

# **MULTIPLE REPRESENTATIONS OF ELEVATION FOR DYNAMIC PROCESS MODELING**

by

**Laércio Massaru Namikawa**  
July 28, 2006

A dissertation submitted to the  
Faculty of the Graduate School of  
the State University of New York at Buffalo  
in partial fulfillment of the requirements for the  
degree of  
Doctor of Philosophy

Department of Geography

Copyright by  
Laércio Massaru Namikawa  
2006

# DEDICATION

To Mariza, Nicholas, and Martin.

# ACKNOWLEDGMENTS

## Dissertation Committee

**Prof. Christian Renschler** Department of Geography, State University of New York at Buffalo and National Center for Geographic Information and Analysis (NCGIA), major adviser

**Prof. David Mark** Department of Geography, State University of New York at Buffalo and National Center for Geographic Information and Analysis (NCGIA)

**Prof. Michael Woldenberg** Department of Geography, State University of New York at Buffalo and National Center for Geographic Information and Analysis (NCGIA)

## Outside Reader

**Prof. Marcus Bursik** Department of Geology, State University of New York at Buffalo

## Grant

**National Science Foundation:** Research Project ITR/AP+IM: Information Processing for Integrated Observation and Simulation Based Risk Management of Geophysical Mass Flows, ITR-0121254

# CONTENTS

DEDICATION	iii
ACKNOWLEDGMENTS	iv
CONTENTS	v
LIST OF TABLES	xv
LIST OF FIGURES	xvii
ACRONYMS	xix
ABSTRACT	xxi

## INTRODUCTION

<b>Chapter 1. INTRODUCTION</b>	<b>1</b>
1.1. OBJECTIVES	1
1.2. DISSERTATION STRUCTURE	3
1.2.1. Part I - Review	4
1.2.2. Part II - Methodology	5
1.2.3. Part III - Validation and Conclusions	7

## PART I

### REVIEW

<b>Chapter 2. DEFINITION OF MODELS</b>	<b>9</b>
2.1. DEFINING MODELS	9
2.2. USING MODELS	11
2.3. CLASSIFYING MODELS	12
2.3.1. Types of Real World Entities	12
2.3.2. Basic Formal Ontology Sub-Ontologies	13
2.4. DEFINING MODELS OF PROCESSES	14
2.4.1. Processual Entities of SPAN Ontology	14
2.4.2. Types of Process Models	16
2.4.3. Process Models and Time Dependency	17
2.4.4. Environmental Process Models	17
2.5. DEFINING MODELS OF OBJECTS	18
2.5.4. Geographic Objects	19
2.6. MODELING GEOGRAPHIC REALITY	20
2.6.1. Modeling Process	20
2.6.2. Modeling Geographic Objects	21
2.6.2.1. <i>Ontological Level</i>	22
2.6.2.2. <i>Conceptual Level</i>	23
2.6.2.3. <i>Representation Level</i>	25
2.6.2.4. <i>Implementation Level</i>	26
2.7. SUMMARY	27

<b>Chapter 3. DEFINITION OF MULTIPLE REPRESENTATIONS OF</b>	
ELEVATION	29
3.1. DEFINING DIGITAL GEOGRAPHIC DATA REPRESENTATION	29
3.1.1. Restrictions to Represent Geographic Data in Digital Format	30
3.1.2. Source of Multiple Representations	32
3.1.3. Current Use of Multiple Representations	32
3.2. DEFINING ELEVATION	33
3.3. MODELING REPRESENTATIONS OF ELEVATION	35
3.4. SOURCES FROM ONTOLOGICAL LEVEL	35
3.5. SOURCES FROM CONCEPTUAL LEVEL	36
3.5.1. Mapping Reality to Fields or Objects	37
3.5.1.1. <i>Mapping to Geo-Fields</i>	37
3.5.1.2. <i>Mapping to Geo-Object</i>	37
3.5.2. Acquisition Time	38
3.5.3. Reference Surface	38
3.5.4. Quality of the Representation	39
3.5.5. Measuring Instrument	40
3.5.5.1. <i>Traditional Survey</i>	40
3.5.5.2. <i>Global Positioning System</i>	41
3.5.5.3. <i>Photogrammetry</i>	41
3.5.5.4. <i>Radar Interferometry</i>	42
3.5.5.5. <i>Light Detection And Ranging</i>	42

3.5.6. Raw Data Processing	43
3.5.7. Ground Control Points	43
3.5.8. Noise Filtering	44
3.5.9. Phase Unwrapping	44
3.6. SOURCES FROM REPRESENTATION LEVEL	44
3.6.1. Representations of Elevation Geo-Fields	44
3.6.1.1. <i>Triangular Irregular Network</i>	45
3.6.1.2. <i>Contour Lines</i>	45
3.6.2. Representations of Elevation Geo-Objects	45
3.7. SOURCES FROM IMPLEMENTATION LEVEL	46
3.7.1. Storage Data Structure	46
3.7.2. Manipulation Algorithms	46
3.8. MULTIPLE REPRESENTATIONS OF ELEVATION ISSUES	47
3.8.1. Fusion	47
3.8.2. Estimation of Data Quality	48
3.8.3. Defining Fitness for Use	49
3.8.3.1. <i>Resolution</i>	49
3.8.3.2. <i>Measuring Instrument</i>	50
3.8.3.3. <i>The Best Representation of Elevation</i>	51
3.9. SUMMARY	51



## PART II

### METHODOLOGY

<b>Chapter 4. INTEGRATION OF MULTIPLE REPRESENTATIONS OF</b>	
ELEVATION AND PROCESS MODELS	53
4.1. DEFINING DYNAMIC SPATIALLY EXPLICIT PROCESS MODEL	54
4.1.1. Dynamic Process Model	54
4.1.2. Spatially Explicit Process Model	55
4.2. LINKING ENVIRONMENTAL PROCESS MODELS AND GIS	56
4.2.1. Existing Linkage Types	56
4.2.2. Selection of Coupling Type	57
4.3. LINKING SPATIALLY EXPLICIT PROCESS MODEL TITAN2D TO GIS	58
4.3.1. Parallel Adaptative Mesh Grid Godunov Scheme	59
4.3.2. Parameters Derived from Elevation	60
4.3.3. Variable Resolution	60
4.3.4. Estimation of Derivatives of Elevation	61
4.4. A TITAN2D SIMULATION SESSION	63
4.4.1. Data Parameters	65
4.4.2. Computational Parameters	66
4.4.3. Flow Pile Parameters	67
4.5. SELECTING A GIS FUNCTION LIBRARY FOR TIGHT COUPLING	68
4.5.1. Representation of Elevation in TerraLib Spatial Database	70
4.5.1.1. <i>Representations of Spatial Entities</i>	71
4.5.1.2. <i>Vector Format</i>	73

4.5.1.3. <i>Raster Format</i>	74
4.5.2. Representation Information	75
4.5.3. Representation Metadata	78
4.5.3.1. <i>Representation Description</i>	79
4.5.3.2. <i>Resolution</i>	80
4.5.3.3. <i>Acquisition Date and Time</i>	80
4.5.3.4. <i>Representation Production</i>	81
4.5.3.5. <i>Representation Accuracy</i>	82
4.5.4. Representation Query	82
4.6. IMPLEMENTING THE NEW LINK TO A PROCESS MODEL	83
4.6.1. Link Requirements	84
4.6.1.1. <i>Adaptative Grid</i>	84
4.6.1.2. <i>Temporal Query</i>	85
4.6.2. Implementation Details	85
4.6.2.1. <i>Cache of Queries</i>	86
4.6.2.2. <i>Prediction of Queries</i>	86
4.6.2.3. <i>Creation of Derived Representations</i>	86
4.6.3. Ranking Representations	87
4.6.4. C++ Classes of the Implementation	88
4.7. A SIMULATION SESSION USING NEW IMPLEMENTATION	92
4.8. SUMMARY	96

<b>Chapter 5. ADDITIONAL METHODS FOR INTEGRATION OF MULTIPLE</b>	
REPRESENTATIONS OF ELEVATION AND PROCESS MODELS	99
5.1. COMPUTING SPATIALLY DISTRIBUTED ACCURACY	100
5.1.1. Global Accuracy	101
5.1.2. Computing Spatially Distributed Accuracy	103
5.1.2.1. <i>Descriptive Statistics of Elevation</i>	104
5.1.3. Cluster Analysis	105
5.1.3.1. <i>Z-Score</i>	105
5.1.3.2. <i>Gaussian Kernel Filter</i>	106
5.1.3.3. <i>Critical Value</i>	106
5.1.4. Classified Probability Map	107
5.1.5. Example of Accuracy Map	108
5.2. COMPUTING A MEASURE OF SIMULATION PERFORMANCE	111
5.2.1. Existing Comparison Methods for Dynamic Processes	112
5.2.2. Quantitative Comparison Method	114
5.2.2.1. <i>Logistic Regression</i>	114
5.2.2.2. <i>Independent Variable Definition</i>	115
5.2.2.3. <i>Quantitative Performance Measure Definition</i>	118
5.3. SUMMARY	118

## PART III

### VALIDATION AND CONCLUSIONS

<b>Chapter 6. VALIDATION OF THE INTEGRATION OF MULTIPLE</b>	
REPRESENTATIONS OF ELEVATION AND PROCESS MODELS	121
6.1. CREATING MULTIPLE REPRESENTATIONS IN TERRALIB DATABASE	121
6.1.1. Tool to Import Representations of Elevation	121
6.1.1.1. <i>Metadata File Description</i>	122
6.1.2. Importing Elevation for Colima	123
6.1.2.1. <i>SRTM DEM</i>	124
6.1.2.2. <i>ARIA DEM</i>	126
6.1.3. Importing Elevation for San Bernardino	128
6.1.3.1. <i>NOAA-CSC DEM</i>	129
6.1.3.2. <i>SRTM DEM</i>	130
6.2. SIMULATION OF COLIMA BLOCK-AND-ASH FLOW EVENT	131
6.2.1. Event Description	132
6.2.2. Parameters for Simulation	133
6.2.3. Simulation Results Using Current Implementation	135
6.2.4. Simulation Results Using New Implementation	136
6.2.5. Comparison of Simulation Results at Colima	137
6.2.5.1. <i>Flow Footprint</i>	138
6.2.5.2. <i>Pseudo-Distance</i>	138
6.2.5.3. <i>Performance Measure of Simulation Using Current Implementation</i>	139
6.2.5.4. <i>Performance Measure of Simulation Using New Implementation</i>	140

6.3. SIMULATION OF SAN BERNARDINO DEBRIS FLOW EVENT	141
6.3.1. Event Description	141
6.3.2. Parameters for Simulation	143
6.3.3. Simulation Results Using Current Implementation	144
6.3.4. Simulation Results Using New Implementation	148
6.3.5. Comparison of Simulation Results at San Bernardino	149
6.3.5.1. <i>Performance Measure of Simulation Using Current Implementation</i>	150
6.3.5.2. <i>Performance Measure of Simulation Using New Implementation</i>	151
6.4. SUMMARY	152
<b>Chapter 7. CONCLUSIONS</b>	155
7.1. MODELS	156
7.1.1. Models of Processes	157
7.1.2. Models of Objects	158
7.1.3. Modeling Processes	158
7.1.4. Modeling Objects	158
7.2. MULTIPLE REPRESENTATIONS OF ELEVATION	159
7.2.1. Issues Related to Multiple Representations of Elevation	160
7.3. FRAMEWORK FOR INTEGRATION OF MULTIPLE REPRESENTATIONS OF ELEVATION AND PROCESS MODELS	161
7.3.1. Implementation Optimizations	163
7.3.2. Ranking Representations	164
7.4. ADDITIONAL METHODS FOR MULTIPLE REPRESENTATIONS OF	165

ELEVATION	
7.4.1. Spatially Distributed Accuracy	165
7.4.2. Performance Measure	166
7.5. VALIDATION OF THE MULTIPLE REPRESENTATIONS FRAMEWORK	166
7.6. EXTENSIONS TO PROPOSED FRAMEWORK	168
7.6.1. Other Geographic Objects	168
7.6.2. Storage and Use of Samples, Contour Lines, and TINs	168
7.6.3. Validation and Storage of Simulation Results	169
7.6.4. Fusion of Representations of Elevation	169
7.7. SUMMARY	170
REFERENCES	173

# LIST OF TABLES

4.1. Description of data parameters	66
4.2. Description of computational parameters	67
4.3. Description of flow pile parameters	68
4.4. Description of columns describing a representation	80
4.5. Description of columns describing production of a representation	81
5.1. Descriptive statistics of difference between ARIADEM and SRTMDEM	109
5.2. Area of clusters for Gaussian kernel filters	110
6.1. Field names and example values of metadata file	124
6.2. Pile parameters for simulation of block-and-ash flow event	135
6.3. Computational parameters for simulation block-and-ash flow event	136
6.4. Logistic regression model for simulation at Colima using current implementation	141
6.5. Logistic regression model for simulation at Colima using new implementation	142
6.6. Piles parameters for simulation of debris flow	145
6.7. Computational parameters for simulation of debris flow	146
6.8. Logistic regression model for simulation at San Bernardino using current implementation	152
6.9. Logistic regression model for simulation at San Bernardino using new implementation	152





# LIST OF FIGURES

2.1. Classification of SPAN and SNAP sub-ontologies	14
4.1. Sequence for simulation using the current link	64
4.2. TITAN2D graphical user interface	65
4.3. Simplified TerraLib database schema	72
4.4. Example of the contents of TerraLib metadata tables	73
4.5. TerraLib database schema for lines points table	74
4.6. TerraLib database schema for storage of raster representation	75
4.7. Contents of TerraLib metadata tables	77
4.8. Contents of “represattrib” table	79
4.9. Internal structure of “TeRepresentationInfo” class	89
4.10. State diagram of insertion of a representation into a TerraLib database	89
4.11. Internal structure of classes that provide interface for process models	91
4.12. State diagram of operations to process a representation of elevation query	92
4.13. Sequence for simulation using the new link	95
5.1. Example of a spatially distributed accuracy map	111
5.2. Example of pseudo-distance	117
6.1. Colima Volcano in October 28, 2003 and its location in Mexico	125
6.2. Metadata file of SRTM elevation for Colima Volcano area	126

6.3. Representations of elevation for Colima Volcano area	127
6.4. Metadata file of ARIA elevation for Colima Volcano area	128
6.5. Metadata file of NOAA-CSC elevation for San Bernardino area	130
6.6. Representations of elevation for San Bernardino area	131
6.7. Metadata file of SRTM elevation for San Bernardino area	132
6.8. Extent of Colima Volcano block-and-ash flow event	134
6.9. Summary of results from simulation of block-and-ash flow event using current implementation	137
6.10. Summary of results from simulation of block-and-ash flow event using new implementation	138
6.11. Colima event flow footprint, Delaunay triangulation of flow edge points, and flow medial axis	139
6.12. Pseudo-distance from distance to flow footprint edges and center	140
6.13. Extent of San Bernardino County debris flow	143
6.14. Summary of results from simulations of debris flow using current implementation and starting Pile 1 and Pile 2	147
6.15. Summary of results from simulation of debris flow using current implementation and starting Pile 1	148
6.16. Summary of results from simulation debris flow using new implementation	150
6.17. Pseudo-distance debris flow event	151

# ACRONYMS

API	Application Program Interface
ARIA	Arizona Regional Image Archive
BFO	Basic Formal Ontology
BLOB	Binary Long Object
DCM	Digital Cartographic Model
DEM	Digital Elevation Model
DLM	Digital Landscape Model
DTM	Digital Terrain Model
EGM	Earth Gravitational Model
FGDC	Federal Geographic Data Committee
GIS	Geographic Information Systems
GPS	Global Positioning System
IEEE	Institute of Electrical and Electronic Engineers
IFOMIS	Institute for Formal Ontology and Medical Information Science
INEGI	Instituto Nacional de Estadística, Geografía e Informática
IRTF92	International Terrestrial Reference Frame of 1992
LINS	Laser Inertial Navigation System
NAD27	North American Datum of 1927
NAD83	North American Datum of 1983
NSSDA	National Standard for Spatial Data Accuracy
RMSE	Root Mean Square Error

SNAP	sub-ontology of BFO - derived from snapshot in time.
SPAN	sub-ontology of BFO - derived from spanning time.
SQL	Standard Query Language
SRTM	Shuttle Radar Topography Mission
TIFF	Tag Image File Format
TIN	Triangular Irregular Network
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
WGS84	World Geodetic System of 1984

# ABSTRACT

A major benefit from the existence of widely available geographic information is the capability to simulate events for prediction, analysis, and decision-making. A reliable simulation of a geographic event is essential for the search of solutions for problems with environmental and socio-economic impacts from the local to the global level. However, the question of how to select the best among the available geographic information has not been answered.

This dissertation's objective is to develop an integrated framework that allows simulations of dynamic geographic phenomena by using multiple representations of geographic information. These simulations will be calibrated and tested against actual geographical events. Therefore, reliable simulations will be enabled by using available multiple sources of information in an effective way.

Since simulations of a geographic phenomenon require the existence of models constructed from entities of the real world, an analysis of modeling should use notions from Ontology. Ontology provides the framework that accounts for all existing entities. An information systems ontology that distinguishes entities by their modes of existence in time is used to classify geographic entities in process and objects. Therefore, the distinction between models of process and model of objects is defined.

A representation of elevation is a model of an object. Modeling, the procedure to create a model from the real world entity, produces multiple representations of elevation. The framework proposed in this dissertation handles multiple representations of elevation

and provides a linkage to a model of process. The main feature of the linkage is the use of information about elevation modeling to select the best representation of elevation based on the spatial and temporal setting of the simulated event.

The framework is validated through simulations of geophysical mass flow events. Simulations are executed by a process model of geophysical mass flow linked to multiple representations of elevation. Simulation results are compared to another simulation that uses only one representation of elevation. Since quantitative methods for comparison of geophysical mass flow simulations do not exist, a method using logistic regression was developed and used in the validation.

Simulated events are the block-and ash flow event, which occurred in April, 1991, at the Colima Volcano, Mexico, and the debris flow event, which occurred in December, 2003, in San Bernardino County, California. Comparisons of simulation results indicate that the use of multiple representations of elevation yields better results than the use of one representation only; therefore, the integrated framework proposed in this dissertation has the potential to provide reliable simulations of geographic phenomena when multiple representations of elevation are available.

# **Chapter 1**

## **INTRODUCTION**

Reliable simulations of geographic phenomena are important because the simulations are used for prediction, analysis, and decision-making on problems with environmental and socio-economic impact from the local to the global level. Simulations require geographic information that becomes more widely available everyday. Although the availability simplifies the simulation procedure, issues related to how well the information matches with the simulation requirements have not been considered. This question is important since more than one source of geographic information may be available for a simulation.

### **1.1 OBJECTIVES**

This dissertation aims to allow simulations of geographic phenomena to exploit the full potential of available geographic information by enabling simulations to use more than one source of information. In other words, the objective is to use multiple representations of the same geographic object to simulate a geographic phenomenon.

The specific geographic object targeted is elevation since only simulations of physical geographic phenomena are considered here. Elevation provides a description of Earth's surface where the gravitational force drives movement of mass; therefore, elevation influences most of the physical geographic phenomena.

A simulation relies on a model describing how it operates and on another model describing the existing conditions. The first model is defined here as a model of process and the later is a model of object. The model of process analyzed in this dissertation is a dynamic spatially distributed model of geophysical mass flows and the model of object is the representation of elevation. The focus is on the model of object, with the model of process being used to present one application using representations of elevation.

Simulations of geophysical mass flows are used to predict risk areas for landslides and mass flows, such as those related to volcanic activities. They can predict areas that can be affected, thereby helping in the preparedness for a disaster by allowing for creation of evacuation plans. Therefore, improving the results of simulation has the potential to save lives in areas with high probability of occurrence of natural disasters.

This dissertation's main hypothesis is that the use of multiple representations of elevation improves a simulation since the model of the process will use the model of the object that provides the best available information for a time and space setting.

To test the hypothesis, a framework to link multiple representations of elevation with dynamic process models is developed. The framework stores representations of elevation in a database together with information describing how the representations were created. An implementation of the link to the geophysical mass flow model is developed to test the framework for simulation of debris flows and block-and-ash flows.

In addition, a quantitative comparison method between results of simulations of geophysical mass flows is required to test the hypothesis. The comparison method is developed and used to prove that the hypothesis is valid through the comparison between



simulation results from a framework using one representation and simulation results from the proposed framework using multiple representations.

## **1.2 DISSERTATION STRUCTURE**

The dissertation is composed of three parts: review, methodology, and validation. In the first part, Chapter 2 reviews models, with focus on models used in the simulation of environmental processes. Also in Part 1, Chapter 3 presents reviews of multiple representations, elevation, and multiple representations of elevation, including existing uses of multiple representations.

In Part 2, the concepts reviewed in Part 1 are used to develop the framework to integrate multiple representations of elevation and process models. Chapter 4 presents the framework and the specific details required by a computationally intensive simulation of a process model. Chapter 5 concludes Part 2 by presenting methods that are essential for an effective use of multiple representations of elevation.

A validation of the methodology of Part 2 is presented in Part 3. Chapter 6 provides analysis of simulations of environmental processes using the proposed methodology.

A detailed description of each chapter is presented next.

### **1.2.1. Part I - Review**

Chapter 2 presents a review of models. Since models are simplified versions of the reality, the definition of models requires a framework for understanding what comprises reality. Ontology provides such a framework given that it is concerned with everything that exists and classifies all the entities of reality.

For geographic models, the classification of reality entities leads to the classification of models into models of objects and models of processes. Once models are classified, the chapter describes modeling; the procedure to create models from reality. Modeling of processes is briefly described given that it is not the main focus of the dissertation. Modeling of objects, on the other hand, is exhaustively described from the abstraction level of reality to storage in computer format.

Chapter 3 uses the modeling of objects presented in Chapter 2 to describe how multiple representations arise. In this chapter, the selection of the focus on elevation is explained through examples of applications using elevation and other representations derived from elevation.

Given that the focus is on representations of elevation, the factors that lead to the existence of multiple representations of elevation are explored. Some of the factors include the reference surface, the techniques for measuring elevation and for processing raw data, and the computational algorithms.

In addition, Chapter 3 describes the issues related to the existence of multiple representations. The issues include: how multiple representations maybe fused, how

accuracy of each representation of elevation can be estimated, and how to select the representation that best suits an application.

### **1.2.2. Part II - Methodology**

Chapter 4 presents details of the proposed framework to handle multiple representations of elevation. The core of the framework is the use of a database to store representations of elevation and their metadata. Although the main purpose of the framework is the linkage of representations of elevation to dynamic spatially distributed process models, the framework is applicable for any use where geographic information is required and there is more than one representation from which the query can be answered.

The framework presented in Chapter 4 includes optimizations required for simulation of geophysical mass flows since these dynamic spatially distributed process models are computationally intensive. The optimizations include a cache of previous queries for fast retrieval and the generation of information derived from elevation only when queried to minimize the existence of unnecessary representations.

The framework presented in Chapter 4 acknowledges a problem that arises when a query is made to a database with multiple representations. Spatial queries are seldom exact, for example, a query for elevation at a location can be answered by any representation of elevation that contains that location. Therefore, a ranking system for representations according to the query is required to allow the selection of one representation. The ranking systems should be specified by the application querying the

database. A ranking system for queries constraining resolution and date and time of the simulated event is proposed for the case of geophysical mass flows.

Chapter 5 presents methods that, although they are only indirectly related to the simulation of dynamic spatially distributed processes, are essential for the manipulation of multiple representations of elevation.

The first method is related to accuracy of a representation. The existence of multiple representations of elevation allows the definition of spatially distributed accuracy in contrast with global measures of accuracy for individual representations of elevation. The principles used to define elevation accuracy for all locations are descriptive statistical analysis and clustering analysis. These statistical tools are attached to representations of elevation. In addition, some methods search for errors in representations of elevation using spatial autocorrelation principles.

The second method presented in Chapter 5 is targeted to measure performance of simulation of dynamic spatially distributed processes. Given that more than one representation of elevation is available, comparison of results using various combinations of representations of elevation would be required to define the representations best suited for a simulation.

The performance measure method is quantitative in contrast to the existing qualitative methods for geophysical mass flows that are dependent on modeler subjectivity. The method is specific for geophysical mass flows since performance is measured by comparing the footprint of a real flow event and the simulation results. The performance value is obtained from the logistic regression where the dependent variable

is the probability of a location being inside the real event flow footprint and the independent variable is related to simulation results.

### **1.2.3. Part III - Validation and Conclusions**

Chapter 6 presents the results of the simulation of geophysical mass flows using the proposed framework at two locations, Colima in Mexico and San Bernardino in California. The simulated event in Colima is the block-and-ash flow that occurred in April 1991. In San Bernardino, the simulated event is the debris flow that occurred in December 2003.

Finally, Chapter 7 presents the conclusion and summary of the dissertation.



# **PART I: Review**

## **Chapter 2**

### **DEFINITION OF MODELS**

In this chapter a study of models for environmental systems analysis is presented, with the definition of terms and concepts used in the following chapters. Sections 2.1 and 2.2 present an introduction to models. Section 2.3 provides a definition of models using ideas from Ontology and the intersection of the definition with concepts of models in geography. In addition, the difference between model and representation for the scope of the dissertation is stated. In Sections 2.4 and 2.5, the types of models derived from the use of concepts from an information systems ontology, the Basic Formal Ontology, are presented. Procedures for modeling geographic reality are presented in Section 2.6.

#### **2.1 DEFINING MODELS**

The shortest answer to the question of what models are is to define models as simplified versions of the reality as perceived by a human observer. This definition covers all the uses of the word model and could be substituted for by another term, representation. However, for clarity, in this text representation is not used in the same sense. Here representation is a model in a computational environment. For example a representation of elevation is a digital model of Earth's surface elevation.

Since models are related to entities of the real world, a framework that accounts for all existing entities and organizes them is needed. Ontology provides such a framework, given that it is “the science of what is, of the kinds and structures of the objects, properties and relations in every area of reality” (Smith 2004). Ontology provides “a definitive and exhaustive classification of entities in all spheres of being” (Smith 2004) by accounting for all individual instances of entities, the types of categories entities belong to, and relations that hold between entities. Reality in Ontology is seen as constituted by entities and these entities are simplified into models.

In geography, models became the center of attention given during the 1960’s quantitative revolution (Martin and James 1993) and are often associated with the vague and imprecise definition given by Chorley and Haggett (1967). Chorley and Haggett’s definition accepts models as being theories, laws, hypotheses, structured ideas, roles, relations, equations, synthesis of data, and reasoning. In this definition, models are more likely to refer to processes and relations of the real world and less likely to refer to the entities that participate in these processes. However, Geographic Information Systems (GIS) models are usually representations of the participants of processes.

Given that there are different definitions of model, for clarity in this dissertation, models are used with a qualifier, such as in process model. When models are used without a qualifier, they refer to any simplified versions of the reality as perceived by a human observer.



## **2.2 USING MODELS**

Models are used due to the impossibility of having complete access, in time and space, to the phenomenon under study. Even if complete access were possible, any analysis would not be feasible due to the complexity of reality. It is the reduction of complexity and the abstraction of models that make reasoning about the phenomenon possible.

The purposes of all models fall into one or more categories: to make surrogate measurements; to test hypotheses about reality mechanisms; and/or to fill gaps in knowledge about reality.

Models allow measurements to be made that are not feasible in the real world. For example, a road map is a model of the road network where distances between cities can be measured without the need to travel the route.

Models are useful to test a hypothesis because in the real world it is often not possible to modify the parameters of a phenomenon under study. The use of models allows manipulation of parameters and verification of results to conclude if the hypothesis is valid or not.

Since complete access to a phenomenon is not possible and only samples about it are available, models are used to fill in the gaps between the samples and to extrapolate outside the space and time limits of the samples.

## **2.3 CLASSIFYING MODELS**

Since models are defined by a vague and imprecise statement, a categorization of models is required. In this study, the classification is based on the idea that models are related to entities in the real world with an ontological categorization of entities providing the basis to define types of models. However, since the focus of this work is on models of physical entities manipulable by computers, the use of categorization from information systems ontology is appropriate. In information systems, ontology is the definition of terms and relations applicable to the system based on the same methods used for Ontology in Philosophy (Smith 2004).

### **2.3.1. Types of Real World Entities**

The Basic Formal Ontology (BFO) is the most domain-neutral information systems ontology and is being developed at the Institute for Formal Ontology and Medical Information Science (IFOMIS) at the University of Leipzig (Grenon and Smith 2004) (currently IFOMIS is located at Saarland University). In BFO, the dichotomy that arises from the modes of existence of entities in time is treated by endurant entities and perdurant entities (Grenon and Smith 2004).

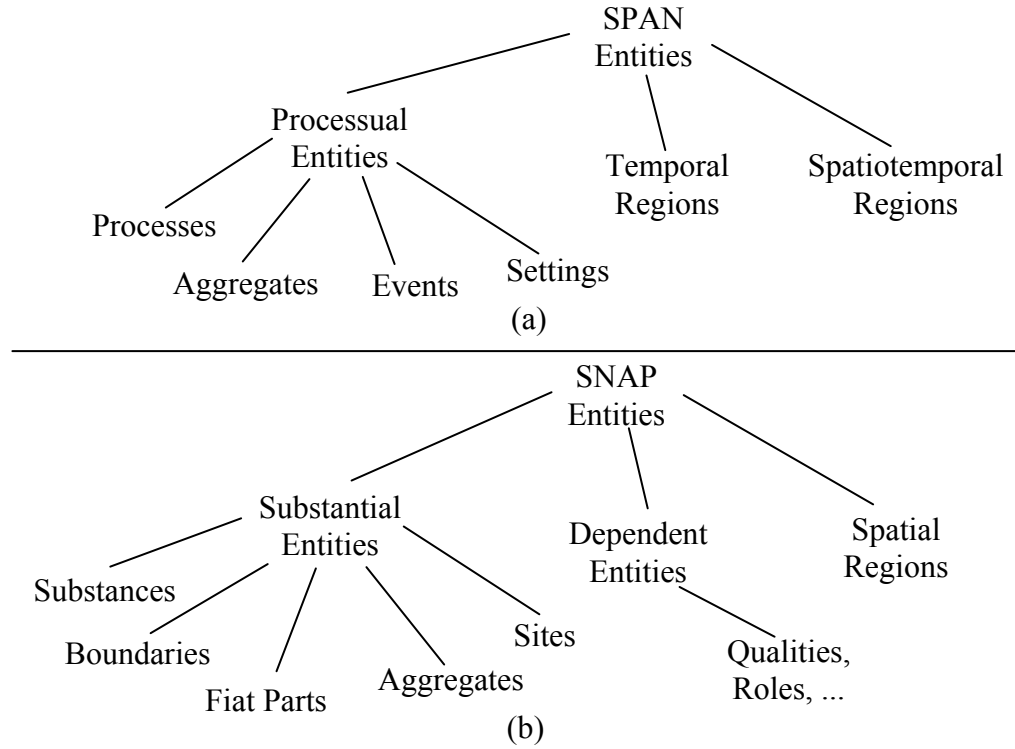
Endurants are entities of the world that although they change with time, they exist continuously and are “wholly present (...) at any time at which they exist” (Feng et al. 2004). Continuant is another term used to qualify these entities due to their continuous existence (Grenon and Smith 2004). Examples of endurants are a chair, a cat, and a tree.

Occurrents are four-dimensional entities, i.e., they occur in space and time. Occurrents are “only partially present at any time at which they exist” (Feng et al. 2004).

Existence of occurrents depends on time since they require a sequence of temporal parts or phases. A perdurant entity is an occurrent that is not instantaneous (Grenon and Smith 2004). Beginnings and endings are examples of instantaneous occurrents since they are parts of occurrent; they are occurrent because they exist as phases of another occurrent. Examples of perdurants are the building of a chair, the growing of a cat, and the evapotranspiration of a tree.

### **2.3.2. Basic Formal Ontology Sub-Ontologies**

In Basic Formal Ontology (BFO), occurrents and endurants specify the top most classification (Feng et al. 2004), and define the SNAP (derived from snapshot, for endurants) and SPAN (derived from entities that span time, for occurrents) sub-ontologies (Grenon and Smith 2004), that are connected by trans-ontological relations (Bittner and Smith 2003). Participation is one of the trans-ontological relations and it connects an endurant to an occurrent (Grenon and Smith 2004). The classification of the entities of SPAN and SNAP ontologies is presented in Figure 2.1.



**Figure 2.1** Classification of (a) SPAN ontology and (b) SNAP ontology (adapted from Grenon and Smith (2004)).

The use of endurants and perdurants notions leads to distinction of models in models of objects for endurants and models of processes for perdurants.

## 2.4 DEFINING MODELS OF PROCESSES

SPAN ontology as presented by Grenon and Smith (2004) provide a comprehensive definition of a process that is used in this work. In SPAN ontology processes are perdurants and thus they only exist in totality through a sequence along time. Process is one of the processual entities of SPAN ontology.

### 2.4.1. Processual Entities of SPAN Ontology

A processual entity is an occurrent located at temporal regions and spatiotemporal regions. Other processual entities are fiat parts of processes, aggregates of processes, events, and settings (Grenon and Smith 2004) (also see Figure 2.1a).

*Process* is a processual entity extended in space-time that does not have discontinuities (Grenon and Smith 2004). Bona fide boundaries correspond to boundaries of an entity due to discontinuities in the reality. Processes have bona fide boundaries at their beginnings and endings only. A process is connected to an enduring through participation trans-ontological relation.

When a processual entity extended in space-time that has discontinuities, it is an *aggregate of processes* (Grenon and Smith 2004). Since views of reality can be done through different levels of granularity, a process at one level of granularity can be an aggregate of processes at another level. The composition allows existence of temporal and spatiotemporal gaps.

*Events* are processual entities that exist in an instant of time only (Grenon and Smith 2004). Events are bona fide and fiat boundaries of processes. Bona fide boundaries exist where there are discontinuities in the process, such as at beginnings and endings. Fiat boundaries are boundaries defined by human intervention. A fiat boundary of a process exist where there are relevant transitions in the interior of the process.

Finally, *settings* create demarcations in the four-dimensional space (Grenon and Smith 2004).

### 2.4.2. Types of Process Models

Once process is defined according to the SPAN ontology, the next step is to classify the types of process models for clarity. There is no definitive classification and different authors classify based on different criteria, with the additional complexity that process are seldom defined or when defined definitions meanings are different. Therefore, in this study, processes models are classified based on algorithm compression as presented by Burrough et al. (1996). Algorithm compression defines how the internal structure of the model emulates a process. The types of algorithm compression are rule-based, empirical, deterministic, and stochastic (Burrough et al. 1996).

*Rule-based process models* are logical models, that is, a model where the state at the next step of the process is defined by a logical combination of conditions before and at the current temporal and spatiotemporal region of the process. Burrough et al. (1996) define rule-based models as time-independent models, however, the definition given here contemplates time dependency, subsequently it is valid as a type of process model.

*Empirical process models* deduce the process structure through regression analysis. Regression is executed on available information about the process at various times as the process evolves. Burrough et al. (1996) define empirical models as time-independent, but by making time one of the independent variables of the regression, time dependency is considered, thus an empirical process model is defined.

*Deterministic process models* assume that the evolution of a process is known and can be described by mathematical equations that relate the participant endurants. The deterministic equations are defined from theories previously established.

Finally, *stochastic process models* are differentiated from deterministic models by the equations used internally. The equations reflect stochastic behavior of the process where only probabilities are known for the next state of the process given the current condition.

### **2.4.3. Process Models and Time Dependency**

The classification given here was adapted from Burrough et al. (1996) to account for the definition of process as being extended in space and time. The definition of Burrough et al. (1996) consider that some processes are in a steady state.

### **2.4.4. Environmental Process Models**

Environmental processes are the processes related to Earth's environment that occur in a geographical scale. There are a large number of such processes from various domains, such as physical, chemical, and biological models. Couclelis (2002) includes natural and human, biotic and abiotic, atmospheric, oceanic, terrestrial and socioeconomic models. These processes are distributed over spatiotemporal regions and are complex. Complexity arises from the interactions among the various processes at different levels of granularity, since processes are not independent.

An additional classification for environmental process models is based on how spatiotemporal regions are treated. When entities are assumed to be homogeneous at the spatiotemporal region under study a lumped process model is defined (Maidment 1993). Spatially distributed models consider entities heterogeneous (at least at the level of granularity under consideration) and requires values for all locations under study (Steyaert 1993).

## 2.5 DEFINING MODELS OF OBJECTS

Objects are endurants entities (see Section 2.3.1). Objects are related to processes by the trans-ontological relation participation. Subsequently, an object participates in a process. SNAP ontology is composed of substantial entities, dependent entities, and spatial regions (Grenon and Smith 2004) (see Figure 2.1a).

The substantial entities category is composed of substances, the fiat parts of substances, boundaries, aggregates of substances, and sites (Grenon and Smith 2004).

*Substances* are substantial entities that: do not depend on other substantial entities for their existence; hold qualities that may change; have allocation in space; and are self connected wholes delimited by bona fide boundaries (Grenon and Smith 2004). These *boundaries* are substantial entities and they are lower-dimensional parts of substances, for example, a surface in a 3-dimensional substance and a line in a 2-dimensional substance.

When boundaries are defined by humans, these boundaries delimit the substantial entities *fiat parts of a substance*, and when a substantial entity is composed of separate substances, at a given level of granularity, it is an *aggregate of substances* (Grenon and Smith 2004). The substantial entity *site* is any region of space that can be occupied by a substance (Grenon and Smith 2004).

A *dependent entity* depends on an independent substantial entity for its existence (Grenon and Smith 2004). Examples of dependent entities include qualities, functions, roles, and conditions. The parts of the absolute space where enduring entities are located form a *spatial region* (Grenon and Smith 2004).



### **2.5.1. Geographic Objects**

Using SPAN ontology, a geographic object is defined in this study to be a substantial entity or a dependent entity.

A geographic object as substance is delimited by bona fide boundaries. The definition of bona fide boundaries in geographic objects depends on the level of granularity (Smith and Mark 1998b). In the highest detailed level, the only existing geographic object is the planet Earth with other geographic entities being fiat parts of Earth.

At other levels of granularity certain degrees of human arbitration are needed to define the bona fide boundaries of a geographic substance entity. For instance, ocean may have a bona fide boundary at a level of granularity inherited from a measuring instrument that averages observations from one squared kilometer.

Since a substantial entity boundary delimits a substance and fiat parts of a substance, boundaries as geographic objects are important because they may have their spatial location changed by processes.

Most geographic objects are fiat parts of the substance Earth since the delimitation of these objects results from human intervention. A geographic object may be an aggregate of other geographic objects at one or more levels of granularity.

Given that there are regions of space that can be occupied by other geographic objects, these regions are geographic objects. Finally, dependent entities of geographic

objects are their attributes obtained by measurements and all values associated to geographic objects are values of their dependent entities.

## **2.6 MODELING GEOGRAPHIC REALITY**

The feasibility of modeling an environmental process depends on the definition of relevant participants of the processes and spatiotemporal regions in a systematic approach. A system can be isolated, open, or closed.

In an isolated system there is no exchange of matter or energy with the system's exterior, while an open system does exchange matter and energy with its surroundings. A closed system exchanges energy only with its exterior environment. All environmental processes occur in an open-system; however, analysis requires an isolated system to be possible. Actually there are no closed systems except for formal logic structures (Oreskes et al. 1994).

Modeling of processes involves making assumptions that permit for a closed system. Those assumptions include: input parameters to be completely known, observations to be unbiased of inferences, and scaling to be linear (Oreskes et al. 1994). Once the assumptions are accepted, the closed system's relevant participants and interactions can be selected and modeled.

### **2.6.1. Modeling Process**

The components of a process model can be adapted from the specifications for a mathematical model in physics given by Hestenes (1987) and using the definitions from SNAP and SPAN ontologies. The components are the names of participant endurants, the

descriptors of the enduring properties, the mathematical or logical equations of the model, and the interpretations of the model.

Approaches for modeling using ontology have been proposed by Reitsma (2004) and by Feng et al. (2004). Reitsma (2004) proposes a framework to extend existing software while surface hydrology domain processes are analyzed under SPAN and SNAP ontologies by Feng et al. (2004).

### **2.6.2. Modeling Geographic Objects**

Modeling geographic objects consists of: making measurements of their geometry and of the values of their dependent entities; and structuring measurements in the form of models for querying and manipulation. If data are used to refer to raw measurements, information can be used to refer to data organized by the modeling method. Furthermore, geospatial data modeling is widely used to refer to modeling of geographic objects. Although geographic object models can be of other kinds, in this study they refer to models in digital format manipulable by electronic computers.

Modeling of geographic objects has been extensively discussed (Goodchild 1992; Laurini and Thompson 1992) and generally follows a sequence of abstraction levels from reality to the representation stored in a computer. The abstractions levels vary with different approaches since the geospatial data modeling procedure is a guide instead of a formula. The sequence usually has the following levels: ontological level, conceptual level, representation level, and implementation level.

#### **2.6.2.1. Ontological Level**

At the ontological level the portion of geographical reality that will be under study is selected and cognitively approached. Although most of the proposals for using ontologies for geospatial data are for integration of heterogeneous spatial data bases (Fonseca et al. 2002; Agarwal 2005), ontologies also provide the formal framework for defining geographic reality entities from an application point of view (Kuhn 2001).

However, ontology of the geographic reality is not easily achievable due to geographic objects being distinct from other entities in the following characteristics: geographic objects are intrinsically tied to space inheriting its mereological (related to wholes and their parts), topological, and geometric properties; categorization of geographic objects is dependent on size or scale; existence of geographic objects depends on individual or cultural perception; and boundaries of geographic objects are not well defined because the location and structure of the boundary depends on the identification of the object type (Smith and Mark 1998a).

It must be noted that although usually geospatial object modeling does not use ontology explicitly, the ontological level is always present. A modeler is applying the ontological level step when selecting the intended use by defining categories, qualities, and interrelations, which correspond to categories of substantial entities, their dependent entities, and intra- and trans-ontological relations. For example, categories selected by a geomorphologist depend on intended use. In geomorphology, the surface of the Earth is one of the main geographic entities. However, a fluvial geomorphologist would prefer to use a drainage basin category in contrast to the choice of a geomorphologist studying soil erosion who would choose to use a slope units category instead (Mark 1979).

Choices made at the ontological level propagate to geographic object models. For example, symbols in a traditional topographic map reflect choices made on the ontological level. Distortions are required to enhance the entities of geographical reality that were more important to the map maker (Monmonier 1996). A map scale is defined based on its intended use. Subsequently, intended use defines the level of granularity of the reality to be considered. For example, large scale maps are mostly used for engineering purposes, where man made structures have individual representations, while a small scale map is more a work of art in which individual structures are represented by either an aggregation of individual representations or by a cartographic symbol. Note that large scale here refers to maps where the ratio of model to reality is 1:24,000 or greater and small scales are maps where the ratio is 1:500,000 or smaller (Monmonier 1996).

#### **2.6.2.2. Conceptual Level**

The conceptual level defines the mathematical formalism used to represent the geographical reality entities in categories selected at the ontological level. Geographical reality can be mathematically interpreted as composed of fundamental elements in the form of the tuple  $T$  from (2.1) (Goodchild 1992):

$$T = \langle x, y, h, t, z_1, z_2, \dots, z_n \rangle \quad (2.1)$$

where:

the  $x, y$  pair and the elevation  $h$  define a location;

$t$  defines time; and

$z_1, z_2, \dots, z_n$  are the values of  $n$  variables at that location at time  $t$ .

Note that although Goodchild (1992) defines  $h$  as elevation,  $h$  is better defined in the context of this study as the value on the third component of the Cartesian coordinate system  $XYZ$ . Using  $h$  instead of  $z$  avoids confusion with the  $z$  values of attributes. The variables represent the dependent entities of the geographic object, as defined in Section 2.5., such as land cover and concentration of a pollutant.

Geographical reality is mapped at the conceptual level to either geographical fields (geo-fields) or geographical objects (geo-objects). If the most detailed level of granularity is used, geo-field represents a continuous spatial region of Earth, the only existing substance, and geo-objects are representations of the other substantial entities: fiat parts, aggregates, boundaries, and sites.

A geo-field  $f$  is mathematically defined by (2.2). (Câmara 1995):

$$f = [R, V, \lambda] \tag{2.2}$$

where:

$R$  is a region of space;

$V$  is a domain of values of the dependent entities; and

$\lambda$  defines the mapping between locations in  $R$  to values in  $V$ .

The mapping  $\lambda$  of (2.2) must ensure that every location of  $R$  is mapped to a value.

A geo-object  $g_o$  must be unique and is defined by (2.3) (Câmara 1995):

$$g_o = [a_1, a_2, \dots, a_n, s_1, s_2, \dots, s_n] \quad (2.3)$$

where:

$a_1, a_2, \dots, a_n$  are the values of  $n$  variables  $A_1, A_2, \dots, A_n$ ; and

$s_1, s_2, \dots, s_m$  are the  $m$  locations where  $g_o$  exists.

This definition allows a geo-object to be represented in different cartographic projections, scales, and at different times.

The values of a geo-field and of a geo-object are obtained through empirical measurement of the geographical reality. Measurement consists of assigning “numerals to objects or events according to rules” (Stevens 1946). Different rules of assignment leads to the different measurement scales: nominal, ordinal, interval, and ratio (Stevens 1946). The domain of a geo-field  $V$  and the domain of the values  $a_n$  of a geo-object depend on measurement scale.

### **2.6.2.3. Representation Level**

At the representation level, geo-fields and geo-objects are associated to either raster or vector representations (Câmara 1995). The use of raster representation in GIS is rooted in the remote sensing field while vector representation derives from spatial science (Couclelis 1992).

Raster representation uses a regular tessellation which makes access to values at any location simple. A relative location is defined by row and column indices in the same

way as in a mathematical matrix. Absolute location is defined to include a basis location and regularity parameters. The most common raster representation uses a tessellation of rectangular patches, usually named a rectangular regular grid. Other regular shaped polygons are seldom used and are limited to some implementations that optimize coverage of the space, such as hexagonal grids. The use of hexagons is due to hexagonal patches being more common in nature because three edges corners are more stable than a four edge corners; therefore, for large networks, the average number of edges per polygon approaches six (Woldenberg 1972).

Vector geometric representation uses a hierarchy of geometric primitives. The most basic primitive is a 2-dimensional point defined by a pair of coordinates in the **X** and **Y** directions. The segment primitive is composed of a pair of 2-dimensional points, where it is assumed that the points are connected by a straight line. The assumption of a straight line connecting the points implies that values associated to the segment are constant along the spatial region intersected by the line. A line primitive is composed of a non-empty set of ordered points, where every point in the set is connected to the next point by a straight line. A polygon primitive is obtained when a set of line primitives has two points sharing the same coordinates and encloses a spatial region. Values inside the polygon are constant. A polygon is allowed to contain another polygon.

#### **2.6.2.4. Implementation Level**

Raster and vector representations are manipulated by the computational system at the implementation level. Implementation level goals are storage space optimization and fast access time.



Raster representation is stored as a sequential set of binary codes with a descriptor that contains, at least, information about spatial extent and size of regular patches. Compression algorithms and tree-like data structures can be used to optimize storage space and access time. Algorithms to access raster data directly, by blocks, and by resolution level are also part of the implementation level.

Vector representation is stored using a hierarchical structure that corresponds to the geometric primitives. Spatial indexing techniques are used for efficient access of elements from vector representation (Laurini and Thompson 1992).

## **2.7 SUMMARY**

This chapter presented a study of models for environmental systems analysis. Models of processes and models of objects were defined using ideas from information systems ontology. The referenced ontology is the Basic Formal Ontology. Objects are endurant entities of the SNAP sub-ontology and processes are perdurant entities of the SPAN sub-ontology. Modeling of objects and processes were presented, with emphasis on modeling of geographic objects. Representation was defined as the object model in digital format. This chapter provided the support to study multiple representations in the next chapter.



## **Chapter 3**

# **DEFINITION OF MULTIPLE REPRESENTATIONS OF ELEVATION**

In this chapter a study of multiple representations of elevation is presented. First, Section 3.1 provides a definition of digital geographic data representation and why multiple representations are possible. In addition, some examples of the current uses of multiple representations of geographic objects are presented. Then in Section 3.2, elevation is defined and the importance of elevation in environmental studies is presented through examples of applications and other representations derived from elevation. Modeling of elevation is presented in Section 3.3, with the highlighting of sources for multiple representations in Section 3.4 through 3.7. Finally in Section 3.8 issues related to the use of multiple representations are presented.

### **3.1 DEFINING DIGITAL GEOGRAPHIC DATA REPRESENTATION**

In this section digital geographic data representation is defined and sources of multiple representations are presented. Also, current uses of multiple representations of geographic objects are presented.

The objective of geographic object modeling is to obtain a representation. The representation may take an analog form, such as “a scale-down replica of a mountain or

an airplane wing for use in wind tunnel experiments” (Steyaert 1993). A digital representation is a model of an entity of real world in digital format manipulable by computer. When the entity being represented is an endurant it is represented by a data representation. Digital geographic data representation is the data representation of a geographic object in digital format.

Digital representation is constrained by the computational environment. Any data in a computational context has to be finite and discrete which imposes requirements for types of data and data values.

The finiteness constraint requires the specification of the expected bounds for the data. The bounds in the type of data define the expected range of values, such as between negative 32767 and positive 32767 for a 16-bit integer. The amount of data that is also constrained by finite requirements since the size of storage is finite.

The discreteness restriction defines the smallest difference between values that can be discerned for a data type. There are computational techniques to manipulate numbers with adaptative precision (Shewchuk 1997), however, they are useful for some geometric algorithms and would not benefit geographic data modeling.

### **3.1.1. Restrictions to Represent Geographic Data in Digital Format**

Restrictions on representing geographic data in digital format include: minimum distance between different locations, minimum difference in value, maximum size of region under study, and maximum amount of data.

Restrictions of digital format are related to the computational types used to represent a value. For the example of restriction on minimum distance between different locations, consider that the largest coordinates in Universal Transverse Mercator (UTM) projection are in tens of millions of meters. The Institute of Electrical and Electronic Engineers (IEEE) floating point standard IEEE Standard for Binary Floating-Point Arithmetic (IEEE 754) specifies the single precision using 32 bits, with 23 bits for fraction part and 8 bits for exponent (Standards Committee of the IEEE Computer Society 1985). The number of significant digits for the fraction is the number of digits of 8,388,608 (equivalent to  $2^{23}$ ), that is 7. Therefore, if a 32-bit floating point is used to represent UTM projection coordinates, locations would be in the worst case scenario up to one meter from the actual location.

When an integer type is used to represent values, the minimum difference in values will be limited to one. Integer type can also be used to define a discrete partition of space and the maximum number of partitions will be limited to 65,536 if a 16-bit integer type is used.

The amount of data that can be stored depends on the size of available storage. The storage will always be finite even when its limit seems too large. Digital data requires processing power for tasks such as querying and transformation. Computational processing capabilities are also finite under a time constraint, since tasks have to be executed within a time interval.

### **3.1.2. Source of Multiple Representations**

Geographical reality is dynamic and representations are dependent on the “point-of-view” - intentions, previous experience, and concepts owned by the representation builder. In addition, differences in technologies for sampling real world play a big role.

Multiple representations of geospatial data arises from the various choices available at each of the steps of modeling procedure: ontological, conceptual, representation, and implementation levels. At every step in the modeling process the modeler creates a new path that leads to different representations of the same entity of the real world. More detailed description of the choices will be given with examples for multiple representations of elevation in Sections 3.4 to 3.7.

### **3.1.3. Current Use of Multiple Representations**

Existing approaches to handle multiple representations of geographic data are targeted to integrating representations from different geographic databases and to create maps of certain scale from geographic data of different scales.

The objective of integrating different databases is to be able to create a unique database schema. This integration can be accomplished by creating a new database from the existing ones, adding a centralized schema, or adding relationships between the schemas of the existing databases (Balley et al. 2004). The solution proposed by Balley et al. (2004) is flexible, given that it allows the use of a new object that represents the unified objects at the multiple original schemas and also for the use of relation in the new schema to link the multiple representations at the original schemas.

In an ideal situation, generalization techniques could generate maps at any scale from unique geographic data with all details required for the output maps. However, this situation does not seem to be possible. The solution for map producers is to generate maps at a given scale from one of the various representations available. Buttenfield and Hultgren (2005) propose a solution for generating maps by producing a Digital Landscape Model (DLM) that stores geographic data representations, as similar as possible to the captured data, including the level of granularity of the capture. The Digital Cartographic Model (DCM) derives maps of certain scale selecting representations from DLM and using generalization techniques.

Balley et al. (2004) identify some of the problems that are not addressed by current uses of multiple representations of geographic data. They are the lack of redundant data treatment and for inconsistencies between representations. In addition, Buttenfield and Hultgren (2005) include the limits of transforming from one resolution to another and the creation of metadata, especially for already existing data, since most of the history about the transformation procedures applied to data have been lost.

### **3.2 DEFINING ELEVATION**

Elevation is an enduring entity that depends on the enduring substance Earth. Elevation value at one location is the distance along the vector that is normal to a reference surface, and passes through that location on Earth's surface. The reference surface is based on Earth's gravity potential.

Elevation participates in most environmental processes since gravitational force influences the movements of mass of all physical entities. In addition, elevation influences atmospheric temperature and the surface normal direction defines exposition of the surface to Sun radiation and weather systems.

A set of attributes, dependent endurant entities of Earth's surface, are obtained from elevation. They describe Earth's surface in attributes that are more appropriate for some of the environmental processes. Li et al. (2005) separates the attributes in geometric, hydrologic, and visibility categories.

Geometric attributes includes computation of areas and distance on the surface, volumes between the surface and another arbitrary surface, surface slope and aspect, and surface curvatures.

Hydrologic attributes are related to the processes of transportation of water and sediments over the surface. Wilson and Gallant (2000) describes various hydrologic attributes, including flow direction and flow accumulation, catchment area, wetness index, and stream power index.

Visibility attributes includes viewshed and solar radiation attributes. In viewshed, locations that are visible from a set of locations are defined. Solar radiation is computed from the surface's normal direction, Sun direction, and time.



### **3.3 MODELING REPRESENTATIONS OF ELEVATION**

A digital representation of elevation is commonly identified by the acronym DEM, which stands for Digital Elevation Model (Doyle 1978; Maune 2001; El-Sheimy et al. 2005). The first known digital representation of elevation was named Digital Terrain Model (DTM) (Miller and Laflamme 1958). However, since terrain has a wider meaning compared to elevation the later term is preferred. Discussions on the meaning of DEM, DTM, and other terms used to refer to the digital representation of elevation are found in Maune (2001), El-Sheimy et al. (2005), and Li et al. (2005). In this study, for clarity, DEM refers to the representation of elevation and the representations derived from a DEM are referred to by their names, such as representation of slope and representation of curvature.

At the source of multiple representations of elevation are the choices available at each step of the modeling elevation procedure. Modeling elevation follows the same steps as modeling geographic objects from selecting the approach by defining the purpose of the elevation model up to the selection of data structures for computational storage.

### **3.4 SOURCES FROM ONTOLOGICAL LEVEL**

Choices that are made at the ontological level for a DEM define how reality is measured to provide elevation data. At this level, the intended use of the DEM is defined, imposing the definition of which surface elevation is modeled and the requirements for data collection.

The elevation surface is usually considered the bare Earth, without vegetation and without man-made features (Maune et al. 2001). There are exceptions for some man-made structures and road-like features that are part of the elevation surface. These features are made mostly with dirt and rocks, such as embankments and areas of cut and fill for roads. This definition of elevation is appropriate for hydrological applications. However, if the intended use of the DEM is for visibility applications, vegetation and man-made structures must be represented.

An example of requirements defined by intended use is the level of detail of the data collection. If a DEM will be used for detailed construction plans of roads, elevations are required to be represented at a higher level of detail than if the DEM would be used to create orthophotos from remotely sensed images.

### **3.5 SOURCES FROM CONCEPTUAL LEVEL**

At the conceptual level, the mapping between geographical reality and geo-fields or geo-objects is defined. Reality is sensed through samples obtained by a data collection technique that includes measurement. Elevation is measured on a ratio scale, one of the measurement scale types defined by Stevens (1946). A ratio scale domain of values is the set of real numbers, the scale has a true zero, and conversion of values of the scale requires only the multiplication of the scale value by a constant (Stevens 1946). Therefore, the domain of values for elevation is the set of real numbers and the true zero is defined at the reference surface.

The selection of the data collection technique to provide values of elevation is tied to the required level of detail defined at the ontological level, and each one of the available data collection and processing techniques will lead to a different representation of elevation.

### **3.5.1. Mapping Reality to Fields or Objects**

Since geographical reality can be mapped to either geo-fields or geo-objects, each of the choices will lead to a different representation of elevation.

#### **3.5.1.1. Mapping to Geo-Fields**

Using the definition from Chapter 3 - Conceptual Level - an elevation geo-field  $f$  is defined by  $f=[R,V,\lambda]$ , where  $R$  is a geographic region,  $V$  is the domain of real numbers, and  $\lambda$  defines the mapping between locations in  $R$  to elevation values in  $V$ . The mapping  $\lambda$  must ensure that every location of  $R$  is mapped to a value of elevation. Since it is not feasible to sample all location of  $R$ , the mapping  $\lambda$  includes some sort of interpolation.

#### **3.5.1.2. Mapping to Geo-Objects**

Using the definition from Chapter 3 - Conceptual Level - an elevation geo-object  $g_o$  is defined by  $g_o=[a_1,a_2,...a_n, s_1,s_2,...s_m]$ , where  $a_1,a_2,...a_n$  are the values of  $n$  variables  $A_1,A_2,...A_n$ ,  $s_1,s_2,...s_m$  are the  $m$  locations, and one of the  $A$  variables is elevation. The  $m$  locations are points where measurements of elevation were made or locations derived during the processing of raw data.

### **3.5.2. Acquisition Time**

Reality is dynamic and thus elevation changes with time, by natural processes such as plate tectonics, water erosion, volcanic eruptions, landslides, and by anthropogenic processes related to land use change, such as mining and road construction. Using radar interferometry, Wadge et al. (2002) found 85 meters increase in elevation after an eruption at Soufrière Hills Volcano, Montserrat, in some valleys, due to deposits of pyroclastic flows. Results from simulation of processes will vary depending on time of acquisition of the data used to create the DEM, even when changes are not as dramatic as in areas affected by volcanic eruptions.

### **3.5.3. Reference Surface**

The reference surface is not unique, although based on Earth's geoid, the gravity equipotential surface at the mean sea level. The geoid can be approximated by leveling or by an ellipsoid. The leveling procedure consists of using line-of-sight instruments to measure the angle between the perpendicular direction to the local gravity and a reference location (Li et al. 2005).

Technological and practical limitations imply that only an approximation of the gravity equipotential surface by leveling can be obtained. Furthermore, the reference gravity equipotential surface is based on mean sea level. Local currents and prevailing wind direction and strength affect sea level and adjustments are required to define the equipotential surface. Therefore, there are many different gravity equipotential surfaces available.

The reference surface can use an ellipsoid that approximates Earth's geoid. Reference ellipsoid is not unique given that for different locations on Earth different definitions of the ellipsoid adjust better to the geoid. For example, the ellipsoid used by the World Geodetic System of 1984 (WGS84) was defined to be the best fit for the whole Earth. This ellipsoid is within 4 meters above and below the geoid model of the WGS84 Earth Gravitational Model (EGM) for 93% of Earth's surface (Kumar 1993).

#### **3.5.4. Quality of the Representation**

Quality is a generic term that depends on the context and in a geographic representation context, it is more difficult to define given that physical characteristics of geographical reality can not be directly assessed (Veregin 1999). The quality of a representation is related to the accuracy, precision, consistency, and completeness of the representation. For elevation representation, accuracy is defined as the measure of its quality.

Accuracy is a measure of how different the representation is in relation to the real world entity (Veregin 1999). Some of the metrics developed to summarize the differences are the Root Mean Square Error (RMSE) and the vertical accuracy at 95% confidence (Daniel and Tennant 2001). DEM accuracy is difficult to be assessed since there is no independent model of the real world to test our digital model against (Carter 1988). The "true value" of elevation is just a representation that is considered to have higher accuracy than the one having its accuracy defined. In addition, accuracy of DEMs is dependent on the Earth's surface characteristics and the measurement techniques, thus errors are not randomly distributed over the entire DEM (Lee et al. 1992).

Accuracy standards are deficient since they define measures of accuracy on random sample points. The National Standard for Spatial Data Accuracy (NSSDA) published in 1998 by the Federal Geographic Data Committee (FGDC) states that accuracy is calculated using a minimum of 20 check points distributed for the whole area covered by the DEM (FGDC 1998). Therefore, the stated accuracy may not represent the real accuracy.

### **3.5.5. Measuring Instrument**

Each measuring instrument has its own method, precision and accuracy. Furthermore, remote sensing instruments measure values over an area, with sizes depending on the sensor resolution and the distance from the sensor to Earth's surface. Some of the instruments and techniques used to measure elevation are traditional survey, Global Positioning System (GPS), photogrammetry, radar interferometry, and Light Detection And Ranging (LIDAR).

#### **3.5.5.1. Traditional Survey**

Position in traditional survey techniques is estimated based on other known positions. The elevation at a location can be estimated by measuring the vertical angle from a known location and the 2-D coordinates of the location are estimated from horizontal angles and distances to a pair of known location (Li et al. 2005).

### **3.5.5.2. Global Positioning System**

GPS measurements are made using a range intersection technique, where the distance from a location is defined by the distance to at least three satellites orbiting Earth (Li et al. 2005). Distance to a satellite is calculated from the time lag required for an electromagnetic signal to travel the distance. Clocks at the GPS receiver and satellite must be synchronized and accurate for a precise measurement. In addition, the speed at which an electromagnetic wave travels is affected by atmosphere.

Therefore, GPS elevation values can be up to tens of meters different from the correct value. However, GPS measurements can be improved to obtain accuracy at the centimeter level by using more than 3 satellites, measuring the same location for a long period of time, employing additional GPS measurement at another location, and correcting with information from other well known locations (Renschler et al. 2002).

### **3.5.5.3. Photogrammetry**

Photogrammetric measurements are made on a pair of images with overlapping that provides a reproduction of the stereo geometry used for the acquisition of the images. Measurement of parallax between the same targets on the images can be related to the distance from the targets on Earth's surface to the image acquisition platform (Molander 2001; El-Sheimy et al. 2005).

The accuracy of elevation extracted from photogrammetric techniques depends on the resolution of the image and on the distance of Earth's surface to the platform, with values of accuracy below 10 centimeters being possible (El-Sheimy et al. 2005).

#### **3.5.5.4. Radar Interferometry**

The Radar Interferometry technique consists of emitting an electromagnetic wave and using the difference in the phase of the signal returned from Earth's surface to compute the distance of the platform to the ground (Rosen 2000).

The most comprehensive data collection using radar interferometry was the Shuttle Radar Topography Mission (SRTM). The mission, flown on the Space Shuttle Endeavour on Flight STS-99 during February 2000, collected data to create elevations for land between 60 degrees north latitude and 54 degrees south latitude in two different frequencies, the C-band and X-band (van Zyl 2001; NASA 2002; DLR 2003).

The target vertical accuracy of elevation from SRTM is an error of 16 meters at the 90% confidence level for C-Band (van Zyl 2001). Rodríguez et al. (2005), in regards to global assessment, found that accuracy is between 5 and 8 meters. The X-band coverage is half of the C-band with the same vertical accuracy, achieved when systematic errors are eliminated (Rabus et al. 2003).

#### **3.5.5.5. Light Detection And Ranging**

A Light Detection And Ranging (LIDAR) system is composed of a laser range finder, a Global Positioning System (GPS) receiver, and a Laser Inertial Navigation System (LINS) (Adams and Chandler 2002). The laser scans the terrain surface perpendicular to the flight direction. Data from LINS, GPS and laser range are processed to generate 3-dimensional coordinates.



A typical laser beam used in airborne LIDAR projects a footprint of 24 to 60 cm in diameter on the ground at a distance of 1,219 meters (4,000 ft.), using a wavelength between 1.053 and 1.064  $\mu\text{m}$  (Hodgson et al. 2003). Vertical accuracy can be as high as 0.10 meters, but depends on the accuracy of sensor components, laser ranging and scan angle measurements, ground survey control points, land cover, and data processing (Hill et al. 2000).

### **3.5.6. Raw Data Processing**

Data acquired by instruments require processing to obtain elevation values. Examples of factors influencing the processing of instruments data include the use of ground control points, noise filtering, and phase unwrapping processing.

### **3.5.7. Ground Control Points**

Ground control points provide reference locations used to fix all the measured distances to Earth's surface. The selection and accuracy of each control point will vary, thus the possibility of multiple representations.

For example, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), onboard National Aeronautics and Space Administration (NASA) Terra satellite, acquires stereo images suitable to generate DEMs (Welch 1998). Eckert et al. (2005) found differences in elevation of 2 meters using two different data sets of control points for creating DEM from ASTER stereo images at a site in Switzerland.

### **3.5.8. Noise Filtering**

Noise filtering is essential in radar systems, since signals returned from reflections have a multiplicative effect, generating speckle noise (Rosen 2000). Some of the available filters to remove speckle noise are Frost, Lee, and Kuan filters (Lopes et al. 1993), with different results, thus generating multiple representations of elevation.

### **3.5.9. Phase Unwrapping**

Phase unwrapping processing is used in radar interferometry, consisting of transforming the phase information of raw data to differences in distance, and can be accomplished by different procedures (Lanari et al. 1996; Sansosti et al. 1999), with each technique having the potential to generate a different representation of elevation.

## **3.6 SOURCES FROM REPRESENTATION LEVEL**

Elevation geo-fields and geo-objects are associated to either vector or raster representations at the representation level.

### **3.6.1. Representations of Elevation Geo-Fields**

Representations of elevation geo-fields are most frequently associated to raster geometry that uses rectangular patches organized in a rectangular regular grid. However, vector geometries are also used, such as triangular irregular network (TIN) and contour lines.

#### **3.6.1.1. Triangular Irregular Network**

A TIN is a vector geometry used to represent DEMs (Kumler 1994; Li et al. 2005). The basic geometric primitive of a TIN is a triangle. Elevation inside a triangular patch is defined by elevation values at its vertices. TINs are used since carefully constructed TINs, that is, when they include breaks in the continuity of the surface, such as at ridge and valley lines, require less storage space than a regular grid of rectangular patches to represent the surface accurate enough to calculate the surface's geomorphometric parameters (Mark 1975).

#### **3.6.1.2. Contour Lines**

Contour lines represent DEM geo-fields by vector line geometries. Elevation between two adjacent contour lines is defined to be inside the range of elevations of the two contour lines. In a contour lines representation, some locations that are significant to describe the surface are also represented with its elevation.

### **3.6.2. Representations of Elevation Geo-Objects**

The most common representations of elevation geo-objects are the sample points in a contour map. Each sample point is a geo-object with a value associated with the attribute elevation.

In addition, surface networks represents Earth's surface through a graph representation with critical points as vertices and critical lines as edges (Wolf 2004). Critical points are the peaks, pits, and passes of a surface. Critical lines are the ridge and

course lines. Although surface networks representations are seldom used, they have the potential to improve applications of DEMs such as visualization (Rana 2004).

### **3.7 SOURCES FROM IMPLEMENTATION LEVEL**

The implementation level consists of the storage data structure and the algorithms to access and manipulate elevation data.

#### **3.7.1. Storage Data Structure**

The raster storage of real numbers scale of DEMs must have the definition of how to interpret the elevation data stored in a cell using the extent and resolution information. Three different possibilities exist for this interpretation (Wise 2000b): the pixel model, where the value stored at the cell is the same for the whole rectangular region of the cell; lattice model, where value is known only at the center of the cell; and lattice model with points at cell corners, where a cell has different values at its vertices.

The lattice models require interpolation methods to estimate the values at locations inside the cell that are neither at the center (for the cell centered lattice model) nor at the vertices (for lattice model with values at vertices) of the cell.

#### **3.7.2. Manipulation Algorithms**

Manipulation algorithms for DEMs may be classified into two types, interpolation and transformation. Interpolation algorithms are used to define values at locations where no explicit values are assigned. They can be used to transform between different

representations of geo-fields such as creating a raster geo-field from a contour map. Some transformation algorithms are also used to make conversions between DEM representations, such as converting from a raster geo-field to a TIN, from contours to TIN, from raster to contours.

Aggregation algorithms are a type of transformation that converts from one raster to another raster, where the size of the cell in the destination is larger than in the source. Given that DEMs are representations of continuous surfaces, partial derivatives of the DEM are useful for many applications. Transformation algorithms to calculate the partial derivatives are based on finite elements and differentiate from each other based on the size and type of the neighborhood used by the algorithm.

### **3.8 MULTIPLE REPRESENTATIONS OF ELEVATION ISSUES**

When there are multiple representations of elevation, it is expected that advantages exist for applications that use elevation. For example, fusion of representations of elevation would allow for creating a new representation incorporating data from the source representations. In addition, comparison between different representations would be possible. Also, the representation that better suits a given process model could be determined.

#### **3.8.1 Fusion**

Fusion integrates data from different representations to produce a new representation. Some fusion techniques handle integration of database schemas and the

geometric representation of geo-objects (Balley et al. 2004; Mancarella et al. 2004; Duckham and Worboys 2005).

Hovenbitzer (2004) describes the fusion of DEM from states in Germany where vertical accuracies range from 0.5 to 2 meters to create a Germany wide DEM at the 1:25,000 scale. Fusion techniques includes the use of elevation estimated using maximum likelihood, where values at different representations are used to calculate the conditional probability (Ye et al. 2003). Podobnikar (2005) proposes fusion by two techniques: interpolation with elevations from all available representations; and weighted sum of all available DEMs.

### **3.8.2. Estimation of Data Quality**

Each one of the available DEM representations has its sources of uncertainty in elevation values. Therefore, the quality of a DEM varies and is due to choices made at the data modeling steps.

Accuracy requirements and effective values of accuracy for some elevation sources were presented in Section 3.5. For simulation of processes a spatial distribution of errors is useful given that the outcomes from the simulation would have local uncertainties associated with the prediction.

Multiple representations can be used to estimate spatially distributed uncertainty through statistical analyses. Elevation is a random variable and descriptive statistics is used to define the probability distribution of elevation values at one location. For one DEM, the elevation value is defined by the value and the confidence of that value being

the true value. Given that errors in a DEM are not randomly distributed, clusters of high values of uncertainty are more correct. Rogerson (2001a) proposes a clustering detection in a rectangular regular grid that is suitable for identifying clusters of high uncertainties and could be used in the DEM context.

### **3.8.3. Defining Fitness for Use**

The criteria used to define the fitness-for-use of a data representation for simulation of an environmental event are based on the type of the process model, scale of the simulation, and time of the simulated event. In an ideal case, any process model has been tested and the sensitivity to each of the factors that influence the simulation results has been defined. However, when the process models are tested, only instrument characteristics and resolution are tested.

#### **3.8.3.1. Resolution**

The most studied characteristic of DEM for its influence on simulation is resolution, with most of the studies being on simulation of hydrological processes. The conclusion of the studies is that increasing resolution improves predictions in hydrological models up to a certain limit and increased resolution after the limit does not provide advantages.

For example, Garbrecht and Martz (1994) state that resolution must be relative to the size of the drainage features that are intended to be extracted. Horritt and Bates (2001) calibrate a flooding model using the observation of a flood and find that the flood

prediction increases with resolution up to a size relative to resolution used in calibration and no improvement is achieved with higher resolutions.

Moglen and Hartman (2001) use the highest resolution DEM as the basis to calculate the true hydrological parameters. When comparing this “true” with parameters extracted from coarser resolutions, a linear relation where differences increases with coarser resolution is found for 44 watersheds in Maryland.

Clarke and Burnett (2003) compare positional accuracy of streams extracted from USGS 10 meter and 30 meter resolution DEMs and the 10 meters resolution gives a better result when comparing with streams from 1:24000 scale topographic maps.

#### **3.8.3.2. Measuring Instrument**

Comparison of drainage network, watersheds and prediction of hydrological response from DEM created from the same source data and different interpolation methods indicates that the smoothest DEM is usually the best (Wise 2000a).

A comparison between different measuring instruments indicates that, at 30 meter resolution, for a watershed at Mahantango Creek, Pennsylvania, a DEM from radar interferometry is outperformed by a standard Level 1 United States Geological Survey (USGS) 7.5 minutes DEM (Kenward et al. 2000). Accuracy of a Level-1 DEM is a RMSE of 7 meters, with maximum permitted of 15 meters (USGS 2003). The result of the study is due to random noises of the radar interferometry technique while USGS DEMs are derived using photogrammetry.



Although differences in DEM from topographic maps and from interferometry radar are present due to forest cover, a volcanic flow model is affected more by errors in elevation of areas with a slope less than 4 percent and on stream channels (Stevens et al. 2002)

#### **3.8.3.3. The Best Representation of Elevation**

The most fit-for-use DEM is the one with the appropriate resolution, the best characteristics of the measuring instrument, and the most correct collection date. Collection date impact will depend on the geomorphic processes in the region. The selection of the appropriate DEM will depend on availability of metadata describing the steps of the data modeling.

### **3.9 SUMMARY**

Multiple representations arise from the choices made during data modeling, starting from the intended use to the computational implementations of algorithms to derive information. DEMs, the representations of elevation, vary from the selection of the reference surface and the distance measurement techniques, to algorithms used to calculate derivatives of elevation.

Examples of the use of multiple representations show that there is no application that takes advantage of multiple representations to simulate environmental processes. Requirements for the effective use of multiple representations are the existence of metadata and process models that state their limitations. Metadata should describe data

modeling steps. Process models using DEM should have a sensitivity analysis to resolution, accuracy, and other characteristics of DEMs and their derived attributes.

## **PART II: Methodology**

### **Chapter 4**

#### **INTEGRATION OF MULTIPLE REPRESENTATIONS OF ELEVATION AND PROCESS MODELS**

In this chapter the proposed framework to link process models to multiple representations defined in Chapter 3 is presented. The objectives of the framework are to use multiple representations in simulation, to minimize generation of additional representations, to provide fast retrieval of information from representations, and to provide information from the representation best suited for process model requirements.

The focus is on representations of elevation since elevation is the base representation required for the type of process models targeted by the framework. However, process models also require representations derived from elevation and those are also handled by the framework. Therefore, the multiple representations term is used to refer to any set of representations of one geographic object and the multiple representations of elevation term is used to refer to representations of the geographic object elevation.

The purpose of the use of multiple representations is the simulation of dynamic spatially explicit environmental process models. These models are presented in Section 4.1, and why simulations of these processes are more demanding than simulations of

other processes is explained. In Section 4.2 the types of linkage between process models and Geographic Information Systems (GIS) are presented and the choice of linkage type used in the framework is justified. The targeted process model and the current implementation of the link of the process model to the GIS are presented in Sections 4.3 and 4.4. The GIS library used in the implementation of the framework is presented in Section 4.5, along with the details of how multiple representations and its metadata are stored. Details of the new linkage implementation are presented in Section 4.6. Finally, in Section 4.7 the session for simulating the process model is presented.

## **4.1 DEFINING DYNAMIC SPATIALLY EXPLICIT PROCESS MODEL**

The focus of the framework is integration with dynamic spatially explicit environmental models, the most demanding process models in terms of computational and representation requirements.

### **4.1.1. Dynamic Process Model**

Process models are models of perdurants, as defined in section 2.3.1. A perdurant occurs along a time sequence; therefore the values of a process are dependent on time. A steady state process model describes the process where derivatives in time of the process are zero and algebraic equations are enough to calculate the process state (Kemp 1993; Steyaert 1993).

Dynamic process models are defined as models of processes in their transient states. The equations that describe the model include at least one derivative in time. Computational solutions for the differential equations must choose a time resolution that better approximates the computational and mathematical solution. While the simplest computational solution uses a fixed time difference (Burrough 1998), more sophisticated ones predict time differences at every step of the computation through the speed of propagating wave (Pitman et al. 2003; Patra et al. 2005).

#### **4.1.2. Spatially Explicit Process Model**

Process models occur in spatiotemporal regions and how the regions are represented in the equations of the model distinguishes between lumped and spatially explicit process models, as presented in Section 2.3.2.2.

A lumped model represents spatially distributed values by their averages over the analysis area (Kemp 1993; Maidment 1993; Brimicombe 2003). In modeling of hydrological processes, the lumped approach divides the analysis areas into units, such as watersheds, where values are averaged and the simulation provides a unique value for the whole watershed (Maidment 1993). Lumped model equations may be expressed in difference, ordinary differential, or simple algebraic equations.

In a spatially explicit process model, values vary with their location; therefore, the model equations include derivatives of space (Kemp 1993). The simulation results of these models are also distributed over the region. The most common computational

solutions for derivatives in space use finite element and finite difference methods (Mitasova and Mitas 2002)

## **4.2 LINKING ENVIRONMENTAL PROCESS MODELS AND GIS**

Simulation of environment process models requires values of spatial entities to solve process equations. GIS manipulates representations of spatial entities; therefore, a linkage between environmental process models and GIS is essential.

### **4.2.1. Existing Linkage Types**

The modes of integrating environmental processes and GIS are loose coupling, tight coupling, and embedded coupling (Brimicombe 2003).

*Loose coupling* integrates the process model and GIS by exchanging common files (Fedra 1993), either files generated by the GIS to provide lumped or spatially distributed values or results from the simulation to be analyzed and visualized by the GIS. A common human-computer interface can be built. However, GIS and process models are independent software systems.

*A tight coupled integration* requires that a new software system be written. The new software has access to the process modeling capabilities and to the representations in the GIS. GIS and process models are required to have an interface that allows access to the implementation of the numerical solutions of the process models and to the values stored in the GIS representations.

An *embedded coupled integration* is obtained when either a GIS or a process model has the capabilities of the other. Thus when a process model can be implemented using the available features of a GIS, the process model is embedded in the GIS. Similarly a GIS is embedded in a process model if a process models has the capabilities of a GIS.

#### **4.2.2. Selection of Coupling Type**

Although embedded coupling seems ideal, a GIS developer can not predict all possible implementations required to solve process equations and a process model can not support all types of representation that may become available. Software systems using embedded coupling are limited in their functionalities. For example, PCRaster (Burrough 1998) does model processes, but requires the use of a fixed resolution rectangular regular grid and a fixed time step.

Loose coupling has the advantage of keeping the process model and GIS independent and is effective when the amount of values exchanged is small, such as in lumped process models. When the process model is spatially explicit, the tight coupling is preferred since values can be directly accessed from GIS representations. The proposed framework is targeted to tight coupling of environmental process models and GIS.

### **4.3 LINKING SPATIALLY EXPLICIT PROCESS MODEL TITAN2D TO GIS**

The proposed framework benefits spatially explicit process models that require multiple representations of elevation. TITAN2D is the process model that requires elevation and its derivatives at different resolutions (Patra et al. 2005). In addition, TITAN2D simulates processes that drastically change the surface. Therefore representations of elevation with different measurement date and time influences its results. In this section, a brief description of TITAN2D is first presented, along with the current implementation that uses only one representation of elevation. The new implementation using multiple representations is presented later in Section 4.3.

TITAN2D is a tool for simulation of geophysical mass flows developed by a multidisciplinary group at The State University of New York at Buffalo that includes the departments of Geology, Mathematics, Mechanical and Aerospace Engineering, and Geography (Patra et al. 2005). The process model of TITAN2D is an averaged granular flow model based on laws of conservation of mass and momentum (Sheridan et al. 2005).

The governing equations of the flow model form the system of partial differential equations given by the depth-averaged mass balance, X-direction momentum balance, and Y-direction momentum balance equations presented by (4.1) (Patra et al. 2005):



$$\begin{cases} \partial_t h + \partial_x(hu) + \partial_y(hv) = 0 \\ \partial_t(hu) + \partial_x\left(hu^2 + \frac{\beta}{2}g_z h^2\right) + \partial_y(huv) = hg_x - h\beta \operatorname{sgn}(\partial_y u) \partial_y(hg_z) \sin(\varphi) - \frac{u}{\sqrt{u^2 + v^2}} \left[ hg_z \left(1 + \frac{u}{r_x g_z}\right) \right] \tan(\delta) \\ \partial_t(hv) + \partial_x(huv) + \partial_y\left(hv^2 + \frac{\beta}{2}g_z h^2\right) = hg_y - h\beta \operatorname{sgn}(\partial_x v) \partial_x(hg_z) \sin(\varphi) - \frac{v}{\sqrt{u^2 + v^2}} \left[ hg_z \left(1 + \frac{v}{r_y g_z}\right) \right] \tan(\delta) \end{cases} \quad (4.1)$$

where:

**XYZ** is the Cartesian coordinate system, with plane **XY** parallel to the basal surface;

**h** is the flow layer thickness;

**u** is the component of the flow velocity field in downslope direction;

**v** is the component of the flow velocity field in across-slope direction;

**r<sub>x</sub>** and **r<sub>y</sub>** define the radius of curvature of the local basal surface;

**g<sub>x</sub>**, **g<sub>y</sub>**, and **g<sub>z</sub>** define the local gravity vector;

**β** is a function of the so-called pressure coefficient;

**φ** is the flowing pile internal friction angle; and

**δ** is the basal friction angle.

The solution of the governing equations uses the parallel adaptative mesh Godunov scheme (Patra et al. 2005). Local direction of gravity (**g<sub>x</sub>**, **g<sub>y</sub>**, and **g<sub>z</sub>**) and the radius of curvature (**r<sub>x</sub>** and **r<sub>y</sub>**) of the local basal surface are estimated from the first and second order derivatives of the elevation.

### 4.3.1. Parallel Adaptative Mesh Grid Godunov Scheme

The Godunov scheme represents and evaluates partial differential equations using finite volumes (Patra et al. 2005). The adaptative mesh characteristic of the equation solver implies that the size of the volume base are subjected to local refinement of the

mesh to minimize approximation errors (Patra et al. 2005). Parallelism is supported in the solution by distributing volumes among available processors (Pitman et al. 2003).

#### **4.3.2. Parameters Derived from Elevation**

The adaptative scheme used for the numerical solution requires elevation and its derivatives to be given at different resolutions for the different times of the simulation. TITAN2D obtains local values from DEMs stored in the data structure of an open source GIS through a custom developed Application Program Interface (API) (Patra et al. 2005). The selected open source GIS is the Geographic Resources Analysis Support System (GRASS) (Mitasova et al. 1995). An API provides an interface to a computational system to allow services and data to be exchanged with another computational system.

The API developed for GRASS allows TITAN2D to access GRASS data structure for DEMs. DEMs in GRASS are stored in a binary file complemented by a header file. The header file is a text file with information about georeferencing, extent, and grid cell resolution. Subsequently, the API provides TITAN2D with elevation and its derivatives of elevation at various resolutions.

#### **4.3.3. Variable Resolution**

The solution employed in the current implementation to provide variable resolution is to use a fixed resolution representation and to use interpolation and aggregation algorithms to compute the different resolutions (Pitman et al. 2003). The

methods used in the API for interpolation and aggregation are bilinear interpolation and the nearest neighbor aggregation method.

For a rectangular regular grid, the bilinear interpolation method estimates the value at a point as the distance-weighted average of values at the closest four grid cell centers. Bilinear interpolation performance is just slightly inferior when compared to more computationally expensive methods such as bicubic interpolation and kriging (Rees 2000).

In addition, the nearest neighbor interpolation method considers the value at one point as the same of the closest grid cell center; therefore the nearest neighbor interpolation method is not appropriate when a smooth surface is required as in a TITAN2D simulation. Therefore, the best method for interpolation to be used for TITAN2D is bilinear interpolation.

The nearest neighbor method for a rectangular regular grid estimates the value at a location by assuming that the value is the same as the value of the grid cell that is closest to the required location. The method was selected given that it is the simplest aggregation method.

#### **4.3.4. Estimation of Derivatives of Elevation**

The derivatives of elevation used to define the local direction of gravity and the radius of curvature of the local basal surface are estimated from a rectangular regular grid of elevation using a third-order finite difference method. The method performs better, or with minimal difference, when compared to more computationally expensive methods

(Skidmore 1989; Jones 1998). Horn (1981) developed the method by using weights on a previous implementation by Sharpnack and Akin (1969). First order partial derivatives are estimated using Horn's method and are given by (4.2) (Horn 1981):

$$\begin{aligned}\left[\frac{\partial z}{\partial x}\right]_{i,j} &= \frac{(z_{i+1,j+1} + 2z_{i+1,j} + z_{i+1,j-1}) - (z_{i-1,j+1} + 2z_{i-1,j} + z_{i-1,j-1})}{8\Delta X} \\ \left[\frac{\partial z}{\partial y}\right]_{i,j} &= \frac{(z_{i+1,j+1} + 2z_{i,j+1} + z_{i-1,j+1}) - (z_{i+1,j-1} + 2z_{i,j-1} + z_{i-1,j-1})}{8\Delta Y}\end{aligned}\quad (4.2)$$

where:

$[\delta z/\delta x]_{i,j}$  is the partial derivative in  $X$  direction at grid cell  $i,j$ ;

$[\delta z/\delta y]_{i,j}$  is the partial derivative in  $Y$  direction at grid cell  $i,j$ ;

$z_{*,*}$  are elevation values at cells in grid cell  $i,j$  neighborhood;

$\Delta X$  is the grid cell size (grid original resolution) in  $X$  direction; and

$\Delta Y$  is the grid cell size (grid original resolution) in  $Y$  direction.

Second order derivatives are estimated by the same method, replacing elevation by slope, and is presented by (4.3):

$$\begin{aligned}\left[\frac{\partial^2 z}{\partial x^2}\right]_{i,j} &= \frac{(s_{i+1,j+1} + 2s_{i+1,j} + s_{i+1,j-1}) - (s_{i-1,j+1} + 2s_{i-1,j} + s_{i-1,j-1})}{8\Delta X} \\ \left[\frac{\partial^2 z}{\partial y^2}\right]_{i,j} &= \frac{(s_{i+1,j+1} + 2s_{i,j+1} + s_{i-1,j+1}) - (s_{i+1,j-1} + 2s_{i,j-1} + s_{i-1,j-1})}{8\Delta Y}\end{aligned}\quad (4.3)$$

where:

$[\delta^2 z/\delta x^2]_{i,j}$  is the curvature in  $X$  direction at grid cell  $i,j$ ;

$[\delta^2 z/\delta y^2]_{i,j}$  is the curvature in  $Y$  direction at grid cell  $i,j$ ;

$\Delta X$  is the grid cell size (grid original resolution) in  $X$  direction; and

$\Delta Y$  is the grid cell size (grid original resolution) in  $Y$  direction.

$s_{*,*}$  are slope values at cells in grid cell  $i,j$  neighborhood, calculated by (4.4):

$$s = \sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2} \quad (4.4)$$

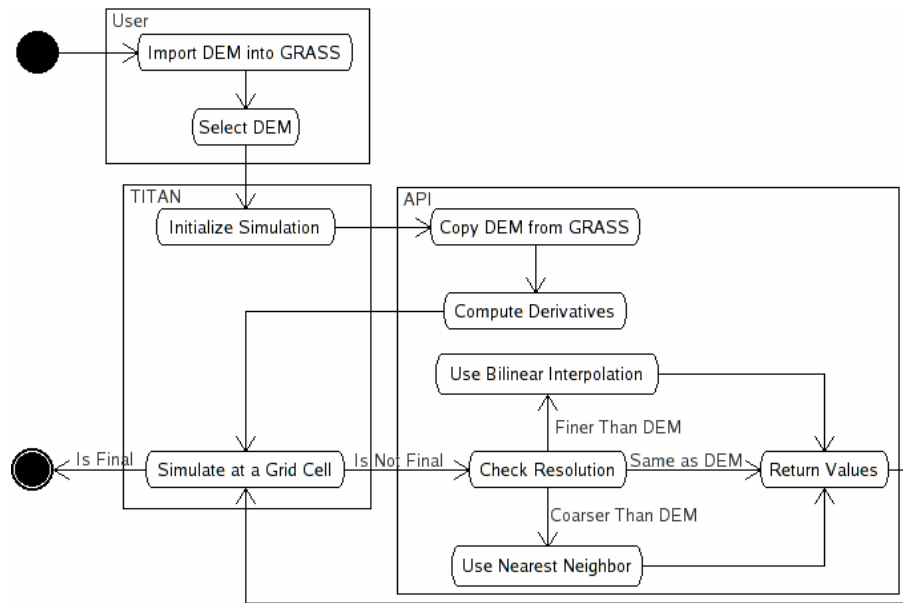
#### 4.4 A TITAN2D SIMULATION SESSION

A session for simulation of a geophysical mass flow using the current link between TITAN2D and DEMs stored in GRASS is presented in Figure 4.1 and has the following sequence:

1. The representation of elevation in a rectangular regular grid format is imported into GRASS. The process creates the binary and text files that implement the rectangular regular grid. The most common way to import is to obtain the grid in binary or text file formats and use GRASS importation tools;
2. The user selects the file with the representation of elevation to be used in the simulation and other simulation parameters;
3. A copy of the raster representation is created in the computer memory, since access of data from memory is faster than from the file system;
4. First and second order derivatives are computed at the same resolution of the representation of elevation using the finite difference method and made available in memory;
5. For every step of the simulation, at each cell of the adaptative grid, the required values of elevation and their derivatives are queried at a resolution defined by the cell size and:
  - a. Values queried at resolutions that are finer than the original resolution (resolution of the representation of elevation) are estimated using the bilinear interpolation method. The method

estimates the value of a position inside a rectangular cell using the values at the four corners of the cell.

- b. Values queried at resolutions that are coarser than the original resolution are estimated using a nearest neighbor method, where the value is estimated by assuming it to be the same as the value of the rectangular regular grid cell that is closest to the required location.



**Figure 4.1** Sequence for simulation of a geophysical mass flow using the current link between TITAN2D and DEMs stored in GRASS.

TITAN2D simulation parameters are set using a graphical interface presented in Figure 4.2 and are grouped according to data parameters, computational parameters, and flow pile parameters. A brief explanation of the parameters is given here, based on the TITAN2D user's manual (Geophysical Mass Flow Group 2004).

GIS Information Main Directory:

GIS Sub-Directory:

GIS Map Set:

GIS Map:

Use GIS Material Map? ☐ Yes

Simulation Directory Location:

Number of Processors:

Computational Mesh Points in Y-Direction:

Number of Piles:

Scale Simulation? ☐ Yes

If Scaled, Length Scale [m]:

Maximum Number of Time Steps:

Maximum Time [sec]:

Time [sec] between Results Output:

Adapt the Grid? ☐ Yes

Visualization Output:

First/Second Order Method: ☐ Second

Minimum x and y location:

Maximum x and y location:

Email Address:

(a)

Material 1 of 1: all materials

Internal Friction Angle (deg):

Bed Friction Angle (deg):

(b)

Information for Pile Number 1

Thickness of Initial Volume,  $h(x,y)$ :  $P^2(1 - ((x-xc)/xr)^2 - ((y-yc)/yr)^2)$

Maximum Initial Thickness,  $P$  (m):

Center of Initial Volume,  $xc, yc$  (UTM E, UTM N):

Major and Minor Extent,  $majorR, minorR$  (m, m):

Orientation (angle [degrees] from X axis to major axis):

(c)

**Figure 4.2** TITAN2D graphical user interface is composed of 3 individual windows for: (a) data, computational, and some flow pile parameters, (b) flow pile friction parameters and (c) additional flow pile parameters.

#### 4.4.1. Data Parameters

These parameters select the information from GIS and the region to be used in the simulation. Table 4.1 describes the data parameters.

**Table 4.1** Description of the data parameters required to run a simulation using TITAN2D.

Parameter	Description
GIS Information Main Directory	These parameters are used to select the representation of elevation to be used in the simulation.
GIS Sub-Directory	
GIS Map Set	
GIS Map	
Minimum X and Y Location	The limits of the region to be used in the simulation.
Minimum X and Y Location	

#### 4.4.2. Computational Parameters

These parameters set the computational environment to run the simulation. The computational parameters are described in Table 4.2.



**Table 4.2** Description of the computational parameters required to run a simulation using TITAN2D.

<b>Parameter</b>	<b>Description</b>
<b>Simulation Directory Location</b>	The directory where simulation results are stored.
<b>Number of Processors</b>	The number of processors to be used in a parallel computing environment.
<b>Computational Mesh Points in Y-Direction</b>	The initial number of cells of the numerical solution grid in Y-direction. The number of cells in X-direction is automatically calculated to maintain the shape of a cell square.
<b>Adapt the Grid</b>	If the adaptative grid is to be used or not.
<b>Scale Simulation</b>	If the governing equations are to be scaled or not. If scaling is used, equation values are scaled by pile height, gravity, and length scale. In addition, if scaled, the length scale must be defined. The scale length is the expected total run out length of the flow.
<b>First/Second Order Method</b>	If the first or second order solution method is to be used to calculate values of the computational grid cell. Second order solution assumes values to vary linearly across the cell and provides results that are more accurate than first order solution (consider values to be constant in a cell) at a higher computational cost.
<b>Maximum Number of Time Steps</b>	The number of time steps which, when this number is reached, the simulation is stopped. Since the length of time steps is variable, the maximum number provides a stopping criterion to avoid filling up computational storage space.
<b>Maximum Time</b>	The time in seconds which, when this time is reached, the simulation is stopped.
<b>Visualization Output</b>	The format of the results output used to visualize the flow.
<b>Time Between Results Outputs</b>	The time in seconds between outputs of results for visualization.

#### 4.4.3. Flow Pile Parameters

These parameters define the characteristics of the mass moving during simulation.

The flow pile parameters are described in Table 4.3.

**Table 4.3** Description of the flow pile parameters required to run a simulation using TITAN2D.

<b>Parameter</b>	<b>Description</b>
<b>Number of Piles</b>	Since more than one pile can be simulated at the same time, this feature can be used to simulate events that start at different locations. Their flows meet later or change the shape of the initial pile.
<b>Maximum Initial Thickness</b>	The initial pile shape is defined by a paraboloid calculated from the maximum thickness (in meters) and the size of the paraboloid major and minor axis (in meters).
<b>Major and Minor Extent</b>	
<b>Center of Initial Volume</b>	The location of the initial pile center.
<b>Orientation</b>	The orientation of the pile major axis, defined in degrees from X-direction to the major axis.
<b>Use of GIS Material Map</b>	If the spatially distributed basal friction angle is to be used.
<b>Internal Friction Angle</b>	The internal friction angle of the pile in degrees.
<b>Bed Friction Angle</b>	The basal friction angle in degrees. If local basal friction angle is needed the spatially distributed material map should be used. In this case, the local internal and basal friction angles are defined for each material of the material map.

## **4.5 SELECTING A GIS FUNCTIONS LIBRARY FOR TIGHT COUPLING**

Tight coupling of environmental process models and GIS require process models to be open, that is, process models can be modified to provide linkage to representations in a GIS. In addition, GIS should also be open to provide access to its representations.

Most commercially available GIS have an open interface to access representations. However, the access is restricted to the portions of the software selected

by the vendor. Furthermore, access is provided through interfaces that hide implementation, thus the algorithms are hidden and can not be audited. In addition proprietary GIS cost is high since the “market is an oligopoly in which two companies ... have a market share of 50%” (Camara and Fonseca 2006).

Although the cost of a proprietary GIS is not a problem for most researchers, simulation of environment processes must benefit everybody, especially populations of hazardous areas. Since most hazardous areas are located in less economically favored regions, local population are less likely to be able to afford the high costs of software. Therefore, the selection of which GIS to use in this study was restricted to open software GIS. Câmara and Onsrud (2004) present a discussion on the various existing open software GIS and how they are developed.

An open software GIS that facilitates applications to access representations stored in relational databases through library functions, TerraLib (available at [www.terralib.org](http://www.terralib.org)), was selected for the implementation of the linkage. TerraLib is being jointly developed by various groups headed by the Image Processing Division at INPE (Brazilian National Institute for Space Research) (Câmara et al. 2000). The advantages of using TerraLib are the availability of the library source code, the ability to create query functions using algorithms that are independent of data structures, and the use of database and GIS functions that have already been implemented.

TerraLib is built using C++ programming language since the language support to object-orientation and generic programming facilitates a collaborative development environment (Câmara et al. 2000). Object-oriented computer languages define classes

that are abstractions of data types with functions that manipulate the contents of the class (Stroustrup 2000). Generic programming relies on containers that contain classes and generic algorithms (algorithms that do not depend on the implementation details of the classes in the container) that manipulate containers (Stroustrup 2000).

#### **4.5.1. Representation of Elevation in TerraLib Spatial Database**

A database is a collection of data and a spatial database has a collection of representations of geographic objects. Spatial databases differ from traditional databases by representing entities that are far more complex than simple records of business applications (Egenhofer et al. 1999).

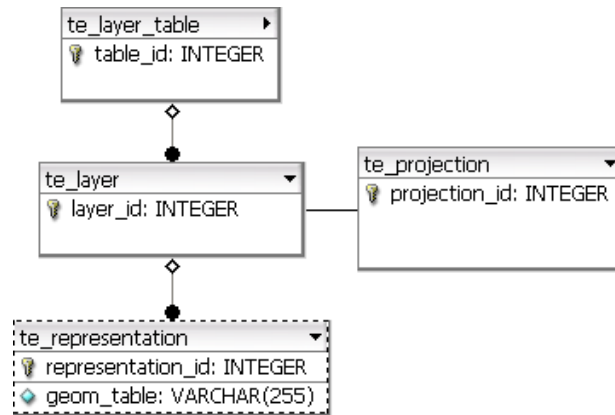
Relational databases are the most used type of a database. A relational database is collection of relations that are implemented as tables of values, where each table row corresponds to a representation of an entity of the real world and each column corresponds to an attribute (Elmasri and Navathe 2000).

The Standard Query Language (SQL) is the language to interact with the relational database, allowing definition, query and update of representations (Elmasri and Navathe 2000). A spatial database provides spatial query capabilities through a SQL modified to manipulate spatial entities by extending SQL through incorporation of spatial relationships and operations (Egenhofer 1994).

#### 4.5.1.1. Representations of Spatial Entities

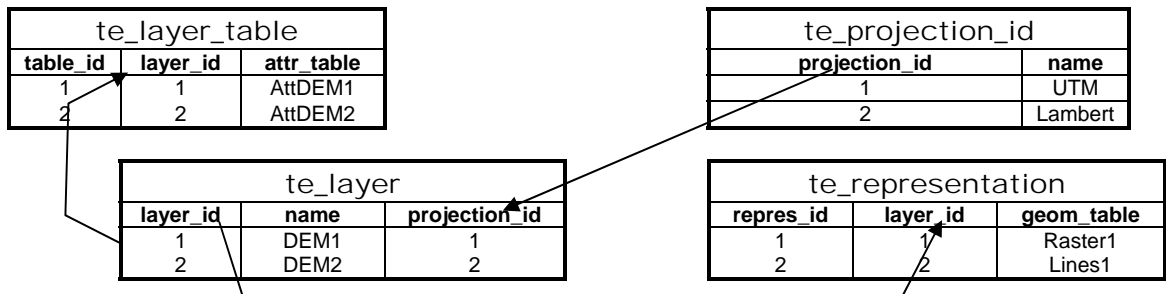
In TerraLib representations of spatial entities are stored in the tables of the database using either the capability of handling spatial representations of spatial database or a Binary Long Object (BLOB) type in a non-spatial database (Vinhas and Ferreira 2005). Database management systems that are supported by TerraLib include spatial and non-spatial, commercial and open software, relational databases.

The structure of a TerraLib database is stored in metadata tables, therefore, there are two types of tables in the database, metadata tables and data tables (Vinhas and Ferreira 2005). Representations are stored in data tables. In the structure of a TerraLib database representations are aggregated in layers if they share some common attribute and geographic space. A layer has a cartographic projection and a set of representations (Vinhas and Ferreira 2005). “**te\_layer\_table**”, “**te\_layer**”, “**te\_projection**”, and “**te\_representation**” tables are the metadata tables describing storage of representations and they are presented in Figure 4.3. Note that all TerraLib metadata tables use the “**te\_**” prefix.



**Figure 4.3** Simplified TerraLib database schema showing only names of columns that are relevant to the framework. Connections indicates that one “**te\_layer\_table**” table is linked to one or more “**te\_layer**” tables, one “**te\_layer**” table is linked to only one “**te\_projection**” table, and one “**te\_layer**” table is linked to one or more “**te\_representation**” tables.

Figure 4.3 shows that one representation of elevation is stored in the table identified by the column named “**geom\_table**” of the “**te\_representation**” table. The cartographic projection information of the representation is stored in the “**te\_projection**” table. All representations of a layer have the same cartographic projection. Since representations of elevation may have different projections due to choices at the conceptual level of modeling--differences in reference surface, measuring instrument, and raw data processing--each layer has only one representation of elevation. Figure 4.4 presents contents of metadata tables when the database stores two representations of elevation.












**Figure 4.4** Example of the contents of TerraLib metadata tables when two representations of elevation are stored. One representation is stored in “**Raster1**” table (“**Raster**” prefix indicates that format of the representation is a rectangular regular grid), its layer name is “**DEM1**”, has an attribute table “**AttDEM1**”, and its cartographic projection name is “**Lambert**”. Connections indicate the links between tables presented in Figure 4.2.





Representation of elevation can be vector or raster representations according to the choice made at the implementation level of the modeling procedure (Section 3.3.4). The selected type of representation is stored in tables identified by the column “**geom\_table**” of the “**te\_representation**” table.

#### 4.5.1.2. Vector Format

Contour line representation is a vector format representation and the representation is stored in a table with the schema for lines presented in Figure 4.5a. The sample points that complement contour lines are stored in another table, with the schema from Figure 4.5b.

Lines_(layer_id)	
	geom_id: INTEGER
	object_id: VARCHAR(255)
	num_coords: INTEGER
	lower_x: DECIMAL(24,15)
	lower_y: DECIMAL(24,15)
	upper_x: DECIMAL(24,15)
	upper_y: DECIMAL(24,15)
	ext_max: DECIMAL(24,15)
	spatial_data: BLOB

(a)

Points_(layer_id)	
	geom_id: INTEGER
	object_id: VARCHAR(255)
	x: DECIMAL(24,15)
	y: DECIMAL(24,15)

(b)

**Figure 4.5** TerraLib database schema for (a) lines table and (b) points table, when spatial geometries are not supported by the database. Elevation of contour lines and elevation of sample points are stored in another table (identified by the column “attr\_table” of “te\_layer\_table” table, see Figure 4.3) that links “object\_id” to an elevation value.

Since the basic geometric primitive of a Triangular Irregular Network (TIN) is a triangle, TINs are stored in a table using the schema defined for geometries of type polygon. The schema is similar to the schema presented for lines in Figure 4.5.

#### 4.5.1.3. Raster Format

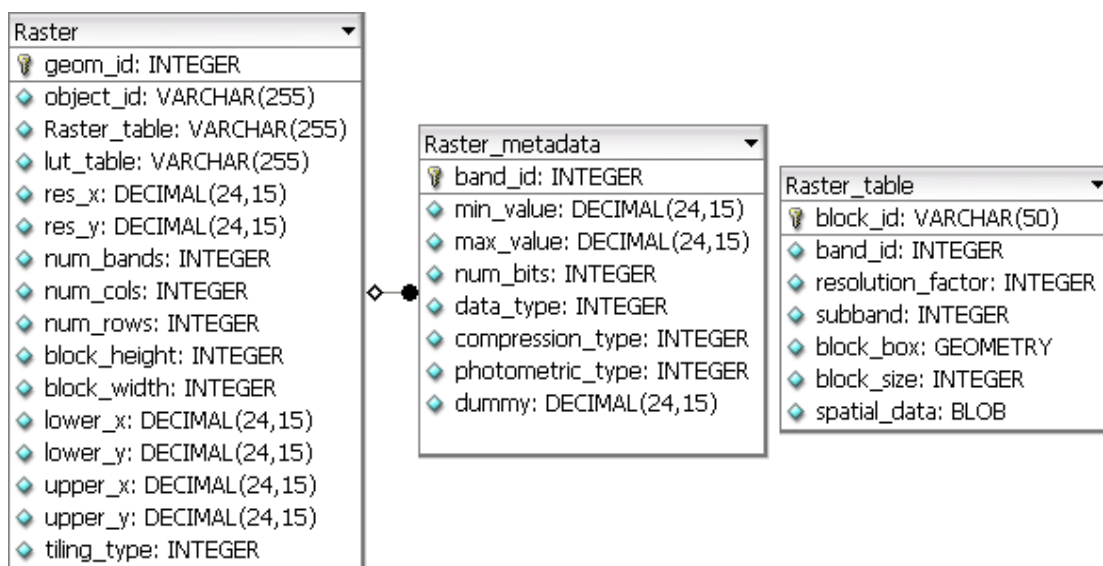
Raster representations of elevation are stored in three types of tables. Two of the tables store descriptions about the raster representation and the third stores the elevation values. Raster type tables store general information about elevation such as resolution and spatial extent. “**Raster\_metadata**” table stores information about minimum and maximum values in addition to the compression technique used to store values.

More than one “**Raster\_metadata**” table may exist for a raster representation. For example, the raster representation can be used to store a color image with red, green, and blue components instead of elevation. Elevation values are stored in the BLOB type



column “**spatial\_data**” of a “**Raster\_table**” table. More than one “**Raster\_table**” table may be used if the raster representation is divided in blocks.

The blocked storage is used to allow fast access to portions of the raster. A detailed description of raster geometries in TerraLib for handling imagery is presented by Vinhas et al. (2003). Figure 4.6 presents TerraLib schemas used to store elevation in a raster representation.



**Figure 4.6** TerraLib database schema for storage of raster representation. More than one “**Raster\_metadata**” table exists if the raster has more than one value for a location, such as in a color image. Values may be stored in more than one “**Raster\_table**” table if the block is defined to be smaller than the raster geometry. Connection indicates that one “**Raster**” table is linked to one or more “**Raster\_metadata**” tables.

#### 4.5.2. Representation Information

The previous section covered storage of elevation representations in a TerraLib database. The next step is to query the representations to obtain elevation values at locations where environmental processes occur for simulations. Representation of

elevation in TerraLib tables described in Section 4.5.1 and existing functions of the TerraLib library would be enough to provide elevation values and provides the same functionalities that exist in other GIS.

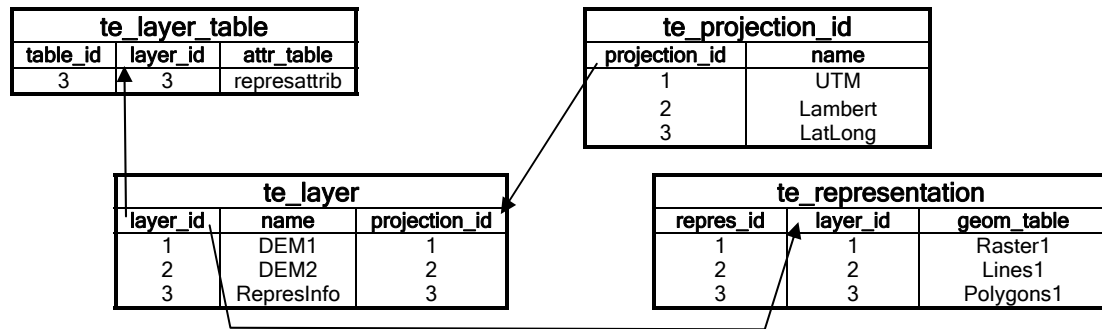
However, since each representation has its own cartographic projection, a transformation of projection would be required at each request. Also, information about the choices made during elevation modeling is not available for the queries. A solution using a representation that stores elevation representation extent and modeling choices is proposed here to solve these problems.

The proposed solution takes advantage of TerraLib query functions with spatial restrictions. A new representation using polygonal geometries that correspond to the spatial extent of each elevation representation is created and stored in a new layer, the Representation Information (“**RepresInfo**”) layer.

Since a layer has one cartographic projection only, “**RepresInfo**” layer must have one projection that is suitable for the whole Earth. The projection using latitude and longitude values is the most convenient and it is identified here by “**LatLong**” name. The datum of this projection must be the one that best fits the whole Earth; therefore, the WGS-84 geoid was selected.

For each representation of elevation a polygonal representation of the regions where it has valid values is created. The polygons are stored in the “**Polygons1**” table as indicated by the column “**geom\_table**” of the “**te\_representation**” table.

External tables may be linked to TerraLib database tables. A table that contains detailed information about the choices made during the modeling of elevation is created and linked to the polygonal geometry through the “**RepresInfo**” layer. This table (“**represattrib**” table) with attributes of each elevation representation is linked to the “**RepresInfo**” layer by the column “**attr\_table**” of the “**te\_layer\_table**” table. Figure 4.7 presents an example with the contents of metadata tables when two elevation representations are stored and their information is also stored.



**Figure 4.7** Example of the contents of TerraLib metadata tables when two elevation representations are stored and their information is stored in “**RepresInfo**” layer. Information about the representations of elevation is divided in geometric (stored in “**Polygons1**” table), and attributes (stored in “**represattrib**” table). Connections indicate the links between tables presented in Figure 4.2.

When elevation at one location is queried, a spatial query is executed in the “**RepresInfo**” layer to search for the polygons stored in “**Polygons1**” table that contains the location. Additional information about the representations of elevation, i.e. their metadata, is available in the “**represattrib**” table, thus the query may be refined to include choices made during elevation modeling steps. In the next section, metadata used in this framework is described.

### 4.5.3. Representation Metadata

Metadata attached to a DEM should describe all the choices made during modeling of the elevation. Existence of metadata associated with a representation of elevation is crucial for proper use of the representation. For example, if the metadata does not provide the information about when elevation was measured, and the application assumes that the elevation corresponds to values at another time and date, the difference in value between expected by application and provided by the representation may be in the range of tens of meters, as seen in the example of change in elevation after a volcanic eruption.

A standard metadata, such as the Content Standard for Digital Geospatial Metadata, published by the United States Federal Geographic Data Committee, usually includes information about the representation producer, source, geographic reference, and availability (Goodchild and Longley 1999; Guptill 1999). However, adoption of standard metadata has been slow due to the high workload of documenting existing representations. Thus there are few DEMs with a complete metadata. Due to this constraint, for the proposed framework, only information that is useful for selecting representation of elevation is included in the metadata.

Storage of metadata about the representations of elevation is executed using the “**represattrib**” table (as described in the previous section). The attributes of the “**represattrib**” table are presented in Figure 4.8.

represattrib	
🔑	repres_id: VARCHAR
💎	description: VARCHAR
💎	resx: DOUBLE
💎	resy: DOUBLE
💎	measure_datetime: DATETIME
💎	production_datetime: DATETIME
💎	production_method: VARCHAR
💎	production_source: VARCHAR
💎	measure_unit: VARCHAR
💎	measure_name: VARCHAR
💎	horizontal_accuracy: DOUBLE
💎	vertical_accuracy: DOUBLE
💎	source_scale: DOUBLE

**Figure 4.8** Contents of “**represattrib**” table that are used to describe metadata about representations of elevation. The column “**repres\_id**” indicates the representation of elevation name. Note that the term column is used for consistency with the definition of columns in relational databases, where column refers to the names of attributes of a representation (see Section 4.5.1).

Each of the attributes describing metadata (presented in Figure 4.8) is briefly described next.

#### 4.5.3.1. Representation Description

The attributes used to describe the representation are the database columns (see Section 4.5.1 for the definition of database column) named “**repres\_id**”, “**description**”, “**measure\_name**”, and “**measure\_unit**”. These attributes are described in Table 4.4.

**Table 4.4** Description of the attributes used to provide description of a representation.

Column Name	Description
<b>repres_id</b>	The “ <b>repres_id</b> ” column identifies to which representation of elevation the metadata is about. The column contains the name of the layer linked to the representation of elevation.
<b>description</b>	The “ <b>description</b> ” column is used to provide a brief report about the representation. The contents are not used in queries and are intended for human interpretation only. Therefore, there are no requirements on how to fill it in.
<b>measure_name</b>	Given that the representations stored may include information derived from elevation, such as slope and curvatures, this column is used to identify the type of information. The valid contents used in this framework were defined by use and they are: “Elevation”, “Slope”, and “Curvature”. Other names can be added depending on the usage of the framework.
<b>measure_unit</b>	This column is used to identify the measure unit. For this framework, the valid names are: “Meters” and “Feet”. Other names can be added if a new application requires it.

#### 4.5.3.2. Resolution

These attributes describe the resolution of the representation in X and Y directions. They are meaningful only if a rectangular regular grid is being used to store the representation of elevation. The attributes are the database columns “**resx**”, for resolution in X direction, and “**resy**”, for resolution in Y direction. Units of resolution values are the same as the “**RepresInfo**” layer projection, i.e. they are in decimal degrees.

#### 4.5.3.3. Acquisition Date and Time

The “**measure\_datetime**” attribute is used to store the date and time when the elevation was measured. In cases where the representation is a derivative measure (such as slope and curvature), the database column contents must be the information of the

original source, given that the use of this column is to keep track of changes in the original date and time due to anthropogenic or geomorphic processes.

#### 4.5.3.4. Representation Production

These attributes are used to describe how measures were transformed to produce the representation. They are the database columns “**production\_datetime**”, “**production\_method**”, “**production\_source**”, and “**source\_scale**”. The attributes are described in Table 4.5.

**Table 4.4** Description of the attributes used to provide description about the production of a representation.

Column Name	Description
<b>production_datetime</b>	Identifies when the representation was created. Content of the column is used to keep track of different implementations since better techniques of modeling may become available and existing representations may be improved.
<b>production_method</b>	Describes the techniques or algorithms used to create the representation. Since the number of techniques and algorithms is small, the contents of the column can be used to infer certain characteristics of the representation.
<b>production_source</b>	Describes the measurement instrument. Typical values are: “ <b>Photogrammetry</b> ”, “ <b>Contour Lines</b> ”, “ <b>LIDAR</b> ”, “ <b>Synthetic Radar</b> ”, and “ <b>Interferometry</b> ”. When the representation is derived from another existing one, the content is the name of the source representation.
<b>source_scale</b>	Indicates the scale of the original source of information. For example, for contour lines from a topographic map, the content is the map scale. Stored value is assumed to be the divider of the fraction, for example, for a 1:24,000 scale, the stored value is 24,000.

These database columns provide the minimum description about production. Additional information, such as the particular equipment and its calibration parameters, can be added to the “**represattrib**” table if they are required by another application.

#### **4.5.3.5. Representation Accuracy**

These attributes describe the declared accuracy for the representation. They are the “**horizontal\_accuracy**” and “**vertical\_accuracy**” database columns. The values are typically taken from the representation specification. For example, the elevation from Shuttle Radar Topography Mission (SRTM) has a vertical accuracy requirement of 16 meters at the 90 % confidence level (van Zyl 2001). The values of the database columns should be RMSE values at the 90 % confidence level preferably, with units being the same as the “**measure\_unit**” database column.

#### **4.5.4. Representation Query**

Representations stored in a TerraLib data base can be queried through using a standard SQL query or a spatial SQL query. For example, a standard query to retrieve the name of the layers with vertical accuracy better than 10 meters would be:

```
SELECT repres_id FROM represattrib WHERE vertical_resolution < 10 AND  
measure_unit = Meters
```

The representation satisfying the query can be accessed through the layer named “**repres\_id**”.



To add a spatial constraint to the query, a spatial query is required. For example, the same query presented in the previous example, with an added constraint that it must contain a location given by coordinates 150, 260 would be:

```
SELECT repres_id FROM represattrib WHERE vertical_resolution < 10 AND  
measure_unit = Meters AND contains(Point(150,260))
```

TerraLib has a query processor to handle queries to the database (Ferreira et al. 2003), thus providing a convenient integration of the spatial query to the application. In the proposed framework, the query processor is used to select the appropriate elevation representation based on the available metadata. Details of the implementation are given in Section 4.6.2.

## **4.6 IMPLEMENTING THE NEW LINK TO A PROCESS MODEL**

Using the current implementation of TITAN2D linked to GIS, the results of simulations using TITAN2D are dependent on the selection of the base representation of elevation.

However, the selected representation of elevation may not be the best one. For example, if the representation resolution is too coarse, a simulated flow may bypass a ridge that is too thin to be represented at that resolution. On the other hand, if the resolution is too fine, the computational resources will not be optimally used and the time to compute may be too long. For instance, for an area of 10 by 10 kilometers, a rectangular regular grid with 1 meter resolution will require processing of up to 100 million cells at every step of the simulation while a 100 meter resolution would require processing of only 10 thousand cells.

In addition, the date and time of the measurement of the selected representation of elevation may be from after the simulated event, and the surface may have changed drastically due to the event or posterior processes.

The following sections describe how the framework handles the adaptative grid, and the date and time of the simulated event. In addition details of the implementation are presented.

#### **4.6.1. Link Requirements**

The requirements for the link are to handle the TITAN2D adaptative grid and support simulations where date and time of the real event are specified.

##### **4.6.1.1. Adaptative Grid**

Multiple representations of elevation provide a better solution for the adaptative grid of the numerical solution of the governing equations. A coarse resolution rectangular regular grid may provide values for the whole area and finer resolution grids may be accessed to provide values where the flow goes through small valleys or where ridges control the direction of the flow.

The proposed framework handles adaptative mesh by selecting the representation with the resolution that is closest to that required by TITAN2D and applying interpolation or aggregation algorithms to provide the required resolution. The interpolation method used is the bilinear interpolation (the same as used in the current implementation). For aggregation, a method that estimates value by averaging all grid

cells covered by a cell with the queried resolution is used. This aggregation method is used since it gives a surface that is smoother than the nearest neighbor used in the current implementation.

#### **4.6.1.2. Temporal Query**

Simulations of geophysical mass flows are used either for prediction or for studying characteristics of particular cases. In any of the cases, the representation of elevation that was created from measurement that occurred close in time to the flow case should be preferred to provide values for the TITAN2D simulation.

The framework uses metadata about the representation of elevation to select the most appropriate representation for the simulation of a particular case. Current TITAN2D does not offer the possibility to specify when the simulation case happens. Thus the interface was modified to include the information of when the real world case started and when it ended.

#### **4.6.2. Implementation Details**

The framework proposed here is general. However, TITAN2D is computationally intensive and the framework needs optimizations to allow simulations to be executed without lasting longer than the current implementation. A cache of queries and prediction of queries are implemented for the required optimization.

#### **4.6.2.1. Cache of Queries**

The objective of the queries cache is to avoid making a query to the TerraLib database for every location where a value is required by TITAN2D. The implementation includes a cache of previous queries. In the cache, the parameters of the query and the values that satisfied the query are stored. When a new query is made, its parameters are checked for a match with previous queries in the cache. If there is a match, the value is obtained from the query and not from TerraLib database, thus improving the performance of the implementation.

#### **4.6.2.2. Prediction of Queries**

TITAN2D adaptative mesh requires values at various resolutions. However, the resolutions are multiples of a base resolution, thus it is very likely that if a query is made at a given resolution at a location, another query at the same resolution will be made for a different location. The implementation avoids making a new query to the TerraLib database by storing all the values of a representation and not only the value for a location.

#### **4.6.2.3. Creation of Derived Representations**

TITAN2D requires values of slope and curvature at various resolutions. Representations of slope and curvature are created when the query can not be satisfied by the cache or by the TerraLib database. New representations are created from an elevation representation with the same resolution as TITAN2D.

Slope and curvature are estimated using a third-order finite difference method presented in Section 4.3.4. The new representations of slope and curvature are stored in TerraLib database with the metadata describing the estimation method and the source representation.

### **4.6.3. Ranking Representations**

When the query specifies only resolution, the best representation for the query is the one with the closest resolution to query resolution. When the query specifies only the time of the case to be simulated, the answer to the query is not straightforward. In addition, if the query involves resolution and time, the answer is even more complicated. Thus a system that ranks representations according to how well they satisfy the query is necessary.

If there are only representations created from measures made before the required time, then the best representation is the one closest in time to the query time. When there are representations with time of measurement before and after the query time, the best representation can not be easily defined.

There are no rules to define the best representation, thus the process model must specify a ranking based on its own requirements. However, there are no cases when process models were tested for sensitivity to time of acquisition. The framework allows the process model to specify the ranking system; however, a default system is provided.

The default system ranks higher representations acquired before the time constraint of the query, with ranking scores getting lower as the time difference becomes

larger. When the acquisition time is just after the query time, the rank is lower than the rank for just before. In addition, the ranking scores of representations acquired after the query time decrease faster as the difference gets larger when compared to representations acquired before.

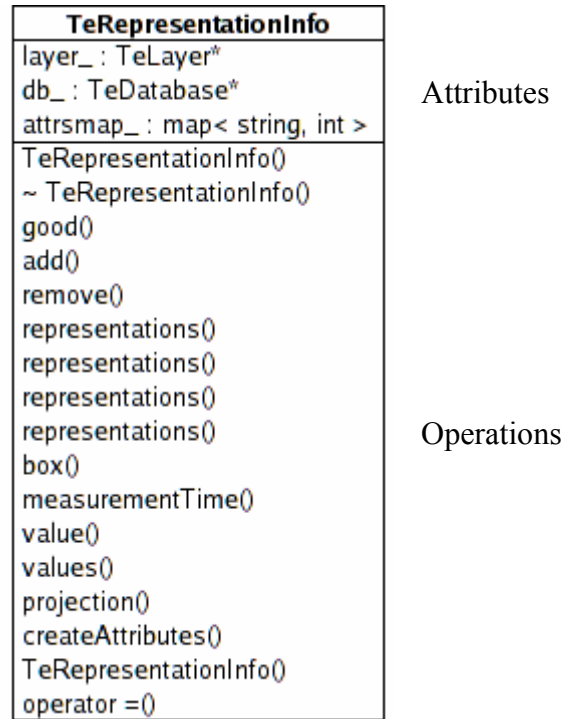
When resolution and time of the case are constraints of the query, the solution proposed in the framework is to add the ranking scores given by resolution and time.

For other possible constraints, such as including vertical and horizontal accuracies, the process models must specify a ranking system. The framework provides a way of using the ranks. Thus it is flexible and does not impose one ranking system that may not be adequate for all uses of multiple representations.

#### **4.6.4. C++ Classes of the Implementation**

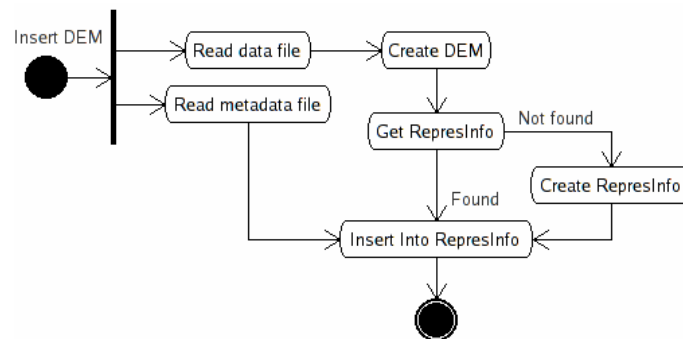
The C++ computer language classes that implement the linkage are the “**TeRepresentationInfo**”, “**TeModelInterface**”, “**TeGlobalModelInterface**”, “**TeModelQuery**”, and “**TeModelQueries**” classes.

The “**TeRepresentationInfo**” C++ class facilitates reading and writing of representations and their information into TerraLib database tables and it is presented in Figure 4.9.



**Figure 4.9** Internal structure of the “**TeRepresentationInfo**” class that handles representations of elevation and their information in TerraLib database tables.

The “**TeRepresentationInfo**” class is used in the program that inserts a representation in TerraLib database. The structure of this program is presented in Figure 4.10.



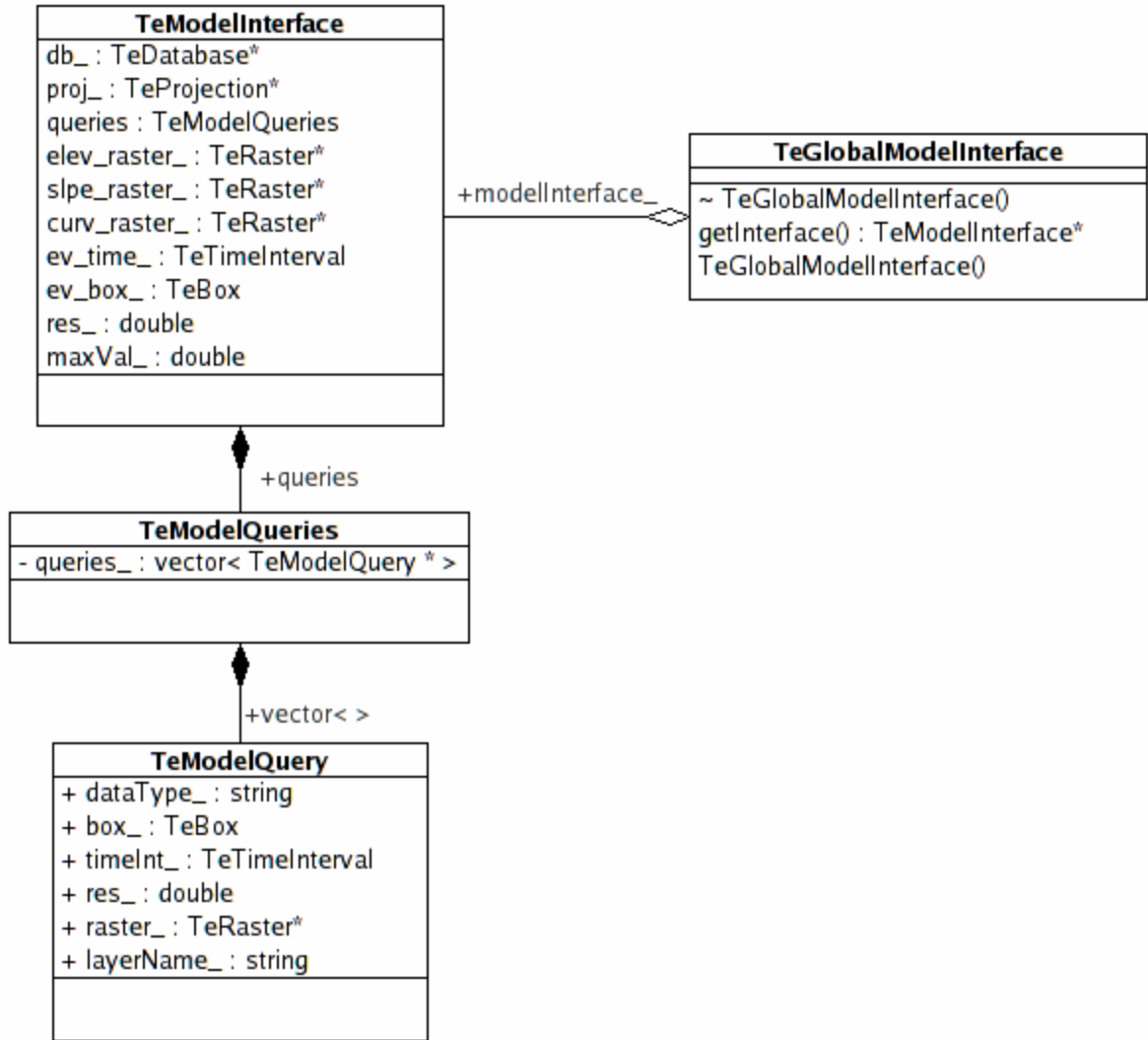
**Figure 4.10** State diagram of the program that inserts a representation of elevation into a TerraLib database. The “**Create RepresInfo**” and “**Insert Into RepresInfo**” states use “**TeRepresentationInfo**” class.

The “**TeModelInterface**”, “**TeGlobalModelInterface**”, “**TeModelQuery**”, and “**TeModelQueries**” C++ classes provide the optimizations required by process models. These classes are presented in Figure 4.11.

“**TeModelInterface**” implements the interface to process models while “**TeGlobalModelInterface**” provides the unique global access to the interface. A unique global access assures that the interface to process models is not duplicated.

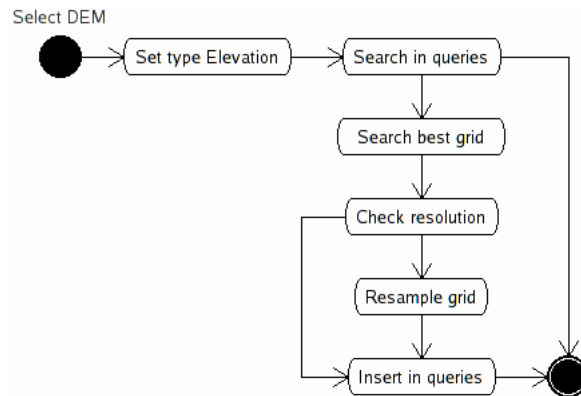
“**TeModelQueries**” class implements a cache of queries where each cache entry is implemented by “**TeModelQuery**” class.





**Figure 4.11** Internal structures of the “**TeModelInterface**”, “**TeGlobalModelInterface**”, “**TeModelQuery**”, and “**TeModelQueries**” classes that provide interface for process models and their relations.

The classes are when a process models queries for a representation in TerraLib database. The structure of the part of a program that is querying for a representation is presented in Figure 4.12.



**Figure 4.12** State diagram of the operations of the TeModelInterface class when a representation of elevation is requested by a process model. If a derivative of elevation is required, the proper type is set and, if not available, a new representation is computed from elevation (when first derivatives are required) or from the first derivatives of elevation (when second derivatives are required).

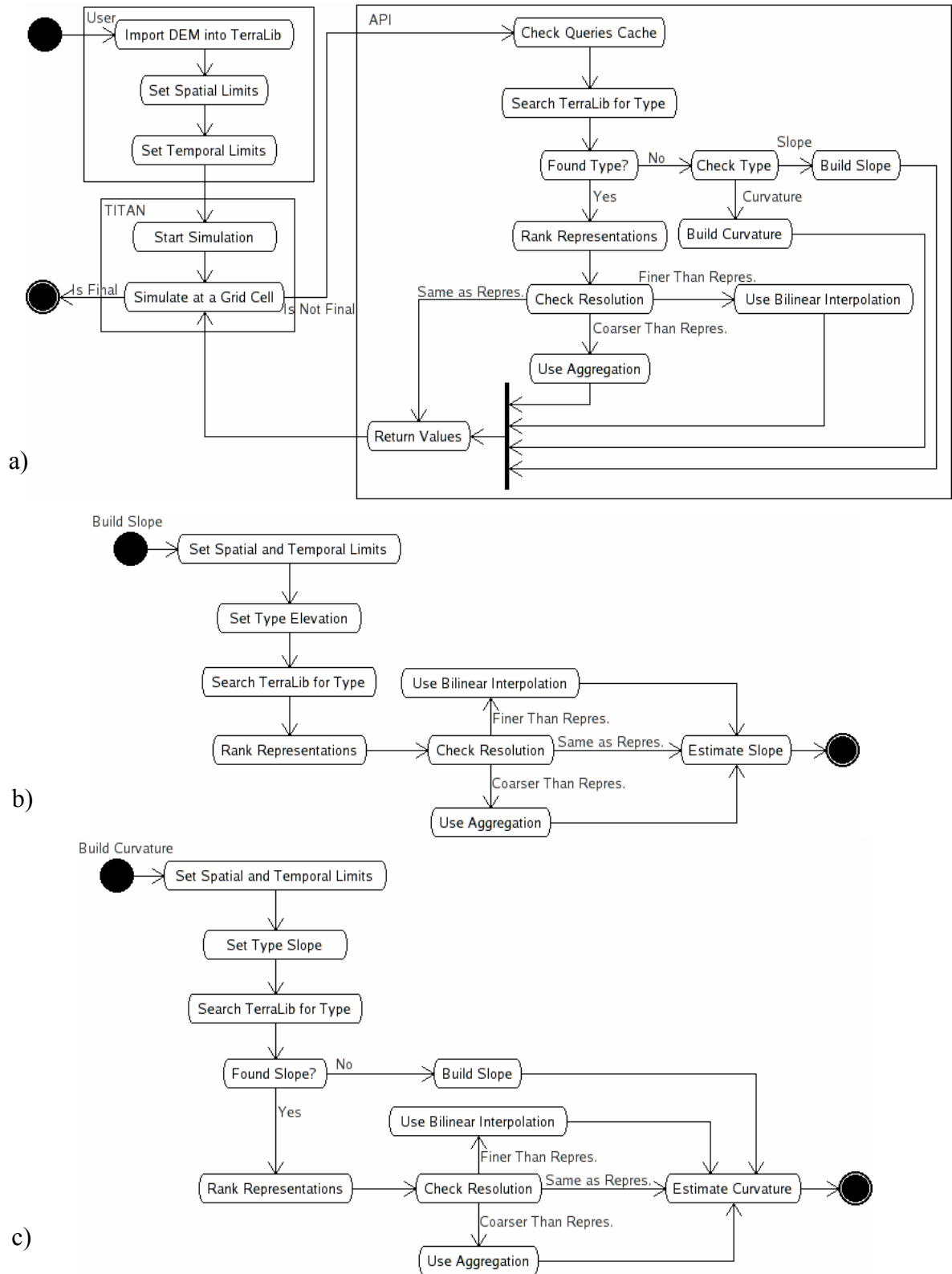
## 4.7 A SIMULATION SESSION USING THE NEW IMPLEMENTATION

The new implementation of TITAN2D obtains values of elevation and its partial derivatives from raster representations of elevation stored in the TerraLib database. The sequence for a session for simulation of a geophysical mass flow is presented in Figure 4.13 and is described by:

1. The representations of elevation, in a rectangular grid format are imported into the TerraLib database;
2. The user defines the temporal and spatial limits of the simulated geophysical mass flow, that is, when and where the simulated event happened. The temporal limits are given by the initial and the final date and time of the event. The spatial limits require definition of the projection used to define coordinates of the limits. Thus the user first defines projection and then the coordinates of event spatial limits;

3. For every step of the simulation, at each cell of the adaptative grid, required values of elevation and their derivatives are queried at a resolution defined by the cell size and constrained by the date and time of the simulated event, and:
  - a. If the query exists in the queries cache, the representation stored in the cache is used to provide the requested value;
  - b. If the queries cache does not contain the query, search the database for all representations that contain a valid value of the requested type, i.e. elevation, slope or curvature. For each of these representations, a value is assigned based on the ranking system presented in Section 4.6.3;
  - c. If the highest ranked representation has resolution that is equal to the query resolution, the representation is used to provide the requested value. The query and the representation are stored in queries cache;
  - d. If the highest ranked representation resolution is different to the resolution of the query, a new representation with resolution equal to the queried one is created, used to provide the requested value, and stored in the cache. If representation resolution is coarser than the queried one bilinear interpolation method is used to create the new representation. Otherwise aggregation method is used to generate the new representation. In addition the new representation is stored in the TerraLib database, with representation information created using the highest ranked metadata information modified to describe the change in resolution;

- e. If the queried type is slope and there is no representation that satisfies the query, the query is changed to search for elevation and processed. The representation of elevation that satisfies the new query is used to create a representation of slope using the finite difference method (see section 4.4.4.3). The new representation of slope is used to provide the requested value, and stored in the cache. Representation information is created based on the representation of elevation, changed to describe the slope generation method, and stored with the new representation in TerraLib database.
- f. If the queried type is curvature and there is no representation that satisfies the query, the same procedure applied to slope is used, using representation of slope instead of elevation.



**Figure 4.13** Main sequence for simulation of a geophysical mass flow using the new link between TITAN2D and the multiple representations of elevation (a). Also the subsequences for building slope (b) and curvature(c).

## 4.8 SUMMARY

A framework to link dynamic spatially explicit environmental models was presented in this chapter, with details of the implementation of the framework to link TITAN2D, the process model used as an example of process model, and a GIS that was built using TerraLib.

The framework supports storage of representations in a relational database. Representations are queried by the process model, using spatial and metadata constraints, to provide values that best match the requirement to solve its differential equations.

The queries are supported by the capabilities of TerraLib to handle spatial SQL queries, thus allowing queries to include location and metadata information. Representations in the framework are stored in database tables. In addition, a set of tables with metadata and spatial extent of each representation is created to support spatial SQL queries.

The framework supports the concept of generating new representations only when an application requires it. Thus in the case implemented in this framework, representations with different resolutions are created when a query needs it. In addition, dynamic spatially explicit environmental models use differential equations in space and the solution for these equations require slope and curvature values. New representations of slope and curvature, derived from elevation, are created when queried using a finite element estimation method.

Due to the requirements that TITAN2D be computationally efficient, given the heavy load of numerical computations, optimizations were implemented. The optimizations are achieved using a cache of previous queries and the values satisfying those queries. In addition, values of all locations for a selected representation are stored in the cache.

The problem that arises when querying multiple representations is how to rank when more than one representation satisfies the query. The solution proposed in the framework is to let the process model specify the ranking system. In the implementation of the framework for TITAN2D, the default ranking for resolution and time of measurement ranks higher for resolution matching the query and decreasing as the difference increases and ranks higher for representations derived from measurements taken just before the query time. The time rank decreases as the time difference increases with a higher decrease if the measurement took place after the time required in the query.

In the next chapter, additional methods required to improve the use of multiple representations are presented. They include the definition of a spatially distributed measure of accuracy based on the existence of multiple representations and a methodology to measure performance of simulation results to allow comparison of the effects of process models and representations.





## **Chapter 5**

# **ADDITIONAL METHODS FOR INTEGRATION OF MULTIPLE REPRESENTATIONS OF ELEVATION AND PROCESS MODELS**

In this chapter, additional methods that were developed and can be used to improve the use of multiple representations are presented. They include definition of a spatially distributed measure of accuracy based on the existence of multiple representations and a methodology to measure performance of simulation results.

The spatially distributed accuracy, when stored in a database and linked to a representation of elevation, should be queried by an application to constrain the search for a representation of elevation by accuracy, for specific locations. For example, some applications may require higher accuracy at the bottom of the valleys while others require higher accuracy along ridge lines. The accuracy in the method presented in Section 5.1 is calculated using clusters where inaccuracies are higher. An accuracy map is created in Section 5.1.5 using the proposed method as an example since the method is not used in the next chapter. Section 5.1 presents the measures of accuracy currently used, where accuracy is given by a global value, a value that is constant for the whole where the representation of elevation has valid values.

Measurement of simulation performance is required to allow comparison of the effects on process models of multiple representations of elevation. Most existing

comparison methods for simulation of dynamic spatially distributed processes are qualitative only, where the modelers describe the performance as good, since they feel that the simulation results compared on a map are similar to the real event. Since the comparison is intuitive and biased, there is a need for a quantitative comparison method, such as the method presented in Section 5.2. The proposed method is used in the next chapter to verify if the use of multiple representations of elevation improves results from simulation of geophysical mass flows.

## **5.1 COMPUTING SPATIALLY DISTRIBUTED ACCURACY**

The existence of multiple representations of elevation allows quantification of accuracy at each location. The basis is to consider elevation at a point as a random variable and each representation can provide a value that can be used to define the distribution of values at that location. Therefore, for a given representation, for the elevation at one point, a confidence value is also attached.

The confidence value can be stored in the TerraLib database, linked to the elevation representation. Subsequently, any application that requires the probability of the value at a location being correct can query and either use the value or decide if the confidence is good enough for the application. Note that previous information about characteristics of the representation, such as stated accuracy is not used here.

The method does not attempt to find errors in the representation. Errors in a representation of elevation are searched through methods based on spatial autocorrelation. Errors are assumed to be random and any deviation from the

autocorrelation model is supposed to be an error. For example, there are techniques using principal components analysis (Lopez 1997), semivariogram and fractal analysis (Brown and Bara 1994), geostatistics (Fisher 1998; Kyriakidis et al. 1999), and wavelets (Falorni et al. 2005). Note that none of the methods cited above can find systematic errors that change all elevation values by a constant value. Although these systematic errors may not change results of a simulation that considers relative elevations only, the same errors may have a major impact in uses of elevation where absolute values of elevation are important, such as in aircraft navigation applications

### 5.1.1. Global Accuracy

Global accuracy of existing representations of elevation is given either by Root Mean Square Error (RMSE) at randomly selected locations or by vertical accuracy at 95% confidence interval (Daniel and Tennant 2001). They are calculated taking into consideration that there are “true values” of elevation at test locations. Note that horizontal accuracy is very difficult to be evaluated in regular grid representations of elevation, since it requires test locations to be recognizable in one of the cells of the grid (Daniel and Tennant 2001). Thus, the vertical accuracy is the one that characterizes a representation of elevation.

RMSE is calculated by (5.1) (USGS 1998):

$$RMSE = \sqrt{\frac{\sum_1^n (h_r - h_t)^2}{n}} \quad (5.1)$$

where:

$h_r$  is the representation value;

$h_t$  is the “true value”; and

$n$  is the number of test locations.

The United States Geological Survey (USGS) uses RMSE accuracy for its DEM for the whole United States and the 30- by 30-meter resolution. For example, a Level-1 DEM has the target for RMSE accuracy set to be better than 15 meters (USGS 1998).

RMSE is supposed to be representative for the whole area covered by the DEM. The USGS DEMs are required to use at least 28 locations to define the DEM RMSE, with 8 of the locations at the edges of the DEM (USGS 1998).

Vertical accuracy at 95% confidence interval is the preferred accuracy measure for the National Standard for Spatial Data Accuracy (NSSDA) calculated from RMSE and, for normally distributed errors in elevation, the accuracy measure is the RMSE value multiplied by 1.96 (95% confidence interval for normal distribution), calculated for at least 20 locations (FGDC 1998).

Although neither USGS nor FGDC state it, the number of locations is most likely defined from the sample size for confidence on the mean. The confidence is related to the number of samples are required to ensure that the confidence that the RMSE is within a maximum allowed error and is a pre-defined value.

The sample size for mean is defined by the square of the ratio between the product of confidence and standard deviation and the maximum expected error (McGrew and Monroe 2000). If the confidence is 1.96 (95% in a normal distribution), the ratio between standard deviation and the maximum error is 10.2 for a sample size of 20 (28

locations minus 8 locations on edges for USGS DEM). Therefore, if the standard deviation of the RMSE is 10 meters, the maximum error for 95% confidence is 1 meter.

The main problem with using global accuracy values is that it assumes that the differences between the representation values and the “true values” are randomly distributed. Due to the characteristics of the techniques used to generate the representation, there are areas where differences are larger than in others. Vertical accuracy depends on land cover type for elevation acquired with LIDAR (Hodgson et al. 2005) and with relief and slope for SRTM DEMs (Falorni et al. 2005). Therefore, a spatially distributed accuracy is desirable.

### **5.1.2. Computing Spatially Distributed Accuracy**

Although the distribution of errors in representation of elevation is known to be not random and correlated with relief and slope of the surface (Fisher 1998), attempts to quantify the spatial distribution of accuracy from the correlations has been validated only for particular locations and datasets. For example, the Carlisle’s (2005) study areas were in Snowdonia, North Wales, United Kingdom and Mestersvig, northeast Greenland; and Gao’s (1997) study areas were in Virginia, USA.

The existence of multiple representations of elevation allows the approach to create an accuracy surface, the spatial distribution of accuracy. The approach proposed here is based on defining the confidence of a representation and on spatial clustering. Since errors in representation of elevation are spatially autocorrelated (Lopez 2000), the

method proposed here aims to create areas with uniform accuracy generated from a classification of clusters that have statistically significant high values of error.

#### **5.1.2.1. Descriptive Statistics of Elevation**

Descriptive statistical analysis of elevation assumes that elevation at one location is not a true value, but a random variable. From observations at that particular location, a set of elevation values is measured and interpreted to describe the random variable elevation through a probability distribution. Thus given one representation of elevation, the value of elevation at one location has a confidence defined by the probability distribution.

For example, if there are four representations of elevation and, at some location their values are 996, 1,000, 1,002, and 1,005 meters, the mean and standard deviation are 1,000.75 and 3.8 meters respectively. Mean and standard deviation are used to calculate the standard score (z-score--the difference from the mean in standard deviation units) of any of the representations at that location. The z-score for the representation with an elevation equal to 996 meters is -1.26, indicating that, using a Student's distribution, the probability that a measure of elevation will be 996 meters or lower is 0.15. For the representation with an elevation value equal to 1,005 meters, the z-score is 1.13, thus the probability that a measure will be or exceed 1,005 meters is 0.17.

Therefore for each of the representations, at every location, the probability that the elevation value is correct can be defined. Note that the probability distribution is assumed to be a normal distribution.

The proposed method requires at least three representations of elevation at one location. If there are only two, the probability will be the same for the whole area since the z-score is the same (0.707) for any pair of values that are different. In addition, the probabilities change with the number of available representations. For example, if there were 3 representations instead of 4 in the example above, the probability of the representation with an elevation value of 996 meters changes from 0.15 to 0.20.

### **5.1.3. Cluster Analysis**

Errors in elevation are spatially autocorrelated, allowing cluster analysis to define areas with uniform accuracy. The cluster detection method is based on finding the significant peaks of a surface that results from the application of a smoothing filter, based on the Gaussian kernel, on a rectangular grid with normally distributed values (Rogerson 2001a). The Gaussian kernel must be applied on a standardized measure, thus the z-score is calculated on the grid with the probability values. The clusters of significant high and low values of probability are found where the smoothed grid value exceeds the critical value that depends on the filter smoothing factor and the total number of grid cells.

#### **5.1.3.1. Z-Score**

The z-score is calculated for the probability by first computing its mean and standard deviation for the whole grid. Next, at every point of the grid, the z-score ( $z$ ) is calculated from the difference between the probability value at the location point and the mean for the grid. Finally, the z-score is calculated by the ratio between the difference and the standard deviation.

### 5.1.3.2. Gaussian Kernel Filter

Given that the distribution of probabilities on the grid is assumed to be normal, the Gaussian kernel filter is used to smooth the local random differences in the probabilities. The filter is implemented by applying weights calculated according to the distance of each neighbor cell of the probability grid to the center cell. The weights  $w_{ij}$  to be applied to neighbor cells  $j$  of center cell  $i$  are calculated using (5.2) (Rogerson 2001a):

$$w_{ij} = e^{\frac{-d_{ij}^2}{2\sigma^2}} / \sqrt{\pi\sigma} \quad (5.2)$$

where:

$\sigma$  is the standard deviation of the Gaussian kernel and

$d_{ij}$  is the distance between cell  $i$  and cell  $j$ .

Note that since the z-score is being used, the distance considered here is also in standardized distance, i.e. distances are measured in cell units. The standard deviation of the Gaussian kernel is selected based on compromises: reduction of local random variation; preservation of clusters of high variability; being convenient for filter implementation; and definition of critical value for cluster detection.

### 5.1.3.3. Critical Value

The critical value  $M^*$  for a grid with total area  $A$ , at a significant level  $\alpha$  is approximated by (5.3) (Rogerson 2001a):

$$M^* = \sqrt{-\sqrt{\pi} \ln\left(\frac{4\alpha(1+.81\sigma^2)}{A}\right)} \quad (5.3)$$



The approximation of  $M^*$  is valid only if  $A$  is smaller than 10000 or  $\sigma$  is not smaller than one, and if  $A$  is greater than 10000, the approximation can be used if  $\sigma_i/\sqrt{A} > 0.01$ , where  $\sigma_i = \sqrt{\sigma_0^2 + \sigma^2}$  and  $\sigma_0$  is equal to 10/9 for most cases (Rogerson 2001a).

In most representations of elevation,  $A$  is greater than 10000, and the  $\sigma$  is selected between 1 and 4. In these conditions, the restrictions for using the approximation of  $M^*$  are not satisfied. However, the critical value obtained by using the approximation is only slightly smaller than the value calculated using the exact formulation (Rogerson 2001a), and the differences in cluster area sizes using the two critical values are small. Therefore, the critical value obtained by the approximation can be used for detecting the clusters.

#### 5.1.4. Classified Probability Map

The areas of the probability grid where the critical value is exceeded are extracted and a value of accuracy is calculated for each of the areas. The options for accuracy measures are using the mean value of the probability used to define the clusters and the use of differences between the representation values and the mean value of all representations to estimate the RMSE inside each of the areas.

The use of probability values has the advantage of including the probability distribution of elevation at each location. However, the measure is not usually used. Thus, the estimation of the RMSE using the mean value of elevation at every location inside the cluster is preferred.

### **5.1.5. Example of Accuracy Map**

An example of the use of multiple representations of elevation to create a spatially distributed accuracy map is provided here since the method is not used in the next chapter. In this example of an accuracy map only two representations were available. Thus, the method was extended to use difference between the representations instead of standard score values. The cluster analysis and the subsequent steps were the same as proposed.

The representations of elevation used in the example are the DEM from the Shuttle Radar Topographic Mission (SRTM) and the DEM from the Arizona Regional Image Archive (ARIA). Detailed description of these DEMs is presented in Chapter 6. A rectangular regular grid of differences between ARIA and SRTM DEMs is first calculated.

The Gaussian kernel filter must be applied on a standardized measure, i.e. the rectangular regular grid values must be z-score values at individual locations. Z-scores are valid only if the distribution of differences between ARIA and SRTM DEMs is a normal distribution. Table 5.1 summarizes the descriptive statistics of the differences at all grid points.

**Table 5.1** Descriptive statistics of difference between ARIADEM and SRTMDEM.

<b>Statistic</b>	<b>Value</b>
<b>Number of Points</b>	220,840
<b>Mean Value (meters)</b>	-10.37
<b>Standard Deviation (meters)</b>	16.87
<b>Skewness</b>	-0.556
<b>Kurtosis</b>	4.690
<b>Minimum (meters)</b>	-139.01
<b>Maximum (meters)</b>	75.68

Given that the difference is calculated subtracting SRTMDEM from ARIADEM, the mean value in Table 5.1 being negative indicates that elevations in SRTMDEM are on average higher than in ARIADEM. The other values in Table 5.1 confirm that the distribution of differences is close to a normal distribution. Absolute value of skewness is smaller than one indicating a near symmetric distribution. Kurtosis greater than 3 indicates a peaked distribution.

The z-score grid was calculated using the standard deviation and mean from Table 5.1 by computing how many standard deviations each location of the differences grid is from the mean.

The z-score grid was smoothed by applying a Gaussian Filter. The standard deviation of the filter was defined by comparing the percentage of areas inside clusters generated when using the filter with standard deviation 1, 2, and 3. The results, shown in Table 5.2, indicate that the size of clusters for standard deviation 1 is too small and the

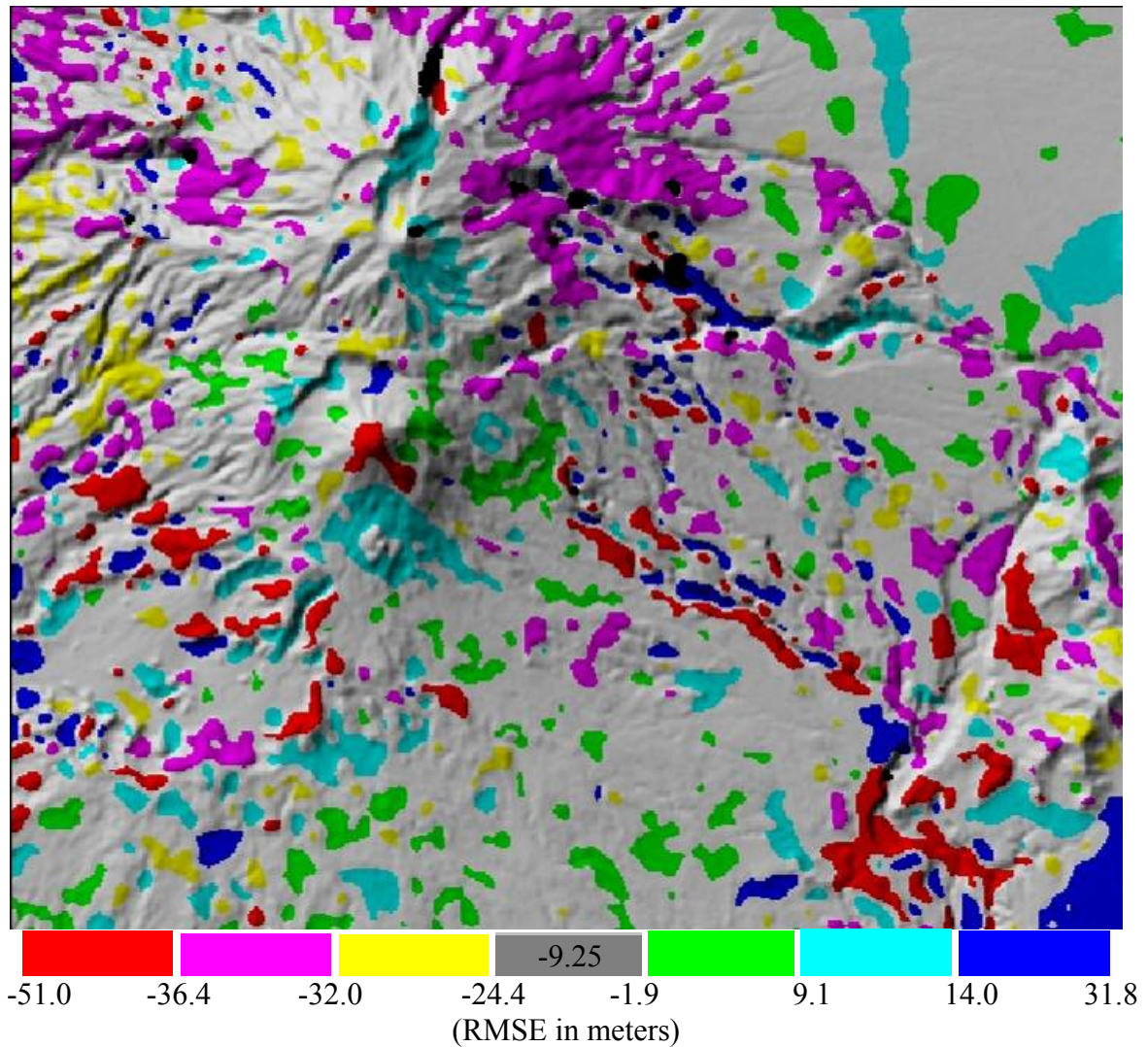
size for standard deviation 3 is too large. Therefore, the value of the filter standard deviation was selected to be equal to 2.

**Table 5.2** Area of clusters for Gaussian kernel filters with standard deviation 1, 2, and 3.

Standard Deviation	Area of Clusters (% of Total)
1	4.48
2	15.16
3	22.70

The critical value for detecting clusters with 95% certainty ( $\alpha$  equal to 0.95) is 4.9, using (5.3) with the standard deviation of the Gaussian filter  $\sigma$  equal to 2 and  $A$  equal to 220,840. Note that the critical value is relative to the z-score of the differences between the two DEMs. Also, since the difference is considered, negative values indicate where there are significant clusters of negative differences.

For each cluster the mean value of difference was calculated and assigned. Figure 5.1 presents the resulting accuracy map.



**Figure 5.1** Example of a spatially distributed accuracy map created using detection of clusters of high values of difference between SRTM and ARIA DEMs. Clusters have difference values indicated in the color scale, with the mean difference values inside each of the clusters. Background in grey is a shaded relief image and the mean difference is -9.25 meters for these areas that are outside clusters. Areas SRTM has no elevation value are presented in black.

## 5.2 COMPUTING A MEASURE OF SIMULATION

### PERFORMANCE

Simulation of spatially distributed dynamic process offers challenges not only for information acquisition and computation of the simulation, but also for comparison of the

simulation results with observations from the real world. Often the amount of information available from a real case is much smaller than the results from a simulation.

Fieldwork to collect data from the real event can only gather a small part of the information set. For example, geophysical mass flows information are presented by descriptions of observations at some locations and a summary map with the flow footprint.

Given these characteristics, a performance measure that compares the large amount of information from simulation of geophysical mass flows with the limited information available in a footprint map is proposed here.

The proposed method considers the characteristics of the real event and of the process model simulation to generate a measure of performance based on logistic regression. The dependent variable in the regression is the probability of being inside the real event flow footprint. The independent variables are the flow height in the model simulation and a measure of distance based on the real event flow footprint.

### **5.2.1. Existing Comparison Methods for Dynamic Processes**

Comparison of outputs from dynamic processes models is more difficult than for outputs from steady-state models processes since for the later the comparison is made using one time snapshot where sample values from the simulation and from the simulated event can be collected and compared. Therefore, common comparison methods for simulation of dynamic processes are quantitative methods using a summary of values at one location and qualitative assessment.

Quantitative analysis in a few locations is commonly used in hydrological models. The key locations are located at the watershed discharge point, where a summary of the hydrological process for the whole basin can be measured. For example, total water runoff volume for a period of time can be used to compare results of a soil-hydrology-vegetation model simulation using various Digital Elevation Models (DEMs) (Kenward et al. 2000).

Qualitative analysis is not a reliable method, since it involves subjectivity. The method consists in drawing the simulated results and the observations from the simulated event together and analyzing for similarity. A hydrological model is usually compared using hydrographs (2-D plots of water discharges at one location as a function of time) and involves verifying if the peaks coincide in time, if they have the same volume, or if they have the similar shapes.

For the simulation of geophysical mass flows, comparison at one location only is not enough since the most important simulation result is where areas affected by the flow are located. Qualitative analysis is executed by creating a map of areas reached by the flow and the footprint built from field work observations, and comparing their shapes. The method has been used for simulation of a pyroclastic flow which occurred at Unzen Volcano in Japan (Takahashi and Tsujimoto 2000), simulation of lava flows at Mount Etna, Italy (Crisci et al. 2004), and simulation of pyroclastic flow at Merapi Volcano, Indonesia (Itoh et al. 2000).

### **5.2.2. Quantitative Comparison Method**

The lack of a quantitative method for comparison of simulations of spatially distributed process models and the need to compare results from simulations using multiple representation of elevation lead to designing a quantitative method that, although it is targeted for the comparison of geophysical mass flow simulations, it can also be applied to other similar process models.

The main characteristic of the proposed method is the use of logistic regression. Regression analysis is suited for quantitative comparison since the relation between variables is quantified by a mathematical equation.

#### **5.2.2.1. Logistic Regression**

Linear regression is used when the relationship between dependent and independent variables is linear and requires regression errors (difference between values predicted by the regression and observed values) to have zero mean, constant variance, independent regression residuals, and normal distribution of errors around the regression line (Rogerson 2001b). On the other hand, the use of linear regression is not appropriate when the dependent variable has a saturation value (Sutton et al. 1997).

Therefore, linear regression can only be used for comparison between simulations using different representations. In that case, sample locations are selected and values from simulation results at those locations define the values of linear regression variables.



Logistic regression defines the probability of occurrence of a categorical variable given an independent variable. Since the footprint from field work provides a categorical value, i.e. a location is either inside or outside the flow footprint, and the simulation result is a numerical variable, logistic regression is the appropriate comparison method for geophysical mass flows. The equation for the probability of a logistic regression is given by (5.4) (Rogerson 2001b):

$$p = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} \quad (5.4)$$

where:

$p$  is the probability of the categorical variable;

$x$  is the independent variable; and

$\alpha$  and  $\beta$  are the parameters of the regression.

$\alpha$  and  $\beta$  are calculated by minimizing the sum of the squared deviations between observed and predicted values using the probability equation (Rogerson 2001b).

$\alpha$  and  $\beta$  provide the quantitative measure for comparison, since for  $\alpha + \beta x = 0$  the probability is 0.5. Thus, in the case of geophysical mass flows,  $\alpha$  and  $\beta$  define the value of  $x$  that makes the probability of a location being inside or outside the flow footprint from the field work equal.

#### 5.2.2.2. Independent Variable Definition

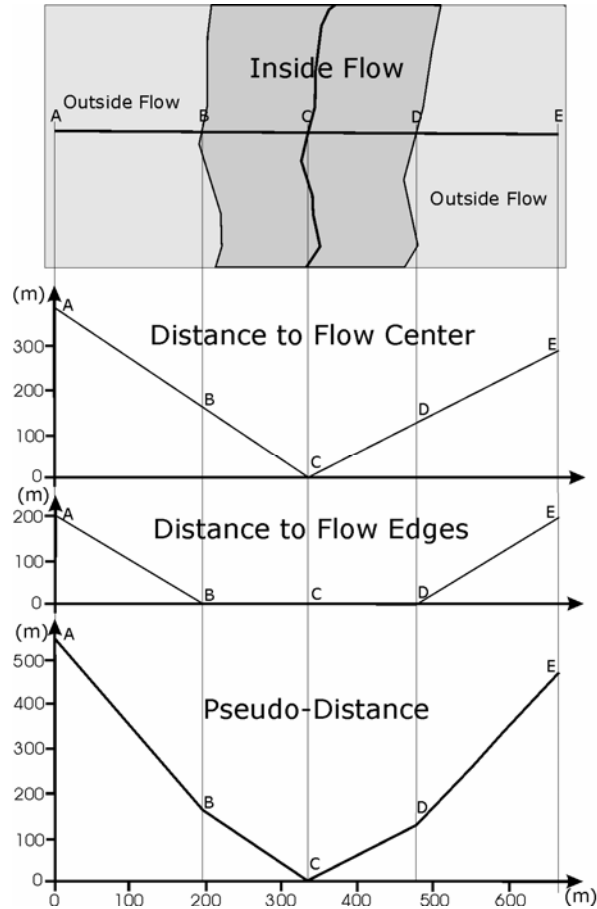
The value of the logistic regression independent variable  $x$  defines the probability of a given location being inside the flow footprint estimated from fieldwork. The variable  $x$  is related to results from the simulation of the test case.

The simulation provides values describing the dynamic process at every time and at every location. For geophysical mass flows, the simulation provides height and momentum of the pile. Since momentum is mass times velocity, for values at the limit of numerical simulation, momentum is not a stable variable. On the other hand, pile height usually decreases smoothly away from the center of the pile. Thus the independent variable  $x$  should be related to the height of the moving pile at all locations.

Since the comparison is made with the flow footprint, which indicates the extent of the pile at all times of the event, the pile height for the comparison measure should represent a similar feature. Instead of considering the pile height at a given time, the independent variable is related to the maximum pile height at all times during a simulation.

If the maximum pile height only is used to define the independent variable, the comparison measure will not include where the pile is higher. Therefore, location where heights are measured should be included in the independent variable  $x$ . Since it is assumed that the pile is higher at the center of the flow footprint, distance to the center of the pile is a candidate to be included.

However, it is desirable for a good simulation to have pile heights higher inside the flow footprint than outside. The proposed solution uses pseudo-distance. The pseudo-distance for a location inside the flow footprint is the distance to the center of the flow and pseudo-distance for a location outside the flow is the distance to the nearest edge of the flow added to the distance to the flow footprint center.



**Figure 5.2** Pseudo-distance for a portion of a flow footprint, along the line from point A to point E (line AE). Points B and D are located at the intersections between the flow footprint edges and the line AE. Point C is located at the intersection between the flow footprint medial axis and the line AE. Pseudo-distance is obtained by adding the distance to flow center and the distance to flow edges.

Combining the pseudo-distance with the pile height to define the independent variable can be done in many ways. However, in this method the choice was made to use pseudo-distance as another independent variable. The solution avoids subjectivity that each of the ways to combine pseudo-distance and pile height would include. Therefore, the equation for the probability of the logistic regression to define the performance measure is given by (5.5):

$$p = \frac{e^{\alpha + \beta_1 x_1 + \beta_2 x_2}}{1 + e^{\alpha + \beta_1 x_1 + \beta_2 x_2}} \quad (5.5)$$

where:

$p$  is the probability of the categorical variable;

$x_1$  is the independent variable pile height (in length units);

$x_2$  is the independent variable pseudo-distance (in length units);

$\alpha$  is the constant of the regression;

$\beta_1$  is the parameter of the regression associated with pile height; and

$\beta_2$  is the parameter of the regression associated with pseudo-distance.

### 5.2.2.3. Quantitative Performance Measure Definition

The exponential of  $\beta_1$  and  $\beta_2$  define the increase and decrease in the odds of a location being inside the flow footprint given the location's pile height and pseudo-distance to the footprint. Given that  $\beta_1$  is associated with the pile height, the performance measure is defined by the exponential of  $\beta_1$ .

The performance is better, that is, the footprint and the simulation results will be more similar, when the measure is higher. The value of performance measure for the best case situation can not be stated given the dependence on the shape of the footprint.

## 5.3 SUMMARY

The existence of multiple representations requires additional tools for analysis and at the same time allows analysis of the representations of elevation that was not possible before.

Multiple representations of elevation allow the definition of spatially distributed accuracy. Previously, only global measures of accuracy were attached to representations of elevation. In addition, some methods searched for errors in representations of elevation using spatial autocorrelation principles.

The method proposed in Section 5.1 uses cluster analysis to define regions where inaccuracy is higher. For each of the regions, the mean and standard deviation of either differences between representations or a probability of the value of elevation being correct are calculated. An example of the method to create a spatially distributed accuracy map was provided since the method is not used in the next chapter.

The example showed that an accuracy map can also be obtained with two representations of elevation. In this case, differences between the representations are used instead of standard score.

Simulation results vary for each representation of elevation or each combination of representations used. Since the qualitative and the quantitative measures that exist are not suited for spatially distributed dynamic process, a new quantitative measure of performance was presented in Section 5.2.

The main characteristic of the performance measure is the use of field work data to provide a categorical variable for a logistic regression. In addition, the method is specifically designed for models of geophysical mass flows. The independent variables of the logistic regression are the maximum pile height for all locations and a pseudo-distance measure of those locations to the center and the edges of the real event flow footprint. Finally, the performance measure is given by the ratio between the parameter

associated with pile height and parameter associated with pseudo-distance in the logistic regression.

Although the comparison method assumes that the pile is higher at the center of the footprint, different locations can be defined for specific cases. The method can be extended to other simulations where field data is obtained as a categorical variable and aggregated in a footprint.

In the next chapter, the performance measure method is used to verify if the use of multiple representations of elevation improves results from simulations of geophysical mass flows.

## **PART III: Validation and Conclusions**

### **Chapter 6**

#### **VALIDATION OF THE INTEGRATION OF MULTIPLE REPRESENTATIONS OF ELEVATION AND PROCESS MODELS**

This chapter presents the results of the simulation of geophysical mass flows using the proposed framework at two locations, Colima in Mexico and San Bernardino in California. The simulated event in Colima is the block-and-ash flow which occurred in April 1991. In San Bernardino, the simulated event is the debris flow which occurred in December 2003.

Section 6.1 presents the representations of elevation that are used in the simulations, along with the description of the information required by the representation metadata. In addition, a tool created to facilitate the storage of elevation in the database is presented.

The simulation of the Colima Volcano block-and-ash flow is presented in Section 6.2, with the description of the simulated event and the results from the simulation using the new implementation with the proposed framework and using the current implementation. Comparison of the results is executed using the performance measure presented in Chapter 5.

Section 6.3 is similar to Section 6.2, differentiated by presenting the debris flow in San Bernardino instead of the block-and-ash flow of Colima Volcano.

## **6.1 CREATING MULTIPLE REPRESENTATIONS IN TERRALIB DATABASE**

The representations of elevation are stored in the TerraLib database using the framework presented in Chapter 4. The example implementation used in this study handles representations of elevation in raster format only. Therefore, representations in a regular rectangular grid are imported and made available in the database. Extensions to this implementation can be built using existing TerraLib library functions to handle import of other type of representations, such as Triangular Irregular Networks (TINs) and contour lines.

The computational tool created to facilitate the importation of elevation grids is presented in Section 6.1.1. The representations available for the simulation of geophysical mass flows that were imported are presented in Sections 6.1.2 (Colima) and 6.1.3 (San Bernardino).

The elevation representations include freely available ones only to emulate simulations that are required to be done with few resources.

### **6.1.1. Tool to Import Representations of Elevation**

Storage of representations of elevation requires the values at the cells of the grid and information about the modeling procedure. In the proposed framework they are



stored in a spatial database. Although it is possible to use database functions to import the representations, an import tool was created to facilitate the procedure.

The main tasks of the tool are to read files with elevation values and files with metadata describing the modeling procedures and insert the representations of elevation and their information in the database tables described in Chapter 4.

The file formats for elevation values that are processed by the tool include text file, binary file, and Tag Image File Format (TIFF) file. Information about the projection is also processed from these files. Spatial extent and projection information are used to create the spatial extent of the valid values for the representation. The spatial extent and modeling information are stored in the “**RepresInfo**” layer (see Chapter 4). Description of the metadata file is presented in the following section.

#### **6.1.1.1. Metadata File Description**

A file with metadata must exist for each representation of elevation to be imported. The file must contain the same fields of the “**represattrib**” table (see Chapter 4). The pair field name and field values must be in the same line, with the field values inside quotes (“ ”). Comments can be inserted in the file, at lines that do not have field names, and must begin with double slashes (/). The required field names are presented in Table 6.1, with examples of field values.

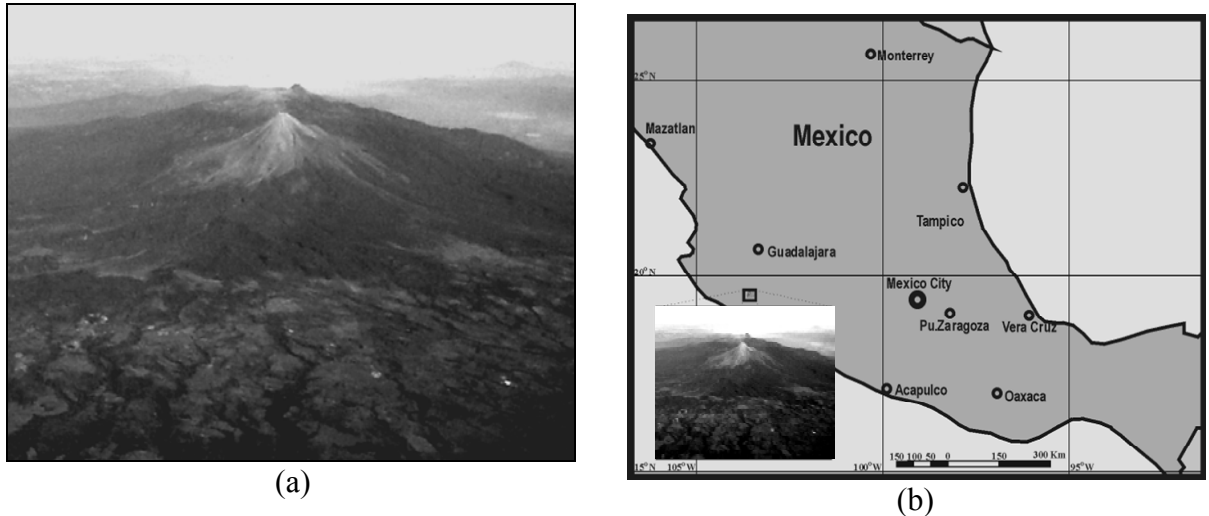
**Table 6.1** Field names of metadata file describing a representation, with examples of field values.

Field Name	Example
DESCRIPTION	"A short description..."
MEASURE_DATETIME	"1963/11/09 0:0:0"
PRODUCTION_DATETIME	"2006/01/26 0:0:0"
PRODUCTION_METHOD	"Name of the method"
PRODUCTION_SOURCE	"Source information"
SOURCE_SCALE	"2400"
MEASURE_UNIT	"Feet"
MEASURE_NAME	"Elevation"
HORIZONTAL_ACCURACY	"0.5"
VERTICAL_ACCURACY	"2"

Note that the scale values omit the “1:” part. In the example of Table 6.1, the scale is 1:2,400. The horizontal accuracy is the accuracy of the measurement position in the projected map units. The vertical accuracy is the stated accuracy of the instrument or technique.

### 6.1.2. Importing Elevation for Colima

Colima Volcano is located in Mexico, in the western portion of the Trans-Mexican Volcanic Belt, at the southern edge of the Colima graben. The volcano summit is located at approximate geographic coordinates 19° 31’ north and 103° 37’ west, at an elevation of 3,860 meters above sea level (Saucedo et al. 2004). Figure 6.1 shows Colima Volcano in October 28, 2003 and its location in Mexico.



**Figure 6.1** Colima Volcano on October 28, 2003 (a) and its location in Mexico (b).  
(Photo by the author)

Two representations of elevation for Colima, freely available, are the one obtained from the Shuttle Radar Topography Mission (SRTM) and the representation available from Arizona Regional Image Archive (ARIA - [aria.arizona.edu/](http://aria.arizona.edu/)).

#### 6.1.2.1. SRTM DEM

DEMs derived from SRTM are available from the United States Geological Survey (USGS) repository for unfinished research Digital Elevation Model (DEM). The unfinished DEM was selected to avoid the changes of the finishing procedure, in which voids are filled and water bodies are included (Slater et al. 2006). Elevation values are available at 3-arcsecond resolution and they are stored in files comprising a one degree squared area (Slater et al. 2006).

The file with elevation covering the Colima Volcano area is the N19W104.hgt (available for download at <ftp://edcscgs9.cr.usgs.gov/pub/data/srtm/>), with elevation

values in meters at every 3-arcseconds between geographic coordinates 19° north, 104 ° west to 20 ° north, 103 ° west, using the WGS 84 reference frame.

Since a rectangular raster representation of elevation is needed for storage in the database, the following manipulations are made on N19W104.hgt file:

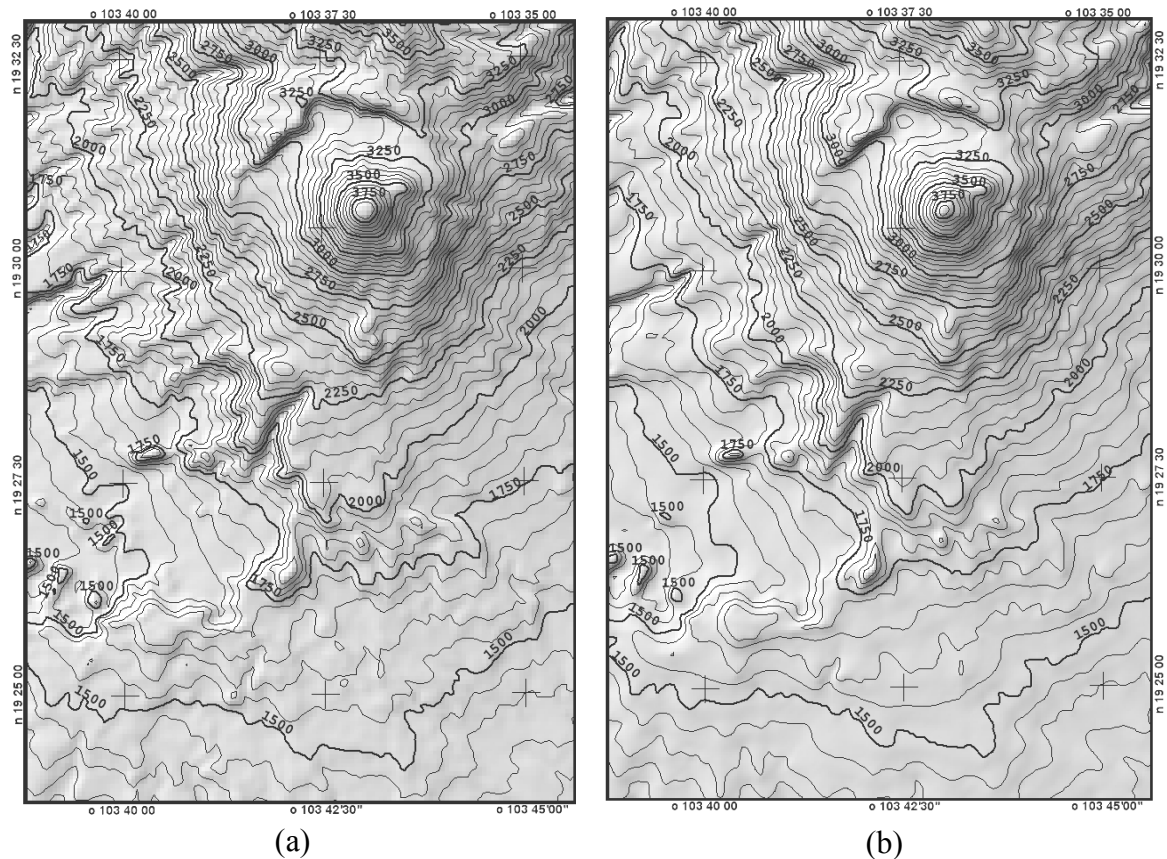
1. The file was converted to a sample points file, i.e. a file with the triple geographic coordinates and elevation value for each of the values of file;
2. The target projection was defined as Universal Transverse Mercator (UTM) projection, zone 13, using the International Terrestrial Reference Frame of 1992 (ITRF92) since this projection was selected by the mapping agency of Mexico, the Instituto Nacional de Estadística, Geografía e Informática (INEGI);
3. A TIN was created from the sample points; and
4. The rectangular grid was created using linear interpolation inside triangular patches. The resolution of the grid was defined as 90-meter in X and Y directions given that 3-arcseconds corresponds to approximately 89 meters in X direction and 91 meters in Y direction in the selected UTM projection for the Colima Volcano area.

The SRTM elevation in raster format was imported into TerraLib database, using a metadata file with the contents presented in Figure 6.2.

```
DESCRIPTION "Elevation from SRTM, reprojected to UTM, 90-meter resolution"
MEASURE_DATETIME "2000/02/1 0:0:0"
PRODUCTION_DATETIME "2006/01/26 0:0:0"
PRODUCTION_METHOD "Linear Interpolation"
PRODUCTION_SOURCE "SRTM 3arc-sec"
SOURCE_SCALE "25000"
MEASURE_UNIT "Meters"
MEASURE_NAME "Elevation"
HORIZONTAL_ACCURACY "5"
VERTICAL_ACCURACY "16"
```

**Figure 6.2** Metadata file associated to the SRTM elevation for the Colima Volcano area.

The values of the metadata file were obtained from the specifications of SRTM DEMs (van Zyl 2001) and the conversion procedure. Figure 6.3 (a) presents SRTM DEM using contour lines and a shaded background image.



**Figure 6.3** Representations of elevation for Colima Volcano: (a) DEM from SRTM, resampled from 3-arcsecond resolution to 90-meter resolution; (b) DEM from ARIA at 60-meter resolution. DEMs are displayed using contour lines and shaded background image. Contour lines are in meters. Cartographic projection is UTM, zone 13, datum ITRF92.

#### 6.1.2.2. ARIA DEM

DEMs are available from ARIA in files with elevation values at a rectangular grid with 60 meter resolution. The projection is UTM zone 13, with North American Datum of 1927 (NAD27). A pair of DEMs from ARIA is needed to cover the Colima Volcano area. They were downloaded (from <http://aria.arizona.edu/>) and used to create a new

rectangular grid that was reprojected to UTM zone 13 using the ITRF92 reference system. The DEM was imported in the TerraLib database, using a metadata file with the contents presented in Figure 6.4.

```
DESCRIPTION "Elevation from Arizona Image Archive - ARIA. Source data is  
probable one, from INEGI topographic maps"  
MEASURE_DATETIME "1995/12/01 0:0:0"  
PRODUCTION_DATETIME "2006/01/26 0:0:0"  
PRODUCTION_METHOD " Interpolation"  
PRODUCTION_SOURCE "Contour Lines"  
SOURCE_SCALE "50000"  
MEASURE_UNIT "Meters"  
MEASURE_NAME "Elevation"  
HORIZONTAL_ACCURACY "20"  
VERTICAL_ACCURACY "15"
```

**Figure 6.4** Metadata file associated to the ARIA elevation for the Colima Volcano area.

Although ARIA DEMs have metadata files, information required for the multiple representations framework is poor. For this example, it was assumed that the DEMs were derived from contour lines of topographic maps from INEGI. The topographic maps are available at 1:50,000 scale and the following ones cover the Colima Volcano area: San Gabriel - E13B24, Ciudad Guzman - E13B25, Comala - E13B34, and Cuauhtemoc - E13B35. The maps were derived from photos taken during flights in December 1995 (INEGI 1999c; INEGI 1999a; INEGI 1999d; INEGI 1999b).

Therefore, the values of measure date and time, and scale of the topographic maps are used in the DEMs metadata files. In addition, it was assumed that the DEMs are similar to USGS level 1 DEMs, and vertical accuracy was assumed to be the maximum permitted, that is, 15 meters (USGS 1998). The horizontal accuracy was assumed to be 20 meters since a 1:24,000 topographic maps has a 10 meters positional accuracy

(Goodchild 1996). Note that the positional accuracy of a topographic map is associated to the available drawing capabilities; therefore, it is correct to assume that when scale is halved the positional error will double.

Figure 6.3b presents ARIA DEM using contour lines and a shaded background image.

### **6.1.3. Importing Elevation for San Bernardino**

The debris flow in San Bernardino County, California, affected the area along Waterman Canyon. DEMs covering the area are freely available from the National Oceanic and Atmospheric Administration (NOAA) Coastal Services Center (CSC), USGS repository for unfinished research DEM, and the USGS Seamless Data Distribution System.

The USGS Seamless Data Distribution System has the following DEMs for the San Bernardino area: the National Elevation Dataset (NED) at 1-arcsecond and 1/3-arcsecond resolutions, and the finished SRTM at 1-arcsecond and 3-arcsecond resolutions. However, none of the DEMs were used. The finished SRTM were not used since the unfinished DEMs were preferred due to changes of the finishing procedure--voids are filled and water bodies are incorporated in the finished DEMs (Slater et al. 2006). The NED 1-arcsecond DEMs were not used because the source measurement date and time are not available. Since the NED 1/3-arcsecond DEMs are interpolated from the 3-arcsecond DEMs, they were also not used.

#### 6.1.3.1. NOAA-CSC DEM

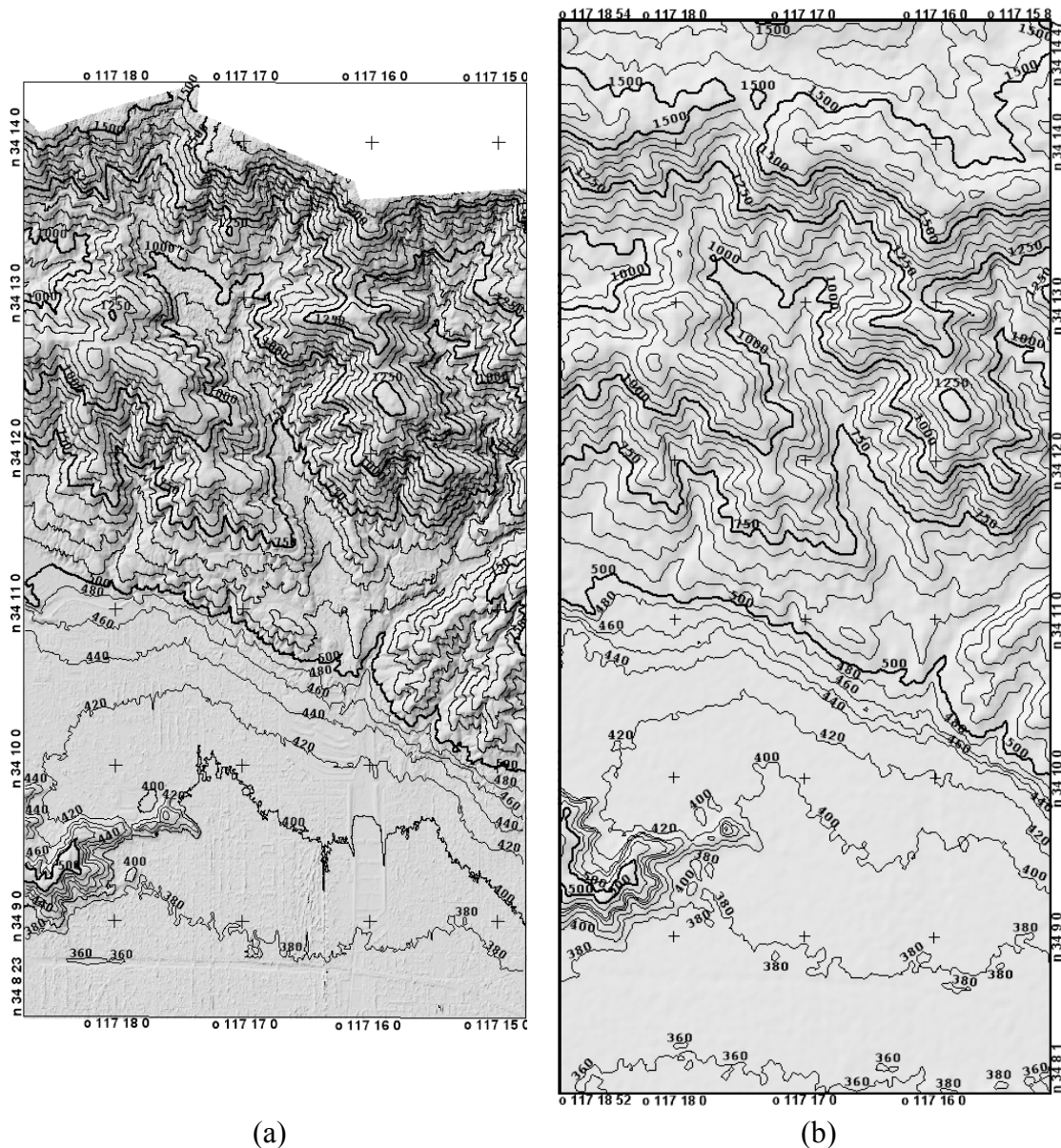
The DEM from NOAA-CSC was generated from interferometric radar using band-X, with 1 meter vertical accuracy at non-vegetated areas and horizontal accuracy of 2.5 meters (NOAA 2006). The file with elevation for the Waterman Canyon was downloaded (from <http://www.csc.noaa.gov/TCM/>) and imported into the TerraLib database. The metadata file was created based on the available information and its contents are presented in Figure 6.5.

```
DESCRIPTION "Elevation from CSCNOAA. Interferometry from X-band (160 MHz  
bandwidth, center frequency of 9.7 GHz) and P-band (350 MHz)"  
MEASURE_DATETIME "2002/10/20 22:52:0"  
PRODUCTION_DATETIME "2006/01/26 0:0:0"  
PRODUCTION_METHOD "Interferometry Processing"  
PRODUCTION_SOURCE "Interferometry"  
SOURCE_SCALE "2400"  
MEASURE_UNIT "Meters"  
MEASURE_NAME "Elevation"  
HORIZONTAL_ACCURACY "2.5"  
VERTICAL_ACCURACY "1.5"
```

**Figure 6.5** Metadata file associated to the NOAA\_CSC elevation for the San Bernardino area.

Scale was estimated from the DEM resolution and approximated to a standard map scale, 1:2400. Accurate measurement date and time is not available since the metadata file provided by NOAA-CSC contains the information for all flights with no specification of which flight data were used to produce the DEM for the area. Therefore, the measure date and time was assumed to be the same as the first flight. Figure 6.6a presents SRTM DEM using contour lines and a shaded background image.





**Figure 6.6** Representations of elevation for Waterman Canyon, San Bernardino County, CA: (a) DEM from CSC-NOAA at 3-meter resolution; (b) DEM from SRTM, resampled from 1-arcsecond resolution to 30-meter resolution. DEMs are displayed using contour lines and shaded background image. Contour lines are in meters. Cartographic projection is UTM, zone 11, datum NAD83.

### 6.1.3.2. SRTM DEM

DEMs derived from SRTM are available from the same USGS repository of the data for Colima (<ftp://edcsgs9.cr.usgs.gov/pub/data/srtm/>). The file with elevation covering the San Bernardino area the N34W117.hgt, with elevation values in meters at

every 1-arcsecond between geographic coordinates 34° north, 118° west to 35° north, 117° west, using the WGS 84 reference frame. The file was processed using the same procedure used for the SRTM DEM for Colima. The SRTM elevation was imported into TerraLib database, using a metadata file with the presented in Figure 6.7.

```
DESCRIPTION "Elevation from SRTM, reprojected to UTM, 30-meter resolution"
MEASURE_DATETIME "2000/02/1 0:0:0"
PRODUCTION_DATETIME "2006/01/26 0:0:0"
PRODUCTION_METHOD "Linear Interpolation"
PRODUCTION_SOURCE "SRTM 3arc-sec"
SOURCE_SCALE "25000"
MEASURE_UNIT "Meters"
MEASURE_NAME "Elevation"
HORIZONTAL_ACCURACY "5"
VERTICAL_ACCURACY "16"
```

**Figure 6.7** Metadata file associated to the SRTM elevation for the San Bernardino area.

Figure 6.6b presents SRTM DEM using contour lines and shaded background image.

## **6.2 SIMULATION OF COLIMA BLOCK-AND-ASH FLOW EVENT**

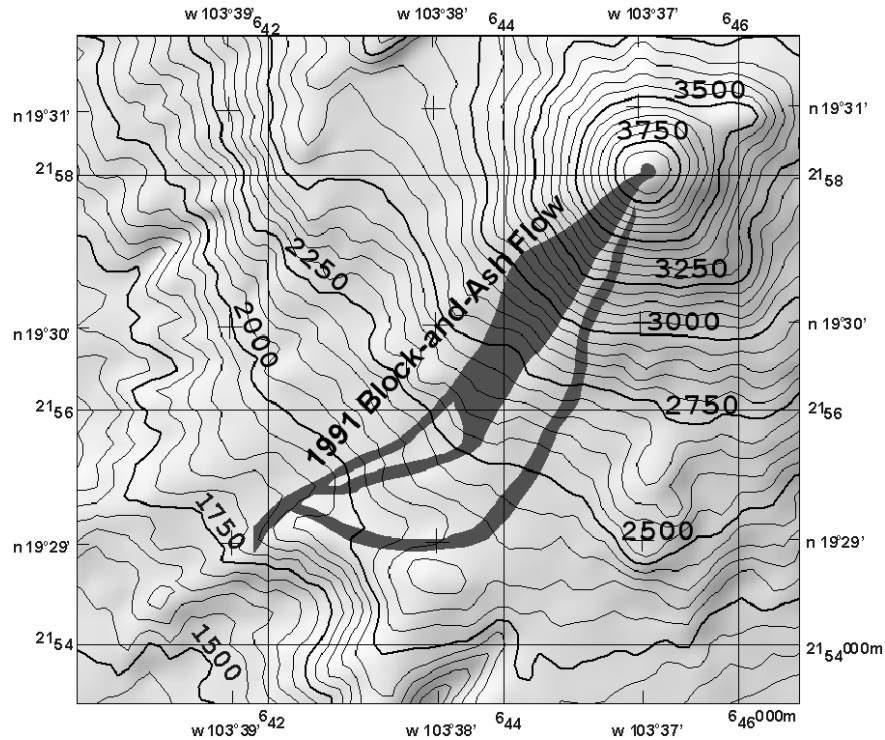
This section describes the simulation of the Colima Volcano block-and-ash flow event which occurred in April of 1991. The location of Colima Volcano is presented in Section 6.1.2. The volcano is a classic composite volcano, with slope angles over 30 degrees from the summit to about 2 km to 3 km from it (Saucedo et al. 2004). Records of eruptive activity at Colima Volcano are available since the 16<sup>th</sup> century, with previous ones being lost due to Spanish conquest, and support the claim that Colima is the most active volcano in Mexico (Gonzalez et al. 2002).

### 6.2.1. Event Description

The block-and-ash flow event of April 16-17, 1991 was a Merapi-type flow. In a Merapi-type flow, pyroclastic flow is generated by the collapse of the lava dome due to gravitational force. Large lava blocks are crushed into smaller particles as they flow down the slope. The smaller particles travel as fluids suspended by gases released during the blocks' breakup (Takahashi and Tsujimoto 2000). The Colima flow occurred in the following sequence (Rodriguez-Elizarraras et al. 1991) (p.400-401):

1. Extrusion of lava started on March 1, 1991, reaching the southern rim of the summit on April 15, 1991;
2. On April 16, 1991, at 16:00 hours, local time, a series of ash clouds were observed descending the volcano at a distance up to 4 km due to several Merapi-type block- and ash-flows;
3. On the morning of April 17, 1991, a viscous lava flow descended, inducing several rockfalls.
4. The various flows were controlled by topography, filled up a channel, and overtopped the channel ridge. The estimated volume of deposits from block-and-ash flow is  $0.8 \times 10^6 \text{ m}^3$ .

Figure 6.8 shows the extent of the flows obtained from the sketch map by Rodriguez-Elizarraras et al. (1991). Note the lower part of the flow ending in a sharp north-south line (next to 1,750 meters contour line label). The feature is due to the flow reaching the edge of the sketch map. Note also that the flow starts at the volcano summit since the sketch map superimposes the lava flow on the block-and-ash flow.



**Figure 6.8** Extent of Colima Volcano block-and-ash flow event which occurred on April 19-20, 1991. Adapted from sketch map by Rodriguez-Elizarraras et al. (1991).

### 6.2.2. Parameters for Simulation

A simulation of the April 1991 block-and-ash flow event at Colima was executed for validation of TITAN2D, using elevation and derivatives extracted from the DEM derived from the Shuttle Radar Topography Mission (SRTM) reprojected to 90 meter resolution (Rupp et al. 2006). The pile parameters for simulation of the event which was considered to obtain the best match with field data are presented in Table 6.2.

**Table 6.2** Pile parameters for simulation of April 1991 block-and-ash flow event at Colima. Coordinates are in UTM projection meters, Northern hemisphere, zone 13, datum ITRF92.

Parameter		Value
<b>Volume (cubic meters)</b>		0.14x10 <sup>6</sup> from paraboloid defined by 30 meters maximum thickness and 55 meters long major and minor axis
<b>Internal friction angle (degrees)</b>		37
<b>Basal friction angle (degrees)</b>		20
<b>Starting location coordinate</b>		260 meters southwest of the summit - 645,080 East, 2,157,920 North
<b>Simulation region</b>	<b>Lower left coordinate</b>	638,820 East, 2,145,960 North
	<b>Upper right coordinate</b>	648,810 East, 2,161,260 North

The difference in volume between that estimated in the field work (0.8x10<sup>6</sup> m<sup>3</sup>) and the simulation is due to the volume from field work being the aggregated volume of all flows which occurred during the two-day period, while the simulation flow is equivalent to the one main flow (Rupp et al. 2006).

The starting point was moved from the location at 644,956 m East, 2,157,970 m North, used by Rupp et al. (2006), to the location at 645,080 m East, 2,157,920 m North. The change in starting location was required given that the location given by the coordinates used by Rupp et al. (2006) is not located at 260 meters southwest of the summit due to georeferencing differences between elevation representations. The location given by the coordinates used here is located 260 meters southwest of the Colima Volcano summit in the ARIADEM and SRTMDEM.

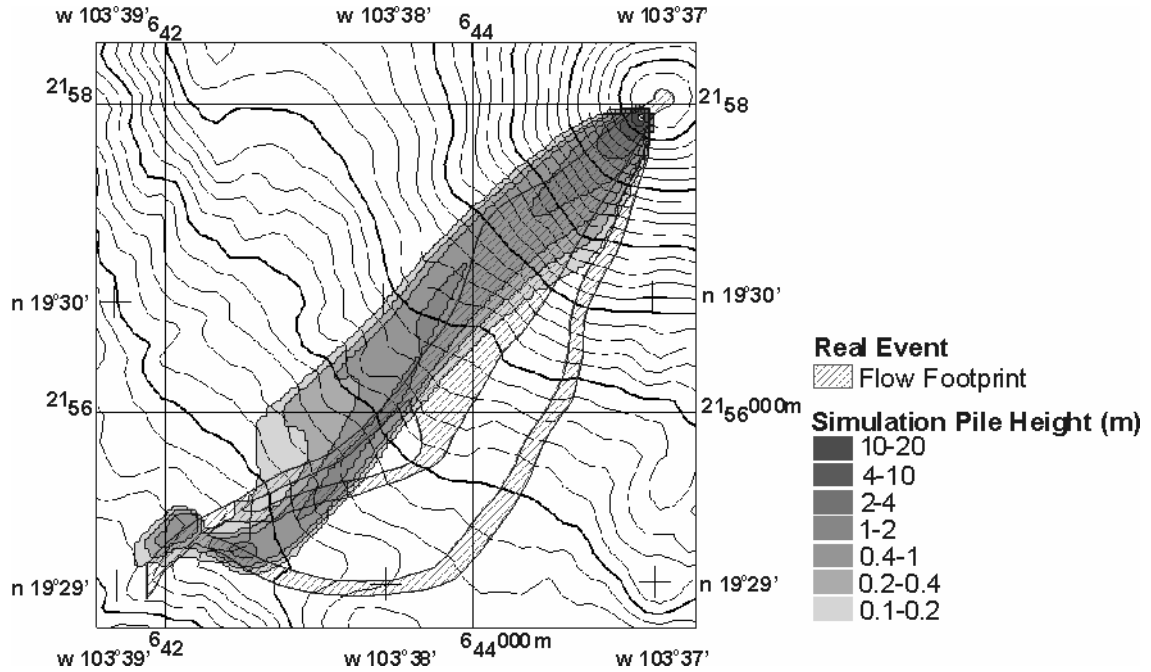
### 6.2.3. Simulation Results Using Current Implementation

The current implementation of TITAN2D was used to simulate the block-and-ash flow event at the Colima Volcano using the elevation and derivatives extracted from the DEM from ARIA and the parameters from Section 6.2.2. The computational parameters of the simulation are presented in Table 6.3.

**Table 6.3** Computational parameters for simulation of April 1991 block-and-ash flow event at Colima.

Parameter	Value
Number of processors	1
Number of points in the computational mesh	300
Use adaptative grid	Yes
Scale governing equations	Yes
Scale length (meters)	4000
Solution method	First order
Maximum number of time steps	8000
Maximum time (seconds)	4000
Time between output of results (seconds)	1

The summary of the results was generated using the procedure presented in Chapter 5 and is presented in Figure 6.9.

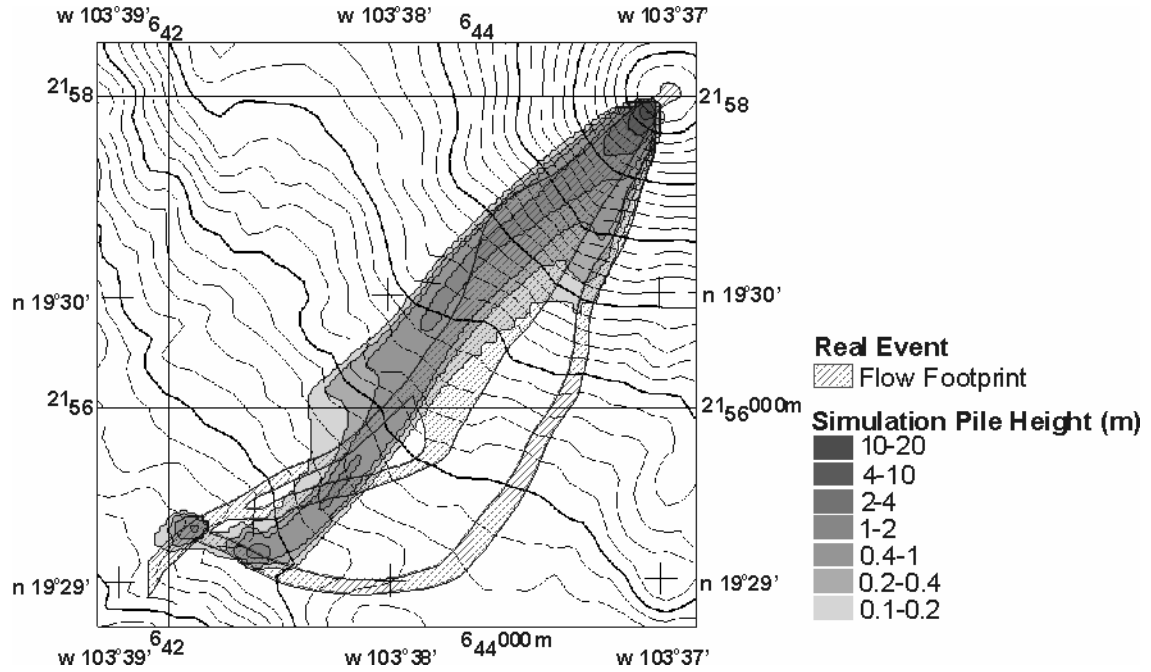


**Figure 6.9** Summary of results from simulation of block-and-ash flow event at Colima Volcano using current implementation of TITAN2D. Pile height values are the maximum values achieved during simulation at each location.

#### 6.2.4. Simulation Results Using New Implementation

The new implementation of TITAN2D and the current standard version were each used to simulate the block-and-ash flow event at Colima Volcano. Each simulation used the same parameters. The elevation and derivatives were extracted from the multiple representations framework, accessing DEMs stored in the TerraLib database. The available DEMs were the SRTM and the ARIA DEMs described in Section 6.1.2.

The summary of the results was generated using the procedure described in Chapter 5 and is presented in Figure 6.10.



**Figure 6.10** Summary of results from simulation of block-and-ash flow event at Colima Volcano using new implementation of TITAN2D. Pile height values are the maximum values achieved during simulation at each location.

### 6.2.5. Comparison of Simulation Results at Colima

The results from the simulations were compared to the field work data using the method presented in Chapter 5. The performance of each simulation is measured using the measure obtained from the logistic regression given by (5.5) (see Chapter 5), where the dependent variable is the probability of a location being inside the flow footprint and the independent variables are the pile height and a pseudo-distance at the same location. Pseudo-distance is the distance to the center of the flow footprint added to the distance to the edge of the footprint.

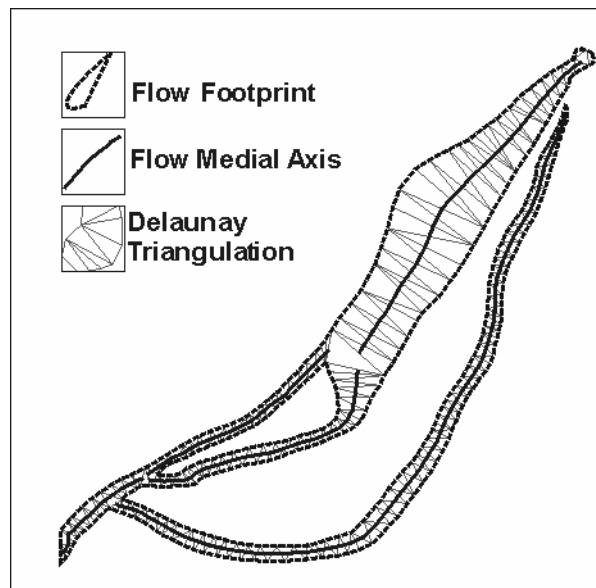


#### 6.2.5.1. Flow Footprint

The footprint of the block-and-ash flow event was obtained from the sketch map by Rodriguez-Elizarraras et al. (1991) presented in Figure 6.8.

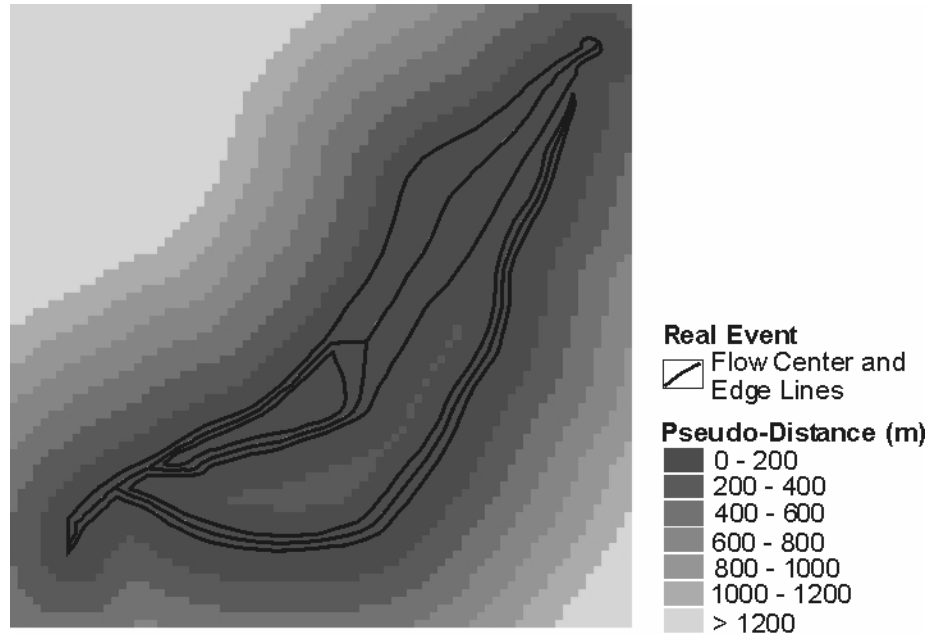
#### 6.2.5.2. Pseudo-Distance

The pseudo-distance is given by the sum of distance to the center and distance to the edge of flow footprint. The center of the flow is the medial axis of the flow footprint, referred to also as skeleton lines (van der Poorten and Jones 2002). Skeleton lines were created using Delaunay triangulation of the points along the edges of the flow. The line formed by connecting the points on the middle of triangle edges that are not on the flow edge is a skeleton line. Figure 6.11 presents the elements used to calculate the pseudo-distance for the Colima event.



**Figure 6.11** Colima event flow footprint, Delaunay triangulation of flow edge points, and flow medial axis (skeleton line).

Pseudo-distance calculated from distance to the flow footprint edges and center is presented in Figure 6.12.



**Figure 6.12** Pseudo-distance calculated from distance to flow footprint edges and center.

#### 6.2.5.3. Performance Measure of Simulation Using Current Implementation

Pseudo-distance (presented in Figure 6.12) and maximum pile height from simulation (presented in Figure 6.10) were sampled at regular intervals (every 50 meters in **X** and **Y** directions) to provide values of independent variables for the logistic regression given by (5.5). Samples located at more than 500 meters from the edges of the footprint were not used in the regression to avoid bias towards locations where pile heights are too low. The bias exists due to the shape of the flow footprint which leads to a large portion of the rectangular sample area to include locations with close to zero pile heights.

The regression analysis sample size was 3,637, with 724 samples inside the flow footprint (20% of total samples). Table 6.4 presents the logistic regression model using samples of pseudo-distance and maximum pile height as the independent variables, and the probability of being inside the flow footprint as the dependent variable (from Eq. 5.5).

**Table 6.4** Logistic regression model for performance measure of simulation at Colima using the current implementation.

Variable	$\beta$	Significance	$\text{Exp}(\beta)$
Pile Height	2.192	0.000	8.952
Pseudo Distance	-0.012	0.000	0.988

Performance measure is given the odds of a location being inside the flow footprint, that is, the exponential of  $\beta$ . Therefore, for the simulation at Colima using the current implementation, the performance is 8.952.

#### **6.2.5.4. Performance Measure of Simulation Using New Implementation**

Pseudo-distance and maximum pile height from simulation were sampled at the same regular intervals as in the previous section to provide values of independent variables of the logistic regression. Furthermore, the same criterion to exclude samples based on distance to flow footprint edges was applied.

Therefore, the sample size and number of samples inside the flow footprint were equal to the previous section--a sample size of 3,637 with 724 samples inside the flow footprint. Table 6.5 presents the logistic regression model using samples of pseudo-distance and maximum pile height as independent variables, and the probability of being inside flow footprint as dependent variables.

**Table 6.5** Logistic regression model for performance measure of simulation at Colima using new implementation.

<b>Variable</b>	<b><math>\beta</math></b>	<b>Significance</b>	<b><math>\text{Exp}(\beta)</math></b>
<b>Pile Height</b>	2.512	0.000	12.334
<b>Pseudo Distance</b>	-0.012	0.000	0.988

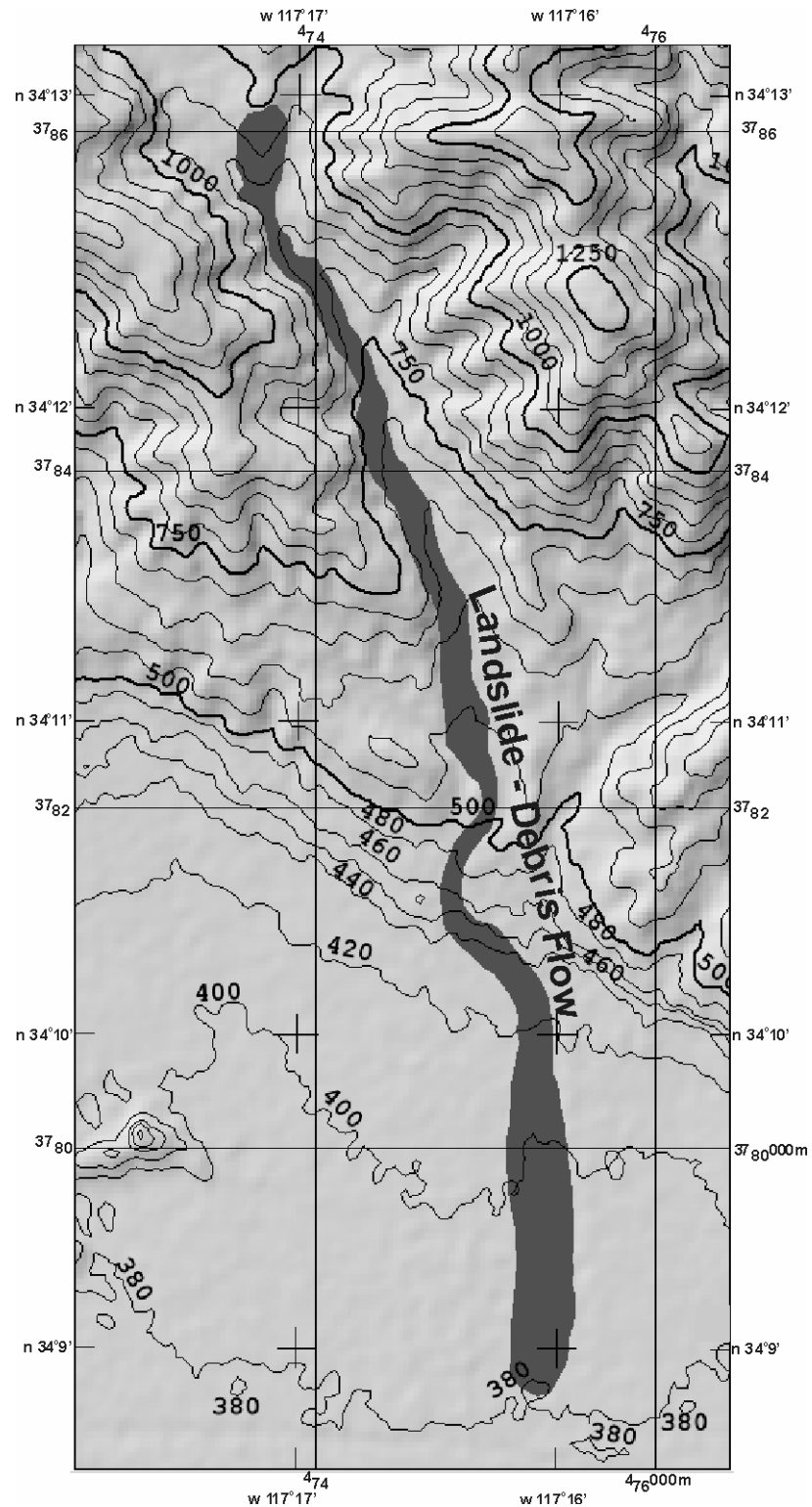
Table 6.5 shows that the performance measure for the simulation at Colima using the new implementation is 12.334. Therefore, the use of the new implementation yields a better performance for simulation of block-and-ash flow than using the current implementation. This result suggests that the use of multiple representations of elevation increases the performance of the simulation of the event which occurred at the Colima Volcano.

### **6.3 SIMULATION OF SAN BERNARDINO DEBRIS FLOW EVENT**

#### **6.3.1. Event Description**

On December 25, 2003 at around 13:00 hours, Pacific Standard Time, a debris flow was triggered by 24 hours of heavy rain and flowed along a secondary canyon and then into Waterman Canyon, killing 14 people and destroying two bridges before stopping at about 3.5 miles north of the center of San Bernardino City in California (Angel et al. 2003).

Figure 6.13 presents the extent of the San Bernardino County debris flow. The affected areas were mapped by the California Office of Emergency Services (2006).



**Figure 6.13** Extent of San Bernardino County debris flow debris flow event which occurred on December 25, 2003. Adapted from sketch map by the California Office of Emergency Services (2006).

### **6.3.2. Parameters for Simulation**

The pile parameters for the simulation of the San Bernardino debris flow used here were obtained from a study to define the stopping criteria for a geophysical mass flow (Yu et al. 2005). The stopping criteria are needed given that the numerical simulation will continue to compute results even when the real event would have stopped due to the finite size of the flow materials.

Since Yu et al. (2005) do not provide the friction angles that best match the field data, their values were arbitrarily selected. Furthermore, Yu et al. (2005) define two initial piles. For this example of multiple representation usage, only one of them is selected and used for analysis.

Although the simulation results are compared to field data, since this study intends to demonstrate the differences between uses of elevation representations, the influence of the pile parameters in the comparison is not an issue. The parameters of the piles are presented in Table 6.6.

**Table 6.6** Pile parameters for simulation of San Bernardino County debris flow event which occurred on December 25, 2003. Coordinates are in UTM projection meters, Northern hemisphere, zone 11, datum NAD83. The difference between piles is on their starting location.

<b>Parameter</b>		<b>Value</b>
<b>Volume (cubic meters)</b>		0.31x10 <sup>6</sup> from a paraboloid defined by 20 meters maximum thickness and 100 meters long major and minor axis
<b>Internal friction angle (degrees)</b>		22.8
<b>Basal friction angle (degrees)</b>		8.21
<b>Starting location</b>	<b>Pile 1 coordinate</b>	474,130 East, 3,787,860 North
	<b>Pile 2 coordinate</b>	475,160 East, 3,786,730 North
<b>Simulation region</b>	<b>Lower left coordinate</b>	471,000 East, 3,776,990 North
	<b>Upper right coordinate</b>	476,790 East, 3,789,500 North

### 6.3.3. Simulation Results Using Current Implementation

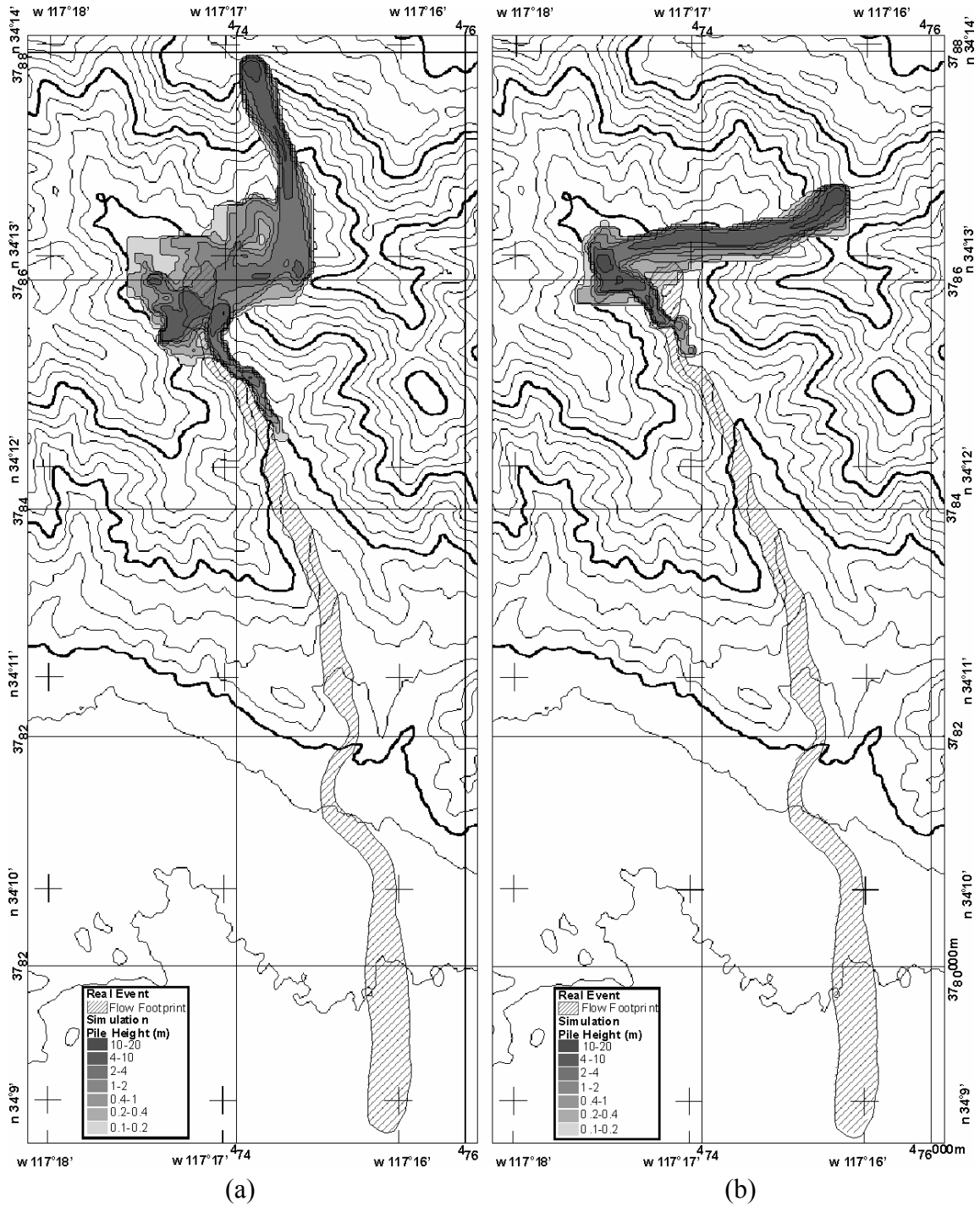
The current implementation of TITAN2D was used to simulate the debris flow event at San Bernardino using the elevation and derivatives extracted from a DEM derived SRTM and the parameters of Section 6.3.1. The computational parameters of the simulation are presented in Table 6.7.

**Table 6.7** Computational parameters for simulation of San Bernardino County debris flow.

<b>Parameter</b>	<b>Value</b>
<b>Number of processors</b>	1
<b>Number of points in the computational mesh</b>	300
<b>Use adaptative grid</b>	Yes
<b>Scale governing equations</b>	Yes
<b>Scale length (meters)</b>	4000
<b>Solution method</b>	First order
<b>Maximum number of time steps</b>	8000
<b>Maximum time (seconds)</b>	4000
<b>Time between output of results (seconds)</b>	1

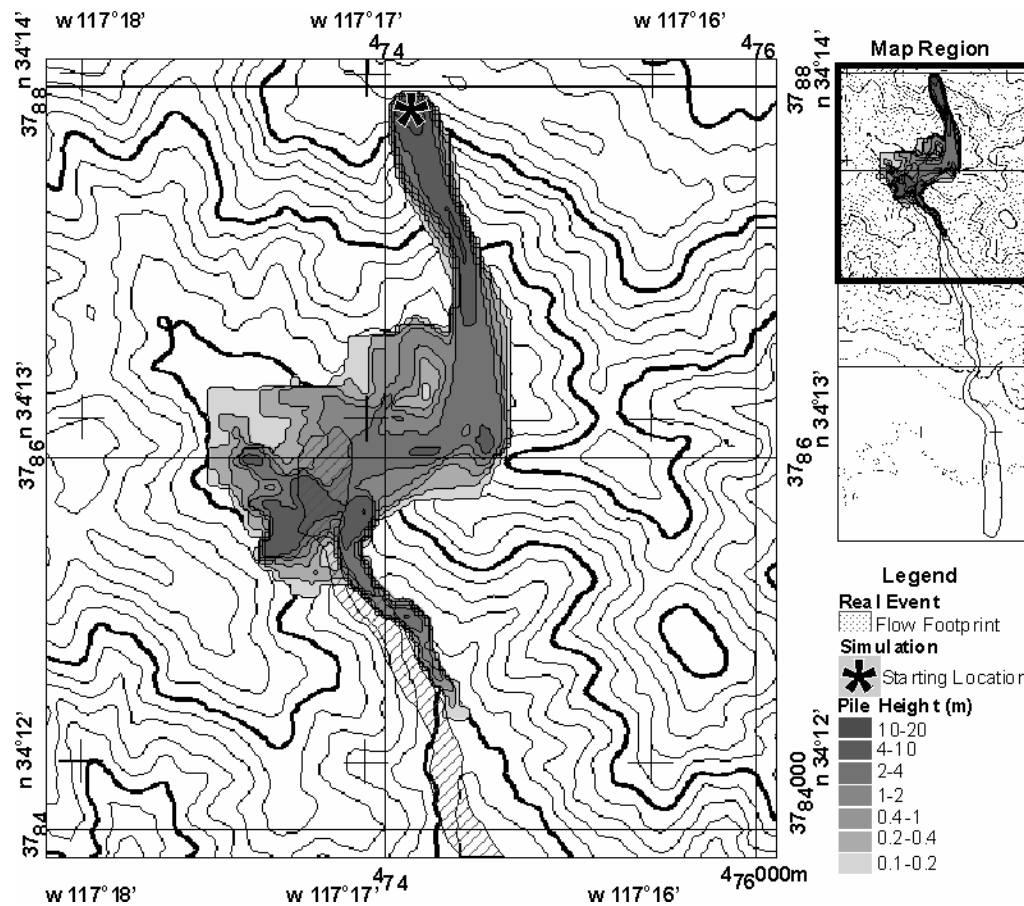
Summaries of simulation results were generated using the procedure presented in Section 5.2.2.2 and are presented in Figure 6.14.





**Figure 6.14** Summary of results from simulation San Bernardino County debris flow using current implementation of TITAN2D and: starting Pile 1 (a); and starting Pile 2 (b). Pile height values are the maximum values achieved during simulation at each location.

Pile 1 was selected for analysis given that it extends almost 1000 meters more than the simulation using Pile 2, which indicates a better starting location. However, both simulations reach only a small portion of the extent of the debris flow mapped by the California Office of Emergency Services. Figure 6.15 presents the results from simulation using Pile 1 at the portion intersecting the extent of the mapped flow.

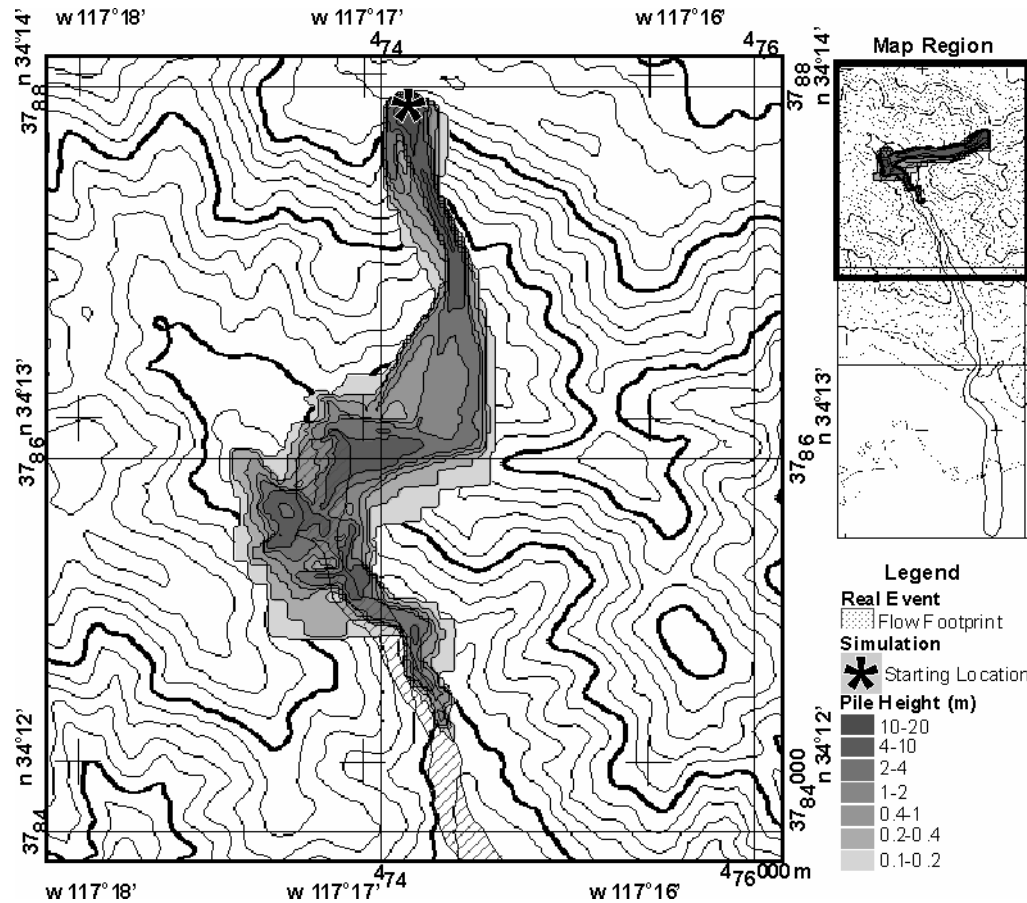


**Figure 6.15** Summary of results from simulation San Bernardino County debris flow using current implementation of TITAN2D and starting Pile 1. Map shows the region where simulation flow intersects the extent of the mapped flow. Pile height values are the maximum values achieved during simulation at each location.

#### **6.3.4. Simulation Results Using New Implementation**

The new implementation of TITAN2D simulated the debris flow event at San Bernardino with the same parameters used for simulation using the current version of TITAN2D and Pile 1. The elevation and derivatives were extracted from the multiple representations framework, accessing DEMs stored in the TerraLib database. The available DEMs were the SRTM and the NOAA-CSC described in Section 6.1.

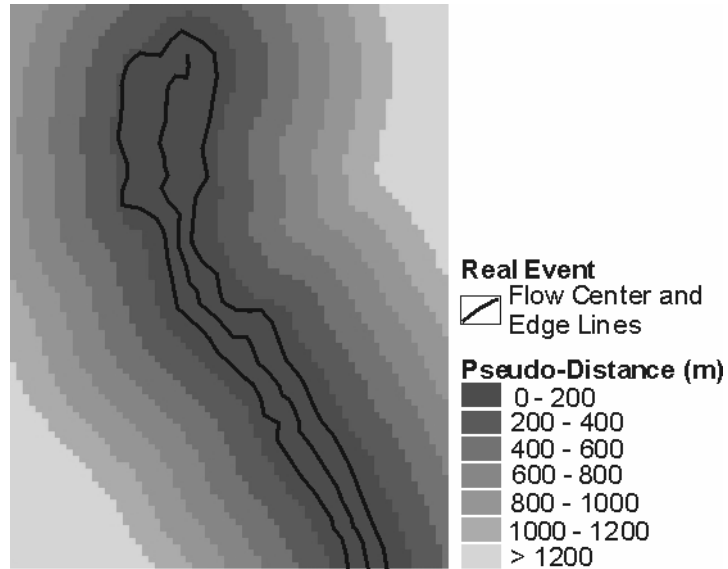
The summary of the results was generated using the procedure presented in Chapter 5 and is presented in Figure 6.16.



**Figure 6.16** Summary of results from simulation San Bernardino County debris flow using new implementation of TITAN2D and starting Pile 1. Map shows the region where simulation flow intersects the extent of the mapped flow. Pile height values are the maximum values achieved during simulation at each location.

### 6.3.5. Comparison of Simulation Results at San Bernardino

The results from the simulations were compared to the field work data using the method presented in Chapter 5 and used in Section 6.2 for Colima block-and-ash flow simulation. The flow footprint was obtained from the sketch map presented in Figure 6.13. The center line was created using Delaunay triangulation. The pseudo-distance calculated by adding the distance to the center of the flow footprint and the distance from the edge of the footprint is presented in Figure 6.17.



**Figure 6.17** Pseudo-distance calculated from distance to center and distance from edge of flow footprint.

#### **6.3.5.1. Performance Measure of Simulation Using Current Implementation**

In a similar way to the performance measure for simulation at Colima, pseudo-distance and maximum pile height from simulation (presented in Figure 6.15) were sampled at regular intervals (every 50 meters in the **X** and **Y** directions). Also, the same criterion to consider only samples located at less than 500 meters from the flow footprint was used. In addition, only samples inside the area shown in Figure 6.15 were used since the simulated flow covers only the north most portion of the flow footprint mapped by California Office of Emergency Services (presented in Figure 6.13).

The values of the independent variables from pseudo-distance and pile height and the dependent variable probability of being inside the flow footprint were analyzed by logistic regression given in (5.5). The regression analysis sample size was 5,242, with 877 samples inside the flow footprint (17% of total samples). Table 6.8 presents the logistic regression model.

**Table 6.8** Logistic regression model for performance measure of simulation at San Bernardino using the current implementation.

<b>Variable</b>	<b><math>\beta</math></b>	<b>Significance</b>	<b>Exp(<math>\beta</math>)</b>
<b>Pile Height</b>	0.495	0.000	1.640
<b>Pseudo Distance</b>	-0.016	0.000	0.984

The performance measure for the simulation using the current implementation, given by the exponential of  $\beta$ , is 1.640.

#### **6.3.5.2. Performance Measure of Simulation Using New Implementation**

The same procedure used in the previous section was applied to the results of the simulation at San Bernardino using the new implementation. Table 6.9 presents the logistic regression model.

**Table 6.9** Logistic regression model for performance measure of simulation at San Bernardino using the new implementation.

<b>Variable</b>	<b><math>\beta</math></b>	<b>Significance</b>	<b>Exp(<math>\beta</math>)</b>
<b>Pile Height</b>	0.529	0.000	1.698
<b>Pseudo Distance</b>	-0.017	0.000	0.983

The performance measure for the simulation at San Bernardino using the new implementation is 1.698. Therefore, in a similar way to the simulation of the block-and-ash flow at Colima Volcano, the use of new implementation yields a better performance for simulation of debris flow at San Bernardino County than using the current implementation.

## 6.4 SUMMARY

The multiple representations of elevations framework was validated through comparison of results of geophysical mass flow simulations using TITAN2D. The current implementation of TITAN2D uses only one representation of elevation for a simulation while the new implementation uses multiple representations for the same simulation.

Results from the simulation of the block-and-ash flow which occurred at Colima Volcano in April 1991 indicates that the use of multiple representations yields results that are more similar to the footprint of the flow surveyed by the field work than results achieved using one representation. In this simulation, representations of elevation and their derivatives originated from SRTM and ARIA DEMs are selected to provide values for simulation based on the best one at every time step and at every location where the flow runs. The definition of best representation is based on a ranking system that considers resolution of the representation and the date and time of the measurement.

Results for the San Bernardino debris flow simulation also indicates that the use of multiple representations yields better results. The representations of elevation and their derivatives originated from the SRTM and NOAA-CSC DEMs.

Although the performance of simulations using the new implementation is better when compared to simulations using the current implementation, it must be noted that the flow footprints used in the performance measurements may not represent the real event footprint. A geophysical mass flow is expected to follow the channels and the footprint of Colima block-and-ash flow shows that the flow ran on sides of a channel in some areas. The footprint of San Bernardino debris flow also indicates that the flow ran on sides of

the canyon. However, even with the problem with the flow footprint, the use of multiple representations of elevation has the potential to improve simulations of geophysical mass flows because more than one representation is available and representations with different resolution and acquisition date and time would allow a better match with the requirements of the event simulation.

In addition to the simulations, a computational tool to import DEMs into the TerraLib database was developed and presented in this chapter. The tool inserts the representation of elevation and creates the representation information used for querying the database. The representation information is created from a custom metafile associated with each imported DEM.



# **Chapter 7**

## **CONCLUSIONS**

This dissertation presents an integrated framework to use multiple representations of elevation for simulation of environmental process. In order to build the framework, a study of models was presented. Innovative approaches were used to define models using notions from Ontology which facilitated the categorization of models into models of objects and models of processes. The representations of elevation are models of object and representations of environmental processes are models of process.

The reasons for the existence of multiple representations of elevation were presented by analyzing the modeling procedure, i.e. the method used to create a model of elevation. Next the framework to store information about modeling and the representations of elevation in a spatial database was developed. The framework allows queries of values of elevation using restrictions based on the modeling information.

With the availability of more than one representation of elevation, methods to take advantage of and to solve problems that arise from this extra information were developed. They include a method to create spatially distributed accuracy information based on cluster analysis thereby improving the quality of accuracy information from the existing global measures of accuracy.

When simulation is executed in a multiple representations setting, its results differ for each representation or combination of representations used. Thus a measure of

performance of each simulation is necessary. The proposed method is adequate for dynamic spatially distributed process models since the simulated events have far less information than the results of simulation. Logistic regression is used to assess how well the footprint of the real event matches the numerical results of simulation.

The other problem arises when a process model queries representations for a value and more than one representation can provide the value. A ranking system was proposed, with ranking criteria defined according to the characteristics of the process model and of the simulated event.

The framework for the use of multiple representations of elevation for simulation of environmental process models was implemented to provide values of slope and curvature to a geophysical mass flow model. A comparison of results using the framework and using one representation of elevation was executed for the simulation of two events. The comparison shows that the use of multiple representations of elevation improves the performance of the simulations.

A summary of the topics of the dissertation is given next, followed by the additional features that could extend the use of the framework.

## **7.1 MODELS**

Models were defined as simplified versions of reality entities. Reality entities are classified using ideas from Ontology, in particular from the Basic Formal Ontology. BFO is an information systems ontology that structures reality entities by using the SNAP and the SPAN sub-ontologies.

Endurants of SNAP are objects and perdurants of SPAN are processes. Consequently, models are classified into model of objects and model of processes. Objects have continuous existence although changing with time. Processes occur in space and time requiring a temporal sequence for their existence. SNAP and SPAN sub-ontologies are connected by trans-ontological relations, such as participation. In the participation relation, an object participates in a process.

### **7.1.1. Models of Processes**

Models of processes were classified based on how the internal structure of the model emulates a process. Rule-based process models define the state of the next step of the process by a logical combination based on the current state of the process. Empirical process models deduce the process mechanism through regression analysis. Deterministic process models use mathematical equations defined from established theories to describe the evolution of a process. Stochastic process models use equations that incorporate the stochastic behavior of a process.

Environmental processes are processes related to the Earth's environment that occur on a geographical scale. Models for environmental processes may consider values and states homogeneous inside a spatiotemporal region or consider them to vary inside the region. In the first case, the models are defined as lumped environmental process models. Since variation is considered in the second case, the models are defined as spatially distributed environmental process models.

### **7.1.2. Models of Objects**

Objects are related to processes through the participation relation. Objects are enduring entities and are classified in substances, boundaries, fiat parts, aggregates, and sites. In addition, there are dependent objects, i.e. they depend on independent objects for their existence. Dependent entities include qualities, functions, roles, and conditions.

Geographic objects cover all types of enduring entities. They are substances, boundaries of substances, fiat parts of substances, aggregates of other geographic objects, sites, and dependent entities. For example, elevation is a dependent geographic object of the substance Earth.

### **7.1.3. Modeling Processes**

A process model has names of participant endurants, descriptors of endurants properties, mathematical or logical equations, and interpretations of the model. The techniques for modeling processes were not discussed given that the focus is on the objects.

### **7.1.4. Modeling Objects**

Modeling of geographic objects consists of making measurements and structuring these measurements for queries and manipulation, thus creating a model. The object models that are analyzed in this study are restricted to models in digital format. The word representation was defined to refer to an object model in digital format.

The steps of modeling follow abstraction levels from reality to representation. The abstraction levels are the ontological, conceptual, representation, and implementation levels.

## **7.2 MULTIPLE REPRESENTATIONS OF ELEVATION**

Of the various representations of geographic objects, this study selected representations of elevation due to the importance of elevation in environmental studies. Using the ideas from SNAP ontology, elevation is an enduring entity that depends on the enduring substance Earth for its existence. Elevation is related to environmental processes through the trans-ontological relation participation since Earth's gravitational force drives movements of mass present in processes.

A detailed analysis of the restrictions imposed by digital representations to represent geographic data revealed that the digital world imposes representations to be finite and discrete. Therefore, representations have a minimum distance between geographic locations, a minimum difference between values of representations, a maximum size of study region, and a maximum amount of data in a representation.

Multiple representations exist due to different choices made during the modeling procedure. The current approaches to handle multiple representations are concerned with integration of representations from diverse spatial databases and with creating maps from different scales.

Choices made for modeling elevation at the ontological level include the definition of the elevation surface, bare Earth or Earth with vegetation and man-made

structures, and requirements for data collection, such as level of detail. Since measurements are made at the conceptual level, different dates and times, and different measuring techniques lead to different representations. At the representation level, the use of different vector and raster geometric representations leads to the diversity of representations of elevation, such as contour lines and surface networks. The implementation level choices are related to storage structures and manipulation algorithms and, since there are various options, each option creates a different representation of elevation.

### **7.2.1. Issues Related to Multiple Representations of Elevations**

By acknowledging the existence of multiple representations of elevation, additional analysis and manipulations become possible. They include fusion of representations, quality analysis, and the definition of the best representation for a given application.

The purpose of fusion is to integrate multiple representations to produce a new representation of elevation. Some of the current techniques include fusion using weighted interpolation and the use of conditional probability.

The quality of a representation of elevation can be defined using multiple representations to provide the values of elevation for comparison. In addition, when elevation is defined as a random variable, each representation of elevation provides a sample; therefore, the random variable elevation can be defined by descriptive statistics.

The issue of best representation depends on the application of elevation, and the question is: which representation of elevation is the most fit-for-use among the available representations for a given application. Each representation has a resolution, measuring technique, and date and time of measurement. These factors must be considered to define the best representation based on the specifications given by the simulated process model.

### **7.3 FRAMEWORK FOR INTEGRATION OF MULTIPLE REPRESENTATIONS OF ELEVATION AND PROCESS MODELS**

Given that there are multiple representations of elevation, a framework that integrates multiple representations to process models was presented. The targeted types of process models are the dynamic spatially explicit process models. These models are the environmental process models where values are spatially distributed and the term dynamic is used to emphasize the explicit time dependency of the process model. Therefore, equations for this type of process model include derivatives of time and space.

The values of spatial entities are provided by Geographic Information Systems (GIS). Of the available methods to link GIS to process models, the tight coupling was selected since values are directly accessed from representations stored in GIS. The other types of integration (embedded and loose coupling) are not adequate for the computationally intensive requirements of dynamic spatially explicit process models.

The choice of which GIS to use was driven by the need to access implementations of spatial algorithms. Since only open source software provides this capability, TerraLib, an open software GIS, was selected. TerraLib provides functions to access

representations stored in relational databases. When the database management system handles spatial representations, TerraLib uses the database functions. Otherwise, TerraLib uses Binary Long Object (BLOB) type to store representations in the database.

In addition to the representation of elevation stored in TerraLib standard way, metadata about the representation is stored using a database table and the spatial capabilities of TerraLib. This representation of information contains the area where valid values exist in the representation of elevation and values describing modeling procedure including date and time of the measurement, techniques used during measurement and production, and accuracy of the representation.

The database tables that were designed in this framework allow queries using Standard Query Language (SQL) with spatial extensions. The queries are used as a basis for linkage with process models.

The specific dynamic spatially explicit process model linked to the GIS is a geophysical mass flow model, TITAN2D. The computational implementation of TITAN2D uses a parallel adaptative mesh Godunov scheme, using values of local bed friction angle, internal friction angle, local surface slope and curvature. The adaptative mesh characteristic of TITAN2D requires spatial values at different resolutions. In addition, since the process model is used to simulate an event, the time and date of the simulated event is considered.



### 7.3.1. Implementation Optimizations

Optimizations were implemented to support the computationally intensive characteristic of TITAN2D. They include a cache of past queries and prediction of future queries. The cache of past queries provides fast access to a query result if the same query was previously executed by avoiding database query. The cache is implemented by the local storage of the pair query and its result. Since TITAN2D uses an adaptative grid, it is possible to predict that locations near a current query will be queried at the same resolution during the computation. The prediction is implemented by the same cache through the storage of the whole representation instead of only the value at the location requested by the query.

In addition to optimizations, the implementation creates representations derived from elevation. The solution of flow equations in TITAN2D requires values of slope and curvature. These values are provided by the implementation through the generation of representations of derivatives of first and second order from the representation of elevation. The new representations, computed using finite difference method, are stored in the database as new representations, inheriting metadata from the representation of elevation. If the same database is used for another flow simulation, the new representations will provide the slope and curvature values without having to derive them from the representation of elevation.

### **7.3.2. Ranking Representations**

A solution to select the best representation for a query is used in the implementation. The selection is necessary since a query may be satisfied by more than one representation of elevation. The solution is based on the process model requisites of resolution and of time and date of the event. Representations satisfying the query receive a value in a ranking system. The highest ranked representation of elevation is selected to provide the queried values.

Representations acquired before the query date and time constraint are ranked higher, with ranking scores getting lower as the time difference becomes larger. When the acquisition time and date is just after the query time, the rank is lower than the rank for just before. In addition, the ranking scores of representations acquired after the query time decrease faster as the difference gets larger when compared to representations acquired before. In addition, rank values are given based on resolution of the representation. Representations with resolution values that are closer to the queried resolution are ranked higher than other representations. The overall rank of a representation is the sum of the rank based on resolution and on time and date.

To apply this to other process models a more appropriate ranking system must be defined and used. The ranking system may include any of the information available for representation. For example, accuracy can be used to define the rank of a representation.

## **7.4 ADDITIONAL METHODS FOR MULTIPLE REPRESENTATIONS OF ELEVATION**

Multiple representations of elevation allow analyses of elevation that were not possible previously. At the same time, they impose the need for additional tools to handle results from the use of multiple representations of elevation. In the first category, a tool to analyze representations of elevation to extract spatially distributed accuracy of a representation was presented. For the second type of tools a methodology to estimate the performance of a simulation of a dynamic spatially distributed process models was presented.

### **7.4.1. Spatially Distributed Accuracy**

The proposed method to estimate the spatially distributed accuracy of a representation is based on defining elevation as a random variable described by a normal distribution with a mean and a standard deviation. A given representation could have the accuracy at every location defined by the standard score (Z-score), i.e. by how many standard deviations the value is away from the mean.

However, elevation is spatially autocorrelated and accuracy can be more realistically defined using cluster analysis. The objective of cluster analysis is to find regions where values are higher statistically than the rest of the region. The method proposes to associate values of accuracy for the statistically significant regions of high values by averaging the accuracy inside these regions.

### **7.4.2. Performance Measure**

Results of simulation of processes vary for each representation or combination of representations used. The existing qualitative and the quantitative measures of simulation performance are not suited for the spatially distributed dynamic processes. Therefore, a new quantitative measure of simulation performance that is adequate for geophysical mass flows was developed.

The method is based on using a footprint of the areas affected by the flow in the real event as the dependent variable of a logistic regression analysis. The dependent variables of the regression are the maximum pile height of simulation at all times and a pseudo-distance to the flow footprint center line. The pseudo-distance is defined by the sum of distance to center line and the distance to footprint edges, thus the distance increases more rapidly outside the footprint.

The performance measure is given by the ratio between the parameter associated with pile height and the parameter associated with pseudo-distance in the logistic regression. Logistic regression was selected given that it correlates the qualitative information from the real event footprint with values of pile height from the simulation.

## **7.5 VALIDATION OF THE MULTIPLE REPRESENTATIONS OF ELEVATION FRAMEWORK**

The framework for the using multiple representation of elevation for process models was validated through simulation of geophysical mass flows. The simulated

events were the block-and-ash flow event which occurred at Colima Volcano in Mexico and a debris flow that occurred at San Bernardino County in California.

DEMs for the simulation were imported into the TerraLib database using a tool implemented to facilitate the creation of the database tables. Information about the modeling of the imported representation is inserted into the appropriate tables by importing a metadata file. The values of elevation are imported from a text file, binary file, or a Tag Image File Format. In addition, the tool handles the cartographic projection of representations.

The simulated event at Colima occurred during April 16-17, 1991. The DEMs available were the SRTM DEM at 3-arcsecond resolution and the ARIA DEM at 60-meter resolution. The parameters of the simulation were obtained from a study for validation of TITAN2D. The comparison of results of the simulation indicates that the use of multiple representation yields results more similar to the field work footprint than the results achieved with a single representation, when using the performance measure tool.

The validation at San Bernardino County simulated a debris flow that occurred in San Bernardino County, California, on December 25, 2003. The DEMs available for the area are from SRTM DEM at a 1-arcsecond resolution and NOAA-CSC DEM at a 3 meter resolution. Parameters for this simulation are only available from a study to define the stopping criteria for the simulation. Although the parameters may not be the most adequate ones, results from simulation indicate that the performance of simulation using

multiple representations is better than the results from simulation using only the SRTM DEM.

## **7.6 EXTENSIONS TO PROPOSED FRAMEWORK**

### **7.6.1. Other Geographic Objects**

Although the study focused on implementing the framework to handle representations of elevation it can be easily extended to include other geographic objects. For example, representations of land use that vary with time and can be obtained through different techniques can use the same structure developed for elevation. A land use map has resolutions defined by the mapping technique and will vary with time and date of acquisition, being similar to elevation.

### **7.6.2. Storage and Use of Samples, Contour Lines, and TINs**

In addition, support to use other types of representations of elevation besides a raster grid can be implemented. In such implementations, elevation values could be obtained from contour lines, TINs, or any other representation. For example, if elevation is requested at a location on a contour line, the query answer is given by the value associated with the line. If the location is between lines, then an interpolation procedure could be applied and the value estimated. In this case, the query can select which interpolator is more appropriate.

Furthermore, the proposed framework is flexible enough to support requirements from applications using values from the representations. The efficiency of the framework is obtained by the ability to generate derived representations only when requested by a query. The existing solutions require the generation a-priori of the representations that are expected to be necessary. Since the derivation can be executed only locally, the framework provides support to avoid the generation of values at locations that will not be queried.

### **7.6.3. Validation and Storage of Simulation Results**

The use of storage in spatial databases with associated representation information allows for the handling by the GIS of new representations created by a simulation. For example, in the case of geophysical mass flows, elevation will be different for the simulated region at the end of simulation, and a new representation of elevation can be inserted in the database. The new representation of elevation would have metadata indicating that it is the result of simulation and that it was measured at the time and date when simulation stopped. The accuracy of the new representation would be low, but once the result is validated, the accuracy could be higher.

### **7.6.4. Fusion of Representations of Elevation**

A possible solution for a query that can not be satisfied by a representation of elevation is to create a new one from the available representations. This solution requires the implementation of a fusion method. Fusion methods include simple linear

interpolations between representations to more complex methods that consider the process that accounts for the difference in representations. The simple linear interpolation could be used when only one feature of the representations is not equal and the process that changes elevations is unknown. For example, if representations' measurement times and dates differ, the values at a location can be defined by linear interpolation between values of the representations. If the process that changes elevation with time is known, the model of the process can be used to define a more accurate value of elevation from the values of the representations.

## **7.7 SUMMARY**

The major contribution of this dissertation is the ability provided by the proposed framework for a more effective use of existing representations of elevation. The focus was on the use of the framework for simulation of dynamic spatially distributed process models although the framework can be used for other applications that require values of multiple representations of geographic objects.

In addition, a method to create spatially distributed accuracy for representations of elevation was presented. This accuracy measure provides a better description of the representation than the existing global measures. Furthermore, spatially distributed accuracy is useful in studies involving quantification of the influence of the accuracy in results derived from representations of elevation.

Another contribution is the method to measure performance of a simulation. It replaces qualitative measures with a quantitative one, eliminating bias and personal



interpretation of the qualitative approaches. The proposed quantitative method is also suitable for sensitivity analysis of simulation parameters.

In relation to ranking representation of elevations when more than one satisfies a query, this dissertation proposes that the process model should define the ranking system. The proposed ranking system is adequate only for the simulation of geophysical mass flows and combines ranks due to resolution and to time and date of elevation measurement.

Finally, the proposed solution uses open software so that modelers can use the implementation to link multiple representations of elevation to their process model. In addition, the framework implementation can be customized to accommodate specific requirements of the process model. When the simulation is used for decision-making the implementation can be inspected to assure that there are no errors or bias toward a solution.



# REFERENCES

- Adams, J. C. and J. H. Chandler (2002). "Evaluation of lidar and medium scale photogrammetry for detecting soft-cliff coastal change." Photogrammetric Record **17**(99): 405-418.
- Agarwal, P. (2005). "Ontological considerations in GIScience." International Journal of Geographical Information Science **19**(5): 501–536.
- Angel, W., S. Hinson and R. Herndon, Eds. (2003). Storm Data. 45(12). Asheville, NC, National Climatic Data Center.
- Balley, S., C. Parent and S. Spaccapietra (2004). "Modelling geographic data with multiple representations." International Journal of Geographical Information Science **18**(4): 327-352.
- Bittner, T. and B. Smith (2003). Granular Spatio-Temporal Ontologies. 2003 AAAI Symposium: Foundations and Applications of Spatio-Temporal Reasoning (FASTR), AAAI Press.
- Brimicombe, A. (2003). GIS, Environmental Modelling and Engineering. London, England; New York: Taylor and Francis.
- Brown, D. G. and T. J. Bara (1994). "Recognition and Reduction of Systematic-Error in Elevation and Derivative Surfaces from 7-1/2-Minute Dems." Photogrammetric Engineering and Remote Sensing **60**(2): 189-194.
- Burrough, P. A. (1998). "Dynamic Modelling and Geocomputation," Geocomputation: a primer, eds. P. A. Longley, S. M. Brooks, R. McDonnel and B. Macmillan. Chichester, England: John Wiley and Sons.
- Burrough, P. A., R. v. Rijn and M. Rikken (1996). "Spatial Data Quality and Error Analysis Issues: GIS Functions and Environmental Modeling," GIS and Environmental Modeling: Progress and Research Issues, eds. M. F. Goodchild, L. T. Steyaert, B. O. Parks, C. Johnston, D. Maidment and M. Crane. New York: John Wiley and Sons.
- Buttenfield, B. P. and T. Hultgren (2005). Managing Multiple Representations of "Base Carto" Features: A Data Modeling Approach. Auto-Carto 2005.

- California Office of Emergency Services. (2006). "Waterman Canyon with Pre and Post Fire special Flood Hazard Areas." Retrieved 06/20/2006, 2006, from <http://www.oes.ca.gov/Operational/OESHome.nsf/978596171691962788256b350061870e/1AECEB42C3A24B87C88256E0D0079A036?OpenDocument>.
- Câmara, G. (1995). "Modelos, linguagens e arquiteturas para bancos de dados geográficos". Ph.D. diss., Applied Computing, INPE.
- Camara, G. and F. Fonseca (2006). "Information Policies and Open Source Software in Developing Countries." Journal of the American Society for Information Science and Technology: accepted.
- Câmara, G. and H. Onsrud (2004). "Open Source GIS Software: Myths and Realities," Open Access and the Public Domain in Digital Data and Information for Science: Proceedings of an International Symposium, eds. J. M. Esanu and P. F. Uhler. Washington: The National Academies Press.
- Câmara, G., R. C. Souza, B. M. Pedrosa, L. Vinhas, A. M. V. Monteiro, J. A. Paiva, M. T. Carvalho and M. Gattass (2000). TerraLib: Technology in Support of GIS Innovation. 2nd Brazilian Symposium on GeoInformatics, São Paulo, Brazil.
- Carlisle, B. H. (2005). "Modelling the Spatial Distribution of DEM Error." Transactions in GIS **9**(4): 521-540.
- Carter, J. R. (1988). "Digital Representations of Topographic Surfaces." Photogrammetric Engineering and Remote Sensing **54**(11): 1577-1580.
- Chorley, R. and P. Haggett, Eds. (1967). Models in Geography. London, England: Methuen.
- Clarke, S. and K. Burnett (2003). "Comparison of digital elevation models for aquatic data development." Photogrammetric Engineering and Remote Sensing **69**(12): 1367-1375.
- Couclelis, H. (1992). "People manipulate objects (but cultivate fields): beyond the raster-vector debate in GIS," Theories and methods of spatio-temporal reasoning in geographic space, eds. A. U. Frank, I. Campari and U. Formentini. Berlin, Germany: Springer.
- Couclelis, H. (2002). "Modeling frameworks, paradigms, and approaches," Geographic Information Systems and Environmental Modeling, eds. K. C. Clarke, B. O. Parks and M. P. Crane. Upper Saddle River: Prentice Hall.

- Crisci, G. M., R. Rongo, S. Di Gregorio and W. Spataro (2004). "The simulation model SCIARA: the 1991 and 2001 lava flows at Mount Etna." Journal of Volcanology and Geothermal Research **132**(2-3).
- Daniel, C. and H. Tennant (2001). "DEM Quality Assessment," Digital Elevation Model Technologies and Applications: the DEM Users Manual, ed. D. F. Maune. Bethesda: American Society for Photogrammetry and Remote Sensing.
- DLR. (2003). "Shuttle Radar Topography Mission - SRTM." Retrieved 01/16/2003, 2003, from [http://www.dfd.dlr.de/srtm/index\\_en.htm](http://www.dfd.dlr.de/srtm/index_en.htm).
- Doyle, F. J. (1978). "Digital Terrain Models: An Overview." Photogrammetric Engineering and Remote Sensing **44**(12): 1481-1485.
- Duckham, M. and M. Worboys (2005). "An algebraic approach to automated geospatial information fusion." International Journal of Geographical Information Science **19**(5): 537-557.
- Eckert, S., T. Kellenberger and K. Itten (2005). "Accuracy assessment of automatically derived digital elevation models from aster data in mountainous terrain." International Journal of Remote Sensing **26**(9): 1943-1957.
- Egenhofer, M. J. (1994). "Spatial SQL: a query and presentation language." IEEE Transactions on Knowledge and Data Engineering **6**(1): 86-95.
- Egenhofer, M. J., J. I. Glasgow, O. Günther, J. R. Herring and D. Peuquet (1999). "Progress in Computational Methods for Representing Geographical Concepts." International Journal of Geographical Information Science **13**(8): 775-796.
- El-Sheimy, N., C. Valeo and A. Habib (2005). Digital Terrain Modeling: Acquisition, Manipulation And Applications. Norwood: Artech House Publishers.
- Elmasri, R. and S. B. Navathe (2000). Fundamentals of Database Systems. Reading: Addison-Wesley.
- Falorni, G., V. Teles, E. R. Vivoni, R. L. Bras and K. S. Amaratunga (2005). "Analysis and characterization of the vertical accuracy of digital elevation models from the Shuttle Radar Topography Mission." Journal of Geophysical Research-Earth Surface **110**(F2).

- Fedra, K. (1993). "GIS and Environmental Modeling," Environmental Modeling with GIS, eds. M. F. Goodchild, B. O. Parks and L. T. Steyaert. New York: Oxford University Press.
- Feng, C.-C., T. Bittner and D. M. Flewelling (2004). "Modeling Surface Hydrology Concepts with Endurance and Perdurance," Giscience 2004. Berlin, Germany; Heidelberg, Germany: Springer-Verlag.
- Ferreira, K. R., L. Vinhas, G. R. d. Queiroz, R. C. Souza and G. Câmara (2003). The Architecture of a Flexible Querier for Spatio-Temporal Databases. VII Brazilian Symposium on GeoInformatics, Campos do Jordão, Brazil.
- FGDC (1998). Content Standards for Framework Land Elevation Data. Reston, Virginia, Federal Geographic Data Committee - c/o U.S. Geological Survey: 33.
- Fisher, P. (1998). "Improved Modeling of Elevation Error with Geostatistics." GeoInformatica **2**(3): 215-233.
- Fonseca, F. T., M. J. Egenhofer, P. Agouris and G. Camara (2002). "Using Ontologies for Integrated Geographic Information Systems." Transactions in GIS **6**(3): 231-257.
- Gao, J. (1997). "Resolution and accuracy of terrain representation by grid DEMs at a micro-scale." International Journal of Geographical Information Science **11**(2): 199-212.
- Garbrecht, J. and L. W. Martz (1994). "Grid Size Dependency of Parameters Extracted from Digital Elevation Models." Computers and Geosciences **20**(1): 85-87.
- Geophysical Mass Flow Group (2004). TITAN2D User Guide - Release 1.010105. University at Buffalo, NY.
- Gonzalez, M. B., J. J. Ramirez and C. Navarro (2002). "Summary of the historical eruptive activity of Volcan de Colima, Mexico 1519-2000." Journal of Volcanology and Geothermal Research **117**(1-2): 21-46.
- Goodchild, M. F. (1992). "Geographical Data Modeling." Computers and Geosciences **18**(4): 401-408.
- Goodchild, M. F. (1996). "The Spatial Data Infrastructure of Environmental Modeling," GIS and Environmental Modeling: Progress and Research Issues, eds. M. F. Goodchild, L. T. Steyaert, B. O. Parks, C. Johnston, D. Maidment, M. Crane and S. Glendinning. Fort Collins: GIS World Books.

- Goodchild, M. F. and P. Longley (1999). "The Future of GIS and Spatial Analysis," Geographical Information Systems, eds. P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind. New York: John Wiley & Sons.
- Grenon, P. and B. Smith (2004). "SNAP and SPAN: Towards Dynamic Spatial Ontology." Spatial Cognition and Computation **4**(1): 69-104.
- Guptill, S. C. (1999). "Metadata and Data Catalogues," Geographical Information Systems, eds. P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind. New York: John Wiley & Sons.
- Hestenes, D. (1987). "Toward A Modeling Theory Of Physics Instruction." American Journal of Physics **55**(5): 440-454.
- Hill, J. M., L. A. Graham and R. J. Henry (2000). "Wide-area topographic mapping and applications using airborne light detection and ranging (LIDAR) technology." Photogrammetric Engineering and Remote Sensing **66**(8): 908–909, 911–914, 927, 960.
- Hodgson, M. E., J. Jensen, G. Raber, J. Tullis, B. A. Davis, G. Thompson and K. Schuckman (2005). "An evaluation of lidar-derived elevation and terrain slope in leaf-off conditions." Photogrammetric Engineering and Remote Sensing **71**(7): 817-823.
- Hodgson, M. E., J. R. Jensen, L. Schmidt, S. Schill and B. Davis (2003). "An evaluation of LIDAR- and IFSAR-derived digital elevation models in leaf-on conditions with USGS Level 1 and Level 2 DEMs." Remote Sensing of Environment **84**: 295–308.
- Horn, B. K. P. (1981). "Hill Shading and the Reflectance Map." Proceedings of the IEEE **69**(1): 14-47.
- Horritt, M. S. and P. D. Bates (2001). "Effects of spatial resolution on a raster based model of flood flow." Journal of Hydrology **253**(1-4): 239-249.
- Hovenbitzer, M. (2004). The Digital Elevation Model 1:25.000 (DEM25) for the Federal Republic of Germany. XX Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS), Istanbul.
- INEGI (1999a). Ciudad Guzman E13B25. Mexico City, Mexico, INEGI, Instituto Nacional de Estadística, Geografía e Informática.

- INEGI (1999b). Comala E13B34. Mexico City, Mexico, INEGI, Instituto Nacional de Estadística, Geografía e Informática.
- INEGI (1999c). Cuauhtemoc E13B35. Mexico City, Mexico, INEGI, Instituto Nacional de Estadística, Geografía e Informática.
- INEGI (1999d). San Gabriel E13B24. Mexico City, Mexico, INEGI, Instituto Nacional de Estadística, Geografía e Informática.
- Itoh, H., J. Takahama, M. Takahashi and K. Miyamoto (2000). "Hazard estimation of the possible pyroclastic flow disasters using numerical simulation related to the 1994 activity at Merapi Volcano." Journal of Volcanology and Geothermal Research **100**(1-4): 503-516.
- Jones, K. H. (1998). "A Comparison of Algorithms Used to Compute Hill Slope as a Property of the DEM." Computers and Geosciences **24**(4): 315-323.
- Kemp, K. K. (1993). Environmental Modeling with GIS: A Strategy for Dealing with Spatial Continuity, NCGIA - National Center for Geographic Information and Analysis: 142.
- Kenward, T., D. P. Lettenmaier, E. F. Wood and E. Fielding (2000). "Effects of digital elevation model accuracy on hydrologic predictions." Remote Sensing of Environment **74**(3): 432-444.
- Kuhn, W. (2001). "Ontologies in support of activities in geographical space." International Journal of Geographical Information Science **15**(7): 613-631.
- Kumar, M. (1993). "World Geodetic System 1984: A Reference Frame for Global Mapping, Charting, and Geodetic Applications." Surveying and Land Information Systems **53**(1): 53-56.
- Kumler, M. P. (1994). "An Intensive Comparison of Triangulated Irregular Networks (TINs) and Digital Elevation Models (DEMs)." Cartographica **31**(2): 1-99.
- Kyriakidis, P. C., A. M. Shortridge and M. F. Goodchild (1999). "Geostatistics for conflation and accuracy assessment of digital elevation models." International Journal of Geographical Information Science **13**(7): 677-707.
- Lanari, R., G. Fornaro, D. Riccio, M. Migliaccio, K. P. Papathanassiou, J. R. Moreira, M. Schwabisch, L. Dutra, G. Puglisi, G. Franceschetti and M. Coltelli (1996). "Generation of digital elevation models by using SIR-C/X-SAR multifrequency



- two-pass interferometry: The Etna case study." IEEE Transactions on Geoscience and Remote Sensing **34**(5): 1097-1114.
- Laurini, R. and D. Thompson (1992). Fundamentals of spatial information systems. London, England: Academic Press.
- Lee, J., P. K. Snyder and P. F. Fisher (1992). "Modeling the Effect of Data Errors on Feature-Extraction from Digital Elevation Models." Photogrammetric Engineering and Remote Sensing **58**(10): 1461-1467.
- Li, Z., Q. Zhu and C. Gold (2005). Digital Terrain Modeling: Principles and Methodology. Boca Raton: CRC Press.
- Lopes, A., E. Nezry, R. Touzi and H. Laur (1993). "Structure Detection and Statistical Adaptive Speckle Filtering in SAR Images." International Journal of Remote Sensing **14**(9): 1735-1758.
- Lopez, C. (1997). "Locating some types of random errors in Digital Terrain Models." International Journal of Geographical Information Science **11**(7): 677-698.
- Lopez, C. (2000). "Improving the Elevation Accuracy of Digital Elevation Models: A Comparison of Some Error Detection Procedures." Transactions in GIS **4**(1): 43-64.
- Maidment, D. R. (1993). "GIS and hydrologic modeling," Environmental Modeling with GIS, eds. M. F. Goodchild, B. O. Parks and L. T. Steyaert. New York: Oxford University Press.
- Mancarella, P., A. Raffaetà, C. Renso and F. Turini (2004). "Integrating knowledge representation and reasoning in Geographical Information Systems." International Journal of Geographical Information Science **18**(4): 417-447.
- Mark, D. M. (1975). "Computer Analysis of Topography: A Comparison of Terrain Storage Methods." Geografiska Annaler. Series A, Physical Geography **57**(3/4): 179-188.
- Mark, D. M. (1979). "Phenomenon-based data-structuring and digital terrain modelling." Geo-Processing **1**: 27-36.
- Martin, G. J. and P. E. James (1993). All Possible Worlds: A History of Geographical Ideas. New York: John Wiley & Sons, Inc.

- Maune, D. F., ed. (2001). Digital Elevation Model Technologies and Applications: The DEM Users Manual. Bethesda: American Society for Photogrammetry and Remote Sensing.
- Maune, D. F., S. M. Kopp, C. A. Crawford and C. E. Zervas (2001). Introduction to Digital Elevation Model Technologies and Applications: The DEM Users Manual, D. F. Maune. Bethesda: American Society for Photogrammetry and Remote Sensing.
- McGrew, J., J. Chapman and C. B. Monroe (2000). An Introduction to Statistical Problem Solving in Geography. Dubuque: McGraw-Hill.
- Miller, C. L. and R. A. Laflamme (1958). "The digital terrain model - Theory and application." Photogrammetric Engineering **24**: 433-442.
- Mitasova, H. and L. Mitas (2002). "Modeling Physical Systems," Geographic Information Systems and Environmental Modeling, eds. K. C. Clarke, B. O. Parks and M. P. Crane. Upper Saddle River: Prentice Hall.
- Mitasova, H., L. Mitas, W. M. Brown, D. P. Gerdes, I. Kosinovsky and T. Baker (1995). "Modeling Spatially and Temporally Distributed Phenomena - New Methods and Tools for Grass Gis." International Journal of Geographical Information Systems **9**(4): 433-446.
- Moglen, G. E. and G. L. Hartman (2001). "Resolution effects on hydrologic modeling parameters and peak discharge." Journal of Hydrologic Engineering **6**(6): 490-497.
- Molander, C. W. (2001). "Photogrammetry," Digital Elevation Model Technologies and Applications: The DEM Users Manual, ed. D. F. Maune. Bethesda: American Society for Photogrammetry and Remote Sensing.
- Monmonier, M. (1996). How to lie with maps. Chicago: University of Chicago Press.
- NASA. (2002, December 20, 2002). "Shuttle Radar Topography Mission." Retrieved 01/16/2003, 2003, from <http://www.jpl.nasa.gov/srtm/>.
- NOAA. (2006). "NOAA Coastal Services Center - Topographic Change Mapping." Retrieved 06/23/2006, 2006, from <http://www.csc.noaa.gov/TCM/>.

- Oreskes, N., K. Shrader-Frechette and K. Belitz (1994). "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences." Science **263**(5147): 641-646.
- Patra, A. K., A. C. Bauer, C. C. Nichita, E. B. Pitman, M. F. Sheridan, M. Bursik, B. Rupp, A. Webb, A. Stinton, L. M. Namikawa and C. Renschler (2005). "Parallel Adaptative Numerical Simulation of Dry Avalanches over Natural Terrain." Journal of Volcanology and Geothermal Research **139**(1-2): 1-21.
- Pitman, E. B., C. C. Nichita, A. Patra, A. Bauer, M. Sheridan and M. Bursik (2003). "Computing granular avalanches and landslides." Physics of Fluids **15**(12): 3638-3646.
- Podobnikar, T. (2005). "Production of integrated digital terrain model from multiple datasets of different quality." International Journal of Geographical Information Science **19**(1): 69–89.
- Rabus, B., M. Eineder, A. Roth and R. Bamler (2003). "The shuttle radar topography mission - a new class of digital elevation models acquired by spaceborne radar." ISPRS Journal of Photogrammetry and Remote Sensing **57**(4): 241-262.
- Rana, S., Ed. (2004). Topological Data Structures for Surfaces: An Introduction to Geographical Information Science. Chichester, England: John Wiley & Sons, Ltd.
- Rees, W. G. (2000). "The accuracy of Digital Elevation Models interpolated to higher resolutions." International Journal of Remote Sensing **21**(1): 7-20.
- Reitsma, F. E. (2004). "A New Geographic Process Data Model." Ph.D. diss., Department of Geography, University of Maryland.
- Renschler, C. S., D. C. Flanagan, B. A. Engel, L. A. Kramer and K. A. Sudduth (2002). "Site-specific decision-making based on RTK GPS survey and six alternative elevation data sources: Watershed topography and delineation." Transactions of the Asae **45**(6): 1883-1895.
- Rodriguez-Elizarraras, S., C. Siebe, J. C. Komorowski, J. M. Espindola and R. Saucedo (1991). "Field Observations of Pristine Block-Flow and Ash-Flow Deposits Emplaced April 16-17, 1991 at Volcan-De-Colima, Mexico." Journal of Volcanology and Geothermal Research **48**(3-4): 399-412.

- Rodríguez, E., C. S. Morris and J. E. Belz (2005). "A Global Assessment of the SRTM Performance." Photogrammetric Engineering and Remote Sensing **72**(3): 249-260.
- Rogerson, P. A. (2001a). "A Statistical Method for the Detection of Geographic Clustering." Geographical Analysis **33**(2): 215-227.
- Rogerson, P. A. (2001b). Statistical Methods for Geography. London, England: Sage Publications.
- Rosen, P. A. H., S.; Joughin, I.R.; Li, F.K.; Madsen, S.N.; Rodriguez, E.; Goldstein, R.M. (2000). "Synthetic aperture radar interferometry." Proceedings of the IEEE **88**(0018-9219): 333-382.
- Rupp, B., M. Bursik, L. Namikawa, A. Webb, A. K. Patra, R. Saucedo, J. L. Macías and C. Renschler (2006). "Computational modeling of the 1991 block and ash flows at Colima Volcano, México," Neogene-Quaternary continental margin volcanism: A perspective from México, eds. C. Siebe, J. L. Macias and G. J. Aguirre-Díaz. Boulder: Geological Society of America.
- Sansosti, E., R. Lanari, G. Fornaro, G. Franceschetti, M. Tesauro, G. Puglisi and M. Coltelli (1999). "Digital elevation model generation using ascending and descending ERS-1/ERS-2 tandem data." International Journal of Remote Sensing **20**(8): 1527-1547.
- Saucedo, R., J. L. Macias and M. Bursik (2004). "Pyroclastic flow deposits of the 1991 eruption of Volcan de Colima, Mexico." Bulletin of Volcanology **66**(4): 291-306.
- Sharpnack, D. A. and G. Akin (1969). "An Algorithm for Computing Slope and Aspect from Elevations." Photogrammetric Engineering **35**(3): 247-248.
- Sheridan, M. F., A. J. Stinton, A. Patra, E. B. Pitman, A. Bauer and C. C. Nichita (2005). "Evaluating Titan2D mass-flow model using the 1963 Little Tahoma Peak avalanches, Mount Rainier, Washington." Journal of Volcanology and Geothermal Research **139**(1-2): 89-102.
- Shewchuk, J. R. (1997). "Adaptive Precision Floating-Point Arithmetic and Fast Robust Geometric Predicates." Discrete and Computational Geometry **18**(3): 305-363.
- Skidmore, A. K. (1989). "A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model." International Journal of Geographical Information Systems **3**(4): 323 - 334.

- Slater, J. A., G. Garvey, C. Johnston, J. Haase, B. Heady, G. Kroenung and J. Little (2006). "The SRTM Data Finishing Process and Products." Photogrammetric Engineering and Remote Sensing **72**(3): 237–247.
- Smith, B. (2004). "Ontology," Blackwell Guide to the Philosophy of Computing and Information, ed. L. Floridi. Malden: Blackwell: 155–166.
- Smith, B. and D. M. Mark (1998a). Ontology and Geographic Kinds. International Symposium on Spatial Data Handling, Vancouver, Canada.
- Smith, B. and D. M. Mark (1998b). Ontology and Geographic Kinds. 8th International Symposium on Spatial Data Handling (SDH'98), Vancouver, International Geographical Union.
- Standards Committee of the IEEE Computer Society (1985). IEEE Standard for Binary Floating-Point Arithmetic, Institute of Electrical and Electronics Engineers: 14.
- Stevens, N. F., V. Manville and D. W. Heron (2002). "The sensitivity of a volcanic flow model to digital elevation model accuracy: experiments with digitised map contours and interferometric SAR at Ruapehu and Taranaki volcanoes, New Zealand." Journal of Volcanology and Geothermal Research **119**(1-4): 89-105.
- Stevens, S. S. (1946). "On the Theory of Scales of Measurement." Science **103**(2684): 677-680.
- Steyaert, L. T. (1993). "A Perspective on the State of Environmental Simulation Modeling," Environmental Modeling with GIS, eds. M. F. Goodchild, B. O. Parks and L. T. Steyaert. New York: Oxford University Press.
- Stroustrup, B. (2000). The C++ Programming Language. Boston: Addison-Wesley.
- Sutton, P., C. Roberts, C. Elvidge and H. Meij (1997). "A comparison of nighttime satellite imagery and population density for the continental united states." Photogrammetric Engineering and Remote Sensing **63**(11): 1303-1313.
- Takahashi, T. and H. Tsujimoto (2000). "A mechanical model for Merapi-type pyroclastic flow." Journal of Volcanology and Geothermal Research **98**(1-4): 91-115.
- USGS (1998). Standards for Digital Elevation Models - Part 2: Specifications, U.S. Department of the Interior, U.S. Geological Survey, National Mapping Division: 70.

- USGS. (2003). "USGS Digital Elevation Model Data." Retrieved 01/16/2003, 2003, from [http://edc.usgs.gov/glis/hyper/guide/usgs\\_dem](http://edc.usgs.gov/glis/hyper/guide/usgs_dem).
- van der Poorten, P. M. and C. B. Jones (2002). "Characterisation and generalisation of cartographic lines using Delaunay triangulation." International Journal of Geographical Information Science **16**(8): 773-794.
- van Zyl, J. J. (2001). "The Shuttle Radar Topography Mission (SRTM): A breakthrough in remote sensing of topography." Acta Astronautica **48**(5-12): 559-565.
- Veregin, H. (1999). "Data Quality Parameters," Geographical Information Systems, eds. P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind. New York: John Wiley & Sons.
- Vinhas, L. and K. R. Ferreira (2005). "Descrição da TerraLib," Bancos de Dados Geográficos, eds. M. Casanova, G. Câmara, C. Davis, L. Vinhas and G. R. d. Queiroz. Curitiba, Brazil: Editora MundoGEO.
- Vinhas, L., R. C. Souza and G. Câmara (2003). Image Data Handling in Spatial Databases. V Brazilian Symposium on GeoInformatics, Campos do Jordão, Brazil.
- Wadge, G., B. Scheuchl and N. F. Stevens (2002). "Spaceborne radar measurements of the eruption of Soufrière Hills Volcano, Montserrat," The Eruption of Soufrière Hills Volcano, Montserrat, from 1995 to 1999, eds. T. H. Druitt and B. P. Kokelaar. London, England: The Geological Society of London.
- Welch, R. J., T. Lang, H. Murakami, H. (1998). "ASTER as a source for topographic data in the late 1990s." IEEE Transactions on Geoscience and Remote Sensing **36**(4): 1282-1289.
- Wilson, J. P. and J. C. Gallant (2000). "Digital Terrain Analysis," Terrain Analysis: Principles and Applications, eds. J. P. Wilson and J. C. Gallant. New York: John Wiley & Sons.
- Wise, S. (2000a). "Assessing the quality for hydrological applications of digital elevation models derived from contours." Hydrological Processes **14**(11-12): 1909-1929.
- Wise, S. (2000b). "GIS data modelling - lessons from the analysis of DTMs." International Journal of Geographical Information Science **14**(4): 313-318.

- Woldenberg, M. J. (1972). "The Average Hexagon in Spatial Hierarchies," Spatial Analysis in Geomorphology, ed. R. J. Chorley. London, England: Methuen.
- Wolf, G. W. (2004). "Topographic Surfaces and Surface Networks," Topological Data Structures for Surfaces: An Introduction to Geographical Information Science, ed. S. Rana. Chichester, England: John Wiley & Sons, Ltd.
- Ye, C. S., B. M. Jeon and K. H. Lee (2003). "Digital elevation model combination using triangular image warping interpolation and maximum likelihood." International Journal of Remote Sensing **24**(18): 3683-3689.
- Yu, B., K. Dalbey, A. Webb, M. Bursik and A. Patra (2005). Numerical Issues in Computing Inundation Areas Over Natural Terrains Using Svage-Hutter Theory. Geophysical Mass Flow Group. Buffalo, NY.