Pergamon

Comput., Environ. and Urban Systems, Vol. 19, No. 5/6, pp. 321–339, 1996

0198-9715/95 $9.50 + 0.00

S0198-9715(95)00017-8

SEMANTIC MODELING FOR THE INTEGRATION OF GEOGRAPHIC INFORMATION AND REGIONAL HYDROECOLOGICAL SIMULATION MANAGEMENT

Vincent B. Robinson[1] *and D. Scott Mackay*
Department of Geography, University of Toronto, Toronto, Ontario, Canada M5S 1A1

ABSTRACT. *Semantic modeling has proven valuable in the experimental development of the KnowledgeBased Land Information Manager and Simulator (KBLIMS) which integrates the management of geographic information and a hydroecological simulation system. Semantic models provide the abstractions such as aggregation, specialization, and generalization necessary for managing the structural and behavioral properties of Regional HydroEcological Simulation System (RHESSys), especially the spatial and temporal elements. It is shown how semantic modeling allows for explicit consideration of semantic heterogeneity as a result of domain evolution. The seamless integration is given expression by the visual spatial query system which provides the disciplinary scientist with the ability to deal directly with landscape elements such as hillslopes, stream valleys and watersheds rather than polygons or pixels. In the end this approach provides tools that are responsive and adaptive to the demands of disciplinary science.* Copyright © 1996 Elsevier Science Ltd.

INTRODUCTION

Understanding changes in environmental patterns and processes and predicting their effects at the ecosystem to the biosphere scale requires the incorporation of spatially explicit simulation modeling into environmental research. Thus, it has become evident that geographic databases and ecosystem modeling require further development to support research at regional to global scales (Michener et al., 1994). This paper discusses the use of semantic models to develop tools suitable for managing geographic data for an ecosystem modeling application.

The importance of semantics is easily overlooked when combining geographic information systems (GIS) and modeling tools. Moore et al. (1993) emphasize the importance of using canonical data models to combine GIS and simulation models. The importance of representing spatial distribution of environmental information in terms of physically meaningful patches is illustrated by approaches that use hillslopes (Band et al., 1991) and stands (Running & Coughlan, 1988) to represent large, complex systems. Recent activity directed towards integrating GIS and environmental models (e.g. Moore et al., 1993; Lathrop et al., 1994; Aspinall, 1994; Mackay et al., 1994) illustrates the information management needs of the scientific community.

[1] Tel: (905) 828-5465; email: vbr@geog.utoronto.ca or vbr@eratos.erin.utoronto.ca

In addition, one of the major challenges for future environmental information management and analysis is development of distributed analytical environments (Stafford et al., 1994).

In order to make use of distributed information, it is necessary that the semantics (or meaning) of each networked database be represented at a level accessible by specific application programs. One such application program is the Regional HydroEcological Simulation System (RHESSys). The parameterization and execution of RHESSys requires knowledge of assumptions implicit in its procedures. While this may present a limited burden to an experienced RHESSys-user in a single-user, single-processor environment, it leads to problems in a multiuser, multiprocessor system where changes in the domain are likely. Ecological models themselves are evolving as well as their data requirements. This domain evolution often results in the semantic heterogeneity problem whereby the local meaning of concepts is not globally consistent, known, or understood. Domain evolution can create semantic heterogeneity problems in distributed and nondistributed environments (Ventrone & Heiler, 1991). These are many of the issues being examined in development of the KnowledgeBased Land Information Manager and Simulator (KBLIMS).

KNOWLEDGEBASED LAND INFORMATION MANAGER AND SIMULATOR (KBLIMS)

Applications such as forest ecosystem modeling demand management of geographically-based information detailing complex interactions between climatic, topographic, hydrologic, pedological and ecological processes. Mackay et al. (1993) describe a knowledge-based approach for a geographically-based information system designed to reason and infer higher order concepts from observations and relations between lower order attributes. KBLIMS is a system for managing spatiotemporal simulations of ecological processes organized around a watershed-based model of terrain. It has been used to demonstrate that an object-oriented spatial database for watersheds can be easily organized as a graph and exploited for building a query system (Mackay et al., 1994). KBLIMS includes modules for the extraction of a watershed representation directly from grid digital elevation models, and an object-based information system allowing the selection, browsing, navigation among, and query of watershed objects using a visual spatial query tool. Procedural calls to a simulation system provide answers based on simulation results. The simulation system is based on the Regional Hydroecological Simulation System (RHESSys).

RHESSys is designed to compute carbon, water and nitrogen budgets over complex terrain. It includes versions of forest ecosystem process simulation models, climate models and hydrologic models which include a variation of the TOPMODEL basin hydrology model (Beven & Kirkby, 1979), MT-CLIM a mountain climate simulation model (Running et al., 1987), and FOREST-BGC a stand-level simulation model of water, carbon and nitrogen balance in coniferous forests. FOREST-BGC has been used to simulate hydrologic balance and primary productivity for a forest near Missoula, Montana, (Running & Coughlan, 1988) and to simulate the effects of climatic change on the regional carbon balance of a forest in Montana (Running & Nemani, 1991). In addition, Scuderi (1993) used the model at treeline sites in the Sierra Nevada to identify the relative importance of precipitation and temperature to tree growth.

RHESSys parameterizes and executes the model over extensive areas that may have strong spatial variability in important model parameters. Spatially continuous parameter fields are aggregated into discrete landscape units using a mean value for the parameters with each landscape unit. Landscape units are defined using techniques that incorporate spatial structure and variability of system behaviour within the simulation of watershed runoff and soil–water dynamics (Band, 1989). Watersheds and their subcomponents, i.e. hillslopes and stream channels, are extracted from digital terrain models. Implicit in the definition of hillslopes and stream links are topological relationships between links at junction points, and between links and hillslopes. The topology as defined by Band (1989) is a topology limited to spatially relating hillslopes with stream links and other hillslopes. To support the object-based management vital to development of KBLIMS (Mackay et al., 1991)Mackay et al. (1993) developed extensions to a full topology which includes watersheds and catchments.

Furthermore, the information is organized and accessible in reference to distinct, identifiable landscape units. Object-oriented and knowledgebased techniques are exploited to provide suitable tools for simulating ecosystem processes on the basis of high order geographic features such as hillslopes, streams and forest stands. Ecological simulation processes are organized around a watershed-based model of terrain with an infinite hierarchy of watersheds, hillslopes and stream links. Spatial representation at each level is organized into a network, consisting of a set of stream links, stream–to–hillslope adjacency links, and hillslope–to–hillslope adjacency links as illustrated in Fig. 1. This graph structure is the basic organizing principle for the object-oriented spatial database (Mackay et al., 1993).

DATA MODELS AND SEMANTIC MODELS

Geographic applications require data models that support spatial data types and special-purpose functions for spatial query processing (Orenstein & Manola, 1988). Traditional data models include the relational data model (Codd, 1970) and the Entity-Relationship (ER) model (Chen, 1976). The limitations of the ER (Worboys et al., 1990) and relational model (Kent, 1979) are attributable to a lack of abstraction mechanisms whereby individual concepts are combined with other individual concepts to increase the information content of the database. The early data models do not allow the semantics of the database to be readily expressed in the schema, but require it to be specified by the database designer and consciously applied by the user (Hammer & McLeod, 1981). In addition, geographic information must be qualified with respect to its location where it is valid, the time at which it is valid, and its accuracy (Roman, 1990).

Many problems with early data models are overcome with semantic data models. Semantic data models directly incorporate more semantics of the database into the database schema, by providing a declarative syntax and incorporating formal abstraction mechanisms. Other features of semantic data models include specialized functions and abstract data types (ADT). In addition, a semantic level of data modeling has been achieved and applied to geographic applications through extensions to the Entity-Relationship model (EER) (Gogolla & Hohenstein, 1991), extended first order logic (Roman, 1990), logic programming (Mohan & Kashyap, 1988), and semantic networks (Atzeni & Parker, 1988).

ABSTRACTION MECHANISMS

A database represents a specific slice of reality at a specific time. Its reality is completely determined if it is known which objects exist, the attributes of those objects, and the relationships between objects. A complete abstract model is a set of all models of state descriptions which are true with respect to the database reality for all time(s) represented by

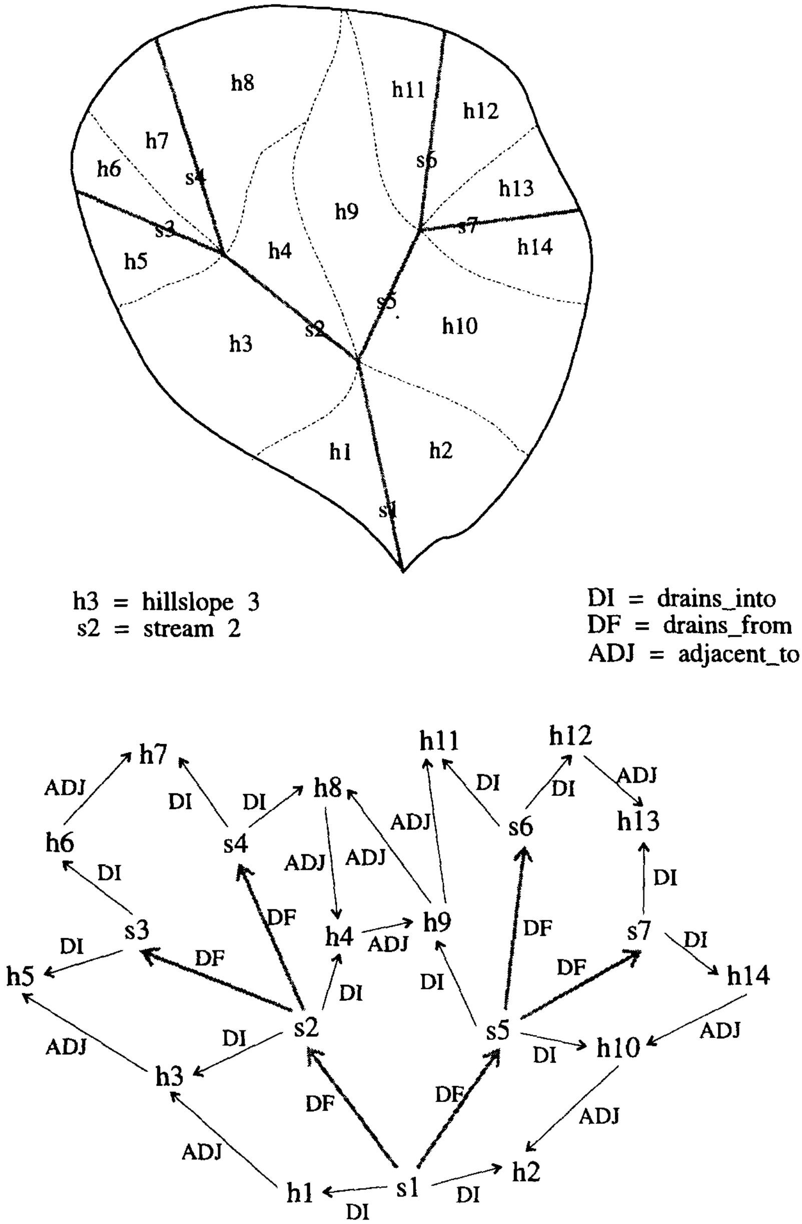

FIGURE 1. Spatial representation of hillslopes and streams as a network of spatial relations.

the abstract model. The abstract model is determined by the database schema (Biller & Neuhold, 1978). In order for an application program to read and correctly interpret information in a database, the schema must contain enough semantic information to describe what the information means in terms of the database. The representation ability of the schema (and database) is determined by expressiveness and semantic relativism. Expressiveness represents the power of the model structures to represent concepts and be interpreted as such concepts, and of the power of the model to represent the behaviours of concepts. Semantic relativism is the ability of the database to accommodate different views of the same information, or allow itself to be manipulated in a manner appropriate to a specific application (Geller et al., 1991).

Semantic relativism and expressiveness are captured by the concept of a database view. A view is a particular biased mode of regarding data and operations of this data. A view is a database that imports all its data from other databases. Consider two databases each of which describes a set of landscape units, and each database represents a unique region in space. Furthermore, each database manages a different configuration of the simulation models and parameterization procedures. A query on the entire region requires the formation of a particular view which has a copy of some information from both databases, with modification to support the structure of the query. The system needs to have some way of accessing the meaning of information in each database, and be able to correctly interpret it. This is accomplished by having each database provide the necessary structural and behavioral details about its information in its schema. This semantic information is organized through abstraction mechanisms, generalization, aggregation and classification.

Database views are captured by various abstraction mechanisms. Abstraction mechanisms provide structuring discipline to the data modelling task. Smith & Smith (1977) suggest that the structuring offers: (1) effective integration of important views of different users; (2) data independence as the database evolves; (3) reduced complexity in managing highly structured models; (4) a systematic approach to database development; and (5) assumptions about high level structure of domain problems allowing for efficient implementations. It has been argued that these benefits require the use of generalization and aggregation abstractions (Smith & Smith, 1977).

Generalization and specialization abstractions

Through generalization a class of individuals can be thought of generically as a single, named class. For example, a set of object classes such as Hillslopes, Stream and Divide can be generalized as the object class named Topographic. The opposite of generalization is specialization. Using the same example, Hillslope, Stream and Divide are all specializations of Topographic. A generalization/specialization hierarchy can be specified in terms of a relationship between each specialized class and their respective generic class(es) using the binary relation *isA*. The power of *isA* comes from its formal mathematical foundation. The *isA* relation is transitive (Atzeni & Parker, 1988) as in:

$$if\ X\ isA\ Y \text{ and } Y\ isA\ Z \text{ then } X\ isA\ Z \tag{1}$$

To illustrate the utility of equation (1) in the formal definition of database properties, consider the binary predicate *isA* which describes the *isA* relation. Given the facts:

isA(Landscape-unit, Hillslope) and

isA(Hillslope, Topographic)

about the relationship between Landscape-units and Hillslopes, used interchangeably by Band et al. (1991), the following fact can be deduced from equation (1):

isA(Landscape-unit, Topographic)

which retains the truths about Topographic objects in the definition of Landscape-units. Every instance of Landscape-unit is also an instance of Topographic and is known as an *extensional isA* constraint.

The development of *isA* hierarchies permits systematic development of database classes and incorporation of differing views that do not violate existing database assumptions.

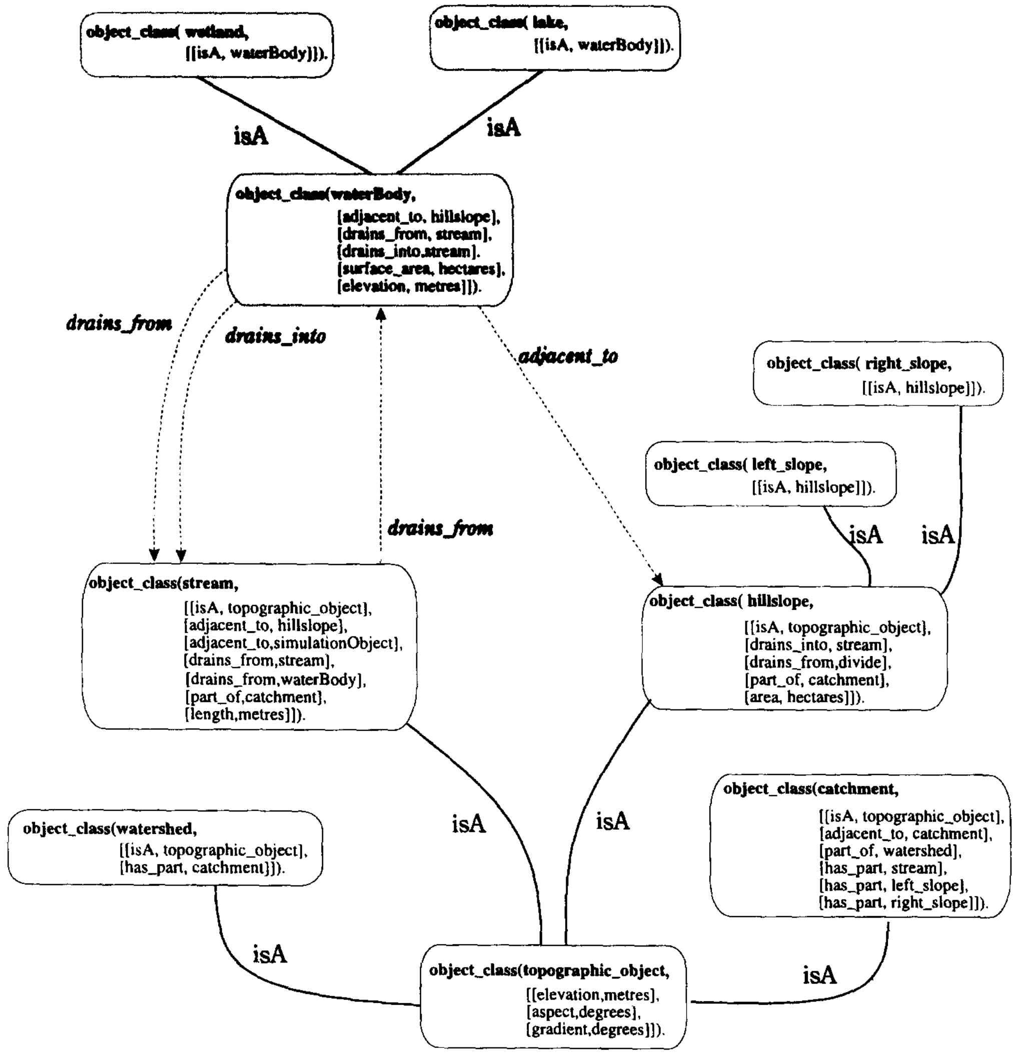

FIGURE 2. Example of *isA* hierarchies used to manage semantic heterogeneity through KBLIMS. The unshaded area is a representation that is applicable to both mountainous and nonmountainous situations. The shaded area is an extension for those regions where waterbodies such as lakes and wetlands are to be represented and managed.

Figure 2 illustrates an example of an *isA* hierarchy drawn from KBLIMS. In an application dealing only with streams in a mountainous region, there are no waterbody objects represented (Mackay et al., 1993). In that case the representation would be the same as depicted in Fig. 2 but would not include those objects and relations in the shaded areas. When an application in a nonmountainous area with lakes and wetlands was developed it was a simple matter to add the waterbody, lake and wetland objects as an *isA* hierachy without violating existing database assumptions.

Aggregation and decomposition

The *isA* relation does not adequately capture the semantics of space. Assume that space and time are viewed as attributes of a geographic object. In other words, a geographic object is qualified by its location in space, spatial reference to other geographic entities, and by it location in time. Thus, spatial and temporal relationships can be incorporated into a data model by using aggregation. Aggregation refers to an abstraction in which a relationship between objects is considered as a higher level object (Smith & Smith, 1977). Aggregation has always played an important role in data modelling. The semantic network is influential in this area since it can explicitly account for the meanings of attributes and relations not captured in the relational model.

Aggregation is used to express relationships between objects and describe attributes of an object. Such an aggregation relation in KBLIMS may have the form:

object_class(object, [[attribute, value]]).

This triplet declares that an individual (*object*) has a relationship (*attribute*) to some value or other individual (*value*). In KBLIMS the representation for the definition of hillslope is:

object_class(hillslope,

[[*isA*, **topographic_object**],	**object_class(topographic_object,**
[drains_into, stream],	[[elevation, metres],
[drains_from, divide],	[aspect, degrees],
[part_of, catchment],	[gradient, degrees]]).
[area, hectares]]).	

The hillslope description inherits properties of the topographic_object description. Note that in the triplet *attribute* may refer to a relationship such as *isA* or to a characteristic such as *area* or *aspect*.

Characteristics, such as *area* and *aspect*, refer to static attributes of an object class. In the above example, *area* is assumed invariant over time for hillslope objects, since hillslope development generally occurs over a greater time-scale than is captured in the database. Attribute-properties refer to dynamic properties of a class. For instance, leaf area index is in continuous flux throughout the growing season, especially for deciduous trees and grasses. This suggests a further qualification of the attribute-properties to include time, in this case the period of time during which a given leaf area index value is valid.

An important advantage of aggregation is the ability to reason about complex objects through hierarchical relations. In KBLIMS there are hierarchical relations such as *part_of* and *has_part* which correspond to Bannerjee et al. (1987) *is-part-of* and *has-part*. The *part_of* is an upward hierarchical relationship between a simple object and a complex object. In KBLIMS this relation is used to describe the relationship between watershed, catchment, stream and hillslope where hillslope and stream are component parts of a catchment which in turn is related upward to watershed (Fig. 3).

Downward hierarchical relationships are expressed with *has_part*. This downward relation between watershed, catchment, stream, left_slope and right_slope where catchment *has_part* of a stream, and the left_slope and right_slope of a hillslope is illustrated in Fig. 4. Note that the characteristic attributes of left_slope and right_slope are inherited from hillslope.

Gogolla and Hohenstein (1991) define a **component** operator which allows for the reference of complex entities to other entities, sets, or lists. In KBLIMS the **aggregate** operator

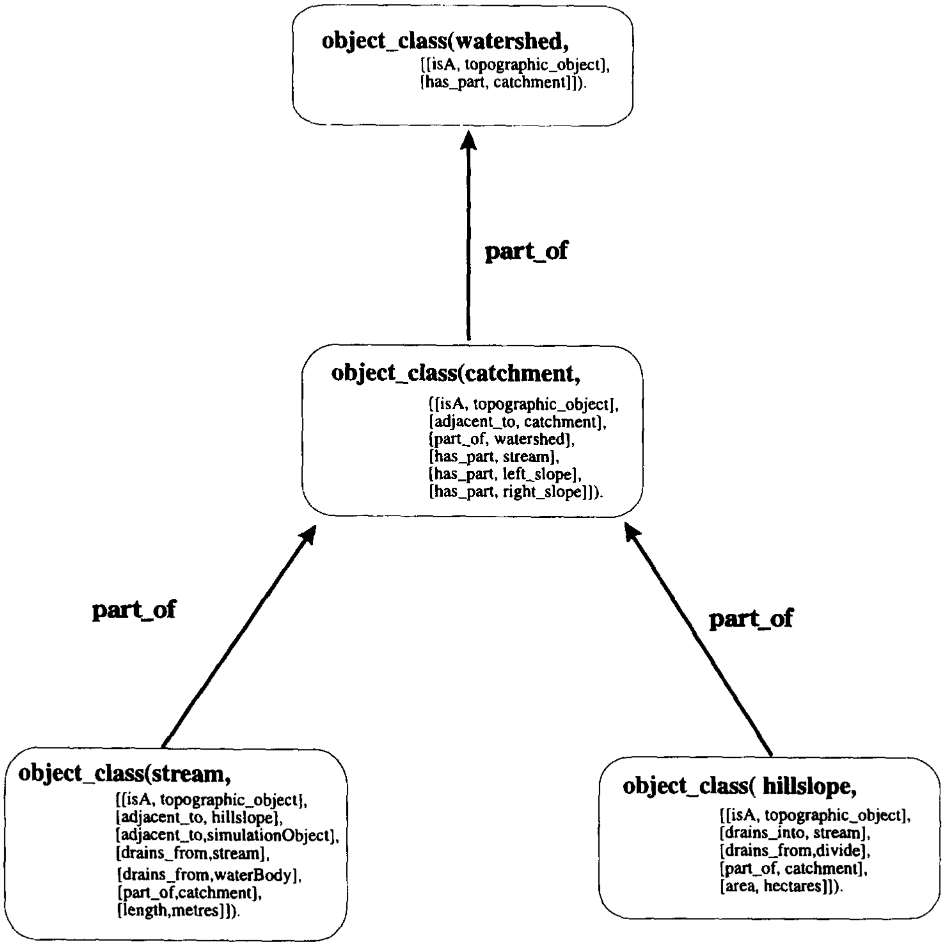

FIGURE 3. The upward relation *part_of* is used to describe the relationship between watershed, catchment, stream and hillslope.

performs much the same function by producing a set of objects with pointers back to component objects. As illustrated in Fig. 5 aggregation exploiting the *isA* relation is used to carry sets of objects to simulation functions. In this case simulationObject is the complex object made up of the simpler objects—hillslope, objectLAI, objectSoil, and simulated.

While complex objects allow for a limited expression of spatial relationships, spatial relations are needed in order to complete the representation of space, and functional relationships between spatial entities. Gogolla and Hohenstein (1991) used relations such as *lies-in* and *flows-into* to express spatial relationships. Similarly, in a hydroecological context, relations such as *drains_into* and *drains_from* express important spatial relationships. The problem with relations like *drains_into* is that they contain both spatial and functional information. Although they are expressive and well understood, i.e. as natural language, the functional component may not be explicitly known. Consider a binary predicate *drains_into* that is defined to represent the fact that an object flows into another. For example, consider the fact:

drains_into(streamLink1, streamLink2)

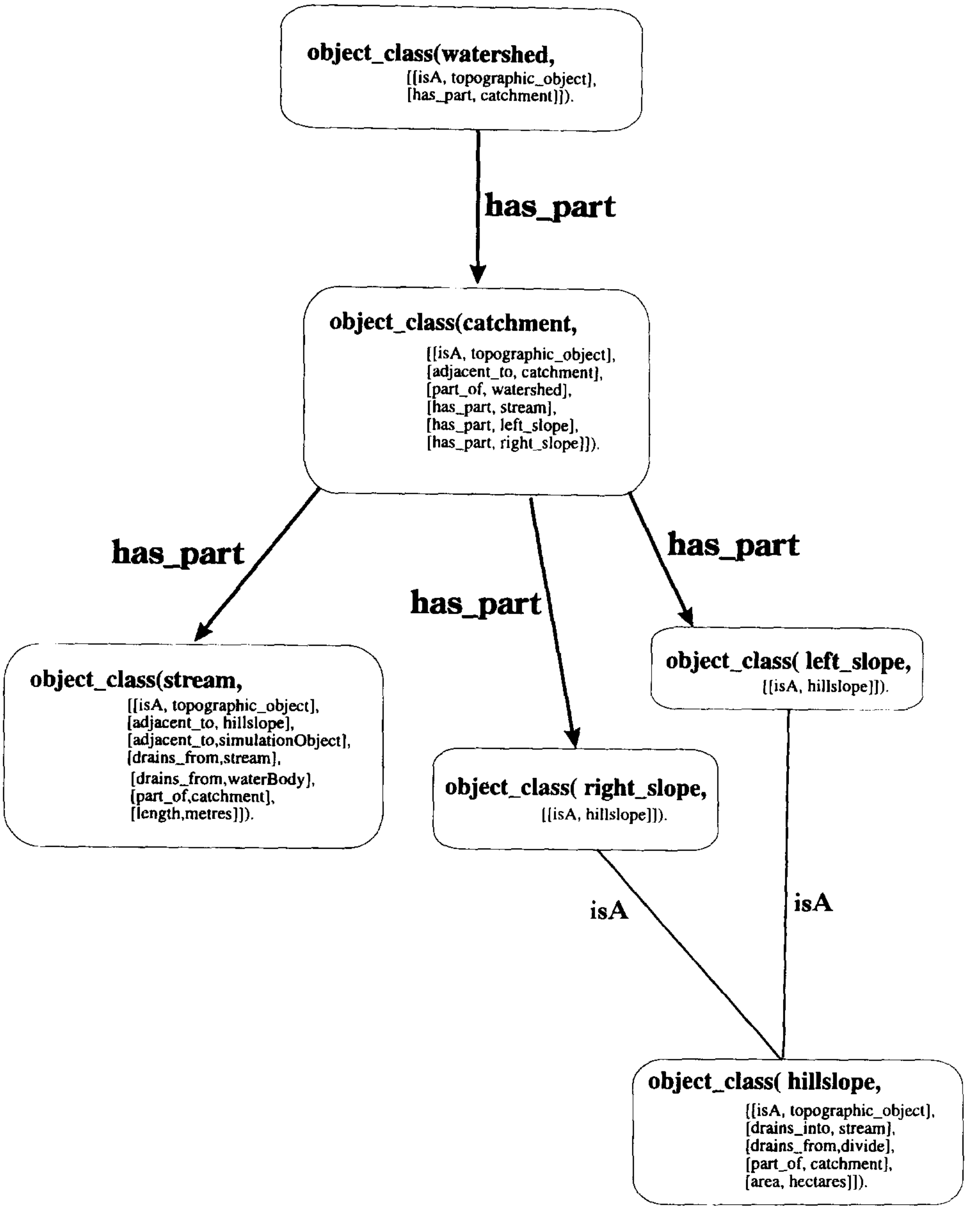

FIGURE 4. The *has_part* relation describes downward hierarchical relationships between watershed, catchment, stream, left_slope and right_slope where catchment *has_part* of a stream, and the left_slope and right_slope of a hillslope.

which is intended to state that streamLink1 flows into streamLink2. The implied functional constraint between stream links is incomplete without specifying both topological relationships between streamLink1 and streamLink2, and the physical constraints which govern the behaviour of the system (i.e. water flows downhill). Relations such as these might better be specified if they are built from simpler relations (Mohan & Kashyap, 1988; Roman, 1990). Equation (2) demonstrates how *drains_into* can be built from known topology (i.e. the relation *connected*) and the elevation of each stream link with e and f being the respective

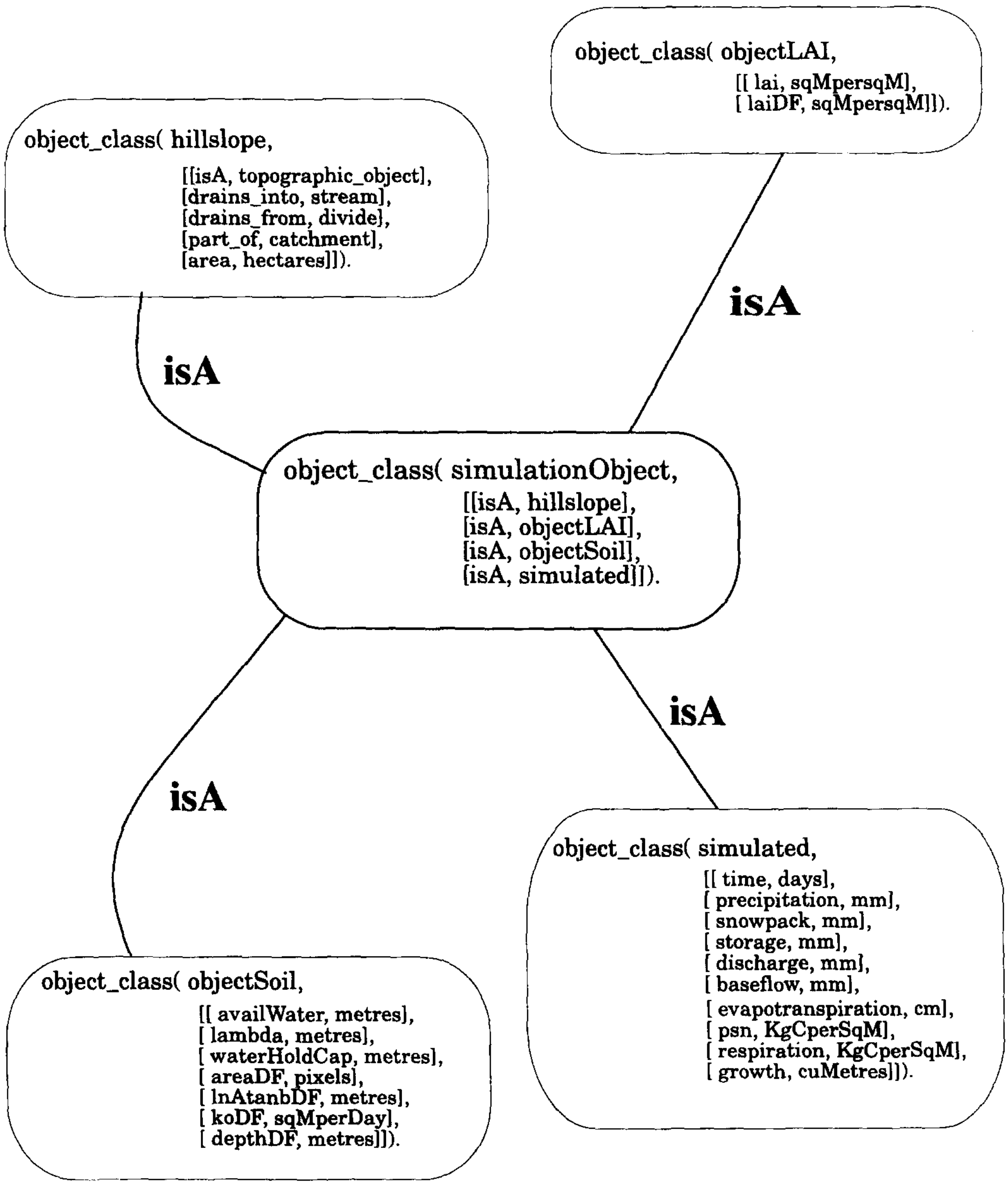

FIGURE 5. In KBLIMS the aggregation abstraction is used to exploit the *isA* relation to form simulationObject as a complex object composed of the set of simpler objects—hillslope, objectLAI, objectSoil, and simulated.

elevation values.

$$
\begin{gathered}
(\forall\ x, y\ \in\ \text{streamLink}) \\
\neg[\mathit{equal}(x,y)]\ \wedge\ \mathit{connected}(x,y)\ \wedge \\
\mathit{elevation}(x,e)\ \wedge\ \mathit{elevation}(y,f)\ \wedge\ e > f \\
\Longrightarrow\ \mathit{drains\_into}(x,y)
\end{gathered}
\tag{2}
$$

Equation (2) explicitly defines spatial and functional conditions on the relation *drains_into*.

Classification and instantiation

A class is a collection of entities that have similar properties. Each entity of a class is an instantiation of that class. For example, a particular landscape-unit entity has specific values associated with each of its attributes. The instantiation of class Landscape-Unit, landscape-unit-1, might have the following attributes:

(landscape-unit-1, #intervals, 5)
(landscape-unit-1, interval-area, [13.0, 266.0, 359.0, 33.0, 14.0])
(landscape-unit-1, hydrologic-index, [2.5, 4.5, 6.5, 8.5, 10.5])
(landscape-unit-1, hydraulic-conductivity, [0.938, 0.938, 0.938, 0.938, 0.938])
(landscape-unit-1, saturation-capacity, [0.4, 0.5, 0.6, 0.7, 0.8])
(landscape-unit-1, leaf-area-index, [7.02, 11.51, 10.46, 11.54, 9.32])

where [ ] denotes a list.

Classification is further useful for partitioning a Universe of Discourse (UoD) into groups of object classes. Hendrix (1979) suggests the use of spaces as groups of nodes and arcs in a semantic network, bundled together. Spaces allow for the representation of metainformation. To express the fact that all spatial objects in KBLIMS are specified with known location, topological relationships to other spatial objects, and geometric properties, the concept spatialObject is specified:

concept spatialObject *isA* object
attribute
topology = topologicalProperty
location = coordinate
geometry = geometricProperty;

SPATIAL AND TEMPORAL REPRESENTATION

Space and time can be modeled as attributes of geographic entities, using the aggregation abstraction. However, specific spatial and temporal semantic issues remain. In order for a relation to be useful for managing spatial and temporal domains, it has to incorporate both topological and functional components. This requires formal models of space and time (Roman, 1990).

Time is typically modelled as a one-dimensional space (Orenstein & Manola, 1988; Roman, 1990). However, its incorporation into the semantics of geographic applications such as hydroecological simulation, should be in the form of constraints which provide qualification on the use of specific concepts. In KBLIMS TemporalProperty provides temporal qualification to a property which may be an attribute of simulationObject. TemporalProperty is essentially a relationship between a property and a time:

$$temporalProperty(\text{property, validTime}).$$

It is assumed that it is a property for which we want to maintain an historical record. Since time is a property, a relationship is defined between two properties. In other words a property which relates two properties. For example,

concept simulationObject *isA* system
attribute

```
        observedProperty = observation;
    concept observation isA [physicalQuantity, temporalProperty];
    concept temporalProperty isA property
        source
         historicalProperty = property
        destination
         temporalQuantification = validTime;
    concept validTime isA time
        timeDimension
         validTimeDimension = <second hour day week year>
```

which shows that each simulationObject has as an observed property temporalProperty, and that temporalProperty is quantified as validTime where validTime refers to the time in the real world not transaction time (Jenson et al., 1992).

SPATIAL DECOMPOSITION

To decompose a spatial domain into subspaces it is necessary to consider how information is apt to be used in an application, how queries may be structured, and how problem generalization may occur. In RHESSys each landscape-unit is treated as an independent object and is scale dependent. A landscape-unit in a small region might be based on the hillslope partition. In a mesoscale situation the partitions could be aggregated into catchments to form landscape-units. At regional scales the aggregation could be at the watershed level. Effective information management over all scales of simulation may be achieved by partitioning the spatial domain into regions of connected objects. This can be achieved by explicitly storing all spatial relationships at all levels in the hierarchy of complex and simple objects. Alternatively this objective can be achieved storing the topology between simple objects and using information about the spatial hierarchy to assemble regions of connected, complex objects. The latter approach is likely to lessen the volume of data stored and requires the system have some knowledge of spatial hierarchies and how to manipulate them.

Topology (Beguin & Thisse, 1979) and logical relationships (Reiter & Mackworth, 1989) are important tools for representing spatial relationships between geographic objects. Some semantic models, tailored for geographic applications, use point-set topology (e.g. PROBE; Orenstein & Manola, 1988). Point-set topology is suggested by Egenhofer and Franzosa (1991) as a tool for formalizing the semantics of spatial relations. For RHESSys applications, point-sets provide a means of determining the topology between objects represented as sets of pixels on raster imagery. In order to represent topology between hillslopes a sentence may be defined as:

$$
\begin{gathered}
(\forall\ w,\ x \in \text{hillslope}) \\
(\forall\ y,\ z \in \text{pixel}) \\
\neg[\mathit{equal}(w,x)] \ \wedge\ \mathit{point\text{-}set\text{-}of}(w,p)\ \wedge\ \mathit{point\text{-}set\text{-}of}(x,q)\ \wedge \\
\mathit{in}(y,p)\ \wedge\ \mathit{in}(z,q)\ \wedge\ \mathit{adjacent}(y,z) \\
\Longrightarrow\ \mathit{adjacent}(w,x)
\end{gathered}
\tag{3}
$$

where *point-set-of* is a predicate which returns a list of all points associated with a hillslope partition, *in* tests for set containment of a value within a list, and *adjacent* is a predicate which tests adjacency between two objects of the same type. Equation (3) assumes that hillslopes do not overlap in space, a condition which may be included with a fact such as *disjoint(p,q)*. However, this assumption is better incorporated into the system as an integrity constraint on the definition of hillslopes rather than a constraint on every query using hillslopes.

With hillslope topology defined, topological relationships between higher order objects can be defined exploiting hillslope topology and related spatial hierarchy. For example, topology between catchments is defined as:

$$\begin{array}{c}(\forall\ w,\ x\ \in\ \text{hillslope})\\(\forall\ y,\ z\ \in\ \text{catchment})\\\neg[equal(y,z)]\ \wedge\ part\_of(w,y)\ \wedge\ part\_of(x,z)\ \wedge\ adjacent(w,x)\\\Rightarrow\ adjacent(y,z)\end{array} \tag{4}$$

An expression similar to equation (4) can be used to retrieve the topology of watersheds from the topology of catchments or hillslopes.

SEMANTIC HETEROGENEITY AND INTEGRITY CONSTRAINTS

Specifying integrity constraints

Integrity constraints (IC) apply to both static and dynamic aspects of a model. An IC is static if true in all states of a UoD. It is dynamic if verification of its truth requires at least two states (Wieringa et al., 1989). A state is a set of knowledge about objects and their attributes at a particular instance in time. There are analytical, empirical or deontic ICs. Analytical constraints follow from the definition of concepts. For example, *aspect* $\in$ *degrees*, is an analytical constraint that aspect is measured in degrees. There is nothing precluding changing this constraint to, *aspect* $\in$ *radians*. This changes both the meaning of the attribute *aspect*, and its use in an application program.

An empirical constraint describes the behaviour of a UoD (Wieringa et al., 1989). An example is as follows:

$$0.0 \leq \text{hydrologic-index} \leq 15.0 \tag{5}$$

where equation (5) sets a constraint on the range of possible values for the hydrologic similarity index. A value of 16.0 would be a deviation in the behaviour of the UoD from its norm, and reported as a violation of the constraint.

Since space and time are treated as attributes of geographic objects, spatial and temporal constraints are specified as any other constraint. For example, a constraint *spatial-disjoint* can be used to verify that all topographic objects, at a given partition level, do not overlap in space. *Spatial-disjoint* can be defined as:

$$\begin{array}{c}(\forall\ w,\ x\ \in\ \text{topographic\_object})\\point\text{-}set\text{-}of(w,p)\ \wedge\ point\text{-}set\text{-}of(x,q)\ \wedge\\\left[(\forall\ y\ \in\ \text{pixel})\ \neg in(y,p)\ \wedge\ in(y,q)\right]\\\Rightarrow\ spatial\text{-}disjoint(w,x).\end{array} \tag{6}$$

Spatial-disjoint can then be used to detect when two hillslopes overlap, as in the following empirical constraint:

$$\begin{array}{c}(\forall\ w,x\ \in\ \text{hillslope})\\\neg spatial\text{-}disjoint(w,x)\ \Rightarrow\ overlap(w,x)\end{array} \tag{7}$$

When the violation of the constraint is detected it can be reported to the system or system user.

A deontic constraint describes the permissible behaviour of a UoD, as defined by some agent (Wieringa et al., 1989). In KBLIMS a deontic constraint might be:

> *Landscape-unit partitions minimize within unit variance of slope gradient and aspect, and maximize between unit variance of slope gradient and aspect.*

This expresses a fundamental assumption of the RHESSys simulation system, and therefore, is a constraint on the semantics of the database. However, a particular application program could violate the deontic constraint. Violation of deontic assumptions in application code is permissible, but the system should be able to detect them and inform the user of the problem. In the case of violation of this constraint, the user might be informed that the spatial distributions of evapotranspiration and net primary productivity are not predictable from the soil water–vegetation density–climate relationship (Nemani & Running, 1989).

Semantic heterogeneity

The parameterization and execution of RHESSys simulations requires knowledge of assumptions implicit in its procedures. While this presents a limited burden to a skilled user in a single-user, single-processor environment, it leads to problems in a multiuser, multiprocessor system where changes in the domain are likely. Domain evolution can result in a semantic heterogeneity problem whereby the local meaning of concepts is not globally consistent, known or understood. In fact, domain evolution can produce semantic heterogeneity problems in either singular or distributed database environments (Ventrone & Heiler, 1991).

In ecological process modeling it is often useful to compare simulation results obtained for differing scales. For example, let us consider a RHESSys configuration consisting of two databases. In the *database-1* landscape-units are defined on the basis of entire watersheds (i.e. 10s of square kilometres) while in *database-2* simulation is based on hillslopes (i.e. 10s of hectares). Band et al. (1991) have shown that one cannot assume that important processes of hydroecological simulation are scale independent. Therefore, it cannot be assumed that both databases drive precisely the same model configurations. It must be expected that there will be semantic differences between *database-1* and *database-2*. For example, at one scale lakes are explicitly represented, while as one scales up, the lakes are incorporated into terrestrial land units and thus are not explicitly represented. These semantic differences will also be reflected in adaptations, or versions, of the process models. If this heterogeneity were not recognized by KBLIMS local assumptions could be violated negating the results of one or both analyses. Local assumptions must be translated and specified at the schema level, rather than embedding them into the program code for the simulation models.

In a distributed scenario simulations may be run covering an area where *database-1*, a mountainous watershed with no waterbodies, is used in conjunction with *database-2*, a lowland watershed with several waterbodies (lakes and wetlands). This heterogeneity must be accommodated during transactions requiring information from *database-1* and *database-2* (e.g. aggregation of process information). Simultaneous application of information from both *database-1* and *database-2* could violate local assumptions. Local assumptions must be translated and specified at the schema level. If constraints are formally specified where they are globally accessible and interpretable, then reconciliation of the semantics can be made at the local level prior to routing a query to the global level (Weishar & Kerschberg, 1991).

Mackay et al. (1993) describe the object-based representation of KBLIMS with reference to a mountainous region, Soup Creek, Montana. This representation does not allow for the representation of waterbodies other than streams. Therefore, a query for all upstream links is a relatively trivial, but recursive, query process. This was reasonable considering the context of their application. Until recently, KBLIMS had been developed and applied only in the context of regions like Soup Creek. However, in nonmountainous areas there is less topographic relief and waterbodies such as lakes and/or wetlands may be common. In this context, KBLIMS performs the same recursive query processing, skipping embedded lakes, then returns to fill them in. In other words, the data types are heterogeneous since the relation between lake and stream has no recursive meaning. It also means that the query has to know of the lake–stream relation then find stream–stream relations to do transitive closure. This is accomplished using the representation in Fig. 2 and the inferred relation *becomesA*. The inverse of *isA* waterBody is *becomesA* lake (or wetland).

BecomesA is inferred by generating literals *isA*(child1,parent)

isA(child2,parent),

then reversing the relationship to *becomesA*(parent, child1)

becomesA(parent, child2).

This strategy is used to address heterogeneity of the data model without cumbersome management of many specialized algorithms. Therefore, we suggest that domain evolution has led to the ability of KBLIMS to address semantic heterogeneity.

SEMANTIC MODELING AND INTELLIGENT QUERY

To the user, it is at the query level where tangible results of semantic modeling for integrating geographic information and hydroecological simulation should be evident. KBLIMS is based on the notion of a query model which executes a set of user-defined or system-defined queries. A typical operation might be to compute the annual total aggregate transpiration for hillslopes connected to a particular divide. To the user this operation is viewed as a query on a graphical object on the interface, then as a selection of some operation on the retrieved object(s) from a pull-down menu. Graphical objects correspond with specialized relations in the knowledgebase. These relations are associated with object class schemes that allow for both a coarse-grained and a fine-grained view of objects in the knowledgebase. The end-user need not explicitly parameterize and run simulation models, although access to low-level tools such as the simulation system is provided. Typical use of the simulation system is managed by the knowledgebase using its metaknowledge. This allows for integration of 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.

The semantic models and object-based notation provide the underpinning upon which the visual spatial query system is based. It supports the query model and facilitates its expression. Fig. 6 illustrates how semantic modeling supports the visual query ability of KBLIMS for the Turkey Lakes watershed in Ontario, Canada. Each query is deductive so a user defines 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. In Fig. 6 the set of objects are the waterbodies and streams

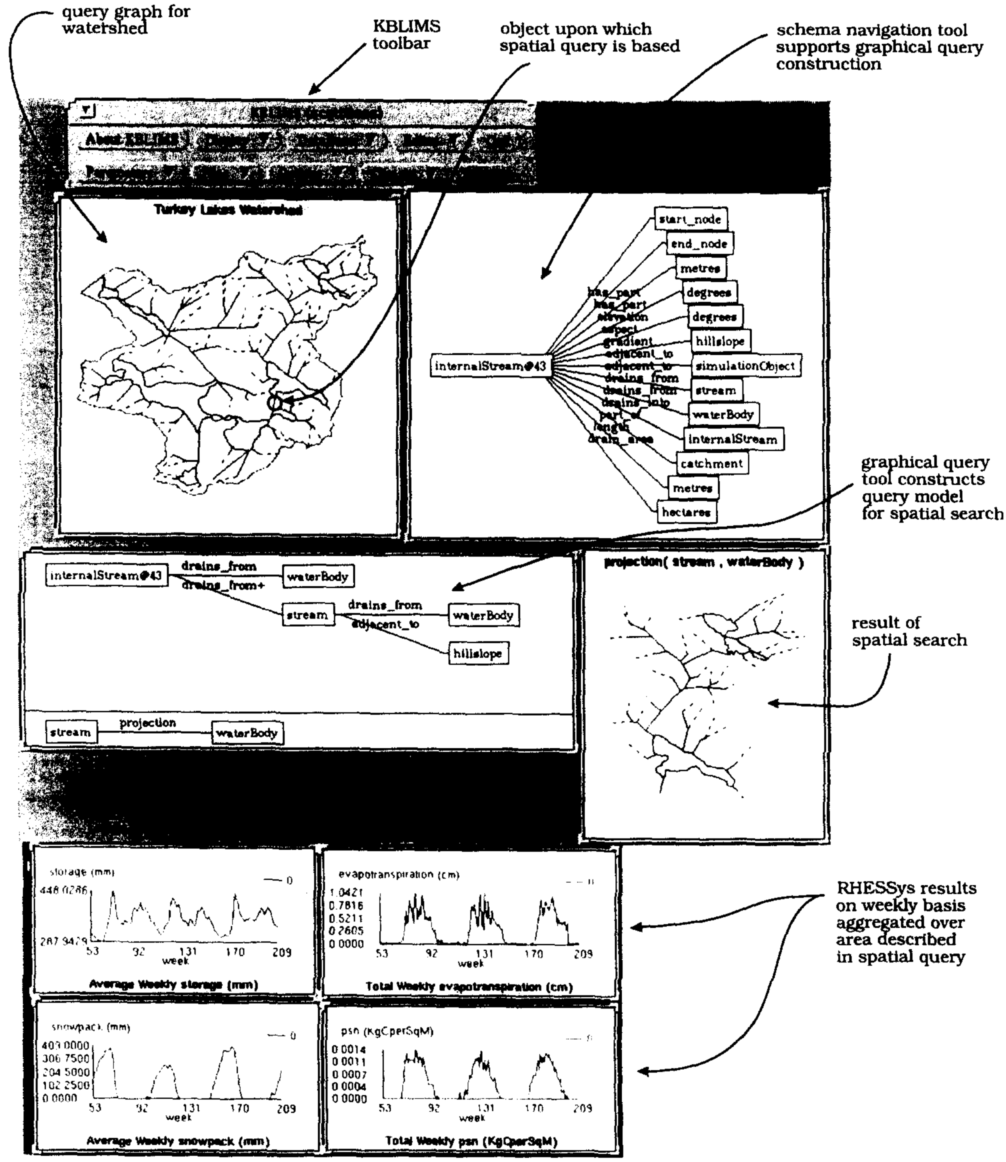

FIGURE 6. Example of a spatial query and simulation session with KBLIMS.

that drain into stream link #43 and the hillslopes adjacent to those streams. The action is a simulation and aggregation query where a RHESSys simulation is conducted over the objects retrieved as a result of the spatial search [i.e. the projection(stream, waterbody)] and aggregated over that area on a weekly basis rather than daily. The user is able to construct the query using the schema navigation tool which allows the navigation through the schema as well as provides that capability of graphically constructing the spatial query.

CONCLUDING DISCUSSION

Semantic modeling has proven to be valuable in the experimental development of KBLIMS which integrates the management of geographic information and a hydroecological simulation system. Semantic models provide the abstractions necessary for managing the structural and behavioral properties of RHESSys, particularly the spatial and temporal aspects. The generalization–specialization abstraction is used to expand upon existing information by adding concepts which inherit the specifications of previously defined concepts. Space and time are treated as attributes of geographic objects. And are managed through aggregation, by incorporating spatial and temporal relations into an aggregation hierarchy. Spatial relations having only spatial information (e.g. topological) are semantically clear and easily interpreted. However, spatial relations which incorporate implicit functionality between objects (e.g. drains_into) incorporate both spatial and functional semantics. This is accomplished by construction of complex relations from simpler relations. Since space and time are attributes of geographic objects, spatial and temporal constraints are specified in the same way that constraints are specified for non-spatial attributes. The specification of constraints is essential to maintenance of system integrity and production of reliable results from simulation transactions.

The KBLIMS project is part of a research strategy to study issues of heterogeneous GIS in a controlled environment, derived from a homogeneous system. KBLIMS development functions as a testbed for experimenting with transactions over a heterogeneous geographic information and RHESSys. Domain evolution is a significant source of semantic heterogeneity in all information systems. The domain of KBLIMS has evolved from strictly a mountainous domain with no provision for waterbodies other than, exclusively, streams, to include a distinctly nonmountainous domain with representation that does include waterbodies such as lakes and wetlands. However, both representations are valid for their respective domains and are handled due to effective semantic modeling. Also, RHESSys incorporates many implicit assumptions into the program code, i.e. assumed knowledge on the part of the user. In a distributed system, these assumptions would be incorporated locally but would differ from one subspace to another, depending on the information processing needs of each subspace. A query over a set of subspaces has to be able to read and interpret the information in each subspace. Thus, this information must be specified in each local subspace's database schema, using a data model. This is also true in a nondistributed setting where differing versions of the same model may exist. Such domain evolution is actually quite common in a scientific simulation modeling environment.

Semantic modeling provides the essential foundation to support the ability to manage ecological/spatial objects in a transparent manner throughout a simulation experiment. In conjunction with a graphical query interface, the object-based notation provides the disciplinary scientist with the ability to deal directly with landscape elements with which scientists are comfortable with. In this case there are hillslopes, stream valleys, and watersheds rather than polygons or pixels. Object models relieve the disciplinary scientist from the necessity of having to directly link and manage software for numerical simulation and geographical analysis. This provides tools that are more responsive and adaptive to the demands of disciplinary science.

ACKNOWLEDGEMENTS– This work was partially supported by an operating grant from the Natural Sciences and Engineering Research Council (NSERC), an NSERC Doctoral Fellowship, and a Tri-Council EcoResearch Doctoral Fellowship.

REFERENCES

Aspinall, R. J. (1994). GIS and spatial analysis for ecological modelling. In W. K. Michener, J. W. Brunt and S. G. Stafford (Eds), *Environmental Information Management and Analysis: Ecosystem to Global Scales* (pp. 377–396). London: Taylor and Francis.

Atzeni, P. & Parker, D. S. (1988). Formal properties of net-based knowledge representation schemes. *Data and Knowledge Engineering*, *3*, 137–147.

Band, L. E. (1989). A terrain-based watershed information system. *Hydrological Processes*, *3*, 151–162.

Band, L. E., Peterson, D. L., Running, S. W., Coughlan, J. C., Lammers, R. B., Dungan, J. & Nemani, R. (1991). Forest ecosystem processes at the watershed scale: basis for distributed simulation. *Ecological Modelling*, *56*, 151–176.

Bannerjee, J., Kim, W., Kim, H. J. & Korth, H. F. (1987). Semantics and implementation of schema evolution in object-oriented databases. *Proceedings ACM SIGMOD*, 311–322.

Beguin, H. & Thisse, J. F. (1979). An axiomatic approach to geographical space. *Geographical Analysis*, *11*, 325–341.

Beven, M. & Kirkby, M. J. (1979). A physically based, variable contributing model of basin hydrology. *Hydrological Sciences Bulletin*, *24*, 43–69.

Biller, H. & Neuhold, E. J. (1978). Semantics of databases: the semantics of data models. *Information Systems*, *3*, 11–30.

Chen, P. P. (1976). The entity-relationship model—toward a unified view of data. *ACM Transactions on Database Systems*, *1*, 9–36.

Codd, E. F. (1970). A relational model for large shared data banks. *Communications of the ACM*, *13*, 377–387.

Egenhofer, M. J. & Franzosa, R. D. (1991). Point-set topological spatial relations. *International Journal of Geographical Information Systems*, *5*, 161–174.

Geller, J., Perl, Y. & Neuhold, E. (1991). Structure and semantics in OODB class specifications. *ACM SIGMOD Record*, *20*, 40–43.

Gogolla, M. & Hohenstein, U. (1991). Towards a semantic view of an extended entity-relationship model. *ACM Transactions of Database Systems*, *16*, 369–416.

Hammer, M. & McLeod, D. (1981). Database descriptions with sdm: a semantic data model. *ACM Transactions on Database Systems*, *16*, 369–416.

Hendrix, G. G. (1979). Encoding knowledge in partitioned networks. In N. V. Findler (Ed.), *Associative Networks: Representation and Use of Knowledge by Computers* (pp. 51–92). New York: Academic Press.

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

Kent, W. (1979). Limitations of record-based information models. *ACM Transactions on Database Systems*, *4*, 107–131.

Lathrop, R. G., Aber, J. D., Bognar, J. A., Ollinger, S. V, Casset, S. & Ellis, J. M. (1994). GIS development to support regional simulation modelling of north-eastern (USA) forest ecosystems. In W. K. Michener, J. W. Brunt and S. G. Stafford (Eds), *Environmental Information Management and Analysis: Ecosystem to Global Scales* (pp. 431–452). London: Taylor and Francis.

Mackay, D. S., Band, L. E. & Robinson, V. B. (1991). An object-oriented system for the organization and representation of terrain knowledge for forested ecosystems. *Proceedings GIS/LIS '91*, 617–626.

Mackay, D. S., Robinson, V. B. & Band, L. E. (1993). An integrated knowledge-based system for managing spatiotemporal ecological simulations. *AI Applications*, *7*(1), 29–36.

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

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

Mohan, L. & Kashyap, R. L. (1988). An object-oriented representation for spatial information. *IEEE Transactions on Software Engineering*, *14*, 675–681.

Moore, I. D., Turner, K. A., Wilson, J. P., Jenson, S. K. & Band, L. E. (1993). GIS and land-surface-subsurface modeling. In M. F. Goodchild, B. O. Parks and L. T. Steyaert, (Eds), *Environmental Modeling with GIS* (pp. 196–230). New York: Oxford University Press.

Nemani, R. & Running, S. W. (1989). Testing a theoretical climate-soil-leaf area hydrologic equilibrium of forests using satellite data and ecosystem simulation. *Agricultural and Forest Meteorology*, *44*, 245–260.

Orenstein, J. A. & Manola, F. A. (1988). PROBE spatial data modeling and query processing in an image database application. *IEEE Transactions on Software Engineering, 14*, 611–629.

Reiter, R. & Macckworth, A. K. (1989). A logical framework for depiction and image interpretation. *Artificial Intelligence, 41*, 125–155.

Roman, G. C. (1990). Formal specification of geographic data processing requirements. *IEEE Transactions on Knowledge and Data Engineering, 2*, 370–380.

Running, S. W. & Coughlan, J. C. (1988). A general model of forest ecosystem processes for regional applications. I. Hydrologic balance, canopy gas exchange and primary production processes. *Ecological Modelling, 42*, 125–154.

Running, S. W. & Nemani, R. (1991). Regional hydrologic and carbon balance responses of forests resulting from potential climate change. *Climatic Change, 19*, 349–368.

Running, S. W., Nemani, R. & Hungerford, R. D. (1987). Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evapotranspiration and photosynthesis. *Canadian Journal of Forest Research, 17*, 472–483.

Scuderi, L. A. (1993). A 2000-year tree ring record of annual temperatures in the Sierra Nevada. *Science, 259*, 1433–1436.

Smith, J. M. & Smith, D. (1977). Data abstraction: aggregation and generalization. *ACM Transactions on Database Systems, 2*, 105–133.

Stafford, S. G., Brunt, J. W. & Michener, W. K. (1994). Integration of scientific information management and environmental research. In W. K. Michener, J. W. Brunt and S. G. Stafford (Eds), *Environmental Information Management and Analysis: Ecosystem to Global Scales* (pp. 3–19). London: Taylor and Francis.

Weishar, D. & Kerschberg, L. (1991). D/K packets as a means of supporting heterogeneity. *ACM SIGMOD Record, 20*, 69–73.

Wieringa, R., Meyer, J. J. & Weigand, H. (1989). Specifying dynamic and deontic integrity constraints. *Data and Knowledge Engineering, 4*, 157–189.

Worboys, M. F., Hearnshaw, H. M. & Maquire, D. J. (1990). Object-oriented data modeling for spatial databases. *International Journal of Geographical Information Systems, 4*, 369–483.

Ventrone, V. & Heiler, S. (1991). Semantic heterogeneity as a result of domain evolution. *ACM SIGMOD Record, 20*, 16–20.