) then R ($\tilde{A}_{\min}, \tilde{B}_{\max}$) $\notin$ {*Disjoint*, *Meet*}, and vice versa.

$$\begin{array}{c} \\ \tilde{A}_{\min} \\ \tilde{A}_{\max} \end{array} \begin{array}{c} \begin{array}{cc} \tilde{B}_{\min} & \tilde{B}_{\max} \end{array} \\ \begin{pmatrix} C(\tilde{A}_{\min}, \tilde{B}_{\min}) & R(\tilde{A}_{\min}, \tilde{B}_{\max}) \notin \{D, M\} \\ -- & CV(\tilde{A}_{\max}, \tilde{B}_{\max}) \end{pmatrix} \end{array}$$

Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Covers* ($\tilde{A}_{\max}, \tilde{B}_{\max}$) and *Contains* ($\tilde{A}_{\min}, \tilde{B}_{\min}$) (1). We suppose now that *Disjoint* ($\tilde{A}_{\min}, \tilde{B}_{\max}$) or *Meet* ($\tilde{A}_{\min}, \tilde{B}_{\max}$) (2). If (2) then R ($\tilde{B}_{\max}, \tilde{B}_{\min}$) $\notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1 and (1), there is a contradiction and (2) cannot be true.

Rule 14: Let $\tilde{A}$ and $\tilde{B}$ two regions with broad boundaries, if *Covers* ($\tilde{A}_{\max}, \tilde{B}_{\max}$) and *Inside*($\tilde{A}_{\min}, \tilde{B}_{\min}$) then *Inside*($\tilde{A}_{\min}, \tilde{B}_{\max}$), and vice versa.

$$\begin{array}{c} \\ \tilde{A}_{\min} \\ \tilde{A}_{\max} \end{array} \begin{array}{c} \begin{array}{cc} \tilde{B}_{\min} & \tilde{B}_{\max} \end{array} \\ \begin{pmatrix} I(\tilde{A}_{\min}, \tilde{B}_{\min}) & I(\tilde{A}_{\min}, \tilde{B}_{\max}) \\ -- & CV(\tilde{A}_{\max}, \tilde{B}_{\max}) \end{pmatrix} \end{array}$$

Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Covers* ($\tilde{A}_{\max}, \tilde{B}_{\max}$) and *Inside* ($\tilde{A}_{\min}, \tilde{B}_{\min}$). We suppose now that R ($\tilde{A}_{\min}, \tilde{B}_{\max}$) $\notin$ {*Inside*} (1). Additionally, Inside is a transitive relation: *Inside(A,B)* and *Inside(B,C)* ➔ *Inside(A,C)* (2). By considering definition 1 and (2), since R ($\tilde{B}_{\max}, \tilde{B}_{\min}$) $\in$ {*Contains*, *Covers*, *Equal*} and *Inside* ($\tilde{A}_{\min}, \tilde{B}_{\min}$) then *Inside* ($\tilde{A}_{\min}, \tilde{B}_{\max}$) (2). Thus, (1) cannot be true.

Rule 15: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries, if *Covers* ($\tilde{A}_{\max}, \tilde{B}_{\max}$) and *R*($\tilde{A}_{\min}$, $\tilde{B}_{\min}$) $\in$ {*Disjoint*, *Meet*} then R ($\tilde{A}_{\min}, \tilde{B}_{\max}$) $\notin$ {*Contains*, *Covers*, *Disjoint*, *Equal*}, and vice versa.

$$\begin{array}{c} \\ \tilde{A}_{\min} \\ \tilde{A}_{\max} \end{array} \begin{array}{c} \begin{array}{cc} \tilde{B}_{\min} & \tilde{B}_{\max} \end{array} \\ \begin{pmatrix} R(\tilde{A}_{\min}, \tilde{B}_{\min}) \in \{D, M\} & R(\tilde{A}_{\min}, \tilde{B}_{\max}) \notin \{C, CV, D, E\} \\ -- & CV(\tilde{A}_{\max}, \tilde{B}_{\max}) \end{pmatrix} \end{array}$$

Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Covers* ($\tilde{A}_{\max}, \tilde{B}_{\max}$) and R($\tilde{A}_{\min}, \tilde{B}_{\min}$) $\in$ {*Disjoint,Meet*} (1). We suppose now that R ($\tilde{A}_{\min}, \tilde{B}_{\max}$) $\in$ {*Contains*, *Covers*, *Disjoint*, *Equal*} (2). If (2) then R ($\tilde{B}_{\max}, \tilde{B}_{\min}$) $\notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1 and (1), there is a contradiction and (2) cannot be true.

Rule 16: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries, if *Covers* ($\tilde{A}_{\max}, \tilde{B}_{\max}$) and *R*($\tilde{A}_{\min}$, $\tilde{B}_{\min}$) $\in$ {*Equal*, Covered by} then *R*($\tilde{A}_{\min}, \tilde{B}_{\max}$) $\in$ {*Covered by, Inside*}, and vice versa.

$$\begin{array}{c} \\ \tilde{A}_{\min} \\ \tilde{A}_{\max} \end{array} \begin{array}{c} \begin{array}{cc} \tilde{B}_{\min} & \tilde{B}_{\max} \end{array} \\ \begin{pmatrix} R(\tilde{A}_{\min}, \tilde{B}_{\min}) \in \{E, CVB\} & R(\tilde{A}_{\min}, \tilde{B}_{\max}) \in \{CVB, I\} \\ -- & CV(\tilde{A}_{\max}, \tilde{B}_{\max}) \end{pmatrix} \end{array}$$

Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Covers*($\tilde{A}_{\max}, \tilde{B}_{\max}$) and R ($\tilde{A}_{\min}, \tilde{B}_{\max}$) $\in$ {*Equal, Covered by*} (1). We suppose now that R ($\tilde{A}_{\min}, \tilde{B}_{\max}$) $\notin$ {*Covered by, Inside*} (2). If (2) then R ($\tilde{B}_{\max}, \tilde{B}_{\min}$) $\notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1 and (1), there is a contradiction and (2) cannot be true.

Rule 17: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries, if *Covers* ($\tilde{A}_{\max}, \tilde{B}_{\max}$) and *Overlap* ($\tilde{A}_{\min}$, $\tilde{B}_{\min}$) then *R*($\tilde{A}_{\min}, \tilde{B}_{\max}$) $\in$ {*Covered by*, *Inside*, *Overlap*}, and vice versa.

$$\begin{array}{c} \\ \tilde{A}_{\min} \\ \tilde{A}_{\max} \end{array} \begin{array}{c} \begin{array}{cc} \tilde{B}_{\min} & \tilde{B}_{\max} \end{array} \\ \begin{pmatrix} O(\tilde{A}_{\min}, \tilde{B}_{\min}) & R(\tilde{A}_{\min}, \tilde{B}_{\max}) \in \{CVB, I, O\} \\ -- & CV(\tilde{A}_{\max}, \tilde{B}_{\max}) \end{pmatrix} \end{array}$$

Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Covers* ($\tilde{A}_{max}, \tilde{B}_{max}$) and *Overlap* ($\tilde{A}_{min}$, $\tilde{B}_{min}$) (1). We suppose now that R ($\tilde{A}_{min}, \tilde{B}_{max}$)$\notin$ {*Covered by*, *Inside*, *Overlap*} (2). If (2) then R ($\tilde{B}_{max}, \tilde{B}_{min}$)$\notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1 and (1), there is a contradiction and (2) cannot be true.

Rule 18: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries, if *Meet* ($\tilde{A}_{max}, \tilde{B}_{max}$) and *Meet* ($\tilde{A}_{min}$, $\tilde{B}_{min}$) then *Meet*($\tilde{A}_{min}, \tilde{B}_{max}$) and *Meet*($\tilde{A}_{max}, \tilde{B}_{min}$), and vice versa.

$$\begin{array}{c|cc} & \tilde{B}_{min} & \tilde{B}_{max} \\ \hline \tilde{A}_{min} & M(\tilde{A}_{min}, \tilde{B}_{min}) & M(\tilde{A}_{min}, \tilde{B}_{max}) \\ \tilde{A}_{max} & M(\tilde{A}_{max}, \tilde{B}_{min}) & M(\tilde{A}_{max}, \tilde{B}_{max}) \end{array}$$

Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Meet* ($\tilde{A}_{max}, \tilde{B}_{max}$) and *Meet* ($\tilde{A}_{min}$, $\tilde{B}_{min}$) (1). We suppose now that R ($\tilde{A}_{min}, \tilde{B}_{max}$)$\neq$ *Meet* (2) and R ($\tilde{A}_{max}, \tilde{B}_{min}$)$\neq$ *Meet* (3). If (2) then R ($\tilde{B}_{max}, \tilde{B}_{min}$)$\notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. Thus, (2) cannot be true. In the same way, if (3) then there is a contradiction because R ($\tilde{B}_{max}, \tilde{B}_{min}$)$\notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1 and (1), there is a contradiction and (3) cannot be true.

Rule 19: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries, if *Meet* ($\tilde{A}_{max}, \tilde{B}_{max}$) and *Disjoint* ($\tilde{A}_{min}, \tilde{B}_{min}$) then R($\tilde{A}_{min}, \tilde{B}_{max}$)$\in$ {*Meet*, *Disjoint*} and R($\tilde{A}_{max}, \tilde{B}_{min}$)$\in$ {*Meet*, *Disjoint*}, and vice versa.

$$\begin{array}{c|cc} & \tilde{B}_{min} & \tilde{B}_{max} \\ \hline \tilde{A}_{min} & D(\tilde{A}_{min}, \tilde{B}_{min}) & R(\tilde{A}_{min}, \tilde{B}_{max}) \in \{M, D\} \\ \tilde{A}_{max} & R(\tilde{A}_{max}, \tilde{B}_{min}) \in \{M, D\} & M(\tilde{A}_{max}, \tilde{B}_{max}) \end{array}$$

Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Meet* ($\tilde{A}_{max}, \tilde{B}_{max}$) and *Disjoint* ($\tilde{A}_{min}, \tilde{B}_{min}$) (1). We suppose now that R ($\tilde{A}_{min}, \tilde{B}_{max}$)$\notin$ {*Meet*, *Disjoint*} (2) and R ($\tilde{A}_{max}, \tilde{B}_{min}$)$\notin$ {*Meet*, *Disjoint*} (3). If (2) then there R ($\tilde{B}_{max}, \tilde{B}_{min}$)$\notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1 and (1), there is a contradiction and (2) cannot be true. In the same way, if (3) then R ($\tilde{B}_{max}, \tilde{B}_{min}$)$\notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1 and (1), there is a contradiction and (3) cannot be true.

Rule 20 : Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries, if *Overlap* ($\tilde{A}_{max}, \tilde{B}_{max}$) then R ($\tilde{A}_{max}, \tilde{B}_{min}$)$\notin$ {*Equal*, *Inside*, *Covered by*}, and vice versa.

$$\begin{array}{c|cc} & \tilde{B}_{min} & \tilde{B}_{max} \\ \hline \tilde{A}_{min} & -- & -- \\ \tilde{A}_{max} & R(\tilde{A}_{max}, \tilde{B}_{min}) \notin \{E, I, CVB\} & O(\tilde{A}_{max}, \tilde{B}_{max}) \end{array}$$

Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Overlap* ($\tilde{A}_{max}, \tilde{B}_{max}$). According to definition 1, any region with a broad boundary $\tilde{A}$ should respect the principal following condition: *Equal*($\tilde{A}_{max}, \tilde{A}_{min}$), *Contains*($\tilde{A}_{max}, \tilde{A}_{min}$) *or Covers*($\tilde{A}_{max}, \tilde{A}_{min}$) (1). We suppose now that R ($\tilde{A}_{max}, \tilde{B}_{min}$)$\in$ {*Equal*, *Inside*, *Covered by*} (2). By considering definition 1, if (1) and (2) then R ($\tilde{B}_{max}, \tilde{B}_{min}$)$\notin$ {*Contains*, *Covers*, *Equal*}. Thus, there is a contradiction with *definition 1*.

Rule 21: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries, if *Overlap* ($\tilde{A}_{max}, \tilde{B}_{max}$) and *Contains* ($\tilde{A}_{min}, \tilde{B}_{min}$) then R ($\tilde{A}_{min}, \tilde{B}_{max}$) $\in$ {*Overlap*, *Inside*, *Covered by*} and *Contains*($\tilde{A}_{max}, \tilde{B}_{min}$), and vice versa.

$$\begin{array}{c|cc} & \tilde{B}_{min} & \tilde{B}_{max} \\ \hline \tilde{A}_{min} & C(\tilde{A}_{min}, \tilde{B}_{min}) & R(\tilde{A}_{min}, \tilde{B}_{max}) \in \{O, I, CVB\} \\ \tilde{A}_{max} & C(\tilde{A}_{max}, \tilde{B}_{min}) & O(\tilde{A}_{max}, \tilde{B}_{max}) \end{array}$$

Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Overlap* ($\tilde{A}_{max}, \tilde{B}_{max}$) and *Contains* ($\tilde{A}_{min}, \tilde{B}_{min}$) (1). We suppose now that R ($\tilde{A}_{min}, \tilde{B}_{max}$)$\notin$ {*Overlap*, *Inside*, *Covered by*} (2) and R

$(\tilde{A}_{max}, \tilde{B}_{min}) \neq$ Contains (3). By considering definition 1 and *Contains*$(\tilde{A}_{min}, \tilde{B}_{min})$, if (2) then $R(\tilde{B}_{max}, \tilde{B}_{min}) \notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. Thus, (2) cannot be true because there is a contradiction. In the same way, if (3) then $R(\tilde{B}_{max}, \tilde{B}_{min}) \notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1 and *Contains*$(\tilde{A}_{min}, \tilde{B}_{min})$, (3) cannot be true because there is also a contradiction.

Rule 22: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries, if *Overlap* $(\tilde{A}_{max}, \tilde{B}_{max})$ and $R(\tilde{A}_{min}, \tilde{B}_{min}) \in$ {*Overlap*, *Meet*} then $R(\tilde{A}_{min}, \tilde{B}_{max}) \in$ {*Overlap, Inside, Covered by*} and $R(\tilde{A}_{max}, \tilde{B}_{min}) \in$ {*Overlap, Covers, Contains*}, and vice versa.

$$\begin{array}{c|cc} & \tilde{B}_{min} & \tilde{B}_{max} \\ \hline \tilde{A}_{min} & R(\tilde{A}_{min}, \tilde{B}_{min}) \in \{O, M\} & R(\tilde{A}_{min}, \tilde{B}_{max}) \in \{O, I, CVB\} \\ \tilde{A}_{max} & R(\tilde{A}_{max}, \tilde{B}_{min}) \in \{O, CV, C\} & O(\tilde{A}_{max}, \tilde{B}_{max}) \end{array}$$

Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Overlap* $(\tilde{A}_{max}, \tilde{B}_{max})$ and $R(\tilde{A}_{min}, \tilde{B}_{min}) \in$ {*Overlap*, *Meet*} (1). We suppose now that $R(\tilde{A}_{min}, \tilde{B}_{max}) \notin$ {*Overlap, Inside, Covered by*} (2) $R(\tilde{A}_{max}, \tilde{B}_{min}) \notin$ {*Overlap, Covers, Contains*} (3). If (2) then $R(\tilde{B}_{max}, \tilde{B}_{min}) \notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1, there is a contradiction and (2) cannot be true. In the same way, if (3) then $R(\tilde{B}_{max}, \tilde{B}_{min}) \notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1, there is contradiction and (3) cannot be true.

Rule 23: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries, if *Overlap* $(\tilde{A}_{max}, \tilde{B}_{max})$ and *Equal* $(\tilde{A}_{min}, \tilde{B}_{min})$ then $R(\tilde{A}_{min}, \tilde{B}_{max}) \in$ {*Overlap, Inside, Covered by*} and $R(\tilde{A}_{max}, \tilde{B}_{min}) \in$ {*Overlap, Covers, Contains*}, and vice versa.

$$\begin{array}{c|cc} & \tilde{B}_{min} & \tilde{B}_{max} \\ \hline \tilde{A}_{min} & E(\tilde{A}_{min}, \tilde{B}_{min}) & R(\tilde{A}_{max}, \tilde{B}_{min}) \in \{I, CVB\} \\ \tilde{A}_{max} & R(\tilde{A}_{max}, \tilde{B}_{min}) \in \{CV, C\} & O(\tilde{A}_{max}, \tilde{B}_{max}) \end{array}$$

Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Overlap* $(\tilde{A}_{max}, \tilde{B}_{max})$ and *Equal* $(\tilde{A}_{min}, \tilde{B}_{min})$ (1). We suppose now that $R(\tilde{A}_{min}, \tilde{B}_{max}) \notin$ {*Inside, Covered by*} (2) $R(\tilde{A}_{max}, \tilde{B}_{min}) \notin$ {*Covers, Contains*} (3). If (2) then $R(\tilde{B}_{max}, \tilde{B}_{min}) \notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1 and *Equal* $(\tilde{A}_{min}, \tilde{B}_{min})$, there is contradiction and (2) cannot be true. In the same way, if (3) then $R(\tilde{B}_{max}, \tilde{B}_{min}) \notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1 and *Equal* $(\tilde{A}_{min}, \tilde{B}_{min})$, (3) cannot be true because there is also a contradiction.

Rule 24: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries, if *Overlap* $(\tilde{A}_{max}, \tilde{B}_{max})$ and *Inside* $(\tilde{A}_{min}, \tilde{B}_{min})$ then $R(\tilde{A}_{min}, \tilde{B}_{max}) \in$ {*Inside*} and $(\tilde{A}_{max}, \tilde{B}_{min}) \in$ { *Contains, Covers, Overlap*}, and vice versa.

$$\begin{array}{c|cc} & \tilde{B}_{min} & \tilde{B}_{max} \\ \hline \tilde{A}_{min} & I(\tilde{A}_{min}, \tilde{B}_{min}) & I(\tilde{A}_{max}, \tilde{B}_{min}) \\ \tilde{A}_{max} & R(\tilde{A}_{max}, \tilde{B}_{min}) \in \{C, CV, O\} & O(\tilde{A}_{max}, \tilde{B}_{max}) \end{array}$$

Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where *Overlap* $(\tilde{A}_{max}, \tilde{B}_{max})$ and *Inside* $(\tilde{A}_{min}, \tilde{B}_{min})$ (1). We suppose now that $R(\tilde{A}_{min}, \tilde{B}_{max}) \notin$ {*Inside*}(2) $R(\tilde{A}_{max}, \tilde{B}_{min}) \notin$ {*Contains, Covers, Overlap*} (3). If (2) then $R(\tilde{B}_{max}, \tilde{B}_{min}) \notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition, there is a contradiction and (2) cannot be true. In the same way, if (3) then $R(\tilde{B}_{max}, \tilde{B}_{min}) \notin$ {*Contains*, *Covers*, *Equal*} or (1) is false. By considering definition 1, (3) cannot be true because there is also a contradiction.

<table>
<tr>
<td>Rule 25: Let $\widetilde{A}$ and $\widetilde{B}$ two simple regions with broad boundaries, if Overlap ($\widetilde{A}_{max}, \widetilde{B}_{max}$) and Covers ($\widetilde{A}_{min}, \widetilde{B}_{min}$) then R ($\widetilde{A}_{min}, \widetilde{B}_{max}$) $\in$ {Inside, Covered by, Overlap} and R($\widetilde{A}_{max}, \widetilde{B}_{min}$)$\in$ {Covers, Contains}, and vice versa.</td>
<td>$$\begin{array}{c c} & \begin{array}{cc} \widetilde{B}_{min} & \widetilde{B}_{max} \end{array} \\ \begin{array}{c} \widetilde{A}_{min} \\ \widetilde{A}_{max} \end{array} & \begin{bmatrix} CV(\widetilde{A}_{min}, \widetilde{B}_{min}) & R(\widetilde{A}_{min}, \widetilde{B}_{max}) \in \{I, CVB, O\} \\ R(\widetilde{A}_{max}, \widetilde{B}_{min}) \in \{CV, C\} & O(\widetilde{A}_{max}, \widetilde{B}_{max}) \end{bmatrix} \end{array}$$</td>
</tr>
<tr>
<td colspan="2">Proof: Let $\widetilde{A}$ and $\widetilde{B}$ two simple regions with broad boundaries where Overlap ($\widetilde{A}_{max}, \widetilde{B}_{max}$) and Covers ($\widetilde{A}_{min}, \widetilde{B}_{min}$) (1). We suppose now that R ($\widetilde{A}_{min}, \widetilde{B}_{max}$)$\notin$ {Inside, Covered by, Overlap} (2) R($\widetilde{A}_{max}, \widetilde{B}_{min}$)$\notin$ {Covers, Contains} (3). If (2) then R ($\widetilde{B}_{max}, \widetilde{B}_{min}$)$\notin$ {Contains, Covers, Equal} or (1) is false. By considering definition 1 and (1), (2) cannot be true because there is a contradiction. In the same way, if (3) then R ($\widetilde{B}_{max}, \widetilde{B}_{min}$)$\notin$ {Contains, Covers, Equal} or (1) is false. By considering definition 1 and (1), (3) cannot be true because there is also a contradiction.</td>
</tr>
<tr>
<td>Rule 26: Let $\widetilde{A}$ and $\widetilde{B}$ two simple regions with broad boundaries, if Overlap($\widetilde{A}_{max}, \widetilde{B}_{max}$) and Disjoint($\widetilde{A}_{min}, \widetilde{B}_{min}$) then R ($\widetilde{A}_{min}, \widetilde{B}_{max}$)$\notin$ {Equal, Contains, Covers} and R($\widetilde{A}_{max}, \widetilde{B}_{min}$)$\notin$ {Equal, Covered by, Inside}, and vice versa.</td>
<td>$$\begin{array}{c c} & \begin{array}{cc} \widetilde{B}_{min} & \widetilde{B}_{max} \end{array} \\ \begin{array}{c} \widetilde{A}_{min} \\ \widetilde{A}_{max} \end{array} & \begin{bmatrix} D(\widetilde{A}_{min}, \widetilde{B}_{min}) & R(\widetilde{A}_{min}, \widetilde{B}_{max}) \notin \{E, C, CV\} \\ R(\widetilde{A}_{max}, \widetilde{B}_{min}) \notin \{E, CVB, I\} & O(\widetilde{A}_{max}, \widetilde{B}_{max}) \end{bmatrix} \end{array}$$</td>
</tr>
<tr>
<td colspan="2">Proof: Let $\widetilde{A}$ and $\widetilde{B}$ two simple regions with broad boundaries where Overlap ($\widetilde{A}_{max}, \widetilde{B}_{max}$) and Disjoint ($\widetilde{A}_{min}, \widetilde{B}_{min}$) (1). We suppose now that R ($\widetilde{A}_{min}, \widetilde{B}_{max}$)$\notin$ {Equal,Contains,Covers} (2) R($\widetilde{A}_{max}, \widetilde{B}_{min}$)$\notin$ {Equal, Covered by, Inside} (3). If (2) then R ($\widetilde{B}_{max}, \widetilde{B}_{min}$)$\notin$ {Contains, Covers, Equal} or (1) is false. By considering definition 1 and (1), (2) cannot be true because there is a contradiction. In the same way, if (3) then R ($\widetilde{B}_{max}, \widetilde{B}_{min}$)$\notin$ {Contains, Covers, Equal} or (1) is false. By considering definition 1 and (1), (3) cannot be true because there is also a contradiction.</td>
</tr>
<tr>
<td>Rule 27: Let $\widetilde{A}$ and $\widetilde{B}$ two simple regions with broad boundaries, if Overlap ($\widetilde{A}_{max}, \widetilde{B}_{max}$) and Covered by ($\widetilde{A}_{min}, \widetilde{B}_{min}$) then R ($\widetilde{A}_{min}, \widetilde{B}_{max}$)$\in$ {Inside, Covered by} and R($\widetilde{A}_{max}, \widetilde{B}_{min}$)$\in$ {Covers, Contains, Overlap}, and vice versa.</td>
<td>$$\begin{array}{c c} & \begin{array}{cc} \widetilde{B}_{min} & \widetilde{B}_{max} \end{array} \\ \begin{array}{c} \widetilde{A}_{min} \\ \widetilde{A}_{max} \end{array} & \begin{bmatrix} CVB(\widetilde{A}_{min}, \widetilde{B}_{min}) & R(\widetilde{A}_{min}, \widetilde{B}_{max}) \in \{I, CVB\} \\ R(\widetilde{A}_{max}, \widetilde{B}_{min}) \in \{CV, C, O\} & O(\widetilde{A}_{max}, \widetilde{B}_{max}) \end{bmatrix} \end{array}$$</td>
</tr>
<tr>
<td colspan="2">Proof: Let $\widetilde{A}$ and $\widetilde{B}$ two simple regions with broad boundaries where Overlap ($\widetilde{A}_{max}, \widetilde{B}_{max}$) and Covered by ($\widetilde{A}_{min}, \widetilde{B}_{min}$) (1). We suppose now that R ($\widetilde{A}_{min}, \widetilde{B}_{max}$)$\notin$ {Inside, Covered by} (2) R($\widetilde{A}_{max}, \widetilde{B}_{min}$)$\notin$ {Covers, Contains, Overlap} (3). If (2) then R ($\widetilde{B}_{max}, \widetilde{B}_{min}$)$\notin$ {Contains, Covers, Equal} or (1) is false. By considering definition 1 and (1), (2) cannot be true because there is a contradiction. In the same way, if (3) then R ($\widetilde{B}_{max}, \widetilde{B}_{min}$)$\notin$ {Contains, Covers, Equal} or (1) is false. By considering definition 1 and (1), (3) cannot be true because there is also a contradiction.</td>
</tr>
<tr>
<td>Rule 28: Let $\widetilde{A}$ and $\widetilde{B}$ two simple regions with broad boundaries, if Contains ($\widetilde{A}_{max}, \widetilde{B}_{max}$) and Disjoint($\widetilde{A}_{min}, \widetilde{B}_{min}$) then R ($\widetilde{A}_{min}, \widetilde{B}_{max}$)$\notin$ {Contains, Covers, Equal} and Contains($\widetilde{A}_{max}, \widetilde{B}_{min}$), and vice versa.</td>
<td>$$\begin{array}{c c} & \begin{array}{cc} \widetilde{B}_{min} & \widetilde{B}_{max} \end{array} \\ \begin{array}{c} \widetilde{A}_{min} \\ \widetilde{A}_{max} \end{array} & \begin{bmatrix} D(\widetilde{A}_{min}, \widetilde{B}_{min}) & R(\widetilde{A}_{min}, \widetilde{B}_{max}) \notin \{C, CV, E\} \\ C(\widetilde{A}_{max}, \widetilde{B}_{min}) & C(\widetilde{A}_{max}, \widetilde{B}_{max}) \end{bmatrix} \end{array}$$</td>
</tr>
</table>

<table>
<tr><td colspan="2">Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where Contains ($\tilde{A}_{max}$, $\tilde{B}_{max}$) and Disjoint ($\tilde{A}_{min}$, $\tilde{B}_{min}$) (1). We suppose now that R($\tilde{A}_{min}$, $\tilde{B}_{max}$)∈ {Contains, Covers, Equal} (2) and R($\tilde{A}_{max}$, $\tilde{B}_{min}$)∉ {Contains} (3). If (2) then R ($\tilde{B}_{max}$, $\tilde{B}_{min}$)∉ {Contains, Covers, Equal} or (1) is false. By considering definition 1 and (1), (2) cannot be true because there is a contradiction. In the same way, if (3) then R ($\tilde{B}_{max}$, $\tilde{B}_{min}$)∉ {Contains, Covers, Equal} or (1) is false. By considering definition 1 and (1), (3) cannot be true because there is also a contradiction.</td></tr>
<tr><td>Rule 29: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries, if Covers ($\tilde{A}_{max}$, $\tilde{B}_{max}$) and Covers ($\tilde{A}_{min}$, $\tilde{B}_{min}$) then R ($\tilde{A}_{min}$, $\tilde{B}_{max}$)∈ {Inside, Covered by, Equal, Overlap, Covers} and R($\tilde{A}_{max}$, $\tilde{B}_{min}$)∈ {Covers, Contains}, and vice versa.</td><td>$$\begin{array}{cc} & \begin{array}{cc} \tilde{B}_{min} & \tilde{B}_{max} \end{array} \\ \begin{array}{c} \tilde{A}_{min} \\ \tilde{A}_{max} \end{array} & \left(\begin{array}{cc} CV(\tilde{A}_{min}, \tilde{B}_{min}) & R(\tilde{A}_{min}, \tilde{B}_{max}) \in \{CVB, O, E, CV, I\} \\ R(\tilde{A}_{max}, \tilde{B}_{min}) \in \{CV, C\} & CV(\tilde{A}_{max}, \tilde{B}_{max}) \end{array} \right) \end{array}$$</td></tr>
<tr><td colspan="2">Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where Covers ($\tilde{A}_{max}$, $\tilde{B}_{max}$) and Covers ($\tilde{A}_{min}$, $\tilde{B}_{min}$) (1). We suppose now that R ($\tilde{A}_{min}$, $\tilde{B}_{max}$)∉ {Covered by, Overlap, Equal, Covers, Inside} (2) R($\tilde{A}_{max}$, $\tilde{B}_{min}$)∉ {Covers, Contains} (3). If (2) then R ($\tilde{B}_{max}$, $\tilde{B}_{min}$)∉ {Contains, Covers, Equal} or (1) is false. By considering definition 1 and (1), (2) cannot be true because there is a contradiction. In the same way, if (3) then R ($\tilde{B}_{max}$, $\tilde{B}_{min}$)∉ {Contains, Covers, Equal} or (1) is false. By considering definition 1 and (1), (3) cannot be true because there is also a contradiction.</td></tr>
<tr><td>Rule 30: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries, if Ru1e 29 and Covers($\tilde{A}_{min}$, $\tilde{B}_{max}$) Then R($\tilde{A}_{max}$, $\tilde{B}_{min}$)∈ {Covers}, and vice versa.</td><td>$$\begin{array}{cc} & \begin{array}{cc} \tilde{B}_{min} & \tilde{B}_{max} \end{array} \\ \begin{array}{c} \tilde{A}_{min} \\ \tilde{A}_{max} \end{array} & \left(\begin{array}{cc} CV(\tilde{A}_{min}, \tilde{B}_{min}) & CV(\tilde{A}_{min}, \tilde{B}_{max}) \\ R(\tilde{A}_{max}, \tilde{B}_{min}) \in \{CV\} & CV(\tilde{A}_{max}, \tilde{B}_{max}) \end{array} \right) \end{array}$$</td></tr>
<tr><td colspan="2">Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where Rule 29 and R ($\tilde{A}_{min}$, $\tilde{B}_{max}$) ∈ {Covers} (1). We suppose now that R($\tilde{A}_{max}$, $\tilde{B}_{min}$)∉ {Covers} (2). By considering definition 1 and (1), if (2) then R ($\tilde{B}_{max}$, $\tilde{B}_{min}$)∉ {Contains, Covers, Equal} or (1) is false. Thus, (2) cannot be true because there is a contradiction.</td></tr>
<tr><td>Rule 31: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries, if Ru1e 29 and R ($\tilde{A}_{min}$, $\tilde{B}_{max}$)∈ {Inside} then R($\tilde{A}_{max}$, $\tilde{B}_{min}$)∈ {Contains}, and vice versa.</td><td>$$\begin{array}{cc} & \begin{array}{cc} \tilde{B}_{min} & \tilde{B}_{max} \end{array} \\ \begin{array}{c} \tilde{A}_{min} \\ \tilde{A}_{max} \end{array} & \left(\begin{array}{cc} CV(\tilde{A}_{min}, \tilde{B}_{min}) & R(\tilde{A}_{min}, \tilde{B}_{max}) \in \{I\} \\ R(\tilde{A}_{max}, \tilde{B}_{min}) \in \{C\} & CV(\tilde{A}_{max}, \tilde{B}_{max}) \end{array} \right) \end{array}$$</td></tr>
<tr><td colspan="2">Proof: Let $\tilde{A}$ and $\tilde{B}$ two simple regions with broad boundaries where Rule 29 and R ($\tilde{A}_{min}$, $\tilde{B}_{max}$) ∈ {Inside} (1). We suppose now that R($\tilde{A}_{max}$, $\tilde{B}_{min}$)∉ {Contains} (2). If (2) then R ($\tilde{B}_{max}$, $\tilde{B}_{min}$)∉ {Contains, Covers, Equal} or (1) is false. By considering definition 1 and (1), (2) cannot be true because there is a contradiction.</td></tr>
</table>

Appendix 3: Demonstrations of the possible topological relationships between regions with broad boundaries resulted from an integration process

In this appendix, we prove the results obtained in Section 5.7. For the next proofs, we use the following terminology:

Let ∂A_1 the boundary of A_1, $\overset{\circ}{A}_1$ its interior and $\overline{A_1}$ its closure
∂A_2 the boundary of A_2, $\overset{\circ}{A}_2$ its interior and $\overline{A_2}$ its closure
∂A_n the boundary of A_n, $\overset{\circ}{A}_n$ its interior and $\overline{A_n}$ its closure
∂B_1 the boundary of B_1, $\overset{\circ}{B}_1$ its interior and $\overline{B_1}$ its closure
∂B_2 the boundary of B_2, $\overset{\circ}{B}_2$ its interior and $\overline{B_2}$ its closure
∂B_n the boundary of B_n, $\overset{\circ}{B}_n$ its interior and $\overline{B_n}$ its closure

A3.1 Disjoint

Let *Disjoint*(A_1, B_1), *Disjoint*(A_2, B_2),..and *Disjoint*(A_n, B_n) with A_1, A_2,.. and A_n the available heterogeneous of *A* and B_1, B_2,.. and B_n the available heterogeneous representations of *B*. According to Section 5.7, the final geometries should conform to the specifications of the next matrix:

$$\begin{pmatrix} \text{Disjoint } (I_A, I_B) & \text{--} \\ \text{--} & \{\textit{Disjoint, Meet, Overlap}\}\ (U_A, U_B) \end{pmatrix}$$

- *Disjoint* (I_A, I_B)

In order to prove that *Disjoint*(I_A, I_B), we should demonstrate that $(\overline{A_1} \cap \overline{A_2} \,..\cap \overline{A_n}) \cap (\overline{B_1} \cap \overline{B_2} \ldots \cap \overline{B_n}) = \varnothing$

(1) $\forall\ x \in (\overline{A_1} \cap \overline{A_2} \,..\cap \overline{A_n})$, we have x $\in \overline{A_1}$, x $\in \overline{A_2}$,.. and x $\in \overline{A_n}$.
Is-it possible for *x* to be an element of $(\overline{B_1} \cap \overline{B_2} \ldots \cap \overline{B_n})$?

(2) If $x \in \overline{B_1}$ then, there is a contradiction because $\overline{A_1} \cap \overline{B_1} = \varnothing$
(3) If $x \in \overline{B_2}$ then, there is a contradiction because $\overline{A_2} \cap \overline{B_2} = \varnothing$
(4) If $x \in \overline{B_n}$ then, there is a contradiction because $\overline{A_n} \cap \overline{B_n} = \varnothing$

(5) According to (2), (3) and (4), $x \notin (\overline{B_1} \cap \overline{B_2} \ldots \cap \overline{B_n})$

Finally, (1) and (5) means that $(\overline{A}_1 \cap \overline{A}_2 .. \cap \overline{A}_n) \cap (\overline{B}_1 \cap \overline{B}_2 \ldots \cap \overline{B}_n) = \varnothing$ and so *Disjoint*$((A_1 \cap A_2 .. \cap A_n), (B_1 \cap B_2 \ldots \cap B_n))$ that we write *Disjoint* (I_A, I_B) (*I* for intersection). In addition, (1) and (5) show that $(\overline{A}_1 \cup \overline{A}_2 .. \cup \overline{A}_n) \not\subset (\overline{B}_1 \cup \overline{B}_2 \ldots \cup \overline{B}_n)$ because $(\overline{A}_1 \cap \overline{A}_2 .. \cap \overline{A}_n) \not\subset (\overline{A}_1 \cup \overline{A}_2 .. \cup \overline{A}_n)$ and $(\overline{A}_1 \cap \overline{A}_2 .. \cap \overline{A}_n) \not\subset (\overline{B}_1 \cup \overline{B}_2 \ldots \cup \overline{B}_n)$

- {*Disjoint, Meet, Overlap*}(U_A,U_B)

➢ For *Overlap* (U_A,U_B), we should prove that
if $((\overset{\circ}{A}_1 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)) \cup (\overset{\circ}{A}_2 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)) .. \cup (\overset{\circ}{A}_n \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n))) \neq \varnothing$ then
$((\overset{\circ}{A}_1 \cup \overset{\circ}{A}_2 .. \cup \overset{\circ}{A}_n) \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)) \neq \varnothing$.

Let $x \in (\overset{\circ}{A}_1 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)) \cup (\overset{\circ}{A}_2 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)) .. \cup (\overset{\circ}{A}_n \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)))$, we have $x \in (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)$

(1) If $\overset{\circ}{A}_1 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) = \varnothing$ then ($x \in (\overset{\circ}{A}_2 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) .. \cup (\overset{\circ}{A}_n \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)))$ else there is a contradiction. Indeed, $x \in \overset{\circ}{A}_2$ or $x \in \overset{\circ}{A}_3$ or,.., $x \in \overset{\circ}{A}_n$

(2) If $\overset{\circ}{A}_2 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) = \varnothing$ then ($x \in (\overset{\circ}{A}_1 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) .. \cup (\overset{\circ}{A}_n \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)))$ else there is a contradiction. $x \in \overset{\circ}{A}_1$ or $x \in \overset{\circ}{A}_3$ or,.., $x \in \overset{\circ}{A}_n$

(3) If $\overset{\circ}{A}_n \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) = \varnothing$ then ($x \in (\overset{\circ}{A}_1 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) \cup (\overset{\circ}{A}_2 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) .. \cup (\overset{\circ}{A}_{n-1} \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)))$
else there is a contradiction. $x \in \overset{\circ}{A}_1$ or $x \in \overset{\circ}{A}_2$ or .. $x \in \overset{\circ}{A}_{n-1}$.

(1), (2) and (3) mean that $x \in (\overset{\circ}{A}_1 \cup \overset{\circ}{A}_2 .. \cup \overset{\circ}{A}_n)$.
$((\overset{\circ}{A}_1 \cup \overset{\circ}{A}_2 .. \cup \overset{\circ}{A}_n) \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)) \neq \varnothing$ and so *Overlap*$((A_1 \cup A_2 .. \cup A_n), (B_1 \cup B_2 \ldots \cup B_n))$ that we write *Overlap* (U_A,U_B).

➢ For Meet (U_A,U_B): a *Meet* relationship is possible between unions if there is only intersection between their boundaries, we suppose that interiors does not intersect so we should prove that
if $((\partial A_1 \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)) \cup (\partial A_2 \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_2)) .. \cup (\partial A_n \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n))) \neq \varnothing$ then
$((\partial A_1 \cup \partial A_2 .. \cup \partial A_n) \cap ((\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n))) \neq \varnothing$

Let $x \in (\partial A_1 \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)) \cup (\partial A_2 \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)) .. \cup (\partial A_n \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)))$, we have $x \in (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)$

(1) If $\partial A_1 \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)= \varnothing$ then ($x \in (\partial A_2 \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)$ $.. \cup (\partial A_n \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)$)) else there is a contradiction. Indeed, $x \in \partial A_2$ or $x \in \partial A_3$ or, .., $x \in \partial A_n$.

(2) If $\partial A_2 \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)= \varnothing$ then ($x \in (\partial A_1 \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)$ $.. \cup (\partial A_n \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)$)) else there is a contradiction. $x \in \partial A_1$ or $x \in \partial A_3$ or, .., $x \in \partial A_n$.

(3) If $\partial A_n \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)= \varnothing$ then ($x \in (\partial A_1 \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n) \cup (\partial A_2 \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)$ $.. \cup (\partial A_{n-1} \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)$))) else there is a contradiction. $x \in \partial A_1$ or $x \in \partial A_2$ or, .., $x \in \partial A_{n-1}$

(1), (2) and (3) mean that $x \in (\partial A_1 \cup \partial A_2 \; .. \cup \partial A_n)$.
Indeed, $((\partial A_1 \cup \partial A_2 \; .. \cup \partial A_n) \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)) \neq \varnothing$ and so
Meet(($A_1 \cup A_2 \, .. \cup A_n$), ($B_1 \cup B_2 \ldots \cup B_n$)) that we write *Meet* (U_A,U_B).

- For *Disjoint* (U_A,U_B), a Disjoint relation is possible between unions if there is no intersection respectively between their boundaries and interiors, we should prove that If $((\overset{\circ}{A}_1 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)) \cup (\overset{\circ}{A}_2 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)) .. \cup (\overset{\circ}{A}_n \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)))=\varnothing$ and $((\partial A_1 \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_2)) \cup (\partial A_2 \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_2)) .. \cup (\partial A_n \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_2)))=\varnothing$ then $((\overline{A}_1 \cup \overline{A}_2 \, .. \cup \overline{A}_n) \cap (\overline{B}_1 \cup \overline{B}_2 \ldots \cup \overline{B}_n)) =\varnothing$

(1) Let $x \in (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)$, then
If $x \in \overset{\circ}{A}_1$, there is a contradiction because $\overset{\circ}{A}_1 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)=\varnothing$
If $x \in \overset{\circ}{A}_2$, there is a contradiction because $\overset{\circ}{A}_2 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)=\varnothing$
If $x \in \overset{\circ}{A}_n$, there is a contradiction because $\overset{\circ}{A}_n \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)=\varnothing$

Indeed, $x \notin (\overset{\circ}{A}_1 \cup \overset{\circ}{A}_2 \, .. \cup \overset{\circ}{A}_n)$ and so
$((\overset{\circ}{A}_1 \cup \overset{\circ}{A}_2 \, .. \cup \overset{\circ}{A}_n) \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n))=\varnothing$

Let y $(\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)$, then
If $y \in \partial A_1$, there is a contradiction because $\partial A_n \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)=\varnothing$
If $y \in \partial A_2$, there is a contradiction because $\partial A_2 \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)=\varnothing$
If $y \in \partial A_n$, there is a contradiction because $\partial A_n \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n)=\varnothing$

(2) Indeed, $y \notin (\partial A_1 \cup \partial A_2 \, .. \cup \partial A_n)$ and so
$((\partial A_1 \cup \partial A_2 \, .. \cup \partial A_n) \cap (\partial B_1 \cup \partial B_2 \ldots \cup \partial B_n))=\varnothing$ (6)

(1) and (2) mean that $((\overline{A}_1 \cup \overline{A}_2 .. \cup \overline{A}_n) \cap (\overline{B}_1 \cup \overline{B}_2 \ldots \cup \overline{B}_n)) = \emptyset$ and so Disjoint $((A_1 \cup A_2 .. \cup A_n),\ (B_1 \cup B_2 \ldots \cup B_n))$ that we write *Disjoint* (U_A, U_B).

A3.2 Contains/Inside

Let *Contains*(A_1, B_1), *Contains*(A_2, B_2),..and *Contains*(A_n, B_n) with A_1, A_2,.. and A_n the available heterogeneous of A and B_1, B_2,.. and B_n the available heterogeneous representations of B. Then, we have $\overline{B}_1 \subset \overline{A}_1$, $\overline{B}_2 \subset \overline{A}_2$,.. and $\overline{B}_n \subset \overline{A}_n$. The final geometries should conform to the specifications of one of the next matrices:

➔ For *Contains* (A, B)

$$\begin{array}{cc} & \begin{array}{cc} I_B & \qquad\qquad\qquad\qquad U_B \end{array} \\ \begin{array}{c} I_A \\ \\ U_A \end{array} & \begin{bmatrix} Contains\ (I_A, I_B) & -- \\ \\ -- & Contains\ (U_A, U_B) \end{bmatrix} \end{array}$$

➔ For *Inside* (A, B)

$$\begin{array}{cc} & \begin{array}{cc} I_B & \qquad\qquad\qquad\qquad U_B \end{array} \\ \begin{array}{c} I_A \\ \\ U_A \end{array} & \begin{bmatrix} Inside\ (I_A, I_B) & -- \\ \\ -- & Inside\ (U_A, U_B) \end{bmatrix} \end{array}$$

- *Contains* (I_A, I_B)

In order to prove that *Contains*(I_A, I_B), we should demonstrate that $(\overline{B}_1 \cap \overline{B}_2 \ldots \cap \overline{B}_n) \subset (\overline{A}_1 \cap \overline{A}_2 .. \cap \overline{A}_n)$

$\forall\ x \in (\overline{B}_1 \cap \overline{B}_2 \ldots \cap \overline{B}_n)$, we have $x \in \overline{B}_1$, $x \in \overline{B}_2$,.. and $x \in \overline{B}_n$

(1) If $x \notin \overline{A}_1$ then, there is a contradiction because $\overline{B}_1 \subset \overline{A}_1$
(2) If $x \notin \overline{A}_2$ then, there is a contradiction because $\overline{B}_2 \subset \overline{A}_2$
(3) If $x \notin \overline{A}_n$ then, there is a contradiction because $\overline{B}_n \subset \overline{A}_n$

(1), (2) and (3) mean $x \in \overline{A}_1$, $x \in \overline{A}_2$ and $x \in \overline{A}_n$; so $x \in (\overline{A}_1 \cap \overline{A}_2 .. \cap \overline{A}_n)$. Indeed, $(\overline{B}_1 \cap \overline{B}_2 \ldots \cap \overline{B}_n) \subset (\overline{A}_1 \cap \overline{A}_2 .. \cap \overline{A}_n)$.

Finally, we have the intersection of the closures of A's representations contains that of B's representations. Then, we can conclude *Contains*$((A_1 \cap A_2 .. \cap A_n), (B_1 \cap B_2 \ldots \cap B_n))$ that we write *Contains* (I_A, I_B).

- *Contains* (U_A,U_B)

In order to prove that *Contains*(U_A,U_B), we should demonstrate that $(\overline{B}_1 \cup \overline{B}_2 \ldots \cup \overline{B}_n) \subset$ $(\overline{A}_1 \cup \overline{A}_2 .. \cup \overline{A}_n)$

$\forall \; x \in (\overline{B}_1 \cup \overline{B}_2 \ldots \cup \overline{B}_n)$, we have $x \in \overline{B}_1$, $x \in \overline{B}_2$,.. or $x \in \overline{B}_n$. In addition, we have $\overline{B}_1 \subset \overline{A}_1$, $\overline{B}_2 \subset \overline{A}_2$,.. and $\overline{B}_n \subset \overline{A}_n$

(1) If $x \notin \overline{A}_1$ then $x \in \overline{B}_2$ ..or $x \in \overline{B}_n$, else there is a contradiction because $\overline{B}_1 \subset \overline{A}_1$

(2) If $x \notin \overline{A}_2$ then $x \in \overline{B}_1$ ..or $x \in \overline{B}_n$, else there is a contradiction because $\overline{B}_2 \subset \overline{A}_2$

(3) If $x \notin \overline{A}_n$ then $x \in \overline{B}_1$ or $x \in \overline{B}_2$, else there is a contradiction because $\overline{B}_n \subset \overline{A}_n$

(1), (2) and (3) mean $x \in \overline{A}_1$, $x \in \overline{A}_2$ or $x \in \overline{A}_n$; so x $(\overline{A}_1 \cup \overline{A}_2 .. \cup \overline{A}_n)$ Indeed, $(\overline{B}_1 \cup \overline{B}_2 \ldots \cup \overline{B}_n) \subset (\overline{A}_1 \cup \overline{A}_2 .. \cup \overline{A}_n)$.

Finally, we have the union of the closures of A's representations contains that of B's representations. Then, we can conclude *Contains* (($\overline{A}_1 \cup \overline{A}_2 ... \cup \overline{A}_n$), ($\overline{B}_1 \cup \overline{B}_2 \ldots \cup \overline{B}_n$)) that we write *Contains* (U_A,U_B).

A3.3 Covers/Covered by

Let *Covers*(A_1, B_1), *Covers*(A_2, B_2),..and *Covers*(A_n, B_n) with A_1, A_2,.. and A_n the available heterogeneous of *A* and B_1, B_2,.. and B_n the available heterogeneous representations of *B*. Then, we have $\overset{\circ}{B}_1 \subset \overset{\circ}{A}_1$, $\overset{\circ}{B}_2 \subset \overset{\circ}{A}_2$,.. and $\overset{\circ}{B}_n \subset \overset{\circ}{A}_n$ (1). In addition, we have $\partial A_1 \cap \partial B_1 = \partial A_1 B_1 \neq \varnothing$, $\partial A_2 \cap \partial B_2 = \partial A_2 B_2 \neq \varnothing$,.. and $\partial A_n \cap \partial B_n = \partial A_n B_n \neq \varnothing$. According to Section 5.7, the final geometries should conform to the specifications of the next matrix:

➔ For *Covers* (*A*, *B*)

$$\begin{array}{cc} I_B & U_B \\ \left[\begin{array}{cc} \{\textit{Covers, Contains}\}\ (I, I) & -- \\ -- & \textit{Covers}\ (U_A, U_B) \end{array}\right] \end{array}$$

➔ For *Covered by* (*A*, *B*)

$$\begin{array}{c|cc} & I_B & U_B \\ \hline I_A & \{\textit{Covered by, Inside}\}\ (I_A, I_B) & -- \\ U_A & -- & \textit{Covered by}\ (U_A, U_B) \end{array}$$

- *Covers* (U_A,U_B)

In order to prove that *Covers*(U_A,U_B), we should demonstrate that $(\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) \subset (\overset{\circ}{A}_1 \cup \overset{\circ}{A}_2 .. \cup \overset{\circ}{A}_n)$ and $(\partial A_1 \cup \partial A_2 .. \cup \partial A_n) \cap (\partial B_1 \cup \partial B_2 .. \cup \partial B_n) \neq \varnothing$.

(1) $\forall\ x \in (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)$, we have $x \in \overset{\circ}{B}_1$, $x \in \overset{\circ}{B}_2$,.. or $x \in \overset{\circ}{B}_n$

If $x \notin \overset{\circ}{A}_1$ then $x \in \overset{\circ}{B}_2$ ..or $x \in \overset{\circ}{B}_n$, else there is a contradiction because $\overset{\circ}{B}_1 \subset \overset{\circ}{A}_1$
If $x \notin \overset{\circ}{A}_2$ then $x \in \overset{\circ}{B}_1$ ..or $x \in \overset{\circ}{B}_n$, else there is a contradiction because $\overset{\circ}{B}_2 \subset \overset{\circ}{A}_2$
If $x \notin \overset{\circ}{A}_n$ then $x \in \overset{\circ}{B}_1$ or $x \in \overset{\circ}{B}_2$, else there is a contradiction because $\overset{\circ}{B}_n \subset \overset{\circ}{A}_n$

(2) Indeed, $x \in \overset{\circ}{A}_1$ or $x \in \overset{\circ}{A}_2$ ..or $x \in \overset{\circ}{A}_n$ and so $x \in (\overset{\circ}{A}_1 \cup \overset{\circ}{A}_2 .. \cup \overset{\circ}{A}_n)$
Because (1), (2) means that $(\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) \subset (\overset{\circ}{A}_1 \cup \overset{\circ}{A}_2 .. \cup \overset{\circ}{A}_n)$.

$\forall\ x/\ x \in (\partial B_1 \cup \partial B_2 .. \cup \partial B_n)$ and $x \in \partial A_1 B_1$ or $x \in \partial A_2 B_2$ ..or $x \in \partial A_n B_n$, we have $x \in \partial B_1$, $x \in \partial B_2$,.. or $x \in \partial B_n$.

If $x \in \partial B_1$ then $x \in \partial A_1$, else there is a contradiction because $\partial A_1 \cap \partial B_1 = \partial A_1 B_1 \neq \varnothing$
If $x \in \partial B_2$ then $x \in \partial A_2$, else there is a contradiction because $\partial A_2 \cap \partial B_2 = \partial A_2 B_2 \neq \varnothing$
If $x \in \partial B_n$ then $x \in \partial A_n$ else there is a contradiction because $\partial A_n \cap \partial B_n = \partial A_n B_n \neq \varnothing$

Indeed, $x \in (\partial A_1 \cup \partial A_2 .. \cup \partial A_n)$ and so $(\partial A_1 \cup \partial A_2 .. \cup \partial A_n) \cap (\partial B_1 \cup \partial B_2 .. \cup \partial B_n) \neq \varnothing$ **(3)**

Finally, (1), (2) and (3) mean that *Covers* $((\overline{A}_1 \cup \overline{A}_2 ... \cup \overline{A}_n)$, $(\overline{B}_1 \cup \overline{B}_2 \ldots \cup \overline{B}_n))$ that we write *Covers* (U_A,U_B).

- {Contains, Covers}(I_A, I_B)

  - Contains(I_A, I_B)

In order to prove that *Contains* (I_A, I_B), we should demonstrate that $(\overset{\circ}{B}_1 \cap \overset{\circ}{B}_2 \ldots \cap \overset{\circ}{B}_n) \subset (\overset{\circ}{A}_1 \cap \overset{\circ}{A}_2 .. \cap \overset{\circ}{A}_n)$ and $(\partial A_1 \cap \partial A_2 .. \cap \partial A_n) \cap (\partial B_1 \cap \partial B_2 .. \cap \partial B_n) = \varnothing$.

$\forall\ x/\ x \in (\partial B_1 \cap \partial B_2 .. \cap \partial B_n)$ but $x \notin \partial A_1 B_1$ and $x \notin \partial A_2 B_2$ ..and $x \notin \partial A_n B_n$, we have $x \in \partial B_1$, $x \in \partial B_2$,.. and $x \in \partial B_n$.

If $x \in \partial B_1$ then $x \in \partial A_1$, else there is a contradiction because $x \notin \partial A_1 B_1$
If $x \in \partial B_2$ then $x \in \partial A_2$, else there is a contradiction because $x \notin \partial A_2 B_2$
If $x \in \partial B_n$ then $x \in \partial A_n$ else there is a contradiction because $x \notin \partial A_n B_n$

Indeed, $x \notin \partial A_1$, $x \notin \partial A_2$,.. and $x \notin \partial A_n$; so $x \notin (\partial A_1 \cap \partial A_2 .. \cap \partial A_n)$. Then, $(\partial A_1 \cap \partial A_2 .. \cap \partial A_n) \cap (\partial B_1 \cap \partial B_2 .. \cap \partial B_n) = \varnothing$ **(1)**

Now, $\forall\ x/\ x \in (\overset{\circ}{B}_1 \cap \overset{\circ}{B}_2 \ldots \cap \overset{\circ}{B}_n)$, we have $x \in \overset{\circ}{B}_1$, $x \in \overset{\circ}{B}_2, \ldots$ and $x \in \overset{\circ}{B}_n$.

If $x \notin \overset{\circ}{A}_1$ there is a contradiction because $\overset{\circ}{B}_1 \subset \overset{\circ}{A}_1$
If $x \notin \overset{\circ}{A}_2$ there is a contradiction because $\overset{\circ}{B}_2 \subset \overset{\circ}{A}_2$
If $x \notin \overset{\circ}{A}_n$ there is a contradiction because $\overset{\circ}{B}_n \subset \overset{\circ}{A}_n$

Indeed, $x \in (\overset{\circ}{A}_1 \cap \overset{\circ}{A}_2 .. \cap \overset{\circ}{A}_n)$; and so $(\overset{\circ}{B}_1 \cap \overset{\circ}{B}_2 \ldots \cap \overset{\circ}{B}_n) \subset (\overset{\circ}{A}_1 \cap \overset{\circ}{A}_2 .. \cap \overset{\circ}{A}_n)$ **(2)**

(1) and (2) mean that *Contains* $((\overline{A}_1 \cap \overline{A}_2 ... \cap \overline{A}_n), (\overline{B}_1 \cap \overline{B}_2 \ldots \cap \overline{B}_n))$, that we write *Contains* (I_A, I_B).

- Covers(I_A, I_B)

In order to prove that *Covers* (I_A, I_B), we should demonstrate that $(\overset{\circ}{B}_1 \cap \overset{\circ}{B}_2 \ldots \cap \overset{\circ}{B}_n) \subset (\overset{\circ}{A}_1 \cap \overset{\circ}{A}_2 .. \cap \overset{\circ}{A}_n)$ and $(\partial A_1 \cap \partial A_2 .. \cap \partial A_n) \cap (\partial B_1 \cap \partial B_2 .. \cap \partial B_n) \neq \varnothing$.

With **(2)**, we have $(\overset{\circ}{B}_1 \cap \overset{\circ}{B}_2 \ldots \cap \overset{\circ}{B}_n) \subset (\overset{\circ}{A}_1 \cap \overset{\circ}{A}_2 .. \cap \overset{\circ}{A}_n)$.

Now, $\forall\ x/\ x \in (\partial B_1 \cap \partial B_2 .. \cap \partial B_n)$ and $x \in \partial A_1 B_1$ and $x \in \partial A_2 B_2$ ..and $x \in \partial A_n B_n$, we have $x \in \partial B_1$, $x \in \partial B_2, ..$ and $x \in \partial B_n$.

If $x \notin \partial A_1$, else there is a contradiction because $x \in \partial A_1 B_1$
If $x \notin \partial A_2$, else there is a contradiction because $x \in \partial A_2 B_2$
If $x \notin \partial A_n$ else there is a contradiction because $x \in \partial A_n B_n$

Indeed, $x \in \partial A_1$, $x \in \partial A_2, ..$ and $x \in \partial A_n$; so $x \in (\partial A_1 \cap \partial A_2 .. \cap \partial A_n)$. Then, $(\partial A_1 \cap \partial A_2 .. \cap \partial A_n) \cap (\partial B_1 \cap \partial B_2 .. \cap \partial B_n) = \varnothing$ **(3)**

(2) and (3) mean that *Covers* $((\overline{A}_1 \cap \overline{A}_2 ... \cap \overline{A}_n), (\overline{B}_1 \cap \overline{B}_2 \ldots \cap \overline{B}_n))$, that we write *Covers*(I_A, I_B).

A3.4 Overlap

Let *Overlap*(A_1, B_1), *Overlap*(A_2, B_2),..and *Overlap*(A_n, B_n) with A_1, $A_2, ..$ and A_n the available heterogeneous of A and B_1, $B_2, ..$ and B_n the available heterogeneous representations of B. Then, we have $\overset{\circ}{B}_1 \cap \overset{\circ}{A}_1 \neq \varnothing$, $\overset{\circ}{B}_2 \cap \overset{\circ}{A}_2 \neq \varnothing, ..$ and $\overset{\circ}{B}_n \cap \overset{\circ}{A}_n \neq \varnothing$. In addition, we have $\partial A_1 \cap \partial B_1 = \partial A_1 B_1 \neq \varnothing$, $\partial A_2 \cap \partial B_2 = \partial A_2 B_2 \neq \varnothing, ..$ and $\partial A_n \cap \partial B_n = \partial A_n B_n \neq \varnothing$. According to Section 5.7, the final geometries should conform to the specifications of the next matrix:

$$\begin{array}{cc} & \begin{array}{cc} I_B & U_B \end{array} \\ \begin{array}{c} I_A \\ U_A \end{array} & \begin{bmatrix} \{\textit{Overlap, Meet, Disjoint}\}\ (I_A, I_B) & -- \\ -- & \textit{Overlap}\ (U_A, U_B) \end{bmatrix} \end{array}$$

- {*Overlap, Meet, Disjoint*}(I_A,I_B)

In this case, we should prove that the relationship $R(I_A,I_B) \neq$ {*Contains, Inside, Covers, Covered by*}.

- $R \neq$ Contains(I_A, I_B)

In order to prove that *Contains* (I_A, I_B), we should demonstrate that $(\overset{\circ}{B}_1 \cap \overset{\circ}{B}_2 \ldots \cap \overset{\circ}{B}_n) \not\subset (\overset{\circ}{A}_1 \cap \overset{\circ}{A}_2 .. \cap \overset{\circ}{A}_n)$

Let x/ $x \notin (\overset{\circ}{B}_1 \cap \overset{\circ}{B}_2 \ldots \cap \overset{\circ}{B}_n)$ and $x \in (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)$, then $x \in \overset{\circ}{B}_1$ or $x \in \overset{\circ}{B}_1$ or .. $x \in \overset{\circ}{B}_n$

(1) If $x \notin \overset{\circ}{B}_1$, then $x \notin \overset{\circ}{A}_1$, else there is a contradiction because $\overset{\circ}{B}_1 \cap \overset{\circ}{A}_1 \neq \varnothing$
(2) If $x \notin \overset{\circ}{B}_2$ then $x \notin \overset{\circ}{A}_2$, else there is a contradiction because $\overset{\circ}{B}_2 \cap \overset{\circ}{A}_2 \neq \varnothing$
(3) If $x \notin \overset{\circ}{B}_n$ then $x \notin \overset{\circ}{A}_n$ there is a contradiction because $\overset{\circ}{B}_n \cap \overset{\circ}{A}_n \neq \varnothing$

(1), **(2)** et **(3)** show that if $x \notin (\overset{\circ}{B}_1 \cap \overset{\circ}{B}_2 \ldots \cap \overset{\circ}{B}_n)$ then $x \notin (\overset{\circ}{A}_1 \cap \overset{\circ}{A}_2 .. \cap \overset{\circ}{A}_n)$. Consequently, $(\overset{\circ}{B}_1 \cap \overset{\circ}{B}_2 \ldots \cap \overset{\circ}{B}_n) \not\subset (\overset{\circ}{A}_1 \cap \overset{\circ}{A}_2 .. \cap \overset{\circ}{A}_n)$.

In conclusion, $R \neq$ *Contains*(I_A, I_B). The same demonstration may be made for $R \neq$ *Inside*(I_A, I_B), i.e. it is required to demonstrate that $(\overset{\circ}{A}_1 \cap \overset{\circ}{A}_2 .. \cap \overset{\circ}{A}_n) \not\subset (\overset{\circ}{B}_1 \cap \overset{\circ}{B}_2 \ldots \cap \overset{\circ}{B}_n)$.

Since the interior of the first intersection I_A (or I_B) cannot be inside the second intersection I_B (or I_A). It is possible to conclude that *Covers* and *Covered by* are also impossible

- For *Overlap* (U_A, U_B) we should prove that

if $((\overset{\circ}{A}_1 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)) \cup (\overset{\circ}{A}_2 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)) .. \cup (\overset{\circ}{A}_n \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n))) \neq \varnothing$ then $((\overset{\circ}{A}_1 \cup \overset{\circ}{A}_2 .. \cup \overset{\circ}{A}_n) \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)) \neq \varnothing$.

Let $x \in (\overset{\circ}{A}_1 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)) \cup (\overset{\circ}{A}_2 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)) .. \cup (\overset{\circ}{A}_n \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)))$, we have $x \in (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)$

(1) If $\overset{\circ}{A}_1 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) = \varnothing$ then ($x \in (\overset{\circ}{A}_2 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) .. \cup (\overset{\circ}{A}_n \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)))$ else there is a contradiction. Indeed, $x \in \overset{\circ}{A}_2$ or $x \in \overset{\circ}{A}_3$ or,.., $x \in \overset{\circ}{A}_n$

(2) If $\overset{\circ}{A}_2 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) = \varnothing$ then ($x \in (\overset{\circ}{A}_1 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) .. \cup (\overset{\circ}{A}_n \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)))$ else there is a contradiction. $x \in \overset{\circ}{A}_1$ or $x \in \overset{\circ}{A}_3$ or,.., $x \in \overset{\circ}{A}_n$

(3) If $\overset{\circ}{A}_n \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) = \varnothing$ then ($x \in (\overset{\circ}{A}_1 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) \cup (\overset{\circ}{A}_2 \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n) .. \cup (\overset{\circ}{A}_{n-1} \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)))$)
else there is a contradiction. $x \in \overset{\circ}{A}_1$ or $x \in \overset{\circ}{A}_2$ or .. $x \in \overset{\circ}{A}_{n-1}$.

(1), (2) and (3) mean that $x \in (\overset{\circ}{A}_1 \cup \overset{\circ}{A}_2 .. \cup \overset{\circ}{A}_n)$.
$((\overset{\circ}{A}_1 \cup \overset{\circ}{A}_2 .. \cup \overset{\circ}{A}_n) \cap (\overset{\circ}{B}_1 \cup \overset{\circ}{B}_2 \ldots \cup \overset{\circ}{B}_n)) \neq \varnothing$ and so
Overlap$((A_1 \cup A_2 .. \cup A_n), (B_1 \cup B_2 \ldots \cup B_n))$ that we write *Overlap* (U_A, U_B).

A3.5 Meet

In this case, we assume that *Meet*(A_1, B_1), *Meet*(A_2, B_2),..and *Meet*(A_n, B_n) with A_1, A_2,.. and A_n the available heterogeneous of A and B_1, B_2,.. and B_n the available heterogeneous representations of B. Then, we have $\overset{\circ}{B}_1 \cap \overset{\circ}{A}_1 = \varnothing$, $\overset{\circ}{B}_2 \cap \overset{\circ}{A}_2 = \varnothing$,.. and $\overset{\circ}{B}_n \cap \overset{\circ}{A}_n = \varnothing$. In addition, we have $\partial A_1 \cap \partial B_1 = \partial A_1 B_1 \neq \varnothing$, $\partial A_2 \cap \partial B_2 = \partial A_2 B_2 \neq \varnothing$,.. and $\partial A_n \cap \partial B_n = \partial A_n B_n \neq \varnothing$. According to Section 5.7, the final geometries should conform to the specifications of the next matrices:

- *Meet*(I_A, I_B):

$$\begin{array}{cc} & \begin{array}{cc} I_B & U_B \end{array} \\ \begin{array}{c} I_A \\ U_A \end{array} & \left(\begin{array}{cc} \textit{Meet}\ (I_A, I_B) & \text{--} \\ \text{--} & \{\textit{Overlap}, \textit{Meet}\}\ (U_A, U_B) \end{array}\right) \end{array}$$

- *Disjoint*(I_A, I_B)

$$\begin{array}{cc} & \begin{array}{cc} I_B & U_B \end{array} \\ \begin{array}{c} I_A \\ U_A \end{array} & \left(\begin{array}{cc} \textit{Disjoint}\ (I_A, I_B) & \text{--} \\ \text{--} & \{\textit{Overlap}, \textit{Meet}\}\ (U_A, U_B) \end{array}\right) \end{array}$$

The demonstrations of *Meet* (I_A, I_B) and *Disjoint* (I_A, I_B) are identical to those presented in the *Overlap* case (see Section A3.4 of the appendix).

The demonstrations of *Overlap*(U_A, U_B) and *Meet*(U_A, U_B) are identical to those presented in the *Disjoint* case (see Section A3.1 of the appendix)

Appendix 4: Extrait de la convention de cotutelle

CONVENTION DE COTUTELLE DE THÈSE

PRÉAMBULE

Conformément aux dispositions et modalités arrêtées dans la «Convention-cadre de cotutelle de thèse» signée entre la CPU, la CDEFI et la CREPUQ le 18 octobre 1996, et mise à jour en mars 1997,

LA PRÉSENTE CONVENTION EST CONCLUE ENTRE :

L'établissement français : l'Université **Blaise Pascal**

représenté par son président, M. Pascal Albert ODOUARD

ET

Le Centre National du Machinisme Agricole du Génie Rural, des Eaux et des Forêts,
Etablissement Public à caractère Scientifique et Technologique (EPST),
désigné ci-après par " Cemagref ",
ayant son siège, parc de Tourvoie, 92160 Antony, France
représenté par son Directeur Régional de Clermont Ferrand, Monsieur Didier Mechineau, agissant au nom et pour le compte du Directeur Général du Cemagref,

ET

L'Université Laval représentée par le vice-doyen de la Faculté des études supérieures, Monsieur Gérard Charlet, qui agit à titre de représentant de la vice-rectrice aux études.

Elle concerne :

M^{me} ou M. **Lotfi Bejaoui**
Née ou né le 17-10-1981

De nationalité Tunisienne

<u>MODALITÉS ADMINISTRATIVES</u>

ARTICLE 1 - Inscription

(Le doctorant s'inscrit obligatoirement, simultanément à temps complet dans les deux établissements.)

- Le doctorant est inscrit :

1) à <u>l'Université Blaise Pascal</u>
   au doctorat, spécialité sciences pour l'ingénieur
   à compter de la rentrée universitaire 2005-2006

ET

2) à l'Université Laval, programme de doctorat en sciences géomatiques
   à compter de la session hiver 2006

- **Droits d'inscription et de scolarité**

  Le doctorant ne paiera les droits d'inscription et de scolarité que dans un seul des deux établissements partenaires, à savoir dans l'établissement universitaire où il effectue son séjour d'études et de recherche, comme convenu ci-après par année ou par session(s) :

- 1re année ou session(s)	12 mois à l'université Laval : H-06, E-06, A-06
- 2e année ou session(s)	18 mois à l'université Blaise Pascal : H-07, E-07, A-07, H-08
- 3e année ou session(s)	6 mois à l'université Laval : E-08, A-08

ARTICLE 2 - Scolarité et thèse

- Le **sujet de thèse** déposé par le doctorant est :

« Spécification de contraintes d'intégrité spatio-temporelles : application à la modélisation des systèmes d'information agri environnementaux »

- La **durée prévisionnelle** de la scolarité et des travaux de recherche du doctorant est normalement de trois ans. Elle pourra être prolongée par avenant avec l'accord des deux établissements, sur proposition conjointe des deux directeurs de thèse.

- Le doctorant effectue **sa scolarité et ses travaux de recherche** en alternance entre les deux établissements, par périodes déterminées d'un commun accord entre les deux directeurs de thèse selon les modalités prévisionnelles suivantes :

  - périodes prévisionnelles dans l'établissement français :
    Janvier 2007- Juin 2008

  - périodes prévisionnelles à l'Université Laval :
    Les sessions d'hiver, d'été, d'automne 2006 et celles d'été et d'automne 2008

Lors de son séjour en France, le doctorant M. Lotfi Bejaoui aura un bureau au sein de l'Unité de Recherche Technologies et Systèmes d'information pour les agrosystèmes, du Cemagref de Clermont Ferrand et bénéficiera de l'ensemble des moyens de travail (notamment informatiques et documentaires) de cette Unité.

Lors de son séjour au Canada, le doctorant M. Lotfi Bejaoui aura un bureau au sein du Département des Sciences géomatiques et pourra accéder aux équipements de la Chaire CRSNG de recherche industrielle en bases de données géospatiales, localisée au Centre de recherche en géomatique de l'Université Laval.

- **La protection du sujet de thèse ainsi que la publication, l'exploitation et la protection des résultats de recherche issus des travaux de recherche** du doctorant dans les deux établissements seront assujetties à la réglementation en vigueur et assurées conformément aux procédures de chaque pays engagé dans la cotutelle.

  Lorsque nécessaire, les dispositions relatives à la protection des droits de propriété intellectuelle feront l'objet d'une annexe particulière à la présente convention.

ARTICLE 3 - Couverture sociale et responsabilité civile

MODALITÉS PÉDAGOGIQUES

ARTICLE 4 - Directeurs de thèse

Le doctorant effectue sa scolarité et ses travaux de recherche sous la responsabilité conjointe d'une directrice ou d'un directeur de thèse en France et d'une directrice ou d'un directeur de thèse à l'Université Laval, les deux personnes ayant déjà établi une collaboration :

- à l'Université Blaise Pascal, le directeur de thèse est :

  Monsieur Michel Schneider

- à l'Université Laval, le directeur de thèse est :

  Monsieur Yvan Bédard

  Les deux directeurs de thèse s'engagent à exercer pleinement la fonction de tuteur auprès de la doctorante ou du doctorant. Ils exercent conjointement les compétences attribuées en France et à l'Université Laval à une directrice ou à un directeur de thèse.

ARTICLE 5 - Déroulement de la scolarité

- **Activités pédagogiques de la doctorante ou du doctorant**
  (préciser les cours, séminaires, etc., dans chacun des établissements)

  Dans l'établissement français :

  2 modules "Sciences Pour l'Ingénieur" de 15 heures au choix.

À l'Université Laval :

La géomatique et ses référentiels (SCG-66672) (scolarité probatoire), SIG et analyse spatiale (SCG-66673) (scolarité probatoire), Séminaire (SCG-60430), Recherche préliminaire (SCG-65825), Examen de doctorat (SCG-65912), Conception de bases de données SIG (SCG-64738), Réalisation d'application en SIG (SCG-64739).

- **Examen de doctorat**

Après concertation entre les deux directeurs de thèse, et compte tenu des acquis du doctorant validés lors de sa scolarité antérieure, la préparation et le contenu de l'examen de doctorat québécois sont adaptés comme suit dans le respect des objectifs du programme ou de la formation

L'examen de doctorat sera conforme à la procédure en vigueur au programme de doctorat en sciences géomatiques à la faculté de foresterie de de géomatique de l'Université Laval.

ARTICLE 6 - Soutenance

- La thèse donne lieu à une **soutenance unique,** reconnue par les deux établissements.
- L'admission à la soutenance de thèse est décidé sur avis conjoint des directeurs de thèse, et fait intervenir une évaluation par au moins deux rapporteurs, extérieurs à l'établissement de soutenance. Les rapporteurs sont désignés conjointement par les deux établissements concernés.

- Le **jury de soutenance** est composé de scientifiques désignés à parité par les deux établissements partenaires. Il comprend obligatoirement les deux directeurs de thèse auxquels s'ajoute au moins un professeur de chacun des deux établissements partenaires. S'y ajoute aussi au minimum, dans le respect de la procédure d'évaluation de l'Université Laval, une examinatrice ou un examinateur externe aux deux établissements.

- **Autres aspects**

  - Le doctorant soutiendra sa thèse au **Québec** à l'Université Laval.

  - La soutenance devrait avoir lieu en Décembre 2008 .

  - La thèse sera rédigée et soutenue en langue Française.

  - Le résumé de la thèse sera rédigé et présenté en langue Française.

N.B. La doctorante ou le doctorant est tenu de rédiger soit la thèse, soit le résumé, en langue française; il est tenu de soutenir la thèse ou de présenter le résumé oral en langue française. Pour toute autre précision quant à la rédaction de la thèse et à la soutenance, veuillez consulter le guide intitulé « Le mémoire et la thèse : de la rédaction à la diplomation », qui est accessible en ligne à l'adresse suivante : www.fes.ulaval.ca <http://www.fes.ulaval.ca>.

ARTICLE 7 - Délivrance des deux diplômes

Sur avis favorable du jury de soutenance,
l'établissement français : l'Université Blaise Pascal

s’engage à conférer à Monsieur Lotfi Bejaoui
le grade de docteur et à lui délivrer le diplôme correspondant.

ET

l’Université Laval s’engage à conférer à Monsieur Lotfi Bejaoui
le grade de Ph.D. et à lui délivrer le diplôme correspondant.

Le libellé de chaque diplôme fera mention de la collaboration de l’établissement partenaire ainsi que de la cotutelle.

ARTICLE 8 - Dépôt, signalement et reproduction de la thèse

Dans chaque pays, ils seront effectués selon la réglementation en vigueur, en particulier celle de l’Université Laval.

SIGNATURES

Le doctorant **Monsieur Lotfi Bejaoui**	*Date*

<u>Pour l’établissement français</u>

Le directeur de thèse **Monsieur Michel Schneider**	*Date*
Directeur régional de Clermont Ferrand - Cemagref **Monsieur Didier Mechineau**	*Date*
Le responsable de l’école doctorale **Monsieur Philippe Mahey**	*Date*
Le président **Monsieur Pascal Albert Odouard**	*Date*

<u>Pour l'Université Laval</u>

Le directeur de thèse **Monsieur Yvan Bédard**	*Date*
Le directeur du programme de doctorat **Monsieur Jean-Jacques Chevalier**	*Date*
Le vice doyen de la Faculté des études supérieures **Monsieur Gérard Charlet**	*Date*

Lotfi BEJAOUI

Qualitative topological relationships for objects with possibly vague shapes

Résumé

Dans les bases de données spatiales actuellement mises en oeuvre, les phénomènes naturels sont généralement représentés par des géométries ayant des frontières bien délimitées. Une telle description de la réalité ignore le vague qui caractérise la forme de certains objets spatiaux (zones d'inondation, lacs, peuplements forestiers, etc.). La qualité des données enregistrées est donc dégradée du fait de ce décalage entre la réalité et sa description.

Cette thèse s'attaque à ce problème en proposant une nouvelle approche pour représenter des objets spatiaux ayant des formes vagues et caractériser leurs relations topologiques. Le modèle proposé, appelé *QMM model* (acronyme de Qualitative Min-Max model), utilise les notions d'extensions minimale et maximale pour représenter la partie incertaine d'un objet. Un ensemble d'adverbes permet d'exprimer la forme vague d'un objet (ex : a region with a *partially* broad boundary), ainsi que l'incertitude des relations topologiques entre deux objets (ex : *weakly* Contains, *fairly* Contains, etc.). Cette approche est moins fine que d'autres approches concurrentes (modélisation par sous-ensembles flous ou modélisation probabiliste). Mais elle ne nécessite pas un processus d'acquisition complexe des données. De plus elle est relativement simple à mettre en œuvre avec les systèmes existants de gestion de bases de données.

Cette approche est ensuite utilisée pour contrôler la qualité des données dans les bases de données spatiales et les entrepôts de données spatiales en spécifiant des contraintes d'intégrité par l'intermédiaire des concepts du modèle QMM. Une extension du langage de contraintes OCL (Object Constraint Language) a été étudiée pour spécifier des contraintes topologiques impliquant des objets ayant des formes vagues. Un logiciel existant (outil OCLtoSQL développé à l'Université de Dresden) a été étendu pour permettre la génération automatique du code SQL d'une contrainte lorsque la base de données est gérée par un système relationnel. Une expérimentation de cet outil a été réalisée avec une base de données utilisée pour la gestion des épandages agricoles. Pour cette application, l'approche et l'outil sont apparus très efficients.

Cette thèse comprend aussi une étude de l'intégration de bases de données spatiales hétérogènes lorsque les objets sont représentés avec le modèle QMM. Des résultats nouveaux ont été produits et des exemples d'application ont été explicités.