RCC*-9 and CBM*

Eliseo Clementini[1] and Anthony G. Cohn[2]

[1]University of L'Aquila, Information Engineering, L'Aquila, Italy
eliseo.clementini@univaq.it
[2]University of Leeds, School of Computing, Leeds, UK
a.g.cohn@leeds.ac.uk

Abstract. In this paper we introduce a new logical calculus of the Region Connection Calculus (RCC) family, RCC*-9. Based on nine topological relations, RCC*-9 is an extension of RCC-8 and models topological relations between multi-type geometric features: therefore, it is a calculus that goes beyond the modeling of regions as in RCC-8, being able to deal with lower dimensional features embedded in a given space, such as linear features embedded in the plane. Secondly, the paper presents a modified version of the Calculus-Based Method (CBM), a calculus for representing topological relations between spatial features. This modified version, called CBM*, is useful for defining a reasoning system, which was difficult to define for the original CBM. The two new calculi RCC*-9 and CBM* are introduced together because we can show that, even if with different formalisms, they can model the same topological configurations between spatial features and the same reasoning strategies can be applied to them.

1 Introduction

The modeling of topological relations in Geographical Information Systems (GISs) and spatial databases has been a central topic of research since the early 90s. Three models have played a very important role, both in terms of theoretical developments and practical applications: the 9-intersection model (9IM) [14], RCC-8 [11], and CBM [5]. Regarding their modeling capabilities, RCC-8 is able to represent topological relations between regions, while 9IM and CBM are able to represent topological relations between spatial features of any dimensionality. With respect to reasoning capabilities, composition tables were defined for RCC-8 and 9IM [12] (for regions only), while composition tables for the CBM were never developed. Having composition tables for all kinds of spatial data types is essential for several tasks, e.g. for spatial query optimization [1]: applying the constraints of the tables, it is possible to discover contradictions in the query expression before the real processing of the query actually starts.

The RCC family of calculi [8] uses a logical approach for the representation of qualitative topological relations. The calculi were developed with regions as the primitive spatial entity and the connection relation as the primitive topologic relation between regions, from which other relations can be defined. RCC-8, the most representative calculus of the family, can model eight topological relations between regions of the plane: there is a one-to-one correspondence with the eight topological relations that are definable with the 9IM between 2D simple regions. As remarked in [18], there exist few attempts to express topological relations between features of lower dimensions than the embedding space, such as lines in R^2, due to the difficulties of dealing with different types. In [19], Galton introduced an axiomatic system for multidimensional mereotopology, using primitives for 'part' (P) and 'boundary' (B). Gott's "INCH" calculus dealt with closed sets of points of uniform dimensionality [21] using a single primitive binary relation INCH ("includes a chunk of"). See further analysis in [9].

CBM [5] is a model for expressing topological relations between regions, lines, and points. It was especially defined for expanding the querying capabilities of database query languages towards spatial data. The operators of CBM have been adopted by the Open GeoSpatial Consortium (OGC) [23] and implemented in all spatial database systems. CBM relations can find an equivalent expression in terms of Egenhofer matrix-based methods [15] and vice versa. In particular, as it was shown in [3], CBM is more expressive than 9IM and equivalent to the Dimensionally-Extended 9-Intersection Model (DE+9IM) [3]. Despite its success in spatial databases and in the standardization process, CBM had little impact in the Qualitative Spatial Reasoning (QSR) community, due to the absence of a strong logical formulation and in particular its lack of composition tables. As pointed out in [16, 22], CBM is difficult to compare to logical calculi such as the RCC and no reasoning rules have been defined for it. The definitions of CBM were dependent on the dimension of the features participating in the relation. For example, a *cross* between a line and a region had a different definition from a *cross* between two lines. This meant it was not possible to find a single composition table for the calculus: at best, it would have been possible to find composition tables for each group of relations, that is, for region/region relations, for line/region relations, and so on, as proposed in [22].

In this paper, we aim at establishing a bridge between RCC and CBM, by defining an extension of RCC-8 that is capable of modeling topological relations between spatial features of any dimensionality and an extension of CBM that is capable of reasoning. To achieve this goal, a new calculus of the RCC family is defined, called RCC*-9, able to deal with features of various dimensions, not just regions[1]. A modification of CBM, called CBM*, is introduced that maps straightforwardly onto calculi of the RCC family and allows a composition table for reasoning to be found. Finally, it is shown that the two new calculi, RCC*-9 and CBM*, are able to model the same topo-

[1] The reason for the asterisk in the name is that it is not just a change in the number of relations. There is also a substantial change in the spatial primitives the new calculus is able to deal with. Further, we will also need in the paper to introduce a coarser calculus which we called RCC*-7, since an RCC-7 already exists [20].

logical configurations, even though they are defined in different ways, and they are compared with 9IM.

In Section 2, we recapitulate the definition of the geometric data model on which we base our work. In Section 3, we briefly recall the definitions of CBM. In Section 4, we introduce CBM* and discuss the changes between CBM and CBM*. In Section 5, we introduce the logical calculus RCC*-9. In Section 6, we define the spatial reasoning system of both RCC*-9 and CBM*. In Section 7, we discuss how to express CBM* relations and RCC*-9 relations in terms of 9IM. In Section 7, we make some concluding remarks.

2 Definition of geometric features

In this paper, we will adopt the same terminology of the OGC where various point-sets of the plane R^2 are called features, distinguishing between simple features and complex features [23]. The OGC simple feature model definitions were in turn taken from [4]. In the following, we briefly recall those definitions. First of all, features are classified with respect to their dimension: regions of dimension 2, lines of dimension 1, and points of dimension 0.

Let x be a two-dimensional point-set.

Def. 1. The interior x° of x is defined as the union of all open sets contained in x.

Def. 2. The closure $\bar{x}$ of x is defined as the intersection of all closed sets containing x.

Def. 3. The boundary ∂x of x is defined as the set difference between its closure and its interior, i.e., $\bar{x} - x^\circ$.

Def. 4. The exterior x^- of x is defined as the set difference $R^2 \setminus \bar{x}$.

Def. 5. x is regular closed if $x=\overline{x^\circ}$.

Def. 6. A simple region is a regular closed (non-empty) two-dimensional point-set x with a connected interior and connected exterior.

Def. 6 implies that a simple region is homeomorphic to the closed unit disk. A simple region does not have holes and is connected. If we remove the constraint of connected exterior from the definition, we obtain regions with holes [13]. In OGC simple feature specifications, regions with holes are implemented with the Polygon spatial data type. If we remove the constraint of connected interior, we obtain complex regions, that is, regions with holes and separations. Complex regions are implemented in OGC feature model with the MultiPolygon spatial data type.

Def. 7. A simple line is a closed (non-empty) one-dimensional point-set x defined as the image of a continuous mapping f:[0,1] $\rightarrow R^2$, such that $\forall t_i, t_j \in [0,1]$, $t_i \neq t_j$, $f(t_i) \neq f(t_j)$.

In other words, a simple line is the mapping of the unit interval in the plane with no self-intersections. A simple line can be described as the trace of a pencil on a sheet of paper without detaching the pencil and by not passing twice on the same position. The initial and final point of the simple line, defined as $f(0)$ and $f(1)$, are called the endpoints of the line.

Topologically, a simple line embedded in R^2, being a one-dimensional set, has an empty interior. As common practice in GIS [15] and in OGC standards as well, the boundary ∂x of a line x is considered to be the set of its endpoints and the interior of the line the difference, $x^{\circ}=x \setminus \partial x$. In this paper, we will adopt these definitions of boundary and interior of a line feature. In OGC feature model, simple lines are implemented with the Polyline spatial data type.

From Def. 7, if we remove the constraint of no self-intersections, we obtain lines with self-intersections. A particular case of a line with self-intersections is the closed ring, where $f(0)=f(1)$. If a one-dimensional point set can be obtained as the union of several mappings from the unit interval to the plane, then we obtain the concept of a complex line. A complex line can be made of several disjoint components. A complex line in OGC feature model is implemented with the MultiPolyline spatial data type.

A simple point is a zero-dimensional element of the embedding space. A complex point is the union of a finite number of simple points. Following the OGC convention, we assume that point features have an empty boundary. Simple and complex point features are implemented in OGC standards with the Point and Multipoint spatial data types, respectively.

3 CBM

One of the basic ideas behind CBM [5] was to provide an easy spatial extension of the tuple relational calculus [7] to express queries such as:

$\{x \mid \exists y\ [River(x) \wedge Region(y) \wedge cross(x,y) \wedge y=$ ‘Abruzzo’$]\}$

The above CBM expression corresponds to the query “Retrieve all the rivers that cross the Abruzzo region”. The topological relations of CBM can be applied not only to simple variables but to the boundaries of geometric features. Boundaries are extracted by the three operators *b* (boundary – the closed line representing the boundary of a simple region), *f* (from – the first endpoint of a line), *t* (to – the second endpoint of a line) [2]. For example, the following queries can be expressed in CBM:

$\{x \mid \exists y \exists z\ [River(x) \wedge Mountain(y) \wedge Sea(z) \wedge in(f(x),y) \wedge y=$ ‘Apennines’ $\wedge\ touch(t(x),z) \wedge z=$ ‘Adriatic’$]\}$

$\{x \mid \exists y\ [Road(x) \wedge Region(y) \wedge cross(x,y) \wedge overlap(x,b(y)) \wedge y=$ ‘Abruzzo’$]\}$

The above expressions correspond to “Retrieve all the rivers that rise in the Apennines mountains and flow into the Adriatic sea” and “Retrieve all the roads that cross the Abruzzo region and have a part of the road along the region’s boundary”.

The five topological relations of CBM are named *disjoint*, *touch*, *in*, *cross*, *overlap*. The definition of these relations are (the ‘dim’ operator evaluates to 0, 1, 2 depending whether the argument is a 0-, 1-, or 2-dimensional point set):

Def. 8. $disjoint(x,y) =_{def}\ x \cap y = \emptyset$

[2] These boundary extraction operators were specifically defined to extract the first and last endpoint of a directed line. More generally, when the direction is not known or there are more than two endpoints (such as in the case of complex lines) or no endpoints (such as in the case of closed rings), a generic boundary operator b is used that extracts the boundary of the feature.

Def. 9. *touch*$(x,y) =_{def}\ x°\cap y°=\varnothing \ \wedge\ x\cap y\neq\varnothing$
Def. 10. *in*$(x,y) =_{def}\ x\cap y=x \ \wedge\ x°\cap y°\neq\varnothing$
Def. 11. *cross*$(x,y) =_{def}\ \dim(x°\cap y°) < \max(\dim(x°),\dim(y°)) \ \wedge\ x\cap y\neq x \ \wedge\ x\cap y\neq y$
Def. 12. *overlap*$(x,y) =_{def}\ \dim(x°\cap y°) = \dim(x°) = \dim(y°) \ \wedge\ x\cap y\neq x \ \wedge\ x\cap y\neq y$

The relations can be applied to all geometric types, either simple or complex [4]. They were implemented by the OGC feature model as a set of functions with names `Disjoint`, `Touches`, `Within`, `Crosses`, `Overlaps`. Additionally, the converse function of `Within` was called `Contains` and the function `Equals` was defined as `Within` and `Contains` at the same time [23]. An expression of the relational tuple calculus extended with the five topological relations and the three boundary operators can be expressed by the Egenhofer matrix-based methods. Conversely, any instance of the DE+9IM can be expressed by an expression of CBM [3].

4 CBM*

In this section, we introduce a modification of CBM, called CBM*, for which it is easier to find an equivalence in terms of calculi of the RCC family and to find a composition table for reasoning. The basic relations of CBM* have a slightly different meaning from the corresponding relations of CBM. We assume the following definitions (we adopt the same names followed by a *) accompanied by a qualitative explanation of the meaning:

Def. 13. *disjoint**(x,y), the two features are disjoint:

*disjoint**$(x,y) =_{def}\ x\cap y=\varnothing$

Def. 14. *touch**(x,y), the two features intersect, but their interiors are disjoint (and it excludes containment):

*touch**$(x,y) =_{def}\ x°\cap y°=\varnothing \ \wedge\ x\cap y\neq\varnothing \ \wedge\ x\cap y\neq x \ \wedge\ x\cap y\neq y$

Def. 15. *in**(x,y), feature x is part of feature y (it excludes equality):

*in**$(x,y) =_{def}\ x\cap y=x \ \wedge x\neq y$
*in**$^{-1}(x,y) =_{def}\ x\cap y=y \ \wedge x\neq y$
*equal**$(x,y) =_{def}\ x=y$

Def. 16. *cross**(x,y), the interiors of the two features intersect, but at least one feature's boundary does not intersect the other feature:

*cross**$(x,y) =_{def}\ x°\cap y°\neq\varnothing \ \wedge\ (\partial x\cap y=\varnothing \ \vee\ x\cap\partial y=\varnothing)$

Def. 17. *overlap**(x,y): the interiors of the two features intersect and also each feature's boundary intersects the other feature (and it excludes containment).

*overlap**$(x,y) =_{def}\ x°\cap y°\neq\varnothing \ \wedge\ \partial x\cap y\neq\varnothing \ \wedge\ x\cap\partial y\neq\varnothing \ \wedge\ x\cap y\neq y \ \wedge x\cap y\neq x$

The proof that the relations of CBM* make a jointly exhaustive and pairwise disjoint (JEPD) set is readily obtained from the decision tree (see Fig. 1). Let us comment in more detail upon the differences between CBM and CBM* definitions. The *disjoint* and *disjoint** relations are the same. The *touch** relation is more restrictive than the *touch* relation, since cases where one feature is entirely contained inside the boundary of another one are excluded and are instead classified as *in** (see Fig. 2). The *in** relation takes over the cases ruled out by *touch** and excludes the case of equality between the two features: therefore, an explicit *equal** relation is needed in

CBM*. The *cross** and *overlap** relations take into account the remaining cases with a different criterion for partitioning the cases with respect to the original *cross* and *overlap* relations. In Fig. 2, we can see the differences with some representative configurations. The *overlap* relation between a region and a line was not possible in CBM, while the relation *overlap** between a region and a line corresponds to a real case.

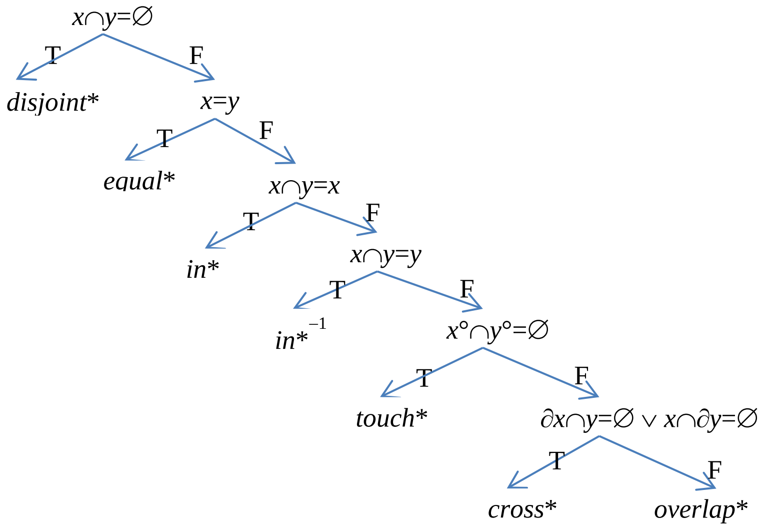

Fig. 1. Decision tree for the relations of CBM*.

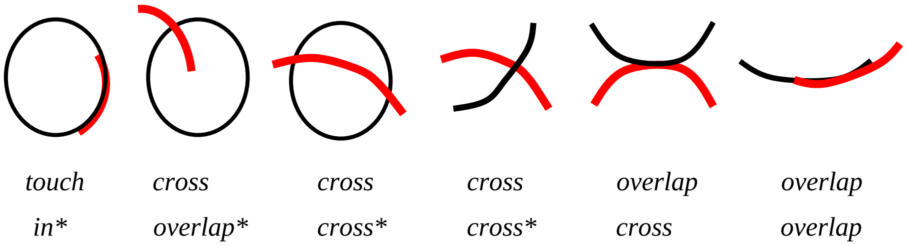

Fig. 2. Some differences between CBM and CBM* relations.

5 Definition of RCC*-9

In Cohn and his coauthors' work, the spatial primitive entities of the calculus are regions [8, 11]. The primitive spatial entities of the proposed calculus RCC*-9 are instead generic spatial features, without forcing an interpretation in terms of regions, lines, or points. As discussed in Section 2, in topology a feature of co-dimension bigger than zero (such as a line or a point in R^2) does not have an interior. One consequence is that a line in R^2 cannot have a non-tangential proper part (see also Galton's work [18]). The RCC definitions work when the universe of discourse contains regions of dimension R^n, for any $n>0$. But the definitions do not work for points or for universes of discourse containing regions of mixed dimensionality[3].

The boundary of an interval is made up of its two endpoints. A non-tangential proper part of an interval is another interval that is inside the first one and that does not connect with the endpoints of the first one. Adopting the "usual" GIS definitions [4, 15], non-tangential proper parts of lines embedded in R^2 can be defined as a mapping from one-dimensional intervals to the plane. In this way, we can find RCC*-9 definitions of topological relations that apply to all kinds of spatial features.

Analogously to RCC-8, we consider a primitive *connected* relation between two features $\mathsf{C}(x,y)$. There are several models for RCC in the literature; here, for consistency with CBM*, we take our universe of discourse to be closed regions (possibly disconnected), closed lines (also possibly disconnected), and sets of isolated points. $\mathsf{C}(x,y)$ is interpreted as being true when x and y have at least one point in common. The *connected* relation enjoys two axioms:

$\mathsf{C}(x,x)$,

$\mathsf{C}(x,y) \rightarrow \mathsf{C}(y,x)$.

From the primitive *connected* relation, other relations are consequently defined. These are as in RCC8 except as noted. The *disconnected* relation is defined as:

Def. 18. $\mathsf{DC}(x,y) =_{def} \neg\mathsf{C}(x,y)$

The *part* relation between x and y is defined by saying that the feature x cannot be connected to features disconnected from y:

Def. 19. $\mathsf{P}(x,y) =_{def} \forall z\ [\mathsf{C}(z,x) \rightarrow \mathsf{C}(z,y)]$

The *proper part* relation excludes the case of equality between the two features:

Def. 20. $\mathsf{PP}(x,y) =_{def} \mathsf{P}(x,y) \wedge \neg\mathsf{P}(y,x)$

The *equal* relation is defined as:

Def. 21. $\mathsf{EQ}(x,y) =_{def} \mathsf{P}(x,y) \wedge \mathsf{P}(y,x)$

In the original RCC, the *overlap* relation was defined as: $\mathsf{O}(x,y) = \exists z\ [\mathsf{P}(z,x) \wedge \mathsf{P}(z,y)]$. Such a definition sufficed to refine the *connected* relation and make a distinction between the *overlap* and the *externally connected* relation. In RCC*-9, when we remove the limitation that features are regions only, the fact that there is a common part belonging to the two features x and y would not suffice to identify a new relation. In essence, the $\mathsf{O}(x,y)$ relation would coincide with the $\mathsf{C}(x,y)$ relation, since the common part could be a line or a point. Therefore, we need to find another definition for the *overlap* relation. The *externally connected* relation in RCC-8 was defined as $\mathsf{EC}(x,y) = \mathsf{C}(x,y) \wedge \neg\mathsf{O}(x,y)$. This means that the EC relation cannot be defined simply by negating O. Further, in RCC-8, the *non-tangential proper part* relation needed the EC relation for its definition, which was $\mathsf{NTPP}(x,y) = \mathsf{PP}(x,y) \wedge \neg\exists z[\mathsf{EC}(z,x) \wedge \mathsf{EC}(z,y)]$.

To overcome the above issues, we need to introduce a new topological primitive and we choose the *boundary* relation $\mathsf{B}(x,y)$, expressing the fact that feature x is the

[3] Different mereotopologies (such as RCC) take a different semantic stance as to what kinds of spatial entities are allowed. See Cohn and Varzi [10] for an extended discussion and analysis on this issues and a comparison of the different approaches, as well as axiomatisations of mereotopologies allowing boundaries (though the *cross* relation considered in this paper is not defined there).

boundary of feature *y*. The type of *x* must be of different type to that of *y*. For a line *y*, *x* is the set of its endpoints[4]. If *y* is a simple region, then *x* is the closed line representing *y*'s boundary; if *y* is a complex region (holed or multipiece), then *x is a set of lines*. This effectively also introduces several kinds of spatial entities, so that our intended universe of discourse now consists of *regions* (2D entities), 1D lines (such as *boundaries of regions*), and sets of isolated points (boundaries of lines). The *boundary* relation obeys the following axiom:

$\mathsf{B}(x,y) \rightarrow \mathsf{PP}(x,y)$.

Hence, we give a new definition of the *non-tangential proper part* relation:

Def. 22. $\mathsf{NTPP}(x,y) =_{def} \mathsf{PP}(x,y) \wedge \forall y_1\, [\mathsf{B}(y_1, y) \rightarrow \mathsf{DC}(x, y_1)]$

The definition is illustrated in Fig. 3. The feature *x* is a proper part of *y* and does not touch the boundary of *y*. Such a definition of NTPP, though it is different, has exactly the same semantics as the original RCC definition in the case of regions.

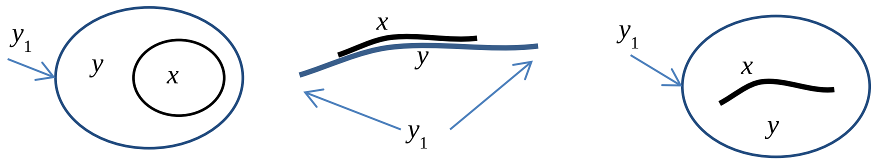

Fig. 3. Illustrations of the NTPP definition of RCC*-9. (Note that in the middle illustration, *x* actually is part of *y*, but it is drawn alongside it for clarity of illustration; we use the same convention in later figures as well).

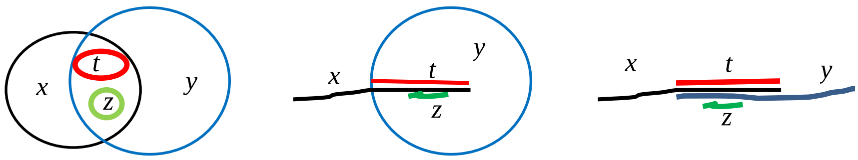

Fig. 4. Illustration of the O relation.

The new definition for the tangential proper part relation is:

Def. 23. $\mathsf{TPP}(x,y) =_{def} \mathsf{PP}(x,y) \wedge \neg\, \mathsf{NTPP}(x,y)$

We can now give a new definition of the *overlap* relation, which is more restrictive than the corresponding definition of RCC-8:

Def. 24. $\mathsf{O}(x,y) =_{def} \exists z[\mathsf{NTPP}(z,x) \wedge \mathsf{NTPP}(z,y)] \wedge \exists t[\mathsf{TPP}(t,x) \wedge \mathsf{TPP}(t,y)]$

The above definition of *overlap* expresses the fact that there is a common non-tangential proper part belonging to the two features and a common tangential proper part as well. The second part of the rule would not be necessary for regions, but it is necessary for lines (see Fig. 4), otherwise also cases of *cross* (see later on) would be regarded as *overlap*.

---

[4] Since B is a relation rather than a functor in RCC*-9, if *y* is a closed ring and its boundary is empty, it means that there is no value *x* for which $\mathsf{B}(x,y)$ is true. Similarly $\mathsf{B}(x,y)$ is never true when *y* is a point.

As in RCC-8, as a refinement of O(x,y), the *partially overlap* relation corresponds to excluding the inclusion of one feature into the other one:

Def. 25. $\mathsf{PO}(x,y) =_{\text{def}} \mathsf{O}(x,y) \wedge \neg\mathsf{P}(x,y) \wedge \neg\mathsf{P}(y,x)$

Considering the new domain of spatial features instead of only regions, there are two other kinds of connection that are not included in the overlap definition, namely, the *externally connected* and the *cross* relations. We use the following definition for the *externally connected* relation (which differs from the RCC8 one):

Def. 26. $\mathsf{EC}(x,y) =_{\text{def}} \mathsf{C}(x,y) \wedge \neg\mathsf{O}(x,y) \wedge \forall z\ [[\mathsf{P}(z,x) \wedge \mathsf{P}(z,y)] \rightarrow [\mathsf{TPP}(z,x) \vee \mathsf{TPP}(z,y)]]$

Fig. 5 depicts the EC definition in case of two regions, a region and a line, and two lines. The whole common part z needs to be a tangential proper part of x or y (this is ensured through the universal quantifier $\forall z$). In the case of a line x and a region y in Fig. 5, the common part z is a tangential proper part of y. Also in the case of the two lines, the common part z is a tangential proper part of y. The EC relation maintains the same semantics as RCC-8 for (2D) regions.

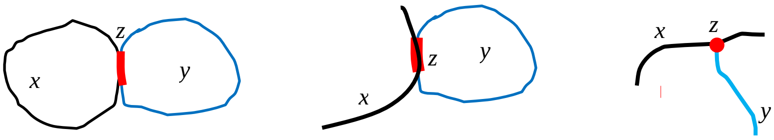

Fig. 5. Illustrations of the EC relation.

Finally, we add the definition of *cross*, which corresponds to the remaining kind of connection and is not included in the previous ones (see Fig. 6)[5]:

Def. 27. $\mathsf{CR}(x,y) =_{\text{def}} \mathsf{C}(x,y) \wedge \neg\mathsf{O}(x,y) \wedge \neg\mathsf{EC}(x,y)$

Fig. 6. Cases of the CR relation.

The inverse relations of the asymmetric *part* relation and its specializations are defined as:

Def. 28. $\mathsf{Pi}(x,y) =_{\text{def}} \mathsf{P}(y,x)$

Def. 29. $\mathsf{PPi}(x,y) =_{\text{def}} \mathsf{PP}(y,x)$

Def. 30. $\mathsf{NTPPi}(x,y) =_{\text{def}} \mathsf{NTPP}(y,x)$

Def. 31. $\mathsf{TPPi}(x,y) =_{\text{def}} \mathsf{TPP}(y,x)$

[5] Note that the *cross* relation is between a region and a line or a pair of lines; one could also imagine a scenario where two regions “cross” each other (so that they form a kind of “fat cross”; this is not an instance of the *cross* relation, but just of the PO relation – but see Galton [17] for definitions of relations specialising PO in this way.

For completeness with respect to the original RCC family of calculi, a DR relation (*discrete*) is defined as:

Def. 32. DR(x,y) $=_{def}$ EC(x,y) $\vee$ DC(x,y)

The 9 relations DC, EC, PO, TPP, NTPP, TPPi, NTPPi, EQ, and CR form a provably JEPD set of relations and are the base relations of RCC*-9. A hierarchical implication structure of all the relations defined above is given in Figure 7. To show that Figure 7 correctly reflects the implication hierarchy of the relations is mostly straightforward from the definitions. The only cases which are not trivial are the subsumption of O by C, and of P and Pi by O. We also define the JEPD set DC, EC, PO, PP, PPi, EQ, and CR, which we name RCC*-7 and, as we shall see below, corresponds to CBM*.

It is important to stress the fact that the changes we have made to some definitions of RCC-8 to obtain RCC*-9 are alternative definitions of RCC-8 relations to accommodate multi-type features. There is no change of meaning for these relations if we apply them to (2D) regions. RCC*-9 introduces the new CR relation, which can only hold when one of the entities is a 1D entity.

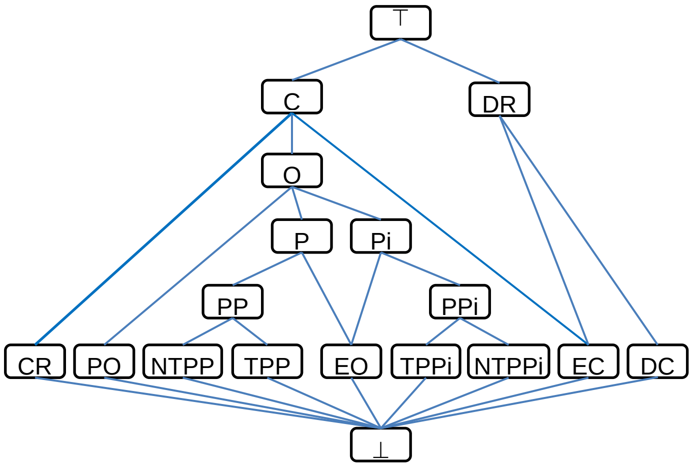

Fig. 7. The subsumption hierarchy of RCC*-9 relations. The lines indicate semantic inclusion – i.e., whenever two relations are linked, the lower one implies the upper one.

6 Spatial reasoning

We introduced in Section 4 a modified CBM, called CBM*, and in Section 5 an extension of RCC-8, called RCC*-9. The latter is logically defined in FOPC in terms of a single primitive relation C(x,y) whereas the former's relations are all taken as primitive and each have their own semantic definitions. It is provable that the two systems can express the same topological relations, as illustrated in Table 1. We can observe a direct correspondence of the base relations of CBM* with the 7 relations of RCC*-7. In CBM*, there are no base relations expressing RCC*-9 relations NTPP and TPP, which can be expressed by a logical formula involving boundaries of features.

Table 1. Correspondence between CBM* and RCC*-9.

RCC*-9	CBM*
DC(*x*,*y*)	*disjoint**(*x*,*y*)
EC(*x*,*y*)	*touch**(*x*,*y*)
PP(*x*,*y*)	*in**(*x*,*y*)
CR(*x*,*y*)	*cross**(*x*,*y*)
PO(*x*,*y*)	*overlap**(*x*,*y*)
PPi(*x*,*y*)	$in^{*-1}(x,y)$
EQ(*x*,*y*)	*equal**(*x*,*y*)
NTPP(*x*,*y*)	*in**(*x*,*y*) ∧ *disjoint**(*x*,*b*(*y*))
TPP(*x*,*y*)	*in**(*x*,*y*) ∧ ¬*disjoint**(*x*,*b*(*y*))
NTPPi(*x*,*y*)	$in^{*-1}(x,y)$ ∧ *disjoint**(*b*(*x*),*y*)
TPPi(*x*,*y*)	$in^{*-1}(x,y)$ ∧ ¬*disjoint**(*b*(*x*),*y*)
C(*x*,*y*)	¬*disjoint**(*x*,*y*)
P(*x*,*y*)	*in**(x,y) ∨ *equal**(*x*,*y*)
Pi(*x*,*y*)	$in^{*-1}(x,y)$ ∨ *equal**(*x*,*y*)
O(*x*,*y*)	*overlap**(*x*,*y*) ∨ *in**(*x*,*y*) ∨ $in^{*-1}(x,y)$ ∨ *equal**(*x*,*y*)
DR(*x*,*y*)	*disjoint**(*x*,*y*) ∨ *touch**(*x*,*y*)

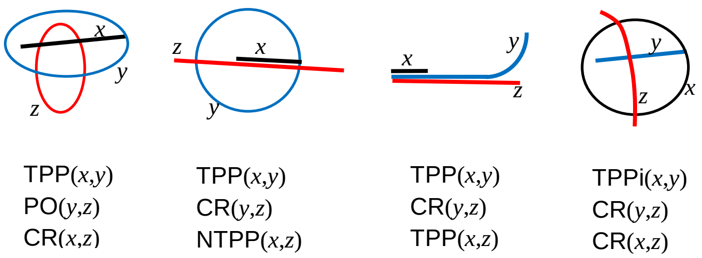

Fig. 8. Some new cases of composition involving the CR relation.

Given the correspondence between the CBM* and RCC*-9, we proceed to find the composition tables for these calculi. The composition tables contain the basic rules to perform qualitative spatial reasoning with such calculi (see, for example, [9]). Given the relation $r_1(x,y)$ and the relation $r_2(y,z)$, the composition is the relation $r_3(x,z)$. The composition table gives all the possible results of composition for each combination of relations. Such results are expressed as disjunctions of the basic relations. For RCC*-9, the results of compositions are those reported in Table 2. Such a table is a direct extension of the composition table of RCC-8 [11], that is, if we restricted Table

2 to regions, we would re-obtain the composition table of RCC-8: the CR relation cannot hold between two regions.

In general, the proof of composition tables is difficult, especially when the semantics of the calculus depends on higher-order constructs such as sets [24]. There are two aspects to proving that a composition table is correct: (1) showing that each disjunct in each cell is necessary; (2) showing that there are no missing disjunctions. The former is usually achieved by demonstrating (i.e. providing a model such as a figure) of each combination of r_1, r_2 and a disjunct from r_3[6]. Showing that there are no missing disjuncts, given an axiomatic theory of the calculus, can be achieved by proving a theorem that r_1 and r_2 imply r_3 for each cell[7]. If we consider that the RCC*-9 composition table is an extension of the RCC-8 table, one way of finding a proof is 'by difference', that is, limiting the analysis to the new cases involving the CR relation only. We found in total 89 new compositions that can be instantiated in R^2 involving the CR relation: see Fig. 8 for a sample of them. The fact that no other cases with CR are possible can be proved with a theorem for each entry, but by using redundancy elimination techniques as in [2], the actual number of entries that need to be proved can be reduced significantly. Alternatively, a proof could be developed with a semi-automatic reasoner as proposed in [24]. Besides formal proofs, in future work we also plan to apply heuristics such as in [6], where composition tables can be filled up by running tests on random data sets made up of points, polygons, and polylines.

Table 2. Composition table for RCC*-9.

r_1 \ r_2	DC	EC	PO	TPP	NTPP	TPPi	NTPPi	EQ	CR
DC	no info	DR, PO, PP, CR	DR, PO, PP, CR	DR, PO, PP, CR	DR, PO, PP, CR	DC	DC	DC	DR, PO, PP, CR
EC	DR, PO, PPi, CR	DR, PO, TPP, EQ, TPPi, CR	DR, PO, PP, CR	EC, PO, PP, CR	PO, PP, CR	DR	DC	DC	DR, PO, PP, CR
PO	DR, PO, PPi, CR	DR, PO, PPi, CR	no info	PO, PP, CR	PO, PP, CR	DR, PO, PPi, CR	DR, PO, PPi, CR	PO	DR, PO, PP, PPi, CR
TPP	DC	DR	DR, PO, PP, CR	PP	NTPP	DR, PO, TPP, EQ, TPPi, CR	DR, PO, PPi, CR	TPP	DR, PP, PO, CR
NTPP	DC	DC	DR, PO, PP, CR	NTPP	NTPP	DR, PO, PP, CR	no info	NTPP	DR, PP, PO, CR
TPPi	DR, PO, PPi, CR	EC, PO, PPi, CR	PO, PPi, CR	PO, TPP, EQ, TPPi	PO, PP, CR	PPi	NTPPi	TPPi	PO, PPi, CR
NTPPi	DR, PO, PPi, CR	PO, PPi, CR	PO, PPi, CR	PO, PPi, CR	O, CR	NTPPi	NTPPi	NTPPi	PO, PPi, CR
EQ	DC	EC	PO	TPP	NTPP	TPPi	NTPPi	EQ	CR
CR	DR, PO, PPi, CR	DR, PO, PPi, CR	DR, PO, PP, PPi, CR	PP, PO, CR	PP, PO, CR	DR, PPi, PO, CR	DR, PPi, PO, CR	CR	no info

[6] This is what we actually did to find the RCC*-9 composition table, that is, finding configurations like those in Fig.8 satisfying each result of the table.

[7] An automatic proof of RCC-8 composition table based on encoding RCC-8 in an intuitionistic propositional calculus has been proposed in [2].

From the composition table of RCC*-9, we can infer the composition table of CBM*. First, we need to find an intermediate result: the composition table of RCC*-7. This is just a reduced version of the composition table of RCC*-9 that is obtained by making the union of relations TPP and NTPP and of relations TPPI and NTPPi. From the composition table of RCC*-7 and from the correspondences between CBM* and RCC-7 (Table 1), we can obtain as an almost immediate result the composition table for CBM* (Table 3) by simple renaming of the relations.

Table 3. Composition table for CBM*. Adopted abbreviations: *di*=*disjoint*, *to*=*touch*, *ov*=*overlap*, *eq*=*equal*, *cr*=*cross*.

r_2 / r_1	*di**	*to**	*ov**	*in**	in^{*-1}	*eq**	*cr**
*di**	*no info*	*di*, to*, ov*, in*, cr**	*di*, to*, ov*, in*, cr**	*di*, to*, ov*, in*, cr**	*di**	*di**	*di*, to*, ov*, in*, cr**
*to**	*di*, to*, ov**, in^{*-1}, *cr**	*no info*	*di*, to*, ov*, in*, cr**	*to*, ov*, in*, cr**	*di*, to**	*di**	*di*, to*, ov*, in*, cr**
*ov**	*di*, to*, ov**, in^{*-1}, *cr**	*di*, to*, ov**, in^{*-1}, *cr**	*no info*	*ov*, in*, cr**	*di*, to*, ov**, in^{*-1}, *cr**	*ov**	*di*, to*, ov*, in**, in^{*-1}, *cr**
*in**	*di**	*di*, to**	*di*, to*, ov*, in*, cr**	*in**	*no info*	*in**	*di*, to*, in*, ov*, cr**
in^{*-1}	*di*, to*, ov**, in^{*-1}, *cr**	*to*, ov**, in^{*-1}, *cr**	*ov**, in^{*-1}, *cr**	*ov*, in**, in^{*-1}, *eq*, cr**	in^{*-1}	in^{*-1}	*ov**, in^{*-1}, *cr**
*eq**	*di**	*to**	*ov**	*in**	in^{*-1}	*eq**	*cr**
*cr**	*di*, to*, ov**, in^{*-1}, *cr**	*di*, to*, ov**, in^{*-1}, *cr**	*di*, to*, ov*, in**, in^{*-1}, *cr**	*in*, ov*, cr**	*di*, to**, in^{*-1}, *ov*, cr**	*cr**	*no info*

7 Comparison with 9-intersection

In this section, we compare the calculi CBM* and RCC*-9 with 9IM [15] and DE+9IM [3]. This is useful for practical reasons to easily implement the relations of the proposed calculi in OGC-compliant systems. We use the `Relate` function defined in the OGC Simple Features Specification [23]. The function returns true if the two features satisfy the topological relation corresponding to the string parameter. Such a string represents a set of values for 9IM matrix by rows: characters allowed in the string are ‘F’ for an empty intersection, ‘T’ for a non-empty intersection, and ‘*’ for ‘don’t care’. The value ‘T’ in the string of the `Relate` function can be specialized to the values 0, 1, 2 to express the dimension of the intersection set: this corresponds to the DE+9IM matrix introduced in [3]. Table 4 summarizes the correspondence between CBM* and 9IM. The equivalent expressions of 9IM can be easily inferred from CBM* definitions.

Table 4. Correspondence between CBM* relations and 9IM relations.

CBM*	9IM
*disjoint**(*x*,*y*)	`Relate(x,y,"FF*FF****")`
*touch**(*x*,*y*)	`Relate(x,y,"FTT***T**")` ∨ `Relate(x,y,"F*TT**T**")` ∨ `Relate(x,y,"F*T*T*T**")`
*in**(*x*,*y*)	`Relate(x,y,"**F**F***")` ∧ ¬ `Relate(x,y,"TFFFTFFFT")`
*cross**(*x*,*y*)	`Relate(x,y,"T**FF****")` ∨ `Relate(x,y,"TF**F****")`
*overlap**(*x*,*y*)	`Relate(x,y,"TTTT**T**")` ∨ `Relate(x,y,"T*T*T*T**")`
in^{*-1}(*x*,*y*)	`Relate(x,y,"******FF*")` ∧ ¬ `Relate(x,y,"TFFFTFFFT")`
*equal**(*x*,*y*)	`Relate(x,y,"TFFFTFFFT")`

Table 5. Correspondence between RCC*-9 and 9IM.

RCC*-9	9IM
DC(*x*,*y*)	`Relate(x,y,"FF*FF****")`
EC(*x*,*y*)	`Relate(x,y,"FTT***T**")` ∨ `Relate(x,y,"F*TT**T**")` ∨ `Relate(x,y,"F*T*T*T**")`
NTPP(*x*,*y*)	`Relate(x,y,"*FF*FF***")`
TPP(*x*,*y*)	`Relate(x,y,"*TF**F***")` ∨ `Relate(x,y,"**F*TF***")` ∧ ¬ `Relate(x,y,"TFFFTFFFT")`
CR(*x*,*y*)	`Relate(x,y,"T**FF****")` ∨ `Relate(x,y,"TF**F****")`
PO(*x*,*y*)	`Relate(x,y,"TTTT**T**")` ∨ `Relate(x,y,"T*T*T*T**")`
NTPPi(*x*,*y*)	`Relate(x,y,"***FF*FF*")`
TPPi(*x*,*y*)	`Relate(x,y,"***T**FF*")` ∨ `Relate(x,y,"****T*FF*")` ∧ ¬ `Relate(x,y,"TFFFTFFFT")`
EQ(*x*,*y*)	`Relate(x,y,"TFFFTFFFT")`

We can see in Table 4 that CBM* relations do not need the dimension of the intersection set to find equivalent expressions. Therefore, it is possible to find equivalent expressions of CBM* queries in terms of 9IM without the need to resort to the much expressive DE+9IM. The CBM relations needed the dimension to find equivalent expressions in the DE+9IM matrices, being 9IM matrix alone not sufficient. In [3], it was proved that CBM is equivalent to DE+9IM in terms of the number of topological configurations that the models are able to distinguish. Though it is out of the scope of this paper, it is provable that CBM* is equivalent to 9IM in terms of number of topological configurations. In this sense, CBM* can be considered weaker than CBM because CBM* does not include the possibility of checking the dimension of intersections. Of course, this is not a real weakness of CBM* since an operator to check dimension could be easily added to the calculus to recuperate the ability of checking set dimension. Given the correspondence between CBM* and RCC*-9 (Table 1), we can express RCC*-9 relations in terms of 9IM matrices by using Table 4 to obtain Table 5.

8 Conclusions and further work

An extension towards multidimensional mereotopology [18] has been advocated for a long time. RCC*-9 is our contribution to address this issue. We defined RCC*-9 by modifying the definition of the basic relations of RCC-8 and adding two new relations, namely, a new primitive B(*x*,*y*) to express that *x* is *boundary* of *y* and CR(*x*,*y)* for the defined *cross* relation. The variables of RCC*-9 no longer range just over regions, but features (or sets of features) of dimension 2, 1, or 0, embedded in R^2. These changes extend rather than change[8] the semantics of RCC-8, since if we consider only regions, then RCC*-9 collapses to RCC-8. The composition table of RCC*-9 with respect to the composition table of RCC-8 presents the relation CR as an additional possible result of composition, but it does not affect the already present results – i.e. each entry in the composition table for RCC*-9 is either the same, or a superset of the corresponding RCC-8 composition table entry (except for the rows and columns labelled by CR, which are new).

In this paper, we also introduced the CBM*, a modified version of CBM where we lose the possibility of distinguishing the dimension of set intersections. CBM* definitions do not depend on the type of features, e.g., a *cross* between two lines has the same definition of the *cross* between a line and a region. With the new definitions, it is possible to obtain a single composition table for all features.

Finally, we provided the usual basis for a reasoning system for a qualitative calculus, i.e. a composition table, for the new calculi, extending the earlier composition tables from simple regions to the case of generic spatial features. Another interesting aspect that we discussed is how to find equivalent expressions of both calculi in terms of 9IM, which is essential to enabling a straightforward implementation in OGC-compliant systems.

Further work is needed to provide a formal proof of the correctness of the composition tables. Another issue that is not covered in this paper is the study of the cognitive

[8] Strictly, it only extends RCC8 if we consider the 2D interpretation of RCC8: RCC8 can be interpreted in any dimension ≥ 2; in principle the definitions here may apply to regions of other dimensions but we have not investigated this yet.

adequacy of the group of relations inside CBM* and RCC*-9 models. It would be interesting to find out the differences in subjective perceptions especially of the previous CBM calculus versus the new CBM* calculus. Finally, an assessment of how the calculi behave for complex features and for higher dimensional spaces remains to be done.

Acknowledgements. Authors wish to express their gratitude to the referees for their useful suggestions and to Paolo Fogliaroni for his comments to a previous version of the manuscript. The financial support of EU projects RACE (FP7-ICT-287752) and STRANDS (FP7-ICT-600623) is gratefully acknowledged.

References

[1] W. G. Aref and H. Samet, "Optimization Strategies for Spatial Query Processing," in *17th International Conference on Very Large Databases*, Barcelona, Spain, 1991, pp. 81-90.

[2] B. Bennett, *Logical Representations for Automated Reasoning about Spatial Relationships*. Ph.D. Thesis, School of Computer Studies, University of Leeds, 1997.

[3] E. Clementini and P. Di Felice, "A Comparison of Methods for Representing Topological Relationships," *Information Sciences*, vol. 3, pp. 149-178, 1995.

[4] E. Clementini and P. Di Felice, "A Model for Representing Topological Relationships Between Complex Geometric Features in Spatial Databases," *Information Sciences*, vol. 90, pp. 121-136, 1996.

[5] E. Clementini, P. Di Felice, and P. van Oosterom, "A Small Set of Formal Topological Relationships Suitable for End-User Interaction," in *Advances in Spatial Databases - Third International Symposium, SSD '93*. vol. 692, D. Abel and B. C. Ooi, Eds. Berlin: Springer-Verlag, 1993, pp. 277-295.

[6] E. Clementini, S. Skiadopoulos, R. Billen, and F. Tarquini, "A reasoning system of ternary projective relations," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 161-178, 2010.

[7] E. F. Codd, "A relational model for large shared data banks," *Communications of the ACM*, vol. 13, pp. 377-387, 1970.

[8] A. G. Cohn, B. Bennett, J. Gooday, and N. Gotts, "Qualitative Spatial Representation and Reasoning with the Region Connection Calculus," *GeoInformatica*, vol. 1, pp. 275-316, 1997.

[9] A. G. Cohn and J. Renz, "Qualitative Spatial Representation and Reasoning," in *Handbook of Knowledge Representation, 1*, F. v. Harmelen, V. Lifschitz, and B. Porter, Eds.: Elsevier, 2007, pp. 551-596.

[10] A. G. Cohn and A. C. Varzi, "Mereotopological connection," *Journal of Philosophical Logic*, vol. 32, pp. 357-390, 2003.

[11] Z. Cui, A. G. Cohn, and D. A. Randell, "Qualitative and Topological Relationships in Spatial Databases," in *Advances in Spatial Databases - Third International Symposium, SSD '93*. vol. 692, D. J. Abel and B. C. Ooi, Eds. Berlin: Springer-Verlag, 1993, pp. 296-315.

[12] M. J. Egenhofer, "Deriving the composition of binary topological relations," *Journal of Visual Languages and Computing,* vol. 5, pp. 133-149, 1994.

[13] M. J. Egenhofer, E. Clementini, and P. Di Felice, "Topological relations between regions with holes," *International Journal of Geographical Information Systems,* vol. 8, pp. 129-142, 1994.

[14] M. J. Egenhofer and R. D. Franzosa, "Point-Set Topological Spatial Relations," *International Journal of Geographical Information Systems,* vol. 5, pp. 161-174, 1991.

[15] M. J. Egenhofer and J. R. Herring, "Categorizing Binary Topological Relationships Between Regions, Lines, and Points in Geographic Databases," Department of Surveying Engineering, University of Maine, Orono, ME Technical Report, 1990.

[16] N. Gabrielli, *Investigation of the Tradeoff between Expressiveness and Complexity in Description Logics with Spatial Operators.* Ph.D. Thesis: University of Verona, 2009.

[17] A. Galton, "Modes of overlap," *Journal of Visual Languages and Computing,* vol. 9, pp. 61-79, 1998.

[18] A. Galton, "Multidimensional Mereotopology," in *Proceedings of the Ninth International Conference on Principles of Knowledge Representation and Reasoning (KR2004),* D. Dubois, C. Welty, and M.-A. Williams, Eds. Whistler, BC, Canada, June 2-5, 2004: American Association for Artificial Intelligence, 2004, pp. 45-54.

[19] A. P. Galton, "Taking dimension seriously in qualitative spatial reasoning," in *Proceedings of the Twelfth European Conference on Artificial Intelligence (ECAI'96),* Budapest, Hungary, 11th - 16th Aug 1996, 1996.

[20] A. Gerevini and J. Renz, "Combining topological and size information for spatial reasoning," *Artificial Intelligence*, vol. 137, pp. 1-42, 2002.

[21] N. M. Gotts, "Formalizing Commonsense Topology: The INCH Calculus. pp.," in *Proceedings of the Fourth International Symposium on Artificial Intelligence and Mathematics*, 1996.

[22] A. Isli, L. Museros Cabedo, T. Barkowsky, and R. Moratz, "A Topological Calculus for Cartographic Entities " in *Spatial Cognition II* vol. LNCS 1849 Berlin: Springer, 2000, pp. 225-238.

[23] OGC Open Geospatial Consortium Inc., "OpenGIS Simple Features Implementation Specification for SQL." vol. OGC 99-049, 1999.

[24] S. Wölfl, T. Mossakowski, and L. Schröder, "Qualitative constraint calculi: Heterogeneous verification of composition tables " in *20th International FLAIRS Conference (FLAIRS-20),* pp. 665-670, 2007.