1994

Integration of Scientific Information Management and Environmental Research

Susan G. Stafford

James W. Brunt

William K. Michener

1

Integration of scientific information management and environmental research

Susan G. Stafford, James W. Brunt and William K. Michener

A new dynamic between science and technology is changing the management and analysis of environmental information. As a result of the expanding scope of environmental research, developers of environmental databases must now address difficult and diverse issues: the wide variety of data being considered for inclusion; the size and complexity of databases being collected; their design, development and utilization; the different scales of data collection; increasingly sophisticated analytical requirements; and the integration of data from various sources and disciplines. Increased attention to spatial variability, scale and the integration of basic and applied research to address societal issues has several implications for the management of scientific information. These new research directions have created a need for rapid and easier data analysis; timely, broad-scale, high-resolution data; new analytical approaches; better sampling resolution in space; and a shift in focus from data to information to knowledge. This will require the use of remote sensing, geographic information systems (GIS), spatial statistics and other methods. Several challenges are discussed for future environmental information management and analysis systems: distributed analytical environments, database management systems, integrated GIS and remote sensing in the time domain, user interfaces, visualization software and knowledge discovery, improved spatial sampling resolution and standardization.

Introduction

Historically, environmental research has been conducted as small-scale studies involving one or a few investigators in a single discipline and funded for relatively short periods. Recently, increased understanding of small-scale environmental patterns and processes, coupled with burgeoning interest in landscape, regional and global patterns and processes, has led to the development and funding of studies addressing broad-scale, long-term questions. For example, the Long-Term Ecological Research Program (LTER), initiated by the National Science Foundation in the early 1980s, supports long-term investigations that could not be addressed effectively in the normal 1–3-year funding cycle of most US granting agencies (Franklin *et al.*, 1990). These and related programmes generally involve experiments and monitoring lasting years, decades or

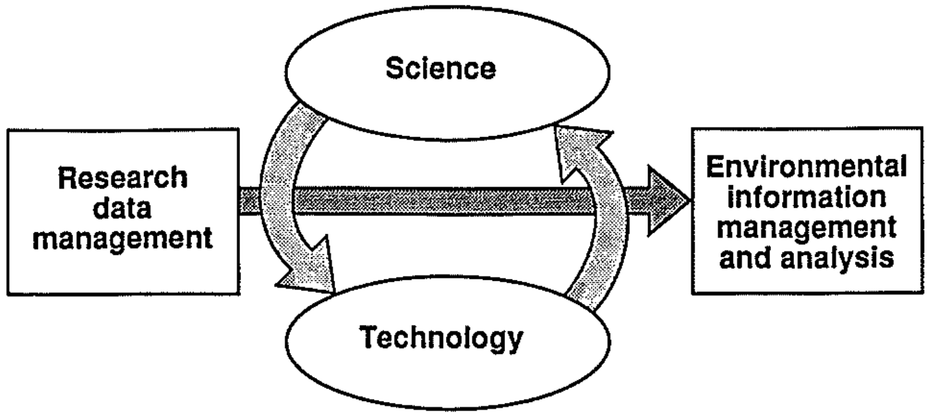

Figure 1.1 A new dynamic between science and technology has forced evolution of the way in which environmental information is now managed and analyzed.

centuries (e.g., Harmon's (1992) 200-year study of log decomposition), and they frequently include many scientists from several disciplines.

As a result of the expanding scope of environmental research, developers of environmental databases must now address difficult and diverse issues: the wide variety of data, both spatial and non-spatial, being considered for inclusion; the size and complexity of databases being created; their design, development and utilization; the different spatial and temporal scales of data collection; increasingly sophisticated analytical requirements; new systems (e.g., Gophers) to provide information about and access to databases; and the integration of data from various sources and disciplines. Concomitantly, attention to management of environmental data within scientific organizations has increased. The intensified focus on effective management of environmental data is based on the premise that data, like people, machines and capital, are a significant organizational resource and product (Michener, 1986).

Coupled with new research directions in the environmental sciences and with improved technologies, data management will evolve far beyond its traditional scope (Figure 1.1). Data management within scientific organizations historically has been driven by the need to support specific tasks (e.g., data entry, archiving, security and quality assurance), many of which may be viewed simplistically as scientific custodial services. The importance of fully integrating data management into the research process – beginning with the design of data format, collection and documentation; following through in data collection, quality assurance, analysis and interpretation; and concluding with synthesis, review and publication – has been emphasized in previous workshops (Stafford *et al.*, 1986; Gorentz, 1992). We see data management within scientific organizations necessarily expanding and evolving to emphasize the timely and effective transformation of data into information and provision of that information to scientists, managers, policy makers and the public. We anticipate that this evolution will lead to the emergence of 'scientific information management' as a discipline. The focus on 'information' will ensure that both management and research are fundamental components of this activity.

In the following discussion, we examine how both the environmental sciences and computational technologies are changing and the implication of these changes for scientific information management. Finally, we describe challenges and research needs in scientific information management.

New Directions in Environmental Research

Increased attention to spatial variability and scale (Levin, 1992) and the emergence of co-directed research (integrated basic and applied research to address societal issues) in environmental science (Sharitz *et al.*, 1992; Swanson and Franklin, 1992) have several implications for the management of scientific information.

Spatial Variability and Issues of Scale

Characteristic scales define the space and time over which processes can be detected, as well as the characteristic dimensions of geographic phenomena. A phenomenon may therefore appear homogeneous at one spatial scale, but heterogeneous at another (Nellis and Briggs, 1989). Many natural systems show hierarchical organization, with nested patterns and processes occurring over a wide range of characteristic spatial and temporal scales (Allen and Starr, 1982; O'Neill *et al.*, 1988; Bradshaw, 1991). Concepts embedded in hierarchy theory highlight the importance of identifying appropriate scales for investigation, and recognizing that new patterns and processes may emerge as the scale of observation is varied (Allen and Starr, 1982). For example, climatic fluctuations over relatively long periods (decades) and large areas (106 m^2) are associated with changes in phenology and population dynamics that produce broad-scale patterns in plant communities. At smaller scales, wind and lightning associated with localized thunderstorms may be responsible for many of the patterns observed in forest stands and ecosystems (Gosz *et al.*, unpublished data).

To address phenomena that depend on spatial pattern, we must find ways to better quantify patterns of variability in space and time and to understand the causes and consequences of pattern and how patterns vary with scale (Levin, 1992). Achieving this will require the use of remote sensing, spatial statistics and other methods.

Co-directed Research

Co-directed research may be defined as the integration of basic and applied research to better address issues with management, policy and societal implications. Foci of co-directed research include natural resource management and sustainable development, environmental monitoring, environmental risk analysis and restoration ecology.

Natural resource management has received much attention recently. In the Pacific Northwest, for example, a new paradigm for forestry is being defined that focuses on forest management primarily for ecosystem structure, diversity and function, rather than for timber and commodity production (Forest Ecosystem Management Assessment Team, 1993). The International Geosphere-Biosphere Programme (Rosswall *et al.*, 1988) and the Sustainable Biosphere Initiative of the Ecological Society of America (Lubchenco *et al.*, 1991; Gosz, this volume) are focusing the efforts of the environmental research community on developing the knowledge and technologies for supporting more sustainable development.

Better management of our natural resources requires that we develop environmental monitoring programmes that allow us to identify and track the condition of specific resources at appropriate spatial and temporal scales. For example, the United States Environmental Protection Agency is launching a multi-year, multi-hundred-million dollar programme to assess status and trends of the earth's natural resources. This programme, the Environmental Monitoring and Assessment Program (EMAP), will focus initially at a national level and eventually expand to global proportions (Environmental Protection Agency, 1991). The United States Geological Survey has begun a more modest programme, the National Water Quality and Assessment Program (NAWQA), to monitor the nation's water resources (Hirsch *et al.*, 1988). Similarly, many new strategies for management of forest resources in the Pacific Northwest may be viewed as broad-scale experiments, the outcome of which can be followed through long-term monitoring.

The success of new resource management strategies and their effects on ecosystems and industrial development cannot be assessed without precise methods of analysing environmental risk. The relatively new field of environmental risk analysis seeks to develop a robust capability for predicting ecosystem risk that can be used for management, policy development and validation of ecosystem models (Whyte and Burton, 1980).

Research in restoration ecology (Jordan *et al.*, 1987; National Research Council, 1992) focuses on applying fundamental ecological principles to the restoration of natural ecological patterns and processes in degraded habitats.

Implications of New Research Directions

These new research directions have created a need for rapid and easier data analysis; timely, broad-scale, high-resolution data; new analytical approaches; better spatial sampling resolution; and a shift in focus from data to information to knowledge. Recent events like the Forest Summit in Portland, Oregon (Forest Ecosystem Management Assessment Team, 1993) and the

Four-Corners virus outbreak in New Mexico (Parmenter *et al.*, 1993) are good examples of the need to alter our trajectory from a data-rich to an information-rich science. The Forest Summit, for example, required panels of experts to integrate and synthesize vast amounts of data into meaningful information within 60 days in order to produce scientifically credible options for policies guiding management of forest and aquatic resources in the Pacific Northwest. Rapid analysis of data from disparate sources was critical in relating the Four Corners virus outbreak to the increasing population densities of rodents.

In order to address issues of scale and spatial variability, data are being collected over a broader spectral range, at finer spatial sampling resolution, and at shorter sampling intervals (Wickland, 1991). Most data used in analysing the spatial dependence of surface and atmospheric phenomena are obtained by satellite- and aircraft-mounted, multispectral sensors. Environmental scientists are increasingly employing geographic information systems (GIS) and remotely sensed data to examine questions at multiple spatial and temporal scales. Data density, which depends jointly on spatial and temporal resolution, is a technical consideration. Continuous increases in computing and storage capacities have shifted studies toward smaller temporal and spatial scales. As a result, processes operating at relatively fine spatial and temporal scales can now be investigated with finer spectral resolution over increasingly large areas (Davis *et al.*, 1991).

According to one estimate, the amount of data in the world doubles every 20 months (Frawley *et al.*, 1992). Earth-observing satellites launched in this decade will generate one terabyte of data every day – more than all previous missions combined. The Earth Observing System (EOS) will create roughly the equivalent of the entire 17-year Landsat archive every two weeks (Marshall, 1989). Similarly, the federally funded Human Genome project will store thousands of bytes for each of the several billion genetic nucleotide bases (National Research Council, 1988). Moreover, scientists are loath to discard any data for fear that a scientist in the future may need exactly that data set to test a specific hypothesis (French *et al.*, 1990; Silberschatz *et al.*, 1991). As a result, the gap between data generation and data understanding is growing (Frawley *et al.*, 1992), and our ability to gather and store data is much more advanced than our ability to manage, analyse and interpret them.

The new analyses needed for examining spatially correlated data require abandoning many of the traditional experimental designs developed primarily for agricultural experimentation (e.g., focusing on Type I and Type II errors). New experimental designs, broad-scale, spatially explicit sampling programs, geostatistical tools, and other methods that will allow us to better identify patterns and link them with processes must be developed. The extensive databases generated by EOS and other programs highlight the need for better tools that will support time-series analyses of spatial data.

Evolving Technologies Applicable to New Directions in Environmental Research

Several technologies that have been developed and improved during the past decade are now being applied increasingly to environmental research. Major advances include greatly increased desktop computing power, new analytical approaches and software, new database designs and the proliferation of remote sensing and GIS.

Advances in Computer Technology

The 1980s saw a dramatic convergence in capabilities of mini-, micro- and mainframe computers. The introduction in 1981 of the IBM personal computer (PC), based on the Intel 8086 computer chip, propelled microcomputers into a high profile. IBM extended PC capabilities initially with the Intel 80286, a 16-bit processor, and then the Intel 80386, a 32-bit processor, providing faster data access and combining multiple instruction executions in clock time (Faust *et al.*, 1991). Minicomputers and workstations have achieved mainframe speeds, and distributed processing strategies via networked PCs have fundamentally altered computing environments. In the short time between the development of the VAX 11/780 minicomputer in the late 1970s and the present, we have gone from a realizable speed of 1 million instructions per second (MIPS) to about 50 MIPS on current-generation single-CPU workstations.

The rate of development of new computer hardware tends to be underestimated, whereas that of software improvements usually is overestimated. Nevertheless, according to recent projections for hardware development (Frank *et al.*, 1991), processing speed will approximately double each year; prices for main memory will decrease about 50 per cent biannually; hard-disk access times will increase very slowly; the capacity of hard disks will increase as prices decrease; and communication networks for the exchange of large volumes of data will proliferate.

In the near future, we anticipate faster communication networks with speeds of 100 megabits/second (Mb s^{-1}); faster CPUs, with 500–1000 MIPS; more efficient devices that can store terabytes; higher resolution monitors; and higher level programming languages, both visual and object-oriented. Networks will allow complete transparency among machines world-wide, and desktop access to supercomputers will become common. By the late 1990s, the specifications for a personal GIS workstation may be a CPU with 500–1000 MIPS; 500 megabytes of main memory; 5 gigabytes of storage space on hard disks and an additional 50 gigabytes on optical disk; a workstation screen 2000 × 2000 pixels; and a communication device with a transfer rate of 100 Mb s^{-1}.

Development of software for these hardware platforms is expensive and time-consuming and lags significantly behind hardware improvements. Popular 'new' programming languages such as C and Pascal are almost 20 years old. Current operating systems (VMS, MS-DOS and UNIX), also were developed 10–20 years ago to circumvent obstacles made obsolete by rapidly changing hardware technologies. UNIX, for example, with its cryptic short commands, was invented to reduce typing when input was commonly by teletype (Winkler, 1986).

Although software development lags far behind, extensible software systems promise relief. Visual programming environments like Khoros (Rasure *et al.*, 1990), and database management systems (DBMS) and analytical systems like S+ (Becker *et al.*, 1988) allow application building and customization for specific projects.

Analytical Approaches

New emphasis on spatial scale and structure has increased characterization and modelling by spatial statistics (Robertson, 1987). Spatial variation can be modelled in a variety of ways by covariance, variograms or fractal analysis. Meteorologists and oceanographers are able to use the covariance or the power-spectrum approach (the Fourier transform of the covariance), since they can overcome some of the inherent limitations of these techniques by assuming stationarity of their data (Davis *et al.*, 1991). Semi-variograms have been used to characterize spatial variability in topography (Dubayah *et al.*, 1989), solar radiation (Dubayah *et al.*, 1990), surface albedo (Webster *et al.*, 1989), spectral vegetation indices (Davis *et al.*, 1989), soil properties (Robertson *et al.*, 1988) and vegetation (Pastor and Broschart, 1990). Precipitation (Lovejoy and Schertzer, 1985), soil properties (Burrough, 1983a, 1983b) and land use (Milne, 1991) have recently been analysed with fractals. Explicit modelling of spatial statistical structure should improve the interpolation of point data, optimize sampling designs for field collections and contribute to our understanding of underlying processes (Davis *et al.*, 1991).

New research directions have also created the need for better resolution in spatial sampling in order to understand phenomena at relevant spatial scales. For example, in order to predict spotted owl habitat, Ripple *et al.* (1991) decomposed landscape pattern by scale (coarse to fine), using the wavelet transform. Phenomena were matched to their appropriate scale of investigation: nest tree to the tree level; thermal cover to the stand level; prey base and old-growth patch connectivity to the landscape level; and conifer cover to the subregion level. This technique has numerous potential applications in landscape ecology, where analysis may focus on several scales of pattern simultaneously.

Remote Sensing and GIS

Remote sensing and GIS technology, which support the computerized capture, management, display and analysis of spatial data (Thapa and Bossler, 1992),

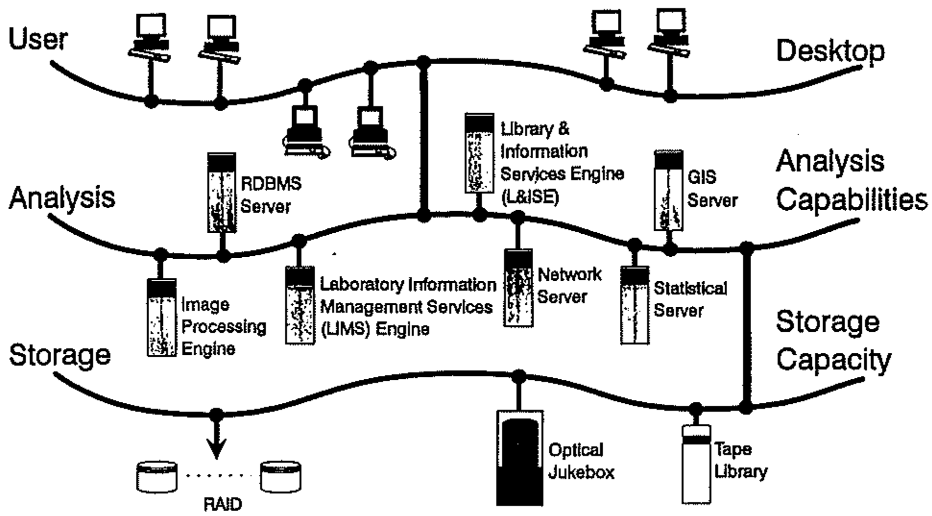

Figure 1.2 Analytical and storage capabilities accessible from the future user's desktop. Note: RAID: Redundant Array of Inexpensive Disks.

have opened new frontiers in our ability to monitor and manage natural resources – from mapping soil mycorrhizal mats (Griffiths *et al.*, 1993) to analysing regional phenomena (Burke *et al.*, 1991). GIS technology has migrated to smaller machines with ever-increasing processing power and ever-decreasing costs (Juhl, 1989). Unfortunately, although many view GIS as a panacea solving data volume and management problems, more important problems still underlie GIS, including the lack of adequate DBMS software to manage scientific data effectively and the lack of sophisticated statistical software and other analytical tools within the GIS environment.

Challenges for Future Environmental Information Management and Analysis Systems

Distributed Analytical Environments

A primary challenge in this area is creation of a highly interactive metacomputer composed of workstations with graphics accelerators, high-speed networks, terabyte data vaults and supercomputers that can address multidisciplinary 'grand challenge' problems. In the future, scientific organizations will likely use multiple file and computer servers with ease from their desktop computers. File servers may be dedicated to one or more specific tasks (e.g., statistics, a DBMS, library support, or image processing; Figure 1.2). Hierarchical storage servers will provide transparent access to huge amounts of optical disk and tape storage previously maintained off-line.

Organizations may have heterogeneous DBMSs differing in their capabilities and structure and dispersed over several sites. The heterogeneity, distribution and sheer number of databases in the public and private domains make it imperative to share access to these databases while maintaining autonomy of individual systems. Thus, new techniques are needed that support the interoperability of autonomous databases without requiring their global integration (Scheuermann *et al.*, 1990).

In the system envisioned for the future, a user submits a database query, completely unaware of and unconcerned with the location and format of the data. The software automatically locates the data across any number of DBMS, processes them and returns the answer. Security, error correction and rollback are independent of user, DBMS, vendor and location. Multimedia, hypertext and hypermedia technologies will be increasingly used for creating, organizing and disseminating on-line information (Mohageg, 1992) and for combining text and graphics with working models and imagery.

The complex nature of spatial, as compared with record-oriented, data complicates data distribution and communication for distributed GIS. Network technology ultimately will develop to the point where intermachine communication is no longer an issue. The challenge will be in providing software to utilize this type of network efficiently. Thus, scientists in Albuquerque, New Mexico, will be able to interactively access soils data stored on the United States Geological Survey server in Denver, Colorado, and climate data stored on the National Climate Data Center server in Asheville, North Carolina.

Database Management Systems

The promise of a DBMS lies in its ability to provide timely and consistent data, to enable users to access those data directly without technical assistance and to evolve easily to meet changing user requirements (Curtice, 1986). Environmental research at the project level requires integrated computational services that provide access to sophisticated database, statistical and numeric operations and to technical text support (Conley *et al.*, 1986). These operations could be supported collectively by a scientific DBMS; however, most commercial DBMS software is record-oriented and directed specifically at business solutions.

If the next-generation database applications are to meet the needs of environmental scientists, they are unlikely to resemble current business-oriented DBMSs (Silberschatz *et al.*, 1991). Recently, considerable research has focused on developing general-purpose DBMS prototypes that present some advantages over relational DBMS (RDBMS) in representing complex (i.e., spatial) data, which do not lend themselves easily to representation or efficient management within a record-oriented system. One such prototype is POSTGRES (for post Ingres) (Pollock and McLaughlin, 1991). Additional research in this area focuses on extending current RDBMS technology and developing object-

Table 1.1 Advanced database management system (DBMS) strategies with potential for scientific information management.

Extended relational:	incorporates new features in the relational model, such as inheritance, user-defined functions and data types, and procedures for data types and nested relations.
Object-oriented:	accesses a complex object as a single concept and performs operations on that object in the DBMS, rather than in an external application program (Dittrich, 1986); supports inclusion of 'unconventional' data types, such as text, voice, graphics or CAD (computer-assisted drafting) (Scheuermann *et al.*, 1990).
Extensible:	utilizes a 'toolkit' approach to tailor DBMS internals (e.g., storage structures, query optimizers) to meet application requirements that are inadequately served by existing DBMSs; supports easier customization of DBMSs for scientific applications.
Logic-based languages:	support access to relational RDBMS by providing a consistent interface between the tools and the RDBMS (Hessinger, 1987); e.g., Structured Query Language (SQL).

oriented and extensible DBMSs and logic-based languages (Table 1.1) (Shoshani, 1990). The overall challenge will be integrating the best of each of these techniques to manage scientific data.

Integrated GIS and Remote Sensing in the Time Domain

Overcoming the technological obstacles associated with jointly processing image and GIS data, such as converting between vectors and rasters and jointly displaying and updating digital imagery and maps, has already entailed much work. Continued efforts in this area should lead to the evolution of integrated geographic information systems (IGIS) (Davis *et al.*, 1991). Since not all physical processes can be mapped directly from remotely sensed data or existing maps, an important role of IGIS will be the generation of such information by appropriate modelling based on more than one information source (Risser, 1986). The coupling of satellite measurements with other spatial data has tremendous potential for describing earth surfaces, predicting future conditions and validating physical representations produced solely through remote sensing or GIS. One key impediment to total integration of GIS and remote sensing is that current GIS technology does not adequately represent the time domain, whereas remote sensing samples in both space and time. Causal relationships among events or entities are embedded in temporal information, and data therefore should be interpreted in a temporal context (Snodgrass and Ahn, 1986).

Errors associated with remotely sensed and GIS data can significantly impact the confidence with which decisions are made (Lunetta *et al.*, 1991). GIS provides new information without simultaneously establishing its reliability. Uncertainty is propagated and transformed each time a conceptual or physical model is constructed in GIS processing (Bedard, 1987). Lack of concern for error in spatial databases reflects the lack of a standard framework for analysing the propagation of modelling error (Lanter and Veregin, 1992).

As these technologies continue to improve, the potential of IGIS analysis will come to be limited only by our conceptual understanding of the phenomena under investigation and their representation in spatial databases (Davis *et al.*, 1991).

User Interfaces

The improvement of interfaces within and among image processing, GIS, DBMS, expert systems, models and statistical software will facilitate analyses significantly. To meet the needs of environmental scientists, future systems will emphasize sharing data and tools among disparate software environments, rather than computing power. More powerful and more efficient languages and user interfaces are needed that can lead less computer-literate users systematically through common data management and information processing without requiring them to master the intricacies of the DBMS and the computer operating system. Ideally, the further development of higher level languages will allow the investigator to focus on the problem at hand (Peled, 1987).

There is also a strong need for better management of critical information associated with models and model input and output - that is, model metadata. This includes keeping good records of model evolution and of the lineage of data as they are used to produce publications or as input to other models or analyses that may cascade to further models. The bottom line is clearly linking the data with the models. Intuitively designed user interfaces can help expedite this.

Visualization Software and Knowledge Discovery

Visualization tools have been developed and primarily utilized by physical scientists. One of the challenges for environmental researchers will be the use of such tools for extracting information from the massive volumes of data being collected. Visualization software will be a required part of the environmental scientist's toolkit.

Knowledge discovery, the use of computers to extract implicit and potentially useful information from data, carries potential for tackling some of the enormous data analysis challenges of the next generation of environmental scientists. Knowledge discovery is expensive even with today's computers and requires additional knowledge regarding the form and content of data.

Visualization software, knowledge discovery algorithms and expert systems will necessarily be used increasingly to facilitate the processing of vast amounts of data into the information needed by scientists, managers and policy makers.

Improved Spatial Sampling Resolution and Standardization

In most environmental research, data are collected in plots of 1 m^2 or less (Kareiva and Anderson, 1988; Brown and Roughgarden, 1990). Few studies entail collection of pattern- and process-oriented data on the ground that can be readily related to the scale of data sampled by Landsat Thematic Mapper (30 m × 30 m) or SPOT (10 m × 10 m, panchromatic; 20 m × 20 m, multispectral). Thus, there is currently a wide discrepancy between the spatial scales of data collected by remote sensing and those collected on the ground that can be used for verification and validation. This gap likely will decrease as scientists take advantage of new airborne remote sensing systems that sample data at the submetre scale.

Nevertheless, several impediments to scaling up from the plot to broader scales likely will remain. For example, several networks (including the LTER Network, Global Change Sites funded by the National Park Service, and National Oceanic Atmospheric Administration Marine Sanctuaries) have been established to examine ecological patterns and processes at the scale of the continental USA. Certain attributes of primary production are examined at most or all of the sites; however, specific parameters being measured, temporal and spatial resolution, field and laboratory protocols, and data management and analysis generally differ significantly among sites. In addition, numerous biomes within any given network are generally not sampled, and spatial replication within biomes is rare or non-existent. Expanding individual networks to encompass more of the inter- and intra-biome variability, as well as integrating individual networks into a 'network of networks', could facilitate broad-scale research by ensuring better spatial coverage and encouraging data comparability and information exchange (Bledsoe and Barber, 1993). We need to focus attention for intersite comparison at the information stage, however, in part because the big differences among biomes and ecosystems may make standardized data collection much less useful than standardized approaches to information assembly and management. Spatial and temporal variability, standardization of monitoring programmes, laboratory and analytical protocols, and data and information management practices likely will continue to challenge us.

Conclusion

The virtual explosion in computing power and accessibility has significantly revolutionized the scientific environment. Managing scientific information has gone far beyond a mere 'custodial service'. Users of scientific databases work in a great variety of environments, from a single personal computer or workstation to supercomputers to heterogeneous networked systems. Network technology ultimately will develop to the point where intermachine communication is no longer an issue, and the potential for analysis will be limited only by our conceptual understanding and our ability to frame the questions.

We face challenges in monitoring and modelling global change. Of these, two interrelated problems are key: (1) combining remote sensing and GIS, and (2) determining the nesting of observations required to bridge scales between short-term, fine-scale and long-term, broad-scale measurements and process models (Rosswall *et al.*, 1988). Tight linkages between research and application will be critical to monitoring the environmental and social consequences of global change (Davis *et al.*, 1991).

Because large and diverse data sets, documents and images are becoming common in both business and science, the gap between business and scientific software is closing. DBMS software that will be useful to both business and scientific applications should result. A properly instituted and supported DBMS can increase the productivity of environmental researchers and enhance the quality of environmental science (Risser and Treworgy, 1986), reducing the time scientists devote to learning details of hardware and software. This level of DBMS for a large project must be supported with a great deal of forethought and ongoing commitment.

Technologically advanced societies frequently cannot transform new ideas into successful products quickly (Strome and Lauer, 1977). Acceptance of technology in any field is complex, and decades often elapse before practical applications of research are achieved. Barriers to communication among engineers, environmental scientists, computer scientists, policy makers and others further delay product development.

A sociology of information management is developing that is as critical as the hardware and software issues. Because bridges must be built between researchers and data managers, training a new breed of scientists and scientific information managers will be a prerequisite to the conscientious efforts required to close these gaps. We anticipate the emergence of 'scientific information management' as a discipline with both a management and a research component, emphasizing the timely and effective transformation of data into information and knowledge for scientists, managers, policy makers and the public.

Acknowledgement

We acknowledge National Science Foundation support from DEB No. 90-11663 to the H. J. Andrews LTER project and BSR No. 88-110946 support to the Sevilleta LTER. This chapter benefited from discussions with Fred Swanson. This is paper 2974, Forest Research Laboratory, Oregon State University and contribution 48 to Sevilleta LTER, University of New Mexico.

References

Allen, T. F. H. and Starr, T. B., 1982, *Hierarchy, Perspectives for Ecological Complexity*, Chicago: University of Chicago Press.

Becker, R. A., Chambers, J. M. and Wilks, A. R., 1988, *The New S Language: A Programming Environment for Data Analysis and Graphics*, Pacific Grove, California: Wadsworth and Brooks/Cole Advanced Books and Software.

Bedard, Y., 1987, Uncertainties in land information systems databases, *Proceedings, Auto-Carto*, **8**, 175–84.

Bledsoe, C. and Barber, M., 1993, *Ecological Network of Networks*, Washington, DC: US MAB Secretariat, US Department of State.

Bradshaw, G. A., 1991, 'Hierarchical analysis of pattern and processes of Douglas-fir forests', unpublished PhD dissertation, Corvallis, Oregon: Department of Forest Science, Oregon State University.

Brown, J. H. and Roughgarden, J., 1990, Ecology for a changing earth, *Bulletin of the Ecological Society of America*, **71**, 173–88.

Burke, I. C., Kittel, T. G. F., Lauenroth, W. K., Snook, P., Yonker, C. M. and Parton, W. J., 1991, Regional analysis of the central Great Plains, *BioScience*, **14** (10), 685–92.

Burrough, P. A., 1983a, Multiscale sources of spatial variation in soil, I: The application of fractal concepts to nested levels of soil variation, *Journal of Soil Science*, **34**, 577–97.

Burrough, P. A., 1983b, Multiscale sources of spatial variation in soil, II: A non-Brownian fractal model and its application in soil survey, *Journal of Soil Science*, **34**, 599–620.

Conley, W., Slator, B. M., Anderson, M. P. and Sitze, R. A., 1986, Designing and prototyping a scientific problem solving environment: the NMSU science workbench, in Michener, W. K. (Ed.) *Research Data Management in the Ecological Sciences*, Belle W. Baruch Library in Marine Science, No. 16, pp. 383–409, Columbia, South Carolina: University of South Carolina Press.

Curtice, R. M., 1986, Getting the database right, *Datamation*, 1 October, 99–104.

Davis, F., Dubayah, R., Dozier, J. and Hall, F., 1989, Covariance of greenness and terrain variables over the Konza Prairie, *Proceedings IGARSS '89*, 1322–5.

Davis, F. W., Quattrochi, D. A., Ridd, M. K., Lam, N. S.-N., Walsh, S. J., Michaelson, J. C., Franklin, J., Stow, D. A., Johannsen, C. J. and Johnson, C. A., 1991, Environmental analysis using integrated GIS and remotely sensed data: Some research needs and priorities, *Photogrammetric Engineering and Remote Sensing*, **57** (6), 689–97.

Dittrich, K., 1986, Object-oriented database systems: The notion and the issues, in *Proceedings of the International Workshop on Object-oriented Database Systems*, pp. 2–4, Washington, DC: IEEE. Computer Society Press.

Dubayah, R., Dozier, J. and Davis, F. W., 1989, The distribution of clear-sky radiation over varying terrain, *Proceedings IGARSS '89*, 885–8.

Dubayah, R., Dozier, J. and Davis, F. W., 1990, Topographic distribution of clear-sky radiation over the Konza Prairie, Kansas, *Water Resources Research*, **267**, 679–90.

Environmental Protection Agency, 1991, *An Overview of the Environmental Monitoring and Assessment Program, EMAP Monitor, January 1991*, EPA-6001 M-90/022, Washington, DC: US EPA, Office of Research and Development.

Faust, N. L., Anderson, W. H. and Star, J. L., 1991, Geographic information systems and remote sensing future computing environment, *Photogrammetric Engineering and Remote Sensing*, **57** (6), 655–68.

Forest Ecosystem Management Assessment Team, 1993, *Forest Ecosystem Management: An Ecological, Economic, and Social Assessment, Report of the Forest*

Ecosystem Management Assessment Team, 1993-793-071, Washington, DC: US Government Printing Office.

Frank, A. U., Egenhofer, M. J. and Kuhn, W., 1991, A perspective on GIS technology in the nineties, *Photogrammetric Engineering and Remote Sensing*, **57** (11), 1431 6.

Franklin, J. F., Bledsoe, C. S. and Callahan, J. T., 1990, Contributions of the long-term ecological research program, *BioScience*, **40** (7), 509 23.

Frawley, W. J., Piatetsky-Shapiro, G. and Mathews, C. J., 1992, Knowledge discovery in databases: an overview, *AI Magazine*, fall, 57–70.

French, J. C., Jones, A. K. and Pfaltz, J. L., 1990, Summary of the final report of the NSF Workshop on Scientific Database Management, *Sigmod Record*, **19** (4), 32 40.

Gorentz, J. B. (Ed.), 1992, *Data Management at Biological Field Stations and Coastal Marine Laboratories*, Report of an invitational workshop, W. K. Kellogg Biological Station, Michigan: Michigan State University April 1990.

Griffiths, R. P., Bradshaw, G. A. and Caldwell, B. A., 1993, 'Distribution of ectomycorrhizal mats in coniferous forests in the Pacific Northwest', presentation at an international symposium: The Functional Significance and Regulation of Soil Biodiversity, East Lansing, Michigan, May.

Harmon, M. E., 1992, *Long-term Experiments on Log Decomposition at the H. J. Andrews Experimental Forest*, USDA Forest Service General Technical Report PNW-GTR-280.

Hessinger, P. R., 1987, DBMS: Adding value to vanilla, *Datamation*, 1 March, 50 4.

Hirsch, R. M., Alley, W. M. and Wilbur, W. G., 1988, *Concepts for a National Water-Quality Assessment Program*, US Geological Survey Circular 1021.

Jordan, W. R., III, Gilpin, M. E. and Aber, J. D. (Eds), 1987, *Restoration Ecology*, Cambridge, UK: Cambridge University Press.

Juhl, G., 1989, GIS technology coming of age, *American City and County*, April, 50 4.

Kareiva, P. and Anderson, M., 1988, Spatial aspects of species interactions: the wedding of models and experiments, in Hastings, A. (Ed.) *Community Ecology*, New York: Springer Verlag.

Lanter, D. P. and Veregin, H., 1992, A research paradigm for propagating error in layer-based GIS, *Photogrammetric Engineering and Remote Sensing*, **58** (6), 825 33.

Levin, S. A., 1992, The problem of pattern and scale in ecology, *Ecology*, **73** (6), 1943 67.

Lovejoy, S. and Schertzer, D., 1985, Generalized scale invariance in the atmosphere and fractal models of rain, *Water Resources Research*, **21**, 1233–50.

Lubchenco, J., Olson, A. M., Brubaker, L. B., Carpenter, S. R., Holland, M. M., Hubbell, S. P., Levin, S. A., MacMahon, J. A., Matson, P. A., Melillo, J. M., Mooney, H. A., Peterson, C. H., Pulliam, H. R., Real, L. A., Regal, P. J. and Risser, P. G., 1991, The sustainable biosphere initiative: an ecological research agenda, *Ecology*, **72** (2), 371–412.

Lunetta, R. S., Congalton, R. G., Fenstermaker, L. K., Jensen, J. R., McGuire, K. C. and Tinney, L. R., 1991, Remote sensing and geographic information system data integration: error sources and research issues, *Photogrammetric Engineering and Remote Sensing*, **57** (6), 677–87.

Marshall, E., 1989, Bringing NASA down to earth, *Science*, **244** (16 June), 1248 51.

Michener, W. K., 1986, Data management and long-term ecological research, in Michener, W. K. (Ed.) *Research Data Management in the Ecological Sciences*, Belle W. Baruch Library in Marine Science, No. 16, pp. 1 8, Columbia, South Carolina: University of South Carolina Press.

Milne, B. T., 1991, Lessons from applying fractal models to landscape patterns, in Turner, M. G. and Gardner, R. H. (Eds) *Quantitative Methods in Landscape Ecology*, pp. 199–235, New York: Springer Verlag.

Mohageg, M. F., 1992, The influence of hypertext linking structures on the efficiency of information retrieval, *Human Factors*, **34** (3), 351–67.

National Research Council, 1988, *Mapping and Sequencing the Human Genome*, Report of the Committee on Mapping and Sequencing the Human Genome, Washington, DC: National Research Council.

National Research Council, 1992, *Restoration of Aquatic Ecosystems: Science, Technology and Public Policy*, Washington, DC: National Academy Press.

Nellis, M. D. and Briggs, J. M., 1989, The effect of spatial scale on Konza landscape classification using textural analysis, *Landscape Ecology*, **2**, 93–100.

O'Neill, R. V., Krummel, J. R., Gardner, R. H., Sugihara, G., Jackson, B., DeAngelis, D. L., Milne, B. T., Turner, M. G., Zygmunt, B., Christensen, S. W., Dale, V. H. and Graham, R. L., 1988, Indices of landscape pattern, *Landscape Ecology*, **1**, 153–62.

Parmenter, R. R., Brunt, J. W., Moore, D. I. and Ernest, S., 1993, *The Hantavirus Epidemic in the Southwest: Rodent Population Dynamics and the Implications for Transmission of Hantavirus-associated Adult Respiratory Distress Syndrome (HARDS) in the Four Corners Region*, a special report to the Federal Centers for Disease Control and Prevention and the New Mexico Department of Health, Publication No. 41, Albuquerque, New Mexico: Sevilleta Long-Term Ecological Research Program.

Pastor, J. and Broschart, M., 1990, The spatial pattern of a northern conifer–hardwood landscape, *Landscape Ecology*, **4**, 55–68.

Peled, A., 1987, The next computer revolution, *Scientific American*, **257** (4), 57–64.

Pollock, R. J. and McLaughlin, J. D., 1991, Data-based management system technology and geographic information systems, *Journal of Surveying Engineering*, **117** (1), 9–26.

Rasure, J., Williams, C., Argiro, D. and Sauer, T., 1990, A visual language and software development environment for image processing, *International Journal of Imaging Systems and Technology*, **2**, 183–99.

Ripple, W. J., Bradshaw, G. A. and Spies, T. A., 1991, Measuring forest landscape patterns in the Cascade Range of Oregon, USA, *Biological Conservation*, **57**, 73–88.

Risser, P. G., 1986, *Spatial and Temporal Variability of Biosphere and Geosphere Processes: Research Needed to Determine Interactions with Global Environmental Change*, Paris: ICSU Press.

Risser, P. G. and Treworgy, C. G., 1986, Overview of research data management, in Michener, W. K. (Ed.) *Research Data Management in the Ecological Sciences*, Belle W. Baruch Library in Marine Science, No. 16, pp. 9–22, Columbia, South Carolina: University of South Carolina Press.

Robertson, G. P., 1987, Geostatistics in ecology: Interpolating with known variance, *Ecology*, **68**, 744–8.

Robertson, G. P., Huston, M. A., Evans, F. C. and Tiedge, J. M., 1988, Spatial variability in a successional plant community: Patterns of nitrogen availability, *Ecology*, **69**, 1517–24.

Rosswall, T., Woodmansee, R. G. and Risser, P. G. (Eds), 1988, *Scales and Global Change: Spatial and Temporal Variability in Biospheric and Geospheric Processes, Scientific Committee on Problems of the Environment* (SCOPE) Report 35, New York: Wiley.

Scheuermann, P., Yu, C., Elmagarmid, A., Manola, F., Rosenthal, A., Garcia-Molina, H., McLeod, D. and Templeton, M., 1990, Report on the workshop on heterogenous database systems held at Northwestern University, Evanston, Illinois, December, 1989, *Sigmod Record*, **19** (4), 23–31.

Sharitz, R. R., Boring, L. R., Van Lear, D. H. and Pinder, J. E., III, 1992, Integrating ecological concepts with natural resource management of southern forests, *Ecological Applications*, **2** (3), 226–37.

Shoshani, A., 1990, Emerging and new technologies, in French, J. C., Jones, A. K. and Pfaltz, J. L. (Eds) *Scientific Database Management, Report of the Invitational NSF*

Workshop on Scientific Database Management, Computer Science Report No. TR-90-22, pp. 13–21, Charlottesville, Virginia: University of Virginia.

Silberschatz, A., Stonebraker, M. and Ullman, J. (Eds), 1991, Database systems: Achievements and opportunities, *Communications of the ACM*, **34** (10), 110–20.

Snodgrass, R. and Ahn, I., 1986, Temporal databases, *Computer*, September, 35-42.

Stafford, S. G., Alaback, P. B., Waddell, K. L. and Slagle, R. L., 1986, Data management procedures in ecological research, in Michener, W. K. (Ed.) *Research Data Management in the Ecological Sciences*, Belle W. Baruch Library in Marine Science, No. 16, pp. 93–114, Columbia, South Carolina: University of South Carolina Press.

Strome, W. M. and Lauer, D. T., 1977, An overview of remote sensing technology transfer in the US and Canada, *Proceedings, 11th International Symposium on Remote Sensing of Environment*, Vol. 1, pp. 325–31, Ann Arbor, Michigan: Environmental Research Institute of Michigan.

Swanson, F. J. and Franklin, J. F., 1992, New forestry principles from ecosystem analysis of Pacific Northwest forests, *Ecological Applications*, **2** (3), 262–74.

Thapa, K. and Bossler, J., 1992, Accuracy of spatial data used in geographic information systems, *Photogrammetric Engineering and Remote Sensing*, **58** (6), 835-41.

Webster, R., Curran, P. J. and Munden, J. W., 1989, Spatial correlation in reflected radiation from the ground and its implications for sampling and mapping by ground-based radiometry, *Remote Sensing of Environment*, **29**, 67–78.

Whyte, A. V. T. and Burton, I. (Eds), 1980, *Environmental Risk Assessment*, prepared by the Scientific Committee on Problems of the Environment (SCOPE) of the International Council of Scientific Unions (ICSU), New York: Wiley.

Wickland, D. E., 1991, Mission to planet earth: The ecological perspective, *Ecology*, **72** (6), 1923–45.

Winkler, C., 1986, Getting from here to there: No way, *Datamation*, 1 June, 202–5.