3: GEOGRAPHICAL INFORMATION SYSTEMS METHODOLOGY

3.1. INTRODUCTION

Geographical Information Science is an interdisciplinary field advancing our basic understanding of digital geographical information and how to best utilise this information. It bridges geography, computer science, psychology, mathematics, philosophy, statistics, surveying, and earth sciences, among other disciplines (Goodchild, 1992; Wright, et al., 1997 *cited in* Kumar 2003). GIScience deals with the fundamental issues, which stem from the creation, storage, integration, analysis, modelling, distribution and dissemination of geographic information (Longley et al, 2001) in other words a Geographical Information System. In this thesis the geographical information system is used to arrive at one or numerous solution(s) to planning an ecological network. Utilising the ability of Geographical Information Systems (GIS) to integrate biophysical and socio-economic data (Eastman et. al. nd.). This is particularly useful and important in the effort to protect the environment. These systems are improving daily with Geographical Information Science driving the technology at the cutting edge (Goodchild 1997).

All aspects of the environment occur at a location, be it above, below or on the surface of the earth. The idea behind GIS is to model the earth by relating everything to location, the objects and processes between them, thus modelling the environment. This enables the modelling of various scenarios. Ascertaining the best solutions to the problem at hand or to possible problems in the future. A major part of the initial stage in developing an ecological network, is determining the criteria, which are bound to their location be it directly or indirectly by their location on or in the earths surface.

To produce these scenarios number of things are required amongst which are hardware, software, spatial data, data management and analysis procedures (methods), people and organisation and most of all, the problem (designing ecological networks) (Davis 2001, Heywood et al. 1998). This thesis will however not tackle all these aspects but rather focuses primarily on the analysis procedures, and the criteria, necessary to create a process to design an effective ecological network. An assumption is made that the data is available. Often the best processes and models get thwarted because the data is not available, incorrect or incomplete. However for any model the data required must be known, including what the required output is. Thus it will be stated what data is ideally required but no further investigation will be made as to whether the data is available or not.

To develop any model one must understand the problem (described in chapter 2) and analyse it to be able to isolate what the criteria, constraints and factors are. These will be further analysed in chapter 5. In this chapter a GIS process for developing a model is discussed in

further detail with respect to those parts of the process, which are core to this thesis. These being: the importance of scale and Cartographic modelling.

3.2. BUILDING A GIS MODEL

"… a Model is an abstraction of reality used to gain conceptual clarity and improve understanding of a system." — Berry 1993

The world is an intensely complicated body with millions of processes and systems which operate on it to make up the whole. Many of these systems and processes remain as "black boxes". Scientists are still in the process of trying to understand theses intricate relationships. GIS attempts to model this understanding of the world. The systems and processes need to be simplified so that they can be captured into a finite computer database (Goodchild 1993). This forms part of a GIS, which is really a simplified model of the real world (Berry, 1993; van Wyk, 2000; Longley *et. al.* 2001). When a modeller follows a process to model the real world and in turn uses that model to find solutions to application specific problems, many of our problems can be lessened.

A good base to follow in converting real world objects to GIS entities is to follow Peuquets (1990, *cited in* Heywood, 2000) four-stage model (figure 3.1).

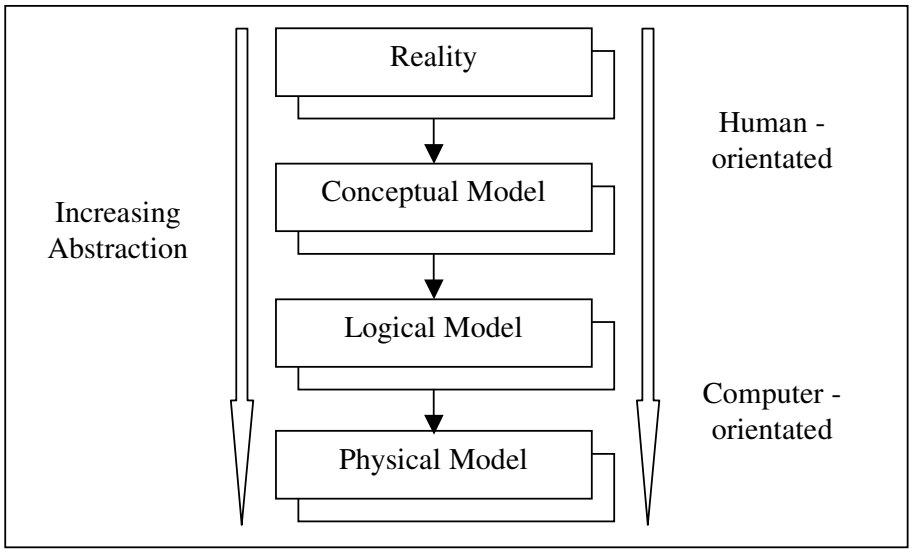

Figure 3.1: Levels of abstraction (From: Longley *et. al.* 2001)

The first stage—Reality—is to identify those spatial entities from the real world which are important, and decide how to represent them.

In the second stage—Conceptual model—the decision is taken as to which spatial data model (raster, vector or a combination) to store and represent the entities and finally your model in.

There is a move to try and combine the two models conceptually this has been done but very few software packages can truly combine the two.
The third stage—Logical model—is to instruct the computer as to how to recreate the entities.
The fourth stage—Physical model—is to instruct the computer as to the data structure (i.e. how to store the data). These stages are alternatively known as levels of abstraction (generalisation or simplification) (Longley *et. al.* 2001)

By expanding Peuquet's (1990, *cited in* Heywood, 2000) four-stage model; a process to convert the real world into a model (figure 3.2), can be developed. This can subsequently be used to find solutions to one or more problems. This Generic GIS process model (figure 3.2) may be summarised as: the point of departure is structuring the problem (chapter 1 & 2); converting real world objects into GIS entities; creating your database (beyond the scope of this thesis); developing (chapters 5 & 6) and using a model to find possible answers to your problem. Within this chapter only those parts of the process, which have been highlighted in figure 3.2, will be discussed in depth.

3.2.1. Structuring the Problem

This is the first stage in solving any problem (Heywood *et. al.* 1998; Puth 1996). In systems development methodologies, the systems analysis phase is like structuring the problem, which is regarded as the most critical phase (Whitten & Bentley, 1998 pg117). Similarly as early as 1968, the American Marketing Association (*cited in* Puth 1996) concluded that the most important part of solving any problem, is defining the problem (the first stage of figure 1). The problem—designing ecological networks—and its many issues, have been discussed in chapters 1 and 2.

If this is done correctly (to the fullest extent) then the next stage – Type of Model – the analyst should easily be able to answer the following problems in modelling: whether a raster or vector format or a combination, should be used i.e. what data model should be used (stage 2 figure 1) and at what scale the entities need to be captured?
This is a complicated question in relation to ecological networks.

3.2.2. Type of Model

There are many types of data models that have been used in GIS (Longley *et. al.* 2001). These range from simple mapping tools (CAD), images, raster, spaghetti (vector), topological (vector), networks and most recently to object models. Each of these models is at least loosely based on the conceptual, discrete object, or field and logical vector, or raster data models (Goodchild 1993; Longley *et. al.* 2001).

There are two basic spatial data models which computers can handle and display: Raster and Vector (Heywood et al 1998). The debate between using raster or vector is an ongoing one.

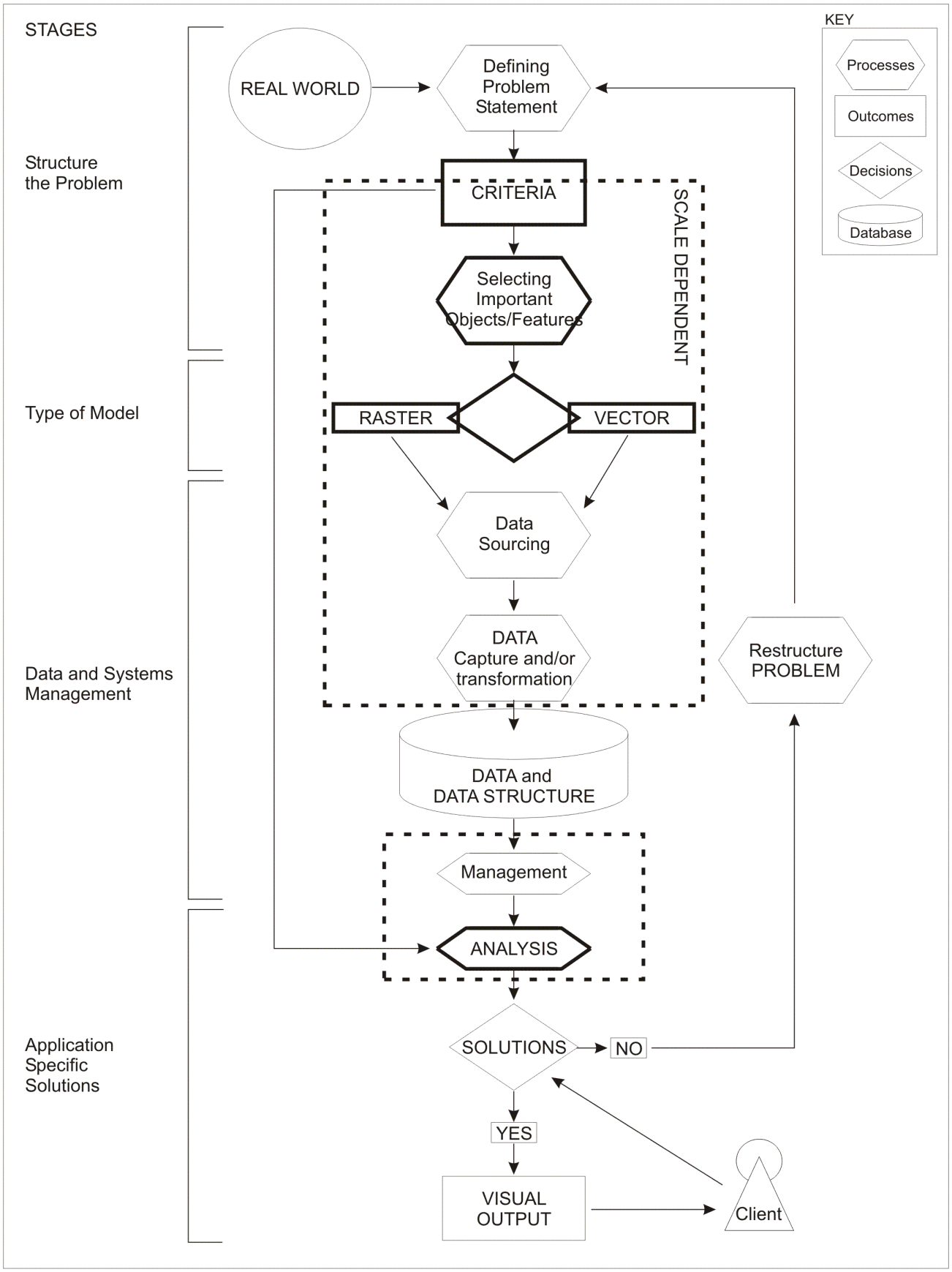

Figure 3.2: Generic GIS Process Model

Raster systems are excellent in processing continuous surfaces, performing complicated calculations in the form of 'map algebra', but are confined to the basic unit in the grid (cell, pixel, tessellation). Vector systems are excellent for inventories, maintenance (feature level), feature specific searches, thematic studies and topological queries. Essentially vector systems are good with feature specific processes, but are lacking in surface processing utilities. Vector data is focused on modelling discrete features with precise shapes and boundaries. Raster data is focused on modelling continuous phenomena and images of the earth. Thus the two spatial data models compliment each other.

The perception that each entity can be captured as a finite object / feature poses some problem. The real world consists of finite objects and "infinite" objects. The finite objects pose no problem in this case. The "infinite" objects or objects with fuzzy boundaries pose more of a challenge. Accepting that a purpose of a GIS model is to model the real world as closely as possible, then logic dictates that these infinite or fuzzy bounded entities should be captured as such. Real world objects such as soil types don't have finite boundaries, as a result these types of entities aren't generally captured correctly. This problem is currently being researched extensively (Assimakopoulos J. H., *et. al.*, 2003; A-Xing Z, *et. al.*, 1996; Dragicevic & Marceau 2000; Urba ski, 1999; Urba ski & Szymelfenig, 2003; amongst others) with Zadeh having done some pioneering work as early as 1965 with his work on *fuzzy sets* and *fuzzy logic.* This research must be considered during data capture to ensure that the model represents reality as closely as possible. It does however complicate some of the analysis in later stages of the modelling process.

Ideally a combination of both raster and vector systems, with the provision for fuzzy boundaries, would be best suited to the design of ecological networks. Goodchild (1993), Zeiler (1998), and Longley *et. al.* (2001) amongst others, provide clear focus of the different spatial data models. More detail into data models is beyond the scope of this thesis. The scale must be decided in conjunction with the data model as these is scale dependent.

3.2.2.1. Importance of Scale

> "*The importance of scale issues has even led some researchers to propose a new science of scale (Goodchild and Quattrochi, 1997; Meentemeyer and Box, 1987). Goodchild and Quattrochi (1997) suggested that a full science of scale would include: which measures or properties are invariant with respect to scale; methods to change scale; measures of the impact of scale change; scale as a parameter in process models; and implementation of multi-scale approaches.*" (Joao 2002)

This is beyond the scope of this thesis, but scale certainly plays a very important role and cannot be disregarded.

Any representation of reality is a simplification and is always smaller than the reality it represents (Monmonier 1991, *cited in* Heywood *et. al.* 1998; Keats, 1982 *cited in* Heywood *et. al.* 1998), by how much is indicated by the scale (Heywood *et. al.* 1998). Scale can be simply defined by the ratio of the distance on the map to that same distance on Earth (Martin, 1991; Longley *et. al.* 2001). In GIS, scale is defined as stated above. Considering that, GIS is flexible it allows a zooming function. The maximum (logical) scale at which data should be displayed for interpretation / visualisation purposes is determined by the scale at which data is captured. During editing, it is sometimes necessary to zoom into node level, forcing the display scale far larger than the scale of the source data or scale at which it was captured.

The effect of scale defines how much data and information (geographical and thematic detail) can be captured in a specific theme (Joao 2002). As scale increases, (as it tends towards 1:1) there is greater scope to include more detail; this concept is referred to as scale-related generalisation (Heywood *et. al.* 1998). Generally as scale increases, so the spatial extent decreases. That is, the spatial extent of the study area will often determine how much detail is possible (Joao 2002) i.e. scale. Scale affects data and analysis; it does not work in isolation. Scale affects processes; these can change from one hierarchical level to another (figure 3.3) as well as within a level. For example, all the processes which effect an *Acacia karoo* tree (at level 3 in figure 3.3) are not the same as those which effect a thicket of *Acacia karoo* trees (level 2) e.g. Giraffe eating the leaves of the tree. A same principal applies to Savanna (level 1). Although some of the processes may be the same, and effect all levels, e.g. rain. What effects one tree may indirectly effect the thicket or the Savanna and visa versa. It follows then that if only one scale is considered some critical processes may be neglected. This sounds like a simple enough concept. However “Scale is one of the most fundamental, yet poorly understood and confusing concepts, underlying research involving geographic information” (Montello and Golledge, 1999, *cited in* Joao 2002).

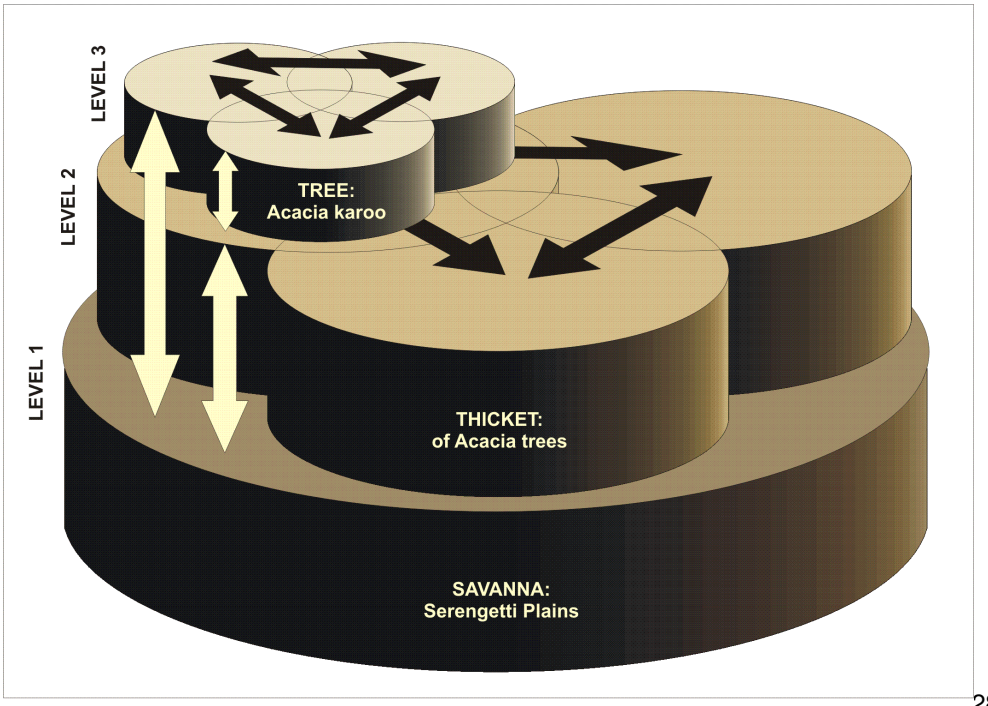

Figure 3.3: Scale and Processes

Goodchild and Quattrochi (1997) state: "…the generalisation associated with scale can be a way of obscuring information that is vital to [making] a correct decision, particularly when information must be integrated from a number of sciences and sources." Oppositely, they (Goodchild and Quattrochi, 1997) mention earlier that generalisation can add information in special cases where geographic patterns only become apparent at a distance, or hidden from unaided human perception (Mackey B, nd). This may be particularly useful in modelling the environment which functions at multiple levels and multiple scales. This notion supports the concept that "…the ecological volume of a landscape is far greater then its surface area." (Mackey B, nd).

A study by Joa˜o (2002) on the influence of scale effects on EIA's (Environmental Impact Assessments), found that changes in scale could affect the results of the EIA's. Hence changes in scale will also affect the design of Ecological Networks. Similarly, according to O'Neill and King (1998) "if you move far enough across scale, the dominant processes change. It is not just that things get bigger or smaller, but the phenomena themselves change."

The theorists Dutilleul (1998a and 1998b) and Martinez and Dunne (1998) discuss how unstable systems seem stable; Hinckley et al (1998) consider how bottom-up control gets inverted into top down control; and in the environment the importance of competition is reduced, and climate dominates (O'Neill and King, 1998) due to changes in scale. At the same time in application studies in aquatic ecosystems (Pahl-Wostl, 1998), terrestrial vegetation (Lertzman and Fall, 1998; Parker and Pickett, 1998), and geomorphologic dynamics the same changes are observed.

Take for example a group of trees in an orchard, for a vector model they can be captured as a group of point entities in a vegetation survey on a farm (figure 3.4a); as a single polygon at the scale of land use patterns across a province (figure 3.4b); or as a group of polygons at a very large scale where individual trees are considered habitats (figure 3.4c).

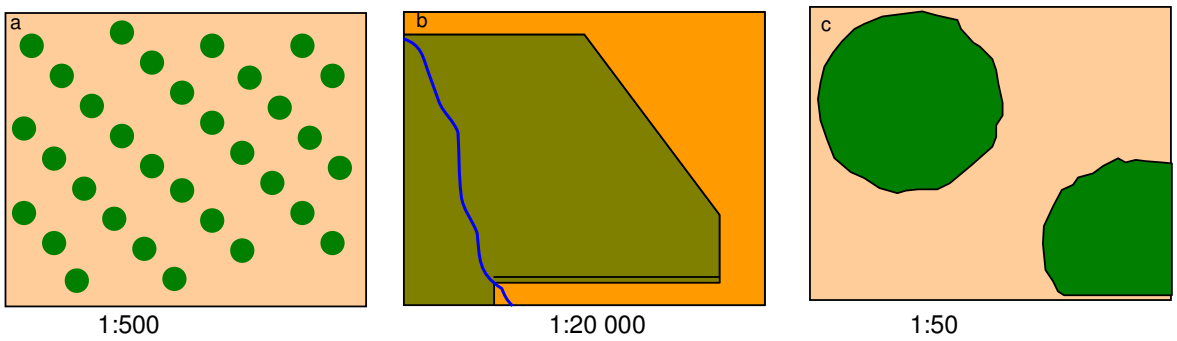

Figure 3.4: The capture scale of an Orchard

In most cases individual trees (figure 3.4a) would not be captured in a raster system as vector systems are perfect for capturing individual features. However, if the scale of the problem is as illustrated in figure 3.4b. A raster system might be a better option, depending on whether there are requirements to perform map algebra or not. A similar problem is found with all entities, although if the capture scale is determined then the entity type (point, line or polygon) usually follows directly.

The amount of research, on a meso- and topo scales has been little up untill now (Mackey nd). But it is an accepted principle that appropriately scaled information is required (USGS 2002, Whigham & Young, 2001; Mackey, nd) to underpin any study. Especially those regarding the complex decision-making needed to conserve Biodiversity (Mackey B, nd). Biodiversity is a dynamic system, if captured at one scale level (resolution), the explanation of system dynamics may not be applicable on another scale level (Huggett 1980). It is also possible that changes at scale B, can be explained by processes on scale A. As such it is important to capture data such that it can be utilised at multiple scales (USGS, 2002) especially when analysing the environment, which is multi-scaled by nature (Mackey B, nd).

From the discussion it is apparent that scale is important. It cannot be ignored or brushed aside in any stages of the model building process. It can have a major effect on the results of any model or analysis. Especially in this case where ecological networks are multi-scaled, multi-levelled phenomenon.

3.2.2.2. The Third Dimension and Time

Other decisions which have to be made include: whether the model should be 2D or 3D; and whether or not it is necessary to make the model time dependent? These in turn bring complications along, which need to be considered.

The decision as to whether the model should be 3D or 2D is usually a simple one. Only in a limited number of applications is it required that a 3D model is created. Due to the complexity of formulating such a model and the high processing power needed to perform even basic queries. Many software packages are not yet capable to model the third and even fourth dimensions. However many offer ‘3D’ fly thru’s and views. Some authors (Heywood. 2000, Wise. 2000) consider these as 2.5 dimensions and not 3 dimensions. Raper (1991, cited in Heywood. 2000) points out in the future 3-D modelling will be an integrated part of GIS (Longley and Batty (eds.) 2003). At this stage, with the limited capabilities of analysis in the third dimension, the benefits for the inclusion of the third dimension are not significant. Thus the third dimension will be excluded from the model.

The next task is modelling time. Many modellers ignore time as a dimension (the fourth) in their models. Only indicating the creation date in their metadata. The addition of the temporal dimension to the spatial dimensions is essential in order to accurately analyse dynamic

geographic phenomena (Dragicevic & Marceau, 2000). Due to problems in delays between *real world time*, *updating time, cartographic time and database (transaction) time* (as defined by Snodgrass, 1992, *cited in* Dragicevic & Marceau, 2000) the temporal dimension has not yet been fully integrated into the GIS framework. However, effort at a conceptual level has been made to improve temporal dynamic representation in GIS (Dragicevic & Marceau, 2000). Time and its scalability plays a significant role in the environment. In the short term (diurnal fluctuations), mid term (seasonal fluctuations), long term (Melancivic cycles), and the extended long term (geological time), time scales. The concept of an ecological network stems from the dynamics of the environment and movement of species (with time). Thus it is crucial to include time as an element, in the model.

3.2.3. Data and Systems Management

After answering the primary questions about what data model to use, the developer needs to source the data, capture and transform (stage 3 figure 3.2) it so that it may be stored in the data warehouse, ready for use. This part of the development is time consuming, tedious and often expensive (Longley *et. al.* 2001). The data capture alone can incur up to between 60 – 85% of the GIS budget (Longley *et. al.* 2001) and 50 – 80% of the project time (Heywood et. al. 1998). Despite this, this stage is often neglected in the planning and development of the GIS project / process with dire consequences (Heywood et. al. 1998). As such, this aspect of building a working model forms an important part. However, since the aim of this thesis is to develop a conceptual prototype, the actual data is not necessary. The capture of spatial data and its transformation is not included within the scope of this thesis. The assumption is made, that the data is available and may be used. This is done so that the available data does not determine the inputs to the model but that the ideal inputs maybe determined.

3.2.4. Application Specific Solutions

The first step of this stage—Analysis—is possibly, after determining the criteria, the most important. This is where the power of GIS lies. It is here where the criteria are assessed, analysed, and combined to result in possible solutions to the problem. Once the solution(s) have been found, these need to be readied, to be presented to the “client” (stakeholders, participants). This process is called visualisation. This can result in maps, graphs, charts, tables, fly-thru, posters etc. If, however the solutions are not acceptable then the problem needs to be redefined, in a feedback system.

3.2.4.1. Criteria

Criteria form the starting blocks of both, deciding what objects to include in the model, as well as the method of analysis, not forgetting the scale at which this occurs. It is this, which determines which methodology to use to analyse and combine the data in order to reach a solution. A good method to use to structure the concepts and criteria, is to use a concept

model (see chapter 6). The concept model inherently defines the scale of the model. There after, it is easier to convert these ideas and criteria to a cartographic model.

3.2.4.2. Cartographical Modelling

Cartographic modelling has become an accepted method for processing spatial information (Heywood et. al. 1998), through “Map Algebra” (as described by Tomlin 1983 cited in Heywood et. al. 1998) and “Mapematics” (as described by Berry 1987 cited in Heywood et. al. 1998; Berry, 1993). Tomlin, Steinitz and colleagues (Burrough 1986) showed that it is possible to create an unlimited number of tailor made map-processing capabilities that can be used to tackle many analytical problems (Burrough 1986). This is achieved by combining a number of simple analytical procedures in a specific combination (a command sequence) to perform a specific task (figure 3.5). The number of combinations tends towards infinity. Tomlin, Steinitz and colleagues called the command sequences a cartographic model and the process cartographic modelling (Burrough 1986).

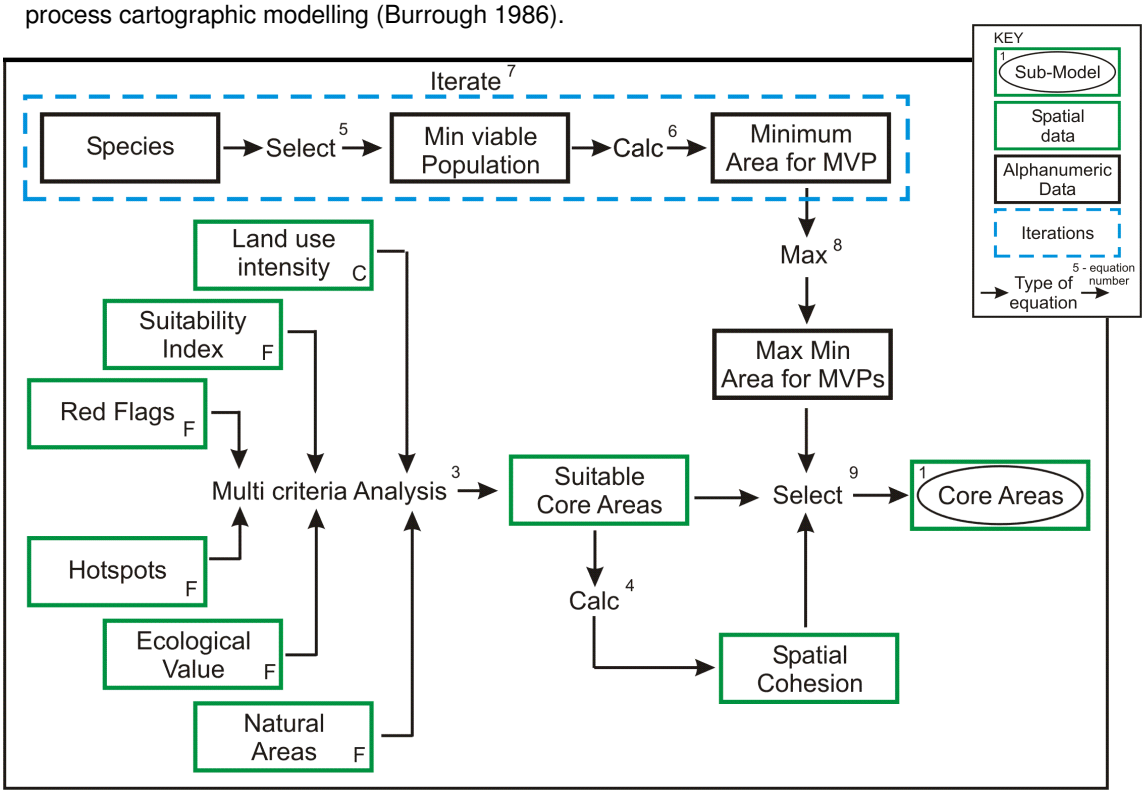

Figure 3.5: Excerpt from the Cartographic Model

For the purpose of this thesis, the above explanation will suffice. Using cartographic modelling to base the analysis on is beneficial, as it allows numerous criteria and processes to be integrated. Most importantly, it records the process in such a manner that it is not software dependent. Hence, the cartographic model for the design of ecological networks will form a core result of the thesis.

3.2.4.3. Solutions

The second step in the *application specific solutions* stage, deals with whether a solution has been reached. But more importantly, if this solution is feasible and an acceptable option. If not, the process needs to be repeated. In most cases there is some reworking and re-analysis to reach the best possible solution. This process should not be seen as a linear one but rather as a cyclic process with numerous iterations continuing until such a time where the solution is acceptable. Or in some cases where the time constraints govern the number of iterations. If the solution is acceptable and feasible the process is continued into the next step, which is *visualisation*. This is ultimately what the client (policy maker etc) sees. Visualisation is highly important. It is a large area of research – growing daily. At the recent (August 2003) International Cartographic Conference, held in Durban, two whole sessions were dedicated to Visualisation with one dedicated to "visualisation for the blind". It is sufficient to note that visualisation is very important but further discussion is beyond the scope of this thesis.

3.3. SUMMARY

Working through one of many modelling processes, a model is created which should solve the problem at hand—designing an ecological network. If done well, can be used to solve unforeseen problems, which may occur in the future. The environment is a extremely complicated and dynamic body, with finite and infinite objects with fuzzy boundaries, three-dimensional, scale, and time dependent processes. The ideal would be, to be able to model the world in all its dimensions – this is not possible. Cartographic modelling facilitates—what is deemed the greatest advantage of GIS according to Goodchild (1993)—the ability to handle and integrate multiple models. As multi-scale dimensions are included, so the power of these models increase. However "It is important to note that no matter how well the model is built, it is always an abstraction of reality." (van Wyk, 2000).

Now, that the method is understood. The criteria must be defined. In chapter 4 an exploration of case studies weeds out the common criteria. Followed by an assessment of their importance (based on expert opinion) in chapter 5. From this the analytical process (the model) is determined and documented in chapter 6.