The Object-oriented Approach to Geographic Data Modeling for Spatial Data Integration

By

Xiuxia Liu

A thesis submitted to
The Faculty of Graduate Studies and Research
In partial fulfillment of
The requirements for the degree of

Master of Arts

Department of Geography and Environmental Studies
Carleton University
Ottawa, Ontario
August, 2005

Your file Votre référence
ISBN: 0-494-10056-7
Our file Notre référence
ISBN: 0-494-10056-7

ABSTRACT

This research examines the strategies and mechanisms of the object-oriented approach to modeling geographic reality for integrating spatial data. A multidisciplinary and multinational research team, led by the Geomatics and Cartographic Research Center at Carleton University, is devoted to creating innovative approaches for combining geospatially referenced data. An object-oriented approach has been adopted to enable spatial information to be managed, interpreted, manipulated, and understood in a more efficient and more understandable way.

In order to explore and test the strengths and weaknesses of the new approach, a case study on modeling the impacts of climate variability on southern ocean ecological dynamics was developed. Diverse data were obtained from a variety of governmental agencies and academic organizations.

This thesis discusses the potential application of the object-oriented approach to geographic data modeling for spatial data integration.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. D. R. Fraser Taylor, my supervisor, for his support, guidance, encouragement, advice, and patience throughout my study.

This thesis has also benefited from the advice and guidance from many other individuals. Particularly, I would like to thank Dr. Scott Mitchell at Carleton University, Dr. Daniel Vergani from Centro Nacional Patagónico (Argentina), Peter L. Pulsifer at Carleton University, and Dr. Doug King at Carleton University.

The research could not have been accomplished without the support from many source data providers. In particular, I would like to thank Dr. Daniel Vergani, National Snow and Ice Center (NASA, USA), Dr. Jose Xavier (British Antarctic Survey, UK), Dr. Bernie McConnell & Dr. Michael Fedak (Sea Mammal Research Unit of the University of St. Andrews, UK), the Alfred Wegener Institute (Germany), Rob Raskin (NOAA, USA), the National Geophysical Data Center (USA), and the World Data Center.

I am grateful for financial support from Social Sciences and Humanities Research Council of Canada (SSHRC) and for software support from ESRI, UMN, and DM Solutions. I am also thankful to all team members who contribute to the Cybercartography and the New Economy Project.

Special thanks to my husband Xiachen D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ADD – Antarctic Digital Database

AGI – Association for Geographic Information

API – Application Programming Interface

CAD – Computer Assisted Design

CASE – Computer Aided Software Engineering

CNP – Centro Nacional Patagónico

COM – Component Object Model

ENSO – El Nino Southern Oscillation

ESRI – Environmental Systems Research Institute

E-R – Entity-Relationship

GIS – Geographic Information System

GUI – Graphic User Interface

HTML – Hypertext Markup Language

MAGI – Maryland Automated Geographic Information

OGC – Open Geospatial Consortium Inc.

OMG – Object Management Group

SCAR – Scientific Committee on Antarctic Research

SDI – Spatial Data Infrastructure

SST – Sea Surface Temperature

UML – Unified Modeling Language

VRML - The Virtual Reality Modeling Language

W3C – World Wide Web Consortium

WMS – Web Map Service

XMI – XML Metadata Interchange

XML – Extensible Markup Language

CHAPTER 1 INTRODUCTION

1.1 Problem statement, objectives, and scope

This thesis is about data modeling. Although the word *model* is freely used, the basic interpretation of *model* is consistent: "a miniature representation of a thing" (Webster Dictionary, 1913); a theoretical depiction of a problem to "be analyzed by an algorithm" (Jones, 1996); "a simplified representation of reality" that can be used to "simulate a process, understand a situation, predict an outcome, or analyze a problem" (AGI, 2004). A data model is an abstraction of the real world which describes how data are represented in an information system or database management system (Wikipedia Dictionary, 2004). A data model tells an information or database management system how to simulate a process, understand a situation, or analyze a problem through a simplified representation of the reality. It is independent of a computer system and its associated data structures. In terms of the application domains, data models can be classified into office, engineering, bibliographic, image and multimedia, and geographic data models among others (Worboys, 1995).

This thesis considers data modeling in the geographic context. A geographic data model defines the vocabulary for describing and reasoning about things that are located on the earth (Morehouse, 1999). A conventional map is an early geographic data model with projection and symbology as its vocabularies. In the computing era, a geographic data model defines the rules on how to interpret complex geographic reality into predetermined numbers to be stored in a computer and how to organize database records and store spatial and non-spatial data (Goodchild, 1992, 1993; Bonham-Carter, 1994; Heywood et al., 2002). These rules must be understood before creating or translating certain custom data models. The main types of data model include the raster data model, the vector data model, the CAD data model, the relational data model, the object-oriented data model, and the object-relational data model. The first three focus on the spatial data of geographic features, while the last three emphasize the attribute data of geographic features (Bonham-Carter, 1994).

The overall purpose of geographic data modeling is to integrate associated data into a geographic information system (GIS) for carrying out spatial analysis, performing interactive queries, producing maps, and finally representing and simulating real world processes. From this point of view, attribute data, are sometimes more interesting to scientists because they can be used to address temporal components of geographic features. This might be one of the reasons for another name for a geographic data model – a database data model, which has been also seen as the heart of any GIS (Worboys, 1995) and is used to describe how geographic features and processes are represented and simulated in a digital computer. The process of geographic data modeling includes three top-down levels: conceptual data modeling, logical data modeling, and physical data modeling. The conceptual data model is independent of any specific computational paradigm; the logical data model is embedded in a particular computational paradigm (e.g. relational, object-oriented); the physical data model corresponds to implementation and is the ultimate representation of the real-world phenomena as computational processes inside a physical computer.

The relational model in the geographic context, using the relational technology constructed by Codd (1970) for computer science, is currently dominant in GIS. The geographic model in the geographic context follows the basic principle of the relational model – all information is represented by data values in relations. In a geographic relational model, spatial data are held in conjunction with attribute data (Zeiler, 1999): the spatial data are stored in binary files, while the attribute data are stored in tables with a number of rows equal to the number of features in the binary tables and joined by unique feature identification numbers (keys). A major advantage of the geographic relational data model is its flexibility. It is easy for users to customize feature tables either through simply adding fields or linking external database tables. In addition, the model has a sound theoretical base in mathematics.

The well-developed relational technology was primarily designed to handle simple business data. Its application in complex, multi-dimensional spatial data, however, faces a number of problems (Worboys, 1995; Korte, 2000; Zeiler, 1999; Heywood et al., 2002; Twumasi, 2002). One major problem of the relational approach is, like tackling the business data in a bank, it treats only the static-oriented aspect of geographic information (Worboys, 1995). The relational approach causes geographic features to be mapped into homogenous collections of points, lines, and polygons with "generic" behavior. For instance, both a road and a stream are represented as lines. Sometimes, however, it is desirable to support the special behaviors of geographic features: streams flow downhill, but not uphill; the flow of the merged stream is the addition of the two upstream flows. Additionally, this model always produces some data redundancy, which may result in a sluggish implementation. The range of data types of this model is limited. The data model construction process, from conceptual modeling, through logical modeling, to physical modeling leaves us with an important problem – information may be lost in the data modeling process from one model level to another (Twumasi, 2002).

Challenges that have been involved in integrating spatial data exist in many forms. For example, spatial data have become more and more complex with increasing user demands. The level of abstraction, which is the central concept of a model, needs to be improved somehow. The previous modeling methods to structure the real world leave much to be desired. The real world is not made up of cells, or coordinates joined by straight line segments, but of geographic objects with each containing its own attributes and behaviors. In addition, dealing with multiple dimensionalities, an important aspect of the realistic view of the real objects to enhance the understanding of the knowledge domain has been a weakness in the previous models.

The objected-oriented approach, which originally emerged in computer science, has been applied to GIS. The object-oriented approach in the modeling context is sometimes divided into two models: the object-oriented data model and the object-relational model, although they largely share the same idea. The emergence of the object-relational model meets the market requirements that until now have been dominated by the relational approach. The main principles of the object-relational model come from the object-oriented world (Stonebraker & Brown, 1999). The value of object orientation, and the GIS expressions of object-oriented technology (generally including object-oriented programming and object-oriented database management), have been widely discussed in many publications (Egenhofer & Frank, 1992; Goodchild, 1993; Worboys, 1995; Berry, 1996; Korte, 2000; Zeiler, 1999; Heywood et al., 2002). The concept of "object" might be a significant reason to incorporate this approach into GIS. In GIS, the modeling object can be the real world geographic object. The real world can be seen as collections of geographic objects which exhibit a wide range of behavior. Through the modeling process, each object relevant to an application domain is represented by a corresponding object in the data model (Twumasi, 2002). Instead of representing objects as lines and nodes, as with the relational model, the object data model stores objects in logical groups known as "classes". For example (Demartino and Hrnicek, 2001), water main lines and water service lines are objects defined as two unique line classes in an object data model; fittings and hydrants are two distinctive point classes in an object data model. These are not simply generic lines and nodes stored in a layer, defined by attached attributes and sorted by queries. Object-orientation enhances the level of abstraction and makes it closer to our perception of the real world, offering a mechanism for expressing our understanding of knowledge domains. The multi-dimensional aspect of the real world objects can be achieved in the computer world.

Object-orientation not only changes the traditional "computer view" of the real world, but also makes the process of modeling more efficient and lessens the information loss from one level to another. An object data model can be planned and designed using computer-aided software engineering (CASE) tools and the Unified Modeling Language (UML). This development process makes the transition from one step to the next more efficient. The transition of information can be relatively safer since UML, the major visual modeling language, brings different levels of models close to each other.

Because of the performance limitations of computer hardware and database software of the time, the object approach to GIS, as promoted in the late 1980s (Oosterom & Bos, 1989; Egenhofer & Frank, 1989), was not implemented until relatively recently and most commonly at the level of conceptual modeling (Raper & Livingstone, 1995; Borges et al, 1999; Gordillo et al, 1999; Bian, 2000; Gartner et al., 2001; Frihida et al., 2002). More recently, with the development of GIS technology, the object-oriented approach, integrating some advantages of the relational database, has been widely applied into geographic data modeling throughout the whole process (Raza, 2001; Twumasi, 2002;

Carvalho et al, 2004; Blongewicz, 2004).

This thesis addresses the specific question of the need and the mechanisms of applying the object-oriented approach to geographic data modeling for spatial data integration. A case study on the southern elephant seals, introduced in Chapter 1 and detailed in Chapter 3, is used to examine the strengths and weaknesses of this modeling approach. An object-oriented geographic data model, the result of the case study, is developed and provides a template which can be used to assemble, manage, and analyze, and query in GIS. This research has the following goals:

- To discuss the principles, strategies and mechanisms of the object-oriented approach to geographic data modeling. The principles of currently existing geographic data models are also explored.
- To design a custom data model for the southern elephant seals in the Antarctic Peninsula Region using the object-oriented approach within a GIS environment. The object-oriented approach extends the power of marine spatial analyses by incorporating behavior in data. This data model can be used as a common template for assembling and publishing the existing data and additional data collected in the future by other marine scientists.
- To provide a friendly cartographic interface for users. Text and numerical data represent the two main historic media types. Text provides users with the most detailed information in almost any circumstances at almost any time; while

numerical data are more quantitative. But does everybody need (or want) to understand (or predict) the great detail behind a certain phenomenon? Probably not. Graphic visualization of texts and data may represent a middle level of "scientific" knowledge and cater to a broad audience. Although cartographic visualization of data continues to be more and more flexible (for example, one map can be reproduced from another without having to know its projection or scale), geographic data modeling represents a solid approach for transferring geographic data to users. The resulting cartographic interface allows users to manipulate, retrieve, analyze, and present geographically referenced data.

- To promote networking and data sharing through established standards and the Internet within the Antarctic scientific community. Antarctic scientific work is carried out across a wide range of disciplines and is executed through a large number of organizations. The Internet, now a part of everyday life, is an efficient tool for data sharing.

The scope of this thesis is basically within the framework of data modeling in GIS technology, without incorporation of the fundamentals of computing theory. Although GIS has a strong link to computer system modeling, many GIS researchers or practitioners are not computer scientists. Computer software packages, however, facilitate the wide use of data modeling in geographic information processing.

The primary focus of this research is to discuss geographic data modeling. The impact of climate variability in the southern ocean ecosystem as indicated by southern elephant seals is on-going research in itself. This research is not intended to investigate the relationship among climatological, biological, and oceanographic parameters, but to look at the results of scientific research, and to examine and integrate existing geospatially referenced data from different data suppliers using state-of-the-art GIS. The study area is the Antarctic Peninsula. The time scale covers a 15-year period from 1980 to 1994.

1.2 Background

The advent and spread of modern technologies has greatly enhanced human views of the world. Satellites enable human beings to look at the world from space at different scales. Computer science and engineering help people visualize and analyze patterns behind huge volumes of data. Information can be communicated through the Internet in almost real-time. Cartography, as an important discipline dealing with spatial data integration, has been revolutionized by the changes in modern technologies. Data can be easily collected from space using remote sensors; then stored and manipulated in a digital computer and eventually transformed through the Internet. Viewing and analyzing geographically referenced data can be dynamic and interactive.

The research issues around the science of mapping can be considered from two perspectives - the technical and the conceptual. Technical research mainly involves the application of advanced technology in mapping, while conceptual research mainly studies the theoretical issues related to cartography. Some of these can be based on the result of technological applications. In the history of cartography, most research investigates only one or two narrower issues. The books, "Visualization in Modern Cartography" (MacEachren & Taylor, 1994), "Interactive and Animated Cartography" (Peterson, 1995), "Multimedia Cartography" (Cartwright et al., 1999), and "Maps and the Internet" (Peterson, 2003) represent the development of cartography in both the technical aspect and the conceptual aspect during the past decade. More recently, Taylor (1997, 2003) challenged the definition and proposed a new term to capture all aspects of the changes in cartography. In his chapter (Taylor, 2003), Taylor defined the term Cybercartography as "the organization, presentation, analysis, and communication of spatially referenced information on a wide variety of topics of interest and use to society in an interactive, dynamic, multimedia, multi - sensory, and multidisciplinary format". Although it is still in its infancy, the contributions of this theoretical construct to the science of mapping can be foreseen. Perhaps for the first time, psychologists are involved in making maps and work directly with users to analyze and test the user satisfaction needs. Thus, map interface design decisions, are based not only on designers' intuition but also on user needs analysis. In this sense, these maps can be called "users' maps". A detailed discussion of this topic falls outside the scope of this thesis, and a comprehensive description of cybercartography can be found in "The Concept of Cybercartography" (Taylor, 2003).

The Cybercartographic Atlas of Antarctica Project was created in order to analyze, develop, and test the value of the new paradigm of Cybercartography for spatial data integration. The Cybercartographic Atlas of Antarctica Project aims to create an innovative new methodology to integrate, discover, utilize, present, and distribute existing scientific data and information about the Antarctic Region to a wide variety of users (e.g. scientists, decision makers, and the general public) (Pulsifer, 2003). As one of the research topics for the Antarctic scientific community, the impact of climate variability on the southern ocean marine ecological environment as indicated by southern elephant seals is an important part of the atlas content.

The interaction between climate change and the ocean has been an important topic for the scientific community. The southern ocean marine environment is the major physical factor in the southern ocean ecosystem. Research indicates that several of the physical controls on phytoplankton production are sensitive to climate change, although it is presently impossible to make numerical predictions on this (IPCC, 2001). In the southern ocean ecosystem, top predators such as southern elephant seals are generally considered suitable indicators of environmental stress (Vergani et al., 2001; Vergani et al., 2002), because their reproductive success depends mainly on the available amount and distribution of prey. In the past, ecological modeling has had a temporal emphasis, which reflects the changes at one location over time. Two of the major reasons for this were that older spatial models are "notoriously data-hungry" and "the representation of space in the first generations of spatial computer models was primitive" (Wegener, 2000). The advent of GISs and the surge of theoretical work associated with them have increased the possibilities of aggregating spatial data and representing space.

Addressing environmental problems is one of the strongest and most successful application areas for GIS (Goodchild, 1993; Goodchild, 2000). Within the broad area of environmental problem-solving, applications can be categorized as: mapping, data processing, modeling, and policy-making. More recently, in the marine context, Valavanis (2002) addresses the roles the Internet plays in the cross-disciplinary integration efforts to master the complexity of global environmental problems. The diffusion of raw data through online geospatial databases and the existence of online mapping and management tools can make information available to many organizations, increase public environmental awareness, and enable exchange of scientific ideas.

The application of GIS in environmental problem analysis usually begins with geographic data modeling (Mitasova et al., 1996). Geographic data models populated with data in a GIS allow people to perform spatial sensitivity studies, and to assess how changes in key system variables alter the system's dynamic behavior. The essential goals of marine GIS are discussed in detail by Valavanis (2002). Figure 1.1, modified from Valavanis (2002), demonstrates the functions of GIS and the data flow in marine science.

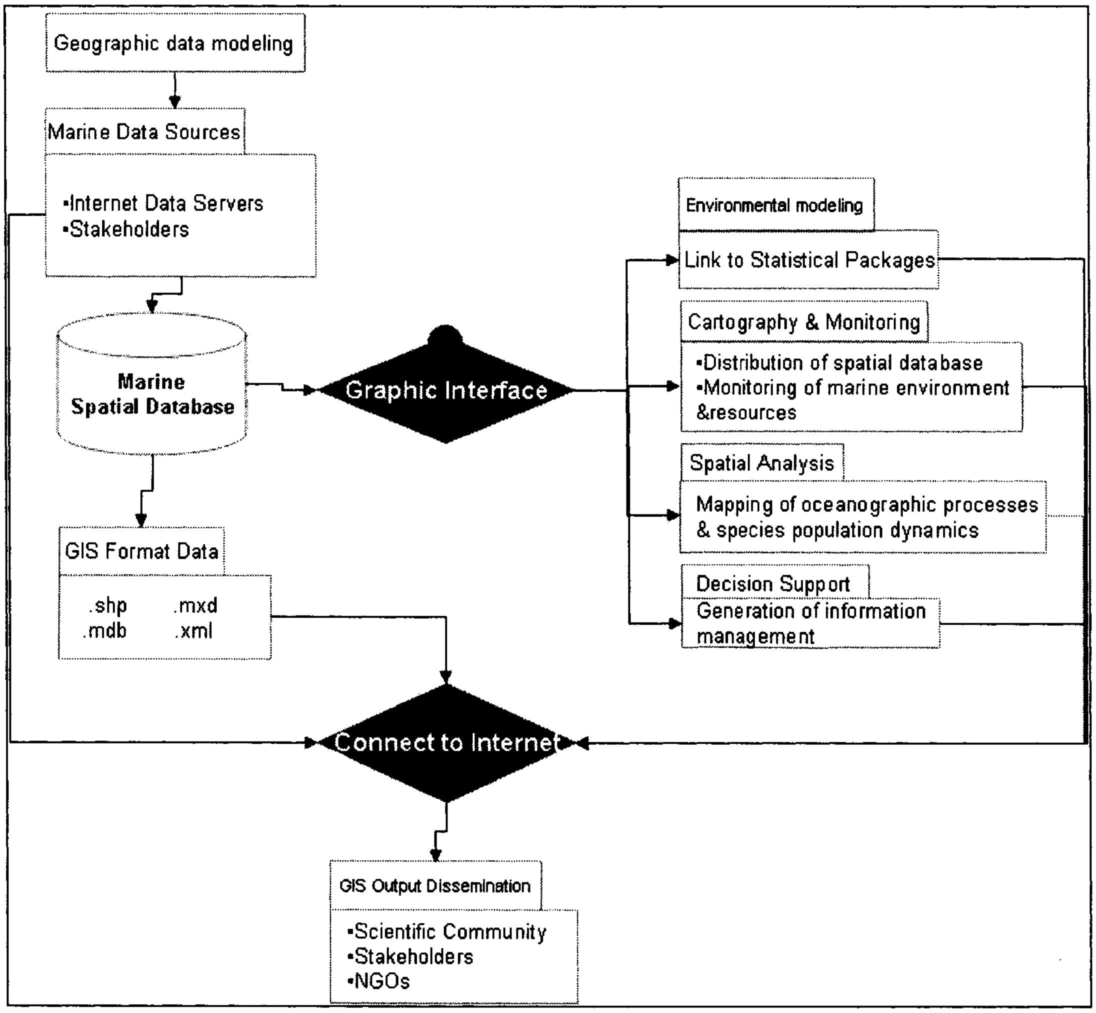


Figure 1.1 Functions of GIS and the data flow in marine science

Geographic data models populated with data in GIS allow a variety of users to perform sensitivity spatial studies, and to assess how changes in key system variables alter the system's dynamic behavior over the Internet (Modified from Valavanis (2002)).

1.3 Research tools

The main tools used in this research include:

ArcGIS-ArcInfo 9.0 (ESRI) – the computing environment for developing and implementing the system

Visio (Microsoft) – providing a friendly environment for designing the model using the unified modeling language (UML)

MapServer (UMN, DM Solutions) – an open source development environment for constructing spatially enabled web applications

Flash (Macromedia) – a powerful computing environment for designing and implementing the dynamic process of spatial information

1.4 Thesis structure

The thesis is organized as follows:

Chapter 1 states the research problem and the central argument of this thesis, introduces the origin of this research, and outlines the major tools to be used.

Chapter 2 explores past and current geographic data models, examines the weaknesses and strengths of those models, and discusses the strategies of the object-oriented geographic data model. The importance of data modeling in GIS applications and the role of GIS in marine environmental problem solving are presented.

Chapter 3 focuses on the analysis, design, and creation of a marine object-oriented GIS data model for representation of southern ocean ecological dynamics as indicated by southern elephant seals. The strategies and techniques of implementation of the object-oriented approach are analyzed. Spatial marine environmental issues relating to southern elephant seals are presented to support the rationale of this research.

Chapter 4 is devoted to testing the characteristics of the object-oriented approach to geographic data modeling and applying the data model to the case study on southern elephant seals. The aspects of the application essentially follow the architecture presented in Figure 1.1, although only certain functions are exhibited. The shortcomings of this approach are also discussed.

Chapter 5 presents the conclusions and recommendations. This chapter summarizes the contribution of the object-oriented approach to geographic data modeling in the field of GIS, and that of this thesis to the area of southern ocean marine environmental problem solving. This chapter also presents some recommendations for future work on the research.

CHAPTER 2 GEOGRAPHIC DATA MODELING IN GIS

2.1 Overview

The real world around us is immensely complex. A common method used by scientists to deal with this complexity is to try to isolate parts of reality either in fact or in theory, and then, to investigate how the parts operate under simplified conditions. This reduction of reality into simplified structures is an entirely subjective product of the mind of the investigator (Chorley & Kennedy, 1971).

In geography, traditionally, the complex geographic space including its geographic features has been represented using three basic entity types (or their combinations) with a spatial referencing system: points, lines, and areas. Points may be used to represent geographic features too small to be represented as areas or lines at the chosen resolution; lines can be used to represent features too narrow to be represented as areas at the chosen resolution, such as glacier lines and sea mammal tracks; areas may be used to represent homogeneous geographic zones such as water bodies and islands. Some three-dimensional surfaces can also be represented through a collection of elevation points or contour lines (Heywood et al., 2002). A network can be created by combining lines and points.

The process of simplifying geographic reality can be thought of as a series of stages of geographic data abstraction (Peuquet, 1984). Data on its own have no meaning, only when interpreted by some kind of data processing system does it take on meaning and become information (Heywood et al., 2002; Wikipedia, 2004). People or computers can find patterns within data to perceive information, which can be used to enhance knowledge. The challenge is the organization and storage of those data, especially in the modern technology era when a myriad of data sources exist but computers cannot be directly applied to handle and display spatial patterns within those data without unambiguous instructions. The conversion of complex geographic reality into finite numbers for computers to store and organize is usually called a geographic data model (Goodchild, 1992, 1993; Bonham-Carter, 1994; Heywood et al., 2002). Geographic data models define the rules to identify features and properties of space. No model can never fully represent the real world but can only be an analogue which has some features and behaviors in common with the real world. In the history of GIS, six types of models have been developed: the raster data model, the vector data model, the CAD data model, the relational data model, the object-oriented data model, and the object-relational data model (Goodchild, 1992, 1993; Bonham-Carter, 1994; Heywood et al., 2002). The first three, which were developed early and also called geographic space models, focus on modeling geometries and locations of geographic entities, while the last three concentrate on modeling the attributes and even behaviors of geographic phenomena. Each of them, except for the CAD model, is examined in different degrees in the following sections. An in-depth review of each model is beyond the scope of this thesis, but can be further extended through the references listed in each section.

2.2 The standard approach

2.2.1 The raster data model

The raster data model represents geographic space by dividing it into a series of units, each of which is limited and defined by an equal amount of earth's surface (Davis, 1996; Heywood et al., 2002) (see Figure 2.1). In the raster world, the basic building block is the individual grid cell, by which the shape and character of an entity is created. Cells are interconnected to create planar surfaces representing all the space of a single area of study. The matrix of cells, organized into rows and columns is called a grid. Geographic features are represented in grid cells or pixels filled with values. In a raster data model, points are represented as a single grid cell (pixel), and the absolute location of the point is somewhere inside the grid cell. Lines are represented as a series of connected grid cells, and each point of the line must occur somewhere within one of the displayed grid cells. In terms of polygons, all points inside the area that is bounded by a closed set of lines must occur within one of the grid cells to be represented as part of the same area.

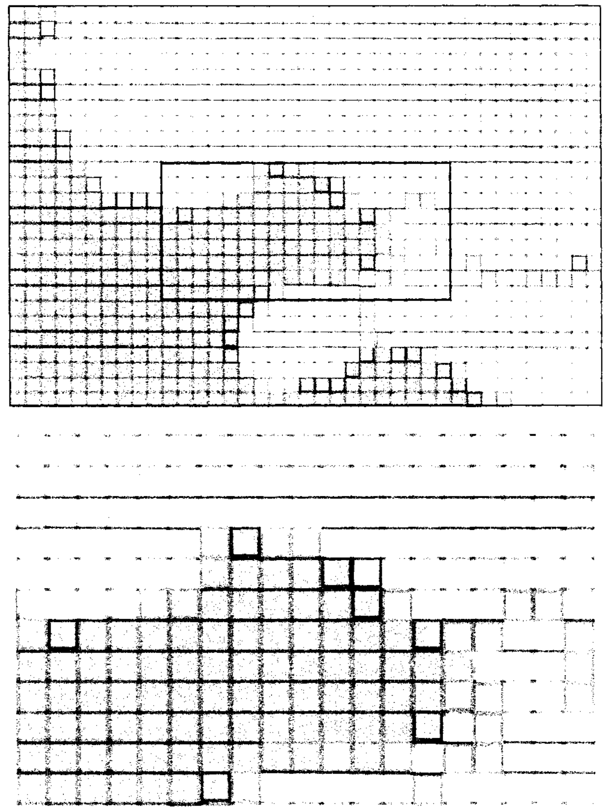


Figure 2.1 Raster view of the Bellingshausen Sea (Antarctica)

The white color represents ice shelf; the blue represents ocean; the yellow represents land. The lower part of the figure is the area in the red rectangle of the upper after being zoomed in twice. As demonstrated in the figure, the raster model represents phenomena by grid cell location in a matrix. The grid cell is the smallest unit of resolution and may vary from centimeters to kilometers depending on the application (by the author).

In comparison with a data model, a data structure deals with more operations on the data in order to store data in a computer efficiently. Often a carefully chosen data structure will allow a more efficient algorithm to be used. A raster data structure, for example, forces a computer to store and reference the individual grid cell values, together with their attributes, coverage names, and legends (DeMers, 1997). Spatial relationships are implicit in the raster model. Raster data is georeferenced by specifying the coordinate system to which a grid is registered, the real-world location of the reference point, and the cell size in real-world distances. The coordinates of the grid cell give the location of a grid cell (see Figure 2.2).

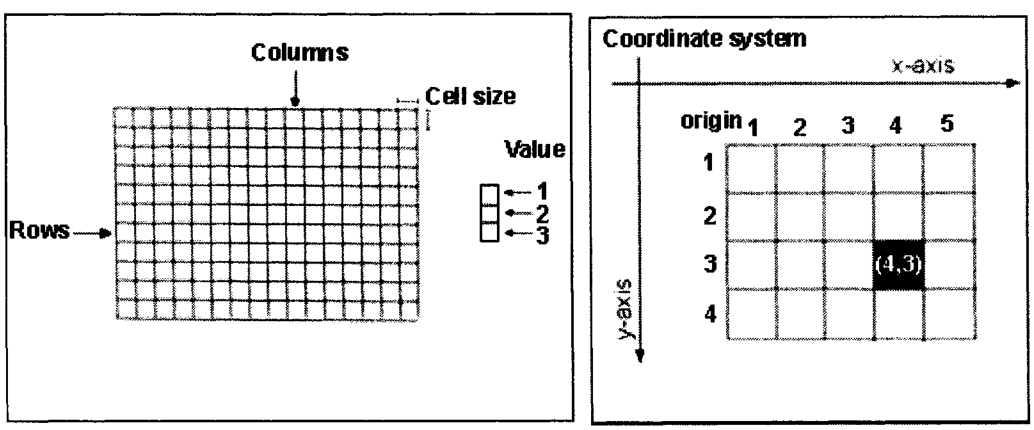


Figure 2.2 The raster model structure

GRID/MAGI, IMGRID, and MAP are models with the raster as their structure (Billah, 2001). In GRID/MAGI data structure, each grid cell is addressed individually and is associated with identically positioned grid cells in all other coverages. This data structure facilitates the multiple coverage analysis for single cells, but it limits the examination of spatial relationships between entire groups or themes in different coverage. In IMGRID, for instance, in order to represent a thematic map of land use that contains four categories – wetland, forest, building, and road, each of these features have to be separated out as an individual layer. Although it causes excessive volume of data stored, there is no limitation of assigning a single attribute value to a single grid cell since each coverage feature is uniquely identified in IMGRID. MAP integrates the two structures discussed previously. The MAP structure allows each thematic coverage to be recorded and accessed separately by map name or title through recording each variable of the coverage's theme as a separate number code or label. This structure may help to perform operations on individual grid cells and groups of similar cells, and the resulting changes in value require rewriting only a single number per mapping unit, simplifying the computations.

In general, the raster model has several advantages including simplicity of concepts and algorithms, easy analysis and modeling, and compatibility with imagery. Spatial data of different types can be overlaid without the need for the complex geometric calculations. Each layer of grid cells, constant in size and generally square, records a separate attribute.

The locations of cells are easily addressed by the row and column numbers (see Figure 2.2), which means spatial coordinates do not have to be explicitly stored for each cell.

Disadvantages of the raster model exist. As Figure 2.3 shows, the primary drawback with raster storage is the prerequisite to store the entire matrix of rows and columns. Storing the entire matrix sometimes may include unwanted data – the pixels surrounding the polygons. Spatial resolution, therefore, may be limited because storage requirements have to increase with increasing resolution. In addition, graphical output may be less pleasing depending on pixel size (compare Figure 2.1 and 2.4); projection transformation is more intricate; the representation of topological relationships is more difficult.

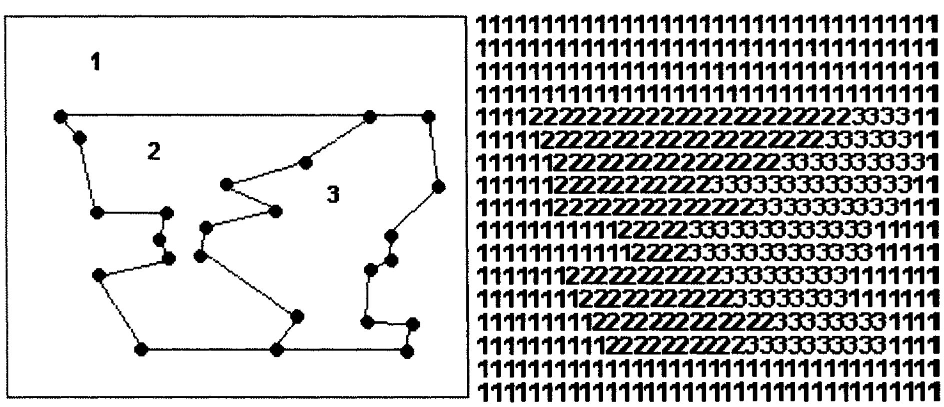


Figure 2.3 The vector data structure (left) and the raster data structure (right)

2.2.2 The vector data model

The vector data model (Figure 2.4) allows the representation of geographic space in a way that may be closer to the familiar analog map (Davis, 1996). It represents spatial location of features explicitly. The shape of a spatial entity in a vector model is stored using two-dimensional Cartesian coordinates: a point is represented as a single XY coordinate pair in the chosen coordinate system; a line is represented as connected series of XY coordinate pairs; a polygon or area is represented as a series of lines with the same beginning and ending point coordinates.

The vector data structure (see Figure 2.3-left) is much more representative of dimensionality as it would appear on a map (DeMers, 1997). Unlike those in the raster structure, geographic entities in the vector structure can be stored without their attributes. Attributes can be stored separately, such as in a database system. Entities and their attributes can be linked at any time information about a map feature is needed.

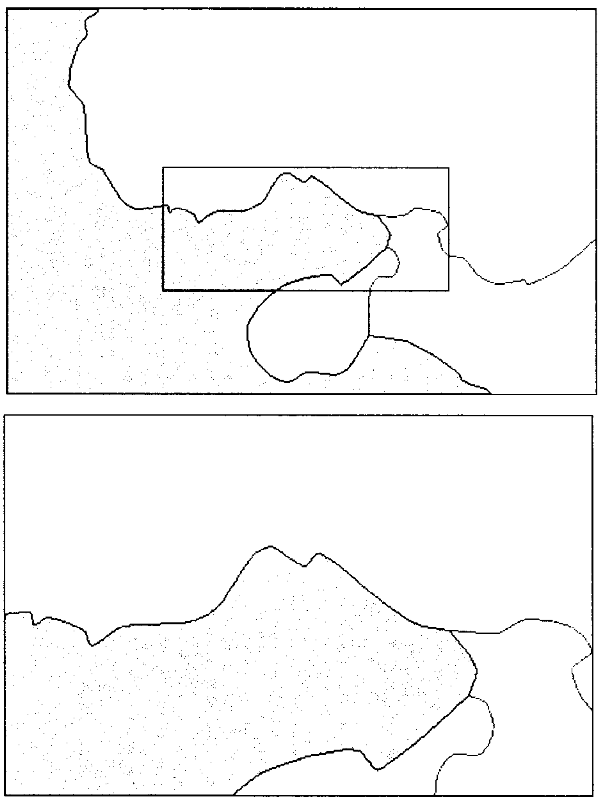


Figure 2.4 The vector view of the Bellingshausen Sea (Antarctica)

The white color represents ice shelf; the blue represents ocean; the yellow represents land. The lower part of the figure is the area in the red rectangle of the upper after being zoomed in twice. As demonstrated on the figure, the vector model represents phenomena by points, lines, and polygons. In comparison with the raster view of the Bellingshausen Sea (Figure 2.1), the graphics out of the vector model remain attractive across scales (by the author).

Two important vector data models are the spaghetti model and the topological model (Bonham-Carter, 1994). The spaghetti model is essentially based on the one-to-one translation from the graphic object on the analog map with a piece of spaghetti, with very short ones for points, longer ones for line segments, and collections for line segments that come together at the beginnings and endings of surrounding areas. Each entity is a single, logical record in the computer, coded as variable length strings (X, Y) coordinate pairs. In CAD where analysis is not the primary purpose, the spaghetti model may be relatively efficient for cartographic display. This lack of topology, which is concerned with the study of topological spaces, however, results in enormous computation, making spatial analysis difficult.

A topological data model, instead, explicitly traces neighboring information into the data structure (Otoo, 1991). The basic entity in topological structure, the line segment, begins and ends when it intersects another line, or when there is a change in direction of the line. The identification number of each line segment is used as a pointer to indicate which set of nodes represents its beginning and ending. Polygons also have identification codes that relate back to the link numbers, reducing the tedious step in looking at, for example, which two polygons are also stored explicitly. The topological model more closely approximates how we as map readers identify the spatial relationships contained in an analog map document. It helps to obtain solutions to common operations in advanced analysis in GIS.

In comparison with the raster model, the vector data model usually requires less storage space in the computer (see Figure 2.3). Since topological relationships are readily maintained and graphical output more closely resembles hand-drawn maps, the vector structure seems to be the system of choice for many GIS users. Performing certain data overlay functions and spatial analyses, however, may be more difficult in a vector system because vector processing requires more sophisticated programming and processing time (Khedker & Dhamdhere, 1994). And vector objects do not represent spatial gradients as well. In a raster system, by virtue of its matrix, it is easy for the computer to identify adjacent features. The strength of the raster structure is for image display and processing because numerical pixel values can be used for image classification or processing operations (Cox, 1995).

The debates on the raster model and the vector model continue on the question related to the geometries and locations of geographic features, leaving the relationships between/among attributes alone. When seeking these relationships, however, it is sometimes more useful to look into the interactions within the real phenomena. The relational model discussed below is one to describe the interactions among attribute data, as opposed to spatial data.

2.2.3 The relational model

The relational model was formally introduced by Codd (1970) and has evolved since then. The relational model represents data in the form of two-dimension tables. A relation is a two-dimensional structure that contains data. A row of the relation is a tuple, while a column is a field or attribute. The tuple is analogous to a data record in a flat file, holding a collection of data items that describe an object. A key is an attribute that uniquely identifies tuples and provides a link between one relation and another. Relations have the following properties (Worboys, 1995): the ordering of tuples in the relation is not significant; tuples in a relation are all distinct from one another; columns are ordered so that data items correspond to the attribute in the relation scheme with which they are labeled.

In the GIS context, the relational model organizes attributes of geographic features into tables (Bonham-Carter, 1994; Foote & Huebner 1996). In the relational model, spatial data are combined with attribute data, which are stored separately in tables with a number of rows equal to the number of features in the binary tables and joined by a common identifier. The main success of the geographic relational model is to enable users to customize feature tables. In this way, not only could fields be added, but the database could be connected to any recognized external tables. The topology within the geographic relational model has a rigid structure and leads to practical results: points know which arcs are connected to them; arcs know which points constitute their origin and destination; arcs which polygons they form the perimeters of.

Since the relational model has its origin in business records (Codd, 1970), its application in the complex geographic data modeling results in a big problem - the behavior of a geographic feature is not able to be supported. In the relational model, geographic features are aggregated into homogeneous collections of points, lines, and polygons with generic behavior. Geographic features in the real world are complex and Twumasi (2002) defines three components for them: geometry, attributes, and behavior. Geometry describes the shape and locations of features; attribute describes the state of a feature at an instance in time; behavior describes features' dynamics, interactions with each other, and the operations on others. Several additional problems arising out of the relational model have been discussed in Chapter 1.

Technology continues to move forward. The advance of computer science has a strong impact in GIS. The object-oriented paradigm, which has shown great progress in computer science, changes the "computer view" of the geographic reality, resulting in the advent of the object-oriented geographic data model. The object-oriented geographic data model, to be discussed in the following sections, reduces the problems of the relational model, and improves the level of abstraction - the central concept of model, through capturing the special peculiarities of geographic data (Worboys, 1995; Tryfona et al., 1997; Twumasi, 2002).

2.3 The object-oriented approach

Object orientation can be described as a strategy for organizing a system as a collection of interacting objects that can combine data and behaviour (Egenhofer & Frank, 1992; Blaha & Premerlani, 1998). Object-orientation is based on the assumption that the real world can be modeled as distinguishable objects that can be grouped together into classes. This paradigm has been applied to many technological areas such as object-oriented programming, object-oriented analysis and design, and object-oriented graphical user interface design. All of these bring considerable advantages to the successful implementation of a GIS. The benefits that come from the object-oriented technology include extensibility, reusability, reduced complexity, and ease of use (Henderson-Sellers & Edwards, 1994; Blaha & Premerlani, 1998). The rest of this section is devoted to discussing the principles of the object-oriented approach and its expression in geographic data modeling.

2.3.1 Foundational concepts

An object can be defined as an identifiable entity, real or abstract, which has a precise role for an application domain (Roy & Clement, 1994; Blaha & Premerlani, 1998). To constitute an entity, something must be identifiable, relevant, and describable. By means of the modeling process, each entity relevant to an application domain is represented by a corresponding object in the data model. The object in the model should have properties that describe the characteristics of the corresponding entity in the real world. In object-oriented systems, each object is defined by three terms: identity, state, and behavior (object = identity + state + behavior). Object identity is the property of an object that uniquely distinguishes it from all other objects (Khoshafian & Abnous, 1995). In an object-oriented system, each object is unique. By introducing a unique identity for each object, different objects can be distinguished from each other without the need to compare their attributes or behavior (Ellmer, 1993). The object identity is usually system-generated, unique to that object and invariant for the object lifetime (Cooper, 1997).

The state of an object is described by the values of its attributes at one moment in time (Fussel 1997; Lee & Tepfenhart 2001). The attribute of an object is actually a named property of the object class that describes a value held by each object of that class. Objects can have a single state throughout their lifetime, or they can go through many state transitions. When an object moves from one state to another, it results in a new version of the object.

Although the concept of behavior may not be able to be implemented as flexibly as how human beings behave, it can be used to describe several aspects of objects: dynamics, interactions and semantics in object-oriented systems (Arbib, 2004). Geographic objects in the real world follow their logical spatial arrangements. For example, a river always flows downhill, not uphill. When two streams merge into one, the resulting stream is the addition of the two upstream flows. With object-oriented programming, we can calculate the volume of water in the downhill stream using a semantic and efficient way.

2.3.2 Abstraction mechanisms in the object-oriented environment

Abstraction is one of the fundamental ways humans cope with complexity. The main purpose of an abstraction is to manage the complexity of the system being modeled. Abstraction focuses on the essential parts of an application while ignoring the details (Blaha & Premerlani, 1998), and plays an important role in modeling what is central to object-oriented system analysis and development. There are four abstraction mechanisms in the object-oriented environment: classification, generalization, association, and aggregation (Egenhofer & Frank, 1992).

Classification is a well-understood form of abstraction. It is a mechanism to organize similar objects into groups or classes (Rambaugh et al., 1991). An object is an instance of a class. A class is an abstract blueprint from which individual object can be made. In this sense, a class is a template for defining the behaviors and properties for a particular type of object. The use of the class is related to such concepts as structure, inheritance and abstract (Graham, 1996). A class provides the structural definition for instances of that class, that is, the names and types of their attributes and methods. When new objects are created as instances of the class, they can access all the attributes and methods of the class to which they belong. This is called inheritance. Some classes are never directly instantiated, that is, they cannot have instances themselves. These classes are called abstract classes. The descendants of abstract classes, however, can have instances. Abstract classes are useful for reducing redundancy of class specification. General specifications can be defined in an abstract class that can be shared by other inheriting classes. For example, the GIS model for a town may include the classes Residence, Commercial Building, and Street. A single instance, such as a residence with the address "150 Lebreton Street North, Ottawa, Ontario" is an instance of the class Residence.

Generalization groups several classes of objects, which have some properties and operations in common, to a more general superclass (Goldberg 1983 in Egenhofer & Frank 1992). It is a relationship between a class (superclass), and its one or more special cases (subclass). The terms subclass and superclass characterize generalization and refer to object types that are related by an is_a relation. For example, the object Residence is a building; Residence is a subclass of building, while building is its superclass. The superclass holds common attributes and methods and the subclass inherits them, adding their own attributes and methods. Objects being instances of the class inherit the attributes and methods specified for the class.

Association is the physical link between objects, and denotes some semantic dependency between the objects. A simple example is the ownership relationship between a parcel and an owner. From parcel's perspective, the ownership relationship captures the fact that it is owned by the owner. Similarly, from the owner's perspective, the relationship captures the fact that he/she owns the parcel. Thus, it is very common to have role names (parcel, owner) associated with a relationship (ownership). Association has another property called cardinality (Lee & Teprenhart, 2001), which describes the number of objects from one class that relate to single objects in the associated class. There are three general kinds of cardinality across an association: one-to-one, one-to-many, and many-to-many. The ownership in this case, for example, is a one-to-one relationship. In other cases, one owner may own several parcels, or one parcel may be owned by several owners.

Aggregation is an abstraction mechanism that allows a more complex object to be composed of one or more components. Aggregation is a kind of association between the whole (composite object) and its parts (component objects). The relationship formed by aggregation is the part_of relationship (Hornsby & Egenhofer, 1998). For example, the class building is an aggregate of all walls, windows, doors, and roofs that are part of it. Composition is a strong form of aggregation in which the composite object controls the lifetime of the parts. Hence, if a composite object was destroyed, it would destroy all of its parts.

2.3.3 Principles - inheritance, encapsulation, polymorphism

Inheritance is the transmission of the properties from superclass to all related subclasses (Harrington, 2000). In generalization hierarchies, the properties and methods of the subclasses depend upon the structure and properties of the superclasses. Properties that are common for superclasses and subclasses are defined only once with the superclasses, and inherited by all objects of the subclass, but subclasses can have additional, specific properties and operations, which are not shared by the superclasses. This concept is very powerful, because it reduces information redundancy and maintains integrity in geographic data modeling. Modularity and consistency are supported since essential properties of an object are defined once and are inherited in all relationships in which it takes part.

Inheritance is either single or multiple. As demonstrated in Figure 2.5, with single inheritance, a subclass (i.e., oceanography) may inherit attributes and methods from a single parent class (line object), usually adding its own attributes and methods. Multiple inheritances mean a subclass (i.e., squid) inherits from multiple superclasses (Biodiversity and Oceanography).

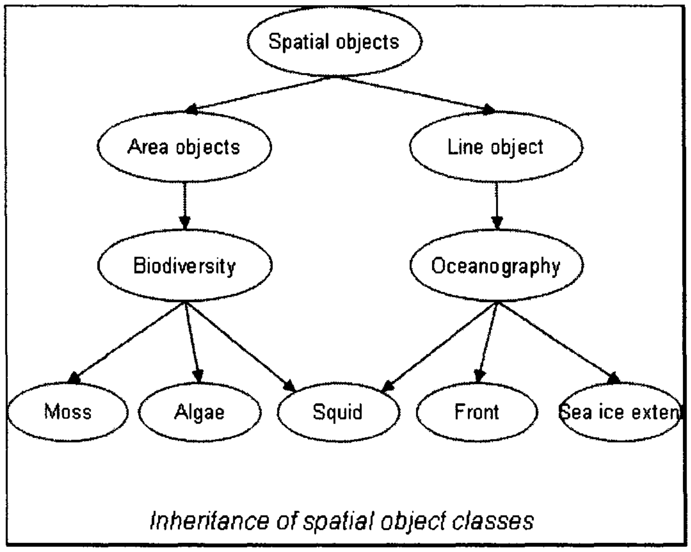


Figure 2.5 Inheritance

Two forms of inheritance of spatial object classes: single (the subclass, i.e. oceanography, may inherit attributes and methods from a single parent class-line objects), and multiple (a subclass, i.e. squid, inherits from the two superclasses - Biodiversity and Oceanography) (by the author).

Encapsulation refers to objects containing data (attributes), and the methods (code) that will manipulate the object (Harrington, 2000). The object's data are accessed only through its methods. The data and code are tied together in an integral unit, working as one. An object class encapsulates both the structure and the behavior in that the characteristics include both attributes and the methods. The organization of that data is known only to the object itself, and is not accessible from outside the object except through a specific and carefully designed interface (Cooper, 1997). Encapsulation thus separates the internal implementation of an object from its public interface. All other details concerning the object are hidden behind its interface. Users do not have to understand the internal implementation of the object (i.e. how its methods are executed). An important benefit of encapsulation is that since the internal implementation is completely hidden within the class, the implementation of a class can be changed without affecting other objects in the system. In other words, the implementation can be extended and modified without affecting the users of the class; this leads to easy software extensibility and reusability (Khoshafian & Abnous, 1995).

One of the goals of object-oriented technology is to reuse code. Polymorphism is a system which allows the same operation to be implemented in different ways by different methods in different classes (Worboys, 1995). Polymorphism makes it possible for two or more objects to respond to the same message differently. Suppose "map feature" is an object. When the user clicks on the "export" icon, the message that is sent to the different objects is the same. However, the actual implementation of what the objects should do varies. Every object must know how to export itself. Different lines of code are activated depending on the type of map feature. The "hidden code" in the data/function bundle first checks the type of feature, and then automatically calls the appropriate routine to export. Polymorphism enhances software reuse by making it possible to implement generic software that will work for not only existing objects, but also for objects to be added later.

2.3.4 Benefits of the object-oriented approach to geographic data modeling

A geographic data model is a representation of the real world that can be used in a GIS to produce maps, perform interactive queries, and execute spatial-based analysis (Zeiler, 1999). Several GIS software vendors provide object-oriented geographic data models, which vary in degrees of sophistication. For example, ESRI ArcGIS applies the object-oriented approach in most of its applications, except for the creation of the physical model which combines the relational technology. The use of the relational approach, however, can be used to integrate raster data into the physical model. GE Smallworld GIS, another traditional spatial technology, also extends the conventional relational database approach by integrating object-oriented design (GE, 2005). Although GIS data models in these systems are not truly object-oriented and are named as object-relational data models in some publications (Fussel, 1997; Zeiler, 1999; Twumasi, 2002), the object-oriented paradigm is central to those particular GIS technologies. A full understanding of the benefits of the object-oriented data model is helpful to the application of GIS in various domains.

The object-oriented approach treats not only the static data – oriented aspects of geographic information, but also the dynamic behavior of the system (Worboys, 1995). Object dynamic behavior has been ignored in traditional GIS systems (Twumasi, 2002), although it holds a critical role in geographic databases. In the past, features of the same type stored in a GIS database were collected together and assigned other attributes to more fully define their nature. This causes geographic features to be mapped into homogenous collections of points, lines, and polygons with "generic" behavior (Chapter 1). With the advent of object-oriented GIS, users are no longer limited to the simple points, lines, and polygons of the past. Lines representing a road, a wall, a shore, or the center of a river need not be functionally equivalent. They do not have the same behavior, so they are not expected to respond to operations in the same way. As a simple example, a river line almost always has one directional flow, while a street usually has two-way flow (Davis, 2000). The successful capture of the special peculiarities (including position, attribute, and behavior) of geographic data is a result of the basic concepts of the object-oriented paradigm – class, as discussed earlier in this chapter. Instead of representing objects as lines and nodes, the object-oriented model stores objects in logical groups - "classes". Each class has predefined properties unique to that object, which are built directly into that object.

The object-oriented model extends the historical two-dimensional perspective on the world to the multi-dimensional. Since GIS has an origin in conventional cartography, most GIS tools tend to take a two-dimensional view of reality. All geographic features we are trying to model, however, have a third dimension or even a fourth dimension. Multidimensional modeling tries, on one hand, to make the data schema closer to our perception, and on the other hand, to improve query performance (Malinowski & Zimanyi, 2004). The object-oriented paradigm provides an excellent mechanism for this purpose. For example, an ice block with location, height, color, direction, and time of observation as attributes can be represented in an instance of a class. Each attribute can be treated as a single dimension so as to be separately represented in the computer and accessed by users. As a result, the beauty of the geographic objects can be fully displayed and the query is more efficient.

Object-orientation enhances the level of abstraction in a way close to our perception of the real world. The concept of abstraction is central to the model. It has been argued that the previous models obscure essential issues about how people think about and manipulate space (Teller, 1996). Suppose a description of the agricultural activity in a certain area is required. One might speak of farms in the region as geometric objects that have owner attributes. Fields could be objects with the crop as attribute, and could inherit from the farms that contain them. This is an example where the object-oriented approach to data is natural: there is clear separation of objects, their representation can be encapsulated, and they form a hierarchy in which representations and attributes can be inherited.

The object-oriented approach makes geographic data modeling more efficient. An object-oriented data model can be planned and designed using computer-aided software engineering (CASE) tools and the UML (see the detail in Chapter 3). CASE tools and UML can automate and simplify the database design process. An object-oriented design is a cumulative process that is graphical, more straightforward, and refined. CASE tools and UML can interpret and portray actual physical processes, workflows, and applications, and then automatically generate a physical database. Developers can demonstrate the interim designs along the way by loading a small amount of data, creating an object-oriented model, and benchmarking the results to see if the design meets the user's needs. In comparison to the relational approach, the object-oriented approach offers a unified representation approach to the system development cycle (analysis, design, and implementation) (Worboys, 1999).

The precise process using the object-oriented approach lessens the information loss from one level to another, during the course of geographic data modeling. The top-down three-level modeling is laborious work. Information is usually lost in mapping from one level to another (Twumasi, 2002). The object-oriented approach handles intelligent objects with precise rules and relationships embedded up front in the database design (Demartino & Hrnicek, 2001). Most applications in the past were developed after the database was designed and data conversion completed. Custom code has to be written and proprietary programs developed and tested to ensure a working application. With an object-oriented GIS, however, the rules and relationships required for applications reside inside the database design itself. Therefore, the number and complexity of custom programming routines that need to be written and run in the background are reduced.

2.4 Principles of geographic data modeling

The data modeling process has been based on the top-down, three-schema architecture: the conceptual data model, the logical data model, and the physical data model (see Figure 2.6) (Elmasi & Nacathe, 1994). “Conceptual” means a definition of the problem; “logical” means a design of a solution to the problem; “physical” means the solution of the problem (Becker, 2000). The conceptual data model, the top-level schema, is independent of the information system. In the conceptual model, objects on the earth with their attributes, including geometric ones, are defined to realize the relations and their cardinalities between different objects. One of the most widely used conceptual level models is the Entity-Relationship (ER) diagram (Chen, 1976). In the last few years, since the object-oriented model matured, the UML diagrams have taken a predominant role especially in computer science. Wide-spread CASE tools based on this language include Rose (Rational), and Visio (Microsoft).

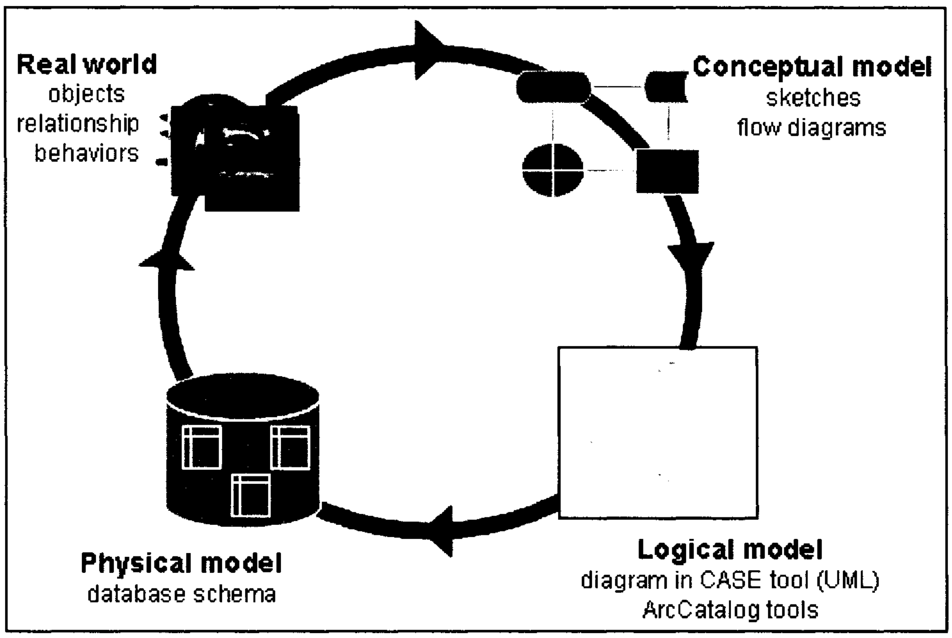


Figure 2.6 The mechanism of geographic data modeling

Based on the real world phenomena observed, three top-down levels include: the conceptual model, the logical model, and the physical model (by the author).

2.5 The role of GIS in addressing environmental problems

2.5.1 Spatial issues in addressing environmental problems

Environmental scientists are being asked to predict how future changes in climate will affect the environment and how the environmental system (e.g. biological communities) will respond at different scales - local, regional, and global scales (Eddy, 1993; Walther et al, 2002; Murphy, 2004). These predictions are needed to allow sensible responses to impending changes in climate and other aspects of the environment (Eddy, 1993).

Ecological processes operate at a variety of scales in space and time. Historically, ecologists focused on changes in time and single sites or small geographic areas. They ignored the spatial dimension of environmental changes to make them simpler and easier to understand. For example, biologists project the effects of climate change in the trend of a population by predicting births, aging, deaths and net migration. Recently, research on understanding spatial heterogeneity of ecological communities has emerged in marine research, and ecologists have applied models to large geographic areas (Hunsaker et al., 1993; Goldberg et al., 1997). The marine environment, for example, is spatially heterogeneous: biomass varies as a result of the heterogeneity in environmental variables, and fish and marine mammals move from one environment to another. Such migration may cause temporal changes in the spatial distribution of biomass and influence prey availability to predation. Many studies have tended to combine effort from other disciplines including geography, physics, biology, and ecology. A better approach to integrate multidisciplinary information for understanding the spatial aspects of ecological processes is needed. GIS has proven to be a powerful tool for multidisciplinary research, and one of the most successful application areas of GIS is that of addressing environmental problems (Goodchild, 1993).

2.5.2 The role of GIS

A GIS is a computer-based information system that enables capture, modeling, manipulation, retrieval, analysis and presentation of geographically referenced data (Worboys, 1995). It is a tool to visualize and analyze spatial data using spatial analysis functions such as search (thematic search, search by region, classification), location analysis (buffer, corridor, overlay), terrain analysis (slope/aspect, catchment, drainage network), flow analysis (connectivity, shortest path), distribution (change detection, proximity, nearest neighbor), spatial analysis/statistics (pattern, centrality, autocorrelation, indices of similarity, topology), and measurements (distance, perimeter, shape, adjacency, direction).

GIS may be used simply as a storage, retrieval, and display system for geographic data, but the real value of GIS lies in its ability to create new data by manipulating existing data or analyzing the relationships among sets of data. Spatial analysis is accomplished by programs which use data retrieved by the management system. The analysis functions to answer questions that may not be explicitly stated in the data (Cox, 1995;

Shekhar & Chawla, 2003). The ability to overlay the geographic data layers is very important to GIS. Overlay is the process of stacking digital representations of various spatial data on top of each other so that each position in the area covered can be analyzed in terms of these data. To overlay the data, the layers must share a common projection and coordinate system to ensure spatial registration. This capability provides GIS with its analytical power – the integration and analysis of a wide variety of data.

2.5.3 Geographic data distribution through the Internet

As a source of knowledge, data are usually expected to be available in a timely fashion. Transferring geospatial data through the Internet provides such as an opportunity. The advent of the Internet and Web-based GIS technologies provides a convenient and efficient way to access and disseminate geographic data (Tsou, 2004). In the development of Web-GIS, currently, the major contribution is cartographic visualization and the application area in environmental science is largely in environmental monitoring. The combination of data collection through the Internet with geospatial analysis tools can largely reduce the high cost and labor compared to the traditional field monitoring and environmental resource management methods.

Several issues come up with the distribution of geographic data through the Internet. "How to share data?" "How to integrate the multi-source data?" "How to deal with geospatial data and convert them into useful knowledge?" These may be the common questions asked by users. Although they are still on-going, some solutions have been developed. One example is OGC Web Map Service (WMS) specification. Through WMS, users can produce complex maps without having to touch the actual data located in remotely and separately distributed databases. OGC WMS is currently the most popular specification for sharing geographic data over the Internet.

2.6 Review

In summary, there are several reasons why a geographic data model proves to be both necessary and important. A geographic data model can allow scientists to process and to inter-relate diverse data types. A successful use of GIS depends on the organization and manipulation of the data, which requires a model of how the phenomena exist in the real world. A data model provides the framework through which a hypothesis may be tested. In addition, the data volumes and integration requirements are being driven by the complexity of environmental problems. It thus becomes advantageous to use a GIS to automate the manual process of gathering and analyzing the wide variety of spatial data needed to make decisions (Alvaro, 1995).

A practical approach is taken in the next chapter to demonstrate the significance and convenience of representing spatial data in an object-oriented model, using a case study on southern ocean marine ecological dynamics.

CHAPTRER 3 AN OBJECT-ORIENTED MARINE DATA MODEL

3.1 Overview

The process of geographic data modeling consists of three stages, increasing in abstraction as one moves from human orientation to implementation in a computer (Li, 2000). This chapter presents the development of an object-oriented geographic data model, starting from the conceptual data model, through the logical data, realizing the model in the database schema. The relevant strategies and techniques used at different levels are discussed respectively.

3.2 Application domain

3.2.1 Southern ocean ecological environmental issues

The southern ocean is biologically rich, and spatially and temporally dynamic. Traditionally, however, as Moloney & Johnston (1996) point out, few marine ecosystem models have been constructed taking into account spatial variation and movements of animals. One of the reasons might be attributed to the difficulties of obtaining marine data. Today, with the development of technology, many invaluable data have been collected through various platforms, such as ships, moorings, and satellites (CSIRO Marine Research, 2003). The accessibility to large amounts of marine data makes it more feasible to build marine models for spatial analysis. Thus, a lot of attention is dedicated to exploring suitable techniques for developing tools for spatial analysis. Maidment (2001) indicates there is a strong increase in the use of GIS in marine resources studies, prompted by such factors as the use of digital elevation models, the rising use of aerial photogrammetry, and so on.

Southern elephant seals *M. Leonina* are one of the top predators in the southern ocean ecosystem. As discussed in Chapter 1, this species has been regarded as an indicator of the impacts of climate change on the southern ocean ecosystem (Vergani & Stanganelli, 1994, 2003; Vergani et al., 2001; Vergani et al., 2002; Vergani et al., 2004). The behavior of seals can be spotted and sampled by measuring devices whether at sea (through satellite telemetry, and time-depth recorders) or on land (through field observation and common measuring devices). Although seals land for breeding and moulting twice a year, they spend most of their time at sea searching for food. Sea ice, one of the most important marine abiotic features in the southern ocean, has been suspected as a significant factor affecting the diving behaviour and foraging activities of seals. Information on the environmental parameters of sea ice, including extent and concentration, can be derived from data collected by satellites and cruises. The domain of GIS concerns georeferenced data, plus integration and analysis procedures, transforming the raw data into meaningful information (Valavanis, 2002).

Due to a better understanding on how life interacts with the Earth's other dynamic processes, "ecology has taken its place among sciences' vital, strategic disciplines"

(USAP, 2001). In Antarctica, several long-term ecological programs are being conducted to research the relationships between/among climate change, ecological migration, and teleconnections in a harsh environment. Among them are programs such as Palmer Long-Term Ecological Research Project (USA) (Ducklow, 2003), Australian Antarctic Science projects (Australia), Centro Nacional Patagónico long-term program (Argentina). The first two can be found in detail through web sites at: http://www.nsf.gov/index.jsp (National Science Foundation); and http://www.aad.gov.au/ (Australian Antarctic Science), respectively. The last one, which contributes to this thesis research, is briefly described in the next section.

3.2.2 The Centro Nacional Patagónico (CNP) long–term research program

The Centro Nacional Patagónico (CNP) long–term research program (Puerto Madryn, Argentina) was established to measure the impact of the EI Nino Southern Oscillation (ENSO) on the southern ocean ecological dynamics as indicated by elephant seals. Field work has been conducted since 1980 at King George Island (62°14'S, 58° 30' W, Antarctica), and biological data (e.g. population fluctuation, weaning mass, suckling length, and birth weight) about population parameters of southern elephant seals have been collected annually. One important effort of this program is to integrate biological and physical data in a geographical and climatological context in order to study the impact of climate change and climate variability on the functioning of the southern ocean ecosystem. To accomplish this, it is essential to establish a robust time series of data for population variables (i.e., weaning mass, total population, male/female population, sex proportion, squid biomass, squid beak, mortality) and climatic parameters (i.e., sea surface temperature, sea ice concentration, and ocean circulation).

The overall goal of this thesis, as mentioned in Chapter 1, is to produce a common structure and a data model template for assembling, managing, and publishing existing and future marine data, using representations of location and spatial extent, along with a means for conducting more complex spatial analyses of the data. It also aims to better understand scientific information on the southern ocean ecological dynamics through cartographic visualization and distributing data over the Internet. A framework is illustrated in Figure 1.1 (Chapter 1).

3.3 Object-oriented analysis

The data modeling process usually starts with project analysis. This section is concerned with requirements analysis of the application domain. Requirements analysis is primary to model any geographic data. Christensen et al. (2001) list several requirements based on their knowledge and interviews with users, including three main types: aspects related to geographic objects, aspects concerning the relation between objects, and aspects related to the events of an object. Part of the object-oriented analysis of the application domain is based on this work of Christensen et al. (2001).

3.3.1 Spatial extent and time series

The spatial location and extent are defined by x, y, z coordinates. Since the earth is nearly spherical, and in most cases, the map is displayed in two-dimensional media, obtaining information from a curved surface to a flat one thus needs a mathematical formula called a projection - the primary factor distinguishing cartography from any other visual discipline. Projections are important in geographic data modeling for several reasons. They make the process of superimposing two or more maps (sometimes called overlay) accurate. They also guarantee measurements on the map (either paper or digital) to be correct, which consequently ensures spatial analysis (e.g. generating contours, histograms) can be conducted in the accurate way. Polar stereographic projection (South) is used as the spatial reference system of the South Polar Region.

The spatial extent in this research covers the Antarctic Peninsula Region, which plays a key role in the southern ecosystem and therefore many scientific research activities have been conducted around it.

Time series is one of the most important factors considered by environmental scientists since they can be used to predict dynamics of real phenomena. Changes within a short period of time are not detected easily, but a relatively wider range of time scale may be significant in obtaining past information and to foretell the future. The southern ocean ecosystem is sensitive to climate change (IPCC, 2001), and looking through the data over a long period of time may help to examine the interval between events within this particular system.

3.3.2 Geographic objects, their logical groups, and their representations

The main geographic objects to be considered in this project include: distributed seals, squid, sea water, and sea ice. The aspects of objects to be considered in geographic data models consist of location, spatial extent, temporal scale, spatial dimension, thematic values, entity and field based data, generalization, constraints, and data quality.

Squid, next to krill, play the most important intermediate role in the food chain within the southern ocean ecosystem (Takashi, 1994). Although squid have seldom been caught alive, field scientists have observed an abundant and diverse squid fauna in the diet of vertebrate predators (Daneri et al., 2000; Piatkowski et al., 2002; Van den Hoff et al., 2003). Research results suggest that the squid *Psychroteuthis Glacialis* is the most important prey of southern elephant seals in terms of numbers, biomass, and frequency of occurrence. Squid in the southern ocean were caught through the fishing gear (e.g. Rectangular midwater trawl, bottom trawl, pelagic trawl, phytoplankton net, small trawl surface net) (Xavier, 1997). All species of squid are grouped into one class – squid. The spatial representation of squid distribution is POINT.

Seal tracks are considered because they demonstrate the feeding routes of seal after breeding and moulting. In order to understand how seals find and exploit food resources, scientists try to trace the movements and dive behaviors of the seals through telemetry such as Satellite Relay Data Logger (McConnell & Fedak, 1996), Geolocation Time-depth Recorder (Jonker & Bester, 1998), and satellite transmitters (Bornemann et al, 2000). All seal tracks are grouped into one class – seal track. The spatial representation of seal track is LINE.

Breeding colony refers to the site where seals regularly land for breeding. Southern elephant seals have a circumpolar distribution (Stanganelli & Vergani, 2000), mainly breeding ranging from Peninsula Valdes (42° 08′ S, 65° 05′ W, Argentina) to King George Island (62° 08′ S, 58° 27′ W, Antarctica) (Laws, 1994; Bornemann et al, 2000; Bradshaw et al, 2002; Vergani et al, 2004). Several breeding sites have been identified by scientists, and they are grouped into one class – colony. The spatial representation is POLYGON.

Sea ice is an important feature in the Antarctic Region. Annual sea ice dynamics affect both ocean structure and circulation (Gradinger & Schnack-Schiel, 1998). For example, sea ice growth during autumn-winter causes salt to be released from the underlying ocean, leading to an increase in seawater density and, occasionally, to deep ocean convention and bottom water formation. Many studies have suggested that Antarctic sea ice extent is related to climate anomalies such as ENSO, which therefore may have an impact on the southern ocean ecosystem (Guinet et al., 1994; Howard-Williams, 2001; Yuan, 2004).

In the case of southern elephant seals, sea ice extent variation may affect their feeding areas since sea ice is an important habitat (Vergani et al., in press). Information about the foraging areas may explain the fluctuation of southern elephant seal populations. The raster structure is used to represent sea ice variation.

Temperature is one of the fundamental variables which determine the physical properties of sea water such as freezing point and thermal expansion (Hunt-Jr et al., 1992). Because the Antarctic Convergence lies in relatively high latitudes, the extreme variations in incoming solar radiation between summer and winter have a pronounced effect on marine life (Jonker & Bester, 1998). Bonner & Walton (1985), in plots of the mean monthly values of radiant energy against primary production and the standing crop (biomass per unit volume) of phytoplankton, showed that the increase and later decline of the phytoplankton biomass and primary productivity reflect the amount of energy received. Satellite-derived measurement of sea surface temperature (SST) is one of the most influential remotely sensed oceanographic parameters providing a quantitative view of thermal features in the ocean, such as fronts. The raster method is used to represent sea surface temperature.

Population parameters and research activity information, to be detailed in the following sections, are organized into attribute tables.

Sea water is the main physical factor in the marine ecosystem. Ocean fronts are considered and represented as LINEs; bathymetry is represented as a raster.

3.3.3 Relations among geographic objects

The aspects of geographic objects to be considered include topological relationships, semantic relationships, and relationship constraints.

Topology, simply stated, is the way in which geographic elements are linked together (Burrough, 1986). Topology is built on two concepts: absolute location and relative location. Absolute location is the actual location of a geographic feature, often expressed as X/Y coordinates. Relative location is the location of a feature in relation to another feature(s). For example, the absolute location of a seal breeding site may be 30°W / 69 ° S. Its relative location may be within the King George Island, west of a penguin breeding site. The construction of topology can be used to calculate the Euclidean distance between features, generating buffers, calculating areas and perimeters.

Semantic relationships emerged in response to the need to incorporate more understanding of the real world in the information that is stored in a database (Storey, 1993). They are data model constructs, connecting a pair of classes and reflecting the characteristics of the specific relationship in application domains. Halper et al. (2003) indicate that semantic relationships are helpful to capture the meaning of data by means of abstractions such as class inclusion, aggregation and association based on the research in linguistics, logic, and cognitive psychology.

Semantic relationships in the spatial context, as mentioned in Chapter 3, can be described as "is_a", "kind_of", "part_of", "has/have", and so on. In the case of southern ocean elephant seals, semantic relationships can be described as follows:

- Alluroteuthis/ Brachioteuthis/ Galiteuthis/ Gonatus/ Kondakovia/ Psychroteuthis "is_a" species of squid
- One marine track device produces zero/one track; one tracking activity needs many devices
- One land research activity produces much land observation information
- Seals go to different feeding grounds from colonies
- Front name is part of ocean front class, that means, if ocean front changes, the name varies

Tables 3.1, 3.2, and 3.3 display the information for each class to be designed into the model.

Land Research Activity
Year
description
leader
member

Seal Track
Species name
Year
Sex
Mass
Age
Reference number

Breeding Colony
Recording Date
Boundary Type

Non-breeding Colony
Recording Date
Boundary Type

Ocean front
Name

Table 3.1 Information on research activity to be recorded

Land Observation Information
local environmental condition
Adult population
Adult male population
Adult female population
Adult sex proportion
diet biomass of Alluroteuthis
diet biomass of Brachioteuthis
diet biomass of Galiteuthis
diet biomass of Gonatus
diet biomass of Kondakovia
diet biomass of Psychroteuthis
beaks of Alluroteuthis
beaks of Brachioteuthis
beaks of Galiteuthis
beaks of Gonatus
beaks of Kondakovia
beaks of Psychroteuthis
total number of new-born
numbers of dead pups
mortality
Female weaning mass
Male weaning mass
Index of Anomaly (IAS)
ENSO
Year

Table 3.2 Information on research activity to be recorded

Psychroteuthis Glacialis
Minimum depth
Maximum depth
Minimum size
Maximum size
Date caught
Number caught
Fishing gears used
Longitude
Latitude

Kondakvia Longimana
Minimum depth
Maximum depth
Minimum size
Maximum size
Date caught
Number caught
Fishing gears used
Longitude
Latitude

Alluroteuthis Antarcticus
Minimum depth
Maximum depth
Minimum size
Maximum size
Date caught
Number caught
Fishing gears used
Longitude
Latitude

Gonatus Antarcticus
Minimum depth
Maximum depth
Minimum size
Maximum size
Date caught
Number caught
Fishing gears used
Longitude
Latitude

Brachioteuthis Sp.
Minimum depth
Maximum depth
Minimum size
Maximum size
Date caught
Number caught
Fishing gears used
Longitude
Latitude

Galiteuthis Glacialis
Minimum depth
Maximum depth
Minimum size
Maximum size
Date caught
Number caught
Fishing gears used
Longitude
Latitude

Table 3.3 Information on six species of squid to be recorded

3.3.4 Consistency constraints

Consistency constraints are the conditions that must hold for the data in the database to be consistent, valid and useful (Cooper & Schiex, 2004). In real world databases, the value of object's attributes is constrained to a range of acceptable values. The enforcement of consistency constraints ensures data collected is entered properly. For example, attribute domains specify the required attribute values that can be input for an object. Examples in the thesis research include maximum size/mass/length constraints, and minimum size/mass/length constraints, with each being specified. They are growth (weight and length) of seals, and variation of sea surface temperature, and ranges of the size and depth of squid. Table 3.4 demonstrates consistency constraints in this thesis research.

Consistency constraints (range)
Length of southern elephant seals (m): 1-7
Weight of southern elephant seals (kg): 1-4000
Size of squid:1-500
Depth of squid caught (m): 0-11000
Sea surface temperature (° C): 0-32

Consistency constraints (code)	
Squid species	**Coded Value**
Alluroteuthis	1001
Brachioteuthis	1002
Galiteuthis	1003
Gonatus	1004
Kondakovia	1005
Psychroteuthis	1006

Table 3.4 Consistency constraints in this thesis research

3.3.5 Multidimensionality

The multidimensional "behavior" of geographic features was neglected in the traditional coverage model which only considers zero-dimensional (POINT), one-dimensional (LINE), and two-dimensional (POLYGON) shapes. In this research, the representation of three-dimensional shape is necessary and used for several groups. The depth of squid caught, for example, reflects the potential depth that seals dive. Together with bathymetry, the depth also reflects some of the environmental conditions where other marine organisms live.

In the case of southern elephant seals, bathymetry and the depth of squid caught are represented in three-dimensions.

3.4 Object-oriented design

The basic idea of object-oriented design, like its related discipline – object-oriented programming, is that the model should be written according to a mental model of the actual or imagined objects it represents. The Wikipedia dictionary (2004) reviews the relevant publications and points out that object-oriented design concentrates on modeling real-world interactions and attempting to create "verbs" and "nouns" which can be used in intuitive ways.

3.4.1 Current conceptual modeling techniques

Data modeling has become a pervasive practice in the database industry. As database technology and design methodology mature, there is an increasing need for modeling techniques to aid in database design. Geographic data modeling, from the standpoint of GIS technology, is part of system modeling. Currently, there exist two general techniques for geographic data modeling: the Entity-Relationship (E-R) approach and the Unified Modeling Language (UML) approach.

3.4.1.1 The Entity - Relationship (E-R) Model

The E-R technique, developed by Chen (1976), is widely used in the design of relational models. E-R modeling is a schematic representation of the constituent parts of a scenario which require storage within a database. The basis of this technique is the entities and the various relationships between them. This technique provides a graphical notation in the form of E-R diagrams. With E-R notations, entities representing geographic objects to be stored within the database can be represented as boxes, linked by lines. The E-R model, a data model for high-level descriptions of conceptual data models, provides a graphical notation in the form of E-R diagrams.

More recently, most object-oriented modelers have adopted the UML, which is a standard notation for expressing object models and is endorsed by leading software companies.

3.4.1.2 The Unified Modeling Language (UML)

The Unified Modeling Language (UML) is briefly introduced in this section. UML is a visual language that can be used in developing software systems (Bennett et al., 1999). It is neither a human language, nor a programming language. It is, however, a formal specification and documentation language for business modeling and other non-software systems. Simply speaking, UML is something like a blueprint that both architects and builders understand well enough to construct a building. UML has been chosen in this project because of its robust semantics, notation definitions, and its wide acceptance.

Formal specification languages such as UML set up their own elements and rules, just like the programming languages that consist of a set of elements and a set of rules to make valid programs. Similar to traditional algebraic notations that define the standard language of mathematics (e.g. A, a, B, b), a data modeling language is made up of a set of notations including lines, rectangles, ovals, and other shapes. Most of the elements of UML are graphical and labeled with words providing additional information. The graphical representation provides multiple views of a system and helps people to understand the model or parts of the model. It is the graphical representation that makes UML a visual specification language rather than a textual one, and may be one of aspects favored by people.

Figure 3.1 shows some of the elements used in the project. In the UML notation an object is shown as a rectangle with two compartments. The top compartment shows the name of the object and its class. The second compartment shows the attributes of the object and their values. A class is represented in UML by a solid-outline rectangle with three compartments separated by two horizontal lines. The top compartment holds the class name and other general properties of the class, the middle compartment lists attributes, and the bottom compartment has operations. An abstract class has its name in italics. A class interface is a specification of properties and methods. Many classes can implement the same interface, which allows a high degree of interoperability and shared behavior among a set of object. An interface is represented in UML as a rectangle with two compartments, the top illustrating the interface name with the keyword <<interface>>, and the bottom part showing the list of operations. The relationship between a class and its interface is shown by means of a refinement.

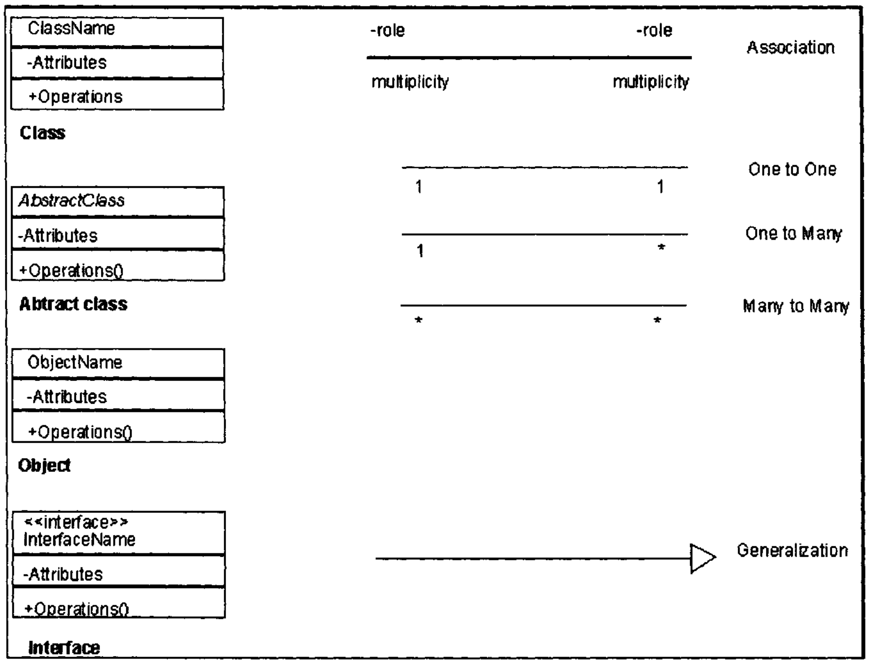


Figure3.1 Part of UML notations used in this case study

More detail can be found in the text, (by the author, based on Visio (Microsoft))

"<< >>", stereotype, tag for interface, attributes, or operations

"-", visibility (Private)

"+", visibility (Public)

Association is a relationship between two or more classes. It is represented in UML as a line connecting two classes with the association name above the line. This kind of association is called binary association in UML. The association role is illustrated at both ends of the line next to the class. Association cardinality is shown above the association line near the appropriate class. An association may have attributes and operations like a class (In ArcInfo, it is called a relationship class). Generalization is a relationship between a superclass and its subclass, with a triangle on the end of the line pointing to the parent class. Complete UML elements, semantics and formal descriptions can be found in Bennett et al. (1999).

Currently, the tools supporting UML are available as commercial software packages, such as Rational Rose (IBM) and Visio (Microsoft).

3.4.2 The object-oriented UML conceptual model

The object-oriented conceptual model was built using UML based on the previous analysis. The geographic features sharing common attributes are grouped into one class; relationships are established and organized into relationship classes. Visio (Microsoft) is used to draw UML diagrams. Part of the conceptual model is demonstrated in Figure 3.2, and will be compared with the next level of model – the logical model.

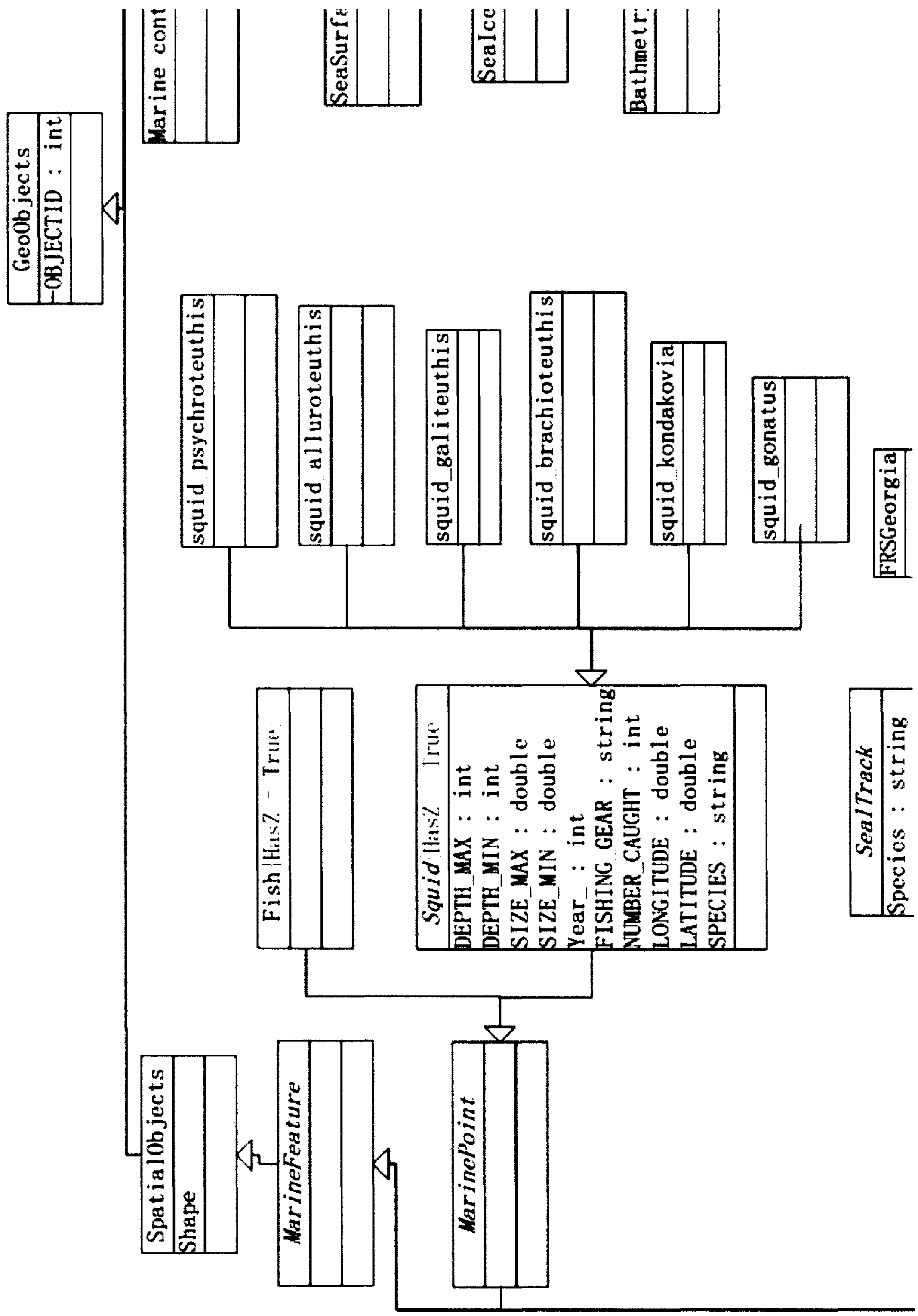


Figure 3.2 Part of the object-oriented conceptual model

The complete model can be found in Appendix A.

3.5 Implementation

This section is devoted to the implementation phase. The conceptual model designed in the previous section is translated into a logical data model and implemented in ArcInfo as a physical data model. The strategies and techniques involved in this implementation are briefly described as follows.

3.5.1 Methodology

Currently, the implementation of an object-oriented data model is usually accomplished through an extension of relational database management systems, such as Geodatabase (ESRI) and PostgreSQL (POSTGRES). Geodatabase or PostgreSQL adds numerous enhancements to the straight relational data model, including support for arrays (multiple values in a single column), inheritance (child-parent relationships between tables), and functions (programmatic methods invoked by SQL statements). The classes are two-dimensional tables which store related pieces of data; object-instances refer to rows; object-attributes refer to columns.

The general strategy, as illustrated in Figure 3.3, is to create a geographic database using UML and CASE tools involving designing a logical model and generating the database schema. The UML logical model, to be explained in Section 3.5.5.2, is created using a specialized tool –Visio (Microsoft), and then exported to an intermediate format – XML Metadata Interchange (XMI) file. The UML logical model is created under the ArcInfo

UML Model workspace. The CASE tools subsystem of ArcInfo (ESRI) will read the model and create the database schema (Zeiler, 1999).

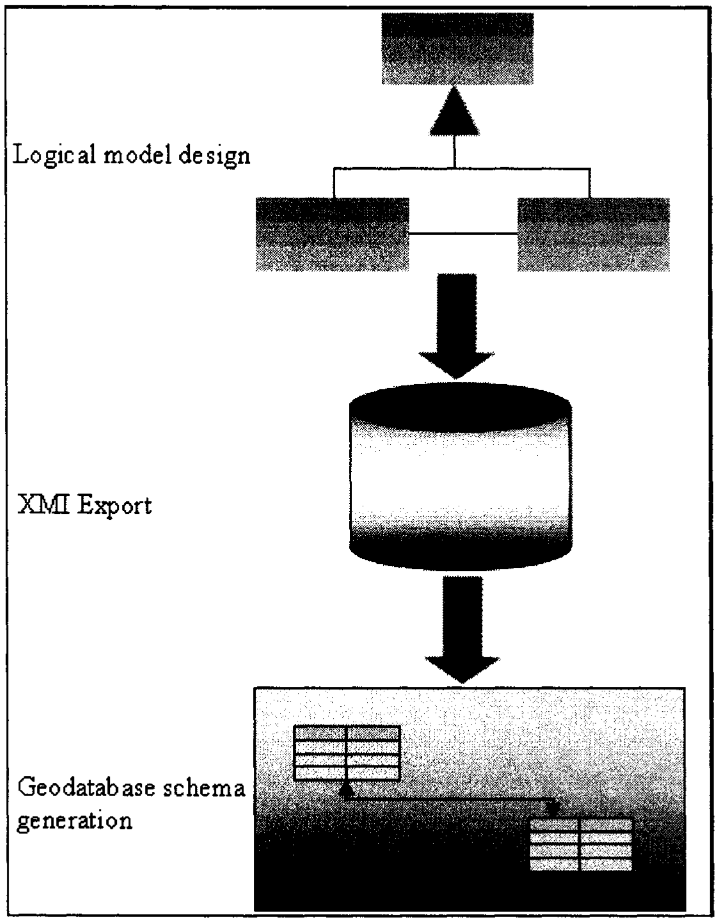


Figure 3.3 The framework of implementing the data model using ArcInfo (ESRI)

The logical model is designed in UML using Visio Professional (Microsoft), and then exported as XMI with XMI Export component. The physical model (schema) is finally generated in Geodatabase) (ESRI) (modified from Zeiler (1999)). Further information can be found in this chapter.

3.5.2 Geodatabase

Spatial and attribute aspects of geographic data have been traditionally stored separately. The georelational model, for instance, uses the principle that spatial data is stored in indexed binary files while attribute data are stored in INFO tables. The Geodatabase extends the georelational model and supports a model of topologically integrated object-oriented feature classes (Booth et al., 2002).

The Geodatabase, built on a technology framework known as ArcObjects, brings the physical representations of geographic features closer to their actual real world counterparts. It is possible, for example, to create mains, hydrants, and valves and to define a number of characteristics for each, such as fields, validation rules, relationships, and subtypes. The CASE tools (see a subsequent section of this Chapter) facilitate the creation of blueprints of the structure of the Geodatabase using the UML. With UML class diagrams, each Geodatabase element and the relationships among these elements can be represented and seen clearly.

3.5.3 Component Object Model (COM) (Microsoft)

Reusability and maintainability are important to any software development, which is a costly and time-consuming process. Reusability ensures the reuse of pieces of code, even in circumstances that the original developer did not foresee. Early resolutions to the first-generation object-oriented GIS technology involved the creation of class libraries, suffering from several limitations including difficulty in sharing binary components, problems of updating components without recompiling (ESRI, 2000; Zeiler, 2002). Component-based approaches are currently widely used by software engineers to counteract those problems. The key to the success of components is that they implement in a very practical way many of the object-oriented principles now commonly accepted.

COM is a Microsoft technology for software components (binary units of reusable code) to enable cross-software communication. The interfaces of COM are how COM object components communicate with each other. Since COM links software component with one another at the binary level, the reuse of software components, therefore, can be performed across any different languages and platforms which meet COM specifications. C, C++, Delphi, and Visual Basic are programming languages with COM interface bindings (Microsoft, 2004).

Technically, COM software is built using COM-aware components. Each COM component exposes its functionality through one or more interfaces. The different interfaces supported by a component are distinguished from each other using interface IDs. Access to components is made through the methods of the interfaces, allowing techniques such as inter-process or even inter-computer programming. All COM components must implement the standard IUnknown interface. Within ArcInfo, ArcMap, ArcCatalog, and ArcScene are based on ArcObjects, which are built using COM technology (ESRI, 2000; Zeiler, 2002).

3.5.4 Computer Aided Software Engineering (CASE) tools

The need for automated tools to help the software developer has been realized since the early days of writing software (Yu & Wright, 1997). In the beginning, the concentration was on program support tools such as compilers, assemblers, and macro processors. However, as computers became more powerful and the software grew larger and more complex, the range of support tools began to expand and the need for a broader notion of software development became apparent. The tools used to aid the development and maintenance of software are known as Computer Aided Software Engineering (CASE) tools.

UML, as described above in this chapter, is a modeling and specification language to specify, visualize, construct, and document the artifacts of an object-oriented software-intensive system under development (Bennett et al., 1999). The strong connection between object-oriented languages and UML is helpful in the construction phase of a system, and CASE tools can support this linkage through automatic generation of program code.

CASE tools can be used for managing database design and geodatabase implementation integrating with diverse data sources. In general, this approach allows the developer to model and edit the UML diagrams and then export the design to UML/XML that can be used to automatically generate database structure within the database management system. Since ArcInfo (ESRI) has very limited capabilities to make structural changes in a geodatabase, a CASE tool allows for more efficient database configuration and maintenance. After modeling the geodatabase and exporting it to XML, the code can be evaluated with the ESRI semantics checker before being imported into ArcCatalog. A log file is also available to document the process and any errors that occurred. Modeling with a CASE tool reduces the efforts during the development because linkage among disparate data sources, the relationships among classes, and validation can be automated. Resolving the problems of application development and maintenance of software can also reduce the maintenance costs for organizations.

In ArcInfo (ESRI), CASE tools help to create COM classes that implement the behavior of the custom features and database schemas in which custom feature properties are maintained. The CASE tools consist of two major activities: code generation and schema generation. Code generation is used to create the behavior, while schema generation is used to create schema in a geodatabase.

3.5.5 The logical data model

3.5.5.1 Design the logical model in Visio (Microsoft)

Visio (Microsoft) is a data reverse engineering tool which can diagram database systems, along with the ability to reverse engineer database schema.

ArcInfo (ESRI) organizes geographic data into a hierarchy of data objects. These data objects are stored in object classes, feature classes, and feature datasets.

- Object class: a collection of attribute data
- Feature class: a collection of features with the same type of geometry and the same attributes
- Feature dataset: a dataset consisting of a set of one or more FeatureClasses that share the same spatial reference
- Relationship classes: to define relationships between objects in ArcInfo

More detail can be found in Zeiler (1999) and Booth et al (2002).

3.5.5.2 ArcInfo UML Logical Model

ESRI provides the basic object-oriented model in UML that can be loaded into Visio (Microsoft) and extended to create custom models. The ArcInfo UML model diagram contains the object model required for using UML to model the spatial database. As illustrated in Figure 3.4, the object model has four packages: Logical View, ESRI Classes, ESRI Interfaces, and Workspace. These UML packages serve as directories where different parts of the entire object model are maintained. The key task in building a logical data model is to define precisely the set of objects of interest and to identify the relationships between them.

- The Logical View package: the root level and contains the other three packages
- The ESRI Classes package: to contain the portion of the GeoData Access

Components necessary to create object models (Booth et al., 2002). Classes in this package signify components that are used to access spatial databases. Feature classes and object classes in the custom object models will inherit from these classes

- The ESRI Interfaces package: to contain the definition of the interfaces implemented by the components shown in the existing ESRI Classes package
- The Workspace package: to be used to create the custom object and database designs and to serve as a container of spatial and non-spatial datasets such as feature classes, raster datasets, and tables

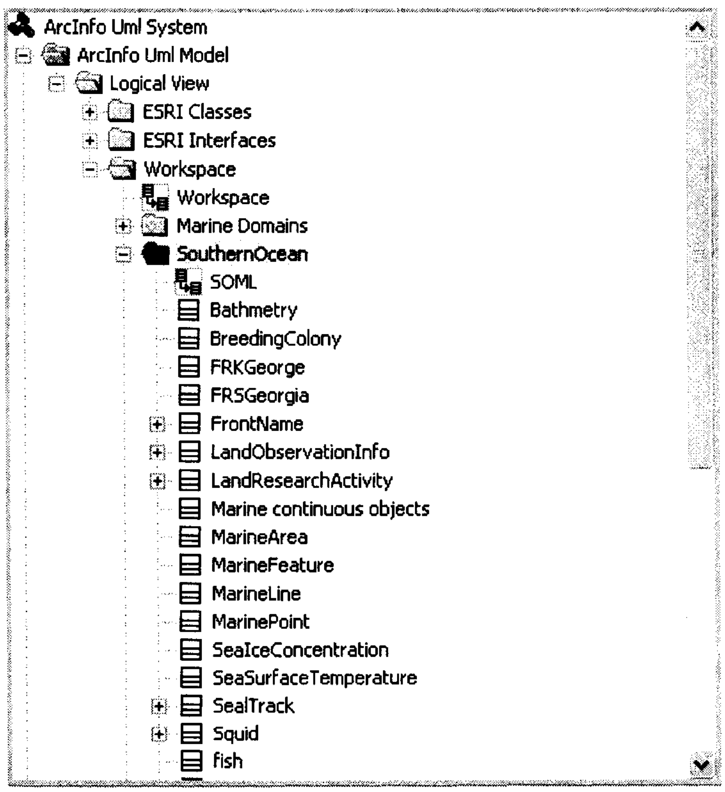


Figure 3.4 The logical data model designed in Visio (Microsoft)

This figure demonstrates part of the architecture of the logical data model (by the author).

3.5.5.3 Designing the feature class models and their relationships

Any custom object model must be created as the extensions of the ESRI data model. There are certain rules, as follows:

- Custom models are created under the ArcInfo "Workspace" package
- Classes, which store non-spatial (or attribute) data, are created as specializations of ESRI object class
- Custom feature classes, which store spatial data, are created as specializations of ESRI feature class. This allows the custom feature classes to inherit the behaviors of ESRI feature classes
- Feature datasets are created as packages under the ArcInfo "Workspace"

Figure 3.5 shows part of the custom feature classes in the logical model, which is also shown as an architecture view in Figure 3.4.

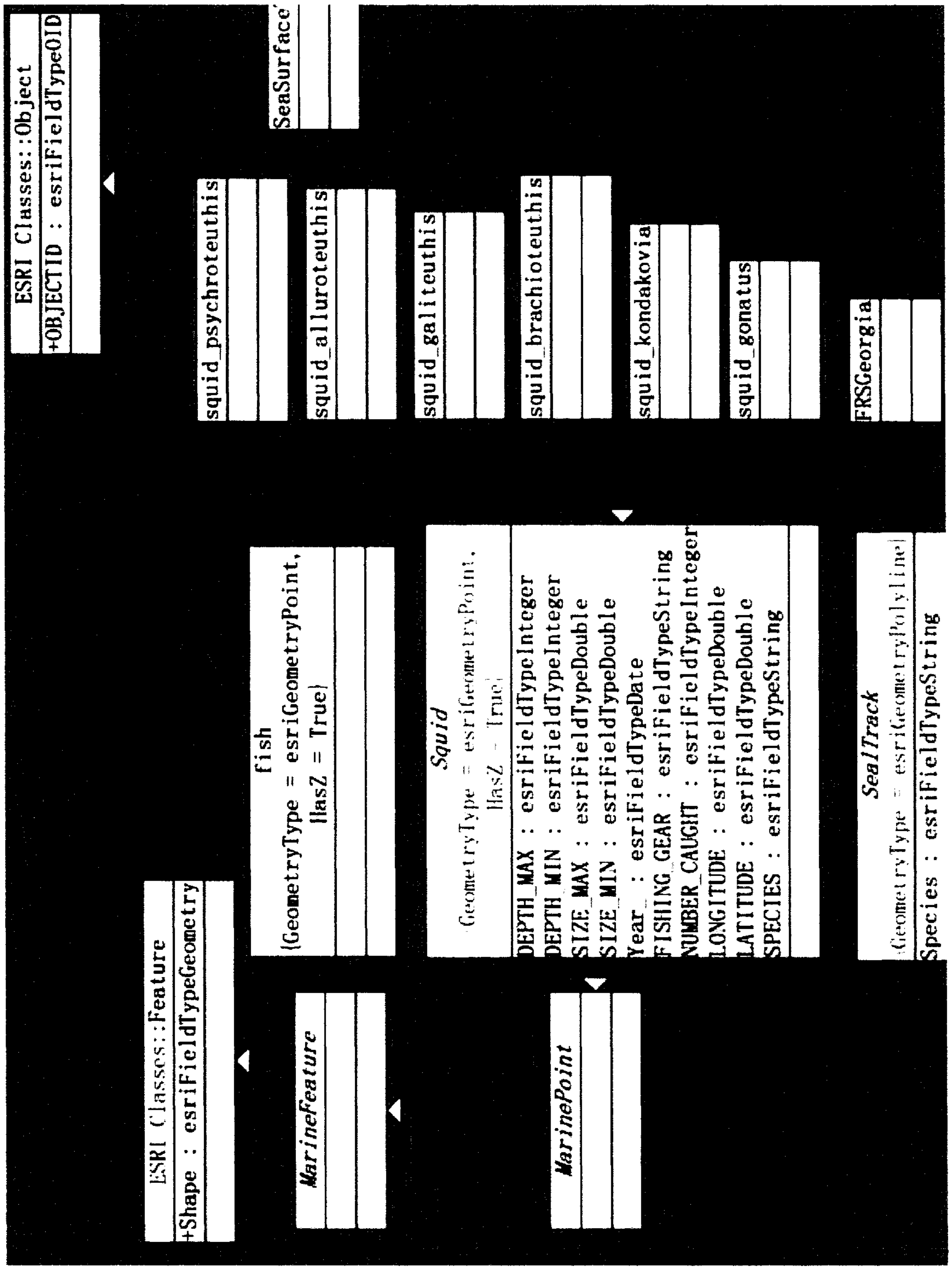


Figure 3.5 Part of the custom feature classes in the logical model

The complete model is in Appendix B.

3.5.5.4 Custom domains and subtypes

This research considers spatial and attribute domains. The spatial domain is used to describe the valid range of coordinates, including x, y, and z values. It limits the maximum spatial extent where the data can reach. The attribute domain is used to describe the properties of geographic features. Two kinds of attribute domain exist: coded value and range value. Care should be taken to choose proper values for both the spatial domain and the attribute domain before input since it cannot be changed once the database is established.

In the previous section, attention was paid to how all objects in one feature class share the same behaviors and attributes. Sometimes, however, there may be special cases. For example, the ocean fronts may share some attributes with one another; each species has its own code number. In such cases, subtypes can be used to distinguish one front from others.

The behaviors of any geographic features are designed using structured relationships, attribute domains, subtypes, validation rules, and default values within a geodatabase framework. These behaviors ensure a more accurate process in modeling the geographic world.

3.5.6 The physical data model

3.5.6.1 Exporting the UML model to XMI

XML (short for Extensible Markup Language) Metadata Interchange (XMI) is an Object Management Group (OMG) standard for exchanging metadata information via XML (Nantajeewarawat et al., 2004). XML is the World Wide Web Consortium (W3C) recommendation for creating special-purpose markup languages, the primary purpose of which is to facilitate the sharing of structured text and information across the Internet. The common use of XMI is as an interchange format for UML, although it can also be used for other purposes, such as defining, interchanging, manipulating and integrating XML data and objects (D'Ambrogio & Iazeolla, 2005). More information on the XMI standard can be found at <http://www.omg.org/technology/documents/formal/xmi.htm>.

In order to automatically translate information such as class, attribute, and operation names from a UML model that was created with Visio (Microsoft) into relational tables, a UML solution is needed. Since the UML solution in Visio (Microsoft) does not expose an object model application programming interface (API) where the elements of UML model can be accessed, the export functionality is provided through a separate dynamically linked library (DLL). XMI Export component (XMIExport.dll) can be invoked and the information contained in static structures (class), components, and deployment diagrams (node elements) can be exported to an XML file compliant with the XMI standard.

In most UML tools, XMI format is generally available either directly or indirectly as a textual means of data archival. However, XMI tags used by a given UML tool are sometimes unique to that tool. This makes information interchange and inter-tool model comparison a challenge.

3.5.6.2 Generating the geodatabase schema

If a system conforms to its specification, verification at the logical level is needed. Verification typically focuses on code inspections, so as to improve system quality and maintainability by ensuring that an implementation conforms to corporate standards (Cooper et al., 2004).

A spatial reference system is defined when generating the physical model. The Scientific Committee on Antarctic Research (SCAR) polar stereographic projection is used. This projection is conformal, in order to preserve the shape of local features. The Y-axis is parallel to the Greenwich meridian; the origin is at the South Pole with the latitude 71°S as the standard parallel in the true scale; the horizontal datum is WGS 84 (SCAR, 2003). Part of the physical model in ArcInfo is illustrated in Figure 3.6.

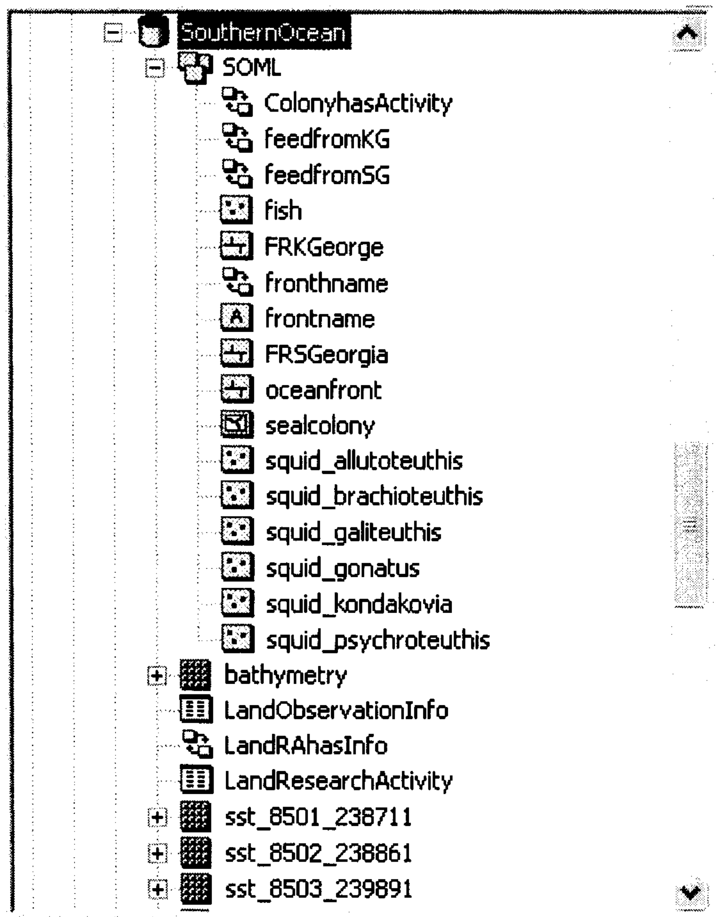


Figure 3.6 Part of the physical model

Sealcolony is represented as POLYGON; FRKGeorge (feeding routes from King George Is.) is represented as LINE; squid_* are represented as POINT; FeedfromKG is a relationship class linking colony and FRKGeorge; frontname is annotation class composite to oceanfront; the lower ones with raster-like icons refer to sea surface temperature data(by the author).

3.6 Review

The object-oriented marine data model is a systematic framework specially designed for creating a marine geospatial ecological database. This model holds one feature dataset (containing several feature classes, relationship classes, and one annotation class), and many raster datasets (containing monthly data of sea ice concentration, sea surface temperature ranging from 1980 to 1994, and southern ocean bathymetry).

In the next chapter, the advantages and disadvantages of the object-oriented data model will be tested through populating it with real marine data collected and modified from several data sources. One application of this marine data model in the southern ocean marine scientific community, as well as in the Cybercartographic Atlas of Antarctica project, is presented and discussed.

CHAPTER 4 THE APPLICATION OF THE OBJECT-ORIENTED MARINE DATA MODEL IN CYBERCARTOGRAPHY

4.1 Overview

Chapter 3 illustrated the phases of building an object-oriented data model. A geographic data model is a framework for synthesizing data from diverse sources to represent a geographic phenomenon. The marine data model designed in Chapter 3 describes the main marine features in the southern ocean: ocean fronts, sea ice concentration, sea surface temperature, foraging routes of seals from King George Island and South Georgia Island, and six species of squid favored by southern elephant seals as their main food. This chapter aims to test the characteristics of the object-oriented approach to geographic data modeling and to present the application of the marine data model in the Cybercartographic Atlas of Antarctica Project.

4.2 Populating the model with data

The strength of a geographic data model lies in providing a framework for scientists to gather and handle geographically referenced data. In other words, further analysis can be done through populating the model with data. A good recognition of data sources may help to understand the potency and limitations of potential research results. In this section, data sources related to the case study of this research are discussed. The study area is the southern ocean around the Antarctic Peninsula Region.

4.2.1 Biological data of southern elephant seals

Biological data observed on land were collected from the Centro Nacional Patagonico (CNP) (see section 3.3.2 in Chapter 3). These data include population of adult seals, adult male, adult female, adult sex proportion, male weaning mass, female weaning mass, total born, dead pups, and diet (six species of squid – Alluroteuthis, Brachioteuthis, Galiteuthis, Gonatus, Kondakovia, and Psychroteuthis). Biological data were originally accumulated based on field work.

4.2.2 Squid

The data sets on squid were obtained from National Environmental Research Council of British Antarctic Survey (Xavier, 1997; Xavier et al, 1999). There are 906 geographic positions of the six species of squid in total. The longitude, latitude, maximum and minimum depth, maximum and minimum size, the date caught, the number caught, and the fishing gear used is all recorded. Since squid are very hard to capture, data sets on squid are highly sporadic and contain large uncertainties.

4.2.3 Seal colonies and seal tracks

The locational data on the colonies of southern elephant seals were obtained from relevant publications and organizations such as Laws (1994), Vergani et al (2004), and

SCAR (2003). The high-fidelity seal track data starting from South Georgia Island are obtained from McConnell & Fedak (1996). The low-fidelity seal track data starting from King George Island were obtained from the Alfred Wegener Institute (Bornemann et al, 2000). Tracking the movements of southern elephant seals was carried out through tagging the seals with special transmitters, which transmit data to satellites.

4.2.4 Satellite-derived sea surface temperature

Sea surface temperature (SST) data were obtained from the National Oceanic and Atmospheric Administration (NOAA) Advanced Very-High Resolution Radiometer (AVHRR) Oceans Pathfinder (version 5.0). This product only consists of SST from 1985 to 2003. Data are available through the Physical Oceanography Distributed Active Archive Center Ocean ESIP Tool (POET) official website: http://podaac.jpl.nasa.gov/poet. The range of valid pixel values is from 16 to 248. The minimum value of SST is mapped into the pixel value 16, while the maximum value is mapped to 248. The relationship between SST and the Pixel value is:

$T = 2 + (P-16)(30/232)$,

where P is the pixel value and T is the value (°C) of SST (Raskin, 2004).

Sea surface temperature data from 1981 to 1984 were derived from the AVHRR Multi-Channel Sea Surface Temperature (MCSST) data set collected from the Jet Propulsion Laboratory (JPL, PO.DAAC), with a spatial resolution of 18 x 18 km. Further documents can be found either through the official website or the PO.DAAC

User Service Office at:

http://podaac.jpl.nasa.gov/poet.

4.2.5 Satellite-derived sea ice concentration

Satellites provide the characteristics of Antarctic sea ice in a large spatial context. Sea ice concentration data were collected from the NASA Goddard Space Flight Center (GSFC). The data were produced using the Nimbus-7 Scanning Multi-channel Microwave Radiometer (SMMR) and Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave/Imager (SSM/I) through the NASA Team algorithm (Cavalieri et al., 2005). The SMMR is a ten channel instrument which can receive both horizontally and vertically polarized radiation and obtain ocean circulation parameters such as sea surface temperatures, wind stress, sea ice extent, and sea ice concentration. The data were generated from brightness temperatures at a grid cell size of 25x25 km, stored in a binary format with one byte per pixel. The sea ice concentration data values are packed into byte format by multiplying the original sea ice concentration floating-point values (ranging from 0 to 1) by a scaling factor of 250. That is, a sea ice concentration value of 0% is mapped into a byte value of 0 and a value of 100% is mapped into a value of 250. Complete details can be accessed either through National Snow and Ice Data Center (NSIDC) User Services or at the official website: http://nsidc.org/data/nsidc-0051.html. For this thesis research, sea ice concentration data were converted from the flat binary format to GeoTIFF format using the remote sensing tool Geomatica 9.0 (PCI Geomatics).

4.2.6 Southern ocean bathymetry

The bathymetric data of the southern ocean around the Antarctic Peninsula Region were obtained from 2-Minute Gridded Global Relief (National Geophysical Data Center). The seafloor topographic data northward of 72°S were derived from satellite altimetry observations combined with carefully, quality-assured shipboard echo-sounding measurements (Smith & Sandwell, 1997). The seafloor topographic data southward of 72°S are from the Naval Oceanographic Office's (NAVOCEANO) Digital Bathymetric Data Base Variable Resolution (DBDBV) (version 4.1). Further information can be accessed either through NOAA Satellites and Information or through the official website: http://www.ngdc.noaa.gov/mgg/fliers/01mgg04.html.

The data mentioned above are respectively input into the geodatabase SouthernOcean created in Chapter 3. Seal colonies, feeding routes from King George Island and South Georgia Island, southern ocean fronts, and squid were imported as shapefiles. Since the spatial representation of the source data on seal tracks starting from South Georgia Island collected from McConnell & Fedak (1996) were POINT, which was different from what was predefined, the class was redefined in ArcInfo (ESRI). The relationshipclass (one term in object-oriented modeling) *feedfromSG* was thus automatically removed because it was designed to link the class *sealcolony* and *FRSGeorgia*. Sea surface temperature data were imported as GeoTIFF raster files. Since reducing the size of the database increases the speed of rendering in ArcInfo (ESRI) applications such as ArcScene and ArcMap, sea ice concentration data were preprocessed and converted into shapefiles before being importing the database. All the populated data share the same spatial reference system – SCAR South Polar Stereographic Projection, which was predefined during the modeling process from the logical level to the physical level.

4.3 Testing the model

This section demonstrates some results that test the characteristics of object-orientation in geographic data modeling discussed in the previous three chapters - identity, properties, behavior, and multi-dimension, which enhance the level of abstraction in a way to match our perception of the real world.

That the "object" in the concept of object-orientation can be the "object" in reality is close to natural human way of thinking. In the real world, most geographic features are different from one another. This is reflected in the object-oriented data model: each object has its own identifier (ObjectID) which is unique within one class. In the case study, the ObjectID within each feature class (shapefile) (such as FRSGeorgia) or raster (SST, _85_238711) is unique. This characteristic is essential for identifying or conducting queries in a geographic database.

"Property" is an important element in the concept of object-orientation. It is also significant in describing or defining any geographic features. In the object-oriented data model, specific properties of each geographic feature are thoroughly portrayed. Figure 4.1 illustrates the properties of shapefile *squid_allutoteuthis*. According to Table 3.3, the species name itself (Field Name, in Figure 4.1) is one of the properties of the shapefile *squid_allutoteuthis*. This model allows the specification of the details of these properties (Field Properties, in Figure 4.1). For example, the field Shape has its sub-properties (e.g. Geometry type, Grid, Spatial Reference, and so on). In addition, each field property can be associated with a suitable data type (e.g. Geometry, Integer, Double, Text, and Boolean).

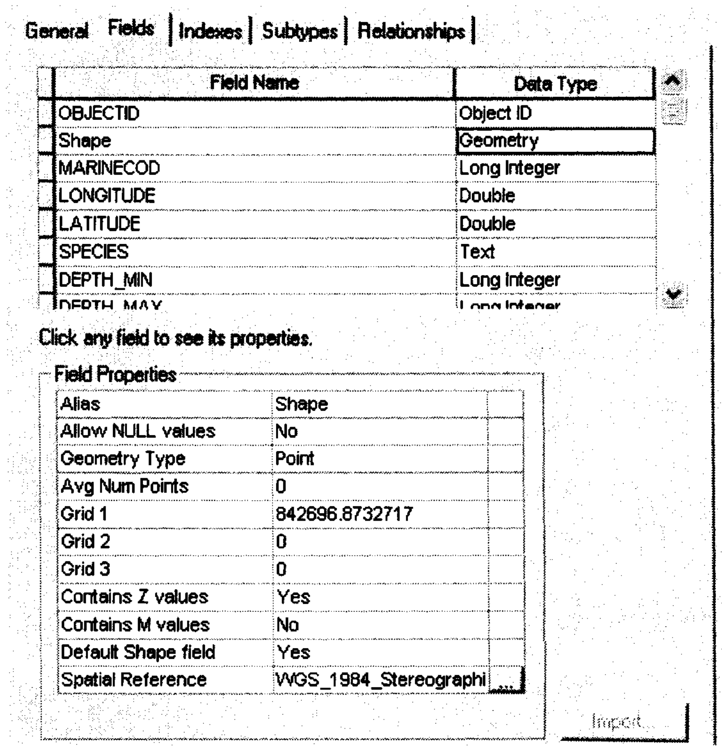


Figure 4.1 Properties of squid_allutoteuthis through object-oriented date modeling

The feature squid_allutoteuthis has several properties (Fields, in the figure), and each property has its sub-properties. For example, the field Shape has type, grid, z values, and spatial reference as its sub-properties (by the author).

“Behavior” distinguishes the object-oriented paradigm from the previous approaches, although the term behavior from the perspective of programming may be different from what is used to describe the geographic features. Object-orientation does describe some behaviors of geographic features through the rules such as constraints, relationships, and subtypes. For example, the depth that squid live in the ocean will not exceed 11,033m (the lowest point on the Earth). A predefined constraint for data input may avoid potential input errors. This constraint in the object-oriented model (in Table 3.4) will restrict any input value for the depth of squid in the future to no more than 11033. Relationships are another rule to describe geographic behaviors. The relationship behavior is based on the predefined relationshipclass. A relationshipclass is an association between two object classes, with one as the origin class and the other as the destination class. Figure 4.2 demonstrates the predefined relationship among the seal colony, seal’s feeding routes, and research activity in the seal colony. According to the graphic interface, the 6 routes (red lines) start from King George Island Colony, the closest colony (blue dot) to the Antarctic continent.

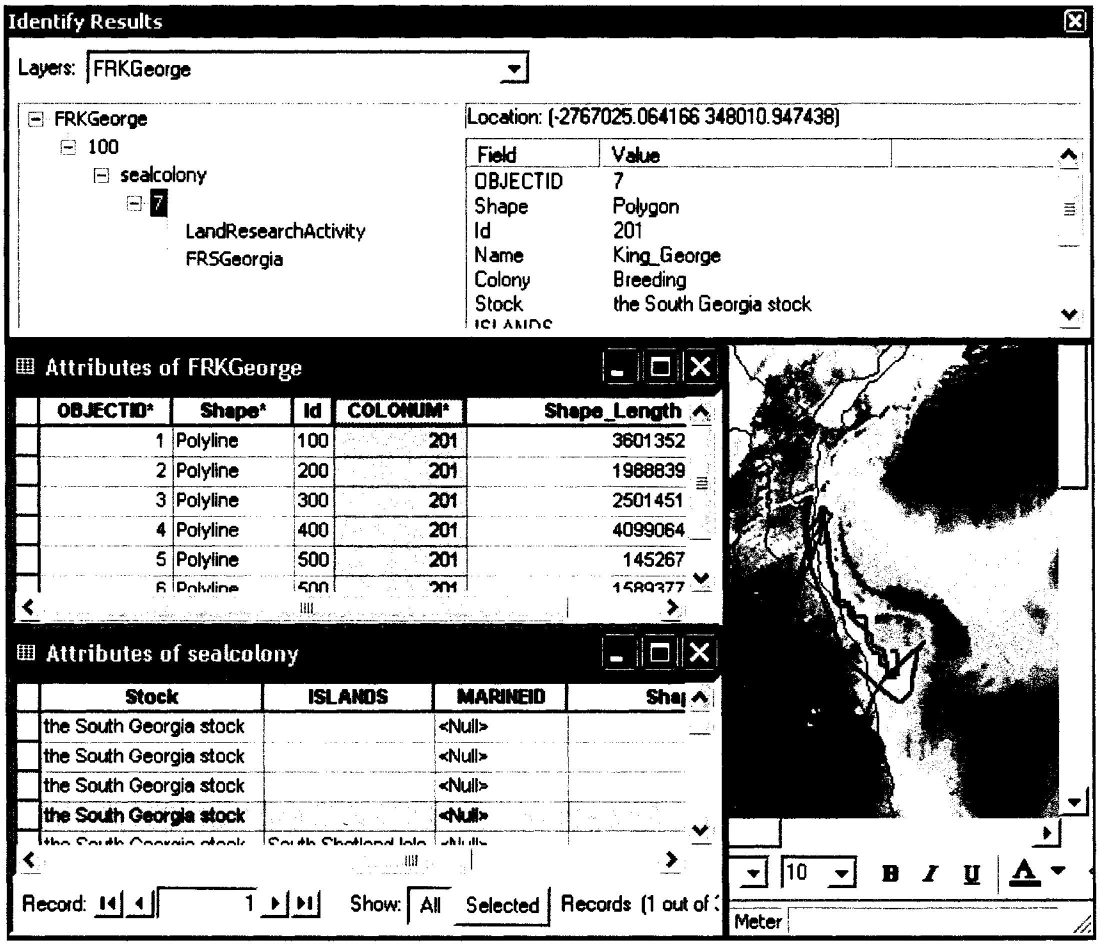


Figure 4.2 The relationship among feeding routes, seal colonies, and research activity

According to the graphic interface, the 6 routes (red lines) start from King George Island Colony, the closest colony (blue dot) to the Antarctic continent among all colonies (by the author).

Geographic features essentially are three-dimensional (X, Y, Z). Traditionally, the third dimension was not really considered in the process of data modeling. In the object-oriented model, the third dimension is represented from the beginning in the conceptual model. Although the third dimension can be displayed by setting a certain field, a previous embedded Z value may allow more rapid 3D rendering. This is especially useful when the data need to be rendered several times. Figure 4.3 illustrates the 3D representations of 6 species of squid in the southern ocean according to Z-values embedded in their geometries.

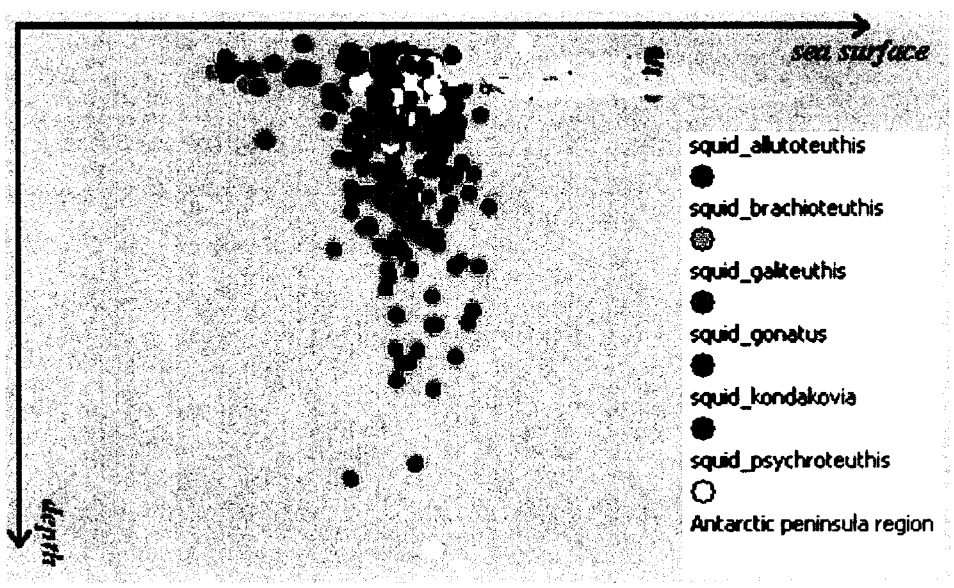


Figure 4.3 3D representations of distribution of squid in the southern ocean according to Z-values embedded in their geometries

4.4 The application in the Cybercartography Atlas of Antarctica Project

Cartography is the science of making maps, primarily functioning as visualization tools for geospatial data that are obtained from measurement and stored in a database. Today, it is rare for maps to be made without mapping software. And the depiction methods have also changed from the static to the interactive, from single media to multiple, from two-dimension to the three/four-dimensional, and from data held by local PCs to data distributed over the Internet. Searching for the best way for the users to understand geospatial data and information is a key purpose for cartography. Cybercartography is a new concept to generalize recent developments in cartography and put the theoretic construct into practice. This section aims to demonstrate the application of the marine model built in the Chapter 3 to the Cybercartographic Atlas of Antarctica Project. Relevant technologies are also discussed.

4.4.1 Spatial analysis using GIS

Many technologies have been involved in the study of natural facts (science). One important application of technology in environmental science is to deal with scientific data, including collecting, processing, analyzing, and publishing data. As discussed in Chapter 1, people always want more data and information to gain knowledge, so the tools that are used to find patterns within data are very important. The tools used by marine ecologists vary in terms of the topics they investigate (USEPA; Smith, 2004). In the spatial context, GIS, together with remote sensing have become common techniques over the last two decades. Among the functions of GIS, spatial analysis may be the main one favored by marine scientists. The rest of this section demonstrates how GIS can be used for spatial analysis. The spatial variations within the foraging routes, the food availability, sea ice concentration, and sea surface temperature data are used for demonstration.

4.4.1.1 The foraging route of southern elephant seals

Southern elephant seals have a circumpolar distribution, mainly breeding on sub-Antarctic islands such as South Georgia, Macquarie and Marion (Stanganelli & Vergani, 2000). Southern elephant seals spend around 90% of their time in sea water (McConnell & Fedak, 1996), with the rest on land. The annual life cycle on land mainly comprises two periods, breeding and moulting (Piatkowski et al., 2002). Adult seals always fast during this time, so they have to store enough energy both for themselves and for feeding their pups. An examination of the foraging behaviour at sea may be helpful to understand the potential food availability in the southern ocean, which is related to the potential energy to be passed onto the next generation of seals.

Foraging behaviour is central to understanding the recent changes of populations of southern elephant seals (Hindell et al., 1991a). Knowledge of the foraging behaviour of marine mammals has been gained by means of tracking the movements of migratory animals. Although the number of relevant techniques has increased, it is still very difficult to obtain the information, especially for the southern elephant seals, because of the large distance they travel at sea. The history of the techniques used to monitor the movements of marine mammals is reviewed by Hindell et al. (1991b), who also investigate the foraging areas of adult southern elephant seals at Macquarie Island (54°30′S, 157°E) using time-depth-temperature recorders. Since then, the foraging behaviors of southern elephant seals from several main breeding colonies - South Georgia, Patagonia, Marion Island, and King George Island, have been sampled respectively by McConnell et al. (1992), McConnell & Fedak (1996), Campagna et al. (1998), Jonker & Bester (1998), and Bornemann et al. (2000).

Information on the movements of southern elephant seals is gained from the data collected by the satellite, which receives transmission from the satellite data loggers or time-depth recorders tagged on the seals. Collecting the data usually requires a long time period as the seals travel a large distance at sea after moulting or breeding. The techniques employed in processing these data are very important to the research results. GIS may provide a good method of processing these data since the movement of the seals can be geographically referenced. In the case of tracking from South Georgia (McConnell & Fedak, 1996), 6970 points, with several parameters (latitude, longitude, dive depth, observation date and time), were recorded. In a GIS, the migration routes can be indicated by different colors for each seal, and direction of the migration routes can be symbolized using arrows. Figure 4.4 illustrates the foraging routes of ten elephant seals from South Georgia Island. The raw numerical data are from McConnell & Fedak (1996). As the part (within the red rectangle) of Figure 4.4 indicates, the original data are in the format of geographic points (6970), which also reflect the travel distance of elephant seals. But are there any other patterns within the data? For example, which route is for a certain seal? How do they migrate? Do all of them return to the starting location? GIS may help to provide such kinds of answers. The points (locations) visualized in the red rectangle of Figure 4.4 are converted into lines (paths) by integrating one spatial ecology analysis script named Animal Movement (Beyer, 2004). Using Arrow at End Line symbols and different colors, the movements of elephant seals and their relationships with other species (e.g. squid) can be examined. Figure 4.4 uses the SCAR polar stereographic projection, with a map scale 1: 50, 000, 000. The power of integrating GIS with the study of animal movements is discussed in detail by Hooge et al. (2000).

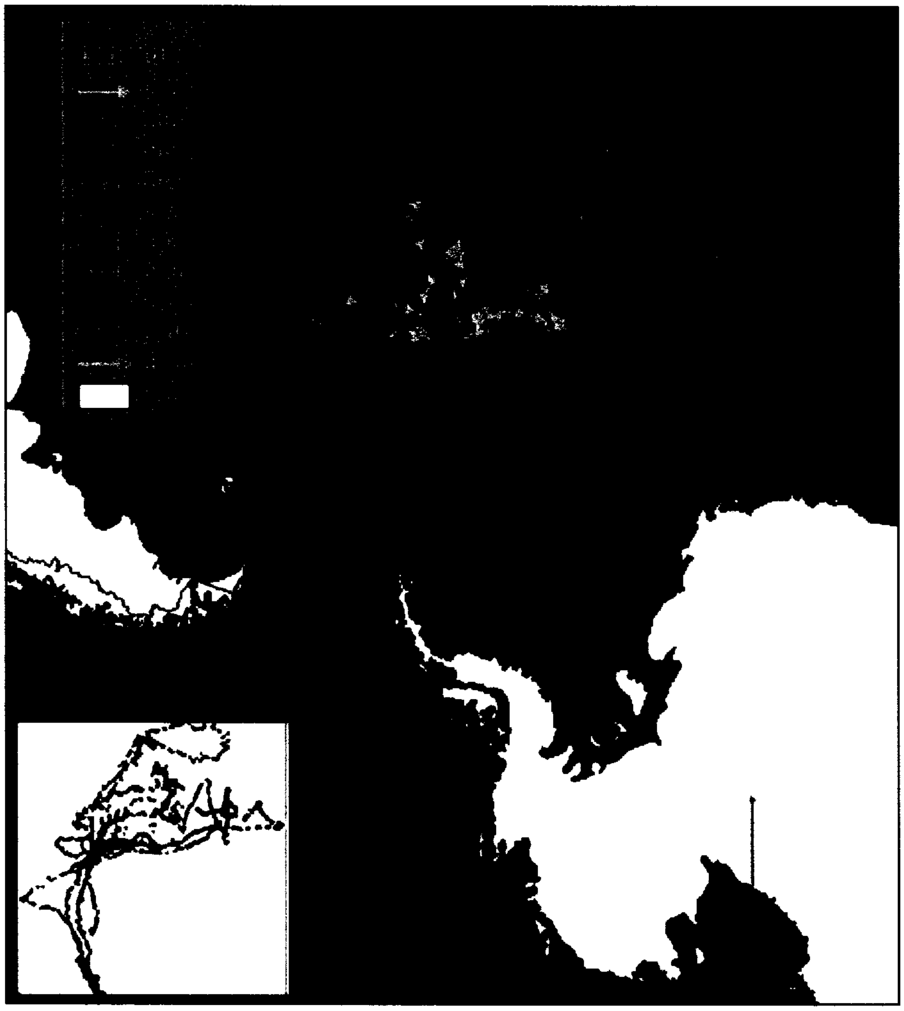


Figure 4.4 Spatial patterns in seal track data revealed by ArcInfo (ESRI)
Projection: SCAR polar stereographic. Map Scale: 1:50, 000, 000.

4.4.1.2 Food availability for southern elephant seals

Food availability is a fundamental factor in the definition of any organism's ecological niche (Hindell et al., 1991a). In the case of southern ocean elephant seals, squid accounts for approximately 70% of the prey, according to the analysis of diet (Rodhouse et al., 1992; Daneri et al., 2000; Piatkowski et al., 2002). The importance of squid in the diet of elephant seals drives biologists to search for the distribution of squid in the southern ocean, especially for the previously mentioned six species. Figure 4.5 demonstrates the spatial distribution of these six species of squid in the southern ocean around the Antarctic Peninsula Region. The raw numerical data were collected from Xavier et al (1999) and input into ArcInfo. The data about the ocean fronts in the southern ocean are also integrated into the map. At this scale, the distribution of geographic locations of squid matches well the locations of the southern ocean fronts and the foraging routes of elephant seals.

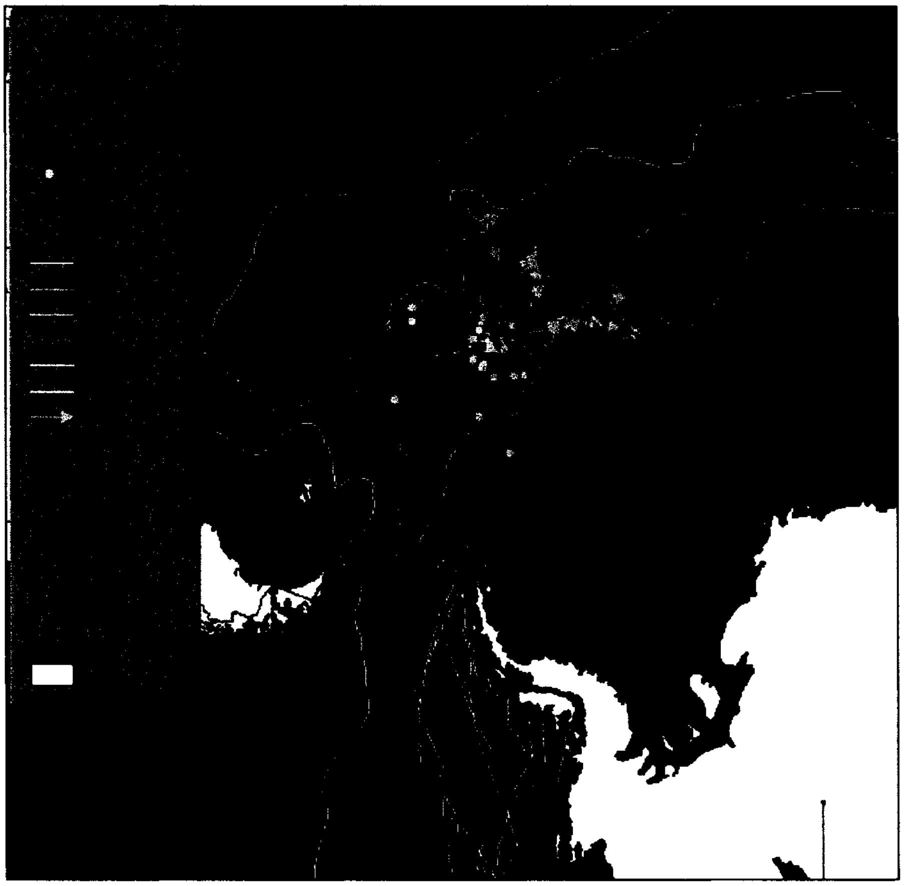


Figure 4.5 Spatial distribution of six species of squid in the southern ocean around the Antarctic Peninsula Region

The raw numerical data are collected from Xavier et al (1999) and input into ArcInfo (ESRI). The data about the ocean fronts in the southern ocean are also integrated into the map. At this scale, the distribution of geographic locations of squid well matches the locations of the southern ocean fronts and the foraging routes of elephant seals (by the author).

Although there are biases in the sampling of squid, information from the squid data indicates the potential routes of the elephant seals. For example, the data within the foraging routes from South Georgia Island indicate the overall mean maximum dive depth is 359 m and the deepest dive is 1595 m in depth (McConnell & Fedak, 1996). This fits well with the result of frequency analysis of the species Kondakowski Longimana using ArcInfo (ESRI), which suggests the maximum depth of this species is 309.83m and the deepest dive is 1010m (see Figure 4.6). Such analyses may address some research questions. For example, why are the elephant seals most interested in the squid species Psychroteuthis Glacialis, whose spatial distribution is not as wide as Galiteuthis Glacialis?

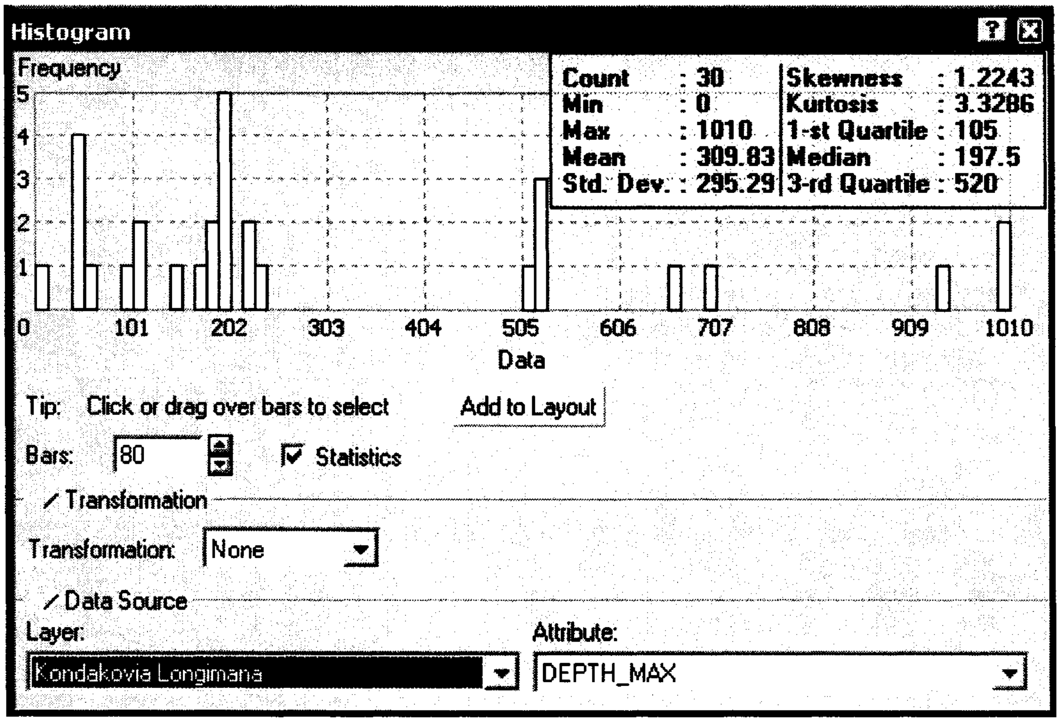


Figure 4.6 One of the results of analysis of squid data

The mean maximum depth of this species Kondakowski Longimana is 309.83m and the deepest dive is 1010m. The information fits with the research result by McConnell & Fedak (1996), who found the overall mean maximum dive depth of southern elephant seals is 359 m and the deepest dive is 1595 m in depth (by the author).

4.4.1.3 The variation within ocean surface properties

Satellite image data may provide rich information on the large extent of the earth system. In the southern ocean, the variation of ocean surface properties (such as sea surface temperature, ocean color, and bathymetry) may be analyzed to examine the potential distribution of marine species. Bradshaw et al (2004) investigate at-sea distribution of female southern elephant seals through analysis of the variation in southern ocean surface properties, including sea surface temperature, sea surface height anomaly, ocean color, bathymetry, sea ice concentration, and their associated gradients.

Sea surface temperature anomalies (SSTA) in the 'Nino 3.4' region are used to define the occurrence of ENSO events. Monthly data, which were obtained from the Climate Prediction Center (NOAA), represented departures from the 1971-2000 climatic mean values (Smith & Reynolds, 1998). The average of the monthly SSTA is calculated as the Index of Anomaly Strength (IAS). If the value of IAS is:

within the range (0.5, +infinity), then the year is defined as an EI Nino year;

within the range (-0.5, 0.5), then the year is defined as a Normal year;

within the range (-infinity, -0.5), then the year is defined as La Nina year (Vergani et al, 2004). According to the GIS processing results, there are significant variations in the Antarctic sea ice concentration and sea surface temperature in the southern ocean around the Antarctic Peninsula Region from 1980 to 1994, when the values of seal population parameters also varied. Figure 4.7 demonstrates the variation of Antarctic sea ice concentration (left) and sea surface temperature (right) separately in a Normal year (1986), a La Nina year (1984), and an EI Nino year (1982). In Figure 4.7 (left), the range of sea ice concentration becomes higher from bright yellow (0.0%) to dark yellow (100%). Figure 4.8 illustrates the statistical analysis of sea ice concentration based on pixel values using Geostatistical Analyst package of ArcInfo. In Figure 4.7 (right), the value of the sea temperature becomes higher from the dark purple (2 °C) to the bright blue (11.37°C).

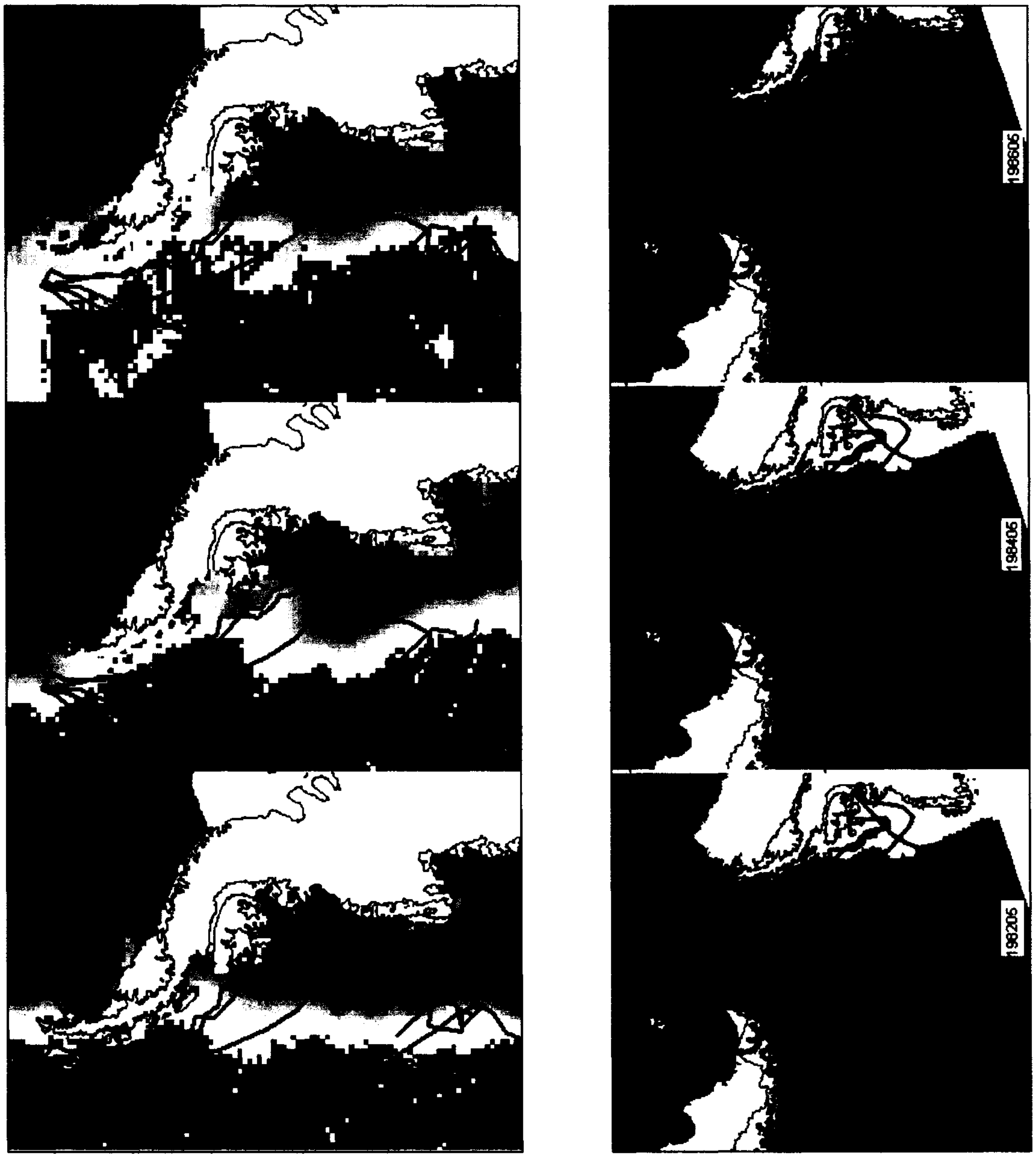


Figure 4.7 Variation of Antarctic sea ice concentration (left) and sea surface temperature (right) in Normal year (1986), La Nina year (1984), and EI Nino year (1982)

In Figure 4.7 (left), the range of sea ice concentration becomes higher from the bright yellow (0.0%) to dark yellow (100%). In Figure 4.7 (right), the value of the sea temperature becomes higher from the dark purple (2 °C) to the bright blue (11.37°C). The left and right figures are in the same projection with different map scale (by the author).

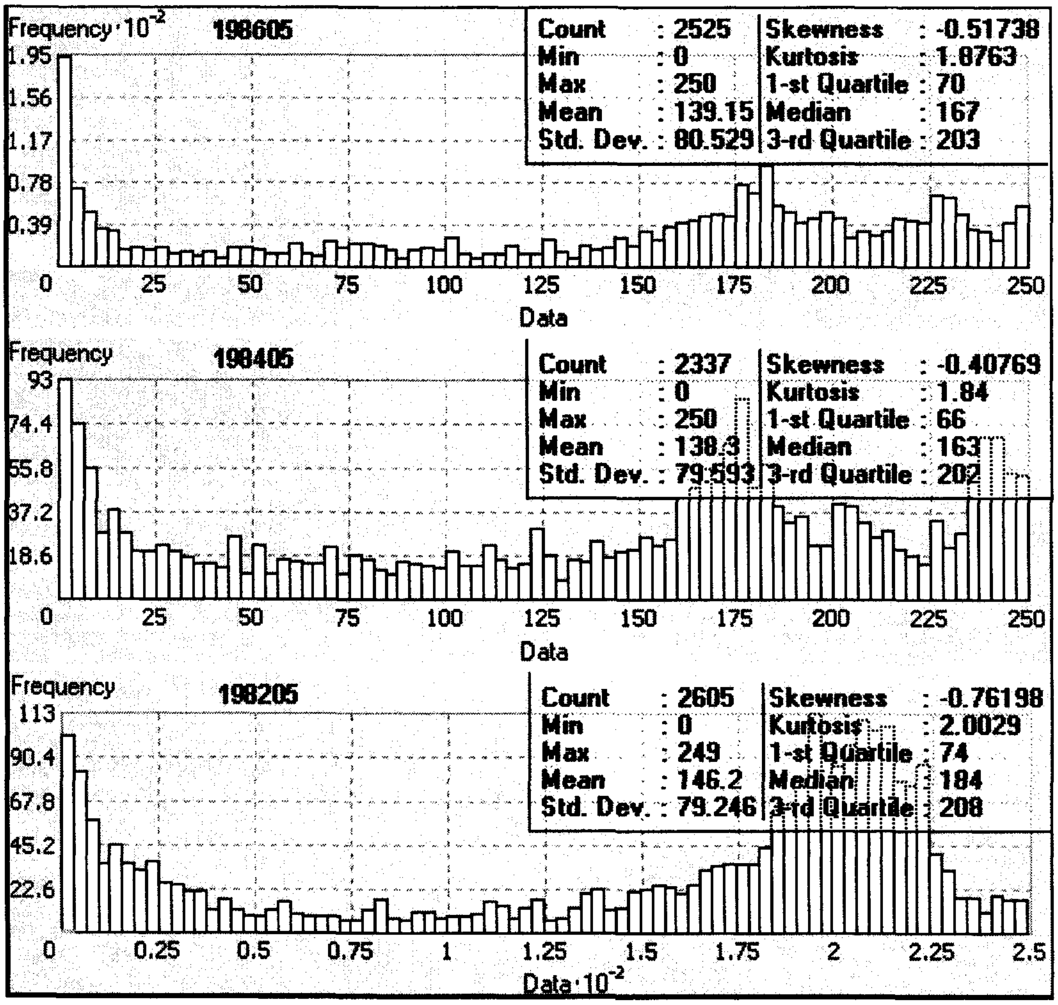


Figure 4.8 Result from statistical analysis of sea ice concentration based on pixel values using ArcInfo (ESRI)

It reveals the significant variation among the same month of three years: Normal year (1986), El Nino year (1982), and La Nina year (1984) (by the author).

GIS is considered as a common-use technical tool in the environmental sciences, although it has limitations. What sets GIS apart from other tools is its explicit and accurate geographic referencing of data and linking attribute data to geographic referenced locations (Paul et al., 2003). In the case of southern elephant seals, identifying the geographic location of both species and physical environmental elements is even more important since the elephant seals move around a large area in the southern ocean (Hindell et al., 1991a & 1991b).

4.4.2 Computer animation as a geographic information user interface

Static maps as main geographic information user interfaces can be used to gain insight into spatial relations and patterns. When the volume of spatial data to be displayed becomes massive, designing an imaginative approach is necessary to keep the map readable (Kraak, 1999). This section discusses the use of animation in displaying geographic data and information.

In the case of southern elephant seals, looking into spatial variation within sea ice concentration data and SST data is useful to examine the potential relationships between these two important physical marine environmental parameters and the population status of elephant seals. For non-specialists, a computer animation interface may provide them with a straightforward information package to assist in understanding the changes of sea ice concentration and SST from month to month, and from year to year.

To accomplish this, the physical data model was populated with data, and processed in ArcInfo using functions such as reprojection, data classification, and symbolization. The processed data were thereafter exported into different maps in JPEG format, and then imported into Flash (Macromedia) as different keyframes. An ActionScript was coded to interact with the animated buttons by the mouse. Figure 4.9 is a screenshot of the computer animation of the monthly spatial variation of Antarctic sea ice concentration and corresponding sea surface temperature in the same region for the time series of 1985 to 1986. The three color ramps demonstrate the variations within bathymetry, sea ice concentration, and sea surface temperature. For example, in comparison to the extent of sea ice in January (austral summer), sea ice extent in August (austral winter) grows several times larger. At the same time, the coverage of the purple color in August, which represents lower temperature, is much more extensive than that in January, especially around the Weddell Sea. The growth rate of sea ice is also different between different areas. For example, the growth rate of sea ice around the Weddell Sea is much faster than other regions, which may indicate special oceanographic conditions.

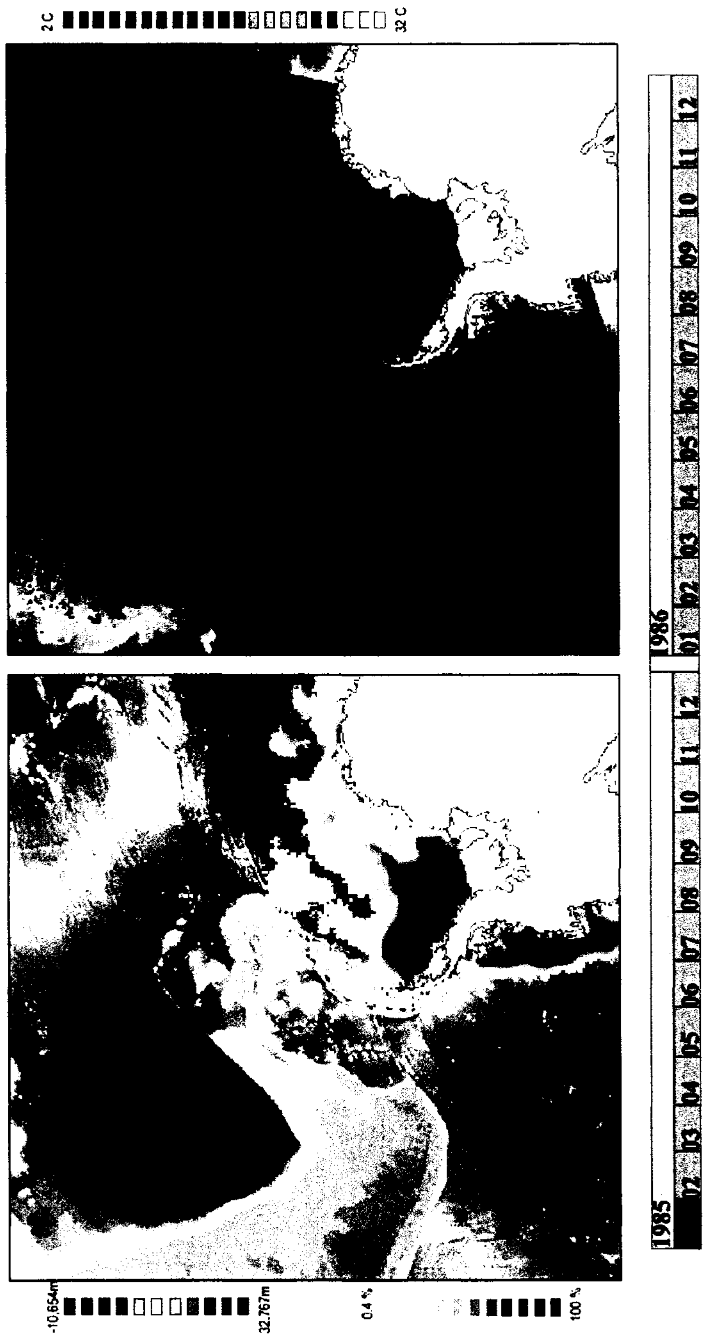


Figure 4.9 A screenshot of the computer animation about the monthly-based spatial variation of Antarctic sea ice concentration in corresponding to sea surface temperature in the same region at the time series from 1985 to 1986

The three color ramps the variations within bathymetry, sea ice concentration, and sea surface temperature (by the author).

It may not be necessary to use geographic data modeling to create a simple computer cartographic animation. In order to design a complex one, however, geographic data modeling is essential. That is, the complex spatial data must be first projected and processed using a GIS tool to ensure accurate spatial representation. In the case of the southern ocean ecosystem, it is not possible to ensure the exact geographic location of several elements through 15 years without previous geographic data modeling.

4.4.3 Access to geographic information over the Internet

The application of the Internet in geographic information science is a ground-breaking research, resulting in many research issues such as data sharing, copyright, and dissemination security.

One of the purposes of the Cybercartographic Atlas of Antarctica Project is to enable a variety of users to explore geospatial information on the remote Antarctic Region. The Internet perhaps is currently the best choice to accomplish this. At present, the popularity of the Internet is increasing and Internet use is becoming an integral part of our society. Hypertexts and hypermedia have made the Internet a powerful means for people to access and exchange information.

In order to deliver geospatial information to many concurrent users, both within an organization and externally on the web, a powerful Internet mapping solution is needed. Although commercial software packages exist and open source packages are available, web mapping packages are not as popular as the desktop GIS. In 2000, there were only 30 web mapping solutions provided by different GIS vendors (Peng & Tsou, 2003). Four of the most popular are ArcIMS (ESRI), MapGuide (AutoDesk), WebMap Server (Geomedia), and MapXtreme (MapInfo). In the open source GIS community, there are four others: GeoServer (GNU GPL), GIServer (Goncalves), MapServer (UMN, DM Solutions), and QuickWMS (Goncalves).

Although commercial packages are currently preferred by many corporations because of the low risk they provide (e.g. regular update of versions, simplified technical support, and more software documentation), interest in open source software packages is increasing rapidly because they are free, both in terms of price and access to source code (Sondheim, 2004). The MapServer (UMN) system, for example, includes MapScript that allows free and popular scripting languages such as PHP, Perl, and Python to access the MapServer C API. It has been used by several projects (e.g. King George GIS, Northern Shrimp Interactive map, Iowa Tornado Database, Acorn MapServer).

The implementation of an on-line atlas requires not only network infrastructures (e.g. high-speed communication channels) to move geospatial data but also software architecture to provide interactive functions and applications. In the case of southern elephant seals, the marine data model is populated with ecological data, processed in ArcInfo, and then exported as shapefiles and GeoTIFF files embedded in the SCAR polar stereographic projection. In order to publish data through the Internet, all data have to be moved onto a server with installed web mapping server packages. MapServer - Chameleon (UMN - DM Solutions) was used in this thesis research.

Figure 4.10 is a screenshot of the graphic user interface (GUI) of the web mapping application for the southern ocean ecosystem. The GUI handles users' input from the keyboard or mouse. The request from the users is sent to the server, which connects the data, then processes the data based on the request. The responses are sent back to the users either as a graphical representation or potentially other forms (e.g. URL, numerical data, geospatial data formats). The users can read the graphic results or go to other web pages as the response indicates.

In Figure 4.10, 14 layers (FRSGeorgia, FRKGeorge, Allutoteuthis, Brachioteuthis, Galiteuthis, Gonatus, Kongdakovia, Psychroteuthis, socenfront, Grids, sealcolony, Antarctica, SST-Seaice, and Bathymetry) are defined in the mapfile, the basic configuration mechanism for MapServer. The bathymetry layer is designed as the first one in the mapfile, while the grid layer is designed as the last. The reason behind this is that, like a stack, the top layer will be displayed as the bottom one in the GUI presentation. Since the bathymetry layer is a raster image, displaying it as the top layer in the GUI will hinder any other layer. The grid is added to depict the lines of latitude and longitude. Designing the grid as the top of the GUI will help users understand the geographic location of any elements they are interested in.

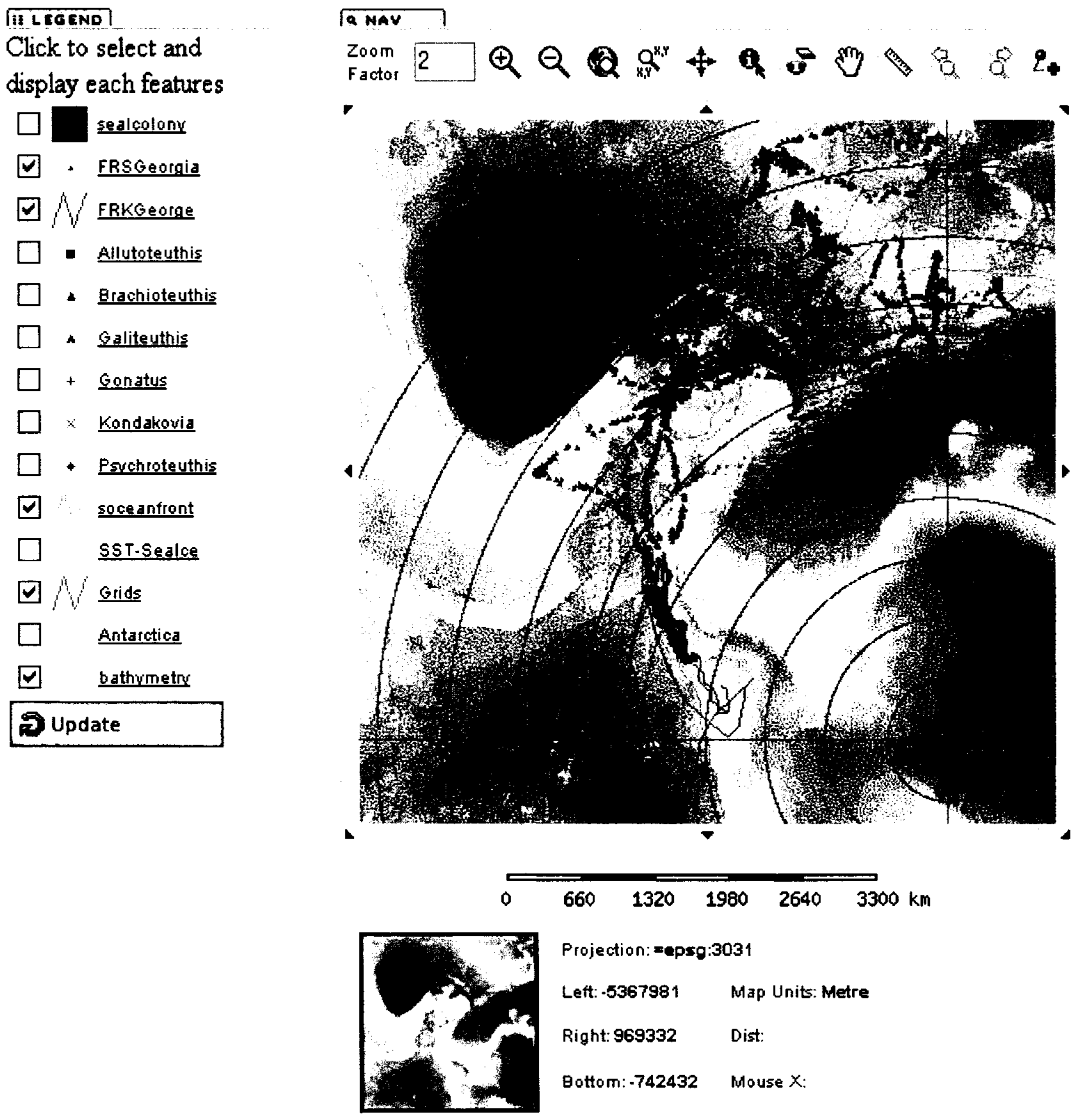


Figure 4.10 A screenshot of the graphic user interface (GUI) of interactive web mapping application in the southern ocean ecosystem

The GUI handles users' input from the keyboard or mouse. The request from the users will be sent to the server, which connects and the data, then processes the data based on the request. The responses will be sent back to the users either in the representation of graphics or in others (e.g. URL, numerical data, geospatial data formats). The users can read the graphic results or go to other web pages as the response indicates (by the author).

Among the 14 layers, bathymetry and SST-Seaice (the integration of sea surface temperature and sea ice concentration) are designed as raster format. The variation of the sea floor topography is represented using a color ramp. The graphic interface, sometimes, might be not enough for gaining an in-depth insight. Providing additional information may be a useful approach. In the Internet mapping application, this can be achieved through hyperlinks. In this case, each layer is linked to another page which provides additional information (e.g. the complete layer title, abstract, web connection, data source URL, metadata URL, Spatial Reference System, and Queryability). Figure 4.11 demonstrates the additional information related to the layer FRSGeorgia.

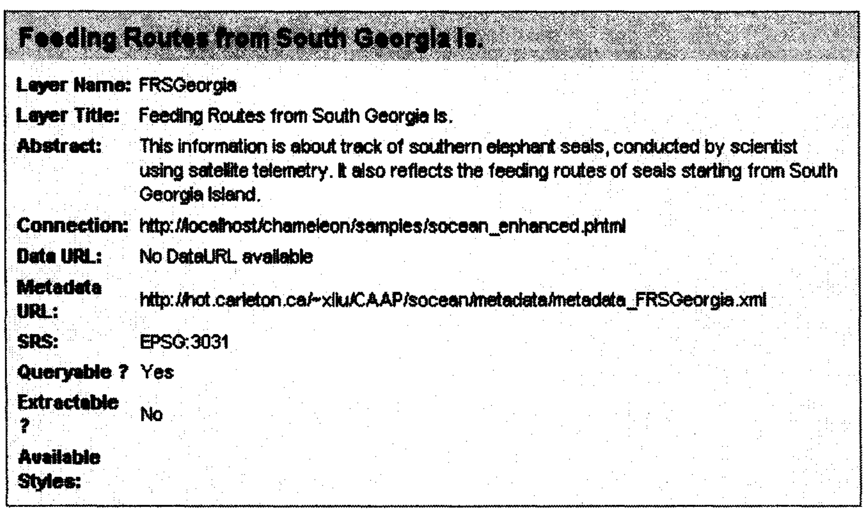


Figure 4.11 Additional information related to the layer FRSGeorgia

The layer SST-Seaice is a special one because of its complexity. As discussed in section 4.5.1 in this chapter, 180 months of Antarctic sea ice concentration and sea surface temperature data are involved. Web navigation problems will occur, if all the data are input as layers of this GUI. One solution is, according to the additional information the layer provides, to make a link to another page – the page showing the spatial and temporal variations within the sea ice concentration and sea surface temperature data using computer animation techniques. All other 12 layers in the GUI are vector-based. The vector-based files are usually much smaller (as discussed in Chapter 3). The smaller the size of files, the faster they can be downloaded from the server over the Internet.

The web mapping application inherits many of the functions the desktop GIS. It is interactive, identifiable, browseable, and searchable. With the navigation buttons (e.g. Zoom In/Out, Full Extent, Recenter, Identify Feature, and Pan, Measure) above the map, users can interactively and quickly browse geographic information from different perspectives. Users can take a closer look at a particular area and point at features to find out more about them. Obtaining a regional perspective can be accomplished by zooming in, which is one of the main functions for storing more details. For example, the movements of elephant seals from King George Island are currently depicted as some small lines on the map. In order to take a closer look, the navigation button Zoom In can be used to make a rectangle around the certain area. The map server will send the response based on the coordinates of the rectangle and show all the information within it. Further information about the movements can be obtained through identifying the features (lines in this case) and clicking the layer name (FRKGeorge) in the Legend.

Since the map can be moved around at a display size larger than that of the screen, a reference map may be needed to provide users with information on the whole map area. For example, where is the center of the map? What part of the map is currently showed? The users also can decide the part of the map they are interested in from the reference map by means of the two buttons (Zoom In and Pan) below. Some other tradition elements (scale bar, projection, original coordinates, and map unit) for any maps can be found on the web mapping interface as well.

The successful application of distributing and transforming geospatial data and information benefits from the object-oriented modeling technology, which has been developed over the past 20 years (Orfali and Harkey, 1997). Java / JavaScript, PHP, and Perl, which are commonly used in web mapping applications, are all object-oriented programming (or script) languages. For example, if a user points at an on-line picture with a mouse and then clicks the mouse button, the location of the mouse pointer on the computer screen will be transmitted as picture coordinates. This can be achieved through JavaScript programming. This facility is suitable for cartographic maps, which have geographic coordinates. It enables users to select and submit geographic locations interactively. By adopting the object-oriented modeling technology, the web map server can handle rich and complex requests from the users. An in-depth discussion is beyond the scope of this thesis but can be found in Peng & Tsou (2003).

Currently most web mapping applications are 2D-based. The spatial extent of the objects in the real world, however, is three-dimensional (Chapter 3). In order to demonstrate the more realistic view of the geographic elements, interactive 3D mapping has been researched for several years in such fields as visualization, GIS, cartography, remote sensing, film, multimedia, and engineering. The Virtual Reality Modeling Language (VRML), which stems from the object-oriented paradigm, is the first easily distributable means for interactive 3D integrating and mapping spatial objects. Similar to Hypertext Markup Language (HTML) in many ways but different in the end results, VRML is a programming technology for use on the Internet, describing 3D scenes. VRML files, created by writing text files, can be loaded into any browser with a VRML plug-in (Lemay et al., 1996). 3D scenes can be moved around in a three-dimensional cyber space within a VRML browser.

From the perspective of the implementation phase, writing large amounts of code is not desirable, especially for complex geographic features. The effort of designing a building in 3D space for city planners may be less than that of creating 100 geographic points with geographic coordinates. To make a compound 3D geographic scene, object-oriented data modeling significantly reduces the labor. In the case of this research, 904 geographic points on 6 species of squid needed to be coded. The object-oriented data model designed previously was used and further processed in ArcInfo (ESRI). The values of X/Y are the geographic coordinates in SCAR polar stereographic projection. The values of Z are the pre-embedded Z multiplied by 1000 for exaggeration in order to illustrate the difference of the depth. Few codes are modified within the VRML file.

Figure 4.12 is a screenshot of the 3D view of squid designed above. Appearing in the Internet Explorer (Microsoft) (or others like Netscape Navigator and Mozilla) with the Cortana VRML Client (Parallel Graphics) plug-in, the spatial distribution of the main prey of the southern elephant seals, 6 species of squid, can be explored virtually. The plane image in the browser shares Antarctic topography, southern ocean bathymetry, and horizontal spatial distribution of squid. The navigation buttons (distributed on the left and bottom of the cyber space) are the tools to be utilized. For instance, the button *Turn* can be used to move the scene around in any direction, and the button *Plan* can be used to take a closer or farther look at the scene.

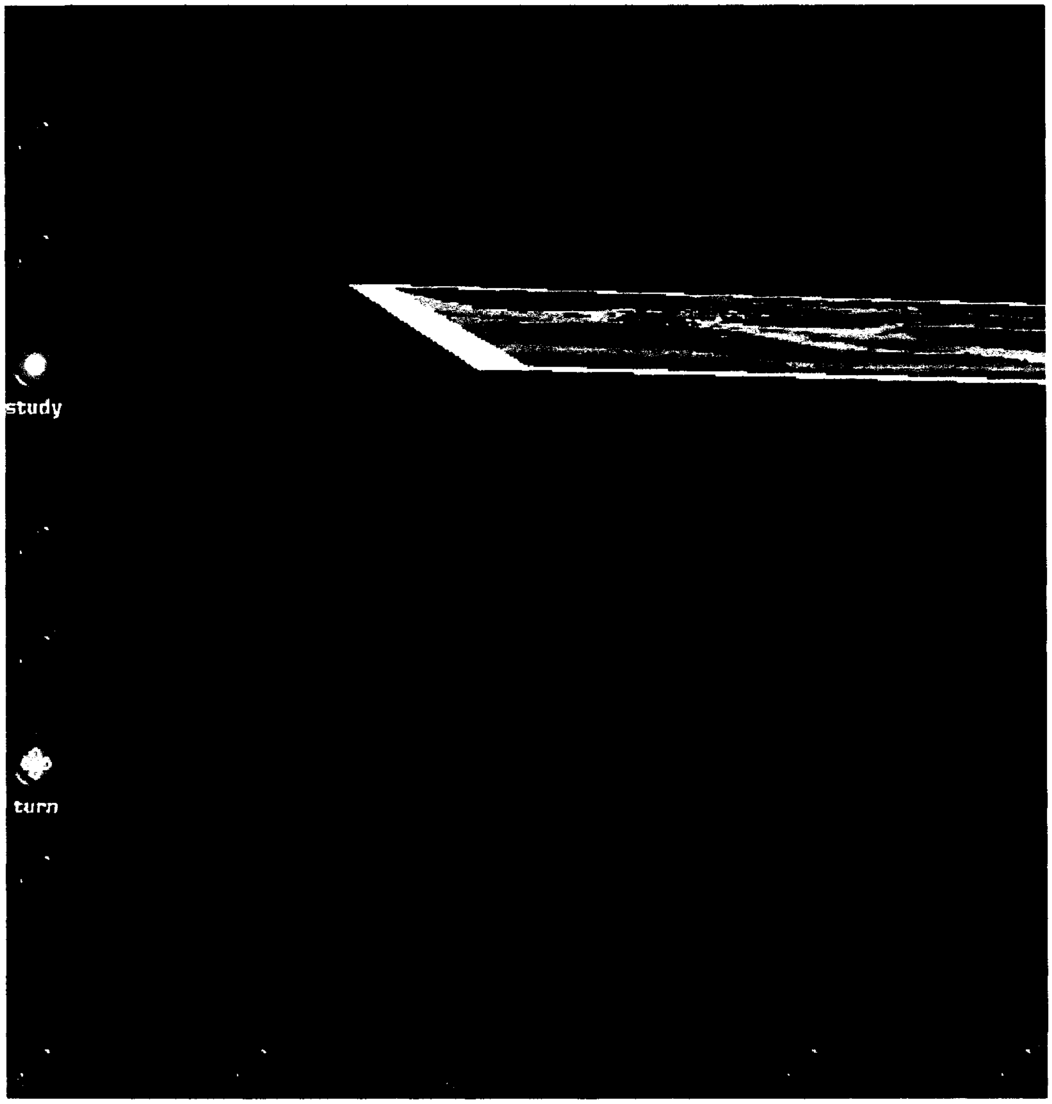


Figure 4.12 A screenshot of the 3D view (VRML) of squid distribution in the southern ocean

Appearing in the Internet Explorer (Microsoft) (or others like Netscape Navigator and Mozilla) with the Cortana VRML Client (Parallel Graphics) plug-in, the spatial distribution of the main prey of the southern elephant seals - 6 species of squid can be explored virtually. The flat image in the browser is about the Antarctic topography, southern ocean bathymetry, and horizontal spatial distribution of squid. The navigation buttons (distributed on the left and bottom of the cyber space) are the tools to be utilized. For instance, the button Turn can be used to move the scene around in any direction, and the button Plan can be used to take a close or distant look at the scene (by the author).

4.6 Review

This chapter tested the importance of the object-oriented approach to geographic data modeling which were discussed in Chapter 3. It also demonstrated the use of the object-oriented model in the integration of spatial data on southern ocean ecological dynamics. Through object-oriented modeling, marine environments, including biotic and abiotic factors of southern elephant seals, were integrated, mapped, and finally displayed to the end users using different graphical interfaces. A comprehensive review, including the contributions and limitations, is presented in the next chapter.

CHAPTER 5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

A geographic data model is an abstraction of the real world from a certain perspective, in order to integrate a variety of spatial data and address certain spatial issues. Geographic data modeling typically refers to the process of making spatial variations discrete and converting geographic reality in spatial dimensions into a finite number of computer database records. Searching for an optimal way for spatial data integration such as Cybercartography is the research focus of this thesis, and three factors were considered during this process of modeling: the degree of approximation to reality, the use in applications, and the efficiency of designing a model. This research investigated how the object-oriented approach helped to model geographic features and examined the importance of data modeling in geography-related fields.

A primary objective of this thesis was a discussion of the principles, strategies and mechanisms of object-orientation for data modeling in the spatial context. Issues related to the object-oriented approach to geographic data modeling have been discussed at various levels of abstraction in the previous chapters, ranging from a general overview to considerations of conceptual modeling and implementation. A central element of these discussions was the advantages of object-orientation in geographic data modeling for spatial data integration. Although the object-oriented paradigm is not new and may not always be the best choice for individual modeling applications, object-oriented modeling does bring several advantages over other approaches:

- Object-orientation enhances the abstraction level of the computer "view" of geographic reality

  Conceptually, the object-oriented approach is a more intuitive way to model spatial objects. It forces the computer to view the real world as individual objects with three major components - unique identity, properties, and dynamic behavior. The combination of these three components enhances the human view of reality: every object is primarily distinctive and active, with the rich and varied properties that may sometimes be shared with one another. The object-oriented approach integrates the advantages of its predecessors – the raster model, the vector model and the relational model. It not only captures the characteristics of the raster model and the vector model in representing the geometry of geographic features, but also inherits the functions of the relational model in revealing the relationships among spatial phenomena. The object-oriented paradigm has a large impact on the development of geographic information processing tools, which also makes the modeling process more flexible.

- The object-oriented approach increases the degree of efficiency in the implementation phase

  In practice, the object-oriented approach makes the process of geographic data modeling more efficient. Inheritance, for example, reduces the effort required in designing and managing geographic elements. Encapsulation allows the implementation details of one object to be hidden from others because the object contains data and method. This means that the implementation of the object can be changed without affecting others, allowing the implementation to be more flexible, and reducing complexity overall.

The use of CASE tools in object-oriented modeling automates the generation of program code. Since UML has a strong linkage with the object-oriented language, geographic data modeling can be processed in CASE tools using UML, a graphical and expressive visual language. In addition, the similar "appearances" of the models from the conceptual level, then to the logical level, and finally to the physical level helps to detect potential errors during the course of planning the projects.

The efficiency, flexibility, and automation of the object-oriented approach lessen the information loss from one level to the next in the process of geographic data modeling.

The importance of object-oriented modeling in spatial data integration such as Cybercartography should not be ignored when discussing issues of geographic data modeling. This research demonstrates a successful application of object-oriented modeling in the integration of marine spatial data for analysis, data management, and data transformation. In the application domain of marine GIS, as noted by Barlett (2000), an important lesson learned from collective experience is the importance of scrupulous data modeling before trying to implement a geographic database. The marine object-oriented data model created in Chapter 3 can serve as a "fill-in-the-blank" framework for relevant scientists to integrate and synthesize their marine geospatial data in the future. The marine data model, populated with data, can be used for further spatial analysis using GIS tools.

Another key contribution of the object-oriented approach for spatial data integration was demonstrated in the graphic display of spatial data. Visualization has been an essential process in data integration for geographic applications. One of the key mechanisms in the display of spatial data through 2D computer animation, 3D virtual reality, and the Internet is the object-oriented paradigm, which allows those processes to be more understandable and more efficient.

5.2 Limitations and recommendations

Although the research results revealed that object-orientation was a promising means to deal with geographic data modeling for complex spatial data integration, the object-oriented design and implementation described here is by no means the only way to model real world objects. A number of limitations were realized and discussed as follows.

Similar to many other applications of computer techniques, the limitations of object-oriented data modeling are largely argued from an implementation perspective. Take the object identity as an example. Although it can be changed manually, the automatic generation of objectID only makes the ID unique within the same class. The uniqueness of the objectID can only be achieved through manually revising the initial values. The developers of modeling techniques in the future may try to find a way to automate the generation of ObjectID for the whole model. It will help, especially for any complex applications that include a large volume of complicated data, to greatly reduce the effort of planning projects.

The methods in connecting the logical model to the physical model need to be improved. Raster data is an important data source in addressing spatial issues. With the current object-oriented GIS such as ArcInfo (ESRI), it is hard to directly transform raster data from the logical model to the physical model. This creates some problems such as the increase of the workload of data modeling and the reduced integrity of spatial data.

The traditional problems of vector-based GIS in dealing with spatial and statistical analysis still exist in object-oriented GIS. Lacking functions to conduct powerful analysis is a big barrier that hinders many GIS from being widely used in solving environmental problems. This problem limits the utility of GIS to visualize the past more than the future. It is recommended that the object-oriented paradigm be used to provide one solution using the "behavior", which was primarily proposed to address the dynamic of the real object.

Interoperable web mapping technology, which is an important tool for integrating and distributing spatial data, still needs to be advanced. The Open Geospatial Consortium (OGC) Web Map Service (WMS) specification offers a standard client-server interaction protocol that each map server implements as a common interface for accepting requests and returning responses. There is still difficulty in accessing the available OGC web map servers over the Internet. Take the Internet web mapping case discussed in section 4.5.2 (Chapter 4) as an example: King George Island is an important breeding site for the layer sealcolony, and its geographic data can be found on the Internet from a scientific database (the KGIS Project, Germany). A direct use of this data for the current project is ideal because the data would be automatically updated when they are changed by the data provider over the Internet. This failed in this research because of the difference in the two map projections. The map projection of King George Island data used by the KGIS Project is WGS/UTM zone 21, while the projection used by this research is EPSG 3031 (SCAR South Polar Stereographic). The future specifications may be able to integrate geographic data in various projects in real time (i.e., through on-the-fly reprojection). The other shortcoming of current web mapping technology is that, even if the geographic data in the remote servers can be accessed and visualized, the output of data on the client-side is unqueryable because of data type issues. Technology to retrieve the data in the remote server(s) just like retrieving the data in the local PC should be considered, in order to enhance the interoperability of geospatial data.

Spatial data integration through data modeling is essential to solve dynamic spatial problems. Investigations showed that the object-oriented approach benefited this process in many ways. Future success of spatial data integration through object-oriented data modeling relies on resolving the problems described above, increasing the robustness, efficiency, and ease of data transformation in the modeling procedures.

REFERENCES

AGI - Association for Geographic Information (2004). GIS Dictionary http://www.geo.ed.ac.uk/agidexe/term?1327 (cited Mar. 2005).

Arbib, M. (2004). Behavior.
< http://pespmc1.vub.ac.be/ASC/BEHAVIOR.html> (cited Sept 2004)

Alvaro, M. (1995). A GIS Data Model Prototype. ESRI International User Conference Proceedings. US: ESRI.

Becker, S.A. (2000). Perspective and Abstraction. The Journal of Conceptual Modeling (17), 1-9.

Bennett, S., Skelton, J., and Lunn, K. (1999). UML (Eds). Italy: LEGO.

Berry, J.K. (1996). Object-oriented technology and its GIS expressions. The Compiler: 14 (3), 16-19.

Beyer, H. L. (2004). Hawth's Analysis Tools for ArcGIS.
<http://www.spatialecology.com/htools> (cited Mar. 2005).

Bian, L. (2000). Object-oriented representation for modeling mobile objects in an aquatic environment. International Journal of Geographical Information Science 14 (7), 603-623.

Billah, M. (2001). Prospect and efficiency of GIS data models in planning applications: a case study on Khulna city. Bachelor Thesis (Khulna).
< http://masumbillah.tripod.com/thesis/> (cited Mar. 2005).

Blaha, M., and Premerlani, W. (1998). Object-oriented modeling and design for database applications. New Jersey: Prentice Hall.

Blongewicz, M.J. (2004). Sediment budget and coastline evolution. ESRI User Conference (San Diego).

Bonham-Carter, G. F. (1994). Geographic Information Systems for Geoscientists: Modeling with GIS (pp. 25-50). New York: Pergamon.

Bonner, W.N. (1985). Birds and mammals – Antarctic seals, In W.N. Bnner, and D.W. Walton (Eds), Antarctica (Key environments) (pp. 202-222). UK: Pergamon Press Ltd.

Bonner, W.N., and Walton, D.W.H. (1985). Marine habitats – Introduction. In Bnner, W.N. & D.W.Walton, Antarctica (Key environments) (Eds, 133-153). UK: Pergamon Press.

Booth, B., Crosier, S., Clark, J., and ManDonald, A. (2002). "Building geodatabases with CASE tools." Building a Geodatabase (Eds). USA: Environmental Systems Research Institute.

Borges, K.A.V., Laender, A.H.F., and Jr, C.A.D. (1999). Spatial data integrity constraints in object-oriented geographic data modeling, ACM GIS'99, 1-6.

Bornemann, H., Kreyscher, M., Ramdohr, S., Martin, T., Carlini, A., Sellmann, L., and Plotz, J. (2000). Southern elephant seal movements and Antarctic sea ice. Antarctic Science 12 (1), 3-15.

Bradshaw C.J.A., Higgins, J., Micael, K.J., Wotherspoon, S.J., and Hindell, M.A. (2004). At-sea distribution of female southern elephant seals relative to variation in ocean surface properties. Journal of Marine Science (61), 1064-1027.

Bradshaw, C.J.A., Hindell, M.A., Michael, K.L., and Sumner, M.D. (2002). The optimal spatial scale for the analysis of elephant seal foraging as determined by geo-location in relation to sea surface temperatures. Journal of Marine Science (59), 770-781.

Burrough, P.A. (1986). Principles of geographic information systems for land resources assessment (p182). Oxford University Press.

Campagna, C., Quintana, F., Le Boeuf, B.J., Blackwell, S., and Crocker, D.E. (1998). Diving behavior and foraging ecology of female southern elephant seals from Patagonia. Aquatic Mammal (24), 1-11.

Cartwright, W., Peterson, M.P., and Gartner, G. (1999). Multimedia Cartography. Berlin: Springer-Verlag.

Carvalho, A., Fitzpatrick, K., and Hudson, R. (2004). Advancement of an enterprise coastal GIS in Martin County, Florida. ESRI User Conference (San Diego).

Cavalieri, D., Parkinson, C., Gloerson, P., and Zwally, H.J. (2005). Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I passive microwave data (1979 – 1994). USA: National Snow and Ice Data Center. Digital media and CD-ROM.

Chen, P.P. (1976). The Entity-Relationship Model: toward a unified view of data.

Communications of the ACM, 13(6), 377-387.

Chorley, R.J., and Kennedy, B.A. (1971). Physical geography-a system approach (pp. 1). London: Prentice-Hall.

Christensen, A.F., Tryfona, N., and Jensen, C.S. (2001). Geographic data modeling: Requirements and research issues in geographic data modeling. Proceedings of ACM international symposium on Advances in geographic information systems (9), 2-8.

Codd, E.F. (1970). A relational model of data for large shared data banks. Communications of the ACM, 13 (6), 377-387.

Cooper, R. (1997). Object databases: An ODMG approach. International Thompson Computer Press.

Cooper D., Khoo, B., Konsky, B.R.V., and Robey, M. (2004). Java implementation verification using reverse engineering. Proceedings of conference on Australian computer science (26), 203-212.

Cooper, M., and Schiex, T. (2004). Arc consistency for soft constraints. Artificial Intelligence (154), 199-227.

Cox, A.B. (1995). An overview to geographic information systems. Journal of Academic Librarianship 21 (4), 237-249.

CSIRO MARINE RESEARCH (2003). The Southern Ocean's role in the climate system. The Facts (61), 1-4.
<http://www.marine.csiro.au/LeafletsFolder/pdfsheets/62sthnocean.pdf>

D'Ambrogio, A., and Iazeolla, G. (2005). Metadata-driven design of integrated environments for software performance validation. The Journal of Systems and Software (76), 127-146.

Daneri, G.A., Carlini, A.R. and Rodhouse, P.G.K. (2000). Cephalopod diet of the southern elephant seal, Mirounga Leonina, at King George Island, South Shetland Islands. Antarctic Science 12 (1), 16-19.

Davis, B.E. (1996). "Raster and Vector Data." GIS: A Visual Approach (Chapter 4). Santa Fe: Onword Press.

Davis, K.M. (2000). Object-oriented modeling of rivers and watersheds in geographic information systems. Master Thesis (Online). University of Texas at Austin. <http://www.crwr.utexas.edu/gis/gishydro01/library/Davis/rpt00_7.pdf>

Demartino, S., and Hrnicek, E. (2001). Object oriented GIS 101: New GIS data model features intelligent objects. CE News. <http://www.woolpert.com/asp/articles/ObjectOrientedGIS101.asp>

DeMers M.N. (1997). "Cartographic and GIS Data Structures." Fundamental of Geographic Information Systems (Chapter 4). New York: Wiley & Sons.

Ducklow, H.W. (2003). Palmer long-term ecological research project: climate change, ecological migration, and teleconnections in an ice-dominated environment. U.S. Antarctic Program (43-45).

Eddy, J. (1993). Environmental research: what we must do. In Goodchild, M. F., B. O. Parks, and L. T. Steyaert (Eds.), Environmental Modeling With GIS (pp. 1). New York: Oxford University Press.

Egenhoder, M.J., and Frank, A.U. (1992). Object-oriented modeling for GIS. URISA Journal 4(2), 3-19.

Elmasri, R., and Navathe, S.B. (1994). Fundamentals of Database Systems. California: Addition-Wesley Pub.

Ellmer, E. (1993). Basic principles and concepts of object orientation. Vienna: Institute of Applied Computer Science and Information Systems. <http://citeseer.nj.nec.com/ellmer93basic.html, accessed on November 3, 2001>

ESRI (2000). Developing Applications with ArcInfo: An Overview of ArcObjects. ESRI White Paper, 1-22.

Foote, K.E., and Huebner, D.J. (1996). Database Concepts. The Geographer's Craft Project. The University of Colorado at Boulder. <http://www.colorado.edu/geography/gcraft/contents.html>

Frihida, A., Marceau, D.J. and Theriault, M. (2002). Spatial-temporal object-oriented data model for disaggregate travel behaviour. Transactions in GIS 6(3), 277 -294.

Fussel, L. M. (1997). Foundations of Object-Relational Mapping. <http://www.chimu.com/publications/objectRelational/objectRelational.pdf> (cited Nov 2001)

Gartner, H., Bergmann, A. and Schnidt, J. (2001). Object-oriented modeling of data sources as a tool for the integration of heterogeneous geoscientific information. Computer and Geosciences (27), 975-985.

GE (2005). Smallworld core spatial technology data modeling.
<http://www.gepower.com/prod_serv/products/gis_software/en/downloads/sw_core_datamodel_training.pdf>

Goldberg, G.A., Piontkovski, S.A., and Williams, R. (1997). Zooplankton field heterogeneity formation on the mesoscale: elements of the theory and empirical characteristics. Ecological Modelling (96), 41-49.

Goodchild, M.F. (1992). Geographical data modeling. Computers and Geosciences 18(4), 401-408.

Goodchild, M. F. (1993). The state of GIS for environmental problem-solving. In Goodchild, M. F., Parks, B.O. and Steyaert, L.T. (Eds.), Environmental Modeling With GIS (pp. 8-15). New York: Oxford University Press.

Goodchild, M. F. (2000). Foreword. In Wright, D., and Bartlett, D. (Eds.), Marine and Coastal Geographical Information Systems. London: Taylor & Francis.

Gordillo, S., Balaguer, F., Mostaccio, C., and Neves, F.D. (1999). Developing GIS applications with objects: A design patterns approach. GeoInformatics 3 (1), 7-32.

Gradinger, R. and Schnack-Schiel, S.B. (1998). Potential effect of ice information on Antarctic pelagic copepods: Salinity induced mortality of Calanus propinquus and Metridia gerlachei in comparison to sympagic acoel turbellarians. Polar Biology(20), 139-142.

Graham, I. (1996). Basic concepts. In Object-Oriented Methods (pp.1-28). Cambridge: Addison-Wesley.

Guinet, C., Jouventin, P., and Georges, J-Y. (1994). Long-term population changes of fur seals Arctocephalus gazelle and Arctocephalus tropicalis on subantarctic (Crozet) and subtropical (St. Paul and Amsterdam) islands and their possible relationships to EI Nino Southern Oscillation. Antarctic Science 6(4), 473-478.

Halper, M., Liu, L., Geller, J., and Perl, Y. (2003). Frameworks for incorporating semantic relationships into object-oriented database systems. Concurrency and Computation: Practice and Experience (15), 1337-1362.

Harrington, J.L. (2000). Object-oriented database design clearly explained. USA: Academic Press.

Henderson-Sellers, B., and Edwards, J. M. (1994). Book two of object-oriented knowledge: The working object. Sydney: Prentice Hall.

Heywood, I., Cornelius, S. and Carver, S. (2002). Geographical Information Systems. London: Pearson Education.

Hindell, M.A., Slip, D.J. and Burton, H.R. (1991a). The diving behaviour of adult male and female southern elephant seals, *Mirounga Leonina* (Pinnipedia: Phocidae). Australian Journal of Zoology (39), 595-619.

Hindell, M.A., Burton, H.R., and Slip, D.J. (1991b). Foraging areas of southern elephant seals, Mirounga Leonina, as inferred from water temperature data. Australian Journal of Zoology (42), 115-128.

Hooge, P.H., Eichenlaub, W.M., and Solomon, E.K. (2000). Using GIS to analyze movements in the marine environment. Alaska Science Center, 1-21.

Howard-Williams, C. (2001). Climate change as a unifying theme in Antarctic research. Antarctic Science 13(4), 353.

Harrington, J. L. (2000). Object-Oriented Database Design. Academic Press.

Hunsaker, C. T., Nisbert, R. A., Lam, D. C. L., Browder, J. A., Baker, W. L., Turner, M. G., and Botkin, D.B. (1993). Spatial models of ecological systems and processes: the role of GIS. In Goodchild, M.F. (Ed), Environmental modelling with GIS (pp. 248-264). New York: Oxford University Press.

Hunt-Jr, G.L., Priddle, J., Whitehouse, M.J., Veit, R.R., and Heywood, R.B. (1992). Changes in seabird species abundance near South Georgia during a period of rapid change in sea surface temperature. Antarctic Science 4(1), 15-22.

IPCC (2001). Climate change 2001: Impacts, adaptation, and vulnerability. IPCC Third Assessment Report. <http://www.grida.no/climate/ipcc_tar/wg2/index.htm> (cited Jan. 2005)

Jones, C.V. (1996). Visualization and optimization. Boston: Kluwer Academic Pub.

Jonker, F.C., and Bester, M.N. (1998). Seasonal movements and foraging areas of adult southern female elephant seals, Mirounga Leonina, from Marion Island. Antarctic

Science 10 (1), 21-30.

Khedker, U.P., and Dhamdhere, D.M. (1994). A generalized theory of bit vector data flow analysis. ACM transactions on programming languages and systems, 16 (5), 1472-1511.

Khoshafian, S., and Abnous, R. (1995). Object Orientation: Concepts, Analysis and Design, Languages, Databases, Graphical User Interfaces and Standards. New York: John Wiley & Sons.

Korte, G. (2000). Trends in spatial database technology. GIS Vision (11). <http://www.gisvisionmag.com>

Kraak, M-J (1999). Cartography and use of animation. In Cartwright, W., Peterson, M.P., and Gartner, G. (Eds.), Multimedia Cartography (pp. 173-180). Germany: Springer.

Laws, R.M. (1994). History and present status of southern elephant seal populations. In Le Boeuf, B.J., and Laws, R.M. (Eds.), Elephant seals: population ecology, behaviors, and physiology (pp. 49-65). University of California Press.

Lee, R. C., and Tepfenhart, W. M. (2001). UML and C++: A practical guide to object-oriented development. Prentice-Hall.

Lemay, L., Couch, J., and Murdock, K. (1996). 3D Graphics & VRML 2.0. SAMS.

Li, R., (2000). Data models for marine and coastal geographic information systems. In Wright, D.J., and Bartlett, D.J. (Eds.), Marine and Coastal Geographical Information Systems (pp. 25-36). London: Taylor & Francis.

MacEachren, A.M., and Taylor, D.R.F. (1994). Visualization in Modern Cartography. Oxford: Elsevier Science Ltd.

Maidment, D.R. (2001). ArcGIS Hydro Data Model. ESRI User Conference in San Diego.
<http://www.crwr.utexas.edu/gis/gishydro01/introduction/intro2001.html>
(cited Nov 2004)

Malinowski, E., and Zimanyi, E. (2004). Representing spatiality in a conceptual multidimensional model. Proceedings of the 12th annual ACM international workshop on Geographic Information Systems, 12-22.

McConnell, B.J., Chambers, C., and Fedak, M.A. (1992). Foraging ecology of southern elephant seals in relation to the bathymetry and productivity of the southern ocean. Antarctic Science 4 (4), 393-398.

McConnell, B. J., and Fedak, M.A. (1996). Movements of southern elephant seals. Canadian Journal of Zoology (74), 1485-1496. (Also by personal contact)

Microsoft (2004). Component Object Model Technologies.
<http://www.microsoft.com/com/default.mspx> (cited Mar. 2005).

Mitasova, I., Visoovcova, J., and Hajek, M. (1996). Approaches to data modeling in GIS. Presented at Lab of Geoinformatics (Bratislava).
<http://e-collection.ethbib.ethz.ch/cgi-bin/show.pl?type=inkonf&nr=65>
(cited Mar. 2005)

Moloney, C.L. and Johnston, S.J. (2002). Using decision analysis to evaluate candidate OMPs for the South African west coast rock lobster fishery. South African Journal of Science (98): 461-464.

Moore, J.K., Abbott, M.R., and Richman, J.G. (1999). Location and dynamics of the Antarctic Polar Front from Satellite Sea Surface Temperature data. Journal of Geophysical Research (104) C2, 3059-3073.

Morehouse, S. (1999), Preface. In Zeiler, M. (Ed), Modeling the World. USA: Environmental Systems Research Institute Inc.

Murphy, E. (2004). Ocean ecosystem dynamics. UK: British Antarctic Survey.
<http://bsauasc.nbs.ac.uk:8080/~nma/testsite/Science/Programmes/oed/>
(cited Mar. 2005).

Nantajeewarawat, E., Wuwongse, V., Anutariya, C., Akama, K., and Thiemjarus, S. (2004). Toward reasoning with unified modeling language diagrams based on Eextensible Markup Language declarative description theory. International Journal of Intelligent Systems (19), 89-98.

Oosterom, P.V., and Bos, J.V.D. (1989). An object-oriented approach to the design of geographic information systems. Computer Science (409), 257-269.

Orfali, R., and Harkey, D. (1997). Client/Server Programming with Java and CORBA. New York: Wiley Computer Pub.

Otoo, E.J. (1991). Realizing a spatial topological data model in a relational database management system. Ottawa: Carleton University.

Paul, J.F., Copeland, J.L., Charpentier, M., August, P.V., and Hollister, J.W. (2003).

Overview of GIS applications in estuarine monitoring and assessment research. Marine Geodesy (26), 63-72.

Peng, Z, and Tsou M. (2003). “Commercial web mapping programs.” Internet GIS (pp. 377-402). New Jersey: John Wiley & Sons.

Pulsifer, P. (2003), The Cybercartographic Atlas of Antarctica Project. <http://www.carleton.ca/gcrc/caap/> (cited Mar. 2005).

Peterson, M.P. (1995). Interactive and Animated Cartography. New Jersey: Prentice-Hall.

Peterson, M.P. (2003). Maps and the Internet. Cambridge: Elsevier Press.

Peuquet, D.J. (1984). A conceptual framework and comparison of spatial data models. Cartographica 21 (4), 66-113.

Piatkowski, U., Vergani, D.E., and Stanganelli, Z.B. (2002). Changes in the cephalopod diet of southern elephant seal females at King George Island, during El Nino – La Nina events. Journal of Marine Biology 82 (3972), 1-4.

Pulsifer, L.P. (2003). The Cybercartographic Atlas of Antarctica Project: towards implementation. Presented at the second International Antarctic GIS Workshop at Freiburg, Germany.

Raza, A. (2001). Object-oriented temporal GIS for urban applications. Master Thesis. Enschede: The International Institute of Aerospace Survey and Earth Sciences (ITC).

Rambaugh, J., Blaha, M., Premerlani, W., Eddy, F., Lorensen, W. (1991). Object-oriented modeling and design. New Jersey: Prentice Hall.

Raper, J, and Livingstone, D. (1995). Development of a geomorphological spatial model using object-oriented design. International Journal of Geographical Information Systems, 9 (4), 359-383.

Raskin, R (2004), PO.DAAC User Service Office, Personal contact.

Rodhouse, P.G., T.R. Arnbom, M.A. Fedak, J. Yeatman, and A.W.W. Murray (1992), Cephalopod prey of the southern elephant seal, Mirounga Leonina, Canadian Journal of Zoology (70): 1007-1015.

Roy, D., and Clement, G. (1994). Object-oriented database technology for geographic applications: A case study. Proceedings of the Canadian Conference on GIS (1), 664-677.

SCAR (2003). <http://www.scar.org/> (cited Jan 2005).

Shekhar, S., and Chawla, S. (2003). Spatial databases: a tour. New Jersey: Peterson Education.

Smith, B.F. (2004). Environmental interrelationships. In Smith, B.F. (Ed.), Environmental Science. New York: McGraw-Hill.

Smith, T.M., and Reynolds, R.W. (1998). A high resolution global sea surface temperature climatology for the 1961-90 base period. Journal of Climate (11), 3320-3323.

Smith, W.H.F., and Sandwell, D.T. (1997). Global Sea Floor Topography from Satellite Altimetry and Ship Depth Soundings. Science 277 (5334), 1956-1962.

Sondheim, M. (2004), Open source geospatial software. The JUMP Project (1-6). <http://www.jump-project.org/assets/Case_for_Open_Source_Geospatial.pdf>

Stanganelli, Z.B., and Vergani, D.F. (2000). Searching for causes of southern elephant seal population declines. In Davison D. (Ed.), Antarctic Ecosystem (pp. 291-297). Caxton Press.

Stonebraker, M. and Brown, P. (1999). Object-Relational DBMSs: Tracking the Next Great Wave. San Francisco: Morgon Kaufmann.

Storey, V.C. (1993). Understanding semantic relationships. The International Journal on Very Large Data Bases 2 (4), 455-488.

Takashi, O. (1994). The importance of the southern ocean cephalopod fauna. Antarctic Science (6) 2, 135.

Taylor, D.R. F. (2003). The concept of cybercartography. In Peterson, M. (Ed), Map and the Internet (pp: 403-418). Cambridge: Elsevier.

Teller, S. (1996). Application challenges to computational geometry. Technical report (Princeton University). <http://www.cs.princeton.edu/~chazelle/taskforce/CGrepot.ps> (cited Jan 2005)

The Alfred Wegener Institute, Germany. Arktis & Antarktis. (Also exhibited in Bonn) <http://www.awi-bremerhaven.de/AWI/KAH97/index.html> (cited Mar. 2005).

Tryfona, N., Pfoser, D. and Hadzilacos, T. (1997). Modelling the behaviour of geographic objects: An experience with the Object Modelling Technique. Proceedings of the 9th International Conference on Advanced Information Systems and Engineering, Computer Science Vol. 1250, Berlin: Springer-Verlag.

Tsou, M-H (2004). Integrating web-based GIS and image processing tools for environmental monitoring and natural resource management. Journal of Geographical Systems (6), 155-174.

Twumasi, B.O. (2002). Modeling spatial object behaviors in object-relational geodatabase. Master's Thesis. Enschede: International institute for geo-information science and earth observation (ITC).

USAP (2001). Long-term ecological research. US Antarctic Program. <http://www.nsf.gov/od/opp/antarct/treaty/projsum01/html/tofc.html> (cited Nov 2004)

USEPA (U.S. Environmental Protection Agency). Terminology Reference System. <http://www.epa.gov/trs/> (cited Mar. 2005)

Van den Hoff, J., Burton, H., Davies, R. (2003). Diet of male southern elephant seals (Mirounga Leonina L.) hauled out at Vincennes Bay, East Antarctica. Polar Biology (26), 27-31.

Valavanis, V.D. (2002). Marine geographic information systems. In Valavanis, V.D. (Ed), Geographic information systems in oceanography and fisheries (pp. 1-33). London: Taylor & Francis.

Vergani, D.F., Stanganelli, Z.B. and Bilenca, D. (2001). Weaning mass variation of southern elephant seals at King George Island and its possible relationship with 'El Nino' and 'La Nina' events. Antarctic Science 13 (1), 37-40.

Vergani, D.F., Stanganelli, Z.B., Carlos, J., Holsbeek, L., and Decleir, H. (2002). Climate variability and global warming impact on Antarctic ecosystem structure and functioning. Presented at Centro Patagonico (Puerto Madryn).

Vergani, D.F., Stanganelli, Z.B., and Bilenca, D. (2004). Effects of El Nino and La Nina events on sex ration of southern elephant seals at King George Island. Marine Ecology Progress Series (268), 293-300.

Vergani, D.F., Stanganelli, Z.B., and Labraga, J.C. (In press). Linking changes on elephant seal parameters at King George Island (1985-1995) with climatic variations in the Bellingshausen Sea during El Nino-La Nina event. 9p. (Also by personal contact)

Walther, G.R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J.M., Guldberg, O.H., and Bairlein, F. (2002). Ecological responses to recent climate change. Nature (146), 389-395.

Webster Dictionary, (1913), Model, pp: 934
<http://machaut.uchicago.edu/cgi-bin/WEBSTER.sh?WORD=model>
(cited Mar. 2005).

Wegener, M. (2000). Spatial Models and GIS. In F.A. Stewart, and Wegener, M. (Eds.), Spatial Models and GIS – New Potential and New Models (pp. 1-20). UK: Taylor & Francis.

Wikipedia Dictionary, (2004), Data model.
<http://en.wikipedia.org/wiki/Data_model> (cited Mar. 2005).

Worboys, M. F. (1995). GIS: A Computing Perspective. London: Taylor & Francis.

Worboys, M. F. (1999). Relational databases and beyond. In Longley, P.A., Goodchild, M.F., Maguire, D.J., and Rhind, D.W. (Eds.), Geographic information systems (pp. 373-384). New York: Wiley.

Xavier, J.C. (1997). Distribution of cephalopods in the Southern Ocean, Sub-Antarctic and Sub-Tropical waters. Degree Thesis of Marine Biology and Fisheries. Portugal: University of Algarve. (Also by personal contact)

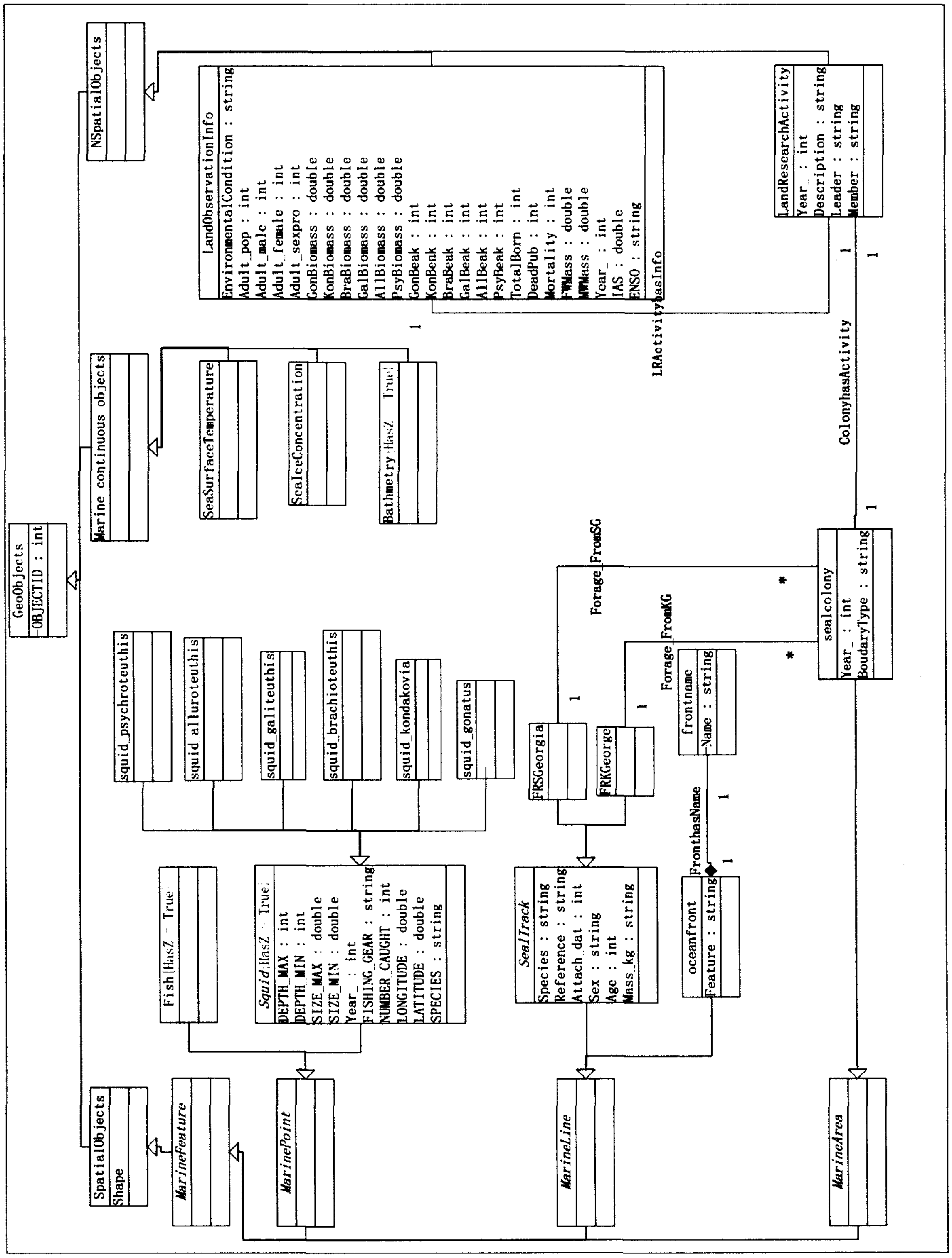
Xavier, J.C., Rodhouse,P.G., Trathan, P.G., and Wood, A.G. (1999). A geographical information system (GIS) atlas of cephalopod distribution in the Southern Ocean. Antarctic Science 11(1), 61-62. (Also by personal contact)

Yuan, X., and Martinson, D.G. (2000). Antarctic sea ice variability and its global connectivity. Journal of Climate (13), 1697-1717.

Zeiler, M. (1999). Modeling our world. USA: Environmental systems research institute.

Zeiler, M. (2002). Exploring ArcObjects (Vol. 1). USA: ESRI.

APPENDIX A: THE CONCEPTUAL MODEL



APPENDIX B: THE LOGICAL MODEL