HYBRID GIS AND REMOTE SENSING IN ENVIRONMENTAL APPLICATIONS

Kurt Fedra and Milan Kubat
International Institute for Applied Systems Analysis
Laxenburg, Austria

RR-93-20
December 1993

Hybrid Geographical Information Systems reprinted from *EARSel Advances in Remote Sensing* (1992) **1**(3):89–100.
A Global Change Impact Assessment System: GIS, Models and Expert Systems reprinted from *Proceedings of the 25th International Symposium on Remote Sensing and Global Environmental Change* (1993) ISSN: 1068–9281, pps. I-657–I-668.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
Laxenburg, Austria

Hybrid Geographical Information Systems reprinted with permission from *EARSel Advances in Remote Sensing* (1992) **1**(3):89–100.

A Global Change Impact Assessment System: GIS, Models and Expert Systems reprinted with permission from *Proceedings of the 25th International Symposium on Remote Sensing and Global Environmental Change* (1993) ISSN: 1068–9281, pps. I-657–I-668.

Printed by Novographic, Vienna, Austria.

Foreword

Geographical information systems (GIS) are a relatively new and rapidly developing class of computer applications. They show considerable potential in a growing number of application domains, regional and environmental planning and management being one of them. The integration of GIS methods adds a key technology for spatial analysis to the set of tools of applied systems analysis, and environmental systems analysis in particular.

IIASA's Advanced Computer Applications (ACA) project develops and implements environmental information and decision support systems that bring together key technologies such as data base management, simulation and optimization modeling, computer graphics, expert systems, and geographical information systems.

The two papers combined in this volume address the use of GIS and satellite imagery in environmental applications. Hybrid GIS, that combine both vector and raster information, are the topic of the first report that was presented at the *EARSel Workshop on Relationship of Remote Sensing and Geographical Information Systems* which took place in Hannover, Germany, in September 1991. The integration of satellite imagery in a global change information system that also uses rule-based expert systems methods, together with GIS data structures and tools, is the theme of the second part. This paper was presented at the *25th International Symposium – Remote Sensing and Global Environmental Change, Tools for Sustainable Development*, held in Graz, Austria, in April 1993.

Both papers illustrate some of the ACA project's attempts to explore and advance the use of new information technology for applied systems analysis.

Kurt Fedra
Leader
Advanced Computer Applications Project

HYBRID GEOGRAPHICAL INFORMATION SYSTEMS [1]

Kurt Fedra and Milan Kubat

Advanced Computer Applications (ACA)

International Institute for
Applied Systems Analysis (IIASA)
A-2361 Laxenburg, Austria

ABSTRACT

Geographical Information Systems (GIS) are gaining increasing importance and widespread acceptance as tools for information and decision support systems for infrastructure, natural resources and environmental management and spatial analysis, and urban and regional development planning. GIS assists in the preparation, analysis, display, and management of geographical or geo-referenced data.

This paper describes experience gained at IIASA's Advanced Computer Applications Project (ACA) with a hybrid GIS, GRASS, that combines vector based representation formats and functions with raster or cell-grid based components.

Following a short description of GRASS itself, including its main features and functionality, the use and integration of the system is then described in a number of practical applications. Based on the hybrid approach, and data sets digitized or pre-processed with GRASS, a number of implementations of dedicated GIS components and tools within the framework of several model-based information and decision support systems are then presented and illustrated. Examples include environmental information systems, air and water quality models, and an expert system for environmental impact analysis.

The paper finally addresses issues of generic versus customized tools, user interface considerations, efficiency, integration and portability, that practical experience in the use of hybrid GIS suggest as major topics for future developments and research.

Introduction

Geographical information systems (GIS) are computer based tools to capture, manipulate, process and display spatial or geo-referenced data [Burrough, 1986]. They contain both geometry data (coordinates and topological information) and attribute data, ie., information describing the properties of geometrical objects such as points, lines, and areas [Armstrong and Densham, 1990].

[1] Paper presented at the EARSeL Workshop on Relationship of Remote Sensing and Geographical Information Systems, September 16-18, 1991, University of Hannover, Germany

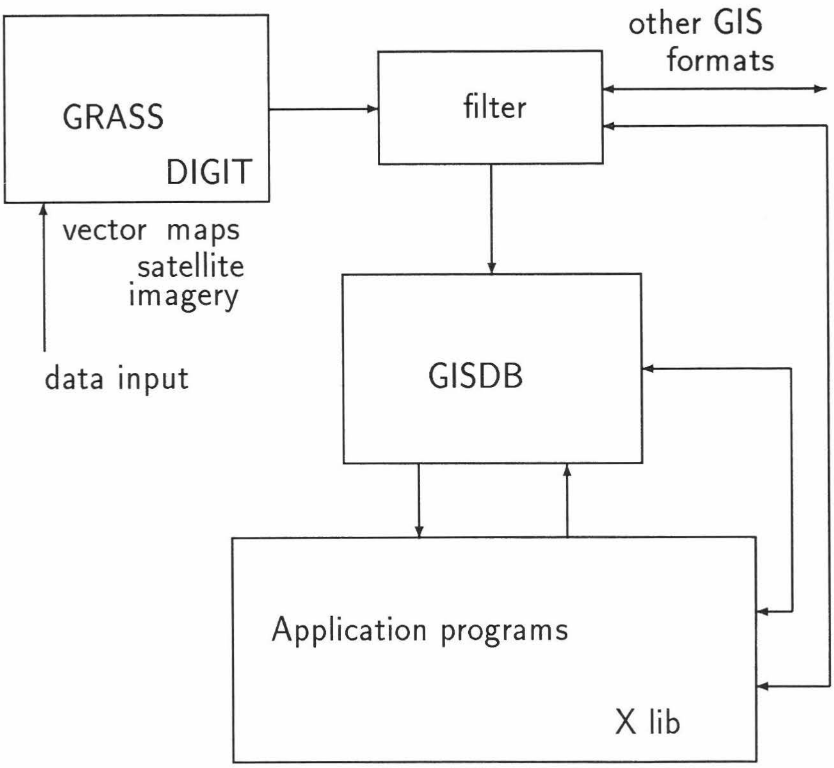

Figure 1: Using GRASS for preparing geographical data eventually used in some other system

Data sources for GIS's are printed maps that need to be digitized or scanned, and remote sensing systems such as aerial photography or satellite imagery [Estes et.al., 1986]. Less common data input systems include video or various surveying methods with direct digital output.

Data storage in GIS is most commonly based on vectors, or rasters, ie., regular grids, and quadtrees as an intermediate format [Anthony and Corr, 1987; Ibbs and Stevens, 1988]. Systems that employ more than one format can be called hybrid systems [Visin et.al. 1990].

Like with any other tool, different forms of storage have different advantages and disadvantages in terms of precision, storage efficiency, and implications for processing. Hybrid systems recognize this fact and attempt to take advantage of it [Tilley and Sperry, 1988; Sena, 1989, Peuquet, 1983; Piwowar et.al. 1990].

And like any other tool, GIS are built for a purpose; this purpose will shape their design and functionality. GIS are powerful tools for planners and decision makers concerned with problem domains that include a spatial dimension such as urban or regional planning, environmental management, transportation planning, etc [Parker, 1988; McGregor, 1988].

Since numerical simulation models and expert systems are also tools for decision support in the above areas, and, at the same time, another possible source of spatial data, their combination with GIS holds promise for a very powerful class of integrated tools. This integration in dedicated information and decision support systems requires features that are best met with a hybrid approach [Armstrong, 1990].

The following sections first describe GRASS, which stands for Geographical Resource Analysis Support System. This public domain hybrid GIS was developed by the US Army Corps of Engineers for land management, environmental impact analysis, and similar applications [Gardels, 1988; GRASS, 1988; Canters and Decleir, 1989; Shapiro et.al. 1989]. We then discuss how GRASS can be integrated in dedicated information systems, and describe data formats and management structures and specific tools for embedded GIS functions. This philosophy is illustrated with a number of examples for environmental information and decision support systems.

GRASS: a hybrid GIS

This section describes experience gained at IIASA's Advanced Computer Applications Project (ACA) with a hybrid GIS, GRASS, that combines vector based representation formats and functions with raster or grid cell based components.

GRASS can be used for preparing geographical data eventually used in some other system, or for processing and analyzing geographical data, including displaying the results. In the former case, geographical data are imported or digitized, converted, edited, analyzed and processed in GRASS and then exported to some other system, for instance a modeling or decision support system or an expert system, where they build a geographical information data base (Figure 1). For analysis, GRASS offers a number of analytical tools and functions for problem oriented displays. With the release of version 3.0, the system consists of nearly 200 different computer programs implementing a broad range of functionality. However, all analysis functions in GRASS are based on grid cell files, ie., a raster format.

The Grid-cell format

Grid cell files are a central element in GRASS. GRASS grid cell data are stored as a matrix of grid cells. Each grid has an absolute spatial reference in terms of its position and size. Each cell is assigned a single integer attribute value called the category number. Each cell is stored in the file as one to four 8-bit bytes of data.

The physical structure of a cell file can take one of three formats: uncompressed, compressed, or reclassed. The uncompressed cell file actually looks like an $N \times M$ matrix. Each byte or set of bytes for multi-byte data represents a cell of the map layer. The physical size of the file, in bytes, will be $rows \times columns \times bytespercell$. The compressed format uses a run-length encoding schema to reduce the amount of disk required to store the cell file. Run-length encoding means that sequences of the same data value are stored as a single byte repeat count followed by a data value. A reclass map layer does not contain any data, it contains references to another map layer along with a schema to reclassify the categories of the referenced map layer. The reclass cell file itself contains no directly useful information. The reclass information is stored in the cell header file.

In addition to the cell data file itself, there is a number of support files for the grid cell map layer. The files which comprise a grid cell map layer all have the same name, but each resides in a different database directory. These files are: grid cell header file, map layer category information, map layer color table and map layer history information. The cell file itself has no information about how many rows and columns of data it contains, or which part of the earth the layer covers. This information is in the cell header file. The format of the cell header depends on whether the map layer is a regular map or a reclass layer. The regular map layer cell header contains information describing the physical characteristics of the cell file. The cell header has the following fields: projection, UTM zone, geographic boundaries, resolution, format containing bytes-per-cell information and compress indicator. A cell header for a reclass cell contains the name of the referenced cell file and the category reclassification table.

The cell category file contains a title for the map layer, the largest category value which occurs in the data, and one line label for each category. The cell color table file contains one line of a color description for each category of data. The colors are represented as levels of red, green, and blue. The cell history file contains historical information about the cell file like creator, date of creation, comments, etc. The history file is generated automatically along with the cell file.

Data input and loading

The basic problem of any GIS is data acquisition. In principle, there are two possibilities: either import data, which means load them from some storage medium and convert them into the specific format used by the GIS, or to digitize them.

Another possibility may be scanning them, which is in principle very close to importing them. The result of importing data is either a vector-based or a grid cell based information. To use the grid cell based analytical tools, vector-based information resulting from digitizing has to be converted into the grid cell form. The only additional information needed for this process is the size of cells which need to be created. In most cases, this step of data processing means loosing some accuracy, since it is not feasible to set the cell size small enough and, at the same time, keep the size of the whole data set in some reasonable ranges.

In GRASS, there are a number of routines that convert some standard grid based data formats into an internal GRASS representation. Formats supported include Landsat Multispectral Scanner data, Landsat Thematic Mapper data, SPOT data, to name few of the satellite data, or Digital Terrain Elevation Data eg., produced by the US Defense Mapping Agency, as an example for elevation data.

In addition to these formats, any raster file containing latitude/longitude grid data and a GRASS specific ascii text file can be imported. Also, any vector based information may be converted into a grid cell based one. The USGS Digital Line Graph format and a GRASS specific ascii text format are supported.

These standard GRASS features for data loading and conversion may be extended by the user on two different levels:

- external - converting available data into any format, that is supported by current GRASS software, or
- internal - write a filter that does all the user specific readings and uses the GRASS library for building data structures consistent with GRASS.

GRASS contains a number of libraries for individual purposes, so that the user should be able to implement even a more complex additional functionality without much effort.

Projections

The recommended projection in the GRASS system is the Universal Transverse Mercator grid system, UTM. GRASS uses a number of different, predefined versions or spheroids, each with its respective regional advantages and disadvantages and distortions. GRASS includes utilities for converting individual latitude/longitude points as well as latitude/longitude grid cell based data into UTM.

The basic limitation regarding UTM consist in the fact that the system supports only processing information within one of 60 UTM grid zones (between 84° N and 80° S) at a time. If the user wants to work with a larger area than one single UTM zone, the UTM zone has to be extended or the information has to be split into pieces belonging to individual UTM zones, that have to be processed separately. If the source data contain no UTM information, the system can treat them in a simple x-y coordinate system, without any specific projection. Alternatively, data can be fitted to a UTM zone if a sufficient number of reference points can be found.

GRASS contains utilities for marking points with known latitude/longitude information, computing the parameters of the rectification and carrying out the rectification.

Attributes and colors

The basic information a grid cell holds are the data value or attribute classification. Data converted from their vector-based representation use a classification based on the attribute labels of the vector data. The grid cell value or attribute translates into a color for the display. Thus, the number of available colors on a given graphics systems imposes some practical limits on the number of categories that can be used simultaneously. Otherwise some categories have to share colors, which may make it difficult to distinguish them on the screen.

Should it be necessary to have more categories than the number of available colors supports on a given system (eg., 256 for an 8 bit color frame buffer) and to be able to distinguish all of them, GRASS supports a numeric displaying utility, that creates an ascii file containing a matrix of individual cell-file category numbers. Another solution for this situation is splitting the whole information into pieces containing only a part of all categories or using a mask which filters a group of available categories. As far as the use of the graphic monitor for the display of cell-based information is concerned, a number of utilities is available in GRASS for assigning user specific or random colors to individual categories, zooming and picking individual data items, selecting graphic terminals, setting appropriate window parameters etc. There is also a utility that generates the maximum number of distinct colors for a given data set, ie., automatically stretches the contrast.

As one of the possible data sources for GRASS satellite imagery can consist of a number of bands. If these bands can be directly interpreted as red/green/blue components of the color image, a GRASS utility may be used for building the composite colors. There are many methods of multispectral classification, but generally, they fall into three groups: supervised classification, unsupervised classification and combinations of supervised and unsupervised classification. GRASS 3.0 only supports a simple unsupervised classification. It is implemented as a two-pass process, where the user defines the number of initial classes, minimum class size, minimum class separation, percentage of convergence, maximum number of iterations, sampling intervals and bands belonging to red, green and blue color component.

Grid cell files can be patched together from several tiles of arbitrary shape, so that they combine information from several sources covering the entire area of interest. The system is able to patch individual tiles of any shape on the basis of the description of the tile position, tile size and the fact that empty cells can be simply distinguished from those that already contain information. A simple precedence rule applies for multiple valid data on the same location. The patching utility of GRASS can be also used for processing files containing non-overlapping information (eg., land use categories) covering the same general region. This is actually a way one can create a more complex geographical layer on the basis of several simpler information sources.

Any information that is imported into GRASS or digitized in GRASS is provided with a description regarding size and position of the tile in its header file, so that when the user decides to patch some specific tiles, no additional information is needed.

Processing and analytical functions

The goal of all the features described in the previous paragraph is to create a complete geographical information layer, which means information that has the right position in some coordinate system, contains features specific for the layer, contains description of individual geographical features, color specification of individual geographical features, scale etc. Additionally a layer may contain information regarding the sources that have been used for creating the layer, like reference to another group of layers containing the individual color components, registration points, imported raw data, source vector data etc.

Features described in the following part are oriented at processing, combining and analyzing individual layers. There are a number of utilities in GRASS for processing and combining layers, that are of different complexity. The basic challenge for the user is in the right choice of utilities, which usually means the simplest possible utility.

GRASS contains a number of complex functions, for instance interpreting arithmetic formulas or interpreting rules, that are able to solve very complex tasks but may need a huge amount of processor time to fulfill the task with the required flexibility.

The basic grid cell GRASS functions can be divided into two groups:

- those that process only one data file and carry out operations like: increasing area sizes by several cells; assigning unique labels to individual areas; assigning a value to a cell that depends on cell values in the neighborhood; computing distances from specific areas; resampling, rescaling or reclassing and
- those that use several layers of grid cell information as input and compute, for instance, coincidences of specific classes, filter specific zones of interest or create cross products of multiple sources.

A special feature of GRASS is the ability to set a mask that dynamically specifies the current zone of interest for subsequent processing. The more complex functions are oriented either by some kind of sophisticated operation like processing a number of sources on the basis of arithmetic formulas, or even on the basis of some rules, or they are oriented by processing of special kinds of information. An example of such information are the elevation data. In this particular case, it is possible to compute the terrain surface, slopes, optimal paths or visibility in the terrain. All the resulting cell-raster based information can be displayed on the graphic monitor in two or three dimensional projection. In addition to displaying the results on the screen, their ascii counterpart, ie., a matrix of individual cell values, can be created and displayed numerically. This representation of information is also suitable for exporting results. The colors can be displayed in fix or float mode, which means in a mode with preallocated colors or in a mode that allocates colors dynamically as they are needed. Since GRASS supports only processing of information in grid cell representation, and it can be useful to keep the information in its vector representation, there are conversion routines between the two formats.

The set of GRASS functions is completed by a number of functions for copying, renaming, removing, grouping informational layers and statistical functions, that compute cell numbers of individual classes, areas covered by individual classes etc. Since the average size of a processed grid cell file is rather big, it may be advisable to use all the data files in their compressed form.

A dedicated GIS implementation: ACA's GISDB

While GRASS itself is sufficiently modular and uses a standardized I/O structure, it is still a general purpose tool with a rather specific character-based user interface with only a limited set of graphic drivers available in Release 3.0

To integrate geographical data and embed selected GIS functionality into more complex information and decision support systems together with simulation and optimization models and expert systems within one common environment and graphical user interface, a specialized database **GISDB** and related display and analysis functions **xacalib** has been developed at ACA as part of an **Applications Interface Tool Kit.**

The basic advantage of implementing a dedicated GIS database is that one is able to process all geographical data in a unique way in all systems dealing with geographical information. It allows to design formats and tools that are adapted to the well defined set of tasks required. All functions, and in particular, all display oriented functions, can be tightly embedded in the style of the applications, sharing a common graphical user interface and environment.

The individual functions processing geographical data have been integrated into an appropriate library and their functionality is oriented at preprocessing geographical data, reading binary GIS data from disc, finding data items possessing specific characteristics, displaying individual GIS objects and their groups and additional features, like for instance changing map resolution or picking map elements displayed on the screen.

All data integrated in the database are preprocessed by a special filter converting data from GRASS into a database specific binary data format or the binary data are created on the basis of some model results. Alternative filters that convert data from other GIS systems or standard GIS data export formats are built as required. The necessary additional information specifying colors used, attributes, and properties is located in a header file. All displaying functions that operate on the GISDB data structures are X Windows based, built directly with X lib.

The binary files are designed for fast loading. The system uses a strategy of RAM databases as much as possible, ie., the entire binary map data are loaded at startup time whenever possible. Where map data sets are too large to be handled effectively in RAM even by powerful virtual memory workstations, a multi-step approach loading data selectively and on demand is used. Data can be partitioned in terms of layers of different resolution in tiles covering one resolution layer.

All data that cover one area create a map. A map is divided into overlays containing individual or groups of features in a topological structure. To improve performance, the overlays can be split into smaller sections containing only one part of the overlay information called tiles. Tiles neighbor each other without overlapping. Tiles can be created for all resolutions implemented.

Individual overlays consist of geographical objects called map elements. Like in GRASS 3, the map elements can be of type line, area or grid cell. An additional map element type supported by the geographical database is a point. This map element type was added to the GRASS 3 set of types in order to process point objects like pollution sources, measurement stations etc.

Every map element must have one attribute. The attribute defines the relation between a map element and its properties and between a property and the map elements it refers to. Each attribute structure has references in both directions, towards map elements and towards properties. If a map element needs more than one attribute a hierarchy is required: an attribute must be defined which has the required attributes as sub-attributes. An attribute should have at least two sub-attributes, or one property, or one of each.

A property structure contains the basic level of non-coordinate information associated with map elements. The property structure must be able to contain various types of information. For this reason the property structure contains a type field and a property information, which are stored as strings. The interpretation of these strings is the responsibility of the application program. The property structure also contains references to all attributes of which it is a property. The properties may create a basis for object oriented data processing. In this view, the property contains the classification function and appropriate parameters as class information that are to be used for processing relevant elements or objects, that inherit these class properties.

It is important to note that the GISDB structure covers both vector based as well as raster based, ie., grid cell objects. Thus, from a user point of view, overlays that are either vector or raster based can be mixed and combined freely, without any specific consideration by the user. This is particularly important for the easy integration of high-resolution maps and the output from spatially distributed environmental models calculating, for example, concentrations fields of pollutants.

Application examples

The hybrid GIS database structure and functions described above are implemented in a number of environmental information and simulation systems, ranging from local systems to regional, national, and global applications.

All systems integrate simulation models and expert systems components with the GIS system and standard data base elements. A detailed description of these systems has been published in Fedra and Reitsma, 1990; Fedra, 1990; Fedra, 1991 and Fedra and Diersch, 1989.

A typical example system is an environmental information system for the City of Hannover. It combines a GIS component with simulation models for specific problems such as groundwater or air pollution.

The GIS forms a central component, providing a variety of map types for use in the system. Maps or overlays include simple line features such as the city boundaries or complex topical maps as background for the spatially distributed models, including model input data sets. Examples would be a landuse map, the geological map, or a biotope map, stored in vector format (Figure 2), or groundwater head and groundwater recharge, stored as grid cell files. Similar to the model input files, model results, ie., computed groundwater heads or concentration fields of pollutants from air, ground- or surface water models, can also be stored as grid cell files.

Another raster format integrated in the GIS is a SPOT satellite image of the city area. The satellite image is stored and treated as a 'true' raster, ie., only color numbers and the attribute data are stored rather than the original multi-spectral data.

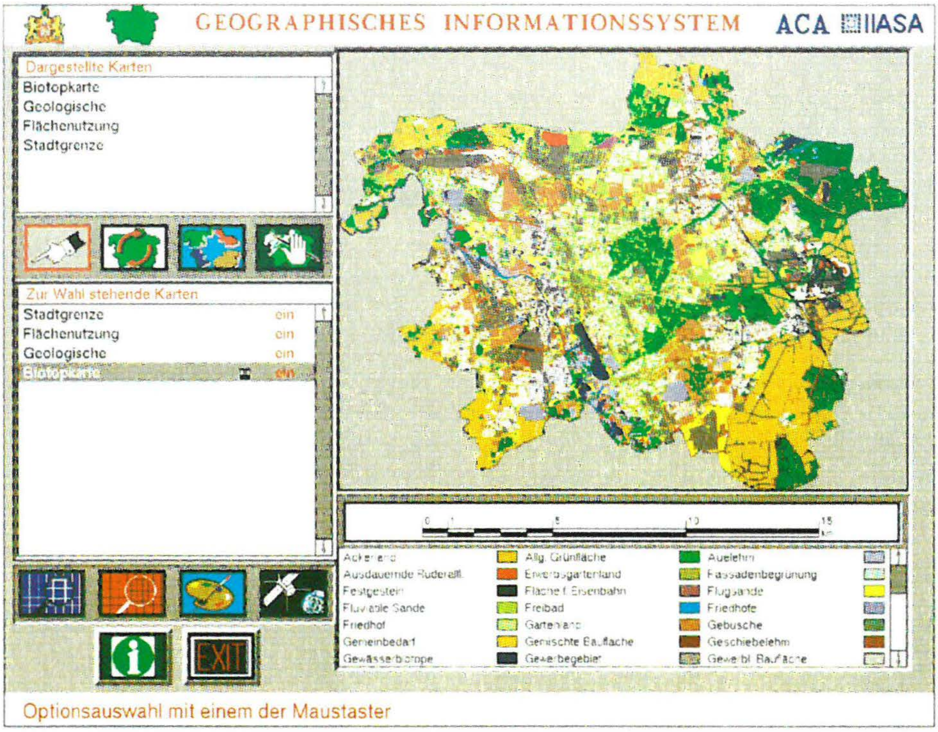

Figure 2: A vector-based topical map: the biotope map

While most of the maps fully cover the entire area, an interactive map editor allows to select individual features from a given map for an overlay (Figure 3). For example, from the full area coverage of the landuse map, only the road and rail network can be extracted as an overlay for the biotope map, eg., for transportation corridor analysis.

A color editor offers the possibility to adjust the display color and style of a given feature so that any arbitrary combination of features and overlays will result in a well designed display, highlighting the important features (Figure 4).

From a user point view, all these different maps are equivalent, the user is not necessarily aware of their structural differences. Composite maps can be generated by overlaying the various maps or subsets of features of the maps (Figure 5), and the GIS offers the possibility of zooming into any sub-area down to the limits of the resolution of the data base (Figure 6). Here, the differences between vector and raster formats become obvious.

Another special feature that is not normally supported in standard GIS is the display of time series data. Grid cell files or maps representing different points in time, for example output from dynamic simulation models or historic development stages, for example of population development, can be animated under interactive control. Examples are average monthly global climate data or the world population development, based on UN data and

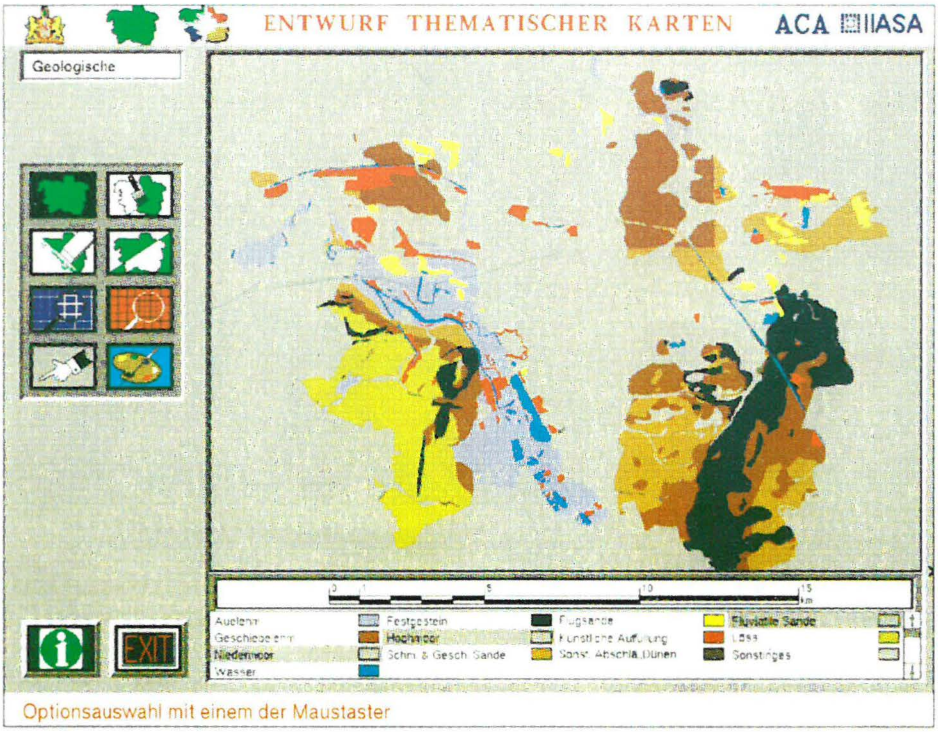

Figure 3: The map editor allows to select or deselect individual features of the map from its legend

projections covering 1950 to 2025, that are part of CLIMEX, a global GIS and expert system for climate impact analysis.

Here the display and any subsequent numerical analysis are based on a dual representation format. To speed up the display, a pre-processes raster image is loaded for the animation. The analysis however, is based on the underlying detailed numerical data with their full resolution.

Discussion

Both raster and vector formats have their respective advantages, such as precision at high resolution and low storage requirements for vectors, and ease of processing and display for rasters.

Integrating GIS functionality with simulation models and expert systems in a decision support environment requires a number of specific features that are best met by hybrid systems. Combining the advantages of raster and vector formats into one system allows the application to choose the most appropriate format for each task.

Figure 4: The color editor can modify individual feature display colors to design a visually attractive map for any set of features

For the user, the differences between vector and raster are largely transparent. It is the system that selects the appropriate format, and in fact, both formats may be used simultaneously, eg., using a vector version for display, but the corresponding grid cell file for numerical analysis, or displaying a pre-processed raster but using the underlying raw data for numerical data retrieval.

While hybrid systems introduce some redundancy, and may complicate the systems development and management task, we believe that they are easier to use and more efficient in their overall performance. Problem specific formats using multiple or hybrid representation formats, and a problem oriented design can be important elements in integrated information systems with embedded GIS functionality.

References and Selected Bibliography

Anthony, S.J. and Corr, D.G. (1987) Data Structures in an Integrated Geographical Information System. ESA Journal, Vol.11/1988, Vol.12, 69-72.

Armstrong, M.P. and Densham P.J. (1990) Database Organization Strategies for Spatial

Figure 5: A composite map showing elements of several topical maps on top of the SPOT satellite image

Decision Support Systems. Int. J. Geographical Information Systems, Vol.4, No.1. 3-20.

Burrough, P.A. (1986) Principles of Geographical Information Systems for Land Resources Assessment. Monographs on Soil and Resources Survey No. 12. Clarendon Press. Oxford. 193p.

Canters, F., Decleir, H. (1989) The World in Perspective. A Directory of World Map Projections. John Wiley. 181p.

Estes, J.E., Sailer, C., Tinney, L.R. (1986) Applications of Artificial Intelligence Techniques to Remote Sensing The Professional Geographer, May. 133-141

Fedra, K. (1991) A Computer-Based Approach to Environmental Impact Assessment International Institute for Applied Systems Analysis, RR-91-13.

Fedra, K. (1990) Interactive Environmental Software: Integration, Simulation and Visualization. International Institute for Applied Systems Analysis, RR-90-10.

Fedra, K., Reitsma, R.F. (1990) Decision Support and Geographical Information Systems. International Institute for Applied Systems Analysis, RR-90-9.

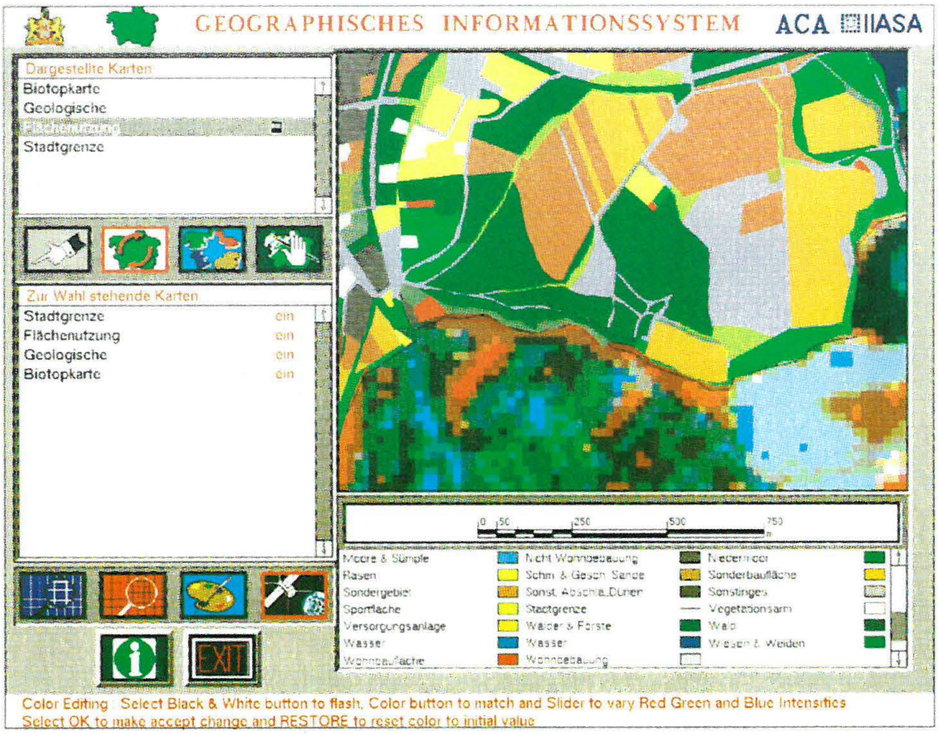

Figure 6: A detail from the above map

Fedra, K. and Diersch, H.J. (1989) Interactive Groundwater Modeling: Color Graphics, ICAD and AI. Proceedings of the International Symposium on Groundwater Management. Benidorm, Spain, IAHS Publication No. 188. pp. 305-320.

Gardels, K. (1988) GRASS in the X-Windows Environment. Distributing GIS data and technology. In: GIS/LIS '88, Accessing the World. Proceedings, Third annual international conference, San Antonio. Vol 2, 751-758.

GRASS 3.0 User's Manual (1988) U.S. Army Construction Engineering Research Laboratory, 1988.

Ibbs, T.J., Stevens, A. (1988) Quadtree Storage of Vector Data. Int. J. Geographical Information Systems, 1988, Vol.2, No. 1, 43-56.

McGregor, D. (1988) Geographic Information System Trends. In: GIS/LIS '88, Accessing the World. Proceedings, Third annual international conference, San Antonio. Vol 2, 911-914.

Parker, H.D. (1988) A Comparison of Current GIS Software: The Results of a Survey. In: GIS/LIS '88, Accessing the World. Proceedings, Third annual international conference, San Antonio. Vol 2, 911-914.

Peuquet, D.J. (1983) A Hybrid Structure for the Storage and Manipulation of Very Large Spatial Data Sets. Computer Vision, Graphics, and Image Processing 24, 1983, 14-27.

Piwowar, J.M., LeDrew, E.F. and Dudycha, D.J. (1990) Integration of Spatial Data in Vector and Raster Formats in a Geographic Information System Environment. Int. J. Geographical Information Systems, Vol.4, No.4., 429-444.

Sena, M.L. (1989) Seeing Double. Computer Graphics World, July 1989, 141-147.

Shapiro, M., Westervelt, J., Gerdes, D., Higgins, M., Larson, M (1989) GRASS 3.0 Programmer's Manual. U.S. Army Construction Engineering Research Laboratory, 1989.

Tilley S. and Sperry, S.L. (1988) Raster and Vector Integration. Computer Graphics World, August 1988, 73-76.

Visin, A.M., Carosella, M., Busillo, M., Campanelli, L. (1990) Hybrid Processing of Images for Territory Monitoring. Conf. proceedings 5th internat. conf. image analysis and processing. Progress in image analysis and processing. pp.637-639.

A GLOBAL CHANGE IMPACT ASSESSMENT SYSTEM: GIS, MODELS AND EXPERT SYSTEMS[1]

Kurt Fedra and Milan Kubat

Advanced Computer Applications (ACA)
International Institute for
Applied Systems Analysis (IIASA)
A-2361 Laxenburg, Austria

ABSTRACT

To support comparative analysis of global change scenarios, and provide a common data base and information system for impact assessment, IIASA's Advanced Computer Applications group is developing CLIMEX, a global GIS impact assessment and expert system.

The software system integrates numerous global data sets relevant to environmental and climate change, including IIASA's global climate data base, but also base data such as global elevation, soils, population, and landcover, some of them directly or indirectly derived from satellite imagery and remote sensors. It also includes model generated data sets such as GCM results or output from IIASA's global agricultural models.

The system can also integrate selected models, such as the BIOME vegetation and carbon budget model, or RIVM's IMAGE model, and in particular, its landuse component. For regional and local environmental assessment of generic problems, a number of simulation models for air, surface and groundwater, and coastal marine water quality are implemented, that all use embedded GIS integrating vector and raster data, including satellite imagery.

For the assessment of global change scenarios on both a global and a regional scale, CLIMEX integrates a rule based expert system that translates a given scenario into a set of indicators covering environmental as well as socio-economic aspects. The expert system in combination with the global GIS offers the possibility to use complex rules for classification and assessment. Both qualitative rules as well as algorithmic components in the form of various models and spatial statistics can be combined to develop a very flexible and powerful assessment system.

Introduction

Research into global change, considering climatic, environmental, and socio-economic aspects, requires and involves very large amounts of data and information, much of it in the form of geo-referenced data or topical maps. Data sets with global coverage, including the output from general circulation models and other models of various aspects of

[1]Paper presented at 25th International Symposium - Remote Sensing and Global Environmental Change. Tools for Sustainable Development. 4-8 April 1993. Graz. Austria.

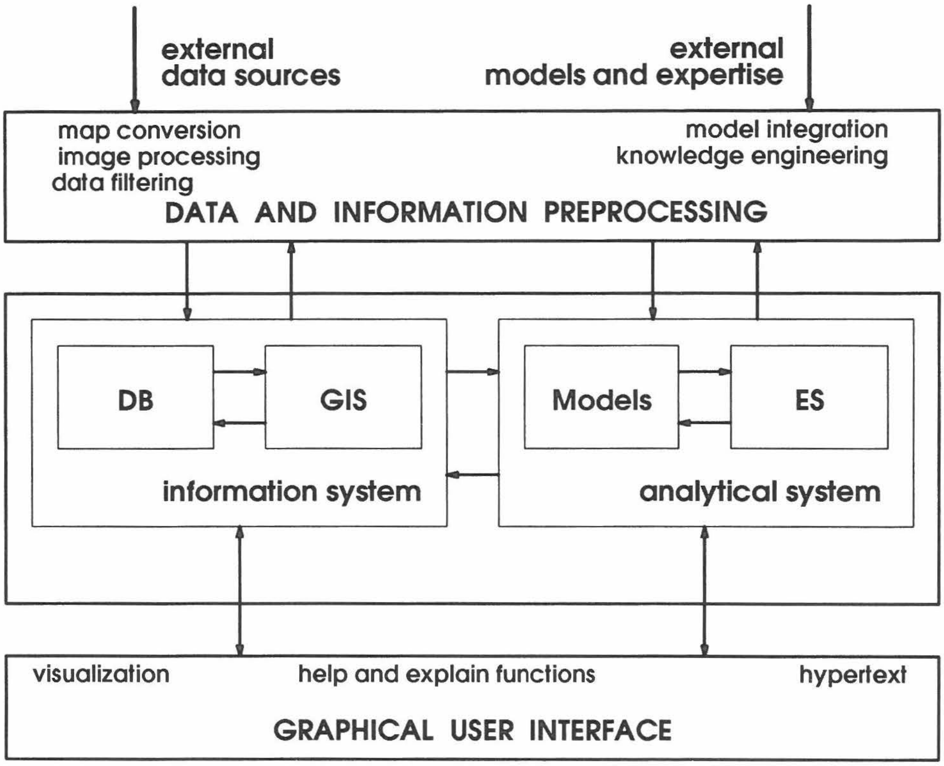

Figure 1: Basic architecture of CLIMEX

global change need to be analyzed, combined and compared. Hypotheses about likely consequences and impacts of global change need to be formulated and tested. The effort in data conversion and pre-processing, to get the numerous data sets into a format that lends itself to direct comparison, can be considerable. And the effort to derive new, useful information from the data, analyze, compare, and interpret them may involve a substantial part of basic data manipulation.

CLIMEX was designed to provide a consistent and easy-to-use framework to manipulate, analyze, and visualize global data sets. The system integrates global geographical or geo-referenced data with an interactive graphical user interface. This includes a hypertext help and explain system, as well as a rule-based expert system for data analysis and global change impact assessment (Figure 1).

The system also serves as a framework for global change models such as the IMAGE model (Rotmans et al., 1991), or BIOME, a global vegetation and carbon budget model

(Esser, 1991). For these models, CLIMEX provides linkages to its data base for the model input data required. It also provides the tools to interactively define model scenarios, using the expert system to check for consistency of assumptions and input data. The hypertext system, working as a user-support system together with the expert system, makes the models more accessible and easy to use. Finally, the visualization and further analysis of model results is supported both at the global GIS level as well as through model specific functionality, including animated model output or the presentation of specific graphs and diagrams in addition to the global maps generated.

CLIMEX provides a flexible framework for global change research. In addition to integrating a large amount of data and information in an open architecture that allows to add more data as they become available easily, and the possibility to integrate models of global change with its data bases and visualization tools, it also provides a set of tools for the design and testing of hypotheses on the environmental and socio-economic impacts of global change.

The Data Bases

In the core of the CLIMEX system is a global GIS and data base. The data collection, based on similar initiatives and data compilation efforts by several other institutions, includes a large set of themes. In addition to basic political and administrative boundaries, it covers physiographic data such as elevation or soils, vegetation, climate and hydrological data, and socio-economic information, eg., on population, agriculture, various resources, and energy. As a special kind of data, output from GCMs and similar global models are stored, defining scenarios of possible climate, vegetation, etc.

Several of the available data sets with global coverage are derived from satellite imagery. This includes, first of all, Monthly Generalized Global Vegetation Index from NOAA - 9 Polar Orbiting Satellite and surface temperature data set from NASA satellite NIMBUS. Another data set derived from satellite outputs and integrated in CLIMEX contains chlorophyl concentration in water. This data set is based on results of Nimbus - 7 Coastal Zone Color Scanner. Besides data sets that have been directly derived from satellite sources, as is the case with some other data sets used in CLIMEX eg., Vegetation and Cultivation Intensity, authors applied the satellite imagery for comparison and qualitative analysis in order to determine the reliability of the data gathered.

In general, data sets with a global coverage can have one of the following structures:

- Country specific data; this includes many of the basic statistical data compiled on country level, eg., by various UN organizations or the World Bank. A typical example of this data type implemented in CLIMEX are population data.
  In principle, data could also be available on a sub-national level or for regions composed of several countries, eg., for administrative entities such as provinces, or for physiographical entities such as river basins.

- Point data; these come from measurement stations, eg., for climate variables or river flow. In most cases, these data are time series of observations over longer periods of time.
  Examples in CLIMEX include IIASA's Climate data base or selected river flow data from the River Flows Data Base of the Global Runoff Data Centre established by the Federal Republic of Germany in Koblenz.
- Gridded data; these can be model generated, as in the case of GCMs, result from the spatial interpolation of point data, or satellite derived as mentioned above.
  As a special case to take advantage of their granularity in various projections, GCM data are stored and processed in quad-tree format.
  In CLIMEX this includes GCM results, some of the satellite derived data, raster versions of vector maps such as the FAO/UNESCO soil maps, and spatially interpolated climate data from IIASA's global climate data base.

Currently, the system includes data on more than 50 themes or coverages, where several of those are available as time series, eg., GCM results, and the population data, which are available on a monthly basis and for a number of years respectively.

The second large block of information in the CLIMEX system is the knowledge base for impact assessment and the construction of derived indicators. This includes the definition of variables (in CLIMEX called descriptors), rules, explanatory hypertext files, and a number of simple numerical models, eg., to estimate irrigation water requirements, linked to the rules as part of a logical inference system.

GIS: Mapping and Visualization

The data sets in CLIMEX are accessible through a graphical user interface which offers a hierarchical selection of topics, and displays the corresponding data set in the form of a topical map (Figure 2). For better orientation, political boundaries can be displayed as an auxiliary overlay for all maps.

The basic mapping and display system uses spatial information prepared or preprocessed by other GIS such as Arc/Info or GRASS. Some of the basic features of this dedicated and embedded GIS functionality are described in Fedra and Kubat, 1992.

The system handles both vector and raster data simultaneously. Arbitrary zooming is supported, and a special color editor allows the user to modify the colors associated with legend entries (Figure 3). For performance reasons, large data sets may be tiled, or be available at different resolutions that are switched according to the zooming factor. A map editor supports the selection of subsets of features from any map (eg., a single soil type or group of soils from the global soil map, or an elevation band, vegetation class, climate zone, etc).

In the case of a time series of maps, a special tool with a tape deck interface is offered, that gives the user control over the animation sequence. For the specific case of

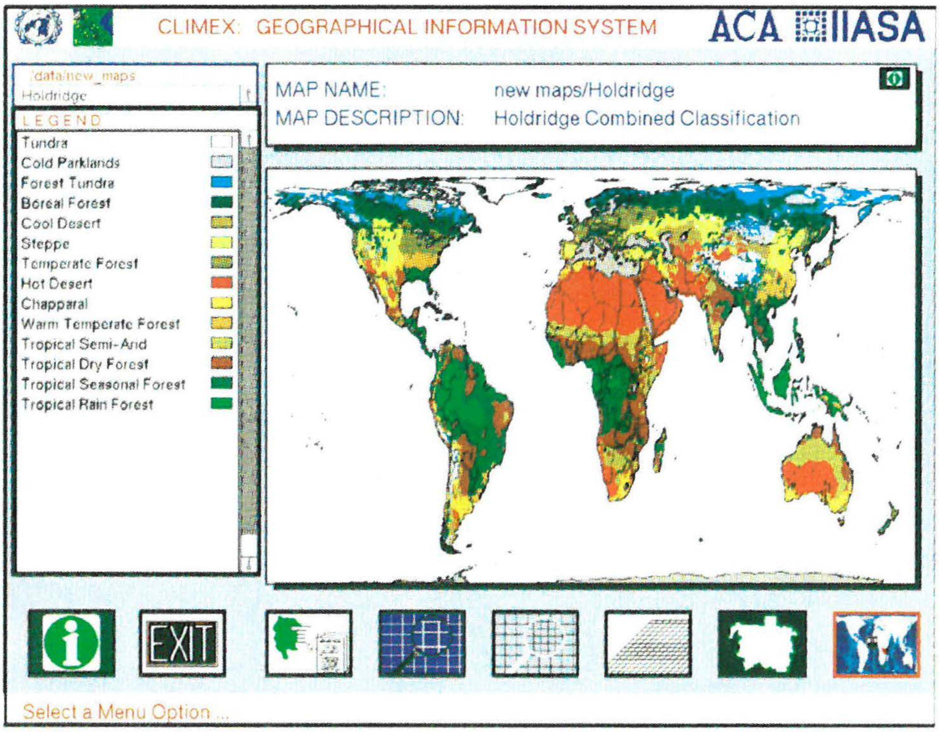

Figure 2: Graphical user interface of CLIMEX

GCM results and the corresponding data (temperature, precipitation, cloudiness) from the climate data base, a direct comparison of model output and the reference climate defined by the observation data is made possible by the simultaneous display of four map windows (Figure 4). For point data, the stations are displayed; picking a station will display the corresponding data (Figure 5). Where more data sets are available than can usefully be displayed simultaneously, the user can interactively select and define the graphs he wants to display.

The Expert System: spatial reasoning

The expert system in CLIMEX serves a number of purposes. Embedded in the interface of the models integrated, it assists the user in defining consistent model scenarios. Its major

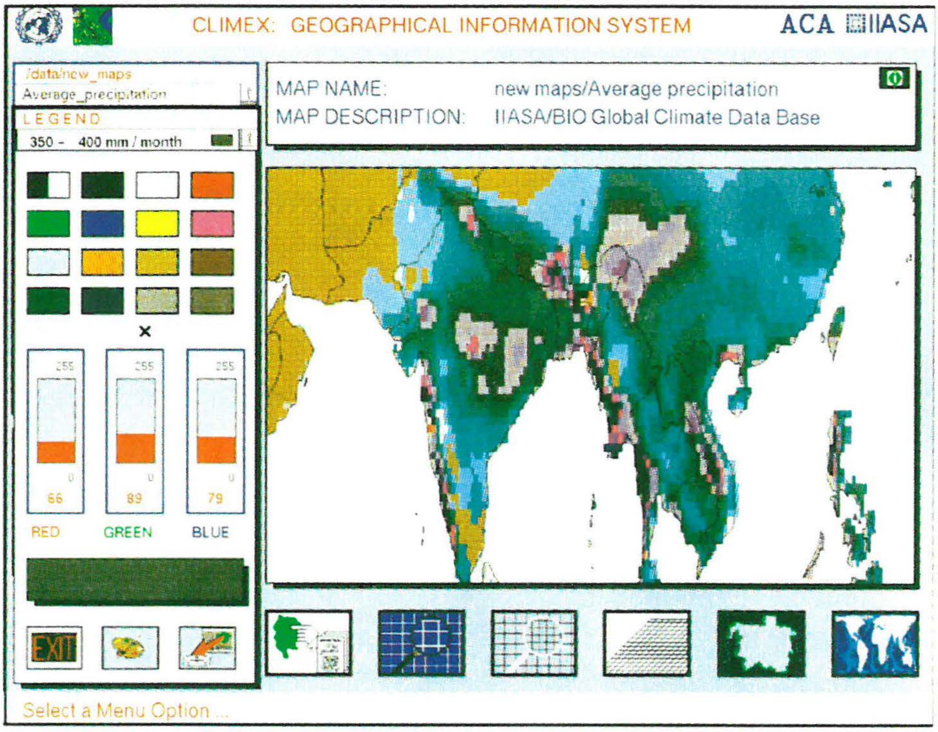

Figure 3: CLIMEX supports arbitrary zooming and a color editor allows the user to modify the color associated with legend entries

role, however, is to derive new, user defined indicators from the available data.

A typical example would be an indication of potential water resources problems: we could formulate a set of rules that include population growth, the current ratio of precipitation and potential evapotranspiration, expected changes according to climate change scenarios, dependency of the economy on irrigated agriculture, etc. The resulting set of rules would then be evaluated for a number of spatial objects, depending not only on the degree of spatial resolution required but also on the nature and resolution of the information used. The result would again be a global map of this new indicator, possibly for alternative scenarios of population growth, climate change, economic development, etc.

In more theoretical terms, the expert system can reason with indicators that may be spatially distributed, that is, describing spatial objects. These could be points (like measurement stations or even cities on a global scale), polygons (most typically countries,

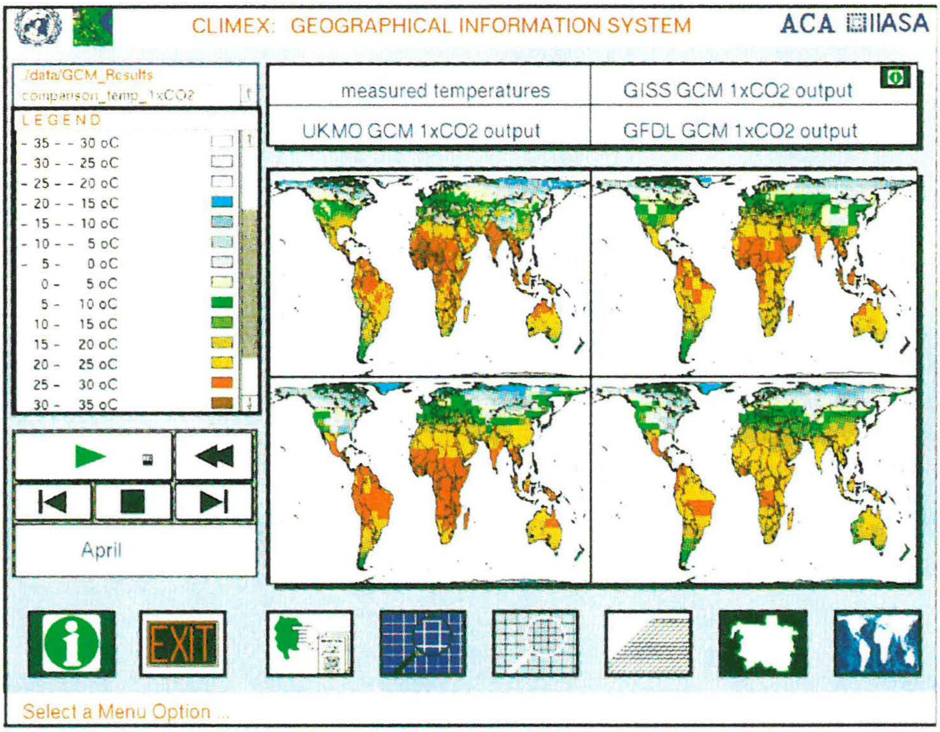

Figure 4: Tape deck interface gives the user control over the animation sequence and simultaneous display of four map windows makes direct comparison possible

but also supersets and subsets, ie., regions and provinces, or river basins, etc.), lines or corridors (eg., an area along a river or a coastline) and grid cells.

The reasoning may involve any or all indicators that apply to this spatial object which may require sampling or interpolation and aggregation. The inference may also refer to other spatial objects, for example bordering or in the "neighborhood" of the primary object. And since some of the descriptors are dynamic, a historical dimension can be included as well.

A specific form of using the expert system in CLIMEX is for the assessment of climate change scenarios in terms of regional impacts. The user first selects a region, such as a major river basin, and then selects or defines a climate scenario. This can be based on a GCM scenario or be created by the user, specifying, for example, a relative increase in temperature and precipitation, or a rise in sea level, etc.

The assessment then uses a system of checklists that guides the user to a set of prob-

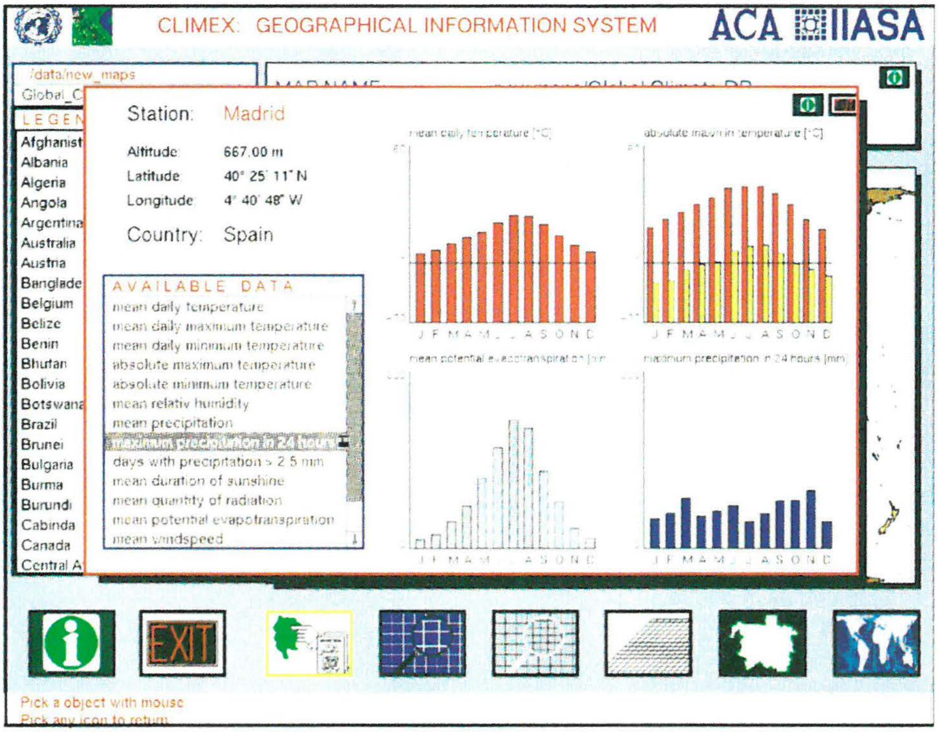

Figure 5: Simultaneous display of four point data sets

lems (derived indicators) for assessment. Problems are grouped into major themes, such as natural environment, water resources, agriculture and food, industry and the energy sector, or population, health and socio-economic impacts. Each problem involves a number of questions that refine the basic scenario assumptions or provide additional region-specific data.

The assessment itself is based on a combination of simple models and rules. The models are triggered to provide individual estimates for key variables; for example, irrigation water demand based on a set of FAO estimation methods. Using a distributed parallel computing scheme, they are run as an integrated part of the expert system inference mechanism.

CLIMEX uses a straight-forward syntax, combining an object oriented design for the descriptors, the basic elements in the inference procedure, with near natural language rules. For a detailed description of the expert system component, its architecture and an application for environmental impact assessment, see Fedra et al. (1991).

As the basic object of the system, descriptors are the concepts or terms used in the knowledge representation; they are linked through the rules that allow to derive values for a given descriptor from other descriptor values through the standard arithmetical and logical operations.

Descriptors are defined as part of the knowledge base of the system. The definition includes basic characteristics such as name, type, and units of measurement, and a list or range of legal values the descriptor can take. Depending on the descriptor, these values can be symbolic, numeric, or both. Descriptors also know about methods they can use to establish their values in a given context. These methods include questions to ask of the user, data base or GIS references, rules, and references to complex numerical functions and simulation models that can be used to obtain an appropriate value. In the interactive dialogue, the user can choose between different methods; priorities of methods are also defined in the descriptor definitions and can be dynamically modified through rules. Finally, descriptors can have alternative sets of definitions, to be used depending on the context and under rule control.

```
DESCRIPTOR
reservoir_lifetime                      # descriptor name
T S                                     # descriptor type
U years                                 # unit of measurement
V short      [  0, 25] /                # legal values
V medium     [ 25, 50] /
V long       [ 50,150] /
MODEL                                   # model interface
resed                                   # model name
T local_wait                            # interface type
I project_stream_location /             # required input descriptors
I average_depth /
O reservoir_lifetime /                  # model output
O retention_time /
ENDMODEL
R 11405 / 11406 / 11407 /               # alternative method:
R 11408 / 11409 / 11410 /               # rules
Q What is the expected lifetime,        # alternative method:
Q in years, of the reservoir?           # ask the user
ENDDESCRIPTOR
```

An example of the simple rule syntax that is used in conjunction with the hybrid numerical and symbolic descriptors introduced above would be:

```
RULE 1010582 #USLE final round: erosion_potential
IF     [  rain_factor          == low
```

```
      OR rain_factor         == very_low ]
AND   soil_slope_factor  == high
THEN  erosion_potential      =  medium
ENDRULE

RULE 1040516
IF           eutrophication_potential == very_large
THEN         floating_weeds INCREASES marginally
ENDRULE
```

Rules can result in the absolute assignment of descriptor values, their relative, incremental modification, or they can be used to control the inference strategy depending on context.

Also, the user can call up a knowledge base browser, that allows to navigate in the tree structure of the knowledge base within the context of individual problems. The browser can descend the inference tree, displaying sets of rules referring to a list of descriptors and allow to inspect individual descriptor definitions.

The possibility to integrate models in place of rules in an expert system and at the same time use embedded rule-based components in models provides a very rich repertoire of building blocks for interactive software systems.

The flexibility to use, alternatively or conjunctively, both qualitative symbolic and quantitative numerical methods in one and the same application allows the system to be responsive to the information at hand, and the users requirements and constraints. This combination of methods of analysis, and the integration of data bases, geographical information systems, and hypertext, allows to efficiently exploit whatever information, data and expertise is available in a given problem situation.

Model Integration

Simulation models of various aspects of global change are another major element of the CLIMEX system. Here the system provides a framework for the integration of models. This integration can be done at different levels.

In the easiest case, data bases and GIS capabilities are used to process the output from models. This is done, for example, for selected GCMs which provide medium-resolution gridded results, or for IIASA's global agricultural model system, that produces output on a national or regional level.

A tighter level of coupling directly integrates models into the overall software system. An example is the IMAGE model (Rotmans, 1991), where pre- and postprocessor functions for the model are included (Figure 6). Here both model input and model output are included in the data bases and GIS, and specific editing functions for the definition of model scenarios are available. These also use the expert system to check the consistency of user defined scenario assumptions. Another example of an integrated model is

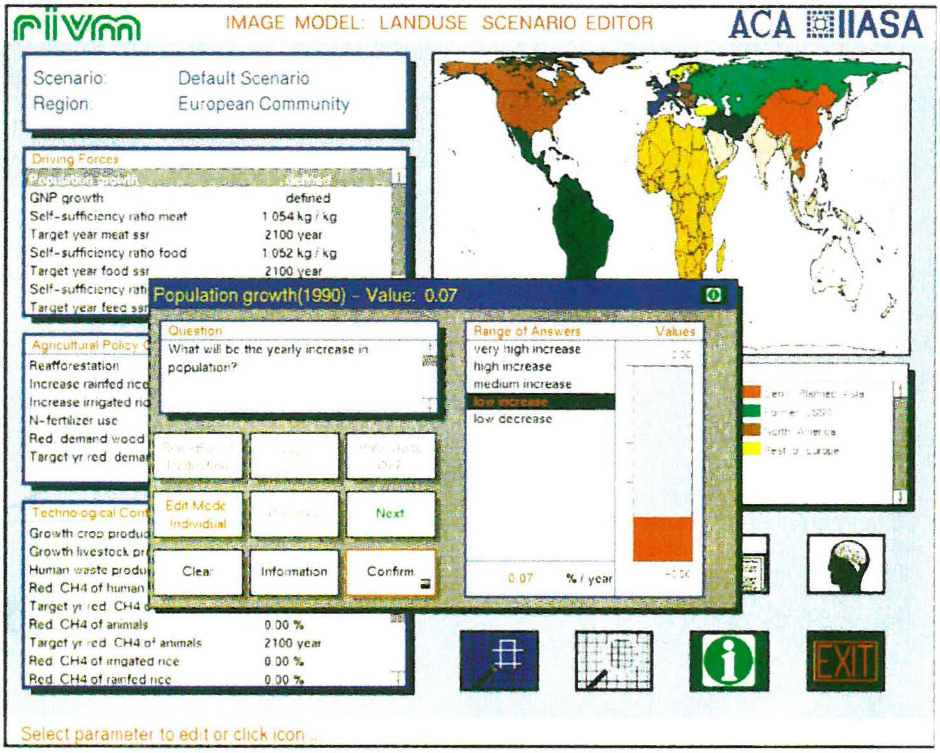

Figure 6: Model IMAGE directly integrated into CLIMEX framework

OBM/CEBM (Esser, 1991). The OBM describes and quantifies the pathways of carbon through the terrestrial biosphere. Carbon transfer rates and transfer coefficients, respectively, are calculated from environmental variables such as mean annual temperature, average annual precipitation, CO_2 - concentration of the atmosphere, soil fertility, related vegetation formation, and land use category.

Finally, the expert system itself can incorporate numerical simulation models as mentioned above.

Discussion

One major problem for anyone trying to understand global change is the sheer volume of complex and often contradictory data on the subject. The Climate Impact Assessment

Expert System (CLIMEX), created by IIASA's Advanced Computer Applications Project, is designed to help users make better sense of this information. Users of CLIMEX can sort through and compare vast amounts of climate-related data, create their own climate change scenarios, and get some idea of the impact on a given region.

CLIMEX is both a repository of often disparate information related to climate change and a formal structure for comparative assessment of various climate scenarios. It combines a global GIS and an extensive climate data base with a rule-based expert system designed for impact assessment on a regional scale.

The GIS component of CLIMEX draws together information from many different sources that is global in scale and relevant to climate change study. The data sets of the global GIS, while of interest in themselves, are designed to provide input to CLIMEX's analysis component. The analysis itself is based on a combination of simple models and a rule-based expert system. The expert system's formalism provides rigor and flexibility at the same time. Each and any conclusion must be based on a set of well-defined precursors and their well-defined interdependence. At the same time, the near-natural language and sufficient set of operators provided by the rule syntax offers considerable expressive power. All the system's workings are open for critical inspection. Rules can be easily understood and modified, and the menu-driven tutorial style of dialogue with the system makes it very user friendly.

The same idea of ease of use, direct accessibility, and ease of understanding is behind the visualization in the GIS component. Together they build a powerful set of tools that can help to make better sense of complex data in the area of climate change research.

References

Rotmans, J., Kreileman E., Den Elzen, M. (1991) IMAGE: An Integrated Model to Assess the Greenhouse Effect National Institute of Public Health and Environmental Protection Bilthoven, Netherlands

Esser, G. (1991) Osnabrück Biosphere Model: structure, construction, results. In: Esser, G. and Overdieck, D. [eds.], Modern Ecology: Basic and Applied Aspects. Elsevier. 679-709

Fedra, K., Winkelbauer, L. and Pantulu, V.R. (1991) Expert Systems for Environmental Screening. RR-91-19, 169 pp., International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria.

Fedra, K. and Kubat, M. (1992) Hybrid Geographical Information Systems. Proceedings of EARSeL Workshop on Relationship of Remote Sensing and Geographical Information Systems. September 16-18, 1991. Hannover. Germany. EARSeL Advances in Remote Sensing, Vol.1, No.3-VII.