

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 and store the software/data, and the users. For most marine 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 the 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 GIS for some reason). Aims for GIS projects can range from being relatively simple (*I wish to make a map showing my sampling locations in relation to water depth*) 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 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 Marine 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 marine 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 marine biology 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 with your data. 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 what is important for determining the distribution of that species. While this has always been easy to do for things like water depth, by plotting the locations on a chart, by using a GIS you can look at a wider range of variables and often ones which are not as clearly represented on paper charts (such as the slope of the seabed).

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 under 10m deep, 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 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. In addition, you can calculate new variables within a GIS. For example, you can use a GIS to calculate slope and aspect of the seabed from water depth information. Similarly, you can estimate where fronts are by analysing how similar the water temperatures are at one location to those that surround it.

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 then 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 water depth at each location, the slope of the seabed and the direction it faces, 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 provide important information when deciding where and how to collect your data in the first place. Given that working in the marine environment is often complex and expensive, it is important to get your data collection right the first time. 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 the right range of water depths, or that you have not sampled a specific part of your study area properly, or that you missed out a location which has a specific combination of different variables, or that your sampling locations were too close together or too far apart. While creating a GIS at the planning stage 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.

Who Is This Book Aimed At?

This book is written by a marine biologist for other marine biologists who wish to learn enough about GIS to start using it in their research. As a result, it aims to keep things simple and to specifically look at them in a practical marine biological context. This means that this book is probably most useful for those marine biologists who have never used GIS before, or who have dabbled with it but been put off because it looked so scary and complicated. However, it also contains information which will be of use to the more experienced user, particularly in section two, which consists of a *'How To...'* reference guide for doing specific tasks using GIS.

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 marine 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 many marine biologists are 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. In short, this book is aimed at people like myself when I first started to trying to use GIS within my own research as a graduate student. If I had had this book then, it would have made my life so much easier and would have allowed me to concentrate on doing my research rather than wrestling with how to translate advice for terrestrial geographers into something that a marine biologist could actually use.

What Does This Book Cover And What Does It Not Cover?

This book will introduce you to the concepts behind a GIS, how to apply these to marine biological research, what sort of data can go into a GIS, how to get data into and out of a GIS, how to set up a project in such a way that it will work, and details of how to do a number of common tasks that you may want to do when using GIS in marine biological research. It will also provide some very basic advice about trouble-shooting GIS projects.

It will not cover how to do more specialist tasks, detailed background of how a GIS works or how to write GIS software (there are already plenty of other books which provide this information if you become interested in these areas). In addition, it is not a marine biology textbook and assumes that you are a marine biologist who wants to know how to use GIS in their research rather than someone wanting to learn about marine biology (again there are already plenty of good books out there covering that topic). However, from time to time, it may provide some advice about how to collect data in a way which will make using them in a GIS easier.

Finally, while this book contains examples of how to do tasks using ArcGIS 10.1 software from ESRI (as this is currently the most commonly used GIS software in an academic environment), much of the information and concepts provided here can be applied to GIS projects constructed using other GIS software from ESRI and, indeed, GIS software from other sources. As a result,

this book does not represent a user guide or instruction manual for any specific GIS software, nor is this book endorsed by any specific GIS software or company. Similarly, any advice in this book should not be seen as an endorsement of any specific GIS software or companies. For example, a brief consideration of different GIS software is provided for those who have not decided on which GIS software to use, but the inclusion or exclusion of any specific GIS software in this section does not represent any form of recommendation as to whether a specific GIS software will or will not be suitable for your specific needs.

The TOL Approach Used To Teach GIS In This Book:

Most training courses in life sciences tend to focus on teaching practical skills using a Knowledge-Based Learning (KBL) approach. In KBL, you are provided with generic information or knowledge, and you are left to work out how to apply it to the specific circumstances you might encounter in your daily research. While this allows people from a wide variety of backgrounds to be taught all at once, it can make it difficult to learn how to apply this knowledge to your own specific area of research. To use a motoring analogy, this is akin to teaching someone to drive by teaching them how the throttle and clutch of an internal combustion engine can be used to control the transfer of power to an axle and expecting them to be able to use this knowledge to be able to drive a variety of different vehicles. While this may give a firmer foundation to the knowledge of how driving is related to the functioning of the internal combustion engine, this is not the best approach for learning how to use driving as a practical skill in every day life.

In terms of GIS, this KBL approach means that most training materials tends to describe what individual tools do or to describe individual concepts in GIS in a very generic form, and then rely on you to work out which ones you need to put together to do a specific task in your own area of interest. While you can learn GIS in this way, it requires you to learn an awful lot of stuff, some of which you will never actually need to know. This is not only very slow but can also be very confusing. It also leads to many people having great difficulty in seeing how GIS could be used in their own, every day activities, and therefore to give up on it before they even get started. For example, it can be very hard to see why being able to work out how to calculate the size of the human population in the catchment area for a store is even remotely relevant to marine biological research. However, while you might never guess it until you know a lot about GIS, this is basically the same task as working out how many sightings of a particular species of whale, or other organism, occurred in each grid cell of a particular study area, something which is likely to be of more interest to a marine biologist.

This book takes a different approach. This is called Task Oriented Learning (TOL). In the TOL Approach, detailed instruction sets are provided of how to do specific tasks that are likely to be encountered in a given situation. This allows the user to complete the entire task from a single set of instructions. When similar steps are required to complete other related tasks, this information is presented again, in full, in a separate detailed instruction set. Through this, you become familiar with how the same or similar steps are used in different specific situations, and can infer how these processes can be generalised to other related situations.

In the driving analogy used above, this is akin to teaching someone to drive a variety of different vehicles so that they can use them in their every day lives. Once they know how to drive several different vehicle types, it is much more likely that they can generalise this knowledge to work out how to drive an unfamiliar type of vehicle if they need to. If they wish, or if they later become sufficiently interested, they can also use this knowledge as the basis for exploring the wider generalities of how internal combustion engines work. If not, they can simply just keep driving around in whatever vehicle they might need to use without worrying too much about what exactly is happening under the bonnet.

The TOL Approach involves providing instructions not in the simple linear list format traditionally used to teach practical skills in life sciences. Instead, they are provided in a nested hierarchical format in instruction sets. A flow diagram forms the backbone of each instruction set. This flow diagram details the primarily level of instructions for a specific task and highlights key steps in the process, while an accompanying paragraph of text provides the secondary level of specific instructions required to complete each key step. This not only reduces the risk that you will get lost and accidentally miss a specific instruction causing you to jump from midway through one step to another without completing it properly, it also means that you will find it easier to work out which instructions relate to specific key steps and, therefore, which blocks of instructions can be transferred as working units and combined with others similar blocks of instructions to create instruction sets for related, but unfamiliar, tasks. In addition, this nested hierarchical approach also allows you to combine entire instruction sets for individual tasks by adding an additional ‘summary’ flow diagram to identify which individual instructions sets need to be followed and in what order they need to be carried out in (see chapter twenty for more details). As a result, you can use the basic building blocks taught as part of the TOL approach to build up easy-to-follow instruction sets for what, at first, may seem like very complex tasks.

In terms of this book, taking the TOL approach means that while there are a number of background chapters, the key component is the instruction sets found in section two and which represents a *‘How To...’* reference guide. In this section, which makes up the bulk of the book, you will find instruction sets that outline how to do a range of different tasks in GIS which you are likely to need to do as part of your research in marine biology without getting too bogged down in the theory and generalities behind GIS. Each consists of a complete set of step by step instructions on how to do a specific task from a specific, and clearly defined, starting point to an equally-well defined end point. This allows you to start using GIS in your work as soon as possible, and it is only by using GIS in your research that you will be able to see its benefits and be encouraged to use it on a regular basis.

Once you have started using GIS in your research, you will hopefully start to see how you can customise the instruction sets provided here to do other similar tasks by tweaking the way in which individual steps are carried out. At this stage, you may start to notice that you are using some of the same steps to help complete different tasks. You will then hopefully start to see how you can work out how to do tasks for which instruction sets are not provided here by combining steps from different tasks in new ways. From there, it is only a short step to understanding how to create your own instruction sets and how to join individual instruction sets together to do more complex tasks which you may need to do in your research, as outlined in chapter twenty.

If you can achieve this, you will have fully mastered the use of GIS. If you do not reach this stage, do not worry, you can still use GIS in marine biological research simply by looking up the instructions for what you wish to do in the *'How To...'* section of this book as and when you need them.

How To Use This Book:

This book is divided into three sections. The first section consists of chapters one to ten and provides background information about using GIS in your research. These aim to provide you with a simple introduction to various aspects of using GIS in your research, and are designed to provide you with enough information to get you going rather than give you a detailed understanding of the subjects it covered (this information is readily available elsewhere if you are interested). The second section consists of chapters eleven to twenty and represents a *'How To...'* reference guide with instruction sets for doing specific tasks which are likely to be commonly used by marine biologists. The final section provides additional useful reference material for using GIS in marine biology.

While this book can be read in a sequential manner, it is written in such a way that there is no need for you to read all the chapters or to read them in a sequential manner. This is particularly true of section two, which is set out as a reference guide for doing specific tasks rather than to teach you about GIS. Therefore, by its very nature, it is extremely repetitive as the same steps may appear in more than one task. This means that it would be very boring to read from start to finish. Instead, it is better to dip in and out of this section as and when you need to find out, or remind yourself, how to do specific tasks.

As a result, if you wish, you can leap straight to chapter six (*Starting A GIS Project*), or chapter fourteen (*How To Add Data To A GIS Project*) to allow you to get started as soon as possible. This is particularly relevant if all you wish to do for the time being is make a map for a presentation, report or publication. However, you may find it useful to at least flick through some of the background chapters before you start (e.g. chapter five - *Types of GIS data layers*, or chapter two – *Common Concepts And Terms In GIS*, and especially chapter four – *The Importance Of Projections, Coordinate Systems And Datums In GIS*) and to read them more fully at some point. In addition, you are not expected to memorise everything. The book will always be there for you to come back to, so settle for remembering where you can find out the information you need rather than trying to remember it all.

Once you are a bit more familiar with using GIS, you will find that you can start taking bits from different parts of the book and bringing them together in new ways to develop your GIS skills further. There is nothing wrong with this, and, in fact, it is the best way to become effective at using a GIS in marine biology research. For those who are less certain of what to do, there are direct instructions on how to do most of the basic tasks that you are likely to be interested in doing. If you wish to start learning how to use GIS in marine biology, but do not have any suitable data to play around with, you can download a dataset from http://www.gisincology.com/Data_For_Supplementary_Workbooks.htm. These data are based on a real data set from a research project looking at the occurrence of cetaceans in northeast Scotland and have been put

together to accompany the exercises outlined in the companion volume to this book called *An Introduction To Using GIS In Marine Biology: Supplementary Workbook One*.

NOTE: As with many things in GIS, there may be more than one way to do the tasks outlined in this book. However, the instructions presented here should work under most circumstances and usually allow the less experienced user to get going in the simplest possible way. While every effort has been made to ensure that these instructions are complete and error-free, as with all technical books, some errors may have slipped through undetected. Whenever I become aware of any such issues, I will post corrections on http://www.gisincology.com/corrections_for_books.htm rather than waiting to correct them in the next edition of this book.

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 marine biology, you can use a normal database to extract all records of a particular species or to calculate an average length of a fish species from a set of trawls.

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 water 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 the 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. As a result, a GIS can be used to join data together based on their spatial relationships within and between data sets in ways not possible within other types of databases (figure 1). It is this power to compare and manipulate data based on their spatial relationships which makes GIS such powerful tools for marine biological research.

The information on spatial distribution is held in the 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, but 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, specialist 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 represented 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 location environment, such as water depth, sea temperature and seabed sediment type represented in a variety of other ways (figure 1).

Examples of possible data layers in a GIS project

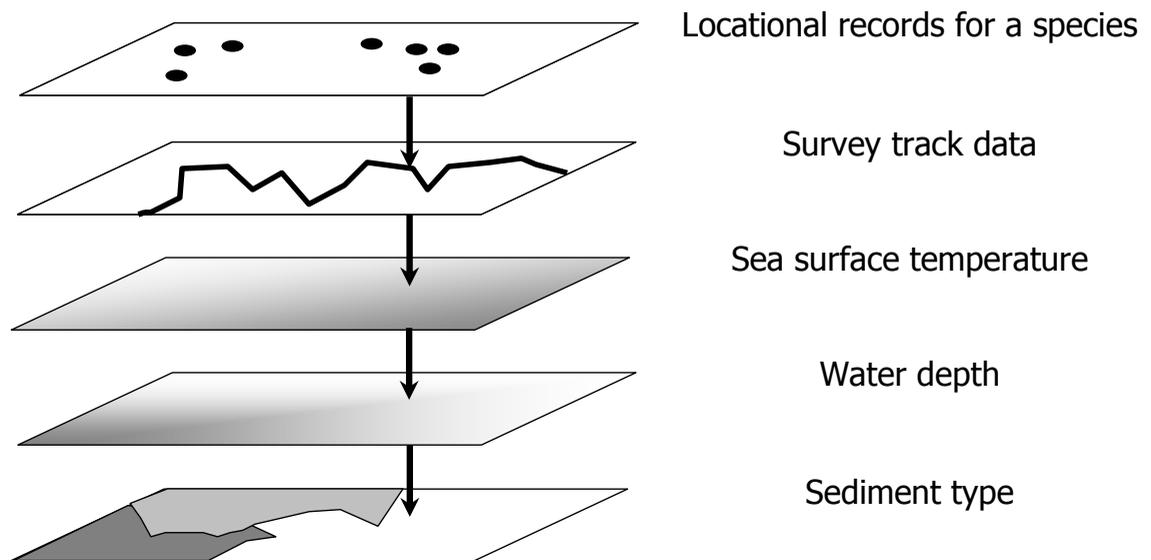


Figure 1. An illustration of how the real world is represented in a GIS project as a series of data layers, each of which contains information about a specific real world characteristic. This could include information about the environment, such as water depth, information about the distribution of organisms 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, 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, but also to investigate the spatial relationships between them in a variety of ways. 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.