We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

6,600
Open access books available

177,000
International authors and editors

195M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
1. Introduction

This chapter deals with a vector model for fuzzy data sets within a GIS and with the raster and vector combination to refine vector objects. The model will be useful to represent information such as natural phenomenon (forest, desert, geology, separation between mountain and valley, rain, inundation, wind, etc.), social phenomenon (population density changes, poverty, etc.) or physical phenomenon (hurricane, etc.), which are diffused in space (Guesgen, 2000). The use of raster information will be useful to split vector units when the vector division reaches its limits.

The main sources of raster data are raw images (airborne and satellite images) and results of treatments (geostatistical, pixel classification, etc.). Vector information is often obtained by manual measurements (using GPS receptor for example), or is the result of the vectorisation of a raster treatment (such as classification, etc.).

The main goal is to manipulate vector information instead of raster because their manipulation within GIS engenders many problems. Indeed, the raster data are not adapted to GIS treatment (Benz, 2004) because of a lack of contextual information, the size of the data and the time consuming algorithm to produce information. The raster information is not split into identified objects and the vector representation is more flexible and gives the possibility to be easily combined with other information layers. For these reasons, we aim to split a wide image into several small units and convert the raw image information into a vector of features characterizing the unit in a raster view.

Another observation is that the vector data structures are not adapted to model fuzzy information in GIS (Shneider, 1999). More than 10 years later, the only way to use fuzzy representation in GIS is to build a raster map (Bjorke, 2004), computed with different raster or vector sources (Kimfung, 2009), (Ruiz, 2007). The only approach dealing with fuzzy vector representation uses a series of regular buffers inside and outside a polygon boundary.
to represent different belts of membership function values (Lewinski, 2004). The main drawback of this approach is that for most of the natural phenomena that engender fuzzy data (such as forest, population movement, physical phenomena like hurricane, etc.) the membership function doesn’t have the same behaviour in each spatial direction.

We propose in this chapter a vector model adapted to fuzzy information without the previous limitations. The chapter is organized in XX sections including this introduction. Section 2 details the fuzzy vector representation by successively introducing 
(i) a state of the art for fuzzy information representation within GIS 
(ii) the proposed fuzzy model 
(iii) an illustration of the model. Section 3 deals with the use of raster information to split vector units within certain hypothesis. Then a conclusion is given in section 4.

2. Fuzzy vector representation

2.1. State of the art

In 1997, in the conclusions of the special issue of Spatial Data Types for Database Systems in Lecture Notes in Computer Science the authors underline the importance of specific data structures for GIS and the lack of adaptation of the existing one to several data including fuzzy data. As if many propositions have been made, more than 10 years later the advance in this field are not relevant. The fuzzy modeling is only approached with strict sets. Historically the first approaches to solve the problem were to arbitrary decide of a strict border between fuzzy sets and to model them with classical data structures. But the raise of more complex problematics integrating many parameters reached the limits of this model which have guided the conception of the datastructures within the GIS.

The first studies about 2D fuzzy sets have been made by Peter Burrough in 1986 (Burrough, 1986) 106. After this, many studies with the definition of 2D fuzzy operators or fuzzy spatial relation (Bjorke, 2004), (Kimfung, 2009) have been made. But all this studies are using a raster representation of the data (Zhu, 2001), (Mukhopadhyay, 2002), (Sunila, 2004), (Bjorke, 2004), (Guo, 2004), (Ruiz, 2007), (Sawatzky, 2008), (Sunila, 2009), (Gary, 2010), (Wolfgang, 2011) and it was still the same in 2010. In all this studies it was necessary to rasterise the data in order to apply the fuzzy operators. This implies a loss of precision when choosing a scale of analysis, a time consuming process and the lack of fuzzy representation of sources data. This last point has been underline by GIS community as a main drawback (Altman, 1994), (Shneider, 1999), (Fisher, 2000), (Cross, 2001), (Yanar, 2004), (Kainz, 2011).

The actual challenge is then to directly deal with the vector data in order to improve precision, reduce computation time and to ensure a better abstraction of the data. Nowadays, only marginal studies include a vectorial approach to fuzzy spatial problems (Benz, 2004) (karimi, 2008). These approaches consist on a series of regular buffers around a strict polygon to represent different level of the membership function values. The main drawback of these approaches is that the regular evolution of the fuzzy membership function in each direction doesn’t translate the reality of most of the observed phenomenon.
2.2. Objective and data

In the study case, used to illustrate this chapter, we built a fuzzy representation of different forest in Guadeloupe Island. A forest is typically the kind of data adapted to the modelling with fuzzy sets. Indeed, the transition between two kinds of forest is mostly a gradient than a strict transition. Moreover, the gradient depends on many parameters and could be relatively short if environmental conditions change quickly or long if there is a smoother change. In some particular conditions there is a strict border for a forest, for example at the interface with agriculture or if a road, river, crest etc. interact with the forest. In any case, the transition gradient is locally defined and not uniform in every direction.

The classes are semantically and numerically defined using floristic information collected over 47 areas (about 250 m² each, Fig. 1-a). This step is based on a Principal Component Analysis (PCA, Fig. 1-b) and an ascending hierarchical clustering method (AHC, Fig. 1-c). These first steps allow regrouping the floristic information into significant clusters by sorting the numerous parameters describing the 47 areas. The AHC is used to select the number of classes and also serves as base for the semantic definition of the classes given through an ontology (Fig. 1-d) (Jones, 2002), (Kavouras, 2005), (Fonseca, 2002, 2006, 2008), (Baglioni, 2008), (Gutierrez, 2006). The ontology is useful for a high level use of the fuzzy representation of the forest and particularly when integrating it to a shared conceptual layer (Eigenhofer, 1991), (Cruz, 2005), (Bloch, 2006), (Grandchamp, 2011).

Figure 1. Clustering on floristic data and class definition
This step allows defining 14 kinds of forest over the main part of Guadeloupe Island (Fig. 1-d and e).

### 2.3. Description of the treatments

The previous step allows labelling each of the 47 areas according to the 14 classes (Fig. 2-a). But the fuzzy representation of the whole territory is not possible using floristic information because we didn’t have this information for any other area. Under the hypothesis that the environment directly influences the formation of the forests we project the 47 areas in a topographic and environmental space. This space includes information such as general ground occupation, elevation, exposition, slope, humidity or latitude. Each of this data is stored within a vector information layer and their fusion leads to the division of the territory into elementary areas called Vector Units (VU, Fig. 2-b). Each of the 47 ground truth areas is contained in a VU.

![Figure 2. Learning on topographic and image data based on decision tree](image)

Each VU represents a uniform area regarding each of the fused layers and we add to these features information extracted from a raster view (Raster Unit RU) of the area (texture and colour characterization using co-occurrence matrices, Law filters, Gabor filters, Hue moments, fractal dimension, etc.). We use satellite images (IKONOS, Spot5 and Quickbird) and airborne images. A total of about 20 features including topographic and image features are used.

This step of characterisation of the VU is the first Raster-Vector cooperation. It includes a partition of the image according to heterogeneous vector data and not to spectral or structural properties of the image. This approach is more simple and quicker than image segmentation and the returned VU have a semantic signification. Moreover, the adjunction of raster and vector information allows combining a theoretical uniformity of an area and a raster view of the reality.

With these 47 labelled VU we are now able to classify the whole territory in a fuzzy way. Indeed, we apply a supervised classification based on a decision tree (Fig. 2-b) to obtain the different kind of forest.

The decision tree is built after a learning step based on both topographic and image features. We use different kind of decision trees such as functional decision trees (FT) (Gama, 2005)
and C4.5 decision tree (Quinla, 1993). The best results were statistically obtained with FT, so we keep this approach to illustrate the method in the rest of the chapter.

The fuzzy model requires the computation of a membership function for each elementary unit (VU). This membership function is derived from the reliable coefficient returned by the decision tree. Indeed, each VU is analysed using the FT and a reliable coefficient is returned for each class. Commonly a strict classification using DT will choose to label the VU with the label of the class having the highest value of reliable coefficient.

By retaining only the highest value, we totally ignore the gradient nature of the transitions. By keeping all values we are able to build a fuzzy representation of each class and different representations of the resulting map by defining rules for the transitions. Moreover, in case of wide transition area, this approach could be used to reveal full transition classes.

Fig. 3 shows the membership function of each VU to different classes among the 14 identified classes of Guadeloupean forest. The lower the membership function value is, the darker is the colour. We remark that some classes are clearly localised, such as class 9 which represents high mountain forest around the sulphur mine, or class number 4 which is a typical forest kind oriented to the west and where a dry hot wind is blowing or else class 1. But other classes are more dispersed over the whole territory. This reveals wide transitions between classes. We remind that these maps are not raster data but each coloured element is a VU.

![Figure 3. Membership function](image-url)
2.4. The fuzzy model

Now we have all necessary information to build the fuzzy model of the forest. The simplest model is to store the vector of the membership degree values in each VU. This is also the more precise model because we don’t lose any data. But this model is not easily useful. In order to simplify the representation of the different transition gradients we decide to build different belts of membership degree values for each class. The spatial and topologic information used in the fuzzy classification process ensure the spatial coherence and compactness of the membership degree values. The models will differ from the number of belts and also the value of the threshold between the belts. These values will influence the treatments in two ways: (i) the more belts there is, the more precise are the results and the more time consuming are the treatments, (ii) the values of the threshold allow focusing on some parts of the transition.

So we will now see different ways to set the belts. Fig. 4 and Fig. 5 show different fuzzy representations of the class number 9. The differences are linked to the number of belts (5 for Fig. 4 and 10 for Fig. 5) and their positions: uniformly distributed (left), centred on most representated values (middle), centered on highest membership degree values (right).

Figure 4. Different fuzzy vectorial representation of a class 9 with 5 buffers
The choice of the number and positions of the belts is made by the user depending on the application. These values could be set manually or automatically computed using rules such as: natural breaks, equal intervals, geometrical interval, standard deviation, etc. This fuzzy data structure allows a fine and faithful representation of the heterogeneous evolution of the class in each direction. Indeed this model translates this property because the heterogeneity of the different information layers used (resulting in VU with totally different shapes) leads to a non-uniform evolution of the membership degrees in each direction.

Moreover, the choice of the thresholds influences the topology of the model and some resulting belts could be a unique connected polygon or on contrary a set of disjointed polygons.

2.5. Validation and strict view of the model

The only way to validate the model is to compare it to an existing model. However there isn’t any fuzzy model of the studied classes. So in order to validate the model we are going to compare a strict view of the model with an existing strict classification of the concerned forests. The referred classification is an ecological map obtained manually by biologists in 1996 (Rousteau, 1996).

Fig. 6 shows in left the reference map and in right the map obtained with a conversion of the fuzzy map into a strict map. The conversion has been made in the simplest way by
attributing to each VU the class label of the class having the highest membership degree value. Areas in white in the map are not taken into account in the classification. The main confusion is made between classes 9 and 12 in the east part of the Island.

Fig. 6. Strict classification comparison

Fig. 7 shows a zoom of the map on a particularly complex area: the National parc. We observe the similarity of the two maps with a localisation of the forest at the attended places. Fig. 7 -c) shows the differences (in black) between the ecologic map and the strict classification. The differences are localized at the limits between each kind of forest and this is exactly what is criticable in a strict classification of diffuse data. Taking into account that the map on the left (Fig. 7 -a) has been made manually with arbitrary decision concerning the limits of the forest and that the map in the middle (Fig. 7 -b) is the result of a complex and complete modelling, learning and classification process we can estimate that the results are of good quality.

Fig. 7. Resulting strict classification zoomed on Natural parc of Guadeloupe
Now we are interested on the transition between classes and want to focus on wide transition area in order to eventually decide to create transition classes. Fig. 8 shows different transitions between classes. We display in black all areas were the membership degree has a value under a fixed threshold. So in the first map, only few areas have a reliable coefficient under 0.5, and in the third map more areas but relatively few are under 0.9 or 0.95. These results tend to show that this uncertainty is relative low concerning the classes and that the transition gradient is high, that is to say that we quickly go from one forest to another.

We also observe that the width of the transition is not the same between every forest (for example compare transition between blue and green classes and between red and orange classes). This illustrates the good and attended property of the model which should translate the reality of the transition which is heterogeneous.

This first part of the chapter shows a new fuzzy model to represent data having diffuse and uncertain limits. The advantage of this model is that (i) it doesn’t require a rasterisation of the data which implies a loss of precision (ii) it doesn’t treat every transition in the same way and allows different transition width (iii) the model is enough flexible to be adapted for different applications. The building of the model uses a raster-vector cooperation to define the features during the classification process but the model is based on a vector representation.

It has been successfully apply to forest modelling and classification and gives results close to a reference manual classification made by biologists.

3. Vector unit refinement using raster information

3.1. Principle

This step aims to refine the Vector Units (VU) by using the Raster information computed on each Raster Unit (RU) separately (Grandchamp, 2009), (Grandchamp, 2010).

At this step, we consider that the VU couldn’t be refined using external vector information. This could be the case if all vector layers have been used to produce the VU.
or if the adjunction of any other vector information haven’t any sense (for semantic reasons or else) (Gahegan, 2008). For example if we split environmental VU with administrative borders. We also remind that a VU is a uniform area according to some specific vector information and that we expect that the VU will also have a uniform visual aspect. If not, or if we are looking for some specific area within a VU we have to split it into sub-VU.

So the only way to split the VU into sub-VU is to use the Raster information (satellite images, airborne images, etc.). Let us now consider two cases leading to the same process. Firstly we are looking for some specific objects within a VU and we hope their spectral or structural information is enough discriminating to detect them. Secondly we detect a non homogenous VU (regarding criterions detailed later) and we hope the homogenous sub-areas are semantically significant. In both cases, we make the hypothesis that the spectral or structural information contained in the raster view of the unit is helpful to split it into sub-VU.

The process is based on an homogenous criterion computed on some specific features describing the spectral or structural information of the image. As a VU has a semantic signification and a small size it’s often not divisible using its raster view if there is no external phenomenon which interact with the unit. Nevertheless, if the unit has been altered it’s often composed of a main sub-VU representing the most important part of the VU, one (or at most two) sub-VU having a semantic definition (in the case of ecological units, they are linked to landslide, deforestation, etc.) and a reject sub-unit composed of non identified or useless area (such as shadow).

The localisation and identification of the sub-VU are done in two main ways. Firstly in a supervised way, where both localisation and identification are done at the same time by a semantic and numeric definition of each sub-VU. We use in this case a classifier where each class is a sub-VU. Secondly in a semi-supervised or unsupervised way by applying a clustering method where at most the number of clusters is set. In this way, the identification must be done in a second step.

3.2. Process

We will now present the details of this Raster-Vector cooperation. Starting from a raw image (Fig. 8-a), and a vector splitting of the space (Fig. 8-b), we fuse both information (Fig. 8-c). The image is an extract of a very high resolution IKONOS satellite image (1m resolution) allowing an accurate classification (Gougeon, 2001) of vegetation (Junying, 2005), (Yu, 2006), (Johansen, 2007).

We compute an homogeneous coefficient on each Raster Unit (RU) (Fig. 8-d) in order to localise and later split the non uniform ones. The homogeneous coefficient is based on the measure of the compactness of the feature vectors computed on each pixel of the RU. This vector is composed of mean and standard deviation of each Color band (Angwine, 1998)
and textures features (over each color band too): co-occurrence matrices (Gotlieb, 1991), Gabor filters (Manjunath, 2010), Laws filters (Laws, 1980), Hue moments (Hu, 1962) and fractal dimension (Mandelbrot, 1977). There is a total of 25 normalized features including geometrical, statistical, frequential and fractal description of the RU (Abadi, 2008), (Scarpa, 2006).

Then the decision to split or not the VU depends on an empirical threshold of the homogeneity coefficient. In this case, the threshold has been set to 0.6 leading to one non homogenous unit (Fig. 8-d) with a coefficient of 0.37.

Figure 9. Non uniform VU localisation

Now we will localize and extract the sub-VU by analyzing the spectral and structural information contained in the RU. In this case, a semi-supervised classification is applied on the non homogenous RU extraction (Fig. 9 a). We use a K-Nearest-Neighbour (KNN) algorithm and set the number of classes to 3, with the aim to obtain classes corresponding to shadow, main ecosystem and secondary ecosystem. The Fig. 9-b) shows the classification results after a filtering post-treatment to eliminate isolated pixels. The class repartition is as follow: class 1 in blue (81%), class 2 in red (13,8%) and class 3 in yellow (5,2%).

The interesting sub-RU is identified has being the class 3 and corresponding to the a deforestation (Fig. 9-c). The identification has been made manually taking into account expert knowledge and environmental information. For example, this sub-VU could be a landslide or a deforestation, the choice has been made according to the slope of the Unit. Then a sub-VU is produced (Fig. 9-d).
The use of raster information to both evaluate the homogeneity of the VU and to split the VU into sub-VU is an efficient way to produce new vector information. The predivision of the image with the VU combined with the homogeneity evaluation considerably reduce the complexity of the classification problem by replacing the classification of the whole image by a selection of reduced classifications applied to non homogeneous RU.

The main problem of this process is the threshold the user has to fixe to decide to split or not a RU. A common way to use this tool and which avoids fixing the threshold is to sort the units according to their homogeneity value and to let the operator take the decision. Indeed, as a post identification of the sub-VU must be done by an operator, the best way is to let him appreciate the relevance of a sub division.

4. Conclusion

This chapter presents in a first part a vector model for fuzzy data sets within a GIS and in a second part a raster and vector cooperation to divide the space into elementary units represented in a vector way and called Vector Units (VU). The fuzzy vector model offers a new way to represent data which have diffuse borders without the limitations of classical approach which are based on a rasterisation of the information or on a regular evolution of the transitions in each direction. This model tries to combine both the flexibility of the vector representation and the accuracy of the raster representation. This fuzzy model allows revealing full transition classes that a strict model doesn’t allow. The raster-vector cooperation is linked to the mixing of both raster features (colour or textures) and vector features (humidity, elevation, etc.).
In a second part, this chapter deals with the use of raster features to split VU into sub-VU. The raster features are used to evaluate the homogeneity of the VU and in case of insufficient homogeneity to localise the region of interest in order to produce new sub-VU.

These two raster-vector cooperations are ways to produce flexible and useful information. They have been successfully used to classify the forest of a Caribbean Island named Guadeloupe and to localise landslide within its Natural parc.

Some improvement could be done concerning the model and particularly concerning the thresholds to define the different fuzzy belts. But the possibility to store the membership degree vectors in each VU allows producing as much fuzzy models as required.

Concerning the use of raster information to subdivide VU, a comparison (or a combination) with object oriented classification could be done (Maillot, 2008), (Forestier, 2008), (Hudelot, 2008). Object oriented classifications directly start with the raw images divided into elementary raster units (RU) according to a segmentation process (such as Watershed). The main drawback of this approach is the lack of semantic of the RU and the difficulty to retrieve semantically defined objects.

**Author details**

Enguerran Grandchamp  
*University of Antilles and Guyana, France, Guadeloupe*

**References**

Abadi, M., Phd Thesis, 2008, Couleur et texture pour la représentation et la classification d’images satellite multi-résolutions,


Bjorke, J. T., 2004,Topological relations between fuzzy regions: derivation of verbal terms, Fuzzy Sets and Systems 141 449–467

Bloch I. et al., 2006, “Modeling the spatial relation between for objects with very different spatial extensions”, In Proceedings of Reconnaissance des Formes et Intelligence Artificielle (RFIA’06), Tours, France,

Cruz, I., Sunna, W., and Ayloo, K., 2005, “Concept-level matching of geospatial ontologies”, In Proceedings of GISPlanet, Lisbon, Portugal,


Fonseca, F., Egenhofer, and al., 2002, “Using ontologies for integrated geographic information systems”, In Transactions in GIS, vol. 6, pp. 231–57,


Gary L., 2010, New fuzzy logic tools in ArcGIS 10, Raines and al., ESRI Communication,


Gougeon, A., and al, 2001, Individual Tree Crown Image Analysis – A Step Towards Precision Forestry. Natural Resources Canada,

Grandchamp, E., 2009, GIS information layer selection directed by remote sensing for ecological unit delineation, IGARSS

Grandchamp,E., 2010, Raster-vector cooperation algorithm for GIS, GeoProcessing

Grandchamp,E., 2011, Specification for a Shared Conceptual Layer in GIS, GeoProcessing


Johansen, K., and al., 2007, Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification, Remote Sensing of Environment, Vol. 110, No. 1., pp. 29-44


Junying, C., Qingjiu, T., 2005, Vegetation classification model based on high-resolution satellite imagery, Proceedings of SPIE, Remote sensing of the environment, 19-23 August, Guiyang City, China


Mukhopadhyay, B., 2002, Integrating exploration dataset in GIS using fuzzy inference modeling, GISdevelopment


Rodriguez, M. A., and Egenhofer, M., 2003, “Determining semantic similarity among entity classes from different ontologies”, In IEEE Transactions on Knowledge and Data Engineering, vol. 15, pp. 442–56,

Rousteau, A., 1996, Carte écologique de la Guadeloupe. 3 feuilles au 1/75.000ème et notice (36 p.). Conseil Général de la Guadeloupe, Office National des Forêts et Parc National de la Guadeloupe.

Ruiz, C., and al, 2007, The Development of a New Methodology Based on GIS and Fuzzy Logic to Locate Sustainable Industrial Areas, 10th AGILE International Conference on Geographic Information Science
Sawatzky, D., and al., 2008, Spatial Data Modeller, Technical Report,
Scarpa, G. Haindl, M., and Zerubia, J., 2006, Hierarchical finite state modeling for texture segmentation with application to forest classification. Research Report 6066, INRIA, France,
http://www.loa-cnr.it/Papers/FOIS98.pdf, last access : 17/01/2011,
Viegas, R., and Soares, V., 2006, “Querying a Geographic Database using an Ontology-Based Methodology”, In Brazilian Symposium on Geoinformatics (GEOINFO 2006), pp. 165-170, Brazil
Wolfgang Kainz, Fuzzy Logic and GIS, book chapter 1. Web document, last access 01/09/2011
ESRI International User Conference
Zandbergen, P.A., 2008, “Positional Accuracy of Spatial Data: Non-Normal Distributions and a Critique of the National standard for Spatial Data Accuracy”, In Transactions in GIS, vol. 12, issue 1, pp. 103–130,