AN APPROACH TO ADVANCED ECOSYSTEM MODELING USING VECTOR AND RASTER DATA

Stanley A. Morain
Department of Geography and Earth Data Analysis Center
and
Amelia Budge
Earth Data Analysis Center
University of New Mexico
Albuquerque, New Mexico, USA 87131-6031

Commission VII, Working Group 5

KEY WORDS: Key Words: Ecosystems, Relational Databases, Object Databases, Modeling, Self-adaptation, satellite measurements, genetic algorithms, fuzzy rules, virtual reality, visualization

ABSTRACT

Complexity science asserts that ecosystems organize into delicate relationships that teeter on the edge of *Order* and *Chaos*. This edge is maintained and governed by imprecise “rules” of nature (i.e., it is fuzzy), but it can be modeled. The challenge in the next few decades will be to learn how to apply spectral, spatial, and temporal technologies so as to mimic the self-adaptive behavior of natural systems. An even greater challenge will be learning how to integrate gene sequencing data, genetic network functions, cellular automata, and advance computational strategies into cohesive data structures for model development. Advanced ecosystem modeling should move in directions that develop genetic algorithms that: (a) emulate self-adaptive behavior through rule-based, stimulus-response mechanisms; (b) are capable of "learning" through artificial neural networks; and, (c) permit evolutionary changes to be animated and visualized in a virtual reality framework. This paper conceptualizes a genetic algorithm that sandwiches a relational database of ecosystem and environmental parameters between two object databases that integrate plant genetics and atmospheric conditions with ecosystem responses.

INTRODUCTION

One of the biggest challenges in biogeography is to link temporal and spatial scales that range from cellular and molecular on the one hand, to whole landscapes on the other. Nature progresses in complexity from bottom to top, but has evolved from stimuli operating from top to bottom. The first level of complexity begins with DNA chemistry and spreads through many functions that characterize individuals and species. The second level involves interactions that lead to communities that then interact with the environment and feedback to level one processes. Finally, the third level of complexity involves atmospheric, geophysical, and hydrospheric processes that trigger a species' adaptive responses to these external stimuli.

The need to develop self-adaptive ecosystem models will increase as human impacts on natural environments intensify. Human activity sets in motion accelerated feedback loops that govern ecosystem reactions and adaptations that may never be fully understood. *Homo sapiens* cannot interpret or predict the importance of their own economic or political imprints on landscapes because they cannot separate their perceptions of landscapes from the biogeochemical forces that drive those systems. Humans are aware of their place in nature, but do not understand how systems evolve in response to their presence. Instead of resource management schemes whose outcomes are biased by human expectations (as for example, the strategy that aims toward *desired future conditions*), modelers should provide managers with tools that simulate evolution in natural systems.

This paper conceptualizes a genetic algorithm for monitoring ecosystem change (specifically changes that can be credited to human impacts and climate warming), and that, in turn, affect the speed and directions of plant species migration and distribution. The concept employs satellite measurements for three purposes: (a) as a source of "continuous" data that may contain fractal dimensions and n^{th} order periodicities like the type recently described by Stølum (1996); (b) as a means for change detection at the Earth's surface; and © as part of the input for articulating fuzzy rules directing self-adaptive behavior. Other *in-situ* measurements are needed to develop eco-physiological rules; and, both types of rules need to be combined into object and relational databases that can interact and adapt on their own to stimuli.

ENABLING TECHNOLOGIES

The construct adopted here is referred to as a genetic algorithm to emphasize that it integrates genetic information with an approach for monitoring observable landscape changes. It provides a framework for integrating retrospective data ...and future data...into an archive for more realistic resource management. Object and relational databases would be assembled from distributed, electronic sources, and integrated into functional, interactive data sets to which dynamic rules are applied. To work properly, it will require artificial neural networks that can process data elements whose instantaneous attributes could assume any of several conditions, at the same time. Clearly, this concept pushes the frontiers of database technology beyond today's capabilities.

Status of Genome Research

If a genetic algorithm template can be created, static data from past research could be integrated with future data. All that has been learned about an ecosystem would thus find service as active digital information. The intent is to develop a digital library beginning with each species' genetic attributes and functions, and ending with its community interactions and environmental responses. The integration of species libraries would form the larger and more complex communities and ecosystems in which they reside, and would allow dynamic data to be inserted and updated, as needed. At present, the template is crudely articulated. There are, for example, quantum leaps in knowledge required to translate genetic base pairs into their control over physiologic responses...but this is where the most rapid technology is evolving.

The human genome project was formulated ten years ago. It was considered an almost impossible job because of its complexity. Today, this genetic map is nearing completion, and the race is on to map other species. Techniques are evolving that apply to whole groups of related species, eliminating the need to map each one separately. These maps are becoming commonplace. More importantly, the networks and functions behind the genes are gradually being unraveled.

Although research is currently driven by a desire to understand the human complex, the research community believes that other forms of life will soon be examined. By the end of this century, genome research is expected to migrate toward agricultural and pharmacological species, as well as to endangered and keystone species.

The digital information that underlies genetic functions is represented by a simple string of proteins symbolized as G/A/T/C (guanine, adenine, thymine, cytocine). This string is the fundamental data structure of any organism's biology. Because of this commonality, base pairs can be linked to the physiologic and metabolic functions they control, and in theory, at least, every species can be integrated into an ecosystem model through these object attributes. Once the linkages are mapped, as complicated as they are, the data structure permits abstraction to higher levels of system integration..

Status of Ecosystem Research

Carpenter et al. (1995) say that predicting responses of ecosystems to perturbation is among the greatest challenges in ecology. Initial research has focused on biogeochemistry and chemical stressors of ecosystems, but the scope is expanding to include community dynamics and ecosystem processes. Most people are familiar with simple laboratory ecosystems like an aquarium, and many have learned how hard it is to create a self-adaptive system using only a few species. Nonetheless, these simple systems provide data that cannot be measured easily in the field. Time compression and cost savings are the usual arguments for these models. Time series experiments that would take generations to complete in the field can be programmed in the laboratory, and their results used to develop rule-based systems. Model systems serve not only as links between levels of complexity, but also to field experiments containing environmental noise.

Over the next decade, ecologists expect more research on model systems, tending toward more complex and realistic assemblages. Experiments are moving outdoors to controlled environmental facilities. Results from experiments on CO_2, for example, represent one kind of research. There is no doubt that rising concentrations of global CO_2 will change the abundance and composition of some plant communities. But, the relationships are not linear. CO_2 response is determined by genetic functions that are influenced by nutrient levels, water supply, and temperature--all of which are themselves influenced by changing CO_2.

Field experiments using uncontrolled natural conditions are always subject to question. History shows that one or two landmark studies can alter paradigms for decades, only to fade on the strength of another study that refocuses the research agenda. Complexity is the current focus. Before data collected using earlier paradigms are lost they should be integrated into digital libraries.

Landscape Ecology and GIS

Landscape ecology is a relatively young field of research, but one that is familiar to spatial and spectral analysts. It is concerned with the spatial dynamics of organisms, materials, and energy. It promotes models and theories that focus on spatial heterogeneity and an examination of data collected at different geographic scales; and, it recognizes that landscapes are patchworks of natural and cultural elements interacting to form a new class of ecosystem based on human supremacy. Figure 1 is a typical output from this field of inquiry.

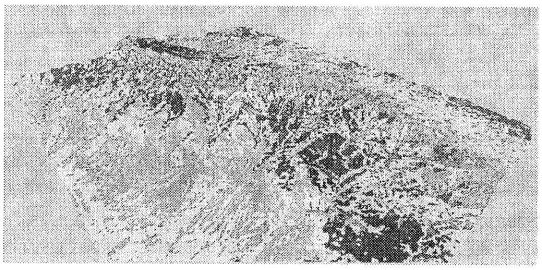

Figure 1. Vegetation Cover Over Terrain Model

GENETIC ALGORITHM TEMPLATE

A genetic algorithm that is based on biological genomes, but which is also sensitive to environmental attributes and human impacts is shown in Table 1. This approach would convert as many field and laboratory observations and measurements as are available for an ecosystem into binary sequences representing organisms interacting with their environments; and developing the rules required to emulate their stimulus-response reactions.

As conceptualized here, binary attributes are referred to as "cybergenes" that can be arranged in chromosome-like fashion into "cyberchromes." Through the tool-box of rules, system properties would be changed according to observed stimulus-response patterns beginning with genetic networks and progressing through the terrestrial environment to atmospheric properties. For example, an increase in mutation rates resulting from increased ultraviolet radiation, would be accountable in the genetic

Table 1. Template for a Genetic Algorithm

Atmospheric Data in Object Database		
Field of Inquiry	**Measurement Parameters**	**Cybergenes/Cyberchromes**
Tropospheric gases	concentrations and cycles of O_3, CO_2, No_x, So_x; trace gas chemistry,..	0/100101/01110010/011001/00010010011/ 10101/....
Aerosols	concentrations, cycles,...	000100110/00111/1111/100/00/0/...
Solar radiation	quality, intensity, spectrum, cycles,..	11/1111/00/100/0001/0000000/...
Environmental Data in Relational Database		
Field of Inquiry	**Measurement Parameters**	**Cybergenes/Cyberchromes**
Human impacts	grazing; agriculture; mining; timbering; settlement pattern; transportation; point sources; non-point sources; fertilizer and pesticide inputs; civil boundaries...	10/11100/00111000/10001100/1/0/0/00011/ 11000011111/000001/...
Near ground atmosphere	precipitation cycles and amounts; humidity patterns; temperature norms, extremes, and cycles; snow depth; degree days...	001/11011/1110/000/10111/0/11/00/000/1/ 0/00000/...
Model systems; Field experiments	species interactions under controlled, semi-controlled, and semi-natural conditions	fuzzy rules; neural nets,...
Species integration	community composition; species interactions under natural conditions	fuzzy rules; neural nets; individual species' binary tags
Topographic and edaphic controls	soil moisture regime; soil depth; lithology; slope; aspect...	11000/001010/1/1/0/00/111/001100/11100/ 100/...
Taxonomic Date in Object Database		
Field of Inquiry	**Measurement Parameters**	**Genes/Chromosomes**
Species' attributes	physiologic and metabolic functions; phenotype; range of variability; geographic range	000/1/000111/11101/000110111/001/00010/ 0110/11001111/...(includes individual species binary tag)
Genetic networks	stimulus/response; signal path dynamics...	Kinetic models & electrical circuit simulations
Genomes	base pairs; gene sequences	G/A/T/C strings

algorithm, and would be monitored by satellite UV sensors. The rules would allow the system to be scaled for local or regional applications, and might then be integrated into global models.

The approach outlined in Table 1 conceptualizes how the internal complexities of natural genomes might be integrated with observable environmental attributes to simulate dynamical systems. It represents a logically structured "bit map" for coding and storing complex, interactive knowledge about an ecosystem as strings of binary data that could then be machine-processed in a virtual reality environment to monitor adaptive reactions. As new, or better knowledge became available, the cybergenes could be changed or lengthened; or, whole new cyberchromes could be added as data from different scientific areas became available. Perhaps the most important feature of such a genetic algorithm is that the rules and neural networks governing self-adaptive behavior would reduce the level of human bias now common in ecosystem models.

Current ecosystem models are based largely on an assumption of linearity that accounts for neither environmental perturbations nor identifiable periodicities. Instead, they project evolutionary trajectories based on the straight-line portion of time functions representing "today", and which are assumed will continue far into the future. Indeed, there are many practicing professionals who are convinced that resource "management" equates to keeping everything as it is.

In short, current models are not designed to self-adapt. They do not ride the fuzzy edge of *Order* and *Chaos* to maintain high levels of structure and function without sacrificing flexibility (Kauffman, 1993). Evolutionary trajectories of species and ecosystems are not random, but neither are they linear. Evolution is normally such a slow process that adaptive reactions are only observable *after the fact*. Some critical questions therefore are..."how can the science community gather and process data fast enough to drive a genetic algorithm?"; "how can fuzzy rules and artificial neural networks be incorporated to simulate an ecosystem's internal "learning" process?"; and, "how can routinely collected satellite environmental data be used in a time-dimension to visualize dynamic processes?"

This last question is especially interesting for the remote sensing and GIS communities because modelers will have neither the variety nor the spatial continuity of *in-situ* data necessary to drive dynamical models over large areas. As an alternative to discontinuous, *in-situ* measurements, scientists can search remotely acquired sensor data for ecological and genetic information, and thus approach the exercise from satellite to ground. Pixel digital numbers (DNs) from time-series data would become "windows" through which genetic and ecologic systems are linked..

ROLE OF REMOTE SENSING AND GIS

Genome maps and genetic networks provide the foundation for understanding how species respond and adapt to environmental stimuli. Another part of natural complexity involves measuring the attributes that constitute those stimuli, and understanding how they vary in time and space. Remote sensing is perhaps the best means for obtaining many of these data because satellite sensors can collect uniform global data that are applicable at local scales and can thereby provide time series comparability. Similarly, airborne sensors collect fine resolution data at the local level but can be flown over analogous sites around the world to obtain comparable geographic data. For developing self-adaptive models, however, the greatest asset of satellite measurements is that they represent repetitive, synoptic, calibrated data sets that satisfy the criterion for "continuity" in rule-based models.

Remotely acquired data come in both raster (image) and vector (sounder) formats depending upon sensor design and measurement technique. Both kinds are directly employable in Geographic Information Systems (GIS), which have proven their value as a means for managing and displaying relationships between data sets. Traditional data collected in the field will not be replaced by remotely acquired data because these, too, are directly employable as point, line, and polygon attributes. The combination of remote and ground-based data types constitute the universe of data that could reside in a relational database sandwiched between the object databases. All of these, in turn, would be employed to form a genetic algorithm that would describe an ecosystem. Figure 2 illustrates this design.

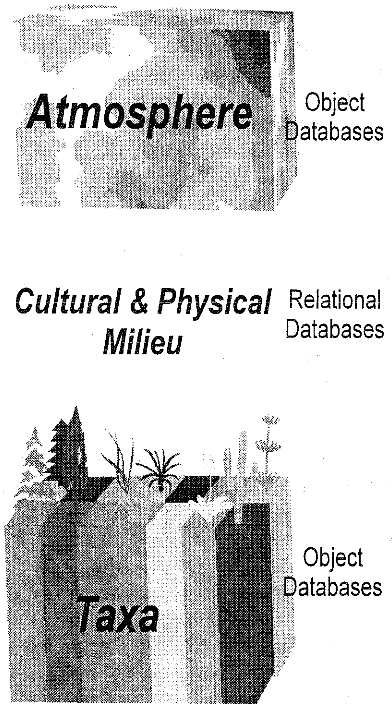

Figure 2. Object and Relational Databases

A critical requirement of any relational GIS is that all data sets be accurately co-registered. In past, this has been accomplished by geocorrecting images using ground control points and by improving the pointing accuracy and pixel-to-pixel registration of sensors. Another technique is to use Global Positioning Systems (GPS) capable of geolocating raster, point, line, and polygon features with accuracies in the centimeter to meter range. GPS is the most recent addition to the spatial data tool kit. It represents the enabling technology for merging data sets into spatially coherent systems, and offers the hope that at least static object and relational database elements will be compatible in space and time.

Latitude, longitude, and elevation coordinates for each raster element are simultaneously recorded through triangulation from a constellation of satellites at the time of data collection. As the technology gets incorporated into future aerial and satellite sensors, the technique should become pervasive in field biology from pinpointing specimen collection sites to recording physical environmental attributes. At present, however, almost none of the millions of site and specimen records have GPS coordinates. While these older data would be useful in model building, they may be beyond use in GIS terms because they would impart spatial inaccuracies into models that would disrupt self-adaptive processes. The dilemma of rectifying or replacing these early records over the next few decades, in time to be useful for biodiversity and environmental modeling, places further importance on remotely acquired data sets, and argues in favor of automating traditional field biology techniques (Morain, 1993).

Remote Atmospheric Measurements

Remote sensing technology has developed rapidly since the mid 1960s (Morain and Budge, 1996). Some atmospheric data sets that began as sensor experiments have progressed into operational programs managed national and international agencies (e.g. National Oceanic and Atmospheric Administration [NOAA], and METEOSAT). Experimental global data sets from the Very High Resolution Radiometer (VHRR), for example, were initiated on NOAA/POES-1 in 1970 and progressed to an advanced sensor (AVHRR) on NOAA-6 in 1978. Despite its name, AVHRR is neither advanced nor high resolution. It collects five channels of visible, near infrared, and thermal infrared data primarily for cloud cover and cloud formation information. In its local area coverage (LAC) mode, its spatial resolution is ≈ 1.1 kilometer. Calibrated retrospective data from this sensor are now available from 1985-1991 (NOAA/NESDIS, 1992).

Retrospective AVHRR measurements represent one of many available time-series data sets. Other satellite data sets include solar irradiance, aerosol content, trace gas species and concentrations, water vapor, and land use. For broad area climate change research, these data are conveniently, and perhaps only, acquired by satellite-based sensors, or by other remote platforms. To prototype a process that self-adapts to the stability resident in *order*, on the one hand, but which can also respond to the *chaos* of entropy on the other, would be a quantum step for global change research programs. There is arguably no higher reward for collecting data from aerial and satellite platforms than one which links the emergent processes of nature with resulting patterns in the landscape, and which accomplishes this without introducing human bias.

The Earth Radiation Budget Instrument (ERBI) was also inaugurated in 1978 on NASA's Nimbus-7, but has since flown on a dedicated platform (ERBS) in 1984 and as part of the payload on NOAA/POES 9-10. Its purpose is to record radiation budget, aerosol, and ozone data for global climate change research. It measures monthly and seasonal variations in radiation balance at regional scales on a ≈ 50 kilometer grid.

Many other atmospheric measurements are being collected from satellites. These are mainly water vapor, trace gas and aerosol concentrations, ultraviolet radiation, radiation budget, total irradiance, wind, and temperature profiling. Limb and vertical sounding measurements, and measurements with specific tropospheric, mesospheric, or stratospheric depth sensitivities are among the many data collection strategies. For a variety of logistical and technical reasons (but mainly because of budget constraints), few of these data, even for the troposphere, have been assembled for experimentation in ecosystem models. The point being made here, however, is that remote sensing systems of the future will be collecting data sets that can be used to characterize biological systems *in addition* to their primary use in developing meteorological predictions and global circulation models.

Most of the current ecosystem experiments and landscape studies still rely on *in-situ*, near-ground measurements because remotely acquired data seem too inaccessible to incorporate into research designs. Sensor systems of the future should provide a continuous stream of spatially and temporally contiguous data sets to augment these discontinuous *in-situ* measurements, which should in turn hasten the development of rule-based models.

Remote Terrain Measurements

AVHRR data are also a striking example of secondary uses for satellite observations. They are often analyszed for their spectral content in terrestrial studies of vegetation index patterns; and to analyze principal components of long time-series data sets to map and monitor severe landscape changes like fire and deforestation (Eastman and Fulk, 1993). There is much more that might be done with these data, if the pixels were spatially correctable with greater accuracy using GPS coordinates. Future AVHRR sensors will have this capability, allowing the data to be analyzed pixel-by-pixel for subtle and gradual landscape changes, and as continuous data that might drive one or more modeling rules.

Beside AVHRR, there are numerous sensors already collecting terrain data in a variety of spectral wavelengths with ground sampling distances ranging between one and thirty meters. They operate from aircraft and space altitudes employing visible, infrared, and microwave frequencies. Most are opto-mechanical or electro-optical scanners, radiometers, radars, or lidars, or digital cameras. Measurement strategies and recording techniques have undergone ten to twenty years of basic and applied research, and more recently, operational use. Some of them offer relatively high pointing accuracy and/or pixel-to-pixel registration, but most suffer from imprecise pixel locating ability. As with atmospheric sensors, the next generation of sensors acquiring terrain data will be equipped with simultaneous GPS recording that will allow accurate spectral and temporal data merges. The technology is trending toward multiresolution capabilities in the spatial, spectral, and temporal domains.

Next-generation sensor systems are being designed to enhance their utility for ecosystem modeling. In addition to better geolocational capabilities, hyperspectral sensors are being developed to record data in hundreds of bands and with bandwidths ≤5 nanometers. Calibrated data cubes generated from such sensors should enable much finer analyses of biological signatures, and depending upon platform altitude should have ground sampling distances on the order of meters to kilometers. In the best circumstances, individual species and microhabitats will be discernable for modeling. With repeated coverage, time sequences of these signatures might permit analyses to link object databases with relational coverages of other physical parameters.

Sensing Programs

The future of remote sensing data collection is ensured by needs recognized by Heads-of State and by the fact that both government and commercial interests in the technology are driving development costs downward. In the next two years, NASA plans to inaugurate its Earth Observing System (EOS). It is designed as a fifteen-year series of low Earth orbiting satellites collecting reflectance, emission, and absorption spectra for atmospheric, hydrospheric and biospheric phenomena. Similarly, the European Space Agency (ESA) plans to launch the Environmental Satellite (ENVISAT) and the Japanese will launch their Advanced Earth Observing Satellite (ADEOS). All are multinational research and applications missions aimed at future operational systems. Many other nations also are contributing smaller satellites and sensors for more specific data collection programs. By 2010, there are scheduled to be more than 60 polar orbiting and geostationary satellites gathering Earth data. Sadly, there is not yet a broadly scoped program to integrate these data with ground-based and object data into holistic models of ecosystem evolution. In the authors' view it is time to begin discussing and planning these system integrations.

Data Integration

System and data integration will necessarily involve reanalysis of the datasets themselves. Even though data are calibrated for detector sensitivity, sensor, drift, and other aberrations within a sensor's lifetime, the entire collection of data from a suite of like sensors operating over long time frames need to be systematically calibrated between sensors.

Another aspect of integration will involve substituting real data, where possible, for mathematically derived values in current models. With specific regard to global circulation models, the new long-term data sets from EOS and supplementary data sets now being developed that span the current century could provide critical testbeds for improving the accuracy of climate models. Progress in model development and improvement can be accelerated by comparisons of this kind, and through the continued intercomparison of the results of different models.

Most important of all in the integration process, however, is the need to develop strategies that link remotely acquired data about the environment to the genetic networks of individuals and genotypes, and to then program how species ought to react to changing environments. For example, measurements of ultraviolet radiation flux gain their greatest application today as a predictor of ozone concentrations related to the ozone hole. Those same measurements, and others like them, may be useful as stimuli for biological mutations and species evolution, especially in high altitude areas.

CONCLUSIONS

Rapid scientific progress is being made in genome characterization, gene response mechanisms, learning processes, and the complexity of self-organizing systems. Remote sensor data sets, employed in raster and vector GIS and that integrated genome and atmospheric data as objects, could provide the framework for a genetic algorithm. As these databases began to assemble and integrate a sufficiently wide diversity of object and relational variables, rules could be developed directing their interactions based on field and laboratory observations. The complexity of hundreds of plant species interacting as individuals and as communities will never be fully understood, but perhaps society can learn enough about how ecosystems operate to at least emulate how they will react to human and natural stimuli.

REFERENCES

Carpenter, S.R., S.W. Chisholm, C.J. Krebs, D.W. Schindler, and R.F. Wright, 1995. Ecosystem Experiments, Science, 269, pp. 324-327.

Eastman, J.R. and M. Fulk. 1993. Long Sequence Time Series Evaluation Using Standardized Principal Components. Photogrammetric Engineering and Remote Sensing, 58(8), pp. 1307-1312.

Kauffman, S. 1993. The Origins of Order: Self-Organization and Selection in Evolution. New York: Oxford University Press. 709pgs.

Morain, S.A. 1993. Emerging technology for Biological Data Collection and Analysis. Annals, Missouri Botanical Gardens, 80 (2), pp.309-316.

Morain, S.A. and A.M. Budge. 1996. Earth Observing Platforms and Sensors. Vol. 1 of Manual of Remote Sensing (R.A. Ryerson, editor-in-chief). Bethesda, MD: American Society for Photogrammetry and Remote Sensing. CD-ROM.

NOAA/NESDIS. 1992. Global Change Database: Volume 2, Experimental Calibrated Global Vegetation Index from NOAA's Advanced Very High Resolution Radiometer, 1985-1991. Boulder, CO:National Geophysical Data Center. CD-ROM.

Stølum, H-H. 1996. River Meandering as a Self-Organization Process. Science, 271, pp. 1710-1713.