United Nations

E/CONF.98/19/Add.1

Economic and Social Council

Distr.: General

26 June 2007

Original: English

Ninth United Nations Conference on the Standardization of Geographical Names
New York, 21 - 30 August 2007
Item 12 (f) of the provisional agenda*
Toponymic data files: National gazetteers

Beyond the Polygon

Submitted by Australia**

---

* E/CONF.98/1

** Prepared by Greg Windsor, Secretary Committee for Geographical Names in Australasia.

SUMMARY

Gazetteers have been maintained in various forms and content by each of the naming jurisdictions in Australia. In the past, these gazetteers have tended to contain point feature references only, although many of the process used during the naming of a place clearly indicate the extent of the subject features.

Recent discussion at CGNA meetings has included the need for the extent of features to be established and held as spatial data.

This paper will discuss the options considered and the possible directions to be followed in the creation of feature extents.

INTRODUCTION

Until the mid 1960's, concepts of regionally based geography were normal. Phenomena were mapped by hard and fast depictions. The natural subtle transitions of nature were not recognised. This regional based geography lent itself to the virtual environment and the polygon became the typical depiction method for quantifying known natural phenomena in a spatial context.

Geographers and Spatial Information experts alike have recently questioned the polygon's ability to depict the continuous nature of the natural environment. The development of geocomputational techniques such as fuzzy modelling and neural statistics has allowed these experts to model natural phenomena in a way that mimics nature. Further, these techniques provide an enhanced level of detailed analysis previously not thought possible.

To ensure nomenclature organisations within Australia and New Zealand remain relevant, the CGNA have studied the development of the geocomputation and are looking at issues surrounding spatial uncertainty and how geocomputational techniques can offer solutions in this area.

AN INTRODUCTION TO GEOCOMPUTATION

To understand the current theory and motivation behind geocomputation one must first look at a revolution which occurred in the discipline of geography in the late 1960's. Geographers of the time like Burton, Gregory, Chorley, Haggett, Cole, King and Harvey started to realise that traditional regionally based geography could not adequately describe the phenomena they wished to model. They recognized that a more systematic approach was needed to research and describe the complex natural structures they observed. The approach they adopted was based on sound scientific principles and sought to describe these natural structures through mathematics and statistics. This new approach has commonly been referred to as "quantitative geography". (Taylor P, 2003, p 7)

The introduction of computers in the 1970's allowed more sophisticated research to be carried out in this area and geographers could start to explore the seemingly endless amounts of geographic scenarios available in nature. (Taylor P, 2003, p 7)

The 1970's, however, also saw the first steps towards the development of technical computer based mapping systems; commonly referred to as Geographical Information Systems (GIS). GIS was seen as an excellent tool for migrating traditional hardcopy cartographic products into the digital environment. Its structure mimicked the layered approach of traditional map making practice and its architecture and design was relatively easy to describe to managers and decision makers (Couclelis H, 1998, Geocomputation in Context, p 19). GIS Constructs used for vector representations of spatial features, like point, line and polygons, were also perfect for the depiction of regional based geography. However, up until the mid 1990's, GIS tools did not exist which could adequately depict the landscapes and models described by quantitative geography. (Gahegan M, Geocomputation, Online)

Demand for GIS spawned a resource hungry development cycle which, to the detriment of quantitative geographical applications, saw expertise funnelled away from various related disciplines.

In the 1990's Geographers and GIS specialists alike started to recognise the limitation of GIS to show the complexity of real life geography. Investigations into techniques to depict and model these complex phenomena saw somewhat of a renaissance in quantitative geography. This, matched with the vast amount of geographical data now easily obtained from the GIS sector, created a new discipline commonly referred to as Geocomputation (Gahegan M, Geocomputation, Online)

SPATIAL UNCERTAINTY

It is rare in nature to see spatial phenomena or features with hard and fast boundaries. Most real world feature boundaries are not crisp but more transitional in nature. Mapping agencies all around the world are looking at different methodologies to depict and determine the extent of these real world features with transitional boundaries, commonly referred to as 'fuzzy features'.

An authoritative definition for the term fuzzy feature is hard to establish due to the nature of idiom being described. Different mapping agencies have different mapping programs, as such, it is hard to characterise the term to one type of data theme or structure. One definition suggests that a fuzzy feature is a depiction of a real world object with an uncertain extent or multiple possible interpretations of extent. (Ordnance Survey, 2005)

To date, most attempts to depict spatial feature extents have centred on a Boolean approach. This approach, as shown in the Venn diagram in figure 1, suggests that two intersecting regions, A and B, can only be in three states. These states are in union (A or B), intersecting (A and B) or inverse union.

This Boolean approach polygonises any feature being mapped and suggests that a single phenomenon can only reside within this polygon. The boundary of the polygon depicts the cessation of the given phenomenon and the commencement of a new phenomenon. As such, this approach fails to accurately depict the real life transition of certain natural phenomena or features into their neighbouring landscape.

Examples of these fuzzy features are many and varied and include the transition of vegetation, soil and geological feature types as well as the depiction of everyday topographical features like mountains, sand dunes and valleys etc.

FUZZY MODELLING

The deficiency of 'conventional set theory', which is based on Boolean sets, to adequately represent natural phenomena is well recognised. In the 1960's Lofti Zadeh proposed that the precision of most objects decreases as examination increases (Fisher P, 2002, p163). As such, if one looks hard enough, imprecision and uncertainty can be found in most parts of the natural world, it can include measures of time, temperature, speed and spatial phenomena, just to name a few. Since this time techniques have been developed to deal with this uncertainty. Probably one of the best known methods in this area is that of "fuzzy logic" which is based around the idea of the "fuzzy set".

Unlike conventional set theory, the fuzzy set allows for vagueness by permitting phenomenon to partly belong to any given set, this partial membership is generally termed the "degree of belonging". The degree of belonging can be measured and is represented in the range from 0, which depicts no belonging, to 1, which indicates complete belonging. Figure 2 shows a plot that depicts a typical transition of a fuzzy set from (0) degree of belonging through to (1) complete belonging compared with the traditional Boolean set. (Fisher P, 2002, p170)

This theory has been recognised as an effective method for depicting fuzzy features by spatial analysis experts like G. Foody and A. X. Zhu. Figure 3 shows the method that these specialists utilised whereby areas of uncertainty are shown in raster format by allocating a degree of belonging to each individual pixel. This diagram shows that pixels close to a known area of a certain phenomena have a high degree of belonging, usually 0.9 or alike. The degree of belonging then decreases the further the pixel is away from this known area until the pixel has low values such as 0.1. These pixels generally abut the known area of a different phenomenon or a known area where the original phenomena does not occur. By using fuzzy logic in this spatial context the represented transition between phenomena is smooth as in nature rather than abrupt as depicted by the use of conventional set theory.

DETERMINING THE DEGREE OF BELONGING

There is somewhat of a paradox when it comes to ascertaining the degree of belonging for any given pixel. One has to attribute, with some degree of certainty, a value of uncertainty against each of these pixels. Unless the methodology behind the allocation of the value of the degree of belonging is sound, the whole process could be seen as subjective and offer very little benefit to that of the traditional conventional set theory. The following information briefly outlines various geocomputational methodologies for determining the degree of belonging for raster pixels in areas of uncertainty.

Neural Networks

In 1995 Giles M Foody described a technique which utilised a Neural Network to detect sub-pixel composition in remotely sensed images. The Neural Network was trained by recording reflectance values of pixels over areas where ground surveys had been carried out and the sub pixel composition were known. This recorded data is then applied to the whole image to detect pixels over areas of uncertainty. These pixels are then attributed with a degree of belonging value. This value is based on a match between the reflectance value of the sample pixel and the pixel in the uncertain area. (Foody, Atkinson, 2002, pp 48-50)

The Mixture Density Network

The Mixture Density Network (MDN) described by Bishop in 1995 (in, Foody, Atkinson, 2002, pp 48-50) is based on a similar methodology to that of the Neural Network. The application is trained in much the same way except a probability distribution is applied to each pixel rather than a single degree of belonging value.

The probability distribution describes how consistent different degrees of belonging values are with the observed reflectance.

The MDN is seen to be a superior solution to that of the Neural Network. The Neural Network allocates one single value to each pixel even though the remotely sensed data is not detailed enough to derive this information with any certainty. In comparison, the MDN provides an estimated range of probable values thus giving a truer reflection of the ambiguity in the data being described. (Foody, Atkinson, 2002, pp 48-50)

Triangular Fuzzy Coordinates

Both the Neural Network and the MDN methodologies are designed to be used on remotely sensed data where some intelligence can be gathered on sub pixel composition in reference to uncertainty from reflectance values of pixels in the image.

In some circumstances, however, information about spatial uncertainty is not that readily available. Consider a topographical depiction of a mountain. Perception of a mountain's extent is subjective. Height above the surrounding landscape, slope and area are just some of the variables considered when making decisions on the extent for such a feature. Certainty can be given, in some instances, to the area which is the mountain and to the area which is not the mountain; however, a buffer generally remains between these two areas which is deemed to be fuzzy.

In 1998 Allan Brimicombe put forward a methodology that looked past trying to place a crisp boundary against fuzzy features. Brimicombe suggested that the transitional nature of a fuzzy feature's boundary should be mapped. As such rather than a definite boundary the feature should have a buffer zone. Further, that this zone should be based on a normally distributed probability curve. This suggests that in 95% of the time the fuzzy feature boundary could fall anywhere within this buffer zone; however, the probability is greater that the boundary will occur in the centre of the zone rather than at the outside. Figure 4 demonstrates Brimicombe's concept, the pyramidal shape reveals where the point is most likely located versus least likely.

Brimicombe's method is best suited to buffers in a vector environment, however, the principle he puts forward can be easily transposed to a fuzzy set in a raster environment.

Fuzzy membership can be depicted as normally distributed transition function, shown in figure 5, as a cumulative normal distribution.

This depiction matches nature in that it removes the sharp abrupt boundary and replaces it with a smooth transition, however, it is consider simplistic by some experts. (Fisher P, 2002, p 171)

SPATIAL FUZZY MODELLING CASE STUDIES

A number of projects have been completed that utilise a fuzzy model technique to derive sub pixel composition. Three case studies are briefly described below.

Fuzzy Land Information in Environment Remote Sensing (FLIERS) Project

The FLIERS project investigated a class of cereal crops in an agricultural region west of Leicester in the United Kingdom. It aimed to estimate the sub pixel composition of imagery covering this area.

An initial ground survey was conducted over areas of known uncertainty with their composition being recorded. This information was compared to a Landsat TM image (band 4) over the same area. The Neural Network was "taught" using the field data and the reflectance values from the Landsat image. The output from the Neural Network was then recorded and applied to the rest of the image covering the crop area. (Manslow J, Nixon, 2002, in *Uncertainty in remote Sensing and GIS* p 48)

Soil Land Inference Model (SoLIM)

Prof. A-Xing Zhu of the Department of Geography University of Wisconsin developed a system to conduct soil surveys and compile and depict data on soil types. As with the FLIERS project the system centres around the concept of using a fuzzy model in a raster environment.

Zhu suggests that soil type is determined by the mixture of environmental factors and time. As such, Zhu commences any given survey with a knowledge acquisition phase. This phase determines the "soil formative environmental conditions" for the survey area. This information is initially used to determine what datasets are required to establish the soil types for that area.

Zhu then uses a Neural Network technique to allocate the degree of belonging for each pixel. The Neural Network is trained by ground surveys and advice from local soil scientists. Unlike the FLIERS project which simply measured one phenomenon, the partial fuzzy membership values can represent a range of soil classes. Figure 6 summarises this process.

Zhu also presents a technique which allows the dataset to be converted back to a vector representation. Zhu's "Hardening" method converts the fuzzy boundaries back to crisp polygons by combining pixels according to the class which has the highest membership value. (Longley, Goodchild, Maguire, Rind, 2001, p 340)

Vampula Soil Maps (Finland)

A paper presented by Sunila, Laine and Kremenova outlines another method for the production of soil type maps. Unlike Zhu, Neural Networks were not used to determine the sub pixel composition of the areas of uncertainty between known soil type regions.

Sunila, Laine and Kremenova proposed a technique known as kriging to estimate the areas of uncertainty between known soil type regions. This technique utilises a ground survey close to the edge of each known polygon. From this known data an interpolated boundary for the area of uncertainty is derived. The pixel composition in this area is then determined by the kriging process. Hence, the fuzzy membership values for each pixel are then proportionate to the distance they are away from each neighbouring known area. Figure 7, depicts this process.

This technique was used to construct a soil map for Vampula, which is a coastal region in Finland. Sunila, Laine and Kremenova suggest that the resulting map clearly depicts the areas of imprecision along the boarder of each known soil type region.

THE FUTURE FOR FUZZY MODELLING IN A TOPOGRAPHICAL ENVIRONMENT

The use of Fuzzy Modelling techniques in remote sensing applications is becoming more widespread. Reflectance values generated from imagery products provide a foundation for these tools. However, the determination of fuzzy extents for topographical features is in its infancy. The limiting factor is the estimation of the degree of belonging for each pixel.

The SOLIM project offers some advice in this area. This project trained the Neural Network through knowledge acquisition rather than reflectance values. Logical relationships and definitions derived from experts were used as a basis for the interrogation of spatial data.

Using this same approach it is conceivable that topographical fuzzy feature extents could be derived through the use of Neural Networks, providing these logical relationships and definitions are ascertained for topographical feature types.

Practical rather than theoretical limitations are also seen as obstacles in developing "Spatial" Fuzzy Modelling applications. Currently most GIS tools are not designed to handle multi-valued cell information and are therefore restrictive in the attribution of sub pixel composition. These practical limitations are currently being addressed and it is not inconceivable that in the near future standard GIS packages will have built in "fuzzy systems" which will handle sub pixel land cover classifications. (Sherren K, 2003, p 42)

As these practical issues are addressed uses for spatial fuzzy models will grow exponentially. Wherever spatial uncertainty is observed in the natural world it will be theoretically possible to depict it in the virtual world.

SO WHAT DOES THIS MEAN FOR CGNA

Firstly, CGNA should acknowledge that there are and will be many ways to depict fuzziness with polygonisation only being one method. So, rather than developing standards on how to draw polygons CGNA should consider the development of logical definitions which can be applied to any number of depiction methods.

In 2002 David Blair first outlined a methodology which enables a logical description of feature types to be built using a binary approach applied to a standard set of semantic components. To demonstrate how the system works, Blair lists the semantic components and definitions for the major category titled 'hypsographic features'. This information is shown in the table below.

Hypsographic Feature's Semantic Components (Blair 2002)	
Semantic Component	Definition
ELEVATED	rises above its surrounds, and is therefore a raised relief feature
INDEPENDENT	is perceived as hypsographic in its own right, not as part of a larger feature
EXTENDED	is perceived as having significant length or extent

ISOLATED	rises conspicuously from level surrounds
STABLE	is not subject to obvious short-term deformation and relocation
HORIZONTAL	is perceived as having no significant gradient, and is characterised by absence of vertical aspect
BROAD	is perceived as having significant width
RAISED	has a further elevation within the context of a larger relief feature
PROJECTED	extends further from a larger relief feature in either the lateral or vertical dimension
OPEN	is bounded on only one side by a larger relief feature
DEEP	is characterised more by depth than by breadth
APICAL	is recognised as the uppermost part of a larger relief feature
TALL	has major elevation, arbitrarily set at >300m
SHEER	is characterised by extreme gradient

A feature type definition is then built by applying yes (+) or a no (-) response to each of the semantic components. For example, the feature type Mountain could be described as + Elevated, + Independent, - Extended, - Isolated, + Tall. Whilst a Hill could be described as + Elevated, + Independent, - Extended, - Isolated, - Tall, + Stable.

This taxonomy provides a logical method for describing fuzzy feature types. This logical approach allows the cartographer to view each feature type in a consistent manner.

Further, if the semantics used to describe the feature types are quantifiable, for example, the semantic of "TALL" has an elevation level of greater than 300metres. The taxonomy could be conceivably used in a knowledge acquisition phase to derive fuzzy topographic features in a Neural Network. That is, providing the relevant spatial datasets are available.

CONCLUSION

In the mid 1980's mapping agencies all around the world had to make decisions on how spatial data was to be collected and stored. Most of these saw vector representation through the use of polygons as the optimal solution.

In the following decades, Geographers and Spatial Analysts began to identify various limitations in polygon representations. Not least of which was the polygon's inflexibility to depict spatial uncertainty.

Geocomputation offers solutions to depict this uncertainty through fuzzy modelling and neural statistical techniques. These solutions offer advantages over the traditional polygon depiction by allowing an accurate depiction of the gradual transition between naturally occurring phenomena. They also allow further detailed analysis of remotely sensed images through sub pixel attribution.

David Blair's work in the development of taxonomy to describe feature types offers an opportunity to introduce a system that standardises the depiction of fuzzy features. Taken to its furthermost extreme this system may one day, through the use of geocomputational applications, provide the basis for an automated application for the determination of extent of these features.

DIAGRAMS

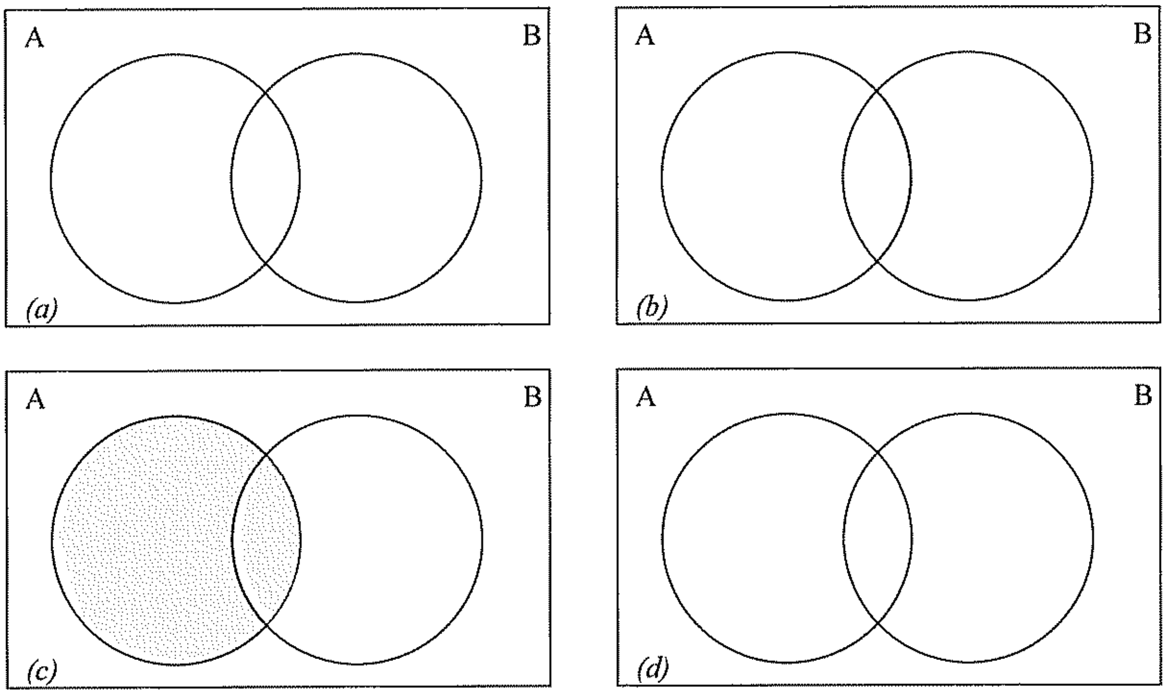


Figure 1. Venn diagram of two regions: (a) shows region A and B in the rectangular universe they occupy: (b) shaded area shows the intersection of A and B: (c) shaded area shows the union of A and B: (d) shaded area shows the inverse intersection of A and B. (Fisher P, 2002, p165)

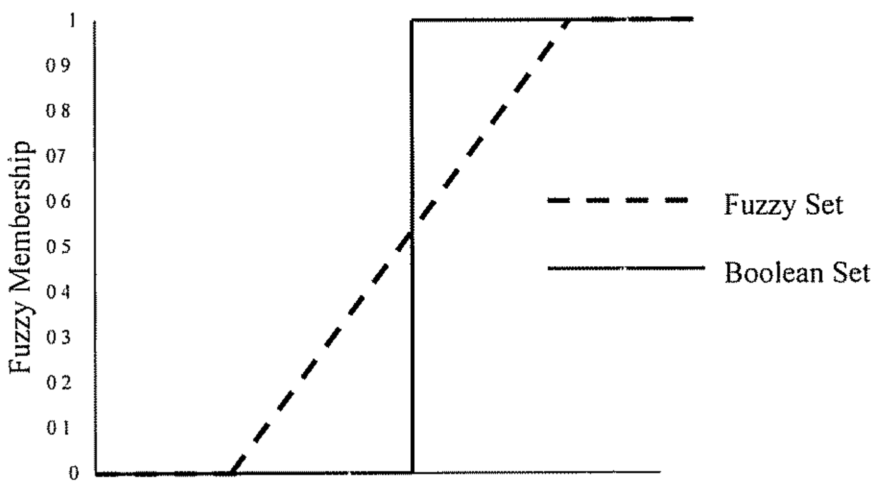


Figure 2. A plot of Boolean and fuzzy sets in terms of fuzzy membership (Fisher P, 2002, p240)

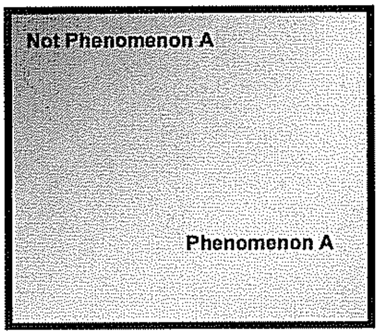


Image showing a spatial phenomenon (A) in green through to an area in red depicting an absence of this phenomenon

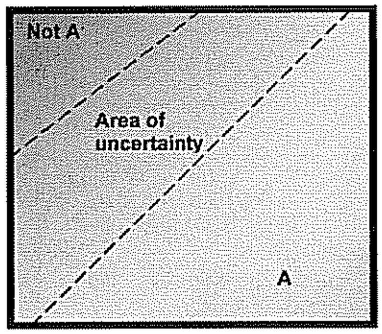
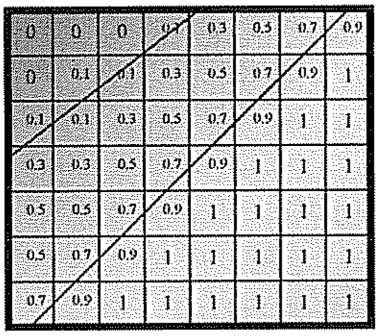
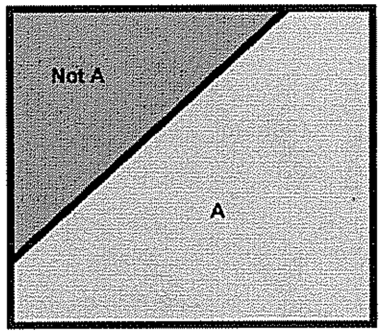


Image showing area of uncertainty between phenomenon A and area known not to be phenomenon A



Raster (Fuzzy Set0
Image showing raster representation of phenomenon A and area known not to be phenomenon A Pixels are allocated a value between 0 and 1 1 indicates full belonging. 0 represents no belonging



Vector (Boolean Set)
Image showing vector representation of phenomenon A and area known not to be phenomenon A Boundaries estimated by cartographic methods

Figure 3. Diagram showing a raster 'fuzzy set' and a vector 'Boolean set' depiction of a fuzzy feature

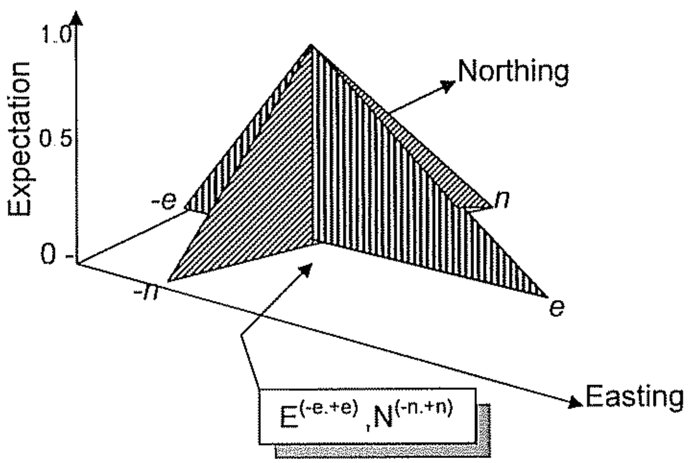


Figure 4 – Illustration of a two –dimensional triangular fuzzy coordinate (Brimicombe A, 1998, p144)

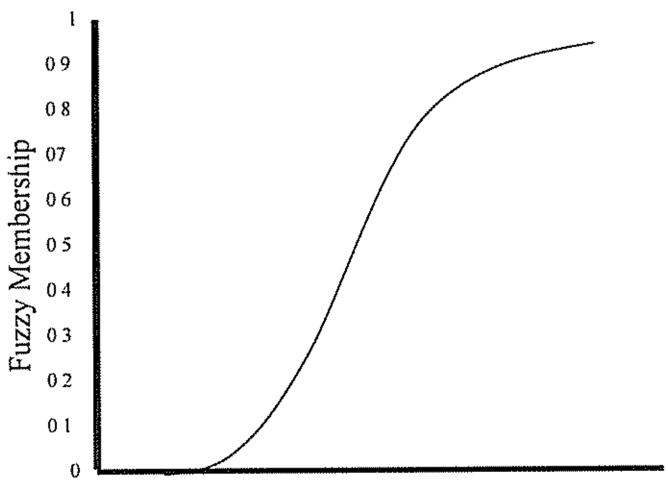


Figure 5. Fuzzy membership displayed a cumulative normal distribution.

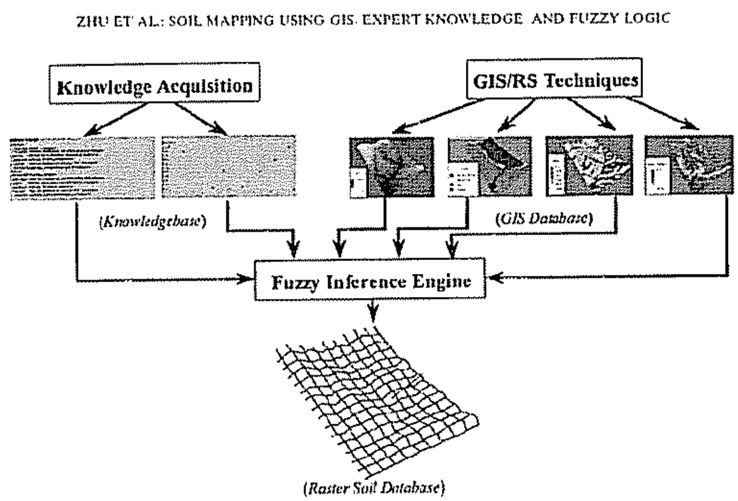


Figure 6. – Zhu's Soil Inference Process. The knowledgebase contains information on soil-environmental relationships. The GIS contains spatial data on soil formative environmental conditions. The fuzzy inference engine combines the relationship in the knowledgebase with the GIS data to produce a raster soil database of the study area. (Zhu, 2001, p 1467)

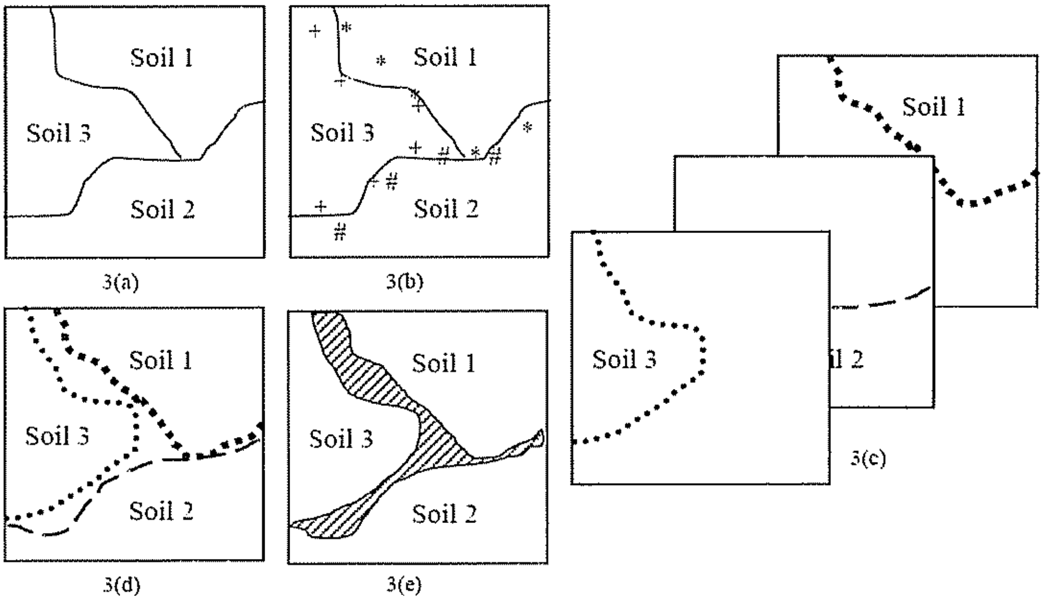


Figure 7. Concepts of Kriged Soil Maps. 3(a)Crisp polygon boundaries shown on soil map. 3(b) Samples are taken to ascertain areas of uncertainty. 3(c) Known areas derived. 3(d) Area of uncertainty derived. 3(e) Kriging used to derive pixel composition in area of uncertainty. (Sunila, Laine and Kremonova, 2004)