X WORCAP
20 e 21 de outubro
Instituto Nacional de Pesquisas Espaciais - INPE
São José dos Campos - SP

Towards a Dynamic Geospatial Database Model

Karine Reis Ferreira[1,2], Gilberto Camara[2], Antônio Miguel Vieira Monteiro[2]

[1]Programa de Doutorado em Computação Aplicada – CAP, INPE

[2]Divisão de Processamento de Imagens – Instituto Nacional de Pesquisas Espaciais (INPE) – São José dos Campos – SP – Brasil

{karine,gilberto,miguel}@dpi.inpe.br

Abstract. *This document aims to present: (1) an analysis of distinct geospatial data dynamics and real application demands on representing and querying spatio-temporal data; and (2) a critical review of ten spatio-temporal database models proposed in the literature during the past two decades. This is an initial part of an ongoing work whose objective is to define a new database model, called Dynamic Geospatial Database Model (DyGeo Model), able to represent and query different geospatial data dynamics and so to support different kinds of spatio-temporal applications.*

Keywords: *dynamic geospatial data, spatio-temporal database model, dynamic geographic information systems.*

1. Introduction

The recent technological advances in geospatial data collection, such as Earth observation and GPS satellites, wireless and mobile computing, radio-frequency identification (RFIDs), and sensor networks, have motivated new types of applications which handle spatio-temporal information. Examples include monitoring of animal tracking and oil spill on the ocean, land parcel changes, as well as environmental change monitoring based on satellite images. To meet this demand, it is necessary to represent dynamic geospatial information in spatial databases and geographical information systems (GIS).

Static geospatial information is represented in GIS following well-established ideas. These ideas include object-based and field-based models [Couclelis 1992] [Goodchild 1992], vector and raster data structures, topological operators [Egenhofer and Franzosa, 1991], spatial indexing as well as spatial joins and operations [Rigaux et al., 2002]. In recent years, database management systems (DBMS) have been extended to handle 2D static geospatial information and there has been a major effort to standardize basic components for such data [OGC 2006].

However, there is no consensus on how to represent spatio-temporal information in computational systems. According to Worboys [2005], there are four stages in introducing temporal capacity into GIS and most current proprietary technologies are in stage zero, that is, they do not deal with spatio-temporal information. In GIS literature, there are many proposals of spatio-temporal database models. Nevertheless, Pelekis et al. [2004] consider that most existing models are application-oriented, focusing on particular aspects of spatio-temporal data. So, they are not general enough to be a basis for a new generation of dynamic geographical information systems.

Therefore, this work aims to present: (1) an analysis of distinct geospatial data dynamics and real application demands on representing and querying spatio-temporal data; and (2) a critical review of ten spatio-temporal database models proposed in the literature during the past two decades. This is an initial part of an ongoing work whose objective is to define a new database model, called Dynamic Geospatial Database Model (DyGeo Model), able to represent and query different geospatial data dynamics and so to support different kinds of spatio-temporal applications.

2. Dynamic Geospatial Data

Based on the dichotomy, discrete-objects (geo-objects) and continuous-fields (geo-fields), to represent geospatial data [Couclelis 1992] [Goodchild 1992], dynamic geospatial data can be represented by:

1. Geo-objects which vary over time
    a. Geo-object whose geometry is fixed but its non-spatial attributes change over time;
    b. Geo-object whose geometry changes discretely over time and whose non-spatial attributes also can change;
    c. Geo-objects whose geometries change continuously over time and whose non-spatial attributes also can change;
2. Geo-fields which change over time

Regarding geo-objects which change over time, the difference between discrete and continuous geometry changes is pointed out by Galton [2004] when he explains the difference between *bona fide* and *fiat* object behavior over time. Bona fide objects are grounded in features of physical reality, such as rivers and forest regions, and fiat objects are the artificial products of human cognitive acts, such as municipality limits and land parcels. So, he says "Both these objects might change over time, but typically the bona fide entity will undergo gradual change whereas the fiat entity undergoes sudden change (as a result of the boundary being redrawn from time to time)." In this work, "gradual change" is called continuous change and "sudden change" is called discrete change.

Guting and Schneider [2005] also talked about this difference, saying that "Regarding kinds of changes, a major distinction concerns *discrete* changes and *continuous* changes. Classical research on spatio-temporal database has focused on discrete changes of all the spatial entities. In contrast, the term moving objects emphasizes the fact that geometries change continuously."

In order to illustrate the main features of each geospatial data dynamic presented above, the following sections present four real applications and their demands on representing and querying dynamic geospatial information. They are: (1) Dengue Fever

Monitoring; (2) Municipal Management; (3) Movement Monitoring; and (4) Amazon Deforestation Monitoring.

2.1 Geo-Objects which Change over Time

Universities and research institutes in Brazil have been involved in a cooperative project called SAUDAVEL which aims at building a surveillance system to control, warn and intervene in epidemic and endemic diseases, like **Dengue Fever** and Leptospirosis [Monteiro et al. 2002]. The central experiment of this project is being carried out in a Brazilian city called Recife. Mainly, it consists in giving out egg traps for *Aedes aegypti* and *Aedes albopictus* mosquito in different locations around the city and in counting the number of infected eggs from each trap weekly. Then, this data is processed together with environmental information, resulting in risk maps for public health interventions.

In this first application, each egg trap can be considered as a fixed geosensor, that is, a sensor which collects information at different times associated to a fixed location. The location of each trap does not change over time, only its attributes, such as number of infected eggs. So, each egg trap can be represented by a geo-object whose geometry is fixed but its non-spatial attributes change over time. Besides that, some important queries associated to this application are: (1) *What was the monthly infected egg average for each trap?* (2) *Which month presented the biggest number of infected eggs?* (3) *When and where were more than 80 infected eggs collected by each trap?* (4) *How many eggs were collected in the summer season?* (5) *Which district had the biggest/smallest number of infected eggs?* Figure 1 shows a set of egg traps (represented by red points) and two time series generated by one of them.

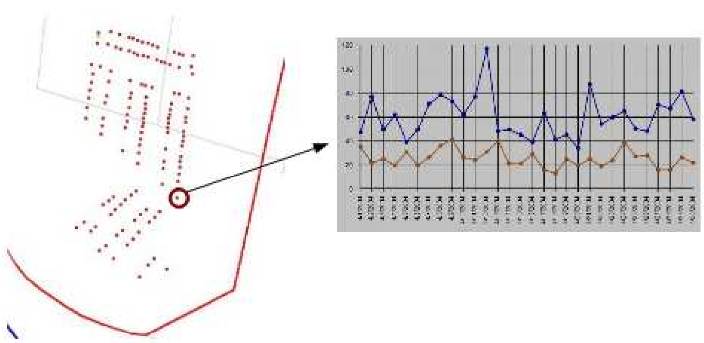

Figure 1. SAUDAVEL egg traps.

Municipal management applications deal with municipality related issues, such as urban land parcels and municipal limit changes. In this application, each urban parcel boundary as well as each municipal limit change discretely over time and their non-spatial attributes can also vary (e. g. the municipal government and the parcel owner). As an instance, Figure 2 shows changes in Rondônia’s municipality limits. In this example, three municipalities “Costa Marques”, “São Francisco do Guaropé” and “Seringueiras” had the same limits from 2001 to 2004, and then, on the first day of 2005 they suddenly changed due to new laws.

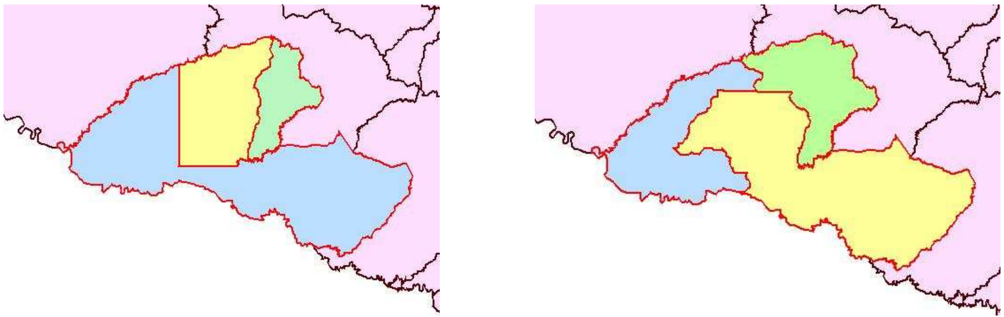

Figure 2. Rondônia's municipality limits in 2001 (left picture) and in 2005 (right picture). Legend: blue polygon is "Costa Marques" municipality; yellow is "São Francisco do Guaropé" and green is "Seringueiras".

In this case, each municipality or each urban land parcel can be represented as a geo-object whose geometry changes discretely over time. Some queries associated to them are: (1) *What municipalities does river r1 cross?* (2) *How many hectares were deforested in each municipality?* (3) *What municipalities were created or extinguished in 2005?* (4) *What land parcels were merged to create parcel p2?* All these questions must consider geo-object variations over time. For example, the first question must return a time series which maps each time to municipalities crossed by river r1 at that time and the second question, a time series which maps each time to deforested hectares for each municipality.

Movement Monitoring refers to applications which monitor and analyze object motions, such as animal, vehicle and person movement. These kinds of applications consist in tracking objects by getting their locations as well as other information such as animal temperature and vehicle velocity at different times. In this case, the object locations vary continuously over time and the concept of trajectory is very important. Some related queries are: (1) *Where was object o1 at time t5?* (2) *When did object o1 enter a specific region r10 and how long did it stay in this region?* (3) *When and where did objects o1 and o2 meet each other (considering a meeting when the distance between two objects is less than 2 meters)?* (4) *Where and when was there a spatio-temporal cluster of objects?*

The Brazilian Amazon deforestation has been monitored since 1988 by INPE through a project called PRODES. It is responsible for calculating Amazon deforestation and for identifying deforested regions in each year through satellite images, by using a well-established methodology [INPE 2008]. Each deforested region evolves continuously and nonlinearly over time and this evolution must be represented in order to allow a specialist to refine its analysis by recognizing patterns of deforested regions [Silva et al. 2005] and how these patterns evolve over time [Motta et al. 2009] [Bittencourt et al. 2008]. A real example of a deforested region evolution is shown in Figure 3.

In this case, each deforested region can be represented by a geo-object whose geometry changes continuously over time and some important queries are: (1) *What was the state of a specific deforested region like in 2003?* (considering that this specific deforested region was observed in 2002 and in 2005, but not in 2003) (2) *What was the area and perimeter variation over time of a specific deforested region?* (3) *How did a* *specific deforested region evolve over time between 2000 and 2008?* (4) *How did the deforested regions that started less than 2 kilometers far from river r1 evolve over time?* (5) *When did a specific deforested region reach municipality x?*

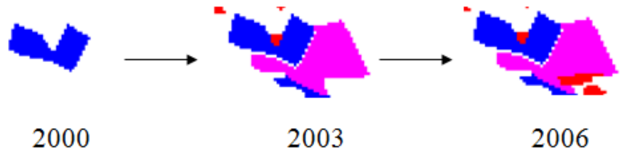

Figure 3. Evolution of a deforested area. Source (INPE, 2008).

2.2 Geo-Fields which Change over Time

Besides the polygonal representation of each deforested region, PRODES project also generates sets of classified images to represent deforestation process. Figure 4 shows an example of the deforestation process in a specific region in Amazon, based on four classified images from different times. These images can be better represented by a geo-field which change/evolve over time since geo-object concept does not exist in this case.

Some important queries associated to it are: (1) *Given a pixel or cell, how has the forest status been varying in this cell over time?* (2) *What was the deforestation in this specific region like in 2001?* (considering that there is no classified image from 2001.) (3) *How many hectares were deforested in this specific region over time?*

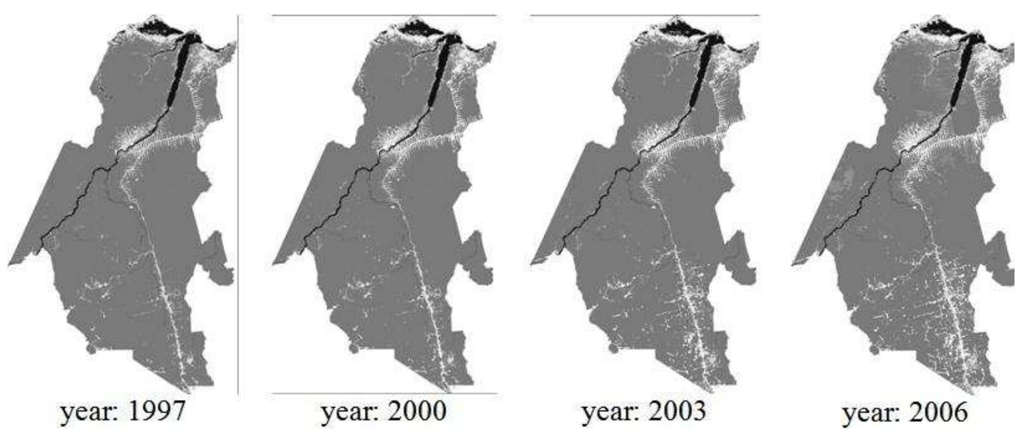

Figure 4. Sequence of four classified images from different years that represent the deforestation process in a specific region in Amazon rainforest. In these images, there are basically three classes: river (dark gray), forest (gray), and deforested area (light gray). Source [INPE, 2008].

3. Spatio-Temporal Database Models: A Critical Review

During the past two decades, many spatio-temporal database models have been proposed in GIS literature. This section presents a critical review of ten models which propose an ontology of space and time and its representation through data types, relationships and operations thereon. They are well-known models which have high number of citations in GIS literature and are shown in Figure 5.

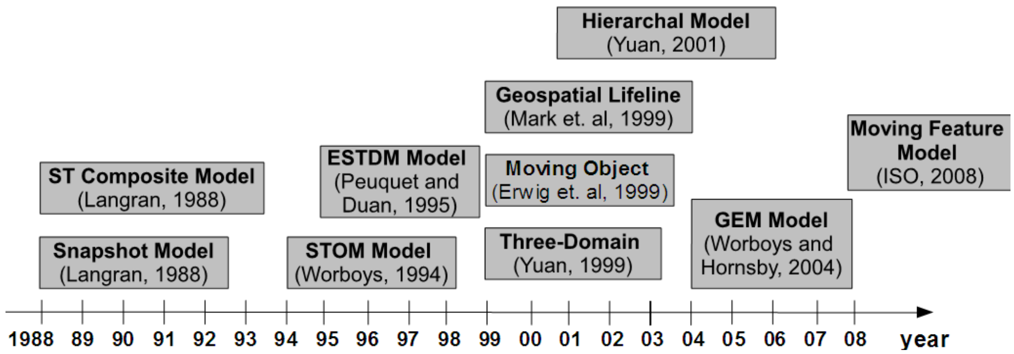

Figure 5. Spatio-temporal database models.

The Time-Slice Snapshot [Langran 1988] is the simplest model of them. This model works with a set of snapshots, where each one is a raster layer which represents a state of the real world at a given time, like a photograph. Each snapshot is a collection of temporally homogeneous units and there are no explicit temporal relations among snapshot. It has two main limitations: (1) operations among snapshots must compare them exhaustively; and (2) redundant storage because a complete snapshot is produced at each time slice, duplicating all unchanged data.

The Space-Time Composite (STC) model [Langran 1988] is an evolution of Snapshot model, by considering vector objects which change over time instead of raster time-slice layers. The mechanics of this model begin with a base layer which represents the objects at some starting time. After this, each change decomposes the space over time into increasingly smaller fragments (objects with geometries) with its own distinct history. Despite being very simple, it is important because it introduces the idea of representing spatial objects which vary over time.

The Unified Spatio-Temporal Object Model (STOM), proposed by Worboys [1994], defines basically two spatio-temporal data types, *ST-simplexes* and *ST-complexes*, and a set of operations over them, such as *ST-Union*, *ST-Intersection* and *ST-Difference*. ST-simplex is an ordered pair<S, T>, where S is a simplex data type and T is a bitemporal element (BTE). A simplex is either a single point, or a finite straight line segment or a triangular area. And BTE is a temporal data type composed of event and transaction time. At last, a ST-complex is a finite set of ST-simplexes. The main disadvantage of the STOM model is not to consider changes in object attributes, that is, in the textual and numerical extents of geographical objects.

The main idea of Event oriented Spatio-Temporal Data Model (ESTDM) [Peuquet and Duan 1995] is to group changes by time of occurrence, ordering changes in locations within a predetermined geographical area. The time associated with each change, called event, is stored in increasing order from initial time t_0 to the latest time t_n. The set of changes C_i recorded for any time t_i consists of the set of each location (x, y) which changed since t_{i-1}, and its new value v. Its two main characteristics are: (1) the events are recorded when changes occur, that is, in any temporal resolution; (2) a value v is recorded only when it is different from the last one found along the scan line. So, this model does not have the two limitations of Snapshot model because it stores only the changed cells by each event. Besides that, it defines a very simple event concept, without exploring concepts related to it, such as, semantics or relationships.

The Three-domain model [Yuan 1999] mainly focuses on how to represent geo-objects which vary over time in a relational database system by using normalized tables and a spatial graph as well as on how to query them by using SQL language. The proposed database schema consists in four tables, one for each domain (semantic, temporal and spatial) and another for the domain link. It can also be implemented in spatial DBMS, as PostGIS and Oracle Spatial, by using its support to deal with spatial information. It is a simple model, without defining spatio-temporal data types and operations. It only uses the data types and query language provided by DBMS.

Moving Object defines a robust algebra, data types and operations, in two levels of abstraction, abstract and discrete, to deal with *moving objects*. Moving Object refers to entities whose geometries change continuously over time, such as, cars, aircraft, ships, mobile phone users, polar bears, hurricanes, forest fires, or oil spills in the sea [Guting and Schneider, 2005]. The authors propose an algebra with two main data types, *moving points* and *moving regions*, and a set of auxiliary types, such as moving real and moving int. Besides that, this algebra defines a set of operators over these data types, such as trajectory, distance, direction, and velocity. Its principal disadvantage is not to consider geo-fields which vary over time.

The Geospatial Lifeline Model [Mark et al. 1999] defines a geospatial lifeline concept which models an individual movement as a time-stamped record of locations. The basic element of lifeline data is a triple *<Id, Location, Time>*, where *Id* is a unique identifier of the individual, *Location* is a spatial descriptor (such as a coordinate pair, a polygon and a street address), and *Time* is the time stamp when the individual was at that particular location. Besides that, this model proposes different types of trajectories or movement approximations, such as, *threads*, *beads*, *necklaces*, and *convex hulls* [Hornsby and Egenhofer 2002]. Depending on the desired granularity and on the application type, distinct types of trajectories are essential. For example, in animal tracking monitoring, the convexhull trajectory is necessary in order to define an animal habitat. So, although this model does not define operations over moving objects, it defines important different types of trajectories. In the Moving Object model only the linear or thread trajectory is extracted from moving points (through the operator trajectory).

The Hierarchal model, proposed by Yuan [2001], provides an interesting way of organizing, using hierarchical layers, dynamic geographical phenomena which posses both field and object characteristics. It is based on a sequence of snapshots called state layers. Therefore, it has redundant storage problem like the Snapshot model. Besides the snapshots, this model also stores the objects which represent the phenomena. These objects are extracted from the state layers. Thus, these two representations of phenomena, geo-fields and geo-objects, are used to improve the spatio-temporal query processing and operations. Finally, this model also defines the concepts of *event* and *process* only to organize the data layers in different levels.

The Geospatial Event Model (GEM), proposed by Worboys and Hornsby [2004], is interesting because it introduces an event concept and relationships between events and geo-objects in a model based on spatial objects. It defines two kinds of relationships, *object-event* and *event-event*, following the idea that an event can affect or be associated to one or more objects or events of different types. Some examples of object-event relationships are *splitting* and *merger* (An event that creates/destroys a boundary between objects). Some examples of event-event relationship are *initiation* and *termination* (The occurrence of event A starts / terminates event B). However, it is a model which defines only data types but not operations over them.

The Moving Feature Model, proposed by the International Organization for Standardization (ISO), defines a conceptual schema for *moving feature* [ISO 2008]. The term feature refers to an abstraction of real world phenomena and moving feature refers to features whose geometries move over time. This schema includes a set of classes, attributes, associations, and operations which provides a common conceptual framework to deal with feature geometry which moves as a rigid body. Therefore, it supports changes of location, translation and rotation of a feature, but not other change types, such as, the feature deformation and changes in non-spatial attributes of a feature. The main advantage of this model is to define a generic type called one-parameter geometry which represents the variation of feature geometry with respect to any single variable, such as pressure, temperature, or time. However, its main disadvantages are not to consider feature geometry deformation and changes in feature non-spatial attributes. For instance, due to these limitations, it is not possible to represent an oil slick moving on the ocean.

4. CONCLUSION

This work presents: (1) an analysis of distinct geospatial data dynamics and real application demands on representing and querying spatio-temporal data; and (2) a critical review of ten spatio-temporal database models proposed in the literature during the past two decades. This work reviews ten models which propose an ontology of space and time and its representation through data types, relationships and operations thereon. They are well-known models which have high number of citations in GIS literature.

This is an initial part of an ongoing work whose objective is to define a new database model, called Dynamic Geospatial Database Model (DyGeo Model), able to represent and query different geospatial data dynamics and so to support different kinds of spatio-temporal applications. We believe that the study presented in this work is crucial before building the DyGeo model.

References

Bittencourt, O.; Camara, G.; Vinhas, L.; Motta, J. (2007) *Rule-based Evolution of Typed Spatio-temporal Objects*. In: IX Brazilian Symposium in Geoinformatics (GeoInfo), Campos do Jordão.

Couclelis, H. (1992) *People manipulate objects (but cultivate fields): beyond the raster-vector debate in GIS*. In: FRANK, A. U.; CAMPARI, I.; FORMENTINI U. (Eds). Theory and Methods of Spatio-Temporal Reasoning in Geographic Space. Berlin: Springer-Verlag, pp. 65–77.

Egenhofer, M. and Franzosa, R. (1991) *Point-Set Topological Spatial Relations*. International Journal of Geographical Information Systems, v. 5, pp. 161-174.

Galton, A. (2004) *Fields and Objects in Space, Time and Space-Time*. Spatial Cognition and Computation Journal, v. 4, n. 1, pp. 39-68.

Goodchild, M. F. (1992) *Geographical Data Modeling*. Computers and Geosciences, v. 18, pp. 401-408.

Guting, R. H. and Schneider, M. (2005) *Moving Objects Databases*. San Francisco, CA: Morgan Kaufmann, p. 389.

Guttag, J. V.; Horowitz, E.; Musser, D. R. (1978) *Abstract Data Types and Software Validation*. Communications of ACM, v. 21.

Horsnby, K. and Egenhofer, M. (2000) *Identity-Based Change: A Foundation for Spatio-Temporal Knowledge Representation*. International Journal of Geographical Information Science, v. 14, pp. 207-224.

Hornsby, K. and Egenhofer, M. J. (2002) *Modeling Moving Objects over Multiple Granularities*. Annals of Mathematics and Artificial Intelligence. v. 36, n. 1-2, pp.177-194.

INPE. (2008) *Monitoramento da Floresta Amazônica Brasileira por Satélite (Monitoring the Brazilian Amazon Forest by Satellite)* [online]. Report, São José dos Campos: INPE, Available from: http://www.obt.inpe.br/prodes [Accessed 15/05/2010].

International Standard Organization (ISO). (2008) *Geographic information — Schema for moving features* (ISO 19141).

Langran, G.; Chrisman, N. R. (1988) *A Framework For Temporal Geographic Information*. The International Journal for Geographic Information and Geovisualization, v. 25, n. 3, pp. 1-14.

Mark, D.; Egenhofer, M. J.; Bian, L.; Hornsby, K.; Rogerson, P.; Vena, J. (1999) *Spatio-temporal GIS analysis for environmental health using geospatial lifelines*. In: 2nd International Workshop on Geography and Medicine (GEOMED'99). Proceeding....

Monteiro, A. M. V.; Carvalho, M. S.; Assunção, R.; Vieira, W.; Ribeiro, P. J.; Davis, C.; Regis, L. (2004) *SAUDAVEL: Bridging the Gap between Research and Services in Public Health Operational Programs by Multi-Institutional Networking Development and Use of Spatial Information Technology Innovative Tools*. Available at: <http://www.dpi.inpe.br/saudavel/documentos/ArtigoSAUDAVELAgo2004.pdf>. Access at: 27/05/2010.

Motta, J.; Camara, G.; Escada, I.; Bittencourt, O.; Vinhas, L. (2009) *Case-Based Reasoning for Eliciting the Evolution of Geospatial Objects*. Conference on Spatial Information Theory COSIT 2009, Aber Wrach, France.

Open Geospatial Consortium (OGC) (2006) *OpenGIS Implementation Specification for Geographic information - Simple feature access - Part 1: Common architecture*. OGC, 2006. 95 p. (OGC 06-103r3). Available at: <www.opengeospatial.org/>. Access at: 27/05/2010.

Pelekis, N.; Theodoulidis, B.; Kopanakis, I.; Theodoridis, Y. (2004) *Literature Review of Spatio-Temporal Database Models*. The Knowledge Engineering Review, v. 19, n. 3, p. 235-274.

Peuquet D. J. and Duan, N. (1995) *An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data*. International Journal of Geographical Information Science, v. 9, n. 1, pp. 7-24.

Rigaux, P.; Scholl, M.; Voisard, A. (2002) *Spatial Databases: With Application to GIS*. San Francisco, USA: Morgan Kaufmann.

Silva, M.; Camara, G.; Souza, R. C.; Valeriano, D.; Escada, I. (2005) *Mining Patterns of Change in Remote Sensing Image Databases*. Fifth IEEE International Conference on Data Mining. Houston,TX, USA.

Worboys, M. F. (1994) *A Unified Model for Spatial and Temporal Information*. The Computer Journal. v. 37, n. 1.

Worboys, M. and Hornsby, K. (2004) *From Objects to Events: GEM, the Geospatial Event Model*. Lecture Notes in Computer Science, v. 3234, pp. 327-343.

Worboys, M. (2005) *Event-oriented approaches to geographic phenomena*. International Journal of Geographical Information Science, v. 19, n. 1, pp. 1-28.

Yuan, M. (1999) *Use of a Three-Domain Representation to Enhance GIS Support for Complex Spatio-temporal Queries*. Transaction in GIS, v.3 n.2, pp.137-159.

Yuan, M. (1991) *Representing Complex Geographic Phenomena in GIS*. Cartography and Geographic Information Science, v. 28, n. 2, pp.83-96, 2001.oulic, R. and Renault, O. (1991) "3D Hierarchies for Animation", In: New Trends in Animation and Visualization, Edited by Nadia Magnenat-Thalmann and Daniel Thalmann, John Wiley & Sons ltd., England.