

# Study of spatial fusion of geographical entities and quantitative information in accordance with their imprecision

Application to agricultural information in Observox

**Cyril de Runz and Eric Desjardin**

CReSTIC  
Université de Reims Champagne-Ardenne  
Reims, France  
{cyril.de-runz, eric.desjardin}@univ-reims.fr

**Herman Akdag**

LIP6  
University of Paris 6  
Paris, France  
Herman.Akdag@lip6.fr

**Abstract**— The article deals with imprecise geographical entities modelled according to the fuzzy set theory for both spatial and quantitative information. It presents the issues of fusion of fuzzy geographical objects according to its storing modes (raster, vector) in a mutualised geographical information system (GIS). We study the aggregation of imprecise spatial entities and its impact in the imprecise description of quantities associated to space. It exposes the current choices in the Observox Project managing multiple sources of information.

**Keywords:** *fusion; fuzzy representation; imprecise geographical entities, fuzzy spatial object storage, mutualized GIS*

## I. INTRODUCTION

In the sustainable development context, the AQUAL project (a State-Region Project in the Champagne-Ardenne, France) highlights the need of a monitoring environment for the study of agricultural practices and of their pressure on the water resources in the Vesle basin. It is called Observox and it exploits data coming from heterogeneous sources: satellite images, land registry, statistical data, Corine Land Cover and other European data. Those data have different qualities; they are often uncertain, imprecise and/or incomplete. They do not have the same spatial representation (vector or raster). Thus, their combinations or fusions imply many problems. In this article, we focus on the fusion of imprecise spatial entities modeled as fuzzy entities.

The multiple sources give information on the same space but do not split it in the same partition. Most of the time, the boundaries of the same spatial entity will differ according to the source. Thus, spatial entities provided by the combination of all the sources will better be modeled as imprecise spatial entities such as fuzzy entities. In fact, according to Fisher (2005), vagueness is well represented by fuzzy sets.

The first goal of our project is to store multivariate data into a GIS. According to this goal, it is fundamental to store merging results in a unique spatial representation mode. Our project deals with vector and raster spatial representation mode. Thus, the choice of a unique spatial representation

mode is essential. Fonte (2006) proposes an approach for the conversion between raster and vector using fuzzy geographical entities.

In fact, if the vector representation is selected, then raster information will be over-interpreted, but if it is the raster representation, the vector information will be under-interpreted. For instance, the choice of raster representation mode does not allow unrealistic interpretation and facilitates the merging, but the generalization and the study at a larger scale will be more difficult and the storage of raster entities will be more complex. In vector mode, objects can be stored as UGML files (Morris and Petry, 2006) or using multiple representations (Dilo, de By R.A. and Stein, 2004). In the two hypotheses, the conversion (raster-vector) and the combination of the spatial entities produce fuzzy spatial entities.

However, the resulting fuzzy spatial entities may partially overlap. Moreover, quantitative information describing each entity should be studied depending on the local region of interest, and so the propagation of quantitative information should be considered. Therefore, an objective is to merge them in order to obtain an appropriate spatial vision for the study of the phytosanitary product diffusion pressure on the state of water in the Vesle basin. The merging operation should consider the spatial imprecision but also the quantitative information propagation. In this article, we present a study of choices (raster/vector, merging function) needed for the fusion of imprecise geographical entities and quantitative information.

This paper is a preliminary study for the building of the observatory. It would present our choice at the beginning of the project.

This paper is organized as follows. In the section 2, the modeling approaches for geographical entities are presented and their storing in a GIS. The section 3 illustrates implications in their fusion. The last section is devoted to the conclusion.

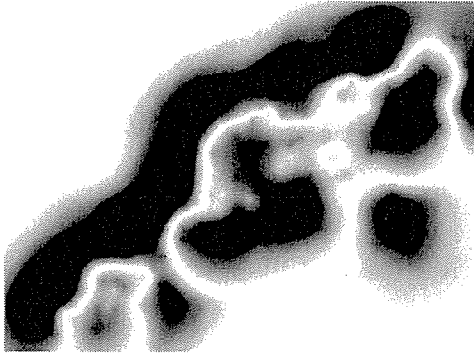


Figure 1. Example of a vague region (Dilo et al., 2004)

## II. SPATIAL IMPRECISION AND REPRESENTATION

This section is devoted to imprecise spatial data representation and their storage. Using spatial fuzzy sets is an usual way to represent imprecision (Zadeh, 1965; Klir and Yuan, 1995). This section highlights the modeling of fuzzy objects and its storage in a GIS.

### A. Fuzzy geographical objects

Fisher (1996) presents a comparative study between crisp sets and fuzzy sets in order to model landscape. The firsts simplify the modeling but could amplify errors. The others make complex the models and treatments. In (Dilo, De By and Stein, 2004), the concept of vague objects is introduced. Fig. 1 illustrates fuzzy regions: the grey scale of a point corresponds to its membership value ("black" is used for "inside", "white" for "outside" and intensity of any other color corresponds to the value of the membership degree of the location to the region).

Quantitative data attributed to a fuzzy region could also be imprecise and represented by a fuzzy set or a confidence index value.

There are two representation modes for storing spatial data: raster and vector. In order to take into consideration the imprecision, those two modes are adapted in order to record fuzzy spatial data.

### B. Fuzzy spatial object storage as raster

A classic in image processing for representing a fuzzy region after a fuzzy segmentation is to affect to each pixel the membership value computed for this pixel and the studied region. An example of fuzzy segmentation approach is presented in (Philipp-Foliguet, Bernardes Vieira and Sanfourche, 2002).

According to this representation, a raster representation of a fuzzy spatial object could be a matrix for which each cell value is computed using the membership function associated to the studied region. Guesguen and Hertzberg (2001) propose some algorithms in order to define fuzzy raster map. Duff and Guesguen (2002) develop the approach.

### C. Fuzzy spatial object storage as vector

There are two possible ways to store fuzzy data as vectors but both are based on multi-representation principle. For Dilo, de By and Stein (2004), "to store a primitive vague region it suffices to store the boundary of its core, together with the

boundary of its support set, as simple vague lines. [They] assume the membership value between the two boundaries decreases linearly". We should remark that this approach propagates, by the linear transformation, not only the imprecision but also the possible over-interpretation according to scale. The Figure 2 illustrates the storage of vague region according to Dilo de By and Stein. (2004).

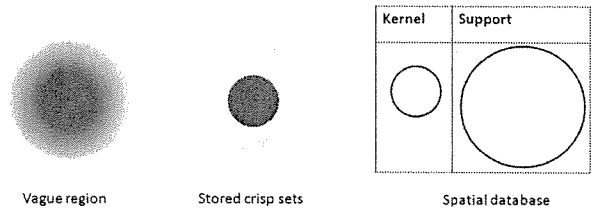


Figure 2. Storage of vague region according to Dilo et al. (2004)

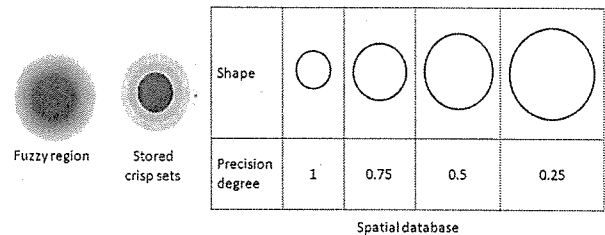


Figure 3. Storage of vague region according to Morris and Petry (2006)

Morris and Petry (2006) use the alpha-cuts principle (definition of crisp subsets on a fuzzy set). They consider a number of alpha-cuts for each fuzzy spatial object (only one if the object has a crisp/precise boundary), and each alpha-cut of that object will represent the boundary of that object with a certain degree of membership. The system may store any number of alpha-cuts for each fuzzy spatial object in accordance with the requiring degrees of precision as illustrated in Figure 3.

Morris and Petry proposed to store fuzzy spatial entities into UGML (Uncertain GML) files that allow us to describe multiple alpha-cut framework within the GML specification. In order to do that, UGML is based on the multiple spatial representations proposed to record spatial configuration according to temporal series.

As for UGML, this approach could simply be adapted in a classical GIS conception by replacing time associated to a spatial configuration by a degree of precision as presented in figure 3. It induces to store many spatial objects in the case of temporal series and fuzzy spatial entities. We should remark that this representation reduces the impact of over-interpretation according to scale.

According to each representation mode, the issue of object combination in the overlap surface arises.

### III. FUSION OF FUZZY GEOGRAPHICAL INFORMATION

The merging of fuzzy geographical entities is more complex than classical topological approach of fusion. Taking imprecision into consideration implies the use of fuzzy combination operators (*t-norm*, *t-conorm*, etc.) for relations between overlapped regions.

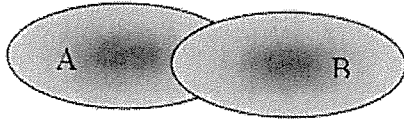


Figure 4. Two overlapped fuzzy regions

There are eight topological relations between simple (classical or crisp) regions. As illustrated in figure 4, the fusion of fuzzy spatial entities is more complex. Tang (2004) proposes to define more topological relations between fuzzy spatial objects. Alboody, Sedes and Inglada (2009) define 152 topological relations between fuzzy regions reduced as kernel, support and boundary. In this paper, we focus on the overlap of fuzzy spatial entities.

#### A. Fusion of fuzzy geographical entities

##### 1) A proposition of fuzzy geographical entity formalization

Let us define a fuzzy geographical entity *A* as an object described by:

- A label or concept  $L_A$  member of an ontology.
- A fuzzy set  $FS_A$  describing its spatial representation. The membership function  $\mu_A$  of  $FS_A$  is defined on  $\mathbb{R}^2$ .
- At least, a quantity (precise or not) valid in the region (e.g. a quantity  $Q_{P,A}$  of a phytosanitary product *P* using by spatial unit measure).

A quantity can be precise and be stored as a positive real number  $q_{P,A}$ . It can also be imprecise and be stored as a fuzzy set  $FQ_A$  (with a membership function  $\gamma_{P,A}$  defined on  $\mathbb{R}^+$ ).

During the merging, even if the quantity is precisely defined, the obtained result may be imprecise. Thus, precise quantity are represented by a singleton in the fuzzy set theory having ( $\gamma_{P,A}(q) = 1$  if  $y = q_{P,A}$ , else  $\gamma_{P,A}(q) = 0$ ,  $q$  belongs to  $\mathbb{R}^+$ ).

##### 2) Aggregation in an overlap region

Now, let us consider two fuzzy geographical *A* and *B* entities that overlap in  $A \otimes B$  (a set of  $x$ ).  $x$  is a location where  $\mu_A(x)$  (and  $\mu_B(x)$ ) takes a unique value in  $\mathbb{R}^+$ .

The fusion  $A \otimes B$  could be view as the aggregation of their spatial representation membership degree, the aggregation of their quantity for each present product and the affectation of a label corresponding to *A* and to *B*.

The label affected to  $x$  could simply be the lowest parent in the ontology shared by *A* and *B* (if the labels of *A* and *B* are different).

The spatial membership function  $FS_{A \otimes B}$  associated for  $A \otimes B$  to the affected label could be either a *t-norm* (as for

example the minimum (1)) or a *t-conorm* (as for example the maximum (2)).

$$\mu_{A \otimes B}(x) = \min(\mu_A(x), \mu_B(x)) \quad (1)$$

$$\mu_{A \otimes B}(x) = \max(\mu_A(x), \mu_B(x)) \quad (2)$$

If (1) is chosen, it could mean that we consider the intersection of *A* and *B*, if it is (2) it may mean that the union of *A* and *B* is used. In order to consider only the overlap region, (1) will be preferred in our application.

The aggregation of a quantity  $Q_{P,A \otimes B}$  (of phytosanitary product in Observox project) of a product *P* (with a membership function  $\gamma_{P,A \otimes B}$ ) should add the fuzzy sets  $FQ_A$  and  $FQ_B$ . In order to conserve all information about quantities, we use the traditional operator of addition in fuzzy set (3).

$$\gamma_{P,A \otimes B}(q) = \sup_{q=z+t}(\min(\gamma_{P,A}(z), \gamma_{P,B}(t))) \quad (3)$$

The adding aspect of (3) allows cumulating the membership degrees if the supports of  $FQ_A$  and  $FQ_B$  are overlapped. It is our first choice to model fusion of agricultural information.

In general, more aggregation operators could be chosen as Truck and Akdag (2009). Detyniecki proposes in (2000) a review of aggregation operators.

##### 3) On the point of view of a specific location $x$

If the objective is to obtain a fuzzy quantity  $Q_{P,x}$  (with  $\gamma_{P,x}$ ) of *P* for each location  $x$  of space in the overlap region between *A* and *B*, we have to consider both the membership of  $x$  to *A* and *B* and also  $FQ_A$  and  $FQ_B$ .

The first option could be to consider the previous approach and thus to obtain  $Q_{P,x}$  using  $Q_{P,A \otimes B}$  and  $FS_{A \otimes B}$  as propose in (4).

$$\gamma_{P,x}(q) = \mu_{A \otimes B}(x) * \gamma_{P,A \otimes B}(q) \quad (4)$$

The second possibility is to consider that the confidence in  $Q_{P,A}$  should be relativized by the membership degree  $\mu_A(x)$  in order to define  $Q_{P,A(x)}$  with its membership function  $\gamma_{P,A(x)}(q)$  as proposed in (5).

$$\gamma_{P,A(x)}(q) = \mu_A(x) * \gamma_{P,A}(q) \quad (5)$$

Using this hypothesis, we define  $Q_{P,x}$  using (6) for the definition of its membership function.

$$\gamma_{P,x}(q) = \sup_{q=z+t}(\min(\gamma_{P,A(x)}(z), \gamma_{P,B(x)}(t))) \quad (6)$$

Those two possibilities can be viewed as upper and lower limits of  $Q_{P,x}$ . In our approach, the first objective is to give indices with a degree of relativity, and we will give the two degrees.

#### B. Fusion according to Storage Mode

In order to reduce the time of analyses and computing process, our system has to minimize the number of merging calculi.

The intersection of fuzzy geographical entities stored as raster will be computed on each raster cell. In accordance with this operator, a misinterpretation (over and lower) of data is possible. Furthermore, the storage of those regions is

space expensive and thus time expensive in analyses and distant access.

In Dilo, de By and Stein (2004) representation mode, the combination should be computed in accordance to the linear diffusion of precision. The overlap region may simply be obtained using a topological division. This approach minimizes the space cost but may increase the time of analyses.

Using the multiple alpha-cut framework, the intersection of two fuzzy region could be viewed as a set of all intersections between the crisp set corresponding to the two fuzzy regions alpha-cuts. It will efficiency be obtained using a topological division. In this approach, the over-interpretation is limited by the combination of imprecision and the pre-minimization due to alpha-cut storage. Furthermore, the access time is reduced in comparison to other modes and the post computing cost is limited. The space cost is intermediary. This is the approach chosen in our project. We prefer it because it presents an intermediary between the two others and it limits the interpretation impact of vector representation mode.

#### IV. CONCLUSION

In this paper, we propose a study of the fusion between imprecise geographical entities in accordance with the storage representation mode.

We highlight the impact of the representation mode for data interpretation and fusion (we focus on the intersection/overlap). Thus, we give a reading grid of the link between data storage, data fusion and data analysis.

The thread of this paper is the future construction of an agricultural practice observatory. In our future work, we would use this study for data storage but we also want to study the propagation of quantitative imprecise information into other topological relations between fuzzy spatial objects.

This paper is a preliminary study for the building of the observatory. It presents our choice at the beginning of the project. In our future work, we would validate our approach and modeling choices by practice and indices construction in the observatory.

#### REFERENCES

- Alboody, A. Sedes, F. and Inglada, J. (2009). Fuzzy Intersection and Difference Model for Topological Relations. *IFSA-EUSFLAT*. Lisbon, Portugal. 1079-1084.
- Dilo, A., de By R.A. and Stein. A. (2004). Definition and implementation of vague objects. *ISSDQ'04*. vol. 1. 139-145.
- Duff, D. and Guesguen H. W. (2002). An evaluation of buffering algorithms in fuzzy GISs, *GIScience*, USA, 2478, 80-92.
- Detyniecki, M. (2000). *Mathematical aggregation operators and their application to video querying*. Paris, France: Université Pierre et Marie Curie (UPMC).
- Guesguen, H.W. and Duff D. (2001). Algorithms for Buffering Fuzzy Raster Maps. *FLAIRS*. Key West, Florida, USA. 543-546.
- Fisher, P. (1996). Boolean and Fuzzy Regions. In P. Burrough and A. Frank (Eds) *Geographic objects with indeterminate boundaries*.(pp. 87-94). London: Taylor and Francis.
- Fisher, P. (2005) Models of uncertainty in spatial data. In P. A. Longley, M. F. Goodchild, D. J. Maguire, and D.W. Rhind (Eds) *Geographical Information Systems. Principles, Techniques, Management and Applications*. (pp.191-205). Chichester: Wiley. Second Edition.
- Fonte, C. (2006). Conversion between the vector and raster data structures using Fuzzy Geographical Entities. *Spatial Accuracy*. Lisbon, Portugal.
- Klir, G.J. and Yuan, B. (1995). *Fuzzy sets and fuzzy logic: Theory and applications*. New York: Prentice Hall.
- Morris, A. (2003). A Framework for modeling uncertainty in spatial databases. *Transactions in GIS*. 7(1). 83-101.
- Morris, A and Petry. F. E. (2006). UGML: an extension of GML to provide support for geographic objects with uncertain boundaries. *Spatial Accuracy*. Lisbon, Portugal. 794-801.
- Philipp-Foliguet, S., Bernardes Vieira M. and Sanfourche M.. (2002). Fuzzy segmentation of color images and indexing of fuzzy regions. *CGIV*. Poitiers. France. 507-512.
- Schmitz, A. and Morris, A. (2006). Modeling and Manipulating Fuzzy Regions: Strategies to Define the Topological Relation between two Fuzzy Regions. *Control and Cybernetics*. vol. 35 (1). 73-95.
- Truck, I and Akdag, H. (2009). A tool for aggregation with words. *Information Science*. 179(14), 2317-2324
- Zadeh, L.A. (1965) Fuzzy Sets. *Information Control*. 8. 338-353