# Spatio-Temporal Databases

Alvin Thai Amruta Sawant Samriddhi Singla

### Background

Classifications of Spatial Information Systems (SIS)

- Geographical Information Systems (GIS)
- Automated Mapping/Facilities Management Systems (AM/FM)
- Land Information Systems (LIS)
- Image Processing Systems



### Background

Geographic Information has the following attributes

- Theme phenomena or objects being observed
- Location
- Time of observation
  - This is often neglected by "modern" (as of 1999) SIS, which stores only spatial and aspatial data



### Background

- SIS Data Contents:
  - Aspatial data thematic data and attributes
  - Spatial data
    - Geometric information
    - Spatial relationship descriptions
      - Topological concerned with concepts like neighborhoods
      - Metric/Algebraic concerned with directions and distances
      - Partial order of spatial objects e.g. "north of," "above"



#### **SIS Data Structures**

Raster - based on the location-oriented field model

- Aspatial attributes are associated with points on a grid, which decomposes data space into cells
- Each attribute represents a separate layer of data

Vector - based on spatial entity-oriented object model

- Aspatial attributes describe entities in 2 or 3 dimensions
- Descriptions are stored with coordinates

# Temporal Database Management Systems (TDBMS)

- Came into existence due to demand for handling time related data
  - Store data relating to time instances
- Traditional database systems only retain the latest state of a modeled system
  - Problem: How do we store and preserve historic data?
- Two major attributes in temporal databases:
  - Valid time time period during which a fact is true in the real world
    - Stored to represent the time the change took place
  - Transaction time time period during which a fact stored in the database is known
    - Denotes when a change was registered in a database

#### **Temporal GIS**

- Role: tracing the lineage of spatial objects and their attributes
- Major functions:
  - Inventory
  - Analysis
  - Scheduling
  - Updates
  - Display
- Approaches to developing a temporal GIS:
  - Modeling changes attribute oriented, topology oriented
  - Modeling time itself



### **Temporal SIS Requirements**

- 1. Representation of Change
- Geographic object has 3 components:
  - Geometry
  - Topology
  - Attribute
- This object can experience change in one, two, three, or none of these components
  - Leads to eight possible spatio-temporal changes that an object may go through



#### **Temporal SIS Requirements**

2. Updating databases - Irregular vs. Regular intervals

3. Maintaining the duration of a status of an object vs recording events that imply status change

4. Storing lifespan (discrete phenomenon) vs temporal differences (continuous phenomenon)

5. Selecting the appropriate way to access data -Spatial vs Temporal vs Attribute priority

# Other aspects of Temporality

- Object identity
  - How do we distinguish one object from another?
- Perspective of time
  - How do we represent and view time?
- Querying temporal data
  - Results must be supported with visualization tools that can aid analysis
- Valid/Transaction time
- Dimensionality
  - How do we represent this information?
- Evolution
  - How do objects change?



# **Spatio-Temporal Modeling**

- Why is time important?
  - Preservation of historical data
- Functions of Usefulness
  - Check for data quality and integrity
  - Evaluation of past performance
  - Analyze past data to determine future trends

#### • Early attempts:

- Attribute timestamping
- Transaction logs
- Versioning



### **Approaches to Spatio-Temporal Modeling**

- Extending existing spatial models with time
- New Geographic Information System
- Incorporating Artificial Intelligence
- Time-based Representation
- 3-Domain Models
- Object-Orientation



#### Extending with Time

- Simplest solution to creating a spatio-temporal model
  - Comes from the need to be consistent with existing solutions
- Timestamp data with date of creation and date of cessation
- Alternative proposals for spatio-temporal database organizations:
  - Grid-based databases additional attribute history for each individual cell
  - Vector-based databases associate interval stamped attributes with locations or attributed dates with locations



# New GIS (Langran)

- Converting existing GIS to support time may not be enough
- Observations of time:
  - Relativity
  - Order of Events
  - Granularity
- Approach: have time and events stored on separate hierarchies
  - Events have locational references
  - Time hierarchy mapped to Event hierarchy
  - Connections between events can show causal relationships



# Incorporating AI (Hermosilla)

- Many spatio-temporal applications benefit or require reasoning capabilities
  - Future predictions
  - Decision-making
  - Discarding unwanted data
- Solution:
  - Common Indexing method to store basic/complex/rule data
  - General purpose query language with standard and spatial operators
  - Updates are directly made to the storage module
  - Achieves full temporal support with the incorporation of absolute and relative time



### Time-based Representation (Peuquet and Wentz)

- Representing time in GIS is a concern, but not a *User* concern.
- User concern: System must provide answers to queries
  - Answering queries with existing data representations can be computationally expensive
- Solution:
  - Capture changes in the environment along a temporal vector
  - Start with an initial state, then record events in a chain-like fashion in increasing temporal order
  - Each event is associated with a list of all changes that occurred
  - Temporal intervals depend on significant-enough changes
  - Changes stored as difference from previous version, or as full map

## 3-Domain Model (Yuan)

- Based on how humans conceptualize their surroundings
- Represent semantics, space, and time separately
- Provide links between them to describe phenomena
- Querying searching through links between entities
- Advantage: can handle movement as well as change



### **Object-Orientation (Wachowicz and Healy)**

- Based on object-oriented paradigm instead of traditional models
- Information is collected and stored as an entity with a unique identifier
- Real-world phenomena represented as complex versioned objects
  - New instance of each object is created for every version of the object
- Events are manifestations of actions that invoke update procedures
- Time is an independent from spatial dimensions



- Atomic Spatial data types (points, lines and polygons) and Temporal data types (events and intervals) can be combined into abstract spatial- temporal data types.
- Logic and Algebra can be used to describe the abstract spatial and temporal data and their relationships with operations for their composition.
- For abrupt changes and slow evolution, there would be a need to represent time both discretely and continuously.
- Including the branching and terminating time, would enhance the reasoning capabilities by allowing alternate timelines.





#### Koeppel and Ahlmer

Propose two techniques for the integration of temporal data into AM/FM systems.

• <u>Data Segmentation</u>, which is built on topological data structure and associated temporal events table that tracks changes. Combine these to get information about the linear network.

 <u>Change detection matrices</u>, which record the differences between two time periods. Each axis represents a time period. Standard matrix operations Standard matrix algebra operations are used to derive change detection matrices for multiple time periods

#### <u>Worboys:</u>

- Defined a spatio-temporal object as a unified object with both spatial and bitemporal extents.
- An elemental spatial object (a point, line segment or triangular area), also called a simplex, is combined with a bitemporal element, to form an ordered pair.
- Many such finite sets of ST-simplexes are combined to form an ST- Complex, on which the query algebra is developed.
- Their traces change discretely, which is why they can't be helpful in representing continuous evolution but, used where changes occur in sudden jump.



#### **Rojas-Vega and Kemp**

- Developed SIDL (Structure and Interface Definition Language) suited for distributed multi-media spatial applications.
- Here the object has structural and interface part, for achieving encapsulation. The structural part an object identifier, conventional attributes, an object component grammar and conceptual relationships are defined, while the interface contains methods operating on the object.
- Here complex object structures can be built to fully model the real life entities and their interactions.
- Time is introduced as separate objects for different models of time, which can be attached to time varying components. The diff models are intervals or points.

#### Yeh and Yeh and de Cambray

- Provide a model for highly variable spatio-temporal data using behavioral functions.
- Behavioral Function forms a spatio-temporal object triplet, which gives the time stamp, spatial data to describe data evolutions.
- Data evolution is described as v, t, c were value timestamp and c is behavioral function.
- Each time when a version is recorded, the function shows how the data is evolving.
- Benefits: less data redundant, resolves data efficiency between the states, allows modeling complex evolution.



- Relationships between Spatio-temporal objects, expresses a valuable information about the interactions of the real-life entities they represent.
- They can be derived by examining the contents of the database or by storing explicitly, depending on the data model and application demand.
- Topological, algebraic and order related spatial relationships are further complemented by temporal relationships, which are either topology-like (e.g., meets,during), or order-like (e.g., before, after).



- Simple Reasoning model by **<u>Guesgen</u>**, based on <u>Allen's temporal logic</u>.
- Four basic one dimensional spatial relationships.
- The relations(left of, attached to, overlapping and inside) can depict, spatial relationships, between multidimensional objects, using orthogonal axes, describing relationships along the individual axes as a tuple.
- Ambiguity can be a problem since some relationships can give unspecified results.
- <u>Hazelton</u> lists possible topological relationships, between objects of various dimensions in a 4-d space.
- Here the 0 to 4 D objects and their interactions are examined in their natural and higher dimensions, with the 4<sup>th</sup> dimension as a time. Two objects are recognized, space like and time like.

#### Egenhofer and Al-Taha

- Examine topological relationships, of 2D objects in a 2D space, while undergoing gradual changes.
- Mathematical formalism (for Scaling, Translation and Rotation) is given by defining topological distance, between 8 possible relationships and observing its properties.
- CTRG is produced from these which connects the relationships with minimal topological distance. This graph is then used to demonstrate the effect of the different kinds of transformations by traversing its nodes.
- Applications are snapshot systems, where change is slow but continuous.



#### Relations







#### <u>Frank</u>

- Uses explicit <u>Calendar like timestamping</u> vital for temporal GIS. Assumes that events are point-like.
- Also, operator "before or Equal" is defined, from which he derives other events to provide a generic, totally ordered and partially ordered time models.
- When used in models that de-couples spatial and event information, these time models add flexibility of reasoning for events where only their ordering is known. Eg: archeology or geology.



#### <u>Dutta</u>

- Uses fuzzy logic to enable the construction of a reasoning model with topological entities and constraints as building block in a point based representation.
- The framework also defines the relationships between the entities and temporal intervals and description of axioms and properties like support, topological equivalence, complement, intersection etc.



#### Access

- Acquisition- Capture, transfer, validation, editing.
- Storage
- Processing .
- Specific languages for querying.
- Good indexing methods to speed up querying.



# Querying

Resemble natural language. Represent powerful operators which, facilitate the analysis.

- 1. Languages
- 2. Operators
- 3. Formalisms.



### **Querying - Languages**

- <u>Semantic complexity</u> is more because of temporal and spatial dimensions, hence languages which go beyond the current mainstream relational ones, are necessary.
- <u>Expressive power-</u> we should determine the queries we expect the language to give answers for. Langran gives space and time dominant queries.

HQL- extension to relational query language DEAL( loops, conditional statements). Recursion makes this powerful. Able to handle spatial temporal queries.

Berman's Geo-Quel, Joseph's PicQuery and Ooi's GeoQL for spatial querying. Ariav's TOSQL and Snodgrass' TQuel , with the latest research culminating in TSQL2 for temporal querying.

#### **Querying - Operators**

- Aspatial operators, spatial metrics, spatial topology
- Temporal topology is analogous to spatial topology.

Worboys proposal- time (single,linearly ordered, unidirectional) dimension, is orthogonal to spatial dimensions. Projecting spatio-temporal primitives, STatom onto the sets of Spatial objects or Time Intervals.

Main idea-->"any spatio-temporal relationships between ST-atoms can always be decomposed into the Cartesian product of separate spatial and temporal relationships"

### **Querying - Operators**

Peuquet and Wenz- Time based data model gives events along a vector, that operates as a time-line. Temporal distance is used which can be an event duration, describes a period when certain conditions remain unchanged. Three groups of temporal operators.

- Linear metrics and topology
- Boolean operators- intersection, union, negation
- Generalization- express events at other resolutions and scales.

Temporal correlation operator- Hypothesis.

STIN operation- Selecting spatio-temporal datasets for analysis.

### **Querying- Formalisms**

To mathematically describe the behaviour of spatio-temporal querying process, first order predicate calculus, modal temporal logic, dynamic logic and non-monotonic logic are formalisms.

Gabbay and McBrien added UNTIL and SINCE, modal operators, to FOL.

Their TRA includes definition of "since" and "until" product to add selection, projection, cartesian product, set difference and union.

Temporal-SQL is designed based on TRA.

#### **Querying- Formalisms**

**Coenen** defines temporal calculus for raster based GIS. Their Event space calculus, is based on time, as another spatial dimension. This allows additions but does not allow delete operations, since it adheres to the concept of, non-destructive assignment found in classical logic.

Using **Allen's temporal logic** and simple arithmetic on tesseral addresses, relationships can be constructed on events and intervals, giving powerful temporal reasoning capabilities. Many spatio-temporal operators can be represented, which extend and redefine components of classic relational algebra.

### Indexing

- Fast access to data depends on the structure of the data stored.
- Complex indexing strategies can help locate the data quickly.
- Well known structures like R-trees, quadtrees etc follow the conventional techniques of building balanced trees.
- Spatio-temporal database has historical information as well.
- More focus should be given to latest data over archival information.



### Indexing methods

Method 1: Salzber and Tsoutras-> hB-Tree and TSB (time-split-B tree). Selecting well paginated indexing techniques with low space needs. Also, gets optimal data clustering on storage media.

Method2: Langran proposes that either space or time dimension will be more dominant. Hence partitioning based on these.



### Indexing methods

- Method 3: using R trees for storing the spatial information in the form of MBR. improving indexing, can be used to create R-tree for an image and then link them in temporal order. MR trees, space saving techniques, can be used for applications which do not undergo drastic changes.
- Method 4: kolovson and Stonebraker created SR trees. Segment trees which reference interval and point data in a single index and merge it with the R tree. Three tactics: Allowing storage of index records in non leaf nodes, variation of node size and preconstruction of the index.

Multidimensional segmented index handle historical data represented by time intervals, with time as a separate dimension.

Basoglu and Morrison- First data structure. Four level hierarchical structure



#### Langran and Chrisman

- 1) Snapshots: Inadequate since, unable to represent events that lead to changes in data, redundant information storage and lack of support to enforce integrity.
- 2) Overlay : Only the changes(overlays) are stored. Superimposing all the current overlays on the basemap.
- 3) STC: Space-time composite. Divides the map into smaller units. Each unit if changes is further split into two objects. Each object is treated differently.



Object oriented paradigm:

Implementation of Temporal change object data model, in which real world entities encapsulate the change components by including the past present and future states.

Operations for creation, change and display are implemented are implemented for each object type.

This however cannot be embedded with most GIS systems, because of its proprietary nature.

#### <u>Teraoka</u>

Describes a data structure based on Multi-dimensional persistent search trees. Extension to the persistent tree. It was developed with a view to comply with the following properties.

- 1. Fast spatial and spatio-temporal searches.
- 2. Ability to handle dynamic temporal data.
- 3. Avoiding the storage of redundant spatial information.



Peuquet and Duan:

- Have outlined ESTDM- Event based Spatial Temporal Data model.
- Uses chain of events, the event vector, to model changes in a raster based thematic maps.
- Each chain represents a single thematic layer of a geographic area.
- A form of compression is used to avoid, storage of unnecessary data points.



Database systems need to provide users with visualization tools to present:

- Contents of Data Collection
- Results of Queries

Standard tools for Spatial Database Systems are:

- Browsers
- Plotters
- Map Displays





- These tools are poorly adapted to display dynamic/temporal information.
- Therefore, we need to design new tools for Spatio-temporal Information Systems.
- Required Features:
  - Karak and MacEachren: Temporal map, which is a representation of changes in geographic reality. A tool that is Visual, Digital or Tactile and can present locational and /or attribute changes over time.
    - Six Dynamic Variables: Display Date, Duration, Frequency, Order, Rate of Change and Synchronization.
  - Muehrcke: Classifies temporal maps into maps showing qualitative, quantitative, composite change and space-time ratio.

- Required Features:
  - Koussoulako and Kraak: Three distinct methods:
    - Static Maps: the temporal component being transcribed graphically by means of variables
    - Series of Static Maps: of progressive time slices
    - Animated Maps: Change is observed through real movement on the map itself
- Animated Maps have been favored throughout literature.



- Case Study: Traffic Flow Analysis on a Static Street Map
  - Need variations that can be easily perceived by human eye.
  - They distinguish between real time and presentation time:
    - Presentation Time being a function of the observation interval and the picture rate, thus enabling the user to follow a course of events by either slowing down or accelerating the display of individual frames.
- There are efforts being made to give users the ability to communicate with the display in a language form.
  - Holmberg: Extends the Classical Map Language (CML) with dynamic variables to handle dynamic features in a map.



- The standard visual variables in CML, size, value, color, texture, grain, shape and orientation, which may be associated with complex map symbols made up from the basic point, line and area symbols, are complemented with dynamic properties for movement, oscillation, pulsation and rotation to form the Dynamic Map Language (DML).
- Dynamic symbols can vary in speed, frequency, amplitude, the number of levels, and rotation and display frequency.
- Trepied: Non-textual languages may prove to be the most useful for querying purposes. The alternatives, natural languages and artificial languages (e.g., SQL) are considered either too difficult to research or poorly adapted to the user.
- Limitations of visual languages are investigated and some limitations are outlined.
- A solution to some of the limitations is proposed to be the use of dynamic icons.

- They are able to represent spatiotemporal point, line and region type objects and can change shape, size and color, move, disappear or be combined with other icons to fully depict the temporal processes in the environment.
- Slocum et al.: Provide a more application-oriented point of view and commit to the development of a prototype visualization system.
  - Users can visualize dynamic data by three different approaches:
    - utilizing separate software packages for individual tasks,
    - developing a library of tools with a development kit, or
    - using a single specific exploration package designed for these purposes.
  - Propose a system which can provide various functionalities in the same system.



- GIS Extensions:
  - Temporal extensions of existing GIS generally aim to provide functionality to handle time from a specific perspective.
  - Most of these extensions are designed to solve a given application specific problem.
  - Indiana Dunes National Lakeshore:
    - Uses historical water quality data in their GRASS (Geographic Resources Analysis Support System) GIS to analyze spatial and temporal water quality parameters, such as chloride concentrations in lakeshore groundwater.
    - Data dating back to 1931 is converted to formats for use in ATLAS\*GRAPHICS, a desktop mapping program, that is capable of quickly displaying simple queries on water quality details.
    - A further conversion to GRASS provides more analytical power via its built-in hypothesis testing tools.

- Yates and Crissman:
  - **SIIASA spatio- temporal database** containing historical local administrative boundaries of East, Southeast and South Asian countries.
  - The database uses Langran's space-time composite model for temporal spatial data, while a versioning method exists for temporal aspatial data.
  - Several existing GIS software products, including ARC/INFO, MGE and MapInfo can serve as the base for the SIIASA database.
- Halls et al. :
  - Go beyond just adding temporal functionality to their features.
  - Developed a system for urban growth analysis from multiple data sources. The goal is to forecast future growth patterns with respect to specific requirements as well as to determine the point at which land saturation has taken place.
  - The system uses a structure that integrates data from different sources, including land use, value, cadastral records and census data recorded over a period of time

- Non-application specific solutions also exist.
- The **TEMPEST** temporal GIS prototype uses a time-based approach to create a tool for the analysis of spatiotemporal dynamics.
  - It incorporates all necessary temporal operators for topological relationships, combination and generalization, while leaving the spatial capabilities to be handled by a shell that interfaces another, conventional GIS.
  - Hence, time-based representation is directly supported within the system, while raster- and vector- based spatial data originates from and is saved to an external GIS.
  - Access is provided via a graphical user interface which offers flexibility, easy modification of the prototype and visual layout of the system components.
- **OOgeomorph** is another such example. It is an object-oriented implementation of a multi-dimensional GIS.

- Scientific Databases:
  - Vast amounts of data are available for researchers in fields such as environmental studies, and thus there is a need to efficiently store, access and analyze this information.
  - One of the most studied problems currently is global change, and several systems have been built specifically to investigate this issue.
  - Beller et al. :
    - Proposed an object-oriented tool set interfacing a conventional GIS (Genamap) that stores raw satellite-derived vegetation index data with temperature and precipitation information for a selected study area.
    - The additional capabilities the system provides include temporal database management built on the Temporal Map Set (TMS) concept; temporal interpolation methods to provide continuity within a TMS even though only a limited number of time slices may be available; the ability to transform existing TMSs into new ones; the use of animation to visualize temporal data, etc.

#### • Hachem et al. :

- GAEA, another object-oriented scientific database management system for global change research.
- It provides the ability to integrate data of multiple formats and includes sophisticated analytical tools integrated within the database manager, and an operator set that is interactively extensible by the user.

#### • Muntz et al. :

- QUEST (QUEries over Space and Time) is a prototype system built to serve as a testbed for "validating various techniques and demonstrating the feasibility and benefits of building information systems for atmospheric and earth science databases".
- In particular, cyclone detection and tracking is of interest.
- System components include a graphical user interface, a database manager, a visualization manager implemented in IDL for plotting, animating and analyzing data, and a query manager.

#### **Tool Sets**

#### • Story and Worboys:

• Describe a design support environment prototype for building spatio-temporal database applications, implemented by temporally extending the CASE tool of the commercially available Smallworld GIS.

#### • Brown et al. :

• Describe a visualization extension toolset for the Geographic Resources Analysis Support System (GRASS) to enable analysis and simulation of spatio-temporal processes.



#### Database issues:

- Spatial databases inherently contain large amounts of information about the environment.Temporal information in STIS, even with no data redundancy further increases the database size.
- Storage costs of this data, even on inexpensive media, may become unmanageable to users.
- A solution to this problem could be:
  - the design of new, innovative data structures with minimal storage requirements.
  - Distributed systems could also enable cost-sharing within or inter-organizations, as well as accommodating user data needs by storing information at locations where it is used the most.
- Another problems is rapid data retrieval when database size is big.
- Another important aspect is the need for strong analytical capabilities. This requires advanced querying, visualization and analytical reasoning abilities integrated within the Database Management System, optimized for the native data types of the system.

- Architectural types:
  - Three different architectures for spatial information systems are:
    - Dual architecture handles spatial and aspatial data in two separate subsystems, being often of proprietary nature, with an integration layer handling communication between the subsystems.
    - Layered architecture is implemented on top of an existing DBMS, with spatial operators added as another layer to operate on the spatial data types.
    - Integrated architectures provide extensible relational or object-oriented database management capabilities on which spatial data types are defined. Querying is hence directly supported by the DBMS.
  - Integrated architectures supports extensions for the time dimension as well, thus it could be preferred for STIS.
  - The former two architectures can still be used: by either developing a triad architecture with a temporal subsystem or layering spatial data types on a Temporal DBMS.

- Legacy systems:
  - Problems may arise due to the incompatibility of data formats between new and old systems.
  - The question that arises then, is how to format/migrate the old to the new system, if that is at all possible.
  - Solution:
    - Build new STIS on existing SIS, i.e., making available spatial information systems temporal.
    - Utilization of a data warehouse that also offers the benefit of enabling several of these legacy systems to be incorporated in a data-supply role.



- Data mining;
  - Interactive, user defined querying of large information bases are often very specific, displaying derived data or looking for the existence of expected patterns.
  - Large databases, often contain unexpected information that is not necessarily explicitly recorded or is searched for by user queries.
  - The extraction of such implicit data or rules about data belong to the field of knowledge discovery and data mining.
  - Two kinds of new rules can be discovered from STIS:
    - Spatio-temporal Evolution rules: that describe processes or state changes of objects over time
    - Spatio-temporal Meta- rules: Rules about rules, which describe changes between two rule-sets generated for static snapshot states of the database.
  - In addition, the application of "standard" data mining techniques, such as generalization, classification and association can provide additional insights into data.

- Data quality:
  - Geographic information captured in digital form contains uncertainty in several ways.
  - They stem from data measurement errors during recording, the discrete representation of numbers in computers, or simply from difficulties in the identification of geographic objects/phenomena at the time of measurement.
  - Imprecision manifests itself in geometrical as well as topological data, as the latter is normally derived from geometric information.
  - The temporal dimension further complicates matters.
  - However, existing methods to resolve imprecision problems, such as visualization, simulation and statistical (probabilistic) procedures could be employed in the same fashion as in the treatment of spatial imprecision, since explicit timestamps in the database are similar to stored geometric data, while implicit temporal data (e.g., before, during) resemble topological relationships.

#### Conclusion

- Spatio-Temporal Information Systems improve on existing spatial information systems by handling temporal information.
- Past, present and future states of the modeled environment in one system make the incorporation of new features possible to surpass current SIS capabilities.
- Most models are Application specific.
- Need more work to get a standard system.



#### Thank You

