ENVIRONMENTAL DATABASES AND GIS

J R G TOWNSHEND

An enormous variety of environmental data is found within existing databases, which are growing rapidly. As yet a relatively small proportion of such data have benefited from the application of GIS technology. This arises because of the high investment needed to convert existing archival data into digital format, because of a lack of awareness of the value of these techniques and because traditional GIS is often unsatisfactory for many of these data, which are inherently three and four dimensional. The diversity of environmental databases also hinders their integration, which is an activity essential for exploiting the full value of their data. The benefits of integrating data sets through GIS is illustrated in five examples: in estimating the availability of natural resources, in improving data capture, in assisting the visualization of changes, in applying physical models and in extrapolating and modelling environmental parameters through the use of sample data. This chapter complements those by Clark, Hastings and Kineman (1991) and Mounsey (1991).

INTRODUCTION

Environmental databases contain an enormous diversity of types of data (Table 49.1), most of which are spatially located either explicitly or implicitly. Consequently the capture, analysis, management and display of environmental data are all activities that can greatly benefit from the application of GIS. Moreover, inherent in the solution of many environmental problems is the need to bring together disparate data sets. Another characteristic of many environmental data sets is that they are often extremely large in volume, arising in part because their collection and maintenance is carried out on a national or even international basis. The size and complexity of these databases makes the requirement for the application of GIS technology all the more necessary. But, perversely, this is also a hindrance to their introduction because of the high investment necessary to convert data in existing archives into digital format.

The benefits of handling environmental data sets within GIS does not mean that all environmental data sets are likely to be stored and handled in a single uniform manner. The establishment of international formats within specific disciplines, coupled with the very different space and time scales of the phenomena being depicted, militate strongly against such standardization.

Despite the high potential of GIS technology for environmental applications, its penetration remains modest in this field in the early 1990s. This relates in part to a lack of awareness of the potential of the technology and in part to the expense in achieving its full operational usage. Many potential users regard GIS as little more than automated map-editing systems. Although this is a mistaken idea, it is undoubtedly true that currently available commercial GIS are most effective when dealing with two-dimensional spatial data (Smith and Paradis 1989; Raper and Kelk 1991 in this volume), whereas many environmental data sets are inherently three dimensional, as in the case of solid geology, or even four dimensional, in the case of

Table 49.1 Examples of geographically referenced environmental data holdings (NERC 1988).

Geological data sets
1. Borehole logs
2. Geochemical records (including stream geochemistry)
3. Geophysical survey data
4. Gravity data
5. Geomagnetic survey
6. Hydrogeological well records

Marine data sets
1. Sea surface temperature
2. Current meter data
3. Wave height/period data records
4. Ocean geophysics
5. Salinity data
6. Side scan sonar data

Ecological data sets
1. Species location data
2. Terrain/land characteristics
3. Location of conservation sites
4. Soil types distribution
5. Biomass data

Hydrological data sets
1. Rainfall data
2. Soil moisture
3. Evapotranspiration
4. River discharge
5. Location of river networks

Atmospheric data sets
1. Air temperature at various heights
2. Air pressure at various heights
3. Atmospheric chemistry
4. Wind speed
5. Humidity

Important ancillary vector data sets
1. Coastlines and political boundaries
2. Digital terrain models
3. Topographic maps.

Important ancillary remote sensing data sets
1. Data from Meteosat/NOAA (National Oceanographic and Atmospheric Administration) AVHRR (Advanced Very High Resolution Radiometer)/DMSP (Defense Meteorological Satellite Program) for climate and meteorological applications.
2. Landsat Thematic Mapper and MSS (Multispectral Scanner System) data, SPOT-HRV (Système Probatoire pour l'Observation de la Terre) and AVHRR data for land survey and monitoring.
3. AVHRR, CZCS (Coastal Zone Color Scanning System) and ERS-1 data for oceanic monitoring.

most atmospheric, marine, and geophysical data sets. Finally – and notwithstanding the low level of awareness of the contribution of GIS *per se* – many users of environmental data have independently developed very sophisticated procedures for the capture, handling and analysis of environmental data. Nowhere is this more apparent than in the field of meteorology where the demands for continuously updated and frequent forecasts have spurred the development and application of highly effective four dimensional data handling systems.

CHARACTERISTICS OF ENVIRONMENTAL DATA SETS

In order to demonstrate the characteristics and diversity of environmental data sets, each one of four main natural science areas will be discussed in turn, followed by an analysis of their differences and similarities (NERC 1988).

Earth science data

Almost all data that are used in the earth sciences are spatially referenced to a point on or below the earth's surface, location being described in terms of the three dimensions of geographical space. Frequently, such data are aggregated into areas defined as polygons or into volumes for the purposes of analysis and visualization. Earth science databases contain a wide variety of geological, geotechnical, geophysical and geochemical parameters, and increasingly include detailed chemical and biological descriptions. Some geophysical data, such as those concerned with seismology and geomagnetism, also have a time dimension. The spatial referencing system used is normally either that of the national topographic mapping agency if on land or latitude and longitude (or UTM grid) if off-shore. Heights and depths variously refer either to a national datum, such as the National Geodetic Vertical Datum of the United States, or to mean sea level. The former is fixed and the latter varies through time and space due to the various local and regional environmental factors that affect sea level (Ellis 1978).

Data are increasingly captured in digital form but, in most national surveys, there is an enormous backlog of analogue data requiring digitizing. Some data are inordinately expensive to digitize such as field notebooks or hand-annotated field maps: others, such as rock specimens, fossil collections and borehole materials, are impossible to digitize though measurements of many of their parameters can of course be collected and stored in digital form. National geological surveys or their equivalent are usually the main repository of systematically stored geological information and, in some countries such as the United Kingdom, some borehole logs and specimens are deposited as a statutory requirement. Smaller collections are also found in universities, museums, industry and local government organizations. Collaboration on a global scale is provided under the World Data Center system (Table 49.2; see also Clark *et al.* 1991 in this volume) for several types of earth science data. These include seismology at the US Geological Survey in Golden, Colorado; recent crustal movements at the International Centre for Recent Crustal Movements in Prague, Czechoslovakia; and Geomagnetism at the British Geological Survey in Edinburgh. Regional collaborations are also under consideration, such as that for the North Sea under the auspices of WEGS (Western European Geological Surveys).

Marine data

Marine data typically refer to the sea surface, the body of the water beneath it, the seabed and the sub-seabed: they include physical, chemical, biological, geological and geophysical parameters (Table 49.3); clearly there is overlap with earth science interests with respect to the latter two categories though separate sets of databases are usually maintained. Marine data are geographically located in three dimensions and are usually referenced by latitude and longitude and depth below the sea surface. For more precise measurements in the vertical dimension, such as sea level itself, a geoidal reference is required. Many of the data sets have an additional time dimension, especially those concerned with the characteristics of the sea itself. The time scales vary enormously, ranging from the high frequencies of sea surface waves, through phenological variations of many biological phenomena to much longer term variations associated with climate change.

Marine data are often stored in the form of summary descriptors, such as averages, in order to characterize an area of the sea surface or volume of the sea body. For example, data are commonly aggregated into 1 degree squares, FAO fisheries squares or Marsden squares (NERC 1988). Some data with a time dimension are presented in a processed form, such as wave spectra. Data such as temperature, salinity, current and wave measurements are readily captured in, or can be converted to, digital form. Other data, such as photographs or sonar and seismic records, can be digitized using raster scanning, but with much greater effort. Biological specimens and geological materials are impossible to digitize.

Capture of marine data sets is carried out in many contrasting ways. Some are collected from commercial shipping whereas others are derived primarily from specific scientific or commercial survey cruises, as in the case of many collections of geological samples of seabed rocks and marine geophysical data.

As a consequence of the diversity of data sets and the underlying requirements for their creation, there are local, national, regional and international organizational frameworks and agreements to ensure the maintenance of data sets and their standards. Table 49.4 illustrates some of the organizations that have responsibilities for marine data. The collection and dissemination of marine data forms an important part of the World Data Center System (Table 49.2).

Terrestrial ecological data

Ecological data sets often include both attributes (e.g. species name, vegetation class and growth stage) along with numerical measurements (e.g. temperature, pH and nutrient levels). Many such ecological measurements are recorded with reference to only two spatial dimensions, usually by means of the locational referencing system of the country within which the data are collected through use of local topographical maps to define location. In international data sets, however, latitude and longitude or the UTM grid are normally used. The vertical dimension (specified in terms of elevation above sea level) is often explicitly recorded or is readily derivable from topographic maps. Measurements made within a soil profile will also

Table 49.2 The World Data Center System (from Allen, 1988 and ICSU 1989).

Type of data serviced	Sponsoring Institution	Location
World Data Center – A		
Coordination Office: US National Academy of Sciences.		
Glaciology (Snow and Ice)	University of Colorado and NOAA/NGDC	Boulder, Colorado
Meteorology	NOAA, National Climate Data Center	Asheville, N. Carolina
Oceanography	NOAA, National Oceanographic Data Center	Washington DC,
Rockets and Satellites	NASA, National Space Science Data Center	Greenbelt, Maryland
Rotation of the Earth	US Naval Observatory	Washington DC
Seismology	US Geological Survey	Golden, Colorado
Marine Geology/Geophysics, Solar-Terrestrial Physics, and Solid Earth Geophysics	NOAA, National Geophysical Data Center	Boulder, Colorado
World Data Center – B		
Operated under the Soviet Geophysical Committee of the Academy of Sciences of the USSR		
World Data Center – B1		
Meteorology, Oceanography, Marine Geology/ Geophysics, Glaciology, Tsunamis, Rockets and Satellites, Rotation of the Earth, Mean Sea Level and Ocean Tides.	USSR State Committee for Hydro-meteorology and Control of the Environment.	Obninsk, USSR
World Data Center-B2		
Solar-Terrestrial Physics and Solid Earth Geophysics	Soviet Geophysical Committee, Academy of Sciences, USSR.	Moscow, USSR
World Data Center – C		
World Data Center – C1		
Earth Tides & Sunspot Index	Royal Observatory of Belgium	Brussels, Belgium
Geomagnetism	Danish Meteorological Institute and British Geological Survey	Copenhagen, Denmark Edinburgh, UK
Glaciology	Scott Polar Research Institute	Cambridge, UK
Recent Crustal Movements	International Centre for Recent Crustal Movements	Prague, Czechoslovakia
Soil Geography and Classification	International Soil Reference and Information Center	Wageningen, The Netherlands
Solar Activity	Observatoire de Paris	Meudon, France
Solar-Terrestrial Physics	Science and Engineering Research Laboratory	Chilton, UK
World Data Center – C2		
Airglow	Tokyo Astronomical Observatory	Tokyo, Japan
Aurora	National Institute of Polar Research	Kaga, Japan
Cosmic Rays	Institute of Physical and Chemical Research	Tokyo, Japan
Geomagnetism	Kyoto University, Ministry of Education	Kyoto, Japan
Ionosphere	Ministry of Posts and Telecommunications	Tokyo, Japan
Nuclear Radiation	Japan Meteorological Agency	Tokyo, Japan
Solar Radio Emissions	Nagoya University, Ministry of Education	Toyokawa, Japan
Solar-Terrestrial Activity	Institute of Space and Aeronautical Research, Ministry of Education	Tokyo, Japan
World Data Center – D		
Astronomy	Beijing Astronomical Observatory	Beijing, China
Geophysics	Institute for Geophysics	Beijing, China
Geology	Chinese Academy of Geological Sciences	Beijing, China
Glaciology and Geocryology	Lanzhou Institute of Glaciology and Geocryology	Lanzhou, China

Table 49.2 *Continued*

Type of data serviced	Sponsoring Institution	Location
World Data Center – D (continued)		
Meteorology	National Meteorological Center	Beijing, China
Oceanography	National Oceanographic and Information Centre	Tianjin, China
Renewable Resources and Environment	Commission for Integrated Survey of Natural Resources	Beijing, China
Seismology	Department of Science Programming and Earthquake Monitoring	Beijing, China
Space Sciences	Chinese Academy of Sciences	Beijing, China

Table 49.3 Types of marine data sets.

1. Ocean biological samples
2. Geological collections of seabed rocks and sediments
3. Current measurements
4. Echo sounding profiles
5. Seismic records
6. Sidescan sonar records
7. Magnetic records
8. Gravity records
9. Earth tide data
10. Plankton records
11. Inter-tidal biological records
12. Sea surface temperature
13. Frequency and location of sea mammals
14. Fisheries data
15. Wind data
16. Bathymetric data
17. Conductivity data
18. Salinity

require a depth dimension, but this is recorded with much greater resolution. The time of data acquisition will usually be recorded but continuously recorded time series outside climatological and hydrological data sets are uncommon. Data sets may be made available in highly aggregated forms, such as the presence or absence of species within grid cells of a specified size, origin and orientation.

However, many types of attribute data are difficult, if not impossible, to manage in digital form. These include actual samples, such as biological reference specimens and soil samples, as well as many paper records. International collaboration with respect to such matters as compilation of global databases, agreement on standards and exchange formats, is apparently not well developed compared with most other environmental disciplines. There are, however, already efforts to coordinate certain types of ecological data such as the global databases on the Status of Biological Diversity, compiled by the Conservation Monitoring Centre under the aegis of the International Union for Conservation of Nature and Natural Resources (Pellew and Harrison 1988); World Data Center-C2 has been extended to include a new centre for Soil Geography and Classification and a centre for Renewable Resources and Environment has been set up in the newly constituted World Data Center-D (Table 49.2). As part of the International Geosphere Biosphere Programme, considerable efforts are being directed towards setting up a global information system for land cover (Rasool and Ojima 1989). Also, as part of the Global Environmental Monitoring System (GEMS) of the United Nations Environmental Programme (UNEP), the Global Resource Information Database (GRID) project has been set up to bring together environmental databases on a common geographical base at both global and more local scales (Mooneyhan 1988).

Atmospheric sciences

Enormous quantities of atmospheric data are collected, processed and managed on a regular daily basis to meet the requirements of operational meteorology, although climate monitoring and modelling are increasingly important users of atmospheric data. Data are referenced to all three spatial dimensions as well as time. The spatial and temporal frequencies of data collection are very variable, especially between land and sea and between developed and developing countries, though satellite data are bringing greater uniformity for some parameters such as sea surface

Table 49.4 Examples of organizations responsible for marine data (NERC 1988).

1. National organizations such as NOAA (National Oceanographic and Atmospheric Administration of the USA) and NERC (the Natural Environment Research Council of the UK).
2. International Oceanographic Commission's (IOC) Working Committee on International Oceanographic Data Exchange (IODE)* which has set up:
   (a) a group on format development leading to a general Formatting System for Geo-referenced Data (GF3)
   (b) National Oceanographic Data Centres (NODC) including Responsible NODC's (RNODC) with responsibilities for specific data sets for the international community. For example, the UK NODC is called MIAS (Marine Information Advisory Service) and has particular responsibility for world waves.
3. International Gravity Bureau at Toulouse.
4. General Bathymetric Chart of the Oceans (GEBCO) under the International Oceanographic Commission and the IHO.
5. FAO Fishery Data Centre in Rome.
6. Major international programmes such as WOCE (World Ocean Climate Experiment), JASIN, IGY.
7. Regional organizations such as the International Council for the Exploration of the Sea (ICES).

* Subsequently renamed the IOC Technical Committee on International Oceanographic Data and Information Exchange (though retaining the same acronym).

temperature and cloud cover. One of the first requirements after data capture is to convert sparse, irregular sets of observations to smooth continuous displays of spatially and temporally referenced meteorological variables for use in numerical weather prediction (NERC 1988).

The strategic and economic requirements for international information on weather conditions, and the need for data for very large areas in producing effective weather forecasting, have led to establishment of the Global Telecommunications System (GTS) under the World Meteorological Organization (WMO). This organization distributes large volumes of atmospheric data in internationally agreed formats to most countries in the world, several times every day. Meteorological data are the responsibilities in World Data Center A of the National Climate Data Center in Asheville, North Carolina and in World Data Centre B they are the responsibility of the USSR State Committee for Hydrometeorology and Control of the Environment at a facility in Obninsk, USSR (Table 49.2).

Overview of the characteristics of environmental databases

The foregoing summary overview has demonstrated the considerable range of environmental data sets. Among the most diverse aspects are the following:

- *Locational referencing*. Although all data are collected in time and space, for some data sets two spatial dimensions provide sufficient locational referencing whereas for many others, especially in the marine and atmospheric areas, three spatial dimensions and the time dimension are required.
- *Longevity of databases*. The longevity of the usefulness of data sets varies from the immediate requirements of weather forecasting, where data become essentially valueless in a few hours, to some geological data banks where observations made in the nineteenth century remain important.
- *Types and formats of data*. Many environmental databases typically contain data in many different forms, including raster and vector, digital and analogue types (the latter, in particular, being manifested in many different forms). Data sets may also often consist of tangible materials such as type examples of biological, soil, or rock specimens.
- *Availability of digital data*. In some fields, such as meteorology and climatology, the majority of data are in digital form but in others this is not the case. In the latter situation, the inherent diversity of data types and the lack of resources

for digitizing data ensure that implementation of a modern GIS capability is considerably curtailed.

- *Degree of integration of data sets.* For meteorology, a highly sophisticated global system of continuous data collection and distribution exists though major weather systems can be well characterized for the purposes of forecasting by a relatively coarse spatial sample of observations, at least as compared with many other environmental phenomena. At a national level, there is often a considerable degree of central management of geological information, though their different data sets are typically imperfectly linked. For ecological information, local and poorly integrated data sets are often the norm.

Some similarities can also be recognized:

- *Growth in size of environmental data sets.* This trend continues because of the addition of new contemporary data and the impact of new, more spatially and temporally comprehensive, methods of data capture – notably remote sensing (Fig. 49.1; see also Davis and Simonett 1991 in this volume). Another factor has been the requirements for creating longer time series of environmental characteristics at global scales associated with growing interest in longer term climate changes and their impacts on the biosphere.
- *Increasing digitizing of data holdings, catalogues and indices.* In all environmental disciplines there is an increasing drive towards digitizing data holdings wherever possible and to improve digital methods of data cataloguing and indexing.
- *Mismatches between requirements for use of environmental data sets and current GIS technology.* In all environmental areas, currently available commercial GIS have substantial limitations: among the most important are their inability easily to deal with hybrid data sets containing both raster and vector data and their failure to handle three- and four-dimensional data sets fully, though techniques are being developed to ameliorate these problems (e.g. Langran 1989; Raper and Kelk 1991 in this volume).
- *Increased inter-linking and spatial integration of data sets.* Environmental data sets are increasingly being linked due to the needs of the environmental sciences and the applications to which they are put (as discussed in the next section), and because of technology-driven improvements in information handling technology. Inter-linking is being facilitated in part through the adoption of common formats and, in particular, through agreements on national and international exchange formats. The latter are especially important in producing spatially more comprehensive data sets. One of the most important of these exchange formats in the environmental field is GF3 (General Format 3) used by the marine community which was formally accepted by the international community over 10 years ago. As Fig. 49.2 indicates, achieving acceptance of the format and maintaining it has involved a considerable amount of international collaboration and organization.

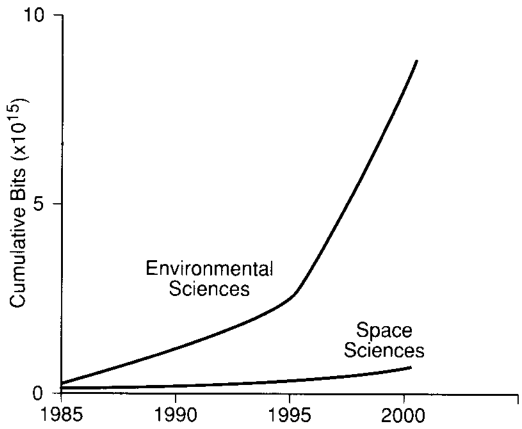

Fig. 49.1 Anticipated growth in data volumes arising from the use of space-borne remote sensing instruments for the environmental sciences compared with the space sciences (Earth System Sciences Committee 1988).

LINKING ENVIRONMENTAL DATA SETS

In all environmental fields, there is increasing interest in problems which by their very nature

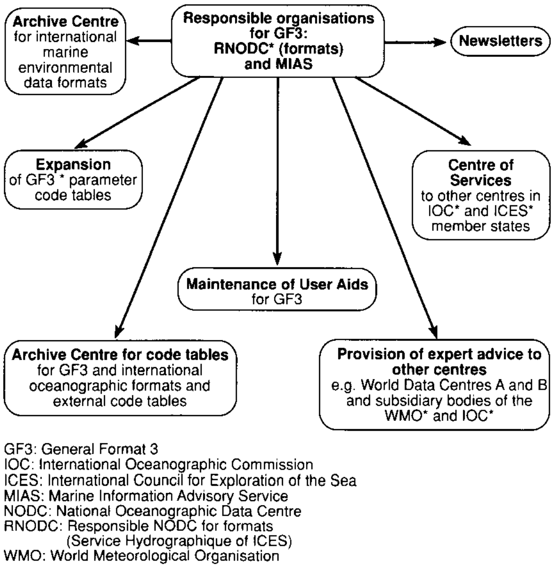

Fig. 49.2 The various functions associated with GF3, the widely used international system for the exchange of marine data (Intergovernmental Oceanographic Commission 1987).

require the integrated use of several different data sets. Three examples can be used to illustrate this point. First, the integrated use of environmental parameters (e.g. surface roughness and soil and vegetation characteristics) with atmospheric parameters is increasingly common within global climate models (GCMs) in order to improve estimates of the partitioning of energy and mass transport at the land surface. Secondly, investigations of environmental geochemistry require integration of geological, stream chemistry and human disease data sets. Thirdly, comprehension of the human impact of climate change requires not only an integration of climatic and vegetation data sets, but also a wide range of socio-economic characteristics as well.

Despite the widely recognized benefits of linking environmental data sets, there are major hindrances which inhibit this activity. Some of those impacting most strongly are discussed below.

Varied procedures of data capture

Generation of spatially comprehensive databases often requires linking many more local data sets. The latter may well frequently use different definitions of attributes and different procedures of data gathering and sampling making the creation of internally consistent data sets difficult (see Mounsey 1991 in this volume). Notable examples where this occurs include the generation of rainfall fields across national frontiers, because of the different ways in which rainfall is measured (Bull 1960), and problems in the creation of globally consistent vegetation maps assembled from numerous local surveys. In the latter example, difficulties arise through variations in the type of classification used (the main types being floristic, physiognomic and ecological), as well as variations in the definition of classes within a given classificatory scheme. This undoubtedly contributes to the considerable variations in both local and global estimates of vegetation types (Table 49.5). An additional cause of error arises from different times of data acquisition. For some geological phenomena, this is relatively unimportant because they change so little. However, in the case of land cover or land use, the phenomena often change rapidly and, moreover, data collection for large areas can extend over several years: consequently, some of the spatial variations within the resultant data set may be spurious.

Table 49.5 Variations in estimates of the global areal extent of major vegetation types.

Vegetation type	Sources 1	2	3	4
	(Area in km^2 × 10^6)			
Forest	48.5	37.4	39.3	49.3
Tundra	8.0	11.7	7.3	11.9
Desert	8.5	4.8	15.6	20.8
Marshland	2.0	3.0	*	2.5
Cultivated	14.0	56.6	17.6	15.9

* Category absent

Sources: 1, Lieth (1975); 2, Hummel and Reck (1979); 3, Matthews (1983); 4, Olson, Watts and Allison (1983).

Lack of compatibility of co-registered data sets

The simple physical co-registration of environmental data sets, through the use of common or linked spatial referencing systems, does not of itself ensure that the interrelationships of phenomena themselves are sensibly described. Data sets may have very different sampling densities or the errors inherent in one data set may be unacceptably high to allow their integrated use – due, for example, to errors introduced during measurement or by generalization. One specific example of this problem is the inadequacy of current USGS digital elevation models for the correction of relief effects in high resolution remotely sensed data (Topographic Science Working Group 1988) because the DEMs have errors sufficiently large to prevent the slope and azimuth of individual Thematic Mapper 30 m pixels to be estimated accurately. However these DEMs would be much more satisfactory for coarse resolution data such as that from the AVHRR with a resolution of 1 km.

Use of polygons to represent heterogeneous phenomena

Polygonal representations of environmental characteristics are common, notably in traditional, geological soils and vegetation maps: a relatively small number of class labels relative to the number of polygons is typically used. Such representations can pose substantial problems, even when used singly, because of the considerable internal variability of the units. In geological maps, the lithological description of a rock unit almost always hides considerable vertical and horizontal variability and, in the case of soil mapping units, the within-class variability is often comparable with (and may even greatly exceed) the between-class variability for some parameters (Webster 1978; Webster and Butler 1976; Burrough 1991 in this volume). In environmental maps, the use of polygons often poses particular problems because of the use of a single set of units to represent multiple environmental parameters and attributes. Thus, in the case of soil maps, each mapped class has to represent a complex three-dimensional combination of physical and chemical soil characteristics. The use of a single set of polygons to represent the variation of multiple attributes and parameters exacerbates substantially the problems associated with polygon overlay. Probably the most extreme case of a single set of polygons being forced to represent a set of heterogeneous environmental properties is found in the 'land system' approach of mapping where the pedological, geological, geomorphological and hydrological properties are all purported to be summarized in a single hierarchical set of areal units (Mitchell 1973).

Diversity of data formats and data types

This is a very common characteristic of environmental data sets and has already been alluded to: it can place a substantial overhead on use of multiple data sets where they have not already been linked.

Maintaining the integrity of data sets through time

The creation of linked data sets depicting changes through time is central to many studies of environmental change. Their creation requires not only the maintenance of consistent measurement and sampling procedures – often accompanied by careful calibration – but should also demand the maintenance of audit trails. One recent example where this was not done occurred in the widely used Global Vegetation Index product of NOAA which depicts vegetation 'greenness' at a global scale (e.g. Justice *et al.* 1986). Subsequent analysis by Goward *et al.* (1991) has revealed considerable unrecorded changes in the specification of this image product; these can introduce considerable errors if used to create longer term time series.

Unlinking data sets

Not only do data sets often need to be linked but, conversely, it is also necessary to ensure that data sets do not suffer from the reverse process and diverge. Specifically, the creation of split systems – involving dual maintenance of copy databases without a clearly established central database manager – can lead to a substantial effort in re-creating a single data set. For example in the United Kingdom, the Water Resources Board produced a copy of the National Well Record Collection and separately maintained it; subsequently the Department of the Environment had to fund the merging of the two databases which had substantially diverged (NERC 1988).

APPLICATIONS OF GIS IN THE ENVIRONMENTAL SCIENCES

Many of the uses of GIS in the environmental sciences will continue to be prosaic, involving more efficient capture, manipulation and display of data in conventional formats. However, the use of GIS opens up a variety of additional analytical opportunities and five case studies are presented in order to demonstrate the contribution of GIS. The role of GIS methods is described successively in estimation of the availability of natural resources, in improving the quality of data capture, in assisting visualization of changes, in applying physical models, in extrapolating and modelling changing environmental parameters through use of sample data, and in providing better access to environmental data.

Use of coarse resolution satellite data for estimating the availability of natural resources

For several years, remotely sensed data have been used for the provision of information concerning the physical environment. More recently, the value of coarse resolution data derived from weather satellites for monitoring the vegetation cover of very large areas has been recognized (e.g. Justice *et al.* 1986; Tucker, Townshend and Goff 1985). The value of these data can be illustrated by a study reported in Millington *et al.* (1989) in which coarse resolution sampled images were used to characterize vegetation types of several countries in southern Africa, using the variation of a spectral vegetation index throughout the year to stratify vegetation types in terms of their woody biomass. Plate 49.1 shows three such images, obtained at different times of year. The pattern of vegetation activity shown by these images is clearly related to the spatial pattern and temporal sequence of rainfall. Using automated classification procedures applied to the multi-temporal data sets, it was possible to separate the area into cover classes (Plate 49.2) and the areas occupied by each class were estimated (Fig. 49.3(a)). Ground survey and previously published results were then used to assign woody biomass values to these classes; statistical estimates of the growing stock (Fig. 49.3(b)) and the mean annual increment were estimated (Fig. 49.3(c)) for the biomass classes of each country. Using an overlay of regional boundaries, estimates were made of the fraction of the growing stock and the mean annual increment in each sub-national region (Fig. 49.3(d)). Figure 49.3(e) outlines the procedures used in making these estimates.

Combining data sets to improve the quality of data capture

The integration of different data sets can greatly assist the data capture process. Work by Mason *et al.* (1988) shows the benefit of incorporating data from topographical maps with remotely sensed data in order to improve the capture of land cover information (Plate 49.3). Specifically, land unit boundaries were linked with remotely sensed data using a knowledge-based approach. The basic system first involved an initial segmentation into polygons using boundaries derived from the topographical map. The segmentation was then revised using information from the remotely sensed data, together with external knowledge embodied as rules in a rule base. This allowed a preliminary identification of the broad classes to which each polygon belonged and this information was then used to guide the classification of the remotely sensed data. Three types of rules were used:

- *Domain consistency rules.* These were concerned with knowledge of a region's characteristics and are used to increase or decrease confidence in the broad classes to which each polygon is assigned.
- *Split rules.* These were used to assess whether a polygon should be split. One rule found to be very important was whether a polygon was homogeneous in terms of its combined spectral and textural properties.
- *Merge rules.* Conversely, these rules were used to assess where congruent regions should be merged. For example, if one region was already assigned to the 'field' class then it was merged with an adjacent class if the merged region was more like a field as measured by a reduction in concavity.

Classification using the remotely sensed data was carried out using a region or polygon-based

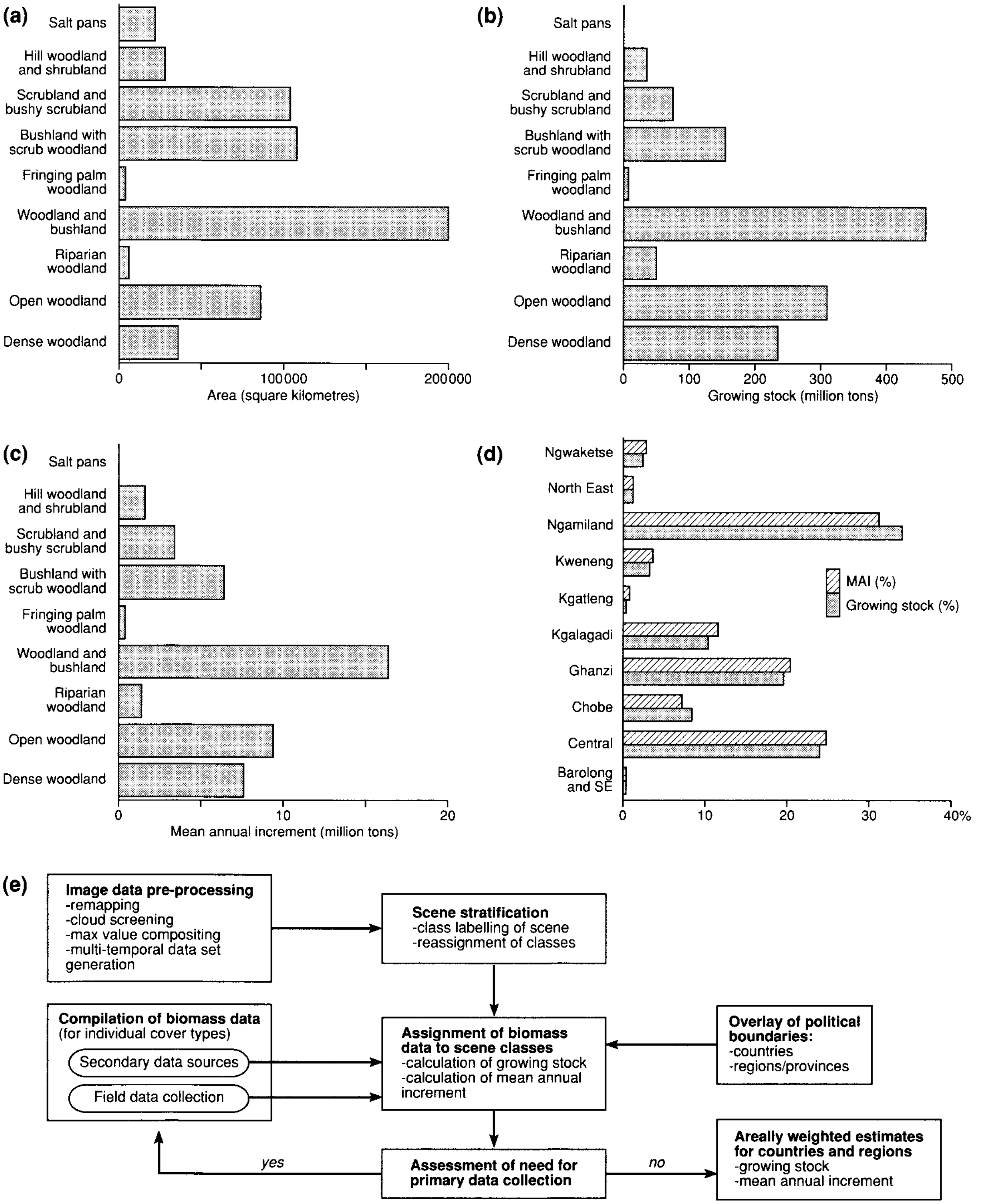

Fig. 49.3 Estimation of fuelwood supplies in southern Africa using remotely sensed data (Millington *et al.* 1989): (a) areal estimates of the extent of biomass classes in Botswana; (b) growing stock of biomass classes for Botswana; (c) mean annual increment for each of the biomass classes for Botswana; (d) biomass attributes by province for Botswana. The percentage of the country's mean annual increment and the percentage of the national growing stock are depicted for each province; (e) overall procedures for estimating fuelwood supplies.

approach rather than a conventional pixel-based approach. Moreover, the class to which a polygon could be allocated was restricted by the broad class to which it had been previously assigned using the rule base. Thus, if a polygon was classified as a field, the class could only be one of the agricultural classes. As a result of applying this approach, areal errors of land cover classification were reduced from 24–29 per cent to 8–9 per cent within the test areas investigated. In Plate 49.3, the benefits of using this approach can be appreciated by comparing the reference data collected on the ground with the results of classification using the per-pixel data and the segmented image. An important methodological point to make about this approach is that the successful integrated use of data sets was achieved through external knowledge and modelling of the phenomena depicted – and not merely by a simple overlay of the data.

Assessment of the visual impacts on the physical environment

In the previous example, remotely sensed data and topographic data were used in an integrated manner to improve the capture of derived land cover information. An alternative benefit of such integration which has become increasingly common is for the visualization of terrain, where remotely sensed data are draped over digital elevation models. This technique not only permits the current appearance of landscape to be depicted but also allows the impact of various changes to be simulated. For example, Quarmby and Saull (1990) carried out an assessment of the impact of residential building on part of rural north Hampshire in the United Kingdom (Plate 49.4). Perspective views were generated from a number of locations and then the impact of screening by trees was investigated by 'growing' belts of trees vertically on top of the digital elevation model. The results showed that, from some viewpoints, this had a significant ameliorating effect but that, from several others, little of the development was obscured.

Incorporation of physical models in a GIS framework

The previous example showed how 'what if' questions could be posed in an informal way in the context of visual impact. However, predictive questions in the environmental sciences often require the use of explicit quantitative models. One example where this has been carried out successfully is in the field of gravitationally induced sediment transport, especially in the cases of pyroclastic surges (Malin and Sheridan 1982), large-scale landslides and small-scale mudflows (Wadge 1988). As one example, Wadge and Isaacs (1988) have developed a procedure based on the integration of digital elevation models with a model for pyroclastic eruptions and have applied it to the Soufriere Hills Volcano on Montserrat in the West Indies. The model has essentially three variables, namely the site of the vent (which is locatable approximately using geophysical observations of pre-eruption activity), the height of the eruption column and an energy decay function. A conceptual representation of this model is given in Fig. 49.4.

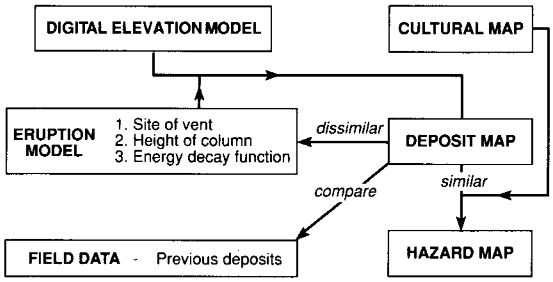

Fig. 49.4 Flow diagram of hazard modelling process (from Wadge and Isaacs 1988).

Application of this model allowed a variety of maps of geological deposits to be generated over the terrain model (Plate 49.5). Comparison of these maps with field data of previous deposits showed good agreement. Constraining the eruption model in terms of parameters (such as the height of the column and the angle which the collapse makes with the horizontal) and using values typical of such conditions allows production of a map showing the resultant expansion of pyroclastic deposits, along with pre-eruption hazards from fumarolic areas. This was used to generate a hazard map (Plate 49.6), which can be used not only as a basis for land use planning but also for planning the sequential evacuation of threatened areas. Wadge (1988) has applied this approach to a number of other gravity flows and slope instabilities and has shown how the use of raster-based GIS operations, combined with

Table 49.6 Examples of GIS operations which can be used for modelling of mass movements within a raster-based GIS (from Wadge 1988).

Class of operation	Specific operation	Example of application
Connectivity	Optimal path identification	Steepest downhill gradient path for the flow paths of debris flows
Characterizing neighbourhoods	Slope determination	First derivative of DEM for landslide stability and dynamic flow models
Overlay	Arithmetic	Calculate deposit thickness in energy balance model (e.g., cone height-original DEM height)
Pixel re-classification	(a) Isolating	Locate and display unstable pixels in landslide potential map
	(b) Contouring	Display thickness of deposit in energy balance models.
Region re-classification	Edge definition	Comparison of boundaries derived from models with those obtained from other sources (e.g. remotely sensed images and/or field work)

appropriate physical equations, can be used to model their various types of behaviour (Table 49.6).

B. Use of stratified sampling system for predictive purposes

Prediction of the behaviour of the environment is often dependent on spatially comprehensive sets of data; these are often not available and limitations in resources or time may preclude collection of new, census-type data. An alternative approach is to use a scheme of land and ecological characterization based on a relatively small sample of the total area. One such scheme is the land classification scheme of the UK's Institute of Terrestrial Ecology (Bunce, Barr and Whittaker 1982), which involves three main phases:

- *Land classification*. In this phase, classes were determined by analysis of 282 attributes recorded from 1228 1 km^2 sample areas drawn from a grid over the whole of the United Kingdom at 15 km by 15 km intersections of the National Grid. The attributes describe the climate, topography, geology and features of human occupancy which were used to define 32 land classes. The procedure allows the identification of indicators which can then be used to assign any other sample 1 km^2 to the appropriate class.
- *Ecological characterization*. In this second phase, field data are added to the initial classification to improve their characterization. The patterns of land use throughout the 1 km squares are mapped, along with ecological information such as hedgerow length and woodland composition.
- *Predictive phase*. Since the numbers of squares in the country belonging to each class are already known, it is possible to estimate any given factor for the whole country based on the conditions within a small sample of squares.

Among the many applications of this approach are the environmental assessment of changes in the farming industry in response to changes in the European Community's Common Agricultural Policy and, consequently, the prediction of impacts on the rural environment (Harvey 1986). Similarly, the area of land which could be available for energy crops was estimated for the Department of Energy and this was combined with existing land use information to determine an optimal allocation of land for wood energy production (Mitchell *et al.* 1983). In more local surveys in South-West England, the land classification scheme was used to provide input for forestry and agricultural models in order to assess the scope for economic integration of wood production in agriculture (Dartington Institute 1986). These various applications have demonstrated that survey, monitoring and modelling in the rural environment can be carried out effectively by combining limited amounts of sampled field survey data with a classification based on readily available environmental characteristics.

IMPROVING THE ACCESSIBILITY OF ENVIRONMENTAL DATABASES

The existence of environmental databases does not guarantee their easy access or availability. This is in part because of questions of confidentiality, cost and inappropriate physical storage mechanisms. It also arises, however, because of the highly dispersed character of many environmental data sets and the varying formats of available digital data. To solve the latter problem, considerable effort has been expended in defining and setting up exchange formats for digital data (see Guptill 1991 in this volume): in some environmental fields, there is now widespread national and even international agreement on such formats. In the field of cataloguing and indexing of data there is much less agreement on standards but recent work, as described below, has done much to demonstrate the potential of an automated graphic approach.

Because of the size and diversity of environmental data sets in terms of their formats, heterogeneity and coverage, users often face major hindrances in simply finding what data are available. Machine-readable catalogues form one partial solution to this problem but these do not necessarily provide easily accessible information on which data are available within a particular area. This problem is particularly acute when information is required from more than one particular data set. One possible solution to this is graphics-based spatial data retrieval, associated with automated graphic indexing. Typically, such experiments include the display at small scale of the environmental attributes, along with information on the availability of more detailed data together with their physical location.

One such indexing facility is the Natural Environment Research Council's Marine Atlas project of the seas around the British Isles (Mason and Townshend 1988; Robinson 1991). Figure 49.5 shows the main components of the system. Data and information from the data holders flow into a central Marine Information System data bank via a process of selection and processing carried out at the discretion of the data holders themselves. There is also a common data bank holding widely used data sets such as coastlines. Only a small sample of data will be included, since the main intent is not to provide a large comprehensive geo-referenced data bank but to provide meta-information (information about information) concerning the availability and quality of environmental data. Graphical display is via VDU displays or may take a number of other forms, such as catalogue lists or traditional atlas displays (Plate 49.7). This particular system is not intended to provide a completely comprehensive catalogue but is intended in part to act as a 'shop window' to make the user community more aware of the various marine data sets that are available. A similar approach has been proposed by Adlam, Clayton and Kelk (1988) in the development of a prototype of a central geosciences data index with a graphics interface. A much more general model and flexible approach for the explicit management of data sets, rather than just their cataloguing, has been suggested by Abel (1989). An interchange format has already been proposed for the exchange of directory-level information about data sets among information systems by NSSDC (1989).

CONCLUSIONS

The examples discussed in the previous section demonstrate the power of GIS for improving the quality of environmental databases and expanding their applications. It is important to restate, however, that the mere existence of these benefits does not in itself ensure the widespread adoption of GIS. Some large and valuable data sets and complex systems for their management already exist to support environmental science and its applications. Converting existing data sets to forms appropriate to GIS and achieving the transition from traditional to modern methods of spatial information handling will be a complex, expensive and time-consuming process: moreover, it will have to compete for resources with many other more glamorous activities. However, without the

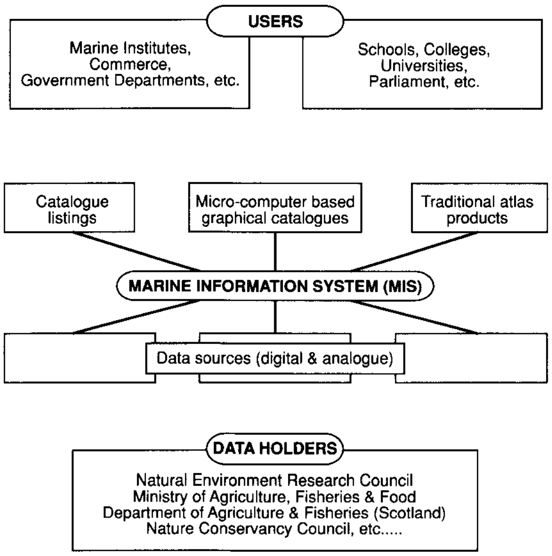

Figure 49.5 Proposed schematic outline of the NERC Marine Atlas (Robinson 1991; Mason and Townshend 1988).

necessary investment in the infrastructure of information handling, our ability to make best use of our data in addressing fundamental environmental questions will be profoundly damaged.

REFERENCES

Abel D J (1989) A model for data set management in large spatial information systems. *International Journal of Geographical Information Systems* **3**: 291–302

Adlam K H, Clayton A R, Kelk B (1988) A demonstrator for the National Sciences Geodata Index. *International Journal of Geographical Information Systems* **2**: 161–70

Allen J H (1988) The World Data Center System, international data exchange and global change. In: Mounsey H M (ed.) *Building Databases for Global Science.* Taylor & Francis, London, pp. 138–53

Bull G A (1960) Comparison of rain-gauges. *Nature* **185**: 437–38

Bunce R G H, Barr C J, Whittaker H A (1982) A stratification system for ecological sampling. In: Fuller R M (ed.) *Ecological Mapping from, Ground, Air and Space.* Institute of Terrestrial Ecology Symposium No. 10. Monk's Wood, Cambridgeshire, pp. 39–46

Burrough P A (1991) Soil information systems. In: Maguire D J, Goodchild M F and Rhind D W (eds.) *Geographical Information Systems: principles and applications.* Longman, London, pp. 153–69, Vol 2

Clark D M, Hastings D A, Kineman J J (1991) Global databases and their implications for GIS. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications.* Longman, London, pp. 217–31, Vol 2

Dartington Institute (1986) *The Potential for Forestry on the Culm Measures Farms of south west England.* Dartington Institute, Dartington Devon

Davis F W, Simonett D S (1991) GIS and remote sensing. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications.* Longman, London, pp. 191–213, Vol 1

Earth System Sciences Committee (1988) *Earth System Science: a closer view.* NASA, Washington DC.

Ellis M Y (1978) *Coastal Mapping Handbook.* United States Government Printing Office, Washington DC

Goward S N, Markham B, Dye D G, Dulaney W, Yang J (1991) Normalized difference vegetation index measurements from the Advanced Very High Resolution Radiometer. *Remote Sensing of Environment* **35**: 257–78

Guptill S C (1991) Spatial data exchange and standardization. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications.* Longman, London, pp. 515–30, Vol 1

Harvey D R (1986) *Countryside Implications for England and Wales of Possible Changes in the Common Agricultural Policy.* Centre for Agricultural Strategy, University of Reading UK

Hummel J, Reck R (1979) A global surface albedo model. *Journal of Applied Meteorology* **18**: 239–53

ICSU (International Council for Scientific Unions) (1989) *Guide to the World Data Center System: part I(a) Updates, Corrections and Additions to Part I.* ICSU, Boulder Colorado Intergovernmental Oceanographic Commission (1987) GF3, A general formatting system for geo-referenced data, vol 2, Technical description of the GF3 Format and code tables. *Intergovernmental Oceanographic Commission, Manuals and Guides* **17**: UNESCO, Paris

Justice C O, Townshend J R G, Holben B N, Tucker C J (1986) Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing* **6**: 1271–318

Langran G (1989) Accessing spatio-temporal data in a temporal GIS. *Proceedings of AUTOCARTO9.* ACSM/ASPRS, Falls Church Virginia, pp. 191–98

Lieth H (1975) Primary production of the major vegetation units of the world. In: Lieth H, Whittaker R H (eds.) *Primary productivity of the Biosphere* (Ecological Studies 14). Springer-Verlag, New York, pp. 203–15

Malin M C, Sheridan M F (1982) Computer-assisted mapping of pyroclastic surges. *Science* **217**: 637

Mason D C, Townshend J R G (1988) Research related to geographical information systems at the Natural Environment Research Council's Unit for Thematic Information System. *International Journal of Geographical Information Systems* **2**: 121–41

Mason D C, Corr D G, Cross A, Hogg D C, Lawrence D H, Petrou M, Tailor A M (1988) The use of digital map data in the segmentation and classification of remotely sensed images. *International Journal of Geographical Information Systems* **2**: 195–218

Matthews E (1983) Global vegetation and land use: new high resolution databases for climate studies. *Journal of Climatology and Applied Meteorology* **22**: 474–87

Millington A C, Townshend J R G, Kennedy P, Saull R, Prince S, Madams R (1989) *Biomass Assessment in the SADCC Region*. Earthscan Publications, London

Mitchell C W (1973) *Terrain Evaluation*. Longman, London

Mitchell C P, Brandon O H, Bunce R G H, Barr C J, Tranter R B, Downing P, Pearce M L, Whittaker H A (1983) Land availability for production of wood energy in Great Britain. In: Strub A, Cartier P, Scleser G (eds.) *Energy from Biomass. Proceedings 2nd European Community Conference, Berlin*. Applied Science, London, pp. 159–63

Mooneyhan D W (1988) Applications of Geographic Information Systems within the United Nations Environmental Programme. In: Mounsey H M, Tomlinson R F (eds.) *Building Databases for Global Science*. Taylor & Francis, London, pp. 315–29

Mounsey H M (1991) Multisource, multinational environmental GIS: lessons learnt from CORINE. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications*. Longman, London, pp. 185–200, Vol 2

NERC (1988) *Geographical Information in the Environmental Sciences*. (Report of the Working Group on Geographic Information), Natural Environment Research Council, Swindon

NSSDC (National Space Science Data Center) (1989) *Directory Interchange Format Manual, Version 1.0*. NASA Goddard Space Flight Center, Greenbelt Maryland

Olson J S, Watts J, Allison L (1983) *Carbon in Live Vegetation of Major World Ecosystems*. US Department of Energy contract No. W-7405–ENG-26. Oak Ridge Laboratory, Oak Ridge Tennessee

Pellew R A, Harrison J D (1988) A global database on the status of biological diversity: the IUCN perspective. In: Mounsey H M, Tomlinson R F (eds.) *Building Databases for Global Science*. Taylor & Francis, London, pp. 330–9

Quarmby N A, Saull R J (1990) The use of perspective views in local planning. *International Journal of Remote Sensing* **11**: 1329–30

Raper J F, Kelk B (1991) Three-dimensional GIS. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications*. Longman, London, pp. 299–317, Vol 1

Rasool S I, Ojima D S (1989) Pilot studies for remote sensing and data management. *International Geosphere Biosphere Program, Global Change Report No. 8.*

Robinson G R (1991) The UK digital Marine Atlas Project: an evolutionary approach towards a Marine Information System. *International Hydrographic Review*, Monaco. **68**: 39–51

Smith D R, Paradis A R (1989) Three-dimensional GIS for the Earth Sciences. *Proceedings of AUTOCARTO9*. ACSM/ASPRS, Falls Church Virginia, pp. 324–35

Topographic Science Working Group (1988) *Topographic Science Working Group Report to the Land Processes Branch, Earth Science and Applications Division*. NASA Headquarters Lunar and Planetary Institute, Houston

Tucker C J, Townshend J R G, Goff T E (1985) African land-cover classification using satellite data. *Science* **227**: 369–75

Wadge G (1988) The potential of GIS modelling of gravity flows and slope instabilities. *International Journal of Geographic Information Systems* **2**: 143–52

Wadge G, Isaacs M C (1988) Mapping the volcanic hazards from Soufriere Hills Volcano, Montserrat, West Indies using an image processor. *Journal of the Geological Society* **145**: 541–52

Webster R (1978) Mathematical treatment of soil information. *Proceedings of the 11th International Congress of Soil Science, Edmonton, Canada*, Volume 3. pp. 161–90

Webster R, Butler B E (1976) Soil classification and survey studies at Ginninderra. *Australian Journal of Soil Science* **14**: 1–24