

2011

A contour tree based spatio-temporal data model for oceanographic applications

Yuliang Chen

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations



Part of the [Social and Behavioral Sciences Commons](#)

Recommended Citation

Chen, Yuliang, "A contour tree based spatio-temporal data model for oceanographic applications" (2011). *LSU Doctoral Dissertations*. 3955.

https://digitalcommons.lsu.edu/gradschool_dissertations/3955

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.

A CONTOUR TREE BASED SPATIO-TEMPORAL DATA MODEL FOR OCEANOGRAPHIC APPLICATIONS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Geography and Anthropology

by

Yuliang Chen

B.S., East China Normal University, 2000

M.S., East China Normal University, 2003

M.S., Louisiana State University, 2009

May, 2011

Acknowledgments

Without the guidance, support, inspiration and help from all of my committee members, colleagues and family, my dissertation would not have become a reality. My sincere thanks and appreciations go to, but not limited to, the following people.

First, I am deeply grateful to my major professor, Dr. Michael Leitner, for his constant guidance, suggestions, understanding and support during the whole period of my study and dissertation research.

Secondly, I want to express my special appreciation and respect to Late Dr. Gregory W Stone, founding director of WAVCIS program at LSU. Both academic and financial support and encouragement from him over the years made this dissertation possible. During my work at WAVCIS lab, his charming personality really impressed me and trained me not just for academic achievements. Unfortunately we lost him before my final defense, but he will be remembered for ever.

I sincerely thank Dr. Lei Wang, Dr. Fahui Wang, Dr. Bijaya B. Karki, and Dr. Haosheng Huang for their valuable comments and advice while sitting in my committee. Dr. Lei Wang's practical suggestions and advice especially helped me a lot in my struggle with the research questions.

I am also indebted to my fellow lab mates at WAVCIS lab for their help on my study and their friendship. Special thanks are extended to Dr. Felix Jose. Thanks for his being willing to read and review the working draft of my dissertation. Besides, Dr. Yixin Luo and Dr. Baozhu Liu gave me a lot of valuable suggestions in my dissertation research.

My biggest grateful appreciation goes to my family members. My parents' everlasting love, encouragement and support always give me motivation. Also acknowledge my parents-in-law for their help and understanding.

The last but most special thank extends to my wife Jie Yu and my daughter Tracy Chen. Thanks for everything my wife did for me during my dissertation research and the time we live together. Tracy is the most precious gift of my life. I would like to dedicate this dissertation to them.

Table of Contents

Acknowledgments.....	ii
List of Tables.....	v
List of Figures.....	vi
Abstract.....	viii
Chapter 1 Introduction.....	1
Chapter 2 Literature Review.....	5
2.1 Spatio-Temporal GIS (ST GIS) Overview.....	5
2.2 Spatio-Temporal GIS Data Model.....	6
2.3 Spatio-Temporal GIS Considerations for Oceanographic Model Data.....	7
2.4 Review on Spatio-Temporal GIS Data Models.....	15
2.4.1 Spatio-Temporal GIS Data Model Survey.....	15
2.4.2 Review on Selected Spatio-Temporal GIS Data Models.....	16
2.5 Conclusion.....	24
Chapter 3 Ontological Analysis of Oceanographic Data.....	25
3.1 Why Ontology?	25
3.2 Ontology Applications in GIS.....	26
3.3 Fundamental Concepts in Ontology.....	29
3.4 Ontological Considerations of Oceanographic Model Data.....	30
3.4.1 Spatial and Temporal Granularity.....	31
3.4.2 Universals-Particulars.....	32
3.4.3 Independent-Dependent.....	33
3.4.4 Endurant-Perdurant.....	34
3.5 Ontologies of Oceanographic Data.....	34
3.6 Conclusion.....	40
Chapter 4 Contour Tree-Based Spatio-Temporal Data Model.....	41
4.1 Main Algorithm for the New Data Model.....	41
4.2 Contour Tree Overview.....	42
4.3 Evolvment Analysis of Contour Trees.....	44
4.4 Temporal Information Integration.....	48
4.5 Spatial Information Integration.....	51
4.6 Spatio-Temporal Interpolation.....	52
4.6.1 Spatial Interpolation.....	53
4.6.1.1 Support Methods for Spatial Interpolation.....	55
4.6.2 Temporal Interpolation.....	58
4.7 Conclusion.....	67
Chapter 5 Data Model Design Implementation and Evaluation.....	68

5.1 Evaluation of the Data Model for Special Challenges.....	68
5.2 Spatio-Temporal Data Model Design.....	70
5.3 A Case Study for the Spatio-Temporal Queries.....	74
5.4 Conclusion.....	81
Chapter 6 Contour Tree Simplification.....	82
6.1 Review on Contour Tree Simplification.....	82
6.2 Temporal Information in Contour Tree Simplification.....	83
6.3 Basic Simplification Operations.....	84
6.4 Rules of Simplification.....	85
6.5 Simplification Algorithm.....	87
6.6 Application of Contour Tree Simplification.....	90
6.7 Conclusion.....	93
Chapter 7 Conclusion.....	95
7.1 Major Findings.....	95
7.2 Contributions.....	96
7.3 Future Research.....	97
References.....	99
Vita.....	105

List of Tables

Table 2.1 List of capabilities and deficiencies of models in supporting oceanographic spatio-temporal considerations.....	23
Table 3.1 SNAP ontologies for oceanographic data and their relations.....	38
Table 3.2 SPAN ontologies for oceanographic data and their relations.....	39

List of Figures

Figure 1.1 An unstructured mesh grid for the Galveston coast, Texas.....	1
Figure 1.2 A snapshot of MIKE21 wave model output for the Gulf of Mexico.....	2
Figure 2.1 Two types of discrete changes: Stepwise constant (top) and erupted discrete (bottom).....	9
Figure 2.2 Periodic change: The time intervals between time steps ($t_1, t_2 \dots t_8$) are equal.....	10
Figure 2.3 Constant changes: Smooth change (top) and irregular change (bottom)	11
Figure 2.4 The eight possible types of spatio-temporal object changes.....	12
Figure 2.5 Examples of the Snapshot Model	16
Figure 2.6 An example of the Space-Time Composite Data Model of urban/rural evolution.....	17
Figure 2.7 Primary elements and the pointer structure in the Event-Oriented Spatio-Temporal Data Model (ESTDM).....	18
Figure 2.8 The conceptual framework of the three-domain model.....	20
Figure 2.9 Example of the STER model for a cadastral application.....	21
Figure 3.1 The main formal categories of SNAP entities.....	35
Figure 3.2 The main formal categories of SPAN entities.....	35
Figure 3.3 Ontologies for oceanographic data extended from SNAP.....	36
Figure 3.4 Ontologies for oceanographic data extended from SPAN.....	37
Figure 4.1 A contour map and its corresponding contour tree.....	43
Figure 4.2 Five types of contour map transformation and their corresponding contour tree evolution.....	46
Figure 4.3 Contour tree branch arc removal.....	47
Figure 4.4 Contour tree branch arc addition.....	48
Figure 4.5 An illustration of how nodes store time information in contour tree evolution.....	50

Figure 4.6 Examples of the spatial interpolation applied in the data model.....	54
Figure 4.7 Ray casting algorithm.....	56
Figure 4.8 MBR algorithm used to speed-up processing time for the ray casting.....	57
Figure 4.9 An example of value assigned to previously recorded value in a discrete change process.....	60
Figure 4.10 An example of value assigned to the nearest value in a periodic change process.....	61
Figure 4.11 The temporal interpolation for a point at a time step not stored during the original model run.....	63
Figure 4.12 An example of the nearest neighbor's value algorithm.....	64
Figure 4.13 An example of linear interpolation.....	65
Figure 4.14 An example of quadratic spline interpolation.....	66
Figure 5.1 UML class diagram of the main class modules.....	72
Figure 5.2 Critical nodes and related UpArc and DownArc in arc CE.....	73
Figure 5.3 Four images selected from the output of MIKE21 wave forecasting model...	76
Figure 5.4 Data model query application work flow chart.....	80
Figure 6.1 Leaf pruning and node reduction.....	84
Figure 6.2 Examples of importance values in contour tree simplification.....	86
Figure 6.3 Hierarchical decomposition of a contour tree.....	87
Figure 6.4 A wave height and direction map from the output of the MIKE21 model for the Gulf of Mexico.....	91
Figure 6.5 Contour tree simplifications.....	92

Abstract

To present the spatio/temporal data from oceanographic modeling in GIS has been a challenging task due to the highly dynamic characteristic and complex pattern of variables, in relation to time and space. This dissertation focuses the research on spatio-temporal GIS data model applied to oceanographic model data, especially to homogeneous iso-surface data.

The available spatio-temporal data models are carefully reviewed and characteristics in spatial and temporal issues from oceanographic model data are discussed in detail. As an important tool for data modeling, ontology is introduced to categorize oceanographic model data and further set up fundamental software components in the new data model.

The proposed data model is based on the concept of contour tree. By adding temporal information to each node and arc of the contour tree, and using multiple contour trees to represent different time steps in the temporal domain, the changes can be stored and tracked by the data model. In order to reduce the data volume and increase the data quality, the new data model integrates spatial and temporal interpolation methods within it. The spatial interpolation calculates the data that fall between neighboring contours at a single time step. The Inverse Distance Weighting (IDW) is applied as the main algorithm and the Minimum Bounding Rectangle (MBR) is used to enhance the spatial interpolation performance.

The temporal interpolation calculates the data that are not recorded, which fall between neighboring contour trees for adjacent time steps. The “linear interpolation” algorithm is preferred to the “nearest neighbor’s value” and “spline” interpolation methods, for its modest accuracy and the simple implementation scheme.

In order to evaluate the support functions of the new data model, a case study is presented with the motivation to show how this data model supports complicated spatio-temporal queries in forecasting applications.

This dissertation also showcases some work in contour tree simplification. A new simplification algorithm is introduced to reduce the data complexity. This algorithm is based on the branch decomposition method and supports temporal information integrated into contour trees. Three types of criteria parameters are introduced to run different simplification methods for various applications.

Chapter 1 Introduction

In modern day oceanographic research, numerical models play an important role in understanding and predicting the behavior of the complex oceanic and atmospheric systems. Using time series of wind vectors (magnitude and direction) as input, the wave models (MIKE 21, SWAN, WAVE WATCH III, etc.) provide the sea state in terms of bulk wave parameters; such as, significant and maximum wave height, mean and peak wave period, wave direction, etc; while the hydrodynamic models provide current speed and direction in a 3D framework (e.g., FVCOM, ADCIRC, ROMS). Oceanographic models are being implemented in forecast/hindcast mode to generate time-dependent parameters for prognostic requirements.

The 3rd generation oceanographic models can use fine spatial resolution for the computational mesh, especially for the coastal zone, to resolve highly complex modification of wave and current fields in the region. The implementation of the unstructured mesh significantly enhances model efficiency by eliminating altogether the time consuming process of nesting regional models with global models. Figure 1.1 provides an unstructured mesh

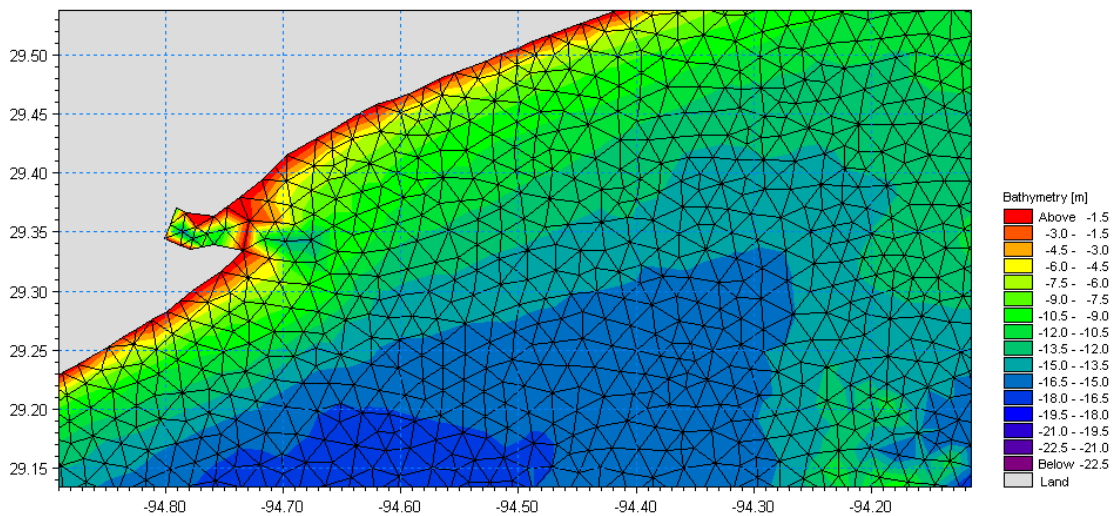


Figure 1.1 An unstructured mesh grid for the Galveston coast, Texas.

implemented for Sabine Bank, northern Gulf of Mexico, to study the wave transformation over the shoal. These 2D unstructured and regular meshes represent fully distributed sea state variables in the model domain, and also by organizing multiple independent frames, represent temporal changes during the computational temporal period.

Figure 1.2 provides a snapshot of the Significant Wave Height (SWH) and Mean Wave Direction (MWD) distribution for the Gulf of Mexico, simulated using MIKE21 spectral wave model. Zones of high wave energy and wave heights are concentrated off Yucatan Peninsula and off the Texas coast, while the north-eastern Gulf of Mexico is

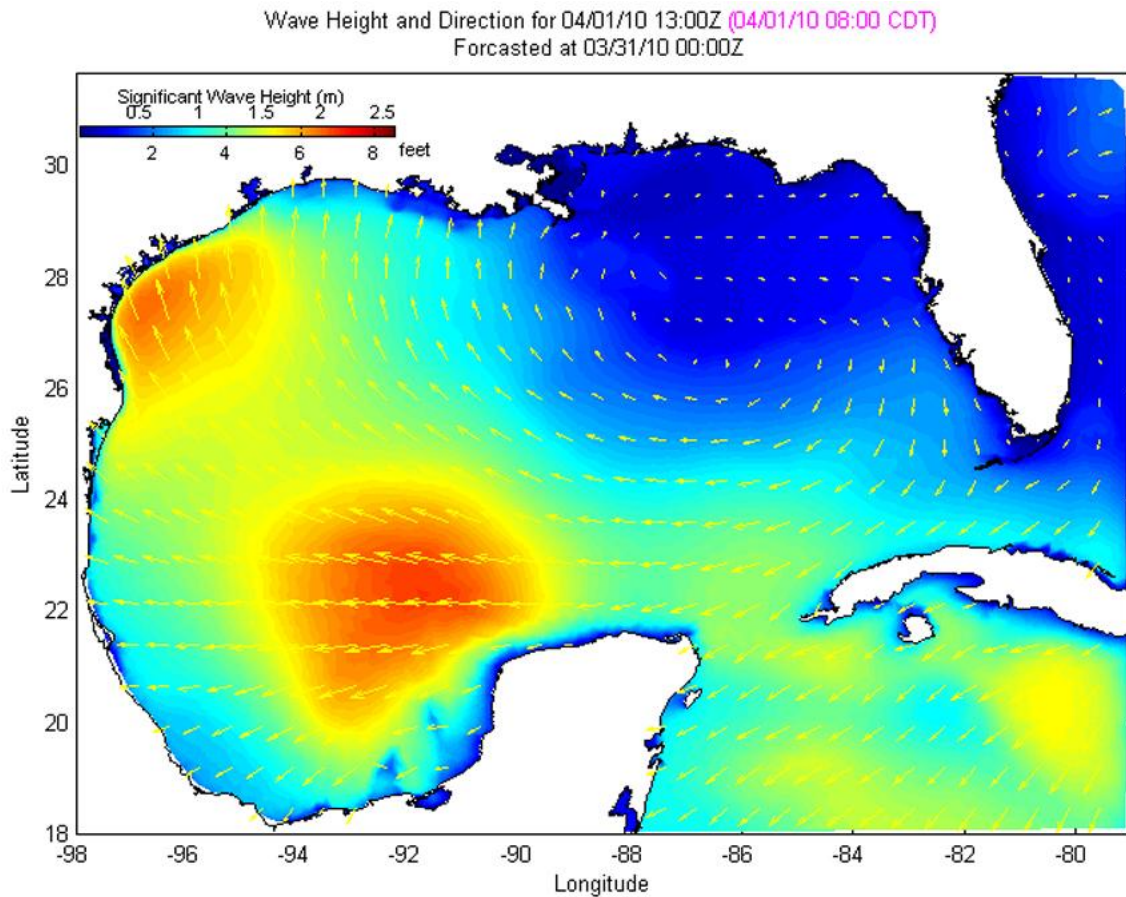


Figure 1.2 A snapshot of MIKE21 wave model output for the Gulf of Mexico: While the color fillings represent significant wave height distribution the yellow arrows provide the mean wave direction (see <http://wavcis.csi.lsu.edu> for an 84 hour wave forecast).

subjected to low energy waves. From this snapshot, the state of the Gulf of Mexico can be described as having two high wave zones and another low wave zone pervading the north eastern section. From a geographic perspective, oceanographic model output is one representation of dynamic geographic phenomena which relates to both space and time.

All wave and current computations yield high data volumes and increase the processing time. When it comes to model data visualization or spatial analysis, there is a need to strike a balance between descriptive details and computational load for particular model applications.

Geographic Information System (GIS) has been emerged as a critical tool in oceanographic modeling study. Traditionally, most GIS processes and oceanographic model implementations have to run separately. On the one hand, GIS software has been used in visualization and analysis of many oceanographic model outputs. On the other hand, a host of marine data application software tries to couple GIS with oceanographic models, in order to either speed-up model development time or to improve a model's representation or visualization. However, coupling GIS with oceanographic modeling software is still a challenging and demanding research.

The biggest challenge in coupling these two entities stems from the issue of the temporal consideration of the data. Conventional GIS emphasizes geospatial information at a single reference time, but oceanographic models usually deal with data having an additional temporal dimension, in which time is a fundamental element related to change. Consequently, the first issue to resolve, when coupling GIS with oceanographic modeling software, is to support the data with temporal information.

A spatio-temporal data model is the fundamental step to solve this issue. This dissertation focuses on developing a spatio-temporal GIS data model in supporting oceanographic model data, because current available data models have limited capability in supporting it.

Following this chapter, Chapter 2 is a literature review on the spatio-temporal GIS data model. Chapter 3 introduces ontological analysis of oceanographic data. Chapter 4 discusses the design of the new data model, namely a contour tree-based, interpolation algorithm integrated spatio-temporal data model. Chapter 5 covers the proposed model's implementation in detail and shows evaluations of the new model. Chapter 6 extends the research of data model on contour tree simplification and the last chapter concludes the major contributions and proposes future research.

Chapter 2 Literature Review

2.1 Spatio-Temporal GIS (ST GIS) Overview

Geographic realm consists of three domains: semantics (themes), time, and space. In GIS application, the basic consideration is to track information about “where,” “what,” and “when” of geographic phenomena (Peuquet 1994). Traditional GIS focuses on geographic objects or geographic phenomena at a specific time point or period. The time domain (“when”) usually is stored or represented as an attribute attached to the other two domains (“what” and “where”). For this reason, such a GIS has limited capabilities in supporting, represent, or record the temporal change or other dynamic information of the objects¹.

Spatio-temporal GIS (ST GIS) extends ordinary GIS with temporal information. A spatio-temporal GIS aims to represent, process, manage, and analyze spatio-temporal data (Yuan 1996). The efficient integration of time with spatial information in GIS is the challenge that a spatio-temporal GIS has to address. Since spatio-temporal GIS considers time as its main domain, it focuses more on the dynamic phenomena that it has to represent.

During the past two decades, spatio-temporal GIS has been a focused research area, with a lot of potential in met-ocean (meteorological and oceanographic) model applications. Research efforts and advancements in spatial-temporal GIS emphasize on the following aspects: Abstract system architectures, conceptual models, temporal operations, database query grammar/algorithm, data structure, graphical user interfaces, etc.

¹ In this dissertation, an object is called “dynamic” when it’s spatial information or attributes change with time.

2.2 Spatio-Temporal GIS Data Model

Among all aspects in spatio-temporal GIS research, the spatio-temporal data model plays a fundamental role. Spatio-temporal data models are abstract structures used to describe how geographic objects with temporal dimension would integrate and are represented in real world. It is “the core of a Spatio-Temporal Information System. Spatial-temporal data models define object data types, relationships, operations, and rules to maintain database integrity” (Pelekis et al. 2005). So when trying to support oceanographic model data in a GIS, the first task is to find an appropriate data model that can represent such model data.

Until recently, various spatio-temporal data models have been proposed. Among those models, although most of them claimed to be general (application-independent) models and have wider application fields, the majority of them are conceptual models and hence the application domains are very much restricted. For example, one of the commonly studied and applied spatio-temporal GISs could be the Land Information Systems (LIS) (Tryfona and Jensen 1999; Claramunt et al. 1998). The Cadastral Information System is one example of LIS, which tracks the change in land parcels with time and related information, such as individual ownership. Many proposed data models (Borges et al. 2001, J. Chen and J. Jiang, Pelekis et al. 2005, Tryfona and Jensen 1999) used LIS as an example to show their strong support in this kind of application. However, LIS could not represent all dynamic geographic phenomena in nature. There are even more complicated geospatial changes. The capability of applying these data models to other applications, such as for oceanographic model data, is very much limited. For example, the geospatial changes in LIS are discrete and easy to track. A direct adaptation of these data models to oceanographic model applications is almost impractical, given the rapid variability of the sea state conditions at the spatio-temporal scale.

So the limitations of the existing GIS models when it comes to marine data visualization and analysis is obvious. Actually, there is no widely accepted and generally applicable, sophisticated commercial spatio-temporal GIS product available thus far.

In effect, there is a need for a new spatio-temporal data model suitable for oceanographic model data. This dissertation focuses the research on developing this kind of spatio-temporal GIS data model, with a scope for managing met-ocean forecast model data. The new data model does not necessary to be generic, but can be customized to this specific application area. In this research, the specific application refers to the output data of numerical oceanographic models for waves and water levels.

2.3 Spatio-Temporal GIS Considerations for Oceanographic Model Data

In spatio-temporal GIS data model research, several fundamental issues need to be considered (Abraham and Roddick 1999). Among them, the most pertinent ones related to oceanographic model data are discussed below.

1) How to define the morphology (metrics) of the geographic phenomena:

Generally, GIS categorizes geographic phenomena into two concepts, namely fields and objects. “Fields usually first identify the spatial and temporal component (the element of the domain) and then associate the (non-spatial) field value”, while “Objects are primarily identified by their non-spatial and temporal characteristics and are then attributed with their spatial (and temporal) extension” (Kemp and Vckovsky 1998). The main difference between the two is that fields are continuous (or homogeneous), while the objects are discrete entities.

In GIS, vector and raster are implemented as the two fundamental data models to support both objects and fields concepts. Though it is possible to convert between fields and objects (raster and vector), the relative merits of vector and raster data for various spatial

analyses is distinct and depends heavily on the analysis purpose. May Yuan (2001a) also pointed out that many geographic phenomena have both object and field characteristics, and typical spatio-temporal data models have difficulty in properly representing them.

Oceanographic model data, such as surface wave condition, extensively cover the specific geographic space (model domain), hence representing a continuous space (homogeneous). But in oceanographic research, a lot of the time the data are monitored from an object perspective, for example, the zones with high wave activity or the track of a hurricane across the domain. Therefore, oceanographic model data have such both object and field characteristics. More studies need to be conducted to carefully select between fields and objects to effectively represent oceanographic model data.

The selection of an object or a field data model fundamentally affects the temporal data model design.

Widely used data models in oceanographic research, such as NetCDF (Network Common Data Form), implement the snapshot model as the temporal data model. The snapshot model is field based and supports high resolution, but cannot represent the discrete objects or track their spatial changes. Another disadvantage is that those field based data models usually have to handle big data volumes which result in data redundancy and extensive computation.

2) Temporal change pattern:

For dynamic geographic phenomena, spatio-temporal changes can be categorized into several types based on temporal change interval patterns, including discrete, periodic, and constant.

- Discrete change interval patterns: Changes (most of the time these changes are significant) occur on irregular triggers (events) without consistent connection between these events (Figure 2.1). The change procedure is ephemeral and also can be considered as instant. The status between the occurrences of change is stable. That means that there are periods of static status for the dynamic objects. Discrete changes can be furthermore divided into two types, including stepwise constant and erupted discrete. In stepwise constant change, the dynamic objects hold the static status for a certain period after the change. Examples of this type of change are, for instance, land parcels change position or size in the cadastral system. In erupted discrete change, the dynamic objects appear during a very short time period, or geographic phenomena happen instantly. Dynamic objects do not exist between the change incidents. Examples of this change could be wild fire, earth quake events, or similar.

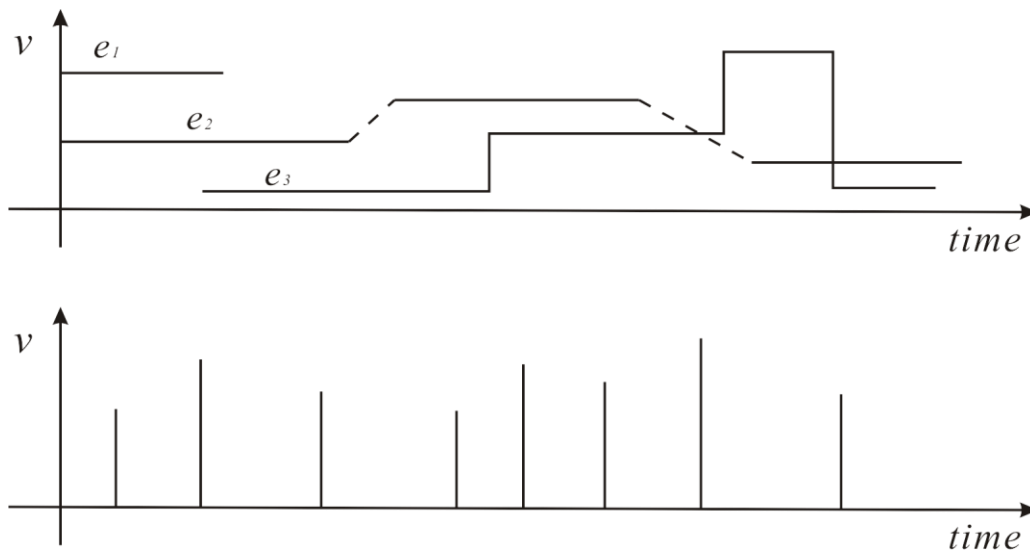


Figure 2.1 Two types of discrete changes: Stepwise constant (top) and erupted discrete (bottom).

- Periodic change: Changes are tracked by even time intervals (Figure 2.2). For example, many census data reflect the population change over a constant (usually 10 years) time interval.

The main difference between discrete and periodic change is that discrete changes can be defined by events, such as wild fire incidents, while periodic changes are defined and measured by even time intervals (periods). The changes that happen in a specific time period are accumulated, and treated as a single event with grouped attributes. Furthermore, in discrete change, events or changes usually are independent, but in periodic change the adjacent changes are related with strong spatio-temporal relationships. For example, the population changes between successive censuses are highly inter-linked. In both type of changes, the interval between two points in time is treated as stable, that means the change that happens in between these two points in time usually is not studied.

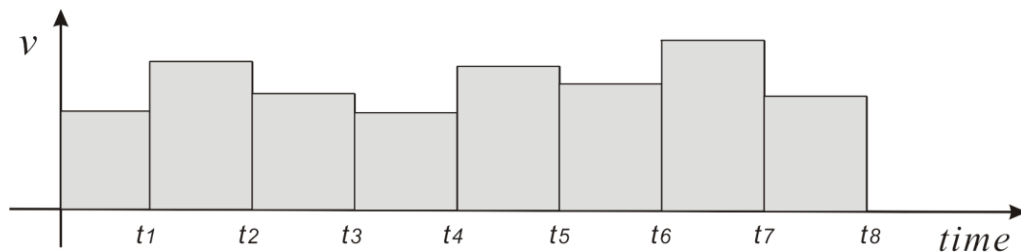


Figure 2.2 Periodic change: The time intervals between time steps ($t_1, t_2 \dots t_8$) are equal.

- Constant change: Unlike the two aforementioned changes, some geographic phenomena are highly dynamic. The change is constant and no discrete separation between events exists. It is also difficult to define the boundary changes based on time intervals. This kind of change is very typical in nature for moving vehicles, wind blowing across a lake, a hurricane advancing along its track, etc. The constant change

can be further divided into two types, namely smooth change and irregular change.

The smooth change is progressive without radical changes, while the irregular change has radical changes. Figure 2.3 demonstrates these two types of changes.

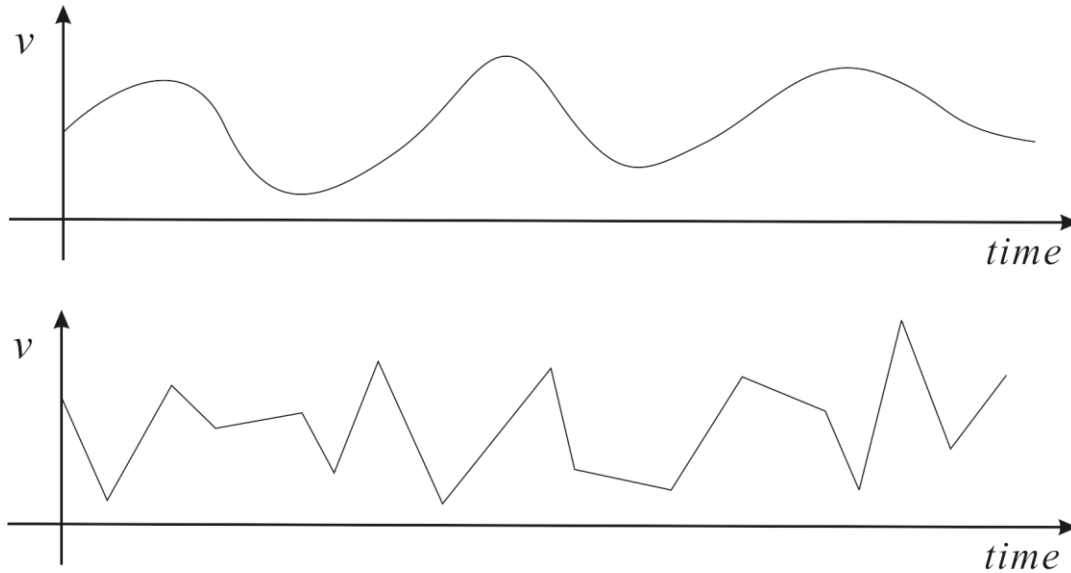


Figure 2.3 Constant changes: Smooth change (top) and irregular change (bottom).

The oceanographic model data evolution pattern can be categorized into the third type, namely constant change. Outputs from an oceanographic model record the met-ocean change for the whole simulation period for a specific study area (domain). The oceanographic changes are constant and it is barely impossible to define any discrete change or periodic change boundaries. Anytime during the modeling period and anywhere in the domain, the geographic phenomena data representation is dynamic.

Generally speaking, the oceanographic data change is smooth in appropriate spatial and temporal resolution. During low time granularity, changes can be irregular. For example, while tracking the high wave zones in the Gulf of Mexico when a hurricane is approaching the continental USA from the Caribbean, and with an observation interval of 6 to 12 hours, the changes are not smooth and regular. If time granularity is selected at a certain fine level,

the irregular changes can be smooth. For the same data change as discussed above for the hurricane, if change is represented every hour, the sea surface water zones of high wave tracking can be treated as smooth.

3) Spatio-temporal change types:

The change with dynamic spatio-temporal data (dynamic objects) involves geometry (position and/or shape), topology, and attributes. These changes can happen separately or simultaneously. As Roshannejad and Kainz (1995) summarized, the change of geographic objects in relation to time can be categorized into eight different scenarios as shown in Figure 2.4. The less complicated change involves only one of three elements, whereas the complicated changes involve two or more.

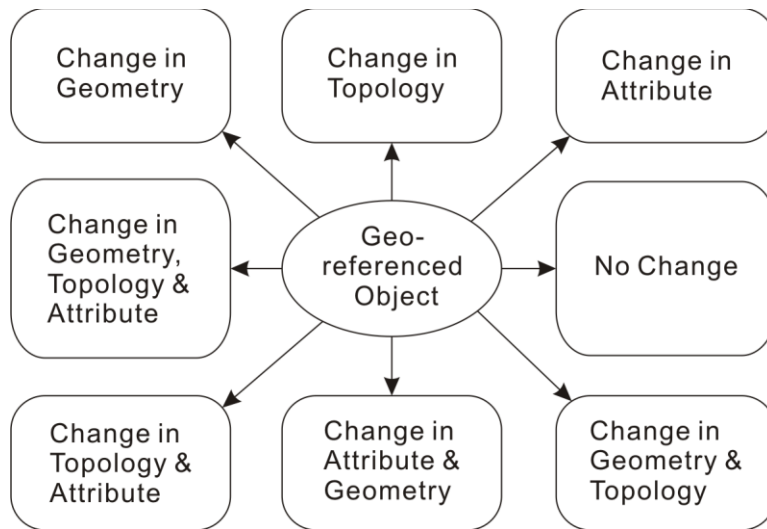


Figure 2.4 The eight possible types of spatio-temporal object changes (Roshannejad and Kainz 1995).

In reality, spatio-temporal changes in oceanographic model data involve geometry, topology, and attribute. And all three changes can occur at the same time. For the geometry change, no fixed/static shapes for a defined oceanographic data object (e.g. high wave zone) exist. The spatial variability of these data, representing geospatial objects, is stochastic in

nature. Since the oceanographic geospatial data movement is highly dynamic, the topology relations between the geospatial objects are also transient. As the wave/current propagates across the domain, the attribute (e.g., wave height) at a fixed position is also constantly changing.

4) Inexactness (uncertainty):

As Peuquet (2001) pointed out, there is inherent inexactness built into all spatial, temporal, and spatio-temporal GIS data models. The inexactness or uncertainty is from the artificial discretization of continuous phenomena in computer representation, both spatially and temporally.

From the spatial side, the inexactness is from the fuzziness of the boundaries or classes. For example, the spatial boundary between areas of different wave height zones is not so crisp. Boundaries can also be fuzzy in the temporal dimension. In the earlier example of high wave zone movement, there is no sharply-defined instant in time under usual circumstances when a high wave changes to a low wave. This change tends to be gradual. Besides fuzziness, there is temporal uncertainty usually derived from the lack of information (Peuquet 2001). For example, the oceanographic model data usually use even intervals in time to represent changes. The changes that happen in between the intervals are not represented and related information is missing.

Previous research (Goodchild et al. 1992, Huang and Lee 2009, Mark and Csillag 1989, Pebesma and Jong 2007, Wilson and Burrough 1999) on data models has tried to control or estimate the inexactness in the spatial and temporal domain. However, with all of the various approaches proposed, there are still no truly workable solutions that have been developed so-far (Hunter 1999). There have been very few researchers, who have addressed

the problem of handling uncertainty in spatio-temporal systems (Peuquet 2001). On a level of practical implementation, linear interpolation has been the predominant approach used for dealing with inexactness in existing space-time databases.

5) Identity persistence:

In the temporal data model research, another important issue is how to handle the persistence of dynamic objects in time, in other words, the data model's ability to manipulate dynamic object identities in time. In particular question that we need to answer is, how we define and track the dynamic object within its lifespan? One challenge in answering this question is, when does a "change" affect an object so as not to be called the same object anymore? The identity persistence plays an important role in data model design, and can heavily affect data storage, processes, and queries. Most of the time, identity persistence is application dependant. In oceanographic data, such questions could relate to the definition of the lifespan (start time and end time) of a high wave zone in a hurricane event. Or, if there is a spatial gap for a wave zone in successive time steps, do they need to be treated as the same instance?

Another critical issue related to identity persistence is that of splitting or merging (unifying) objects. For example, if a high wave zone splits into two or more zones, is it necessary to keep the original identity also for these new instances or entities? What about merging?

All of these fundamental issues or challenges discussed above need to be carefully studied and resolved, when designing new data model for oceanographic model applications.

2.4 Review on Spatio-Temporal GIS Data Models

The critical issues concluded in last section set the standards that require new data model to meet. First of all, with these main issues in the backdrop, currently available spatio-temporal GIS data models will be reviewed and evaluated in detail in the following sections.

2.4.1 Spatio-Temporal GIS Data Model Survey

Spatio-temporal GIS (or Temporal GIS) is a relatively young research area. The groundbreaking research in spatio-temporal GIS data model can be traced back to 1988 (Armstrong 1988, Langran and Chrisman, 1988). After that, especially in the last two decades, various spatio-temporal GIS data models have been proposed including the Snapshot Model (Armstrong 1988), Space-Time Composite (STC) Data Model (Langran and Chrisman 1988), Event-Oriented Spatio-Temporal Data Model (ESTDM) (Peuquet and Duan 1995), and the Three-Domain Model (Yuan 1994, 2001b). With the development of Object Oriented (OO) modeling technique in information science, more and more spatio-temporal GIS data models adopted the OO design concept. These models include: Spatio-Temporal Object (ST-object) (Worboys 1990, 1994), Spatio-Temporal Entity-Relationship (STER) (Tryfona and Jensen 1999), Object-Relationship (O-R) (Claramunt et al. 1998), Spatio-Temporal Object Model (STOM) (Renolen 1997), Geo-OM (Tryfona et al. 1998), and the Spatio-Temporal UML (STUML) (Price et al. 2000), etc. Meanwhile, a number of surveys or reviews about the spatio-temporal GIS data models, either in abstract or in a comprehensive form, have been published.

In the following section of this chapter, several typical spatio-temporal GIS data models are reviewed. The advantages, disadvantages, and main application fields are discussed.

2.4.2 Review on Selected Spatio-Temporal GIS Data Models

1) Snapshot Model (Time Stamp)

The Snapshot Model (Time Stamp) was the first one to be introduced and the most widely applied spatio-temporal GIS data model. It was proposed by Armstrong in 1988. The concept is simple and straightforward. In this model, temporal information is incorporated into the spatial data model by a sequence of time-stamped individual layers (slices). Each layer stores the spatial distribution of all available geographic objects. The layer index and order reflect the time values. Figure 2.5 demonstrates this model.

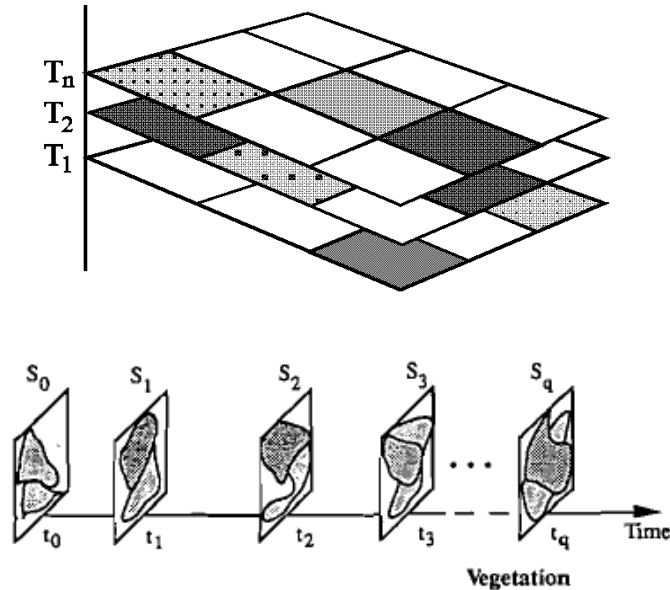


Figure 2.5 Examples of the Snapshot Model (Armstrong 1988). The top graph has constant time intervals between layers; the bottom graph has variable time intervals.

Basically this model simply extends the spatially based GIS data model from a 2-D to 3-D plane by adding time as the third dimension. Each cell within a separate snapshot contains the value for the corresponding location at that time, so the time related information can be easily accessed and retrieved. One big advantage of this model is that it can be conveniently transferred from traditional GIS data (e.g. satellite images).

The disadvantages of this data model are also obvious. First, it introduces a large degree of data redundancy, if the represented geographic phenomenon is not constantly changing everywhere. Second, there are no internal spatial relation logics upon the temporal structure in this model. Therefore, it is hard to track spatial changes in dynamic objects over time.

2) Space-Time Composite (STC) Data Model

This model represents the space as a set of spatially homogenous and temporally uniform objects in a single 2D space (a layer). It casts all objects, in all time slices in the Snapshot Model, on one layer and re-assigns attributes (including the temporal information) to each intersected polygonal mesh (Figure 2.6 shows an example).

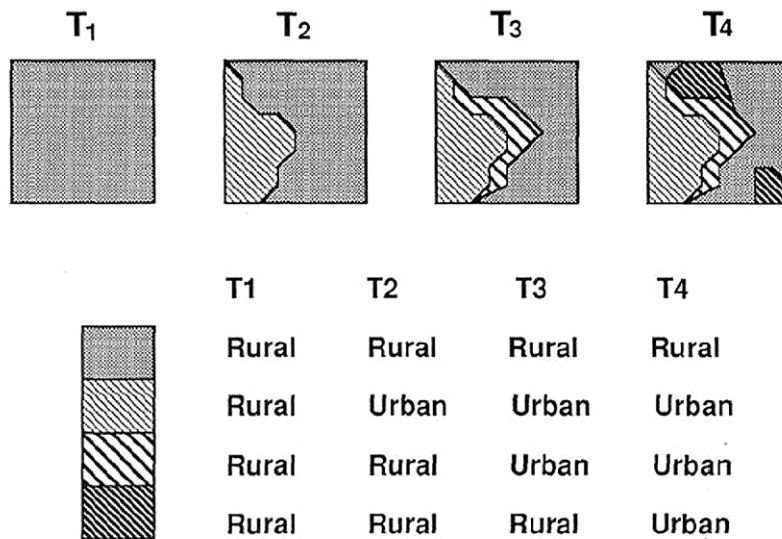


Figure 2.6 An example of the Space-Time Composite Data Model of urban/rural evolution. Each polygon has a distinct attribute related to its time.

The model uses polygon vector data as the basic data type. Compared with the Snapshot Model, data redundancy in the STC model is sharply decreased in most cases. But the data model may have difficulty in supporting the data set with multiple attributes or

objects having high frequent changes. It is impossible for this model to support constant changes. Since the object is divided into new objects each time change happens, there is no constant object identity. Also, the track of the geometry (shape, movement) or topology changes of the dynamic object is very difficult. Another disadvantage is that any updating of this model requires reconstruction of the whole data set.

3) Event-Oriented Spatio-Temporal Data Model (ESTDM)

Unlike the location/objects based models, the Event-Oriented Spatio-Temporal Data Model (ESTDM) stores changes triggered by events in relation to a previous state. A header file contains information about its initial spatial information, the pointer to a base map, as well as, to the first and the last event lists. The base map shows an initial snapshot of the thematic domain in a geographic area. Every event is time-stamped and associated with a list of event components to indicate where changes have occurred. The changes can be tracked by links between event lists. Figure 2.7 shows the main elements and the pointer structure in the ESTDM.

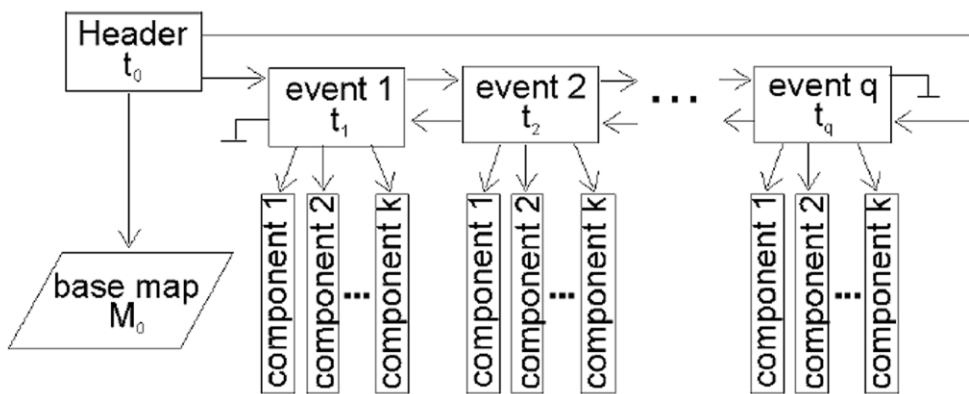


Figure 2.7 Primary elements and the pointer structure in the Event-Oriented Spatio-Temporal Data Model (ESTDM).

This model supports both spatial and temporal queries about attributes with a low data redundancy. The main disadvantage is that this model is based on a raster data model, and hence it is hard to maintain spatial objects' identities and track transitional information of an entity over time. It is possible to transform ESTDM to a vector data model based system, but this requires a substantial redesign of event components. And additional mechanisms are needed to allow event components to keep track of their predefined entities and locations (Pelekis 2005).

4) The Three-Domain Model

Yuan (1994) proposed the three-domain model to track wildfire events. In this model, the dynamic geographic processes and phenomena are represented by three separate domains: Semantic domain, spatial domain, and temporal domain, and the links between these domains (Figure 2.8). Unlike other domains, the separate semantic domain holds objects identities that correspond to human concepts independent of their spatial and temporal location. When implemented, this data model usually uses spatial trees to trace spatial changes.

The major advantage of the three domain model is that there are no pre-defined data schemata. Rather the model will dynamically link relevant objects from the three domains to represent a geographic entity or concept. Compared to other models, the Three-Domain Model is more efficient in tracking movement and attribute changes (semantic), which is an improvement over many of the pre-proposed models. Moreover, it supports more functional queries, such as both spatially and temporally involved queries.

Although this model claims to support spatial topology, there is no related implementation methods defined to deal with the spatial objects relationships. Also, the three domain model is good at supporting discrete changes, but cannot support constant changes.

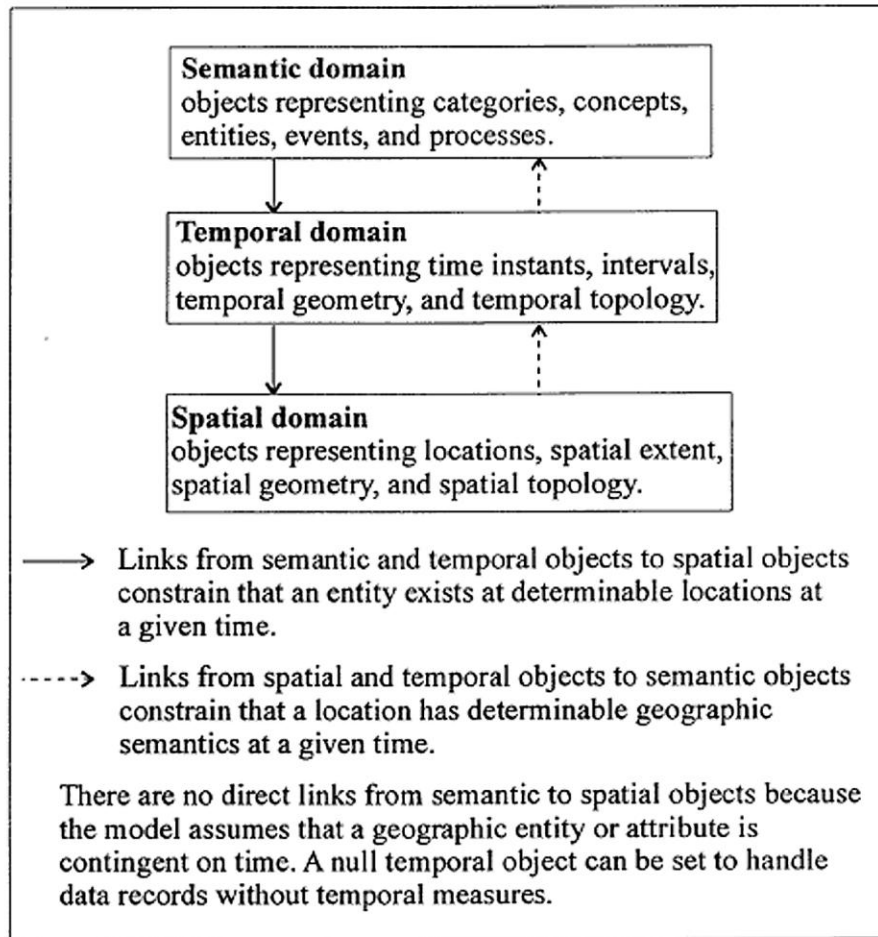


Figure 2.8 The conceptual framework of the three-domain model (Yuan 1996).

5) Objects Oriented Models

With the development of Objects Oriented (OO) modeling technique in information science, more and more spatio-temporal GIS research adopted OO concepts in model development. With the better representation and closer connection between object-oriented technology and geographic objects, the OO based GIS data models have shown strong advantages. The majority of recently developed spatio-temporal data models are based on the OO modeling technique.

Some typical OO based spatio-temporal data models include, the Spatio-Temporal Object (ST-Object), the Spatio-Temporal Entity-Relationship (STER), the Object-

Relationship (O-R), the Spatio-Temporal Object-Oriented (STOO), the Spatio-Temporal Object Model (STOM), the Geo-OM, the Spatio-Temporal UML (STUML), etc.

The typical features of the object-oriented technology include, classes and instances, attributes, interfaces, methods, abstraction and encapsulation, inheritance, polymorphism, dynamic binding, decoupling, etc. OO based spatio-temporal GIS data models incorporate these features during the development and inherit the advantages from OO modeling techniques. For example, the STER model uses three types of entities (spatial entity, temporal entity, and spatio-temporal entity), their related attributes, and relationships between entities as the basic elements. Figure 2.9 is an example. These elements can be cast into feature concepts in OO modeling, such as classes, attributes, and methods, etc.

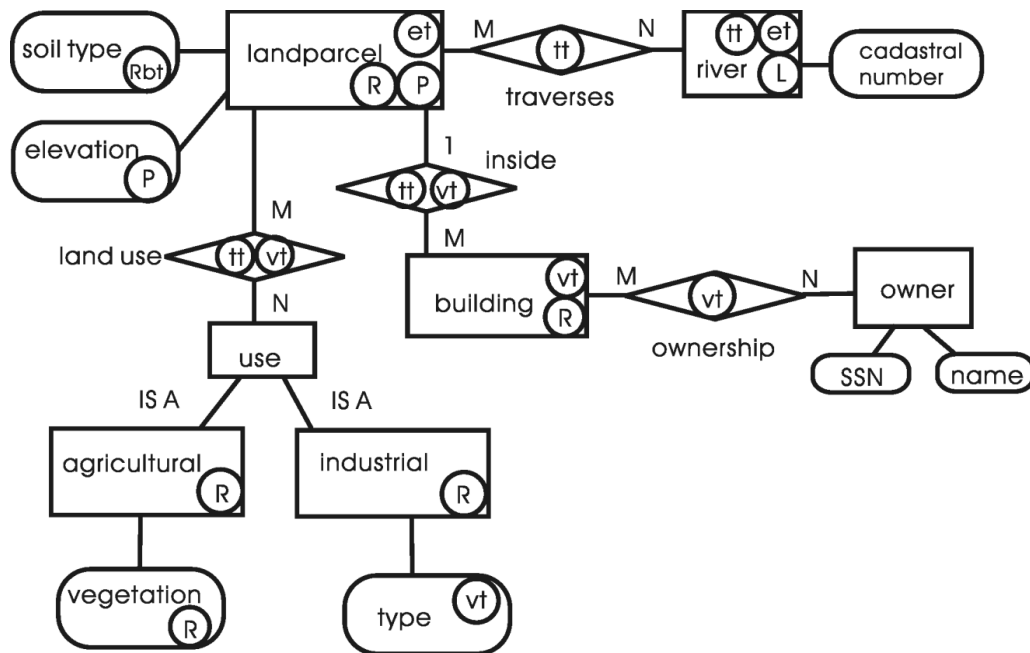


Figure 2.9 Example of the STER model for a cadastral application (Tryfona and Jensen 1999).

Montgomery (1995) concluded that there are four advantages of using the object-oriented approach in spatio-temporal data modeling:

- A single object can represent the whole history of an entity

- Queries can be simple and easy to implement, because they deal with each single object of an entity
- Efficient temporal data handling
- Uniform treatment of spatial and temporal data handling

In addition, most OO based spatio-temporal GIS data models also support all eight types of dynamic object changes.

Meanwhile, there are also some common disadvantages for these OO based models. First, these models usually support vector data very well. When representing raster (homogeneous) data with frequent change, the increased data volume and complexity make OO based models impracticable. Second, most of these models lack the spatial topology support and cannot handle the spatially related queries. Finally, most of these models only support discrete change or periodic change. It is still challenge to track the evolution of dynamic objects in space and time, which is important when representing constant changes.

Although there are many other more data models, the spatio-temporal GIS data models that were reviewed above represent most of the basic and typical models that are available today. Most of others are similar or related to the ones discussed here.

Since this dissertation focuses on oceanographic model data, the evaluation of those models that support all five fundamental issues, mentioned earlier, is most important. Table 2.1 provides a general, comparative view of the criteria.

As can be seen from the Table 2.1, the existing models cannot ideally support all requirements for oceanographic model applications. A new data model modified from the available models and that is dedicated to the oceanographic model data is thus necessary.

Table 2.1 List of capabilities and deficiencies of models in supporting oceanographic spatio-temporal considerations

	Homogeneous data	Constant change	Spatio-temporal change	Inexactness (uncertainty)	Identity persistence
Snapshot	support	not support	attribute change	good at spatial	not support
STC	not support	not support	attribute change	hard to support	not support
ESTDM	support	not support	attribute change	good at spatial	not support
Three-Domain	not support	not support	attribute, geometry, topology ¹	hard to support	support
Objects Oriented Models²	not support	not support	attribute, geometry, topology	hard to support	support

(1: Three-domain model needs to implement a mechanism in topology support; 2: Different objects oriented models vary in support of all features, the best performing models are included in this table.)

The objectives of this research are as follows:

- To study the oceanographic model output data (e.g., wave, water level) that possess a homogeneous characteristic and to acquire related knowledge about this characteristic, which can be utilized to develop a new data model.
- To design and develop a tentative/conceptual spatio-temporal data model dedicated to efficiently represent oceanographic model output data and similar geospatial phenomena as restricted application domains.
- Evaluate the new data model in an application and research the potential optimization methods.

2.5 Conclusion

In this chapter, the oceanographic model is introduced as a spatio-temporal application in the field of GIS. Some special characteristics of the oceanographic data, such as field-object characteristics, highly dynamic change pattern, are challenging features to be implemented into data models. After a careful analysis of the challenges and a comparative review of current spatio-temporal GIS data models, it has been determined that it is necessary to create a new spatio-temporal GIS data model to efficiently represent this special geospatial phenomenon.

Chapter 3 Ontological Analysis of Oceanographic Data

In order to model the world, we need to be able to represent it.

--Kemp and Vckovski (1998b)

In order to develop a suitable data model to represent a geographic phenomenon, the fundamental characteristic information of nature needs to be abstracted. Most recently, ontology has been brought into the discussion on modeling, as an important concept and tool to effectively address this issue.

Within a matter of a few years, ontology has become a hot buzzword in information science and GIS research and it has been mentioned in more and more research areas. Although the concepts and understanding of ontology is still in an ambiguous stage, riddled with a lot of controversy, diverse research and applications have been tentatively attempted. “Understanding of ontology in terms of its scope, details, and purpose varies significantly” (Green and Rosemann 2004). In this dissertation research, ontology is applied as a tool to help abstract oceanographic model data and then assist with data model components design.

This chapter starts out with a literature review on ontology. Then, an analysis of ontology for oceanographic data is discussed, and later ontologies for oceanographic data are documented.

3.1 Why Ontology?

Originally, ontology is a term from the philosophical realm and the initial concept of ontology is related to epistemology research: What is the essence of the physical world and what are the human conceptualizations of reality (Winter 2001)?

In information science (computer science), the ontology concept was first borrowed by Artificial Intelligence (AI) research by McCarthy (1980), and then gradually expanded to

other fields. As Fonseca (2007) concluded, the ontology research in information science can be categorized into two types, including *ontology of information systems* and *ontology for information systems*. The first type supports the validation of tools which are used to create conceptual models. The latter type is mainly concerned with the use of ontology to generate and validate information system components. This is how ontology has been integrated into this research.

Ontology is implemented to describe and explain a specific domain (e.g. sea state, physical geography) using concepts and relationship between concepts. These domains are converted to components in an information system. In this kind of application, ontology is a tool that helps describing a specific world (a domain that the conceptual model intends to represent). In GIS, the geo-ontology (ontology built for geography) should be the essential components of the logical universe for geographic data modeling.

Fonseca (2007) pointed out that “if we are able to build the theories (ontology for information systems) before starting the activity of conceptual modeling, we will build better models,” and “these are the first steps towards building ontology-driven information systems.” This statement clearly explains why ontology is so important in conceptual model development.

3.2 Ontology Applications in GIS

The three main application areas of ontology in GI Science that are apparent from a review of the related literature are:

- Ontology as a standardization procedure for easier translation/access between heterogeneous data.

- Ontology as a systematic approach to capture the universal concepts and meanings that defines the geo-spatial domain to prescribe suitable theories for the discipline.
- Ontology used for the design of data models and information systems to make them better equipped for handling geographic concepts.

In this research, ontology is applied to the data model design. The new data model is based on the concept of ontology and it also belongs to the type of ontology-driven geographic information systems.

Ontology-Driven Information Systems (Guarino 1998) are based on the explicit use of ontology at the development time and/or at run time, which translates ontologies into active information system components. In the specific case of GIS, and its adaption to Ontology-Driven Geographic Information Systems (ODGIS) (Fonseca and Egenhofer 1999), the system is built upon using software components derived from ontologies.

Conventional GIS deals with spatial objects and their relationships. Ontology can help the establishment of correspondences and interrelations among the different domains of spatial entities and their mutual relations (Smith and Mark 1998). Traditional approaches for data model development start from the implementation and representation point of view. Ontology affords the integration of information based on its semantic content instead of dealing primarily with data formats and geometric representations. ODGIS are built using software components derived from various ontologies. Usually these software components are classes that can be used to develop new applications. Being ontology driven, these classes embed the knowledge extracted from ontologies.

In previous ODGIS research, ontology is often employed as a method for identifying categories, concepts, relations, and rules. Ontology has also been employed to define and

conceptualize the knowledge in a domain to make it easier to model and to provide standardized vocabulary and rules for the application of this vocabulary (Agarwal 2005). In this research, proper ontology related to oceanographic data (significant wave height, water level, etc.) and their associated spatial and temporal information, which are required to define these dynamic data, are identified together with their relationships.

Ontology has a close relationship to object-oriented techniques being used in Information Science. ODGIS combines the use of object-oriented techniques and ontologies and delivers an enriched model to appropriately represent geographic entities.

In an ODGIS application, based on different criteria, ontologies can be classified into different types. Below, two classifications are introduced as they are more related to the new data model that will be developed in this research.

Guarino (1997) classified ontologies into four levels according to their dependence on a specific task or point of view:

- *Top-level ontologies* describe high abstract concepts. In ODGIS a top-level ontology describes top-level categories of general geographic objects and phenomena from philosophy. For instance, mereotopology is a theory describing parts and wholes, and their relationships to topology.
- *Domain ontologies* describe the vocabulary/terms related to a generic domain. These ontologies are more specific than the top-level ontology, but still are sufficiently generic to apply to the domain, such as ODGIS for remote sensing or the coastal environment.
- *Task ontologies* describe a specific task or activity, such as a cadastral system in ODGIS.

- *Application ontologies* describe concepts depending on both a particular domain and a task, and are usually a specialization of domain ontologies or task ontologies. In ODGIS these ontologies are created from a combination of higher level ontologies. These ontologies represent the user needs regarding a specific application.

Uschold and Gruninger (1996) distinguished another four types of ontologies depending on the kind of language used to implement them, namely:

- *Highly informal ontologies* are written in a human natural language.
- *Semi-formal ontologies* are expressed in a restricted and structured form of natural language (i.e. using patterns)
- *Formal ontologies* are defined in an artificial and formally defined language.
- *Rigorously formal ontologies* are defined in a language with formal semantics, theories, and proofs of properties such as soundness and completeness.

In this study, application ontologies and highly informal ontologies are applied to the marine environment to better analyze oceanographic data. These ontologies focus on explicating the actions, terms, and relations for the particular application, including oceanographic model data. In order to simplify the definition of ontologies, only the natural language is used.

3.3 Fundamental Concepts in Ontology

For ontology driven systems, several fundamental concepts or relations need to be studied first. In the following section, these relations are introduced and discussed specifically as they relate to oceanographic model data.

First, the ontology distinction for real world entities are identified as follows:

- Universals-Particulars

Universals correspond to the material concepts or patterns in the real world. They are related to types, species, or kinds. Correspondingly, the particulars are related to individuals, instances, etc. As to geographic phenomena, the earth surface with different land covers is universal. The specific area with a certain plant type is particular.

- Independent-Dependent

Independent means that entities can exist without the other entities' ontological support. On the contrary, dependent means that the entities' existence requires other entities' existence, like a pre-condition. For example, rivers are independent entities, while flooding is dependent on rivers. The dependent relationship can be spatial, logical, or any other predefined condition.

- Endurant-Perdurant

Endurant entities and perdurant entities define entities or processes relative to time. For spatio-temporal reality, an endurant entity exists through a certain period of time. In other words, the entity identity does not change for the whole time period. The building, mountain, or people are the examples of endurant entities. On the other hand, the perdurant entities or processes are recognized by the process in a certain period of time, but not by the status at any particular point in time. So, they have to be defined by the continuous time period, such as hurricane tracks or the temperature change during a day.

3.4 Ontological Considerations of Oceanographic Model Data

In this section, the relations mentioned above, based on the domain of oceanographic model data, will now be discussed in more detail. Since time is a major consideration in this

research, another important issue of time related ontology is the spatial and temporal granularity. This will be discussed first.

3.4.1 Spatial and Temporal Granularity

Scale is an important issue in GISs. Geographic categorization and classification is scale dependent. Geographic objects are tied not only to their spatial but also to their temporal context. The ontology consideration in the geographic domain relates to spatial and temporal granularity.

- **Spatial granularity**

Similar to geographic scale, different levels of ontologies can be defined based on spatial granularity. Accordingly, different levels of ontologies represent different information. Low level ontologies correspond to fine granularity related information, and high level ontologies correspond to more general (coarse granularity) information. In practice, granularity often reflects the specific ways of carving up domains of reality associated with different scientific theories (Grenona and Smitha 2004). In this research, levels of granularity are roughly equivalent to levels of details. The spatial granularity is related to scale, but not the same as the scale. The granularity is not just based on spatial scales. For example, the pattern of high wave zones with a big area indicates the dominant sea surface situation for the Gulf of Mexico. It is a continuous field of wave values with a specific range. The high level ontologies can correspond to such information. On the other hand, the specific wave value for a specific spatial point is a discrete independent value. It is not the same as the resolution concept of field data. A new low level ontology is needed to represent this kind of spatial point based undividable entity).

- **Spatio-temporal granularity**

When the temporal dimension is added, the spatial granularity is more complicated and turns into a spatio-temporal granularity. High level ontologies with coarse spatio-

temporal granularity could represent the main processes or the big entity changes. Low level ontologies with fine spatio-temporal granularity could represent small processes or small entity changes. For example, the track of the highest wave zone movement when a hurricane passes across the Gulf Mexico can be represented as a high-level ontology which reflects the hurricane impacts on large spatial extent and big temporal scale in human cognition. A different ontology (low level) is needed when representing met-ocean data change at a point in time or during a short time period at a specific location. The former one is related to a process while the later one is related to a static status or a short period.

In the following section, the relationships mentioned above based on the domain of oceanographic model data are discussed.

3.4.2 Universals-Particulars

For oceanographic model data, the universals are the met-ocean data (wave height, water level, etc.) distribution for the whole study area or a specific region, e.g., the whole Gulf of Mexico. It is a continuous field. The particulars are the met-ocean data values for selected spatial points, e.g., wave heights at a buoy location off the Louisiana coast. If time is considered, the universals are the change in met-ocean data patterns during a time period for the study area. The change in oceanographic data is a continuous dynamic process. The particulars could be the wave height change for a spatial point during a time period, or the wave height value for a spatial point with a certain time stamp.

Bona fide and fiat

From the ontology perspective, the field can be described as *bona fide* and *fiat* based on boundary conditions. *Bona fide* represents the genuine discontinuities of the geographic entities in the world, such as rivers, roads, buildings, etc. They generally have crisp

boundaries. *Fiat* represents the geographic objects which can only be identified in our individual and social conceptualizations, such as state borders, postal districts, soil types, etc. So, *Fiat* could have either crisp boundaries or transitional (also vague, fuzzy) boundaries.

Oceanographic data are viewed as *fiat* with a homogeneous distribution. There are no distinct changes and the boundaries are not clear or crisp, rather they are transitional. It is very hard to define a high wave zone with strict and clear physical borders. In contrast, coast lines can be treated as *bona fide*.

3.4.3 Independent-Dependent

Since the oceanographic data distribution is spatially continuous, they are dependent on ontological relationships. For example, high zones and low zones are inter-dependent and defined by each other. High wave zones (peaks) are generally surrounded by low waves (holes). Similarly, low waves are surrounded by high waves. Apparently, the high-low (peak-hole) relationship exists in a relative sense. Also, it has to be pointed out that in some instances, peaks and holes are formed near to the coast and in that case they are partially enclosed and influenced by land features also.

From a temporal perspective, the process of oceanographic data variability is independent from any other dynamic objects movement. For example, each high wave zone change/evolution is independent of other high wave zones. As mentioned in Chapter 2, the oceanographic data change is constant and continuous, so the dynamic object status at a point in time is inter-related to the data at prior/later points in time. In that case the relationship of the data with prior/later points in time is closely dependent.

3.4.4 Endurant-Perdurant

Based on the spatial and temporal granularity, the enduring and perduring for oceanographic data can vary. When considering wave data from a high level spatial view, wave entities are endurants, representing specific wave heights, whereas the existence of wave entities are persistent. From the high level temporal view, the track of the wave entity movement represents the high/low wave changes for a certain time period, and this can be defined as a process that is perdurant. When the granularity becomes fine, the wave field for a geospatial region is composed of points/grids of wave points. The evolving wave parameters at each point of the field and during a certain time period are a perdurant.

3.5 Ontologies of Oceanographic Data

Since the ontology in this research is an application level ontology, they are usually created from high-level ontologies. SNAP-SPAN (Grenon and Smith 2004) has been selected in this research as the high-level spatio-temporal ontology to be extended to the oceanographic ontology. In the spatio-temporal ontology research, SNAP-SPAN ontology proposes a top-level conceptual framework for enduring and perduring. SNAP includes enduring entities (endurants), such as substance, qualities, roles, functions. SPAN includes perduring entities (perdurants), such as processes and their parts and aggregates. SNAP-SPAN has been used as a basic framework to define sub-ontologies for geographic phenomena before. In general, the spatio-temporal ontology must combine these two distinct types of ontological models in order to support complex domains of the real world.

Figures 3.1 and 3.2 show the formal categories of SNAP and SPAN entities. Each SNAP ontology corresponds to a single point in time, while each SPAN ontology corresponds to a

single time interval. In other words, the SNAP is spatially-based, while the SPAN is temporally-based.

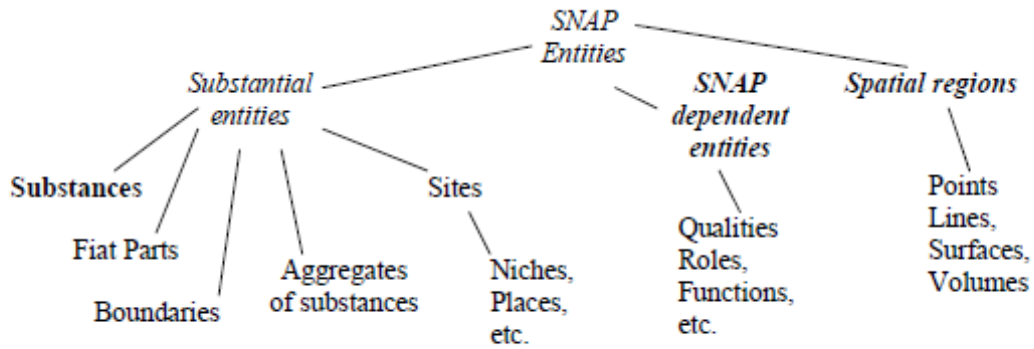


Figure 3.1 The main formal categories of SNAP entities.

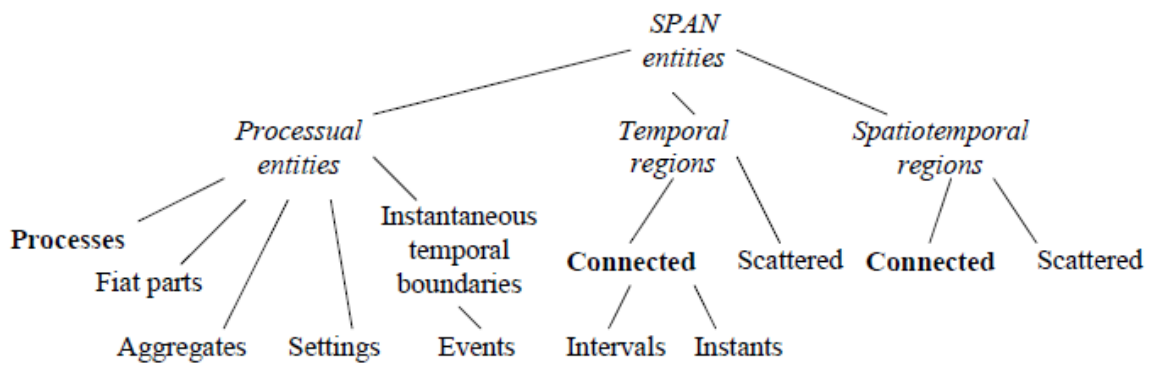


Figure 3.2 The main formal categories of SPAN entities.

In this research, the taxonomies of SNAP and SPAN are used and extended to support defining oceanographic data ontologies. The SNAP entities cover oceanographic data regions at specific points in time. SNAP supports both object and field perspectives of geographic phenomena. The SPAN entities cover dynamic oceanographic processes.

Figures 3.3 and 3.4 show the ontologies for oceanographic data, extended from SNAP and SPAN ontologies.

From a GIS perspective, fields are defined relative to their attributes. In SNAP, a field ontology recognizes one geospatial region with its associated distributions of attributes. These

attributes are recognized as SNAP dependent entities. Attributes in a field are necessarily attached to a specific part of the field and thus to a specific spatial location (Grenon and Smith 2004). When applied to oceanographic data, each ontology recognizes the water body in a geospatial region with a certain range of met-ocean parameter (e.g., wave height) values. These water bodies are substantial entities with fiat boundaries. There are three sub-type ontologies below water body, namely peak, hole, and transition area. They represent the high value, low value, and middle value of the parameters, respectively. Meanwhile, the three sub-type ontologies are dependent on each other. The water body at a spatial point possesses a fine granularity ontology, which represents the undividable, basic element of the SNAP entity. The aggregated ontology represents the water body with a large range of parameter values inside an expanded spatial region. One example could be a big peak water body with multiple small peaks and holes.

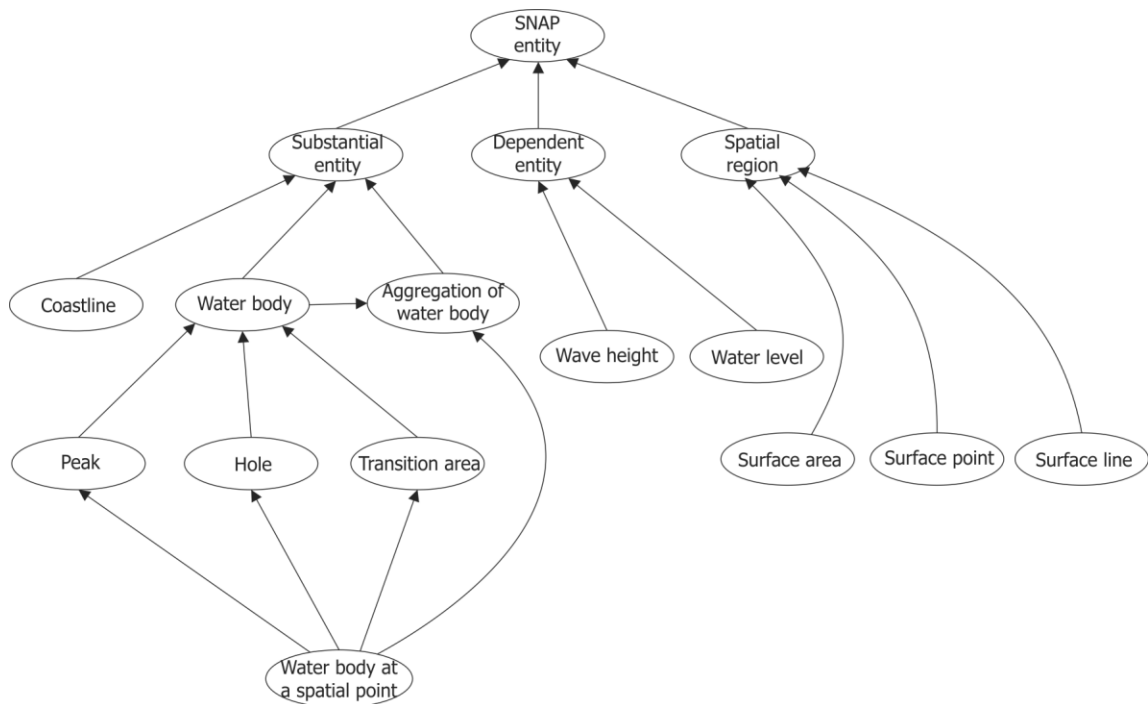


Figure 3.3 Ontologies for oceanographic data extended from SNAP.

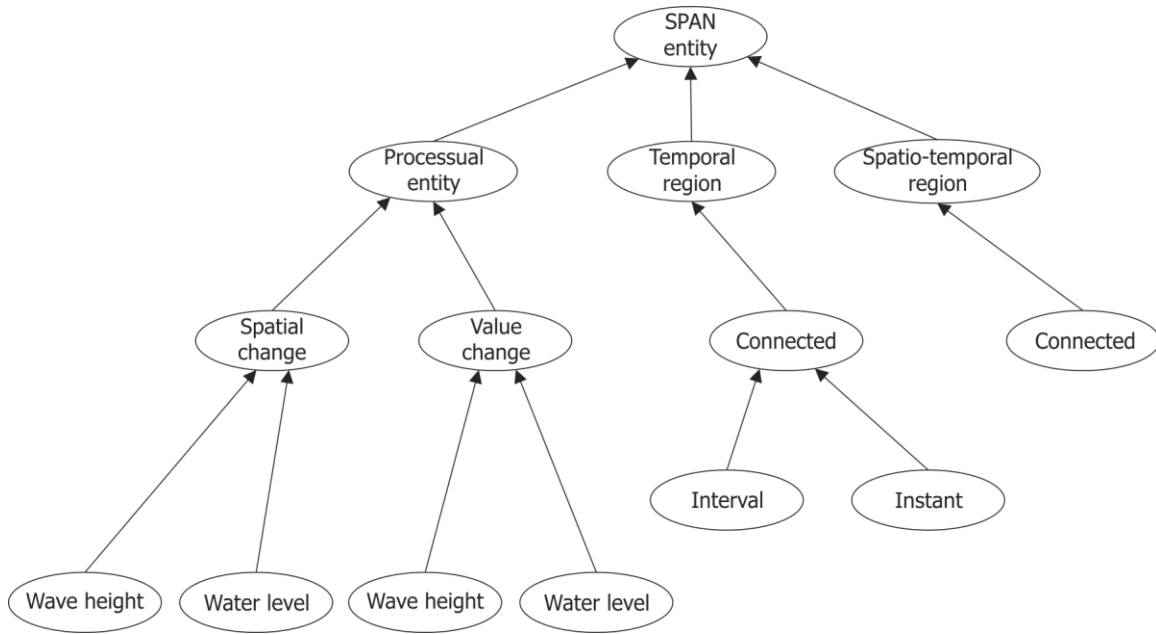


Figure 3.4 Ontologies for oceanographic data extended from SPAN.

In SPAN, either spatial or non-spatial changes can be represented. The processes are decomposed into different categories of substantial changes. Although these changes are associated with extended processes, which precede or follow them, substantial changes are always instantaneous (Grenon and Smith 2004). In other words, geographical changes of oceanographic data can be considered as a series of seamlessly connected instants, if looked at from a fine level of granularity. The “interval” and “instance” ontologies represent two statuses of change in the temporal domain. As discussed in the previous “Endurant-Perdurant” section, the model data changes can be defined in two directions, including spatial changes for a met-ocean parameter value and met-ocean parameter value changes for a fixed location. These two types of processes are represented by two ontologies, namely spatial change and value change. Each of these two ontologies is further subdivided into two types of different parameters including, wave height change and water level change. In case of the spatio-

temporal region, ontology of the “connected” is used to represent the constant change and the continuous field.

Defining ontologies may take a variety of forms. Obviously, an ontology needs a vocabulary of terms and some specifications of their meaning (Uschold et al. 1998). Another component of an ontology is “relations”, which represent the types of interactions between ontologies. Tables 3.1 and 3.2 list the ontologies for oceanographic data and their relations.

Table 3.1 SNAP ontologies for oceanographic data and their relations

Ontology	Specification of meaning	Relation to other ontology
Water body	Object concept of water with spatial extent at a static view	One of the substantial entities in oceanographic data
Peak	The high value zone from human perception	Part-whole of water body
Hole	The low value zone from human perception	Part-whole of water body
Transition area	Water zones between peaks and holes	Part-whole of water body
Aggregation of water body	The spatially aggregated water zone (peaks, holes, transition area, etc.)	Part-whole of water body and water body at a spatial point
Water body at a spatial point	The smallest spatial divide of water body entity with a spatial point dimension	Part-whole of water body (peak, hole, and transition area)
Coastline	Coastline represents the boundary of water body	One of the substantial entities
Wave height	One attribute of water body	Dependent from water body
Water level	One attribute of water body	Dependent from water body
Surface area	Spatial extent of water body	Dependent from water body
Surface point	Spatial point of water body	Dependent from water body
Surface line	Spatial line of water body and coastline	Dependent from water body

Table 3.2 SPAN ontologies for oceanographic data and their relations

Ontology	Specification of meaning	Relation
Spatial change	Spatial change of a water body with specific met-ocean data value or value range	Belongs to processual entity
Value change	Value change for a water body with fixed spatial point or spatial extent	Belongs to processual entity
Wave height (spatial change)	Spatial change for a water body with specific wave height value or value range	Belongs to spatial change
Water level (spatial change)	Spatial change for a water body with specific water level value or value range	Belongs to spatial change
Wave height (value change)	Wave height value change for a water body with fixed spatial point or spatial extent	Belongs to value change
Water level (value change)	Water level value change for a water body with fixed spatial point or spatial extent	Belongs to value change
Connected (temporal region)	Connected temporal region (continuous, not scattered)	Belongs to temporal region
Connected (spatiotemporal region)	Connected spatiotemporal region (continuous, not scattered)	Belongs to spatio-temporal region
Interval	Temporal region with duration between two points in time	Belongs to connected temporal region
Instant	Temporal region at one point in time	Belongs to connected temporal region

As mentioned in the first part of this chapter, ODGIS uses software components (classes) derived from ontologies to develop new applications. Sometimes, these components can be cast or transferred from ontologies directly. However, the defined ontologies in this chapter, particularly concerning oceanographic data, are not sufficiently clear enough to support all classes needed for the new data model development.

The main benefit of these ontologies is to come up with a logical tool to de-structure the spatial and temporal domains of oceanographic data. From the spatial point of view, the dependent relationships of peaks and holes are extracted from the homogeneous field data. From the temporal point of view, the changes are divided into two types of processes, including an ontology with spatial change and an ontology with value change. This spatio-temporal data classification and the ensuing improved data understanding would build the backbone for the new data model and can help substantially in the development of a new data model.

3.6 Conclusion

To improve the understanding of oceanographic model data and its spatio-temporal characteristics, ontology is employed as a method for identifying categories, attributes, and relationships.

As a first step, current ontology applications in GIS have been reviewed. Then, after the fundamental concepts in ontology as related to GIS have been introduced, the key ontological considerations are applied to oceanographic model data. Finally, ontologies for oceanographic data expanding on the SNAP and SPAN taxonomies are created for implementation.

The newly defined ontologies have provided further guidance in the design and development of a new data model for oceanographic data.

Chapter 4 Contour Tree-Based Spatio-Temporal Data Model

GIS generally uses two most fundamental data models in representing geographic phenomena, including field-based and object-based data models. The field-based model deals with evenly distributed grids with related attributes over a geographic region, while the object-base model deals with discrete, identifiable entities in the geographic region. Both models have led to two separate areas of GIS research, namely the definition of spatial operations on discrete objects, and the definition of operations on fields (Coppock and Rhind 1991).

Based on the ontology analysis of the oceanographic model data, provided in Chapter 3, the classified entities (ontologies) can be utilized to design software components to represent marine physical processes. Due to the complicated spatial and temporal change patterns of marine processes, the direct implementation by casting these ontologies to software components (classes) is still not capable enough to fully represent this special geospatial phenomenon. Rather it leads to the development of feasible new data structures and algorithms.

4.1 Main Algorithm for the New Data Model

With the ontology analysis of the oceanographic model data, some basic entities that may help build up data model are also retrieved. From the spatial perspective, an oceanographic model data can be represented by peaks, holes, and transitional zones (refer Chapter 3, Section 3.5, for these entities). These entities can be aggregated with topology relationships to preserve the distribution and variability of a parameter in the spatial domain. From the temporal perspective, a dynamic process can be represented by a series of instants correlated by incremental time steps, and the intervals between these instants.

The main algorithm will be based on contour tree and the OO modeling technique. The dynamic process is represented by a series of contour trees. Each contour tree stores the spatial and temporal information for a snapshot at a given time step. The spatio-temporal topology changes between contour trees are also stored in contour trees. Interpolation techniques are used to estimate the values not represented between contours on the same contour tree and between consecutive neighboring contour trees. The OO modeling technique is used to build contour trees as a data model.

4.2 Contour Tree Overview

An efficient way to represent continuous surfaces is by using isolines (contours). In oceanography, contours have been commonly used in representing bulk wave parameters in a map, including significant wave height, peak wave period, and other parameters like still water level. There are other set of vector data, such as current direction, wave speed, which are not suitable to be represented by contours, due to their directional properties. As mentioned in Chapter 2, in order to simplify the scope and challenge, this research focuses on the data model application to the scalar oceanographic data, which can be effectively represented by contours.

The contour tree is a fundamental and powerful data structure that represents the topological relations of contours on a map. It was first proposed by Boyell and Ruston (1963), in order to model and represent the relationships between contour lines. Another pioneer in contour tree research was Morse (1968, 1969). He used contour trees to find terrain profiles in a contour map.

The contour tree has been widely used in GIS, computer graphics, and scientific visualization research, such as, to index contours (Kreveld et al. 1997, Carr and Snoeyink

2003), to describe terrain (Sircar and Cebrian 1986, Takahashi et al. 1995), to detect features (Takahashi et al. 2004), to extract isosurfaces (Carr and Snoeyink 2003), to simplify data (Chiang and Lu 2003, Carr et al. 2004), to design transfer functions (Takahashi et al. 2004), and to extract contour properties (Kettner 2001).

Figure 4.1 is an example of a simple contour map and its corresponding contour tree. As the figure illustrates, the contour tree uses vertical levels to represent the value range (elevation ranges from 10-70m). The branch (leaf) ends show the local extreme values, with upward leaves representing local maximum values (A, B, and D), a.k.a. peaks and downward

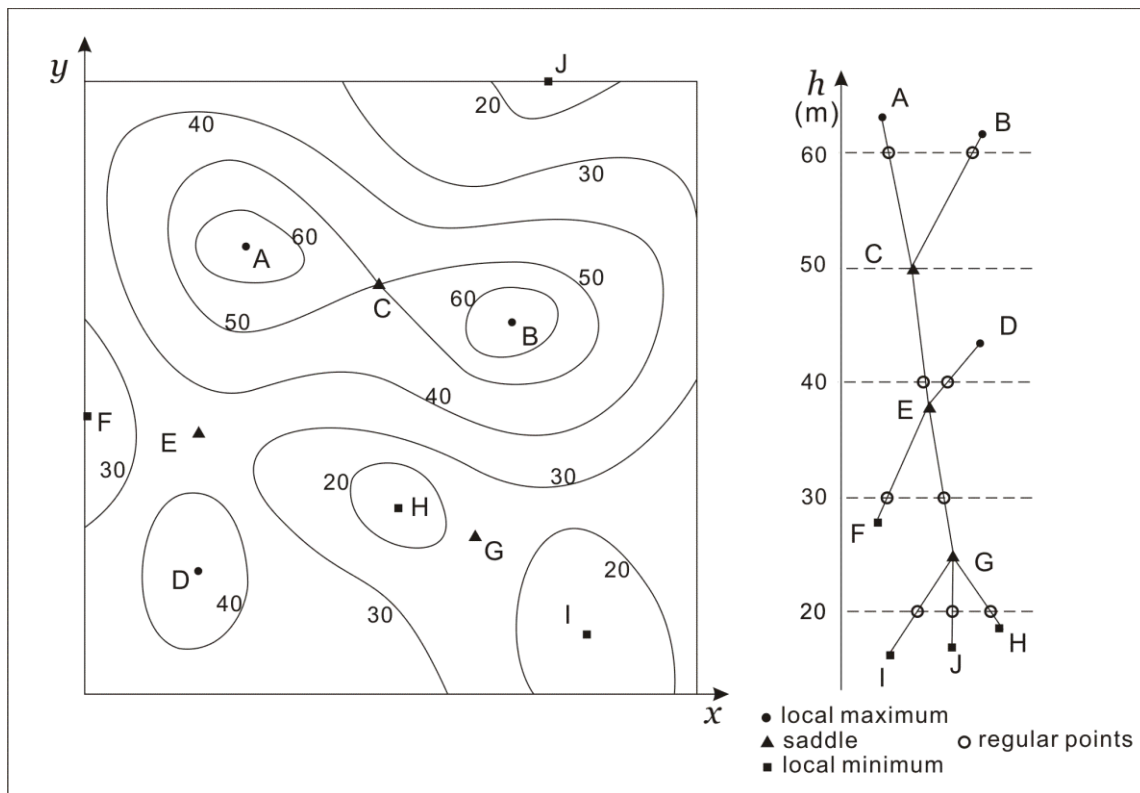


Figure 4.1 A contour map and its corresponding contour tree.

leaves representing local minimum values (F, H, I, and J), a.k.a. holes. These extreme values are connected at connection points (C, E, and G). It is called a “join”, if the points are connected at a lower point. Otherwise, if the points are connected at a higher point, it is called

a “split”. Both joins and splits indicate saddle areas on contour maps. All of these points are called *critical points*, since they capture the maps main topological changes. These points are connected by arcs, which span all interval values between each pair of critical points. Some contour trees also use *regular points* (circles in Figure 4.1) to store non-critical points to represent selected contour values between critical points. ²

From a contour tree, much knowledge or information about the represented geospatial data can be obtained. First, the number of nodes on a contour tree represents the total number of isolines on a contour map. For the same area on the contour map, more contours represent more detailed information. Secondly, the number of leaf nodes represents the number of local peaks or holes. Third, the level of the tree represents the range of elevation. When the contour interval is given, more levels mean more range of elevation. The values of these items can be used to discover new patterns, trends, and relationships that might be hidden deep within large topographic datasets. Basically, this information directly relates to the data volume and structure complexity of the final data model.

Various methods/algorithms have been introduced in computing contour trees for 2D or 3D surfaces (Bajaj et al. 1998; Carr et al. 2000; Pascucci and Cole-McLaughlin 2002; Takahashi et al. 1995; Pascucci et al. 2004; Takahashi et al. 2004). In general, it is not at all a problem to generate contour trees from the contour maps of oceanographic data.

4.3 Evolvment Analysis of Contour Trees

In highly dynamic oceanographic processes, the continuous spatial and temporal changes may involve topology changes. Allen (1983) and Claramunt et al. (1997, 2000) have

² Some research papers also use nodes for points on contour tree. Points are used here for better explanation. In order to discriminate the points used in this paper as spatial conception, nodes are used in the remainder of this paper for those points on a contour tree.

suggested several topology changes with temporal relationships. For oceanographic model data, it may not be necessary to make all of those topology changes as highlighted in those studies. However, the topology changes put forward by these authors can be categorized into five types and these changes can be tracked with the contour tree evolution over the modeling period. Figure 4.2 gives examples of the five changes with the left side showing the contour map changes and the right side showing the corresponding contour tree changes. Each change in this figure is explained below:

- 1) A new component emerges as a local maximum (peak) or local minimum (hole).

Example 1 in Figure 4.2 shows a new small high wave contour (F) emerging besides a pre-existing high wave contour (D). F and D are separated by a saddle G. F and D are further separated from the bigger contours (A and B) by a saddle (E). For the contour trees, a new branch (an arc connects node F) is added to arc ED at joint node G, which results in arc ED changing into an arc EG with two branches DG and FG.

- 2) An existing component (a local maximum/minimum) disappears.

Example 2 in Figure 4.2 shows a small high wave contour (B) disappearing besides high wave contour (A). The saddle node (C) also disappears accordingly. On the contour tree side, the branch (arc BC) is removed, then arc AC and CE transforms into arc AE.

- 3) Two or more existing components merge into a single one component.

Example 3 in Figure 4.2 shows two small high wave contours (A and B) merge into a new one (AB). The saddle node (C) also disappears. On the contour tree, the branch (arc BC) is removed, then arc AC and CE turn into arc A(B)E.

- 4) An existing component splits into two or more new components.

Example 4 in Figure 4.2 shows a small high wave contour (D) being split into two separate contours (D₁ and D₂) with a saddle (G). With respect to the contour trees, arc

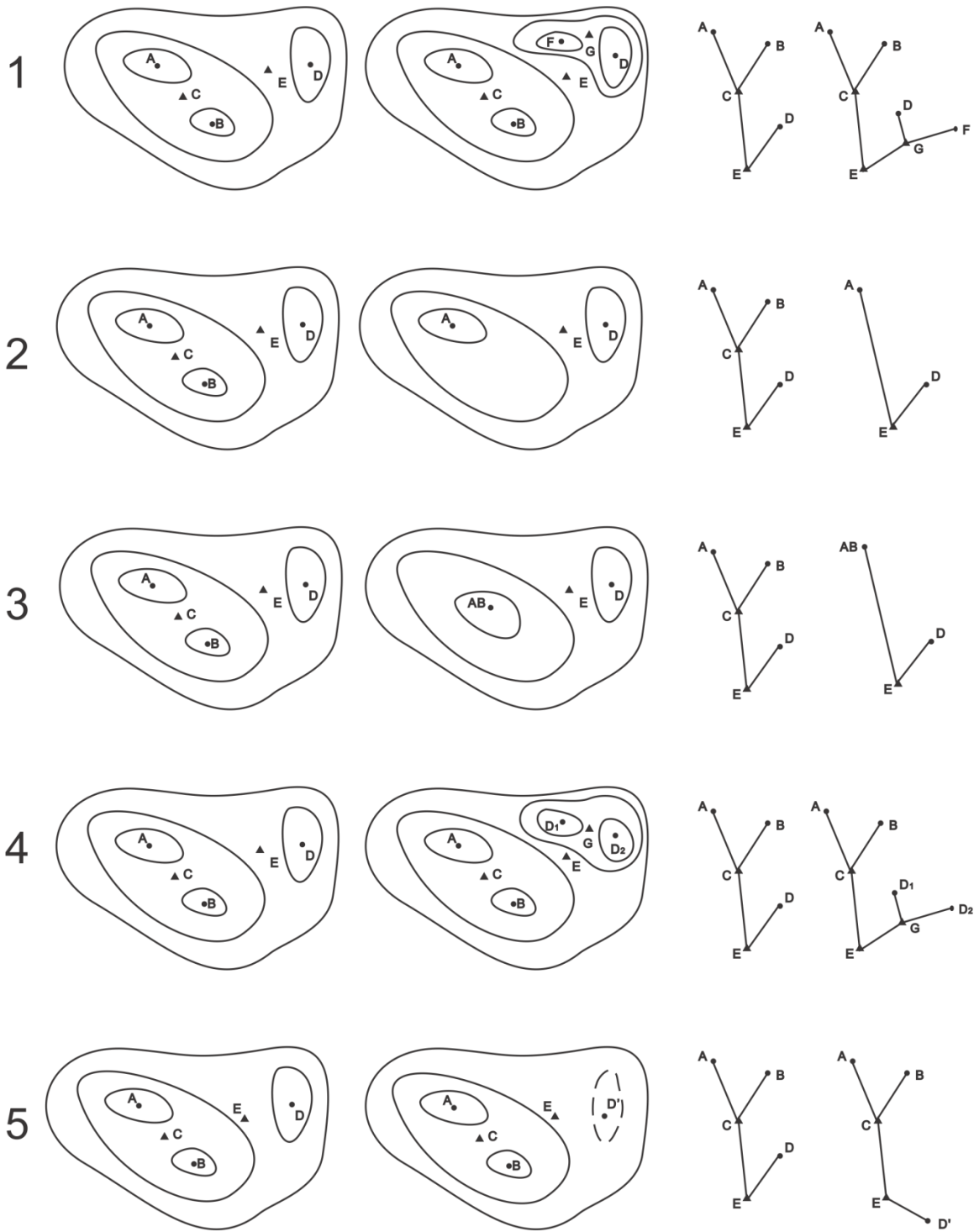


Figure 4.2 Five types of contour map transformation and their corresponding contour tree evolution.

ED turns into a tree with root EG and two branches D_1G and D_2G .

- 5) Topological changes without changing the number of components.

Example 5 in Figure 4.2 shows a local high wave contour (D) transforming into a local low wave contour (D'), which is similar to a peak turning into a hole. On the contour tree side, an upward arc ED changes into a downward arc ED'.

Despite the contour map transformations, if one only focuses on the contour tree changes, examples 1 and 4 reflect the identical results. The same is also true for examples 2 and 3. So, based on the above analysis and observation, the contour tree evolutions can be categorized and simplified into three basic spatio-temporal topological changes, including branch arc removal, branch arc addition, and branch arc direction shift.

- 1) Branch arc removal

The branch arc removal involves an arc with an end node and possible a joint node being removed. If there is more than one arc remaining after the arc has been removed, then the joint node is kept. Otherwise, the joint node is also removed. Figure 4.3 shows examples of these two situations.

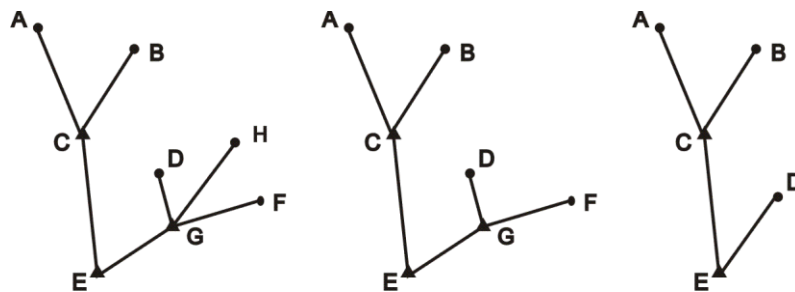


Figure 4.3 Contour tree branch arc removal.

- 2) Branch arc addition

Similar to the removal of arcs, adding a branch arc may or may not involve adding the joint node. The new arc can be added to a pre-existing joint node, or the new arc can be

connected to the middle of an arc with a newly created joint node. As Figure 4.4 shows, the two situations of adding branch arcs are the reverse processes of removing branch arcs.

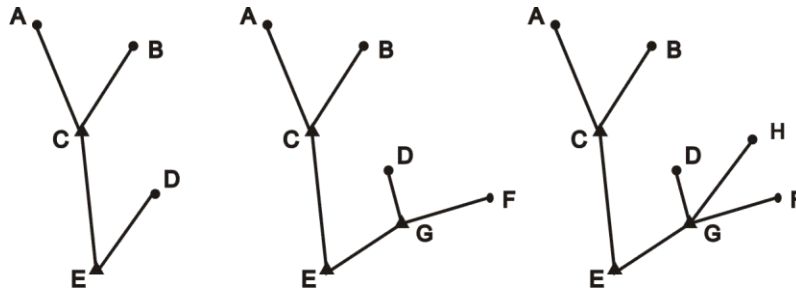


Figure 4.4 Contour tree branch arc addition.

3) Branch arc direction shift

The branch arc direction shift represents a branch end node's value change relative to its joint node. It may occur because the joint node's value changes or branch end node's value changes, or both.

The simplified contour tree topological changes help with the contour map evolution tracking.

According to the analysis above, the spatio-temporal changes of the contour map can be represented by tracking the contour tree changes. Another prominent issue is how to integrate the temporal information into the contour tree. Since the topology changes happen on the arcs with critical nodes, it is feasible to attach temporal information to critical nodes and share this information with other regular nodes on the same arc.

4.4 Temporal Information Integration

In the ontology analysis, the oceanographic data change processes composed of instant ontologies and interval ontologies. In this research, based on the information presented in the ontology analysis, a new spatio-temporal model is designed as a hybrid of the snapshot model and the event driven model.

From a temporal perspective, the time is represented by a series of snapshots of contour trees. Unlike the static view of each time slice, contour trees integrate the topology changes of contours. In the snapshot model, spatial relationships of objects between snaps (time step) are not recorded. Therefore, interpolation between consecutive snaps is needed. In the model developed in this research, similar to the event driven model, the temporal information for each process is integrated into the contour tree nodes and arcs. Each bulk wave entity's (the significant wave height is chosen as an example) variability is defined as a process. The starting time step is the first appearance of the corresponding nodes in the contour tree. The end time step is the last appearance of the corresponding nodes in the contour tree. The time period between the start time step and end time step is the lifespan of that parameter (here the significant wave height), which undergoes consistent transformation in space and time. If each node stores its start time step and end time step, the processes can be tracked. Figure 4.5 illustrates an example of the contour tree evolvment.

- The dynamic process is divided into time slices with even time intervals. Each time snapshot slice records corresponding dynamic information at that time step. A contour tree is created based on the corresponding snapshot geo-spatial information. At time t_0 there is only one peak (A) in the time snapshot map, which is shown as a leaf with end node A in the contour tree.
- At time t_1 , a local peak B appears besides peak A with a saddle C. On the contour tree, a new leaf B-C is added to the leaf with end node A. The time value t_1 is stored within nodes B and C as the starting point of their lifespan.
- As the time progresses from t_1 to t_3 , the contour map evolves. Since the topology relationships do not change, the corresponding contour tree's structure also does not

change. At time step t_4 , the peak B on the contour map disappears. So, the lifespan for leaf B-C is from t_1 to t_3 . The time value t_3 is stored within nodes B and C as the end point of their lifespan.

- All contour trees have leaf B-C during time steps t_1 to t_3 stored with their lifespan values therefore it is possible to retrieve the temporal evolution for leaf B-C from any contour tree.

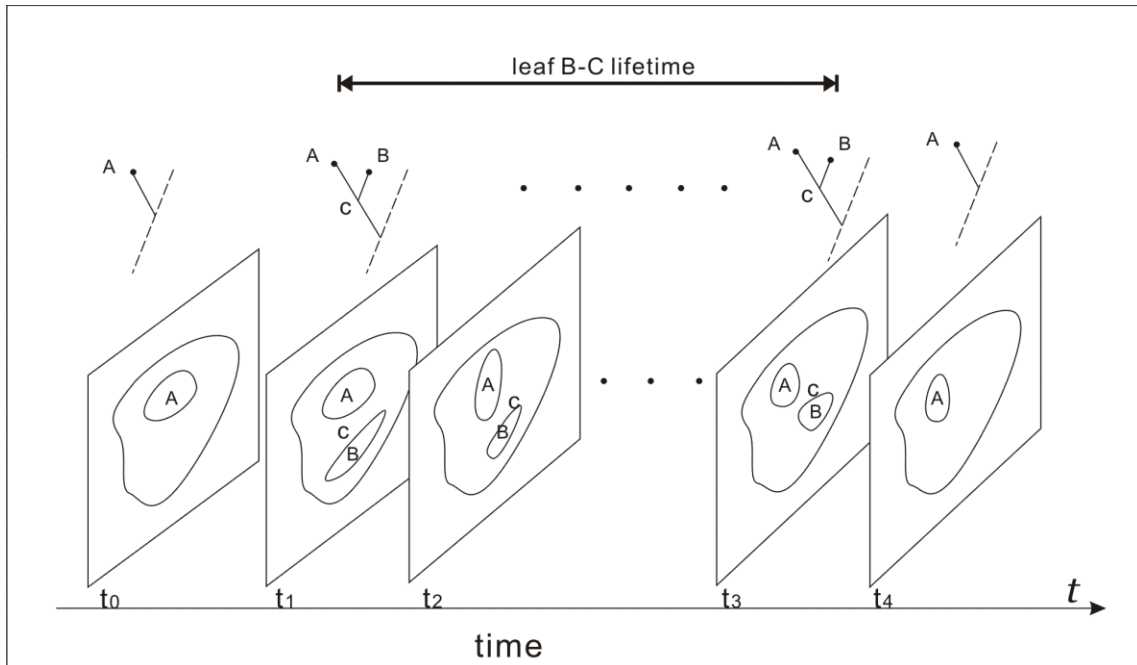


Figure 4.5 An illustration of how nodes store time information in contour tree evolution.

Except for the basic evolution case illustrated above, there are several special cases that need to be taken care of in this data model:

- Node/Arc only appears in one contour tree.

If a branch arc or a critical node just appears in one contour tree, its stored start time and end time are identical and which may denote this as a constant event. However, its lifespan may last from very short (an instance) up to two interval steps of consecutive contour

trees. Considering the continuous nature of oceanographic data, an average value is selected in this data model, which is one time step interval long.

- Arcs merge/Arcs split

As discussed in the previous section, topology changes may affect the identification of objects. But how can we define the merged new arc or the split arcs in relation to the original arcs? This is one of the identity persistence challenges for dynamic objects that are discussed in Chapter 2. In this research, new IDs are assigned to all arcs after merging or splitting events. So the lifespan values are updated and assigned to these new entities. However, how can we define the “new” objects at the application level, which is still independent of this algorithm. Based on the spatial and temporal information (model domain & start and end times), it is easy to identify the same objects at different time steps. For example, a high wave zone splits into two zones (one big, one small). If the big zone’s spatial area is more than a certain value (say 70% of the original one), then that zone is treated as having the same ID as the original zone. However, the small zone is assigned a new ID. So, the issue of identity persistence can be decided by the user at the application level.

4.5 Spatial Information Integration

In addition to the spatial topographic relationships stored in the contour tree, the spatial information of each object (critical and regular nodes in the contour tree) is also stored with each node in the contour tree. Due to the dynamic nature of the oceanographic data, the spatial information (spatial distribution or geometry) for each time slice for a certain attribute value (e.g., wave height) usually is not same. It is necessary to store the non-topology (spatial and attribute) information in each node (both critical and regular) on each contour tree. Then

the non-topology changes can be tracked by comparing this information on adjacent time slices.

For example, in Figure 4.5, node A included in the contour trees stores the spatial information for the local peak contour independently. The marked area for that contour differs and changes with each time slice (from time t_0 to time t_4), which represents the spatial change for the peak contour in time periods t_0 to t_4 . By connecting the node A on each contour tree as a time series, the continuous change in the spatial distribution can be approximately represented.

The basic idea of spatial information integration in this research is similar to the snapshot model, but using object view (vector data) rather than the raster data model.

4.6 Spatio-Temporal Interpolation

Data completeness, as one of the key components of data quality, is a very important issue in GIS. This concept is used to measure how many of the spatial objects and their attributes in the real world are represented in the data model. Brassel et al. (1995) gave a formal definition of completeness as “whether the entity objects within a data set represent all entity instances of the abstract universe.” Here, entity instances mean real world geographic phenomena and entity objects are their digital representation in the data model. In spatio-temporal GIS, data completeness includes both spatial and temporal considerations.

In spatio-temporal GIS studies, the effective representation of time is a fundamental challenge. In most temporal models, time is the fourth element on the 3D view of space. When the time is added to the GIS, one of the major problems in representing dynamic geospatial changes is the information completeness. The ideal model should store or represent all spatial and temporal information of changes linked to dynamic processes.

Even though the object view based model is good at representing discrete data, oceanographic model output data cover continuous space and hence, in order to represent the data not stored by the object model, spatial interpolation is needed. When the GIS data model represents dynamic processes, only data observed at a certain time are recorded. To represent the data falling outside of the stored time points, temporal interpolation is needed. The new spatio-temporal data model integrates both spatial interpolation and temporal interpolation.

4.6.1 Spatial Interpolation

From the definition mentioned above, spatial completeness measures the dynamic objects representation at a certain time step. In conventional GIS, spatial data completeness has been widely studied. To improve the data completeness of the continuous geospatial field, spatial interpolation has been widely applied. Spatial interpolation generally produces estimated values for unobserved positions based on the locations and attributes of pre-selected points in space. This type of spatial interpolation is termed “point-oriented interpolation”. The specific methods of point-oriented interpolation include linear, spline, stochastic, and proximal interpolation methods (Laurini and Thompson, 1992).

As to the contour map based interpolation, one of the most commonly used methods is Inverse Distance Weighting (IDW) developed by D. Shepard (1968). IDW methods are based on the assumption that the interpolating surface should be influenced most by nearby points and less by distant points. The interpolated surface is a weighted average of the scattered points and the weight assigned to each scattered point diminishes as the distance from the interpolated point to the scattered point increases. The value of each estimated point is calculated with the following formula:

$$z_j = k_j \sum_{i=1}^n \frac{1}{d_{ij}^\alpha} z_i \quad (1)$$

Where Z_j is the interpolated value, n is the number of scattered points in the set, Z_i is the observed value at the selected point in the data set, d_{ij} is the distance from the selected point i to the interpolation point j , and k_j is an adjustment parameter to ensure that the weights add up to 1. The parameter α is an arbitrary positive real number called the power parameter, which indicates that the rate of value diminishes with the distance. In a simple linear distance decay, $\alpha=1$ and K_j is:

$$k_j = \frac{1}{\sum_{i=1}^n \frac{1}{d_{ij}^\alpha}} \quad (2)$$

The main spatial interpolation algorithm used in this model is based on IDW. When interpolating a value at a point between contours, the points on enclosing contours are selected as the scatter points. The enclosing contours are the neighboring contours on either side of the point to be interpolated. Figure 4.6 shows two examples of the spatial interpolation method implemented in the model. In order to interpolate the value at point x , two closest points a and b are selected from the contours enclosing point x . Then, the value of x is calculated using the IDW formula mentioned above (left image in Figure 4.6). The right image in Figure 4.6 is another example with three contours enclosing point x . Similar to the

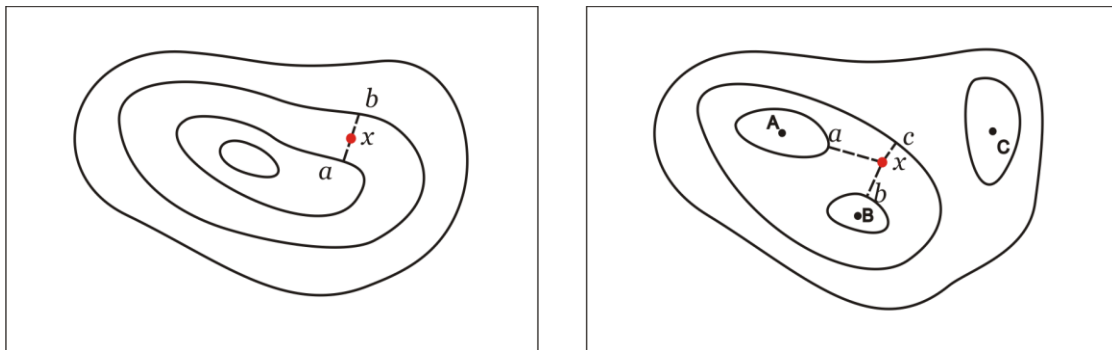


Figure 4.6 Examples of the spatial interpolation applied in the data model.

first example, the three closest points a , b , and c are selected from the contours enclosing point x , in order to interpolate the value at point x .

In this data model, the selection of IDW is based on the following considerations:

- 1) Unlike the general applications of IDW to get interpolation data for the area interpolation, the application in this model is for a single spatial point when a data query is requested. In other words, it is a partial area based interpolation. The resolution or spatial information richness in the wave map is based on the contour numbers and intervals. Accordingly, the enclosing contours involved in data interpolation have the major impact to the precision of the interpolated data quality. More contours with small contour intervals in the area to be interpolated can improve the interpolation precision. Based on the assumption that the interpolated point should be influenced mostly by nearby points and less by distant points, points on the contours closest to the interpolating point should be selected.
- 2) Fewer interpolated points will help the model performance for interpolation and hence the whole data process. Interpolation is a computationally intensive algorithm, especially for high data volumes. Reducing points with acceptable precisions is the first choice to improve the overall performance.
- 3) The data model represents the real world objects in an abstract way. One import rule is to keep the data model as simple as possible. A simple data model has more to do with the data manipulation and is expendable for more sophisticated algorithms.

4.6.1.1 Support Methods for Spatial Interpolation

When applying the algorithm for the point based interpolation, the first step is to determine the contours that enclose the points to be interpolated. This challenge is similar to

the point-in-polygon (PIP) issue in GIS. The two widely applied algorithms in solving the point-in-polygon issue are the ray casting algorithm (Sutherland et al. 1974) and the winding number algorithm (Leland 1975, Worboys and Duckham 2004).

The ray casting algorithm counts the number of times a ray starting from the point P crosses the polygon boundary edges. The point is outside when this "crossing number" is even, otherwise, when it is odd, the point is inside. The winding number algorithm counts the number of times the polygon winds around the point P. The point is outside only when this "winding number" is equal to 0, otherwise, the point is inside. Compared with the ray casting algorithm, the winding number algorithm is more computationally intensive, especially because there are more points along contours than there are along general polygon boundaries. The general winding number algorithm needs to iterate all boundaries of the polygon and may involve costly inverse trigonometric functions, but ray casting algorithm may filter out unnecessary boundaries first). Hence, in this data model the ray casting algorithm is selected. Figure 4.7 demonstrates how the ray casting algorithm works. The ray starting from point x intersects the contours C_1 , C_2 , and C_3 . The crossing

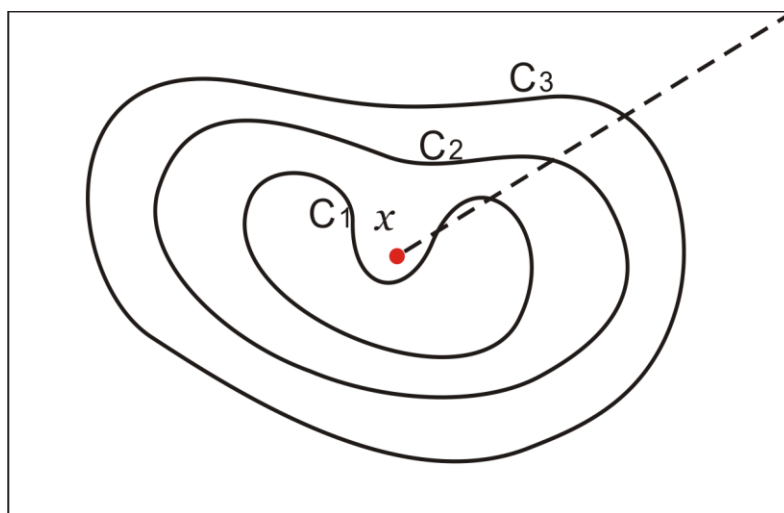


Figure 4.7 Ray casting algorithm.

number for C_1 is two, consequently, point x is outside of C_1 . For C_2 and C_3 , the crossing number is one, so point x is inside of both contours.

In order to accelerate the whole interpolation process, an initial screening approach is employed before the ray casting algorithm, to check whether a point lies in the contour's minimum bounding rectangle (MBR). In this way, the contours involved in the ray casting algorithm calculation can be reduced to a minimum applying this much simpler and faster method. Most of the time, there is only one contour that needs to activate the ray casting algorithm to determine the spatial enclosing relationship to the interpolating point. Figure 4.8 shows examples for the MBR application in this data model. In the left panel, in order to find the enclosing contours of point x , the smallest MBR that encloses x is assigned as contour C_1 's MBR R_x . Then, by applying the ray casting algorithm on C_1 and x , it is found that x is located outside of C_1 , so the enclosing contours of x are determined as C_1 and C_2 . In a second example (right panel of Figure 4.8), the point is inside of the contour and inside its MBR. After the ray casting algorithm is applied to C_2 and y , y is found inside of C_2 . So y 's enclosing contours are same as x , namely C_1 and C_2 .

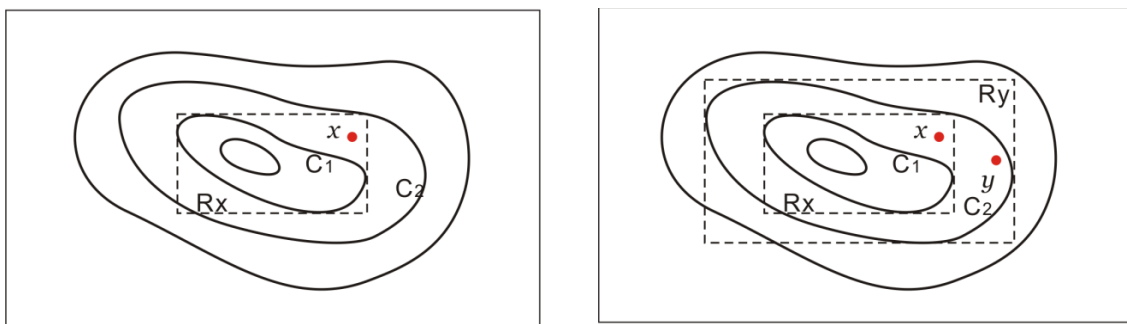


Figure 4.8 MBR algorithm used to speed-up processing time for the ray casting.

In summary, the main suite of algorithms implemented for spatial interpolation is:

1. Scan contour tree to find the contour with smallest MBR containing the point to be interpolated.

2. Apply the ray casting algorithm on the selected contour and point to be interpolated to determine if that point is inside the contour. Then, find the enclosing contours for the selected point.
3. Get the closest points for each selected contour. Apply the IDW algorithm to calculate the interpolated value for the selected spatial point.

4.6.2 Temporal Interpolation

Similar to the spatial completeness, the temporal completeness can be defined as whether dynamic objects within a temporal data set represent all changes of the entity, similar to the progression of time in the real world.

In general, when a GIS data model represents dynamic processes, only data observed at a certain time (snapshot) are recorded. In order to represent the data falling outside of a stored time step, temporal interpolation needs to be considered.

Like the spatial interpolation, the underlying idea behind temporal interpolation is estimating values for unobserved spatial data at time steps, not stored, based during a portion of or the entire modeling period. For two arbitrary time steps t_1 and t_2 , whenever $t_1 < t_2$, there is another time step v such that $t_1 < v < t_2$. Because of the temporal continuity of the data in a continuous world, two data values close to each other are more likely to have similar values than two data values that are farther apart. To improve results, it is suggested that temporal interpolation should be implemented based on the dense time model rather than the discrete time model. Most interpolation techniques are based on this assumption, and therefore only the closest data are used in the computational process or else, the closer data are assigned with higher weights.

The spatial interpolation techniques can be extended to the temporal domain by estimating a new value or set of values between two adjacent time steps. In recent years, many efforts have been put into the development of the temporal interpolation of geographic phenomena. Zhang and Hunter (2000) have discussed and summarized some techniques available for handling the temporal interpolation of spatially dynamic objects.

Most of the spatial interpolation techniques can be applied to the temporal interpolation realm also and time values are considered to be the equivalent of z values in the 3-D space. However, the significant difference between spatial and temporal interpolation is that, with the spatial case, the input parameters are the spatial location (x, y) and the outcome is a z-value. In contrast, for the temporal case, the input parameter is the time (t) itself and the outcome is the value v, or perhaps the spatial location (x, y) of the parameter which changes in time.

As discussed in Chapter 2, there are three types of temporal changes, including discrete, periodic, and constant. Discrete changes occur at particular instants and statuses remain fixed until the next changes. It is possible to explicitly record all changes with spatial and temporal information. Such as in cadastral systems, the land type or land ownership changes are based on human knowledge and can be changed instantly. Each change can be stored when it happens or it can be defined as having happened. In this situation, the temporal data completeness is easy to accomplish. As to periodic changes, since most of the time only the changes at the period boundaries are studied, the status between time boundaries (in period intervals) are considered as stable. In other words, the data completeness does not include changes that happened between boundaries. Therefore, the above two types of changes (discrete, periodic) are not a big concern for data completeness. In the case of

constant changes, it is hard to define boundaries, where at the exact instant time changes happen. In contrast the complete collection of data during the entire temporal period needs to be studied. Moreover, achieving data completeness is a much more challenging task, when comparing to other type of changes.

Although there are different ways to represent geospatial changes, spatio-temporal models can only store data at certain discrete time steps. For discrete and periodic changes, these models have the abilities to support the complete information or close to the complete information. If these models were to store the data at every time step when changes happen, the entire changes could be recorded with complete information. For example, changes in a cadastral system are discrete, if all changes at discrete time steps have been stored, and the cadastral status for other time steps are easy to be interpreted. As Figure 4.9 shows, changes happened at time steps t_1, t_2, \dots, t_5 . The value at time step t_x can be interpreted as the value at any time step recorded in the database at the nearest and earlier time step than t_x . In this figure,

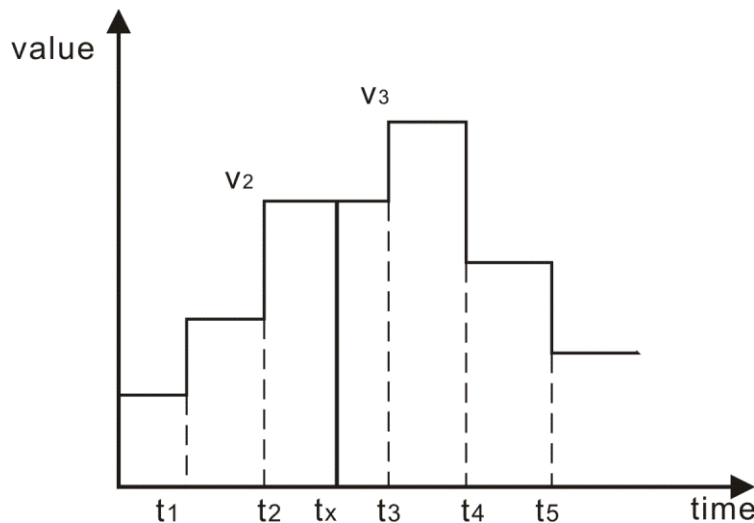


Figure 4.9 An Example of value assigned to previously recorded value in a discrete change process.

the value at t_x is the same as v_2 . This method has been discussed in Laurini and Thompson (1992) and Zhang and Hunter (2000). Using population change as an example for periodic changes, the population values between stored time steps usually are not studied. If necessary, the nearest neighbor values are usually used instead. In Figure 4.10, when the t_x is between t_2 and t_3 but closer to t_3 , then the value at t_x is set to v_3 .

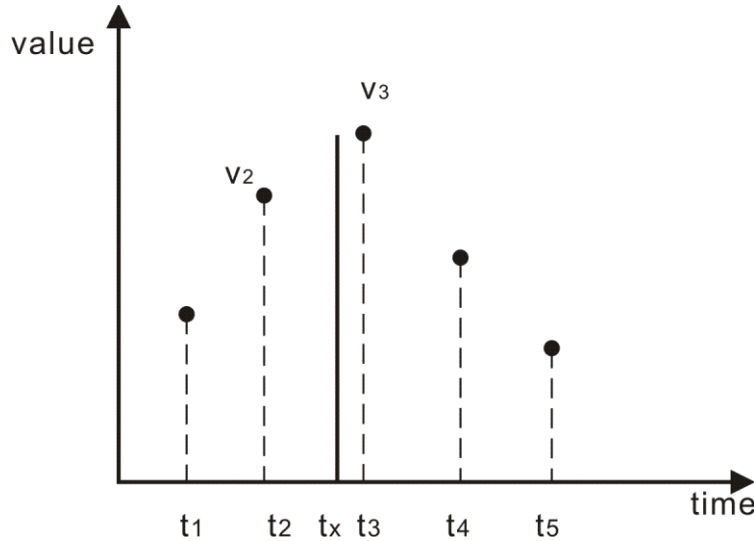


Figure 4.10 An Example of value assigned to the nearest value in a periodic change process.

When representing constant geographic changes, the problem of incomplete information is obvious. Fundamentally, the conflict exists between the linear character of constant change and the discrete records at disconnected time steps. A lot of research efforts have been expended trying to solve this conflict. For example, some researchers suggest blurring or ignoring the changes between recorded time steps. For example, Peuquet and Duan (1995) suggested that the data can be recorded at the time when a significant amount of accumulated changes have been reached since the last recorded change. But most of these approaches still cannot help GIS users to retrieve data, which are not recorded directly. For this reason and in order to redress this problem, suitable temporal interpolation techniques are required.

Similar to data models, current temporal interpolation research is either raster based or vector based. The raster data model based interpolations rely on grid-cell attributes, and therefore the interpolation is more attribute value based, rather than spatial object value based. For this type of data, the interpolation methods mentioned above can be employed directly and relatively simply. For example, we can apply these techniques to temporal interpolation of the thematic values of spatial objects, such as in determining the change in water temperature or in sea level at a specific location.

The vector data model based interpolations are more challenging, since most of the interpolation techniques are attribute based, such as the point based spatial interpolation method mentioned above. Generally, the vector data model is more suitable to represent isolate and independent spatial objects. The field data with constant changes are not suitable for vector data models.

Due to the object-field characteristics of oceanographic data, the interpolation could include both of these two characteristics.

As cited in Chapter 2, the spatio-temporal changes of spatial objects may involve geometry, topology, and attributes. Temporal interpolations can also include any of these three types. Object data models are more suitable to deal with geometry and topology temporal changes, while the field data models are more suitable to do attribute related interpolation.

In this dissertation spatial interpolation algorithms have been successfully integrated with the contour tree as the basic data structure. A similar strategy will be pursued for integrating temporal interpolation.

The fundamental purpose of the temporal interpolation in this data model is to calculate the oceanographic data values between time intervals not recorded by contour trees. Figure 4.11 illustrates the basic concept of a temporal interpolation. The contours' changes are stored in the contour tree at discrete time steps t_2 and t_3 , while a value at point P is needed at time step $t_{2.5}$. First, the values at time steps t_2 and t_3 are interpolated using the spatial interpolation technique mentioned above. If the values from point P are already stored in the contour trees, the interpolation is not necessary. The next step is to use temporal interpolation techniques to calculate the point P values at time step $t_{2.5}$.

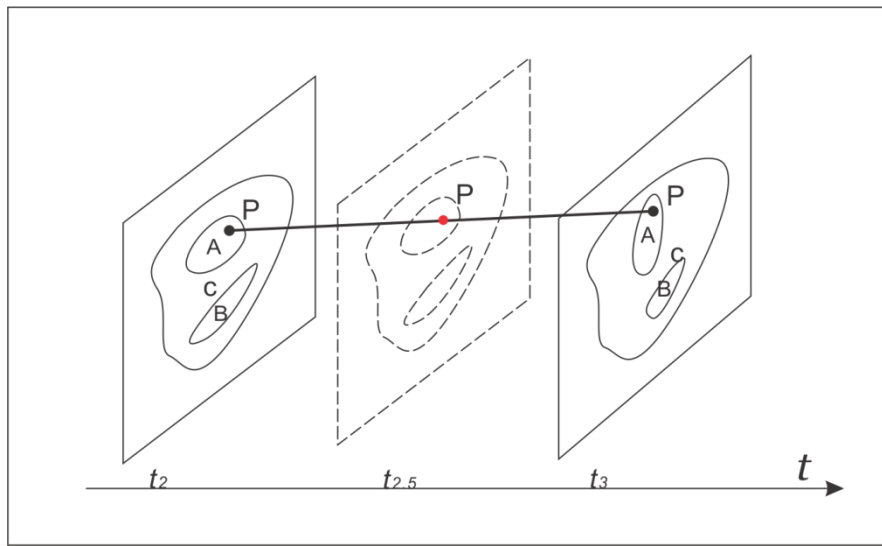


Figure 4.11 The temporal interpolation for a point at a time step not stored during the original model run.

Based on the review of possible interpolation algorithms, three of them are identified as potential solutions, including the nearest neighbor's value, the linear interpolation, and the spline interpolation.

1. Nearest neighbor's value

This method is similar to the nearest neighbor's value mentioned above. Figure 4.12 shows an example. In a continuous spatio-temporal change, statuses at time steps t_1 , t_2 , and t_3

are recorded. In order to interpolate a value V_x between t_1 and t_2 , the time intervals between t_1 to t_x and t_2 to t_x are compared. Then V_x is set to the closer value in time, which, in this example, is V_1 .

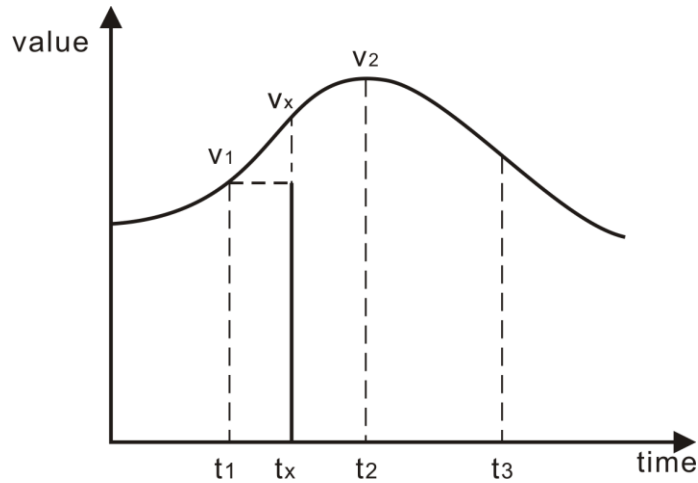


Figure 4.12 An example of the nearest neighbor's value algorithm (t_x is closer to t_1 and so V_x is assigned V_1).

This method actually is trying to treat continuous changes as discrete changes. The value changes between observed time steps are ignored. The advantages for this method are obvious. Since this method only calculates values based on time and not on the actual values, the algorithm is simple, hence it is easy to implement. This can be a disadvantage from the values' accuracy, because in continuous changes the values in the time interval are usually in between the boundary values. But in this method, the interval values are biased and assigned to either value on the time boundaries.

2. Linear interpolation

In this method the interpolated values are calculated as linear polynomials between values on time boundaries. As in Figure 4.13, the two known values are V_1 and V_2 at time steps t_1 and t_2 . In the orthogonal coordinate system the values are two points with

coordinates (t_1, V_1) and (t_2, V_2) . The linear interpolation is the straight line between these points. The value V_x at time t_x in the interval (t_1, t_2) is calculated with the equation:

$$\frac{V_x - V_1}{t_x - t_1} = \frac{V_2 - V_1}{t_2 - t_1} \quad (3)$$

Compared to the nearest neighbor's value interpolation, the linear interpolation gives more accurate values with a simple algorithm.

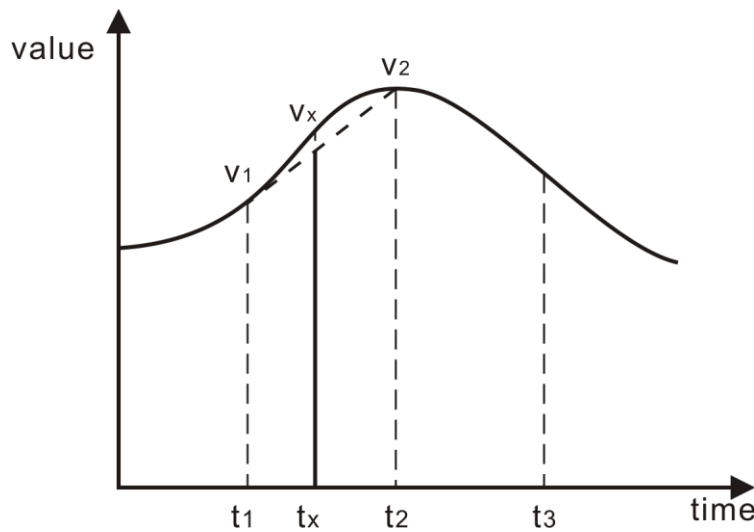


Figure 4.13 An example of linear interpolation (V_x is related to t_x 's distance to t_1 and t_2).

3. Spline interpolation

Spline interpolation takes a set of points and fits multiple piece-wise continuous functions (known as splines) to these points. The most common functions to fit are polynomials. Most of the time, spline interpolation is preferred over general polynomial interpolation, because the interpolation error can be kept small even when using lower degree polynomials for the spline. Thus, spline interpolation avoids the problem of Runge's phenomenon (1901) which occurs, when using higher order polynomials. The most common spline interpolations are linear, quadratic, and cubic splines.

In general, higher precision requires higher order polynomials. But higher order polynomials cause computations to increase significantly, thereby getting worse performance. In most circumstances, cubic function is the minimum number of polynomials to get a smooth curve over all data.

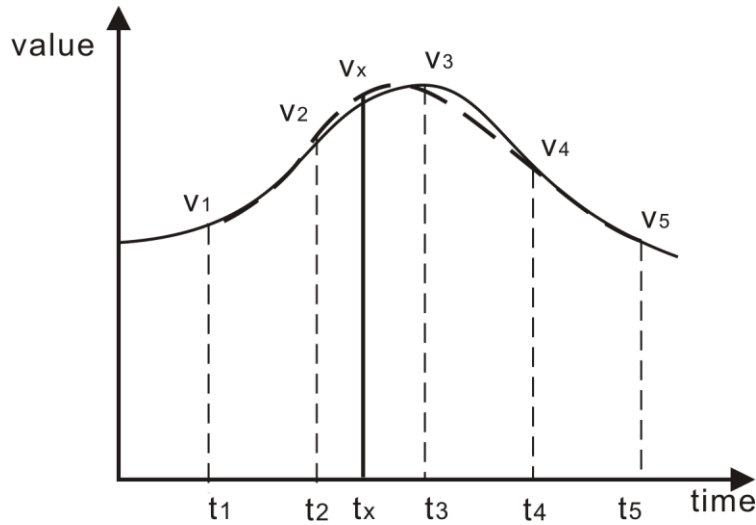


Figure 4.14 An example of quadratic spline interpolation.

Figure 4.14 is an example of the quadratic spline interpolation. The dashed line is the interpolated curve calculated based on five values: V_1 , V_2 , V_3 , V_4 , and V_5 . Except for the linear spline, spline interpolation results in a smooth interpolation among a series of data sets, and should have a better interpolation performance than the other two methods mentioned above.

In this data model, linear interpolation is selected as the interpolation method. The linear interpolation gives more accurate values than the nearest neighbor algorithm. It is a simple algorithm, which requires only a little bit more computing expenses. Spline interpolation may result in more accurate and smooth interpolated values than linear interpolation, but it needs much more complex computations, especially when using higher order polynomials.

The linear interpolation just needs two observed values in two consecutively known time steps, which are usually from before and after the interpolated time step. Spline interpolation usually involves more than two points to compute a reasonable result. That means data from more than two time steps need to be employed. In addition, the spatial interpolation for each point and at each time step, the computation expenses are significantly increased and the performance is heavily and negatively impacted. Based on these considerations, linear interpolation is implemented in this data model as the temporal interpolation method.

In summary, this data model integrates both spatial and temporal interpolation methods. The Inverse Distance Weighting (IDW) is applied as the main algorithm for the spatial interpolation. In addition, the minimum bounding rectangle (MBR) is also used to help the spatial interpolation performance. The linear interpolation algorithm is applied as the main algorithm for multi-time step data interpolation.

4.7 Conclusion

Driven by the ontology analysis of oceanographic model data, a new data model based on contour tree and OO modeling technique is proposed in this chapter. The dynamic process is represented by a series of contour trees. Each contour tree stores the spatial and temporal information for a snapshot at a time step. The spatio-temporal topology changes between contour trees are also stored in contour trees. Interpolation techniques are used to represent the values, which are not represented between contours on the same contour tree and between consecutive neighboring contour trees.

Chapter 5 Data Model Design Implementation and Evaluation

Chapter 4 introduced the main algorithm of the new spatio-temporal data model designed to represent oceanographic data. This chapter explores the proposed model's implementation in detail. In addition, an analysis of this data model in supporting special challenges from oceanographic data is evaluated. A case study will be provided as part of the evaluation of this model in support of spatial and temporal related query applications.

5.1 Evaluation of the Data Model for Special Challenges

In Chapter 2, some special spatio-temporal considerations for oceanographic model data have been discussed: (1) Field-object characteristics of oceanographic model data; (2) constant change pattern; (3) complicated spatio-temporal change types; (4) inexactness (uncertainty); and (5) identity persistence. These considerations are also unique challenges in designing new data models.

Some of these challenges, such as field-object characteristics, have already been discussed in previous chapters. However, other challenges still have not been addressed in details. After the main algorithm has been proposed, it is necessary to evaluate the new data model, especially if it can be shown that the challenges have been met by the new data model. The following discussion covers a detailed analysis of these issues.

1) Support of field-object characteristics in oceanographic model data

As mentioned in the previous chapter, this data model is based on the contour tree — an object view data model, which can be generated from the raw data. All contour tree nodes correspond to isolines, represented as objects. The topological changes and spatial movements can be retrieved in a practical way. But when representing the values not stored as nodes on the contour tree, this data model applies interpolation methods to calculate them. From a GIS

data view model perspective, these values are treated as field. Basically, they are point-based and attribute changes are more easily to be tracked than spatial geometry and topology. In this sense, this data model can reduce data redundancy and simplify the data process.

2) Support constant change pattern

Since it is an object-view based data model, a single object (node) can represent the whole history of an entity (isoline). By using snapshots to record changes in a lifespan with short time intervals, it can represent and track spatio-temporary topological changes. Adding spatial and temporal interpolation for missed values, the spatial and temporal attribute changes can be represented constantly.

3) Support complicated spatio-temporal change types (geometry, topology, and attribute)

Since the contour tree stores topological relationships, the topological change is easily supported. Each node on the contour tree stores spatial information and attributes, and changes in these two parameters are also tractable. Furthermore, the values of these three parameters are stored independently. If more than two types of changes happen simultaneously, then they still can be represented. Apparently, the previous statements only work for entities represented as nodes on contour trees. For the entities which cannot be represented by nodes, these change types cannot be represented. The only exception is that the attribute change for a fixed spatial point position can be represented by the integrated spatial and temporal interpolation.

4) Inexactness (uncertainty)

To solve the inexactness challenge in the spatial and temporal domain, this data model has integrated interpolation algorithms. The interpolated values are not the same as real values, but should be very similar, and eventually will increase the data quality.

5) Identity persistence

The review of spatio-temporal data models in Chapter 2 has shown the advantages of OO based models. One advantage is that one object can represent an entity and manipulate it during its whole lifespan. Since usually other entities or variables are dependent on this object, it is an important improvement in the data model design.

Another critical issue in identity persistence that has been discussed in the previous chapter is the splitting and merging of objects in contour tree evolvement. In oceanographic data, these issues are related to attributes and spatial values. The better solution is to let the application to deal with identity persistence in splitting and merging processes.

In brief, the new spatio-temporal data model can handle the special challenges from representing the complicated geospatial changes as related to oceanographic model data.

5.2 Spatio-Temporal Data Model Design

The main idea and key algorithms of the new data model have been discussed in the previous chapter. More detailed information about the model design will be discussed here. From the perspective of software engineering, the design is illustrated by class models with related functions and attributes. The UML (Unified Modeling Language) class diagram will be utilized in the introduction of the design implementation.

UML is a graphical modeling language and consists of different types of diagrams specifying the structure or behavior of a system. The most commonly used type of diagram in

modeling geographic data is the static class diagram. It describes the structure and static relations of a system (the data model in this research) by using classes, attributes, and methods.

The classes are the basic elements of the class diagram. Each class is represented with a box (rectangle) containing three parts: The name of the class, attributes of the class, and methods or operations that the class can take. The three parts are separated by lines and ordered from top to bottom. Among them, only class names are necessary. Sometimes in order to simplify a class diagram and show the most important parts of the structure, only class names and selected attributes and methods are shown in the diagram.

The UML class diagram in Figure 5.1 shows the conceptual design of the main class modules of the new data model. In order to keep the diagram concise and focused on the most important parts, only class names and important attributes are listed. The main methods and other related information are provided in each class's explanation.

When a user designs detailed modeling in a practical application, the class modules in the conceptual design are often split into a number of subclasses.

The following part is a brief explanation of the major modules in the new data model for oceanographic data.

- STobj

STobj class is the main entrance of the data model. It stores the whole time series of the contour trees in a list of ContourTree objects, and it stores all time spots with data in a list of TimeStep objects.

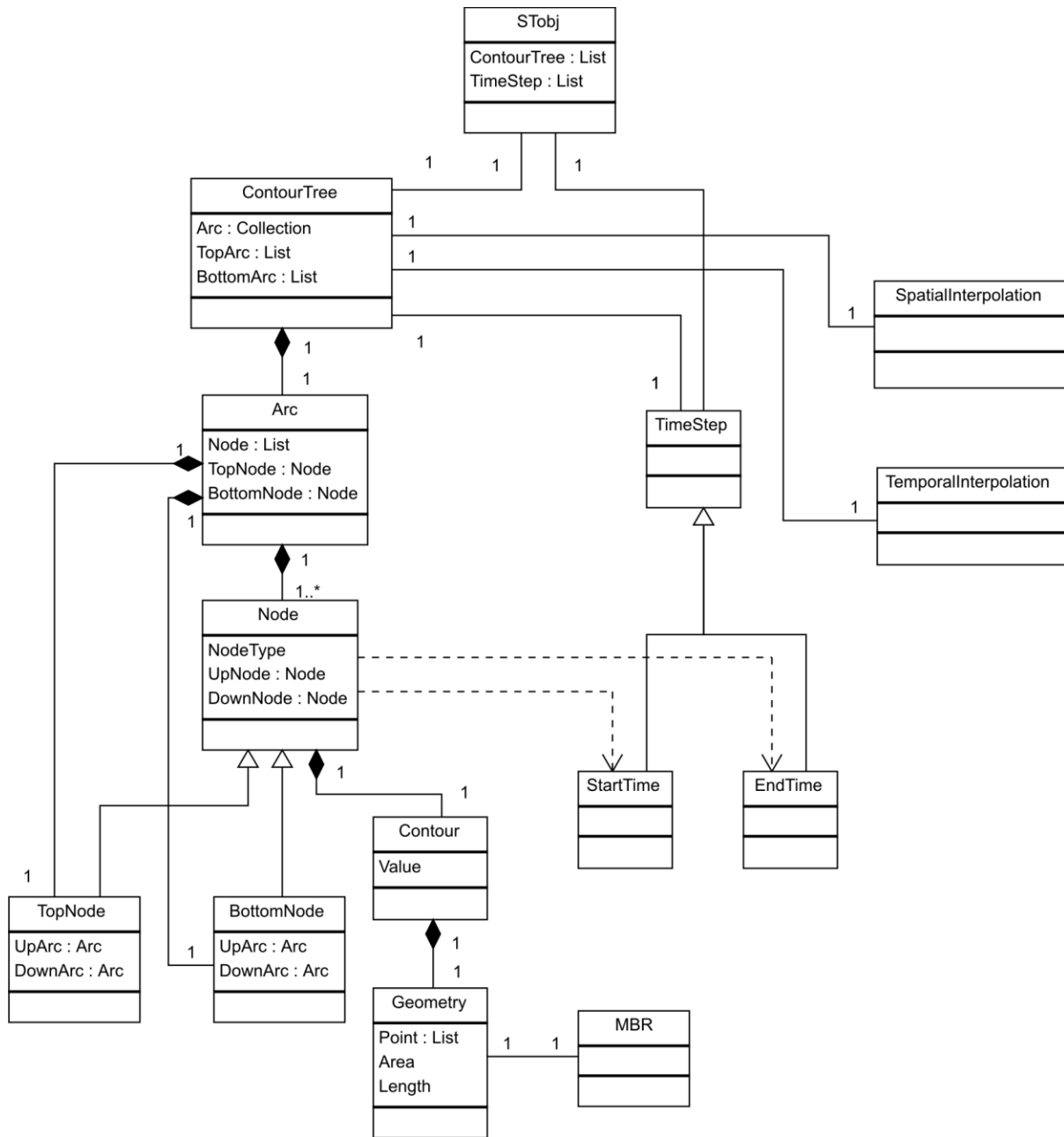


Figure 5.1 UML class diagram of the main class modules.

- ContourTree

The ContourTree class represents a contour tree at a time spot. It has a collection of Arc objects in a contour tree. In these Arc objects, the top and bottom objects in a contour tree are marked separately.

- Arc

An Arc class represents an arc on the contour tree. It includes a list of Node objects and functions to traverse these Nodes. Each Arc class also stores a TopNode object and a BottomNode object, which represent two critical nodes on the arc.

- Node

The node class represents a node on the contour tree. It stores node related information, like contour values and their spatial information (geometry, MBR, etc). The node class has two subclasses, including TopNode and BottomNode. These two subclasses represent the two critical points at the end of each arc. The two subclasses have some attributes and functions to store and visit their connected arcs. These are the upArc for arcs in the direction going upward and downArc for arcs in the direction going downward. In the example in Figure 5.2, Node C has two upArcs(AC and BC) and one downArc(CD) . Based on the analysis of upArc and downArc, the local extreme values (peaks or holes) and saddles can be retrieved.

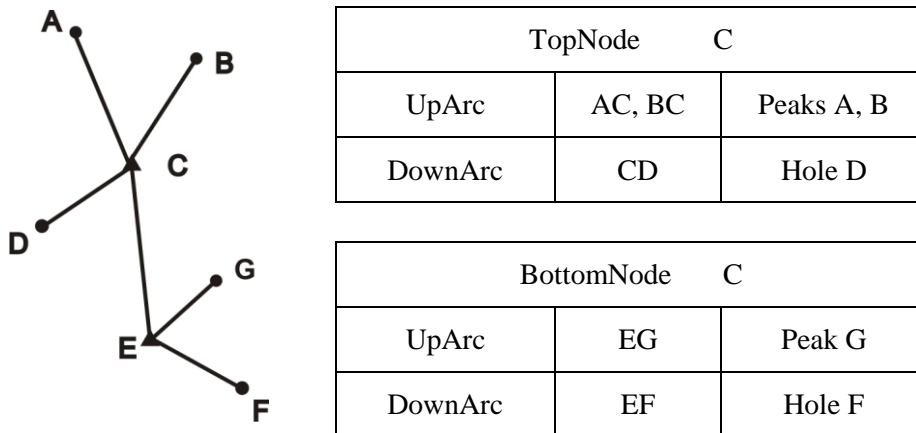


Figure 5.2 Critical nodes and related UpArc and DownArc in arc CE.

- TimeStep

The TimeStep class stores all temporal information related to Nodes and Arcs. The two subclasses, including StartTime and EndTime represent the start time spot and end time spot, respectively.

- SpatialInterpolation & TemporalInterpolation

These two modules include the main spatial interpolation and the temporal interpolation algorithms. These modules invoke related Arcs and Nodes in ContourTree with TimeStep to do the interpolation.

- Contour, Geometry, MBR

The Contour class stores a contour's attributes (e.g., wave height, water level, etc.). It also includes a class representing its spatial information, namely Geometry. The Geometry class stores a contour's spatial extent and related attributes, including area and length. It refers to the MBR class as the minimum bounding rectangle (MBR).

5.3 A Case Study for the Spatio-Temporal Queries

As Pelekis et al. (2005) pointed out, “a rigorous data model must anticipate spatio-temporal queries and analytical methods to be performed in the spatio-temporal Information System.” As part of the evaluation, a case study involving spatio-temporal queries in the application of the data model will be conducted in this subchapter.

Yuan (1997) argued that there are six major types of spatial and/or temporal changes in geographic information. A spatio-temporal data model should support some, if not all, of the queries for these types of changes. The queries of such changes can be categorized into three types, including spatially-based queries, temporally-based queries, and attributes/objects-based queries.

Spatially-based queries refer to queries for a fixed space, including queries of attributes at a specific point in time, or attribute variations during a specific time period. In oceanography research, a typical spatially-based query could be to find changes in the wave height during a certain period at a specific position.

Temporally-based queries refer to queries for a fixed point in time, including queries of spatial locations for certain attributes/objects at a point in time. In oceanography research, a typical temporally-based query could be to find the highest wave in an area at a specific point in time.

Attributes/objects based queries refer to queries of the spatial change and the temporal information of selected attributes/objects. In oceanography research, this type of query could be applied to find the wave height area, the movement track, and the spatial change, such as tracking the wave heights in the hurricane season. This type of query is the most challenging of all three types.

Most of the previously proposed spatio-temporal GIS data models only support part of the query functions. In particular, these models have difficulty in supporting complex spatio-temporal queries, especially the type of queries exemplified above as the third type (Yuan 1999).

The proposed data model in this research has the ability to support all three types of queries mentioned above. In this subsection, a case study will show how the new data model supports these queries. Figure 5.3 includes four images selected from the oceanography model MIKE21 output for the wave height and wave direction forecasting in the Gulf of Mexico. The model output includes a series of images representing the dynamic change during a specific period (from 03/26/2010 0:00 AM to 03/29/2010 12:00 PM in this example). These

four images capture the major changes in wave height conditions. High wave areas in these images are labeled by “A”, “B”, “C”, through “J”.

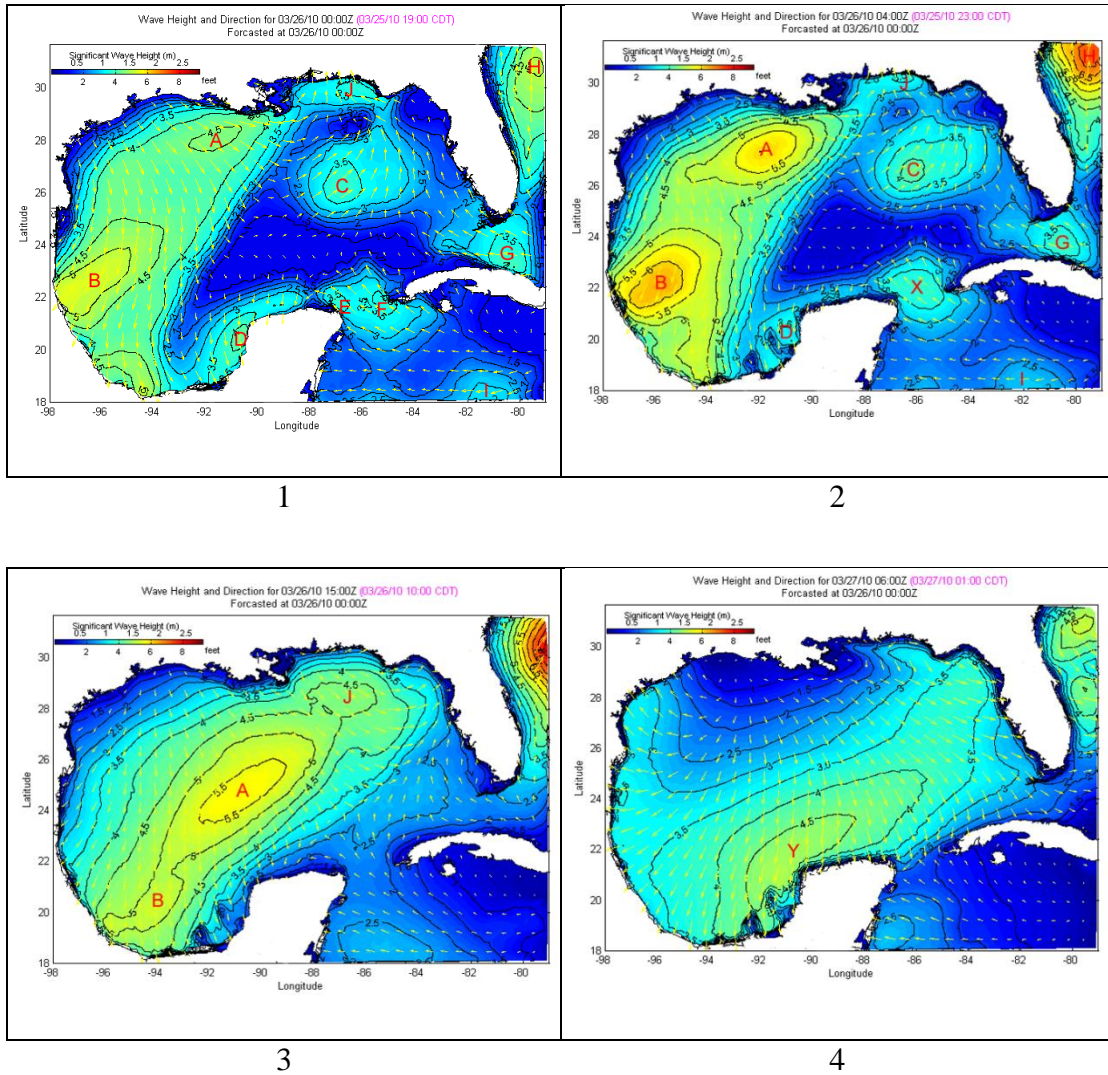


Figure 5.3 Four images selected from the output of MIKE21 wave forecasting model (the forecasting time period is from 03/26/2010 0:00 AM to 03/29/2010 12:00 PM).

The first query is a spatially-based query, namely find the wave height change at a specific position (e.g., at point “J” close to the Louisiana Coast) during the whole model forecast period. The main steps are listed below:

1. Get the contour tree for the first point in time and then find the contours enclosing the spatial point.

2. Perform the spatial interpolation to get the wave height value for the queried point, based on the enclosing contour values. The wave height for the initial point in time is acquired.
3. Apply step 1 and step 2 to the contour trees of the subsequent points in time, iterate these contour trees and get wave height values of the queried spatial point for the whole time period.
4. If the time interval between the points in time are bigger than required, the temporal interpolation for wave height values between adjacent points in times are necessary.
5. Traversing all selected values in time (from first to last) will show the wave height change at that position.

The second query is a temporally-based query, namely find the highest wave value in the Gulf of Mexico at a point in time as depicted in image 2 in Figure 5.3. The main steps are listed below:

1. Get the contour tree at the queried point in time.
2. Iterate peak nodes from this contour tree and then select the nodes from the region of interest in the study area (Gulf of Mexico). Location “H” is removed as it is outside the Gulf of Mexico.
3. Compare the peak nodes on the remaining branches of the contour tree and select the highest peak node.
4. Retrieve the spatial information (position, extent, etc.) for that node.

The last query is an attributes/objects-based query, namely finding the high wave area (e.g., wave height greater than 4 feet) and track their movement and spatial changes. This query includes the geometry and topology changes. In this case, there are two peaks wave

heights greater than 4 feet, including the areas labeled “A” and “B” in Figure 5.3. Both areas subsequently merge into one area, namely “Y”. The main steps are listed below:

1. Select the first point in time in the time series as the starting point for the query.
2. Select the contour tree at the selected point in time. Iterate peak nodes and check if there are contour values ≥ 4 feet. If there are no such peak nodes, retrieve the next contour tree for the next point in time and repeat the contour value search. Keep retrieving contour trees one at a time until peak nodes are found with a contour value ≥ 4 feet.
3. For the arcs that include peak nodes (step 2), select regular nodes with contour values equal to 4 feet. Iterate these nodes and discard nodes falling outside of the study area (Gulf of Mexico). In this example, “H” is removed as it is outside of the Gulf of Mexico.
4. If there are no available nodes left, search for the peak nodes with contour values ≥ 4 feet from the next contour tree for the point in time following (the point in time from) step 1, and continue with step 2 until available peak nodes are found.
5. Search regular nodes on the arcs with identified peak nodes (“A” and “B”) and select those nodes with contour values equal to 4 feet.
6. Get the temporal information (start and end times) of the selected nodes. Retrieve the same nodes on all contour trees between current point in time and end time. In this example, the end time is when peaks “A” and “B” merge into “Y”.
7. Continue from step 2 with the contour tree at the point in time when “A” and “B” merge into “Y”.

8. If the end time is reached or the required peak nodes cannot be found in the contour trees during remain time, then quit the search.
9. Compound the selected nodes for the entire time series and retrieve the spatial (position and extent) and temporal information.

From the above analysis, this data model shows strong support for spatio-temporal queries related to oceanographic model data. This data model can easily accomplish both simple queries related to static attributes and spatial positions and complicated queries related to changes (processes). The work flow chart in Figure 5.4 shows the main steps in processing the queries. The algorithms about contour tree simplification will be discussed in the next chapter.

The contour map in the case study includes relatively few contours and a simple contour tree structures. In oceanography applications, the contour maps may use numerous contours to represent detailed isosurface parameters, which incur far more complicated contour trees. Adding the topology tracking in time, the structure and relationships in the data model may grow extraordinary complex. In order to use the data model efficiently, one practical solution is selecting a modest contour interval value that can represent sufficient isosurface information with acceptable contour tree complexity. Another one is to perform the contour tree simplification which will be discussed in the next chapter.

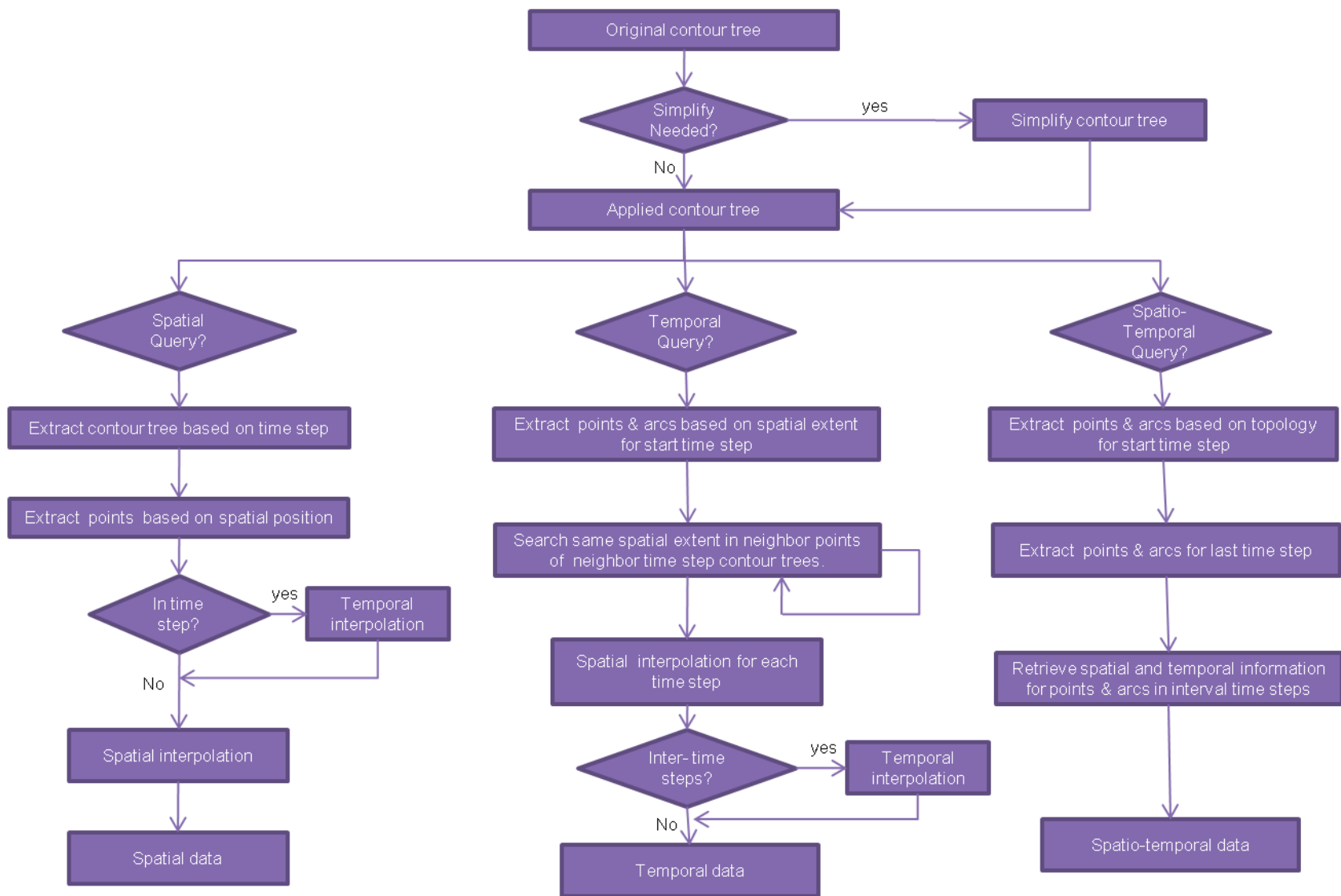


Figure 5.4 Data model query application work flow chart

5.4 Conclusion

This chapter first evaluates specific challenges related to the data model for oceanographic data and then introduces the main class modules of the data model using UML class diagram. Subsequently, the main attributes and functions in implementations are explored, and finally a case study is applied that illustrates how this data model supports complicated spatio-temporal queries in a real-world application.

Chapter 6 Contour Tree Simplification

Due to the existence of complicated geospatial phenomena in oceanography, the data model representing static isosurfaces can produce unmanageable large contour trees. When highly dynamic changes are involved, the temporal information to be integrated into contour trees will be even more complex. Small details increase the size of a contour tree dramatically. However, these details are not always necessary for automatic processing or visualization purposes. It is thus advantageous to simplify contour trees for practical applications in order to reduce its complexity, while keeping the fundamental structures of interest intact.

6.1 Review on Contour Tree Simplification

For different applications, several algorithms of contour tree simplification have been proposed. Carr et al. (2004) explained two basic operations in the contour tree simplification, such as leaf pruning and node reduction. Leaf pruning removes a leaf and the arc incident to the leaf from the contour tree. Removing an arc from the contour tree discards the corresponding contours from further consideration in the rendering process. Node reduction removes regular nodes without changing the essential structure of the contour tree. This does not affect the contours or values in the data set. Pruning and reduction are performed in an order that minimizes the error based on a local geometric measure, with node reduction to be carried out prior to leaf pruning. The geometric measures used in the contour tree simplification include persistence, volume, and hypervolume.

Pascucci et al. (2004) presented a multi-resolution data structure for representing contour trees and an algorithm for its construction. The multi-resolution data structure uses branch decomposition, an efficient way for storing a hierarchy of contour tree simplifications. Pruning a branch from the branch decomposition is equivalent to performing a node pruning

operation in the scheme of Carr et al. (2004). Takahashi et al. (2004) simplified the contour tree by pruning leaves to determine the “most important” isovalues for a volume rendering transfer function. The priority in leaves pruning is based on a new weight value. The new weight value is the product of the volume swept by the isosurface component in the subtree that is discarded and the difference in the scalar field between end critical nodes of the subtree. The new weight value is used as an important measure to simplify the contour tree. Saddles are processed until only a few of them remain, then a transfer function is constructed based on the simplified contour tree.

Zhou and Takatsuka (2008) presented an importance-driven method which combines different measures of importance into a single contour tree simplification pipeline through an importance triangle (ITri). The ITri is set up based on an importance measure vector, whose components include different measures of importance. In addition, the Extended Gaussian Image (EGI) and a map projection method are used to map the importance measure vector to a point onto a 2D plane, in order to differentiate branches of interest from other branches. The contour tree simplification depends not only on the priority value (e.g., the size of the ITri), but on every component of the importance measure vector. So the proposed approach of contour tree simplification is an importance-driven and user-directed process.

6.2 Temporal Information in Contour Tree Simplification

The methods/algorithms of contour tree simplification explored in the above examples only focus on the single static independent contour tree. These methods did not consider contour trees with inter connections, such as the temporal information integrated into contour trees as proposed in this research. These contour trees are connected and organized in a group to represent a series of time-varying data from complex geospatial phenomena. The need for

simplification (contour tree pruning) is rather challenging. In this research, the simplification methods are considered from two sides, namely from the spatially-and the temporally-based sides.

6.3 Basic Simplification Operations

The basic simplification operations define the most direct and fundamental operations on contour trees. The two operations proposed by Carr et al. (2004), leaf pruning and node reduction, are used in this research. Leaf pruning removes a leaf of the tree, changing the tree structure and reducing the complexity of the tree. Node reduction deletes a node (critical node) from the contour tree without changing the essential structure of the contour tree³.

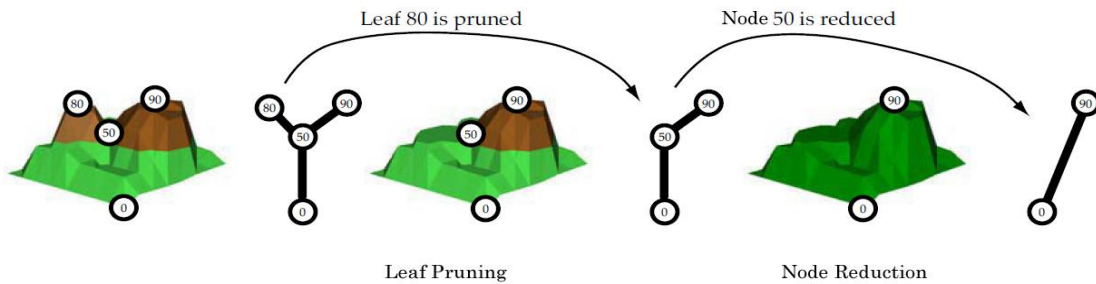


Figure 6.1 Leaf pruning and node reduction (Source: Carr et al., 2004)

Figure 6.1 shows the example of leaf pruning and node reduction. As indicated in this figure, a leaf (branch arc) #80 is pruned from the tree on the left to produce the tree in the middle graphic. The node #50 is removed from the tree in the middle graphic to produce the tree on the right side. After the leaf and node are removed, the left two arcs merge into one new arc. Through these basic operations contour trees can be simplified to any desired size.

³What is taken off here is not the regular node but the critical node. The node reduction only applies to the critical node, not the regular node. This usually happens after the leaf pruning, without directly changing the structure of the contour tree. However, it changes the contour tree structure along with leaf pruning.

6.4 Rules of Simplification

In Carr et al. (2004), and Zhou and Takatsuka's (2008) methods of contour tree simplification, the rules of pruning leafs/arcs and nodes are based on different measures of importance. Carr et al. (2004) used local geometric measures (perimeter, cross-sectional area, volume, and surface area) separately to define the importance values. Zhou and Takatsuka (2008) combined different importance measures together into a single value.

The importance of each contour or region is measured depending on the application (Zhou and Takatsuka 2008). For example, in 3D visualization, the importance is related to the visibility of an object within the volume data.

In this research, the simplification method is also based on the importance values. The importance values are measured based on each leaf (branch arc) of the contour tree. In addition, the simplification method introduces the temporal dimension into the importance value. The importance values used in this research are “contour value difference”, “base contour area”, and “contour profile”. The contour value difference is the difference between the absolute parameter values in end nodes in the branch arc of contour trees. It is defined as:

$$f = (V_{\max}) - (V_{js}) \quad (1)$$

with f : the contour value difference; V_{\max} : the local extreme value; and V_{js} : the value at joined or split nodes. This simplification method describes the 1-D value change. When applied to oceanography, these contour values are wave height, water level, etc. The base contour area is the region enclosed by the contour at the joined or split nodes of the contour tree. It describes the 2-D area (spatial extent) of the selected value. The contour profile is defined by the ratio of the two importance values mentioned above. Formula 2 defines a 3-D property, depicting the trend in contour changes to see whether the enclosed volume (hump or

pit) contours are flat or steep. To distinguish the upward arcs (peaks) from the downward arcs (holes), the importance values of the downward arcs are given negative values.

$$f = \frac{\text{contour value difference}}{\text{base contour area}} \quad (2)$$

Formula 2. The profile value is based on the difference between local values (two end nodes of the arc) and the base contour area (joined or split node)

Figure 6.2 illustrates how these importance values are measured. The extreme local values (v_1 and v_2) and the joined or split nodes (v_0) are selected to calculate the value difference between contours. Regular nodes are not considered. The importance value of the base contour area is calculated from the connecting nodes (joined or split nodes) of each leaf on the contour tree (s_1 , s_2 , and s_3). The contour profile describes the shape property of the

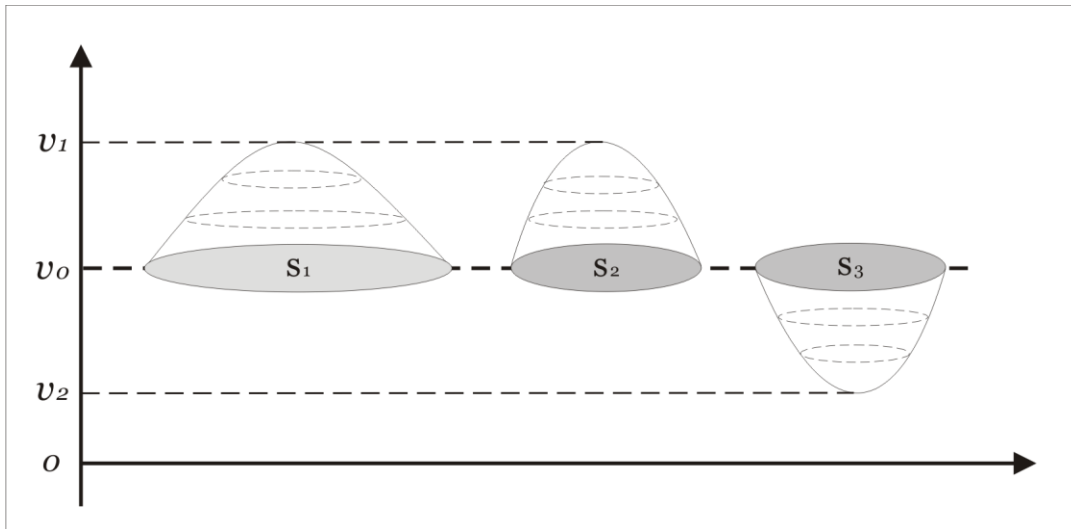


Figure 6.2 Examples of importance values in contour tree simplification. The importance values for three examples are shown, including humps (left and middle) and one pit (right).

- 1) The contour value difference: The left and middle humps have the same value $v_1 - v_0$, whereas the right hump has $v_2 - v_0$.
- 2) The base contour area: While the left hump has a base contour with an area value S_1 , the base contour of the middle hump and the right pit have the same area ($|S_2| = |S_3|$), with S_3 being negative.
- 3) The contour profile: The left hump has a smaller profile value than the middle hump, which indicates a flatter hump. Since the middle hump and the right pit have the same area, the profile value is determined by the contour value v_1 and v_2 .

volume of the enclosed branch contours, including both critical nodes and regular nodes. The large contour profile values (absolute values) indicate a steep and slim hump (S_2) or a pit (S_3), and small contour profile values indicate a flat hump (S_7) or a pit. In oceanography research, the application of the profile can help to define hurricane affected sea surface area and time period.

6.5 Simplification Algorithm

The simplification operation in this research also uses the branch decomposition method to hierarchically represent the contour tree introduced by Pascucci et al. (2004). This method decomposes contour trees into sets of branches with hierarchical relationships (child-parent). As illustrated by Figure 6.3, the contour tree is decomposed into four sets of branches. Branch A1 is the root branch. All other branches (A2, A3, and A4) are connected by a saddle node or joint node. Branches can be removed by a parent-child priority. A1 is the parent of A2 and A4 and since both A2 and A4 have one end node, they are dependent on A1. Similarly, A2 is the parent of A3. A branch can be simplified only when there is no connected child branch. Therefore, branches A3 and A4 can be removed independently. But the branch A2 cannot be removed before A3 has been removed.

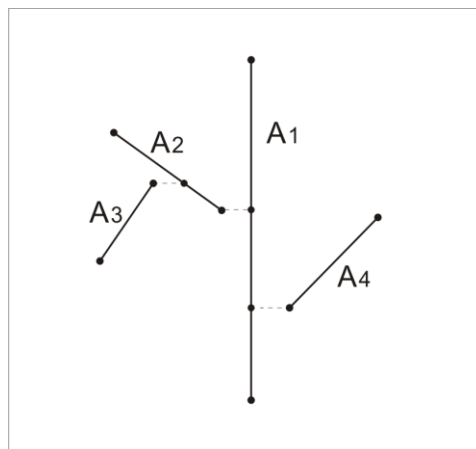


Figure 6.3 Hierarchical decomposition of a contour tree (Pascucci et al. 2004).

The simplification algorithm and the data structure are based on branches. The importance values are stored on the critical nodes of each branch and can be extracted from the branches. The previous simplification methods focus on a single contour tree, but in this data model multiple contour trees are connected by temporal information. The simplification algorithms thus need to take this into consideration. In a dynamic process, the importance values integrated with each branch on contour trees are also changing. For example, an upward branch on a contour tree may represent a high wave zone. The end node value may vary in a series of contour trees related to different time periods. The shape and size of contours represented at the joined node may also change, and this leads to a change in the area. Subsequently, the profile values related to these two values are also changing.

To store the importance value related to the change, a new variable is added to the data model. This new variable is referred to as the *global importance value* (GIV). As mentioned in Chapter 4, the data model presented in this research assigns the same ID to each branch arc identified as the same object in different contour trees. The ID helps tracking the branch arc at different time steps and if the GIV is related to the ID, the importance values change can also be tracked. In this data model, each arc's ID has a corresponding GIV. The GIV stores the highest importance values (lowest if value is negative) of the corresponding arc during the whole lifespan. Based on different applications, the GIVs are selected and calculated from the three importance parameters mentioned above.

For example, if the contour value difference is used as importance value, and the importance values of a branch arc on contour trees for its entire lifespan are $c_1, c_2, c_3, \dots, c_n$ in time series, then the GIV can be defined as:

$$\text{GIV} = \max (|c_1|, |c_2|, |c_3| \dots |c_n|) \quad (3)$$

This formula can also be applied to other importance parameters, including the base contour area and the contour profile.

By comparing the GIV with a preset threshold value, the unnecessary branches are removed from the contour trees. This simplification algorithm can thus be distinguished from other methods in that it does not only focus on the single contour tree, but extends its applicability by considering the links of branches on multiple contour trees. If a branch is selected to be removed from a contour tree at one point in time, all branches with the same ID on all contour trees in this branch's lifespan are removed.

As suggested in Pascucci et al. (2004) and Carr et al. (2004), a priority queue is used as a standard data-structure to store all branch arcs and it always pops out the branch arc with the smallest GIV. The priority queue is used in one of the most important processes to retrieve a valid branch to be simplified. The main steps of this process are described below:

- 1) Iterate all contour trees and calculate the GIV for each branch arc. Store the branch arcs and their GIVs into a priority queue.
- 2) Pop the top element of the queue, which is the branch with the smallest GIV value from the contour trees during their entire lifespans.
- 3) Compare the GIV value with the threshold value and define whether that branch needs to be simplified. If not, discard this element and go back to step 2.
- 4) Fetch the temporal information (starting and end points of time) from that branch. Retrieve the branches from the contour trees between start and end points of time.
- 5) Check if each of these branches can be simplified. The three pre-required criteria to simplify a branch include (a) leaf arcs without any child; (b) the branch only has one end node connected to other arcs at the joined or split node; and (c) not the root branch

arc. If any of these three criteria is not met, discard all of these elements and go back to step 2.

- 6) If branches can be removed, remove these branches from their contour trees. Then update the importance values of the merged branches connected to the removed branches.

The above process ensures that the first branch that represents a valid cancellation from the queue can be found. Unlike the method used by Pascucci et al. (2004), in which nodes are updated only when they are pop out of the queue, in this method the update of importance value related to the merged branches is processed right after the branch is removed.

Applying this process to build contour trees by using “join tree” and “split tree”, a contour tree can be created in which each branch represents a valid topological simplification. Then a simplified contour tree can be extracted from the previously created contour tree.

6.6 Application of Contour Tree Simplification

As Zhou and Takatsuka (2008) concluded, the application of the contour tree simplification is “a graph simplification algorithm applied to the contour tree. Then this simplification is carried back to simplify the input data.”

As mentioned above, three types of importance values are used in the simplification of the oceanographic model data. In the following part, an example is used to show how the application of these importance values works.

Figure 6.4 is a wave height and direction map from the output of the MIKE21 model for the Gulf of Mexico. The contours represent the wave height distribution. The local peak/hole and saddle areas are labeled (A, B ... K, L). Figure 6.5 illustrates the

corresponding contour tree and simplified contour trees with different importance parameters (the contour trees only represent wave height, not wave direction).

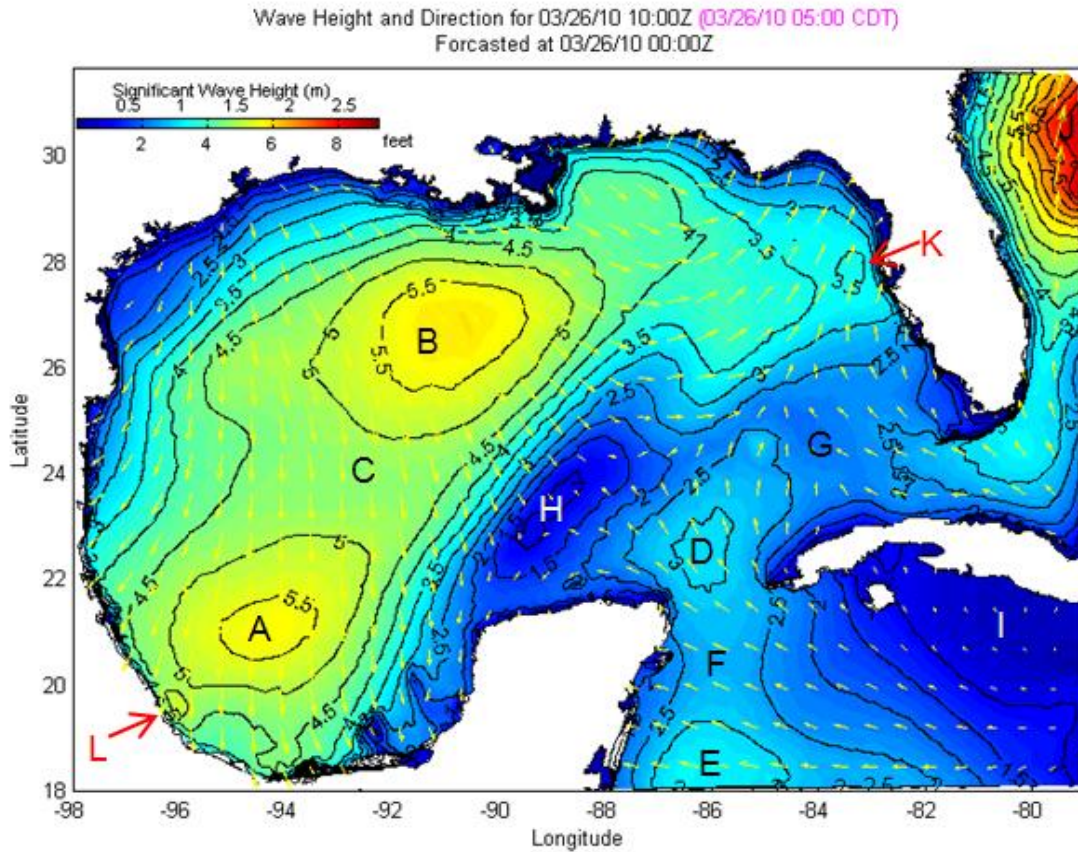


Figure 6.4 A wave height and direction map from the output of the MIKE21 model for the Gulf of Mexico

As shown in the first graphic in Figure 6.5, the original contour tree includes the local peaks A, B, D, E, J, K, and L and local holes H and I.

The second graphic in Figure 6.5 shows the result of the contour value difference based simplification. In this graphic the branches with a wave height difference (end node to joined/split node) of less than 1.5 feet are removed from the contour tree. Compared with the original contour tree, branches with local peaks A, D, L, and K are trimmed. As a result from the contour tree decomposition, branches BC and CG as well as EF and FG are merged. But

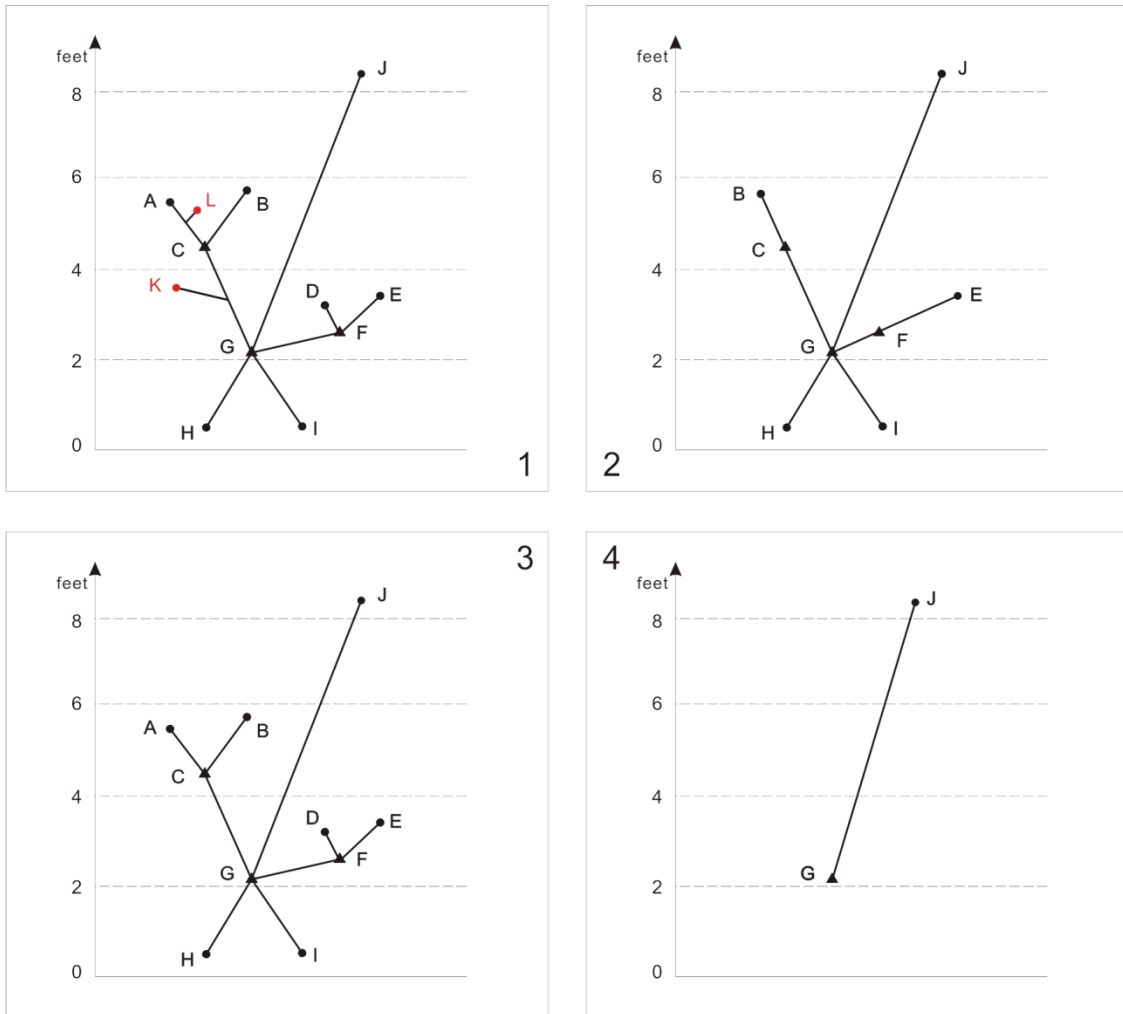


Figure 6.5 Contour tree simplifications: 1. Original contour tree 2. Contour value difference-based simplification 3. Base contour area-based simplification 4. Contour profile-based simplification.

the joined nodes C and F remain for the further construction of the contour tree. Because the temporal information for each independent branch arc on multiple contour trees has to be tracked, it is easier to keep the temporal information with the branch arcs from the original branch structure. The left joined nodes will help track these changes. As discussed above, node reduction is one of the basic simplification operations. In this method, it is applied to help with the simplification process, but it does not physically remove that node's information from the contour tree. In this example, the C and F nodes still hold the information as there

are joined nodes and related arcs (BC and CG, as well as EF and FG). This information is necessary for the rebuild of final contour trees. From an application point of view, this method removes lower wave height peaks and shallower wave height holes and keeps the wave height zones more distinguished. This method does not take into consideration the zone sizes of wave heights.

The third graphic in Figure 6.5 shows the result of the contour tree after contour area based simplification. All branches whose zone size of wave height is less than 300 square feet are removed from the contour tree. So, branches with local peaks L and K are trimmed from the contour tree. From an application point of view, this method removes small size wave height zones and keeps big size wave height zones. It does not consider the wave height change values.

The fourth graphic in Figure 6.5 shows the result of contour profile-based simplification. Branches with small profile values are removed from the contour tree. In this graphic, all original branches are removed from the contour tree except branch JG. From an application point of view, this method considers both wave height values and the spatial extent of the wave height zone. This method removes the flat wave height zones and keeps slim (steep) wave height zones.

Taking it all together, the three simplification methods result in different outputs and are for different applications.

6.7 Conclusion

Due to the complicated spatial distribution and highly dynamic changes in oceanographic data, contour trees representing oceanographic isosurfaces can produce very big size contour trees. A simplification of original contour trees is thus necessary. In this

chapter, a new simplification algorithm is introduced to reduce the complexity of contour trees. This algorithm is based on the branch decomposition method and supports temporal information integrated into contour trees. In order to meet the requirement from various applications, three types of importance values are introduced to be applied to different simplification methods. The practical example illustrates how these importance values affect the output of the contour tree simplifications.

Chapter 7 Conclusion

7.1 Major Findings

In oceanographic research, the output data of wave and hydrodynamic numerical models represent highly dynamic geospatial phenomena. A data model is the fundamental issue in supporting this special application in spatio-temporal GIS. This dissertation focuses the research on a spatial-temporal GIS data model applied to oceanographic model data, especially to homogeneous isosurface data.

Due to the special challenges from this kind of application, such as field-object characteristics, constant temporal change, and complex spatio-temporal change pattern, current spatial-temporal data models show deficiencies in supporting such challenges. A tentative new data model is proposed in this research.

As an emerging method, ontological analysis is applied to the study of oceanographic data. In this research, ontology has been utilized to extract abstract oceanographic data entities to create components for the new data model. When transferring these components to data model design, the contour tree is introduced as the basic data structure for the data model. The contour tree can represent the spatial topological relationships of the isosurface data by using element components of nodes and arcs. By adding temporal information to each node and branch arc of the contour tree, and by using multiple contour trees to represent different time steps in the temporal dimension, changes can be stored and tracked by the data model.

In order to support the field characteristic and reduce the data volume, the new data model integrates spatial and temporal interpolation methods. The spatial interpolation calculates the data falling between neighboring contours at a single point in time. The inverse distance weighting (IDW) is applied as the main algorithm, and the minimum bounding

rectangle (MBR) is used to help the spatial interpolation performance. The temporal interpolation calculates the data which are not recorded by contour trees and falling between consecutive contour trees for subsequent points in time. The linear interpolation algorithm is applied because of its modest accuracy and simple implementation.

The OO (Object-Oriented) design technology has been applied to the model design and implementation. The UML class diagram demonstrates the main objects with the main attributes and functions.

To evaluate the supported functions of the new data model, a case study has been designed to show how this data model supports complicated spatio-temporal queries in every day forecasting applications.

As an attempt to optimize the data model, this dissertation also shows some attempts in contour tree simplification. A new simplification algorithm is introduced to reduce the data complexity. This algorithm is based on the branch decomposition method and supports temporal information integrated with contour trees. Three types of importance values are introduced to run different simplification methods for various applications.

7.2 Contributions

The main contribution of this research is a new data model for spatio-temporal GIS which is superior to current data models and thus can be applied to isosurface data with highly dynamic change patterns. The advantages of this data model are as follows:

- 1) The data model supports the unusual field-object characteristics of oceanographic model data;
- 2) It supports constant temporal change type;

- 3) It supports complex spatio-temporal change patterns (geometry, topology, and attributes) ;
- 4) It supports strong spatio-temporally related queries;
- 5) Overall, it is a balanced data model in data size and complexity. To represent similar geospatial phenomena, this data model has a smaller data volume than raster data and it can represent more entities than traditional vector data with integrated interpolation algorithms.

Additional research about the contour tree simplification explores simplification algorithms for the temporal information integrated into multiple contour trees. The proposed simplification algorithms reduce the data volume and complexity by filtering out less significant data.

Obviously, this new data model can be applied not only to oceanographic data, but other dynamic isosurface data or 3-D data values.

7.3 Future Research

Although this data model supports oceanographic data very well, there are still some worthwhile challenges to pursue. One main problem is that this data model is still incapable of fully supporting object-field representations of the complicated oceanographic data. The data model developed in this research uses nodes in the contour tree to represent objects (contours), but it can hardly represent objects (contours) between these nodes. For example, in a wave height contour map if only contours from 0 to 10 feet with 1 foot interval are stored in the contour tree, this data model will have a difficulty in defining an area to be queried that represents a wave height zone with values in between two contours (e.g., 3.5 feet). It is

possible to accomplish this at the application level, but it is better to support this at the data model level.

Another concern of this data model is that the integrated interpolation algorithms reduce the quality (e.g., accuracy) of the data model and also make the data model more complicated. Optimization attempts are needed to make the data model more practical.

Moreover, although this data model does not aim to support vector data, many met-ocean data used in oceanography research are vector based, such as wave directions. If this data model can integrate or partially support vector data, it will be more applicable.

These three issues are the major shortcomings of this data model and future research should focus on solving these shortcomings.

References

- T. Abraham and J. F. Roddick. 1999. Survey of Spatio-Temporal Databases. *Geoinformatica*, 3(1).
- P. Agarwal. 2005. Ontological considerations in GIScience. *International Journal of Geographical Information Science*, 19(5):501–536.
- M. P. Armstrong. 1988. Temporality in spatial databases. In: *proceedings of GIS/LIS'88*, 2:880-889.
- K. A.V. Borges, C. A. Davis, and A. H. F. Laender. 2001. OMT-G: An Object-Oriented Data Model for Geographic Applications. *GeoInformatica*, 5(3):221-260.
- K. Brassel, F. Bucher, E. M. Stephan, and A. Vckovski. 1995. Completeness. In: S. C. Guptill and J. L. Morrison (Eds.), *Elements of Spatial Data Quality*. Elsevier, New York pp. 81-108.
- C.L. Bajaj, V. Pascucci, D.R. Schikore. 1997. The contour spectrum, in: *Proceedings of Visualization*, pp. 167–173.
- C. Bajaj , M. Kreveld , R. Oostrum , V. Pascucci , and D. R. Schikore. 1998. Contour Trees and Small Seed Sets for Isosurface Traversal. In: *Proceedings of the 13th Annual ACM Symposium on Computational Geometry*.
- H. Carr, J. Snoeyink, and U. Axen. 2000. Computing Contour Trees in All Dimensions. In: *Proceedings of the eleventh annual ACM-SIAM symposium on discrete algorithms*, ACM.
- H. Carr, J. Snoeyink. 2003. Path seeds and flexible isosurfaces: Using topology for exploratory visualization, in: *Proceedings of Eurographics Visualization Symposium*, pp. 49–58.
- H. Carr, J. Snoeyink, and M. van de Panne. 2004. Simplifying flexible isosurfaces using local geometric measures. In: *Proceedings of IEEE conference on Visualization*, pp.497–504.
- J. Chen and J. Jiang. 2000. An Event-Based Approach to Spatio-Temporal Data Modeling in Land Subdivision Systems. *GeoInformatica*, 4(4):387-402.
- Y.J. Chiang and X. Lu. 2003. Progressive simplification of tetrahedral meshes preserving all isosurface topologies, *Computer Graphics Forum*, 22 (3):493–504.
- Y.J. Chiang, T. Lenz, X. Lu, G. Rote. 2005. Simple and Optimal Output-Sensitive Construction of Contour Trees Using Monotone Paths, *Computational Geometry: Theory and Applications*, 30.

- C.Claramunt, C. Parent, S.Spaccapietra, and M. Theriault. 1998. Database Modeling for Environmental and Land Use Changes, *Geographical Information and Planning*, Springer-Verlag, Chapter 20.
- J. T. Coppock, and D. W. Rhind. 1991. The history of GIS. In David J. Maguire, Michael F. Goodchild and David W. Rhind (Ed.), *Geographical Information Systems: principles and applications*. Essex: Longman Scientific & Technical, 1: 21-43.
- F. Fonseca and M. Egenhofer. 1999. Ontology-Driven Geographic Information Systems. In: Medeiros C B (Ed.) *7th ACM Symposium on Advances in Geographic Information Systems*, pp.14-19.
- F. Fonseca, M. Egenhofer, P. Agouris, and G. Câmara. 2002. Using Ontologies for Integrated Geographic Information Systems. *Transactions in GIS*, 6(3):231-257.
- F. Fonseca. 2007. The Double Role of Ontologies in Information Science Research. *Journal of the American Society for Information Science and Technology*, 58(6):786-793.
- M. F. Goodchild, S. Guoqing, and Y. Shiren. 1992. Development and test of an error model for categorical data. *International Journal of Geographical Information Systems*, 6:87-104.
- P. Green and M. Rosemann. 2004. Editorial Preface: Ontological Analysis, Evaluation and Engineering of Business Systems Analysis Methods. *Journal of Database Management*, 15: i-ii.
- P. Grenona and B. Smitha. 2004. SNAP and SPAN: Towards Dynamic Spatial Ontology. *Spatial Cognition & Computation: An Interdisciplinary Journal*, 4(1):69-104.
- N. Guarino. 1997. Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. in: M. Pazienza, (Ed.), *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*, International Summer School, SCIE-97, Frascati, Italy, pp.139-170.
- N. Guarino. 1998. Formal Ontology and Information Systems. in: Guarino N (Ed.), *Formal Ontology in Information Systems*. pp.3-15, IOS Press, Amsterdam, Netherlands.
- Y. Huang and C. Lee. 2009. Efficient evaluation of continuous spatio-temporal queries on moving objects with uncertain velocity. *Geoinformatica*, 14(2):163-200.
- G.J. Hunter. 1999. Managing uncertainty in GIS. in P.A. Longley, M.F. Goodchild, D.J. Maguire, and D.W. Rhind (Eds.), *Geographical Information Systems*, Wiley.
- K. Kemp and A. Vckovsky. 1998a. At the "The Ontology of Fields" specialist meeting, a research initiative under the Varenus Project of National Center for Geographic Information and Analysis (NCGIA). http://www.geocomputation.org/1998/60/gc_60.htm.

- K. Kemp and A. Vckovsky. 1998b. Towards an ontology of fields. In: *the proceedings of the Third International Conference on GeoComputation, Bristol UK*.
- L. Kettner, J. Rossignac, and J. Snoeyink. 2001. The safari interface for visualizing time-dependent volume data using iso-surfaces and contour spectra, *Computational Geometry: Theory and Applications*, 25 (1–2):97–116.
- M. V. Kreveld, R. V. Oostrum, C.L. Bajaj, V. Pascucci, and D.R. Schikore. 1997. Contour trees and small seed sets for isosurface traversal, in: *Proceedings of the 13th ACM Symposium on Computational Geometry*, pp. 212–220.
- G. Langran. 1988. Temporal design tradeoffs. In: *proceedings of GIS/LIS'88*, (Falls Church, VA: ACSM), 2:890-899.
- G. Langran and N. R. Chrisman. 1988. A Framework for Temporal Geographic Information. *Cartographica*, 25(3):1-14.
- R. Laurini and D. Thompson. 1992. *Fundamentals of Spatial Information Systems*. Academic Press: London.
- D. Mark and F. Csillag. 1989. The nature of boundaries on “area-class” maps. *Cartographica*, 26:65-78.
- J. McCarthy. 1980. Circumscription - A Form of Non-Monotonic Reasoning. *Artificial Intelligence*, 1-2(13):27-39.
- L. D. Montgomery. 1995. Temporal Geographic Information Systems Technology and Requirements: Where we are Today, Thesis, Department of Geography, Ohio State University, USA.
- S.P. Morse. 1968. A Mathematical Model for the Analysis of Contour-Line Data. *Journal of the ACM*, 15(2).
- S.P. Morse. 1969. Concepts of Use in Contour Map Processing. *Communication of the ACM*. 12(3):147-152.
- J. Nunes. 1991. Geographic space as a Set of Concrete Geographical Entities. In: D. Mark and A. Frank (Eds.), *cognitive and Linguistic Aspects of Geographic Space*. pp. 9-33, Kluwer, Dordrecht, the Netherlands.
- V. Pascucci, K. Cole-McLaughlin, and G. Scorzell. 2004. Multi-Resolution Computation and Presentation of Contour Trees. In: *Proceedings of the IASTED Conference on Visualization, Imaging and Image Processing (VIIP 2004)*, pp.452–290.
- V. Pascucci and K. Cole-McLaughlin. 2002. Efficient Computation of the Topology of Level Sets. In: *Proceedings of the conference on Visualization '02*.

- E. J. Pebesma, K. D. Jong, and D. Briggs. 2007. Interactive visualization of uncertain spatial and spatio-temporal data under different scenarios: an air quality example. *International Journal of Geographical Information Science*, 21(5):515–527.
- N. Pelekis, B. Theodoulidis, I. Kopanakis, and Y. Theodoridis. 2005. Literature Review of Spatio-Temporal Database Models. *The Knowledge Engineering Review journal*, 19(3): 235-274.
- D.J. Peuquet. 1994. It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers* 84(3):441-61.
- D.J. Peuquet 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.
- D. J. Peuquet. 2001. Making Space for Time: Issues in Space-Time Data Representation. *GeoInformatica*, 5(1)11-32.
- R. J. Price, N. Tryfona, and C. S. Jensen. 2000. Extended SpatioTemporal UML: Motivations, Requirements and Constructs, *Journal on Database Management*, Special Issue on UML, 11(4).
- A. Renolen. 1997. *Modeling Spatiotemporal Information: The Spatiotemporal Object Model*, Norwegian University of Science and Technology, Trondheim, Norway.
- A.A. Roshannejad and W. Kainz. 1995. Handling Identities in Spatio-Temporal Databases. In: *Processing of ACSM/ASPRS 1995 Annual Convention and Exposition Tech. Papers*.
- C. Runge. 1901. Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten, *Zeitschrift für Mathematik und Physik* 46: 224–243.
- O.R. Sorensen, H. Kofoed-Hansen, M. Rugbjerg, and L.S. Sorensen. 2004. A Third-Generation Spectral Wave Model Using an Unstructured Finite Volume Technique, 29th International Conference on Coastal Engineering, pp. 894-906.
- D. Shepard. 1968. A two-dimensional interpolation function for irregularly-spaced data. In: *Proceedings of the 1968 ACM National Conference*. pp. 517–524.
- J.K. Sircar and J.A. Cebrian. 1986. Application of image processing techniques to the automated labelling of raster digitized contour maps, in: *Proceedings of the 2nd International ACM Symposium on Spatial Data Handling*, pp. 171–184.
- B. Smith and D.M. Mark. 1998. Ontology and geographic kinds. In: *Proceedings of International Symposium on Spatial Data Handling*, Vancouver, Canada.

- S. Takahashi, T. Ikeda, Y. Shinagawa, T.L. Kunii, and M. Ueda. 1995. Algorithms for Extracting Correct Critical Points and Constructing Topological Graphs from Discrete Geographical Elevation Data, *Computer Graphics Forum*, 14(3).
- S. Takahashi, G. M. Nielson, Y. Takeshima, and I. Fujishiro. 2004. Topological volume skeletonization using adaptive tetrahedralization. In: *Proceedings of the Geometric Modeling and Processing 2004*, pp. 227–236.
- S. Takahashi, Y. Takeshima, and I. Fujishiro. 2004. Topological Volume Skeletonization and Its Application to Transfer Function Design. *Graphical Models*, 66(1).
- Y. Theodoridis. 2003. Ten Benchmark Database Queries for Location-based Services. *The Computer Journal*, 46(6):713-725.
- N. Tryfona and C. S. Jensen. 1999. Conceptual Data Modeling for Spatiotemporal Applications. *Geoinformatica*, 3(3).
- N. Tryfona, D. Pfoser and T. Hadzilacos. 1998. *Modeling Behaviour of Geographic Objects: Experience with the Object Modeling Technique*, University of Maine, Orono, USA.
- M. Uschold, M. Healy, K. Williamson, P. Clark and S. Woods. 1998. Ontology reuse and application. In: N. Guarino (Ed.), *Formal Ontology in Information Systems (FOIS'98)*, Trento, Italy, (Amsterdam: IOS Press).
- J. P. Wilson and P. A. Burrough. 1999. Dynamic Modeling, Geostatistics, and Fuzzy Classification: New Sneakers for a New Geography? *Annals of the Association of American Geographers*, 89(4):736-746.
- S. Winter. 2001. Ontology: Buzzword or Paradigm Shift in GI Science? *International Journal of Geographical Information Science*, 15(7):587 – 590.
- M. F. Worboys. 1994. “A Unified Model for Spatial and Temporal Information”, *The Computer Journal*, 37(1):27-34.
- M. F. Worboys, H. M. Hearshshow, and D. J. Maguire. 1990. Object-Oriented Modeling for Spatial Databases, *International Journal of Geographical Information Systems*, 4(4).
- X. Xin, S. Shaw, and H. Lin. 2003. Developing a Spatio-Temporal Data Model for an Exploratory Study on the Land Use-Transportation Interaction: a Temporal GIS Attempt. In: *Proceedings of Asia GIS Conference 2003*.
- M. Yuan. 1994. Wildfire Conceptual Modeling for Building GIS Space-Time models, In: *Proceedings of the GIS/LIS'94*, pp. 860-869.
- M. Yuan. 1996. Temporal GIS and Spatio-Temporal Modeling, Published on CD-ROM by the National Center for Geographical Information and Analysis.

M. Yuan. 1997. Modelling Semantical, Spatial, and Temporal Information in a GIS, in M. Craglia and H. Couclelis (Eds.). *Geographic Information Research: Bridging the Atlantic*, CRC Press.

M. Yuan. 1999. Use of a Three-Domain Representation to Enhance GIS Support for Complex Spatiotemporal Queries. *Transactions in GIS*, 3(2): 137-159.

M. Yuan. 2001a. Representing Complex Geographic Phenomena in GIS. *Cartography and Geographic Information Science*, 28(2):83-96.

M. Yuan. 2001b. Use of Knowledge Acquisition to Build Wildfire Representation in Geographical Information Systems. *International Journal of Geographical Information Science*, 11(8).

W. Zhang and G. J. Hunter. 2000. Temporal Interpolation of Spatially Dynamic Object. *GeoInformatica*, 4(4):403-418.

J. Zhou, and M. Takatsuka. 2008. Contour Tree Simplification Based on a Combined Approach. Technical Report, School of Information Technologies, The University of Sydney, Australia. June 2008. <http://sydney.edu.au/engineering/it/research/tr/tr624.pdf>.

Vita

Yuliang Chen received his Bachelor of Science degree in geography in 2000 and Master of Science degree in geographic information science (GIS) in 2003 from the Department of Geography in the East China Normal University. He began his doctoral program in GIS at the Department of Geography and Anthropology at the Louisiana State University in the fall of 2005. Meanwhile, he completed a Master of Science degree in systems science from the Department of Computer Science, Louisiana State University in 2009.