Journal of Zhejiang University SCIENCE
ISSN 1009-3095
http://www.zju.edu.cn/jzus
E-mail: jzus@zju.edu.cn

Spatial/temporal indexing and information visualization genre for environmental digital libraries

CHEN Su-shing[†1], GRUNWALD Sabine[2]
([1]Department of Computer & Information Science & Engineering, University of Florida, Gainesville, Florida 32611, USA)
([2]Department of Soil and Water Science, University of Florida, Gainesville, Florida 32611, USA)
[†]E-mail: suchen@cise.ufl.edu
Received Aug. 5, 2005; revision accepted Sept. 10, 2005

Abstract: Protecting and preserving our environmental systems require the ability to understand the spatio-temporal distribution of soils, parent material, topography, and land cover as well as the effects of human activities on ecosystems. Space-time modelling of ecosystems in an environmental digital library is essential for visualizing past, present, and future impacts of changes occurring within such landscapes (e.g., shift in land use practices). In this paper, we describe three novel features, spatio-temporal indexing, visualization, and geostatistical genre, for the environmental digital library, Environmental Visualization and Geographic Enterprise System (ENVISAGE), currently in progress at the University of Florida.

Key words: Spatio-temporal indexing, Geostatistics, GIS (Geographic Information System), Visualization, Environmental digital library, Spatio-temporal search
doi:10.1631/jzus.2005.A1235 **Document code:** A **CLC number:** TP391

INTRODUCTION

A large number of environmental digital libraries have been designed and implemented recently to effectively manage environmental data such as satellite photos and images, map images, cartographic materials, and so on. Support for storing, distributing, searching for, retrieving, and utilizing this information is imperative in order for it to be used to its maximum potential benefit and effectiveness (Hicks and Tochtermann, 1998). The variability of environmental data may be a major hurdle in developing an environmental digital library. In addition, environmental data are dispersed in heterogeneous digital library systems and are usually large in size. In this paper, we present the development of an environmental digital library, Environmental Visualization and Geographic Enterprise System (ENVISAGE), in progress. Currently we are focusing on soil and landscape properties as a first step towards building the large environmental digital library.

Development of spatio-temporal models in Geographic Information Systems (GISs) was hampered in the past by the large amount of input data required and the complexities of ecosystem processes. Yet with emerging geographic information technologies (GITs), such as high-resolution remote sensing (RS), digital elevation modelling, digital GPS, and geostatistics, explicit spatio-temporal ecological landscape modelling becomes possible. Scientists argue that such high spatio-temporal resolution models yield better results than lumped (coarser resolution) models. Serious scientific questions remain to be answered:

(1) How much space-time complexity (or simplicity) do we need to address ecosystem problems?

(2) How much do we gain by pushing the spatial resolution of above-ground landscape properties into the sub-meter (potentially into the cm) range, while subsurface properties are still too labor intensive and time consuming to collect?

(3) How different or similar are spatial landscape patterns at fine and coarse scale?

Research in the area of explicit space-time modelling of ecosystems focusing on the seamless integration of subsurface and above-ground landscape properties within the space-time continuum is vitally needed.

This paper addresses these research issues using environmental datasets and GIT. We develop a novel space-time data model with emerging GIT capabilities, and improve the reconstruction and scientific visualization of eco-landscapes utilizing geo-statistics and emerging 3D visualization and interaction technologies to support investigative analyses of eco-landscape data. Spatial pattern analyses serve to quantify emerging landscape patterns at different scales. Results are embedded in a Web-based framework which is an object-oriented information system of multimedia objects to provide users access to (3D/4D/5D eco-landscape models) results of research on metadata, photographs, animations, hyperlinks, and auxiliary ecological information. This modular system facilitates also the educational needs at a variety of levels (K-12, undergraduate and graduate students, professionals, lifelong education).

SPATIO-TEMPORAL DATA MODELS

Much research has been conducted in the combined handling of spatio-temporal data in large spatial databases. In this section we discuss space-time data models in GIS.

The general geographic space-time data model is based on the "Where (space)-When (time)-What (attribute)" concept. There are three geographic space dimensions (x: eastward; y: northward, z: downward depth from the soil surface) and time (4th dimension), relegating attribute values to the 5th dimension as shown in Fig.1 (Abraham and Roddick, 1999).

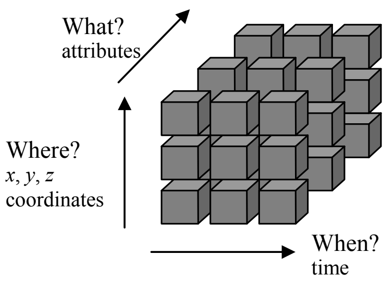

Fig.1 Space-time data model (Abraham and Roddick, 1999)

Generally, continuous geographic space must be divided into discrete units for management and display. The resulting tessellation is taken as a reasonable approximation of reality at the level of resolution under consideration. The simplest geographical data model of reality is a basic spatial entity, which is further specified by attributes and geographical location (spatial coordinates or geometry) and relationships (topology). This model can be further subdivided according to one of the three basic geographical data primitives, namely, a point, a line, or an area (polygon). These are the fundamental units of the vector data model. This entity-based model or object-based model defines the real world as an arrangement of discrete, well-defined objects characterized by their geometrical and topological properties and by their non-spatial attributes.

In contrast, the field-based or space-based model displays the real world as a set of pixels or voxels (volume cells). The field-based model is adequate for modelling natural phenomena that do not show obvious boundaries (e.g., hydro-, pedo-, and geomorphological processes). The original phenomena must be clearly identified; i.e., the spatial variability of attributes in 3D geographic space must be described to optimize the size of geographic entities (e.g., polygons) or fields (pixels and voxels). Most digital subsurface data use the entity data model. For example, the classical approach of soil mapping is to define classes of soils (e.g., Soil Series) and then to identify areas of land (entities) corresponding to these classes, represented by vector boundaries, which contain zones of equal value. Digital soil data provided by the Natural Resource Conservation Service (NRCS) for the conterminous U.S. include the State Soil Geographic Database (STATSGO) and the Soil Survey Geographic Database (SSURGO), which are both based on the entity data model. However, crisp sets allow only binary membership functions (i.e., true or false); an individual is either a member or not a member of any given set as defined by exact limits. The current approach does not allow ambiguities, and too inflexible to take account of genuine uncertainty. We intend to develop a statistical, probability, or fuzzy approach to manage the spatial variation in both the soil-forming processes and in the resulting soils.

BASIC DL SYSTEM TO BUILD THE EXTENSIONS

The project's overall goal is to develop an interactive, Web-based environmental geographic information and visualization system, ENVISAGE. This system is based on an object-oriented digital library (DL) system developed at the Laboratory for IT Enterprises (LITE) of the University of Florida. The system has established a number of Web-based DL services: database query and full text search, user profiling, recommender and notification, ingest, review, role-based access control, OAI (Open Archives Initiative) metadata harvesting, OAI extension to federated search and data mining/information visualization (Chen and Choo, 2002; Kim *et al*., 2003). We develop a space-time data model extending these services for search, data mining, and visualization of spatial and temporal objects. To test and validate the system, we populate the digital library with subsurface properties, topographic, and land-cover data for selected eco-regions collected at different scales by the University of Florida for the U.S. Government. This information system can expand to cover other eco-regions in the U.S. or anywhere else. Our research objectives are as follows:

(1) To develop a space-time data model for DL services, we will enhance the current DL services for spatial and temporal objects by adding content-based indexing to database query and full-text search capabilities. Our current DL services use eXtensible Markup Language (XML) data interchange. We build the space-time data model using GeoXML and OpenGIS formats and implement visualization using GeoVRML and Vis3D.

(2) To ingest soil, topographic, and land-cover data for selected eco-regions and employ 3D ordinary and block kriging (i.e. sampling and interpolating) methods to create continuous eco-landscape models, we evaluate the uncertainty of model predictions using validation and cross-validation of datasets.

OAI harvesting server to harvest several federal data archives for developing our content

Recently the OAI has established a metadata harvesting protocol that promises the interoperability at the service exchange level (Chen and Choo, 2002). We harvest metadata from heterogeneous metadatasets of several websites and digital library collections:

(1) Digital Earth: http://www.digitalearth.gov/;

(2) Environmental Systems Research Institute Inc. (ESRI) Geography Network: http://www.GeographyNetwork.com;

(3) Environmental Protection Agency's (EPA) Geo-Data Clearinghouse: http://www.epa.gov/nsdi/;

(4) U.S. Geological Survey's (USGS) GeoSpatial Data Clearinghouse: http://nsdi.usgs.gov/;

(5) National Oceanic Atmospheric Administration's (NOAA) Geospatial Data Center: http://www.noaa.gov/;

(6) Source of Environmental Representation and Interchange (Sedris): http://www.sedris.org;

(7) Alexandria Digital Library (ADL) at UCSB: http://www.alexandria.ucsb.edu/;

(8) National Imagery and Mapping Agency (NIMA): http://www.nima.mil/;

(9) Natural Resources Conservation Service—NRCS's Soil Information Systems;

(10) STATSGO: http://www.ftw.nrcs.usda.gov/stat_data.html;

(11) SSURGO: http://www.ftw.nrcs.usda.gov/ssur_data.html;

(12) MUIR: http://www.statlab.iastate.edu/soils/muir/;

(13) Soil Series Classification: http://www.statlab.iastate.edu/soils/sc/;

(14) NASIS: http://nasis.nrcs.usda.gov/;

(15) National Soil Survey Center: http://www.statlab.iastate.edu/soils/nsdaf/.

LOVE digital library system

We have implemented our Learning Object Virtual Exchange (LOVE) server for the NSF NSDL (National STEM Education Digital Library) Program with OAI harvesting capabilities. Our digital library (DL) server maintains a registry of harvested metadata which mediates the heterogeneity of various metadatasets. Also, users may submit their own metadata to the DL server with object URL addresses. Since metadata describes intrinsic and extrinsic attributes of information content and collections at an abstract level, the cross walk among various concepts must be established. We download a sufficient amount of datasets for users to visualize online. The tool set will be disseminated to the public so that they can use either the above data resources or their own datasets.

We incrementally digest what has been harvested within our DL server, which contains existing metadatasets created internally or harvested externally from collections. This digesting process is supported by many tools, one of which is the existing search and retrieval capability of a DL server. We implement several automated tools which are useful to this server. First, the digesting process is automated with minimal human intervention with a rule-based system. The second tool is a data mining system to refine harvested metadatasets and to synchronize them with the existing metadatasets (Kim *et al*., 2003). Thirdly crosswalks are established in the metadata registry system to support federated search. Of course, this DL server can be harvested by other service providers so that a collaborative group of DL servers can jointly refine the digesting process. Finally, this DL server is integrated with the 3D visualization tool. We harmonize metadatasets from different sources by building semantic maps between them. We also address the federated search engine's problem in query translation.

LOVE is a full-fledged DL server with search and retrieval capabilities. It is implemented in the Java, Tomcat, Apache, and MySQL environment. Our main design objectives of LOVE as a DL server are adaptivity, interactivity, and openness. Adaptivity is needed to select and customize the learning resources to the learners and to the context in which the learning is taking place. The two aspects exhibit a wide range of variability for digital libraries. Such systems cannot make a priori assumptions on the learner's characteristics, such as educational background, cognitive style, etc., nor about the context and purpose of the learning process but instead must be able to adapt dynamically based on explicit knowledge of aspects that must be maintained independently of the more generic learning content knowledge. Interactivity allows learners to engage and interact with the DL server. The 3D visualization tool is an excellent example. Openness permits new data resources to be added to the existing collections. We feel that the vast amount of federally supported datasets can be exploited and utilized. In a LOVE server, we provide learning objects in standardized forms so that we can index user profiles and learning objects and match them directly. The LOVE architecture consists of the collection managed by a community (e.g., a school district) and several networked services, such as user profiles management, recommender system, and notification system. Recommender and notification systems pull and push the information to users.

EXTENSIONS TO LOVE

This section describes our approaches to extend LOVE to the environmental DL, ENVISAGE.

Combining learning object metadata with geospatial metadata

The IEEE Learning Technology Standards Committee (IEEE-LTSC P1484) has undertaken the initiative of drafting a set of standards among which they define a data model for Learning Object Metadata (LOM). This standard has received the endorsement of other consortiums dealing with educational standards such as ARIADNE, Instructional Management Systems (IMS) Consortium, and Shareable Courseware Object Reference Model (SCORM) for the Advanced Distributed Learning Network (ADL-Net) within the Department of Defense.

Several of these standards are being endorsed by the IMS Consortium, who in addition is developing the Content Packaging Information Model, which describes a self-standing package of learning resources. The IMS Content Packaging Information Model describes data structures that are used to provide interoperability of Internet-based content with content creation tools, learning management systems, and run time environments. The objective of the IMS Content Packaging Information Model is to define a standardized set of structures that can be used to exchange content. These structures provide the basis for standardized data bindings that allow software developers and implementers to create instructional materials that interoperate across authoring tools, learning management systems, and run time environments that have been developed independently by various software developers. The IEEE-LTSC LOM model is an abstract model; however, the IMS Consortium has provided one possible binding specification using pure eXtensible Markup Language (XML) and XML-Schema standards. The XML Schema introduces an unambiguous specification of low-level and intermediate level data types and structures that assure a higher level of interoperability between XML documents.

The Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata (CSDGM) is to provide a common set of terminology and definitions for the documentation of digital geospatial data (FGDC, 1998). This standard is structured in a hierarchy of data elements and compound elements. In this standard there are seven compound elements, which are identification information, data quality information, spatial data organization information, spatial reference information, entity and attribute information, distribution information, and metadata reference information, under the top-level compound element.

We combine the LOM with the CSDGM to develop an environmental metadata model. The current metadata model includes the following ten compound metadata elements, some of which either are optional or can occur multiple times:

(1) General: title, catalog/entry, language, description, keyword, coverage, and aggregation level.

(2) Lifecycle: version, status, and contributor.

(3) MetaMetaData (Metadata of data resources): identifier, catalog/entry, contributor, metadata scheme, and language.

(4) Technical: format, size (bytes), location, requirements (installation, platforms), and duration.

(5) Pedagogical: interactivity type, learning resource type, interactivity level, semantic density, intended end user role, learning context, age range, typical learning time, and description on how to be used.

(6) Rights: cost, copyright, and description of condition of use.

(7) Relation: program and resource (target).

(8) Annotation: person, date, description, and comment.

(9) Classification: taxon and taxon-path.

(10) Geospatial Metadata: identification information, data quality Information, spatial data organization information, spatial reference information, entity and attribute information, distribution information, and metadata reference information.

From this set, the Relation and Classification elements and sub-elements are specifically relevant from the perspective of supporting 3D visualization. The Relation element provides an ideal practice set of visualization routines and parameters of the target LO or resource to which the current LO has access. Their use for navigation between LOs, in a kind of visualization process, is very important. Note that these relations do not have to link to other LOs necessarily, so relations with other resources that may provide associated active-content is a possibility here. The Classification element provides the principal mechanism for extending the LOM model by allowing it to reference a Taxonomy and describe associated taxon-path sub-elements corresponding to the LO. Thus, classification provides for multiple alternative descriptions of the LO within the context and meaning of several Taxonomies.

In our LOVE collection, LO's represent small capsules of knowledge in a form suitable for didactic presentation and for assimilation by learners. We believe that the LO metadata standardization will introduce a large degree of interoperability and re-use, promoting the widespread investment in, and adoption of, educational technology. Each learning object by being highly atomic and complete in capturing a concept or "learning chunk" provides the opportunity for the configuration of a large number of course variations. The resulting fine-grained customization is expected to lead to "just-in time", "just-enough", "just-for-you", training and performance support courseware. This implies the traversal of a subject matter domain in a highly flexible and learner-specific way. However, this flexibility must comply with inter-LO dependencies and restrictions, which in turn will require new goal-driven more intelligent navigation facilities.

There is a variety of different LO's ranging from "metadata LO's", "geospatial LO's" to "course material LOs". We tightly link the geospatial LOs to the 3D scientific visualization. A demonstration library of 3D soil-landscapes based on soils, topographic, land use, and environmental quality data will be developed for representative landscapes in the U.S. This results in a tool set of 3D scientific visualization models. ENVISAGE is readily expandable to include other environmental data and information such as biological, physical and chemical collections.

Spatial temporal modelling approach

We employ spatial modelling techniques to investigate patterns of soil, land-cover, and topographic properties at different scales (ELUD, ELLD, and regional scale). To represent landscape properties at regional scale, we use readily available GIS datasets from federal and state geo-databases for the purpose of visualization. The impact of different support can be identified using block kriging. This technique enables us to more clearly identify the phenomena (or processes) responsible for producing the heterogeneous eco-landscape patterns. To separate the different factors (e.g., human, geologic, pedogenic) responsible for the spatial distribution of landscape properties (e.g., pH, soil organic carbon) we use factorial kriging. This method identifies the spatial variation at different scales (short, medium, and long-range variability). Spatial patterns of soil property data will be compared to available entity-based soil maps from SSURGO and STATSGO in the following paper, respectively.

Spatial data are stored and used in the DL services as a sequence of spatio-thematic object triplet $\{(s, e, p)\}$, where s is a spatial coordinate (x, y, and z where x represents eastward, y is northward, and z represents the depth measured from the soil surface downward), e is elevation (i.e., the altitude above mean sea level with reference base 0 m sea level), and p is a soil or landscape property (e.g. soil texture, soil horizon, phosphorus concentration). To maintain data consistency indexing is used. The spatio-thematic object triplet can be expanded to include time (t). Numerous pedo-geomorphological processes are relatively slow; for example, it takes several thousand years to form a new soil horizon. However, other ecosystem processes are highly dynamic; for example, the transport of nitrate in soils during a rainfall event changes at very short time steps (seconds to minutes).

To ingest soil, topographic, and land-cover data for selected eco-regions and employ 3D ordinary and block kriging to create continuous eco-landscape models, we are currently evaluating the uncertainty of model predictions using validation and cross-validation of datasets. Because soil attributes are labor intensive and costly to collect, robust prediction methods were developed using geo-statistical methods to create field-view subsurface models (e.g., models describing the spatial distribution of soil texture, zinc and arsenic concentrations, and water content. Geo-statistics is based on: (1) the assumption that observations close to each other are more likely to be similar than observations at a larger distance from each other; and (2) intrinsic stationarity, which assumes that the variance of the difference is the same between any two points that are the same distance and direction apart no matter which two points are chosen. The spatial variation is subdivided into 3 components: deterministic variation, spatially autocorrelated variation, and uncorrelated noise. The spatial correlation between observations is described by the semivariance, which is half the average squared difference between paired data values. The fitted semivariogram model identifying the spatial variation of attributes is subsequently used in weighted interpolation methods to predict values at previously unsampled locations. True 3D variogram modelling and interpolations are still rare (Grunwald *et al*., 2000).

SEARCH OF SPATIAL TEMPORAL DATA

We retrieve soil, topographic and land use geo-data for representative landscapes in the U.S. from geo-data clearinghouses via our digital library server with OAI harvesting and search capabilities. Searching based on metadata records allows the user to locate relevant basic data items utilizing the standard descriptive information that has been defined for them (Hicks and Tochtermann, 1998). The metadata search provides users with a way to search on part or all of the archived metadata from the archives harvested (Kim *et al*., 2003). The search service is based on full-text search on the predefined metadata elements from the harvested metadata collection. A simple, keyword-based search is provided, where the title and description metadata elements from all harvested archives are queried. The resulting records returned are ranked based on the relevance to the user's query term.

We also provide an advanced search for users who want more refined results for their search process. The query fields provided are based on the predefined metadata elements and users can sort the resulting records based on the results' relevance to the query term, the date the record was created or updated, and the original archive name. Boolean operators (AND/OR) can be used with the query fields. Users are also provided with the option to choose the archives that they want to query on. The result set from the search process can be further refined with the filter elements.

In addition to metadata search, we maintain an alternate spatial index based on the proposed spatial temporal data model. The availability of alternate indexes can be very useful in searching for soil, land-cover, and topographic data. By using the alternate index, we are able to perform spatial-temporal queries as follows (El-Geresy and Jones, 2000): point query where the search space is a point; line search space queries where the search space is a line; plane search space queries where the search space is a plane, and volume search space queries where the search space is a plane.

Various metadata categories, such as soil, parent material, topographic, and land-cover data, are integrated into this Spatio-Temporal-Context Modelling (STCM) methodology with user-based navigation and searching (i.e., users can dynamically navigate, search, and visualize different categories) in ENVISAE. For example, "Land use" focusing on specific functions of eco-landscapes is derived for each study site. This category of metadata can be searchable according to subcategories in Gainesville, Orlando, or Miami, when users select a particular location, says Florida. It is unlikely that an all-encompassing spatio-temporal modelling and information system cater to all individual requirements. We do not attempt to deliver a generic 3D/4D/5D information system. Instead, we present optimizing space-time representations of eco-landscapes using DL services to better understand ecosystems.

After retrieving the results, we use spatial modelling (kriging) to interpolate retrieved geo-data to create two different types of models: (1) Stratigraphic models with polyhedron face geometry representing taxonomic (e.g. Soil Series and soil horizons) and classified data (e.g. drainage classes), and (2) Solid models with voxel geometry representing gradual changes in soil and landscape properties and environmental quality data (e.g. phosphorus concentrations). Both model types are implemented in 3D geographic space.

A major component of this research is the adoption of state-of-the-art 3D visualization and interaction technologies to support investigative analyses of soil-landscape data. The overall system design is depicted in Fig.2.

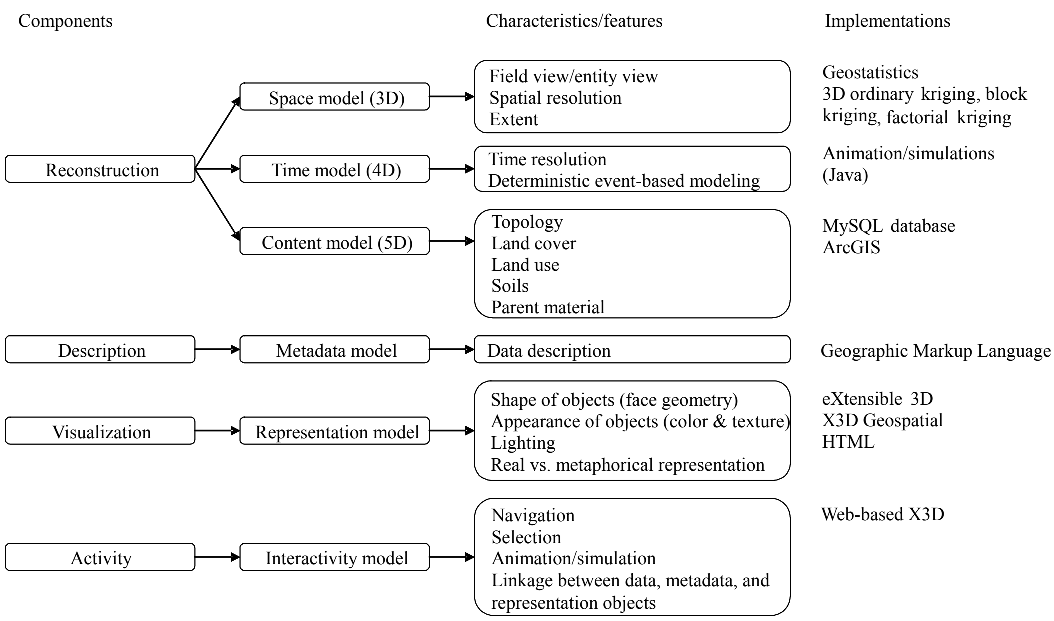

Fig.2 A system design describing the components, models, and characteristics of our approach and how they are implemented

VISUALIZATION

We explain environmental datasets, scientific visualization of eco-landscapes using geostatistics, and analyses of eco-landscape data. Fig.3 shows a summary of our visualization approach.

Environmental datasets

Environmental datasets are derived from this research and public available geodatabases. A versatile use is desirable because the collection of environmental data is labor intensive and costly.

In an ongoing effort in the Santa Fe River Watershed (3585 km^2) in North Central Florida we collect environmental datasets (e.g., nitrogen, phosphorus, soil texture, total carbon, pH, and so on) at 100 locations at four different soil depths resulting in 2400 validation samples. Additionally, at 50 locations we collect 1200 validation samples.

The Florida Wetland's Geo-Database and WebGIS (http://GISWetlands.ifas.ufl.edu) assembles soil physical and chemical data. We standardized and integrated 2130 geo-referenced point observations of 78 different soil physical, chemical, and biological attributes collected in Florida's wetlands from 1987 to present.

The Florida Geographic Data Library (FGDL: http://www.fgdl.com) provides hundreds of GIS data layers from State and Federal Agencies and research project to users including orthophotos, land use derived from Landsat^{TM} remote sensing images, topographic contour data, and other land resource data. FGDL provides one of the richest geospatial collections in the nation.

Over 1100 soil profiles were sampled and analyzed as part of the Accelerated Soil Survey Program at the Environmental Pedology Laboratory, the Soil and Water Science Departments at the University of Florida, which provides a valuable resource of physical, chemical, mineralogical, and morphological soil property information. This dataset currently being standardized and converted from hardcopy to digital geodatabase provides a valuable resource of environmental data.

We use validation dataset to evaluate the uncertainty of model predictions at two sites. Cross-validation is used to assess the uncertainty for all study sites. Statistical measures include the standard predi-

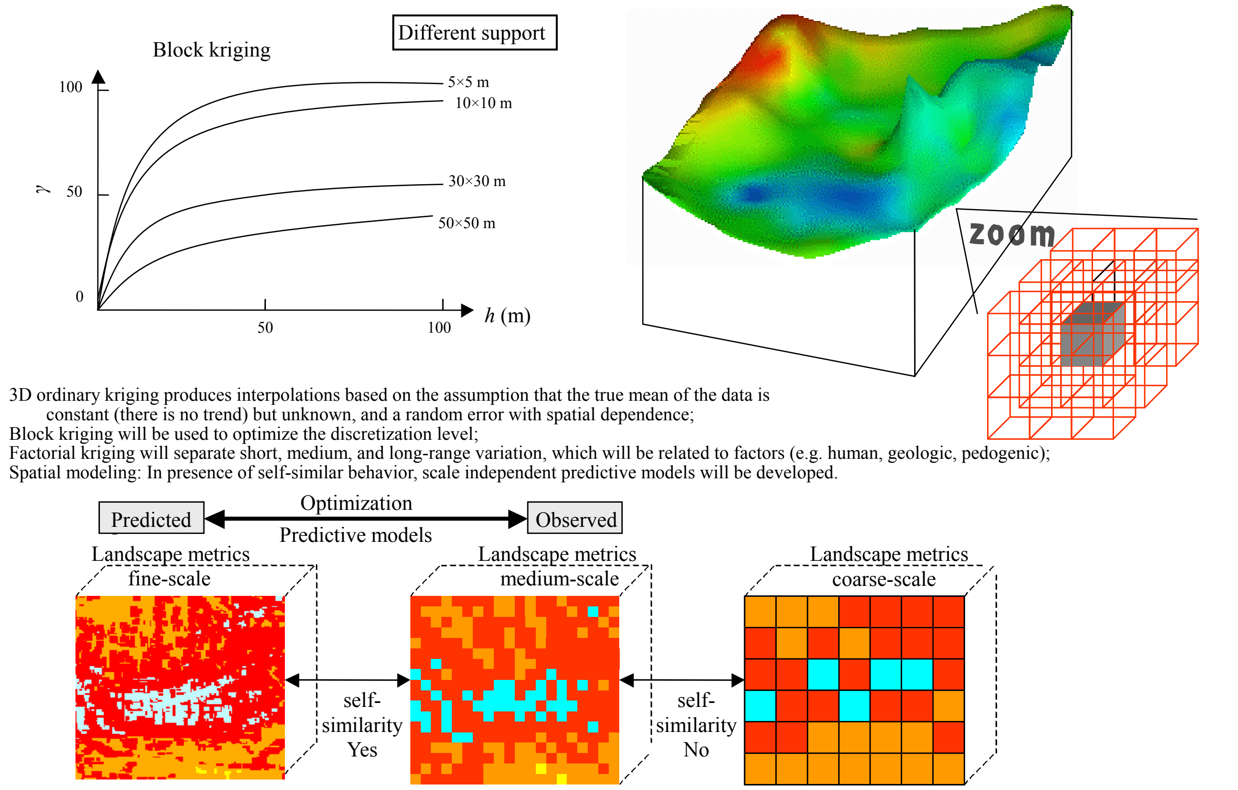

Fig.3 Summary of visualization approach

ction error and the root mean square error. Besides a least-square estimate, the kriging variance is computed, and the resulting confidence interval is compared with actual measurements in the validation step.

Visualization methods

A major component of this research is the adoption of emerging 3D visualization and interaction technologies to support investigative analyses of eco-landscape data. We plan to adopt the evolving Web-based 3D graphics format called eXtensible 3D (X3D). This format succeeds the International Organization for Standardization (ISO)-standard file format known as the Virtual Reality Modelling Language (VRML). VRML is a rich format for 3D Web graphics with advanced support for rendering, lighting, animations, scripting, audio, and video. VRML and X3D were developed by the Web3D Consortium (Beitle *et al*., 2002). The X3D specification includes integrated support for a component called GeoSpatial, which provides projection capabilities, streaming of large terrain grids, and other geospatial services (Reddy *et al*., 2001). The adoption of X3D and the GeoSpatial component can bring a number of unique benefits this project, including:

(1) Web-based: X3D is designed from the bottom up to support the distribution and visualization of 3D models over the Web, which improves the learning capabilities of the system; for example, it improves distance learning.

(2) Interactivity: Important investigative visualization capabilities are included to allow users to manipulate the 3D representation, navigate around it, and query it for further information.

(3) Geospatial support: Through the X3D GeoSpatial component, support is provided for highly accurate landscape visualization of surfaces.

(4) Open standards/open source: International open standards provide solutions for geo-data distribution that are not tied to the proprietary products of certain companies.

(5) Extensibility: X3D supports the creation of new application-specific extensions that plug into the existing core. We will use this feature extensively to provide customized eco-landscape visualization capabilities.

(6) Componentization: X3D is based on a hierarchical structure called profiles that provide application-specific features.

(7) Rich feature set: X3D supports a rich set of features, such as texture mapping, transparency, animation, manipulation, and lighting effects, that will be important in producing high-quality visualizations.

(8) Backward compatibility: X3D is being designed to ensure backward compatibility with VRML to enable easy migration to the new technology and offer continued support for the existing user base and legacy content.

(9) State-of-the-art tools: The adoption of X3D in this project will introduce state-of-the-art 3D graphics technologies to real and virtual classrooms, providing a compelling and exciting global learning environment.

In addition, we are trying to go beyond the GeoSpatial X3D component, which is limited to surfaces (2D plane) and design and specify a new environmental component for X3D that provides sufficiently rich capabilities to handle the space-time-content ecodata gathered in this project. We plan to create a XML schema to support the representation of data, with specific support for entity and voxel-based eco-landscape models. We will build on this toolset to create an immersive digital 3D experience for environmental data that will not only link to the educational component but also demonstrate new research in 3D visualization. We will produce appropriate visualizations of the various eco-data types as well as user interaction metaphors and enable the user to query the scene for further details, such as full metadata descriptions and hyperlinking to auxiliary information.

We believe that the research result is a powerful suite of visualization and interaction tools that utilize state-of-the-art Web-based 3D graphics that can be immediately deployed in the classroom. Our ecosystem data are represented in the form of representation models mimicking natural ecosystems and metaphorical representations. For example, a tree might be represented using face geometry to recreate branches, stems, and leaves. Metaphoric abstraction represents the same tree as a point or a cube. The latter method is attractive for complex and small-scale landscapes exhibiting many interrelated properties. In particular, at the EL Upper Domain level, it is challenging to visualize ecosystem patterns, such as the spatial distribution of above and below-ground landscape properties. Isomorphing, transparency, and level of detail visualization, which varies with scale, based on the grouping of objects into classes, must be explored to optimize scientific visualization of our STC datasets.

To identify the representative land-cover (LC) classes, we process the Landsat ETM7+ scenes utilizing ERDAS Imagine software. We acquire two sets of cloud-free images for each calendar year, corresponding to early-planting and post-harvest stages in order to capture the temporal variability within the study site and, hence, better differentiate croplands from pasture and rangelands. Using single-date imagery, it is possible to obtain only the spectral reflectance pattern for each LC class, and often, differences between the reflectance patterns may not be significant enough to separate them. By using multi-temporal imagery, the temporal changes for each class can be derived in addition to its spectral reflectance pattern. The accuracy of the LC map can be assessed, and the goal is to achieve an overall accuracy of at least 80% or higher. If certain classes are not classified to the required accuracy levels, iterative classification procedures are adopted to generate subclasses. This process is repeated until the overall accuracy of the map is at least 80% or higher. To analyze the AVIRIS dataset, we use gains and offset derived from the calibration targets. The images are calibrated to ground reflectance in the full AVIRIS spectral range by the empirical line method. AVIRIS radiometric calibration factors are calculated by measuring the response of AVIRIS to an integrating sphere (a known target illuminated by a known light source). This calibration is reported to be accurate to within 7%. Atmospheric corrections to the AVIRIS data can be carried out with models, such as MODTRAN II. The atmospherically corrected AVIRIS radiance is compared against the ground target. The differences between the corrected AVIRIS data and the ground measurements with the field spectrometer are used to create a forced-fit function that combines sensor and residual atmospheric corrections.

The variables influencing an ecological process may or may not change with scale. In particular, we attempt to discover whether the spatial landscape patterns at fine scale (ELLD) are reproduced at coarser scale (ELUD and regional scale). To test self-similar behavior we use lacunarity which is the deviation of a fractal from translational invariance. According to (Plotnik *et al*., 1996), lacunarity can be considered a scale-independent measure of the heterogeneity of patterns. They showed that lacunarity can be computed for binary and quantitative datasets. Self-similarity may only exist over a limited range of orders of magnitude. If self-similarity exists for a specific landscape property, landscape-metrics will be used to transcend patterns across scales. Spatial pattern analyses conducted in all selected study areas provide a large range of contrasting landscape characteristics. Landscape patterns are compared, using indices that measure diversity (i.e., heterogeneity), dominance (i.e., deviation from the maximum possible diversity), and dispersion (to capture the tendency for patterns to be contagious (clumped) or uniform). For example, a typical measure of diversity is entropy. Predictive models are developed based on an optimization procedure such as simulated annealing. Our methodology is driven by the fact that currently high-resolution datasets and analyses of above-ground landscape properties (e.g. high-resolution hyperspectral scenes) for large ecoregions are prohibitively expensive. Dense soil sampling covering large ecoregions is labor intensive and costly. A predictive transfer method is capable of computing landscape patterns at fine scale without using dense, exhaustive sampling. Our method would provide predictive capabilities for high-resolution eco-landscape models until sub-surface sensors are developed to measure soil properties continuously so that all end-users can afford high-resolution remote sensing techniques.

Spatial pattern analysis yields a better understanding of the complexity/simplicity of eco-landscapes. Our results would be valuable for the design of future sampling efforts and improve understanding of ecosystems, which is based on knowledge about the spatial variability of landscape properties.

Analyses

1. Reconstruction

The low and high-resolution topographic data are used to describe soil surface elevation changes. Topographic attributes, such as slope, profile and plan curvature, aspect, and flow patterns, are derived for both datasets using ArcGIS software. Land-cover data are superimposed on the land surface.

We use 3D variogram analysis to identify the spatial structure of observed soil properties. This information is used in 3D ordinary kriging to create 3D soil representations. Depending on the content data type, we create entity and field view models. For example, soil horizons can be transformed into virtual polyhedrons (entity view), and bulk density and calcium content can be transformed into virtual voxels (field view). The geostatistical analysis is conducted using EVS-PRO and SPlus software. Variograms guide the spatial discretization level for each soil property, depending on the spatial variability. Block kriging and factorial kriging are employed and predictions compared to ordinary kriging. Block kriging is helpful to identify the spatial discretization level which may vary among soil properties. Fig.4 shows 3D soil layer model and Fig.5 depicts 3D soil-landscape model, respectively.

2. Metadata description

Metadata are provided in Geographic Markup Language (GML), which builds on XML recently adopted by the Open GIS Consortium (OGC). GML provides seamless data sharing and exchange of geo-

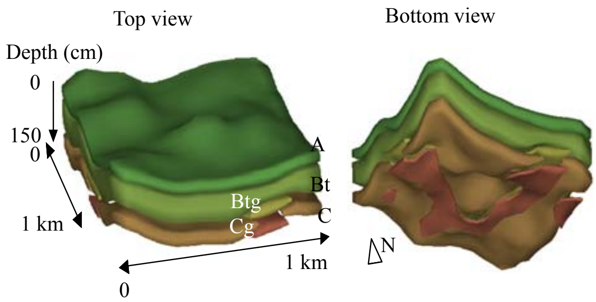

Fig.4 3D stratigraphic model showing the spatial distribution of soil horizons. Soils on elevated areas: Plano, Ringwood and Saybrook Series (fine-silty, mixed, mesic Typic Argiudolls); soils in depressions: Joy and Ossian Series (fine-silty, mixed, mesic Aquic Hapludolls and Typic Haplaquolls, respectively)

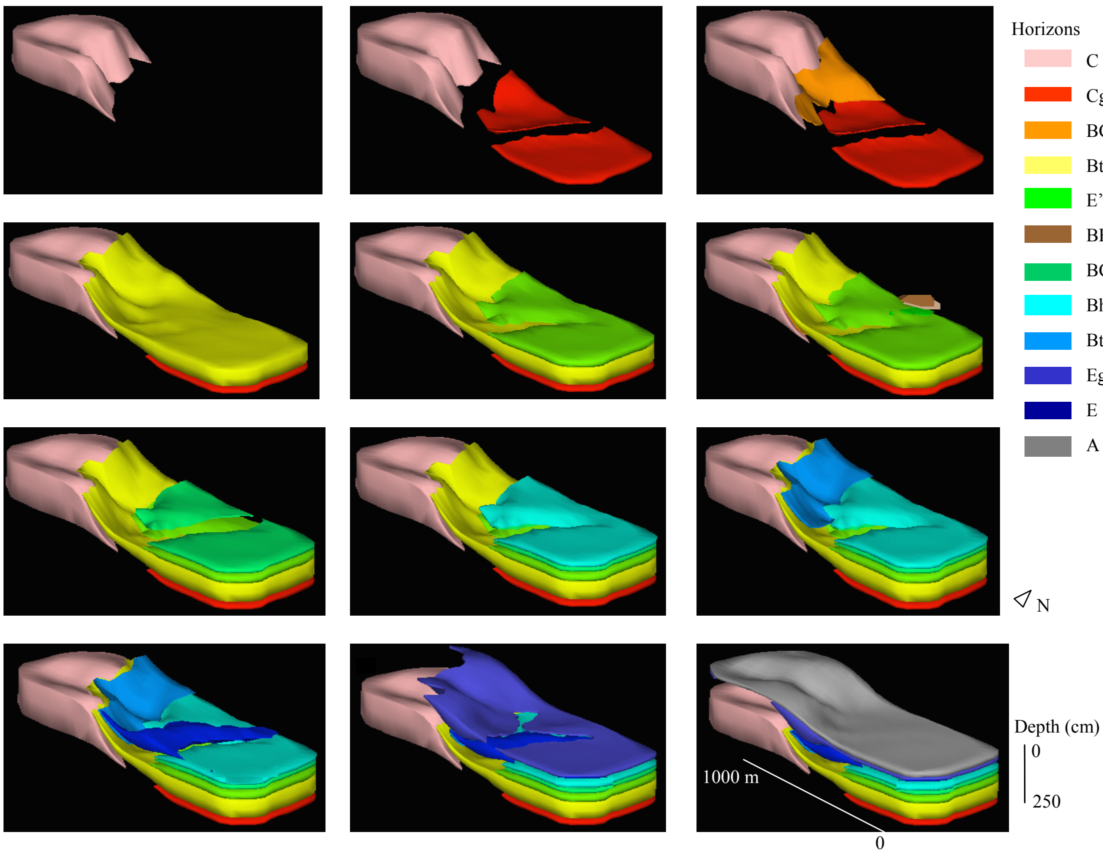

Fig.5 3D soil-landscape model showing the spatial distribution of soil and topographic patterns. The elevations ranging from 70~105 m were represented with a vertical exaggeration factor of 5 to clearly distinguish between upslope and toeslope landscape positions

graphic information via the WWW. We follow the outline of the Federal Geographic Data Committee (FGDC)-approved metadata standard "Content Standards for Digital Geospatial Metadata".

EVALUATION METHODOLOGY

The evaluation approach to asking questions is based on two basic premises drawn from research on technology and learning. Technology per se does not determine learning outcomes. Rather, learning outcomes are influenced by the choices that Internet users, teachers, students, and others make about the organization of teaching and learning tools and choices about content. The role of information technology is to expand the available choices.

The evaluation process enables us to identify weaknesses in the scientific visualization component, transparency of the digital library, and usefulness of ENVISAGE for education resulting in "participatory" design of our tool. There is a tight feedback loop between our proposed research and the evaluation procedure. Table 1 outlines the Flashlight-based methods we are now conducting evaluation of ENVISAGE and the infusion of modelling and visualization into existing courses. We will present the evaluation results of ENVISAGE in the following paper.

The proposed evaluation procedure is an objective method to judge the effectiveness of ENVISAGE for learning and teaching. Most important is to assess if these new technologies (e.g. scientific visualization in form of 3D models) really do improve the outcome (e.g. students learn more effectively). Simply using new technology for the sake of it might not be useful at all. Flashlight provides an activity-centered approach to evaluation. Knowing that technology was used and that outcomes improved (or stayed the same, or got worse) does not help anyone decide anything. To make improvements we need to know how the technology was used (the activity that used it).

RELATED WORK

Currently, no such universal spatio-temporal system exists, but there are a variety of prototypes. Abraham and Roddick (1999) provided a review of spatio-temporal database concepts. Koeppel and Ahl-

Table 1 Methods applied to evaluate ENVISAGE

Goal	Criteria	Method	Evaluation tool	Test groups
Evaluate how the technology was used	Information content and use of the technology	Structural analysis	Questionnaire posted on the ENVISAGE Web page	Internet users
Evaluate speed to access information; speed to perform a specific task	Time to find data and information	Compare NVISAGE to 2 geo-portals and 2 commonly used search engines	Structural experiment (clock)	(1) K-12 students from the Alachua School District (Florida); (2) Undergraduate students (UF); (3) Graduate students (UF)
The empirical study will identify if students better understand or recall environmental information when they are juxtaposed within the 3D environment as compared with the way they currently learn	2D vs 3D scientific visualization	Compare the usefulness of information provided in form of 2D GIS layers to 3D models accessible via ENVISAGE	Empirical study conducted in the classroom	Students enrolled in course SOS6932 "GIS in Land Resource Management", UF (instructor: Sabine Grunwald) and CIS 6950 "Digital Libraries", UF (instructor: Su-Shing Chen)
Evaluate how attractive ENVISAGE is to users (e.g. do first time users come back to revisit the site? Nationalities of visitors etc.)	Number of hits	Counting	WebTracker	Internet users

mer (1993) distinguished between attribute-oriented spatio-temporal databases that track changes in information about spatial entities, and topology-oriented spatio-temporal databases track changes in positional information about features and their spatial relationships. Whigham (1993) proposed a dual-ordered, hierarchical structure, where time and events are represented in their own hierarchies, placed on a spatial background reference. Hermosilla (1994) argued for a temporal GIS with reasoning capabilities based on artificial intelligence. Peuquet and Duan (1995) suggested an Event-Based Spatio-Temporal Data Model (ESTDM) focusing on events represented along a temporal vector in chain-like fashion. Yuan (1997) suggested a three-domain model representing semantics, space and time separately and providing links between them to describe geographic processes and phenomena. The Source of Environmental Representation and Interchange (SEDRIS) developed by the Defense Modelling and Simulation Office (http://www.sedris.org) offered a Data Representation Model, augmented with its environmental data coding specification and Spatial Reference Model, for the representation and seamless interchange of environmental data. Commercial GIS software, such as ESRIs ArcView and ArcGIS, uses the relational database concept. These software packages are rather limited in their data management.

Commonly, GIS are used to create 2D maps to represent environmental data (Osher and Buol, 1998). Other soil-landscape representations use a 2½D design, where soil or land-cover data are draped over a digital elevation model (DEM) (Hogan and Laurent, 1999) to produce a 3D view. Since this technique describes patterns on 2D landscape surfaces rather than the spatial distribution of subsurface attributes (e.g., soil texture, soil horizons), it fails to represent 3D soil-landscape reality. The integration of above and below-ground attribute layers is accomplished using GIS technology. While 3D information technology dominates the video game market and has received wide acceptance in various fields, 3D environmental representations are still rare. Even fewer studies use reconstruction along with virtual reality techniques to portray soil data in 3D space (Grunwald *et al*., 2001).

CONCLUSION

ENVISAGE enables Internet users to access and explore selected ecosystems. Eco-landscape models are useful teaching tools and can be accessed by any instructor. Three-dimensional scientific visualization enables users to learn intuitively about different environmental settings. Metadata describing the source of data and photographs are hyperlinked. We envision that the system will grow by adding more information layers (e.g., bio-data, geologic data) and sites over time. Building an online eco-landscape tool set will be beneficial for education at all levels (K-12, undergraduate students, graduate students, professionals, lifelong education) in both individual and collaborative settings. The resulting product will provide a compartmentalized learning environment based on Learning Objects that are small capsules of knowledge addressing specific educational needs at a variety of levels.

One major advantage of our proposed system is that implementation is based on Open Source software and standards (GML, X3D, GeoSpatial), providing interoperability and standardized exchange of data and information. Our models are not limited to a specific vendor, and access requires no client-based conditions besides Internet access. Our implementation provides interactivity, which is most important for educational tools engaging students in hands-on activities. Three-dimensional scientific visualization can impart huge amounts of data, allows the perception of emergent properties that were not anticipated, and facilitates understanding of both large and small-scale features of eco-landscape data.

We believe that our research develops an innovative approach to integrate soil, land-cover, and topographic data. Three-dimensional geostatistical applications provide more realistic geographic representation of environmental datasets than previously possible. Spatial variability and pattern analyses may allow us to explore scaling behavior of landscape properties across multiple scales in different landscape settings. Innovation is the infusion of geostatistical techniques with spatial pattern analyses. The complexity of ecosystems is X-rayed, making subsurface properties visible.

References

Abraham, T., Roddick, J.F., 1999. Survey of spatio-temporal databases. *GeoInfomatica*, **3**(1):61-69.

Beitle, M., Brutzman, D., Diefenbach, P., Parisi, T., Reddy, M., Silverglate, D., 2002. X3D Architecture and Profiles. Tutorial at the 7th International Conference on 3D Web Technology, Tempe, ZA, p.24-28.

Chen, S., Choo, C., 2002. Extending OAI Metadata Harvesting to Federated Search. ACM/IEEE JCDL 2002, Portland, OR.

El-Geresy, B., Jones, C., 2000. Models and Queries in a Spatio-temporal GIS. *In*: Atkinson, P., Martin, D.(Eds.), GIS and Geocomputation, Innovations in GIS 7. Taylor & Francis, New York, p.27-39.

FGDC (Federal Geographic Data Committee), 1998. Content Standard for Digital Geospatial Metadata, FGDC-STD-001-1998.

Grunwald, S., Barak, K.M., Lowery, B., 2000. Soil landscape models at different scales portrayed in Virtual Reality Modelling Language. *Soil Science*, **165**(8):598-615.

Grunwald, S., Barak, P., McSweeney, K., Lowery, B., 2001. The use of VRML for virtual soil landscape modeling. *Systems Analysis Modelling Simulation*, **41**:755-776.

Hermosilla, L.H., 1994. A Unified Approach for Developing a Temporal GIS with Database and Reasoning Capabilities. Proceedings of EGIS 94, The 5th European Conference and Exhibition on Geographical Information Systems, Paris, France.

Hicks, D.L., Tochtermann, K., 1998. Environmental Digital Library Systems. Proceedings of the 1st GI-Workshops "Hypermedia und Umweltschutz", Metropolis Verlag.

Hogan, M., Laurent, K., 1999. Virtual Earth Science at USGS (U.S. Geological Survey). http://virtual.er.usgs.gov/.

Kim, H., Choo, C., Chen, S., 2003. An Integrated Digital Library Server with OAI and Self-Organizing Capabilities. ECDL 2003, LNCS.

Koeppel, I.J., Ahlmer, S.D., 1993. Integrating the Dimension of Time into AM/FM Systems. Proceedings of AM/FM XVI Int. Annual Conference, Aurora, Colordo.

Osher, L.J., Buol, S.W., 1998. Relationship of soil properties to parent material and landscape position in eastern Madre de Dios, Peru. *Geoderma*, **83**:143-166.

Peuquet, D.J., Duan, N., 1995. An event-based spatio-temporal data model (ESTDM) for temporal analysis of geographical data. *Int. J. of Geographical Information Systems*, **9**:7-24.

Plotnik, R.E., Gardner, R.H., Hargrove, W.W., Prestegaard, K., Perlmutter, M., 1996. Lacunarity analysis: a general technique for the analysis of spatial patterns. *Physical Review E*, **53**(5):5461-5468.

Reddy, M., Iverson, L., Leclerc, Y., Heller, A., 2001. GeoVRML: Open Web-based 3D carography. Proceedings of the International Cartographic Conference (ICC2001), Beijing, China.

Whigham, P.A., 1993. Hierarchies of Space and Time. Spatial Information Theory: A Theoretical Basis for GIS. LNCS 716, Springer, New York.

Yuan, M., 1997. Modelling Semantics, Spatial and Temporal Information in a GIS. *In*: Craglia, M., Couleclis, H. (Eds.), Progress in Trans-Atlantic Geographic Information Research. Taylor & Francis, New York.