4

Geographic data for environmental modelling and assessment

Bradley C. Reed, Jesslyn F. Brown and Thomas R. Loveland

ABSTRACT

Access to consistent and objective environmental data is a prerequisite for modelling. Many (global) models exist that require terrestrial environmental data. They include land-atmosphere interaction models, ecosystem process models, hydrologic models, and dynamic biosphere models; as explained in Chapter 2, these models are usually of a combined inductive-deductive construction. Modellers often face a deficit of appropriate and accurate environmental data to set the initial conditions of parameters in their models, as well as to validate model output. Until recently, there were very few observations of global-scale environmental phenomena from which to construct consistent scientific databases of vegetation, soils, and climate. A significant increase in activity to fill this void has resulted in improved databases of global topography, land cover, soils, and satellite imagery that provides information on vegetation dynamics. These data sets, governed by various use and cost policies, are available from national and international organizations. Gaining access to the appropriate data sets is one hurdle that modellers must clear; another is incorporating the geospatial data into their numerical simulation and forecasting models. Improving the communication between environmental data providers and data users is crucial for improving access to data of the appropriate scale and thematic content, refining data quality, and reducing the technical barriers to working with complex geospatial data.

4.1 INTRODUCTION

Data are the fundamental building blocks of scientific inquiry. At the beginning of the 1990s, there were few objective databases designed to support modelling, especially at a global scale. There were, of course, interpreted maps of vegetation, soils, and climate, often based on broad assumptions, but no direct, or even remote, observations of global-scale environmental phenomena. Recently there has been a significant increase of activity in building geospatial databases to support broad-scale modelling applications. These activities include modelling global topography, mapping regional- and global-scale land cover, and constructing soils databases.

Environmental assessments require data that provide a means to determine the geographic location of various land resources. In addition to describing the land characteristics, assessments often require attributes that describe various properties of the landscape, such as surface roughness, biomass, and slope. As a result, the collection of environmental baseline data involves intensive and costly efforts prior to the initiation of a project. Cost serves as an important incentive for encouraging the use of existing spatial data sets as an alternative to expensive data collection and assembly.

An International Geosphere Biosphere Programme (IGBP) report states that there are two important technical requirements that affect global environmental forecasts and assessments (IGBP 1992); (1) numerical models that account for the complex interactions and feedbacks of the Earth system and (2) validated, geographically referenced environmental data sets to determine parameters for these models. The geographically referenced data include both data for documenting and monitoring global change (such as land and sea temperature, and atmospheric concentrations of carbon dioxide and other trace gases) and data that characterize important forcing functions. The IGBP report concludes with the observation that land data are required by most IGBP core projects, and these data are a critical, but missing, element in models of global ecosystems and hydrology.

In a review of the role of land cover maps, Hall *et al.* (1995) describe the role of environmental data in land process models. For example, in these models, vegetation community composition is used to partition the landscape into functionally different regions. The specific land cover types represented in the database are related to the biological, thermodynamic, or chemical pathways. The seasonal dynamics of land cover are critical because of their influence on patterns of latent heat flux throughout the year, as represented by changes in turbulent exchange parameters, such as surface roughness and radiation exchange variables, such as albedo. They suggest the use of land cover data with functional categories that are directly related to properties such as energy, water, and nutrient cycling. They propose a set of functional classes for use in International Satellite Land Surface Climatology Project (ISLSCP) science initiatives but concede that a single categorization of land cover types is unlikely to meet all modelling requirements.

In Kemp's (1992) review of the role of spatial data in environmental simulation models, she identifies problems of environmental data faced by modelling groups. Although Kemp does not focus on specific data sets, she states that dealing with the range of spatial and temporal scales at which different processes are depicted in models is a formidable task. Kemp suggests that modelling has been hampered by the fact that critical environmental processes operate on many different time and spatial scales, and there may be scale thresholds at which critical processes change. This is complicated by the fact that land cover, soils, and geology are typically mapped by category, but mathematical models commonly use continuous data. As a result, the categorical information must be converted to parameter values (e.g. stomatal resistance, surface roughness, leaf area index) before being input to numerical calculations. This means that researchers must use measurements from limited field or laboratory studies, in combination with their scientific 'judgement', to set parameter values for models; that is, use a combination of induction and deduction as discussed in Chapter 2.

4.2 LAND-ATMOSPHERE INTERACTION MODELLING

Climate models incorporate a greater quantity and higher quality of terrestrial data compared to a decade ago. Both general circulation models (GCM) that are used to estimate global climates under specific conditions, and mesoscale meteorological models, which operate at regional levels, make use of land cover data to set parameters for land-atmosphere interactions. Some of the more detailed parameterizations of land-atmosphere interactions are land surface schemes coupled with climate models. Fluxes of moisture, energy, and momentum at the land-atmosphere boundary are key drivers of climate models. Land surface process models were developed to describe the effects of the environmental state on these fluxes. Results from coupled models demonstrate that the biophysical properties of vegetation and the physical properties of soils are important causal agents acting on regional and global climate (Bonan 1995).

Dickinson (1995) provides an overview of land cover scale issues in climate modelling. Traditionally, single representations of land cover were used to describe large grid cells (e.g. 40 km^2, 1° x 1°, or even 5° x 5°). Land cover data are used to assign the various parameters that characterize the role of vegetation in the model. These parameters are given as data arrays; that is, as a number or numbers for each type of land cover. One of the serious criticisms of present treatments of land processes in climate models is their failure to include many of the most essential aspects of sub grid-scale heterogeneity (Henderson-Sellers and Pittman, 1992). The latter issue can only be solved when land cover databases become available at resolutions higher than the grid cells used in the models. As previously mentioned, validation of the data sets input to the model is critical because poor quality input data will cause inaccurate output.

4.3 ECOSYSTEMS PROCESS MODELLING

There is an increasing recognition of the relationship between atmospheric process modelling and ecosystems processes and functions, including biogeochemistry and net primary productivity. Models of biogeochemistry processes and cycles are being developed and used to assess the influence of biogeochemical cycles on the physical climate system. This, in turn, has inspired significant advances in the understanding of the cycles that are responsible for the chemical compositions of the atmosphere, oceans, and surface sediments. The IGBP Global Analysis and Integrated Modeling Task Force determined that it is essential to quantify the characteristic dynamics of these cycles and their controlling factors in order to improve the understanding of global processes. This means that many types of data are needed for model input and validation of contemporary era models (IGBP 1994).

In biogeochemical modelling, it is important to include land cover attributes describing community composition or vegetation types so that physiological differences in assimilation rates, carbon allocation, and nutrient use efficiency, which influence CO_2 uptake during photosynthesis, can be modelled (Bonan 1995). Matson and Ojima (1990) stress a growing requirement for land cover data in atmospheric chemistry studies because of a need for data that can be used for the synthesis of global N_20, CH_4, and CO_2 fluxes. Models that include detailed canopy physiology, energy exchange, and microbial processes have been developed (e.g., the CENTURY model of Parton *et al.* 1988; BIOME-BGC developed by Hunt *et al.* 1996). These models have been used to evaluate the equilibrium response of ecosystems to doubled atmospheric CO_2 and associated climate change.

Issues concerning appropriate scale of study are also apparent in this type of modelling. Kittel *et al.* (1996) suggest that the accurate representation of the spatial distribution of driving variables and boundary conditions is necessary for simulations of regional to global ecological dynamics. However, developing such representations is difficult due to a lack of regional data sets and because of the coarse resolution needed to cover large domains. When land cover and soils are heterogeneous with respect to the model simulation grid, grid averages may not adequately represent existing conditions. As a result, data must have sufficient spatial detail for the statistical treatment of spatial heterogeneity and spatial coherence among variables that have complex relationships to ecosystem processes. The ability to incorporate sub grid information (i.e., estimates of the percentages of different land cover types within a grid cell) instead of using finer grid cells is necessary to keep the already high computational requirements of large-area simulations at a manageable level. This can only be accomplished, they state, with (1) the development of physically consistent model input data sets, (2) the efficient transfer of data between geographic information systems (GIS) and applications, and (3) the analysis of model results in both geographic and temporal contexts.

4.4 HYDROLOGIC MODELLING

Elevation, elevation derivatives (e.g. slope, aspect, drainage flow direction) and land cover data are required for modelling the physical processes of the hydrologic cycle (Steyaert *et al.*, 1994). However, land cover is frequently treated as uniform throughout a watershed. Vegetation interacts with the atmosphere by its influence on the return of water back to the atmosphere through evaporation and transpiration, thereby exerting a significant impact on the hydrologic cycle by influencing the exchange of energy, water, carbon, and other substances at the land surface-atmosphere interface. The IGBP Biological Aspects of the Hydrologic Cycle (BAHC) core project aims to serve as a catalyst for the improving future hydrologic models. BAHC plans stress the need to explicitly assess and represent the relevant topographic, vegetation, soil, and geologic parameters that control soil moisture and surface and subsurface hydrologic conditions in order to improve the results of hydrologic simulations (IGBP 1993).

4.5 DYNAMIC BIOSPHERE MODELLING

Advances in equilibrium terrestrial modelling have led to an evolution towards a dynamic biosphere model. Tian *et al.* (1998) describe the goals and challenges associated with modelling the terrestrial biosphere and accounting for its truly dynamic character. This 'new generation' model would ideally demonstrate ecosystem dynamics caused by natural or anthropogenic factors, in addition to the changing interactions and feedback among the cycles of energy, water, and biogeochemical elements. Again, data are lacking to help make models more mechanistic and to support model validation. Tian *et al.* (1998) call for measurements of whole ecosystem behavior, such as carbon and water fluxes, in addition to vegetation dynamics and land use change in order to make progress towards a dynamic biosphere model.

4.6 DATA ACCESS

Identifying and securing data that support environmental modelling and assessments is often a frustrating process. Data are commonly produced with a specific customer or application in mind and are typically not easily adapted to the needs of other applications. Often one must deal with vexing problems regarding data format, resolution, and content. In addition, simply finding the best data for an application is time consuming and difficult. For regional-scale applications, it is especially difficult to sort through all of the possibilities for data, many of which are under publicized, incompletely documented, and not fully supported. Consequently, data sets that become widely used are often 'advertised' by word-of-mouth and subsequent exchange of data among modellers. Data for developing countries are especially difficult to obtain as data sources are often outdated, unreliable, or simply unavailable. For many developing countries, the only recourse is to resort to coarser data at regional, continental, or global scales.

The number of primary data sources is limited for continental- and global-scale modelling applications. Although by no means complete, a good start towards acquiring global- or coarse-scale data at little or no cost is through one of the following sources:

- US Geological Survey's (USGS) Earth Resources Observation Systems (EROS) Data Center (edcwww.cr.usgs.gov)
- NASA Goddard Space Flight Center (GSFC) Distributed Active Archive Center (daac.gsfc.nasa.gov)
- United Nations Environment Programme Global Resources Information Database (UNEP-GRID) (grid2.cr.usgs.gov)
- The European Commission Joint Research Centre, Space Applications Institute (JRC, SAI) (www.sai.jrc.it).

Data access policies of the many data providers vary widely but, encouragingly, are starting to become more liberal. For example, UNEP-GRID has a policy that allows U.N. organizations, inter-governmental organizations, scientific and academic organizations, and non-governmental organizations to request data. However, data requests from private firms and individuals will not be filled except in exceptional cases.

The US Government policy calls for cost recovery, with data prices based on the cost of filling user requests rather than on the value of the data. The USGS, one of the major spatial data providers in the US, follows this policy. The USGS is also moving toward a goal of making digital data available at no cost via electronic transfer. However, full no-cost access to electronic data will not be feasible until bandwidth increases for all providers and users

Recently JRC-SAI and the European Umbrella Organisation for Geographic Information (EUROGI) sponsored a joint workshop to address the issue of geographic data policy in the European Union (Craglia *et al.*, 2000). Policies on geographic data vary widely in Europe. With respect to pricing, there is often a distinction between 'essential' data, such as ownership, administrative boundaries and roads, and 'non-essential' data that are value-added. Essential data are often accessible free of charge, while non-essential data may have a charge. Where there are charges for data, the policy usually states something to the effect that price should not deter the use of data. In reality, it is difficult to sort through all of the various data access issues, and this, unfortunately, discourages data usage. With the rapid developments in computing and the Internet, many mapping agencies are being forced to re-examine data policies. Among actions resulting from the workshop, the European Commission seems ready to establish a framework for developing a more cohesive geographic information policy.

4.7 GLOBAL DATABASES

The amount of geographic data available for environmental applications has rapidly increased over the past two decades. The following section describes widely available land-related data that are commonly applied in broad-scale environmental modelling using GIS and remote sensing.

4.7.1 Multiple-theme global databases

In recent years, ISLSCP has performed a lead role in the Global Energy and Water Cycle Experiment (GEWEX). Its primary responsibilities concern land-atmosphere interactions, with special focus on process modelling, data retrieval algorithms, field experiment design, and development of global data sets to supply modellers with data for land-biosphere-atmosphere models. The data sets provide fields for model initialization and forcing functions. In 1994, the ISLSCP Initiative I data sets were produced and published.

The data, distributed both on CD-ROM and online, are organized into five groups:

- Vegetation
- Hydrology and Soils
- Snow, Ice, and Oceans
- Radiation and Clouds
- Near-Surface Meteorology.

The ISLCSP team gathered data from various sources and processed each data layer for spatial and temporal consistency. Generally, the Initiative I global data layers are distributed on a 1° by 1° grid and cover a 24-month period (1987–1988) at monthly time steps. Building on the experience of the first initiative, a project to produce the ISLSCP Initiative II data collection project was launched in 1999 with funding from NASA's Hydrology Program. The project objectives include producing and validating an expanded series of global data sets. Meterological data are produced at a 1° x 1° spatial resolution; topography, soils, and vegetation layers will be produced at finer scales (half- and quarter-degree). These data layers make up a comprehensive set covering a 10-year period from 1986 to 1995. Prototype data are scheduled for release in late 2000, with the publication of final data sets scheduled for 2002. More information and ftp access to data layers can be found on the Internet (http://islscp2.gsfc.nasa.gov/).

Another multi-theme database is the Digital Chart of the World. This data set includes a global base map with many features and attributes including administrative boundaries, coastlines, hydrologic features, and urban features. The database is distributed from the Environmental Systems Research Institute, Inc (ESRI). This is a private company in the US that also distributes geospatial software. This data set is distributed on CD-ROM media for a cost. It can be accessed from the ESRI web page (at http://www.esri.com).

Other multiple-theme global databases include the Global Map 2000 project. The Japan Ministry of Construction first proposed the concept of the Global Map 2000 in 1992, and in 1994 the Geographical Survey Institute of Japan proposed the first specifications for the project. Global Map 2000 currently involves the participation of over 70 countries. The main objective of the project is to bring mapping agencies from across the world together to develop and provide easy and open access to global digital geographic information at a 1:1,000,000 scale, an appropriate scale to ratify and monitor global agreements and conventions for environmental protection, and to encourage economic growth within the context of sustainable development. To be viable, the initial project will use existing global data whenever possible and will also seek to improve the reliability and accuracy of these data sets. The upgrading of existing data is expected to involve editing or replacing parts of it with data from other sources. For more information, see (http://www1.gsi-mc.go.jp/iscgm-sec/index.html).

4.7.2 Heritage global land cover databases

In the 1980s, efforts were undertaken to develop very coarse resolution databases (approximately 1° x 1°) that could be used by researchers working with GCMs. The global databases combined existing maps with different cover classes into coarse-scale mosaics of categorical data. The source maps were sometimes out of date and usually required a translation in class names between the map classification and the desired classification scheme. The procedures to create appropriate land surface data for modelling included extensive reclassification or aggregation of pre-existing vegetation information.

The Olson and Watts (1982) database, generally referred to as the Olson Global Ecosystems map, originally consisted of 49 categories. These categories represented natural, anthropogenic, or natural/modified mosaics of land cover. The Olson database, with a spatial resolution of 0.5° x 0.5°, has undergone continual updating and improvement. The current Olson Global Ecosystems legend now consists of 96 land cover classes. The Matthews (1983) global database consists of two data layers; the primary layer is a 32-category map of potential natural vegetation with a grid resolution of 1° x 1° and the secondary layer includes estimates of the percentage of cultivated land for each grid cell. The two layers can be combined to produce a map that approximates actual land cover. Like the Olson map, the Matthews map was compiled from an exhaustive review of hundreds of source materials, including maps, texts, atlases, and statistical summaries. A third global database is the Wilson and Henderson-Sellers global land cover database (1985). This database has a spatial resolution of 1° x 1° and contains 53 categories of land cover. Each grid cell includes a primary and secondary land cover attribute. The land cover categories consist of physiognomic elements, including canopy density, and have inferences to land use activities.

4.7.3 Global land cover from satellite data

One of the driving forces improving global land data sets was the IGBP-Data and Information System's (DIS) promotion of the development of a 1 km global land data product from the National Oceanic and Atmospheric Administration's advanced very high resolution radiometer (NOAA-AVHRR) (Eidenshink and Faundeen, 1994). The development of the data set, begun in 1992, was sponsored by the USGS, NASA, NOAA, and the European Space Agency (ESA). This data set provides satellite observations processed in a consistent manner into 10-day composite data.

The Land Cover Working Group (LCWG) of the IGBP-DIS was responsible for designing and fostering the development of an improved global land cover data set. Through a series of international workshops, the LCWG finalized a strategy and definition for a global land cover product (DISCover) based on 1 km AVHRR data (Belward 1996). The USGS EROS Data Center (EDC), in partnership with national and international agencies and universities, developed a global land cover characteristics database described in detail by Loveland *et al.* (2000). The database was constructed through the unsupervised classification of continental-scale, time-series normalized difference vegetation index (NDVI) data with extensive post-classification stratification. The global database contains 961 distinct seasonal land cover regions that are translated into a number of more general, predefined classification schemes, such as the USGS Anderson System, IGBP, Olson Ecosystems, and those required by the Biosphere Atmosphere Transfer Scheme (BATS) and Simple Biosphere (SiB) models. The database layers, formatted as raster grid images, are distributed free of charge at (http://edcdaac.usgs.gov/glcc/glcc.html).

The University of Maryland also produced a global land cover product using the same base AVHRR data, but the data were classified using a classification tree approach (Hansen *et al.* 2000). Several metrics derived from temporal AVHRR data, such as minimum red reflectance, peak NDVI, and minimum channel-three brightness temperatures, were used to predict class memberships. The resulting product contains 14 land cover classes. These data are available through the University of Maryland Laboratory for Global Remote Sensing Studies (http://www.geog.umd.edu/landcover/global-cover.html).

The EDC and University of Maryland global land cover products represent significant strides toward resolving the land cover information needs of environmental modellers. These data sets provide consistent global products that can be used to estimate parameters for land-atmosphere interactions with sufficient spatial resolution to describe subgrid cell heterogeneity for most models. The primary weaknesses of the products are those imposed by the limitations of the AVHRR satellite sensor. Many of these weaknesses will be addressed in current efforts at land cover characterization using data from the SPOT Vegetation sensor and the Moderate Resolution Imaging Spectroradiometer (MODIS).

4.7.4 Topographic data

One common input to a wide variety of terrestrial and land-atmosphere interaction models is topographic or digital elevation model (DEM) data. The recently developed global model, GTOPO30, reflects a significant improvement in the quality and resolution of DEMs over previous models (Gesch *et al.* 1999). Elevations are spaced at 30-arc-second intervals (about 1 km) with a reported overall vertical accuracy of 70 m.

A team from the USGS EDC assembled topographic data from eight different sources to derive elevation values for GTOPO30. The primary source was the Digital Terrain Elevation Data (DTED) product of the National Imagery and Mapping Agency. GTOPO30 has no use or redistribution restrictions and is available at low or no charge, depending on the user's choice of access. The global DEM is distributed through the EROS Data Center Distributed Active Archive Center. Users may either download the data electronically (the global dataset is distributed in tiles) or order the data on CD-ROM for a nominal fee via the Internet: (http://edcdaac.usgs.gov/gtopo30/gtopo30.html).

The EDC is also creating a HYDRO1k data set for use in hydrologic and morphometric studies. Development of the HYDRO1k database was made possible by the completion of the 30-arc-second DEM (GTOPO30) at the EDC in 1996 (Gesch *et al.* 1999). A goal of the HYDRO1k project is the systematic development of hydrologically sound derivative products for all the land masses of the globe (with the exception of Antarctica and Greenland). Additionally, topologically sound representations of the Earth's rivers and basins are being developed, which can be used for comparison among rivers at corresponding scales.

The HYDRO1k data sets are being developed on a continental basis and provide a suite of six raster and two vector data sets. The raster data sets are the hydrologically correct DEM, derived flow directions, flow accumulations, slope, aspect, and a compound topographic (wetness) index. The derived streamlines and basins are distributed as vector data sets. The HYDRO1k data sets are available over the Internet at (http://edcdaac.usgs.gov/gtopo30/hydro/index.html).

4.7.5 Soils data

The FAO has produced a global soil texture data set known as GLOBTEX (FAO 1970). This digital database for the Soil Map of the World represents the most comprehensive available global soils data. The source maps were compiled by the FAO and published by UNESCO in 10 volumes between 1970 and 1978. The map references 106 FAO soil types and includes the texture and slope of the dominant soil in each area. Texture is defined in three classes (coarse, medium, and fine) and slope is defined in three classes (0–8 percent, 8–30 percent, and >30 percent slopes).

Several attempts have been made to generalize the FAO data for use in climate models (Zobler 1987; Wilson and Henderson-Sellers 1985). They encoded the FAO maps on 1-degree or half-degree cells and considered only the characteristics of the dominant soil of the cell. In 1984, the FAO map was digitized by the ESRI. The resulting database includes the full spatial detail contained on the original published maps and the associated soils (approximately 20 per cent by area) and included soils (approximately 10 per cent by area) within the map units.

A digital version of the database is available through the Web page for UNEP-GRID at (http://grid2.cr.usgs.gov/data/fts.html).

4.7.6 Global population

Population data sets provide significant information concerning the human-factor in modelling. One data set, the Gridded Population of the World, is distributed by CIESEN (http://www.ciesen.org). The data set includes population for multiple years (1990–1995) at a 2.5 minute grid cell resolution. Another world-wide database of ambient population for the year 1998 has a 30-second resolution. It was made using census counts and then deriving probability coefficients based on a number of information layers (roads, slope, land cover, and night-time lights satellite data). The data set, known as LandScan, was created at Oak Ridge National Laboratory, US Department of Energy and is available from (http://www.ornl.gov/gist/projects/LandScan/landscan.doc).

4.7.7 Satellite data

Although most environmental modelling applications require processed information (e.g., land cover, elevation, or soils), some models are being adapted to directly enter data collected from remote sensing satellites. The choice of which satellite data to use becomes a question of spatial resolution, spectral characteristics of the satellite sensor, the frequency of coverage, and the areal coverage or extent (swath size). Table 3.1 provides a summary of the more commonly used satellite sensors. These sensors vary in swath width from 12 km to 2400 km and in spatial resolution from 1 m to over 1 km. Generally, the cost of these products is inversely proportional to the spatial resolution. The resolution of satellite data used in models must satisfy information requirements through spatial discrimination of land features, providing flexible, subgrid cell parameterization, with a data volume that can be accommodated in computationally complex models. Data sets with fine spatial detail have costly data storage requirements that increase the time required for modelling exercises. Thus, a balance must be found between spatial resolution and associated data volume efficiency.

Most models require data covering large land areas, which until recently meant the use of AVHRR data. These data have been analyzed, scrutinized, and criticized for nearly two decades. There is little doubt that, even with the shortcomings of AVHRR data, studies using this sensor have served as a strong bridge between data providers and environmental modellers. The encouraging results gained from AVHRR studies in the field of land cover characteristics (e.g., Loveland *et al.* 2000; Hansen *et al.* 2000) and the derivation of biophysical parameters from land cover have helped to drive requirements for improved large-area sensors, such as SPOT4 Vegetation and MODIS. MODIS, in particular, is specifically designed to provide new, innovative land products that are especially designed to support modelling applications (Justice 1998). In addition to providing satellite data, the MODIS Land Group is providing higher-level data products that are specifically designed to support the needs of global to regional monitoring, modelling, and assessment, such as leaf area index, net primary production, albedo, and others. For the latest information on MODIS land products, see: (http://modis-land.gsfc.nasa.gov).

4.8 SUB-GLOBAL SCALE DATABASES

In addition to the global efforts described in previous sections, there are numerous environmental data sets that cover large but sub-global regions. Presenting a comprehensive treatment of these databases is beyond the scope of this chapter. However, this section will identify some sources of sub-global data sets and provide examples.

Regional or national datasets cover as many data themes as their global counter parts (i.e. land cover, topography, climate, economic, population), and have varied sources, costs, and distribution policies. These types of data sets are often even more difficult to locate than the global data mentioned above. The advantages they offer may include greater spatial detail, better local and regional consistency, and improved accuracy.

4.8.1 Regional land cover mapping

There are several examples of relatively current regional land cover databases. The National Land Cover Database (NLCD) covering the conterminous US is produced by the USGS EDC and is based on 1992-era Landsat TM imagery. It is available at no charge from (http://landcover.usgs.gov/nationallandcover.html). The NLCD maps land cover into a 21-class legend. The CORINE (coordination of information on the environment) database, produced by the European Environment Agency and JRC, is based on both SPOT and Landsat data and covers most of the European Union. This database is produced from 1990-era data and is available through (http://natlan.eea.eu.int/datasets.html). Both the NLCD and CORINE databases will be updated based on year 2000 satellite data, which will provide valuable information on land cover change during the 1990s.

Another project in Europe, Pan-European Land Use and Land Cover Monitoring (PELCOM) had the objective of creating 1 km resolution land use and land cover data from AVHRR data that could be updated frequently. The project was led from the Netherlands (ALTERRA/Green World Research), and had international project partners from Europe. This land cover data set is available at http://137.224.135.82/cgi/geodesk/geodata/pelcom.html.

The Africover project, sponsored by the Food and Agriculture Organization (FAO) of the United Nations, has a goal of establishing a digital database on land cover for all of Africa at a 1:200,000 scale (1:100,000 scale for small countries). The implementation of the project in East Africa is currently best developed (http://www.africover.org). For a more extensive review of activities in global and regional land cover mapping, see DeFries and Belward (2000).

4.8.2 Topographic databases

There are a number of national-level topographic data sets available. Many of them are accessible from the appropriate government mapping agency. In Australia, for example, the Australian Surveying and Land Information Group (AUSLIG) is responsible for distributing a number of different topographic data sets. They include gridded digital elevation data at several different resolutions (for example, 18-second, nine-second, and three-second resolution data) and point elevation data sets. Data sets may be accessed from their site (http://www.auslig.gov/products/).

In the US, the National Elevation Dataset (NED) has been developed by the USGS by merging high-resolution elevation data into a seamless raster format. The data set has a resolution of one arc-second resolution for the conterminous US, Hawaii, and Puerto Rico and two arc-seconds for Alaska. NED will be updated periodically to incorporate the highest resolution and quality data possible. In areas where 7.5 minute DEM data are available, those data are used to generate NED). Where they are not yet available, coarser resolution DEM data (for example, 30 minute or one degree DEM) products are used to generate NED. Users may order their area of interest interactively, and there is a cost for the data, online at http://edcnts12.cr.usgs.gov/ned/.

4.8.3 Administrative and census data

Administrative data, including political boundaries, administrative features, and place names, are frequently distributed by national mapping agencies. In Britain, the Ordnance Survey produces 1:10,000 scale raster data in tiff format. It is sold in 5 by 5 km tiles. The basic product includes information on place names, houses, streets, rivers, etc. Access to the Ordnance Survey of Britain can be found at http://www.ordnancesurvey.co.uk/. Similarly, in Ireland, the Ordnance Survey Ireland has a mandate to provide three types of map services to Ireland. These are detailed large scale maps at 1:1,000 scale in urban locales, rural areas covered by 1:2,500 and 1:5,000 scale maps, and tourist maps at 1:50,000 scale. Access to all of these products is through http://www.irlgov.ie/osi/.

Much of the US Census Bureau data (http://www.census.gov/) is inherently statistical. However, they distribute a number of geographic products with versions available reflecting the year 2000 census data. The Census 2000 Tiger/Line Files is the public version of the US Census Bureau's digital database of geographic features for the US This product includes Census 2000 tabulation block numbers, address ranges, and governmental unit boundaries. They also distribute a number of maps depicting the geographic units used for census data collection, including census tract outlines, census blocks, voting districts, and state legislative district outlines.

4.8.4 Data clearinghouses

Some organizations have missions to provide data for applications that require regional environmental data. Data clearinghouses, with entry points on the Internet that provide access to data and metadata, are becoming more common as technology has developed. The following examples outline several different types of clearinghouses.

The Africa Data Dissemination Service (ADDS) is managed by the US Agency for International Development (USAID) in collaboration with the USGS to support the Famine Early Warning System Network. The project goals are to identify areas that might experience food supply problems due to famine, flood, or other environmental conditions. The ADDS includes data from 17 sub-Saharan African countries. Many multi-theme data sets are distributed from the ADDS (located at http://edcintl.cr.usgs.gov/adds/adds.html). These include satellite data, administrative boundaries, digital elevation models, hydrology, roads, and vegetation. Tabular statistics by country are also available on agriculture and precipitation.

As mentioned earlier, the UNEP-GRID provides data to assist the international community and individual nations to make sound decisions on environmental planning and resource management. The GRID network consists of several nodes. The Geneva node (http://www.grid.unep.ch/) supports access to over 200 data sets, both global and sub-global in geographic coverage. Multiple data themes are included, such as administrative boundaries, atmosphere, biodiversity, oceans, climate, and soils. The North American GRID node (http://grid2.cr.usgs.gov/) node also distributes data sets on population and administrative boundaries.

The Australian Department of the Environment and Heritage supports its own data clearinghouse, the Australia Spatial Data Directory (at http://www.environment.gov.au/search/databases.html). Data sets from a variety of themes, such as air, biodiversity, land conservation, and industry, are all linked from this location. The site also supports an online map-making function which will add layers interactively depicting features such as hydrology, infrastructure, conservation areas, and species distributions.

The Inter-American Geospatial Data Network (IGDN), formed by USAID in partnership with the USGS, promotes information infrastructure (especially Internet capabilities) in the Western Hemisphere for electronic access to information on geospatial data. This site (at http://edcsnw3.cr.usgs.gov/igdn/in-dex.html) does not house data sets, but does provide links and searching capabilities to geospatial data providers throughout the Western Hemisphere.

4.9 THE ROLE OF THE END-USER IN THE USGS GLOBAL LAND COVER CHARACTERIZATION PROJECT

Improving communication between data users (e.g., modelers) and the producers of land surface environmental data has been a key component to improving terrestrial data sets. If a data set is to meet the diverse data requirements of a broad environmental modelling community, then avenues for positive communication flow must be opened. One example of incorporating user input is the USGS global land cover characterization project.

User feedback (that is, the comments, suggestions, constructive criticism, and requests from the users of the global land cover database) played a vital role in both setting the requirements for the global land cover effort and assessing the quality and utility of the database through the peer review process (Brown *et al.* 1999). The IGBP global land cover strategy explicitly called for a validation protocol to assess the accuracy of the DISCover product (Belward 1996). This process consisted of three primary activities: (1) peer review of the preliminary continental databases, (2) comparisons with other land cover data sets, and (3) a formal, statistically sound, accuracy assessment. The USGS mapping team used peer review not only to identify and correct classification problems, but to gain feedback on a wider range of database issues, including data utility in applications, suggestions as to data set improvements, and insights into the technical challenges faced by users working with the data. Users provided many suggestions; frequently identifying areas with land cover label problems. The comments, when collaborated using other evidence, were a critical resource for refining the global land cover characteristics database, leading to the generation of a new version, released in 2000. Approximately 10 per cent of the land area has been revised since the first release, largely based on user feedback. In addition, improvements in documentation, alternative projections, and additional classification schemes have all been added to the database because of open communication with users.

The accuracy assessment was performed in accordance with the specifications developed by the IGBP Validation Working Group and approved by the Land Cover Working Group. The approach was a stratified random sample of DISCover and the use of higher spatial resolution satellite imagery (either Landsat TM or SPOT) to test class accuracy (Scepan 1999). After consideration of both time and personnel required to execute the accuracy assessment, a random sample was taken within each of the DISCover classes (excluding snow and ice and water) with a target accuracy of 85 per cent and a confidence level of 95 per cent. A manageable sample size of 25 for each class, based on the expected accuracy and confidence level results in an interval with a range of +/- 0.143 (Scepan 1999).

The validation samples were aggregated into 13 regions and several expert image interpreters (EII) were enlisted for image interpretation for each of the regions. The EIIs assembled for validation workshops and three EIIS were assigned to validate each sample using the high resolution imagery. The VWG protocol specified that at least two of three EIIs must agree on the land cover type before a given DISCover sample was accepted as correct. If all three disagreed and identified three different land cover types for a sample, the DISCover classification was considered to be in error. The average class accuracy was 59.4 per cent with single class accuracies ranging between 40 and 100 per cent. If the accuracies are weighted according to class size, the area weighted accuracy rises to 66.9 per cent (Scepan 1999). For more detail on the validation procedure and related studies see the special issue of Photogrammetric Engineering and Remote Sensing: Global Land Cover Data Set Validation, September 1999.

4.10 SUMMARY

A variety of models have a growing need for improved environmental data sets to better approach biophysical parameterization. Significant advances have recently been made in the development of large-area databases to support such models and environmental assessments. Consistency in database construction, as well as the scale and areal coverage of databases, is improving. This allows issues such as subgrid cell heterogeneity to be better addressed. Rapid improvements in technology, computing, and the Internet, along with the decrease in the cost of disk storage, have changed the arena of land surface mapping. Several large area databases are now available free of charge by means of the Internet. Although recent developments have resulted in improved environmental data, there are still serious impediments to using the data owing to proprietary data formats, technology limitations (i.e. bandwidth), inconsistent data policies, and the inherent complexity of spatial data (e.g. disparate data resolutions, map projections, etc.). It appears that governmental data distribution policy shifts will continue to encourage the availability of high-quality, low-cost data. Even though there is a trend towards improving data, it is necessary for environmental data providers and environmental modelling and assessment researchers to continue to keep lines of communication and collaboration open to improve our understanding of the Earth's environment.

4.11 REFERENCES

Belward, A.S. (ed.), 1996. *The IGBP-DIS Global 1-Km Land Cover Data Set (DISCOVER): Proposal and Implementation Plans.* IGBP-DIS Working Paper No. 13, IGBP-DIS Office, Meteo-France, 42 Av. G. Coriolis, 31057 Toulouse, France, 61 pp.

Bonan, G.B., 1995, Land-Atmosphere Interactions for Climate System Models: Coupling Biophysical, Biogeochemical and Ecosystem Dynamical Processes. *Remote Sensing of Environment*, **51**, 57–73.

Brown, J.F., Loveland T.R., Ohlen D.O. and Zhu Z., 1999, The global land-cover characteristics database: the users' perspective. *Photogrammetric Engineering & Remote Sensing*, **65**, 1069–1074.

Craglia, M., Annoni A. and Masser I. (eds), 2000, In: *Proceedings of the EUROGIEC Data Policy Workshop: Geographic Information Policies in Europe: National and Regional Perspectives*, European Communities, 42 pp.

DeFries, R.S. and Belward A.S. (eds), 2000, Global and regional land cover characterization from remotely sensed data. *Special Issue: International Journal of Remote Sensing*, **21**, 1081–1560.

Dickinson, R.E., 1995, Land Processes in Climate Models. *Remote Sensing of Environment*, **51**, 27–38.

Dobson, J.E., Bright, E.A, P.R., Durfee, R.C. and Worley, B.A., 2000, Landscan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing*, **66**, 849–857.

Eidenshink, J.C. and Faundeen, J.L., 1994, The 1 km AVHRR global land data set: first stages in implementation. *International Journal of Remote Sensing*, **15**, 3443–3462.

Food and Agriculture Organization of the United Nations, 1970–1978, *Soil Map of the World*, Scale 1:5,000,000. Volumes I-X. United Nations Educational Scientific and Cultural Organization, Paris.

Gesch, D.B., Verdin, K.L. and Greenlee, S.K., 1999, New land surface digital elevation models cover the earth. *EOS, Transactions, American Geophysical Union*, **80**, 69–70.

Hall, F.G., Townshend, J.R. and Engman, E.T., 1995, Status of Remote Sensing Algorithms for Estimation of Land Surface Parameters. *Remote Sensing of Environment*, **51**, 138–156.

Hansen, M.C., Defries, R.S., Townshend, J.R.G. and Sohlbert, R., 2000, Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, **21**, 1331–1364.

Henderson-Sellers, A. and Pitman, A.J., 1992, Land-Surface Schemes for Future Climate Models: Specification, Aggregation and Heterogeneity. *Journal of Geophysical Research*, **97**, 2687–2696.

Hunt, E.R. Jr., Piper, S.C., Nemani, R., Keeling, C.D., Otto, R.D. and Running, S.W., 1996, Global Net Carbon Exchange and Intra-Annual Atmospheric CO_2 Concentrations Predicted by an Ecosystems Process Model and Three-Dimensional Atmospheric Transport Model. *Global Biogeochemical Cycles*, **10**, 431–456.

International Geosphere Biosphere Programme, 1992, *Improved Global Data for Land Applications*, J.R.G. Townshend, (ed.) IGBP Global Change Report No. 20, International Geosphere Biosphere Programme, Stockholm, Sweden, 87 pp.
International Geosphere Biosphere Programme, 1993, *Biospheric Aspects of the Hydrological Cycle (BAHC): The Operational Plan.* IGBP Global Change Report No. 27. International Geosphere Biosphere Programme, Stockholm, Sweden, 103 pp.
International Geosphere Biosphere Programme, 1994, *IGBP Global Modeling and Data Activities 1994–1998*. IGBP Global Change Report No. 30, International Geosphere Biosphere Programme, Stockholm, Sweden, 87 pp.
Justice, C.O. (ed.), 1998, The moderate-resolution imaging spectroradiometer (MODIS); land remote sensing for global change research. *TRANS IEEE Geoscience and Remote Sensing*, **36**, no. 4.
Kemp, K.K., 1992, Spatial Models for Environmental Modeling with GIS. In: *Proceedings: 5th International Symposium on Spatial Data Handling*, v. 2. International Geographic Union, Charleston, SC 524–533.
Kittel, T.G.F., Ojima, D.S., Schimel, D., McKeown, D., Bromberg, R., Painter, J., Rosenbloom, T. and Parton, W., 1996, Model GIS Integration and Data Set Development to Assess Terrestrial Ecosystem Vulnerability to Climate Change. In: *GIS and Environmental Modeling: Progress and Research Issues*, M.F. Goodchild, L. Steyart, B. Parks, (eds) GIS World, Inc., Fort Collins, CO: 293–297.
Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L. and Merchant, J.W., 2000, Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, **21**, 1303–1330.
Matthews, E., 1983, Global Vegetation and Land Use: New High Resolution Data Bases for Limited Studies. *Journal of Climatology and Applied Meteorology*, **22**, 474–487.
Matson, P.A. and Ojima, D.S., 1990, *Terrestrial Biosphere Exchange with Global Atmospheric Chemistry: Terrestrial Biosphere Perspective of the IGAC Project*. IGBP Global Change Report 13, International Geosphere Biosphere Programme, Stockholm, Sweden, 103 pp.
Olson, J.S. and Watts, J.A., 1982, *Major World Ecosystem Complex Map*. Oak Ridge, TN: Oak Ridge National Laboratory.
Parton, W.J., Stewart, J.W.B. and Cole, C.V., 1988, Dynamics of C, N, P and S in Grassland Soils: A Model. *Biogeochemistry*, **5**, 109–131.
Scepan, J., 1999, Thematic validation of high-resolution global land-cover data sets. *Photogrammetric Engineering & Remote Sensing*, **65**, 1051–1060.
Steyart, L.T., Loveland, T.R., Brown, J.F. and Reed, B.C., 1994, Integration of Environmental Simulation Models with Satellite Remote Sensing and Geographic Information Systems Technologies: Case Studies. In: *Proceedings: Pecora 12 Symposium on Land Information from Space-Based Systems*, American Society of Photogrammetry and Remote Sensing, Bethesda, MD, 407–417.
Tian, H., Hall, C.A.S. and Qi, Y., 1998, Modeling Primary Productivity of the Terrestrial Biosphere in Changing Environments: Toward a Dynamic Biosphere Model. *Critical Reviews in Plant Sciences*, **15**, 541–557.

Wilson, M.F. and Henderson-Sellers, A., 1985, A Global Archive of Land Cover and Soils Data for Use in General Circulation Models. *Journal of Climatology*, **5**, 119–143.

Zobler, L., 1987, A world soil hydrology file for global climate modelling. p. I-229-244. In: *International Geographic Information Systems Symposium: The Research Agenda*, November 15-18, 1987, Arlington, Virginia, Proceedings, Vol. 1, Association of American Geographers. National Aeronautics and Space Administration, Washington, D.C.

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.