A fuzzy logic approach to urban land-use mapping

Henning Sten Hansen

National Environmental Research Institute,
Ministry of Environment & Energy
Frederiksborgvej 399, DK-4000, Roskilde, Denmark
Phone : +45 46 30 18 07
Fax : +45 46 30 12 12
HSH@DMU.DK

Abstract. The growth of cities represents huge problems for modern societies. Monitoring, analysing and modelling the urban dynamic call for detailed mapping of urban land-use. Traditionally, urban land-use mapping is based on orthophotos and satellite images, but deriving land-use from remote-sensing alone is not satisfactory. The Danish Building & Dwelling Register is a database containing detailed information like year of construction, use, area etc. Therefore, this database provides a useful foundation for urban land-use mapping. To be able to track urban land-use changes over time, we have chosen square cells (100m x 100m) as basic mapping units. Generally, land cover and land-use mapping are based on crisp classification, but in the current project we have applied a fuzzy modeling approach to land-use mapping. Fuzzy classification offers a better choice in urban land-use mapping, because it can indicate the primary, secondary etc. land-use simultaneously. This will offer more meaningful information for planners and a more detailed understanding of the land-use patterns. Based on these principles, a nation wide urban land-use database for the year 1997 is established.

1 Introduction

Metropolitan areas all over the world are growing at unprecedented rates. Many of the farmlands, wetlands and forests have been transformed during the pat 100 years into human settlements, and urban growth rates show no signs of slowing. Urban sprawl reinforces the need to travel and increases independence upon private motorised transport, leading in turn to increased traffic congestion, energy consumption, polluting emissions and noise. However, many cities in Europe have gone through regeneration and recycling of existing structures rather than creating new ones. There have been significant advances in re-use of urban “brownfields”, thus reducing further outward urban expansion, although “brownfields” might constitute a problem in itself – especially if it is contaminated. Monitoring, analysing and modelling the urban dynamic call for detailed mapping of urban land-use.

Although the terms land cover and land-use are often used interchangeably, their actual meanings are quite different. Land cover refers to the surface cover on the ground, whether vegetation, water, build-up areas or other. Contrary land-use refers to the purpose the land serves, for example agriculture, industry or recreation. Land-use applications requires as well baseline mapping as continuing monitoring, since updated information is a prerequisite to know what current quantity of land is in what type of use and to identify the land-use changes from year to year. Thus, easy available data sources have high priority.

Urban land-use classification from remotely sensed images has achieved great attention in the past decades, and often land-use data are derived from remote sensing data alone. However, the results are not quite satisfying for detailed land-use detection in urban areas. This is because urban land-use is an abstract concept, defined in terms of function rather than form. Thus the relationship between land-use in urban areas and spectral responses recorded in satellite images is very complex and thus precluding the use of traditional classification approaches. Therefore, several alternative approaches have been suggested. These methods make use of ancillary spatial information like population distribution or road network, providing an improved classification accuracy (Sadler & Barnsley, 1990; Zhan et al., 2000). To avoid these problems, we have within the current project have chosen an approach based on detailed public registers on buildings, which are available for many European countries over a time span of some decades. Due to the temporal dimension within the building registers this approach also support the monitoring of land-use changes from year to year (Hansen, 2001).

Traditional classifications have been based on the old Aristotelian laws of thought: the law of *identity*, the law of *non-contradiction* and the principle of the *excluded middle*. Unlike traditional set theory where set memberships are crisp and binary, fuzzy set theory permits partial membership. The degree of membership is represented by a fuzzy membership value that ranges between zero and one, where the extremes 0 and 1 are the only available membership values in traditional set theory. Therefore, fuzzy classification offers a better choice in urban land-use mapping, because it can indicate the primary, secondary etc. land-use simultaneously. This will offer more meaningful information for planners and a more detailed understanding of the land-use patterns. Finally, it can easily be converted to crisp classification by simply labelling the primary classes to the land-use classes.

The aims of the current paper are to make a nation wide urban land-use database based on fuzzy sets and generally available public registers.

2 Data

Earlier, topographic maps were one of the main sources of land-use information, but satellite images and orthophotos have recently been used as additional sources of information. However, as mentioned above the physical reflectance processes of remote-sensing are not really recording land-use, but rather land cover.

The current study is based on primarily three national data sets, of which the Danish Building & Dwelling Register (BBR in Danish) plays the most important role. This database contains detailed information about every building in Denmark and has been in operation since 1977 (Hansen & Skov-Petersen, 2000; Daugbjerg & Hansen, 2000). The Building & Dwelling Register is maintained by the Danish local authorities as a part of the general administration of real estates and buildings. The register is among other things used for taxation purposes and therefore updated regularly and thus up to date.

The Building & Dwelling Register uses the following 3 levels of registration:

- Property level - type of ownership, sewage disposal system etc. are stored. A unique *property number* is the primary key at this level.
- Building level - purpose for which the building is used, year of construction, material of outer walls, building area etc. are stored. The individual buildings are identified by a composite key made by *property number* and a *building number*.
- Unit level - area of the unit, number of rooms, kitchen facilities etc. are stored. The dwelling units are identified by *addresses*.

Only the building level is considered within the current project. Besides this, the Building & Dwelling Register contains some additional foreign keys (e.g. municipality number, road code and house number / letter). In the current context, the purpose for which the building is used as well as the building area play key roles, making possible the assignment urban land-use category to the data. The spatial referencing is provided through the address assigned to all buildings in the database.

As mentioned earlier, nation-wide address points are not available now, but from the beginning of year 2003 a national address database with a very high accuracy should be freely available. Currently, the most obvious replacement of real address points seems to be the Danish Address & Road database (DAV in Danish) containing nearly every Danish road. Every road segment is assigned a municipality number, a road code and separate address range fields for each side of the road. Using ordinary address matching techniques, the records of the agricultural databases can be assigned x-y co-ordinates and presented as point events.

In order to fit into ArcView's Address concepts, a new string type column called Address is added to the Building & Dwelling database and the values of this column is calculated using Avenue syntax by the following expression:

```
Address = RoadName ++ HouseNumber,
```

where HouseNumber is a string type column because it may contain a letter as well.

Residential	
110	Farm houses
120	Detached houses
130	Terraced houses
140	Flats
150	Students hostels
160	Homes for children/young persons
190	Other residential buildings
Industry	
210	Buildings for agriculture, market gardens etc.
220	Factories and repair shops
230	Power stations, waterworks etc.
290	Other buildings for production
Trade and administration	
310	Transport buildings and garages
320	Offices and trade buildings
330	Hotels, restaurants and other service
390	Other buildings for office and business
Public institutions and culture	
410	Theatres, cinemas, museums etc.
420	Schools and research inst.
430	Buildings for health care
440	Kindergarten and children's homes
490	Other buildings (military barracks, prisons etc)
Recreation and leisure	
510	Summer houses
520	Holiday camps and youth hostels
530	Sports centres, swimming bath etc.
540	Allotment houses
590	Other buildings for leisure

Table 1. Building use-codes in the Building & Dwelling register.

The ++ operator indicates insertion of a blank character between two added strings. Most of the address matching is done automatically, but in some cases manual matching is necessary. This is generally due to bad representation of Danish letters 'Æ', 'Ø' and 'Å'. Of-course, there are differences between real addresses located within each building and addresses generated by address matching. However, in the current project these approximate addresses are able to locate the buildings within the square cells, nearly as good as ordinary address points (fig. 1), but a few address points are located in cells adjacent to the "correct" one.

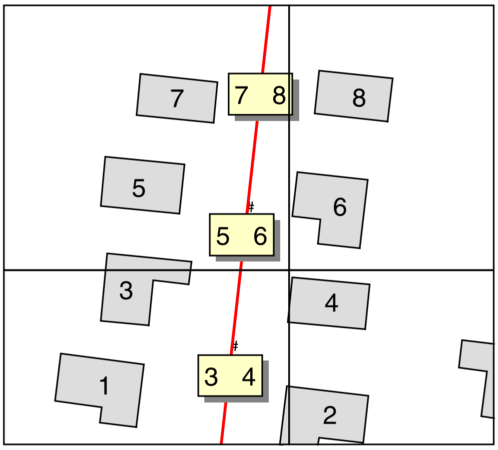

Fig. 1. Comparing geo-coded addresses with real world address points

Finally, the Danish national square grid is used. Grid cells are free of spatial bias and do not change over time as parcels or other administrative units. Hans Skov-Petersen (1999) examined various aggregation strategies and concluded that 100 m grid cells provide a good balance between homogeneity and "empty" cells. Each cell in the national square grid is uniquely identified be a string type key. The key is composed of the UTM-32 co-ordinate of the lower left corner of the grid cell. Thus the key corresponding to the 100 m grid cell with lower left co-ordinates 691200 and 6148300 respectively will be "6912 61483".

The Building & Dwelling Register was acquired from the Association of Local Authorities, whereas the Danish Address & Road database was obtained from the DAV Consortium (encompassing Danish telephone companies, regional natural gas companies and private map producers).

3 Application of fuzzy sets

Spatial phenomena are typically represented by sharply defined points with exactly known coordinates, lines linking a series of exactly known points, and regions bounded by exactly defined boundaries (lines). The properties of the space at the points, along the lines or within the regions are given by attributes, whose values are assumed to be constant over their total extent of the object (Schneider, 1999). The assumption of crisp boundaries fits very well with the internal representation and processing of spatial objects in a computer. Therefore, there has traditionally been a tendency to force reality into crisp objects. This kind of thinking seems to fit well for many anthropogenic objects like houses, roads and immaterial objects like parcels and administrative units. However, concepts of spatial vagueness or spatial indeterminacy is inherent in many geographic data describing natural, social and cultural phenomena like land-use, soil quality, vegetation, unemployment or native language. In these situations, the boundaries need not be lines but might be regions (transition zones). Furthermore, geographical phenomena are more complicated than many other entities because you have to consider both attribute space and geographical space (Burrough & McDonnell, 1998).

Thus the problem of dealing with imprecision and uncertainty is a part of human experience, but until Zadeh (1965) introduced the theory of fuzzy sets there was no strict mathematical method to handle impression and uncertainty. Zadeh (1965) published his work *Fuzzy Sets*, which described the mathematics of fuzzy set theory. This theory, which was a generalisation of classic set theory, allowed the membership functions to operate over the range of real numbers [0, 1]. The main characteristic of fuzziness is the grouping of individuals into classes that do not have sharply defined boundaries. Thus fuzzy sets are useful whenever we have to describe ambiguity, vagueness and ambivalence in models of empirical phenomena.

Fuzzy memberships differ from probabilities primarily in interpretation. Probability theory assumes that only one class or set is present and expresses the degree to which its presence is likely as a probability. The class with the highest probability is interpreted as the actual class. Fuzzy set theory accepts that multiple classes or sets can be present at one place and time and express the degree to which each class or set is present as a membership value. A vector of fuzzy memberships is maintained and classes or sets with non-zero memberships are interpreted as present to some degree.

A fuzzy subset, **A** of **Z**, is defined by a function $\mu_{\mathbf{A}}$

$$\mathbf{A} = \{\ z, \mu_{\mathbf{A}}(z)\ \} \quad \text{for each } z \in \mathbf{Z}. \qquad (1)$$

The relation $\mu_{\mathbf{A}}(z)$ is called a fuzzy membership function (MF^F) defining the grade of membership the object z in **A** and $z \in \mathbf{Z}$ indicates that z is an object contained in **Z**. For all **A**, $\mu_{\mathbf{A}}(z)$ takes on the values between and including 0 and 1.

A fuzzy membership function is thus an expression defining the grade of membership of z in **A** – i.e. a function that maps the fuzzy subset **A** into a membership value. Various types of fuzzy membership functions have been defined (fig. 2). The traditional Boolean membership function is represented by figure 2.a.

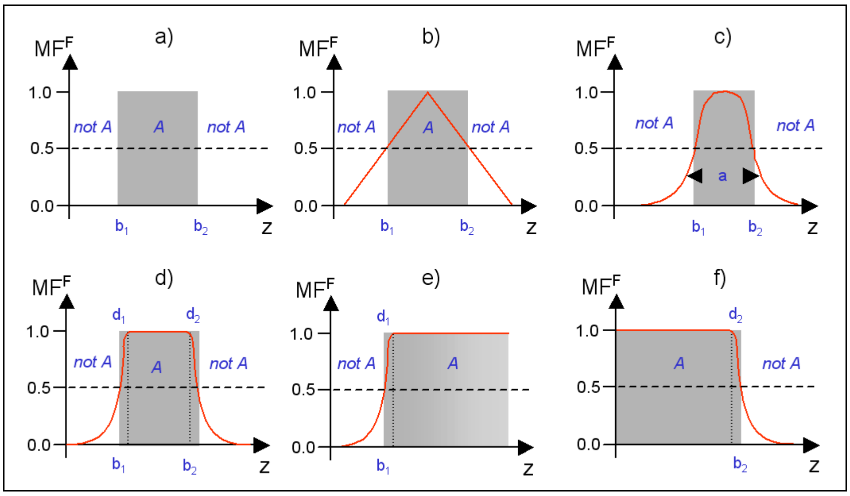

Figure 2. Selected fuzzy membership functions
(after Burrough & McDonnell, 1988)

Contrary to these membership functions figure 2.d has a central core and upper and lower transition zones. It seems reasonable to place the transition zones so the crossover points of fuzzy sets lie at the boundaries of the corresponding Boolean set (Burrough & McDonnell, 1998). The width of the transition zones d_1 and d_2 are rather difficult to choose, but currently we rely on some guidelines in traditional urban land-use mapping. Mathematically, this fuzzy membership function can be expressed by the following three equations:

$$MF^F(z) = \frac{1}{1+\left(\frac{z-b_1-d_1}{d_1}\right)^2} \qquad \text{for } z < b_1 + d_1 \qquad (2)$$

$$MF^F(z) = 1 \qquad \text{for } b_1 + d_1 \le z \le b_2 - d_2 \qquad (3)$$

$$MF^F(z) = \frac{1}{1+\left(\frac{z-b_2+d_2}{d_2}\right)^2} \qquad \text{for } z > b_2 - d_2 \qquad (4)$$

Concerning land-use, only the lower boundary has practical importance, and therefore figure 2.e will be used in the current project.

Definition of fuzzy membership values is central to the application of fuzzy sets to any problem. Multiple methods can be used to determine the membership values – e.g. depending on the amount of a priori information available. Two alternative approaches have been defined: The semantic import (SI) approach and the fuzzy c-means. The first relies on expert knowledge where boundary values are chosen by custom, law or an external taxonomy (Burrough, 1996). The second approach uses methods from cluster analysis and numerical taxonomy, and a frequently used technique is known as fuzzy c-means (Bezdek, 1981; Bezdek et al. 1984).

4 Land-use mapping

Urban land-use mapping is a process of simultaneous classification and aggregation. Classification combining separate phenomena into classes is a form of attribute generalisation. Aggregation referring to the combination of point data into polygons representing urban land-use categories is a form of spatial generalisation.

There exist numerous classification systems for land-use and land cover, and many countries and organisations have defined their own classification system. Among the most well known we can mention Land Use Land Cover (LULC) classification system from US Geological Survey (Anderson et al., 1976), and CORINE from the European Union (EEA, 1995). The LULC classification system has 4 levels of which the two upper levels are mandatory. Many other classification systems have their roots in these classification systems. Common to all the classification systems are a hierarchical approach, enabling different levels of details among users and applications. Additional, this approach will make it easier to generalise form detailed to coarser maps. The current land-use classification are partly inspired by the CORINE classification – although it is a land cover classification and partly on the statistical classification systems by EUROSTAT.

Urban land-use classes have been structured hierarchically according to their application fields and levels. At the overall level only three main classes are distinguished. These three classes refer to various aspects of a human's life namely: its *home*, its *work* and its *leisure*. The main argument for this division is, that every individual has to move between those classes everyday or at least some times a week.

- Residential
  This group consists of land-uses that serve as a place of permanent residence. Within this group we have single-family detached houses, terraced houses, flats and student's hostels. Dwellings closely related to business activities like farm houses are assigned to the appropriate business category.

- Business
  Land-uses within this group comprise a wide range of categories having only one thing in common – they serve as working places. Thus this group deals with the manufacturing and distribution of goods, shops and other retail outlets, offices for professional activities, schools and churches,

hospitals and other health facilities, hotels and restaurants, museums and galleries, etc.

- Leisure
  First of all this group covers summer houses, which are very widespread in the coastal zone. Next, sport facilities like football stadiums, sports and exhibition halls, golf courses etc. are of importance within this group.

The hierarchy and its relationships with the use codes in the Building and Dwelling register is illustrated at figure 3. The 3 classes at level one are divided into 17 classes at level two, thus creating a rather detailed urban land-use classification. This urban land-use classification scheme has been developed in cooperation with Hvidberg (2001).

The classification process is done in ArcView 3 (ESRI, 1996a) using the Avenue scripting language (ESRI, 1996b) and is divided into the following steps:

- Based on the classification scheme below an urban land-use class is assigned to every building in the Building & Dwelling Register.

- The UTM co-ordinates of the address point theme is added to the Building & Dwelling Register by using the Avenue requests:
  [Shape].GetCentroid.GetX
  [Shape].GetCentroid.GetY.

- The identification of the cell in which each address point is located can be determined by the following Avenue request:
  (Easting / 100).Trunc.AsString ++ (Northing / 100).Trunc.AsString

- For each unique cell id the total area of each urban land-use class is calculated. Thus a new land-use table is created with one row for each unique cell id and a column for each land-use class. The values in the table represent the summarised area of each land-use class within each cell. Cells with area values less than 200 m^2 are considered non-urban.

- The final step in the database creation process concerns the calculation of fuzzy membership values for each urban land-use class. As mentioned above, there are two different approaches to grouping individuals into fuzzy sets or classes. The approach chosen in the current context is based on the so-called Semantic Import model where a MF is pre-defined without any reference to the data. The class limits are specified based on experiences or definitions before individuals are allocated on the basis of how close they match the requirements of the classes. The fuzzy membership function applied is shown in figure 2.e and described by equation (2), where $b_1 = 75$ and $d_1 = 10$. For each class column in the urban land-use table an additional column is added containing the fuzzy membership values (FMV).

- For visualisation purposes the new urban land-use table is joined to the grid cell theme using the common cell identification field.

LEVEL 1	LEVEL 2	BDR-code
Residential	Detached houses	120
	Terraced houses	130
	Blocks of flats	140, 150
	Misc. residential	190
Business	Farming	110, 120
	Manufacturing	220
	Trade & administration	320
	Hotels & restaurants	330
	Education	420
	Health and social	430, 440, 160
	Culture	410
	Infrastructure	230, 310
	Misc. public service	390, 490
Leisure	Summer houses	510
	Allotment houses	540
	Sport facilities	530
	Other recreation	520, 590

Figure 3. Hierarchy of urban land-use classification

5 Results

Applying the data and methods described above, a land-use database covering all built-up areas in Denmark have been established. Figure 4 shows an extraction from the database focusing on Greater Copenhagen illustrating the traditional crisp classification. The dominant land-use category within each 100m square cell is found by defuzzification – i.e. by assigning the class with highest membership value to each cell. This crisp classification fits very well into the general Danish land-use mapping Ministry of Environment & Energy, 2000).

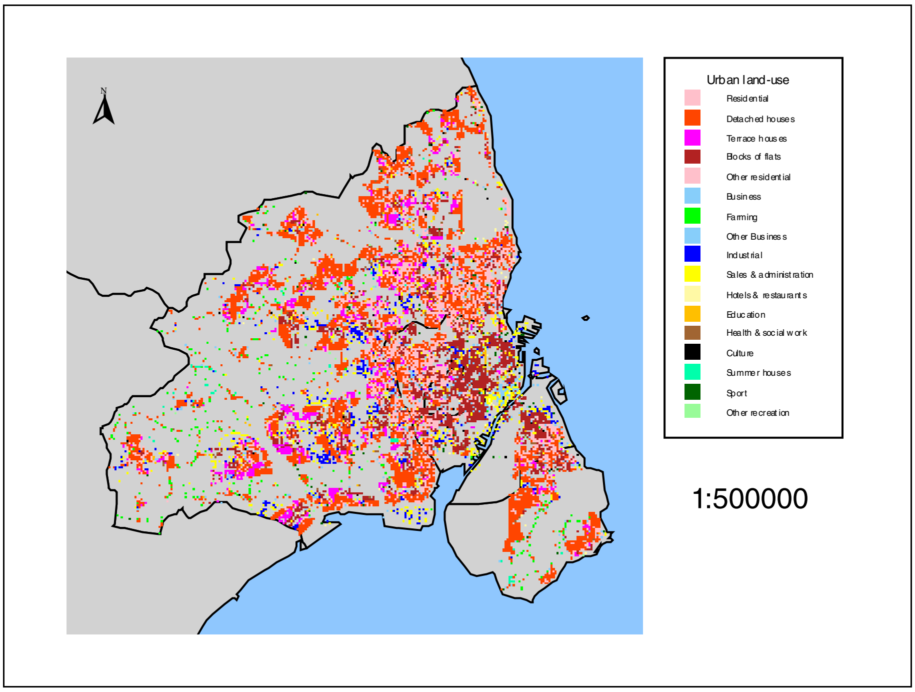

Figure 4. Crisp urban land-use classification based on the max. class principle.

Multiple membership maps are more detailed and can provide better ground for advanced spatial modelling than crisp polygon maps since they contain more information. The fuzzy membership maps themselves can be used separately. The maps of figure 5 illustrate the spatial distribution of the following two land-use categories: manufacturing and trade / administration. Comparing the continuous classification in figure 5.a with the corresponding crisp classification (Industry) in figure 4 reveals the much more detailed map using the fuzzy approach. Thus potential users of the fuzzy urban land-use database will have much more detailed input for advanced analysis and modelling.

If a cell has a fuzzy membership value at or near to 1.0 in one of the urban land-use classes, there is no doubt to which land-use class it belongs. On the other hand, if the fuzzy membership values for the two most dominating classes are near similar, it is by no means clear to which urban land-use class the cell should be allocated. There is a situation of confusion.

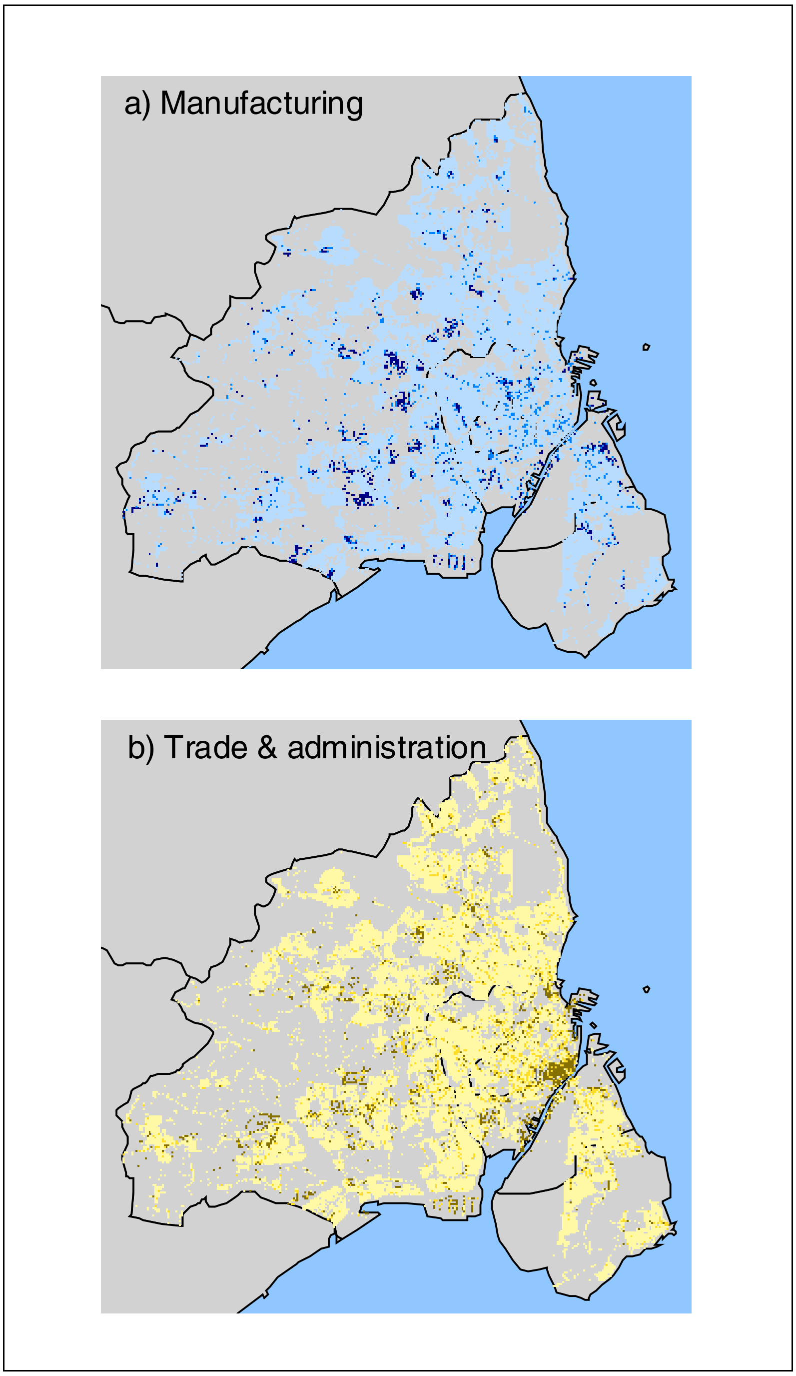

Figure 5. Continuous land-use mapping for selected categories.

The measure most commonly used to inspect interference between multiple membership maps is the so-called confusion index (CI), which is calculated as the difference between the highest and the second highest membership class per cell:

$$CI = 1.0 - (FMV_{max} - FMV_{2nd\ max}) \quad (5)$$

where FMV_{max} represents the highest fuzzy membership value and $FMV_{2nd\ max}$ represents the second highest membership value (Burrough, 1996). If the two values are similar, the confusion index returns a value close to one, indicating high confusion about class membership. If the two values are very different, then the confusion index is closer to zero, indicating less confusion concerning class membership.

Figure 6 shows the confusion index for Greater Copenhagen. The yellow spots refer to cells with rather low confusion. Although this colour is dominating with 16551 out of 23718 cells (70%) there are still a lot of areas with a high degree of confusion.

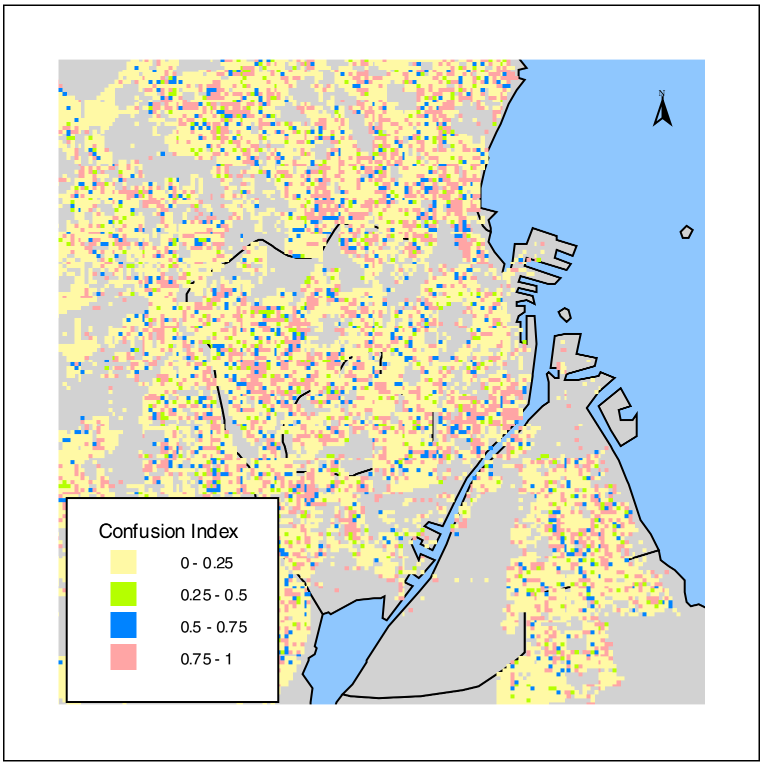

Figure 6. Confusion index

6 Concluding remarks

Traditional urban land-use is based on remotely sensed data like orthophotos and satellite images, but often the results are not quite satisfying for detailed land-use detection in urban areas. Urban land-use generally represents activities inside the buildings and therefore difficult to monitor by remote-sensing methods. The current project illustrates the usefulness of using public building registers to create urban land-use maps. The registers are general available in Scandinavian as well as other European countries. Furthermore, these registers are maintained daily, typically by the municipalities. The Danish Building & Dwelling Register has been in operation since 1977, thus making a valuable data set to track and analyse urban land-use changes during the last 25 years.

Generally land cover and land-use mapping are based on Boolean (crisp) principles assigning only one land-use category to each polygon, although some mapping projects introduce a few mixed classes (Hvidberg, 2001). One obvious way to circumvent these limitations is to apply fuzzy sets, thus providing a continuous classification of urban land-use.

The established urban land-use database provides a much more detailed input to various analysis tasks like monitoring the urban structure and estimating emissions to the atmosphere. The use of the spatial unbiased, temporal staple and easy to use square grid cell increases the range of possible applications of the urban land-use database.

Ongoing research and development are concerned with enhancing the database to cover the whole 25 years period from 1977 to 2002. The main problem in this ongoing project will be to find appropriate addresses representing the beginning of this period. This will provide an excellent foundation for urban simulations using cellular automata methods.

Acknowledgements

I would like to acknowledge the financial support from the former Ministry of Environment & Energy, and helpful suggestions during the course of this research from my colleagues within the Area Information System (AIS) project.

References

Anderson, J. R., Hardy, E. E., Roach, J. T. & Witmer, R. E. (1976). A Land Use and Land Cover Classification System for Use with Remote Sensor Data. *Professional Paper 964, U.S. Geological Survey* , Reston, VA.

Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

Bezdek, J.C. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers and Geosciences*, vol. 10, pp.191 – 203.

Burrough, P.A. (1996). Natural objects with indeterminate boundaries. In Burrough, P.A. & Frank, A.U. (Eds.) *Geographic Objects with Indeterminate Boundaries*. GISDATA 2. Taylor & Francis, London, pp. 3 – 28.

Burrough, P.A. & McDonnell, R.A. (1998). *Principles of Geographical Information Systems*. Oxford University Press, 1998.

EEA (1995). *CORINE Land Cover*. European Environment Agency, 1995.

Daugbjerg, P. & Hansen, K.V. (2000). *Property data*. National Survey & Cadastre, Copenhagen, 2000. (in Danish)

ESRI (1996a). *Introducing ArcView GIS Version 3*. Environmental System Research Institute Inc., Redlands, California.

ESRI (1996b). *Using Avenue*. Environmental System Research Institute Inc., Redlands, California.

Hansen, H.S. (2001). A quasi-four dimensional database for the built environment. In Westort, C.Y. (Ed.) Digital Earth Moving, *Lecture Notes in Computer Science*, vol. 2181, pp. 48 – 59.

Hansen, H.S. & Skov-Petersen, H. (2000). Digital Maps and Public Registers. *NERI Technical Report series no. 330*. 109 pp. National Environmental Research Institute. Roskilde. (in Danish)

Hvidberg, M. (2001). Urban land use mapping using register data and standard GIS. In Bjørke, J.T. & Tveite, H. (Eds. *Proceedings of the 8th Scandinavian Research Conference on Geographical Information Science*, Ås, Norway, pp. 141 – 156.

Ministry of Environment & Energy (2000). *The Area Information System*. Ministry of Environment and Energy, 2000. (in Danish)

Sadler, G.J. & Barnsley, M.J. (1990). Use of population data to improve classification accuracies in remotely sensed images of urban area. *Proceedings of EGIS´90*, Amsterdam, pp. 968 – 977.

Skov-Petersen, H. (1999). Spatial aggregation strategies – application in urban land use mapping. In Stubkjær, E. & Hansen, H.S. (Eds. *Proceedings of the 7th Scandinavian Research Conference on Geographical Information Science*, Aalborg, Denmark, pp. 131 – 142. Aalborg University Press, 1999.

Schneider, M. (1999). Uncertainty management for spatial data in databases: fuzzy spatial data types. In Gütig, Papadias, D. & Lochovsky, F. (Eds.): SSD'99, *Lecture Notes in Computer Science*, vol. 1651, pp. 330 – 351.

Zadeh, (1965). Fuzzy Sets. *Information and Control*, vol. 50, pp. 856 – 865.

Zhan, Q., Molenaar, M. & Gorte, B. (2000). Urban land-use classes with fuzzy membership and classification based on integration of remote sensing and GIS. *Proceedings IAPRS*, vol. XXXIII, Working Group IV/III.2.