CARTOGRAPHIC MODELLING

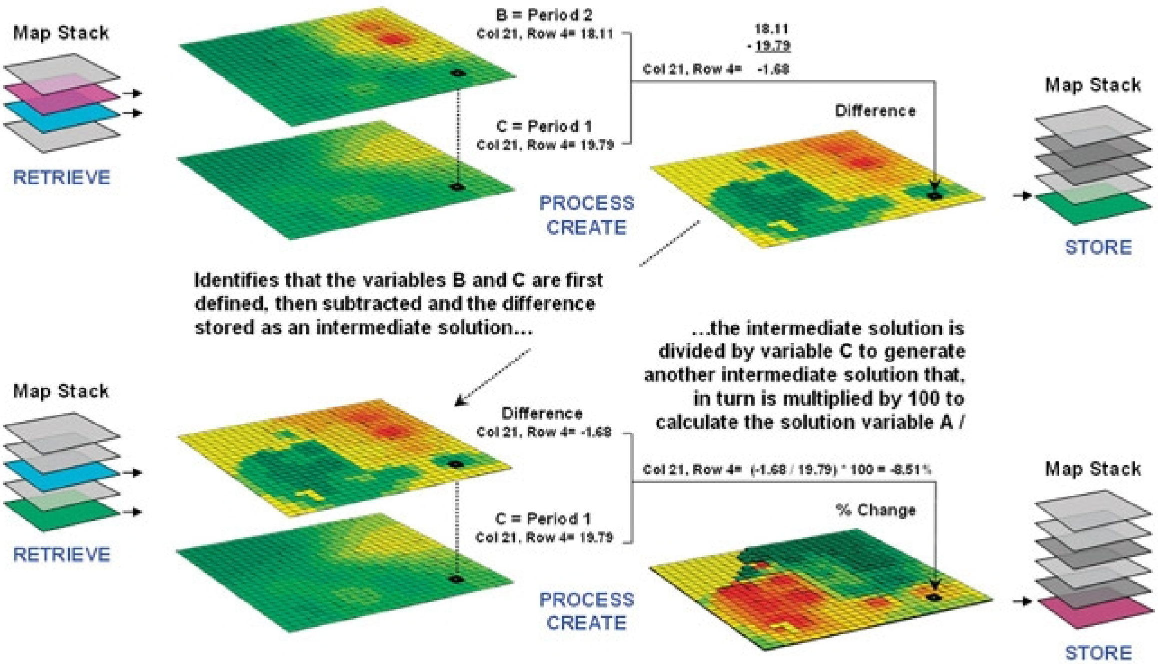
TABLE OF CONTENTS

CARTOGRAPHIC MODELLING

1 Introduction

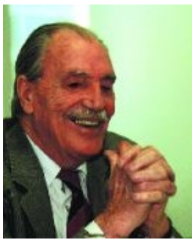
This section looks at Cartographic Modelling. Fortunately, the concepts are not as frightening as the terminology and the application of cartographic models is delivering benefits to spatial analysts worldwide. Why? Because cartographic modelling offers a rigourous procedure for integrating map layers and designing analysis schema for spatial data (Figure 1). This section will require you to demonstrate some straightforward logical reasoning – and a basic understanding of mathematics – as we work through the basic methods of cartographic modelling.

Figure 1. Example of applying a cartographic model to calculate the percentage change in value between two dates.

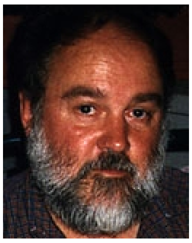


Source: Berry (2004)

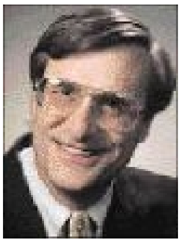
It is difficult to be precise about the origin of cartographic modelling because it is derived from a collection of old ideas, such as sieve mapping and Ian McHarg's *Design with Nature* (1969), that have been organised, augmented and expressed in terms amenable to digital data processing. It is, however, the work described in Tomlin (1983) as 'Map Algebra' and Berry (1987) as 'Map-ematics' which established cartographic modelling as an accepted methodology for the processing of spatial data. Furthermore, by expanding map algebra as a high-level 'natural' computational language cartographic modelling researchers have greatly influenced the GIS and geocomputation communities by, for example, adding time, dynamic processing and the third dimension to the traditionally static, two-dimensional spatial data analysis. As you develop your interest in the topic there is a bibliography in the resources which provides reference to papers on cartographic modelling.



Ian McHarg (1920-2001)



Dana Tomlin



Joseph Berry

2 What is cartographic modelling?

Cartographic modelling, at its simplest, is a generic way of expressing and organising the methods by which spatial variables, and spatial operations, are selected and used to develop an analytical solution with a GIS. Cartographic modelling is based on the concept of data layers, operations, and procedures. The purpose of the method is to create new map layers using existing map layers and operations that are sequenced in procedures. Tomlin (1991) states:

> "*The fundamental conventions of cartographic modelling are not those of any particular GIS. On the contrary, they are generalized conventions intended to relate to as many systems as possible*."

The truth in this statement is illustrated by the number of GIS software products which use the concepts of cartographic modelling in their approach to spatial analysis. For example, in Idrisi you can develop algebraic equations using layers as the variables and spatial operations as the constructs. Tomlin's Map Analysis Package (MAP) was first released in the 1980s and since then the influence of cartographic modelling has extended beyond the university research community and is now integral to public domain and commercial software packages. Cartographic modelling capabilities are found in, for example, ARC/INFO Grid, ArcGIS Spatial Analyst, ERDAS Imagine, GeoMedia Grid / MFWorks, GRASS, Idrisi, PC Raster and the Professional Map Analysis Package (pMAP). In addition, many of these provide toolkits to build models that draw heavily on map algebra, where map algebra is a formal language for analysing spatial data sets. This includes Idrisi, ArcGIS and FME.

A cartographical model is a graphic representation. It uses pictures linked by arrows in a flow chart. Its purpose is to help the analyst organize the necessary procedures as well as identify all the data needed for the study. It also serves as a source of documentation and reference for the analysis. In this way it can support communication between colleagues and contribute to the metadata that accompanies the outputs from spatial analysis.

2.1 Modelling and model inversion

The 'cartographic' dimension to cartographic modelling is, hopefully, self-evident – we are handling spatial data and seek to generate useful information by processing data about individual entities, groups of entities and / or complete layers. Before we proceed we should also have a clear understanding of what is meant by the 'modelling' in cartographic modelling. Modelling, as we shall see, is all about design.

In simple terms, a model is a simplification of reality. This can include simplified representations of the physical landscape, as in a topographic map is a representation of the natural and built environments, and a map of demographics representing people. Equally a model can represent processes and transformations that take place. For example, we can build a model to represent increasing urbanisation and the depopulation of rural areas through the migration of people and the relocation of services. Or we can model the flow of energy and matter through a river catchment by representing the processes of rainfall interception, evaporation, infiltration, subsurface throughflow and channel flow. In building a model we think in terms of identifying the factors that influence the subject of our study. Discharge from a catchment might be the product of the amount of rainfall, the shape and relief of the catchment, vegetation cover and soils, and the presence of barriers to regulate water flow such as a lake. This process of building a model is known as 'forward modelling'. It starts with a conceptual model and then we can attempt to quantify the effects of each factor. For example, through observation we might discover that there is a linear relationship between rainfall and discharge – as we double or triple the rainfall so we see a corresponding increase in discharge.

The really useful thing about models is that we can use them to predict an unknown – either the value of something that we can't measure directly or the future state or condition of the phenomena of interest. Wouldn't it be great if we knew what the weather would be like next weekend or in France next summer so that we could choose the best two weeks for a holiday. If we built a model that shows how different factors – local and global - influence weather patterns then we could predict what is likely to happen and also what might happen if there were changes in one or more factor – the classic 'What if?' simulations.

An alternative use of models is to strive to achieve a predetermined objective. In forward modelling we establish the relationship between an input factor and the output and can then re-run the model using different input values. In a model where A = B + C we can see that if we increase the value of B by 2 then the new value of A (denoted A') will increase by 2 i.e. A' = (B + 2) + C. The alternative use is to fix A and then examine how we might vary B and C to achieve that objective. For example, we might seek to maintain a minimum discharge from our river catchment at all times. As we cannot control the flow of water directly we could do this by making changes to one or more of the factors that influence discharge. Some things we cannot change – rainfall, catchment size and relief – but others we can. For example, we could test what changes in vegetation cover would have to be made to maintain a more uniform flow, or investigate the possibility of increasing the size of a lake e.g. by dredging or building a dam. If we were to introduce additional criteria, such as cost, we could attempt to identify an optimal solution based on multiple criteria. By focusing our analysis on changes to the input factors we have inverted the model and these type of applications are known as 'model inversions'.

We have learned that modelling can support 'simulations' and 'optimisations' and here lies the real strength of cartographic modelling as a method of analysis. Of course, before we get carried away you must remember that the accuracy of your predictions will depend greatly on the quality of your original model of reality and the quality of the data that you use to populate the model.

2.2 Modelling with spatial data layers

We will be using cartographic models extensively in GI analysis. Building a model is worthwhile because it allows you to identify and correct mistakes or change parameters before you start processing data. You should develop a habit of using cartographic models in your own work.

In developing a cartographic model, we find it most useful to begin with the final product and proceed backwards in a step by step manner toward the existing data. This process guards against the tendency to let the available data shape the final product. The procedure begins with the definition of the final product. What information is required? What values will the product have? What will those values represent? We then ask what data are necessary to produce the final product and define each of these data inputs and how they might be measured or derived. The following example illustrates the process:

- Suppose we wish to produce a final product that shows those areas with slopes greater than 20 degrees. What data are necessary to produce such a map?
- To produce a map of slopes greater than 20 degrees, we will first need a map of all slopes. Is a map of all slopes present in our database? If not, we take one step further back and ask another question: what data are necessary to produce a map of all slopes?
- An elevation layer can be used to create a slope map. Does an elevation layer exist in our database? If not, what data are necessary to derive it? The process continues until we arrive at existing (raw) data. The existing data may already be in digital form, or may be in the form of paper maps or tables that will need to be digitized. If the necessary data are not available, you may need to develop a way to use other data layers or combinations of data layers as substitutes.

Once you have the cartographic model worked out, you may then proceed to invert the model, run the inverted model and generate the output data layers.

2.3 Map algebra: modelling conventions

Cartographic model diagrams should adhere to some conventions in terms of symbology. Why? Because a common set of symbols and a standard style of presentation make it quicker to build a model and easier for others to comprehend your model. Just as we use Structured Query Language (SQL) as the language of databases so we construct cartographic models using map algebra. As in algebra[1] we place the final output on the left side of the model, input data on the right and the intermediate data (if necessary) and command elements in the middle. For example, A = B + C. The difference with map algebra is that our inputs are spatial data and so we can apply spatial operations as well

1 Algebra is the branch of mathematics that uses letters and other symbols to represent numbers and quantities in formulae and equations. Modern algebra deals with systems that consist of elements (numbers or geometries) and operations that may be performed with them (such as addition, multiplication or composite transformations). Emphasis is placed on the form, the 'design', of different systems.

as the normal range of algorithms.

Once we have built our model then we need to invert the model before we apply it. The computations are made for each independent element according to mathematical convention. Let us consider a model with one dependent term, A, and 3 independent terms, B, C and D:

A = B + (C * D)

The inverted model becomes:

(C * D) + B = A

and we make calculations in the following order:

multiply C and D to generate the first element

add the first element to B (the second element)

It is important to respect the priorities given by use of brackets and the fact that some arithmetic functions might not be commutative e.g. A - B does not equal B - A.

What are some of the conventions in cartographic modelling? Each different data file type is represented by a different shape (Figure 2). Raster layers are represented by rectangles, vector layers by triangles, values files by ovals, and tabular data by a page with the corner turned down. File names are written inside the symbol. Operations are shown as parallellograms and linked to input and output data layers with arrows. When an operation requires the input of two layers, the arrows from those two layers are joined, with a single arrow pointing to the operation (Figure 3).

Figure 2. Data file types.

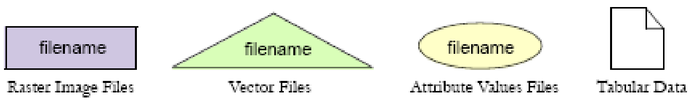
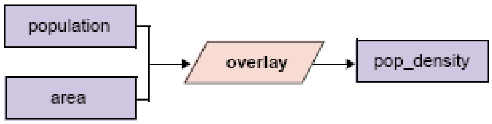


Figure 3 shows a model in which two raster layers, administrative areas and population, are used with an overlay operation and a division calculation to produce a new raster layer of population density. The original model is that population density is a function of population and area. Figure 3 illustrates the inverted model that is used to make the calculation.

Figure 3 A cartographic model to generate population density.



How would the graphics differ if the input and output layers were in vector format?

e-Tutorial Exercise 1

Time: 15 minutes

Construct a cartographic model of the slopes example in sub-section 2.1. The objective is to map 'steep' slopes and you can assume that an elevation map exists in the database.

Hint: first construct a model of the relationship between steep slopes, all slopes and elevation and then invert the model to show how you would make the calculations.

An answer is given in the Answers to e-Tutorials.

In the same way as algebraic operations can be combined to form a complex equation, cartographic models can be constructed to model complex spatial relationships. But before we consider these procedures we will look at how natural language is used in developing a cartographic model.

2.3 Natural language

A logical extension to the use of algebraic notation is the use of natural language. Whist not strictly necessary, the use of natural language opens up cartographic modelling to non-mathematicians.

> "*If a user can express in words the actions that he wishes to perform on the geographical data, why should s/he not be able to express that action in similar terms to the computer?*"

Burrough (1986)

Tomlin (1983) when developing the analogy between cartographic modelling and algebra, recognised the role for natural language to express the logic of spatial analysis. His solution was a GIS with a natural language interface: The Map Analysis Package. In this package each spatial operation is a verb which acts on one or more subject noun(s), each representing a map layer, to create object nouns, or new map layers. Referring to the example in Figure 3, a population map (subject noun 1) is overlaid (verb) on a map of administrative areas (subject noun 2) to create a population density map (object noun). Table 1 lists some verbs of spatial operations used in Tomlin's approach. Note that, unfortunately, not everyone adopts the same terminology and you should be alert to differences between analysts. These inconsistencies and ambiguities are also, regrettably, prevalent in different GIS software.

Table 1. Examples of natural language verbs.

Operation	***Verb***	***Description***
Make a corridor from a linear data set	SPREAD	Renumber all loci with a value reflecting their distance from a given starting point or line
Intersect two polygon networks	OVERLAY	Lay two polygon networks over each other and produce new polygon net
Select according to a condition	EXTRACT	Select specified values and / or ranges of values from one layer to make a new layer

3 Cartographic models and GIS

Once you have created and inverted a cartographic model there is a four stage procedure that leads to the application of that model in a GIS (after Burrough, 1986) :

3.1 Implementing a cartographic model

1. Identify the map layers or spatial data sets which are required.
2. Use logic and natural language to develop the process of moving from the available data to a solution.
3. Set up a flow chart with steps to graphically represent the above process. In the context of map algebra this flow chart represents a series of equations you must solve in order to produce the solution.
4. Annotate this flow chart with the commands necessary to perform these operations within the GIS you are using.

To explore these stages let us consider a supermarket siting example. We can complete stage one of the cartographic modelling process by identifying four data layers:-

land_use

site_status

river_map

roads_map

Stage two is completed by describing, in natural language, a scheme of spatial operations required to identify potential sites for the supermarket. Figure 4 shows how stage three is completed by forming a flow chart to represent the logic in a GIS project. It is sometimes easier to visualise this with thumbnails of the data layers (Figure 5). Note how the appropriate verb (describing a spatial operation) has been added from Table 1. The equation numbers, Eq1, Eq2,... Eq8, relate to equations in Table 2.

Table 2 presents four of the equations it would be necessary to solve as part of the process of finding a suitable site for the supermarket.

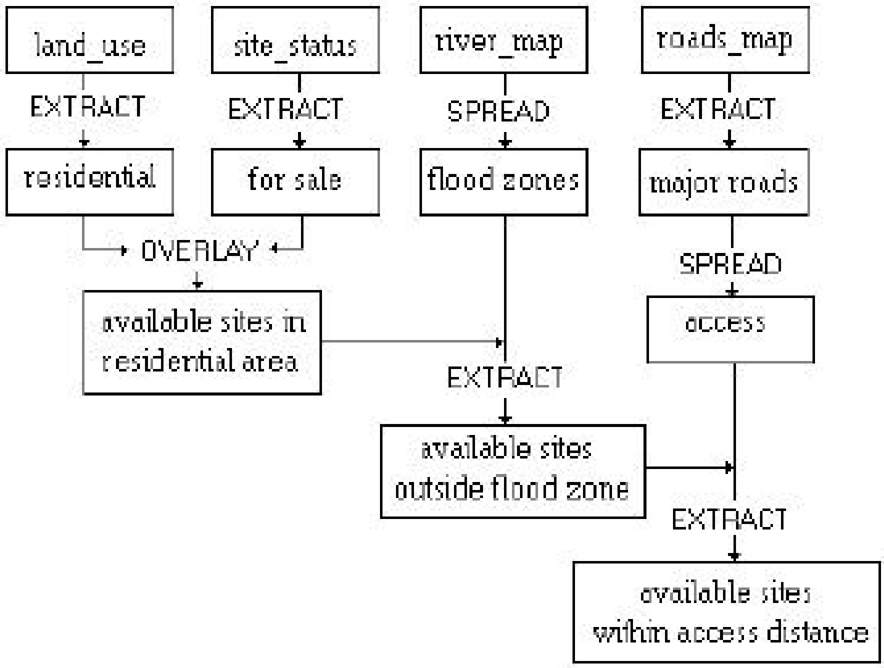


Figure 4. Flowchart of the operations needed to create a map identifying suitable locations for a supermarket.

Figure 5. Flow chart with thumbnails.

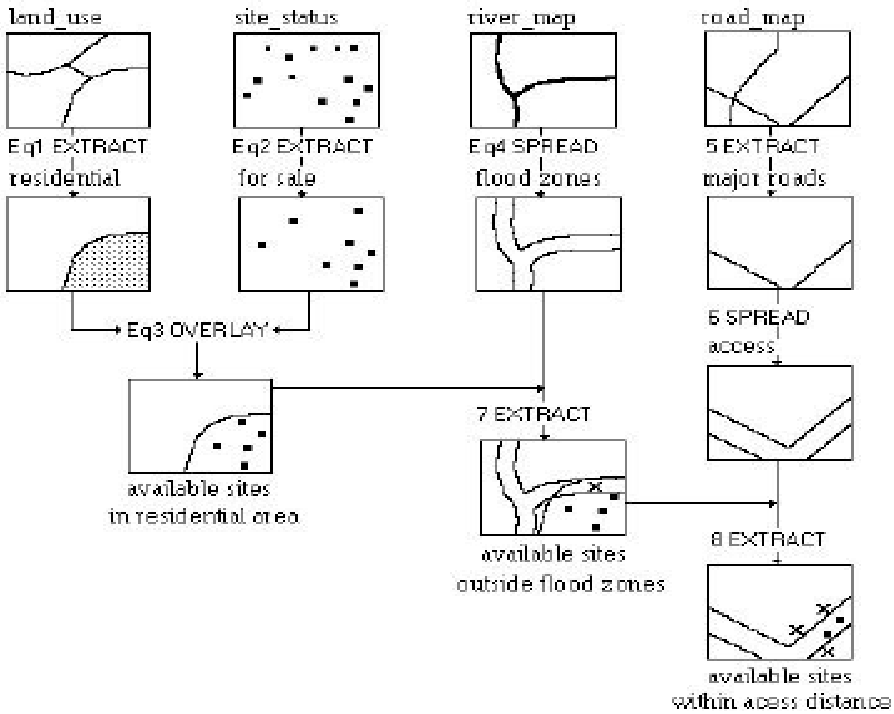


Table 2. Algebraic equations from Figure 5.

From LAND_USE 'extract' RESIDENTIAL	
Eq 1 a – b = c	
where:	a = land_use map
	b = non residential zone
	c = residential
From SITE_STATUS ' extract' FOR_SALE	
Eq 2 d – e = f	
where:	d = site_status map
	e = sites not for sale
	f = sites for sale
'Overlay' RESIDENTIAL and FOR_SALE	
Eq 3 c * f = g	
where:	g = residential sites for sale

From the above exercise it should be apparent that the analytical power of cartographic modelling lies in the ability to combine a series of equations by using the results obtained from one equation as the input for the next. In this way a complex spatial problem can be tackled by breaking down the individual components of the problem into a series of

> **e-Tutorial Exercise 2**
>
> Time: 30 minutes
>
> Referring to Figure 3 and Table 2 complete the missing equations for steps 4, 5, 6, 7 and 8 in the modelling process. Answers are given in the Answers to e-Tutorials.

smaller solvable equations.

The final stage in the modelling process is to annotate the flow chart with the appropriate commands from the GIS package in which it is intended to perform the analysis. In our example these commands are represented by verbs from Tomlin's Map Analysis Package (EXTRACT, OVERLAY and SPREAD). As was mentioned previously, however, different software applications translate these generic statements into their own terminology. Whilst making the running of training courses more profitable this localised terminology does nothing for spatial analysis and makes learning a new package more time consuming. These are trivial issues, however, compared to the benefit to be gained from implementing cartographic models.

> **e-Tutorial Exercise 3**
>
> Time: 30 minutes
>
> Annotate the flow chart in Figure 2 with the appropriate commands from a GIS software package you are familiar with. For this exercise you should assume the available sites are represented as area entities.

3.2 Conditions for applying cartographic models

There are two fundamental conditions required by any spatial analysis package: a consistent data structure and an iterative processing environment.

It has been a dream of mine for some time that I, as a data analyst, could forget about data structures, that I would perform spatial analysis without being concerned about whether a data layer was stored in a raster or vector format. Whilst there are many tools to convert between raster and vector representations it is still not possible to combine raster and vector data layers in a seamless analytical framework. This is because rasterisation and vectorisation are not clean, they might add something or change something in a data layer, and the effects of conversion are unpredictable. So, I am still required to prepare data in a consistent data structure. In general, there are more spatial operators coded to process raster data than vector data and so the raster data model may be preferred by data analysts.

The second condition, the iterative processing environment, logically sequences map analysis operations and involves the following:

1. Retrieving one or more map layer from a database,
2. Processing the data as specified by users,
3. Creating a new map containing the processing results,
4. Storing the new map for subsequent processing.

It is important that each new map is automatically georegistered to the other maps in the database. The output from one operation can then form the input to a later stage of processing. Such cyclical processing provides a flexible structure similar to "evaluating nested parentheticals" in traditional math. Within this structure, you first define the values for each variable and then solve the equation by performing the mathematical operations on the values in the order prescribed by the equation. For example, the equation for calculating percentage change in population starts with population at time 0 (X) and time 0+1 (Y). The difference is calculated by subtraction and then stored as an intermediate solution. The intermediate solution is divided by the initial population (X) to generate another intermediate solution that, in turn, is multiplied by 100 to calculate the solution (A) - the percentage change value.

Input values: X, Y

Intermediate solutions

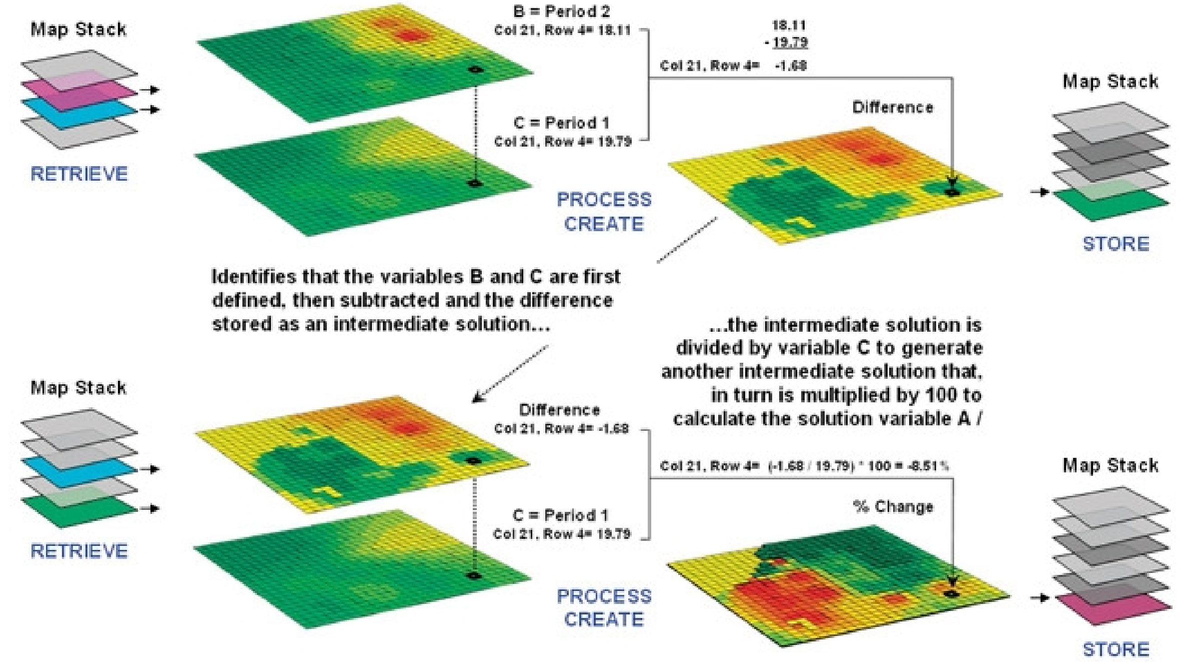
Absolute change in value: $XY_{diff} = Y – X$

Proportionate change in value: $XY_{proportion} = XY_{diff} / X$

Solution: $A = XY_{proportion} \times 100\%$

The same mathematical structure provides the framework for computer-assisted map analysis. The only difference is that the variables are represented by mapped data composed of thousands of georeferenced values. Figure 6 shows a similar solution for calculating the pattern of percentage change in animal activitybut the calculations are performed for each grid cell in the study area. The result is a map that identifies the percent change at each location.

Figure 6. An iterative processing environment, analogous to basic math, is used to derive new map variables (red tones indicate decreased animal activity; green tones indicate increased activity; the example location shows a 8.51% decrease).



Map analysis quantifies the nature of change i.e. magnitude and direction of change in the thematic attribute, and the location of change using the spatial attribute. The characterization of "what" and "where" provides information needed for further environmental analysis, such as determining if areas of large increases in animal activity are correlated with particular cover types or near areas of low human activity.

3.3 Flexible modelling in practice

The mathematical structures and classification schemes form a conceptual framework that's easily adapted to modelling spatial relationships in physical and abstract systems. A major advantage is flexibility. For example, a model for siting a new highway can be developed as a series of processing steps. The analysis may consider economic and social concerns e.g. proximity to high housing density, visual exposure to houses, as well as purely engineering concerns e.g. steep slopes, water bodies. Combining physical and socio-economic concerns as part of an integrated spatial solution is another significant benefit. Furthermore, the ability to simulate various scenarios e.g. steepness is twice as important as visual exposure, and proximity to housing is four times more important than all other considerations, provides an opportunity to embed geospatial information into the decision-making process. By noting how often and where the proposed route changes as successive runs are made under varying assumptions, information on the unique sensitivity to siting a highway in a particular locale is described.

Compare this to a non-model based planning process. In the old environment, decision makers attempted to interpret results bounded by vague assumptions and system expressions of a specialist. Cartographic modelling, however, engages decision makers in an analytic process, because it documents the thought process and encourages interaction. It is the equivalent of a "spatial spreadsheet" that encapsulates the spatial reasoning of a problem and solves it using digital map variables.

4 A final word

What we hope you will have discovered from this section is that cartographic modelling provides a structured approach to GIS design. At present the biggest drawback to cartographic modelling is the lack of standards in the use of algebraic construct and natural language terminology between GIS. Therefore, developing a cartographic model in Idrisi, for example, requires a different set of statements in a different order to those necessary to perform the same operation in another raster GIS. You could probably perform a similar analysis using a vector based GIS, but it might take many more steps and require a different set of commands.

Despite these problems cartographic modelling is probably one of the key areas of GIS which will be refined by future developers. One of the reasons for this is the relentless desire to converge GIS with mainstream IT. It should not go unnoticed that up until now many GIS analysts have done their own thing... and not neccessily adopted industry standard development tools and procedures.

In the past decade, Unified Modelling Language (UML) has emerged as the software blueprint language for analysts, designers, and programmers alike. It is now part of the software trade. The UML gives everyone from business analyst to designer to programmer a common vocabulary to talk about software design. As the Borland software house explains:

> '*UML is applicable to object-oriented problem solving. Anyone interested in learning UML must be familiar with the underlying tenet of object-oriented problem solving - it all begins with the construction of a model. A model is an abstraction of the underlying problem. The domain is the actual world from which the problem comes*'.

At the center of UML are its nine kinds of modelling diagrams and several of these diagrams provide a ready-made framework that can be mapped onto cartographical modelling (Table 3). By learning a new vocabulary GIS developers might find it easier to integrate with other information technologies.

Table 3. Analogies between UML and cartographical modelling.

Diagram		***Relationship to cartographical modelling***
Use case diagrams	describe what a system does.	Identifies objectives
Class diagrams	give an overview of a system by showing its classes and the relationships among them	The conceptual cartographic model
Object diagrams	show instances instead of classes. They are useful for explaining small pieces with complicated relationships	
Sequence diagrams	show how operations are carried out in time	Flow chart
Collaboration diagrams	convey the same information as sequence diagrams, but they focus on object roles	
Statechart diagrams	show the states of the object and the transitions that cause a change in state	Iterative processing environment
Activity diagrams	focus on the flow of activities involved in a single process	Flow chart

Diagram		*Relationship to cartographical modelling*
Component diagrams	are physical analogs of class diagram	
Deployment diagrams	show the physical configurations of software and hardware	Implementation in a GIS

Periodic Self Assessed Exercise 1

Time: 15 minutes

1. Explain the relationship between cartographical modelling and map algebra. Which authors have been responsible for developments in these areas?

2. What is natural language? Why is it important in processing spatial data?

3. What are the four steps required to implement a cartographic model in a GIS software application?

4. Identify three benefits of using cartographical modelling in GIS.

Look back through the text to check your answers.