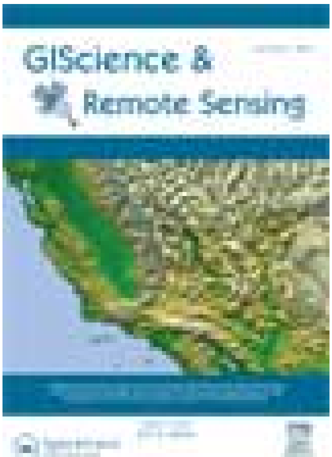
This article was downloaded by: [University of Hawaii at Manoa]
On: 25 December 2014, At: 23:21
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



GIScience & Remote Sensing
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tgrs20

Building a Dynamic, Large-Scale Spatio-temporal Vector Database to Support a National Spatial Data Infrastructure in China

Huibing Wang [a], Tong Zhang [a], Xinming Tang [b] & Yaolin Liu [a]
[a] Wuhan University
[b] Chinese Academy of Surveying and Mapping, Beijing, China
Published online: 15 May 2013.

To cite this article: Huibing Wang, Tong Zhang, Xinming Tang & Yaolin Liu (2010) Building a Dynamic, Large-Scale Spatio-temporal Vector Database to Support a National Spatial Data Infrastructure in China, GIScience & Remote Sensing, 47:1, 135-162

To link to this article: http://dx.doi.org/10.2747/1548-1603.47.1.135

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Building a Dynamic, Large-Scale Spatio-temporal Vector Database to Support a National Spatial Data Infrastructure in China

Huibing Wang
School of Resource and Environmental Science, Wuhan University 129 Luoyu Road, Wuhan, China 430079 and Key Laboratory of Geo-informatics, State Bureau of Surveying and Mapping, Chinese Academy of Surveying and Mapping, 16 Beitaiping Road, Beijing, China 100039

Tong Zhang[1]
State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China 430079

Xinming Tang
Key Laboratory of Geo-informatics, State Bureau of Surveying and Mapping, Chinese Academy of Surveying and Mapping, 16 Beitaiping Road, Beijing, China 100039

Yaolin Liu
School of Resource and Environmental Science, Wuhan University 129 Luoyu Road, Wuhan, China 430079

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

Abstract: This paper introduces a comprehensive approach to support dynamic, large-scale spatio-temporal vector database development. The approach is based on a version-difference spatio-temporal data model that targets common problems in developing large-scale spatio-temporal vector databases. The scenario is to establish a National Fundamental Geospatial Information Dynamic Database (NFGIDD) in China. The proposed methodology was demonstrated by an experimental spatio-temporal database system that consists of a database management platform and a data processing system. The experimental results show that our methods have significant advantages over traditional spatio-temporal data models.

INTRODUCTION

The management of spatio-temporal data is common in geospatial applications. However, regular geographic information systems (GIS) have limited functions to support effective spatio-temporal data queries and analysis. Research on spatio-temporal

[1]Corresponding author; email: geogrid@gmail.com

GIScience & Remote Sensing, 2010, **47**, No. 1, p. 135–162. DOI: 10.2747/1548-1603.47.1.135
Copyright © 2010 by Bellwether Publishing, Ltd. All rights reserved.

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

databases is therefore necessary and critical. Despite a few innovative approaches to spatio-temporal database development (Claramunt and Theriault, 1995; Griffiths et al., 2005; Pultar et al., 2009), there are many technical challenges when integrating temporal and spatial data seamlessly in practical applications. Spatio-temporal databases deal with a wide range of issues including spatio-temporal models, dynamic temporal queries, indexing, and representation issues. At the core of spatio-temporal database research is the development of spatio-temporal data models. Most research efforts have been made to design and develop scalable and efficient spatio-temporal data models (Parent et al., 1999; Griffiths et al., 2004; Beard, 2006). However, little attention has been paid to the implementation issues of spatio-temporal databases. The development of a real-world spatio-temporal database dedicated to real-world geospatial applications may be challenging due to insufficient practical experience and the inherent drawbacks of proposed data models. It is even more challenging to develop large scale spatio-temporal databases such as national geospatial databases.

The development of a national spatial data infrastructure in China may create challenges in data integration, indexing, representation, and most importantly the design of spatio-temporal data models given a large amount of heterogeneous spatio-temporal data. For example, when building a large-scale efficient spatio-temporal database, we may have massive data from different sources created at different historical times. The very first task is to build relations for the same spatial objects from different datasets so as to integrate these data seamlessly. Indexing and version management are critical as well, inasmuch as the query and visualization of massive spatio-temporal data usually are computationally intensive.

This paper differs from other studies of spatio-temporal databases, which have focused on conceptual issue, as it places emphasis equally on the introduction of a spatio-temporal data model as well as the implementation of the proposed model. Implementation of spatio-temporal data models are rare compared to numerous conceptual discussions of spatio-temporal models. This paper seeks to develop scalable and dynamic spatio-temporal vector databases to support the National Spatial Data Infrastructure (NSDI) in China. A real-world geospatial task is to design and build a spatio-temporal vector database for NSDI in China. Obviously, the overwhelming data volume as well as management challenges call for a new comprehensive solution. Our attempts to develop a comprehensive approach are introduced. The approach is based on a version-difference data model that targets common problems in developing a large-scale spatio-temporal vector database. The base version-difference model is augmented by advanced methods to establish dynamic spatio-temporal relations. Dynamic relation techniques are proposed to solve data fusion problems for multiple-version data that are generated from update management operations. Changes can be identified instantly and the inheritance relations for geographic objects are explicitly stored. The methodology is demonstrated by an experimental spatio-temporal database system that consists of a database management platform and a data processing system. Two typical experiments were conducted to examine the effectiveness of the proposed methodology and the prototype system. The experimental results show that our methods have significant advantages over traditional spatio-temporal data models and can be used to develop a large-scale national geospatial database.

SPATIO-TEMPORAL MODELS AND DATABASES

The trend of incorporating temporal information into a spatial database has produced a large number of innovative approaches to building a spatio-temporal database effectively and efficiently. The modeling of temporal information has gained significant considerations in these approaches in comparison with early approaches, which only regarded time-based information as a supplemental component to spatial databases. These early efforts include the Snapshot Model (Armstrong, 1988) and Space-Time Composite Model (Langran and Chrisman 1988). The snapshot model encompasses temporal information with multiple time-stage layers, and each layer is a "snapshot" of time-varying geographical information. The Space-Time Composite Model describes temporality based on individual spatial objects by recording attribute changes. Both models have inherently negative aspects, such as data redundancy and representation problems (Yuan, 1996a).

Subsequent progress has been made to propose event-based (Peuquet and Duan, 1995; Worboys, 2005) and object-oriented data models (Worboys, 1992; Wachowicz and Healey, 1994; Rapper and Livingstone, 1995; Lohfink et al., 2007). Event-based models focus on changes rather than spatial objects per se. Time-staged event series can represent where and when changes take place. Object-oriented data models rely on identifiable spatio-temporal objects and associated attributes as well as relations to model spatio-temporal process following the object-oriented programming paradigm. In parallel, domain-based models are developed to represent semantical, temporal, and spatial objects or conceptual/thematic description separately (Smith, 1994; Yuan, 1994, 1996b; Claramunt and Theriault, 1995). Domain-based models dynamically link separate domains when simulating various changes, thus claiming to have advantages of being highly flexible and dynamic in capturing spatio-temporal changes.

Other models include the history graph model (Renolen, 1996), the spatio-temporal entity-relationship model (Tryfona, 1997), the object-relationship model (Claramunt et al., 1998), and the moving object models (Erwig et al., 1999), as summarized by Pelekis et al. (2004) who also provide a thorough comparison of these models from the perspectives of temporal, spatial, and spatio-temporal semantics as well as query capabilities. Recently, process-based models and ontologies have been proposed to explicitly describe process information that is central to the representation, query, and analysis of spatial-temporal changes (Taciana and Gilberto, 2004; Reitsma and Albrecht, 2005). To summarize, a good spatio-temporal data model must be able to model complex spatio-temporal events and processes, to support efficient spatial and temporal queries, and to manipulate massive spatial-temporal datasets.

Many models are largely conceptual without implementation. Pelekis et al. (2004) have reviewed common spatio-temporal data models and find that less than half of existing models have been implemented. Very few prototypes of spatio-temporal databases are available, not to mention real-world success stories of building complete, reliable, and scalable spatio-temporal databases. The lack of implementational efforts in spatio-temporal database research may be partially due to a variety of technical challenges in developing spatio-temporal databases, such as data preprocessing, indexing, management systems, user interface, query functions, and database maintenance. In comparison with theoretical research (primarily on spatio-temporal data models), real development and implementation issues pose more challenges because the goal is to

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

develop a usable real-world application. Another reason, as Paton et al. (2000) point out, is that most previous research efforts only focus on a small portion of the whole picture, creating a gap between theoretical and practical studies. Even when every small aspect is significantly improved, an operational spatio-temporal database application that consists of a number of components may not function as expected.

Based on progress in the research on spatio-temporal data models, it is possible to develop usable spatio-temporal databases for real-world applications. Early demonstrations were mostly built as extensions to existing GISystems. Such prototypes rely on data storage, visualization, and analysis functions in GISystems to support temporal applications. As Abraham and Roddick (1999) contend, most of these, developed for specific applications, are not appropriate for large scale spatio-temporal database development. Also, only a small number of spatio-temporal data models has been formalized and implemented (Pelekis et al., 2004). Of those having been implemented, most are experimental prototypes (e.g. TEMPEST, MADS, Geo-OM). Technical challenges in modeling spatio-temporal changes hamper the wide adoption of spatio-temporal databases in real-world applications.

A large scale spatio-temporal database application should be able to: (1) support composite spatio-temporal queries in an efficient way; (2) implement dynamic and scalable indexing techniques; (3) premit convenient data updating; (4) represent the evolution of spatio-temporal objects/events effectively; and (5) provide friendly graphical user interfaces. There is no optimal model for all spatio-temporal database applications—each of which has different requirements in modeling, storing, and representing spatio-temporal objects/events. In addition to conventional conceptual research, application-oriented research has gradually gained attention among spatio-temporal database researchers. It is quite reasonable to investigate spatio-temporal database issues from the perspective of real-world applications. This may help identify the strengths and weaknesses of each current data model for a specific application and propose new approaches for addressing remaining issues.

CHALLENGES IN DEVELOPING A LARGE-SCALE SPATIO-TEMPORAL VECTOR DATABASE

A large-scale spatio-temporal database accommodates an extremely large number of spatio-temporal entries and occupies extremely large storage space. The establishment of NSDI involves the development of a large-scale spatio-temporal database. There have been numerous scientific studies and engineering development projects on massive imagery databases (John et al., 1994; Barclay et al., 1999; Ageenko and Fränti, 2000; Castelli and Bergman, 2002), whereas vector-based spatio-temporal databases are rare. This paper primarily focuses on vector database issues.

Spatio-temporal vector data have a series of particular characteristics such as: (1) spatial-temporal dependence and heterogeneity; (2) implications introduced by scale; (3) diverse reference frameworks; (4) representation challenges for dynamic spatio-temporal objects/events; and (5) other special properties such as spatial metadata, security, uncertainty, diverse spatial analysis strategies, and high reliance on visualization.

Incorporating the time dimension to capture geographic dynamics and processes further compounds the challenges for spatio-temporal data organization, representation, and analysis. The volume of spatio-temporal data is usually much larger than traditional spatial datasets due to the incorporation of historical data as a result of frequent update operations. Advanced mechanisms are needed to reduce data redundancy, to enhance access efficiency, and to extract interesting spatio-temporal patterns. In addition to common spatial analysis tools, temporal or spatio-temporal analysis functionalities should be developed and deployed in spatio-temporal vector databases. All of the aforementioned special issues will in turn be reflected in the design and implementation of spatio-temporal databases. Large-scale, high-dimensional spatio-temporal vector databases require special considerations in physical database design, including the data structure of spatio-temporal datasets, the description of dynamic temporal changes, approaches to eliminate data redundancies, efficient indexing techniques, scalable data query capabilities, and appropriate representations. Obviously, the model of spatio-temporal changes and historical data storage, new data processing and change detection, as well as representation of spatio-temporal changes are all essential components of large-scale dynamic spatio-temporal vector databases. Unlike conventional spatio-temporal data model research, implementation issues such as data production/processing tools to ensure the data integrity, consistency, and completeness are important as well.

Globally speaking, the establishment of NSDI since the 1990s has produced many tangible products and services that enable the sharing of geospatial data at different administrative levels in many counties. However, technical challenges of data standards, metadata, and huge data volumes, along with institutional and policy issues are impeding the widespread deployment of NSDI. Most technical challenges are concerned with the development and maintenance of large-scale dynamic spatio-temporal databases. It is encouraging to develop a large-scale dynamic geospatial database framework in which key technical issues are addressed. The common challenges of very large databases demand creative and novel approaches covering the topics including database indexing (Assent et al., 2008; Jin et al., 2009), database performance (Duan et al., 2009), caching (He and Luo, 2008), concurrency control (Borkar et al., 2009), database administration (Mecca et al., 2009), database search/query (Zhang et al., 2009), and other related issues.

Here we introduce a scenario for the building of China's National Fundamental Geospatial Database. The overall goal is to solve the problems of historical GIS data management in the development and management of China's National Dynamic Digital Geospatial Data Infrastructure. The National Fundamental Geospatial Databases of China are critical components of the NSDI in China. The databases collect nationwide fundamental geospatial data at multiple scales, including topographic data, hydrographic data, transportation networks, gazetteer and residential area information. These datasets integrate digital raster graphics (DRG), digital elevation models (DEM), digital vector maps (similar to USGS DLG), gazetteer data, as well as digital orthophoto maps. Since 1994, the State Bureau of Surveying and Mapping (SBSM) of China has been leading the development of the NSDI of China, and especially of the National Fundamental Geospatial Databases. All of these database projects have progressed smoothly and have produced numerous well-organized digital databases serving the needs of socio-economic development. However, China's rapidly growing economy has introduced dramatic landscape and environmental changes in both rural and metropolitan areas. These changes have occurred at a very fast rate, thus posing

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

challenges for updating the National Fundamental Geospatial Databases. Among all the components of National Fundamental Geospatial Databases, it is challenging to develop an efficient updating mechanism for digital vector maps.

In order to ensure data currency, China's National Fundamental Geospatial Databases have undergone frequent data updating operations and have produced a large amount of historical geospatial data. Historical geospatial data have huge potential for applications in trend analysis and decision making. For example, historical cadastral data can be used for land evaluation and transactions. One serious issue is that traditionally historical data were stored separately from current datasets without any relations between different versions of datasets. This created obstacles in using historical data effectively because changes between different states in time are hard to detect.

Historical geospatial data management is central to dynamic digital vector map databases. A well-functioning historical geospatial data management system should permit rapid and reliable geospatial data updating, support efficient queries of temporal changes, and represent dynamic spatio-temporal objects/events/processes effectively. In response to these challenges, SBSM launched a research project aiming at promoting the efficiency and effectiveness of dynamic data management for digital vector maps in the National Fundamental Geospatial Databases (or the "National Fundamental Geospatial Information Dynamic Database [NFGIDD] as they were officially named in 2006). The project aims to design a generic spatial-temporal data model providing a comprehensive framework to monitor, analyze, and forecast spatial-temporal changes over time for a large-scale geospatial database. The model is required to support both fundamental and thematic geospatial information management. Other tasks include the development of a spatio-temporal data management platform and data processing system with an easy-to-use graphical interface. The operational systems can establish dynamic relations between historical data and current data for the NFGIDD and visualize geospatial information changes/evolutions at multiple scales. The overall goal aims at minimizing data redundancy and supporting advanced/rapid spatio-temporal query and analysis. Spatio-temporal changes can be detected by semi-automated techniques (with greater than 85% accuracy). In addition, data processing, database management, and data analysis are important tasks for the NFGIDD project.

METHODOLOGY

In response to the challenges of efficient management for digital vector maps in the National Fundamental Geospatial Databases, we propose a comprehensive methodology, ranging from spatio-temporal data models, data processing, database management, and data storage to spatio-temporal data analysis (Fig. 1). This methodology works as a solution to the routine management and updating of China's National Fundamental Geospatial Databases. The basic idea is to build a multiple-version database at different time stages in which historic changes are considered to be "differences" between versions.

The Version-Difference Model

Spatio-temporal data models play a central role in the entire dynamic database solution (Fig. 1). A version-difference spatio-temporal data model is proposed. This

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

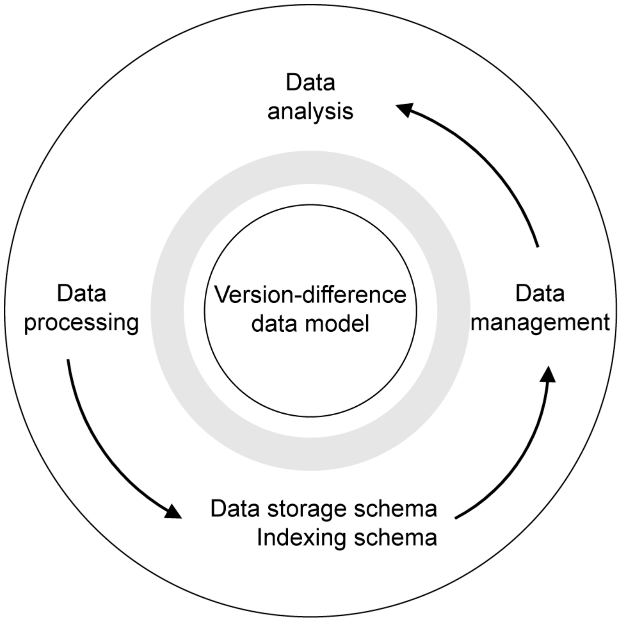


Fig. 1. Components of the comprehensive spatio-temporal database solution.

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

new data model is capable of deriving implicit data at any point in time from explicitly stored data by means of a series of spatio-temporal operations. Compared to the snapshot model, the version-difference model does not store all geospatial data once changes happen. Only base state, current state, along with real changes are recorded. All unchanged spatial objects only have one copy in the database. Each change is labeled with a unique ID. At the conceptual level, this model has three principle entities: *states, events*, and spatial-temporal *relations*. Both valid-time and transaction-time are recorded for each spatial object, thus making the version-difference model bitemporal to better support flexible temporal queries. Figure 2 shows the three principle entities of the version-difference model. *State* is a relatively stable concept, describing time periods during which objects remain unchanged. *Events* describe notable changes of states that occur in a particular time point or in a time duration. And spatio-temporal *relations* establish connections between different states by recording changes between two temporal states.

Events and States. During its lifespan, a spatio-temporal object may have multiple *states*, whereas *events* describe change processes between states. The state of a spatial-temporal object can be categorized into attribute state and spatial state, and the spatial state in turn can be classified as a spatial topological state and a spatial geometry state.

Dynamic Spatial-temporal Relations. In a spatio-temporal database, changes can be monitored and traced if dynamic spatial-temporal relations are explicitly stored. In addition to thematic attribute changes for spatial objects, we emphasize existential changes and spatial property changes. Existential changes describe the event of a spatial object appearing and disappearing. Spatial property changes are concerned with

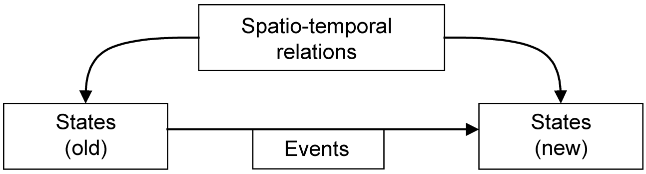


Fig. 2. Three principal entities of the version-difference model.

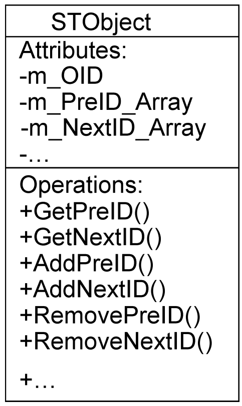


Fig. 3. STObject class.

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

location, shape, size, and orientation changes. These changes may involve dynamic spatio-temporal relations. A disappearing object may produce two new objects and this event indicates the two new objects have inheritance relations with the disappearing object.

A spatio-temporal object (ST-Object) can have multiple predecessors and successors. Figure 3 describes a ST-Object class to represent ST-Object (m_OID as unique ID for STObject) in the real world (an ST-Object class may have other member variables and functions, such as temporal and geometric methods). We use ST-Object relation links to record ST-Object's predecessors and successors. Relation links are derived from the sets of predecessors (m_PreID_Array) and successors (m_NextID_Array). Other regular operations to fetch, add, and delete predecessors and successors are designed as well.

An ST-Object's predecessor or successor may have its own predecessors and successors. When we track one ST-Object over time through relation links, the historical sequence of a change may look like a tree (relation tree, Fig. 4). The information of the relation link and relation tree is useful for tracking ST-Object evolution at various time stages.

Based on spatio-temporal relation trees, we can develop algorithms to detect ST-Object changes over time. Suppose we have input data in the form of two vector maps of the same area at different time stages. Let MAP_0 be the old time vector map and MAP_1 be the current vector map. Let OBJ_0 be an object in MAP_0 and OBJ_1 be an object in MAP_1, accordingly. In order to accelerate the computation speed of these

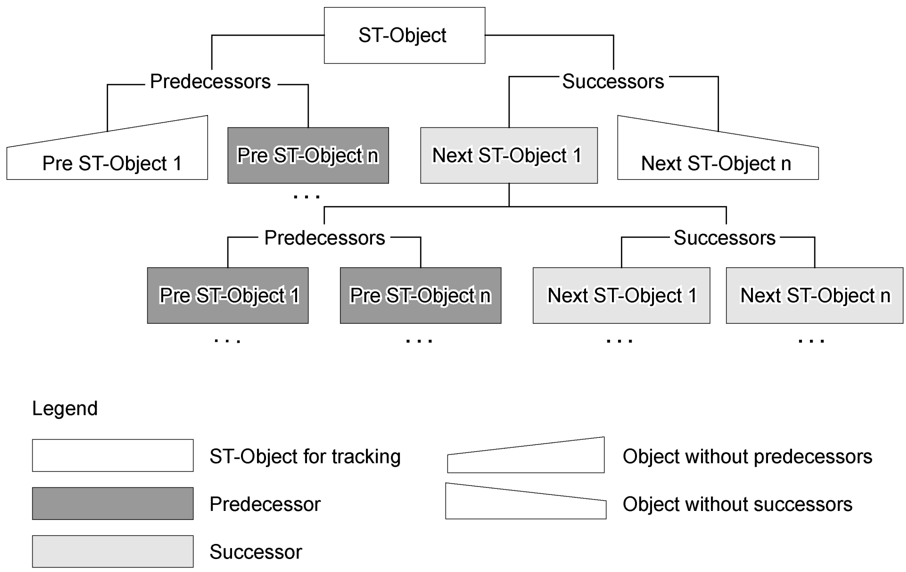


Fig. 4. ST-object relation tree.

algorithms, we first build a regular grid index for the input vector maps. Each object is assigned a grid cell with respect to its coordinates. The algorithms are based on the identity of the ST-Objects. We divide ST-Object changes into four types.

1. BORN: A new ST-Object with new identity appears at time *t*. This change type can be formalized as a numerical expression 0:1.

2. DIE: An ST-Object that existed previously with its own identity disappears at time *t*. This is the inverse type to BORN, and its numerical expression is 1:0.

3. ALTER: An ST-Object changes to another one, becomes a portion of a new ST-Object, or splits into more ST-Objects at time *t*. This type includes three subtypes: ΔALTER (1:1), in which one ST-Object changes to another ST-Object with a new identity; ΔALTER (1:n), in which an ST-Object changes to two or more ST-Objects; and ΔALTER (n:1), in which two or more ST-Objects merge into one ST-Object.

4. KEEP: An ST-Object remains unchanged. This is a special example of the ALTER type.

Below we explain the algorithms for change detection.

Point relation algorithm. Let Δd be the distance of OBJ_0 and OBJ_1; if Δd = 0, then the change type is KEEP (1:1); if Δd < threshold, then the change type is ALTER (1:1); if Δd > threshold, then OBJ_0 is DIE (1:0), OBJ_1 is BIRTH (0:1). If more than one OBJ_0 matches OBJ_1, the change type is ALTER (n:1). If more then one OBJ_1 matches to OBJ_0, the change type is ALTER (1:n).

Line relation algorithm. First, compare the MBRs (minimum boundary rectangle) of OBJ_0 and OBJ_1. If the MBRs intersect, use an Extended Douglas-Peucker Compressing Algorithm to obtain line objects' characteristic nodes. Then, use the point relation algorithm to match characteristic nodes. If definite characteristic nodes match, OBJ_0 and OBJ_1 are considered to be related, or in other words ALTERED. If OBJ_0 is

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

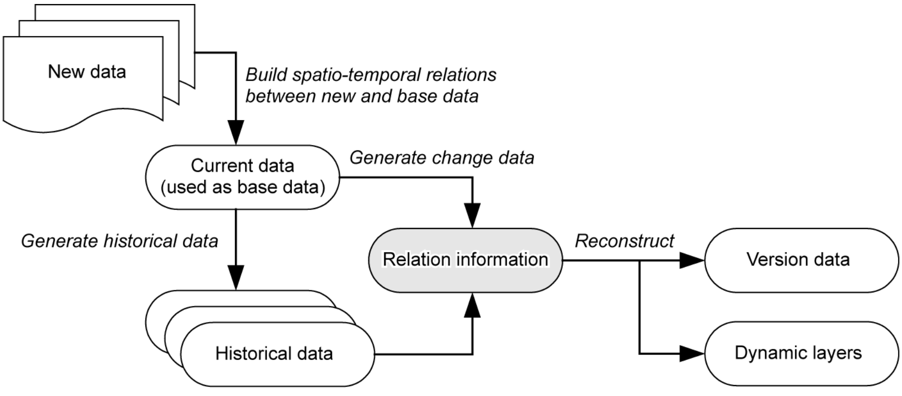


Fig. 5. Logical design of the version-difference data model.

not related to OBJ_1, the change type is DIE (1:0); if OBJ_1 is not related to OBJ_0, the change type is BIRTH (0:1).

Polygon relation algorithm. First, compare the MBRs of OBJ_0 and OBJ_1. If the MBRs intersect, execute the polygon-rasterizing algorithm. OBJ_0 and OBJ_1 are considered to be related (ALTER) if they have definite matched raster pixels; if OBJ_0 is not related, DIE (1:0); if OBJ_1 is not related, BIRTH (0:1).

Manual Detection

The algorithms above can detect most changes and build dynamic relations automatically. However, some detection results may be erroneous. This method can be complemented by human manual editing tools to rectify automated results of spatio-temporal data relation detection.

The change detection procedure imports new data and runs detection algorithms, along with manual editing tools to build dynamic relation data.

Version and Difference

Versioning, as the name implies, explicitly records versions of individual features and objects as they are altered, added, and deleted at different states. A version explicitly records each state of a feature or an object as a row in a table along with important transaction information.

The version-difference model conceptually incorporates three versioning types: (1) on-line or on-the-fly versioning: new versions are reconstructed by other versions and differences from databases when needed; (2) stand-alone or snapshot versioning: users can create a version that includes all data belonging to a specific state at a given point in time (this versioning has a data redundancy issue); (3) currency or default versioning: versioning for the current time state. Figure 5 describes the logical design of the version-difference data model.

Versioning rules are listed as follows: (1) only owners can rename, create, modify, or delete a version; (2) versioning does not take effect until commit operations are

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

performed; (3) each dataset has a default version; (4) owners can restrict privileges of other users; (5) the database should have mechanisms of version verification to prevent updating errors. As for "difference," it is obtained by comparing data from any given two temporal states. Difference tables are created to keep "difference" information in storage. First, object change relations between two states are stored. It is straightforward to record thematic changes if no spatial/existential change is involved. Spatial property changes involve tracking geometry/location differences. For existential changes, we can record: $0 \rightarrow 1$ (object appears), $1 \rightarrow 0$ (object disappears), $1 \rightarrow N$ (the number of object increases from 1 to N), and $N \rightarrow 1$ (the number of objects decreases to one). Difference tables also store dynamic spatio-temporal relations between old and new ST-Objects.

The basic features of the version-difference model are:

1. Minimal storage size. For the version-difference model, we do not need to store all information of every state in the area of interest. Only data for the base state and changes between different states are saved. Each change is labeled with a unique state ID (state ID is incremental) and a changed object is recorded with its predecessors and successors. Datasets may look different at different states.

2. Easy to access data at any point in time. The model provides schema (called version management) to generate data at any point in time using reconstruction operators on the base state.

3. Efficient tracking of spatio-temporal object changes. Because a changed object is recorded with its predecessors and successors, the states of spatio-temporal objects can be easily accessed. Thus, the model can be used for spatio-temporal analysis and forecasting.

Database Storage Schema Design

The underlying driving force for designing the version-difference model is to achieve intricate balance between effective dynamic data management and efficient data storage schema. The first goal, minimizing data redundancy, requires the use of a minimum number of databases to capture all required spatio-temporal dynamic relations and changes. In the version-difference spatio-temporal data model, we designed a four-database schema to realize this goal. The four databases are the Current Database, Change Database, Difference Database, and Version Database.

Current Database. The current database stores the entire spatial-temporal dataset of the current state. Because the current dataset is accessed frequently, we made it as the base state in order to improve database performance. ST-Objects in the Current Database are "activated" as default objects for user manipulation. New data can be processed and stored in the Current Database.

Difference Database (Historical Database). The Difference Database is responsible for storing difference data between historical states and current state. ST-Objects, if remaining the same for both spatial and thematic attributes, would not be saved in the Difference database. The latest state of an object will be stored in the current database once it was changed in an event. Previous changes of states/processes/events will be stored in the Difference Database. Events should be arranged in time order. Differences between the current and historical states will be stored when the conditions or constraints are all met. Given a specific time, it is easy to obtain a previous

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

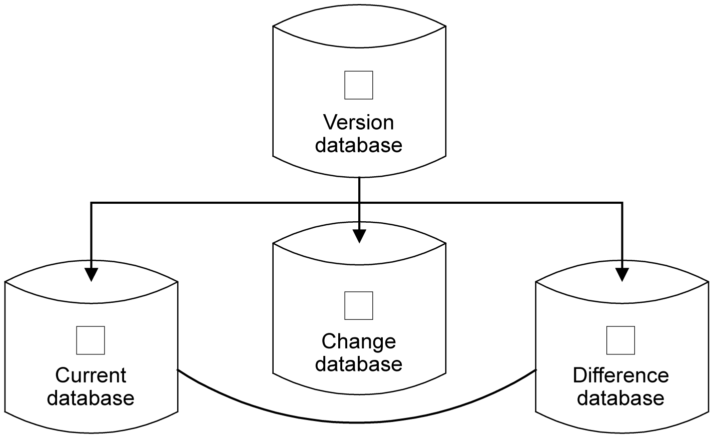


Fig. 6. The four-database storage schema.

state and roll the current spatial-temporal relations back to specified historical point in time.

Change Database. Changes are subject to spatial and temporal constraints, which are preconditions to limit, stop, or trigger a change. Some changes may be unlikely to take place in reality and will be labeled as "fake changes." When "fake changes" occur, the Change Database should discard the change and roll back. The Change Database only stores the information of ST-Object change types, the dynamic relations, and IDs for ST-Objects that changed. No spatial information is stored in the Change Database.

Version Database. A version is a snapshot of the dataset at a given point of time. Users can browse the spatial-temporal data of any time instance and assign it as the default version in the Version Database. It can be regarded as a base state for other version datasets whose time stages are earlier than the new default base version.

Figure 6 shows the relationships among the four logical databases. The relation of the Current Database and Difference Database is stored in the Change Database. The Version Database is created by the Current Database, the Difference Database, and the dynamic relations in the Change Database. In short, the Current Database stores current datasets as the base state; the Difference Database stores the historical dataset of changed objects; the Change Database records the process of history changes; and the Version Database is used to store snapshot datasets at times specified by users.

Each ST-Object is associated with two types of time: valid time and transaction time. The valid time captures the time at which a fact is true in the real world. It is typically represented as an interval [VTB, VTE], which means that the fact is valid from time VTB to the time VTE in the real world. Transaction time is the time at which a transaction happens to save new attribute values or spatial changes in the database. The time interval [TTB, TTE] represents the time span during which changes of a particular transaction are current in the database. Valid and transaction time are orthogonal and each could be independently recorded.

ST-Object relation links are recorded in four tables: the table of relation links, change events, current entity states, and historical entity states. Current entity tables and

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

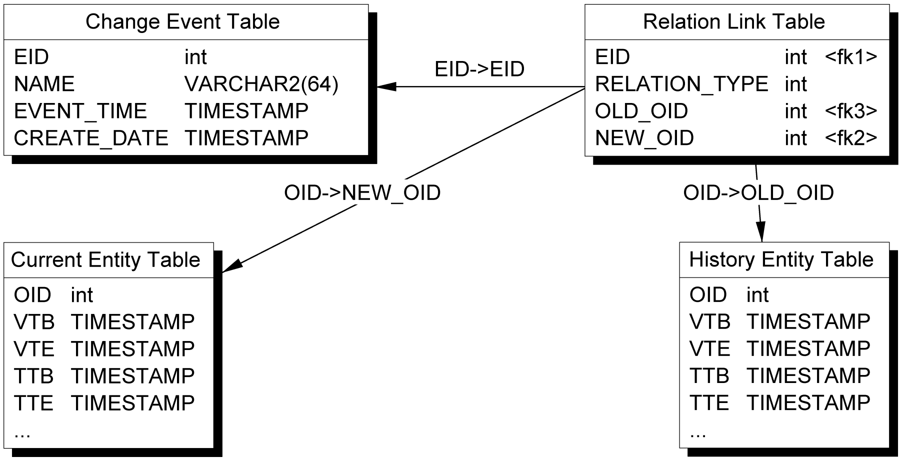


Fig. 7. Database structure of ST-Objects and relation links.

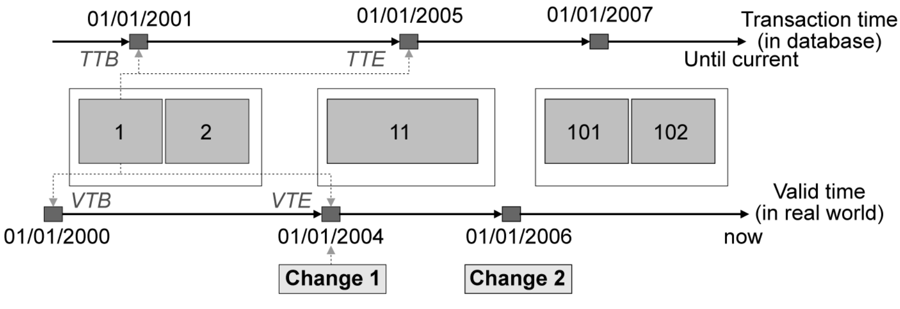


Fig. 8. An example of land parcel changes.

historical entity tables store an ST-Object's current and historical entity data. Change event tables record the event information about an ST-Object changes. Relation link tables record an ST-Object's relation links. The four tables' structures and relations are shown in Figure 7. The database stores unchanged objects only once.

Figure 8 shows an example of land parcel changes. The time granularity is "day," and the current time (CT) is 05/01/2007 (May 1, 2007). Parcels 1 and 2 were merged into one parcel 11 at time 01/01/2004 (January 1, 2004), and then parcel 11 was split into parcels 101 and 102 at the time of 01/01/2006 (January 1, 2006). A parcel's entity and corresponding relation information are stored in Tables 1–4. For parcel 1, its TTB and TTE are 01/01/2004 and 01/01/2005, respectively. In the real word, parcel 1 is valid from 01/01/2000 to 01/01/2004.

A Hierarchical Indexing Schema

There are two common indexing schema, direct and indirect indexing. The former only has one base state *t0*, and thus when updating spatio-temporal objects, all difference files have to be modified. Indirect indexing reduces maintenance overhead

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

Table 1. Change Event Table

EID	NAME	EVENT_TIME	CREATE_DATE[a]	…
1	Change1	01/01/2004	01/01/2005	
2	Change2	01/01/2006	01/01/2007	

[a]CREATE_DATE is transaction time.

Table 2. Relation Link Table

EID	RELATIONTYPE	OLD_OID	NEW_OID	…
1	ALTER(1:n)	1	11	
1	ALTER(1:n)	2	11	
2	ALTER(n:1)	11	101	
2	ALTER(n:1)	11	102	

Table 3. Current Entity Table

OID	VTB	VTE	TTB	TTE	…
101	01/01/2006	NOW	01/01/2007	Until current	
102	01/01/2006	NOW	01/01/2007	Until current	

Table 4. Historical Entity Table

OID	VTB	VTE	TTB	TTE	…
1	01/01/2000	01/01/2004	01/01/2001	01/01/2005	
2	01/01/2000	01/01/2004	01/01/2001	01/01/2005	
11	01/01/2004	01/01/2006	01/01/2005	01/01/2007	

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

while indirect search performance is poor. A hierarchical indexing schema (Fig. 9) is designed to support fast, accurate, and efficient retrieval of spatio-temporal changes. The indexing schema is supposed to develop a data structure to improve the comparison performance for changes between different states. This hierarchical indexing schema combines the two schemas and divides the temporal dimension into a specified number of time slots. Therefore, we can have multiple base states while indirect indexing is implemented within each time slot.

Data Processing and Management

Figure 10 describes the procedure of processing spatio-temporal data and storing data into the four databases. When new data are available, database administrators can initiate the update process by creating dynamic relations and store the new data along with new relations into the Current Database. The differences between current base state and previous base state will be stored in the Difference Database. The Change

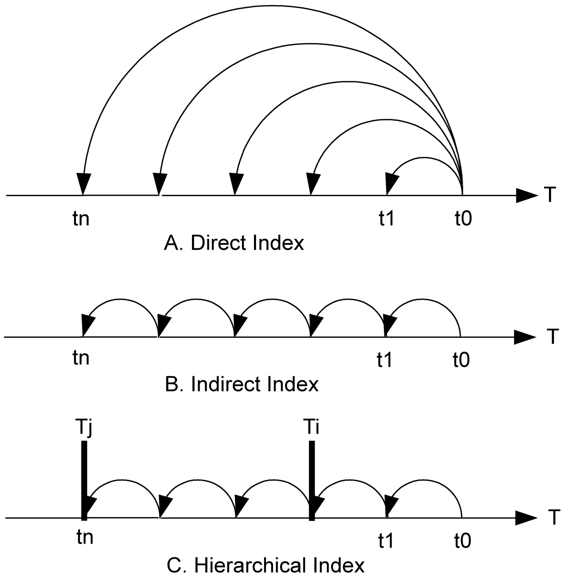


Fig. 9. Hierarchical indexing schema.

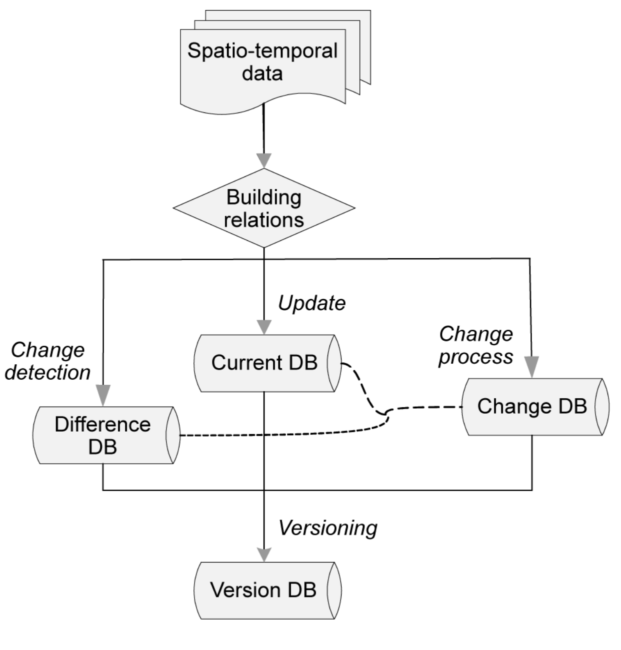


Fig. 10. Data processing procedure.

Database will also be updated to reflect these transactions. New Version Databases will be produced only upon users' request to change any point of time to the base state and make associated datasets as the current version.

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

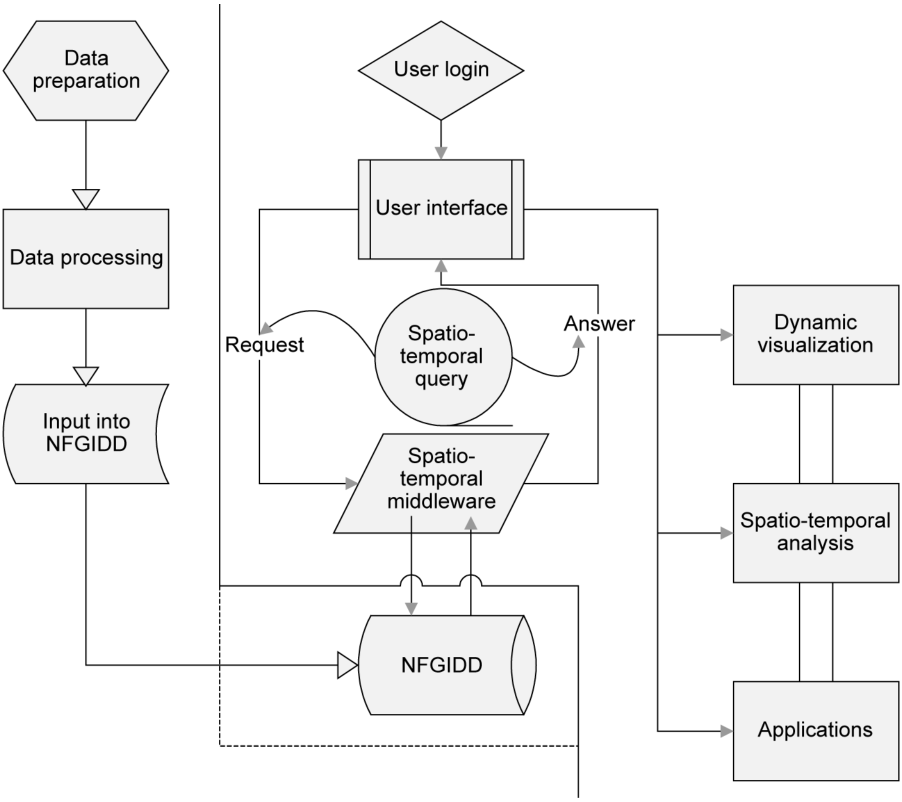


Fig. 11. Architecture of NFGIDD.

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

Two methods for building dynamic data are provided: (1) automated data relation construction, in which changes between new and old datasets are detected via automated dynamic relation algorithms that may not produce 100% correct dynamic relations; and (2) human editing interactive editing tools, which complement the automated method by manual user checking and correcting. Incorrect dynamic relations can be identified by users and then be modified manually. The map-centered interface can highlight all changes at the ST-Object level to facilitate identification of wrong relations.

Spatio-temporal Data Query and Visualization

All previous efforts in developing advanced data models, databases, and data management have facilitated spatio-temporal data querying and visualization. Common functionalities include dynamic spatio-temporal inquiry, history data tracking, change animation, and other visualization tools. Visual spatio-temporal inquiry tools are needed to support exploratory data analysis. Map views are dynamically updated when applying spatio-temporal and thematic filters that focus on specific space subareas or time subintervals. Spatio-temporal querying focuses on the exploration of

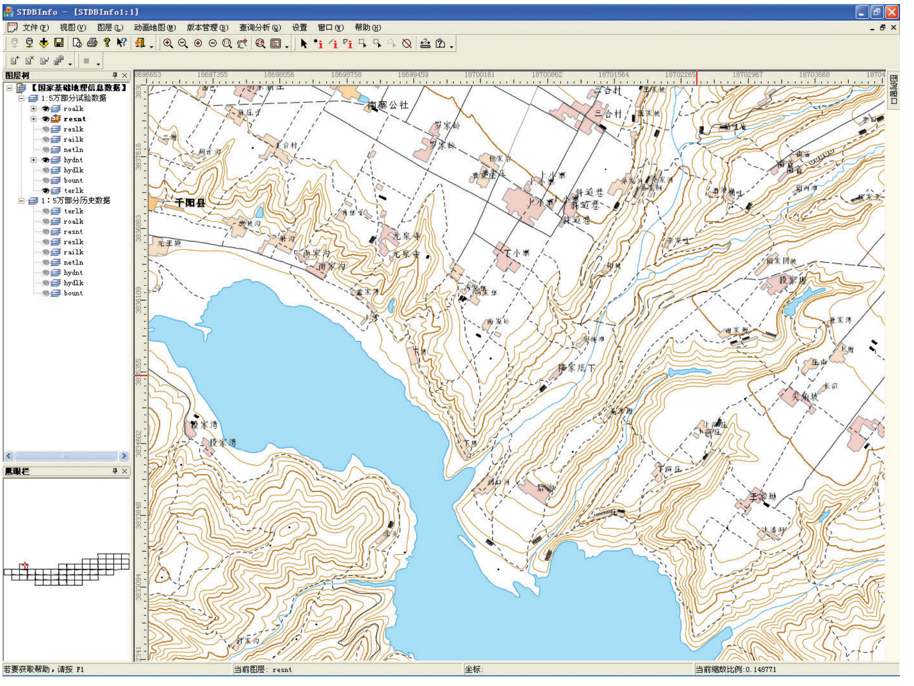


Fig. 12. GUI of STDBInfo system.

time variance in space, such as change event distribution over space and time and the statistics of changes.

Typical tools to support spatio-temporal data query and visualization are interactive animation, or dynamic symbolization. Based on the version-difference data model, we developed a series of dynamic symbols to represent a variety of changes over time. A dynamic symbol library was developed to incorporate these dynamic symbols, which are connected to different types of changes. Basic visual variables such as shape, size, color, transparency, location, and orientation of dynamic symbols are carefully used to devise visually effective tools to reveal spatio-temporal change patterns.

IMPLEMENTATION

The NFGIDD prototype adopts client/server architecture and the low-level software development language Visual C++. The NFGIDD platform consists of two major components, a data management system STDBInfo and a data processing system STDBMaker. The spatio-temporal management system STDBInfo has a graphic user interface (GUI) that is capable of importing data, checking data validity, supporting user and security management, and providing graphic query/visualization tools. STDBMaker was developed to preprocess spatio-temporal data including editing, data transformation (format/projection), change detection, data backup, and the building of relation links. The two systems work seamlessly to support dynamic spatio-temporal

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

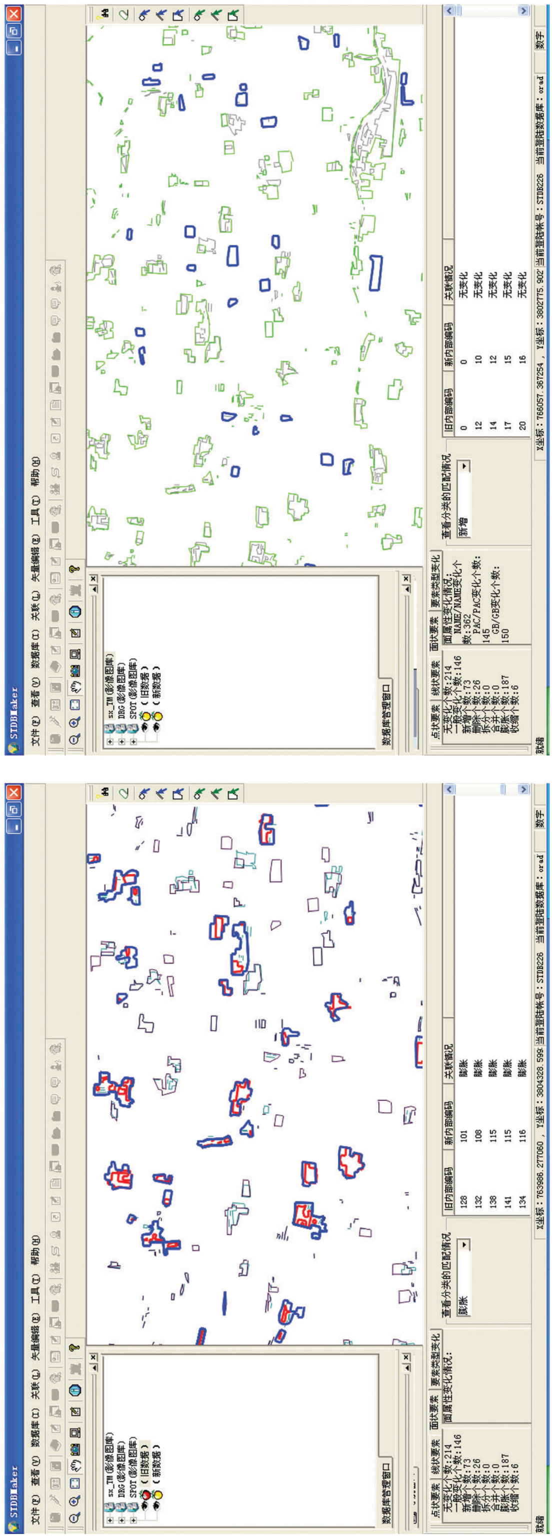


Fig. 13. Representation of changes.

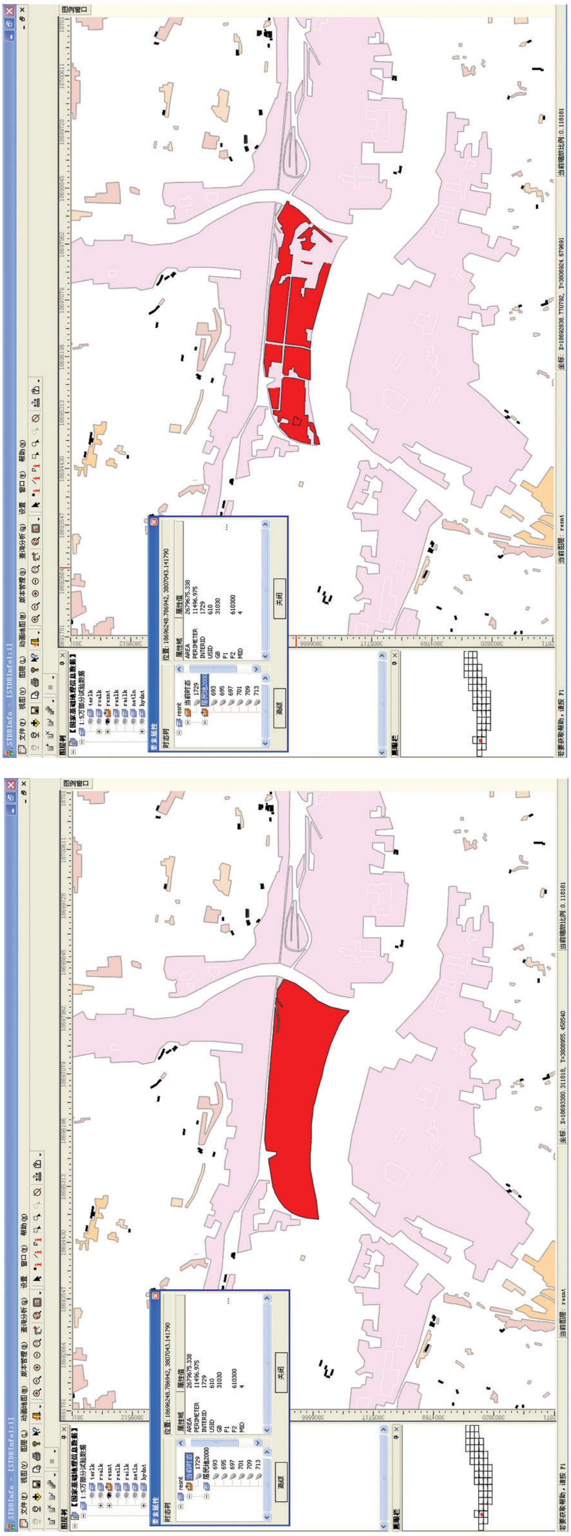


Fig. 14. Historic tracing at the ST-Object level.

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

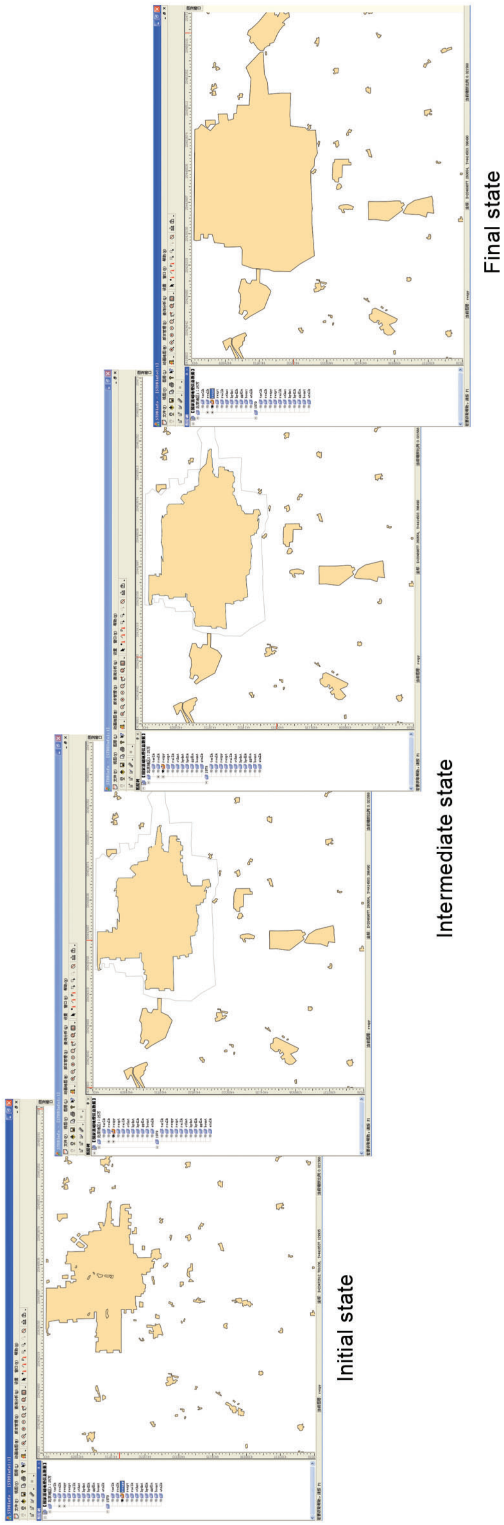


Fig. 15. Snapshot visualization method.

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

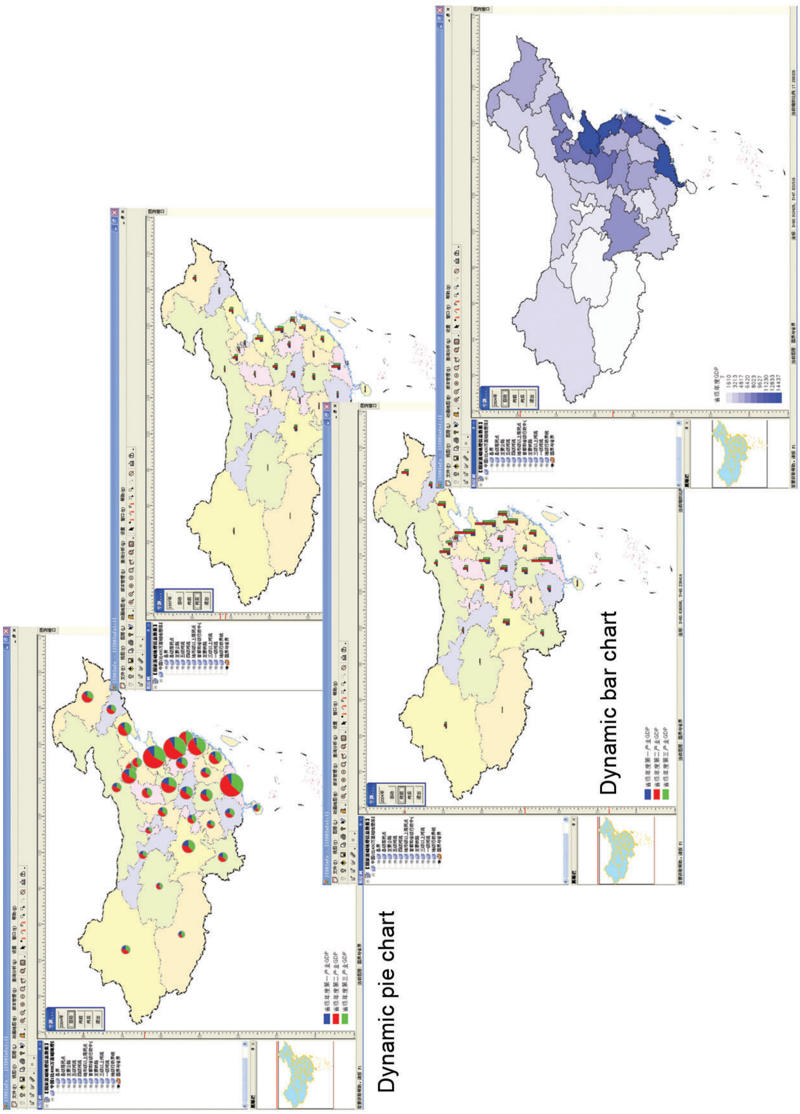


Fig. 16. Thematic mapping in STDBInfo.

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

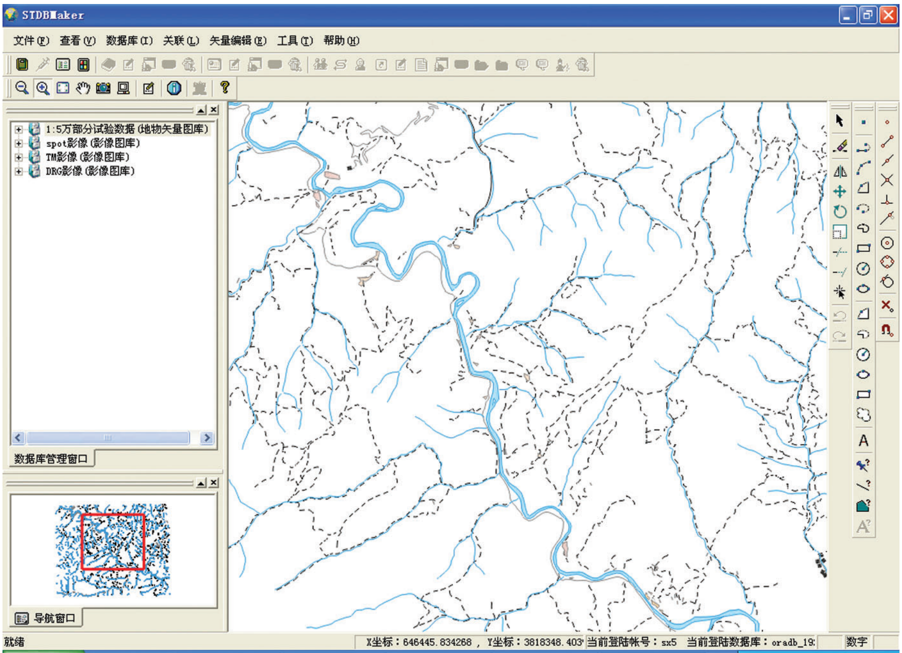


Fig. 17. GUI of the STDBMaker system.

data updating and routine management (as shown in Fig. 11).

The operational system adopts a three-tier framework including database server, spatio-temporal middleware, and user clients. We set up a highly efficient *database server* (SUN machine) and commercial database software (Oracle™) to store spatio-temporal data. *Spatio-temporal middleware* is a spatio-temporal data engine, which mainly serves as a data service interface and bridge between the user and data. It not only responds to user requests and returns answers, but also has other capabilities, such as assigning computation tasks automatically. The middleware provides a layer for the querying of spatio-temporal information, which delivers more efficient and productive processing of spatio-temporal queries. The middleware is deployed on a separate server. The *user clients* tier provides the user interface to respond to user requests and display output results.

Figure 12 shows the main GUI of STDBInfo. The system can easily display the changes between different states. As Figure 13 indicates, the blue objects represent new ST-Objects whereas the red objects represent old ST-Objects. The system also supports statistical analysis for these changes.

In addition to traditional spatial inquiry tools, we have implemented history tracing tools at the ST-Object level through dynamic spatio-temporal relations. Figure 14 shows an example of residential area changes from 1970 to 2000. From it, we can easily find where residential areas have expanded.

Based on China's official cartographic standards and specifications, we developed a set of static symbols as well as dynamic visualization methods to represent spatio-temporal changes effectively. Dynamic visualization methods include the snapshot

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

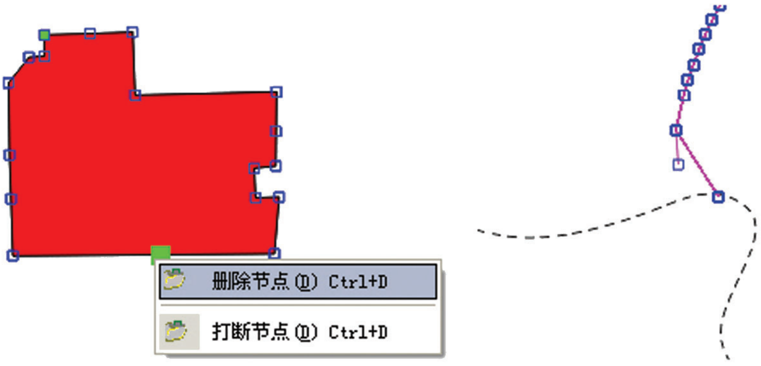


Fig. 18. Data editing tools.

method (representing states at a specific point in time) (Fig. 15), the difference method (animating the differences between time states), the interpolation method, and dynamic symbols to simulate the dynamic changes of geographic events and processes. In addition, common thematic mapping tools, such as choropleth maps and chart maps (bar and pie charts) (Fig. 16) are provided.

The STDBMaker system (Fig. 17) provides advanced data editing tools (Fig. 18) to enable time-constraint management. This means the data editing results can be validated to avoid conflicts between time states.

EXPERIMENTS AND DISCUSSION

Based on our spatio-temporal database prototype, we conducted a few experiments to test the prototype. The first experiment was conducted at the National Geomatics Center of China to build a National Fundamental Geospatial Information Dynamic Database at the scale of 1:250,000. Figure 19 shows the coverage area of these two experiments. The historical river bed changes of the Yellow River were simulated by the dynamic spatio-temporal database technologies. In total, 13 map layers were included to incorporate administrative boundaries, residential areas, transportation, the hydrographic network, topography, land use, and other important elements. And a total of 13 GB vector data (for one version) were used for the experiment. Our spatio-temporal database solution can save up to one-third the storage space compared to the snapshot model.

The second experiment was performed for the second National Land Use Survey Project. Land use changes can be modeled by NFGIDD technologies. Experiments proved that NFGIDD technologies can be successfully applied to routine land information update management. Figure 20 shows a screenshot of a land use management system that adopts the version-difference model and snapshot dynamic visualization technologies.

The two experiments all achieved high accuracy for change detection (>85% using automated detection mechanisms). Human intervention can further promote the accuracy rate. The two systems combined significantly facilitate the development of NFGIDD. Dynamic relations between spatial and temporal data have been

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

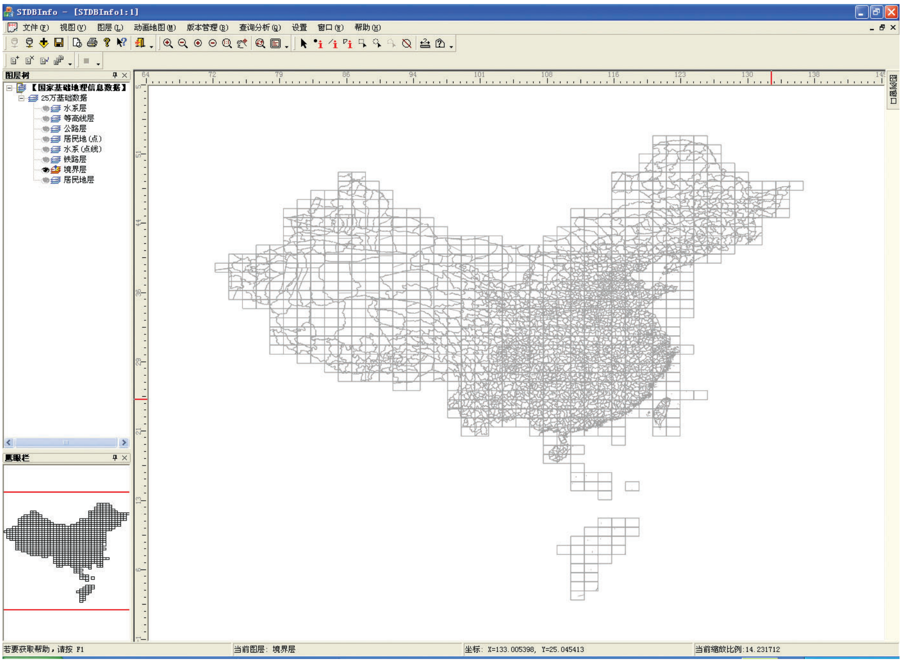


Fig. 19. NFGIDD experiment.

successfully established in integrated spatio-temporal databases. The prototype is able to support highly flexible spatio-temporal applications via dynamic geospatial information management. Our comprehensive methodology has demonstrable advantages over traditional spatial database solutions with respect to history data management and spatio-temporal data querying.

SUMMARY

This paper has introduced our experience in the development of a dynamic, large-scale spatio-temporal vector database to support the NSDI in China. A large portion of the current spatio-temporal models are only conceptual without concrete implementation. Comprehensive spatio-temporal database solutions are lacking. We address both theoretical and practical issues in the development of real-world NSDI applications. The generic methodology has been tested and applied to building an NFGIDD in China. The comprehensive NFGIDD solution includes a generic version-difference spatio-temporal data model, change detection techniques through dynamic relations, data storage schema with four logic databases, dynamic visual analysis, spatio-temporal data pre-processing and editing, and prototype implementation using object-oriented programming models.

The "version-difference" spatio-temporal data model presented in this paper organizes and stores the spatio-temporal data through different versions (states) and corresponding differences, and builds the dynamic relation links and hierarchical index, which can effectively reduce data redundancy and enhance spatio-temporal

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

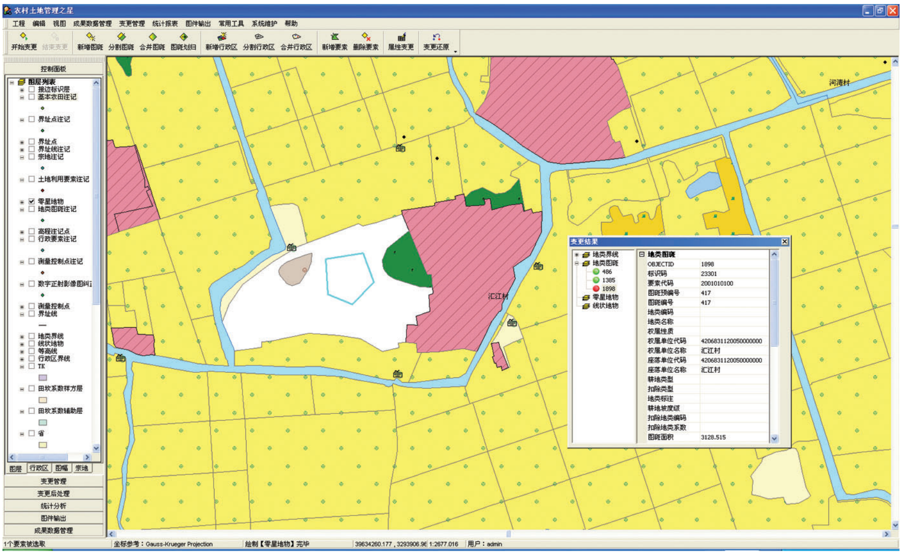


Fig. 20. Land use information system using the version-difference model.

data access. Implementational efforts (NFGIDD prototypes and experiments) based on this model demonstrates the advantages of the model. The dynamic and convenient GIS data maintenance and advanced spatio-temporal visualization can be used in many applications such as environmental monitoring, urban simulation, emergency management, and cadastre management. Further studies will focus on the temporal query language, spatio-temporal topology, and spatio-temporal data mining.

ACKNOWLEDGMENTS

This work was supported by the National Fundamental Surveying and Mapping Project–National Fundamental Geospatial Information Dynamic Database (No. 1460130524207), the National High-Tech Research and Development Plan of China (No. 2006AA12Z214, 2007AAA12Z346-5(2), 2009AA121404), the National 973 Basic Research Program of China (No. 2006CB701304), the Fundamental Research Program of the Ministry of Science and Technology, China (2007FY220300), and the Open Fund of the Chinese Academy of Surveying and Mapping (No. 77735).

REFERENCES

Abraham, T. and J. F. Roddick, 1999, “Survey of Spatio-temporal Databases,” *Geoinformatica*, 3(1):61–99.

Ageenko, E. and P. Fränti, 2000, “Lossless Compression of Large Binary Images in Digital Spatial Libraries,” *Computers & Graphics* 24:91–98.

Armstrong, M. P., 1988, “Temporality in Spatial Databases,” *Proceedings of GIS/LIS’88*, 2:880–889.

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

Assent, I., Krieger, R., Afschari, F., and T. Seidl, 2008, "The TS-Tree: Efficient Time Series Search and Retrieval," In *Proceedings of the 11th International Conference on Extending Data Base Technology* (EDBT 2008), Nantes, France, 252–263.

Barclay, T., Gray, J., and D. Slutz, 1999, *Mircrosoft TerraServer: A Spatial Data Warehouse*. Redmond, WA: Microsoft Research, Technical Report MS-TR-99-29.

Beard, K., 2006, "Modeling Change in Space and Time: An Event-Based Approach," in *Innovations in GIS: Dynamic & Mobile GIS: Investigating Change in Space and Time,* Drummond, J., Billen, R., Forrest, D., and E. João (Eds.), Boca Raton, FL: CRC Press, 55–74.

Borkar, V., Carey, M., Engovatov, D., Lychagin, D., Reveliotis, P., Spiegel, J., Thatte, S., and T. Westmann, 2009, "Access Control in AquaLogic Data Service Platform," in *Proceedings of the 35th ACM SIGMOD Conference*, Providence, RI, June 29–July 2, 2009, 939–946.

Castelli, V. and L. D. Bergman (Eds.), 2002, *Image Databases: Search and Retrieval of Digital Imagery*, New York, NY: John Wiley & Sons.

Claramunt, B. and M. Theriault, 1995, "Managing Time in GIS: An Event Oriented Approach," in *Recent Advances on Temporal Databases*, Clifford. J and A. Atuzhilin (Eds.), Berlin, Germany: Springer, 23–42.

Claramunt, C., Parent, C., Spaccapietra, S., and M. Theriault, 1998, "Database Modeling for Environmental and Land Use Changes," in *Geographical Information and Planning*, Stillwell, J., Geertman, S., and S. Openshaw (Eds.), Berlin, Germany: Springer, 181–202.

Duan, S., Thummala, V., and S. Badu, 2009, "Tuning Database Configuration Parameters with iTuned," in *Proceedings of the 35th International Conference on Very Large Data Bases*, Lyon, France, 2009. 1246 – 1257.

Erwig, M., Guting, R. H., Schneider, M., and M. Vazirgiannis, 1999, "Spatio-temporal Data Types: An Approach to Modeling and Querying Moving Objects in Databases," *Geoinformatica*, 3(3): 265–291.

Griffiths, T., Fernandes, A., Norman, W. P., and R. Barr, 2004, "The Tripod Spatio-Historical Data Model," *Data and Knowledge Engineering*, 49(1):23–65.

Griffiths, T., Fernandes, A., Norman, W. P., Jeong, S.-H., Djafri, N., and K. T. Mason, 2005, "Spatio-temporal Databases in Practice: Directly Supporting Previously Developed Land Data Using Tripod," in *Proceedings of the 21st International Conference on Data Engineering,* 2005, 1135–1136.

He, B. and Q. Luo, 2008, "Cache-Oblivious Databases: Limitations and Opportunities," *ACM Transactions on Database Systems* 33(2):1–42.

Jin, R., Xiang, Y., Ruan, N., and D. Fuhry, 2009, "3-HOP: A High-Compression Indexing Scheme for Reachability Query," in *Proceedings of the 35th ACM SIGMOD Conference*, Providence, RI, June 29–July 2, 2009, 813–826.

John, A. T., Anthony, J. T., and A. K. Mark, 1994, "The Handling of Large Digital Orthophoto Raster Database in GIS," *International Archives of Photogrammetry and Remote Sensing*, 30(4):457–459.

Langran, G. and N. R. Chrisman, 1988, "A Framework for Temporal Geographic Information," *Cartographica*, 25(3):1–14.

Lohfink, A., Carnduff, T., Thomas, N., and M. Ware, 2007, "An Object-Oriented Approach to the Representation of Spatiotemporal Geographic Features," in

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014

Proceedings of the 15th Annual International Symposium on Advances in Geographic Information Systems, Seattle, WA.

Mecca, G., Papotti, P., and S. Raunich, 2009, "Core Schema Mappings," in *Proceedings of the 35th ACM SIGMOD Conference*, Providence, RI, June 29–July 2, 2009, 655–668.

Parent, C., Spaccapietra, S., and E. Zimányi, 1999, "Spatio-temporal Conceptual Models: Data Structures + Space + Time," in *Proceedings of the 7th ACM International Symposium on Advances in Geographic Information Systems*, Kansas City, MO, 26–33.

Paton, N. W., Fernandes, A. A. A., and T. Griffiths, 2000, "Spatio-temporal Databases: Contentions, Components, and Consolidation," in *Proceedings of the 11th International Workshop on Database and Expert Systems Applications*, 851–855.

Pelekis, N., Theodoulidis, B., Kopanakis, I., and Y. Theodouridis, 2004, "Literature Review of Spatio-temporal Database Models," *The Knowledge Engineering Review*, 19(3):235–274.

Peuquet, D. J. and N. Duan, 1995, "An Event-Based Spatiotemporal Data Model (ESTDM) for Temporal Analysis of Geographical Data," *International Journal of Geographical Information Systems*, 9(1):7–24.

Pultar, E., Raubal, M., Cova, T. J., and M. F. Goodchild, 2009, "Dynamic GIS Case Studies, Wildfire Evacuation, and Volunteered Geographic Information," *Transactions in GIS,* 13(1):85–104.

Rapper, J. and D. Livingstone, 1995, "Development of a Geomorphologic Spatial Model Using Object-Oriented Design," *International Journal of Geographical Information Systems*, 9(4):359–384.

Reitsma, F. and J. Albrecht, 2005, "Implementing a New Data Model for Simulating Processes," *International Journal of Geographical Information Science* 19(10):1073–1090.

Renolen, A., 1996, "History Graphs: Conceptual Modeling of Spatio-temporal Data," paper presented at the conference GIS Frontiers in Business and Science, Brno, Czech Republic, International Cartographic Association.

Smith, T. R., 1994, "On the Integration of Database Systems and Computational Support for High-Level Modeling of Spatio-temporal Phenomena," in *Innovations in GIS*, Worboys, M. F. (Ed.), Bristol, PA: Taylor & Francis, 11–24.

Taciana, D. and C. Gilberto, 2004, "Bottom-Up Development of Process-based Ontologies," in *3rd International Conference on Geographic Information Science, GIScience 2004,* Adelphi, MD, October 20–23, 2004.

Tryfona, N., 1997, "Modeling Phenomena in Spatiotemporal Databases: Desiderata and Solutions," *Lecture Notes in Computer Science*, 1460:155–165.

Wachowicz, M. and R. G. Healey, 1994, "Towards Temporality in GIS," in *Innovations in GIS*, Worboys, M. F. (Ed.), London, UK: Taylor and Francis.

Worboys, M. F., 1992, "A Model for Spatio-temporal Information," in *Proceedings of the 5th International Symposium on Spatial Data Handling*, 2:602–611.

Worboys, M. F., 2005, "Event-Oriented Approaches to Geographic Phenomena," *International Journal of Geographical Information Science,* 19(1):1–28.

Yuan, M., 1994, "Wildfire Conceptual Modeling for Building GIS Space-Time Models," in *Proceedings of the GIS/LIS'94*, 860–869.

Yuan, M., 1996a, "Temporal GIS and Spatio-temporal Modeling," National Center for Geographical Information and Analysis (CD-ROM).

Yuan, M., 1996b, "Modeling Semantical, Temporal, and Spatial Information in Geographic Information Systems," in *Geographic Information Research: Bridging the Atlantic*, Craglia, M. and H. Couclelis (Eds.), London, UK: Taylor and Francis, 334–347.

Zhang, N., Agarwal, N., Chandrasekar, S., Idicular, S., Medi, V., Petride, S., and B. Sthanikam, 2009, "Binary XML Storage and Query Processing in Oracle 11g," in *Proceedings of the 35th International Conference on Very Large Data Bases*, Lyon, France, 2009, 1354–1365.

Downloaded by [University of Hawaii at Manoa] at 23:21 25 December 2014