CHAPTER 5

SUMMARY AND CONCLUSIONS

5.1 Research Summary

The objective of this research was to examine how the usefulness and capabilities of a GIS in site suitability assessment and environmental impact prediction could be improved by incorporating (embedding) into a GIS the elements of an expert system. The following motivated the research:

(1) A GIS is seen as a convenient and well-structured information technology for handling large quantities of spatial data and related attributes. However, there is a widespread agreement that it does not provide sufficient capabilities needed to fully support complex spatial decision problems. The main reasons for this are:

- Inappropriate logical foundation which give no room for imprecision in information, human cognition, perception and thought process;
- Low level of intelligence in terms of possessing facilities for utilizing declarative and procedural knowledge;
- Lack of explicit analytical and modelling tools required to adequately explore the solution space of problem-specific and unstructured decision-making tasks.

In addition, the complexity of current GIS technology has generally showed a tendency to divert the process of spatial decision-making away from decision-makers into the hands of highly trained technological specialists and experts.

(2) Efforts to overcome these deficiencies of GIS by integrating them with decision support tools drawn from other disciplines have emerged as an important research area. However, they have not yet reached maturity since the majority of the Spatial Decision Support Systems are still in a research and development stage and the number of operational ones seems to be limited.

(3) The integration of GIS and knowledge-based systems (KBS) was found particularly interesting in site selection and environmental domains. Tasks in this spatial problem solving area are often unstructured requiring heuristics and other knowledge based techniques. This type of integration has become a substitute for purely knowledge driven expert systems because the latter proved to be to limited for a wide range of spatial problem solving tasks. This was mainly due to their inability to interact with spatial data, especially in cases where decision rules depend on geographical location.

Various logical ways of linking GIS and expert systems for spatial problem solving tasks have been identified. Of these, the most frequently used approach was the so-called loosely coupled GIS-KBS integration. Only a few attempts have been made to incorporate elements of knowledge-based techniques into GIS environment and to develop fully integrated spatially enabled Knowledge Based System.

To improve on the above issues, this study aimed to develop a practical approach for the integration of GIS and Knowledge Based Systems (KBS) to support site suitability assessment and environmental impact prediction. In an effort to meet the above requirements a prototype Knowledge-based GIS

(KBGIS) was developed that would be able to anticipate development-environment conflicts at an early stage of project planning.

This prototype KBGIS is based on the evaluation model developed by UNEP/UNCHS (habitat). It is essentially based on a checklist of problems and can be seen as a screening and diagnostic process for the identification of interactions between three sets of mutually related factors, namely location, development actions and environmental settings. The model involves the assignment of qualitative labels to site-specific development-environment conflicts based on the available data on the physical environment and the planned development action, as well as a set of generic rules (facts) for assessing and grading the likely consequences. This model has been used on several occasions as part of environmental planning and management routines to make urban and regional planning more responsive to environmental considerations.

These applications were however based on manual processing techniques. The objective of this research was to reconfigure this model into a KBGIS using automated techniques and computer technology. The conceptual approach and development of the various elements of the proposed system were discussed in chapter 4.

Generally, the whole research revolved around the idea to build an integrated set of computer-assisted procedures to produces a system that could be used as a "consultant" in the process of anticipating possible development-environment conflicts. To fulfil this task it was necessary to integrate the basic functionality of a GIS with elements of a Knowledge Based System.

Different levels of GIS-KBS integration have been suggested in the related literature. This research developed an example of the full GIS-KBS integration in which the elements of KBS techniques are actually one of the subroutines available within GIS.

The design of the prototype KBGIS places the model, data, domain knowledge, as well as the system's knowledge acquisition and reasoning mechanism together in a single GIS environment and within one single application with shared communication routines, common interface and data structure. The role of GIS within the system's implementation environment was to provide visualization tools, data and domain knowledge storage and management capabilities. It was also conceived as a slave to KBS with tasks of retrieving spatial and attribute data from the database ("where" and "what") and passing them to the system's reasoning mechanism (inference engine) for further analysis. The role of the KBS was to fumish the prototype system with easily accessible domain knowledge, as well as with the reasoning capabilities (inference mechanism) for identifying development-environment concerns in the language familiar to users. It was also conceived to act as an intelligent front end capable of controlling and guiding the communications between the user and the system.

5.2 Achievements of the Research

The application of the prototype KBGIS in an existing test area lends credibility to the results of this research. Some of the important achievements are as follows:

Firstly, the research proved that integrating different information technologies - in this case GIS and KBS - is a very useful approach to support screening and diagnostics tasks in site suitability assessment and environmental impact prediction. A further enhancement is the fact that the prototype KBGIS combines the functionality of "conventional" GIS with elements of KBS techniques.

The capabilities of the system to store facts and expert opinions (domain knowledge) and to emulate reasoning processes of experts seemed to have an important impact on the effectiveness and flexibility of the evaluation task in terms of:

1) Providing users with the high speed and reliable expert advice for the problems at hand, and
2) Eliminating the necessity of the involvement of a GIS expert and domain experts on development-environment impacts that would normally be required when using the "conventional" GIS techniques in generating different problem solving solutions.

The successful application of the prototype KBGIS system in a test area, has clearly illustrated:

1) How current GIS can be improved by linking GIS with a domain specific KBS capable of imitating expert reasoning processes in spatial problem solving situations.
2) How the limitations of the purely knowledge based system could be overcome by linking it with GIS tools.

Another achievement of this research, going beyond its basic objectives, was the design of the application specific and GIS-enabled expert system shell. Although fairly limited in scope and functions, it has proven to be capable not only to reproduce the specific domain of expertise it was designed for, but also to be adapted to other applications with a similar conceptual framework. This was made possible by implementing the so-called "plug-in" system architecture characterized by the separation of the system's "Knowledge Acquisition" and its "Consultation" mode.

Another noteworthy capability provided by the prototype KBGIS is the provision of an user-friendly and interactive graphical user interface (GUI) designed to control and guide the communication between the user and the system. The system's Knowledge Acquisition GUI is capable of emulating a so-called "paper and pen" environment and as such it bears a resemblance to an expert driven knowledge acquisition method in which the expert enters required information (facts) into dialogue fields and input boxes without assistance of a knowledge engineer. The KBGIS Consultation mode, on the other hand, operates as an icon-based graphic menu with the capabilities to:

1) Guide users easily through the consultation session;
2) Assist them in defining the problem and considering the possible outcomes; and
3) Allow them to examine the line of the reasoning process and refine it, if necessary.

These abilities make the system easy to implement, thus promoting its usability even by occasional users who usually demand a less complicated problem-solving environment.

The successful integration of the KBGIS modules and components into a single application within ArcView desktop GIS is yet another achievement of this research. It has revealed the possibility of using available customisation and programming utilities of desktop GIS to transform a conceptual knowledge model and include it as one of the analytical functions of a GIS.

5.3 Important Issues Relevant for Environmental Zoning and Construction of an Environmental Zone Map

During the development of the prototype KBGIS several issues related to the establishment of the GIS database for the Environmental Zone Map emerged. Two are elaborated with an intention to provide ideas for further research.

(1) Quality of Data to be used in the construction of an Environmental Zone Map

Based on the experience gained through the construction of the Environmental Zone Map for a test area, it appeared that the efficiency of GIS utilisation in supporting this task largely depends not only on correct choice, but also on quality of data.

As illustrated in the chapter 4, GIS databases required for environmental zoning would normally come from a variety of sources of analogue and digital data, each with its own characteristics, format, scale, positional and attribute accuracy. Accuracy of data is defined in terms of the magnitude of the difference between the value eventually reported and its true value. These differences are errors and they typically range from:

- Positional error in source material usually viewed as a discrepancy that might arise between the type of analogue or GIS data model and the nature of reality that it is seeking to capture;

- Errors in the attributes associated with spatial data.

- The impact of manipulation procedures e.g. digitising, logical consistency of data structure, overlay analyses, image processing etc.

Working with a combination of several different data layers from various sources will invariably result in error propagation. Consequently it would be naive to believe that an error free suitable and simple model could be devised under normal circumstances.

This issue is frequently out of the hands of practitioners involved in the construction of an Environmental Zone Map. From practical experience, it is apparent that either analogue or digital data sets often exist before the environmental zoning task is conceived and this task is usually designed to take advantage of what is available. Thus the issue here is not how the model represents reality (or ground truth) but rather how to understand and work with the existing data representation and uncertainty associated with the given data set (e.g. its confidence limits). Consequently, what a GIS database for environmental zoning seeks to accomplish is not a precise estimate of errors but some confidence that the error levels are not too high to doubt the validity of the results. This issue obviously needs an in-depth investigation since it influences the validity of KBGIS results.

(2) Formal Data Structure of an Environmental Zone Map

In the context of this research, the issue of data structure basically refers to a question of which one (raster or vector) would be better for the construction of an Environmental Zone Map. In practice there is no clear-cut preference and often both are combined to make use of their specific advantages.

For an Environmental Zone Map where it is important to maximize the accuracy of the spatial presentation of environmental features, the vector data structure seems to be more appropriate. On the other hand, the raster data structure has the advantage of being compatible with remotely sensed and other automated data capturing technologies, as well as its computational simplicity of spatial analysis and modelling. Furthermore, the vector data structure has richer data content, which basically means that a larger variety of database queries can be formulated. However, the overlay process within the vector domain can be time consuming and computationally intensive requiring the comparison of many line segments with many others for the purpose of detecting intersections and rebuilding the topologic and feature attribute tables. Sometimes, depending on the number and quality of data sets being aggregated, it can produce a cumbersome composite layer containing a large volume of very small polygons (not to mention slivers typical for the situation where input data sets with common boundaries do not geometrically coincide). On the other hand, the raster data structure provides a more flexible and efficient overlay capability, mostly due to a simpler data structure. However, this data structure has the problem of accuracy of spatial representation. It is well known that regularly spaced shapes rarely distinguish geographic phenomena. Therefore, grid cell in raster-based systems are usually classified as the most common attribute for the relevant cell. This leads to a problem of determining the proper resolution for particular data sets required for environmental zoning. If one selects too coarse a cell size then data may be fairly generalized and, therefore, less accurate. On the other hand, if the cell size is too fine then too many cells may be created resulting in a large data volume, slower processing time and greater request for storage space.

A brief comparison of the two main classes of data structure (vector and raster) sough to emphasize the fact that there are certain advantages and disadvantages associated with each data structure and the awareness of these advantages or disadvantages allows users to select the more appropriate one. It seems that the combination of the vector and raster data structures and their processing capabilities provides the greatest flexibility. However, this is not yet entirely achievable since there is still no GIS system capable of providing tools for integral vector-raster data sets aggregation that is required for the final construction of an Environmental Zone Map. In other words, GIS data must be, firstly transformed to either vector or raster format before they can be aggregated.

Although the research has not handled this issue explicitly it is my contention that future research efforts in enhancing the prototype KBGIS system should explore this theme in a detailed manner so that appropriate solutions for different environmental zoning requirements can be devised.

5.4 Directions for Further Research

As indicated above the prototype KBGIS developed in this research achieved a degree of effectiveness. It is however just a modest start or the first step towards a system with far more functions to support screening and diagnostic tasks in site suitability assessment and environmental impact prediction. Much work remains to be done especially in extending the system functionality and, subsequently improving its capabilities to assist all phases involved in development-environment impact evaluation and related decision-making processes.

The evaluation model applied in this research could be improved by adding an easily accessible repository of knowledge, which would:

1) Indicate the type and extent of mitigation measures that could be applied to overcome potential development - environment conflicts identified by the system for particular development actions at the selected location;

2) Provide recommendations for controlling identified potential conflicts, and for complying with the established standards and legislation.

This appears to be possible by constructing a knowledge browser containing a repository of the aforementioned mitigation measures and established standards or legislation. It would essentially represent a searchable hypertext system of help and explanatory text functions connected to the reasoning mechanism (inference engine).

Furthermore, research efforts should focus on the development of additional evaluation scenarios in order to improve and expand the current capabilities to support the model for site suitability assessment and environmental impact prediction. Two additional evaluation scenarios are considered important for the expansion of the system's capabilities. The first one is a scenario that would use the system's reasoning mechanism and its available GIS data and domain knowledge to assist users in the identification of the most suitable sites for an intended development action within an area of concern. The second could support the determination of development actions that are suitable or permitted at the particular location within a study area. Both scenarios are so-called "What if" types of analysis in order to assist users in the examination of the consequences of different planning proposals.

In addition, future research should also look at possibilities to add a set of computer-assisted procedures capable to automate the process of constructing the Environmental Zone Map. As indicated in chapter 4, this process is currently executed separately. Although GIS can facilitate the environmental zoning process it still remains difficult for a user to perform all the necessary data preparation and complex overlay procedures required for the construction of an Environmental Zone Map. Therefore, in order to make this task more user-friendly further research should focus on a strategy to automate overlay analysis. This will not be an easy task as it obviously aims at anticipating a number of possible options in order to keep the system sufficiently open to support different environmental zoning requirements.

Experience gained through this research reveals that improvements and extensions of the prototype KBGIS should be based on direct involvement of domain experts. It is their view of problems and experience that provides the necessary input in constructing the various domain databases.

The aforementioned especially refers to the provision of a valid and easily accessible repository of domain knowledge required for environmental evaluation tasks supported by the KBGIS. As explained in chapter 4, this domain knowledge is captured and stored in the form of:

1) Impact identification checklists for various development actions, and

2) Development-environment interaction matrix (as a basis for assessment of potential conflicts between the project's development implications and the site-specific environmental factors).

In many cases both the project impact identification checklist and development-environment interaction matrix can be incomplete in its coverage and miss important effects. In some cases they also try to cover such a wide range of implications and/or impacts that it is almost impossible to identify the key environmental concerns. In these cases they depend more on the background, expertise and experience of the people involved in its construction.

This research, although limited in scope, has clearly illustrated that desktop GIS could be efficiently used as an appropriate environment for the development of an intelligent GIS-based DSS capable of assisting unstructured spatial problem solving task.