



UCGE Reports
Number 20379

Department of Geomatics Engineering

**Interactive Visualization to Reveal Activity Patterns of
Marine Mammals and Boat Traffic in the St. Lawrence
Estuary in Quebec, Canada**

(URL: <http://www.geomatics.ucalgary.ca/graduatetheses>)

by

Botao Zheng

May 2013



UNIVERSITY OF CALGARY

Interactive Visualization to Reveal Activity Patterns of Marine Mammals and Boat Traffic in the
St. Lawrence Estuary in Quebec, Canada

by

Botao Zheng

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

May 2013

© Botao Zheng 2013

Abstract

Intensive human activities of commercial shipping, ferries, and tourism have posed cumulative threats to the marine mammals in the St. Lawrence Estuary. In order to understand and quantify the exposure of marine mammals to visiting traffic, an agent-based model was built to simulate the behaviour of whales and boats in a portion of the St. Lawrence Estuary, which covers the Saguenay-Saint Lawrence Marine Park and the proposed Saint Lawrence Estuary Marine Protected Area. This model serves as a decision support tool to test the impact of various zoning and regulatory scenarios proposed by the Park's managers. The objective of this research was to design and implement an interactive Geovisualization software program to manage the large datasets in volumes generated by the agent-based model to reveal, demonstrate and explore the potential activity patterns of marine mammals and boat traffic. This module includes two main components: the Database Module and the Application Module. The Database Module was designed using a standard general-purpose relational database management system. Through an interactive visual user interface, the Application Module offers a series of functionalities to visualize the results of users' queries at both aggregate and disaggregate level. This software successfully allows the agent-based model to demonstrate the activity patterns of boat traffic and marine mammals by: 1) integrating many interactive dimensions at the same time while avoiding the interpretative complexity of multivariate pattern generalization or recognition methods, 2) verifying if the simulated patterns have a good fit with the real world, 3) allowing the Park's managers to perform exploratory analysis for testing their regulation scenarios, and 4) providing a visual spatial support to aid spatial decision making process regarding the sustainable management of the Marine Park.

Acknowledgements

This project was funded by a Strategic Project Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) awarded to Dr. Lael Parrott, University of Montreal and Dr. Danielle Marceau.

Thanks to the financial support received from the University of Calgary and the Government of Alberta.

Thanks to Professor Danielle Marceau, my supervisor for all the academic support, instruction, suggestions and advices. Danielle introduces me to research in the field of Geovisualization.

Thanks to Fang Wang, Nishad Wijesekara, Jean-Gabriel Hasbani in the Geocomputation Lab of Professor Danielle Marceau. These colleagues gave me lots of valuable suggestions.

Table of Contents

Abstract	i
Acknowledgements.....	ii
Table of Contents.....	iii
List of Tables	v
List of Figures and Illustrations	vi
CHAPTER ONE: INTRODUCTION.....	1
1.1 Context of the study.....	1
1.2 The Marine Mammal and Maritime Traffic Simulation (3MTSim) model.....	4
1.3 The issue of large model output datasets	6
1.4 Geovisualization: definition and previous research.....	8
1.4.1 Space-Time Prism model	9
1.4.2 Simple activity patterns and activity pattern surface.....	11
1.4.3 Integrating spatio-temporal databases with Geovisualization.....	13
1.4.4 Time-based database models.....	14
1.4.5 Scope of research in Geovisualization design.....	16
1.4.6 Classification of visual analysis in Geovisualization research.....	16
1.4.7 Recent trends in Geovisualization research.....	17
1.5 Objective of the study.....	18
1.6 Organization of the thesis	20
CHAPTER TWO: METHODOLOGY.....	21
2.1 Datasets generated by 3MTSim.....	21
2.2 Challenges and strategies of software design	21
2.3 Software architecture	22
2.4 Database module.....	24
2.4.1 Database design.....	24
2.4.2 Components of the database module.....	25
2.4.3 Types of datasets in the database module.....	26
2.4.4 The tables in the whale database	27
2.4.4.1 Table SourceDataOfWhaleModel and its data structure	28
2.4.4.2 Table ClassifiedSourceDataOfWhaleModel and its data structure	29
2.4.4.3 Table ActivityOfWhales and its data structure.....	31
2.4.4.4 Table CompositionOfWhaleTypes and its data structure	32
2.4.5 Tables in the boat database.....	33
2.4.5.1 Table SourceDataOfBoatModel and its data structure	33
2.4.5.2 Table ClassifiedDataOfBoatModel and its data structure	36
2.4.5.3 Table StatisticDataOfExcursions and its data structure.....	40
2.4.5.4 Table SailingCharacteristics and its data structure.....	43
2.4.5.5 Table TimeOfObservation and its data structure.....	44
2.4.5.6 Table ActivityDistributionOfBoat and its data structure	46
2.4.6 The file system	47
2.5 Application module.....	47

2.5.1 Functionalities in the Application module	48
2.5.2 Algorithms of the application functions	49
2.5.2.1 Build whale database	49
2.5.2.2 Build boat database	50
2.5.2.3 Composition of whale types in simulation	50
2.5.2.4 Composition of excursions by boat types and home ports	50
2.5.2.5 Point density map of whales	50
2.5.2.6 Point density map of boat type (all activities)	52
2.5.2.7 Point density map of boat type (Whale Observation).....	53
2.5.2.8 Line density map of boat type.....	55
2.5.2.9 Trajectory of whales	56
2.5.2.10 Trajectory of boats.....	57
2.5.2.11 Trajectory characteristics of boat excursions by boat type and port type	57
2.5.2.12 Repartition of boat activities by boat type and port type	58
2.5.2.13 Boat observation on specific whale type	60
2.5.2.14 Analyze speed, duration, and distance of excursion	60
2.6 User interface design	61
2.7 Graph rendering methods.....	63
2.8 Connections between tables in the database module and functionalities of the application module.....	64
CHAPTER THREE: RESULT ANALYSIS	65
3.1 Composition of excursions by boat type and home port, composition of whale types	65
3.2 Point and line density map of boats and whales	67
3.3 Trajectory characteristics of boat excursions.....	72
3.4 Repartition of boat activities by boat type and home port.....	74
3.5 Dynamic visualization of trajectories of whales.....	76
3.6 Dynamic visualization of trajectories of boats	77
3.7 Boat observation on specific whale type	80
3.8 Analysis of speed, duration of excursion and distance of excursion	81
3.9 Software performance.....	82
CHAPTER FOUR: CONCLUSION.....	85
4.1 Further research	87
REFERENCES	89

List of Tables

Table 2.1 Source data structure generated by the whale model.....	28
Table 2.2 Classified source data structure generated by the whale model	29
Table 2.3 Data structure of whale activities.....	31
Table 2.4 Data structure of composition of whale types	32
Table 2.5 Source data structure generated by the boat model	33
Table 2.6 Classified source data structure generated by the boat model	37
Table 2.7 Data structure of boat excursions.....	40
Table 2.8 Data structure of boat sailing characteristics	43
Table 2.9 Data structure of boat observing whales.....	45
Table 2.10 Data structure of boat activity distributions.....	46
Table 3.1 Statistic data on software performance.....	83

List of Figures and Illustrations

Figure 1.1 The Saguenay-St. Lawrence Marine Park and Marine Protected Area.....	3
Figure 1.2 The 3MTSim main user interface (Source: Parrott <i>et al.</i> 2011).....	5
Figure 1.3 Space-Time aquarium with space-time paths (source: Kwan 2000).....	10
Figure 1.4 Simple activity patterns in space-time (source: Kwan, 2000).....	12
Figure 1.5 Activity density patterns in geographic space, with the surface for non-employment activities over the surface of home locations (source: Kwan 2000)	13
Figure 1.6 A temporal grid with a variable-length list attached to each grid cell denoting successive changes (from Langran 1993, cited by Peuquet and Duan 2007).....	15
Figure 2.1 Software architecture.....	23
Figure 2.2 Database structure of database module	27
Figure 2.3 Diagram of the application module	49
Figure 2.4 Calculation of the point density map of whale types	51
Figure 2.5 Calculation of the point density map of boat activity types (all activities).....	53
Figure 2.6 Calculation of the point density map of boat activity types (whale observation)	54
Figure 2.7 Calculation of the line density map of boat types (all activities)	56
Figure 2.8 Definition of movement parameters (Source: Chion <i>et al.</i> 2011)	58
Figure 2.9 Normalization of boat activities of two excursions (Source: Chion <i>et al.</i> 2011)	59
Figure 2.10 Moving window of five time points	60
Figure 2.11 Main interface of the Geovisualization software.....	61
Figure 2.12 Connection between tables and functionalities	64
Figure 3.1 Composition of boat excursions	67
Figure 3.2 Density map of boats.....	68
Figure 3.3 User menu of density map of boats	71
Figure 3.4 The geographic map of the study area which is used for displaying the density locations and trajectories of boat excursions (source: Parks Canada)	72

Figure 3.5 Trajectory characteristics of boat excursions	73
Figure 3.6 Repartition of the boat activities by boat type and port type.....	75
Figure 3.7 A trajectory of a beluga (agent id = 134) in the whole simulation season	77
Figure 3.8 Trajectories of boats by time point.....	79
Figure 3.9 Boat observing a particular type of whale.....	81
Figure 3.10 Analysis of average speed, duration, and distance of excursions.....	82

Chapter One: Introduction

1.1 Context of the study

The St. Lawrence Estuary, in Quebec, Canada, is the world's largest estuary and a major feeding ground for certain North Atlantic whale populations due to the presence of high concentrations of forage species such as krill (*Euphausia superba*) and capelin (*Mallotus villosus*). Every year, many cetaceans, including large whales, migrate to the estuary to feed and build up their energy reserves in preparation for the breeding season. Twelve species of marine mammals can be found in the estuary either regularly or occasionally, usually on a seasonal basis. The concentration and relative predictability of the marine mammals along with the majestic view of the confluence of the Saguenay fjord and the St. Lawrence River and an easy access have made this region one of the most renowned in the world for whale-watching (Scarpaci *et al.* 2008). Nearly 50 whale-watching boats transport over 400,000 tourists onto the estuary every year to encounter whales (Chion *et al.* 2011). Five species are the main targets of these excursions:

- the endangered blue whale (*Musculus*),
- the minke (*Balaenoptera acuturostrata*),
- the humpback whale (*Megaptera novaeangliae*),
- the finback whale (*Balaenoptera physalus*) listed as special concern by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC 2009), and
- the threatened resident beluga (*Delphinapterus leucas*).

The estuary is a major transportation corridor for commercial cargo ships heading to and from the Atlantic Ocean and the industrial centers of the Great Lakes region (Parrott *et al.* 2011). It also contains several ferry routes that cross the river and connect islands to the mainland, totaling more than 40 000 trips per year (Chion *et al.* 2011). Other activities include commercial fishing,

scenery viewing aboard cruise boats, sailing, and kayaking. This high volume of traffic has the potential to adversely affect the coastal and marine environment. Crowds of boats can result in increased risk of collision and whale injuries. They also can cause disturbances during the breeding season. Powerful and close sounds could bring permanent ear damage, internal injuries, and even the death of whales (Acoust *et al.* 2008).

In recognition of the importance of protecting the marine environment and promoting an integrated ecosystem management, the federal government and the province of Quebec created the Saguenay-St. Lawrence Marine Park in 1998, located at the confluence of the Saguenay River and the St. Lawrence Estuary (Fig. 1.1). Covering an area of 1245 km², the park preserves a representative portion of the marine environment for conservation purposes, while encouraging its use for educational, recreational and research purposes (Parrott *et al.* 2011). To ensure the long-term protection and conservation of the marine mammals as well as their habitats and food resources, Fisheries and Oceans Canada is also planning to establish a Marine Protected Area (MPA) of approximately 6000 km² (Fig. 1.1), adjacent to the Saguenay-St. Lawrence Marine Park, where pressure on marine mammals is the greatest (Fisheries and Oceans Canada 2004).

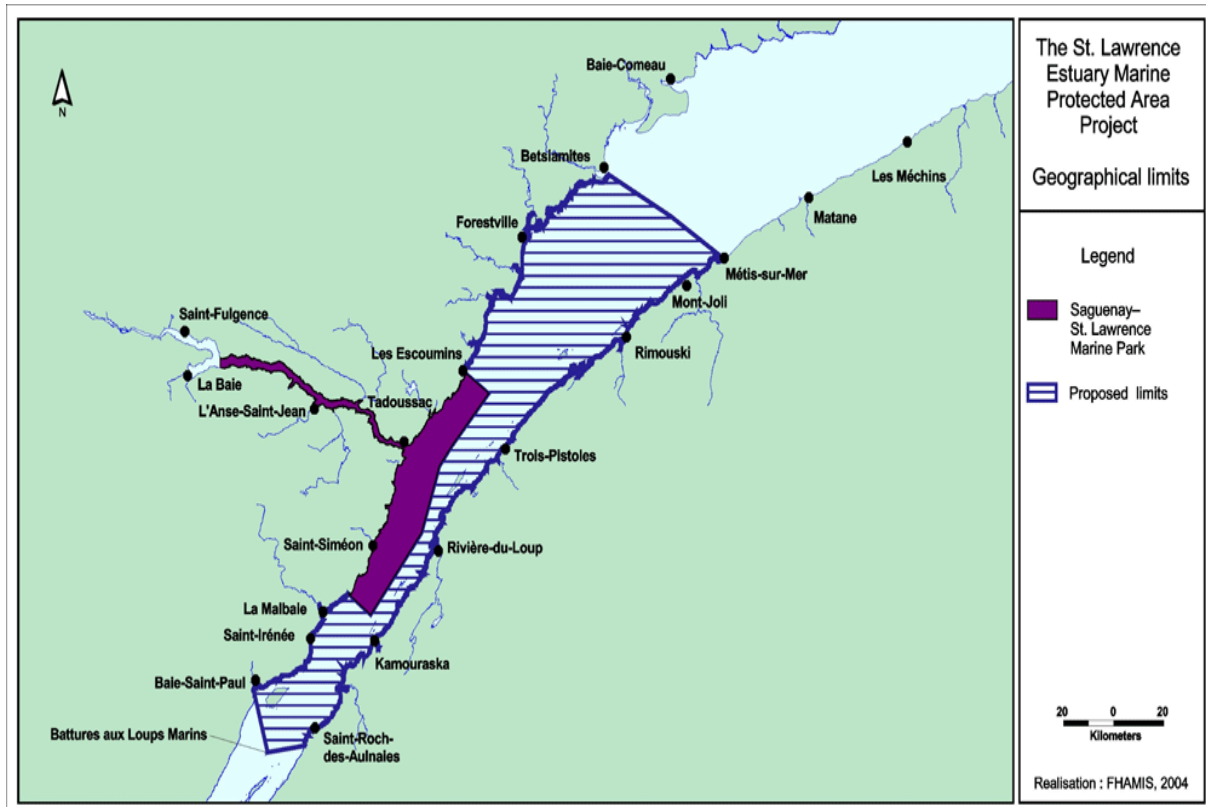


Figure 1.1 The Saguenay-St. Lawrence Marine Park and Marine Protected Area

Source: <http://www.qc.dfo-mpo.gc.ca/gestion-management/images/CarteEstuaire-eng.gif>

The tourism industry is a major driver of the local economies of several small coastal communities surrounding the Park and MPA. The marine transportation industry is also a significant economic contributor, not only to coastal communities but to Canada as a whole. Managers of the Park and the MPA need an adaptive zoning plan to regulate the traffic in order to avoid congestions, ensure safe passages, and protect the ecosystems. At the same time the zoning plan will maintain the region's economic viability. Such a plan should be justified and updated regularly based on new scientific knowledge and via public consultation with stakeholders (Parrott *et al.* 2011). Access to a spatial decision support tool to provide computerized assistance to the Park and MPA managers in the development, evaluation and

selection of proper policies, plans scenarios, projects or interventions to facilitate long-term decision becomes necessary.

1.2 The Marine Mammal and Maritime Traffic Simulation (3MTSim) model

To address these issues, a multidisciplinary project was initiated by a team of Canadian scientists in collaboration with Parks Canada, the Department of Fisheries and Ocean Canada, and the Research and Education Group on Marine Mammals (GREMM) in Quebec to develop a computer simulation model linked to a Geographic Information System (GIS) that could serve to inform and support management decisions in the Marine Park and MPA. An agent-based model, called Marine Mammal and Maritime Traffic Simulation (3MTSim) was built to simulate the spatio-temporal movement of marine mammals and marine traffic in the Marine Park.

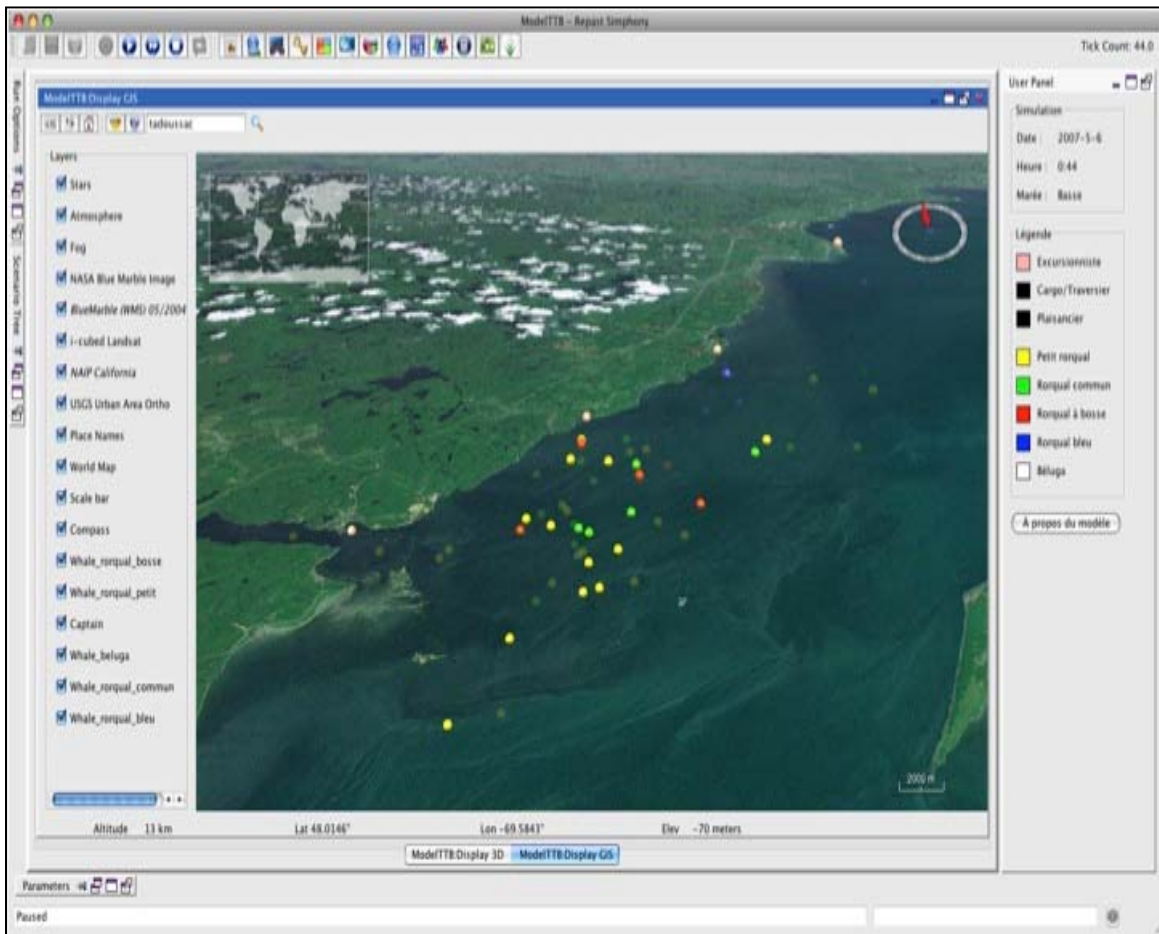


Figure 1.2 The 3MTSim main user interface (Source: Parrott *et al.* 2011)

Figure 1.2 provides a snapshot of the running 3MTSim, which shows an area view of the study area using the WorldWind virtual globe. Coloured dots are boats and whales that move around as a simulation progresses. Bright dots are whales at the surface; pale dots are whales under the surface. The size of the dots has been exaggerated to facilitate viewing. The model is developed in Repast Symphony v.1.2 as a decision support tool to assist management and planning in the Estuary. 3MTSim allows users to simulate and test different zoning scenarios for maritime traffic (e.g., area closures, speed limits, regulations concerning the observation of marine mammals) in order to assess their effects both on navigation dynamics and on marine mammal exposure to

vessels. 3MTSim includes an individual-based model of marine mammal movement patterns that has been elaborated based on existing telemetry data on fin, blue, and beluga whales as well as on land-based theodolite tracking of humpback and minke whales. Different types of vessels, including cargo ships, commercial whale-watching tour boats, pleasure craft and ferries are also modelled individually. The model represents the decision-making process of boat captains as a function of environmental conditions, the contextual setting and their respective goals. 3MTSim quantifies the spatio-temporal movement of marine whales and marine traffic in the Park and MPA and explores a diversity of management solutions through the testing of various zoning scenarios and regulations (Parrott *et al.* 2011).

1.3 The issue of large model output datasets

Simulations are run with 3MTSim for periods of time ranging from one day to a maximum of a whole season (early May to late October) with at most five whale types, thirty boat types, and ten home ports. The simulations generate a large datasets in volumes of spatio-temporal data with a complex structure of 34 interacting attributes such as time, boat type, whale type, location of agents, etc. For example, one simulation of 200 boat excursions over a period of 100 days at a temporal resolution of 1 minute generates 28,800,000 records. To assist the Park and MPA managers in their decisions, one needs to extract meaningful information from this massive dataset, such as

- the spatio-temporal distribution and corresponding density of marine mammals and boat traffic,
- the temporal distribution of boat activities (observing whale, observing geographical features, sailing, observing seals and back to port),
- the various characteristics of the excursions such as sailing speed, length of excursion,

and whale observation time, and

- the characteristics of the spatio-temporal trajectory (turn angles, step length, duration) of marine mammals and boats, and the identification of high risk zones of possible congestions and collisions.

A major challenge for retrieving this information is related to the presence of spatio-temporal attributes (e.g. whale and boat geographical locations over time) along with various non-geographic attributes (e.g. boat type, whale type, home port), which results in a large number of multi-dimensional combinations of attributes. The processing of these attribute combinations cannot be easily automated considering the important diversity of requests that could be made by the Park and MPA managers based on their domain of expertise and their specific needs. For example, managing boat activities requires an understanding of how they are spatially distributed and the consequence of this distribution on the whale movement patterns. In this example, the Park and MPA managers may want to examine the relationship between boat activity and geographic locations over different time periods. They may also be interested in the distribution of different boat types from different home ports at different moments in time (i.e. 11:00 am at a particular day) or time periods (e.g. the whole season). Alternatively, they may wish to look at the distribution of all boat types for the whole season. These examples illustrate the numerous requests of information that could be made involving different combinations of boat and whale type, time, and home port. Traditional statistical data handling and mapping methods alone are not adequate to fulfill the task of retrieving the desired information from the voluminous datasets generated by 3MTSim. While some visualization functions built in ArcGIS could be used to retrieve information, each task requires writing a specific script, which would become highly tedious and inefficient considering the large possible number of tasks that could be requested.

Therefore, there is a need for an interactive approach to extract information from high dimensional spatio-temporal datasets, to allow users to explore the data, and to present the information in a unified view.

1.4 Geovisualization: definition and previous research

Geovisualization is an intuitive and effective tool for interpreting complex multidimensional datasets. It aims at exploring, analyzing, synthesizing and presenting geospatial data by implementing information representation and analysis techniques including cartography, scientific visualization, image analysis, and exploratory data analysis (Dykes *et al.* 1992). It is widely regarded that the earliest Geovisualization research of space-time started in early 1960s, when Hägerstrand introduced a space-time model that included features such as the space-time path and the space-time prism. Due to the limitation of computer hardware and software, Geovisualization was limited to manual graphing method when Hägerstrand introduced his model. During the following 1970s to 1980s, Geovisualization research remained into a stagnant state. The increased interest was not witnessed until 2000, when Kwan applied and improved the Hägerstrand's space-time model and, at the same time, introduced a surface density pattern model. In recent years, the scientific community dealing with large volumes of spatial and temporal data has demonstrated an increased need for Geovisualization techniques.

Geovisualization techniques overcome the challenges faced by traditional statistics by using size, shape, position, color, diagram, motion and other attributes to handle multi-dimensional datasets on a unique display (Dykes *et al.* 2005). Geovisualization is currently growing as a subject of practice and research to interpret spatial-temporal datasets and has been widely applied in space-

related environmental decision support problems to explore and interpret activity patterns in space and time. Applying Geovisualization techniques to extract and demonstrate patterns from large multi-interacting dimension space-time dataset studies has become a trend.

During the last decades, various techniques have been created and applied in large spatial-temporal dataset research to support decision making. The most important are described in the following sections. Some of these techniques have been revised and adapted to the current research while others have served as guidelines for this project's design.

1.4.1 Space-Time Prism model

Spatio-temporal trajectory visualization is an essential requirement in spatial-temporal dataset research. Currently, the most widely applied model is still the 3D Space-Time Prism originally introduced by Hägerstrand in early 1960s. A typical Space-Time Prism contains the space-time-paths of individuals indicating the locations that can be reached in a particular time interval. The prism's height (z-axis) represents time point while the base (x- and y-axis) has a draped geographic map to represent the geographic area, its boundaries and features (Fig. 1.3). A vertical line in the space-time path represents the time a person remains at the same location.

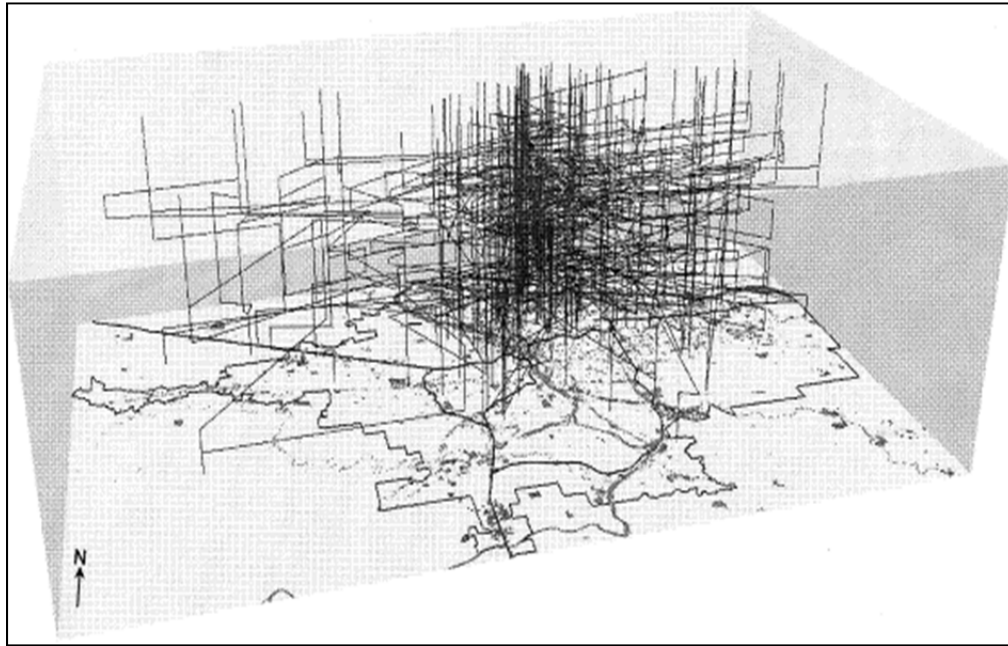


Figure 1.3 Space-Time aquarium with space-time paths (source: Kwan 2000)

The Space-Time Prism was used only sporadically since its conceptualization in the 1970s and 1980s (Kraak 2008). The increased interest in this model started in 2000. In that year, Kwan implemented this model to study the travel pattern differences among gender and different ethnic groups from large spatial and temporal dimensional travel diary datasets (e.g. location, timing, duration and sequencing of trips and activities) collected in the Portland (Oregon) metropolitan region. The model demonstrated different non-employment groups' activity patterns. For example, most of the non-employment activities at lunch hour are found largely in the high density area near downtown. Non-employment activities in the suburban areas tend to take place in the morning, late afternoon or evening.

Throughout the years, this model has been applied and improved to understand movements through space. The original concept was introduced with the limitation of computer graphic technique. Due to the heavy load of manual graphing, the model allowed users to experience a single view. This is not efficient to discover knowledge from a large spatial-temporal dataset. Kraak (2003), using the data of Napoleon's 1812 march into Russia, further revised the space-time-prism model. With a geographic map on the x-and y-axis, and time on z-axis, users are allowed to have multiple linked views by querying the data through time, such as the longitude and latitude. This attribute view offers users the options to define which criteria will be displayed in the prism. More optional views have been shown to help the orientation and navigation in the cube and make the visualization more usable.

The Space-Time Prism is valuable for understanding the activity-travel behavior of (multiple) individuals or groups and it provides valuable reference for designing the visualization of whale and visiting boat trajectories in this project.

1.4.2 Simple activity patterns and activity pattern surface

When large spatial-temporal datasets are used that include numerous individuals or subgroups, the space-time prism does not allow the user to visually compare the multiple trajectories that might be created. Methods that facilitate the inter-group or inter-activity comparison are needed. Kwan (2000), in addition to applying Hägerstrand's space-time prism, introduced Activity Density Patterns (Fig. 1.4) that can be visualized using two approaches.

- a) Simple activity density pattern. This model has its z-axis representing the density of individuals and its x- and y- axes representing the geographic area and features. The

lengths of the vertical lines represent the high or low number of individuals in a particular geographic location.

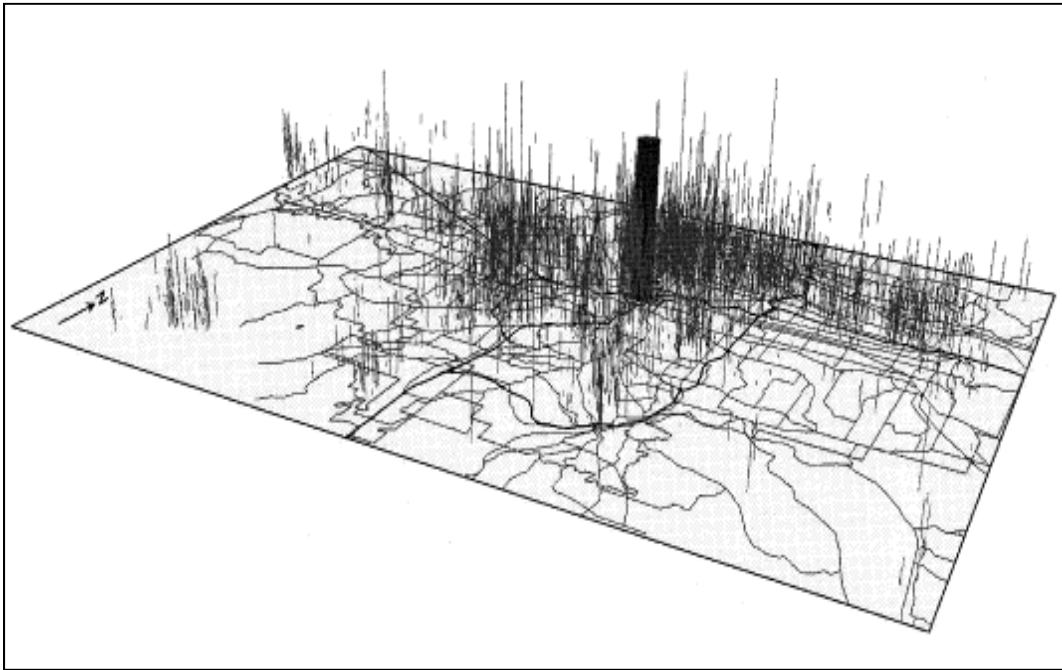


Figure 1.4 Simple activity patterns in space-time (source: Kwan, 2000)

- b) Activity density pattern surface. To generate the density surface, a non-parametric density estimation method called kernel estimation (Gatrell 1994 and Silverman, 1986) is used. The kernel method smoothes the density (total number) of individual and creates a smooth surface with peaks and troughs, which represents the high and low density of individuals of the geographic area (Fig. 1.5). This model demonstrates the high activity and low activity area of different groups of people. A drawback of the simple activity density pattern is that when the densities of several geographic locations are visualized, the vertical lines overlap and become hard to distinguish. In comparison, the activity density pattern surface addresses this challenge. It provides a good reference to visualize the distribution of whales and visiting boats in the study area in 2D graphs.

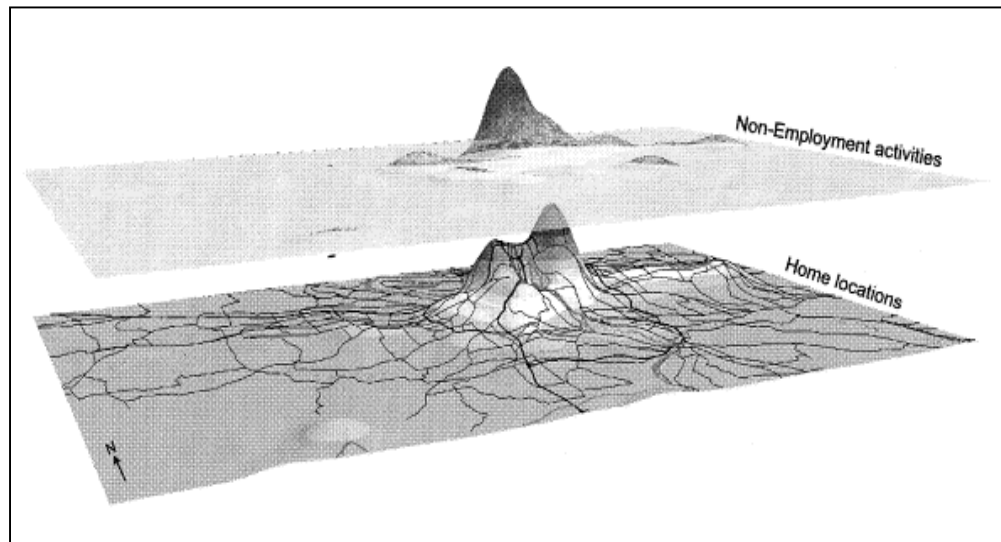


Figure 1.5 Activity density patterns in geographic space, with the surface for non-employment activities over the surface of home locations (source: Kwan 2000)

1.4.3 Integrating spatio-temporal databases with Geovisualization

Spatio-temporal databases deal with applications where data types are characterized by both spatial and temporal semantics (Nikos *et al.* 2004). In order to discover knowledge from a large spatio-temporal dataset, recent Geovisualization research efforts have attempted to achieve an appropriate interaction between visualization techniques and large spatio-temporal databases. Frihida *et al.* (2004), in their GIS Transport Demand Model, integrated the Space-Time Path with a spatio-temporal database handling large amounts of multivariate spatial data. The model successfully detected undesired traffic congestion and could be used to predict traffic flow. In this approach, time path is formulated as a time ordered set composed of activity events and trip events. As sets, the time path and its components can be searched using their respective indexes. This approach provides cast-based animation that mimics continuous individual movement on the street network. Interactive views of space-time paths by a spatio-temporal querying database using respective indexes assist the transformation of large amounts of raw geographic data into

knowledge. This model provided insights about the importance of a well-designed time-based database which is necessary to support time-based queries and intensive background calculations and data aggregation.

1.4.4 Time-based database models

As discussed in Section 1.4.3, it is necessary to integrate a time-based database to Geovisualization to facilitate an appropriate interaction between visualization techniques and large spatio-temporal datasets. The following types of database models provide a good reference for this study.

Langran (1993) proposed a time-based database model that represents each value in a gridded array of discretized locations as a list (Fig. 1.6). A change at a given location is added at the beginning of the list. The result is a set of variable-length lists referenced to grid cells. This method has the advantage of only storing changes related to specific locations (Peuquet and Duan 2007). This model was implemented to store the geographical occurrence of whales and boats in this study (Section 2.4.1).

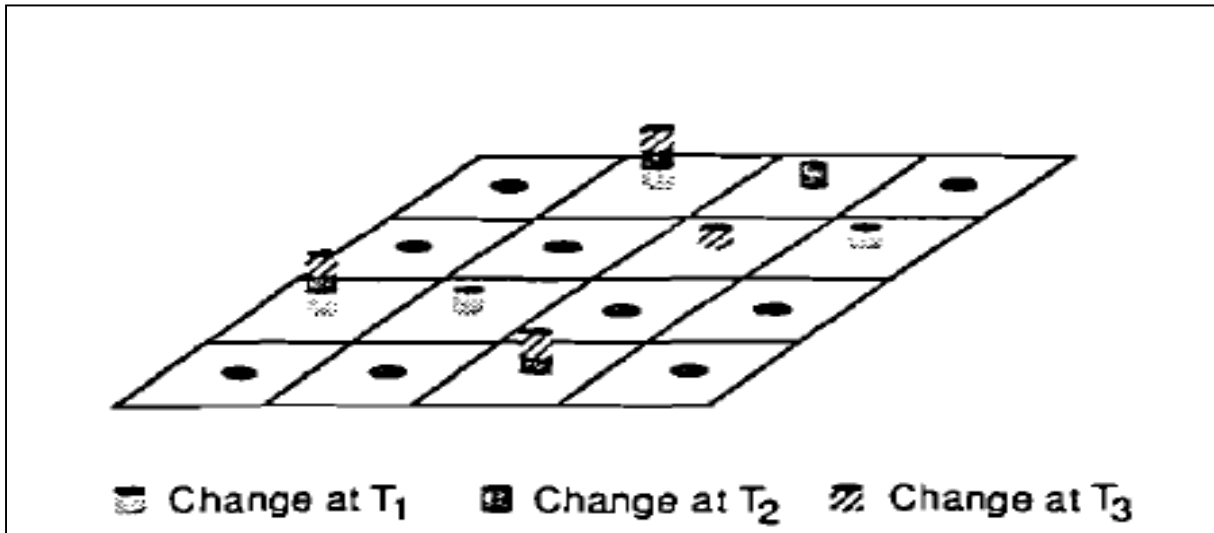


Figure 1.6 A temporal grid with a variable-length list attached to each grid cell denoting successive changes (from Langran 1993, cited by Peuquet and Duan 2007)

Peuquet and Duan (2007) designed a time-based spatio-temporal database, called Event-based Spatio- Temporal Data Model (ESTDM), to facilitate the temporal analysis of geographical data. This model organizes the spatial-temporal datasets based on location in time. Each location in time along the time-line (with its temporal location noted as $t_0, t_1 \dots, t_n$) is associated with a particular set of locations and features. The time associated with each change (i.e., each event) is stored in increasing order from the initial world state of time t_0 to the latest recorded change at time t_n .

This database model was implemented in this project to facilitate two fundamental temporally-based retrieval tasks:

- retrieving location(s) that changed to a given value at a given time or temporal interval, and
- aggregating geographic related data, for example total area, that has changed to a given

value over a given temporal interval.

Most of the retrievals of 3MTSim datasets for the visualization of spatial-temporal patterns falls into the above two categories, so this model was implemented as part of the database design (Section 2.4.1).

1.4.5 Scope of research in Geovisualization design

Dykes *et al.* (2005), in their book entitled ‘Exploring Geovisualization’, summarized and listed the scope of Geovisualization research, which mainly focused on the following four areas:

- 1) determining the limits and advantageous use of both traditional and novel representation methods which create meaningful graphics to represent very large, multivariate spatio-temporal data sets (that may include both three spatial dimensions and time);
- 2) constructing knowledge from geospatial data though integrating interactive visual representation with advanced technologies in Computer Science and computer graphics;
- 3) designing an interface for visually enabled knowledge construction; and
- 4) producing a useable solution that actually works under a circumstance with particular users.

The four areas are not only the main research areas of current Geovisualization, but also the scope of consideration when designing a Geovisualization software. These four areas are considered and implemented in parallel for the design of Geovisualization software in this study.

1.4.6 Classification of visual analysis in Geovisualization research

Visual analysis combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex

data sets (Keim *et al.* 2008). Andrienko *et al.* (2008) schematically classified visual analysis in Geovisualization research into three approaches:

- 1) Traditional direct depiction of each record in a dataset so as to allow the analyst to extract noteworthy patterns by looking at displays and interacting with them. However, since large and complex datasets are becoming increasingly common, the result may become illegible due to visual clutter and massive over plot. The rendering and responses to user interaction may become too slow.
- 2) Deriving and visualizing abstract data summaries computed from the original measurement by aggregation, generalization, and sampling.
- 3) Extracting patterns by applying more sophisticated computational techniques such as those developed for data mining. These techniques visually represent data pattern which are interpreted by the user.

The approaches of above 2) and 3) are intended to support the production of graphics. They are legible, comprehensive and rapid in response in light of the challenges associated with large and complex data sets. Any single of the above three methods is unable to describe and explain the data completely and efficiently. These approaches are implemented for the visual analysis of this study.

1.4.7 Recent trends in Geovisualization research

Knowledge Discovery in Database (KDD) is a conceptual and methodological framework that was proposed to address the issue of transforming large amounts of raw geographic data into knowledge (Frawley *et al.* 1991, Fayyad *et al.* 1996, 2002, MacEachren *et al.* 1999, Gahegan *et al.* 2001, Frihida *et al.* 2004). Geovisualization and KDD have similar goals. However they differ in the extent to which they rely upon human vision or computation methods to process data (MacEachren *et al.* 2010). The primary focus of geovisualization is to use visual representation

to facilitate the identification and interpretation of patterns and relationships in complex data (MacEachren *et al.* 2010). KDD is characterized as a process, in which data mining algorithms (through which patterns are extracted from data) play a central role (MacEachren *et al.* 2010). Since the 2000s, Geovisualization research has been focused on integrating interactive visualization with efficient computation and database processing and KDD for effective problem solving. Frihida *et al.*'s transport demand model developed in 2004 is among the earliest studies to integrate KDD with geovisualization to handle spatio-temporal datasets. The data mining method called disaggregate action-based approach, which was proposed by Kwan in 2000 was implemented to extract and visualize individual space-time-paths in Quebec City, Canada. Rinzivillo (2008) investigated the possibility of using clustering techniques in visual exploration and analysis of large numbers of spatio-temporal trajectories. To assess the degree of (dis)similarity between trajectories, several distance functions were successively applied to cluster the trajectories progressively. The progressive clustering where a simple distance function is applied at each step refined the earlier obtained results. Bringing KDD and Geovisualization methods together has been a long term strategy to extract and represent information from large spatio-temporal datasets. The strategy is also one of the main considerations of the software design of this study.

1.5 Objective of the study

The objective of this project is to develop a Geovisualization software to facilitate the interactive presentation and interpretation of whale and boat activity patterns in the St. Lawrence Estuary from large spatio-temporal datasets generated by the agent-based model 3MTSim. By integrating various novel Geovisualization methods and traditional cartographic techniques, this software

will assist the Marine Park and MPA managers in the development, evaluation, and selection of proper policies and regulations to ensure the sustainable management of this unique region. The objective can be further detailed as follow:

- creating and displaying a point and line density map of boats and whales with zoom in and zoom out functionality;
- dynamically visualizing the trajectories of boats and whales by agent ID or by time point with zoom in and zoom out functionality;
- visualizing the composition of whale types and boat types using pie graphs;
- visualizing the distribution (temporal pattern) of boat activity types; and
- calculating and displaying various statistics, such as the step moving parameters of boat excursions, the time spent at observing a particular type of whale, etc.

The effort of visualization is focused on direct depiction of the original dataset and display of the results calculated from the dataset by selection, aggregation, generalization, and sampling. For example, the space-time path of whales and boats directly depicts the behaviour of whale and boats. The step length, turn angle, bearing and net displacement are displayed to represent the sailing habit/characteristics of boats by summarizing and aggregating each record by boat type and home port from the original generated dataset.

In the long term, it is hoped that this software will make a valuable contribution to the Geovisualization exploration research and assist in the development, evaluation, and selection of proper policies and regulations to ensure the sustainable management of the Marine Park and MPA region in Quebec.

1.6 Organization of the thesis

The remaining of this thesis is organized as follow. Chapter 2 describes the methodology that was used in developing the Geovisualization software in this study. Chapter 3 presents some examples of queries and subsequent result graph visualizations conducted with the software using simulated data obtained from 3MTSim for a given scenario. Chapter 4 draws conclusions of the research project followed by a list of potential improvements and recommendations for further research.

Chapter Two: Methodology

2.1 Datasets generated by 3MTSim

3MTSim generates datasets of up to thirty boat types in terms of size and passenger capacity, up to five whale types in terms of species, up to ten homeports in terms of names, for periods of simulation varying from one day to the whole season (from early May to late October) with a temporal resolution of one minute. The spatial extent of the simulated environment is approximately 43,700 km². The spatial resolution is 100 m x 100 m leading to a total of 2205 x 1983 cells. Thirty-five interrelated attributes are associated to each simulation record including: boat type, boat type name, excursion ID, company ID, captain ID, homeport, boat heading, boat activity, boat speed, visibility, whale type, and whale ID. The simulation can run with any number of whales of different types and any number of excursions of different boat types. For example, one simulation of 200 boat excursions over a period of 100 days at a temporal resolution of 1 minute generates 24,800,000 records of boat activities. The records are written in two separate comma separated values (csv) files. One file contains the records from the whale model. The other file contains the records from the boat model.

2.2 Challenges and strategies of software design

As reviewed in the previous chapter, interactive visual analysis has become a popular method to explore and investigate large spatio-temporal datasets. This process requires the selection and aggregation of datasets by time, space and attribute, the appropriate visual encodings, and the flexibility of interactions between datasets and visualization. The major challenge is that the speed of rendering and responses to user interactions may become too slow. To handle this issue, the software was designed with the following several strategies:

- Design an efficient time-based spatio-temporal database module instead of flat txt format data file to store the large spatio-temporal datasets. This facilitates the interactions between the users and the datasets.
- Provide a flexible interface to allow users to query and visualize the data by subset, for example by boat type, whale type, home time, and different time intervals of the whole travel season. This decreases the volume of data for rendering.
- To match possible queries and the resulting visualizations, the entire raw dataset is split and reorganized into smaller tables. Each table stores the associated attributes needed for the visualization.
- The software is designed as dynamic software that adapts to different parameter setups of the 3MTSim simulation. This means that the software must be robust and able to handle changeable numbers of boat types, excursions, whale types, and homeports.

The visualization of some data requires extensive calculation, for example, the calculation of the average turn angle defined as the number of turns between two straight sailing paths that have a common point connected (Fig 2.8). To calculate the turn angle at any time point, the geographical locations of both the previous and after time points are needed. If each excursion is two hours long, 140 geographical locations will need to be retrieved. With 200 excursions, then the retrieved dataset will include 140*200 records. There will be extensive aggregation of every time point. In order to reduce the response time to users, all functionalities are separated into two independent steps: calculation and rendering. This avoids redundant computations.

2.3 Software architecture

The Geovisualization software consists of two modules, respectively called *Database* and *Application*. The software used for the database module is MySQL 5.0 Community Server, a popular open-source database server with consistent fast performance, high reliability, and ease

of use that can run on more than 20 platforms. It provides a high level of flexibility for this project. The application module was developed in Java JDK 6 with the running environment JRE 6 Update 17. Using an interactive visual interface, this module offers users the functionality to interactively query the database and visualize the spatio-temporal activity patterns of marine mammals and boats for specific periods of time or for the entire tourist season. These two modules are linked through JDBC (Java Database Connectivity) and mysql-connector-java-5.1.6. JDBC is an API for the Java programming language that defines how a user may access a database. It also provides methods for querying and updating the database. Mysql-connector-java-5.1.6 converts JDBC calls into the network protocol used by the MySQL database and allows the application module to interact with MySQL for accessing the datasets in the database. A diagram of the software architecture is shown in Figure 2.1.

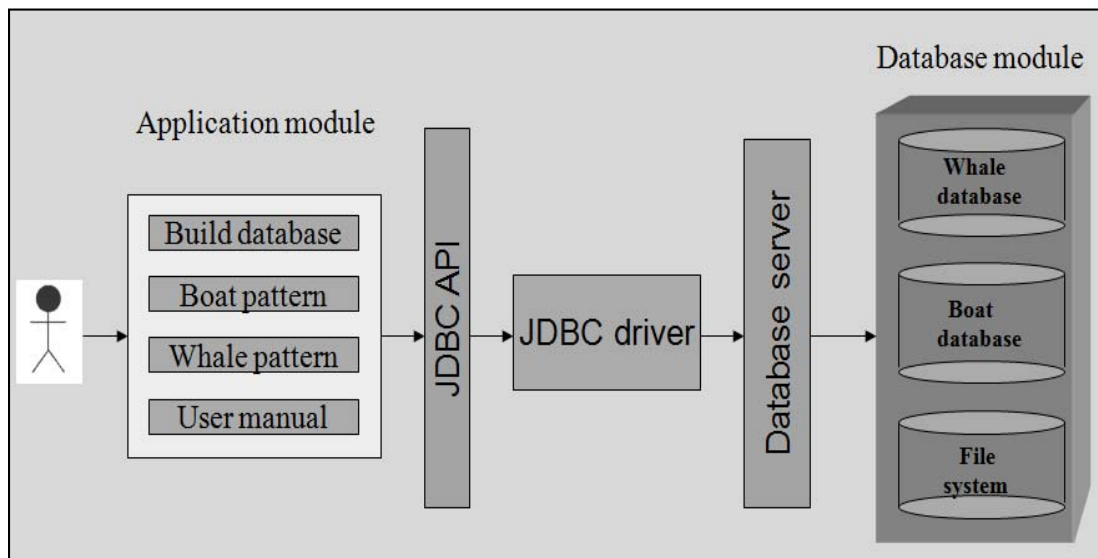


Figure 2.1 Software architecture

2.4 Database module

This section describes the database design and the main components of the database module, the tables in each component, and the attributes in each table.

2.4.1 Database design

The database is designed based on the retrieval tasks of the visualization functions in the application module. All retrieval tasks for visualizing the behaviour pattern of whales and sailing boats in the application module can be classified in the following several types:

- retrieving whale types and corresponding whale agents (ID), boat types and corresponding boat agents(ID) and departing home ports during the whole simulation season;
- retrieving geographical locations change based on whale types, boat types and corresponding departing homeports at a given time or time interval;
- retrieving the sailing speed, duration of excursion, duration of observing particular whale by boat types and home ports during the whole sailing time period;
- retrieving activity types (back to port, not in activity, observing whales, sailing, observing seals, observing landscape features) of boats during the whole sailing time period; and
- retrieving geographic occurrence by whale types, boat types and departing home ports at a given time/time interval.

All of the above retrieval tasks are time related, either at a given time or a given time interval. Consequently, this database fits the design of a time-based spatio-temporal database. Location in time is the primary organizational basis for the datasets. Each location in time along the time-line ($t_0, t_1 \dots, t_n$) is associated with a particular set of features, for example activity type = observing whale, observing whale type is minke, sailing speed is 20 knots, etc.

For the last retrieval task mentioned above, the resulting data of the total occurrence of boats and whales at every grid cell (100*100 m) are stored for further visualization using the grid model presented in Section 1.4.4. That means that only the occurrence data of boats and whales are stored, which speeds up the retrieval process for future visualizations. These data are stored as a flat txt file, and the format of the data is always latitude longitude, number of occurrence.

2.4.2 Components of the database module

The major challenge of the high performance database is to provide low latency retrieval when using a complex selection process that involves very large spatio-temporal datasets. For example, the retrieval of step length, turning angle, bearing and net displacement of all excursions of a particular boat type departing from a particular home port can be time consuming if the database is not well designed. Because the whole dataset generated by 3MTSim can amount to millions of records with the time resolution of one minute, the result of the query is also a very large number of records. In order to speed the query process, large datasets are split into small tables based on whale types, boat types and departing homeports. All these tables are organized into three main components: the *whale database*, the *boat database* and the *file system* (Fig.2.1).

The *whale database* is a relational database containing the whale behaviour data. Some of the data are generated by the whale model of 3MTSim while some are the calculated results for the visualization of whale behaviour in the application module. The data are organized into different tables based on the whale types and the visualization purposes. Tables of this component are described in Section 2.4.4.

The *boat database* is a relational database containing the boat behaviour data. Some of these data are generated by the boat model of 3MTSim while some are the calculated results for the visualization of boat behaviour in the application module. The data are organized into tables based on boat types, departing home ports, and the corresponding attributes to visualize or calculate. Tables of this component are described in Section 2.4.5.

The *file system* is a predefined folder created on the hard disk to store multiple density matrices of dimension 2205 x 1983 in a text format. These matrices are used to visualize the occurrence patterns of whales or boats. Using this approach to store occurrence patterns is more efficient than storing in the database. The files of this component are described in Section 2.4.6.

2.4.3 Types of datasets in the database module

The datasets in the three components: whale database, boat database, and file system can be further classified into the following four types of datasets (Fig. 2.2):

- *Density maps of whales and boats* are large txt files containing matrix representations of occurrence patterns in the File system. The number of corresponding rows and columns represents the occurrence of whales or boats in the cell location. The rows and columns were converted to corresponding latitude and longitude based representations in the 100 m by 100 m resolution of the study area. Saving in a flat text file instead of the database server reduces querying time and improves the performance of the software.
- *Statistical datasets* are tables for storing calculated data of different problem domains. Some tables are for final visualization, some tables contain the data that are queried frequently in many functions. For example, some tables store the pre-calculated total number of boats and the total number excursions for each boat type. These tables are used to visualize the proportion of boat types from different homeports in the simulation.

- *Classified raw datasets* are tables for storing split raw dataset generated by 3MTSim by whale types or boat types. Each table contains only subset of the whole raw dataset.
- *Raw dataset generated by 3MTSim* are the un-processed simulation model outputs stored in the database as a backup for users to browse.

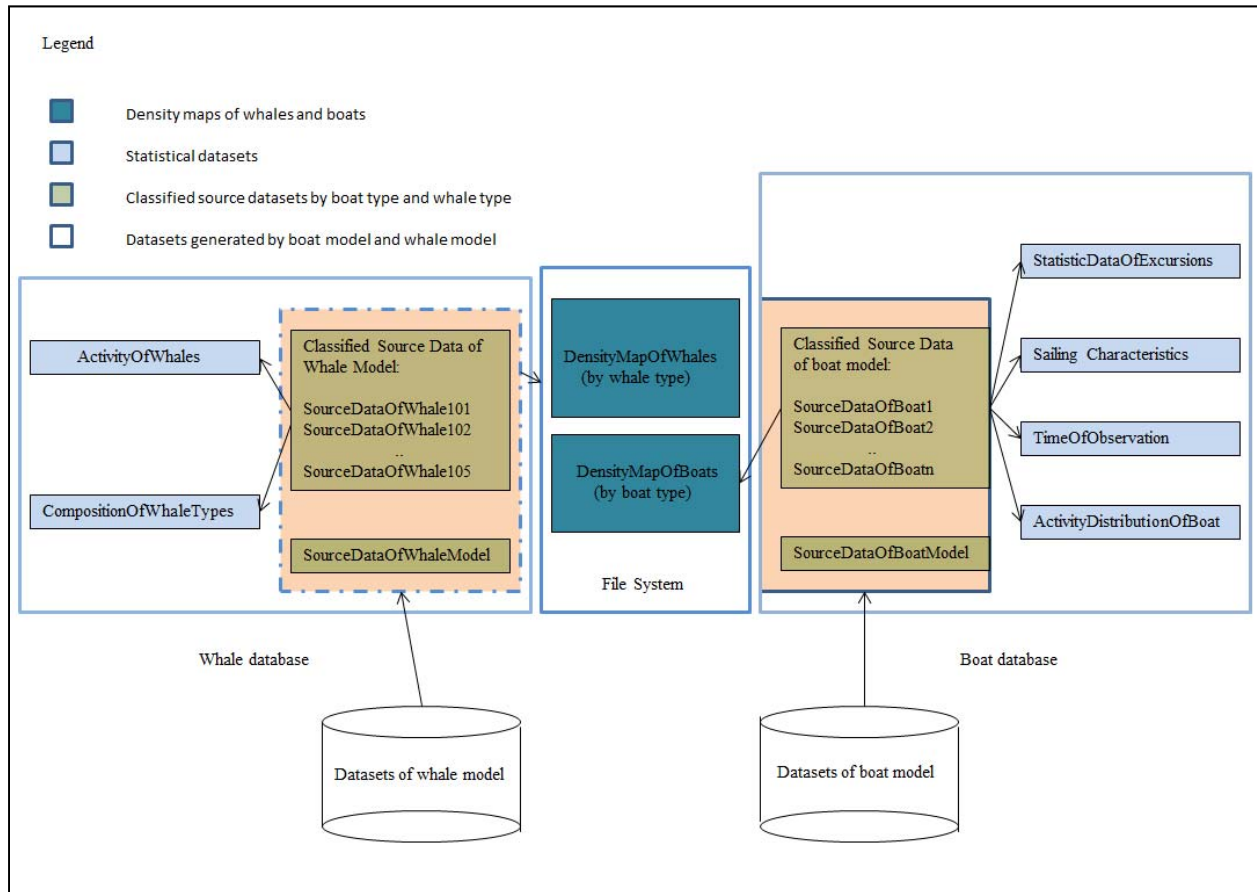


Figure 2.2 Database structure of database module

2.4.4 The tables in the whale database

This section describes all the tables and the data structures in the whale database.

2.4.4.1 Table SourceDataOfWhaleModel and its data structure

The table SourceDataOfWhaleModel contains the raw simulation output data of whales from the whale model of 3MTSim. This table is the first table being created in the whale database and serves as a backup of the raw csv dataset of whale activities generated by 3MTSim. In addition, if any other tables in the whale database collapse, it can be recovered from this table and the original csv dataset does not have to be reloaded and parsed again. The data structure is presented in Table 2.1.

Table 2.1 Source data structure generated by the whale model

Row No.	Attribute	Structure	Definition
1	Id	INT	Index of each record
2	Experimentation	INT	Index of experimentation
3	Replication	INT	Index of replication
4	TICK	DOUBLE	Accumulated passed time in minutes since the start of the replication; the start tick is 0, each tick represents one minute
5	AgentType	INT	Codes of whale types: 101 = minke 102 = fin 103 = blue 104 = humpback 105 = beluga
6	Agent_ID	INT	ID of agent (whale)

7	X_UTM	DOUBLE	Position X of the whale in the UTM zone 19N (NAD83)
8	Y_UTM	DOUBLE	Position Y of the whale in the UTM zone 19N (NAD83)
9	Groupe_Baleine	INT	Group size (total number of agents for each type of whale)
10	Tide	INT	Level of tide

2.4.4.2 Table ClassifiedSourceDataOfWhaleModel and its data structure

Each ClassifiedSourceDataOfWhaleModel table contains the raw simulation dataset associated to each whale type (101 – 105). The code 101 to 105 represents respectively the minke, fin, blue, humpback, and beluga whales. This type of table has the same data structure of the table csvOfWhale, except that the attribute AgentType is the same for each table. For example, the data in the AgentType column of csvwhale101 are all 101, which is the type code of a fin whale. The data in column AgentType of csvwhale105 are all 105, which is the type code of a beluga whale. Splitting csvofwhale into csvwhale101 to scvwhale105 by whale type reduces the size of the dataset to select from and reduces the retrieval time (Table 2.2).

Table 2.2 Classified source data structure generated by the whale model

Row No.	Attribute	Structure	Definition
1	Id	INT	Id of record

2	Experimentation	INT	Id of experimentation
3	Replication	INT	id of replication
4	TICK	DOUBLE	Time Accumulated from 0; each tick represents one minute
5	AgentType	INT	Codes of whale types: AgentType = 101 for csvwhale101(minke) AgentType = 102 for csvwhale102(fin) AgentType = 103 for csvwhale103(blue) AgentType = 104 for csvwhale104 (humpback) AgentType = 105 for csvwhale105(beluga)
6	Agent_ID	INT	ID of agent whale
7	X_UTM	DOUBLE	Position X of the whale in the UTM zone 19N (NAD83)
8	Y_UTM	DOUBLE	Position Y of the whale in the UTM zone 19N (NAD83)
9	Groupe_Baleine	INT	Group size
10	Tide	INT	Level of tide

2.4.4.3 Table ActivityOfWhales and its data structure

The table ActivityOfWhales contains some typical activity attributes of each whale agent such as the activity boundary (minimum and maximum easting, minimum and maximum northing) and the time period stayed in the park and MPA. This table is created when the raw csv dataset is loaded into the database. These attributes will be queried frequently in the application module. The objective of storing these attributes in a separate table is to decrease the querying time from the original large dataset. The data structure is presented in Table 2.3.

Table 2.3 Data structure of whale activities

Row No.	Attribute	Structure	Definition
1	AgentID	INT	ID of whale agent
2	timeStart	INT	Time of coming to the park
3	timeEnd	INT	Time of leaving the park
4	timePeriod	INT	timeEnd – timeStart, period of staying in the park
5	Minx	DOUBLE	Minimum easting
6	maxX	DOUBLE	Maximum easting
7	minY	DOUBLE	Minimum northing
8	maxY	DOUBLE	Maximum northing
9	typeOfWhale	TINYINT	Codes of whale types: 101: minke

			102: fin 103: blue 104: humpback 105: beluga
--	--	--	---

2.4.4.4 Table CompositionOfWhaleTypes and its data structure

The table CompositionOfWhaleTypes stores the total number of whales for each type of whale and its corresponding percentage. The table is created when the raw csv dataset is loaded into the database. The objective of storing this statistical result in a separate table is to speed the query process. The data structure is presented in Table 2.4.

Table 2.4 Data structure of composition of whale types

Row No	Attribute	Structure	Definition
1	TypeOfWhale	SMALLINT(4)	Codes of whale types: 101 = minke 102 = fin 103 = blue 104 = humpback 105 = beluga
2	NumOfWhale	INT	Total number of whales in the simulation experiment
3	Percentage	DECIMAL(3,2)	Percentage of each whale type in the

			simulation experiment
--	--	--	-----------------------

2.4.5 Tables in the boat database

This section lists all the tables and their data structure in the boat database.

2.4.5.1 Table SourceDataOfBoatModel and its data structure

The table SourceDataOfBoatModel contains the raw simulation datasets of boats excursions from the boat model of 3MTSim. This table serves as a backup of the raw csv dataset of boat excursions generated by the 3MTSim. In addition, if any other tables in the boat database collapse, it can be recovered from this table and the original csv dataset does not have to be reloaded and parsed again. The data structure is presented in Table 2.5.

Table 2.5 Source data structure generated by the boat model

Row No.	Attribute	Structure	Definition
1	ID	INT	Row number of each record
2	Tick	DOUBLE	Time (tick) Accumulating from 0, each tick represents one minute
3	BoatType	INT	Codes of boat types 1 = small zodiac, 2 = medium zodiac 3 = large boat, 4 = kayak

			Other boat types could be added if needed
4	ExcID	INT	ID of the excursions
5	CompanyID	INT	The company to which the boat and captain belongs
6	CaptainID	INT	ID of the captains
7	IsFree	INT	0 = the captain is at sea, 1 = the captain is not at sea
8	BoatActivity	INT	Codes of boat activities -1 = going back to port 0 = not in activity 1 = observing whales 2 = sailing 3 = observing seals 4 = observing landscape features
9	X_UTM	DOUBLE	Position X of the boat in the UTM zone 19N (NAD83)
10	Y_UTM	DOUBLE	Position Y of the boat in the UTM zone 19N (NAD83)
11	BoatSpeed	INT	Speed of the boat (knot)
12	BoatHeading	DOUBLE	Heading of the boat (degree)

13	Visibility	INT	Distance of visibility (meter)
14	TouristPeriod	INT	Codes of tourist periods 1 = before July15 2 = July 15-Aug. 15 3 = after Aug. 15
15	Tide	INT	Tide levels: low, flowing, high, ebbing
16	NumOfWhales	INT	Number of visible whales within the visibility extent. All species are aggregated at the moment but this can be changed if needed.
17	TargetSpecies	INT	ID of the targeted species, which are currently observed during an excursion. The values are those proposed by the whale model
18	BoatTypeName	INT	Name of boat types: zodiac, medium zodiac, large boat, kayak, etc.
19	Month	INT	Month of excursion: June, July, August, September and October

20	DayOfMonth	INT	Day of the month of excursion
21	Hour	INT	Hour of the day of excursion
22	Minute	INT	Minute of the hour of excursion
23	PortID	INT	Codes of home ports 1 = Tadoussac, 2 = Bergeronnes, 3 = Escoumins/Anse aux Basques, 5 = Portneuf/mer, 10 = Saint- Siméon
24	HomePortName	VARCHAR(64)	Names of home ports : Tadoussac, Bergeronnes, Escoumins/Anse aux Basques, Portneuf/mer, Saint-Siméon

2.4.5.2 Table ClassifiedDataOfBoatModel and its data structure

Each table ClassifiedDataOfBoatModel contains the raw simulation dataset associated to each boat type (1 to n). The simulation parameters define the number of boat types listed in this dataset. If there are thirty types of boats setup in the simulation, there will be thirty split tables associated to each boat type. This type of table has the same data structure of the table csvofBoat, except that the attribute AgentType is the same for each table. For example, the data in column AgentType of csvboat1 are all 1, which is the type code of the zodiac. The data in column AgentType of csvboat is 2, which is the type code of medium zodiac. Splitting the table

into a number of smaller tables by boat types reduces the retrieval latency of query, because the selections are subsets of the large dataset (Table 2.6).

Table 2.6 Classified source data structure generated by the boat model

Row No.	Attribute	Structure	Definition
1	ID	INT	Row number of each record
2	Tick	DOUBLE	Time (tick) Accumulating from 0; each tick represents one minute
3	BoatType	INT	Codes of boat types 1 = zodiac, 2 = medium zodiac 3 = large boat, 4 = kayak etc.
4	ExcID	INT	ID of the excursions
5	CompanyID	INT	The company to which the boat and captain belongs
6	CaptainID	INT	ID of the captains
7	IsFree	INT	0 = the captain is at sea 1 = the captain is not at sea
8	BoatActivity	INT	Codes of boat activities: -1 = going back to port 0 = not in activity

			<p>1 = observing whales</p> <p>2 = sailing</p> <p>3 = observing seals</p> <p>4 = observing landscape features</p>
9	X_UTM	DOUBLE	Position X of the captain in the UTM zone 19N (NAD83)
10	Y_UTM	DOUBLE	Position Y of the captain in the UTM zone 19N (NAD83)
11	BoatSpeed	INT	Speed of the boats (knots)
12	BoatHeading	DOUBLE	Heading of the boats (degrees)
13	Visibility	INT	Distance of visibility (meters)
14	TouristPeriod	INT	<p>Codes of tourist periods:</p> <p>1 = before July15</p> <p>2 = July 15-August 15</p> <p>3 = after August 15</p>
15	Tide	INT	Four levels of tide (low, flowing, high, ebbing)
16	NumOfWhales	INT	Number of visible whales within the visibility extent. All species are aggregated at the moment but that will evolve.
17	TargetSpecies	INT	IDs of the targeted species that are

			currently observed during the excursion. The values are those proposed by the whale model.
18	BoatTypeName	VARCHAR(64)	Name of boat types, for example, zodiac, medium zodiac, large boat, kayak, etc.
19	Month	TINYINT	Month of excursion: June, July, August, September and October
20	DayOfMonth	TINYINT	Day of the excursion month
21	Hour	TINYINT	Hour of the excursion day
22	Minute	TINYINT	Minute of the excursion hour
23	PortID	INT	Codes of home ports: 1=Tadoussac, 2=Bergeronnes, 3= Escoumins/Anse aux Basques, 5 = Portneuf/mer, 10 = Saint-Simeon There will be more homeports being simulated in the 3MTSim model.
24	HomePortName	VARCHAR(64)	Names of home ports : Tadoussac, Bergeronnes, Escoumins/Anse aux Basques, Portneuf/mer, Saint-Siméon

2.4.5.3 Table *StatisticDataOfExcursions* and its data structure

The table *StatisticDataOfExcursions* contains the sailing characteristics of each boat excursion, such as the sailing boundary of each boat excursion (minimum and maximum easting, minimum and maximum northing), start and end time of each excursion, minimum and maximum speed, company ID, etc. This table is created when the raw csv dataset is loaded into the database. The attributes in this table are queried frequently in the application module. Storing these attributes in a separated table reduces the querying time because the data size to query from this table equals the total number of excursions. However, the size of the raw dataset is about the total number of excursions x the average number of records of each excursion. With a time resolution of one minute, a two-hour excursion includes 120 records. The data structure of this table is presented in Table 2.7.

Table 2.7 Data structure of boat excursions

Row No.	Attribute	Structure	Definition of Attribute
1	ExcID	INT	Excursion ID
2	ExcStart	DOUBLE	Excursion start time
3	ExcEnd	DOUBLE	Excursion end time
4	TypeOfBoat	TINYINT	Codes of boat types 1 = small zodiac, 2 = medium zodiac 3 = large boat, 4 = kayak
5	BoatTypeName	VARChARE(64)	Names of Boat Types

6	ExcPeriod	INT	ExcEnd – ExcStart
7	CompanyID	TINYINT	Company ID of the excursion
8	MinXUTM	DOUBLE	Minimum easting in the UTM zone 19N (NAD83)
9	MaxXUTM	DOUBLE	Maximum easting in the UTM zone 19N (NAD83)
10	MinYUTM	DOUBLE	Minimum northing in the UTM zone 19N (NAD83)
11	MaxYUTM	DOUBLE	Maximum northing in the UTM zone 19N (NAD83)
12	AveSpeed	INT	Average speed (knots)
13	MinSpeed	INT	Minimum speed (knots)
14	MaxSpeed		Maximum speed (knots)
15	ExcLength	DOUBLE	Trajectory length of excursions $\sum_i \sum_j \text{SQRT}_j((XUTM_i - XUTM_j)^2 + (YUTM_i - YUTM_j)^2)$
16	fartherDistance4	DOUBLE	The farthest distance from the start location $\text{MAX}\{ \text{SQRT}((XUTM_i - XUTM_{\text{Start}})^2 + (YUTM_i - YUTM_{\text{Start}})^2) \}$

			$(XUTM_i - XUTM_{Start}) +$ $(YUTM_i - YUTM_{Start}) *$ $(YUTM_i - YUTM_{Start})) \}$
17	farthestEasting	DOUBLE	Farthest XUTM
18	farthestNorthing	DOUBLE	Farthest YUTM
19	CaptainID	DOUBLE	Captain ID of the excursion
20	TouristPeriod	TINYINT	Codes of tourist periods 1 = before July15 2 = July 15-August 15 3 = after August 15
21	startMonth	TINYINT	Start month of excursion
22	startDay	TINYINT	Start day of the start month
23	startHour	TINYINT	Start hour of the start day
24	startMinute	TINYINT	Start minute of the start hour
25	endMonth	TINYINT	End month of excursion
26	endDay	TINYINT	End day of the end month
27	endHour	TINYINT	End hour of the end day
28	endMinute	TINYINT	End minute of the end hour
29	visiblity	INT	Visibility
30	PortID	INT	Codes of home ports : 1 = Tadoussac, 2 = Bergeronnes, 3 = Escoumins/Anse aux Basques,

			5 = Portneuf/mer, 10 = Saint-Siméon
31	HomePortName	VARCHAR(64)	Names of home ports : Tadoussac, Bergeronnes, Escoumins/Anse aux Basques, Portneuf/mer, Saint-Siméon

2.4.5.4 Table SailingCharacteristics and its data structure

The table SailingCharacteristics stores the average step length, turn angle (degree), bearing (degree to north) and the net displacement. The first ‘?’ is replaced by the code of boat type. The second ‘?’ is replaced by the code of home port. The total number of this type of table is the combination of boat types and number of port types. Assume that there are n boat types and m homeports set up for the simulation; then there will be n by m tables in total. The calculation of these data is time-consuming. For example the calculation of turning angle needs the geographic location of three time points: the previous location, the current location, and the next location. The time resolution of the record is one minute. If the duration of the excursion is two hours, then there will be 140 records in total. The objective when creating this table is to speed up the query speed. The data structure of this table is presented in Table 2.8.

Table 2.8 Data structure of boat sailing characteristics

Row No.	Attribute	Structure	Definition

1	ID	INT	Row number of each record
2	boatType	INT	Codes of boat types 1 = zodiac, 2 = medium zodiac 3 = large boat, 4 = kayak etc.
3	Homeport	VARCHAR(64)	Port names boat excursions depart from
4	Feature	VARCHAR(64)	Step Length (m), Turn Angle (degree), Bearing (degree to North), Net Displacement (m)
5	stardardDeviation	DOUBLE	Standard deviation
6	Var	DOUBLE	Variance
7	Median	DOUBLE	Median number of corresponding feature
8	Maximum	DOUBLE	Maximum number of corresponding feature
9	Minimum	DOUBLE	Minimum number of corresponding feature

2.4.5.5 Table TimeOfObservation and its data structure

The table TimeOfObservation stores the duration that a particular boat type spends observing a particular whale type and the duration of the entire excursion. Each record contains the combination of a particular boat type, whale type, and distance between whale and boat. The data structure is presented in Table 2.9.

Table 2.9 Data structure of boat observing whales

Row No.	Attribute	Structure	Definition
1	boatType	INT	Codes of boat types 1 = zodiac, 2 = medium zodiac 3 = large boat, 4 = kayak etc.
2	whaleType	INT	Codes of whale types: 101 = minke 102 = fin 103 = blue 104 = humpback 105 = beluga
3	Distance	INT	Particular distance between boats and whales specified users
4	DuraOfExc	INT	Duration of boat excursion
5	DuraOfObserve	INT	Duration of boat spending at observation
6	Percentage	DECIMAL(3,2)	$(\text{timeOfObserve}/\text{ttlTimeOnSea})^*$ 100

2.4.5.6 Table ActivityDistributionOfBoat and its data structure

The boat activities can be classified into six types. Each type of activity is assigned a type code. The attribute Activity Type in the table (table ActivityDistributionOfBoat) lists all the activity types and their corresponding type code. For example, -1 means going back to port, 3 means observing seals, 1 means watching whales, etc. This table stores the calculated result of the temporal distribution of the six types of activity during an excursion. The number of tables depends on the combination of number of boat and port in the simulation. The data structure of this table is presented in Table 2.10.

Table 2.10 Data structure of boat activity distributions

Row No.	Attribute	Structure	Definition
1	Id	INT	Row number
2	ActivityType	INT	Codes of boat activities -1 = going back to port 0 = not in activity 1 = watching whales 2 = sailing 3 = observing seals 4 = observing landscape features
3	Bin1	INT	1 st bin after normalization
4	Bin2	INT	2 nd bin after normalization

5	Bin3	INT	3 rd bin after normalization
6	Bin4	INT	4 th bin after normalization
7	Bin5	INT	5 th bin after normalization
8	Bin6	INT	6 th bin after normalization
9	Bin7	INT	7 th bin after normalization
10	Bin8	INT	8 th bin after normalization
11	Bin9	INT	9 th bin after normalization
12	Bin10	INT	10 th bin after normalization

2.4.6 The file system

The file system is a default directory called “..\geovisualization\data” created under the current directory of the software package. Density maps are large files with a simple data structure. The following density maps are created and stored in this directory and visualized in the application module:

- Point density maps of resolution 100 *100 m by whale type in the whole simulation season;
- Point density maps of resolution 100*100 m for all activities by boat type and time period;
- Point density maps of resolution100*100 m for the activity ‘whale observation’ by boat type and time period;
- Line density maps of resolution 100*100 m for all activities by boat type, and time period.

2.5 Application module

This section describes the main components of the application module and its functionalities.

2.5.1 Functionalities in the Application module

The functions in the application module are classified into four groups: *Build database*, *Whale pattern*, *Boat pattern* and *User manual* as shown in Figure 2.3. The algorithms of these functions are described in the following section.

The function Build database consists in building, maintaining and managing the relational database; it includes two functions:

- Build boat database, and
- Build whale database.

Whale pattern retrieves and visualizes whale activity patterns through three functions:

- Composition of whale type,
- Point density map of whales by whale type, and
- Trajectory of whales.

Boat pattern retrieves and visualizes boat patterns through seven functions:

- Composition of excursions by boat types and home ports,
- Point and line density map of boats,
- Trajectory characteristics of boat excursions by boat type and port type,
- Boat observation on specific whale type by boat type,
- Repartition of boat activities by boat type and home port type,
- Trajectory of boats,
- Analysis of speed, duration of excursion, and distance of excursion by boat type and home port.

The user manual documents the Geovisualization software for users' reference.

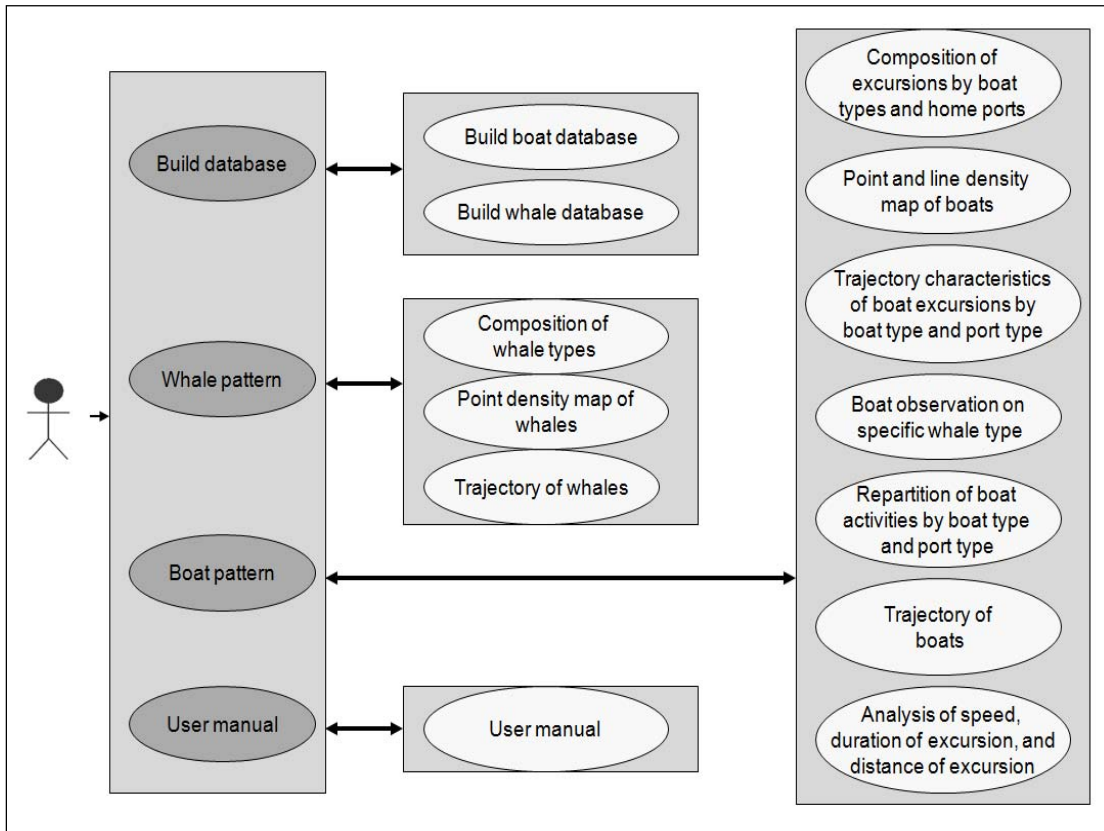


Figure 2.3 Diagram of the application module

2.5.2 Algorithms of the application functions

This section explains the algorithms of each application functions.

2.5.2.1 Build whale database

Build database creates the whale database schema, parses the raw csv dataset of whale, aggregates the whales by whale types, extract typical activity characteristics of each individual whale, and populates the tables csvofwhale (Section 2.4.4.1), csvwhale101 to csv105 (Section 2.4.4.2), statofwhaleactivity (Section2.4.4.3), and whaletypecomposition (Section2.4.4.4) in the whale database.

2.5.2.2 Build boat database

Build boat database creates the boat database schema, parses the raw csv dataset of boat, aggregates the boat excursions by boat types, extracts the typical activity characteristics of excursions (such as the start and end time, maximum activity boundaries, and average speed), and populates the tables csvofboat (Section 2.4.5.1), csvboat1 to csvboatn (Section 2.4.5.2) and statofboatexcursion (Section 2.4.5.3).

2.5.2.3 Composition of whale types in simulation

This function displays the proportion of whale types (table whaletypecomposition, Section 2.4.4.4) and the corresponding number of whales of each whale type. A dynamic color table is built to ensure its extensibility to account for any number of whale types.

2.5.2.4 Composition of excursions by boat types and home ports

This function displays the number of excursions done by particular boat types departing from particular home ports. A dynamic color table is built to ensure its extensibility and to account for any number of boat types, home ports and whale types in a simulation. There are two types of data to display: (1) for chosen boat types, the composition of excursions of each boat type and the composition of home ports the excursions start from; (2) for chosen home ports, the composition of home ports that are used and the composition of excursions of each boat type.

2.5.2.5 Point density map of whales

Density maps are two-dimensional matrices in txt format. Each value in the matrix represents the frequency of occurrences in a particular unit area. The number of columns and rows of each

matrix depends on the resolution of the maps. With a resolution of 100 m, the density map has a dimension of 2205 x 1193.

The point density maps display the density locations of whale occurrence. The density maps of whales are calculated based on the whale type for the whole season, because whales stay in the park and MPA during the whole summer. The algorithm for calculating the frequency of occurrence is illustrated in Figure 2.4.

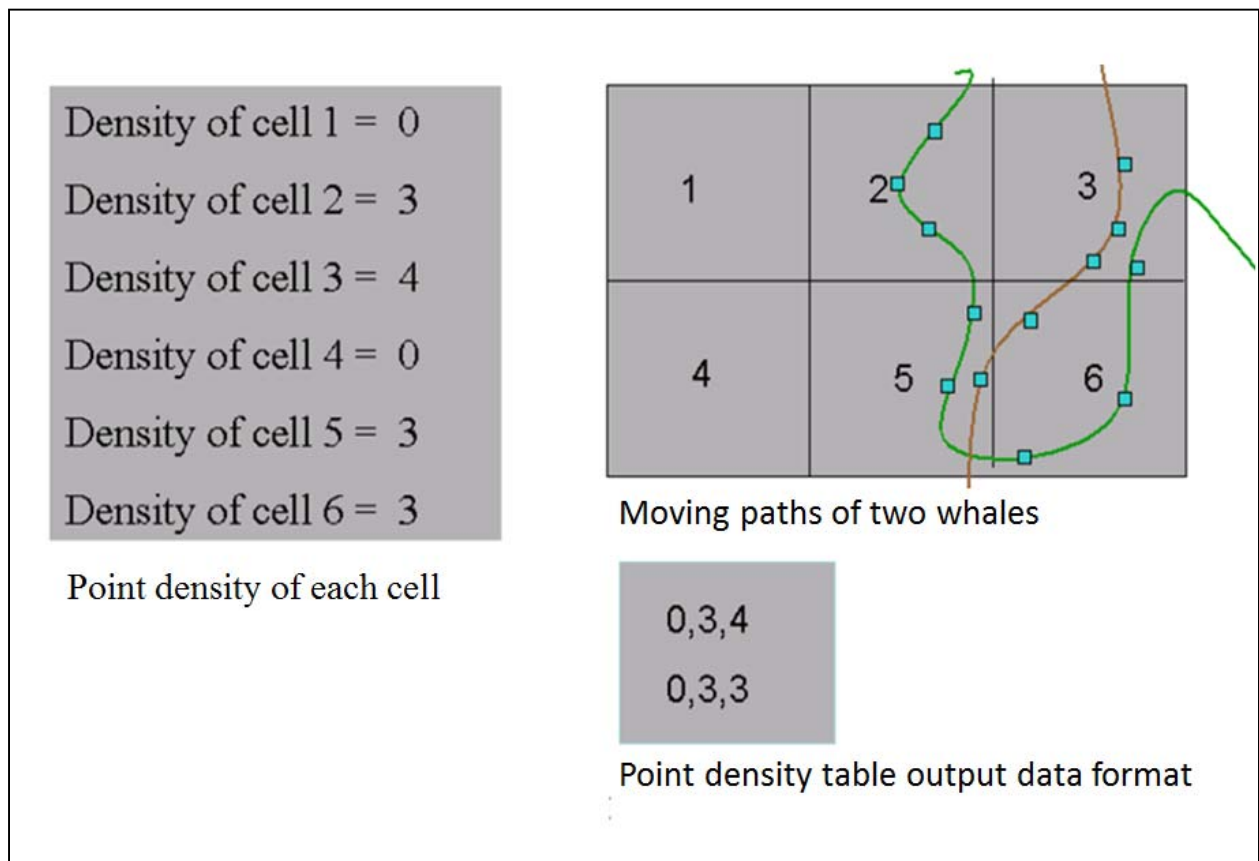


Figure 2.4 Calculation of the point density map of whale types

2.5.2.6 Point density map of boat type (all activities)

This function displays the density locations of excursions by boat type, season and type of boat activities (all type of activities). Each value in these point density maps represents the frequency of occurrence per unit area of a particular boat type. The number of columns and rows of each matrix depends on the resolution of the maps. With a resolution of 100 m, the density map has a dimension of 2205 x 1193. The table is named with the following convention:

pdmAllActR”r?”Boat”b?”StartM”m?”D”d?”H”h?”Min”min?”EndM”m?”D”d?”H”h?”Min”min?”

For example, the table pdmAllActR100Boat1SartM6D15H0Min1EndM9D30H24min59 stores the frequencies of occurrence per 100 m² of boat type 1 (zodiac) from June 15, 00:00:00 to Sept. 30, 24:59:00, in which

1. “pdmAllAct” is the point density map of boats for all activities.
2. “R100” is the spatial resolution (per unit area), $r = 100$.
3. “Boat1” is the boat type = 1 (zodiac), $b = 1$.
4. “SartM6D15H0Min1” is the start time, $m = 6$ (start month), $d = 15$ (start day), and $h = 0$ (start hour) $min = 1$ (start minute).
5. “EndM9D30H0min1” is the end time, $m = 9$ (end month), $d = 30$ (end day), and $h = 24$ (end hour). $Min = 59$ (end minute).

The algorithm for calculating the frequency of occurrence is illustrated on Figure 2.5.

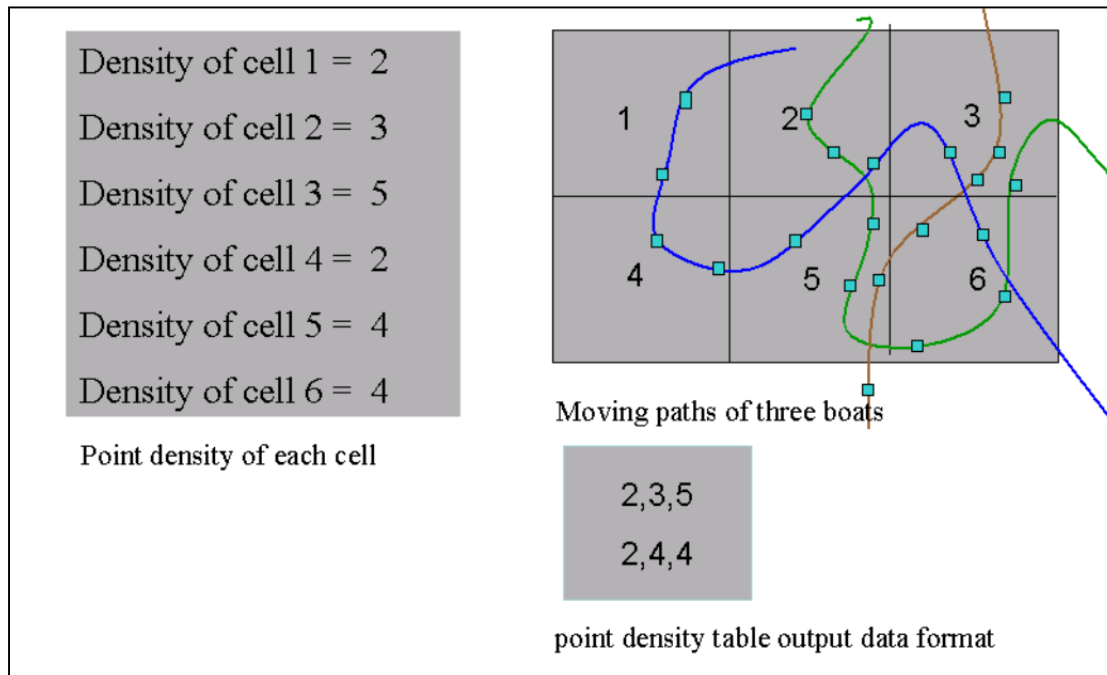


Figure 2.5 Calculation of the point density map of boat activity types (all activities)

2.5.2.7 Point density map of boat type (Whale Observation)

Each value in these density maps represents the frequency of whale observation (boat activity = observing whales). The number of columns and rows of each matrix depends on the resolution of the maps. With a resolution of 100 m, the density map has a dimension of 2205 x 1193. These tables are named according to the following convention:

pdmObserverR”r”Boat”b?”StartM”m?”D”d?”H”h?”Min”min?”EndM”m?”D”d?”H”h?”Min”min?”

For example, the table

pdmObserverR100Boat1SartM6D15H0Min1EndM9D30H24min59 stores the frequencies of observing whales per 100 m² of boat type 1 (zodiac) from June 15,

00:00:00 to Sept 30, 24:59:00, in which:

1. “pdmObserve”: point density map of observing whales
2. “R100”: resolution (per unit area), $r = 100$.
3. “Boat1”: boat type = 1 (zodiac), $b = 1$ (boat type = 1).
4. “SartM6D15H0Min1”: start time, $m = 6$ (start month), $d = 15$ (start day), and $h = 0$ (start hour) $min = 1$ (start minute).
5. “EndM9D30H0min1”: end time, $m = 9$ (end month), $d = 30$ (end day), and $h = 24$ (end hour). $Min = 59$ (end minute).

The algorithm for calculating the frequency of observations is illustrated on Figure 2.6.

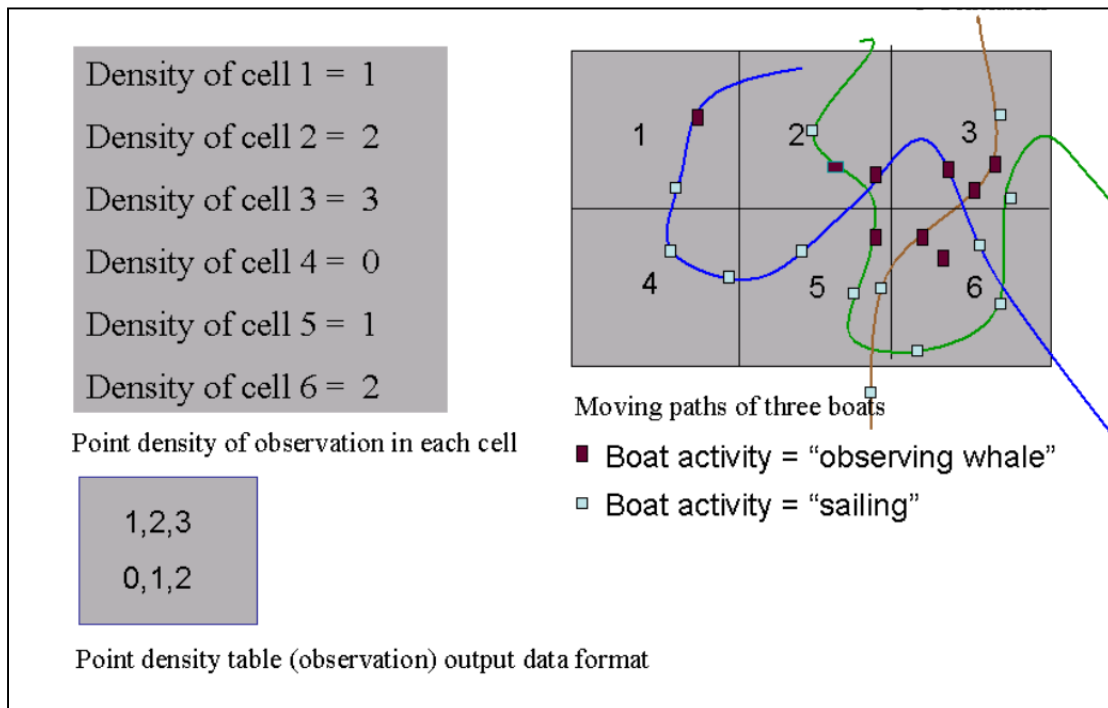


Figure 2.6 Calculation of the point density map of boat activity types (whale observation)

2.5.2.8 Line density map of boat type

Each value in the density maps represents the frequency of occurrence of excursions that forms a line. The number of columns and rows of each matrix depends on the resolution of the maps. With a resolution of 100 m, the density map has a dimension of 2205 x 1193. The locations that form a line represent the high traffic area. These tables are named using the following convention:

ldmVisitR^{r?}Boat^{b?}StartM^{m?}D^{d?}H^{h?}Min^{min?}EndM^{m?}D^{d?}H^{h?}Min^{min?}

For example, the table ldmR100Boat1SartM6D15H0Min1EndM9D30H24min59 stores the frequency of occurrences that forms a line per 100 m² of boat type 1 (zodiac), in which:

1. “ldmAllAct”: line density map of occurrence (visiting).
2. “R100”: resolution (per unit area), r[?] = 100.
3. “Boat1”: boat type = 1 (zodiac), b[?] = 1.
4. “SartM6D15H0Min1”: start time, m[?] = 6 (start month), d[?] = 15 (start day), and h[?] = 0 (start hour) min[?] = 1 (start minute).
5. “EndM9D30H0min1”: end time, m[?] = 9 (end month), d[?] = 30 (end day), and h[?] = 24(end hour). Min[?] = 59 (end minute).

The algorithm for calculating the frequency of occurrence that forms a line is presented on Figure 2.7.

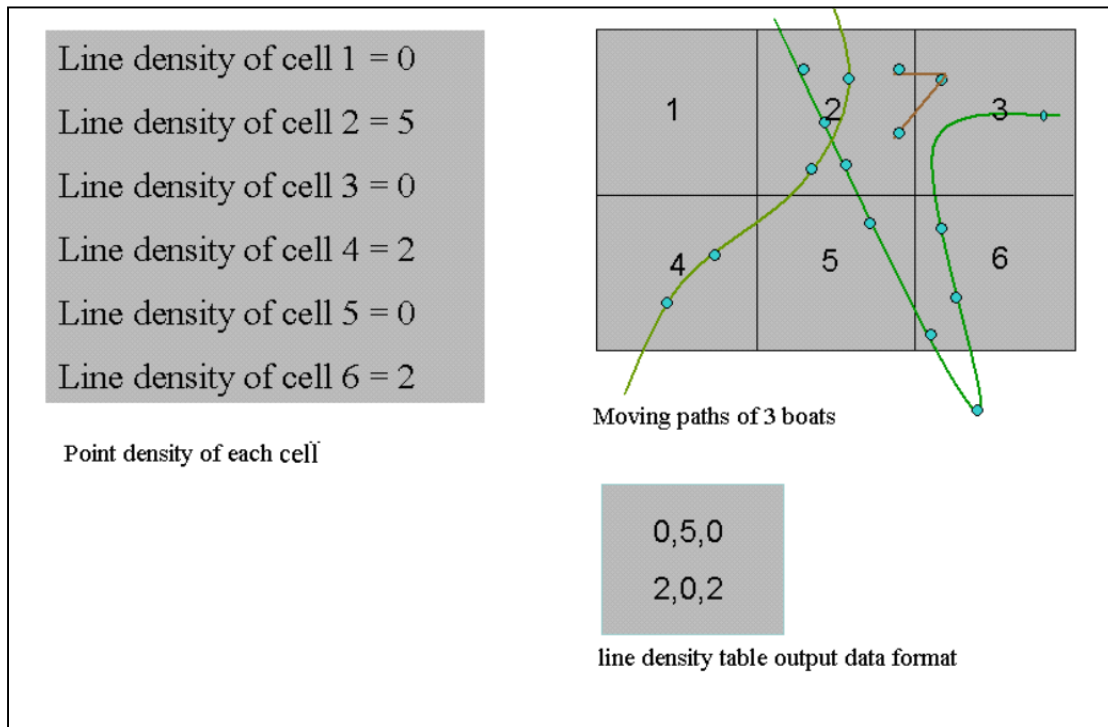


Figure 2.7 Calculation of the line density map of boat types (all activities)

2.5.2.9 Trajectory of whales

This function is to dynamically display the trajectories of any single whale by whale type and whale ID at a specified speed in millisecond; the display can be stopped at any time, because the whales stay in the park during the whole summer and the simulation can be very long. An activity boundary is draw before the simulation to provide an overall view of where the whales are active. The trajectory can be simulated at different time steps to demonstrate the diving behavior. In addition, statistical information regarding the trajectory is displayed for user's reference. Zooming in and zooming out is allowed.

2.5.2.10 Trajectory of boats

This function allows the dynamic display of the trajectories of boat excursions by time point by excursion ID. When displaying by time point, there can be multiple trajectories that are displayed at the same time. When simulating by excursion ID, only a single trajectory is displayed dynamically. Some information regarding the trajectory is displayed for users to refer to. Users are allowed to control the display speed, stop the display at any time, and zoom in and zoom out the trajectory.

2.5.2.11 Trajectory characteristics of boat excursions by boat type and port type

This function is used to calculate the average step spatial parameters (step length, turn angle, bearing and net displacement) of excursions by boat types and homeports. Since the calculation needs to aggregate the parameters of each time point of excursions, the response to the users query is very slow. Calculating for each subset, by boat type and home ports, reduces the calculation and the retrieval latency. In order not to frustrate the users, the calculation and visualization of this function are separated and the user is alerted when the calculation is finished. When the calculation is completed, the user is free to visualize or quit. However the data calculated are already stored in the database (table attributeboat?port?, Section 2.4.5.4) where they can be visualized with no more calculation needed. The concepts of step length, turn angle, bearing, and net displacement are displayed on Figure 2.8.

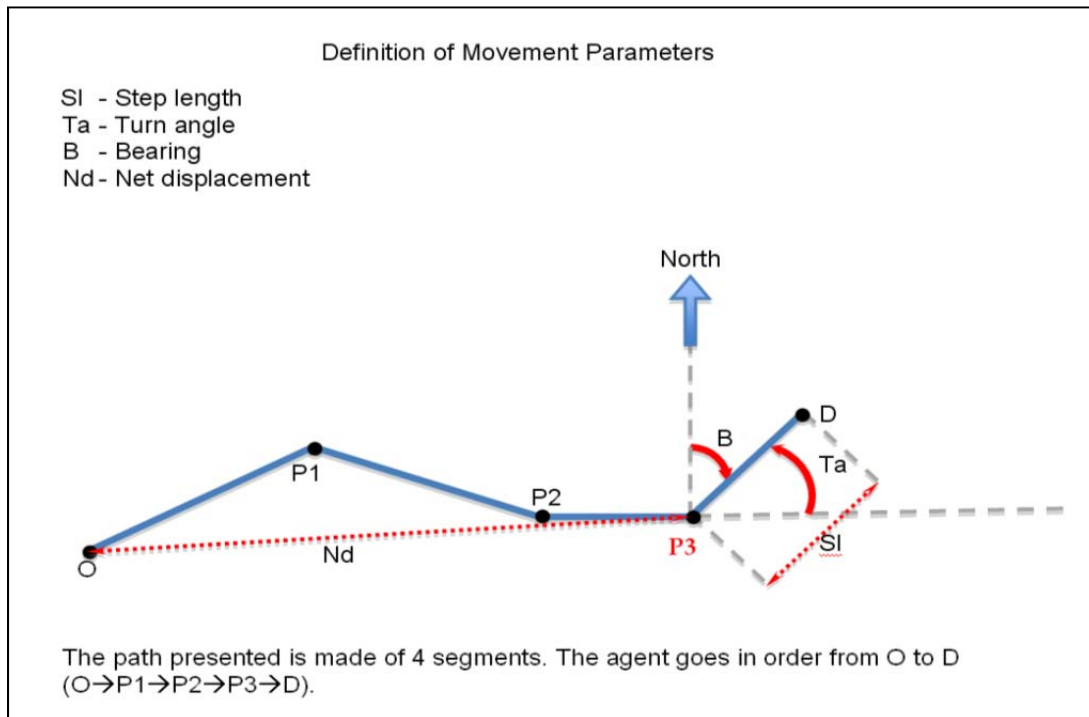


Figure 2.8 Definition of movement parameters (Source: Chion *et al.* 2011)

2.5.2.12 Repartition of boat activities by boat type and port type

This function is used to calculate and visualize the temporal distribution of boat activity types (observing whales, observing seal, observing land features, sailing, back to port, and not in activities). The calculation is performed according to the following steps:

1. Normalize sampled activities of boat excursions into ten different bins, because the duration of the excursions varies, e.g., 120 minutes, 130 minutes, etc. as illustrated in Figure 2.9.

<u>ExcID</u>		10 th min	20 th min	30 th min	40 th min	50 th min	60 th min	70 th min	80 th min	90 th min	100 th min	110 th min	120 th min	130 th min
<u>Exc 1</u>	Activity	2	2	2	1	1	1	1	2	3	2	1	2	-
	Bin	1/10	2/10	3/10	3/10	4/10	5/10	6/10	7/10	8/10	8/10	9/10	10/10	-
<u>Exc 2</u>	Activity	2	2	1	1	2	2	1	1	2	2	2	4	2
	Bin	1/10	2/10	2/10	3/10	4/10	5/10	5/10	6/10	7/10	8/10	8/10	9/10	10/10

Figure 2.9 Normalization of boat activities of two excursions (Source: Chion *et al.* 2011)

2. Replace the sampled activities at each time point by the activity type that has the highest priority in the moving window. The boat activity priorities are assigned as following:

- Priority 1: activity 1 (observing whales),
- Priority 2: activity 3 (observing seals),
- Priority 3: activity 4 (observing geographical features),
- Priority 4: activity 2 (sailing), and
- Priority 5: activity 5 (back to port).

For example, when a technician collects the boat activities at every 10th minute, instead of simply recording the activity at exactly every 10th minute, he/she monitors the boat activity starting from 8th minute to the 12th minute and assigns the record with an activity type observed from 8th minute to 12th minute, which has the highest priority ranking. In this case, a moving window of five is used as illustrated on Figure 2.10. In the following moving window, the boat activity at 10th minute is assigned the value 3 (Observing seals) instead of the value 4 (Observing geographical features), because the value 3 has the highest priority in the moving window of 5 time points.

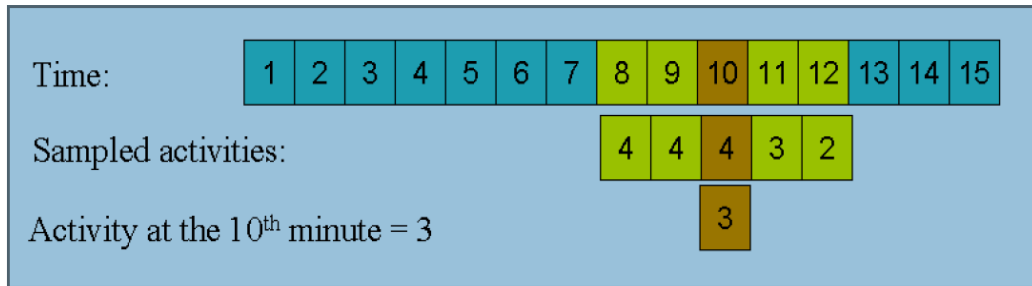


Figure 2.10 Moving window of five time points

3. Calculate the total time durations of different boat activities for ten bins

Multiple tables are created depending on the combination of boat types and home ports; these tables are named using the convention `repartitionboat?port?` (Section 2.4.5.6). For example, the table `repartitionboat2port01` stores the distribution of boat activities of boat type 2 (medium zodiac) departing from home port1 (Tadoussac). The data structure of `repartitionboat?port?` can be found in Section 2.4.5.6.

2.5.2.13 Boat observation on specific whale type

This function is to calculate and display the average percentage of time that a particular boat type spends at observing a particular whale type, excursion length, farthest location from the starting home port, etc. The result is stored in the table `observeonwhale`; the data structure of this table can be found in Section 2.4.5.5.

2.5.2.14 Analyze speed, duration, and distance of excursion

This function is to calculate and analyze the boat average speed, excursion duration and farthest excursion distance a boat can sail. The statistical result is obtained from the table `statofboatattribute` in the boat database. The data structure of this table can be found in Section 2.4.4.3.

2.6 User interface design

The user interface of this software serves as a tool to coordinate multiple visualizations, analysis and presentations of the dataset. It also provides a friendly and easy-to-understand environment for users to interactively explore and analyze the huge information generated by the 3MTSim model, as the example given in section 1.3, one simulation of 200 boat excursions over a period of 100 days at a temporal resolution of 1 minute generates 24,800,000 records. According to the above functions of the application module, the user interface is designed in three layers. Figure 2.11 shows the first and second layer.

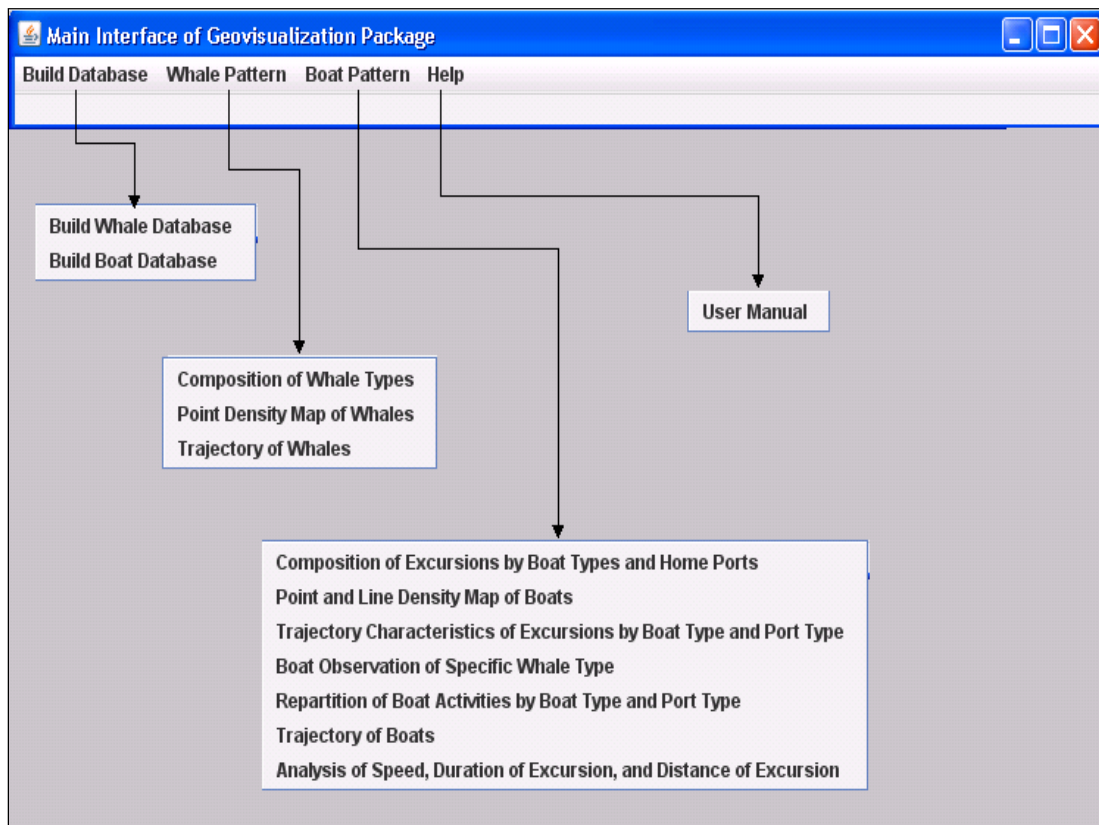


Figure 2.11 Main interface of the Geovisualization software

The first layer is the main user interface, which has four main menus: build database, whale pattern, boat patterns and user manual. This layer guides the users through the different domains of functions. The second layer lists all the functions under the main domains. The third layer provides a visual analysis platform for each function. The platform has all the variables (criteria) listed as groups; a user is allowed to explore the data by choosing variables to form a query. The density map of boats (Figure 3.2) in Section 3.5 is a good example. Through choosing different criteria on this interface, the user is able to identify the high density traffic area.

By reviewing the functions in the application model, the criteria used to form queries can be classified into the following groups:

- Boat types consist of all names of boat types retrieved from the raw dataset; the number of boat types depends on the parameter setup of the simulation.
- Whale types consist of all the names of whale types retrieved from the raw dataset; the number of whale types depends on the parameter setup of the simulation.
- Homeport consists of all names of homeports retrieved from the dataset; the number of homeport depends on the parameter setup of the simulation.
- Density types consist of the type of density maps the package can calculate; there are three types of density maps: the point density and line density of all actions and the point density of watching.
- Activity types consist of all the type of boat activities, which are going back to port, not in activity, observing whales, sailing, observing seals and observing landscape features.
- Time interval (start and end time) for density calculation.

2.7 Graph rendering methods

Graphs are the major integral part of the application package. The graphs implemented must be meaningful and represent the reality of the dataset. The following various types of traditional and novel graphs are combined to display the boat excursion and whale behaviour pattern.

- A pie graph is used to illustrate the proportion of whale types, boat types and home ports set up in the simulation. For example, in a pie graph for boat types, each section represents the proportion of excursions of a particular boat type.
- With the limitation of time for the research project and the heavy workload of 3D data preparation, a 2D cartographic system is implemented for the visualization. Instead of using a vertical axis to represent the time point of the geographical locations, a dynamic clock is displayed to show the time that has passed.
- A bar graph is used for the categorized type of boat activities. Boat activities are categorized into six types: going back to port, not in activity, watching whales, observing landscape features, sailing and observing seals.
- A diagram is employed to define and demonstrate the terminologies that are used to evaluate a captain's sailing habit. The terminologies are step length, turning angles, bearing, and net displacement.
- A geographic map with scatter-plotted whale/boat density locations is used for the point/line density map of boats and whales and the spatio-temporal trajectory of boats and whales.
- A dynamic color table is created to ensure its extensibility to account for any number of whale types, boat types and homeport for the visualization of correspondent proportions and the dynamic simulation of boat trajectory based on starting time.

2.8 Connections between tables in the database module and functionalities of the application module

In both the whale database and boat databases, each table, except the raw output datasets csvofwhale and csvofboat, stores related attributes associated with each particular behavior pattern of boats and whales. These tables are created, displayed and visualized in the application module to reveal the behavior patterns of boats and whales. The connection between these tables and the functionalities of the application module are shown on Figure 2.12.

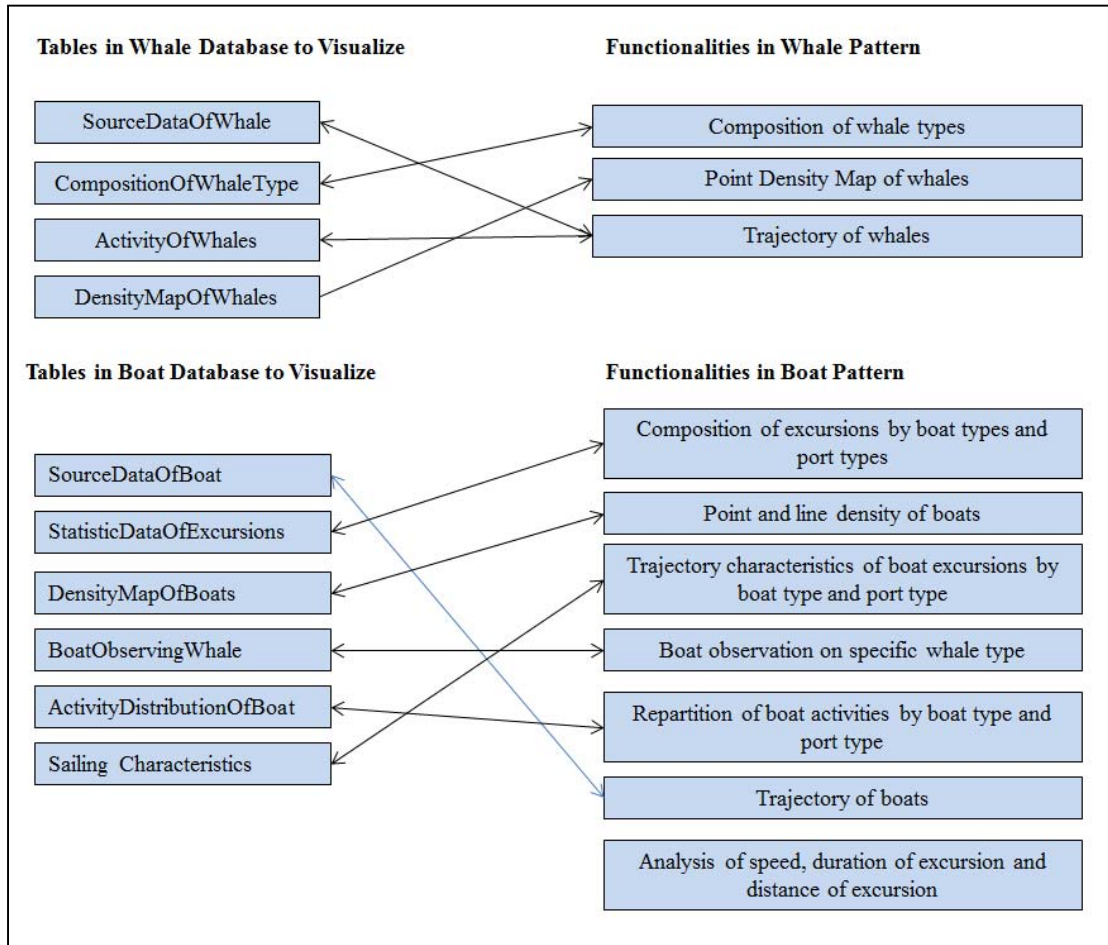


Figure 2.12 Connection between tables and functionalities

Chapter Three: Result Analysis

This chapter provides examples of queries and subsequent graph visualizations conducted with the Geovisualization software using simulated data obtained from 3MTSim for a given scenario. The figures shown here are intended to demonstrate the functionality of the software and should not be considered as definitive results for policy and management purposes.

3.1 Composition of excursions by boat type and home port, composition of whale types

The purpose of this visualization is to display the simulation parameter setup, and get an overall idea about how many boat types, homeports, and whale types are simulated in 3MTSim, and how many excursions of different boat types are from different homeports. The user interface on Figure 3.1 is an example of such visualization, which offers two ways to display the data:

- by choosing the boat types and clicking the button Composition of Port, the pie on the left displays the proportion of excursions by the chosen boat types. The pie on the right illustrates the proportion of homeports the excursions of the chosen type(s) of boat depart from. In addition, the corresponding statistical number and percentage are displayed besides the pie graphs. The background query of the user interface for Composition of Ports is:

```
select count (distinct BoatTypeName) from statofboatexcursion where TypeOfBoat = BoatID;
```
- by choosing the homeports and clicking the Composition of Boats button, the pie on the right displays the proportion of homeports used, the pie on the left displays the proportion of boat types departing from the chosen type(s) of homeport. In addition, the corresponding statistical number and percentage are displayed besides the pie graphs. The background of the user interface of Composition of Ports is:

```
select count (distinct BoatTypeName) from statofboatexcursion where TypeOfPort =  
PortID.
```

Figure 3.1 indicates that for the given scenario simulated with 3MTSim, the user specifies four boat types and six home ports. The results of the query reveal that 464 excursions were made by the type of boat Excursion_Small, accounting for 48% of the total number of excursions. They also indicate that Tadoussac is the preferred home port with 328 excursions departing from that location, which accounts for 34% of the total number of excursions, while only 5% of the excursions depart from Sainte-Rose-du-Nord, which is a smaller port.

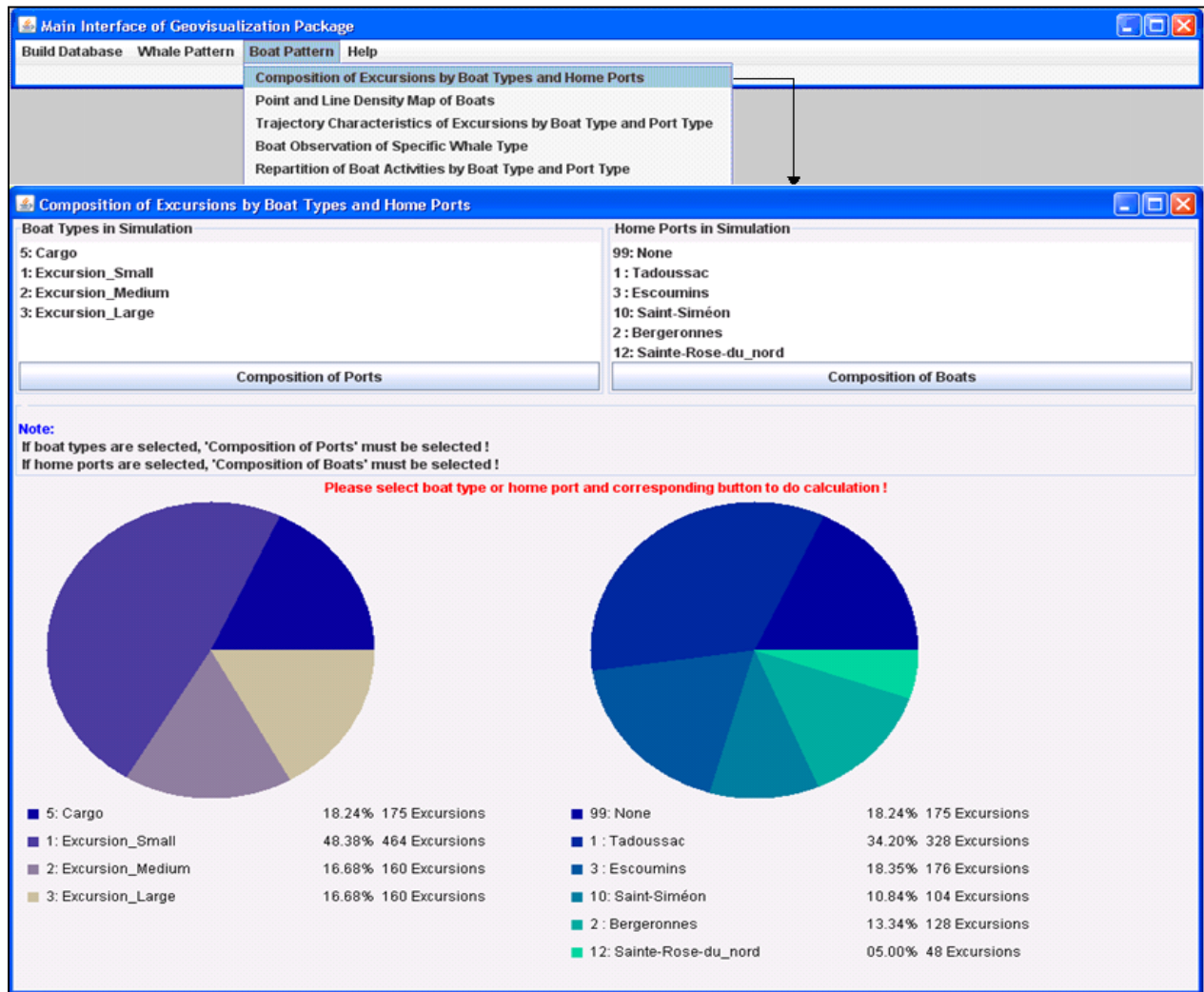


Figure 3.1 Composition of boat excursions

3.2 Point and line density map of boats and whales

The point density map is to visualize the occurrence locations by boat type and whale type. For boat patterns, more specific maps are offered, for example, by activity type (whale observing), by season, by time interval (month, day, hour, minute). The density map of observing whale locates the areas where boats observe whales and it may indicate that this area is also an area of high density of whales. The line density indicates high traffic locations of excursions by boat types and time intervals. This function is only offered for boat pattern extraction.

By combining the results of point density and line density of different boat activities in different season, the Park and MPA team is able to identify the possible congestion and high risk areas or time periods where a particular type of boat may collide with a particular type of whale. Specialist with solid domain knowledge can also detect if the simulated result of 3MTSim matches the real situations retrieved from field data.

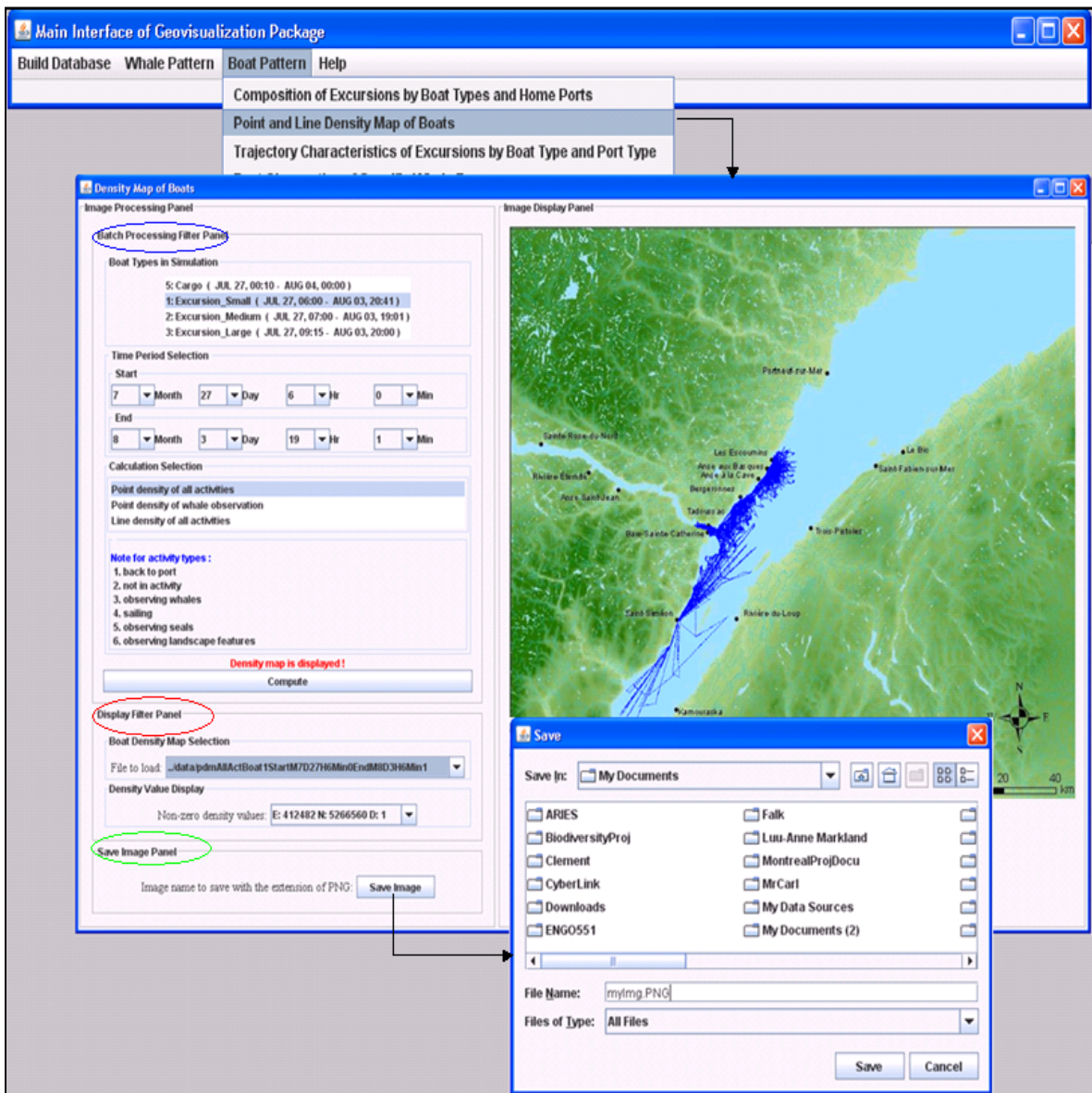


Figure 3.2 Density map of boats

Figure 3.2 provides an example of a density map of boats. The criteria chosen to compute the density occurrence are provided in Figure 3.3, which are boat type(s), start and end of time interval, types of maps (point density of all activities, point density of whale observation, line density of all activities). A density map filter lists all the current density maps existing in the database. After the computation is completed, the density map is stored in the database and the density map filter is updated. By clicking any name in the filter, a density map is draped down on the right geographic map. The geographic map of the study area that is used for overlaying the density locations and trajectories of boat excursions is provided in Fig 3.4. In addition, a drop-down menu displays all the density values (number of occurrence) by latitude and longitude. As we stated in Sections 2.5.2.5, 2.5.2.6, and 2.5.2.7 in Chapter two, the computation and visualization are separated into two steps to reduce the user response time. Some of the background queries for the calculation of the boat density maps are the following:

1. Extract the excursion IDs/whale IDs by boat type. The following queries are to extract types of boat from the statistical table statofboatexcursions and excursion IDs for boat type = 1.

```
select distinct (TypeOfBoat) from statofboatexcursion;
```

```
select distinct (ExcID) from statofboatexcursion where TypeOfBoat = 1
```

2. Extract the excursion lengths of the above extracted ExcIDs by time period, boat activity types etc. The following query is to extract the excursion length (total number of minutes for each excursion IDs. ExcIDArr[countOfExc] is an array of ExcID (excursion id), tickStart – tickEnd = the time period users type on the user interface.

```
select count(*) from csvboat1 where ExcID = ExcIDArr[countOfExc] and Tick <= tickEnd + and Tick >= tickStart;
```

3. Extract the (X_UTM, Y_UTM) location of each excursions by excursion ID or whale ID. The following query example is for boat type = 1, excursion id = 1 and boat activity type

= 1 (observing whale), tickStart - tickEnd = the time period users type on the user interface.

```
select X_UTM, Y_UTM from csvboat1 where id = 1 and BoatActivity = 1 and Tick <= tickEnd and Tick tickStart order by Tick;
```

4. Convert (X_UTM, Y_UTM) to cell ID, (370582, 5237060) is the north-west UTM location of the study area, 100 = the resolution of each cell in the study area.

```
cellNextX =(X_UTM - 370582)/100-1;
```

```
cellNextY =(Y_UTM - 5237060)/100-1;
```

Figure 3.2 illustrates how the point density locations of all types of activities for “Excursion_Small” during July 20, 6:00 am – August 3, 19:01 are retrieved and displayed. The drop-down menus of start and end time (month, day, hour, minute) allow the Park and MPA users to visualize the density location by specifying any time period (different start and end time) or time points (same start and end time). The function is finished in two steps: calculation and visualization. When the calculation is completed, the file name is automatically listed in the area ‘File to load’. A message pops up telling the user that the calculation is completed. By clicking the file data, a density map is displayed. If the map was already created, a user can display the map without re-doing the calculation. An option to facilitate the saving of the map as a PNG file is also provided on the user interface to allow a user to save the map in any folder. By left and right clicking, the density map can be zoomed in or out.

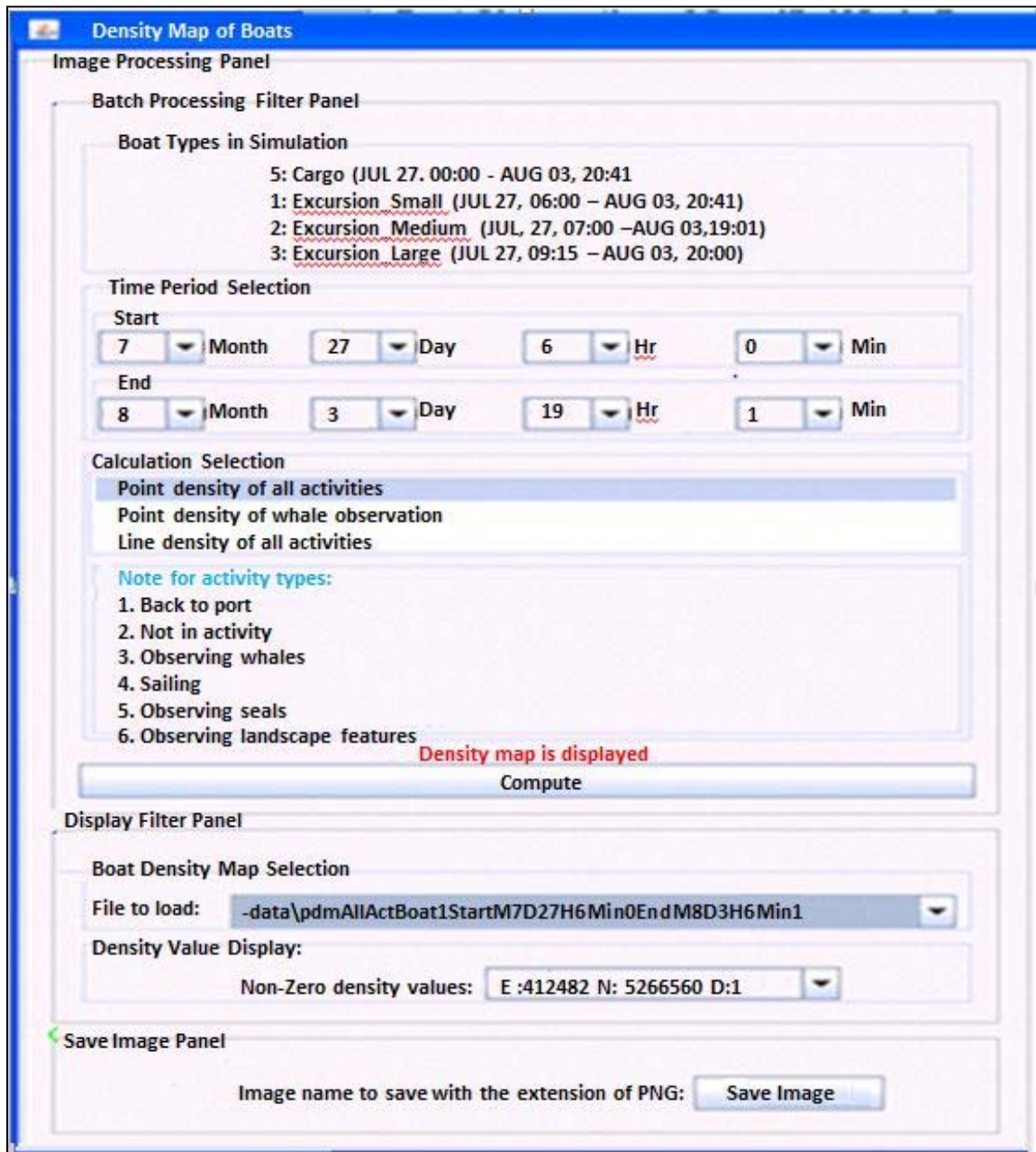


Figure 3.3 User menu of density map of boats

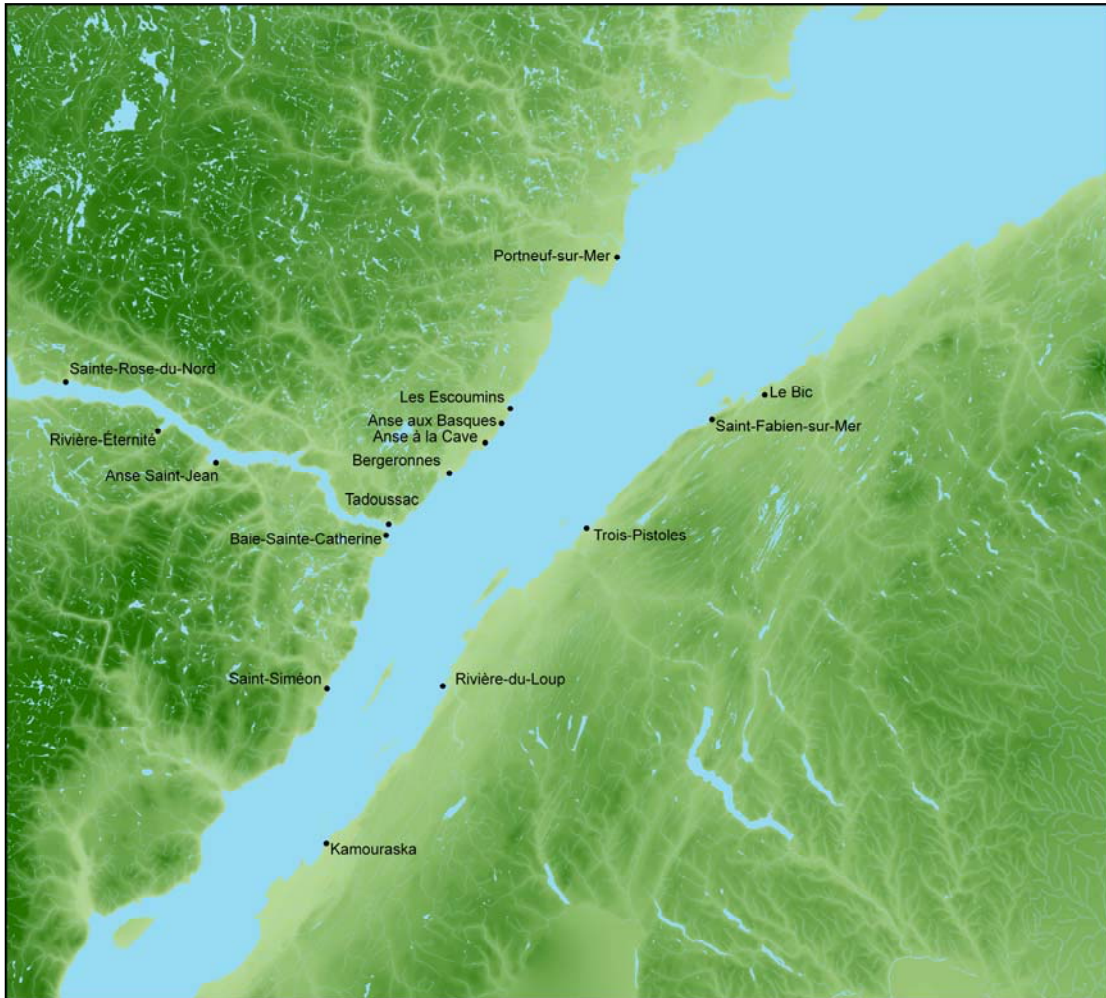


Figure 3.4 The geographic map of the study area which is used for displaying the density locations and trajectories of boat excursions (source: Parks Canada)

3.3 Trajectory characteristics of boat excursions

The purpose of this function is to calculate the trajectory characteristics of boat excursion by step length, turn angle, bearing and net displacement and their corresponding mean, minimum, maximum, medium and standard deviation. The calculated result is displayed in a table as shown on Figure 3.5. A diagram is displayed to define the concept of Step Length, Turn Angle, Bearing, and Net Displacement. To calculate these parameters, each (X_UTM, Y_UTM) and their

previous one/two step locations need to be checked. This function requires extensive calculation. To solve this problem, the calculation is sliced by boat type and home port type; the computation and result display are also separated in two independent steps in order to reduce the user response time as stated in Section 2.5.2.11. Once the computation is completed, a message pops up to inform the user. The user is allowed to display the result or to quit this function; the result is stored in the database to avoid more computation applied to the same subset.

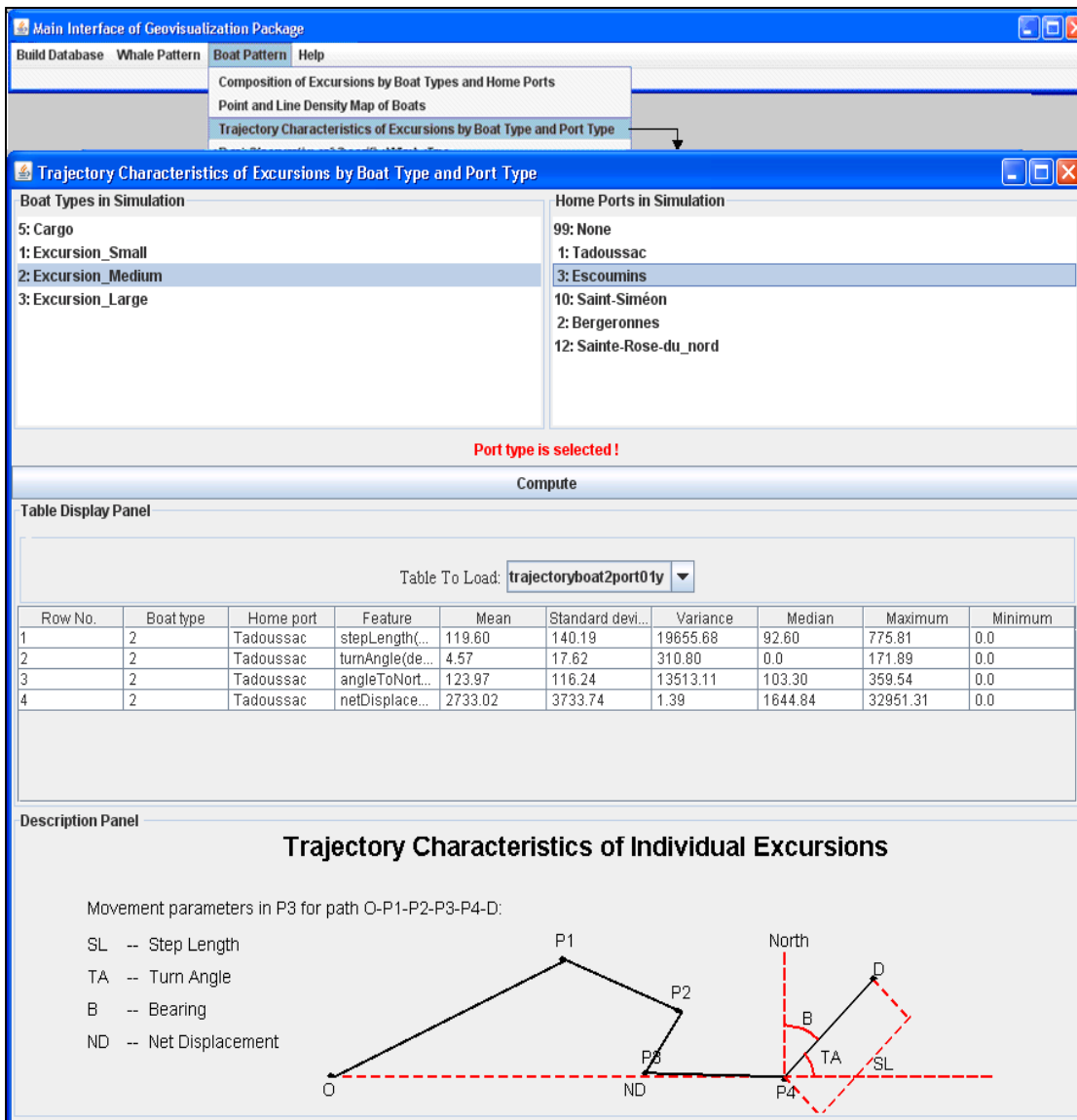


Figure 3.5 Trajectory characteristics of boat excursions

Figure 3.3 indicates that boat type Excursion_Medium departing from Tadoussac make turns at about every 119.60 m. The maximum step length is 775.81 m, while the minimum is 0. A large step length represents high sailing speed, which may result in collision with whales. The average turning angle is around 4.97 degrees, the maximum is 171.0 and the minimum is 0. A large turning angle usually represents rude boat sailing activities.

3.4 Repartition of boat activities by boat type and home port

The purpose of this function is to calculate and visualize the temporal distribution of boat activities (observing whales, sailing, back to port etc.). A traditional bar graph is implemented for the visualization of this function. This function aggregates the activity types (back to port, not in activity, observing whales, sailing, observing seals, observing landscape features) of each time point. As we stated in Section 2.5.2.11, a moving window of 5 is used to assign the activity type to each of time point. This application of a moving window triples the calculation time. To solve this issue, the calculation is sliced into small components to be assigned to different parts of the application. Some calculations are completed when the user opens the interface function; some others are executed after the user selects the boat type and port time. The calculation and the visualization are also separated to reduce the user response time. The main part is executed after the compute button is clicked. The result is stored in the database to avoid additional calculation. As shown in Figure 3.4, a boat type and homeport panels are provided. A data file filter is provided for the user to visualize the existing dataset in the database without calculation. The X axis is divided into ten bins that indicate the different time intervals of the whole excursion. The Y axis shows the percentage of activity types of each bin.

Figure 3.6 reveals that 80% of the boat activities in the first 1st/10 segmentation are watching whales while the remaining 20% of boat activities are sailing. In the middle of the excursions, most of the activities are observing whale. In the last 10th/10 segmentation, around 50% of boat activities are sailing back to port. A visual inspection confirms that the distribution of boat activities from the 3MTSim matches the real behavior pattern collected from the field.

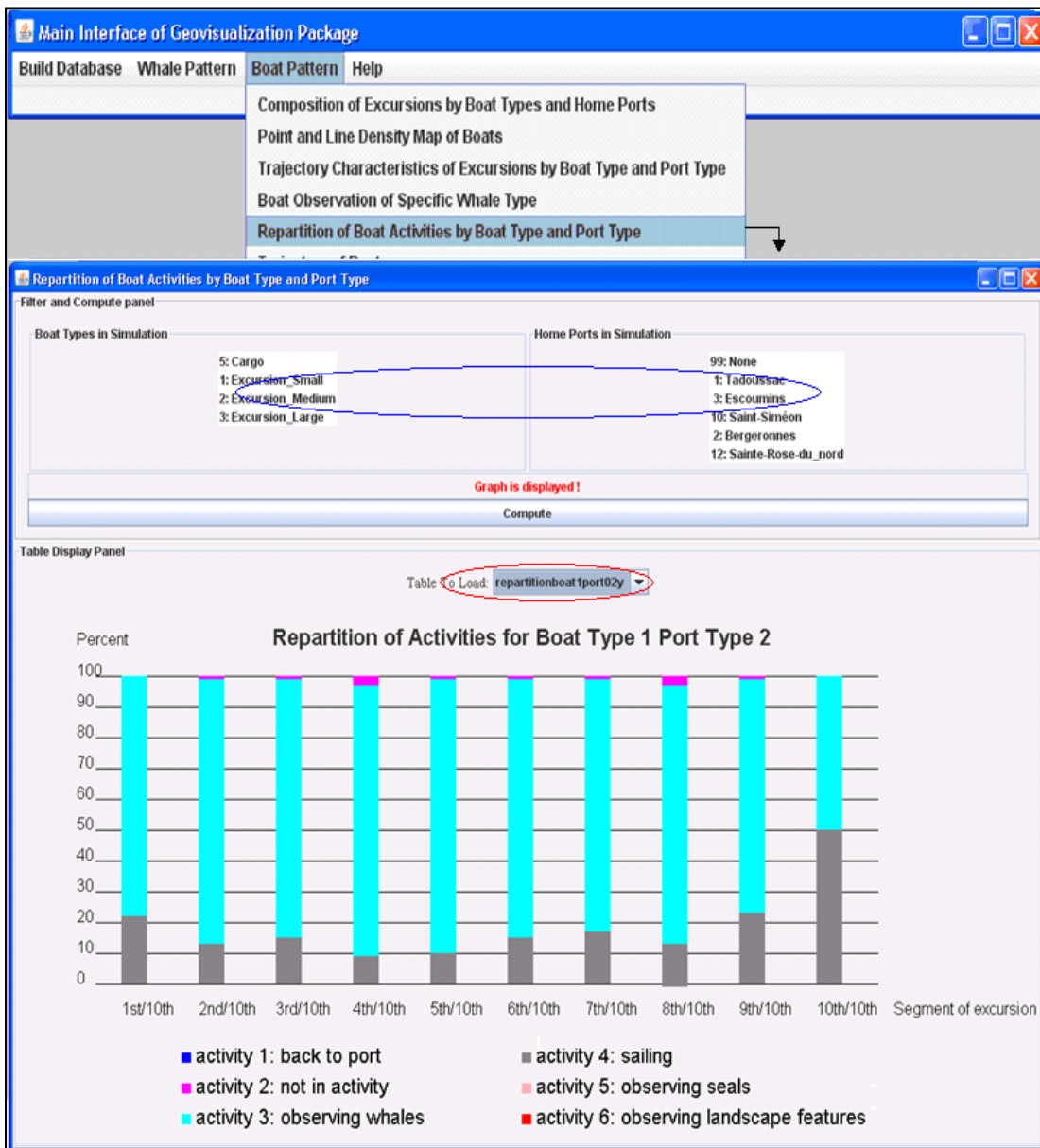


Figure 3.6 Repartition of the boat activities by boat type and port type

3.5 Dynamic visualization of trajectories of whales

The purpose of this function is to dynamically display the trajectories of any single whale by whale type and whale ID at a specified rendering speed in millisecond (Fig. 3.7). The simulation can be long because whales stay in the park and MPA during the whole summer. The simulation can be stopped at any time by clicking the stop button. A user can also choose the number of time steps to simulate. This will help to demonstrate the diving habit of the whale. A boundary is displayed in red before the simulation starts since the users may be interested only in the activity areas, instead of the trajectories. Figure 3.5 provides an example of a trajectory of a Beluga (Whale type code= 105, Agent id = 134). The whale moves within the area delineated by these coordinates: min X_UTM = 425613 m, minY_UTM = 5267067 m. Max X_UTM = 449025 m and max Y_UTM = 5309856 m. The time elapsed during the current simulation is 38 hours and 30 minutes.

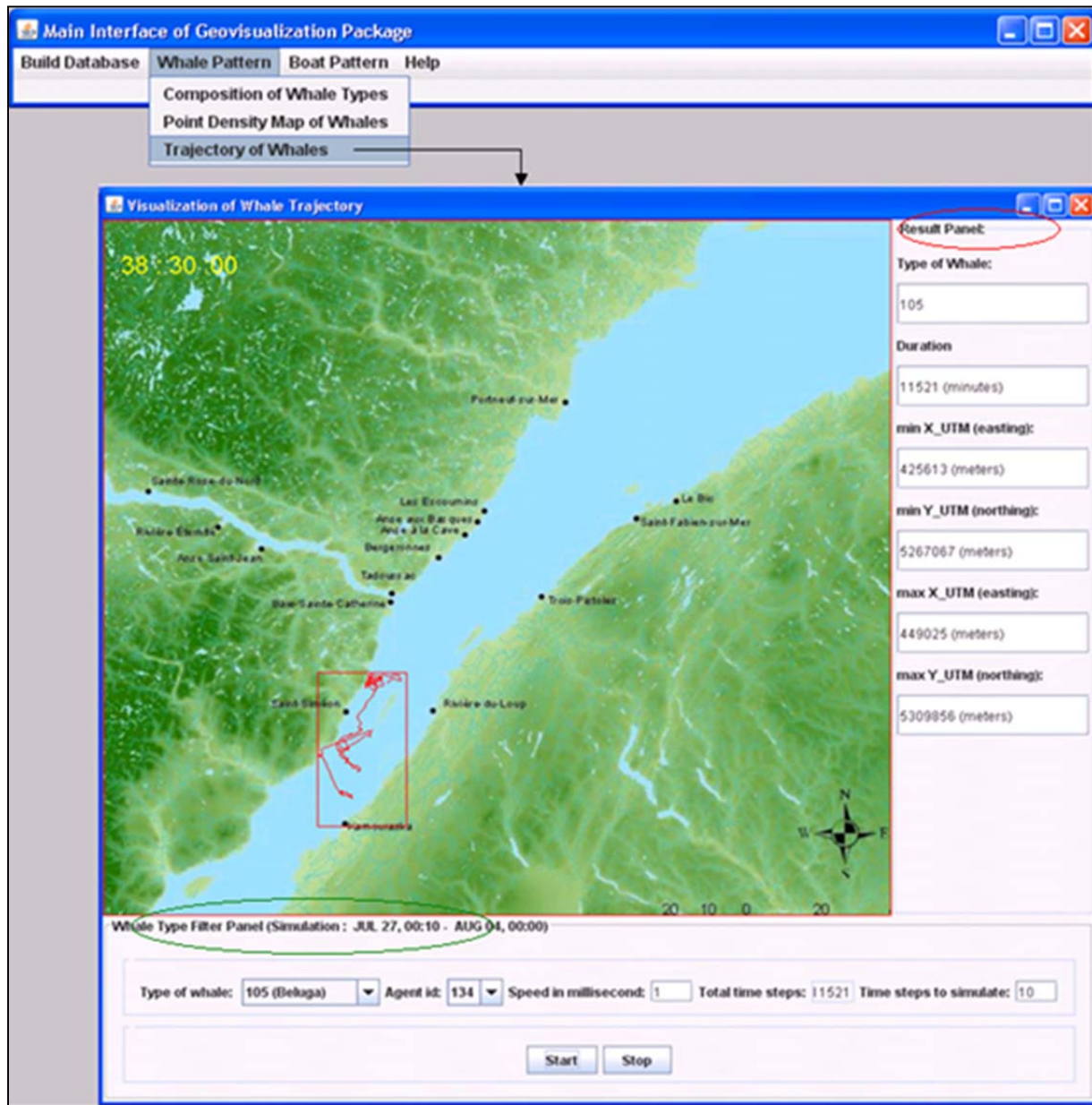


Figure 3.7 A trajectory of a beluga (agent id = 134) in the whole simulation season

3.6 Dynamic visualization of trajectories of boats

The purpose of this function is to dynamically visualize the trajectories of boats by boat type, excursion start time point or excursion ID. Instead of using vertical axes to represent the time period, a clock is used to display the time elapsed. To reduce the number of trajectories to be

displayed simultaneously resulting in overlap, the whole dataset is partitioned into meaningful datasets by boat type, excursion start time, or by excursion ID. Figure 3.6 is an example of boat trajectory visualization. The same geographic map (Fig. 3.4) of the study area used in section 3.2, is employed for illustrating the trajectories. To visualize the trajectories by starting time point, the interface lists all the boat types in the simulation and corresponding starting times of all excursions. After a boat type is chosen, a drop-down menu lists the departing time of all excursions. After a user chooses the departing time, the trajectories of all excursions are dynamically displayed on the map at the right. A user can also choose the computer rendering speed in millisecond. After the simulation is completed, all excursion IDs are displayed in the filter of ExcID panel by a drop down menu. By choosing any one of the IDs, the corresponding trajectory characteristics are displayed for user's reference, such as maximum and minimum northing, easting, average speed start time, and end time of the interval. To demonstrate the time elapsed at each geographical location, a dynamic clock is provided on the left upper corner of the geographic map. Some of the queries used for such display are:

1. select distinct (TypeOfBoat) from statofboatexcursion to extract the number of boat types in the simulation;
2. select distinct BoatID from statofboatexcursion where TypeOfBoat = 1 to extract the number of excursions for each boat type
3. select X_UTM, Y_UTM from csvboat1 where ExcID = arrOfExcID[i] order by tick to extract the (X_UTM, Y_UTM) of each excursion, where $i = 1..n$, n is the number excursions when excursion starts.

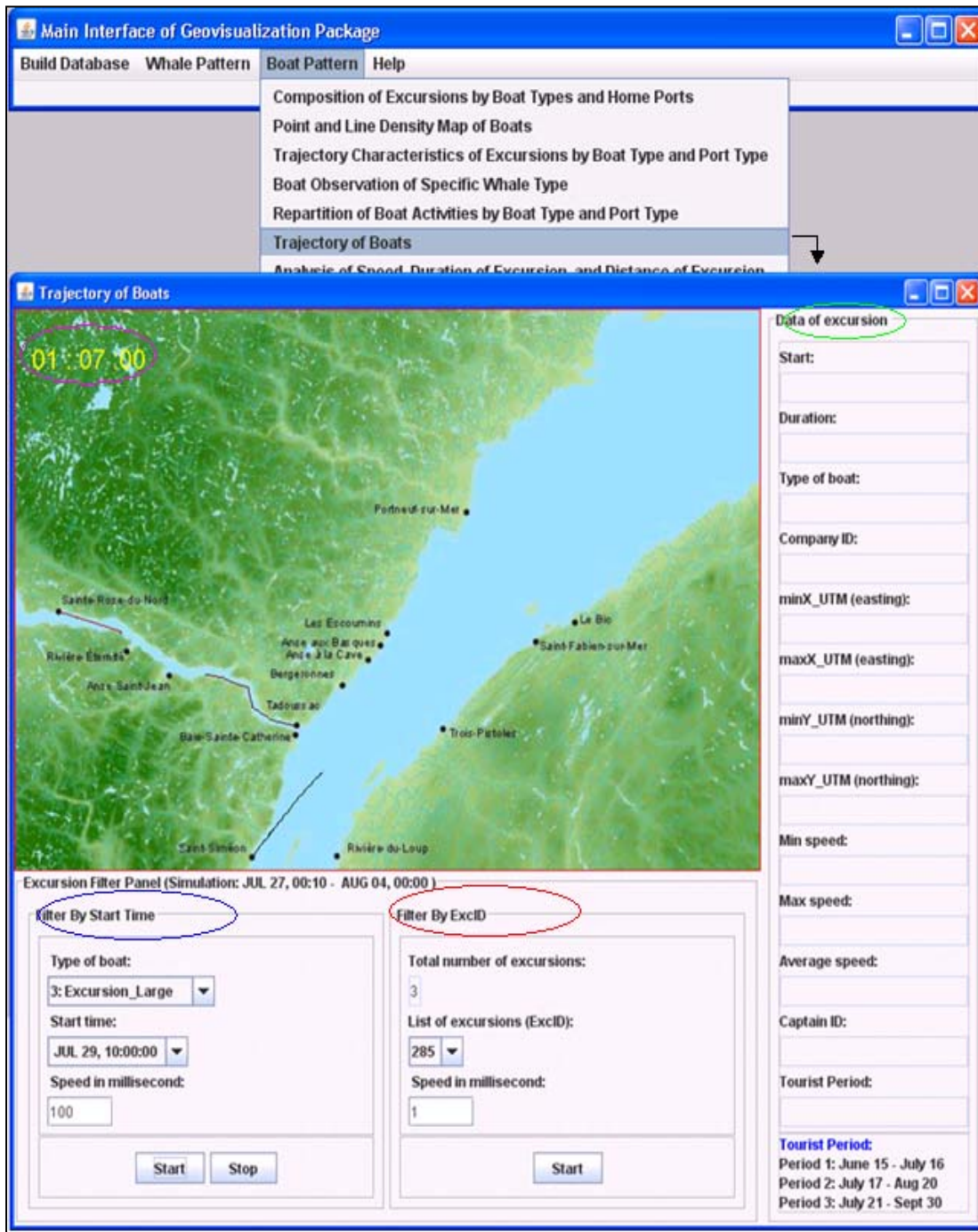


Figure 3.8 Trajectories of boats by time point

Figure 3.8 reveals that three Excursion_larges departed on July 29, 10:00:00 from three different ports: Sainte-Rose-du-Nord, Saint Siméon, and Tadoussac. The speed of the simulation is 100 milliseconds. The elapsed time is one hour and seven minutes. The typical excursion characteristics will be displayed in the statistical panel at the right if an ExcID is clicked.

3.7 Boat observation on specific whale type

This function calculates and displays the percentage of time a particular boat type spends on observing a particular whale type compared to the total time that the boat stays at sea (Table TimeOfObservation). The calculation is performed based on the visibility between the whales and the boat user types. As shown in the following example (Fig. 3.9), there is a total of three boat types and one whale type in the simulation. Large boats with more than 50 seats (boat type =3) spend 28% percent of their total excursion time on observing fin whales within 4000 m, while the medium zodiacs (with 13 to 49 seats) spend 43% of their excursions on observing fin whales within 4000 m.

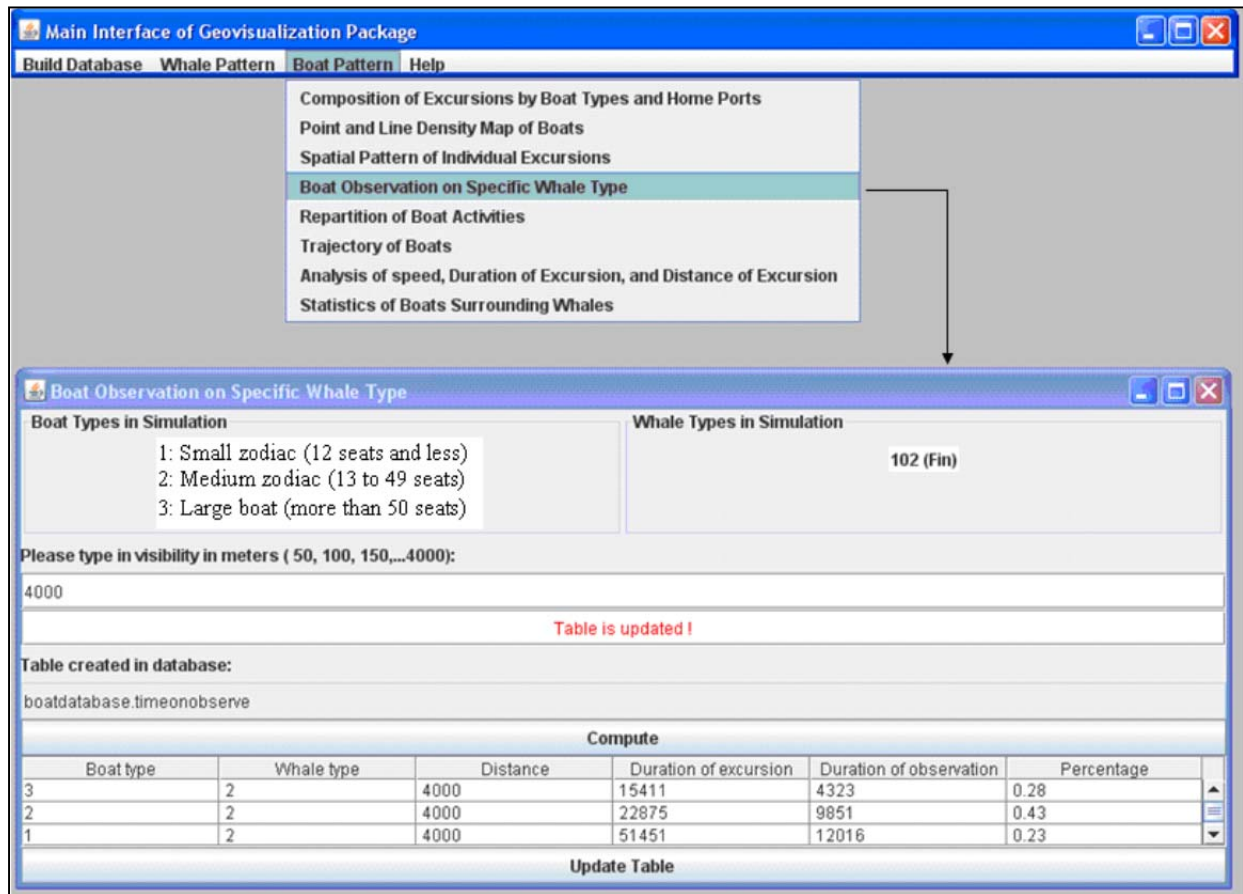


Figure 3.9 Boat observing a particular type of whale

3.8 Analysis of speed, duration of excursion and distance of excursion

This function shows the statistical results of average speed, duration and farthest distance of by individual excursions (Table StatisticDataOfExcursions). Figure 3.10 indicates that there are three types of boats and four homeports in the simulation. The average speed of a small zodiac (excursion ID =1) departing from Tadoussac is 11.15 knots, the minimum speed is 9 knots and the maximum speed is 12 knots. The duration of another small zodiac (excursion ID = 5) departing from Les Escoumins is 35160.13 minutes in the simulation.

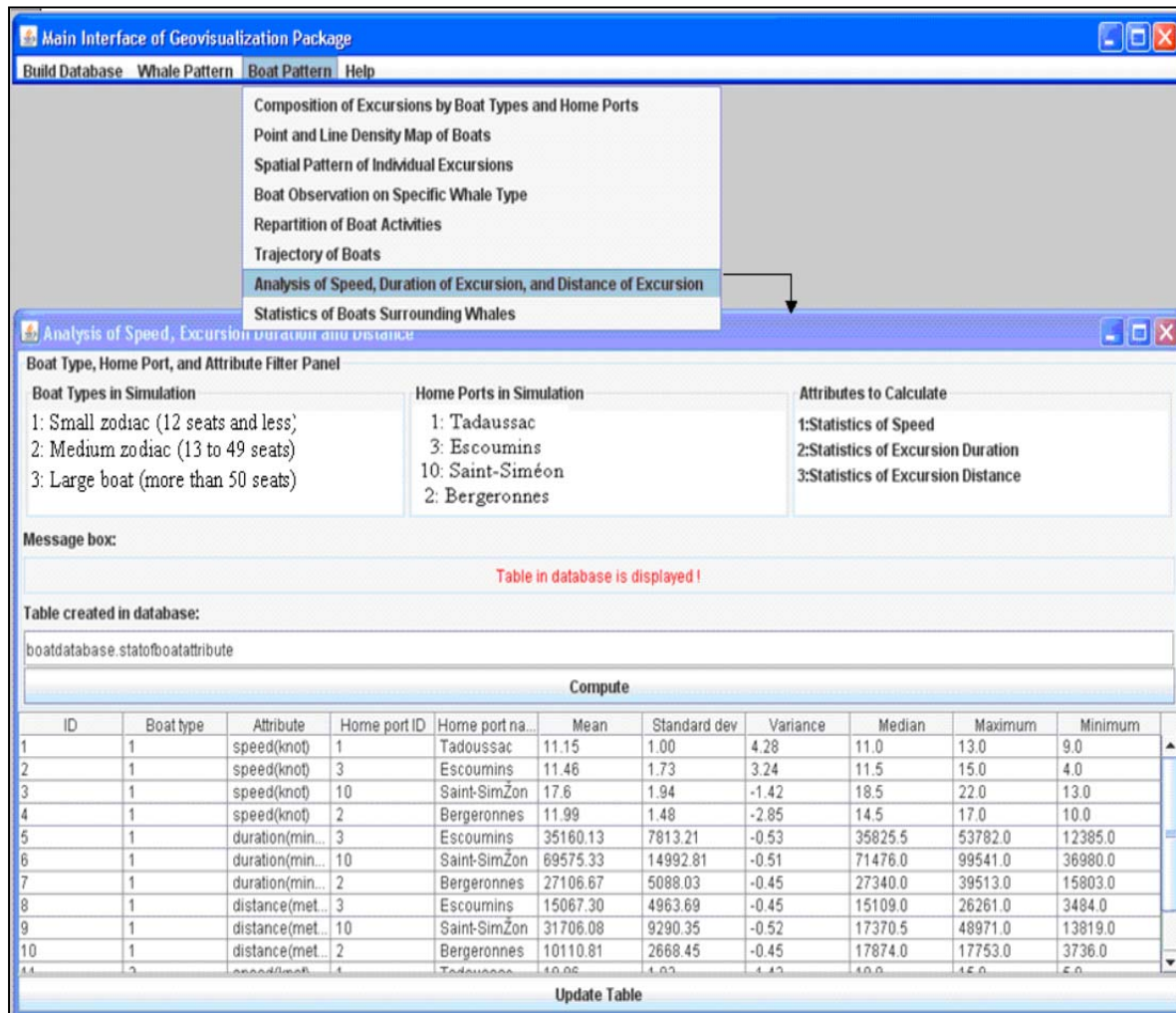


Figure 3.10 Analysis of average speed, duration, and distance of excursions

3.9 Software performance

The software performance was evaluated based on a selection of functions provided on the interface (Table 3.1). The listed performance was obtained from a 100 days simulation data of 4 boat types, 4 port types and 5 whale types, involving 28,000,000 records for boat excursions and 10,000,000 records for whale activities generated by the 3MTSim model. The performance expressed in minutes/seconds is an average of different combinations within the same category.

For example, 4 boat types (Cargo, Excursion_Small, Excursion_Medium, Excursion_Large) and 4 departing port types (Tadoussac, Bergeronnes, Escoumins, SainteRose-Du-Nord) represent 16 combinations. The performance (50 seconds) of function ‘repartition of boat activities by boat type and port type’ is the average of the 16 combinations.

Table 3.1 Statistic data on software performance

Row No	Functions	Performance (minutes/seconds)
1	Build boat database (28,800,000 records)	4 minutes
2	Build whale database (10,000,00 records)	1 minute
3	Composition of whale types	0.01 second
4	Point density map of whales	30 seconds
6	Trajectory of whales	1 second
7	Composition of excursions by boat types and home ports	2 seconds
8	Point and line density map of boats	45 seconds
9	Trajectory characteristics of boat excursions by boat type and port type	1 second
10	Boat observation on specific whale type	0.01 second

11	Repartition of boat activities by boat type and port type	50 seconds
12	Trajectory of boats	0.01 second
13	Analysis of speed, duration of excursion, and distance of excursion	0.01 second
14	User manual	0.01 second

Chapter Four: Conclusion

The Geovisualization software developed in this study allows a user to visualize, interpret and explore the behavior patterns of whales and boats from large datasets (as described in Section 1.3) generated by the 3MTSim model. The result serves as a decision tool both for park managers and MPA teams to develop, evaluate, and select proper policies and for scientist to validate the feasibility of the model by comparing the result with patterns derived from side collected data. However, further research needs to be done to provide an effective presentation of the underlying data and to enable users to explore datasets to identify patterns, associations, trends and so on.

By integrating databases, an interactive user interface, and multiple visualization methods, this Geovisualization software provides an analytical platform allowing the Marine Park and MPA team to explore the data generated by the 3MTSim according to their domain knowledge and expertise. With the interactive user interface, all datasets are split and organized into meaningful small subsets, the barrier of large amount of calculation is solved, and the performance of the software is improved. The various visualization methods, such as pie graphs, bar graphs, diagrams, tables and geographic maps with zooming in/out functionality, and the use of a dynamic clock avoids the interpretative complexity of traditional statistical multivariate pattern generalization or recognition methods. The visualization also facilitates the visual knowledge construction process of users. In addition, integrating a visual-based database reduces the time of selecting, aggregating and sampling, decreasing the response time to users.

As a decision-support tool, the Geovisualization software provides facilities for retrieving and visualizing a series of important behavior patterns of whales and boats from huge interrelated multi-dimensional temporal-spatial datasets, which serves both for the Park and MPA teams to develop, evaluate and select proper policies and for scientists to validate the feasibility of the simulation model by comparing the patterns derived from the model outcomes with the ones derived from site collected data.

The Geovisualization software uncovers the temporal characteristics of the datasets in terms of visually representing it, which has been a focus in recent Geovisualization research. For example, the normalized distribution changes of boat activity types (back to port, not in activity, observing whales, sailing, observing seals, observing landscape features) reveals the temporal patterns of boat sailings.

One key contribution of this study to the field of Geovisualization is that the software is designed to handle very large volumes of data. Most of Geovisualization efforts focus on developing visual tools to represent data; however, this study emphasizes the database design in order to speed up the query response time to users. Another contribution is that an interactive user interface was designed in terms of universal usability. Most Geovisualization tools or software are designed and developed for researchers, data analysts, and experts. This software was designed for park managers and it can also be used by the public. Moving toward achieving universal usability is part of the goal of Geovisualization efforts.

4.1 Further research

Additional functionalities could be implemented to provide a more effective presentation of the spatio-temporal datasets generated by 3MTSim to facilitate the exploration and analysis of data by users.

For the visualization of point/line density map of boats and whales, simply plotting data on a background map is not the most effective way to present abstract information; unseen patterns (Andrienko *et al.* 2008) can be further revealed by using a convolution kernel function, which is a three-dimensional data mining technique that aims at grouping data to patterns based on mutual dis-similarity (Stefan *et al.* 2010).

A functionality to overlay different point/line density maps of related data subset would provide better visual effects for analysis. For example, overlaying a line density map over a point density map of a particular boat type and time period in a unique map would allow users to easily identify the high traffic and low traffic areas than comparing separate two maps.

The visualization functions in this study are focused on the direct depiction of the original datasets and results are calculated from the datasets by selection, aggregation and generalization. More sophisticated computational techniques from data mining and KDD could be implemented to visually assist the knowledge construction by the park and MPA managers. For example, one possible solution for the function trajectory characteristics of boat excursion (Section 3.3) is to use a clustering algorithm to divide the excursions into various classes based on the calculation results of a set of parameters (step length, turn angle, bearing and net displacement). The resulted

classes could represent different sailing habits, rough or gentle. A similar method can also be used for other functions, such as the analysis of speed, duration of excursion and distance of excursion. After the excursions are classified, the behavior of the different classes of excursions (captains) becomes predictable.

References

- Acoust, J. (2008). Passive acoustic detection and localization of whales: Effects of shipping noise in Saguenay-St.Lawrence Marine Park, *The international Journal of Animals and its Recording* 123(6):4109-4117.
- Andrienko, G., N. Andrienko, P. Jankowski , D. Keim D, M-J. Kraak, A. M. MacEachern, S. Wrobel (2008). Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research, *Information Visualization* 7:173-180.
- Andrienko N., G. Andrienko, P. Gatalsky. (2003). Exploratory spatio-temporal visualization: an analytical review, *Journal of Visual Languages and Computing* 14:503-541.
- Andrienko G., N. Andrienko, P. Jankowski, D. Keim, M-J. Kraak, A.MacEachren, S. Wrobel (2007), Geovisual analytics for spatial decision support: Setting the research agenda, *International Journal of Geographical Information Science* 21(8):839-857.
- Andrienko G., N. Andrienko, J. Dykes, M-J. Kraak, H. Schumann (2010). Geospatial visual analytics: focus on time special issue of the ICA commission on geovisualization, *International Journal of Geographical Information Science* 24(10): 1453-1457.
- Anwar M., A. Jeanneret, L. Parrott, and D.J. Marceau (2007). Conceptualization and implementation of a multi-agent model to simulate whale-watching tours in the St. Lawrence Estuary in Quebec, Canada, *Environmental Modelling & Software* 22:1775-1787.
- Blok C., B. Köbben, T. Cheng, Agnes A. Kuterema,A. (1999). Visualization of relationships between spatial patterns in time by cartographic animation, *Cartography and Geographic Information Science* 26 (2):139-151.

- Cartwright W., J. Crampton, G. Gartner, S. Miller, M. Suzette, K. Mitchell, E. Siekierska, J. Woo (2000). Geospatial information visualization, user interface issues, *Cartography and Geographic Information Science* 28(1):45-60.
- Chertov O., K. Alexander M. Alexey, A. Gennady, N. Andrienko, P. Gatal'sky (2005). Geovisualization of forest simulation modeling results: A case study of carbon sequestration and biodiversity, *Computer and Electronics in Agriculture* 49(1):175-191
- Chion C., P. Lamontagne, S. Turgeon, L. Parrott, J.-A. Landry, D. J. Marceau, C. C. A. Martin, R. Michaud, N. Ménard, G. Cantin, S. Dionne (2011). Eliciting cognitive processes underlying patterns of human-wildlife interactions for agent-based modelling, *Ecological Modelling* 222(14):2213-2206.
- Dibiase D. (1990). Visualization in the earth sciences, *Earth and Mineral Science, Bulletin of the College of Earth and Mineral Sciences, Penn State University* 59:13-18 (also: www.geovista.psu.edu/publications/others/dibiase90/swoopy.html).
- Dykes J., A. M. Maceachren, M. Kraak (2005). *Exploring Geovisualization*, Elsevier, The Netherlands.
- Fayyad, U. M., G. Piatetsky-Shapiro, P. Smyth (1996). From data mining to knowledge discovery: An overview In: U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy (eds), *Advances in knowledge discovery and data mining*. AAAI Press/MIT Press, pp. 1-34.
- Fayyad, U. M., G. G. Grinstein, A. Wierse (2002). *Information visualization in data mining and knowledge discovery*. San Francisco, California: Morgan Kaufmann.
- Frawley, W.J., G. Piatetsky-Shapiro, C. J. Matheus (1991). Knowledge discovery in databases: An overview. In: G. Piatetsky-Shapiro, W.J. Frawley (eds), *Knowledge discovery in databases*. AAAI Press/MIT Press, pp 1-27.

- Frihida, A., D.J. Marceau, and M. Theriault (2004). Extracting and visualizing individual space-time paths: An integration of GIS and KDD in transport demand modeling, *Cartography and Geographic Information Science* 31(1):19-28.
- Guo, D. (2007). Visual analytics of spatial interaction patterns of pandemic decision support, *International Journal of Geographical Information Science* 21(8), 859-877.
- Gatrell, A., (1994). Density estimation and the visualization of point patterns. In: Hearnshaw, H.M., Unwin, D.J. (Eds.), *Visualization in Geographical Information Systems*. Wiley, New York, pp 65-75.
- Gahegan, M., M. Wachowicz, M. Harrower, T-M Rhyne (2001). The integration of geographic visualization with knowledge discovery in databases and geocomputation, *Cartography and Geographic Information Science* 28(1): 29-44.
- Hägerstrand, T., Diorame (1982). Path and project, *Tijdschrift voor Economische en Sociale geografie* 73(6):323-339.
- Katharina M., B. Kanishka, K. Hillol (2012). Introduction to data mining for sustainability, *Data Mining and Knowledge Discovery* 24(2):311-324.
- Keim, D., G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, G. Melancon (2008). Visual analytics: definition, process and challenges. In: A. Kerren *et al.* (Eds), *Information visualization*. Springer-Verlag Berlin Heidelberg, pp 154-175.
- Kraak, M.-J. (2008). Geovisualization and time: new opportunities for the space – time cube. In: M. Dodge, M. McDerby and M. Turner. Chichester (Eds), *Geographic visualization: concepts, tools and applications*. Wiley & Sons, pp 293-306.

- Koua, E. L., A. Maceachren, M-J. Kraak (2006). Evaluating the usability of visualization methods in an exploratory geovisualization environment, *International Journal of Geographic Information Science* 20(4):425-448.
- Kwan, M. P. (2000). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set, *Transportation Research Part C* 8:185-203.
- Langran, G. (1993). Issues of implementing a spatiotemporal system, *International Journal of Geographical Information* 7(4): 305-314.
- MacEachren, A. M., M. Wachowicz, R. Edsall, D. Haug (1999). Constructing knowledge from multivariate spatiotemporal data: Integrating geographical visualization with knowledge discovery in database methods. *International Journal of Geographical Information Science* 13(4):311-34.
- MacEachren A. M. (2001). An evolving cognitive-semiotic approach to geographic visualization and knowledge construction, *Information Design Journal* 10:26-36.
- MacEachren A. M. J. H. Ganter (1990). A pattern identification approach to cartographic visualization, *Cartographic* 27:64-81.
- MacEachren, A. M., M. Wachowicz, R. Edsall, D. Hang, R. Masters (2010). Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in database methods, *International Journal of Geographical Information Science* 13(4): 311-334.
- MacEachren, A. M. and M-J. Kraak (2001). Research challenges in geovisualization, *Cartography and Geographic Information Science* 28(1):3-12.

- McCormick B. H., T. A. DeFait, M.D. Brown (1987). Visualization in scientific computing – A synopsis, *IEEE Computer Graphics & Application* 7(7):61-70.
- Nikos Pelekis, BabisTheodoulidis, Ioannis Kopanakis and Yannis Theodoridis (2004). Literature review of spatio-temporal database models, *The Knowledge Engineering Review* 19:235-274. doi:10.1017/S026988890400013X.
- Parrott, L., C. Chion, C. C. A. Martins, P. Lamontagne, S. Turgeon, J. A. Landry, B. Zhens, D. J. Marceau, R. Michaud, G. Cantin, N. Ménard, and S. Dionne (2011). A decision support system to assist the sustainable management of navigation activities in the St. Lawrence River Estuary, Canada, *Environmental Modelling and Software* 26(12): 1403-1418.
- Peng Z–R (2005). A proposed framework for feature-level geospatial data sharing: a case study for transportation network data, *International Journal of Geographical Information Science* 19(4):459-481.
- Peuquet, D.J. and N. Duan (2007). An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data, *International Journal of Geographical Systems* 9(1):7-24.
- Rinner, Claus (2007). A geographic visualization approach to multi-criteria evaluation of urban quality of life, *International Journal of Geographical Information Science* 21(8):907-919.
- Rinzivillo S., D. Pedreschi, M. Nanni, F. Giannotti, N. Andrienko, G. Andrienko (2008). Visually driven analysis of movement data by progressive clustering, *Information Visualization* 7(3): 225-239.
- Silverman, B.W. (1987). Density estimation for statistics and data analysis. *Technometrics* 29(4):495.

- Schilling A., V. Coors, K. Laakso. 15 dynamic 3D maps for mobile tourism applications, Fraunhofer Institute for Computer Graphics, Darmstadt, University of Applied Sciences Stuttgart, Nokia Research Center, Helsinki.
- Scarpaci C., E.C.M. Parsons, and M. Lück (2008). Recent advances in whale-watching research: 2006-2007, *Tourism in Marine Environments* 5:55-56.
- Stefan, P. M. K. Jukka (2010, May). Density calculation for moving points, *13th AGILE International Conference on Geographic Information Science*, Guimarães, Portugal.
- Wiegand T., F. Jeltsch, I. Hanski, and V. Grimm (2003). Using pattern-oriented modelling for revealing hidden information: a key for reconciling ecological theory and application, *Oikos* 100:209-222.
- Wang, X. (2005). Integrating GIS, simulation models, and visualization in traffic impact analysis, *Computers, Environment and Urban Systems* 29: 471-496.