Appears in Mason, C. (Ed.) The First International Workshop on AI and the Environment, International Joint Conference on Artificial Intelligence, Montreal, Canada, 1995.

KnowledgeBased Land Information Manager and Simulator (KBLIMS) For Forested Ecosystem Simulation Management

Vincent B. Robinson and D. Scott Mackay
vbr@geog.utoronto.ca scott@geog.utoronto.ca

Department of Geography
University of Toronto
Toronto, Ontario M5S 1A1
Canada

Abstract. Applications such as forest ecosystem modeling demand management of geographically-based information detailing complex interactions between climatic, topographic, hydrologic, pedological and ecological processes. The Knowledgebased Land Information Manager and Simulator (KBLIMS) is based on the notion of a query model which executes a set of user-defined queries. Each query is deductive so a user may define a simulation experiment by first identifying a set of objects as a spatial query, then specifying some action to be performed on these objects, such as a combined simulation query and aggregation query. The KBLIMS approach has led to a parsimonious coupling of geographic information system(GIS) and ecosystem simulation modeling. The user, *e.g.* an ecologist, need not explicitly parameterize and run simulation models. Typical use of the simulation system is managed by the knowledgebase using its metaknowledge which allows for the integration of either tightly-coupled or loosely-coupled systems.

Introduction

Within the forest ecosystem modeling community there is much attention devoted to developing models for simulating carbon, water and energy flux processes at regional scales. One such effort is the Regional HydroEcological Simulation System (RHESSys) (Band *et al* 1993. RHESSys provides ecological modeling capability at the scale of watersheds. Because of the geographic elements around which RHESSys is organized the use of geographic information systems (GIS) to support RHESSys simulation has been suggested (*e.g.* see Michener *et al*, 1994). However, GIS and simulation models, such as RHESSys organize their information in fundamentally different ways. Objects in a GIS are typically in the form of primitives such as points, lines, polygons, or layers. Objects in a simulation model are defined in terms of system state, mass and energy fluxes, and interaction and dynamics of species or individuals. Management of these two different domains usually means that large, cumbersome files are developed and modeling is conducted in a rigid, almost batch-oriented approach. There is also usually little flexibility in what kinds of geographic queries can be asked the system(s). To bridge the conceptual gap between GIS objects and simulation objects we have designed a hybrid knowledgebase management system that transforms simple observations and relations in collected data into higher order concepts.

This paper describes the KnowledgeBased Land Information Manager and Simulator (KBLIMS), a knowledgebased geographic information system (GIS) approach to organizing both data and simulation models on the basis of distinct, identifiable landscape units such as watersheds. KBLIMS is based on the notion of a query model which executes a set of user-defined or system-defined queries. Each query is deductive so a user may define a simulation experiment by first identifying a set of objects as a spatial query, then specifying some action to be performed on these objects, such as a combined simulation query and aggregation query.

Simulation modelling of ecosystem processes, or information management of the physical characteristics of the landscape, over large areas, is hampered by the extreme variability of the land cover, topography and soils. As an example, the processes of evapotranspiration and photosynthetic productivity of a forest canopy are dependent on surface variables including soil properties, elevation, aspect, solar radiation, humidity, wind speed, and standing vegetation characteristics (*e.g.* species, leaf area index, biomass). In mountainous watersheds, each of these variables show strong dependencies on landscape position such that different factors limit growth on ridges as opposed to valleys, or on north facing as opposed to south facing slopes. Depending on the community structure, soil properties or geomorphic context of a forest stand, it may be possible to ignore certain processes or emphasize others to simulate forest activity. Hence, general models that can accurately simulate forest processes over the full range of conditions encountered over a region are often not tractable. Choosing and structuring the proper simulation model for each of a potentially large number of landscape units is a tedious, and error-prone task. Instead, our approach is based on encoding the basic ecosystem knowledge required to infer the state of the landscape unit, to decide and choose the proper strategy for simulation.

System Design Methodology

Knowledgebased Management. Terrain analysis and simulation modeling tasks require specialized tools which are generally separate from the information management facilities provided by a GIS. We assume that information systems are knowledge-intensive, as are the specialized tools needed to perform tasks such as terrain analysis and ecosystem simulation. Therefore, system components are viewed as types of knowledge which can be incorporated into a hybrid knowledgebased system. The knowledgebase is organized around a query model manager which distinguishes between three types of knowledge: (1) extensional and intensional predicates, (2) structural descriptions, and (3) procedures. (Mackay *et al*, 1994).

Terrain analysis and simulation modeling tools are viewed as procedural knowledge, as they embody specialized knowledge from their respective domains the form of programs. Query models build concrete objects by combining predicates, procedures and abstract structural descriptions. Concrete objects are instantiated counterparts of their respective class descriptions. The transformation from object class descriptions to instantiated objects is accomplished using definitions of ***Class description***, ***isA relation***, ***Query***, *Query model*, and ***Object structure*** discussed in Mackay *et al* (1994).

An object structure reflects the description given in the class schema. The target object structure is instantiated by: (1) retrieving all attribute descriptions from the isA graph, such as the attributes given in the topographic_object and hillslope class descriptions; (2) resolving any attribute cancellations such as replacing the attribute, (*e.g.*, elevation, feet), with the attribute; (3) generating a query model from retrieved descriptions; and (4) populating the object structure with the result.

It is often important to ecologists to retrieve aggregate totals for queries over a number of objects of the same class. To perform aggregation in queries we distinguish between atomic identifiers, and complex identifiers. Each instantiated object of a given class has a unique identifer. Complex identifiers, *e.g.* [1,2], are made up of unique atomic identifiers from some class, and denote aggregates of objects from that class. Results of queries on objects with complex identifiers are obtained by applying the query over each of the objects represented by the atomic identifiers within the complex.

Methods provide access to the database, but hide procedural aspects of queries. System defined methods include models for asserting attributes, retracting attributes, retrieving attributes, and checking constraints on objects. Each method has an external interface which provides a declarative semantics for queries. The system takes a description for an object attribute and automatically generates an implementation query. This strategy forms a distinction between an implementation layer and class description layer. Layering permits changes in low-level procedures without changing the semantics of descriptions defined at a higher level.

Deductive rules are specialized kinds of methods which infer new facts. Deduction refers to the fact that inferences made during a previous transaction are available as explicit facts for subsequent transactions. Furthermore, a user may select, or browse, a set of objects without knowing specifically what kinds of questions to ask about these objects. The system retains retrieved objects and assists in asking more specific questions.

Terrain Analysis. Information used for discovering terrain objects is image and map based. The terrain analysis system translates layer-based geographic information into object-based geographic information. The layer-based model of geographic information represents spatially distributed attributes as a set of data layers each of which defines a distribution of a single attribute over a defined space. The object-based model combines generic spatial objects (points, lines, polygons), explicit spatial relations between objects (lines with left and right polygons) and attributes (*e.g.* area) into a single, unified data model.

Transformation from image data to symbolic elements is a three-step process: (1) extraction of drainage basins and their segmentation into hillslopes and stream links, (2) analysis of the hillslope and stream links, and (3) generation of a objects.Terrain partitioning uses techniques described by Band (1989). This approach can adaptively scale terrain partitioning from a few, large hillslopes to many, small hillslopes for any given drainage basin, by adjusting the constraints on pruning the drainage area transform. It is the scale flexibility that allows layer-to-object transformation without significant loss of land surface heterogeneity meaningful to ecosystem simulation.

In KBLIMS analysis of hillslopes and streams follows Lammers and Band (1990), with extensions to incorporate soils, remotely sensed imagery, and detailed topology (Band and Robinson 1992). Analysis includes geometric computations such as gradient, aspect, and junction angles, topological computations such as stream link connections, stream link to hillslope connections, and hillslope to hillslope connections, or locations of divides, and computation of hillslope distributional information for soil hydraulic properties and leaf area index. This information is incorporated into the knowledgebase (Mackay *et al,* 1994) .

Simulation System. The ecosystem simulation system is based on the Productivity and Hydrology Simulator (PHS) (Band and Robinson, 1992). PHS is derived from the model components of the Regional HydroEcological Simulation System (RHESSys) (Band *et al* , 1993) which includes MTCLIM, FOREST-BGC and TOPMODEL. It is designed to simulate the distribution of forest carbon and water flux, and storage processes over a mountainous landscape. MTCLIM extrapolates daily meteorological observations from base stations using adiabatic lapse rates, orographic gradients, and topographic correction of solar radiation. FOREST-BGC is a stand level model of forest carbon, water and nutrient budgets. It computes daily carbon photosynthesis, respiration, evaporation, snowmelt and runoff production, using site conditions such as current stand characteristics, observed or computed daily meteorological data, and soil properties. In addition, daily and seasonal hydrologic flux process such as snowmelt, interception, runoff and evapotranspiration are computed.

Integrating the knowledgebase and simulation system. Metaknowledge is used to integrate the simulation system and the knowledgebase. Metaknowledge provide a means of describing how the knowledgebase interfaces with the simulation system, including data structuring, procedural calls, and retrieval of results. Current versions of the simulation system are designed to operate both as a stand-alone systems, as well as an integral part of the knowledgebase. System-defined metaknowledge provide a seamless view of integration of system components to the end-user, hence it is not visible at the user interface level. The simulation system operates on simulation objects which are specialized kinds of objects, defined within the knowledgebase to integrate information from topography, remote sensing and soils domains.

Implementation and Testing

The KBLIMS knowledgebase is implemented in Prolog (BIM 1990), PHS is implemented using a combination of Prolog and C, and the terrain analysis package is in C. Users interact with KBLIMS through the Terrain Object Interface (TOI) that is a graphical database interface providing a graphical query facility, a hypertext-like browsing facility for object classes, and text and graphics windows for displaying results.The top level of TOI is accessed from a toolbar which has pull-down menus for displaying windows, accessing databases, editing class descriptions, and accessing application tools (Figure 1). The graphical query tool shows stream objects, stream-hillslope connectivity, and divides representing Turkey Lakes. There is a hierarchical organization to database queries, with the graphical tool providing rapid access to specific regions within a database, and hypertext to move about locally.

Graphical queries are facilitated through a transformation of topological relations between drainage basin objects into a graph. Links between objects are physically stored as binary relations, but are recognized by TOI as graphical objects. A graphical object is queried using a mouse pointer. Each graphical object is classified as a given type (*e.g.* stream, divide, or stream-to-hillslope) each of which is isomorphic to a conceptual object class. Organizing around a relatively small number of binary topological relations provides access to all attributes and other spatial relations, while giving the end-user a simple, uncluttered graphical query tool. Since graphical queries are limited to graphical counterparts of binary topological relations, the association of object classes to these primitive object types allows for accessing any objects within the knowledgebase.

As a testbed of variable land surface and process representation requirements we using the Turkey Lakes watershed in Ontario, Canada. As an example of a watershed scale management problem, we used KBLIMS to determine how a specific catchment within the Turkey Lakes Watershed would respond to clearcutting. By observing discharge differences between a forested baserun simulation and clearcut simulation, inferences can be made about potential effects on water quality *(e.g.* nutrient loading to aquatic environments) in areas downstream of the cleared catchment. Figure 1 show the results of the simulation for the queried catchment. Results indicate that clearcutting, approximated by reducing leaf area index to 1.0, results in greater spring melt runoff and a sustained higher summer baseflow. We can infer that enhanced nutrient loading to the downstream wetland will be accompanied by enhanced anaerobic conditions,which means the nutrients may be lost or unavailable for extended periods of time early in the growing season.

Other potential effects of clearcutting the catchment, such as compaction of soil, would further enhance runoff.

Concluding Comments

The approach taken in KBLIMS has led to a parsimonious coupling of GIS and ecosystem simulation modeling using object-oriented techniques which transform simple observations and relations into higher order concepts based on structural, deductive and procedural knowledge. A terrain analysis system transforms layer-based information to object information for use by the system. The knowledgebase uses a layering strategy to hide implementation details from higher level descriptions. Layering allow changes in low level tools such as simulation models and terrain analysis tools without changing the semantics of the objects which are defined in terms of meaningful hillslope and forest stand units. KBLIMS exploits the notion of a query model which executes a set of user-defined or system-defined queries. Each query is deductive so a user may define a simulation experiment by first identifying a set of objects as a spatial query, then specifying some action to be performed on these objects, such as a combined simulation query and aggregation query. The user, *e.g.* an ecologist, need not explicitly parameterize and run simulation models, although access to low level tools is provided. Typical use of the simulation system is managed by the knowledgebase using its metaknowledge. This allows for integration of either tightly-coupled or loosely-coupled systems designed to operate as integrated or standalone programs respectively, while maintaining a seamless view at the user interface level.

There is considerable benefit of the KBLIMS approach to support resource management and research where geographically-structured simulation modeling is important. The potential exists for a forest manager to sit down with the graphical interface and explore management scenarios in an efficient, effective manner. Land management researchers can also use a KBLIMS approach in developing and testing their models by using the information management tool to maintain a record of results tagged by a set of constraints for simulation. This approach provides, fast, easy access to simulation resources and results without the substantial information management overhead of creating and maintaining cumbersome input files for simulation programs.

Acknowledgements The work was partially supported by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada,a NSERC Doctoral Fellowship and a Tri-Council EcoResearch Doctoral Fellowship.

References

Band, L.E. 1989. A terrain-based watershed information system. *Hydrological Processes* 3:151-62.

Band, L.E., P. Patterson, R. Nemani and S.W. Running. 1993. Forest ecosystem processes at the watershed scale: Incorporating hillslope hydrology. *Agricultural and Forest Meteorology*.

Band, L.E. and V.B. Robinson. 1992. *Development of an Intelligent Land Information System (ILIS) as Part of the System for Hierarchical Experts for Resource Inventories (SHERI) Program*, Final Report, Submitted to Canada Centre for Remote Sensing.

Beven, K.J. and M.J. Kirkby. 1979. A physically based, variable contributing model of basin hydrology. *Hydrological Sciences Bulletin* 24:43-69.

BIM. 1990. *Prolog by BIM*, BIM, Kwikstraat, 4-3078 Everberg, Belgium.

Jenson, C.S., J. Clifford, S.K. Gadia, A. Segev, and R.T. Snodgrass. 1992. A glossary of temporal database concepts. *ACM Sigmod Record* 21(3):35-43.

Lammers, R.B. and L.E. Band. 1990. Automating object representation of drainage basins. *Computers and Geosciences* 16:787-810.

Mackay, D.S., V.B. Robinson, and L.E. Band. 1994. A knowledge-based approach to the management of geographic information systems for simulation of forested ecosystems, in Michener, W.K., Brunt, J.W., and Stafford, S.G. (eds), *Environmental Information Management and Analysis: Ecosystem to Global Scales*, Taylor & Francis, London. p. 511-534.

Michener, W.K., Brunt, J.W., and Stafford, S.G. (eds). 1994. *Environmental Information Management and Analysis: Ecosystem to Global Scales*, Taylor & Francis, London.

Moore, I.D. A.K. Turner, J.P. Wilson, S.K. Jenson, and L.E. Band. 1991. GIS and land surface-subsurface process modeling, in Goodchild, M.F., Parks, B.O. and Steyaert, L.T. (eds), *Environmental Modeling with GIS* , Oxford University Press, New York. p. 196-230.

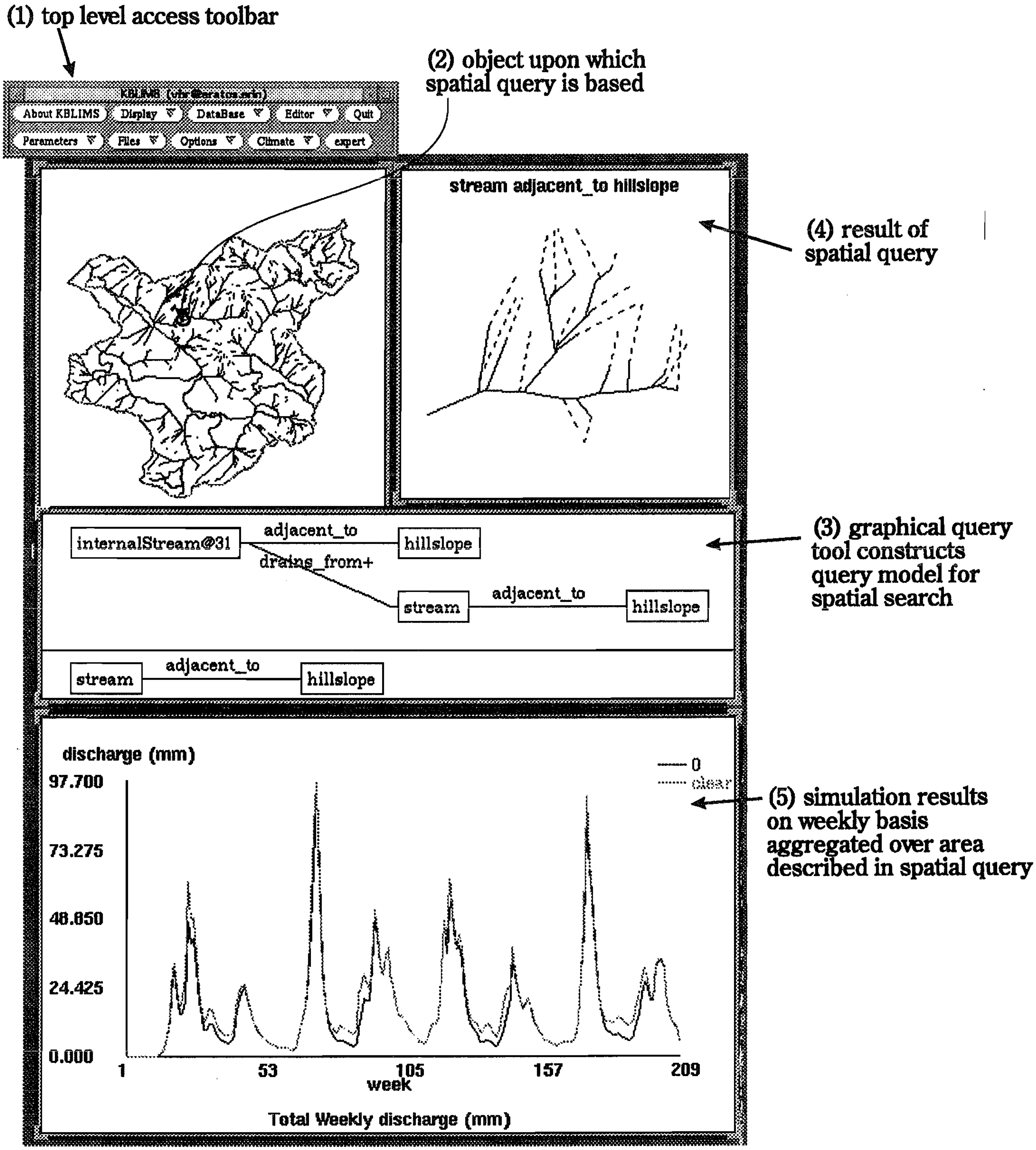

Figure 1. KBLIMS provides multiple windows for interaction with the system.