

Hybrid Geographical Information Systems

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 feature 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, i.e., information describing the properties of geometrical objects such as points, lines, and areas [Armstrong and Densham, 1990].

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.] 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, i.e., 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; Piowar 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 dimensions 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 fea-

tures 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 Declair, 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.

1. GRASS: A Hybrid GIS

This sections 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 (Fig. 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, i.e., a raster format.

2. 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 reclassified. The un-

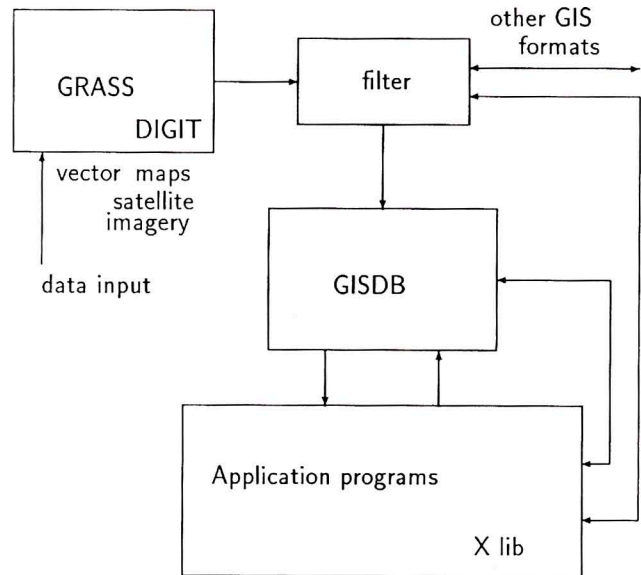


Fig. 1 - Using GRASS for preparing geographical data eventually used in some other system.

compressed 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 use 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 its cell header file.

In addition to the cell data file itself, there is a number of support files or 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 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 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.

3. 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 losing 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.

4. 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 consists in the fact that the system supports only processing information within one of 60 UTM grid zones (between 84° 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.

5. ATTRIBUTE AND COLORS

The basic information a grid cell holds are the data value or the 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 it difficult to distinguish them on the screen.

Should it be necessary to have more categories than the number of available colors support 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 from a given data set, i.e., 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 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 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.

6. 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, 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, coincidence 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 informa-

tions. 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, i.e., 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 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.

7. 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 functionally 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 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 covering 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, i.e., 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 an attitude. The attribute defines the relations between a map element and its properties and between a property and the map elements it refers 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.

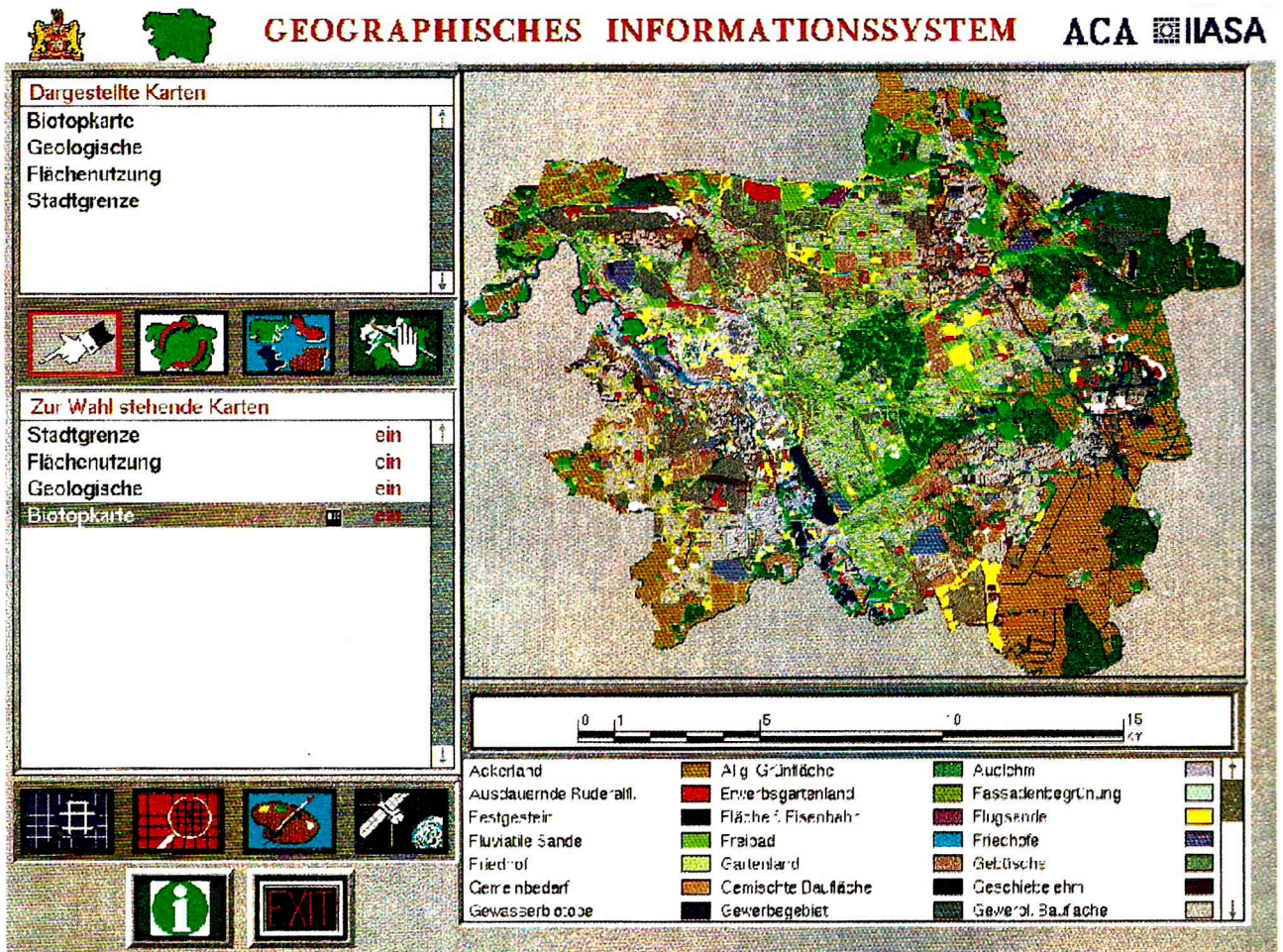
It is important to note that the GISDB structure covers both vector based as well as raster based, i.e., grid cell objects. Thus, from a user point of view, overlays that are either vector or raster based can be mixed and combined freely,

without any specific considerations 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.

8. 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, 1991 and Fedra and Diersch, 1989.



Optionsauswahl mit einem der Maustaster

Fig. 2 - A vector-based topical map: the biotope map.

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.

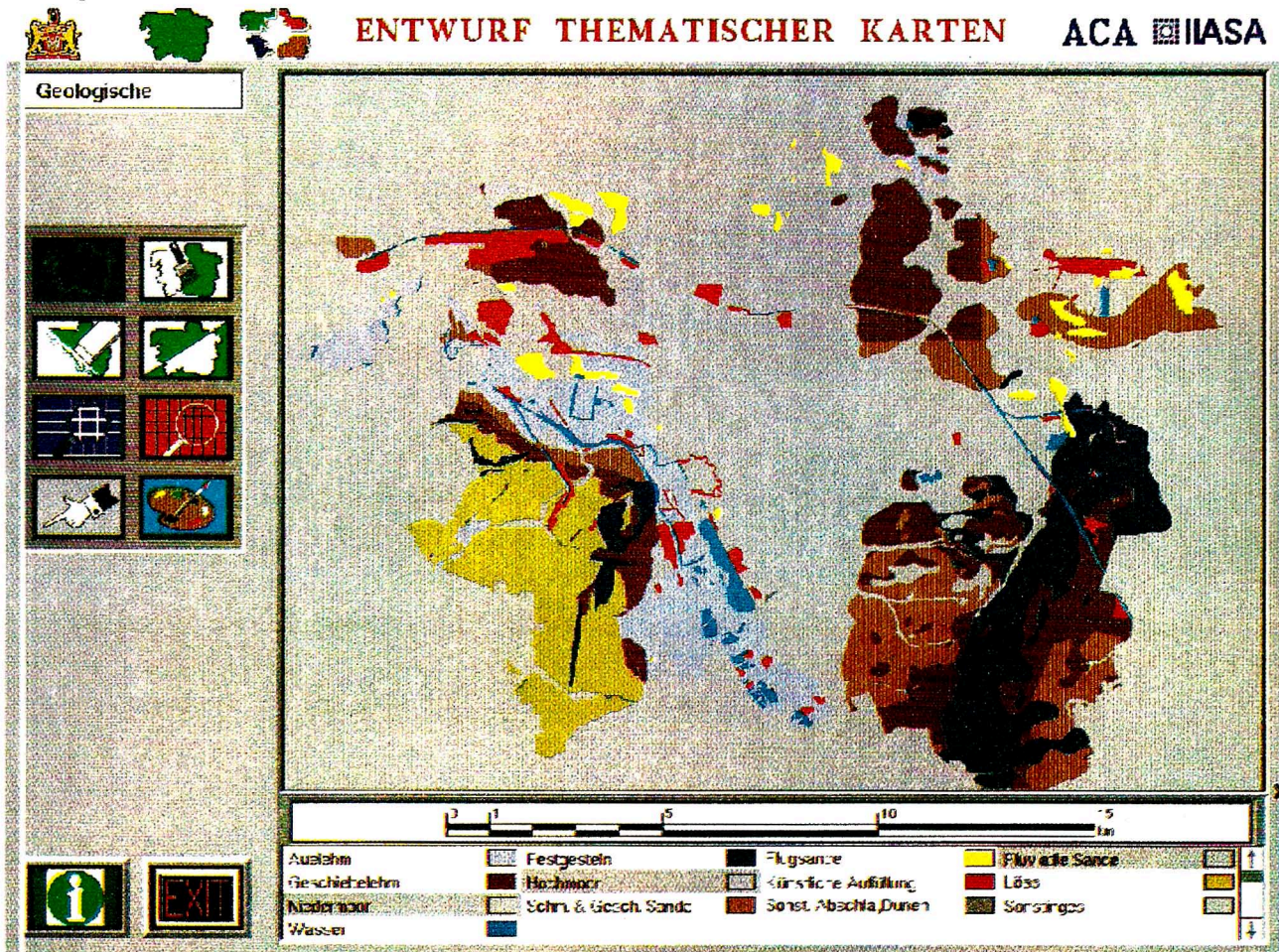
The GIS forms a central component, providing a variety of map types for use in the system. Maps or overlays include simple line features such as the city boundaries or complex topical maps as background for the spatially distributed models, including model input data sets. Examples would be a landuse map, the geological map, or a biotope map, stored in vector format (Fig. 2), or groundwater recharge, stored as grid cell files. Similar to the model input files, model results, i.e., 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, i.e., only color numbers and the attribute data are stored rather than the original multi-spectral data.

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 (Fig. 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, e.g., 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 (Fig. 4).

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

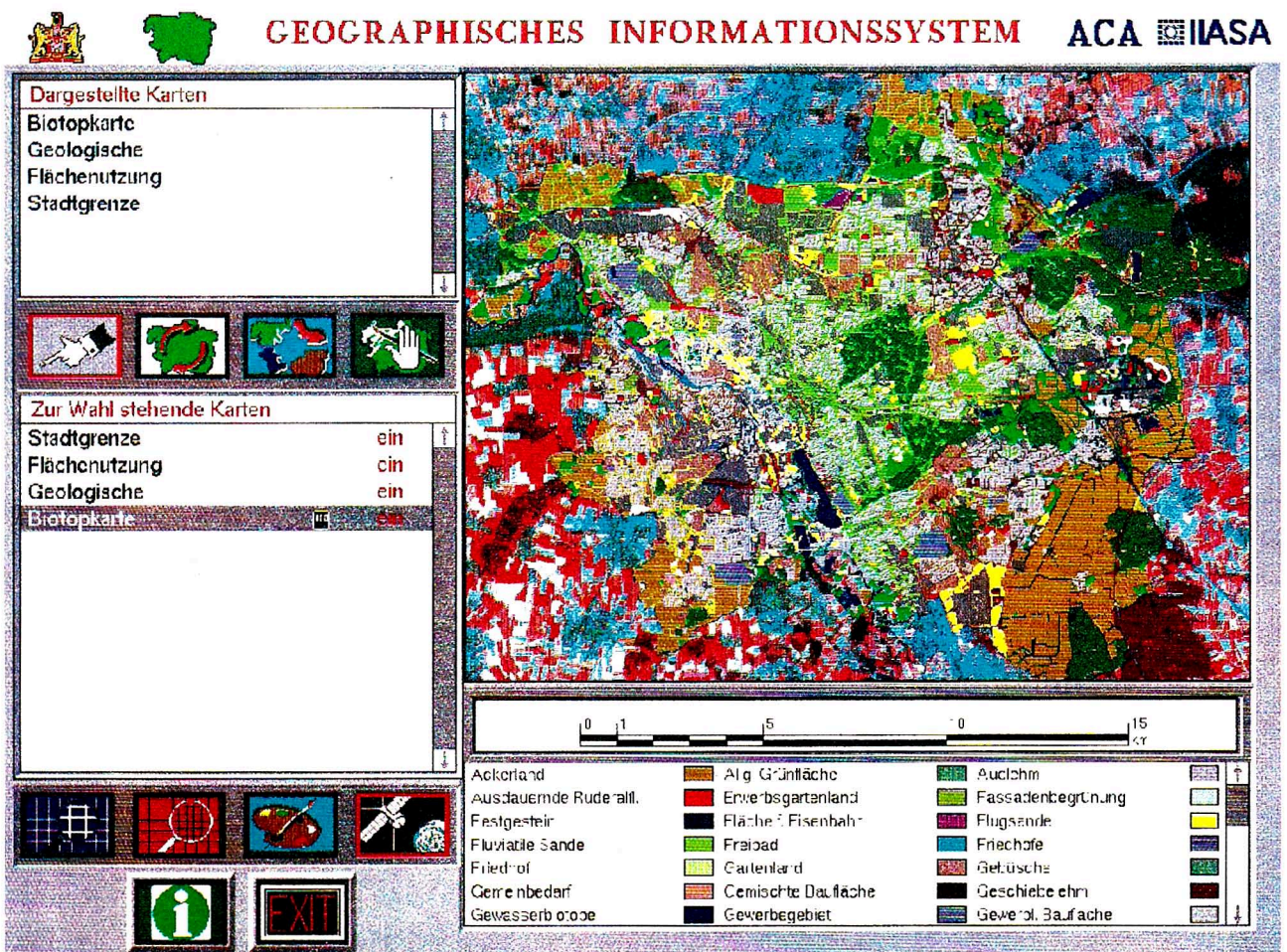


Optionsauswahl mit einem der Maustaster

Fig. 3 - The map editor allows to select or deselect individual features of the map from its legend.

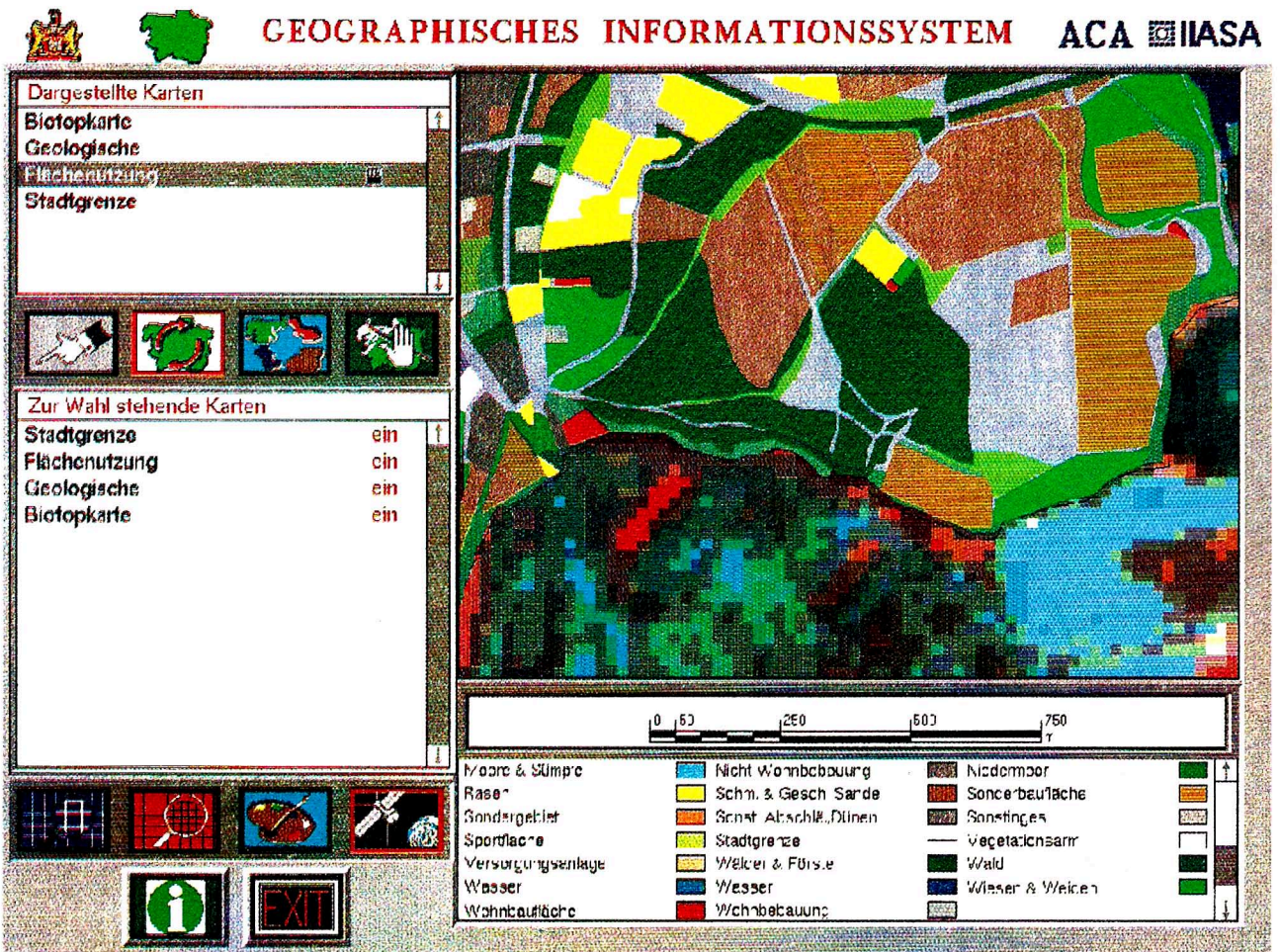


Fig. 4 - The color editor can modify individual feature display colors to design a visually attractive map for any set of features.



Optionsauswahl mit einem der Maustaster

Fig. 5 - A composite map showing elements of several topical maps on top of the SPOT satellite image.



Color Editing : Select Black & White button to flash, Color button to match and Slider to vary Red Green and Blue Intensities
 Select OK to make accept change and RESTORE to reset color to initial value

Fig. 6 - A detail from the above map.

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 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-processed raster image is loaded for the animation. The analysis however, is based on the underlying detailed numerical data with their full resolution.

9. 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 functionally 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 applications to choose the most appropriate format for each task.

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, e.g., 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

- Anthony S.J. and Corr D.G., 1987. "Data Structures in an Integrated Geographical Information Systems". *ESA Journal*, Vol.11/1988, Vol. 12 69-72.
- Armstrong M.P. and Densham P.J., 1990. "Database Organization Strategies for Spatial 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., Decler 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.
- 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.
- Pequet 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., 1989. "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 Territory Monitoring". Conf. proceedings 5th Intern. Conf. Image Analysis and Processing. Progress in image analysis and processing. pp.637-639.