

Introduction

What Is GIS?

A geographical information system (known as a GIS for short) is a database which allows the user to explore spatial relationships within and between data sets. It generally consists of four basic parts. These are the data itself, the software used to explore the spatial relationships, the hardware used to run the software and store the data, and the users. For most biological research, the only parts which you ever really need to think about are the data and the software as you are the user and you will probably be using a standard laptop or desktop computer to run your GIS software and to store your data.

What software and data you use will depend on what you wish to do. Therefore, it is always good to have a clear aim of what you wish to achieve before you create a GIS project (you are free to change your mind later, but be aware that this might mean making a new GIS if this new aim is not compatible with your original one for some reason). Aims for GIS projects can range from being relatively simple (*I wish to make a map showing my sampling locations*) to being highly complex (*I wish to model the influence of climate change on the distribution of marine organisms*). As a result, different GIS projects may have very different software and data requirements. However, all of them will share the same basic principles and concepts regarding how spatial relationships between different sets of data can be explored. It is these basic principles which this book aims to help you understand and work with.

Throughout this book the term GIS will be used for both a single geographic information system (of the type you would use in your research), and collectively to refer to all geographic information systems. This is because using the more correct term of GISs as the collective term for all geographic information systems is unwieldy and it is confusing to continually switch between the terms GIS and GISs.

Why Is GIS Useful In Biology?

While you might only be interested in learning how to create nice maps and figures for presentations, reports and publications (and there is nothing wrong with that), it is worth remembering that many biology research projects inherently have a spatial component that is worth exploring. This can range

from the distribution of sampling sites to survey tracks, capture locations, movements of individual animals and information about the distribution of specific habitat types. As a result, many biological research projects would benefit from the creation of a GIS to explore spatial relationships within and between the data. In particular, while some projects can be done without using a GIS, many will be greatly enhanced by using it (see www.GISinEcology.com/casestudies for some examples of research projects which have used GIS).

The very act of creating a GIS will make you think about the spatial relationships within your data, and will help you formulate hypotheses to test, or suggest new ones to explore. In addition, thinking about your data in a spatial manner will help you identify potential spatial issues and/or biases within it. For example, plotting the spatial distribution of sampling sites may help you see whether sites which are closer to each other are more similar than those which are further apart. This is something which may not be clear if you only look at your data in a simple spreadsheet or database format. If such patterns exist in your data, and are driven by factors other than those you are studying, this may mean that you have something called spatial auto-correlation which violates the assumptions of many statistical techniques. In this case, you will need to deal with this in some way or other. Similarly, by plotting the locations where a specific species of animal has been recorded over different habitat variables, you can start to develop ideas about which of these variables are important for determining the distribution of that species. While this has always been easy to do for things like land elevation, by plotting the locations on a map, by using a GIS you can look at a wider range of variables and often ones which are not as clearly represented on paper maps (such as the direction that a slope faces).

GIS can also be used to make measurements and do calculations which would otherwise be very difficult. For example, a GIS can be used to work out how much of your study area consists of a specific habitat type, or how much of it is over 500 metres above sea level, or has a slope greater than 5° , and so on. Similarly, a GIS can be used to calculate the size of the home range of an individual animal, or the total area occupied by a specific species, or how long your survey tracks are, or how much survey effort was put into different parts of your study area (see exercise four in chapter fourteen). In addition, you can calculate new variables within a GIS. For example, you can use a GIS to calculate slope and aspect of land from information about its elevation (see exercise three in chapter fifteen). Similarly, you can measure the distances between sampling sites, or the most likely path an animal took from one location to another.

GIS can also be used to link data together in the way that is needed for statistical analysis. For example, many statistical packages require all your data to be in a single table, with one line per sample and information about that

sample, and the location where it came from, in different columns or fields. A GIS provides you with a way to easily create such tables and populate it with information, such as the elevation at each location, the direction the slope faces and the distance from features such as the nearest river or lake, from other data sets. This makes preparing your data for statistical analysis much simpler.

Finally, while GIS is mostly used for displaying and/or analysing data, a GIS can also provide important information when deciding where and how to collect your data in the first place. Given that working in the field can be complex and expensive, it is important to get your data collection right the first time, and a bit of planning can go a long way to ensuring that your research is successful. There is nothing worse than spending all your time and money collecting lots of data only to find out when you come to analyse it that you are missing a vital bit of information or coverage. For example, you may find that you have not sampled all the available habitats, or that you have not sampled a specific part of your study area properly, or that your sampling locations were too close together or too far apart. While creating a GIS at the planning stages of your fieldwork will not banish such possibilities completely, if done correctly, it will help you reduce, to a minimum, the risk of these issues cropping up at a later date.

In reality, these are just a few of the many things which can be done using GIS and which are useful to biologists, and there are many other things it can do, too. Really, the usefulness of GIS to biologists is only limited by their imaginations, but in order to see the full potential of GIS, biologists first need to learn how to use it properly. This is because without a good basic foundation in GIS, it is unlikely that a biologist will ever be able to use it to its full potential.

Who Is This Book Aimed At?

This book is aimed at undergraduates, and other novice biological GIS users, who wish to gain a basic foundation in the use of GIS in biological sciences. As a result, it aims to keep things simple and to specifically look at them in a practical biological context. This means that this book is probably most useful for those biologists who have never used GIS before, or who have dabbled with it, but been put off because it seemed a bit scary and complicated.

Almost all the currently available training information for using GIS is, to a greater or lesser extent, designed and written for terrestrial geographers. However, the GIS concept has such a wide scope that no two users will apply it in exactly the same way. As a result, learning how to use a GIS in biological research from information written for terrestrial geographers is like trying to

learn how to fly a helicopter by reading a training manual for a jumbo jet. Sure, they both fly in a manner consistent with the laws of aerodynamics, but you are unlikely to succeed. This means that in the past, many biologists have been put off using GIS in their research from the very start.

However, a GIS is, in fact, a relatively simple, but hugely powerful, tool. It is simply a matter of learning how to use it to answer the questions you are interested in, and that is unlikely to be achieved by doing tutorials which tell you how to work out the best position for a shopping centre from census information and a road map. Therefore, as well as background information on GIS, this book provides six practical exercises which can be worked through either independently or in the classroom. These practical exercises cover specific practical aspects of using GIS in biological research, and taken together, they provide enough experience with GIS to get the novice biological GIS user up and running. Two sets of instructions are provided for each practical exercise, one set based on the leading commercial GIS software package (ESRI's ArcGIS[®] software) and one set based around the leading open source, and so freely available, GIS software package (QGIS or Quantum GIS – available from www.qgis.org). This means that these exercises can be done by anyone interested in learning about using GIS in biological research, regardless of what GIS software resources are available to them.

The Basics Of GIS

The basic concept of GIS is very simple. A Geographic Information System (GIS) is, at heart, a type of database. However, it differs from most normal databases in one key way. Normal databases primarily consist of a series of tables which can be linked together to allow the data within them to be extracted, compared or manipulated based on values in different fields or columns in the individual tables. For example, in a typical type of data set used in biology, you can use a normal database to extract all records of a particular species or to calculate an average mass of individuals in a sampled population.

However, while normal databases are very good at allowing you to manipulate temporal data (meaning that you can select just those data from a particular time period), they have great difficulty in manipulating data in a spatial context. For example, while you can link data on species occurrence to temperature for the month in which it was recorded (if you have such data available), it is not as easy to link it to temperature data from a similar location in space because a normal database cannot easily work out which temperature data point is closest to the location where a species was recorded.

In contrast, a GIS, as well as consisting of a series of tables, also contains information on the spatial distribution of the data. As a result, in a GIS, it is much easier to link data within the tables based on their spatial relationships. For example, the nearest temperature data point to any point where a species was recorded can be identified and linked to it. This means that a GIS can be used to join data together based on their spatial relationships within and between data sets in ways not possible with other types of databases (figure 1). It is this power to compare and manipulate data based on their spatial relationships which makes GIS such a powerful tool for biological research.

The information on spatial distribution is held in a GIS as a series of data layers (analogous to individual tables within a normal database). These data layers not only have the attributes for each record in them, they also have information which defines the areas of the Earth which they represent, and the size, shape and position of any features within it. Based on this information, GIS software can work out how features in different data layers relate to each other. For example, a GIS could be used to work out the values for different habitat or environmental variables, such as water depth or temperature, where a particular species was seen, as indicated by the arrows in figure 1.

As a result, while tables are the key component of normal databases, data layers are the key component of a GIS. Each data layer will represent a specific

source of data in a specific way. For example, one data layer may contain information on the locations where a particular species was recorded during a survey, which might be represented by a series of points, while another may contain information about the route of the survey itself, represented as a line. Others might then contain information about the environment, such as water depth, sea surface temperature and seabed sediment type represented in a variety of other ways (figure 1).

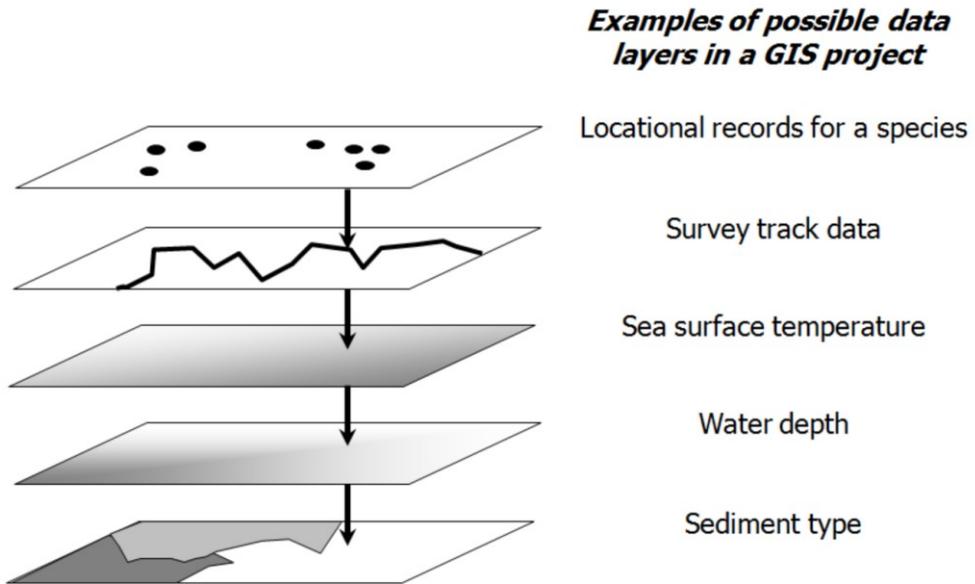


Figure 1. An illustration of how the real world is represented in a GIS project as a series of data layers, each of which contains information about a specific real world characteristic. This could include information about the environment, such as water depth, information about the distribution of organisms in it, such as locational records for a species, or information about human activities, such as the track of a survey through the local area. Due to the way that the data layers are ‘stacked’ or overlaid on top of each other, the information in different layers can be joined together based on their spatial relationships (as indicated by the black arrows) and not just based on common values in a shared field of a table as is the case in normal databases.

By adding data layers that contain the specific information you are interested in, to your GIS, you can start investigating the spatial relationships between them. For example, if you wish to know what water depths different species are found in, you can add one data layer which contains information on the locations where each species were recorded, and another which contains information on water depth. You can then compare the spatial relationships of

these two sources of data to look at which species occur in which water depths.

A GIS is generally created using specialist GIS software, and such software usually provides a series of tools which allow you to not only create, manipulate and edit data layers (see exercises two and three in chapters thirteen and fourteen respectively), but also to investigate the spatial relationships between them in a variety of ways (see exercise four in chapter fifteen). Therefore, the GIS software that you use is a key component of any GIS project. However, different GIS software may contain different tools, and some are better at some tasks than others. As a result, it is important that, where possible, you choose GIS software which is appropriate to your requirements. More information on GIS software can be found in chapter ten.